A COMPARISON OF ERROR PROBABILITIES FOR TWO STANDARD ANALYSES OF VARIANCE IN UNBALANCED TWO-WAY DESIGNS

by

Robert E. Johnson

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1447

August 1983

A COMPARISON OF ERROR PROBABILITIES FOR TWO STANDARD ANALYSES OF VARIANCE IN UNBALANCED TWO-WAY DESIGNS

by

Robert E. Johnson

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

Chapel Hill

1983

Approved by:

David G. Kleinbaum
Advisor

Dana Quade
Reader

M. J. Symons
Reader

ROBERT EARL JOHNSON. A Comparison of Error Probabilities for Two Standard Analyses of Variance in Unbalanced Two-Way Designs (Under the direction of DAVID G. HERR.)

The test of no difference in main effects in two-way factorial designs usually involves a two step procedure: first test for the presence of interaction and, if no significant interaction is detected, proceed with a main effects test. It would be ideal if the interaction test were independent of any main effects test and if the presence of interaction had no effect on the tests for main effects.

In this work joint errors for the two step procedure are studied for each of three tests: (1) an idealized (but not realizable) independent test, (2) STP analysis which tests equality of simple column (row) averages of cell means, and (3) EAD analysis which tests columns (rows) adjusted for tows (columns).

These analyses are first studied in light of their geometric properties and then, with the aid of existing distribution theory and new theory developed here, the error probabilities are compared. EAD is found to be the more powerful test for detecting column (row) differences in an additive model, but a biased test in the presence of interaction.

ACKNOWLEDGEMENTS

David G. Herr has served not only as the director of this dissertation but also as my instructor, colleague, and friend. I wish to express my deepest appreciation for the interest he has shown over the past nine years in my professional development.

I wish to thank Dana Quade for serving as chairman and co-advisor on the dissertation committee. His friendship has been , from the time we worked together on the SENIC project until the present, a sustaining force for me.

A special thanks goes to my family, to whom this work is dedicated. I wish to thank my wife Janice for her love and support and for sharing the job of typing the manuscript, and to my two year old son Joseph for being such a joy.

CONTENTS

Chapter 1

Introduction

One of the most widely used statistical procedures today is the classical analysis of variance (ANOVA). Procedures for calculating sums of squares as test statistics are well known and there is wide agreement on how to interpret the results when the design is balanced. For a balanced a×b two-way fixed effects design, the usual hypotheses of interest are:

H_I: no interaction effects

H_R: no difference in main row effects

H_C: no difference in main column effects.

If the true mean of the (i,j)th cell is μ_{ij} $(i=1,\cdots,a;$ $j=1,\cdots,b)$, then these hypotheses may be parametrically stated as:

H_I: $\gamma_{ij}=0$, $i=1,\cdots,a-1$; $j=1,\cdots,b-1$

H_{R1}: $\alpha_i=0$, $i=1,\cdots,a-1$

H_{C1}: $\beta_j=0$, $j=1,\cdots,b-1$

where the ith row effect is $\alpha_i = \mu_{i\cdot} - \mu_{\cdot\cdot}$, the jth column effect is $\beta_j = \mu_{\cdot j} - \mu_{\cdot\cdot}$, and the (i,j)th interaction effect is $\gamma_{ij} = \mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} - \mu_{\cdot\cdot}$ for:

$\mu_{i\cdot} = \sum_j \mu_{ij}/b$;

$$\mu_{\cdot j} = \sum_i \mu_{ij}/a \; ; \text{ and}$$

$$\mu_{\cdot\cdot} = \sum_i \mu_{i\cdot}/a = \sum_j \mu_{\cdot j}/b \; .$$

Suppose that H_{C1} is the hypothesis of primary interest and that the row factor is included in order to remove any systematic variability due to differences in the row effects. The usual analysis procedure begins with a test of H_I. If the data fail to provide significant evidence against H_I, then test H_{C1}. Accepting H_{C1} further implies that there are no differences in column means over any given row. Rejecting H_{C1} implies that differences in main column effects exist and that these differences will reflect the differences in column means over any given row. If H_I is rejected then the above interpretations of H_C being true or false are not satisfactory since differences in column means need not be consistent over rows in the presence of interaction.

The possible errors incurred with the implementation of the above procedure are summarized in Table 1.1. The errors are indexed by double indices (since two hypotheses are tested) with the first index indicating the type error made testing H_I ('0' means no error, '1' means a type I error, and '2' means a type II error) and the second index indicating the type error made testing H_{C1} ('•' means that H_{C1} is not tested).

The analysis that one might presume is being performed with the prescribed procedure is one where the F statistics for the tests have independent central F distributions

Table 1.1 Possible Errors in Testing Procedure

	Analysis Results ↓	True State of H_I and H_{C1} — H_I True: H_{C1} True	H_I True: H_{C1} False	H_I False: H_{C1} True	H_I False: H_{C1} False
Accept H_I	Accept H_{C1}	no error	type(0,2)	type(2,0)	type(2,2)
	Reject H_{C1}	type(0,1)	no error	type(2,1)	type(2,0)
Reject H_I		type(1,•)		no error	

under their respective null hypotheses. Henceforth we shall refer to this as the independent analysis. Throughout this work we will assume that the marginal tests of H_I and H_{C1} have type I error probabilites of 0.10 and 0.05 respectively.

Consider for example a 3×3 design (based on an example in Kleinbaum and Kupper, 1978) for studying the relationship of patient perception of pregnancy and physician-patient communication to patient satisfaction with medical care. The contrived data shown in Table 1.2 are values of a measure of patient satisfaction with medical care. The row categorization divides the data into three levels of patient worry and the column categorization divides the data into three levels of a measure of affective communication.

These data were generated by using a standard random

Table 1.2 ANOVA Example: The Data

Patient Worry	Affective Communication: High	Medium	Low
Negative	6.67 5.50 4.77 4.29	4.49 4.17 3.86 5.18 4.66 2.52 5.43	4.10 7.51 5.98 5.40
Neutral	3.90 6.27 3.27 4.98 3.68 3.56 5.50	4.61	4.97 4.41 4.21 3.90 3.20 3.73 4.81
Positive	3.08 5.67 4.34 6.62	5.56 5.13 4.47 4.36 3.98 3.67 3.74	8.07 6.24 6.74 4.08

<u>Estimated Cell Means</u>

i \ j	1	2	3	$\hat{\mu}_{i.}$:
1	5.31	4.33	5.75	5.13
2	4.45	4.61	4.18	4.41
3	4.93	4.42	6.28	5.21
$\hat{\mu}_{.j}$:	4.90	4.45	5.40	

number generator and adding the true cell means shown in Table 1.3. The nonsystemic error variance was set to 1.

For the true cell means in Table 1.3, it can be seen that H_I is false and H_{C1} is true. The possible errors in this case are type(2,0) and type(2,1). For the example, the marginal probability of not rejecting H_I is 0.4426. If the design were balanced with five observations per cell, and

Table 1.3 ANOVA Example: True Cell Means

				$\mu_{i.}$:
	5.3	4.4	5.3	5
	4.4	6.2	4.4	5
	5.3	4.4	5.3	5
$\mu_{.j}$:	5	5	5	

thus the same number of observations over all, this probability would reduce to 0.068. Since the marginal probability of rejecting H_{C1} is 0.05, then the probability of a type(2,1) error for the independent analysis is (0.4426)(0.05)=0.0221 or if the balanced design is used, (0.068)(0.05)=.0034.

The independent analysis is one which is in general not realizable with the prescribed procedure. Even with a balanced design, the F statistics are not independent if the same mean squared error is used for each. We could of course use the error sums of squares to make two or more independent estimates of the error variance by ,say, dividing it over groups of cells, however this would result in a loss of power. F statistics which are dependent only by their common denominator we will call singly dependent. For the purpose of comparison with the other analyses, we will make reference to the independent analysis even though one may not exist.

With unbalanced designs many pairs of tests will also have dependent sums of squares as numerators of the

F statistics. F statistics with this dependence in conjunction with a common mean squared error will be called doubly dependent.

The existing computer packages that offer ANOVA procedures for unbalanced designs provide several analyses. Some of these analyses advertise "orthogonality", that is, that the sums of squares add up to the total sums of squares. The hope in using orthogonal sums of squares is that the confounding of effects may in some sense be unraveled. Another "plus" for the orthogonal analysis user is that the F statistics are at most singly dependent. Some of the analyses for unbalanced designs have the property that, unlike the independent analysis, the null distribution of the F statistic for the test of H_{C1} is noncentral. The noncentrality parameter will be unknown unless certain combinations of the cell means are known, which is rarely the case.

We are thus faced with a choice among analyses that are less ideal than the independent one. If the design is balanced, all of the alternative analyses offered by most computer packages are the same. If the design is unbalanced, each possible analysis differs from the independent analysis in some way. The problem is to be able to choose the appropriate analysis for the situation at hand. To do this one needs to know what effect the differences with the independent analysis will have on the error probabilities. In Chapter 2, unbalanced ANOVA is reviewed in general along with the literature on some of the various analyses.

Two of the common analyses for unbalanced designs are, as termed by Herr and Gaebelian (1978), the standard parametric analysis (STP) which is Overall and Spiegel's Method 1 (1969) and the "each adjusted for the other" (so named because the sum of squares for each of the two factors are the "extra" sum of squares explained by the model when the factor is added to the model containing the other factor) which is Overall and Spiegel's Method 2. In SAS (Statistical Analysis System, SAS Institute 1979), for example, one can perform the STP analysis by adding the option SS3 to the MODEL statement and the EAD analysis by adding the option SS2. In general the STP analysis has the property that the $\mathcal{F}$ statistics for H_I and H_{C1} are doubly dependent where they are singly dependent for the EAD analysis. The null distribution for H_{C1} will be central $\mathcal{F}$ for STP but may be noncentral for EAD.

The results of the ANOVA example for STP and EAD are shown in Table 1.4. Per the prescribed procedure, we would not reject H_I using either analysis. Using STP, H_{C1} is not rejected, which constitutes a type(2,0) error. That is, we would conclude that there are no differences in column means. While this is true for the average column means, such is not the case for a given row due to interaction. Using EAD, H_{C1} is not rejected, which constitutes a type(2,1) error. In this case we would conclude that there are differences in column means which are uniform over rows.

Table 1.4 ANOVA Example: Results of the STP and EAD Analyses

STP

SOURCE	DF	SS	F	PROB>F
Affective Communication	2	4.51	1.88	0.168
Patient Worry	2	3.26	1.35	0.271
Interaction	4	5.27	1.10	0.373
Error	36	43.28		

EAD

SOURCE	DF	SS	F	PROB>F
Affective Communication	2	9.78	4.07	0.026
Patient Worry	2	10.00	4.16	0.024
Interaction	4	5.27	1.10	0.373
Error	36	43.28		

It is the uniformity conclusion about the column means that is in error.

The purpose of this work is to compare the STP and EAD analyses to the independent analysis in order to arrive at an assessment of both analyses for various designs and patterns of cell means. This will be done by means of a comparison of error probabilities. In order to calculate these probabilities, we will draw on current distribution theory and make further developments where needed.

To facilitate a discussion of the distributions needed, we introduce the following notation. Let $F(\psi_1,\psi_2,s)$ and $F(\psi_1,\psi_2,d)$ denote bivariate $\mathcal{F}$ distributions where ψ_1 is the noncentrality parameter for the marginal interaction test,

Table 1.5 Distributions Corresponding to STP and EAD Analysis Errors

Error	STP Analysis	EAD Analysis
type(0,1)	$F(0,0,d)$ [1]	$F(0,0,s)$ [1]
type(0,2)	$F(0,\psi_2,d)$ [2]	$F(0,\psi_2,s)$ [2]
type(2,1)	$F(\psi_1,0,d)$ [2]	$F(\psi_1,\psi_2,s)$ [2]
type(2,2)	$F(\psi_1,\psi_2,d)$ [2]	$F(\psi_1,\psi_2,s)$ [2]

[1] The density has been derived, and some computational methods have been developed.

[2] The density has not been derived in the literature (see Chapter 4 for some developments).

ψ_2 is the noncentrality parameter for the marginal column main effects test, 's' denotes single dependence, and 'd' denotes double dependence. The distributions needed to compute the various probabilites are shown in Table 1.5.

In Chapter 2 the current distribution theory is reviewed. We begin with the general gamma-type distributions and move to the bivariate chi-squared and F distributions. Since some of the theory is based on inversions of the characteristic function, we review the characteristic function of the bivariate chi-squared distribution. Finally we consider current computational techniques.

In Chapter 3 we review the geometry of the EAD, STP, and independent analyses. A discussion of the distribution theory specific to this work along with further developments and computational considerations is presented in Chapter 4.

In Chapter 5 we present the comparison of the error probabilities. Under the assumption of no interaction, we consider the effects of single dependence and nonorthogonality on the type(0,1) and type(0,2) errors. When interaction is present we will compare the type(2,1) and type(2,2) errors. Throughout we will illustrate the results with variations on the ANOVA example given in this chapter.

Suggestions for further research are offered in Chapter 6.

Chapter 2

Literature Review

2.1 Nonorthogonal ANOVA

A classical paper dealing with nonorthogonal designs was presented by Yates (1934). He suggested that under the assumption of no interaction, ordinary least squares methods may be used. When this assumption is not made, the recommended procedure is one he termed "the method of weighted squares of means." For the over-parameterized model $Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$ with the usual restrictions $\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$, Searle (1971) showed that the hypothesis tested by the row statistic of this method is the equality of the α_i's. The functional form of this analysis proved cumbersome for practitioners of the pre-computer era, so Yates further suggested an "unweighted means analysis" for which the computations are simplified. Searle showed that this analysis also tests the equality of the α_i's. The corresponding mean square for rows divided by the mean squared error however does not in general have an F distribution as do the variance ratios in the weighted means analysis. (In Chapter 3 we give general conditions on the cell sizes for these statistics to have exact F distributions.)

Eisenhart (1947) presented the "necessary" assumptions for testing hypotheses for a fixed effects design. These assumptions are summarized as:

1. Randomness. The cell means are fixed and the observations are values of random variables distributed about the means.
2. Additivity. This is equivalent to assuming no interaction.
3. Equal variances and zero correlations. The variance-covariance matrix of the vector of observations is $\sigma^2>0$ times the identity matrix.
4. Normality. The vector of observations is an observed random vector whose elements have a jointly normal distribution.

Scheffé (1959) discussed the impact of the violation of one or more of these assumptions. One common assumption that is not made above is that of equal cell sizes. This leads to the question, "What is the impact upon the ANOVA when the cell sizes are not all equal?"

Searle (1971) studied this problem in terms of various decompositions of the total sums of squares via the regression approach. Let $R(\mu,\alpha,\beta,\gamma)$ denote the total corrected sum of squares one obtains when fitting the unrestricted model $Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$ by least squares. Omission of one or more of the terms μ,α,β,γ indicates fitting the model without including these terms. For example, $R(\mu,\alpha)$ refers to the total corrected sum of squares for the model

$Y_{ijk} = \mu + \alpha_i + e_{ijk}$. The method proposed for testing hypotheses about the parameters of the full model involves a technique that is used to test the significance of the additional variation explained by an independent variable in a regression model in the presence of other regressors. Searle showed that the sum of squares $R(\mu,\alpha)-R(\mu) = R(\alpha|\mu)$ tests the hypothesis:

$$H_{R2}\colon\ (1/n_{i\cdot}) \sum_j n_{ij}\mu_{ij} \quad \text{equal for all } i,$$

and $R(\mu,\beta)-R(\mu) = R(\beta|\mu)$ tests

$$H_{C2}\colon\ (1/n_{\cdot j}) \sum_i n_{ij}\mu_{ij} \quad \text{equal for all } j.$$

These hypotheses may also be stated as:

$$H_{R2}\colon\ \alpha_i + (1/n_{i\cdot}) \sum_j n_{ij}(\beta_j + \gamma_{ij}) \quad \text{equal for all } i, \text{ and}$$

$$H_{C2}\colon\ \beta_j + (1/n_{\cdot j}) \sum_i n_{ij}(\alpha_i + \gamma_{ij}) \quad \text{equal for all } j.$$

The sum of squares $R(\mu,\alpha,\beta)-R(\mu,\beta) = R(\alpha|\mu,\beta)$ and $R(\mu,\alpha,\beta)-R(\mu,\alpha) = R(\beta|\mu,\alpha)$ tests

$$H_{R3}\colon\ \delta_i = 0 \quad \text{for all } i, \text{ and}$$

$$H_{C3}\colon\ \varepsilon_j = 0 \quad \text{for all } j,$$

respectively, where:

$$\delta_i = \sum_j (n_{i\cdot} - n_{ij}^2/n_{\cdot j})\alpha_i - \sum_{i' \neq i} \sum_j (n_{ij}n_{i'j}/n_{\cdot j})\alpha_{i'} + \sum_j (n_{ij} - n_{ij}^2/n_{\cdot j})\gamma_{ij} - \sum_{i' \neq i} \sum_j (n_{ij}n_{i'j}/n_{\cdot j})\gamma_{i'j} .$$

A similar definition is given for ε_j . These parameters can also be defined as:

$$\delta_i = \sum_j \{n_{ij}\mu_{ij} - \sum_{i'} (n_{ij}n_{i'j}/n_{\cdot j})\mu_{i'j}\}, \text{ and}$$

$$\varepsilon_j = \sum_i \{n_{ij}\mu_{ij} - \sum_{j'} (n_{ij}n_{ij'}/n_{i\cdot})\mu_{ij'}\}.$$

Each of these hypotheses can be tested under Eisenhart's assumptions 1, 3, and 4 with exact methods. Burdick *et al.* (1974) describe these exact methods from a geometric point of view. They show that the sum of squares for testing any estimable linear hypothesis of μ_{ij}'s can be expressed as the squared length of the perpendicular projection of the data vector onto some vector space corresponding to violations of the hypothesis. The spaces for testing H_{R1} and H_{C1} are shown to be in general nonorthogonal as are the spaces for testing H_{R2} and H_{C2}. The spaces corresponding to H_{R3} and H_{C3} are always orthogonal to the spaces corresponding to H_{C2} and H_{R2}, respectively, but are in general not orthogonal to each other. Conditions for orthogonality of the first pair of spaces are given as functions of the cell sizes. For testing H_{R1} and H_{C1}, the conditions are $1/n_{ij} = r_i + s_j$ where r_i and s_j are real numbers. For testing H_{R2} and H_{C2}, the conditions are $n_{ij} = p_i q_j$ where p_i and q_j are positive real numbers. The latter condition is oft termed the condition of proportionality. Two other analyses are studied by Burdick *et al.* which involve the paired hypotheses H_{R2}, H_{C3} and H_{C2}, H_{R3} . The spaces corresponding to these spaces will be, as stated above, orthogonal.

The practitioner is faced with choosing among these various methods when dealing with unbalanced designs, or so it seems. The question "how to choose?" sparked a series of interesting articles in the psychological research methods literature. Overall and Spiegel (1969) described three methods for testing H_R and H_C . Method 1 is equivalent to the weighted squares of means analysis and the STP analysis. Method 2 is equivalent to testing the sums of squares $R(\alpha|\mu,\beta)$ and $R(\beta|\mu,\alpha)$ which is the EAD analysis. Method 3 is testing the sums of squares $R(\alpha|\mu)$ and $R(\beta|\mu,\alpha)$ or $R(\beta|\mu)$ and $R(\alpha|\mu,\beta)$. They claim that "...there is no way to compare and to say (for example) that Method 1 is the proper generalization (of the equal cell sizes case) while Methods 2 and 3 are not." Their recommendations are:

Method 1: "...should be used if conceptualization of the problem is in general linear regression terms and not in experimental design terms." Furthermore, "...care should be exercised not to interpret significant 'main effects' in the same way that one might interpret a significant main effect from a balanced factorial design."

Method 2: should be used "...whenever the problem is conceived as a multiclassification factorial design and where conventional analysis of variance might have been employed except for unequal cell frequencies...."

Method 3: should be used "...if a logical apriori ordering exists among hypotheses to be tested."

Rawlings (1972) claimed that Method 1 and 2 do not provide tests of the simple hypothesis (H_{R1} or H_{C1}) in the presence of interaction. He further claimed that Method 1 does not result in "exact statistics" since certain orthogonality conditions are not met. (It needs to be pointed out here that the sums of squares for Method 1 are the squared lengths of the perpendicular projections of the data vector onto the spaces described by Burdick *et al.* (1974) that test H_{R1} and H_{C1}. Method 1 not only tests the simple hypotheses, but it is an exact method.) Rawlings recommended an alternative to Method 1, Method 1′, which tests first for interaction and then if nonsignificant, tests for main effects with the interaction sum of squares included with the error sum of squares.

Overall and Spiegel (1973) felt that Rawlings' new method was their Method 2. They also pointed out some difficulties with pooling interaction with error. Rawlings (1973) returned with stating that (in spite of his earlier claims) when there is no interaction, Methods 1, 2, and 1′ all test the simple hypotheses - so use the one with the greatest power. Rawlings felt that Method 1′ will have power at least as large as the other two methods.

Appelbaum and Cramer (1974) treat the problem in terms of mutiple regression as a comparison of competing linear models. They claimed that this resolves the problem in an "obvious manner." These models "are logically independent of the observed numbers of observations per cell...", however "...the analyses will be affected by the cell frequencies." They advocated Method 2 stating that it always gives the "correct estimate". Their algorithm involves three steps:

1. Test for interaction. If significance is found then stop.

Otherwise, proceed to step 2.

2. Test the sums of squares $R(\alpha|\mu,\beta)$ and $R(\beta|\mu,\alpha)$. If both are significant, accept the model $Y = \mu+\alpha+\beta+e$. If one is significant, accept the model $Y = \mu+\alpha+e$ or $Y = \mu+\beta+e$, respectively. Otherwise, proceed to step 3.
3. Test $R(\alpha|\mu)$ and $R(\beta|\mu)$. If neither is significant, then accept the model $Y = \mu+e$. If one is significant, accept the model $Y = \mu+\alpha+e$ or $Y = \mu+\beta+e$, respectively, but interpret "cautiously". Otherwise no conclusion is made.

This algorithm actually constitutes a model selection process, however they imply that accepting the model $Y = \mu+\beta+e$ over $Y = \mu+\alpha+\beta+e$, for example, says that H_R is true. O'Brien (1976) amended Appelbaum and Cramer's algorithm by adding to step 2 the following: when only one of the tests performed is significant, also test $R(\alpha|\mu)$ or $R(\beta|\mu)$ respectively. If this test is also significant, then accept the corresponding model. If the additional test is not significant, "adopt the model $Y = \mu+\alpha+\beta+e$, but cautiously interpret its meaning."

The final article in this series was presented by Herr and Gaebelein (1978). They termed the method of Appelbaum and Cramer as a "regression-oriented view", wheras the method of Burdick *et al.* (1974), for example, is an "effect-oriented view". The testing of H_R and H_C should include the desired interpretation. Five analyses are presented:

1. "STandard Parametric" (STP) analysis tests H_{R1} and H_{C1} .
2. "WeighTed Means" (WTM) analysis tests H_{R2} and H_{C2} .
3. "Each ADjusted" (EAD) analysis tests H_{R3} and H_{C3} .

4. "Hierarchical: A (row) first, then B (column)" (HAB) analysis tests H_{R2} and H_{C3} .
5. "Hierarchical: B (column) first, then A (row)" (HBA) analysis tests H_{R3} and H_{C2} .

The sums of squares for the above analyses are obtained via the methods described by Burdick *et al.* (1974). Herr and Gaebelein pointed out that in SAS (SAS Institute, 1979) the sums of squares for STP are the Type III SS; for WTM two runs are required alternating rows and columns being listed first in the MODEL statement and the Type I SS for the one listed first is used; for EAD the sums of squares are the Type II SS; for HAB, the Type I SS with rows listed first; and for HBA, the Type I SS with columns listed first. They surmised that the practitioner's dilemma is not to choose one of the five analyses, but to choose which of the parametric hypotheses he wishes to test. Once a hypothesis is chosen, the appropriate analysis is determined. It follows that if one wishes to test H_{R1} and H_{C1}, which would have been tested if the design were balanced, then one should use the STP analysis (which is Yates' weighted squares of means analysis and Overall and Spiegel's Method 1 as stated by Hinkelmann, 1975).

Hocking *et al.* (1980) investigated the STP, WTM, and EAD analyses to determine under what conditions each test the simple hypotheses (H_{R1} and H_{C1}). For example, they gave model restrictions that provide sufficient conditions for the analyses to test H_{R1} . These are:

STP: all cells must be filled;

WTM: H_{C1} and H_I are true; and

EAD: H_I is true.

Burdick *et al.* (1974) gave conditions on the cell sizes that are sufficient for the analyses to test H_{R1} and H_{C1}. These are:

STP: $n_{ij} > 0$;

WTM: to test H_{R1} -- $n_{ij} = p_i$,

to test H_{C1} -- $n_{ij} = q_j$,

to test both -- $n_{ij} = p$; and

EAD: same as for WTM.

Under the restriction of no interaction, Hocking *et al.* showed that STP and EAD test the same hypotheses, thus a preference should be based on power. Burdick and Herr (1980) showed that EAD has power at least as large as STP. The proof is in recognizing that the leg of a right triangle is shorter than the hypotenuse. The leg and the hypotenuse correspond to the noncentrality parameters of the $\mathcal{F}$ statistics associated with STP and EAD respectively. (See Figure 6, Chapter 3.)

Two other papers which cite parametric versions of the hypotheses and appeared about the time of Burdick's and Appelbaum and Cramer's were Carlson and Timm (1974) and Kutner (1974). Both recommended testing H_{R1} and H_{C1} for general situations. Carlson and Timm along with Bryce (1975) equated the sums of squares associated with the hypotheses to Searle's $R(\cdot)$ notation. Gianola (1975) revisited a notion discovered by Searle (which has confused many on this issue). He stated that the sums of squares for H_{R1} and H_{C1} would be, as Bryce also noted, $R(\alpha|\mu,\beta,\gamma)$ and $R(\beta|\mu,\alpha,\gamma)$ respectively. But each of these values is zero! He concluded that "since, in the model with interaction, differences between rows or between columns cannot be estimated in the absence of interaction effects, it does not seem reasonable to compute sums of squares for main effects without interaction effects in their expectations." Kutner (1975) resolved Gianola's confusion (or did he?) by pointing out that if the usual "estimability" assumptions, $\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$, are made then $R(\alpha|\mu,\beta,\gamma)$ and $R(\beta|\mu,\alpha,\gamma)$ are nonzero and do provide the appropriate sums of squares. Unfortunately Nelder (1975) added to the confusion by stating that fitting γ_{ij} subject to constraints is undesirable. His reasons were: (1) if interaction is negligible, loss of power occurs, and (2) if interaction is present (not negligible?), the sums of squares correspond to uninteresting hypotheses. Obviously Nelder did not realize what hypotheses STP tests.

The papers by Burdick *et al.* (1974), Herr and Gaebelein (1978), and Hocking *et al.* (1980) among others resolved the problem of how to proceed with hypotheses testing in unbalanced two-way designs. Two recent textbooks bear out the fact that the message has not been heard by all. Keppel (1982) treats the unweighted means analysis much as Yates presented it in his 1934 paper. Kleinbaum and Kupper (1978) mention Yates' unweighted means analysis and present four additional analyses under the heading of "regression approach". The four methods turn out to be HAB, HBA, EAD, and STP (when the estimability assumptions are made). What is confusing is that under the heading of the null hypothesis tested, the same parametric representation appears for each method. It seems they are saying that the same hypotheses are tested under each analysis.

2.2 Central Multivariate Gamma-Type Distributions

A gamma-type multivariate distribution can be constructed by considering random vectors of the form $X_i = (X_{i1}, \ldots, X_{im})'$ for $i = 1, \ldots, n$ where the n vectors are independently and identically distributed as multivariate normal, each having a common variance - covariance matrix V where the diagonal elements of V are all ones. The statistics $S_j = \sum_{i=1}^{n} (X_{ij} - \overline{X}_{\cdot j})^2$, $j = 1, \ldots, m$, are distributed as chi-squared random variables with $n - 1$ degrees of freedom. The joint distribution of $S_1, S_2, \ldots, S_m$ is a multivariate chi-squared distribution (Krishnaiah *et al.*, 1961). The joint distribution of $\sqrt{S_1}, \ldots, \sqrt{S_m}$ is a generalized Rayleigh distribution (Miller, 1964). These distributions are described by Johnson and Kotz (1972) under the general heading of multivariate gamma-type distributions. When $m = 2$ and $\rho = \mathrm{corr}(X_{i1}, X_{i2})$, $i = 1, \ldots, n$, the joint distribution function of S_1, S_2, (the standard bivariate central chi-squared distribution) is given by:

(2.1) $\text{Prob}(S_1 \leq s_1,\ S_2 \leq s_2) =$

$$\sum_{j=0}^{\infty} c_j \text{Prob}(\chi^2_{n-1+2j} \leq (1-\rho^2)^{-1}s_1)\text{Prob}(\chi^2_{n-1+2j} \leq (1-\rho^2)^{-1}s_2)$$

where $c_j = \Gamma((n-1)/2+j)(1-\rho^2)^{(n-1)/2}\rho^{2j}/j!$ (Bose 1935, Finney 1938). A general standard bivariate gamma distribution as defined by Johnson and Kotz is obtained by replacing (n-1) in 2.1 with ν where $\nu>0$. Jensen (1970b) derived a further generalized form by allowing different variance-covariance matrices for the X_i random vectors. Kibble (1941) derived the joint density of S_1, S_2 where $\text{corr}(S_1,S_2) = \rho^2$ and S_j (j=1,2) has a standard central gamma distribution with density function:

(2.2) $p_{S_j}(s_j) = s_j^{\Theta_j-1} e^{-s_j}/\Gamma(\Theta_j)$

for $\Theta_j>0$, $s_j>0$. The joint density of S_1 and S_2 is:

(2.3) $$p_{S_1,S_2}(s_1,s_2) = (s_1s_2)^{\Theta-1} e^{-s_1-s_2} \{\Gamma(\Theta)\}^{-2} \times$$
$$\{1+\sum_{k=1}^{\infty} \rho^{2k}\mathcal{L}_k^{\Theta-1}(s_1)\mathcal{L}_k^{\Theta-1}(s_2)\}$$

where $\Theta=\Theta_j>0$; $0 \leq \rho < 1$; s_1, $s_2 >0$; and

$$\mathcal{L}_k^{\Theta-1}(s) = \left[\Gamma(\Theta)\Gamma(\Theta+k)/k!\right]^{\frac{1}{2}} \sum_{h=0}^{k} (-1)^h \binom{k}{h} s^h/\Gamma(\Theta+h)$$

is the Laguerre polynomial. Sarmanov (1968) labeled 2.3 a symmetrical (bivariate) gamma distribution since $\Theta_1 = \Theta_2$. D'jackenko (1961, 1962a, 1962b) calculated tables for Kibble's result for ρ=0.9 and Θ=2,3,4,5. Sarmanov (1970a, 1970b) constructed an asymmetrical bivariate gamma distribution with density

$$(2.4)\quad p_{S_1,S_2}(s_1,s_2) = s_1^{\Theta_1-1} s_2^{\Theta_2-1} \exp(-s_1-s_2)\ \{\Gamma(\Theta_1)\Gamma(\Theta_2)\}^{-1} \times \{1+\sum_{k=1}^{\infty} a_k \mathcal{L}_k^{\Theta_1-1}(s_1)\mathcal{L}_k^{\Theta_2-1}(s_2)\}$$

where for some $\lambda>0$ and $\lambda<1$,

$$a_k = \lambda^k\left(\frac{\Gamma(\Theta_1)\Gamma(\Theta_2+k)}{\Gamma(\Theta_2)\Gamma(\Theta_1+k)}\right) \text{ and } \Theta_1 \geq \Theta_2 \ .$$

In this case, $\mathrm{corr}(S_1,S_2) = a_1 = \lambda\sqrt{\Theta_2/\Theta_1}$. A generalized symmetrical gamma distribution is obtained by replacing $\rho^2, \rho^4, \cdots$ in 2.3 with a sequence of nonnegative numbers $c_1, c_2, \cdots$ such that $\sum_{k=1}^{\infty} c_k$ converges. Since

$$\int_0^\infty s^{\Theta-1}\exp(-s)\mathcal{L}_k^{\Theta-1}(s)ds = 0 \quad \text{for } k \geq 1, \text{ then}$$

$$\int_0^\infty\int_0^\infty p_{S_1,S_2}(s_1,s_2)ds_1 ds_2 = 1 \quad .$$

Sarmanov (1968) showed that $p_{S_1,S_2}(s_1,s_2) \geq 0$ if and only if $c_1, c_2, \cdots$ are the central moments of a random variable Y for which $P(0 \leq Y < 1)=1$. Under these conditions, $p_{S_1,S_2}(s_1,s_2)$ is a density function.

Jensen (1970a) derived a bivariate chi-squared density for $S_1^* = \sum_i^n X_{i1}^2$ and $S_2^* = \sum_i^n X_{i2}^2$ where each X_{ij} is a standard normal random variable, $\mathrm{corr}(X_{i1},X_{i2}) = \rho_i$, and

$$\mathrm{corr}(X_{ij},X_{i'j'}) = 0 \text{ for } i \neq i' \ (i,i'=1,\cdots,n;\ j,j'=1,2) \ .$$

Note that (X_{i1},X_{i2}) $(i=1,\cdots,n)$ are canonical pairs with canonical correlations $\rho_1,\rho_2,\cdots,\rho_n$. The joint density, derived via the probability integral transformation, is:

$$(2.5)\quad f_{S_1^*,S_2^*}(s_1,s_2) = \sum_{k=0}^{\infty} G_k(\boldsymbol{\delta}) \sum_{p=0}^{k} \sum_{q=0}^{k} b_{pq}(k)\Psi(s_1;n/2+p) \times \Psi(s_2;n/2+q)$$

where $\boldsymbol{\delta}' = (\rho_1^2,\rho_2^2,\cdots,\rho_n^2)$;

$$G_k(\boldsymbol{\delta}) = \sum \cdot \cdot \cdot \cdot \sum_{j_1+\cdots+j_n=k} a_{1j_1} a_{2j_2} \cdots a_{nj_n} \ ;$$

$$a_{ij} = \rho_i^{2j}\ \Gamma(j+\tfrac{1}{2})\ \{\Gamma(j+1)\Gamma(\tfrac{1}{2})\}^{-1}\ ;$$

$$b_{pq}(k) = (-1)^{p+q}\binom{k}{p}\binom{k}{q};\ \text{and}$$

$$\Psi(s;\nu) = (\tfrac{1}{2})^{\nu} s^{\nu-1}\exp(-s/2)/\Gamma(\nu)\ .$$

Note that $\Psi(s;\nu)$ is the gamma density with parameters ν and $\frac{1}{2}$.

2.3 Noncentral Multivariate Gamma Distributions

Johnson and Kotz (1972) described what they called biased Rayleigh distributions for which the marginal distributions of $S_1,\cdots,S_m$ are noncentral χ^2 distributions with n-1 degrees of freedom and noncentrality parameters

$$\sum_{i=1}^{n}(\xi_{i1}-\overline{\xi}_{\cdot 1})^2,\ \cdots,\ \sum_{i=1}^{n}(\xi_{im}-\overline{\xi}_{\cdot m})^2$$

respectively where $E(X_{ij}) = \xi_{ij}$ and $\overline{\xi}_{\cdot j} = \sum_{i=1}^{n}\xi_{ij}/n$. Blumenson and Miller (1963) derived a joint density assuming that $V = (v_{pq})$ is of the Jacobi form, that is, $v_{pq}=0$ if $|p-q|>1$. Miller and Sackrowitz (1967) derived a form of the density of the ratio of a noncentral chi-squared random variable with d degrees of freedom to a central chi random variable with d+1 degrees of freedom. Miller *et al.* (1958)

derived a jointly noncentral chi distribution where X_{i1}, X_{i2} $(i=1,\cdots,n)$ are canonical variates with correlation ρ, and $\mathrm{Var}(X_{ij}) = \sigma^2$ $(i=1,\cdots,n;j=1,2)$. The form of the density is:

$$(2.6)\quad p_{R_1,R_2}(r_1,r_2) = \{2\eta r_1 r_2 \Gamma(n)/\sigma^2(n-2)\}\{2/\rho\omega^2\sigma^2\}^{(n-2)/2} \times \exp\{-(\eta(r_1^2+r_2^2)+A\omega)\} \times \sum_{k=(n-2)/2}^{\infty} 2k\binom{n/2+k-2}{n-3} I_k(2\rho\eta r_1 r_2) \times I_k(\omega r_1) I_k(\omega r_2)$$

where $R_j = S_j^{\frac{1}{2}}$ $(j=1,2)$; $\omega = A/\sigma^2(1+\rho)$;

$A = \sum_{i=1}^{n} \xi_{ij}^2$ $(j=1,2)$; $\eta = \{2\sigma^2(1-\rho^2)\}^{-1}$; and $I_k(\cdot)$ is the modified Bessel function (Watson,1962). The above density function assumes $n>2$ and $\xi_{i1}=\xi_{i2}$ $(i=1,\cdots,n)$. Miller *et al.* further derived the density relaxing the equal expectations assumptions and allowed for unequal variances (but still required equal canonical correlations).

Johnson (1978) derived the jointly noncentral chi-squared distribution of $Y_1 = X_{11}^2$ and $Y_2 = X_{12}^2$, where $\mathrm{Var}(X_{11}) = \mathrm{Var}(X_{12}) = \sigma^2$ and $\rho = \mathrm{Corr}(X_{11},X_{12})$, which has density

$$(2.7)\quad p_{Y_1,Y_2}(y_1,y_2) = 2K_0(y_1y_2)^{-\frac{1}{2}} \exp\{K_1(y_1+y_2+\xi_{11}^2+2\rho\xi_{11}\xi_{12}+\xi_{12}^2)\} \times \sum_{i=1}^{2}\{\exp((-1)^{i+1}2K_1K_2\sqrt{y_1})\cosh(2K_1(K_3\sqrt{y_1}+(-1)^i\sqrt{y_1y_2}))\}$$

where $K_0 = \{8\pi\sigma^2\sqrt{1-\rho^2}\}^{-1}$; $K_1 = -\{2\sigma^2(1-\rho)\}^{-1}$;

$K_1 = \rho\xi_{11} - \xi_{12}$; and $K_3 = \rho\xi_{12} - \xi_{11}$.

Johnson also derived an integral form of the joint density of S_1 and S_2 allowing for arbitrary canonical correlations. Let $Z_j' = (X_{1j},\cdots,X_{nj})$ $(j=1,2)$ and $F_{Z_1,Z_2}(\cdot,\ \ldots,\cdot)$ denote the multivariate normal density function of Z_1 and Z_2. Then the joint density of S_1 and S_2 is (using the notation of Miller *et al.*):

$$(2.8)\quad f_{S_1,S_2}(s_1,s_2) = \int_{\{z_1:|z_1|=\sqrt{s_1}\}}\iint_{\{z_2:|z_2|=\sqrt{s_2}\}}\int F_{Z_1,Z_2}(z_1,z_2)dz_1dz_2$$

where $|z|$ denotes the norm of z, and $\int_{\mathcal{S}}\int$ denotes integration over the set $\mathcal{S}$.

Krishnam (1976) also derived a jointly noncentral chi-squared density which is a transformed version of 2.6.

2.4 Bivariate $\mathcal{F}$ Distributions

A bivariate $\mathcal{F}$ distribution may be constructed by considering the random variables S_0, S_1, S_2, each having a chi-squared distribution with t, r, and s degrees of freedom respectively. S_0 is assumed to have a central distribution and is independent of both S_1 and S_2. The statistics $F_1 = tS_1/rS_0$ and $F_2 = tS_2/sS_0$ each have $\mathcal{F}$ distributions with (r,t) and (s,t) degrees of freedom respectively. It is evident that F_1 and F_2 may not be independent since they share a common denominator, S_0. When S_1 and S_2 are independent, F_1 and F_2 are "singly dependent". When S_1 and S_2 are dependent, F_1 and F_2 are "doubly dependent". A form of the density for singly dependent bivariate central $\mathcal{F}$

distributions with r=s was derived by Finney (1938). Krishnaiah and Armitage (1965) tabulated Finney's distribution for the upper 10th, 5th, 2.5th, and 1st percentiles with $t=1,\cdots,19$ and $r=s=5,\cdots,45$. Kimball (1951) derived an inequality for the singly dependent doubly central $\mathcal{F}$ distribution which states that the dependence between F_1 and F_2 increases the joint probability that F_1 and F_2 are less than some upper bounds over the case when F_1 and F_2 are independent. That is:

$$P(F_1 \le f_1 ; F_2 \le f_2) \ge P(F_1 \le f_1) P(F_2 \le f_2) \ .$$

Olkin (1971) showed that when r=s and $f_1=f_2=f$,

$$P(F_1 \le f ; F_2 \le f) \ge 2P(F_1 \le f) - P(F_2 \le 2f) \ .$$

Jensen (1969b) showed Kimball's inequality true for doubly dependent, doubly central $\mathcal{F}$ distributions. Jensen further derived a form of the density given by:

$$(2.9) \quad p_{F_1,F_2}(f_1,f_2) = \sum_{k=0}^{\infty} G_k(\delta) \sum_{m=0}^{k} \sum_{n=0}^{k} d_{mn}(k) \times \frac{f_1^{(r+2m-2)/2} \, f_2^{(s+2n-2)/2}}{\{1+(r/t)f_1+(s/t)f_2\}^{(r+s+2m+2n+t)/2}} \quad \text{where}$$

$$d_{mn}(k) = (-1)^{m+n} \binom{k}{m} \binom{k}{n} \, r^{(r+2m)/2} \, s^{(s+2n)/2} \times \frac{\Gamma((r+s+2m+2n+t)/2)}{\Gamma((r+2m)/2)\Gamma((s+2n)/2)\Gamma(t/2)} \ ,$$

and $G_k(\delta)$ is as defined in 2.5 .

2.5 Characteristic Function of the Multivariate Chi-squared Distribution

Many of the density functions presented in the previous sections were derived by means of inverting the characteristic function of the distribution. Kibble (1941), for example, derived the joint characteristic function of $X_1^2/2$ and $X_2^2/2$, which is written here as:

$$\phi(t_1,t_2) = \{(1-it_1)(1-it_2) - \rho^2 it_1 t_2\}^{-\frac{1}{2}}$$

where X_1 and X_2 are standard normal random variables and $\mathrm{Corr}(X_1,X_2) = \rho$. Jensen (1969c) showed that the characteristic function for the m-dimensional chi-squared distribution is:

$$(2.10)\quad \phi(T) = |M|^{-v/2} \exp(iv\mu' M^{-1} T\mu) \;,$$

which is defined for real T where:

$M = I - 2iTC$;

T is the m×m diagonal matrix with real elements $\{t_1,\cdots,t_m\}$;

I is the m×m identity matrix;

C is the m×m correlation matrix of the vector of m chi-squared variables;

v is the common degrees of freedom of the m variables; and

$\mu' = (\mu_1,\cdots,\mu_m)$ where μ_j $(j=1,\cdots,m)$ is the common expectation of v independent, unit variance, normal variates whose sum of squares is the jth marginal chi-squared variable.

Lancaster (1958,1963) gave much of the theorical basis needed to invert these characteristic functions in terms of orthogonal systems of polynomials. Gurland (1955) derived the inverse Fourier exponential transform of the function:

$$w(t;g,h) = (1-it)^{-g}\{it/(1-it)\}^{h}$$

which is given by:

$$(2.11) \quad (2\pi)^{-1}\int_{\infty}^{\infty} \exp(-itx)w(t;g,h)dt = h!\Gamma(g)\mathcal{L}_h^{(g-1)}(x)x^{g-1}\exp(-x)\{\Gamma(g)\Gamma(g+h)\}^{-1}$$

where $\mathcal{L}_h^{(g-1)}(x)$ is the Laguerre polynomial of order h. This transform was used by Jensen(1970a) in the derivation of 2.5 and 2.9 .

2.6 Computational Procedures for Evaluating Multivariate Gamma-Type Distributions

The problem considered here is the evaluation of $P(S_1 \leq s_1; S_2 \leq s_2)$ or $P(F_1 \leq f_1; F_2 \leq f_2)$ where S_1 and S_2 are distributed as chi-squared random variables with r and s ($r \leq s$, say) degrees of freedom and $F_1 = tS_1/rS_0$ and $F_2 = tS_2/sS_0$ where S_0 is a central chi-squared random variable with t degrees of freedom and is independent of S_1 and S_2 . The second probability can be derived, given knowledge of the first, as:

$$(2.12) \quad P(F_1 \leq f_1; F_2 \leq f_2) = \int_0^{\infty} P(S_1 \leq rf_1x/t; S_2 \leq sf_2x/t)dP(S_0 \leq x) .$$

Let $X = (X_1' \; X_2')'$ be an (r+s)-dimensional random vector with X_i distributed as normal with mean vector μ_i and dispersion matrix $\sigma^2 I_i$, where I_i is the identity matrix with the same dimension as X_i. Furthermore, let $Cov(X_1,X_2) = \sigma^2 C$ where $C = (D \vdots 0)$ is $r \times s$, D is an $r \times r$ diagonal matrix with diagonal elements $\rho_1, \cdots, \rho_r$, and 0 is an $r \times (s-r)$ matrix of zeros. Suppose that $S_i = X_i' X_i / \sigma^2$ $(i=1,2)$. Then

$$(2.13) \quad P(S_1 \leq s_1; S_2 \leq s_2) = \int_{\{x: S_1 \leq s_1; S_2 \leq s_2\}} \int p_X(x) dx,$$

where $p_X(\cdot)$ is the density function of X.

One method of evaluating 2.13, following Miller *et al.* (1958) and Johnson (1978), is to make a change of variables to polar coordinates and solve the multiple integral explicitly or numerically. In lieu of the densities derived by Jensen (1970a) and Kibble (1941), it appears that quadrature procedures using gamma type weights are the most promising numerical methods. One such procedure is described by Ralston (1965). The quadrature function $Q(s_1, s_2) = \sum_{i=1}^{n} H_i p_X(a_i)$ is used to approximate the probability in 2.13, where H_i and a_i depend only on s_1, s_2, and n. The a's are vectors of zeros of Laguerre polynomials and the H's are known functions of these zeros. The a's and H's have been tabulated for various n by Shao *et al.* (1964).

Krishnaiah (1965), Amos (1972), and Gunst (1973b) considered the evaluation of bivariate doubly central $\mathcal{F}$ probabilities via the bivariate $\mathcal{F}$ density for specific cases where $\rho_i = \rho$ $(i=1,\cdots,r)$. Gunst and Webster (1973a) suggested using the the equicorrelation density to approximate the general case by substituting $(\sum_{i=1}^{r} \rho_i^2/r)^{\frac{1}{2}}$ for each ρ_i. Based of Monte Carlo estimates, they showed that these approximations overestimate the true probabilities when not all ρ_i $(i=1,\cdots,r)$ are equal.

Dutt and Soms (1976) used the results of Gurland (1948) to derive an integral transformation approximation for calculating probabilities of the central multivariate chi-squared distribution. The method involves calculating the real parts of probability integral transformations using characteristic functions of all possible marginal distributions and substituting them into Gurland's density expansion. This procedure appears to be practical only for the cases when $\rho_i = \rho$ $(i=1,\cdots,r)$, $s_1 = s_2$, and $r = s$.

Tan and Wong (1978) presented a promising approximation for the general noncentral case. Let X_1 and X_2 be any nonnegative continuous random variables with density $F(x_1,x_2)$ and the following properties:

(i) $E(X_1^u\ X_2^v)>\infty$ for any $u,v\geq 0$;

(ii) $E(X_i) = m_i>0$ $(i=1,2)$; and

(iii) $\mathrm{Var}(X_i)>0$ $(i=1,2)$.

Define a function $g_{k_1,k_2}(x_1,x_2)$ such that

$$(2.14)\quad \int_0^\infty\int_0^\infty x_1^u x_2^v g_{k_1,k_2}(x_1,x_2)dx_1dx_2 = E(X_1^uX_2^v)$$

for any $u = 0,1,\cdots,k_1$ and $v = 0,1,\cdots,k_2$. The function g is used as an approximation of $F(x_1,x_2)$ where k_1 and k_2 are chosen by the user. Tan and Wong reported that their g used as an approximation to a doubly noncentral gamma distribution is "good" for "moderate" canonical correlations and noncentrality parameters.

McAllister *et al.* (1981a) presented a procedure for computing the probabilities of Jensen's bivariate doubly central $\mathcal{F}$. Their approximation is based on the first few terms of the infinite sum:

$$(2.15) \quad \sum_{k=0}^{\infty} G_k(\delta) \sum_{m}^{k} \sum_{n}^{k} b_{mn}(k) I_2(r/2+m, s/2+n, rf_1/t, sf_2/t, t/2)$$

where $I_2(\cdot)$ is the integral derived by Amos and Bulgren (1972).

2.7 Applications: Simulataneous Hypothesis Testing

Many authors have noted that simultaneous tests of several variance ratios all sharing a common denominator will be dependent. Singly dependent $\mathcal{F}$ tests have been studied by Kimball (1951), Ghosh (1955), Ramachandran (1956), and Hurlburt and Spiegel (1976). Doubly dependent $\mathcal{F}$ tests with common canonical correlations have been studied by Krishnaiah (1964), Jensen (1969a), and McAllister *et al.* (1981b). Jensen (1970a) and McAllister *et al.* (1981b) also allowed for unequal canonical correlations. These authors used the distribution theory in the previous sections as a basis for their discussions.

2.8 Applications: Conditional Type I Error Probilities

Pertinent to this work are the previous attempts at quantifying probabilities relating to the simultaneous testing of any two of the three hypotheses H_R, H_C, and H_I. Hurlburt and Spiegel (1976) investigated the probability of rejecting H_R conditioned on rejecting H_C for singly dependent $\mathcal{F}$ tests. They noted that when the numerator degrees of freedom are small relative to the denominator degrees of freedom, the variance ratios are approximately independent. Their examples also suggest that when all the degrees of freedom are nearly equal, the conditional type I error probability is significantly larger than the unconditional probability.

McAllister *et al.* (1981b) focused their attention on the EAD analysis and reported probabilities obtained from using evaluation techniques (McAllister *et al.*, 1981a) on Jensen's bivariate F distribution. Their investigation allowed for various combinations of canonical correlations. Because they were dealing with the EAD analysis, they made the necessary assumption that there is no interaction (otherwise Jensen's formula would not apply). They advocated consideration of experimentwise error rates when performing multiple F tests of main effects much as one would do when "assessing differences among levels of a main effect."

None of the papers reviewed dealt with the STP analysis directly, the EAD analysis in the presence of interaction, or the relative power of these analyses.

CHAPTER 3

The Geometry of ANOVA and the STP and EAD Analyses

The ANOVA of fixed effects two-way factorial designs is often studied and applied in terms of its relationship to multiple linear regression. To many users the mathematics of ANOVA is a collection of least squares equations and $\mathcal{F}$ ratios with little to help those seeking a basis for understanding. The foundations of linear regression are embedded in the geometry of finite dimensional Euclidean vector spaces. Within this realm an intuitively appealing development of the theory of linear models is possible. For more on this and the history of geometry in linear models, see Herr 1980.

In this chapter we will discuss some of the geometric aspects of ANOVA and define ideas and terminology needed for subsequent chapters.

Consider an $a \times b$, fixed-effects, no-empty-cells design with a total of n observations arranged as an $n \times 1$ vector in lexicographic order, first by rows and then by columns. This data vector, Y, is contained in n-dimensional Euclidean space $\mathbb{R}^n$. The mean vector, $\mu = E(Y)$, will also lie in $\mathbb{R}^n$. Proposing a linear model for Y with respect to the $a \times b$ design corresponds to identifying a subspace, say $\mathcal{M}$, where

μ is assumed to reside. The method of least squares for fitting the linear model is the process of finding the vector in $\mathcal{M}$ which is closest to Y. This vector is the perpendicular projection of Y onto $\mathcal{M}$, and is denoted by $P(Y:\mathcal{M})$, see Figure 1. The space $\mathcal{M}$ is called the "estimation space" since estimates of μ suggested by the model are in $\mathcal{M}$.

Let X be a full rank $n \times s$ matrix and let the span of the columns of X be denoted by $[X]$. If we assume that $\mu = X\beta$ for some unknown but fixed vector of parameters β in $\mathbb{R}^s$, then $[X]$ is an example of a subspace $\mathcal{M}$. Consequently $\mathbb{R}^s$ is called the "parameter space." The vector Y is thus modeled by $\mu = E(Y)$ plus some random error vector, e, and written as $Y = X\beta + e$. If β is chosen to be the vector of expected cell means in a two-way design, then X is the incidence

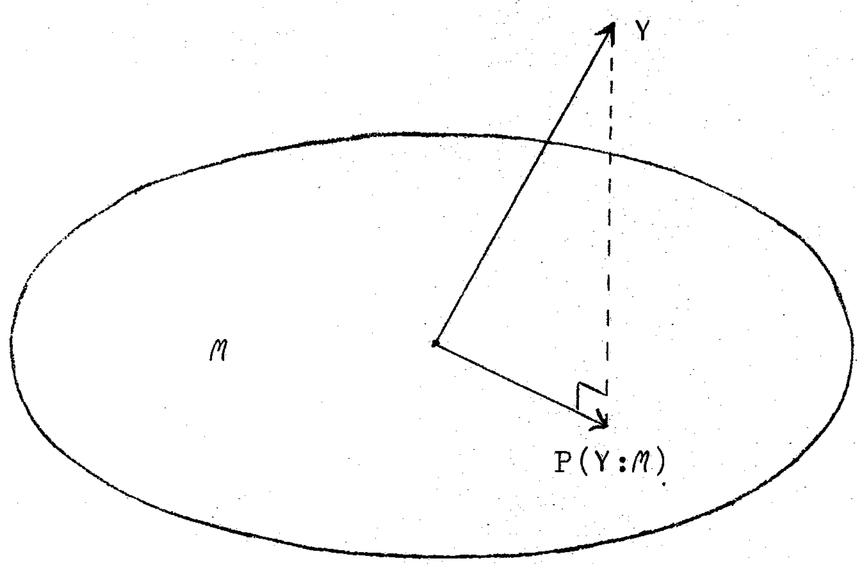


Figure 1: The perpendicular projection of Y onto the vector space $\mathcal{M}$.

matrix (with one (1) in each row corresponding to the appropriate element of $\boldsymbol{\beta}$ and zeros elsewhere) and $s = ab$. In what follows we shall only be concerned with the case in which $\boldsymbol{\beta}$ is the $ab \times 1$ vector of cell means.

The parameter space $\mathbb{R}^{ab}$ can be written as the direct sum of orthogonal subspaces J, A|J, B|J, and AB which are defined below.

J: the subspace spanned by the $ab \times 1$ vector of ones, $\mathbf{1}_{ab}$ (we will also use J for the analogous subspace of $\mathbb{R}^n$ and depend on the context to indicate which J is intended);

A: the subspace corresponding to the row factor and equal to the span of all vectors that could be the true parameter vector if there were no column and no interaction effects;

A|J: the orthogonal complement of J in A+J, or just A since J is a subspace of A;

B: the subspace corresponding to the column factor and equal to the span of all vectors that could be the true parameter vector if there were no row and no interaction effects;

B|J: the orthogonal complement of J in B;

AB: the orthogonal complement of A+B in $\mathbb{R}^{ab}$ and equal to the span of all vectors that could be the difference between the true parameter vector and a parameter vector corresponding to a no interaction model.

Using the orthogonal decomposition of the parameter space $\mathbb{R}^{ab} = J \oplus A|J \oplus B|J \oplus AB$, where $\oplus$ denotes the direct sum of subspaces, we may write the mean vector as $\beta = \beta_J + \beta_{A|J} + \beta_{B|J} + \beta_{AB}$ where for any subspace W of $\mathbb{R}^{ab}$, $\beta_W = P(\beta:W)$. An additive model is one such that $\beta_{AB} = 0$ or $\beta \varepsilon A+B$. The existence of $\beta_{AB} \neq 0$ thus indicates a nonadditive or interactive model.

To make the connection with the usual over-parameterized parametric representation of the linear model, $E(y_{ijk}) = \mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}$, note that $E(y_{ijk})$ is the (i,j)th cell mean which is a coordinate of β and μ, α_i, β_j, $(\alpha\beta)_{ij}$ are the corresponding coordinates of β_J, $\beta_{A|J}$, $\beta_{B|J}$, β_{AB} with the constraints $\sum_i(\alpha\beta)_{ij} = \sum_j(\alpha\beta)_{ij} = \sum_i \alpha_i = \sum_j \beta_j = 0$.

Three hypotheses of interest are:

H_I: no interaction;

H_R: no difference in main row effects; and

H_C: no difference in main column effects.

Although different interpretations are possible, these may be expressed parametrically as:

H_I: $(\alpha\beta)_{ij} = 0$ for $i=1,\cdots,a$ and $j=1,\cdots,b$;

H_R: $\alpha_i = \alpha_{i+1}$ for $i=1,\cdots,a-1$; and

H_C: $\beta_j = \beta_{j+1}$ for $j=1,\cdots,b-1$.

Geometrically we express these hypotheses as:

H_I: $\beta_{AB} = 0$;

H_R: $\beta_A \varepsilon J$ or $\beta_{A|J} = 0$; and

H_C: $\beta_B \varepsilon J$ or $\beta_{B|J} = 0$.

The geometric representation leads us to a natural method for testing the hypotheses. Suppose U is some subspace of $\mathbb{R}^{ab}$ and we hypothesize H_U: $\beta_U = 0$. This means that the squared length of β_U, $\|\beta_U\|^2$, is zero under H_U. Using an estimate of β, say $\hat{\beta}$, large values of $S^* = \|P(\hat{\beta}:U)\|^2$ may be regarded as evidence against H_U. To quantify this evidence however we must know the null distribution of S^*.

Assume $\hat{\beta} \sim N(\beta,\Sigma)$ and let U be a full rank matrix such that $[\mathsf{U}] = U$. Then $\lambda^{-1}S^* \sim \chi^2_{(\dim U)}$ for some positive constant λ if and only if $\lambda^{-1}\mathsf{U}(\mathsf{U}'\mathsf{U})^{-1}\mathsf{U}'\Sigma$ is idempotent (Rao,1968). It is easily shown that this is equivalent to requiring $\mathsf{U}'\Sigma\mathsf{U} = \lambda\mathsf{U}'\mathsf{U}$. Consider also $S_2^* = \|P(\hat{\beta}:V)\|^2$ for some V in $\mathbb{R}^{ab}$. Let V be a full rank matrix such that $[\mathsf{V}] = V$. Then S^* and S_2^* are independent if and only if $\mathsf{U}'\Sigma\mathsf{V} = 0$ (see Chapter 4, Corollary 4.4). If U and V are orthogonal subpaces, then the requirement is

(3.1) $\mathsf{U}'\Sigma\mathsf{V} = \lambda\mathsf{U}'\mathsf{V}$, which is 0 since $\mathsf{U}'\mathsf{V} = 0$.

Let $Y \sim N(X\beta,\sigma^2 I)$ and $\hat{\beta} = (X'X)^{-1}X'Y$, the best linear unbiased estimate of the cell means vector β. Then $\Sigma = \sigma^2(X'X)^{-1}$. Let $\eta' = \sigma^2(n_{11}^{-1}\ n_{12}^{-1} \cdots\ n_{1b}^{-1} \cdots n_{ab}^{-1})$, the diagonal of Σ. Table 3.1 shows for various vector spaces U and V orthogonality conditions on η for 3.1 to hold (see Appendix A for a proof).

From Table 3.1 we see that if one requires that the S^* statistics for H_{R1} and H_{C1} be independent chi-squared random variables, then $\eta \perp (A|J \oplus B|J \oplus AB)$ or $\eta \varepsilon J$ for $a>2$ and $b>2$. If one only requires independence of all three tests, then the same conditions prevail. Finally if one

Table 3.1. Orthogonality Conditions on $\eta = \text{diag}(\sigma^2(X'X)^{-1})$ such that $U'(\sigma^2(X'X)^{-1})V = \lambda U'V$.†

U	V	Orthogonality Conditions	Comments
$A\|J$	$A\|J$	for $a>2; \eta \perp A\|J$ †† for $a=2$; no restrictions	$\lambda^{-1}S^* \sim \chi_{(a-1)}$ *
$B\|J$	$B\|J$	for $b>2; \eta \perp B\|J$ for $b=2$; no restrictions	$\lambda^{-1}S^* \sim \chi_{(b-1)}$
AB	AB	for $a>2, b>2; \eta \perp \mathbb{R}^{ab}\|J$ for $a>2, b=2; \eta \perp A\|J$ for $a=2, b>2; \eta \perp B\|J$ for $a=2, b=2$; no restrictions	$\lambda^{-1}S^* \sim \chi_{(a-1)(b-1)}$
$A\|J$	$B\|J$	$\eta \perp AB$	$\lambda^{-1}S^*, \lambda^{-1}S_2^*$ independent **
$A\|J$	AB	for $a>2; \eta \perp (B\|J \oplus AB)$ for $a=2; \eta \perp B\|J$	$\lambda^{-1}S^*, \lambda^{-1}S_2^*$ independent
$B\|J$	AB	for $b>2; \eta \perp (A\|J \oplus AB)$ for $b=2; \perp A\|J$	$\lambda^{-1}S^*, \lambda^{-1}S_2^*$ independent

† $\lambda = \sigma^2 \sum_i \sum_j n_{ij}^{-1}/ab$, the mean eigenvalue of $\sigma^2(X'X)^{-1}$

†† η "is orthogonal to" $A|J$

* $S^* = ||P(\hat{\beta}:U)||^2$

** $S_2^* = ||P(\hat{\beta}:V)||^2$

simply requires that the test of H_I be a chi-squared test, then $\eta \perp \mathbb{R}^{ab}|J$ or $\eta \varepsilon J$ for $a>2$ and $b>2$. If $\eta \varepsilon J$, then $n_{ij} = n/ab$, that is the design is balanced. Thus departure from equal cell sizes may lead to undesirable properties of the S* statistics.

Some further notes on S* are:

i. λ is unobservable unless σ^2 is known. When σ^2 is unknown the residual sum of squares for regression

$||P(Y: \mathbb{R}^n | [X])||^2$ divided by n-dim([X]) will provide an unbiased independent estimate of σ^2, say $\hat{\sigma}^2$, if $\mu = X\beta$. Hence $F^* = S^*\{ab/(\dim(U)\hat{\sigma}^2 \sum_i\sum_j n_{ij}^{-1})\} \sim \mathcal{F}_{(\dim(U), n-\dim([X]))}$ when $\beta_U = 0$ and $\sigma^2 U'(X'X)^{-1}U = (\sigma^2 \sum_i\sum_j n_{ij}^{-1}/ab)U'U$.

ii. The statistic F* above was originally proposed by Yates (1934) and the resulting analysis is usually referred to as the unweighted means analysis. The conditions listed in Table 3.1 are the conditions for which Yates' test is an exact F test.

iii. As will be shown next, $\lambda^{-1}S^*$ (or F* for unknown σ^2) is the likelihood ratio statistic when the conditions of Table 3.1 hold.

The model $Y = X\beta + e$ may be rewritten as $Y = X\beta_U + X\beta_V + e$ where $U \oplus V = \mathbb{R}^{ab}$ and $U \perp V$. Thus $\mu = X\beta_U + X\beta_V$ as shown in Figure 2. If $\beta_U = 0$, then μ lies in $\hat{V}$, where for any W in $\mathbb{R}^{ab}$, $\hat{W} = (\{Xw: w \varepsilon W\})$, as shown in Figure 3. It is readily seen that $P(\mu: \mathbb{R}^n | \hat{V}) = \mu - P(\mu:\hat{V}) = 0$ under H_U: $\beta_U = 0$. When $Y \sim N(\mu, \sigma^2 I)$ and $\mu = X\beta$, the statistic for the likelihood ratio test of H_U is given by:

$$(3.2) \quad ||P(\hat{\mu}: \mathbb{R}^n | \hat{V})||^2 / \sigma^2$$

for known σ^2 where $\hat{\mu} = P(Y:[X])$. Since $\hat{\mu}\ \varepsilon\ [X]$, then 3.2 may be written as:

$$(3.3) \quad ||P(Y:[X] | \hat{V})||^2 / \sigma^2 .$$

The idea here is to view large values of 3.3 as evidence

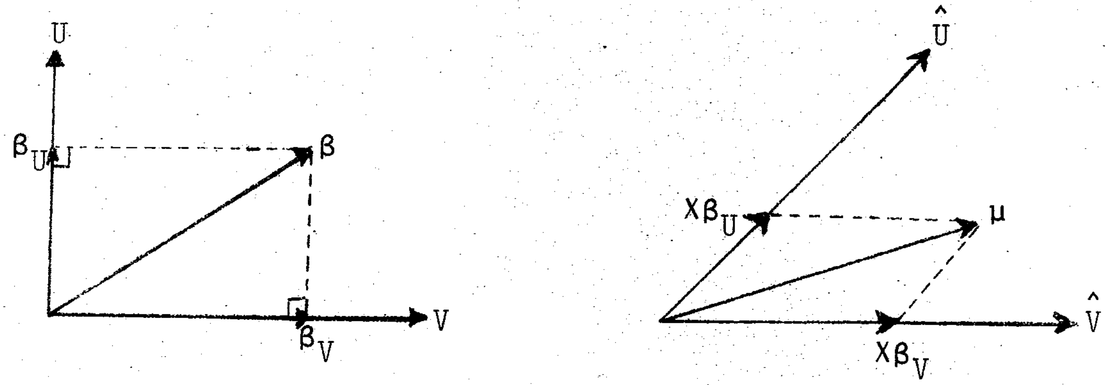


Figure 2: β as sum of components in U and V and μ as sum of components in $\hat{U}$ and $\hat{V}$ where for any W in $\mathbb{R}^{ab}$, $\hat{W} = (\{Xw : w \varepsilon W\})$.

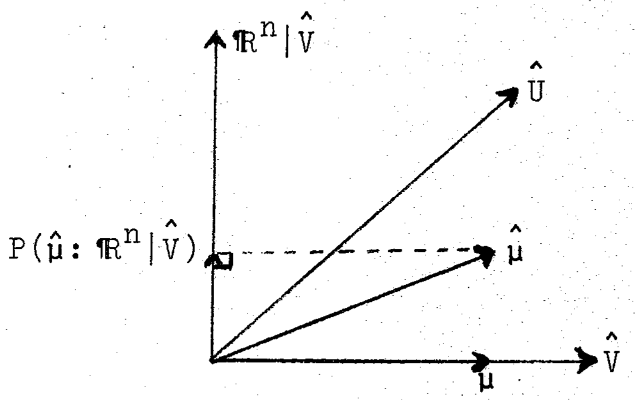


Figure 3. μ as an element of $\hat{V}$ under $H_U:\beta_U=0$ and the perpendicular projection of $\hat{\mu}$ onto the orthogonal complement of $\hat{V}$ in $\mathbb{R}^n$.

---

against H_U (see Figure 3).

It is easily shown that since $U \perp V$, $[X]|\hat{V} = \tilde{U}$ where for any W in $\mathbb{R}^{ab}$, $\tilde{W} = (\{X(X'X)^{-1}w : w \varepsilon W\})$. For W in $\mathbb{R}^{ab}$ the two spaces $\hat{W}$ and $\tilde{W}$ are arrived at via two different transformations from the parameter space to the estimation space. Each transformation, X and $X(X'X)^{-1}$, may be thought of as a bridge from $\mathbb{R}^{ab}$ to $[X]$. The maximum likelihood estimate of $X\beta_W$ is the perpendicular projection $P(Y:\hat{W})$. If $Y \sim N(\mu,\sigma^2 I)$ and $\mu = X\beta$, then the likelihood ratios statistics for H_I, H_{R1}, and H_{C1} are:

H_I: $\sigma^2 S_I = \|P(Y:A\tilde{\uparrow}B)\|^2$;

H_{R1}: $\sigma^2 S_{R1} = \|P(Y:A\tilde{|}J)\|^2$; and

H_{C1}: $\sigma^2 S_{C1} = \|P(Y:B\tilde{|}J)\|^2$.

Furthermore:

$S_I \sim \chi_{(a-1)(b-1)}$;

$S_{R1} \sim \chi_{(a-1)}$; and

$S_{C1} \sim \chi_{(b-1)}$, since

$X(X'X)^{-1}W(W'(X'X)^{-1}W)^{-1}W'(X'X)^{-1}X'(\sigma^2 I)/\sigma^2$ is idempotent. Note that $\|P(Y:\tilde{W})\|^2/\sigma^2 = \hat{\beta}'W(W'(X'X)^{-1}W)^{-1}W'\hat{\beta}/\sigma^2$ which is equal to $\lambda^{-1}\hat{\beta}'W(W'W)^{-1}W'\hat{\beta}$ if $W'(X'X)^{-1}W = \lambda W'W$. Thus if the conditions of Table 3.1 hold, the statistics S_I, S_{R1}, and S_{C1} are the S* statistics discussed earlier.

When considering the three hypotheses H_I, H_R, and H_C, it is common practice to first test H_I and if insufficient evidence is found against it, assume the additive model is correct and test H_R and/or H_C. When the additive model <u>is</u> correct, $\beta = \beta_J + \beta_{A|J} + \beta_{B|J}$. The parameter space is thus reduced to $A|J \oplus B$ and the estimation space is reduced to $\mathcal{M}_a = A\hat{|}J + \hat{B}$. The likelihood ratio statistic for testing H_{R1}: $\beta_{A|J} = 0$ is $\|P(Y:\mathcal{M}_a|\hat{B})\|^2$. Since $\mathcal{M}_a = \hat{A} + \hat{B}$, then $\mathcal{M}_a|\hat{B} = A\hat{\uparrow}B$. The likelihood ratio statistics under the additive model are:

H_I: $\sigma^2 S_I = ||P(Y:\widetilde{AB})||^2$;

H_{R1}: $\sigma^2 S_{R2} = ||P(Y:\hat{A}|\hat{B})||^2$; and

H_{C1}: $\sigma^2 S_{C2} = ||P(Y:\hat{B}|\hat{A})||^2$

when $Y \sim N(\mu, \sigma^2 I)$ and $\mu = X\beta$. Furthermore:

$S_I \sim \chi_{(a-1)(b-1)}$;

$S_{R2} \sim \chi_{(a-1)}$; and

$S_{C2} \sim \chi_{(b-1)}$.

The subscript 2 on S_{R2} and S_{C2} is there because in general these sums of squares are not the same as the earlier sums of squares derived without the additivity assumption. Therefore the two sets of test statistics provide two different analyses. The first (S_I, S_{R1}, S_{C1}) is the STP analysis mentioned in the previous chapters and the second (S_I, S_{R2}, S_{C2}) is the EAD analysis. We now explore the geometric properties of these analyses.

The properties of STP and EAD are summarized in Table 3.2. To demonstrate these properties first consider vector spaces in $\mathbb{R}^n$ which are the results of mappings from $\mathbb{R}^{ab}$ either by X (the best-fit-bridge) or $X(X'X)^{-1}$ (the best-estimate-bridge) (Burdick, *et al.*, 1974). Neither transformation preserves orthogonality under general conditions, but spaces orthogonal in $\mathbb{R}^{ab}$ mapped to $\mathbb{R}^n$ via different bridges will still remain orthogonal. That is if $U \perp V$, then $\tilde{U} \perp \hat{V}$. With this in mind we readily see that certain subspaces are always orthogonal (see Table 3.3) while others may be orthogonal under certain conditions on the cell sizes.

Table 3.2: Geometric Properties of the STP and EAD Analyses

Hypothesis		χ^2 Noncentrality Parameter
STP:	H_I	$\|\|P(\mu:\tilde{A}B)\|\|^2 = \|\|P(X\beta_{AB}:\tilde{A}B)\|\|^2$
		$= 0$ if and only if H_I is true
	H_{R1}	$\|\|P(\mu:\tilde{A}\|J)\|\|^2 = \|\|P(X\beta_{A\|J}:\tilde{A}\|J)\|\|^2$
		$= 0$ if and only if H_{R1} is true
	H_{C1}	$\|\|P(\mu:\tilde{B}\|J)\|\|^2 = \|\|P(X\beta_{B\|J}:B\|J)\|\|^2$
		$= 0$ if and only if H_{C1} is true
EAD:	H_{R1}	$\|\|P(\mu:\hat{A}\|\hat{B})\|\|^2 = \|\|P(X\beta_{A\|J}+X\beta_{AB}:\hat{A}\|\hat{B})\|\|^2$
		$= \|\|P(X\beta_{AB}:\hat{A}\|\hat{B})\|\|^2$ if H_{R1} is true
		$= 0$ if H_{R1} and H_I are true
	H_{C1}	$\|\|P(\mu:\hat{B}\|\hat{A})\|\|^2 = \|\|P(X\beta_{B\|J}+X\beta_{AB}:\hat{B}\|\hat{A})\|\|^2$
		$= \|\|P(X\beta_{AB}:\hat{B}\|\hat{A})\|\|^2$ if H_{C1} is true
		$= 0$ if H_{C1} and H_I are true

Additive Model Properties (H_I true):

$$||P(\mu:\tilde{A}|J)||^2 \leq ||P(\mu:\hat{A}|\hat{B})||^2$$

$$||P(\mu:\tilde{B}|J)||^2 \leq ||P(\mu:\hat{B}|\hat{A})||^2$$

Table 3.3: Orthogonal Subspaces of [X]

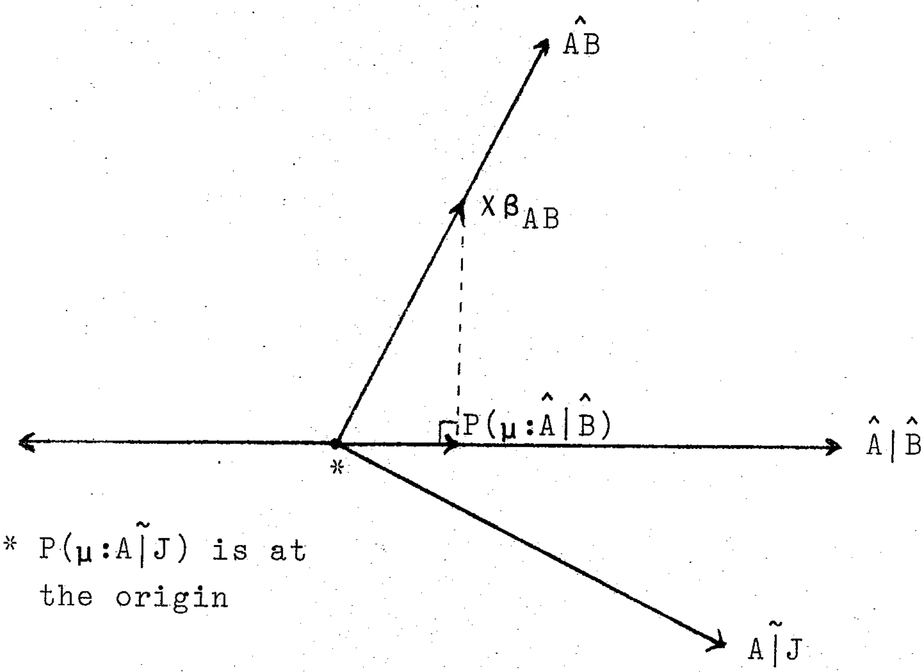
	J	$\hat{A}$	$\hat{B}$	$\widehat{AB}$	$\widetilde{AB}$
J		0*	0	0	0
$\tilde{A}\|J$	0		0	0	
$\tilde{B}\|J$	0	0		0	
$\widetilde{AB}$	0	0	0		
$\hat{A}\|\hat{B}$	0		0		0
$\hat{B}\|\hat{A}$	0	0			0

* 0 denotes orthogonality

The STP analysis has the property that the effect being tested for will not be affected by the presence or absence of the other effects. For example if there is interaction ($\beta_{AB} \neq 0$) then it will show up in $\widehat{AB}$ of the estimation space. We see from Table 3.3 however that $\widehat{AB}$ is orthogonal to both $\tilde{A}|J$ and $\tilde{B}|J$ which is where we would test for row and column effects. Thus the presence of interaction will not affect the component of μ in $\tilde{A}|J$ or $\tilde{B}|J$. The result of this property is that under the null hypothesis H_{R1} or H_{C1}, the χ^2 distribution will be central regardless of the validity of the other hypotheses.

The EAD analysis does not share the above property under general conditions. For example, the presence of interaction may show up in $\hat{A}|\hat{B}$ and/or $\hat{B}|\hat{A}$ because these spaces are not orthogonal to $\widehat{AB}$. The χ^2 distribution under the null hypothesis for H_{R1} and/or H_{C1} may be noncentral with unknown noncentrality parameter. To illustrate this problem we consider two cases concerning H_{R1}.

Case 1: H_{R1} true implies that $X\beta_{A|J} = 0$. Now under H_{R1} $\mu = X\beta_{A|J} + X\beta_B + X\beta_{AB} = X\beta_B + X\beta_{AB}$. Since $X\beta_B \epsilon \hat{B} \perp \hat{A}|\hat{B}$, the noncentrality parameter for S_{R2} is $||P(\mu:\hat{A}|\hat{B})||^2 = ||P(X\beta_{AB}:\hat{A}|\hat{B})||^2$ which is zero when $X\beta_{AB} = 0$ or $\hat{AB} \perp \hat{A}|\hat{B}$ (see Figure 4). Thus one may be more likely to reject H_{R1} when H_{R1} is true using S_{R2} than S_{R1} in the presence of interaction.



* $P(\mu:\tilde{A|J})$ is at the origin

Figure 4: Component of μ in $\hat{A}|\hat{B}$ in the presence of interaction when H_{R1} is true.

<u>Case 2</u>: Suppose $P(\mu:\hat{A}|\hat{B}) = 0$. Then $P(\mu:\hat{A}|\hat{B}) = P(X\beta_{A|J}:\hat{A}|\hat{B}) + P(X\beta_{AB}:\hat{A}|\hat{B}) = 0$ or $P(X\beta_{A|J}:\hat{A}|\hat{B}) = -P(X\beta_{AB}:\hat{A}|\hat{B})$. This suggests that in the presence of interaction, $X\beta_{A|J}$ may not be the zero vector, i.e., H_{R1} may be false (see Figure 5). Thus one may be less likely to reject H_{R1} when H_{R1} is false using S_{R2} than S_{R1} in the presence of interaction.

In the above discussion a key phrase was "in the presence of interaction". How do the STP and EAD analyses compare when the additive model is the correct model? Suppose H_{R1} is true. Then $\mu = X\beta_B$ and $P(\mu:\hat{A}|\hat{B}) = 0$ since $\hat{B} \perp \hat{A}|\hat{B}$. The noncentrality parameters for S_{R1} and S_{R2} are both zero.

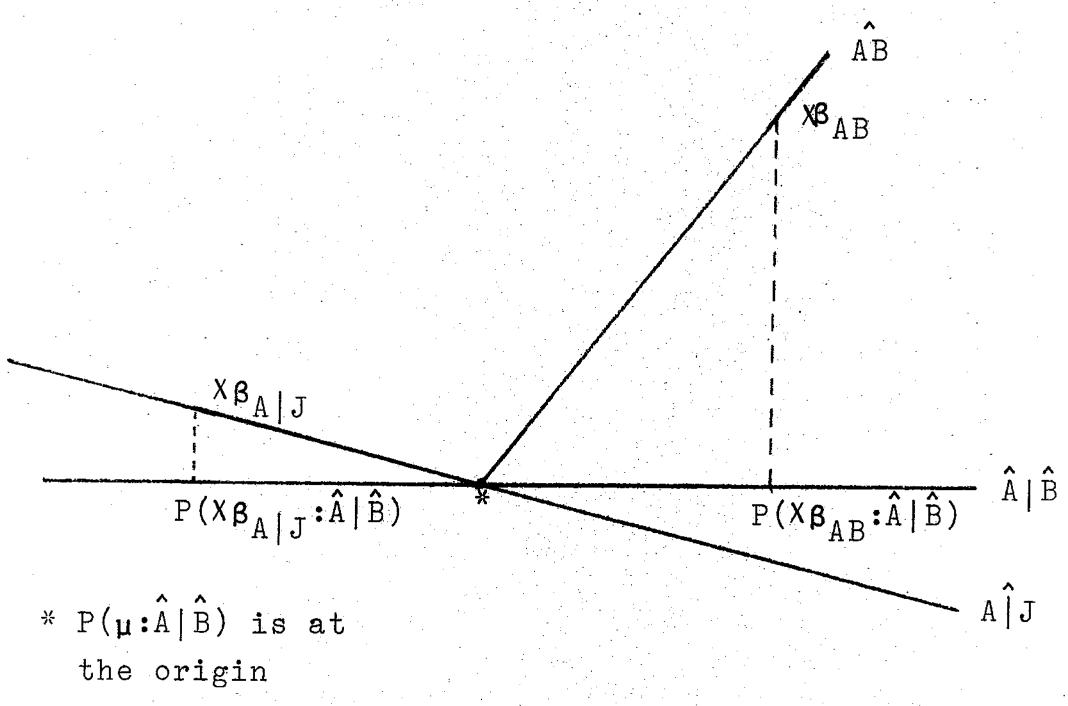


Figure 5: Component of μ in $\hat{A}|\hat{B}$ being zero in the presence of interaction when H_{R1} is false.

Suppose H_{R1} is false. Since $\mu \varepsilon \hat{A} + \hat{B}$ and $\hat{A} + \hat{B} = \hat{A}|\hat{B} \oplus \hat{B}$, then $\mu = P(\mu:\hat{A}|\hat{B}) + P(\mu:\hat{B})$. The noncentrality parameter for S_{R1} is thus $\|P(\mu:A\tilde{|}J)\|^2 = \|P(P(\mu:\hat{A}|\hat{B}):A\tilde{|}J)\|^2$ since $P(\mu:\hat{B}) \varepsilon \hat{B} \perp A\tilde{|}J$. It is easily shown that $\|P(\mu:A\tilde{|}J)\|^2 \leq \|P(\mu:\hat{A}|\hat{B})\|^2$ since $P(\mu:A\tilde{|}J)$ and $P(\mu:\hat{A}|\hat{B})$ represent respectively the leg and the hypotenuse of a right triangle (see Figure 6). Thus under the additive model the probability of rejecting H_{R1} when H_{R1} is false may be greater using S_{R2} then S_{R1}. That is, the EAD analysis provides at least as powerful a test as the STP analysis when the additive model is correct.

A desirable property of any analysis of the main effects is that the sums of squares be independent. This means that the spaces on which Y is projected to obtain the sums of squares must be orthogonal. For the EAD anlysis independence

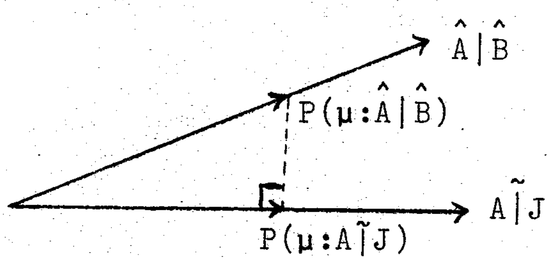


Figure 6: Components of μ in $\hat{A}|\hat{B}$ and $A\tilde{|}J$ (corresponding to the degree of noncentrality in S_{R2} and S_{R1}) when the model is additive.

of S_{R2} and S_{C2} requires $\hat{A}|\hat{B} \perp \hat{B}|\hat{A}$. The well known condition for orthogonality is that the cell sizes are such that $n_{ij} = r_i s_j$ for positive constants r_i $(i=1,\cdots,a)$ and s_j $(j=1,\cdots,b)$. This is the condition of proportionality. An important note to make here is that proportionality does not mean that $\hat{A}|\hat{B}$ or $\hat{B}|\hat{A}$ will be orthogonal to $\hat{AB}$. So S_{R2} and S_{C2} may still be affected by the presence of interaction.

For the STP analysis the condition for $A\tilde{|}J \perp B\tilde{|}J$ is given in Table 3.1 as the condition for $U'(X'X)^{-1}V/\sigma^2 = \lambda U'V = 0$ when $U = A|J$ and $V = B|J$. The requirement is that the vector of reciprocal cell sizes, η, be orthogonal to AB. This may also be written as $n_{ij}^{-1} = r_i + s_j$ for real constants r_i $(i=1,\cdots,a)$ and s_j $(j=1,\cdots,b)$ (Burdick, *et al*, 1974). Note here that this condition does not leave $A\tilde{|}J$ or $\tilde{B}|J$ orthogonal to $\tilde{AB}$ (see Table 3.1 for conditions on η to achieve this orthogonality). Thus S_{R1} and/or S_{C1} may not be independent of S_I, but recall that $A\tilde{|}J$ and $B\tilde{|}J$ are always orthogonal to $\hat{AB}$, the location of any interaction that might be present.

In this chapter we have investigated the geometric aspects of the ANOVA and in particular the STP and EAD analyses. In Chapter 5 we will quantify some of the differences between STP and EAD using the properties mentioned here and the distribution theory developed in Chapter 4.

Chapter 4

Distribution Theory

In this chapter we discuss the distribution theory relevant to testing multiple hypotheses in the fixed-effects, all-cells-filled ANOVA setting. We begin with the chi-squared (χ^2) distribution and the $\mathcal{F}$ distribution. We conclude with a discussion of computational methods for evaluating the cumulative density functions (cdf's) of the mentioned distributions.

4.1 χ^2 Distribution

Definition 4.1 (Incomplete Gamma Function)

Let $\Gamma(\cdot)$ denote the gamma function defined by

$$\Gamma(a) = \int_0^\infty t^{a-1}\exp(-t)dt;\ a>0.$$

The function $\Gamma(\cdot,\cdot)$ defined by

$$\Gamma(a,x) = \{\Gamma(a)\}^{-1}\int_0^\infty t^{a-1}\exp(-t)dt;\ a>0,\ x\geq 0$$

is called the incomplete gamma function.

Clearly $\Gamma(\cdot,\cdot)$ is a cdf, thus $\Gamma(a,0) = 0$ and $\lim_{x\to\infty}\Gamma(a,x) = 1$.

Theorem 4.1 (Noncentral Incomplete Gamma Function)

The series

$$G(x;\nu,\psi) = \begin{cases} 0 & \text{for } x<0, \\ \sum_{k=0}^{\infty} \exp(-\psi/2)(\psi/2)^k\Gamma(\nu/2+k,x/2)/k! & \text{for } x\geq 0 \end{cases}$$

is a cdf in x for $\nu>0$ and $\psi\geq 0$.

Proof

The theorem is a direct result of the properties of the incomplete gamma function (Searle, 1971, p.49).

One family of distributions represented by the non-central incomplete gamma function is the family of χ^2 distributions.

Definition 4.2 (χ^2 Distribution)

A random variable X with cdf $G(x;\nu,\psi)$, $(\nu=1,2,\cdots;\psi\geq 0)$, is said to have a χ^2 distribution, denoted by $\chi^2_{\nu,\psi}$. The parameter ν is the degrees of freedom and ψ is the noncentrality parameter. When $\psi=0$, the distribution of X, denoted by χ^2_ν, is called a central χ^2 and its cdf is given by:

$$G(x;\nu) = \begin{cases} 0 & \text{for } x<0 \\ \Gamma(\nu/2,x/2) & \text{for } x\geq 0 \end{cases} .$$

For the remainder of this chapter we assume that the $n\times 1$ vector Y is distributed as a multivariate normal with mean vector θ and positive definite, dispersion matrix V (that is, $Y\sim N(\theta,V)$).

The following theorems are useful for discussing properties of the χ^2 distribution.

Theorem 4.2

$Y'AY \sim \chi^2_{\nu,\psi}$ where $\nu=\text{rank}(A)$ and $\psi=\theta'A\theta$ if and only if AV is idempotent.

Proof

See Searle (1971, p.57).

Theorem 4.3

$$E(Y'AY) = tr(AV) + \theta' A\theta.$$

Proof

Searle (1971, p. 55). The function tr(A) denotes the trace of A.

Theorem 4.4

$$Var(Y'AY) = 2tr(AVAV) + 4\theta' AVA\theta.$$

Proof

Searle (1971, p.57).

Corollary 4.4

$$Cov(Y'AY,Y'BY) = 2tr(AVBV) + 4\theta' AVB\theta.$$

Proof

First note that $2Cov(Y'AY,Y'BY) = Var(Y'(A+B)Y) - Var(Y'AY) - Var(Y'BY)$.

From Theorem 4.4 we have that:

$$\begin{aligned} Var(Y'(A+B)Y) &= 2tr((A+B)V(A+B)V) + 4\theta'(A+B)V(A+B)\theta \\ &= 2tr(AVAV) + 4\theta' AVA\theta + 2tr(BVBV) + \\ &\quad 4\theta' BVB\theta + 2(2tr(AVBV) + 2\theta' AVB\theta + \\ &\quad 2\theta' BVA\theta) \\ &= Var(Y'AY) + Var(Y'BY) + 2(2tr(AVBV) + \\ &\quad 4\theta' AVB\theta). \end{aligned}$$

Hence $Cov(Y'AY,Y'BY) = 2tr(AVBV) + 4\theta' AVB\theta$.

For the remainder of this chapter let U and W be n×n symmetric matrices with the following properties:

i. $Ux = P(x:[U])$ and $Wx = P(x:[W])$, that is U and W are projection matrices and therefore idempotent.

ii. $rank(U) = p$ and $rank(W) = q$ where, without loss of generality, $p \leq q$; and

iii. $tr(UW) = \sum_{i=1}^{p} \lambda_i$ where $\{\lambda_i\}_1^p$ are the p largest eigenvalues of UW (the other q-p eigenvalues will all be zero). Note that as a property of idempotent matrices, $\lambda \geq 0$ $(i=1,\cdots,p)$.

Also for the remainder of this chapter, let $S_1 = Y'UY/\sigma^2$, $S_2 = Y'WY/\sigma^2$, and $V = \sigma^2 I_n$.

Theorem 4.5

i. $S_1 \sim \chi^2_{p,\psi_1}$ and $S_2 \sim \chi^2_{q,\psi_2}$ where $\psi_1 = ||P(EY:[U])||^2/\sigma^2$

$= \theta'U\theta/\sigma^2$ and $\psi_2 = ||P(EY:[W])||^2/\sigma^2 = \theta'W\theta/\sigma^2$;

ii. $Corr(S_1,S_2) = \{\sum_1^p \lambda_i + 2\theta'UW\theta/\sigma^2\}\{(p + 2\psi_1)(q + 2\psi_2)\}^{-\frac{1}{2}}$; and

iii. S_1 and S_2 are independent if and only if U and W are orthogonal.

Proof

i. Since U and W are idempotent, the result follows from Theorem 4.2.

ii. By Corollary 4.4,

$$\begin{aligned} Cov(S_1,S_2) &= 2tr(\ U(\sigma^2 I_n)W(\sigma^2 I_n))/\sigma^4 + \\ &\quad 4\theta'U(\sigma^2 I_n)W\theta/\sigma^4 \\ &= 2tr(UW) + 4\theta'UW\theta/\sigma^4 \\ &= 2(\sum_1^p \lambda_i + 2\theta'UW\theta/\sigma^4). \end{aligned}$$

By Theorem 4.4,

$$Var(S_1) = 2tr(U(\sigma^2 I_n)U(\sigma^2 I_n))/\sigma^4 + 4\theta'U(\sigma^2 I_n)U\theta/\sigma^4$$

$= 2(p + 2\psi_1).$

Similarly $Var(S_2) = 2(q + 2\psi_2)$. Hence

$$Corr(S_1,S_2) = Cov(S_1,S_2)\{Var(S_1)Var(S_2)\}^{-\frac{1}{2}}$$

$$= \{\sum_1^p \lambda_i + 2\theta' UW\theta/\sigma^2\}\{(p + 2\psi_1)(q + 2\psi_2)\}^{-\frac{1}{2}}.$$

iii. Searle (1971, p.59).

The characteristic function (cf) is defined by

$\phi(t) = \int_{-\infty}^{\infty}\exp(itx)dF(x)$ where $F(x)$ is a cdf. This is useful in deriving distributions by means of the inverse Fourier exponential transform $\int_{-\infty}^{\infty}\exp(-itx)\phi(t)dt$. The cf of a $\chi^2_{\nu,\psi}$ random variable is given by Oberhettinger (1973):

$$(4.1)\quad \phi_{\chi^2_{\nu,\psi}}(t) = (1-2it)^{-\nu/2}\exp(it\psi/(1-2it)).$$

A bivariate cf is defined by

$$\phi(t_1,t_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(i(t_1x_1 + t_2x_2))dF(x_1,x_2)$$

where $F(x_1,x_2)$ is a bivariate cdf. The bivarite cf of S_1 and S_2 is given in Theorem 4.6.

Theorem 4.6 (Characteristic Function of S_1 and S_2)

$$(4.2)\quad \phi_{S_1,S_2}(t_1,t_2) = (1-2it_1)^{-p/2}(1-2it_2)^{-q/2} \times \prod_{j=1}^{p}(1-c\lambda_j)^{-\frac{1}{2}} \times \exp(c_1\sum_{j=1}^{p}\tau_{1j}^2 + c_2\sum_{j=1}^{p}\tau_{2j}^2 + c\sum_{j=1}^{p}\sqrt{\lambda_j}\tau_{1j}\tau_{2j}),$$

where $\tau_{kj} = \theta^*_{kj}(1-c\lambda_j)^{-\frac{1}{2}}$;

θ^*_{1j} is the jth element of $\theta^*_1 = U_1'\theta$;

θ^*_{2j} is the jth element of $\theta^*_2 = W_1'\theta$;

U_1 and W_1 are $n\times p$ and $n\times q$ full column-rank matrices such that $U_1U_1' = U$, $U_1'U_1 = I_p$, $W_1W_1' = W$, $W_1'W_1 = I_q$, and $U_1'W_1 = C = (\mathrm{diag}(\sqrt{\lambda_1},\cdots,\sqrt{\lambda_p}) \vdots 0)$;

$\lambda_j = 0$ for $j>p$;

$c_k = it_k/(1-2it_k)$; and

$c = 4c_1c_2$.

<u>Proof</u>

Let $Z = \begin{pmatrix} U_1' \\ W_1' \end{pmatrix} Y/\sigma$ so that $Z \sim N(\theta^*,V^*)$ where

$$\theta^* = \begin{pmatrix} U_1' \\ W_1' \end{pmatrix}\theta/\sigma = (\theta^*_{11},\cdots,\theta^*_{1p},\theta^*_{21},\cdots,\theta^*_{2q})' \text{ and}$$

$$V^* = \begin{pmatrix} U_1'U_1 & U_1'W_1 \\ W_1'U_1 & W_1'W_1 \end{pmatrix} = \begin{pmatrix} I_p & C \\ C' & I_q \end{pmatrix} .$$

Note that $S_1 = Z'\begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix}Z$ and $S_2 = Z'\begin{pmatrix} 0 & 0 \\ 0 & I_q \end{pmatrix}Z$.

By the definition of ϕ_{S_1,S_2} ,

$$(4.3) \quad \phi_{S_1,S_2}(t_1,t_2) = E(\exp(it_1S_1 + it_2S_2)) = E(\exp(iZ'TZ))$$

where $T = \begin{pmatrix} t_1I_p & 0 \\ 0 & t_2I_q \end{pmatrix}$. Generalizing equation 2.10, 4.3 can be written as:

$$(4.4) \quad \phi_{S_1,S_2}(t_1,t_2) = |M|^{-\frac{1}{2}}\exp(i\theta^{*\prime}M^{-1}T\theta^*) \quad ,$$

where $M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = \begin{pmatrix} (1-2it_1)I_p & -2it_1C \\ -2it_2C' & (1-2it_2)I_q \end{pmatrix}$.

The partitioned representation of M allows its determinant to be written as:

$|M| = |M_{11}||M_{22} - M_{21}M_{11}^{-1}M_{12}|$.

Thus,

$$(4.5)\quad |M| = (1-2it_1)^p|(1-2it_2)I_q - 4it_1it_2C'C/(1-2it_1)|$$
$$= (1-2it_1)^p(1-2it_2)^q \prod_{j=1}^{p} (1-c\lambda_j) \quad .$$

Now M^{-1} can be written as:

$$M^{-1} = \{(1-2it_1)(1-2it_2)\}^{-1} \times$$
$$\begin{pmatrix} (1-2it_2)(I_p - cCC')^{-1} & 2it_1(I_p - cCC')^{-1}C \\ 2it_2(I_q - cC'C)^{-1}C' & (1-2it_1)(I_q - cC'C)^{-1} \end{pmatrix} .$$

Thus,

$$(4.6)\quad i\theta^{*\prime}M^{-1}T\theta^{*} = \{(1-2it_1)(1-2it_2)\}^{-1} \times$$
$$\{it_1(1-2it_2)\theta_1^{*\prime}(I_p-cCC')^{-1}\theta_1^{*} +$$
$$it_2(1-2it_1)\theta_2^{*\prime}(I_q-cC'C)^{-1}\theta_2^{*} +$$
$$2it_1it_2\theta_1^{*\prime}(I_p-cCC')^{-1}C\theta_2^{*} +$$
$$2it_1it_2\theta_2^{*\prime}(I_q-cC'C)^{-1}C'\theta_1^{*}\} ,$$

where $\theta^{*} = \begin{pmatrix} \theta_1^{*} \\ \theta_2^{*} \end{pmatrix}$. By making the substitutions:

i. $\theta_1^{*\prime}(I_p-cCC')^{-1}\theta_1^{*} = \sum_{j=1}^{p} (1-c\lambda_j)^{-1}\theta_{1j}^{*\,2}$,

ii. $\theta_2^{*\prime}(I_q-cC'C)^{-1}\theta_2^{*} = \sum_{j=1}^{q} (1-c\lambda_j)^{-1}\theta_{2j}^{*\,2}$, and

iii. $\theta_1^{*\prime}(I_p-cCC')^{-1}C\theta_2^* = \theta_2^{*\prime}(I_q-cC'C)^{-1}C'\theta_1^*$

$$= \sum_{j=1}^{p}(1-c\lambda_j)^{-1}\sqrt{\lambda_j}\,\theta_{1j}^*\theta_{2j}^*$$

into 4.6 and letting $\tau_{kj} = (1-c\lambda_j)^{-\frac{1}{2}}\theta_{kj}^*$, we have

$$(4.7)\quad i\theta^{*\prime}M^{-1}T\theta^* = c_1\sum_{j=1}^{p}\tau_{1j}^2 + c_2\sum_{j=1}^{q}\tau_{2j}^2 + c\sum_{j=1}^{p}\sqrt{\lambda_j}\,\tau_{1j}\tau_{2j}\ .$$

Upon substituting 4.5 and 4.7 into 4.4, the result is obtained.

We now consider the bivariate cdf of S_1 and S_2. Let

$$(4.8)\quad f_{S_1,S_2}(x_1,x_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(-it_1x_1-it_2x_2) \times \phi_{S_1,S_2}(t_1,t_2)\,dt_1\,dt_2$$

be the probability density function of S_1 and S_2. Jensen (1970a) solved 4.8 for the case where $\psi_1 = \psi_2 = 0$ (both S_1 and S_2 are central χ^2 variables here). This in turn leaves $\tau_{kj} = 0$ $(k=1,2;\ j=1,\cdots,p)$. Jensen first expanded $(1-c\lambda_j)^{-\frac{1}{2}}$ as $\sum_{k=0}^{\infty}a_{jk}c^k$, where

$$a_{jk} = \lambda_j^k\Gamma(k+\tfrac{1}{2})\{\Gamma(k+1)\Gamma(\tfrac{1}{2})\}^{-1} \quad (j=1,\cdots,p;\ k=0,1,2,\cdots).$$

Collecting terms in c^k we have that

$$\prod_{j=1}^{p}(1-c\lambda_j)^{-\frac{1}{2}} = \sum_{k=0}^{\infty}G_k(\boldsymbol{\lambda})c^k$$

where $\boldsymbol{\lambda} = (\lambda_1,\cdots,\lambda_p)'$ and $G_k(\boldsymbol{\lambda}) = \sum_{k_1+\cdots+k_p=k}\cdots\cdots\sum a_{1k_1}a_{2k_2}\cdots a_{pk_p}$. Thus

$$(4.9)\quad \phi_{S_1,S_2}(t_1,t_2) = \sum_{k=0}^{\infty}G_k(\boldsymbol{\lambda})(2c_1)^k(1-2it_1)^{-p/2} \times (2c_2)^k(1-2it_2)^{-q/2}\ .$$

Applying the inverse Fourier exponential transform given in equation 2.11 to 4.9, 4.8 has the form given in equation 2.5 with $\delta = \lambda$.

Removing the restriction of centrality and using Jensen's method, the general cf of S_1 and S_2 is:

$$(4.10) \quad \phi_{S_1,S_2}(t_1,t_2) = \sum_{k=0} \{G_k(\lambda)(2c_1)^k(1-2it_1)^{-p/2} \exp(c_1\sum_1^p \tau_{1j}^2)(2c_2)^k(1-2it_2)^{-q/2} \exp(c_2\sum_1^q \tau_{2j}^2)\exp(c\sum_1^p \sqrt{\lambda_j}\tau_{1j}\tau_{2j})\} .$$

In 4.9 each term of the summation can be factored into two products, each one involving only one of t_1 and t_2. This facilitates the use of 2.11. Unfortunately this algebraic nicety is not possible with the general case of 4.10, at least in its present form. A problem, which to our knowledge remains unsolved, is to find the appropriate inverse Fourier exponential transform to invert 4.10.

4.2 $\mathcal{F}$ Distribution

Definition 4.3 (Incomplete Beta Function)

Let $B(\cdot,\cdot)$ denote the beta function defined by $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$; $a>0$ and $b>0$. The function $I(a,b,x) = B(a,b)^{-1}\int_0^x t^{a-1}(1-t)^{b-1}dt$, for $0\leq x\leq 1$, $a>0$ and $b>0$, is called the incomplete beta function.

It is evident that $I(a,b,0) = 0$ and $I(a,b,1) = 1$.

Theorem 4.7

The series $H(x;\nu,t,\psi) = 0$ for $x<0$

$$= \sum_{j=0}^{\infty} \exp(-\psi/2)(\psi/2)^j I(\nu/2+j, t/2, x^*)/j!$$

$$\text{for } x \geq 0 \quad ,$$

where $x^* = \nu x/(\nu x+t)$, is a cdf in x for $\nu>0$, $t>0$ and $\psi \geq 0$.

Proof

The theorem is a direct result of the properties of the incomplete beta function (Searle, 1971, p.51).

One family of distributions represented by the incomplete beta function is the family of $\mathcal{F}$ distributions.

Definition 4.4 ($\mathcal{F}$ Distribution)

A random variable X with cdf $H(x;\nu,t,\psi)$, ($\nu=1,2,\cdots$; $t=1,2,\cdots;\psi \geq 0$), is said to have an $\mathcal{F}$ distribution, denoted by $\mathcal{F}_{\nu,t,\psi}$, with (ν,t) degrees of freedom and noncentrality parameter ψ. When $\psi=0$, X is said to have a central $\mathcal{F}$ distribution, denoted by $\mathcal{F}_{\nu,t}$, and its cdf is given by:

$$H(x;\nu,t) = \begin{cases} 0 & \text{for } x<0 \\ I(\nu/2,t/2,x^*) & \text{for } x \geq 0 \end{cases} .$$

The usefulness of the $\mathcal{F}$ distribution is perhaps best seen in its relationship to the χ^2 distribution.

Theorem 4.8

If X_1 and X_2 are independent random variables such that $X_1 \sim \chi^2_{\nu,\psi}$ and $X_2 \sim \chi^2_t$, then $F = tX_1/\nu X_2 \sim \mathcal{F}_{\nu,t,\psi}$.

Proof

Bickel and Doksum (1977, p.16).

Let $S_0 \sim \chi^2_t$ and be independent of both S_1 and S_2 as previously defined. Let $F_1 = tS_1/pS_0$ and $F_2 = tS_2/qS_0$ for $t>2$. We then have the following theorem about F_1 and F_2.

Theorem 4.9 (Correlation Coefficient of F_1 and F_2)

(4.11) $\mathrm{Corr}(F_1,F_2) = \mathrm{Corr}_s(F_1,F_2)\{1+\omega_1\omega_2\mathrm{Corr}(S_1,S_2)\}$

where $\mathrm{Corr}_s(F_1,F_2) = \{(1+\omega_1^2)(1+\omega_2^2)\}^{-\frac{1}{2}}$ is the correlation coefficient of F_1 and F_2 when S_1 and S_2 are independent, that is , when F_1 and F_2 are singly dependent; and $\omega_1 = \{(t-2)(p+2\psi_1)\}^{\frac{1}{2}}/(p+\psi_1)$, $\omega_2 = \{(t-2)(q+2\psi_2)\}^{\frac{1}{2}}/(q+\psi_2)$.

Proof

Note that $\mathrm{Cov}(F_1,F_2)= t^2\mathrm{Cov}(S_1/S_0,S_2/S_0)/pq$. Because S_0 is independent of S_1 and S_2,

(4.12) $$\begin{aligned}\mathrm{Cov}(S_1/S_0,S_2/S_0) &= E(S_1S_2)E(S_0^{-2})-E(S_1)E(S_2)E^2(S_0^{-1})\\ &= \mathrm{Cov}(S_1,S_2)E(S_0^{-2})-E(S_1)E(S_2)\times\\ &\quad \mathrm{Var}(S_0^{-1}) .\end{aligned}$$

Now $E(S_0^{-k}) = \int_0^\infty x^{-k}dG(x,t) = \Gamma(t/2-k)\{2^k\Gamma(t/2)\}^{-1}$ for $t>2k$. Therefore $E(S_0^{-1})=\{2(t/2-1)\}^{-1}=(t-2)^{-1}$ for $t>2$; $E(S_0^{-2})=\{4(t/2-1)(t/2-2)\}^{-1}=\{(t-2)(t-4)\}^{-1}$ for $t>4$; and $\mathrm{Var}(S_0^{-1}) = 2/\{(t-2)^2(t-4)\}$. Using the above and $E(S_1) = p+\psi_1$, $\mathrm{Var}(S_1) = 2(p+2\psi_1)$, $E(S_2) = q+\psi_2$, $\mathrm{Var}(S_2) = 2(q+2\psi_2)$, 4.12 may be written as:

(4.13) $$\begin{aligned}\mathrm{Cov}(S_1/S_0,S_2/S_0) &= \mathrm{Cov}(S_1,S_2)\{(t-2)(t-4)\}^{-1} +\\ &\quad 2(p+\psi_1)(q+\psi_2)\{(t-2)^2(t-4)\}^{-1}\\ &= \frac{2(p+\psi_1)(q+\psi_2)}{(t-2)^2(t-4)}\left[\frac{\{(p+2\psi_1)(q+2\psi_2)\}^{\frac{1}{2}}(t-2)}{(p+\psi_1)(q+\psi_2)}\mathrm{Corr}(S_1,S_2)+1\right].\end{aligned}$$

The variance of F_1 is $Var(F_1) = t^2 Var(S_1/S_0)/p^2$ and

$$(4.14)\quad Var(S_1/S_0) = Var(S_1)E(S_0^{-2}) + E^2(S_1)Var(S_0^{-1})$$
$$= 2(p+2\psi_1)\{(t-2)(t-4)\}^{-1} + 2(p+\psi_1)^2\{(t-2)^2(t-4)\}^{-1}$$
$$= \frac{2(p+\psi_1)^2}{(t-2)^2(t-4)}\left\{\frac{(t-2)(p+2\psi_1)}{(p+\psi_1)^2} + 1\right\} .$$

Similarly,

$$(4.15)\quad Var(S_2/S_0) = \frac{2(q+\psi_2)^2}{(t-2)^2(t-4)}\left\{\frac{(t-2)(q+2\psi_2)}{(q+\psi_2)^2} + 1\right\} .$$

Writing $Corr(F_1,F_2)$ as:

$$(4.16)\quad (t^2/pq)Cov(S_1/S_0, S_2/S_0) \times \{(t^2/p^2)Var(S_1/S_0)(t^2/q^2)Var(S_2/S_0)\}^{-\frac{1}{2}} ,$$

using 4.13, 4.14, 4.15, and letting

$\omega_1^2 = (t-2)(p+2\psi_1)/(p+\psi_1)^2$ and $\omega_2^2 = (t-2)(q+2\psi_2)/(q+\psi_2)^2$,

we obtain 4.11 .

The form of 4.11 illustrates how the common denominator of S_0, which affects $Corr_s(F_1,F_2)$, and the dependence of S_1 and S_2, which affects $Corr(S_1,S_2)$, together affect the dependence of F_1 and F_2. It can be shown that $\lim_{t\to\infty} Corr(F_1,F_2) = Corr(S_1,S_2)$.

We now consider the bivariate cdf of F_1 and F_2. First note that the cdf can be written as:

$$(4.17)\quad P(F_1 \le f_1; F_2 \le f_2) = \int_0^\infty P(S_1 \le rf_1x/t; S_2 \le sf_2x/t)\,dG(x;t) .$$

The cdf can be derived, at least up to an integral form, given the bivariate cdf of S_1 and S_2. A general expression for 4.17 has not been found in closed form except for some special cases.

Theorem 4.10

When r, s, and t are positive, even integers and S_1 and S_2 are independent, 4.17 can be written as:

$$(4.18)\quad H(f_1;r,t,\psi_1) + H(f_2;s,t,\psi_2) - 1 + \frac{\exp(-(\psi_1+\psi_2)/2)}{A^{t/2}\Gamma(t/2)} \sum_{k=0}^{\infty} \sum_{k'=0}^{\infty} \frac{\psi_1^k \psi_2^{k'}}{k!k'!2^{k+k'}} \times \sum_{j=0}^{n_1} \sum_{j'=0}^{n_2} \frac{z_1^j z_2^{j'}}{j!\ j'!} \Gamma(t/2+j+j')\ ,$$

where $A = rf_1/t + sf_2/t + 1$; $z_1 = rf_1/tA$; $z_2 = sf_2/tA$; $n_1 = r/2+k-1$; and $n_2 = s/2+k'-1$.

Proof

From elementary probability theory, if ξ_1 and ξ_2 denote two events, then $P(\xi_1\xi_2) = P(\xi_1) + P(\xi_2) - 1 + P(\overline{\xi}_1\overline{\xi}_2)$, where $\overline{\xi}_1$ and $\overline{\xi}_2$ are the respective complementary events. Thus in order to establish 4.18, it will suffice to show that the last term of 4.18 is:

$$(4.19)\quad P(F_1>f_1;F_2>f_2).$$

Since S_1 and S_2 are independent, 4.19 can be written as:

$$(4.20)\quad \int_0^\infty\{1-G(rf_1x/t;r,\psi_1)\}\{1-G(sf_2x/t;s,\psi_2)\}dG(x;t)\ .$$

For even degrees of freedom ν, Abramowitz and Stegun (1964, equation 26.4.5) show that:

$$G(z;\nu) = 1 - \exp(-z/2)\sum_{j=0}^{\nu/2-1}(z/2)^j/j!\ .$$

Thus, applying Theorem 4.1,

$$(4.21)\quad 1 - G(z;\psi,\nu) = \sum_{k=0}^{\infty} \exp(-\psi/2)\{(\psi/2)^k/k!\}\exp(-z/2)\sum_{j=0}^{\nu/2+k-1}(z/2)^j/j!\ .$$

Let $n_1 = r/2+k-1$ and $n_2 = s/2+k'-1$ and note that $g(x)=dG(x;t)/dx = x^{t/2-1}\exp(-x/2)/\{2^{t/2}\Gamma(t/2)\}$. Then using 4.21, 4.20 becomes:

$$(4.22)\quad \int_0^{\infty}\kappa_1(x)\kappa_2(x)g(x)dx$$

where $\kappa_i(x) = \sum_{k=0}^{\infty}\alpha_i(x,k)$ and

$$\alpha_i(x,k) = \exp(-\psi_i/2)\{(\psi_i/2)^k/k!\}\exp(-x/2)\sum_{j=0}^{n_i}(x/2)^j/j!.$$

Making a change of variable $w = Ax$, where $A = rf_1/t + sf_2/t + 1$, and interchanging the integral and summations (see Appendix B for justification of this step), we obtain:

$$(4.23)\quad \sum_{k=0}^{\infty}\sum_{k'=0}^{\infty}\int_0^{\infty}\alpha_1(x,k)\alpha_2(x,k')g(x)dx\ .$$

Each term of 4.23 will be the product of a factor not involving w times a factor of the form:

$A^{-(t/2+j+j')}\int_0^{\infty}w^{(t/2+j+j'-1)}/(2^{(t/2+j+j')})\exp(-w/2)dw =$

$\Gamma(t/2+j+j')/(A^{(t/2+j+j')})\int_0^{\infty}dG(w;t+2j+2j') =$

$\Gamma(t/2+j+j')/(A^{(t/2+j+j')})$. Letting $z_1 = rf_1/tA$ and $z_2 = sf_2/tA$, we obtain the last term of 4.18.

From Jensen's (1970a) bivariate χ^2, we have an expression for 4.17 when S_1 and S_2 are central χ^2 random variables.

Theorem 4.11

When $\psi_1 = \psi_2 = 0$, the bivariate cdf of F_1 and F_2 is given by:

$$(4.24)\quad \sum_{k=0}^{\infty} G_k(\lambda) \sum_{m=0}^{k} \sum_{n=0}^{k} b_{mn}(k) \times$$

$$P_s\{\mathcal{F}_{r+2m,t} \leq rf_1/(r+2m); \mathcal{F}_{s+2n,t} \leq sf_2/(s+2n)\}$$

where $G_k(\lambda)$ is as defined for 4.9; $b_{mn}(k) = (-1)^{m+n}\binom{k}{m}\binom{k}{n}$; and $P_s\{\mathcal{F}_{\nu_1,t} \leq x_1; \mathcal{F}_{\nu_2,t} \leq x_2\}$ denotes the joint cdf of two singly dependent $\mathcal{F}$ variates with distributions $\mathcal{F}_{\nu_1,t}$ and $\mathcal{F}_{\nu_2,t}$.

Proof

Consider the cdf of F_1 and F_2 in the form of 4.17. Using 2.5,

$$(4.25)\quad P(S_1 \leq rf_1x/t; S_2 \leq sf_2x/t) =$$

$$\int_0^{rf_1x/t} \int_0^{sf_2x/t} \sum_{k=0}^{\infty} G_k(\lambda) \sum_{m=0}^{k} \sum_{n=0}^{k} b_{mn}(k) G'(t_1; r+2m) \times$$

$$G'(t_2; s+2n)\,dt_2\,dt_1 =$$

$$\sum_{k=0}^{\infty} G_k(\lambda) \sum_{m=0}^{k} \sum_{n=0}^{k} b_{mn}(k) G(rf_1x/t; r+2m) G(sf_2x/t; s+2n)$$

(the justification for the last step is similar to that made for 4.23, see Appendix B). Therefore the cdf is:

$$(4.26)\quad \sum_{k=0}^{\infty} G_k(\lambda) \sum_{m=0}^{k} \sum_{n=0}^{k} b_{mn}(k) \times$$

$$\int_0^{\infty} G(rf_1x/t; r+2m) G(sf_2x/t; s+2n)\,dG(x;t) .$$

From 4.17, we recognize the integral in 4.26 as $P_s(\mathcal{F}_{r+2m,t} \leq rf_1/(r+2m); \mathcal{F}_{s+2n,t} \leq sf_2/(s+2n))$, thus 4.24 is established.

4.3 Computational Methods

The computation of probabilities from univariate χ^2 and $\mathcal{F}$ cdf's is facilitated by the availability of functions in SAS which were contributed by Hardison, Quade, and Langston (1980). These functions, given degrees of freedom and a noncentrality parameter, can provide the probability given the argument of the cdf, or provide the argument given the probability.

To calculate probabilities of singly dependent $\mathcal{F}$'s, one may use 4.18, for the case of even degrees of freedom, by truncating the infinite series. For the case r=2 and s=4, a limit of 50 for k gave us results accurate to four decimal places for even values of t. The programming was done in SAS. A listing is included in Appendix C.1.

To calculate probabilities of doubly dependent, but central $\mathcal{F}$'s, we employed 4.18 together with 4.24. The method used for calculating values of $G(\lambda)$ is similar to one suggested by McAllister *et al.* (1981a). This method also involves the truncation of an infinite series. For r=2 and s=4, a limit of 20 for k yielded results accurate to three or four decimal places for even values of t. A listing of the SAS program is in Appendix C.2.

Finally we considered a method for evaluating the cdf of doubly dependent, doubly noncentral $\mathcal{F}$'s. Because this distribution has not been found in closed form, we must use an approximate method. None of the methods reviewed in Chapter 2 was deemed optimal for various reasons:

i. (Gaussian quadrature based on orthogonal Laguerre polynomials.) This method proved accurate for the univariate case, but not for the bivariate, doubly central case. We thus assumed the same would be true for the doubly noncentral case.

ii. (Substitution of $\sum_1^r \lambda_i/r$ for each λ_i $(i=1,\cdots,r)$.) As mentioned in Chapter 2, this method overestimates the probabilities.

iii. (Dutt and Som's integral transformation approximation.) This procedure is rather cumbersome and proved accurate only for the cases when $\lambda_i = \lambda$ $(i=1,\cdots,r)$, $r=s$, and $f_1 = f_2$.

iv. (Tan and Wong's moment approximation.) This method is extremely cumbersome since it involves higher order joint moments. Furthermore Tan and Wong (1977) reported that this approximation is not good for "extreme" values of λ_i (values close to 1) or "large" noncentrality parameters.

We will use a simulation method that will take advantage of the normality assumptions. Consider an $a\times b$ fixed-effects design with no empty cells. We can calculate any of the sums of squares in terms of the cell means and cell sizes. Let $\overline{Y}_{ij}$ denote the sample mean for the (i,j)th cell. Under normality, $\bar{Y}_{ij} \sim N(\mu_{ij}, \sigma^2/n_{ij})$. Using a standard normal random number generator (such as the function NORMAL in SAS), we can generate a random sample of cell means and thus a random sample of sums of squares. An independent sample of error sums of squares can be generated by using the inverse χ^2 function with uniform (on the interval (0,1)) random numbers as the arguments.

Suppose ξ is the intersection of the events ξ_1 and ξ_2, that is, $\xi = \xi_1\xi_2$. With knowledge of $P(\xi_1)$ and $P(\xi_2)$ we can approximate $P(\xi)$ in the following manner. Let n_1 be the number of occurrences of ξ in N simulations. Let n_2 be the number of occurrences of $\xi_1\bar{\xi}_2$. Then the statistic

$$(4.27) \quad \hat{P}_\xi = w_1 n_1/N + w_2(P(\xi_1) - n_2/N) ,$$

where $w_1 + w_2 = 1$ and $w_k \geq 0$ $(k=1,2)$, will have the following properties.

<u>Theorem 4.12</u>

i. $E(\hat{P}_\xi) = P(\xi)$;

ii. $\mathrm{Var}(\hat{P}_\xi) = \{P(\xi_1)(1-P(\xi_1))w_2^2+(w_1+(2P(\xi_1)-1)w_2)P(\xi) - P^2(\xi))/N\}$;

iii. $\mathrm{Var}(\hat{P}_\xi)$ is minimal when $w_2 = P(\xi_2|\xi_1) = P(\xi)/P(\xi_1)$.

<u>Proof</u>

i. $E(n_1/N) = P(\xi)$ and $E(n_2/N) = P(\xi_1\bar{\xi}_2)$. Since $P(\xi) = P(\xi_1) - P(\xi_1\bar{\xi}_2)$ and $w_1 + w_2 = 1$, then the expectation of 4.27 is $P(\xi)$.

ii. $\mathrm{Var}(n_1/N) = P(\xi)\{1-P(\xi)\}/N$,

$\mathrm{Var}(n_2/N) = P(\xi_1\bar{\xi}_2)\{1-P(\xi_1\bar{\xi}_2)\}/N$

$$= \{P(\xi_1)-P(\xi)\}\{1-P(\xi_1)+P(\xi)\}/N, \text{ and}$$

$$Cov(n_1/N, n_2/N) = E(n_1 n_2/N^2) - E(n_1/N)E(n_2/N)$$

$$= (N-1)P(\xi)P(\xi_1\overline{\xi}_2)/N - P(\xi)P(\xi_1\overline{\xi}_2)$$

$$= -P(\xi)\{P(\xi_1) - P(\xi)\}/N .$$

Thus,

$$Var(\hat{P}_\xi) = w_1^2 Var(n_1/N)+w_2^2 Var\{P(\xi_1)-n_2/N\} +$$

$$2Cov(w_1 n_1/N, w_2\{P(\xi_1)-n_2/N\})$$

$$= \{P(\xi_1)(1-P(\xi_1))w_2^2+(w_1+w_2)(w_1+(2P(\xi_1)-1)w_2)P(\xi)$$

$$+ (w_1+w_2)P^2(\xi)\}/N .$$

Noting that $w_1 + w_2 = 1$, the result is obtained.

iii. The derivative of $Var(\hat{P}_\xi)$ with respect to w_2 is

$$(4.28) \quad \{2P(\xi_1)(1-P(\xi_1))w_2+ ((2P(\xi_1)-1)-1)P(\xi)\}/N$$

$$= 2\{1-P(\xi_1)\}\{P(\xi_1)w_2-P(\xi)\}/N .$$

Now 4.28 is zero when $w_2 = P(\xi)/P(\xi_1)$. Since the second derivative of $Var(\hat{P}_\xi)$ is $2\{1-P(\xi_1)\}P(\xi_1)/N \geq 0$, $w_2 = P(\xi)/P(\xi_1)$ yields the minimum value of $Var(\hat{P}_\xi)$.

In the context of the problem at hand, ξ_1 is the event that $(F_1 \leq f_1)$ and ξ_2 is the event that $(F_2 \leq f_2)$. $P(\xi)$ is the probability that H_C (no difference in main column effects) and H_I (no interaction) are both accepted, where f_1 and f_2 are the respective critical values.

Since the optimal weight, w_2, is dependent upon $P(\xi)$, we must approximate w_2 iteratively by first letting $\hat{P}_\xi = n_1/N$, calculate w_2, use w_2 in 4.27 to obtain a new $\hat{P}_\xi$, and repeat the process until essentially no change in w_2 occurs. A 95% confidence interval for $P(\xi)$ may be found by the use of $\{\hat{P}_\xi - P(\xi)\}/\{\mathrm{Var}(\hat{P}_\xi)\}^{\frac{1}{2}}$ and the central limit theorem. Due to limitations of program efficiency, 10,000 simulations were used. A listing of the SAS program is in Appendix C.3.

Chapter 5

A Comparison of the STP and EAD Analyses to the Independent Analysis

5.1 Introduction

In this chapter we compare the STP and EAD analyses to the independent analysis, and thus to each other, by way of the hypothesis-testing error probabilities as defined in Chapter 1. We make the following assumptions for all three analyses:

1. H_{C1} is the hypothesis of primary interest,
2. the analysis starts with a test of H_I and concludes with a test of H_{C1} only if H_I is not rejected,
3. the marginal type I error probabilities are set to 0.10 for the test of H_I and 0.05 for the test of H_{C1} (alternatively we may set the type I error probability for the test of H_{C1} conditioned on not rejecting H_I to 0.05), and
5. the design is a 3×3 ECART (equal column and row totals) design with no empty cells.

The comparison will be divided into four cases: H_I and H_{C1} true, H_I true and H_{C1} false, H_I false and H_{C1} true, and both H_I and H_{C1} false. For each case the discussion will begin with an example derived from the example in Chapter 1.

A two-parameter family of ECART designs as shown in Table 5.1 will be used throughout this chapter. ECART designs are useful in that they allow for various patterns of cell sizes, and thus degrees of nonorthogonality, by simple adjustment of the parameters.

Table 5.1 Relative Cell Sizes* for a Two-Parameter Family of ECART Designs

$1+p+q$	$1-2q$	$1-p+q$
$1-2q$	$1+4q$	$1-2q$
$1-p+q$	$1-2q$	$1+p+q$

$-\frac{1}{4} < q < \frac{1}{2}$, $|p| < 1+q$

* cell sizes divided by the average cell size

Various values of p and q were considered. Those used for this study were chosen in order to display the effects of nonorthogonality in what may be viewed as extreme cases according to an eyeball inspection. Indeed values of p and q near zero have minimal effect, but the purpose here is to investigate the effects of nonorthogonality and the extreme cases merely serve to make the effects identifiable. Three designs were chosen to represent those considered and are given in Table 5.2.

Values of the average cell size, say k, were chosen to be 3, 5, and 15 yielding error degrees of freedom 18, 36, and 126 respectively. Here individual cell sizes may be

Table 5.2 Relative Cell Sizes for Three ECART Designs

Design 1 (D1)	1	1	1	p=0
	1	1	1	q=0
	1	1	1	
Design 2 (D2)	.8	1.4	.8	p=0
	1.4	.2	1.4	q=-.2
	.8	1.4	.8	
Design 3 (D3)	1.3	1.4	.3	p=.5
	1.4	.2	1.4	q=-.2
	.3	1.4	1.3	

non-integral.

The independent analysis, as mentioned in Chapter 1, will be assumed to have the properties below:

1. The noncentrality parameter of the F statistic for testing H_I is $k\|\mathbb{P}(\beta:AB)\|^2$ and the noncentrality parameter for testing H_{C1} is $k\|\mathbb{P}(\beta:B|J)\|^2$.
2. The F statistics for testing H_I and H_{C1} are independent.

The first property forces the noncentrality parameters to be that which would be realized for both the STP and EAD analyses if the cell sizes were all replaced by k. The idea is to make the noncentrality parameter (and therefore the tests) not depend on the degree of unbalancedness or upon the presence of other effects. This is accomplished since the vector spaces AB and B|J are orthogonal and in the parameter space which is the same regardless of the cell sizes.

The noncentrality parameters for the independent analysis differ from those of the STP and EAD analyses due to the unbalancedness of the design. As a result the interaction tests may be dissimilar when interaction is present due to the different noncentrality parameters. When interaction is present we will consider an additional analysis which is the independent one with the interaction noncentrality parameter replaced by the interaction noncentrality parameter of the EAD and STP analyses. This is the noncentrality parameter we would attain by reducing k, say to k*. In effect k-k* indicates what reduction in the average cell size would be possible in order to achieve the same power for the test of H_I with a balanced design with known variance.

We now define patterns of true cell means that we can manipulate to yield varying magnitudes of effects. Four patterns were chosen as shown in Table 5.3.

Table 5.3 Patterns of True Cell Means

Pattern 1 (P1):
H_I true and H_{C1} false

c	0	-c
c	0	-c
c	0	-c

Pattern 2 (P2):
H_I true and H_{C1} false

0	c	0
0	c	0
0	c	0

Table 5.3 continued

Pattern 3 (P3):
H_I false, H_{C1} true

d	-2d	d
-2d	4d	-2d
d	-2d	d

Pattern 4 (P4):
H_I false, H_{C1} false

d	c-2d	d
-2d	c+4d	-2d
d	c-2d	d

The noncentrality parameters in terms of the ECART design parameters are given for each of the test statistics in Table 5.4.

Table 5.4 ECART Design Noncentrality Parameters

Analy. Type	H	Geometric[a] NCP	Cell Means Pattern P1	P2 & P3 & P4
Indep.	H_I	$k\|P(\beta:AB)\|^2$	0	$36d^2k$
Indep.	H_{C1}	$k\|P(\beta:B\|J)\|^2$	$6c^2k$	$2c^2k$
STP & EAD	H_I	$\|P(\mu:\tilde{A}B)\|^2$	0	$\frac{1+3q-6q^2-8q^3}{1+q}\,36d^2k$
STP	H_{C1}	$\|P(\mu:B\|\tilde{J})\|^2$	$\frac{18}{\eta_1}\,c^2k$	$\frac{18}{\eta_1+2\eta_2}\,c^2k$
EAD	H_{C1}	$\|P(\mu:\hat{B}\|\hat{A})\|^2$	$\frac{2}{3}(9-4p^2)c^2k$	$2(1-4q^2)\left(c+\frac{6qd}{1+2q}\right)^2k$

$\eta_1=(1+p+q)^{-1}+(1-2q)^{-1}+(1-p+q)^{-1}$ $\quad\eta_2=(1+4q)^{-1}+2(1-2q)^{-1}$

[a] The noncentrality parameter, ψ, is as defined in Chapter 4. Pearson-Hartley charts (1951): $\phi=\sqrt{\psi/m\sigma^2}$, m=degrees of freedom.

Values of c were chosen so that $6c^2$ for P1 and $2c^2$ for P2 are equal to 1/3, 2/3, 1, and 4/3. Two values of d were chosen so that the power of the interaction test using the independent analysis would be approximately 0.66 and 0.99 while being 0.33 and 0.75 using the STP and EAD analyses (k=15). This corresponds to $36d^2$ being 0.5 and 1.5.

As stated before, the discussion of each case will begin with a numerical example. The following items will be displayed:

1. Pattern of true cell means.
2. Data obtained by adding computer generated random errors with standard normal distributions to the true cell means. (If the data indicated rejecting H_I, then a new set was generated.)
3. True values and estimates of μ_{ij}, $\mu_{\cdot j}$, μ_{*j}, and $(\mu_{i*})_{*j}$. (The noncentrality parameters of the F statistics testing H_{C1} under STP and EAD are zero if and only if $\mu_{\cdot 1}=\mu_{\cdot 2}=\mu_{\cdot 3}$ and $(\mu_{i*})_{*j}=\mu_{*j}$ for j=1,2,3 respectively.)
4. ANOVA table including
   i. the marginal p-value which is the univariate probability that F>observed-F,
   ii. the conditional p-value which is the conditional probability that F>observed-F given H_I is not rejected at the 0.10 significance level (for both the marginal and conditional p-values we assume H_I and H_{C1} are both true), and
   iii. the values of the noncentrality parameters.

5. The joint error probabilities for the independent, STP, and EAD analyses that apply given the true cell means.

For each example we use an ECART design with $p=0$, $q=-.2$, and $k=5$.

5.2 Case 1: H_I true and H_{C1} true

Example 1:

True Cell Means

	5	5	5
	5	5	5
	5	5	5
$\mu_{\cdot j}$	5	5	5
μ_{*j}	5	5	5
$(\mu_{i*})_{*j}$	5	5	5

Data

5.24 3.54 5.37 3.50	4.53 3.78 5.45 4.90 4.93 3.98 5.42	5.77 4.97 3.64 4.29
3.61 5.69 4.29 4.10 4.94 4.08 4.44	4.84	5.28 5.18 3.60 5.54 5.15 3.99 3.30
4.22 3.05 6.53 6.04	2.88 3.51 4.82 5.34 4.04 3.86 5.54	5.72 3.23 4.15 3.69

Estimates*

4.41	4.71	4.67	
4.45	4.84	4.58	
4.96	4.28	4.20	
4.61	4.61	4.48	← $\bar{y}_{\cdot j}$
4.58	4.52	4.50	← $\bar{y}_{*j}$
4.53	4.53	4.53	← $(\bar{y}_{i*})_{*j}$

ANOVA

Source	DF	SS	F	Marginal p-value	Conditional p-value	NCP
EAD: Column	2	0.0452	0.02	0.9758	0.9750	0
STP: Column	2	0.1370	0.07	0.9286	0.9254	0
Interaction	4	1.8395	0.50	0.7372		0
Error	36	33.2360				

Joint Error Probabilities	Error	Independent	EAD	STP
	type(0,1)	0.0450	0.0423	0.0375

*Note that each $\bar{y}_{ij} < 5$. This is due to random chance only.

The analysis indicates insufficient evidence to conclude there is interaction. Since neither the STP nor the EAD sum of squares for columns indicate a difference in column effects, we would conclude that H_{C1} should not be rejected. If we had rejected H_{C1}, then a type(0,1) error would have occurred. The probability of a type(0,1) error is largest with the independent analysis and smallest with the STP analysis. Both the STP and EAD analyses may be considered conservative in this context with STP the most conservative.

Table 5.5 contains the type(0,1) error probabilities and conditional probabilities of a type I error rejecting H_{C1} given H_I not rejected. By '$k=\infty$' we mean 'as $k\to\infty$'. Alternatively we may view the corresponding probabilities as those that would be realized if perfect information about the error variance were available. In this case we would perform chi-squared tests instead of $\mathcal{F}$ tests. These probabilities were thus calculated via a bivariate chi-squared distribution.

The probabilities under EAD are invariant to the design. This follows from the single dependence of the F statistics for testing H_I and H_{C1} and from H_I and H_{C1} being true leaving the EAD noncentrality parameters zero. When k is equal to 3, the error probability is smallest. With k=15, the limiting value of 0.0450 is almost reached. This might be expected in light of the correlation coefficient of the two F statistics. The correlation coefficients for

Table 5.5 Type(0,1) Unconditional and Conditional[a] Probabilities of Rejecting H_{C1} Given H_I Not Rejected

		Analysis Type					
		Independent		EAD		STP	
Design	k	Uncon.	Cond.	Uncon.	Cond.	Uncon.	Cond.
D1	3	.0450	.05	.0396	.0440	.0396	.0440
	5	.0450	.05	.0423	.0470	.0423	.0470
	15	.0450	.05	.0443	.0492	.0443	.0492
	∞	.0450	.05	.0450	.05	.0450	.05
D2	3	.0450	.05	.0396	.0440	.0350	.0389
	5	.0450	.05	.0423	.0470	.0375	.0417
	15	.0450	.05	.0443	.0492	.0394	.0438
	∞	.0450	.05	.0450	.05	.0401	.0446
D3	3	.0450	.05	.0396	.0440	.0314	.0349
	5	.0450	.05	.0423	.0470	.0338	.0376
	15	.0450	.05	.0443	.0492	.0356	.0396
	∞	.0450	.05	.0450	.05	.0364	.0404

[a]Conditional probability = (Unconditional probability)÷0.9

k = 3,5,15 are 0.149, 0.076, and 0.022 respectively.

The F statistics of the STP analysis are doubly dependent and thus nonorthogonality affects the error probabilities. Design 2 makes the STP a more conservative analysis than either the independent or EAD analyses. Design 3 leaves STP most conservative compared to the other designs.

The correlation coefficients of the F statistics in STP for Design 1 are the same as for EAD with Design 1; but for Design 2 the correlations are 0.156, 0.080, 0.023 and for

Design 3 the correlations are 0.162, 0.083, and 0.024. For each design, the STP correlation coefficient is (1+f)*(EAD correlation coefficient) where f is proportional to the correlation between the interaction and STP column sums of squares (see Theorem 4.9, Chapter 4). The effects of this correlation can be seen in the reduction of the error probabilities from Design 1. In fact for each k the ratio of the reduction from Design 1 to Design 2 and from Design 1 to Design 3 is 1:1.8 which is similar to the ratio of the sums-of-squares correlations.

A problem with a conservative test ('conservative' in that the assumed α-level is an overstatement) is that its power is less than what could be achieved if the rejection region were increased to bring the true α-level up to the assumed one. For the subsequent cases, we will also display the error probabilities when the critical values of the STP and EAD F statistics for testing H_{C1} are adjusted so as to make the conditional type I errors equal to 0.05.

5.3 Case 2: H_I true and H_{C1} false

Example 2:

True Cell Means

	5.3	5	4.7
	5.3	5	4.7
	5.3	5	4.7
$\mu_{\cdot j}$	5.3	5	4.7
μ_{*j}	5.3	5	4.7
$(\mu_{i*})_{*j}$	5	5	5

Data

6.48 5.29 5.75 4.65	4.43 5.11 4.79 7.19 4.74 4.50 4.23	3.71 5.61 4.38 3.64
3.82 4.92 6.45 7.19 6.13 5.62 5.73	5.10	6.16 3.74 6.27 2.62 5.48 4.19 4.72
5.63 5.86 4.04 6.60	4.39 3.43 4.40 4.99 5.77 4.49 5.32	5.77 5.18 4.16 3.02

Estimates

5.54	5.00	4.34	
5.69	5.10	4.74	
5.53	4.68	4.53	
5.59	4.93	4.54	← $\bar{y}_{\cdot j}$
5.61	4.86	4.58	← $\bar{y}_{*j}$
5.05	4.94	5.05	← $(\bar{y}_{i*})_{*j}$

ANOVA

Source	DF	SS	F	Marginal p-value	Conditional p-value	NCP
EAD: Column	2	8.1655	3.72	0.0340	0.0317	2.70
STP: Column	2	7.8654	3.58	0.0381	0.0311	2.52
Interaction	4	0.3959	0.09	0.9850		0
Error	36	39.5249				

Joint Error Probabilities	Error	Independent	EAD	STP
	type(0,2)	0.6552	0.6623	0.6837 ±.0016

Table 5.6 Column Noncentrality Parameters for Values of c Versus k

Design 1

Cell Means Patterns: P1, P2

Analyses: Independent, EAD, STP

	c:(P1,P2)			
k	$(\sqrt{2}/6,\sqrt{6}/6)$	$(1/3,\sqrt{3}/3)$	$(\sqrt{6}/6,\sqrt{2}/2)$	$(\sqrt{2}/3,\sqrt{6}/3)$
3	1	2	3	4
5	1.67	3.33	5	6.67
15	5	10	15	20

Design 2

Cell Means Pattern: P1

Analysis: EAD

	c			
k	$\sqrt{2}/6$	1/3	$\sqrt{6}/6$	$\sqrt{2}/3$
3	1	2	3	4
5	1.67	3.33	5	6.67
15	5	10	15	20

Analysis: STP

3	0.93	1.87	2.80	3.73
5	1.56	3.11	4.67	6.23
15	4.67	9.33	14	18.7

Cell Means Pattern: P2

Analysis: EAD

	c			
k	$\sqrt{6}/6$	$\sqrt{3}/3$	$\sqrt{2}/2$	$\sqrt{6}/3$
3	0.84	1.68	2.52	3.36
5	1.40	2.80	4.20	5.60
15	4.20	8.40	12.6	16.8

Table 5.6 continued.

Design 2

Cell Means Pattern: P2

Analysis: STP

k	$\sqrt{6}/6$	$\sqrt{3}/3$	$\sqrt{2}/2$	$\sqrt{6}/3$
3	0.56	1.12	1.68	2.24
5	0.93	1.87	2.80	3.73
15	2.80	5.60	8.40	11.2

Design 3

Cell Means Pattern: P1

Analysis: EAD

k	$\sqrt{2}/6$	$1/3$	$\sqrt{6}/6$	$\sqrt{2}/3$
3	0.89	1.78	2.67	3.56
5	1.48	2.96	4.44	5.93
15	4.44	8.89	13.3	17.8

Analysis: STP

3	0.62	1.25	1.87	2.49
5	1.04	2.08	3.11	4.15
15	3.11	6.23	9.34	12.5

Cell Means Pattern: P2

Analysis: EAD

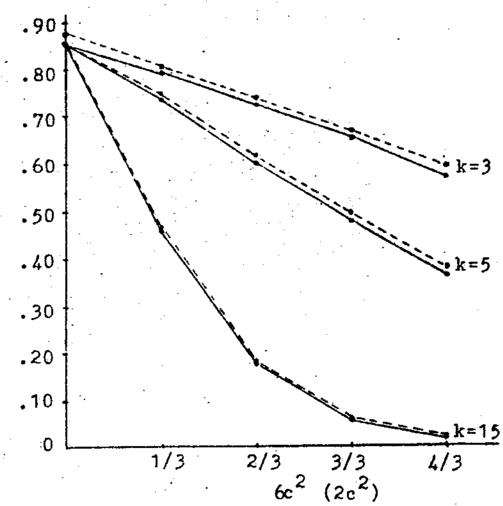
k	$\sqrt{6}/6$	$\sqrt{3}/3$	$\sqrt{2}/2$	$\sqrt{6}/3$
3	0.84	1.68	2.52	3.36
5	1.40	2.80	4.20	5.60
15	4.20	8.40	12.6	16.8

Analysis: STP

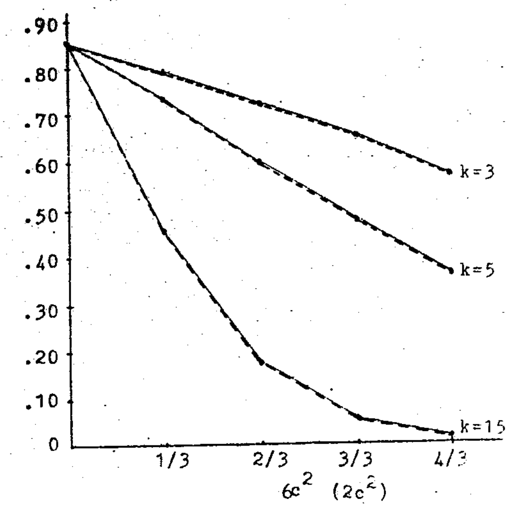
3	0.51	1.02	1.53	2.04
5	0.85	1.70	2.55	3.39
15	2.55	5.09	7.64	10.2

Figure 5.1 Type(0,2) Error Probabilities for k=3, 5, and 15

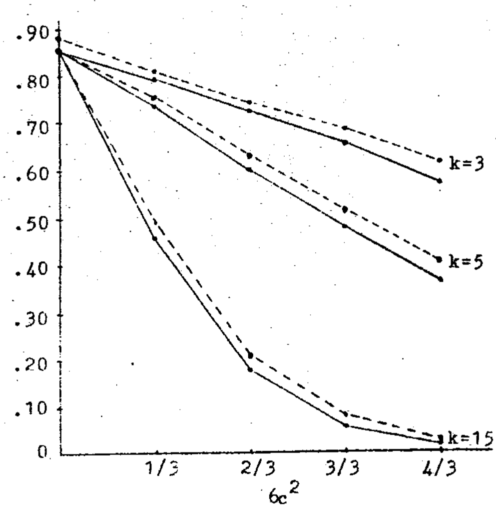
Legend: Vertical- type(0,2) error probability
Horizontal- column noncentrality
Solid line- Independent analysis
top, middle, bottom line- k=3, 5, and 15 respectively
Dashed line-EAD or STP analysis



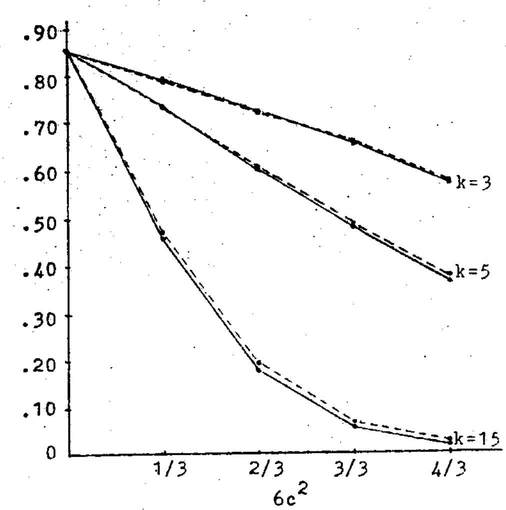
a. EAD and STP: D1,P1 (D1,P2)
EAD: D2,P1



b. Adjusted EAD and adjusted
STP: D1,P1 (D1,P2)
Adjusted EAD: D2,P1

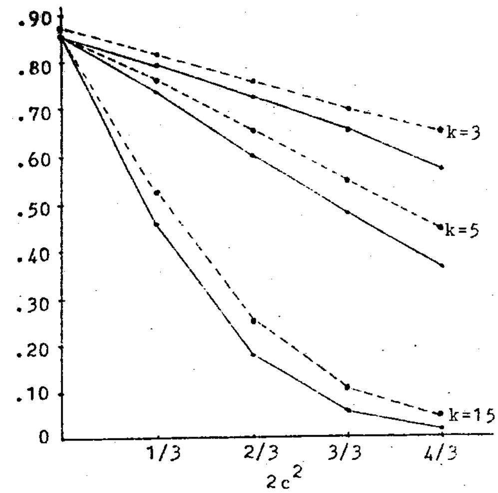


c. STP: D2,P1

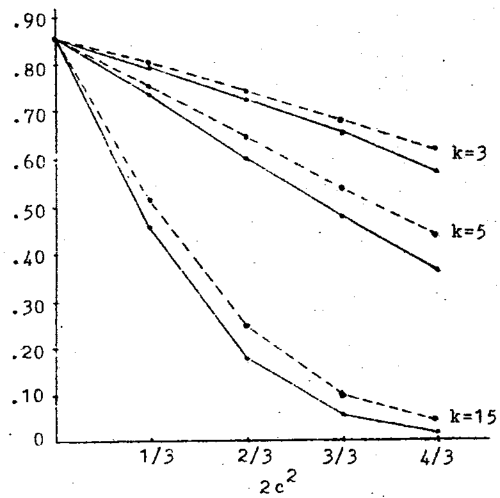


d. Adjusted STP: D2,P1

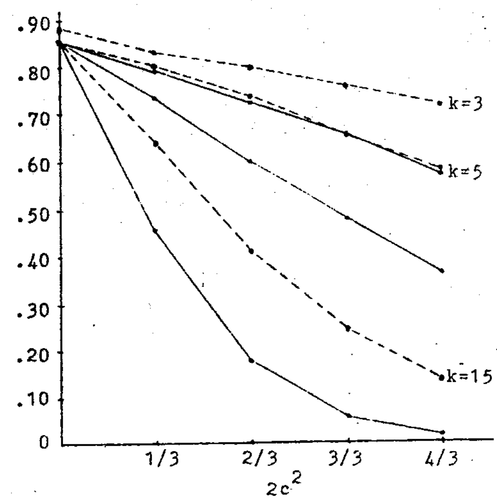
Figure 5.1 continued



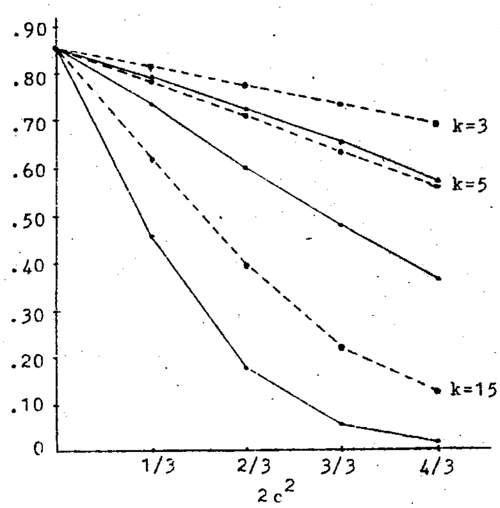
e. EAD: D2,P2 D3,P2



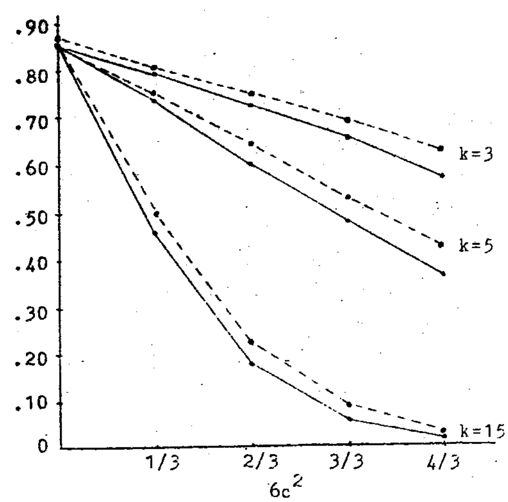
f. Adjusted EAD: D2,P2 D3,P2



g. STP: D2,P2

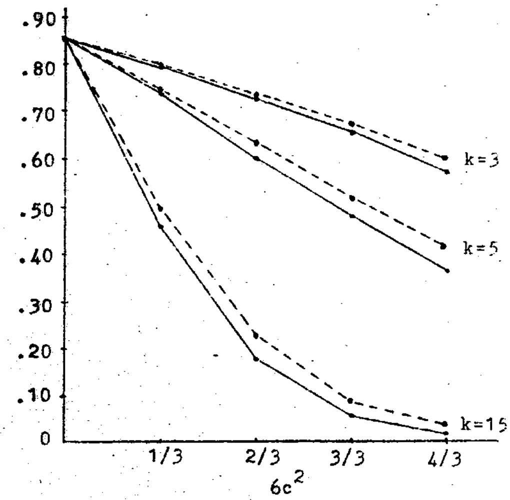


h. Adjusted STP: D2,P2

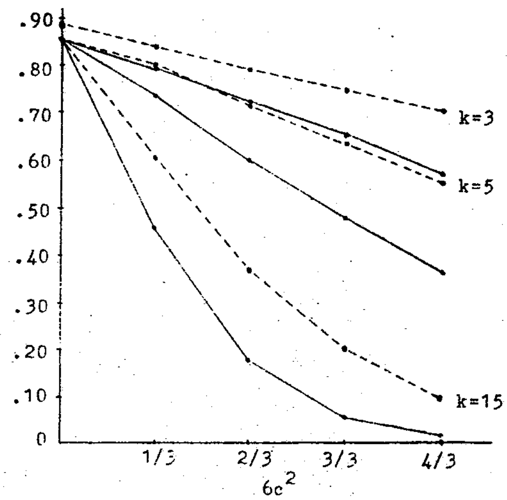


i. EAD: D3,P1

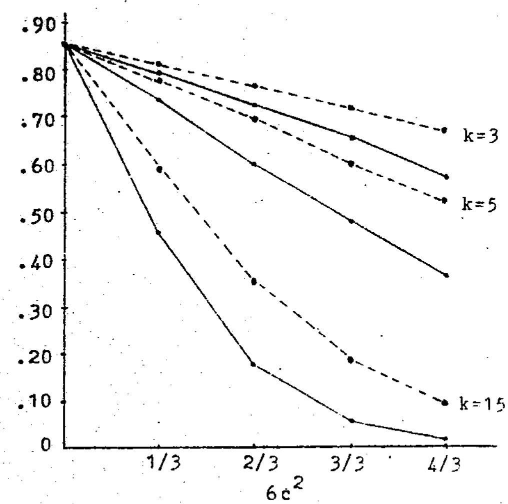
Figure 5.1 continued



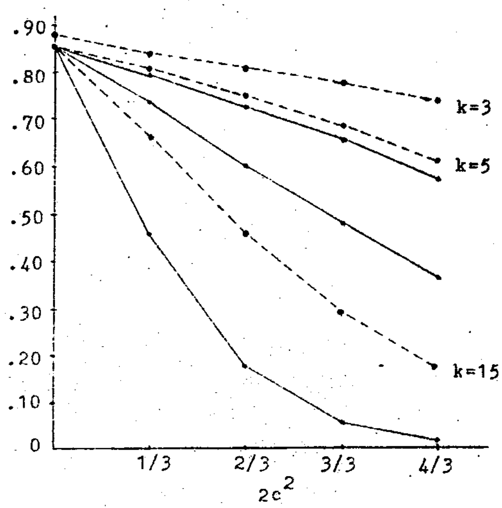
j. Adjusted EAD: D3,P1



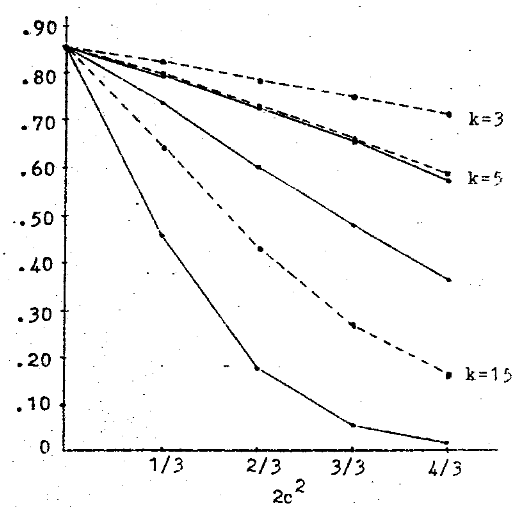
k. STP: D3,P1



l. Adjusted STP: D3,P1



m. STP: D3,P2



n. Adjusted STP: D3,P2

As shown in each of the graphs the type(0,2) error probabilities are greater for EAD and STP analyses than for the independent analysis. The effect of single dependence of the F statistics can be seen in Figure 5.1(a) where the design is balanced. The EAD and STP error probabilities are slightly less when the H_{C1} rejection regions are adjusted. As would be expected the error probabilities decrease with increases in k.

Designs 2 and 3 have two properties not shared by Design 1 which affect the error probabilities. The first, which affects both EAD and STP, is that the nonorthogonality reduces the noncentrality parameter for H_{C1}. This in turn increases the chances of a type(0,2) error. The second, which affects STP only is the dependence between the interaction and column sums of squares induced by nonorthogonality. The second property is most readily seen when cell-means-pattern 2 is used. This illustrates the important point that for a fixed pattern of cell sizes, different patterns of true cell means with H_I true may cause more extreme error probabilities. When the rejection regions are adjusted, the probabilities are reduced, but not dramatically. It also appears that for a fixed cell means pattern, Design 3 gives more extreme error probabilities than Design 2.

5.4 Case 3: H_I false and H_{C1} true

Example 3:

True Cell Means

	5.3	4.4	5.3
	4.4	6.2	4.4
	5.3	4.4	5.3
$\mu_{\cdot j}$	5	5	5
μ_{*j}	4.88	4.52	4.88
$(\mu_{i*})_{*j}$	4.71	4.86	4.71

Data

6.67 5.50 4.77 4.29	4.49 4.17 3.86 5.18 4.66 2.52 5.43	4.10 7.51 5.98 5.40
3.90 6.27 3.27 4.98 3.68 3.56 5.50	4.61	4.97 4.41 4.21 3.90 3.20 3.73 4.81
3.08 5.67 4.34 6.62	5.56 5.13 4.47 4.36 3.98 3.67 3.74	8.07 6.24 6.74 4.08

Estimates

5.31	4.33	5.75	
4.45	4.61	4.18	
4.93	4.42	6.28	
4.90	4.45	5.40	← $\bar{y}_{\cdot j}$
4.81	4.39	5.16	← $\bar{y}_{*j}$
4.70	4.97	4.70	← $(\bar{y}_{i*})_{*j}$

ANOVA

Source	DF	SS	F	Marginal p-value	Conditional p-value	NCP
EAD: Column	2	9.7777	4.07	0.0256	0.0237	3.024
STP: Column	2	4.5097	1.88	0.1679	0.1197	0
Interaction	4	5.2710	1.10	0.3733		6.048
Error	36	43.2810				

Joint Error Probabilities	Error	Independent	EAD	STP
	type(2,1)	0.0034	0.1185	0.0175
	Adjusted:	0.0221		±.0009

The data do not provide sufficient evidence to conclude that interaction is present, thus we would test H_{C1}. The EAD analysis suggests rejecting H_{C1} in favor of, coupled with the assumption of no interaction, accepting that a difference in column main effects exists uniformly over rows. The STP analysis on the other hand suggests that insufficient evidence exists to reject H_{C1}.

Since H_I was not rejected, a type II error occurred. The interaction noncentrality parameter under the independent analysis is 16.2 which yields a power of 0.932. However due to nonorthogonality, the interaction noncentrality parameter decreases to 6.048 for EAD and STP yielding a power of 0.557. Obviously nonorthogonality can severely handicap the EAD and STP analyses' ability to detect interaction.

An average cell size of 1.87 instead of 5 would have made the interaction noncentrality parameter of the independent analysis equal to that of EAD and STP. Alternatively a value of d (for cell means pattern P3 which is used in the example) equal to 0.183 as opposed to 0.3 in the example would reduce the noncentrality parameter to 6.048.

Using the EAD analysis we would commit a type(2,1) error. The EAD column noncentrality is due entirely to the presence of interaction and the nonorthogonality of $\hat{B}|J$ (the space used to test H_{C1}) and $\hat{AB}$ (the space where the interaction resides, see Table 3.2 and Figure 4 in

Chapter 3).

With the STP analysis, H_{C1} is not rejected. Note that the STP column noncentrality parameter is zero. This is due to the orthogonality of $\tilde{B}|J$ (the space used to test H_{C1}) and $\hat{AB}$.

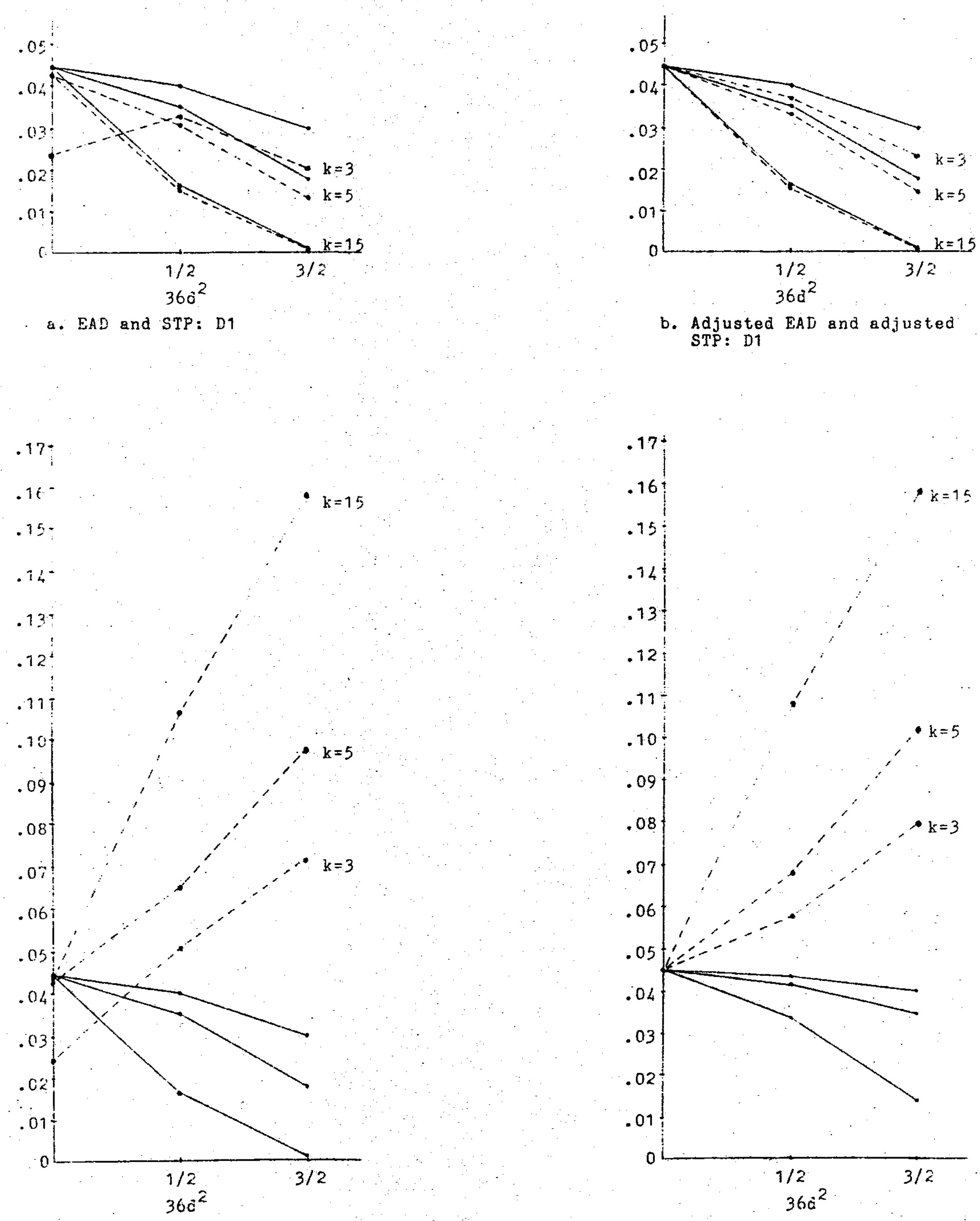
In Figure 5.2 we show the type(2,1) error probabilities for the example designs and cell means patterns. With the independent analysis the noncentralities for interaction are 1.5, 2.5 and 7.5 for k= 3, 5, and 15 when d= $\sqrt{2}/12$ and 4.5, 7.5, and 22.5 for k= 3, 5, and 15 when d= $\sqrt{6}/12$. These are also the EAD and STP interaction noncentralities for Design 1, however for Designs 2 and 3 the noncentralities are approximately 37% of those for the independent analysis. The column noncentrality parameters for EAD are zero with Design 1 and 50% of the EAD interaction noncentrality parameters for Designs 2 and 3.

For the balanced design (D1) we see that the type(2,1) error probabilities are less for EAD and STP than for the independent analysis. Adjusting the column test critical region so to equate the type(0,1) error probabilities increases, although not dramatically, the type(2,1) error probabilities.

For Designs 2 and 3, the EAD error probabilities take on a strikingly different pattern. The probabilities increase with increasing d and with increasing k. As stated earlier, the EAD analysis "thinks" it sees a column effect which is in reality due only to interaction. Increases in

Figure 5.2 Type(2,1) Error Probabilities for for k= 3, 5, and 15

Legend: Vertical- type(2,1) error probability
Horizontal- interaction noncentrality
Solit line- Independent analysis
top, middle, bottom line-
k=3, 5, and 15 respectively
Dashed line-EAD or STP analysis



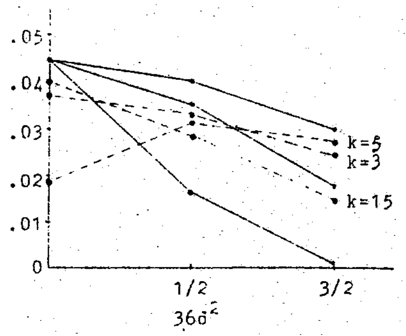
a. EAD and STP: D1

b. Adjusted EAD and adjusted STP: D1

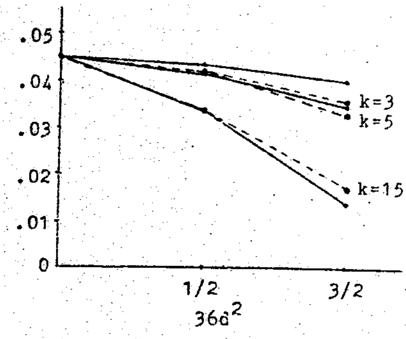
c. EAD: D2 D3

d. Adjusted EAD: D2 D3 (Adjusted Independent)

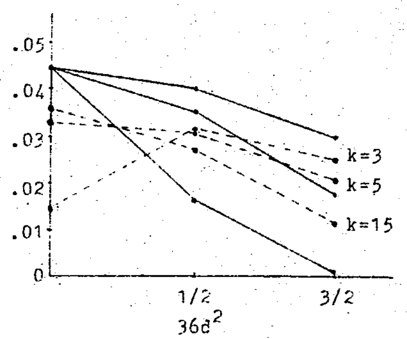
Figure 5.2 continued



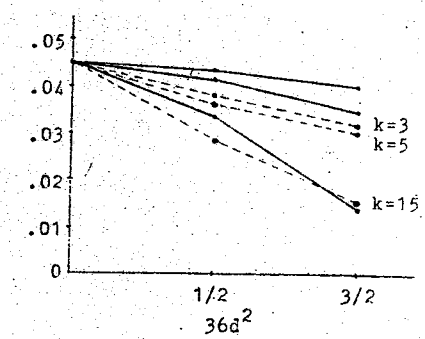
e. STP: D2



f. Adjusted STP: D2
(Adjusted Independent)



g. STP: D3



h. Adjusted STP: D3
(Adjusted Independent)

sample size enhances EAD's ability to detect this phantom column effect. Adjusting the column test's critical region only serves to increase the likelihood of error (most noticeable for k=3).

The STP error probabilities also increase for Designs 2 and 3 over Design 1, but still retain similar patterns, that is, the probabilities decrease with increases in d. It is interesting to note that Design 2 yields larger probabilities than Design 3. First recall that the non-coincidence of $\tilde{AB}$ (the space where interaction is tested) and $\hat{AB}$ will decrease the power of STP's (and EAD's) interaction test. This alone will cause an increase in the type(2,1) error probabilities. However the more nonorthogonal $\tilde{B|J}$ is to $\tilde{AB}$, the more the failure to reject H_I will tend to cause failure to reject H_{C1} (see again Table 5.5). This in turn tends to bring down the type(2,1) error probabilities. The nonorthogonality of $\tilde{B|J}$ and $\tilde{AB}$ is greater with Design 3 than with Design 2. Again it should be noted that a large decrease in the type(2,1) error probability compared to the independent analysis may not be desirable since the power to detect differences in column effects would also be reduced.

5.5 Case 4: H_I false and H_{C1} false

Example 4:

True Cell Means

	5.1	5.3	5.1
	4.8	5.9	4.8
	5.1	5.3	5.1
$\mu_{\cdot j}$	5	5.5	5
μ_{*j}	4.96	5.34	4.96
$(\mu_{i*})_{*j}$	4.97	5.28	5.04

Data

3.29 4.85 3.75 5.38	6.15 5.75 5.08 7.19 7.26 4.35 4.12	6.40 4.74 5.95 3.68
3.61 4.48 5.95 4.64 4.52 4.67 5.04	4.31	4.12 5.82 4.23 3.89 3.71 5.71 3.80
5.76 3.88 5.46 5.35	5.23 4.54 4.45 5.79 6.99 4.88 6.38	6.10 5.84 5.19 6.05

Estimates

4.32	5.70	5.19	
4.70	4.31	4.47	
5.11	5.47	5.78	
4.71	5.16	5.15	← $\bar{y}_{\cdot j}$
4.71	5.50	5.01	← $\bar{y}_{*j}$
4.97	5.28	4.97	← $(\bar{y}_{i*})_{*j}$

ANOVA

Source	DF	SS	F	Marginal p-value	Conditional p-value	NCP
EAD: Column	2	2.0145	1.10	0.3435	0.3371	0.756
STP: Column	2	1.6586	0.91	0.4130	0.3945	1.400
Interaction	4	4.0616	1.11	0.3669		0.672
Error	36	32.9344				

Joint Error Probabilities	Error	Independent	EAD	STP
	type(2,2)	0.5695	0.7663	0.7461
	Adjusted:	0.7489		±.0037

The data do not provide sufficient evidence to reject H_I. Both the EAD and STP tests of H_{C1} suggest not rejecting H_{C1}, thus we would conclude that there are no differences in columns means over the rows.

Since H_I was not rejected, a type II error occurred. Since H_{C1} was not rejected, another type II error occurred. Thus for both analyses, we have committed a type(2,2) error. Figure 5.3 is a plot of the true column means by rows. Although the differences in column means are not all the same over rows, the differences are in the same direction.

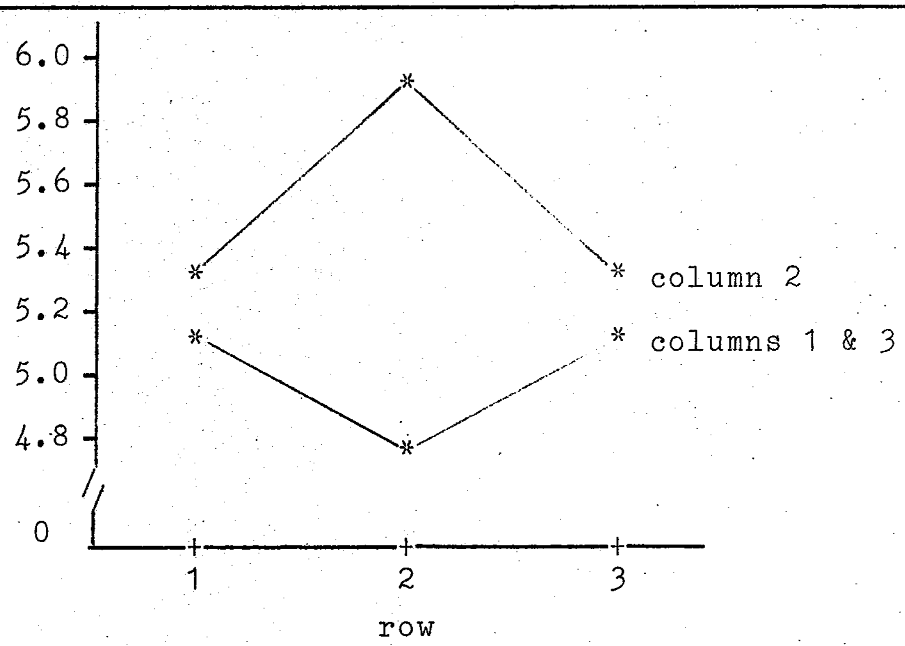


Figure 5.3 True Column Means of Example 4 By Rows

It is open to interpretation whether the interaction that is present is of any practical significance. Surely one would not want to conclude that there are no differences in column means, thus the second type II error is an error still to be protected against.

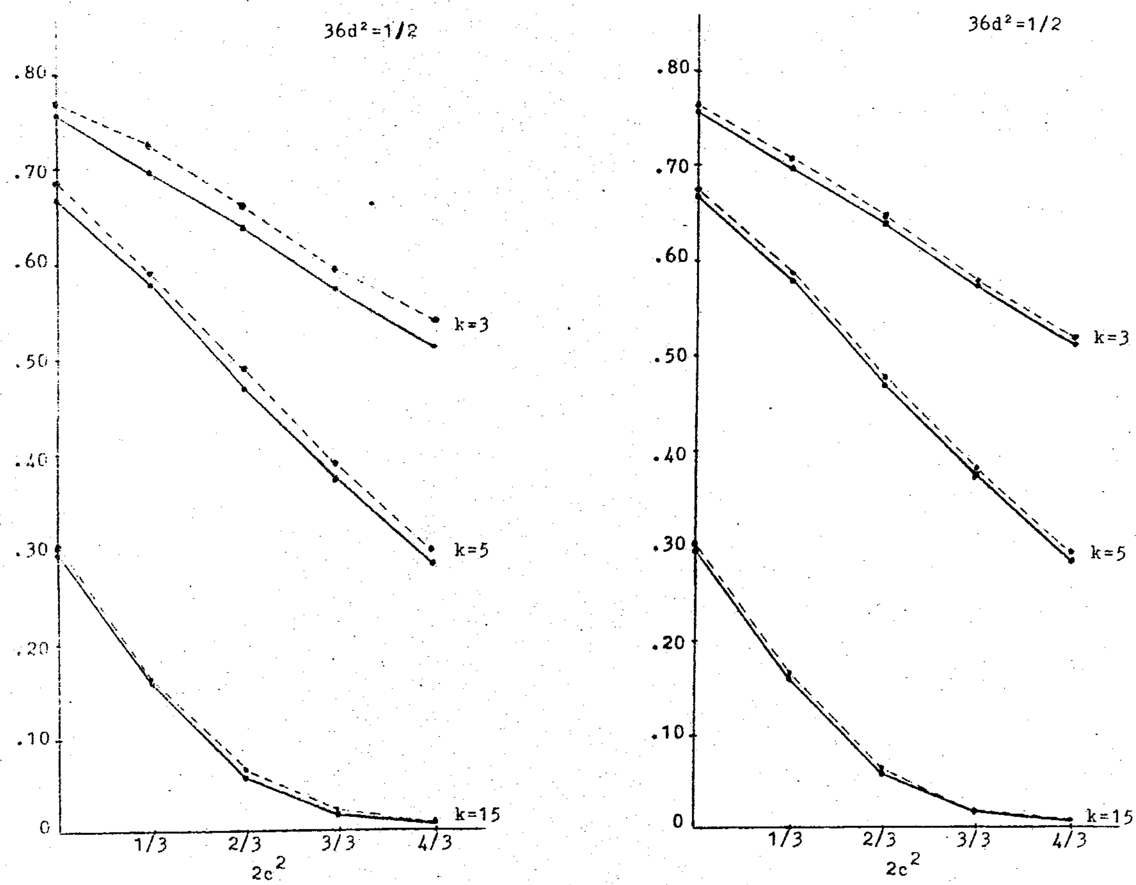
In Figure 5.4 we show the type(2,2) error probabilities. The interaction noncentrality parameters are the same as in Figure 5.2 and the STP column noncentrality parameters are the same as in Figure 5.1. The EAD coulmn noncentrality parameters are given in Table 5.7 for Designs 2 and 3 with cell means pattern 4. Note that we use both $d=-\sqrt{2}/12$ and $d=\sqrt{2}/12$. While both yield the same results for the independent and STP analyses, they yield different results for EAD.

Table 5.7 EAD Column Noncentrality Parameters for Designs 2 and 3 with Cell Means Pattern 4

		k		
d	c	3	5	15
$-\sqrt{2}/12$	$\sqrt{6}/6$	2.09	3.48	10.45
	$\sqrt{3}/3$	3.33	5.55	16.66
	$\sqrt{2}/2$	4.48	7.47	22.40
	$\sqrt{6}/3$	5.58	9.30	27.90
$\sqrt{2}/12$	$\sqrt{6}/6$	0.15	0.25	0.75
	$\sqrt{3}/3$	0.60	1.00	3.00
	$\sqrt{2}/2$	1.11	1.85	5.55
	$\sqrt{6}/3$	1.71	2.85	8.55
$\sqrt{6}/12$	$\sqrt{6}/6$	0	0	0
	$\sqrt{3}/3$	0.15	0.25	0.75
	$\sqrt{2}/2$	0.45	0.75	2.25
	$\sqrt{6}/3$	0.84	1.40	4.20

Figure 5.4 Type(2,2) Error Probabilities for k=3, 5, and 15 and $36d^2$=1/2, 3/2

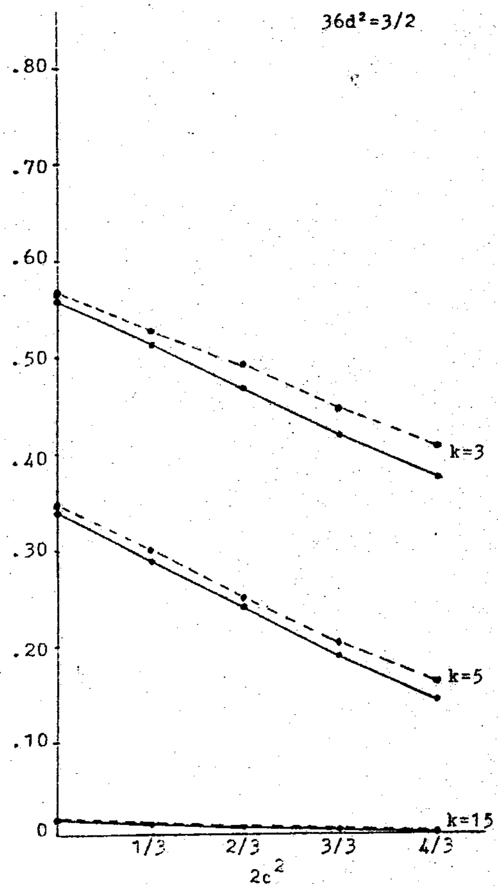
Legend: Vertical- type(2,2) error probaility
Horizontal- column noncentrality
Solid line- Independent analysis
top, middle, bottom line-
k=3, 5, and 15 respectively
Dashed line-EAD or STP analysis



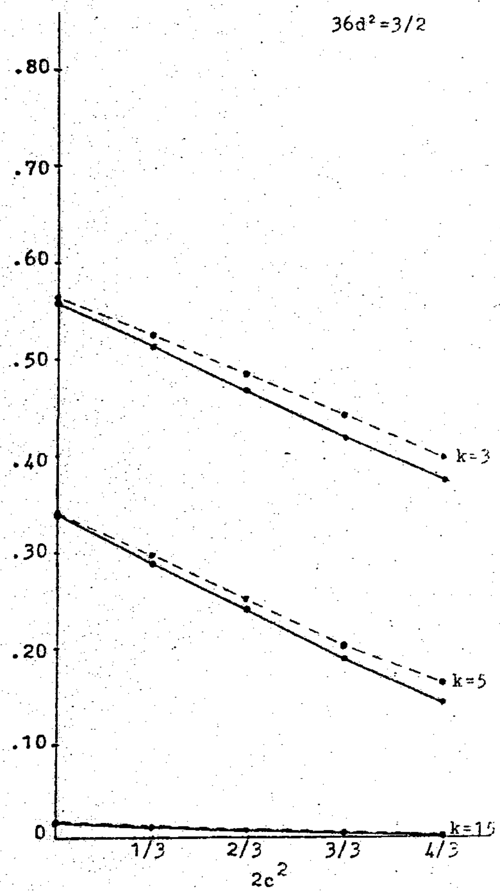
a. EAD and STP: D1

b. Adjusted EAD and adjusted STP: D1

Figure 5.4 continued



c. EAD and STP: D1



d. Adjusted EAD and adjusted STP: D1

Figure 5.4 continued

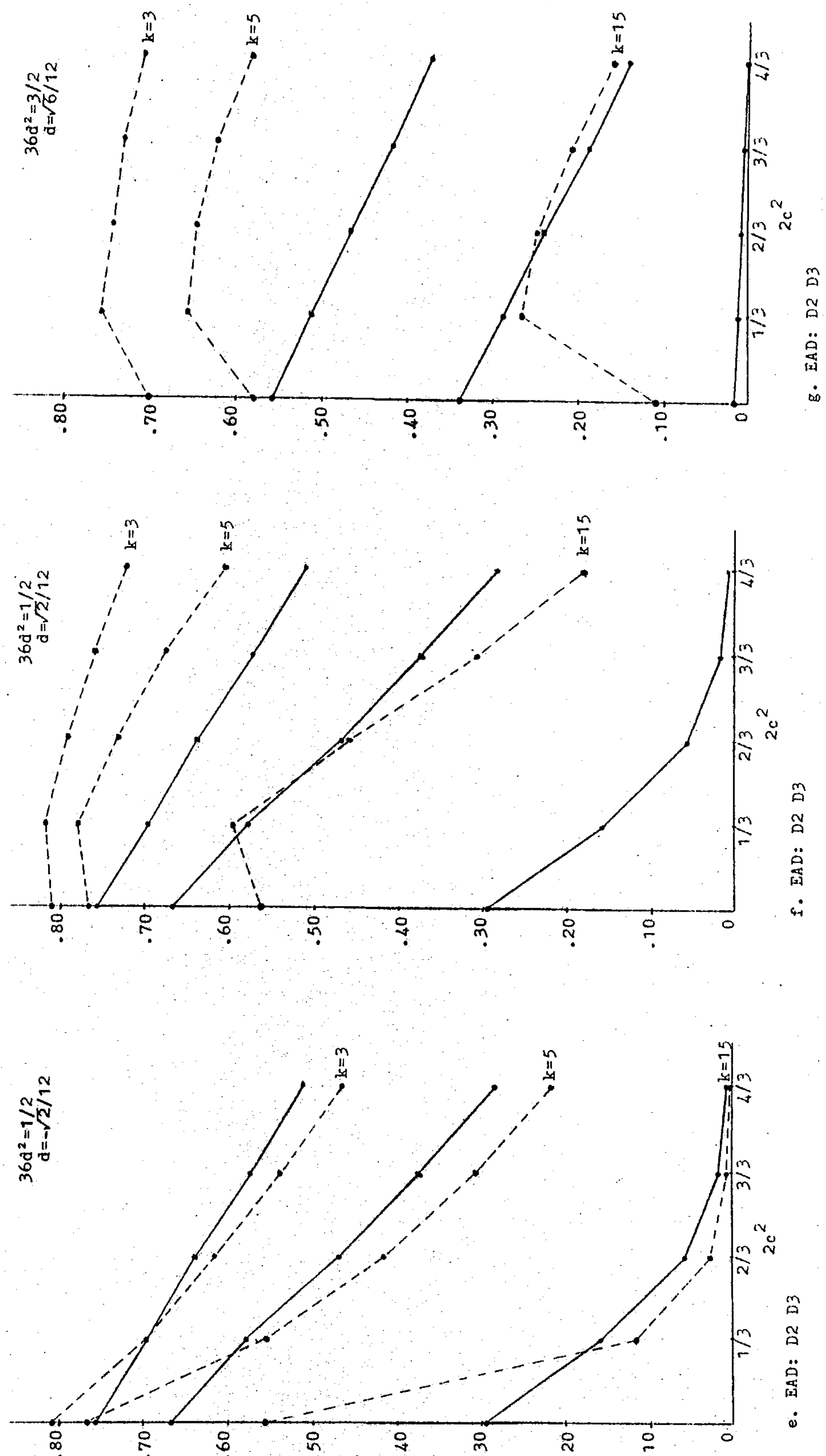
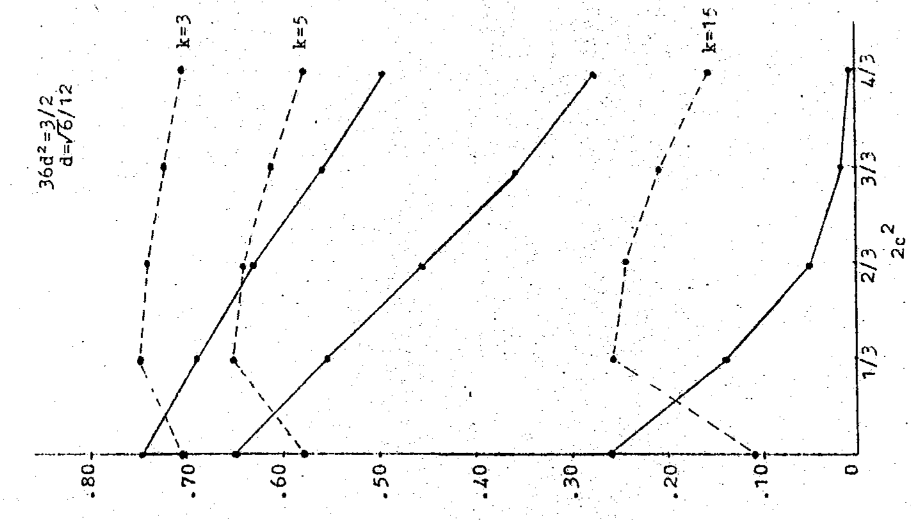
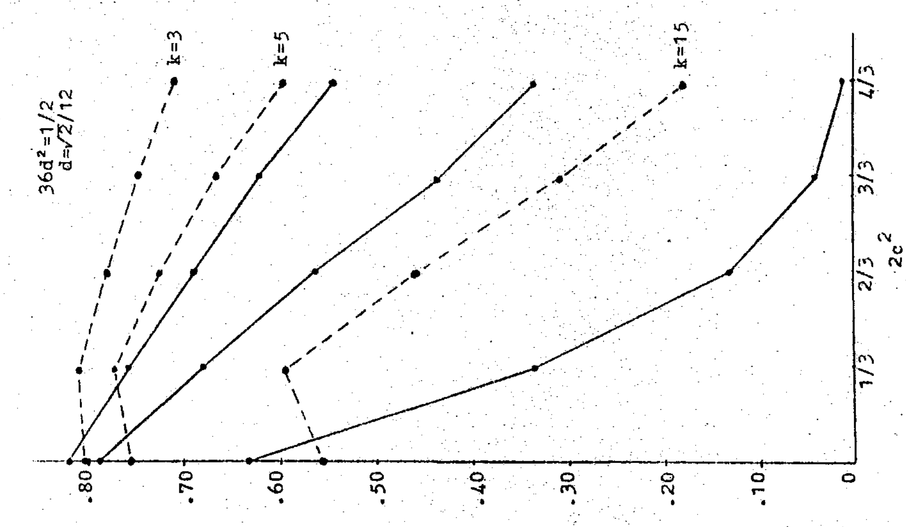


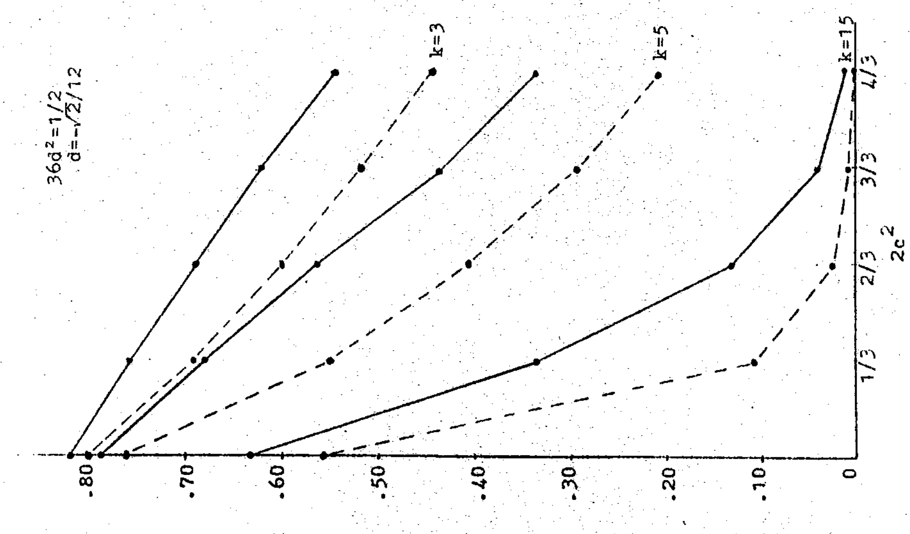
Figure 5.4 continued



j. Adjusted EAD: D2 D3 (Adjusted Independent)

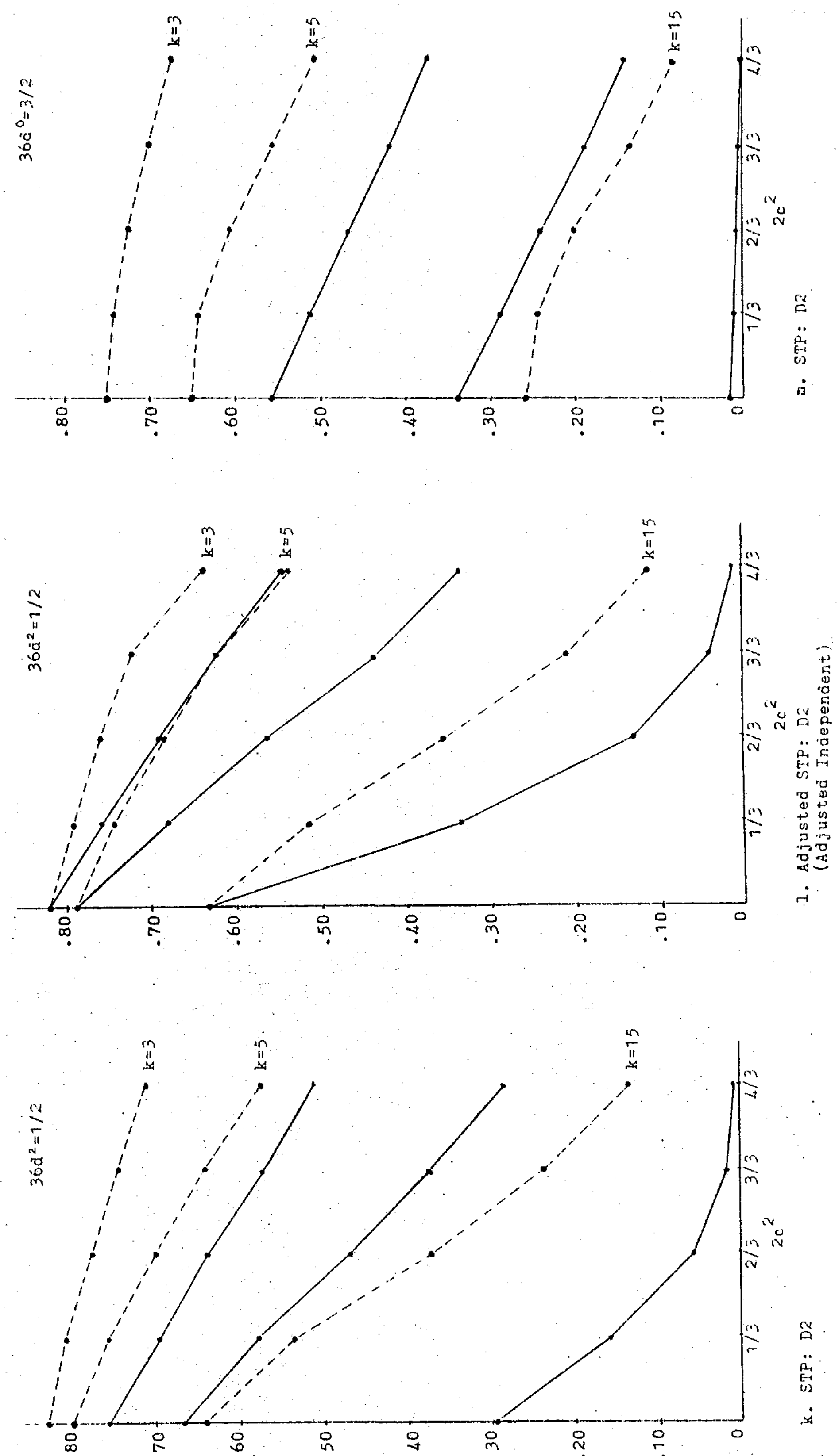


i. Adjusted EAD: D2 D3 (Adjusted Independent)



h. Adjusted EAD: D2 D3 (Adjusted Independent)

Figure 5.4 continued



m. STP: D2

l. Adjusted STP: D2 (Adjusted Independent)

k. STP: D2

Figure 5.4 continued

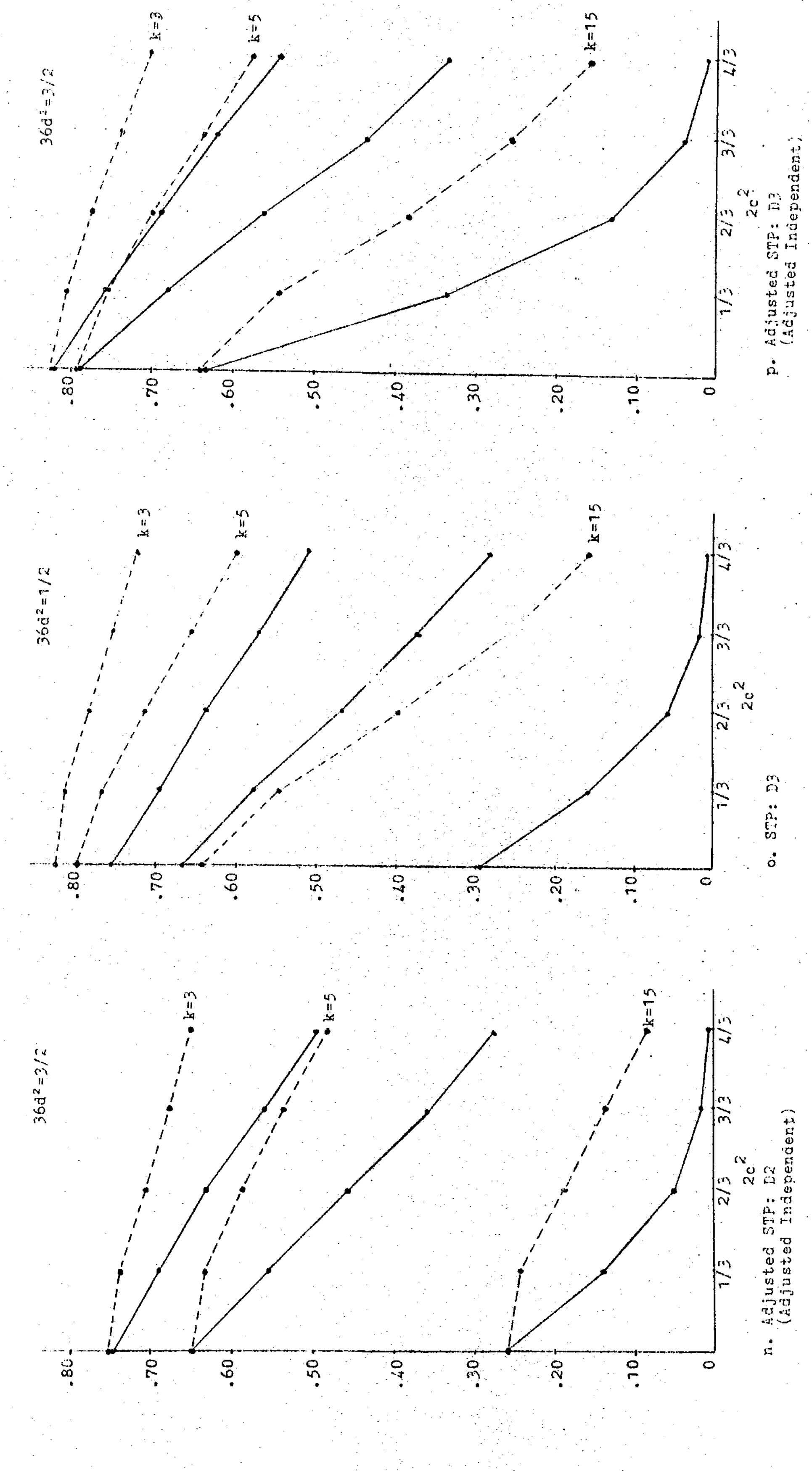
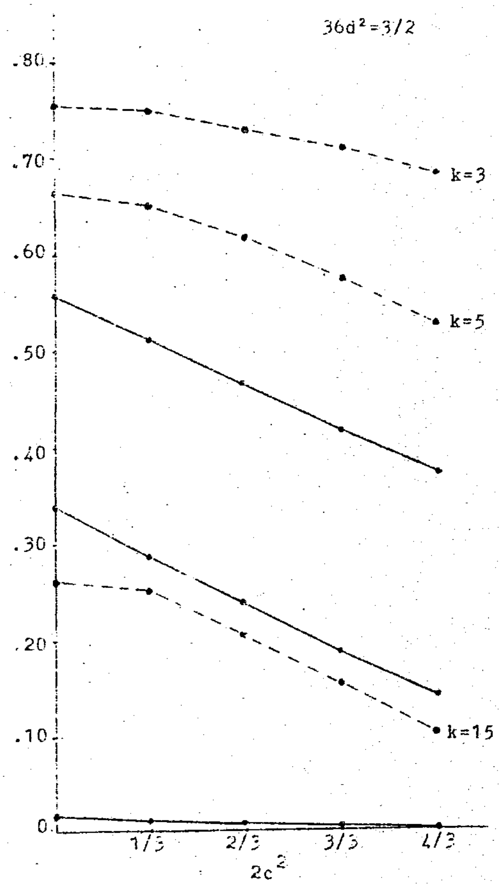
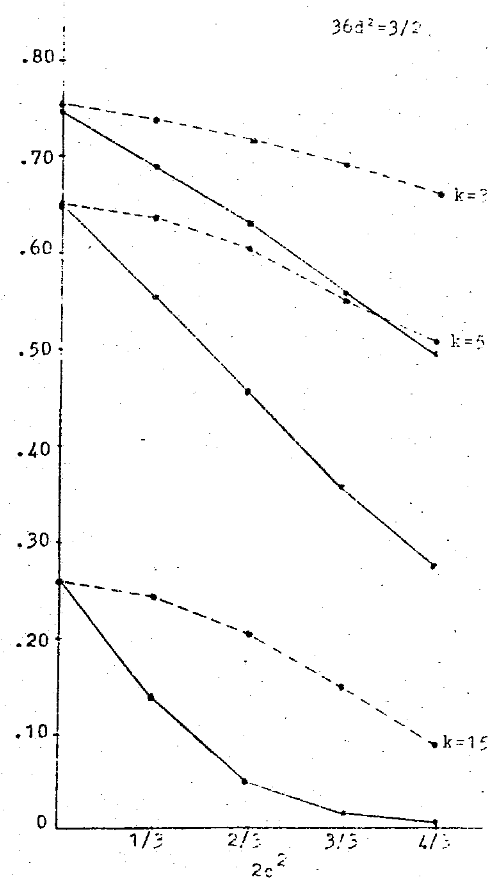


Figure 5.4 continued



q. STP: D3



r. Adjusted STP: D3
(Adjusted Independent)

With the balanced design, the type(2,2) error probabilities for the EAD and STP analyses are greater than the independent analysis probabilities. This difference decreases with increases in k, but it appears to remain about the same over values of the column noncentrality parameter. Increases in the interaction noncentrality parameter increase the differences between the EAD-STP and the independent analyses' error probabilities, at least for small k. Adjusting the critical region for the column test reduces the differences, but not dramatically.

EAD performs quite differently with an unbalanced design depending upon the magnitude and pattern of the interaction. With $d= -\sqrt{2}/12$ the EAD type(2,2) error probabilities are less than those of the independent analysis for sufficiently large values of the column noncentrality parameter. However when the direction is changed so that $d= \sqrt{2}/12$ (but the interaction noncentrality parameter is the same since d^2 is used in its formulation), the type(2,2) error probabilities are dramatically larger than the independent analysis error probabilities. These large differences still prevail when d is increased to $\sqrt{6}/12$. Adjusting the column test critical region and adjusting the interaction noncentrality parameter for the independent analysis does very little to change the differences, although they appear slightly less dramatic.

A key point about the EAD analysis that is illustrated by the graphs is that the error curves may not be monotonically decreasing in the column noncentrality parameter. In fact when d= $\sqrt{6}/12$, the EAD type(2,2) error is most probable for $2c^2$ equal to 1/3. This was predictable since the column noncentrality parameter for EAD is zero for this case (see Table 5.7). This results in the EAD test of H_{C1} being a biased test in the presence of interaction.

The STP type(2,2) error probabilities are consistently larger than the error probabilities for the independent analysis, and dramatically so. Although the differences are reduced when adjusted column test critical regions and adjusted interaction noncentrality parameters are used, the differences are still substantial. The effects of Designs 2 and 3 show very small differences.

Whether the EAD error probability exceeds or falls below the STP probability depends on the noncentrality parameters and the pattern of interaction. Generally EAD performs better than STP when d= $-\sqrt{2}/12$, but not as well as STP, except for small column noncentrality, for d= $\sqrt{2}/12$ or $\sqrt{6}/12$.

5.6 Summary

In the analysis of two-way, fixed effects, all-cells-filled designs where no difference in main column effects (H_{C1}) is the key hypothesis of interest, the procedure begins with a test for interaction (H_I), or an assumption of no interaction, and concludes with a test of H_{C1} only if H_I is not rejected (see Figure 5.5). The possible errors

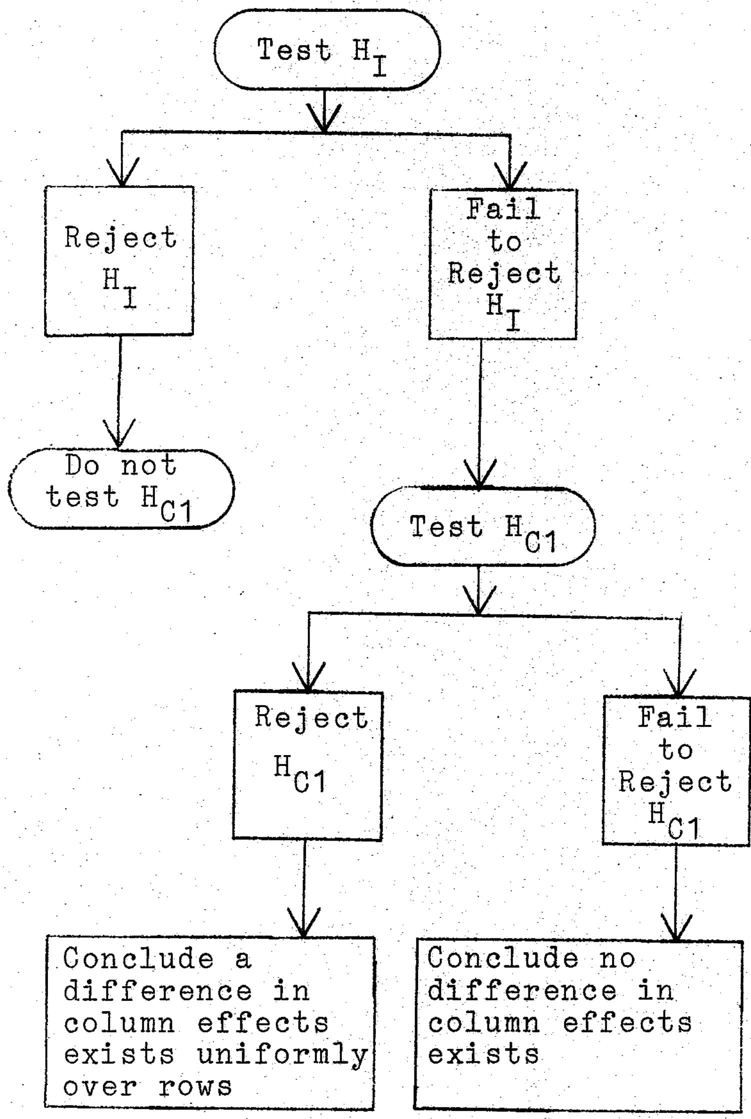


Figure 5.5 Procedure for testing H_I and H_{C1}

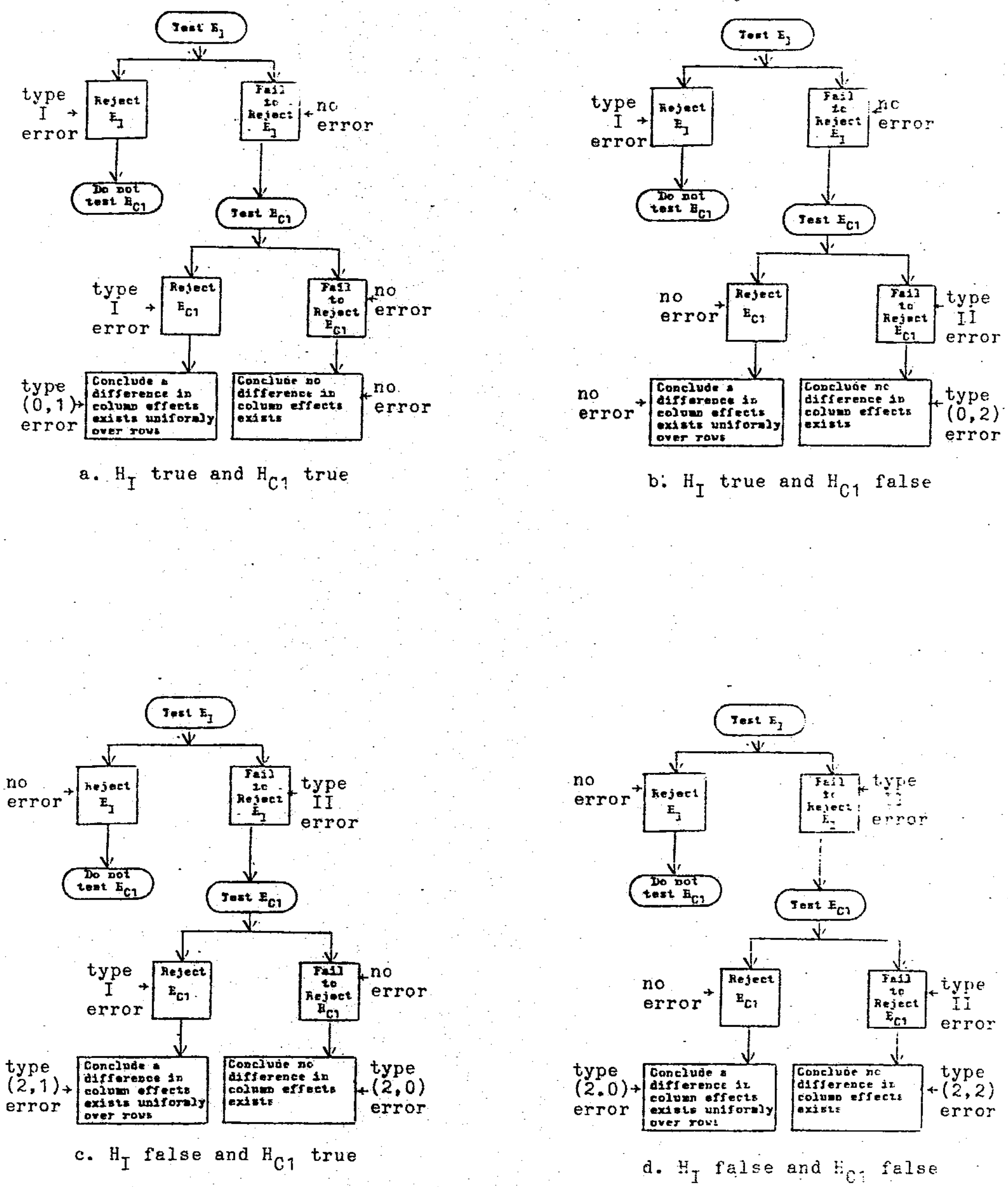


Figure 5.6 Possible Errors in the Testing Procedure

in this two step procedure depends on the truth of H_I and H_{C1} (see Table 1.1 and Figure 5.6).

Two analyses have been considered, EAD (each-adjusted-for-the-other) and STP (standard-parametric). A practioner must make a choice between at least these two analyses. For example, if one has the SAS output from the general linear models program, the choice is between the type 2 sums of squares (EAD) and the type 3 sums of squares (STP). How should one decide?

The answer to the posed question may be that one uses the presence or absence of interaction to decide. If interaction is present, then no choice needs to be made since the main effects test will not be done. If interaction is not present, then the choice is EAD, since EAD is more powerful than STP. Even when both analyses are adjusted so that their respective conditional alpha levels are the same, EAD is still more powerful than STP. This difference in power, or reduction in the type(0,2) error probability, is demonstrated in Figure 5.1. For example, in Figure 5.1(c) with $k=5$ and $6c^2=4/3$ (Design 2 and cell means pattern 1), the STP type(0,2) error probability is 0.40 compared to 0.38 for EAD (Figure 5.1(a)). When cell means pattern 2 is used instead (Design 2, $k=5$, and $2c^2=4/3$), the STP probability is 0.58 (Figure 5.1(g)) compared to 0.44 (Figure 5.1(e)) for EAD.

The fallacy of this "presence of interaction" strategy is that it is unusual to <u>know</u> whether or not interaction is present. Usually one only has evidence provided by an interaction test. It has been stated that there is almost always some interaction. If this be the case, then one should make the decision about which analysis to use based on how they compare in the presence of interaction.

In the presence of interaction, EAD can greatly exceed its assumed α-level. The EAD analysis, with an unbalanced design, "thinks" it sees a column effect even when there is none. This phantom effect is due only to interaction. Increases in sample size only enhance EAD's ability to detect the phantom effect. This is demonstrated in Figure 5.5. For example, in Figure 5.2(c) with k=5 and $36d^2=3/2$ (Design 2 or Design 3), the EAD type(2,1) error probability is 0.098. For Design 1 (balanced design) the probability is 0.013 (Figure 5.2(a)).

The STP type(2,1) error probabilities are also affected by the presence of interaction, but the probabilities generally decrease for increases in the interaction noncentrality and increases in the average cell size (for k=3, there may be some increase in the type(2,1) error probabilities for small interaction noncentrality). For example, the same one noted for EAD, in Figure 5.2(e) with k=5 and $36d^2=3/2$ (Design 2), the STP type(2,1) error probability is 0.027 compared to 0.013 for Design 1 (Figure 5.2(a)).

The power of the EAD analysis is unpredictable in the presence of interaction. In general, the EAD column test is biased in that the power curve may not be minimal under the null hypothesis. The magnitude of the EAD type(2,2) error probability depends heavily on the magnitude and pattern of interaction. For example, when $k=5$, $36d^2=1/2$, and $2c^2=1/3$, the EAD type(2,2) error probability for $d=-\sqrt{2}/12$ is 0.55 (Figure 5.4(h)) and for $d=\sqrt{2}/12$ the probability is 0.775 (Figure 5.4(i)). Note that 0.775 is larger than the value obtained for $2c^2=0$ with $d=\sqrt{2}/2$.

The STP column test is unbiased whether interaction is present or not. The STP type(2,2) error probability is affected by interaction, but not by the pattern of interaction. Whether STP has a smaller or larger type(2,2) error probability than EAD does depend however on the pattern and magnitude of interaction. The probabilities are consistently larger for STP when $d=-\sqrt{2}/12$ (compare Figure 5.4(h) with Figure 5.4(k)) and smaller for STP when $d=\sqrt{2}/12$, with the following exception (compare Figure 5.4(i) with Figure 5.4(k)): EAD has a smaller type(2,2) error probability for values of $2c^2$ close to zero, that is, EAD has better power for detecting minute differences in column effects.

The analysis procedure shown in Figure 5.5 could be modified for the EAD analysis in the presence of interaction by replacing H_{C1} with H': no interaction and no difference in main column effects. Failure to reject H' would lead one to the same conclusion of no difference in main column effects uniformly over rows. Rejection of H' would lead one to conclude that either:

i. no difference in the main column effects exists, but differences in cell means may exist within a row, or

ii. differences in the main column effects exist, but the differences may not be uniform over rows.

The EAD test of H' is in general still biased.

In summary, since the choice of analysis should be made with the awareness of possible interaction, STP should be used unless one <u>knows</u> that interaction is not present. The reasons for this choice are:

i. EAD is a biased test of H_{C1},

ii. the EAD conditional alpha level may exceed the assumed level , but the STP conditional alpha level cannot exceed the assumed level, and

iii. the power of EAD is unpredictable.

If indeed interaction is not present, then some power is lost by choosing STP over EAD with the magnitude of the loss depending on the degree of nonorthogonality of the STP column and interaction sums of squares. The "closer" the design is to being balanced, the smaller the loss in power.

Chapter 6

Suggestions for Further Research

Since interaction is inevitable in most experimental multifactor designs, one must be prepared to deal with it when testing for differences in main effects and upon interpretation of the analysis results. The material presented in Chapters 3 and 4 point out that the STP analysis can be used as a safeguard against interaction creeping into main effects. This does not provide an escape from interaction however when an interpretation of detected differences in main effects is presented. So what patterns of interaction, perhaps such that the power of a test for interaction is small, would be deemed practically insignificant? Should it be such that the differences in cell means are all in the same direction over rows or columns? A study of patterns of interaction that lead to these kinds of differences would be useful with respect to how the various analyses preform.

Even with a small amount of interaction, EAD could be severly affected. Suppose that $P(\mu:\hat{AB})$ lies in a direction which is highly nonorthogonal to $\hat{B}|\hat{A}$. The mean sum of squares for interaction may be small, but the sum of squares for columns under EAD may be larger than expected under H_{C1} due to the nonorthogonality. The same degree of interaction could be represented however by $P(\mu:\hat{A}B)$ lying in a direction that is orthogonal to $\hat{B}|\hat{A}$, if such a direction exists, thus leaving the column sum of squares unaffected. The practitioner would have no way of knowing which is the case since such knowledge would require knowledge of the interaction. A study of the 'worst case' interaction problem needs to be done so that for a particular unbalanced design, a potential EAD user may be forewarned of the degree of possible influence that undetected interaction could play on his main effects tests.

It has been oft suggested to pool, when the additive model is indicated, the insignificant interaction sum of squares into error (e.g. Rawlings 1972). How do the error probabilities for STP and EAD compare when this new error term is used? The column sum of squares for EAD will still be independent of the error sum of squares since $\hat{B}|\hat{A} \perp \tilde{AB}$, but for STP they may be dependent. In either case the new error sum of squares will be noncentral if some interaction is present.

Suppose the purpose of the factorial design and resulting analysis is to provide an estimate of β, the vector of cell means. An obvious unbiased estimate is the vector of sample cell means. This is what one obtains when the full model including interaction terms is used. When might it be more desirable to use the additive model to estimate β? Certainly the estimate would be more precise when there is in fact no interaction, but for what patterns of interaction would this continue to be the case? How does nonorthogonality of the design and overall sample size enter in? A measure of precision that may be used is $E||\hat{\beta}-\beta||^2$ where $\hat{\beta}$ is an estimate of β. Since β is the vector of cell means, this error measure is simply the sum of the mean squared errors of the individual mean estimators. The focus should be on the component of β contained in AB. It is conceivable that this component must lie in some convex subset of AB in order for the additive model to provide the more precise estimate.

The last suggestion refers to the development of a package of information that could be made available to the user of an ANOVA computer program that deals with unbalanced designs. Such possible items of interest might be:

i. The correlation coefficients of the various sums of squares under their respective null hypotheses;

ii. The observed conditional alpha level of any particular statistic given that interaction (and possibly other hypotheses) is not rejected at a user-specified level);

iii. The possible effect of interaction upon the EAD sum of squares (perhaps in terms of the maximum percentage of the interaction noncentrality parameter); and

iv. A more complete analysis of interaction broken up into logically selected directions so that one may determine where the interaction lies.

REFERENCES

Abramowitz, M. and I. A. Stegun (1964). *Handbook of Mathematical Functions*, National Bureau of Standards, Washington, DC.

Amos, D. E. and W. G. Bulgren (1972). Computation of a multivariate F distribution, *Mathematical Computing*, 26:255-264.

Appelbaum, M. I. and E. M. Cramer (1974). Some problems in the nonorthogonal analysis of variance, *Psychological Bulletin*, 81(6):335-343.

Bickel, P. J. and K. A. Doksum (1977). *Mathematical Statistics: Basic Ideas and Selected Topics*, Holden-Day, Inc., San Francisco, California.

Blumenson, L. E. and K. S. Miller (1963). Properties of generalized Rayleigh distributions, *The Annals of Mathematical Statistics*, 34:903-910.

Bose, S. S. (1935). On the distribution of the ratio of variances of two samples drawn from a given normal bivariate correlated population, *Sankhyā*, 2:65-72.

Bryce, G. R. (1975). Letters to the editor: the one-way model, *American Statistician*, 29(1):69-70.

Burdick, D. S., D. G. Herr, W. M. O'Fallon, and B. V. O'Neill, (1974). Exact methods in the unbalanced two-way analysis of variance - a geometric view, *Communications in Statistics*, 3:581-595.

Burdick, D. S. and D. G. Herr (1980). Counter-examples in unbalanced two-way analysis of variance, *Communications in Statistics*, A9(2):231-241.

Carlson, J. E. and K. H. Timm (1974). Analysis of non-orthogonal fixed-effects designs, *Psychological Bulletin*, 81:563-570.

D'jachenko, Z. N. (1961). On moments of bivariate (gamma) distribution, *Izvestiya Vysshykh Uchebnykh Zavedenii Matematika*, 1(20):55-65.

D'jachenko, Z. N. (1962a). On a form of bivariate (gamma) distribution, *Nauchnye Trudy Leningradskoi Lesotekhnicheskoi Akademii*, 94:5-17.

D'jachenko. Z. N. (1962b). Distribution surfaces of (gamma) type, *Trudy VI Vsesoyuznogo Soveshchaniya po Teorii Veroyatnostei i Mathematicheskoi Statistike*, Vilnjus, 389-395.

Dutt, J. E. and A. P. Soms (1976). An integral representation technique for calculating general multivariate probabilities with application to multivariate chi-squared, *Communications in Statistics*, A5:377-388.

Eisenhart, C. (1947). The assumptions underlying the analysis of variance, *Biometrics*, 3(1):1-21.

Finney, D. J. (1938). The joint distribution of variance ratios based on a common error mean square, *Annals of Eugenics, London*, 11:136-140.

Ghosh, M. N. (1955). Simultaneous tests of linear hypotheses, *Biometrika*, 42:441-449.

Gianola, D. (1975). Letters to the editor: hypothesis tests in linear models, *American Statistician*, 29(3):133.

Gunst, R. F. and J. T. Webster (1973a). Density functions of the bivariate chi-square distribution, *Journal of Statistical Computation Simulation*, 2:275-283.

Gunst, R. F. (1973b). On computing critical points for a bivariate chi-square random variable, *Communications in Statistics*, 2:221-224.

Gurland, J. (1948). Inversion formulae for the distribution of ratios, *The Annals of Mathematical Statistics*, 19:228-237.

Gurland, J. (1955). Distribution of definite and of indefinite quadratic forms, *The Annals of Mathematical Statistics*, 26:122-127.

Hardison, C. D., D. Quade, and R. D. Langston (1980). SAS functions to compute chi-squared, F, and T distributions and their inverses allowing for noncentrality and fractional degrees of freedom, *SAS Users Group Conference Proceedings 1980*, SAS Institute, Inc., North Carolina.

Herr, D. G. and J. Gaebelein (1978). Nonorthogonal two-way analysis of variance, *Psychological Bulletin*, 85:207-216.

Herr, D. G. (1980). On the history of the use of geometry in the general linear model, *American Statistician*, 34(1):43-47.

Hinkelmann, K. (1975). Letters to the editor: hypothesis testing in linear models, *American Statistician*, 29(2):110.

Hocking, R. R., F. M. Speed, and A. T. Coleman (1980). Hypotheses to be tested with unbalanced data, *Communications in Statistics*, A9(2):117-129.

Hurlburt, R. T. and K. K. Spiegel (1976). Dependence of F ratios sharing a common denominator mean square, *American Statistician*, 30(2):74-78.

Jensen, D. R. and M. Q. Jones (1969a). Simultaneous confidence intervals for variances, *Journal of the American Statistical Association*, 64:324-332.

Jensen, D. R. (1969b). An inequality for a class of bivariate chi-square distributions, *Journal of the American Statistical Association*, 64:333-336.

Jensen, D. R. (1969c). Limit properties of noncentral multivariate Rayleigh and chi-square distributions, *SIAM Journal of Applied Mathematics*, 17(4):802-814.

Jensen, D. R. (1970a). The joint distribution of quadratic forms and related distributions, *Australian Journal of Statistics*, 12(1):13-22.

Jensen, D. R. (1970b). A generalization of the multivariate Rayleigh distribution, *Sankhyā*, A32(2):193-208.

Joe, G. W. (1971). Comment on Overall and Spiegel's "Least squares analysis of experimental data", *Psychological Bulletin*, 75:364-366.

Johnson, R. E. (1978). On the joint distribution of row and column sums of squares in two-way designs, University of North Carolina at Greensboro MA Thesis.

Johnson, N. L. and S. Kotz (1972). *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley & Sons, Inc., New York.

Keppel, G. (1982). *Design and Analysis: A Researcher's Handbook*, Prentice-Hall, Inc., New Jersey.

Kibble, W. F. (1941). A two-variate gamma-type distribution, *Sankhyā*, 5:137-150.

Kimball, A. W. (1951). On dependent tests of significance in the analysis of variance, *The Annals of Mathematical Statistics*, 22:600-622.

Kleinbaum, D. G. and L. L. Kupper (1978). *Applied Regression Analysis and Other Multivariable Methods*, Wadsworth Publishing Co., Inc., California.

Krishnamoorthy, A. S. and M. Parthasarathy (1951). A multivariate gamma-type distribution, *The Annals of Mathematical Statistics*, 22:549-557.

Krishnaiah, P. R. (1964). On the simultaneous ANOVA and MANOVA tests, *Annals of the Institute of Statistical Mathematics*, 17:35-53.

Krishnaiah, P. R. and J. V. Armitage (1965). Probability integrals of the multivariate F distribution with tables and applications, ARL 65-236, *Aerospace Research Laboratories*, Wright-Patterson Air Force Base, Ohio.

Krishnaiah, P. R. and M. M. Rao (1961). Remarks on a multivariate gamma-type distribution, *The American Mathematics Monthly*, 68:342-346.

Krishnan, M. (1967). The noncentral bivariate chi-distribution, *SIAM Review*, 9:708-714.

Krishnan, M. (1976). The noncentral chi-squared distribution and extensions, *Communications in Statistics*, A5(7):647-660.

Kutner, M. H. (1974). Hypotheses testing in linear models (Eisenhart Model I), *American Statistician*, 28(3): 98-100.

Kutner, M. H. (1975). Letters to the editor: hypothesis tests in linear models, *American Statistician*, 29(3): 133-134.

Lancaster, H. O. (1958). The structure of bivariate distributions, *The Annals of Mathematical Statistics*, 29:719-736.

Lancaster, H. O. (1963). Correlations and canonical forms of bivariate distributions, *The Annals of Mathematical Statistics*, 34:532-538.

McAllister, P. R., R. Lee, and B. S. Holland (1981a). Computation of probabilities from Jensen's bivariate F distribution, *Communications in Statistics*, B10(3): 249-263.

McAllister, P. R. and B. S. Holland (1981b). Use of the experimentwise error rate in simultaneous tests of main effects in two-way disproportionate ANOVA without interaction, Technical Report #5, Temple University.

Miller, K. S., R. I. Bernstein, and L. E. Blumenson (1958). Generalized Rayleigh processes, *Quarterly of Applied Mathematics*, 16:137-145.

Miller, K. S. (1964). *Multidimensional Gaussian Distribution*, John Wiley & Sons, Inc., New York.

Miller, K. S. and H. Sackrowitz (1967). Relationships between biased and unbiased Rayleigh distributions, *SIAM Journal of Applied Mathematics*, 15:1490-1495.

Mood, A. M., F. A. Graybill, and D. C. Boes (1963). *Introduction to the Theory of Statistics*, McGraw-Hill, Inc., New York, New York.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, John Wiley & Sons, Inc., New York.

Nelder, J. A. (1975). Letters to the editor: hypothesis-testing in linear models, *American Statistician*, 30(2): 103.

Oberhettinger, F. (1973). *Fourier Transforms of Distributions and Their Inverses*, Academic Press, Inc., New York.

O'Brien, R. G. (1976). Comment on "Some problems in the nonorthogonal analysis of variance", *Psychological Bulletin*, 83:72-74.

Olkin, I. (1971). Monotonicity properties of Dirichlet integrals with applications to the multinominal distribution and the ANOVA test, Technical Report No. 137, Departments of Operations Research and Statistics, Stanford University.

Overall, J. E. and D. K. Spiegel (1969). Concerning least squares analysis of experimental data, *Psychological Bulletin*, 72:311-322.

Overall, J. E. and D. K. Spiegel (1973). Comments on Rawlings' nonorthogonal analysis of variance, *Psychological Bulletin*, 79:164-167.

Ralston, A. (1965). *A First Course in Numerical Analysis*, McGraw-Hill, Inc., New York.

Ramachandran, K. V. (1956). On the simultaneous analysis of variance test, *The Annals of Mathematical Statistics*, 27:521-528.

Rao, C. R. (1968). *Linear Statistical Inference and Its Applications*, John Wiley & Sons, Inc., New York.

Rawlings, R. R., Jr. (1972). Note on nonorthogonal analysis of variance, *Psychological Bulletin*, 77:373-374.

Rawlings, R. R., Jr. (1973). Comments on the Overall and Spiegel paper, *Psychological Bulletin*, 79:168-169.

Rudin, W. (1976). *Principles of Mathematical Analysis*, McGraw-Hill, Inc., New York.

Sarmanov, I. O. (1968). A generalized symmetric gamma correlation, *Soviet Mathematics Doklady*, 9(2):547-550.

Sarmanov, I. O. (1970a). Gamma correlation process and its properties, *Doklady Akademii Nauk, SSSR*, 191:30-32.

Sarmanov, I. O. (1970b). An approximate calculation of correlation coefficients between functions of dependent random variables, *Mathematical Notes, Academy of Sciences of the USSR*, 7:373-377.

SAS Institute (1979). *SAS User's Guide, 1979 Edition*, J. T. Helwig and K. A. Council editors, SAS Institute, Inc., North Carolina.

Scheffe´, H. (1959). *The Analysis of Variance*, John Wiley & Sons, Inc., New York.

Searle, S. R. (1971). *Linear Models*, John Wiley & Sons, Inc., New York.

Shao, T. S., T. C. Chen, and R. M. Frank (1964). Tables of zeros and Gaussian weights of certain associated Laguerre polynomials and the related generalized Hermite ploynomials, *Mathematical Computing*, 18:598-616.

Tan, W. Y. and S. P. Wong (1977). On the Roy-Tiku approximation to the distribution of sample variance from nonnormal universe, *Journal of the American Statistical Association*, 72:875-881.

Tan, W. Y. and S. P. Wong (1978). On approximating the central and noncentral multivariate gamma distributions, *Communication in Statistics*, B7:227-242.

Watson, G. (1962). *A Treatise on the Theory of Bessel Functions*, University Press, Cambridge.

Yates, F. (1934). The analysis of multiple classification with unequal numbers in different classes, *Journal of the American Statistical Association*, 29:51-66.

Appendix A

Proof of Orthogonality Conditions on η

Definition

For any $p \times r$ matrix U, let u_{i*} be the vector corresponding to the ith row of U for $i=1,\cdots,p$ and let u_{*j} be the vector corresponding to the jth column of U for $j=1,\cdots,r$. Then

$$U = \begin{pmatrix} u'_{1*} \\ \vdots \\ u'_{p*} \end{pmatrix} = (u_{*1},\cdots,u_{*r}).$$

Definition

Let $\sum_{j=1}^{r}\| u_{*r}$ denote the partitioning of U into its individual columns. We say that $U = \sum_{j=1}^{r}\| u_{*r}$. Note that $U = \sum_{j=1}^{r}\| (\sum_{i=1}^{p}\| u_{ij})'$ and $U' = \sum_{i=1}^{p}\| (\sum_{j=1}^{r}\| u_{ij})'$.

Definition (Kronecker product)

Let U be $p \times r$ and V be $q \times s$. The Kronecker product of U and V is

$$U \otimes V = \begin{pmatrix} u_{11}V & u_{12}V & \cdots & u_{1r}V \\ u_{21}V & u_{22}V & \cdots & u_{2r}V \\ \vdots & \vdots & \ddots & \vdots \\ u_{p1}V & u_{p2}V & \cdots & u_{pr}V \end{pmatrix}.$$

Some of the properties of the Kronecker product are listed below (Muirhead, 1982).

i. $(\alpha U)\otimes(\beta V) = \alpha\beta(U\otimes V)$ for any scalars α, β.

ii. If U and V are both $p\times r$ and W is $q\times s$, then
$(U+V)\otimes W = U\otimes W + V\otimes W$.

iii. $(U\otimes V)\otimes W = U\otimes(V\otimes W)$.

iv. $(U\otimes V)' = U'\otimes V'$.

v. $U\otimes V = \sum\|_{j=1}^{r}(u_{*j}\otimes V) = \sum\|_{j=1}^{r}\sum\|_{k=1}^{s}(u_{*j}\otimes v_{*k})$ for U $p\times r$ and V $q\times s$.

<u>Definition</u> (J-product)

Let U be $p\times r$ and V be $p\times s$. The J-product of U and V is
$U\odot V = (\sum\|_{i=1}^{p}(u_{i*}\otimes v_{i*}))'$.

Some of the important properties of the J-product are given in Theorem A.1 .

<u>Theorem A.1</u>

Let U be $p\times r$ and V be $p\times s$.

i. If $r=1$ or $s=1$, then $U\odot V = V\odot U$.

ii. $(\alpha U)\odot(\beta V) = \alpha\beta(U\odot V)$.

iii. $U\odot 1_p = 1_p\odot U = U$.

iv. $U\odot V = \sum\|_{j=1}^{r}\sum\|_{k=1}^{s}(u_{*j}\odot v_{*k}) = \sum\|_{j=1}^{r}(u_{*j}\odot V)$.

v. If U_2 is $p\times r$ and V_2 is $q\times s$, then
$(U+U_2)\odot V = (U\odot V) + (U_2\odot V)$ and $U\odot(V+V_2) = (U\odot V)+(U\odot V_2)$.

vi. Let C be $q\times c$ and D be $q\times d$. If $c=1$ or $s=1$, then
$(U\otimes C)\odot(V\otimes D) = (U\odot V)\otimes(C\odot D)$.

vii. $[(U\otimes C)\odot(V\otimes D)] = [(U\odot V)\otimes(C\odot D)]$.

Proof

i. Suppose r=1. Then $U \odot V = \left(\mathop{\sum\|}_{j=1}^{p} (u_{i1} \otimes v_{i*}) \right)'$

$$= \left(\mathop{\sum\|}_{i=1}^{p} (v_{i*} \otimes u_{i1}) \right)'$$

$$= V \odot U \ .$$

Suppose s=1. Then $V \odot U = U \odot V$.

ii. $(\alpha U) \odot (\beta V) = \left(\mathop{\sum\|}_{i=1}^{p} (\alpha u_{i*} \otimes \beta v_{i*}) \right)' = \alpha\beta \left(\mathop{\sum\|}_{i=1}^{p} (u_{i*} \otimes v_{i*}) \right)'$

$$= \alpha\beta (U \odot V) \ .$$

iii. $U \odot 1_p = 1_p \odot U$ from i. $U \odot 1_p = \left(\mathop{\sum\|}_{i=1}^{p} (u_{i*} \otimes 1) \right)' = \left(\mathop{\sum\|}_{i=1}^{p} u_{i*} \right)'$

$$= U.$$

iv. $(U \odot V)' = \mathop{\sum\|}_{i=1}^{p} (u_{i*} \otimes v_{i*}) = \mathop{\sum\|}_{i=1}^{p} \left(\mathop{\sum\|}_{j=1}^{r} \mathop{\sum\|}_{k=1}^{s} u_{ij} \otimes v_{ik} \right)'$, thus

$$U \odot V = \mathop{\sum\|}_{j=1}^{r} \mathop{\sum\|}_{k=1}^{s} \left(\mathop{\sum\|}_{i=1}^{p} u_{ij} \otimes v_{ik} \right)' = \mathop{\sum\|}_{j=1}^{r} \mathop{\sum\|}_{k=1}^{s} (u_{*j} \odot v_{*k}) \ .$$

Now using the above, $\mathop{\sum\|}_{k=1}^{s} (u_{*j} \odot v_{*k}) = u_{*j} \odot V$. Hence

$$U \odot V = \mathop{\sum\|}_{j=1}^{r} (u_{*j} \odot V) \ .$$

v. $(U+U_2) \odot V = \left(\mathop{\sum\|}_{i=1}^{p} (u_{i*} + u_{2i*}) \otimes v_{i*} \right)' = \left(\mathop{\sum\|}_{i=1}^{p} (u_{i*} \otimes v_{i*}) + \right.$

$$\left. (u_{2i*} \otimes v_{i*}) \right)' = \left(\mathop{\sum\|}_{i=1}^{p} (u_{i*} \otimes v_{i*}) \right)' +$$

$$\left(\mathop{\sum\|}_{i=1}^{p} (u_{2i*} \otimes v_{i*}) \right)' = (U \odot V) + (U_2 \odot V) \ .$$

The remainer of the proof of v is similar.

vi. Suppose c=1. Then $(U \otimes C) \odot (V \otimes D) =$

$$\left(\mathop{\sum\|}_{j=1}^{r} (u_{*j} \otimes c_{*1}) \right) \odot \left(\mathop{\sum\|}_{k=1}^{s} \mathop{\sum\|}_{l=1}^{d} (v_{*k} \otimes d_{*l}) \right) =$$

$$\mathop{\sum\|}_{j=1}^{r} \mathop{\sum\|}_{k=1}^{s} \mathop{\sum\|}_{l=1}^{d} (u_{*j} \otimes c_{*1}) \odot (v_{*k} \otimes d_{*l}) \ .$$

Now $(u_{*j} \otimes c_{*1}) \odot (v_{*k} \otimes d_{*l}) =$

$$\Big(\sum_{i=1}^{p}\| \sum_{i'=1}^{q}\| (u_{ij} \otimes c_{i'1} \otimes v_{ik} \otimes d_{i'l}) \Big)' =$$

$$\Big(\sum_{i=1}^{p}\| \sum_{i'=1}^{q}\| (u_{ij} \otimes v_{ik} \otimes c_{i'1} \otimes d_{i'l}) \Big)' =$$

$$\Big(\Big(\sum_{i=1}^{p}\| u_{ij} \otimes v_{ik} \Big) \otimes \Big(\sum_{i'=1}^{q}\| c_{i'1} \otimes d_{i'l} \Big) \Big)' =$$

$$\Big(\sum_{i=1}^{p}\| u_{ij} \otimes v_{ik} \Big)' \otimes \Big(\sum_{i'=1}^{q}\| c_{i'1} \otimes d_{i'l} \Big)' =$$

$$(u_{*j} \odot v_{*k}) \otimes (c_{*1} \odot d_{*l}) \; .$$

So $(U \otimes C) \odot (V \otimes D) = \sum_{j=1}^{r}\| \sum_{k=1}^{s}\| \sum_{l=1}^{d}\| (u_{*j} \odot v_{*k}) \otimes (c_{*1} \odot d_{*l}) =$

$$\sum_{j=1}^{r}\| \sum_{k=1}^{s}\| (u_{*j} \odot v_{*k}) \otimes \sum_{l=1}^{d}\| (c_{*1} \odot d_{*l}) = (U \odot V) \otimes (C \odot D) \; .$$

The proof is similar for $s=1$.

vii. From vi, $(U \otimes C) \odot (V \otimes D) = \sum_{j=1}^{r}\| \sum_{k=1}^{c}\| (u_{*j} \otimes c_{*k}) \odot (V \otimes D) =$

$$\sum_{j=1}^{r}\| \sum_{k=1}^{c}\| (u_{*j} \odot V) \otimes (c_{*l} \odot D) \; .$$

A different partitioning of the same vectors is:

$\sum_{j=1}^{r}\| (u_{*j} \odot V) \otimes \sum_{k=1}^{c}\| (c_{*k} \odot D) = (U \odot V) \otimes (C \odot D)$. Thus

$$[(U \otimes C) \odot (V \otimes D)] = [(U \odot V) \otimes (C \odot D)] \; .$$

Definition

For U p×r, $\mathrm{vec}(U) = \left(\mathop{\sum\|}_{j=1}^{r} \mathop{\sum\|}_{i=1}^{p} u_{ij} \right)'$. Some of the properties of the vec operator are listed below (Muirhead, 1982) .

i. $\mathrm{vec}(U'WV) = (U\otimes V)'\mathrm{vec}(W)$.

ii. $\mathrm{tr}(U'V) = \mathrm{vec}(U)'\mathrm{vec}(V)$.

iii. $\mathrm{vec}(\alpha U+\beta V) = \alpha\, \mathrm{vec}(U) + \beta\, \mathrm{vec}(V)$.

Theorem A.2

If $W = (w_{ij})$ is diagonal and $\eta = \left(\mathop{\sum\|}_{i=1}^{p} w_{ii} \right)'$, then $\mathrm{vec}(U'WV) = (U\otimes V)'\eta$.

Proof

Consider $\mathrm{vec}(W)'(U\otimes V) = \mathop{\sum\|}_{k=1}^{r} \mathop{\sum\|}_{l=1}^{s} \left(\mathrm{vec}(W)'(u_{*k}\otimes v_{*l}) \right)$.

$$\text{Now } \mathrm{vec}(W)'(u_{*k}\otimes v_{*l}) = \left(\mathop{\sum\|}_{j=1}^{p} \mathop{\sum\|}_{i=1}^{p} w_{ij} \right)\left(\mathop{\sum\|}_{j=1}^{p} \mathop{\sum\|}_{i=1}^{p} u_{jk}\otimes v_{il} \right)' =$$

$$\sum_{j=1}^{p} \sum_{i=1}^{p} w_{ij} u_{jk} v_{il} =$$

$$\sum_{i=1}^{p} w_{ii} u_{ik} v_{il} =$$

$$\left(\mathop{\sum\|}_{i=1}^{p} w_{ii} \right)\left(\mathop{\sum\|}_{i=1}^{p} u_{ik}\otimes v_{il} \right)' =$$

$$\eta'(u_{*k}\otimes v_{*l}) .$$

$$\text{So } \mathrm{vec}(W)'(U\otimes V) = \eta' \mathop{\sum\|}_{k=1}^{r} \mathop{\sum\|}_{l=1}^{s} u_{*k}\otimes v_{*l} =$$

$$\eta' (U\otimes V) .$$

Hence $\mathrm{vec}(U'WV) = (U\otimes V)'\eta$.

Definition

Let U and V be subspaces of $\mathbb{R}^p$ and let W be a subspace of $\mathbb{R}^q$. Let:

$U \otimes W = [\{u \otimes w : u \varepsilon U \text{ and } w \varepsilon W\}]$ and

$U \odot V = [\{u \odot v : u \varepsilon U \text{ and } v \varepsilon V\}]$.

Theorem A.3

If U is p×r, V is p×s, and W is q×t such that

$[U] = U$, $[V] = V$, and $[W] = W$, then

$U \otimes W = [U \otimes W]$ and $U \odot V = [U \odot V]$.

Proof

The proof follows directly from properties i and v of the Kronecker product and ii and v of the J-product.

Theorem A.4

Let $[C_a] = \mathbb{R}^a | [1_a]$ for $a \geq 2$. Then:

i. $[C_a \odot C_a] = \mathbb{R}^a$ if $a > 2$ and

ii. $[C_a \odot C_a] = [1_a]$ if $a = 2$.

Proof

Let ε_j be the vector in $\mathbb{R}^a$ with a 1 in the jth position and 0's elsewhere. Then let $C_a = \mathop{\sum\|}_{j=2}^{a} (\varepsilon_1 - \varepsilon_j)$.

$$C_a \odot C_a = \mathop{\sum\|}_{j=2}^{a} \mathop{\sum\|}_{k=2}^{a} (\varepsilon_1 - \varepsilon_j) \odot (\varepsilon_1 - \varepsilon_k) =$$

$$\mathop{\sum\|}_{j=2}^{a} \mathop{\sum\|}_{k=2}^{a} (\varepsilon_1 + \varepsilon_j \odot \varepsilon_k).$$

Now $\varepsilon_j \odot \varepsilon_k = \begin{cases} \varepsilon_j & \text{if } j=k \\ 0 & \text{if } j \neq k \end{cases}$. So if $a > 2$ then

$[C_a \odot C_a] = [(\varepsilon_1, \mathop{\sum\|}_{j=2}^{a} (\varepsilon_1 + \varepsilon_j))] = \mathbb{R}^a$ and if $a=2$,

$$[C_a \odot C_a] = [\varepsilon_1 + \varepsilon_2] = [1_a].$$

Theorem A.5

Let U be $p \times r$, V be $p \times s$, and W be $p \times p$. $U'WV = kU'V$ for $k = tr(W)/p$ if and only if $vec(W) \perp [U \otimes V] \,|\, [vec(I)]$.

Proof

First note that $vec(U'(W-kI)V) = (U \otimes V)' vec(W-kI)$.

Now $P(vec(W-kI):[U \otimes V]) =$

$P(vec(W):[U \otimes V] \,|\, [vec(I)]) - kP(vec(I):[U \otimes V] \,|\, [vec(I)])$

$+ P(vec(W):[vec(I)]) - kP(vec(I):[vec(I)])$.

Since $P(vec(I):[U \otimes V] \,|\, [vec(I)]) = 0$,

$$P(vec(I):[vec(I)]) = vec(I), \text{ and}$$

$$\begin{aligned} P(vec(W):[vec(I)]) &= (vec(I)' vec(W)/p) \cdot vec(I) \\ &= (tr(W)/p) \cdot vec(I) \ ; \end{aligned}$$

then $P(vec(W-kI):[U \otimes V]) = P(vec(W):[U \otimes V] \,|\, [vec(I)])$.

Hence $U'WV = kU'V \leftrightarrow U'(W-kI)V = 0 \leftrightarrow$

$$(U \otimes V)' vec(W-kI) = 0 \leftrightarrow vec(W) \perp [U \otimes V] \,|\, [vec(I)].$$

Corollary A.5

Suppose $W = (w_{ij})$ is diagonal and $\eta = \left(\sum_{i=1}^{p} \| \, w_{ii}\right)'$. Then $U'WV = kU'V$ where $k = tr(W)/p$ if and only if $\eta \perp [U \odot V] \,|\, J$.

Proof

The proof follows directly from the theorem upon noting that $(U \otimes V)' vec(W) = (U \odot V)' \eta$.

Let A, B, AB, and J be vector subspaces of $\mathbb{R}^{ab}$ such that $A = \mathbb{R}^a \otimes [1_b]$, $B = [1_a] \otimes \mathbb{R}^b$, $AB = [C_a \otimes C_b]$, and $J = [1_{ab}]$. Note that $A|J = [C_a \otimes 1_b]$ and $B|J = [1_a \otimes C_b]$.

Theorem A.6

Let W be diagonal and $\eta = (\sum_{i=1}^{p} \| w_{ii})'$. If the restrictions below on η are met for the various $[U]$ and $[V]$ spaces, then $U'WV = kU'V$ where $k = tr(W)/p$.

	[U]	[V]	Restrictions on η for $U'WV = kU'V$
i.	$A\|J$	$A\|J$	for $a>2$, $\eta \perp A\|J$
ii.	$B\|J$	$B\|J$	for $b>2$, $\eta \perp B\|J$
iii.	AB	AB	for $a>2$ and $b>2$, $\eta \perp \mathbb{R}^{ab}\|J$
			for $a>2$ and $b=2$, $\eta \perp A\|J$
			for $a=2$ and $b>2$, $\eta \perp B\|J$
iv.	$A\|J$	$B\|J$	$\eta \perp AB$
v.	$A\|J$	AB	for $a>2$, $\eta \perp (B\|J \oplus AB)$
			for $a=2$, $\eta \perp B\|J$
vi.	$B\|J$	AB	for $b>2$, $\eta \perp (A\|J \oplus AB)$
			for $b=2$, $\eta \perp A\|J$.

Proof

For each case it will suffice to show that $[U \odot V]|J$ is equal to the mentioned space orthogonal to η.

i. $[U \odot V] = [(C_a \otimes 1_b) \odot (C_a \otimes 1_b)] =$

$$[(C_a \odot C_a) \otimes (1_b \odot 1_b)] =$$

$$\begin{cases} \mathbb{R}^a \otimes [1_b] = A \text{ if } a>2 \\ [1_a] \otimes [1_b] = J \text{ if } a=2 \end{cases} .$$

Hence $[U \odot V] | J = \begin{cases} A|J \text{ if } a>2 \\ \text{null space if } a=2 \end{cases}$.

ii. $[U \odot V] = [(1_a \otimes C_b) \odot (1_a \otimes C_b)] =$

$[(1_a \odot 1_a) \otimes (C_b \odot C_b)] =$

$\begin{cases} [1_a] \otimes \P R^b = B \text{ if } b>2 \\ [1_a] \otimes [1_b] = J \text{ if } b=2 \end{cases}$.

Hence $[U \odot V] | J = \begin{cases} B|J \text{ if } b>2 \\ \text{null space if } b=2 \end{cases}$.

iii. $[U \odot V] = [(C_a \otimes C_b) \odot (C_a \otimes C_b)] =$

$[(C_a \odot C_a) \otimes (C_b \odot C_b)] =$

$\begin{cases} \P R^a \otimes \P R^b = \P R^{ab} \text{ if } a>2 \text{ and } b>2 \\ \P R^a \otimes [1_b] = A \text{ if } a>2 \text{ and } b=2 \\ [1_a] \otimes \P R^b = B \text{ if } a=2 \text{ and } b>2 \\ [1_a] \otimes [1_b] = J \text{ if } a=2 \text{ and } b=2 \end{cases}$.

Hence $[U \odot V] | J = \begin{cases} \P R^{ab} | J \text{ if } a>2 \text{ and } b>2 \\ A|J \quad \text{if } a>2 \text{ and } b=2 \\ B|J \quad \text{if } a=2 \text{ and } b>2 \\ \text{null space if } a=2 \text{ and } b=2 \end{cases}$.

iv. $[U \odot V] = [(C_a \otimes 1_b) \odot (1_a \otimes C_b)] =$

$[(C_a \odot 1_a) \otimes (1_b \odot C_b)] =$

$[C_a \otimes C_b] =$

AB .

Hence $[U \odot V] | J = AB$ (since $AB \perp J$).

v. $[U \odot V] = [(C_a \otimes 1_b) \odot (C_a \otimes C_b)] =$

$[(C_a \odot C_a) \otimes (1_b \odot C_B)] =$

$$\begin{cases} \mathbb{R}^a \otimes [C_b] = [1_a] \otimes [C_b] \oplus [C_a] \otimes [C_b] = B|J \oplus AB \text{ if } a>2 \\ [1_a] \otimes [C_b] = B|J \text{ if } a=2 \end{cases} .$$

Hence $[U \circledcirc V]|J = \begin{cases} B|J \oplus AB & \text{if } a>2 \\ B|J & \text{if } a=2 \end{cases} .$

vi. $[U \circledcirc V] = [(1_a \otimes C_b) \circledcirc (C_a \otimes C_b)] =$

$[(1_a \circledcirc C_a) \otimes (C_b \circledcirc C_b)] =$

$$\begin{cases} [C_a] \otimes \mathbb{R}^b = [C_a] \otimes [1_b] \oplus [C_a] \otimes [C_b] = A|J \oplus AB \\ \qquad \text{if } b>2 \\ [C_a] \otimes [1_b] = A|J \quad \text{if } b=2 \end{cases} .$$

Hence $[U \circledcirc V]|J = \begin{cases} A|J \oplus AB & \text{if } b>2 \\ A|J & \text{if } b=2 \end{cases} .$

Appendix B

Proof of Equation 4.23

We show here that:

$$\int_0^\infty \kappa_1(x)\kappa_2(x)dG(x;t) = \sum_{k=0}^{\infty}\sum_{k'=0}^{\infty}\int_0^\infty \alpha_1(x,k)\alpha_2(x,k')g(x)dx$$

where $\kappa_i(x) = \lim_{n\to\infty} \kappa_i(x,n)$;

$$\kappa_i(x,n) = \sum_{k=0}^{n} \alpha_i(x,k);$$

$$\alpha_i(x,k) = \exp(-\psi_i/2)(\psi_i/2)^k \exp(-x/2) \sum_{j=0}^{n_i} (x/2)^j/j!k!,$$

$$n_i = \nu_i/2 + k - 1 \text{ for } i=1,2;$$

$g(x) = G'(x;t)$; and

ν_i, ψ_i, t, G are as given in Chapter 4.

The following theorems will be called upon (Rudin, 1976).

Theorem B.1

Suppose K is compact, and:

i. $\{f_n\}$ is a sequence of continuous functions on K,

ii. $\{f_n\}$ converges pointwise to a continuous function f on K,

iii. $f_n(x) \geq f_{n+1}(x)$ for all $x \in K$, $n=1,2,3,\cdots$.

Then $f_n \to f$ uniformly on K.

Theorem B.2

If f is continuous on $[a,b]$ then f is Riemann integrable on $[a,b]$.

Theorem B.3 (A weak form of Legesque's dominated convergence theorem.)

Suppose g and f_n $(n=1,2,3,\cdots)$ are defined on $(0,\infty)$, are Riemann-integrable on $[a,b]$ whenever $0<a<b<\infty$, $|f_n|<g$, $f_n \to f$ uniformly on every compact subset of $(0,\infty)$ and $\int_0^\infty g(x)dx<\infty$; then $\lim_{n\to\infty} \int_0^\infty f_n(x)dx = \int_0^\infty f(x)dx$.

Let $f_n(x) = \kappa_1(x,n)\kappa_2(x)g(x)$. We will state without proof that $f_n(x)$ $(n=1,2,3,\cdots)$ and $g(x)$ are continuous on $(0,\infty)$. Then $f_n \to f = \kappa_1(x)\kappa_2(x)g(x)$ for every $x\varepsilon(0,\infty)$ and f is continuous on $(0,\infty)$. Now $f_{n+1}(x)-f_n(x) = \alpha_1(x,n+1) \geq 0$ for every $x\varepsilon(0,\infty)$. Hence for every compact set K on $(0,\infty)$ we have by Theorem B.1 that $f_n \to f$ uniformly on K.

Since f_n and g are continuous on $(0,\infty)$, Theorem B.2 gives us that they are Riemann integrable on $(0,\infty)$.

The summands $\alpha_1(x,k) \geq 0$, thus $\kappa_1(x,n) \geq 0$. Since $\kappa_1(x) = 1-G(x;\nu_1,\psi_1) < 1$, then $\lim_{n\to\infty} \kappa_1(x,n) = \kappa_1(x) < 1$ (where $G(x;\nu_1,\psi_1)$ defined in Chapter 4 is a cumulative probability distribution function). Hence $|f_n| < g$. Furthermore, $\int_0^\infty g(x)dx = G(x;t) < 1$.

We may now use Theorem B.3 to state:

$$\int_0^\infty \kappa_1(x)\kappa_2(x)g(x)dx = \lim_{n\to\infty}\int_0^\infty \kappa_1(x,n)\kappa_2(x)g(x)dx$$
$$= \lim_{n\to\infty}\int_0^\infty \sum_{k=0}^{n} \alpha_1(x,k)\kappa_2(x)g(x)dx$$

$$= \sum_{k=0}^{\infty} \int_0^{\infty} \alpha_1(x,k)\kappa_2(x)g(x)dx \ .$$

Redefine f_n now as $f_n(x) = \kappa_2(x,n)\alpha_1(x,k)g(x)$. By making analogous arguments as above, we may apply Theorem B.3 again to obtain

$$\sum_{k=0}^{\infty} \int_0^{\infty} \alpha_1(x,k)\kappa_2(x)g(x)dx = \sum_{k=0}^{\infty} \lim_{n\to\infty} \int \sum_{k'=0}^{n} \alpha_1(x,k)\alpha_2(x,k')$$

$$g(x)dx$$

$$= \sum_{k=0}^{\infty} \sum_{k'=0}^{\infty} \int_0^{\infty} \alpha_1(x,k)\alpha_2(x,k')$$

$$g(x)dx \ .$$

Appendix C

SAS Programs

C.1

```
* THIS PROGRAM CALCULATES PROBABILITIES OF SINGLY DEPENDENT BUT      ;
* POSSIBLY NONCENTRAL BIVARIATE F DISTRIBUTIONS                      ;

OPTIONS LS=80;
PROC MATRIX;
     G=J.(51,52); H=J.(51,51); L1=J.(51,1); L2=J.(51,1);

*************************************************************;
K=5;                         * AVERAGE CELL SIZE;
T=(K-(1))*9;                 * DENOMINATOR DEGREES OF FREEDOM;
NCP1=2.70;                   * COLUMN NONCENTRALITY PARAMETER;
NCP2=5.40;                   * INTERACTION NONCENTRALITY PARAMETER;
*************************************************************;

     F1=FINV(.95,2,T,0); F2=FINV(.90,4,T,0);


      Q=FPROB(F1,2,T,NCP1);
      P=FPROB(F2,4,T,NCP2);
      N=T#/2;
      C1=2#/T*F1;
      C2=4#/T*F2;
      A=C1+C2+1;
      C1=C1#/A;
      C2=C2#/A;

      G(1,1)=1;
      DO J1=1 TO 50;
         G(J1+1,1)=C1*(N+J1-(1))#/J1*G(J1,1);
      END;

      DO J1=0 TO 50;
         DO J2=1 TO 51;
            G(J1+1,J2+1)=C2*(N+J1+J2-(1))#/J2*G(J1+1,J2);
         END;
      END;

      H(1,1)=1+N*C2;
      DO K1=1 TO 50;
         H(K1+1,1)=H(K1,1)+G(K1+1,1)+G(K1+1,2);
      END;

      DO K2=1 TO 50;
         G0=0;
         DO K1=0 TO 50;
            G0=G0+G(K1+1,K2+2);
            H(K1+1,K2+1)=H(K1+1,K2)+G0;
         END;
      END;

      L1(1,)=1;
      L2(1,)=1;
      DO K1=1 TO 50;
         L1(K1+1,)=L1(K1,)*NCP1#/K1#/2;
         L2(K1+1,)=L2(K1,)*NCP2#/K1#/2;
      END;

      SUM=0;
      DO K1=1 TO 51;
         DO K2=1 TO 51;
            SUM=SUM+L1(K1,)*L2(K2,)*H(K1,K2);
         END;
      END;

      SUM=1-Q-EXP(-(NCP1+NCP2)#/2)#/(A**N) *SUM;
      SUM=P-SUM;
      ALL=NCP1||NCP2||CELL||SUM;
      NOTE NCP1 NCP2  CELL-SIZE    PROB;
      PRINT ALL;
```

C.2

```
* THIS PROGRAM CALCULATES JOINT PROBABILITIES OF DOUBLY DEPENDENT,;
* DOUBLY CENTRAL F DISTRIBUTIONS.;

OPTIONS LS=80;
PROC MATRIX;

***************************************************************;
T=18;                    * DENOMINATOR DEGREES OF FREEDOM;
E1=1#/3;                 * FIRST SQUARED CANONICAL CORRELATION;
E2=1#/15;                * SECOND SQUARED CANONICAL CORRELATION;
***************************************************************;

      B=J.(22,22);H=J.(22,23);MULT=J.(22,1);G=J.(22,23);

      F1=FINV(.95,2,T,0);      F2=FINV(.90,4,T,0);
       DO I=1 TO 22;
          B(I,1)=1;
          DO J=2 TO 22;
             B(I,J)=0;
          END;
       END;

       DO  I=1 TO 21;
          DO J=1 TO 21;
             B(I+1,J+1)=B(I,J)+B(I,J+1);
             IF B(I+1,J+1)=1 THEN GO TO KEEPON;
          END;
KEEPON:   END;

       PROB=0;

       N=T#/2;
       C1=2#/T*F1;
       C2=4#/T*F2;
       A=C1+C2+1;
       C1=C1#/A;
       C2=C2#/A;
       G(1,1)=1;
       DO J1=1 TO 21;
          G(J1+1,1)=C1*(N+J1-1)#/J1*G(J1,1);
       END;
       DO J1=0 TO 21;
          DO J2=1 TO 22;
             G(J1+1,J2+1)=C2*(N+J1+J2-1)#/J2*G(J1+1,J2);
          END;
       END;

       H(1,1)=1;
       DO K1=1 TO 21;
          H(K1+1,1)=H(K1,1)+G(K1+1,1);
       END;

       DO K2=1 TO 22;
          G0=0;
          DO K1=0 TO 21;
             G0=G0+G(K1+1,K2+1);
             H(K1+1,K2+1)=H(K1+1,K2)+G0;
          END;
       END;
       DO K1=1 TO 22;
          DO K2=1 TO 23;
             H(K1,K2)=H(K1,K2)#/A**N;
          END;

       END;

       DO K=0 TO 21;

       IF K = 0 THEN DO;
          FACTOR=1;
          GO TO DONE;
       END;
       IF E1 = 0 AND E2 = 0 THEN DO;
          FACTOR=0;
          GO TO DONE;
       END;
       MULT(1,)=1;
       MULT(2,)=.5;
       DO I=2 TO K;
          MULT(I+1,)=MULT(I,)*(I-(.5))#/I;
       END;
```

```
      IF E1*E2 = 0 THEN DO;
         FACTOR=MULT(K+1,)*(E1+E2)**K;
         GO TO DONE;
      END;
      FACTOR=0;
      DO I=0 TO K;
      FACTOR=FACTOR+MULT(I+1,)*MULT(K-I+1,)*E1**I*E2**(K-I);
      END;

DONE: INTER=0;
      DO M=0 TO K;
         DO N=0 TO K;
            INTER=INTER+(1-(2)*MOD(M+N,2))*H(M+1,N+2)
                  *B(K+1,M+1)*B(K+1,N+1);
         END;
      END;
      PROB=PROB + FACTOR.* INTER;
      END;

      PROB=1-FPROB(F1,2,T,0)-PROB;
      ALL=T||E1||E2||PROB;
      NOTE  DENOM-DF        E1           E2            PROB;
      PRINT ALL;
```

C.3

```
* THIS PROGRAM ESITMATES PROBABILITES , VIA A SIMULATION PROCEDURE,;
* OF BIVARIATE F DISTRIBUTIONS ASSOCIATED WITH THE STP ANALYSIS     ;

*****************************************************************;
ADJUSTED=1;          * 1=YES, 0=NO - ADJUST THE ALPHA LEVEL ? ;
PATTERN=1;           * CELL MEANS PATTERN;
K= 3 ;               * AVERAGE CELL SIZE;
P=0;                 * ECART DESIGN PARAMETER;
Q=.-2;               * ECART DESIGN PARAMETER;
C=0;                 * COLUMN EFFECT PARAMETER;
D=0;                 * INTERACTION PARAMETER;
*****************************************************************;

      MACRO BOB    %  * INSERT AN ASTERISK IN ORDER TO SUPPRESS
                        CHANGING ORIENTATION OF PROB STATEMENT;

      U=5+D//5-(2*D)//5+D//5-(2*D)//5+4*D//5-(2*D)//5+D//5-(2*D)//5+D;
      IF PATTERN=1 THEN U2=C//0//-C//C//0//-C//C//0//-C;
         ELSE U2=0//C//0//0//C//0//0//C//0;
         U=U+U2;
      D=1+Q+P//1-(2*Q)//1+Q-P//1-(2*Q)//1+4*Q//1-(2*Q)//
        1+Q-P//1-(2*Q)//1+Q+P;
      D=K@D;  U=D#U;  STP1=0;   STP2=0;   DF=(K-(1))*9;
      OBS=J.(9,1);
      C= 1 1 // -1 0 // 0 -1 ;
      B= J.(3,1)@C;
      AB=C@C;

      X=INV(DIAG(D));                              * INFORMATION MATRIX;

      B=X*B*INV(B'*X*B)*B'*X;               * COLUMN PROJECTION MATRIX;
      AB=X*AB*INV(AB'*X*AB)*AB'*X;     * INTERACTION PROJECTION MATRIX;

      NCP1=U'*B*U;                   * COLUMN NONCENTRALITY PARAMETER;
      NCP2=U'*AB*U;             * INTERACTION NONCENTRALITY PARAMETER;

         IF NCP1 < 0 THEN NCP1=0;
         IF NCP2 < 0 THEN NCP2=0;
IF ADJUSTED=0 THEN DO;
      F1= FINV(.95,2,DF,0);                   * COLUMN CRITICAL VALUE;
      F2= FINV(.90,4,DF,0);              * INTERACTION CRITICAL VALUE;
END;
```

```
IF ADJUSTED=1 THEN DO;
     IF P=0 AND Q=0 THEN CRITV= 3.386586 3.18931 3.050728;
     IF P=0 AND Q=-.2 THEN CRITV=3.242622 3.061562 2.935930;
     IF P=.5 THEN CRITV=3.229280 3.049529 2.925083;
     IF K=3 THEN KK=1; IF K=5 THEN KK=2; IF K=15 THEN KK=3;
     F1=CRITV(,KK);
     F2=FINV(.90,4,DF,0);
END;
     Q=1-FPROB(F1,2,DF,NCP1);
     P=FPROB(F2,4,DF,NCP2);


     DO I=1 TO 10000;                                    * THE SIMULATIONS;
        DO J=1 TO 9;
           OBS(J,)=NORMAL(0)*SQRT(D(J,)) + U(J,);
        END;


        MSE=CINV(UNIFORM(0),DF,0)#/DF;
        MSC=(OBS'*B*OBS)#/2;

        IF MSC < F1*MSE THEN GO TO SKIPIT;
           MSI=(OBS'*AB*OBS)#/4;
           IF MSI < F2*MSE THEN STP1=STP1+1;
           ELSE STP2=STP2+1;
SKIPIT: END;

      W1=.5;  W2=.5;
      PROB=Q#/2;
      PROB2=0;
      DO WHILE(ABS(PROB-PROB2) > .00005);
         PROB2=PROB;
         PROB= W1*STP1#/10000 + W2*(Q-STP2#/10000);
         W2=PROB#/Q;
         W1=1-W2;
      END;
      CONF_ERR= SQRT((W1+(2*Q-(1))*W2)*PROB -PROB**2 +
                 Q*(1-Q)*W2**2)#/SQRT(10000)*1.96;
BOR   PROB=P-PROB;

      NOTE NCP1   NCP2   DF   PROB;
      ALL=NCP1//NCP2//DF//PROB; ALL=ALL`;
      PRINT ALL;

      PRINT CONF_ERR;
```