KENDALL, ALLEN CLEMENT, JR. Angles Between Subspaces in Two-Way Designs. (1976)
Directed by: Dr. David G. Herr. Pp. 34

The purpose of this work is to examine the notion of near orthogonality of subspaces in two-way statistical designs. The cell mean model is introduced to facilitate the analysis. The hypotheses, H_r: No difference in main row effects, and H_c: No difference in main column effects, are proposed and the ANOVA is introduced as a method of testing these hypotheses. This analysis is explained in terms of the perpendicular projection of the data vector onto a space corresponding to violations of the hypothesis tested.

A design is balanced if all the cell sizes are equal. The design is unbalanced otherwise. For the two hypotheses mentioned above, when the design is unbalanced, several spaces may be chosen for the analyses. $\hat{A}|J$ and $\hat{B}|J$, corresponding to violations of H_r and H_c respectively are selected for examination in this work. When the design is balanced, $\hat{A}|J$ and $\hat{B}|J$ are orthogonal. When the design is unbalanced, $\hat{A}|J$ and $\hat{B}|J$ are, in general, not orthogonal.

Orthogonality and near orthogonality of $\hat{A}|J$ and $\hat{B}|J$ may be checked by calculating the angle, θ, between $\hat{A}|J$ and $\hat{B}|J$. For a 2×3 two-way design, a formula is developed for $\cos^2\theta$ in terms of the cell sizes of the design.

The ECART design is defined to be a two-way design in which all of the row totals are equal and all of the column totals are equal. If a design is formed by permuting the cell sizes of an ECART design, then the new design formed is said to be generated from that ECART design.

It is conjectured that θ for an ECART design is less than or equal to θ for any design generated from that ECART design.

In this work, the above conjecture is proved for 2×3 ECART designs in which two of the cell sizes are equal to the average cell size or in which the total number of observations is 120 or less.

ANGLES BETWEEN SUBSPACES
IN TWO-WAY DESIGNS

by

Allen C. Kendall, Jr.

A Thesis Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Greensboro
1976

Approved by

David G. Herr

Thesis Adviser

APPROVAL PAGE

This thesis has been approved by the following committee of the Faculty of the Graduate School at the University of North Carolina at Greensboro.

Thesis Adviser David G. Herr

Committee Members Richard B. Shaw
Michael Willett
William A. Powers

May 24, 1976
Date of Acceptance by Committee

ACKNOWLEDGMENT

The author wishes to express his gratitude and appreciation to Dr. David G. Herr for his time, patience, instruction and guidance. Appreciation is also extended to Dr. William A. Powers for his aid in programming and to Drs. Richard B. Sher, and Michael M. Willett for their aid in the preparation of this thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

NON-STATISTICAL INTRODUCTION

An experimenter is interested in the effect of several factors, such as age, sex, number of eggs eaten per day, and weight, on blood serum cholesterol. Define as a cell, all observations taken from people who have been identically classified with respect to all factors. In the above case, the observations are blood serum cholesterol counts. An observation may be thought of as an observed value of a random variable, and all of the observations in a cell may be modeled by random variables which are assumed to be independent and identically distributed with unknown mean and variance. The unknown mean and variance are, respectively, the cell mean and cell variance. Suppose one wishes to compare the mean, or average, amount of blood serum cholesterol of two or more cells to each other. For example, compare men over 30 years of age and under175 pounds who eat one egg per day with women over 30 years of age and under 175 pounds who eat one egg per day, and determine if the blood serum cholesterol count is different for these two cells. This kind of comparison is accomplished by an analysis of variance (ANOVA). An ANOVA is an analysis of the measurements in the various cells, for the purpose of determining if these observed differences represent actual differences between cell means or merely reflect random fluctuations in the data. In an ANOVA one is comparing the variability of the data in the cells for the purpose of testing hypotheses about cell means. It is important to note that one does not test hypotheses about cell variances using an ANOVA, per se.

The design of the experiment in the above example is called a four-way design. This is because the cholesterol counts are classified in four ways (i.e. by four factors), these being age, sex, weight, and number of eggs eaten per day. In general, an N-way design is a design in which the observations are classified in N ways (by N factors).

Each factor is partitioned into two or more subclassifications, called levels. For example, sex has two levels, male and female. Each measurement or observation is associated with one and only one of these two levels. Male and female, therefore, partition the set of observations with respect to sex. If a factor has 's' levels, then that factor is used to partition the set of observations into 's' disjoint groups.

Now, suppose one wishes to classify observations with respect to only two factors. For example, suppose in the cholesterol example one only wishes to classify the measurements with respect to age and number of eggs eaten per day. This would be a two-way design. Let the first factor, age, have two levels (30 and below, and over 30) and let the second factor, number of eggs eaten per day, have three levels (none, one, and two or more). This is called a 2 x 3 design. In general, if there are N factors $F_1, F_2, \cdots, F_N$ and n_i is the number of levels of F_i, for $i = 1,2,\cdots,N$, then the design is called an

$n_1 \times n_2 \times \ldots \times n_N$ design.

In the 2 x 3 cholesterol example described above, suppose there are 30 measurements in each level of the first factor. That is, 30 people in each age group are tested. Also, for each level of the first factor suppose there are 10 measurements of blood serum cholesterol taken at each level of the second factor. That is, for each age group 10 people are tested in each of the egg consumption groups. There are 60 measurements taken in all, 10 in each cell. Figure 1 shows the design.

Figure 1. 2 x 3 Design for Blood Serum Cholesterol Experiment.

	None	One	2 or more	Total in Age Group
30 or Under	10	10	10	30
Over 30	10	10	10	30
Total in Egg Group	20	20	20	60 Total Sample Size

Since the first factor divides the array in Figure 1 into rows, it is called the row factor. Similarly, the second factor is called the column factor. Since either factor can be used to divide the array into rows, this labeling is arbitrary. The number of observations in a cell is called the cell size. In Figure 1, notice that all of the cell sizes are the same. When this occurs, the design is called an equal cell size design, or balanced design. If a design is not balanced, then it is called an unbalanced design. One might wish to know if one (or both) of the factors has any effect on the measurements. In attempting to answer questions of this nature it is helpful to assume a mathematical model. One such model for an a x b design would denote by μ_{pq}, the average, or mean, of the population of the (p,q) cell. Then the $r^{\underline{th}}$ observation in the (p,q) cell, y_{pqr}, would be an observation of the random variable Y_{pqr}. This model is called the Cell Mean Model (CMM). A more formal definition of CMM appears in Chapter II.

By using an ANOVA, the following hypotheses may be tested:

H_r: No difference in main row effects;

H_c: No difference in main column effects.

The main row effect is that portion of an observation attributable strictly to the level of the row factor in which it is classified. The main column effect is that portion of an observation attributable strictly to the level of the column factor in which it is classified. In terms of the 2 x 3 blood serum cholesterol example above, H_r and H_c are:

H_r: No difference in that portion of the observations attributable strictly to age;

H_c: No difference in that portion of the observations attributable strictly to egg consumption.

When an a x b design is balanced, there is consensus among statisticians that the main row effects are the averages of the population means in each level of the row factor, the row means, $\mu_{p\cdot}$, for $p = 1,2,\cdots,a$, and the main column effects are the averages of the population means in each level of the column factor, the column means, $\mu_{\cdot q}$, for $q = 1,2,\cdots,b$. The cell means, the row means, and column means are parameters of the mathematical model for the design. In the above case parametric statements of the hypotheses H_r and H_c

are:

$$H_r\colon\ \mu_{1.} = \mu_{2.} = \cdots = \mu_{a.};$$

$$H_c\colon\ \mu_{.1} = \mu_{.2} = \cdots = \mu_{.b}.$$

When the design is unbalanced, however, there is no consensus as to how to define the main row and main column effects. This lack of consensus allows the main row and main column effects to be interpreted in several ways. Each interpretation yields slightly different parametric statements of the hypotheses H_r and H_c. There are at least five such sets of parametric hypotheses of particular interest, discussed in Chapter II. Some of these hypotheses are more easily interpretable, and therefore more desirable than others.

Assume N observations are made. Define the data vector to be the $N \times 1$ vector containing the N observations. The set of all possible data vectors is the sample space. Associated with each hypothesis, H, is a subspace, G, of the sample space, which corresponds to violations of H. In testing H by performing an ANOVA on data from a two-way design, the squared length of the perpendicular projection of the data vector onto G is compared to a stochastically independent estimate of the within-cell variability by means of an F-statistic. When testing H_r and H_c, it is especially convenient if the corresponding subspaces, G_r and G_c, are orthogonal. This prevents effects attributable to one hypothesis from being contaminated by effects attributable to the other hypothesis.

In the case of an unbalanced design, when the parametric hypotheses are easy to interpret, the corresponding subspaces are not, in general, orthogonal. Likewise, when the subspaces are orthogonal, the parametric hypotheses to which the subspaces correspond are not, in general, easy to interpret. Thus, any choice made to use one set of parametric hypotheses and corresponding subspaces will be a compromise.

In the case of a balanced 2×3 design, such as the design of Figure 1, the above mentioned five sets of parametric hypotheses are all mathematically equivalent and the corresponding five sets of spaces are identical. Also, not only are the parametric hypotheses H_r and H_c easy to interpret, but G_r and G_c are orthogonal, so that the ideal situation exists. The analysis in this case is very easy to perform, (Dixon and Massey [2], p.175).

If the design is unbalanced, the analysis may become considerably more difficult, (Burdick, Herr, O'Fallon, O'Neill [1]). Suppose, however, that for H_r and H_c having easily interpreted parametric statements, G_r and G_c are 'nearly' orthogonal. Then, when using the analysis corresponding to these parametric hypotheses, one may interpret the analysis as having main row and main column effects 'almost' uncontaminated by each other, i.e. one may speak of the two main effects as being uncontaminated by each other without being 'far' from truthful.

Consider an unbalanced design with an equal number of measurements in each row (level of the row factor), and an equal number of measurements in each column (level of the column factor). Call this an <u>Equal Column and Row Total design</u> (ECART <u>design</u>). An $a \times b$ design formed by permuting the cell sizes of another $a \times b$ design, D, is called a design generated from D.

One of the five analyses mentioned above uses the hypotheses,

H_r: The weighted means for each row are identical;

H_c: The weighted means for each column are identical.

Here the weighted mean for a row (column) is the average of the cell means for the row (column) with each cell mean weighted by the corresponding cell size. The above analysis is called the Weighted Mean (WTM) analysis. In this analysis the subspaces G_r and G_c are denoted $\hat{A}|J$ and $\hat{B}|J$ respectively. A way of checking for orthogonality of $\hat{A}|J$ and $\hat{B}|J$ is to calculate the angle, θ, between these two spaces. If $\theta = 90^{\circ}$, then the spaces are orthogonal. If θ is almost 90°, then the spaces are almost orthogonal. If it could be shown that θ_E, for an ECART design, is less than or equal to θ_g, for any design generated from that ECART design, then θ_E is the minimum of all the possible θ_g. Assume that $\theta_E \leq \theta_g$. Then if θ_E is close enough to 90° to interpret the analysis of the ECART design as if θ_E were 90°, the analysis of each design generated from that ECART design may be intepreted as if θ_g were 90° with at least the same degree of accuracy for the interpretation.

The conjecture discussed above has not yet been proved for arbitrary ECART designs, but will be shown to hold true for special cases of the ECART design.

CHAPTER II

STATISTICAL INTRODUCTION

Consider an experiment in which the observations collected are classified in two ways (a two-way design). Data thus classified are often arranged in a rectangular array with one factor, called the row factor, represented by the rows of the array, and the other factor, called the column factor, represented by the columns of the array. Figure 1 is a representation of such a design.

Let the design be an $a \times b$ design, and let y_{pqr} be the $r^{\underline{th}}$ observation in the (p,q) cell, $p = 1,2,...,a$; $q = 1,2,...,b$; and $r = 1,2,\cdots,n_{pq}$. The observation y_{pqr} may be thought of as a value of a random variable Y_{pqr} satisfying $Y_{pqr} = \mu_{pq} + e_{pqr}$. Here μ_{pq} is the mean of the population in the (p,q) cell, which is fixed but unknown, and e_{pqr} is the departure of Y_{pqr} from that mean. It is assumed that the e_{pqr} are random variables which are independent and normally distributed with mean zero and variance σ_{pq}^2. This means that the e_{pqr} are identically distributed for $r = 1,2,\cdots,n_{pq}$, and fixed p and q. It is further assumed that $\sigma_{pq}^2 = \sigma^2$ for $p = 1,2,\cdots,a$ and $q = 1,2,\cdots,b$. This further assumption means that all of the e_{pqr} are identically distributed. The Y_{pqr} are therefore assumed to be independent and normally distributed with mean μ_{pq} and variance σ^2. Denote the total number of observations by $n_{..} = \Sigma_{p,q}\, n_{pq}$.

This model may also be written in vector form as $Y = T\beta + e$ where Y, e, β, and T are defined as follows ($'$ denotes transpose):

(1) $Y = (Y_{111}, Y_{112}, \cdots, Y_{11n_{11}}, Y_{121}, Y_{122}, \cdots, Y_{12n_{12}}, \cdots, Y_{ab1}, Y_{ab2}, \cdots, Y_{abn_{ab}})'$,

(2) $e = (e_{111}, e_{112}, \cdots, e_{11n_{11}}, e_{121}, e_{122}, \cdots, e_{12n_{12}}, \cdots, e_{ab1}, e_{ab2}, \cdots, e_{abn_{ab}})'$,

(3) $\beta = (\mu_{11}, \mu_{12}, \cdots, \mu_{1b}, \mu_{21}, \mu_{22}, \cdots, \mu_{2b}, \cdots, \mu_{a1}, \mu_{a2}, \cdots, \mu_{ab})'$,

(4) T is the appropriate transformation matrix which equates the quantities above.

For a 2 x 3 design $T = \begin{bmatrix} j_{n_{11}} & 0_{n_{11}} & 0_{n_{11}} & 0_{n_{11}} & 0_{n_{11}} & 0_{n_{11}} \\ 0_{n_{12}} & j_{n_{12}} & 0_{n_{12}} & 0_{n_{12}} & 0_{n_{12}} & 0_{n_{12}} \\ 0_{n_{13}} & 0_{n_{13}} & j_{n_{13}} & 0_{n_{13}} & 0_{n_{13}} & 0_{n_{13}} \\ 0_{n_{21}} & 0_{n_{21}} & 0_{n_{21}} & j_{n_{21}} & 0_{n_{21}} & 0_{n_{21}} \\ 0_{n_{22}} & 0_{n_{22}} & 0_{n_{22}} & 0_{n_{22}} & j_{n_{22}} & 0_{n_{22}} \\ 0_{n_{23}} & 0_{n_{23}} & 0_{n_{23}} & 0_{n_{23}} & 0_{n_{23}} & j_{n_{23}} \end{bmatrix}$ where, for any positive integer, k, j_k is the $k \times 1$ vector of all ones and 0_k is the $k \times 1$ vector of all zeros.

Consider the 2 x 3 design of Figure 2. An observation, y, of the random vector Y is called the <u>data vector</u> and is an element of $R^{n..}$, the <u>sample space</u>. The range of the transformation matrix T, R_T, <u>is called the estimation space</u>. $R_T^{\perp} = R^{n..}|R_T$ is called the <u>error space</u>. The notation $U|V$ stands for the orthogonal complement of V in $U + V$, and is defined by $U + V = (U|V) \oplus V$ for U and V, subspaces of the same vector space.

Figure 2. Cell Sizes in a 2×3 Design.

		Column Factor			
		Level 1	Level 2	Level 3	
Row Factor	Level 1	n_{11}	n_{12}	n_{13}	$n_{1.}$
	Level 2	n_{21}	n_{22}	n_{23}	$n_{2.}$
		$n_{.1}$	$n_{.2}$	$n_{.3}$	$n_{..}$

$n_{.i} = \Sigma_p n_{pi}$; $n_{i.} = \Sigma_q n_{iq}$; $n_{..} = \Sigma_{p,q} n_{pq}$

In testing hypotheses about the data using an ANOVA, the measure of the within-cell variability mentioned in Chapter I is the squared length of the perpendicular projection of the data vector onto the error space, $R_T^{\perp}$. The subspace corresponding to violations of the hypotheses are subspaces of the estimation space, R_T.

As stated in Chapter I, when a two-way design is unbalanced, H_r and H_c may be interpreted in several ways. Each interpretation may yield different G_r and G_c, and therefore a different analysis. The five most common interpretations are parametrically presented in Table 1 for a 2×3 design, along with a symbol identifying the analysis, definitions of G_r and G_c, and statements as to whether, in general, G_r is orthogonal to G_c, G_r is orthogonal to G_i, and G_c is orthogonal to G_i. Orthogonality is important, because orthogonality of two G-spaces occurs if and only if the associated sums of squares for the corresponding two hypotheses are statistically independent as random variables.

Table 1. Orthogonality in Five Exact Analyses for 2 x 3 Two-Way Designs and Corresponding Parameteric Hypotheses (Herr [4]).

Analysis	G_r	G_c	$G_r \perp G_c$	$G_r \perp G_i$	$G_c \perp G_i$	H_r	H_c
STP	$\tilde{A} \mid J$	$\tilde{B} \mid J$	No	No	No	$\mu_{1.}=\mu_{2.}$	$\mu_{.1}=\mu_{.2}=\mu_{.3}$
EAD	$\hat{A} \mid \hat{B}$	$\hat{B} \mid \hat{A}$	No	Yes	Yes	$\Sigma_q n_{1q}(\mu_{1q}-\mu_{*q})=0$	$\Sigma_p n_{pk}(\mu_{pk}-\mu_{p*})=0; k=1,2$
HAB	$\hat{A} \mid J$	$\hat{B} \mid \hat{A}$	Yes	Yes	Yes	$\mu_{1*}=\mu_{2*}$	AS IN EAD
HBA	$\hat{A} \mid \hat{B}$	$\hat{B} \mid J$	Yes	Yes	Yes	AS IN EAD	$\mu_{*1}=\mu_{*2}=\mu_{*3}$
WTM	$\hat{A} \mid J$	$\hat{B} \mid J$	No	Yes	Yes	$\mu_{1*}=\mu_{2*}$	$\mu_{*1}=\mu_{*2}=\mu_{*3}$

$$\mu_{p.} = (1/3)\Sigma_{q=1}^{3} \mu_{pq};\ \mu_{.q}=(1/2)\Sigma_{p=1}^{2}\mu_{pq};\ \mu_{p*}=\Sigma_{q=1}^{3} \mu_{pq}\left(\frac{n_{pq}}{n_{p.}}\right);\ \mu_{*q}=\Sigma_{p=1}^{2} \mu_{pq}\left(\frac{n_{pq}}{n_{.q}}\right)$$

J_k denotes the space spanned by $\{j_k\}$. For the remainder of this work J will be used to denote $J_{n_{..}}$. A is the space in which β would lie if there were no column effects, and is called the <u>row space</u>. B is the space in which β would lie if there were no row effects, and is called the <u>column space</u>. A basis for A is $\{\begin{pmatrix} j_3 \\ 0_3 \end{pmatrix}, \begin{pmatrix} 0_3 \\ j_3 \end{pmatrix}\}$, and a basis for B is $\{(1,0,0,1,0,0)', (0,1,0,0,1,0)', (0,0,1,0,0,1)'\}$. These are subspaces of R^6, the <u>parameter space</u>. In general if S is a subspace of the parameter space, then let $\hat{S}$ be the subspace of the estimation space defined by $\hat{S} = \{v: v = Tu \text{ where } u \in S\}$ and let $\tilde{S}$ be the subspace of the estimation space defined by $\tilde{S} = \{v: v = T(T'T)^{-1}u$ where $u \in S\}$, where T is the transformation matrix (defined above for a 2×3 design). From this and the definition of $U|V$ the following spaces may be formed for the case of a 2×3 design:

$$\hat{A} = [\{ \begin{bmatrix} j_{n_{1.}} \\ 0_{n_{2.}} \end{bmatrix}, \begin{bmatrix} 0_{n_{1.}} \\ j_{n_{2.}} \end{bmatrix} \}], \text{ so } \hat{A}|J = [\{ \begin{bmatrix} (1/n_{1.})j_{n_{1.}} \\ (-1/n_{2.})j_{n_{2.}} \end{bmatrix} = a\}];$$

$$\hat{B} = [\{ \begin{bmatrix} j_{n_{11}} \\ 0_{n_{12}} \\ 0_{n_{13}} \\ j_{n_{21}} \\ 0_{n_{22}} \\ 0_{n_{23}} \end{bmatrix} = b_1, \begin{bmatrix} 0_{n_{11}} \\ j_{n_{12}} \\ 0_{n_{13}} \\ 0_{n_{21}} \\ j_{n_{22}} \\ 0_{n_{23}} \end{bmatrix} = b_2, \begin{bmatrix} 0_{n_{11}} \\ 0_{n_{12}} \\ j_{n_{13}} \\ 0_{n_{21}} \\ 0_{n_{22}} \\ j_{n_{23}} \end{bmatrix} = b_3\}], \text{ so}$$

$$\hat{B}|J = [\{\begin{bmatrix} (1/n_{.1})j_{n_{11}} \\ (-1/n_{.2})j_{n_{12}} \\ 0_{n_{13}} \\ (1/n_{.1})j_{n_{21}} \\ (-1/n_{.2})j_{n_{22}} \\ 0_{n_{23}} \end{bmatrix}, \begin{bmatrix} (1/n_{.1})j_{n_{11}} \\ (1/n_{.2})j_{n_{12}} \\ (-2/n_{.3})j_{n_{13}} \\ (1/n_{.1})j_{n_{21}} \\ (1/n_{.2})j_{n_{22}} \\ (-2/n_{.3})j_{n_{23}} \end{bmatrix}\}];$$

$$A|J = [\{\begin{pmatrix} j_3 \\ -j_3 \end{pmatrix}\}], \quad \text{so} \quad \tilde{A}|J = [\{\begin{bmatrix} (1/n_{11})j_{n_{11}} \\ (1/n_{12})j_{n_{12}} \\ (1/n_{13})j_{n_{13}} \\ (-1/n_{21})j_{n_{21}} \\ (-1/n_{22})j_{n_{22}} \\ (-1/n_{23})j_{n_{23}} \end{bmatrix}\}];$$

$$B|J = [\{\begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -2 \\ 1 \\ 1 \\ -2 \end{bmatrix}\}], \quad \text{so} \quad \tilde{B}|J = [\{\begin{bmatrix} (1/n_{11})j_{n_{11}} \\ (-1/n_{12})j_{n_{12}} \\ 0_{n_{13}} \\ (1/n_{21})j_{n_{21}} \\ (-1/n_{22})j_{n_{22}} \\ 0_{n_{23}} \end{bmatrix}, \begin{bmatrix} (1/n_{11})j_{n_{11}} \\ (1/n_{12})j_{n_{12}} \\ (-2/n_{13})j_{n_{13}} \\ (1/n_{21})j_{n_{21}} \\ (1/n_{22})j_{n_{22}} \\ (-2/n_{23})j_{n_{23}} \end{bmatrix}\}];$$

$$\hat{A}|\hat{B} = [\{ \begin{pmatrix} 1-(n_{11}/n_{.1})j_{n_{11}} \\ 1-(n_{12}/n_{.2})j_{n_{12}} \\ 1-(n_{13}/n_{.3})j_{n_{13}} \\ -(n_{11}/n_{.1})j_{n_{21}} \\ -(n_{12}/n_{.2})j_{n_{22}} \\ -(n_{13}/n_{.3})j_{n_{23}} \end{pmatrix} \}];$$

$$\hat{B}|\hat{A} = [\{ \begin{pmatrix} 1-(n_{11}/n_{1.})j_{n_{11}} \\ -(n_{11}/n_{1.})j_{n_{12}} \\ -(n_{11}/n_{1.})j_{n_{13}} \\ 1-(n_{21}/n_{2.})j_{n_{21}} \\ -(n_{21}/n_{2.})j_{n_{22}} \\ -(n_{21}/n_{2.})j_{n_{23}} \end{pmatrix}, \begin{pmatrix} -(n_{12}/n_{1.})j_{n_{11}} \\ 1-(n_{12}/n_{1.})j_{n_{12}} \\ -(n_{12}/n_{1.})j_{n_{13}} \\ -(n_{22}/n_{2.})j_{n_{21}} \\ 1-(n_{22}/n_{2.})j_{n_{22}} \\ -(n_{22}/n_{2.})j_{n_{23}} \end{pmatrix} \}].$$

Also notice $\hat{A} + \hat{B} = \hat{A} \oplus (\hat{B}|\hat{A})$, so $\hat{A} + \hat{B}$ is the span of the union of the bases for $\hat{A}$ and $\hat{B}|\hat{A}$. For a further discussion of the above spaces refer to Burdick, Herr, O'Fallon, and O'Neill [1]. Finally, define G_i to be $R_T|(\hat{A} + \hat{B})$, and call it the <u>interaction space</u>. G_i corresponds to violations of

H_i: <u>The row factor and the column factor do not interact</u>. Another statement of H_i which will tend to define interaction is

H_i: <u>The only effect due to the row factor and the column factor is the sum of the two separate effects</u>.

Thus the interaction effect of the $p^{\underline{th}}$ level of the row factor with the $q^{\underline{th}}$ level of the column factor on an observation in the (p,q) cell is the total effect due to the $p^{\underline{th}}$ level of the row factor and the $q^{\underline{th}}$ level of the column factor minus the sum of the two separate effects. A model is said to be additive if there is no interaction.

When a design is balanced, the five analyses are equivalent, so no choice must be made. Thus only unbalanced designs will be considered for the remainder of this work.

As stated in Chapter I, an analysis does not generally have both easily interpreted parametric hypotheses and orthogonality of the associated G-spaces. Here easy to interpret means intuitive. From Table 1, one may see that STP has easy to interpret parametric hypotheses, but is not extremely suitable for analysis because no orthogonality is guaranteed. EAD has orthogonality of G_r and G_c to G_i, but does not guarantee orthogonality of G_r and G_c. Also, notice that the parametric hypotheses are, in general, difficult to interpret. HAB and HBA have mutual orthogonality of G_r, G_c, and G_i guaranteed, but in each analysis, while one of the parametric hypotheses is easy to interpret, the other is the same as in EAD. WTM has the same quality of orthogonality as EAD; in fact, the angle between G_r and G_c is the same for the two analyses, (see Appendix A). But the parametric hypotheses for WTM are easier to interpret than those for EAD.

There are arguments for and against using each of these analyses. Consider the case in which the statistician feels that the best choice for his analysis would be WTM. He would then need to check the angle between $G_r = \hat{A}|J$ and $G_c = \hat{B}|J$ to see if the spaces are orthogonal. This angle, θ, may be obtained by first calculating $\cos^2\theta$ =

$$\frac{n_{1.}n_{2.}}{n_{..}}\left[\frac{1}{n_{.1}}\left(\frac{n_{11}}{n_{1.}}-\frac{n_{21}}{n_{2.}}\right)^2+\frac{1}{n_{.2}}\left(\frac{n_{12}}{n_{1.}}-\frac{n_{22}}{n_{2.}}\right)^2+\frac{1}{n_{.3}}\left(\frac{n_{13}}{n_{1.}}-\frac{n_{23}}{n_{2.}}\right)^2\right],$$

and then solving for θ. The proof of this formula appears in Appendix A.

If θ is close enough to 90° then one may speak of the main row effect and main column effect as being nearly uncontaminated by each other.

CHAPTER III

THEOREMS AND CONJECTURES

For integers a and b, define the least common multiple of a and b, denoted $LCM(a,b)$, to be the smallest positive integer divisible by a and b. Also define $n_{p.}$ and $n_{.q}$ as in Figure 2, for $p = 1,2,\cdots, a$ and $q = 1,2,\cdots,b$.

Theorem 1. *Given an* $a \times b$ ECART *design, then* $n_{..} = K\cdot LCM(a,b)$, *where* K *is a positive integer*.

Proof. Given an $a \times b$ ECART design, $n_{1.} = n_{2.} = \cdots = n_{a.}$ and $n_{1.} + n_{2.} + \cdots + n_{a.} = n_{..}$, therefore $n_{..} = a(n_{1.})$. Also $n_{.1} = n_{.2} = \cdots = n_{.b}$ and $n_{.1} + n_{.2} + \cdots + n_{.b} = n_{..}$, therefore $n_{..} = b(n_{.1})$. This implies that a and b both divide $n_{..}$; therefore, $LCM(a,b)$ divides $n_{..}$. Thus $n_{..} = K\cdot LCM(a,b)$ for some positive integer K.

Corollary 1. *Given a* 2×3 ECART *design*, $n_{..} = 6K$, *where* K *is a positive integer*.

Proof. $LCM(2,3) = 6$.

Consider an $a \times b$ ECART design with arbitrary (but fixed) a,b, and $n_{..}$. Since $n_{..} = an_{i.} = bn_{.j}$, for $i = 1,2,\cdots,a$ and $j = 1,2,\cdots,b$, then $n_{i.}$ and $n_{.j}$ are also fixed. Therefore, since $n_{i1} + n_{i2} + \cdots + n_{ib} = n_{i.}$, for $i = 1,2,\cdots,a$, once the first

$(b-1)$ cell sizes in the i^{th} row have been chosen, the last cell size in that row is determined. Likewise, since $n_{1j} + n_{2j} + \ldots + n_{aj} = n_{.j}$, for $j = 1,2,\cdots,b$, once the first $(a-1)$ cell sizes in the j^{th} column have been chosen, the last cell size in that column is determined. This implies that one is at liberty to choose only $(b-1)$ of the cell sizes in each of $(a-1)$ rows. One may, therefore, only choose $(a-1)(b-1)$ of the $a \cdot b$ cell sizes for an $a \times b$ ECART design. Notice that this implies that for a 2×3 ECART design, the first and second cell sizes in the first row, n_{11} and n_{12}, along with $n_{..}$, determine the entire design.

Now, a 2×3 ECART design has six cells, so $n_{..}/6$ is the average cell size (the average number of observations in each cell). From Corollary 1 note that for a 2×3 ECART design, $n_{..} = 6K$, for some positive integer K. This implies that the average cell size for a 2×3 ECART design is $n_{..}/6 = 6K/6 = K$, for some positive integer K. Each cell size must be $K + d$, where d is an integer such that $d \geq -K$, since a negative cell size is impossible. Let the first and second cell sizes in the first row be $K + u$ and $K + v$ respectively. Then the third cell size in the row is $K - (u + v)$ and the first, second, and third cell sizes in the second row are respectively $K - u$, $K - v$, and $K + (u + v)$. From the above one may observe that $-K \leq u \leq K$, $-K \leq v \leq K$, and $-K \leq u + v \leq K$. Figure 3 exhibits the form of a 2×3 ECART design.

Figure 3. Cell Sizes in a 2×3 ECART Design.

		Column Factor			
		Level 1	Level 2	Level 3	
Row Factor	Level 1	K + u	K + v	K-(u + v)	3K
	Level 2	K - u	K - v	K+(u + v)	3K
		2K	2K	2K	6K

Conjecture 1. The angle, θ_E, between $\hat{A}|J$ and $\hat{B}|J$, for an ECART design, is less than or equal to the angle, θ_g, between $\hat{A}|J$ and $\hat{B}|J$ for any design generated from that ECART design.

At this time no approach to the proof of this conjecture has been successful, nor has a counter example been discovered.

Conjecture 2. The angle, θ_E, between $\hat{A}|J$ and $\hat{B}|J$, for a 2×3 ECART design, is less than or equal to the angle, θ_g, between $\hat{A}|J$ and $\hat{B}|J$ for any design generated from that ECART design.

In the case of Conjecture 2, there are six cells in the ECART design, and so there are $6! = 720$ designs generated from a 2×3 ECART design. It is possible that each of these 720 designs yields a different $\hat{A}|J$ and $\hat{B}|J$, but many of these designs have identical θ_g. Notice from the formula in Chapter II for $\cos^2\theta$ for a 2×3 design, that if the rows of the design (array) are interchanged, $\cos^2\theta$ does not change. This implies that while there are 720 different designs, there are at most 360 different associated angles. For this reason, only 360 of the 720 designs must be examined. One may also notice from the formula in Chapter II for a 2 x 3 design, that if any of the columns of the design (array) are rearranged, $\cos^2\theta$ is unchanged. Since there are three columns, there are 3! = 6 ways that the columns may be rearranged. Therefore, of the 360 designs which may need to be examined, only 60 of them can yield different θ_g. For this reason, one must only examine 60 of the original 720 designs. These designs shall be called the essentially different designs.

From the above, one may observe that, given any 2 x 3 ECART design, by interchanging rows and interchanging columns, an ECART design may be obtained which yields the same angle, θ_g, as the original ECART design, and in which the u and v (in the notation of Figure 3) are both nonnegative integers. For this reason, without loss of generality, assume that u and v are nonnegative integers.

As with Conjecture 1, Conjecture 2 has not been proved or disproved to date. When either of u or v is zero, it can be shown that there are only 11 essentially different designs generated. Under these conditions the conclusion of Conjecture 2 may be obtained.

Theorem 2. For a 2 x 3 ECART design of average cell size K, if u or v is zero, then the angle, θ_E, between $\hat{A}|J$ and $\hat{B}|J$ is less than or equal to the angle, θ_g, between $\hat{A}|J$ and $\hat{B}|J$ for any design generated from that ECART design.

Proof. $\theta_E \leq \theta_g$ if and only if $\cos^2\theta_E - \cos^2\theta_g \geq 0$. The proof of the theorem for only one of the 11 cases will be shown. The other 10 cases have similar proofs. Consider an ECART design of the form

K + u	K	K − u
K − u	K	K + u

and let

K + u	K + u	K − u
K − u	K	K

represent a design generated from it, (note that $v = 0$). $\mathrm{Cos}^2\theta_E - \cos^2\theta_g =$

$$\frac{2u^2}{3K^2} - \frac{9K^2-u^2}{6K}\left[\frac{1}{2K}\left(\frac{K+u}{3K+u} - \frac{K-u}{3K-u}\right)^2 + \frac{1}{2K+u}\left(\frac{K+u}{3K+u} - \frac{K}{3K+u}\right)^2 + \frac{1}{2K-u}\left(\frac{K-u}{2K-u} - \frac{K}{3K-u}\right)^2\right].$$

Now since $K - u \geq 0$ and $u \geq 0$, then $u = xK$, where $0 \leq x \leq 1$. Substituting xK for u, $\cos^2\theta_E - \cos^2\theta_g =$

$$\frac{2x^2}{3} - \frac{9-x^2}{6}\left[\frac{1}{2}\left(\frac{1+x}{3+x} - \frac{1-x}{3-x}\right)^2 + \frac{1}{2+x}\left(\frac{1+x}{3+x} - \frac{x}{3-x}\right)^2 + \frac{1}{2-x}\left(\frac{1-x}{3+x} - \frac{x}{3-x}\right)^2\right].$$

$0 \leq \cos^2\theta_E - \cos^2\theta_g$ if and only if $0 \leq (\cos^2\theta_E - \cos^2\theta_g)6/(9-x^2) = 4x^2(9-x^2) - (4x^2/2 + (3x-x^2)^2/(2+x) + (x^2-5x)^2/(2-x))$. This holds if and only if $0 \leq (36x^2-4x^4)(4-x^2) - 8x^2(4-x^2) - (3x-x^2)^2(2-x) - (x^2-5x)^2(2+x) = 44x^2 + 16x^3 - 44x^4 + 4x^6$. Now, since $0 \leq x \leq 1$, $44x^2 \geq 44x^4$, so $44x^2 + 16x^3 - 44x^4 + 4x^6 \geq 16x^3 + 4x^6 \geq 0$. If $u = 0$ (instead of $v = 0$) one may interchange the first two columns of the ECART design and rename v to be u and u to be v. The above proof then applies.

The procedure in the above proof is too cumbersome to use to prove Conjecture 2. The conjecture may be provable by such a method, but calculations are unwieldy. So far only 11 of the 60 essentially different designs have been successfully shown to produce an angle between $\hat{A}|J$ and $\hat{B}|J$ not less than for the ECART design from which they were generated, but no counter example to Conjecture 2 has been found.

Notice that for a 2 x 3 ECART design (where K is fixed), since $u + v \leq K$ and u and v are both positive integers, (The case where u or v is zero need not be considered, because it has already been considered in Theorem 2), there are only finitely many ordered pairs (u,v), each associated with a different ECART design of average cell size K. If K is known, then each possible ECART design of average cell size K may be found. The above approach (checking $\cos^2\theta_E - \cos^2\theta_g$ to see if it is nonnegative for each essentially different design) may be used for each of these ECART designs.

Theorem 3. <u>For a</u> 2 x 3 ECART <u>design of average cell size</u> K, <u>the angle</u>, θ_E, <u>between</u> $\hat{A}|J$ <u>and</u> $\hat{B}|J$ <u>is less than or equal to the angle</u>, θ_g, <u>between</u> $\hat{A}|J$ <u>and</u> $\hat{B}|J$ <u>for any design generated from that</u> ECART <u>design if</u> $1 \leq K \leq 20$.

Proof. Given the hypothesis of Theorem 3, where $u + v \leq K$ and u and v are positive integers, (If either of u or v is zero, Theorem 3 is equivalent to Theorem 2 and is therefore proved), there are only a finite number of ordered pairs (u,v), for a fixed K. Each ordered pair is associated with an ECART design and 60 essentially different designs generated from that ECART design. Of this finite number of ordered pairs, not all need to be considered. The ECART design defined by (u,v) is simply the ECART design defined by (v,u), with the first two columns interchanged. For this reason, θ_E and the 60 θ_g associated with (u,v) are identical to the θ_E and the 60 θ_g associated with (v,u).

A computer program has been written which, given K, finds each ordered pair (u,v) which needs to be considered, and checks to see if the ECART design associated with (u,v) has the angle between $\hat{A}|J$ and $\hat{B}|J$ less than or equal to the corresponding angle for any design generated from that ECART design by actually calculating θ_E and all of the θ_g. This program appears in Appendix B along with some sample output. For $K = 1,2,\cdots,20$, θ_E is always less than or equal to the associated θ_g, and so Theorem 3 is proved.

The program used for the proof of Theorem 3 might be used to prove an identical theorem for any possible K. However, the length of the output and the cost makes this impractical for large values of K (For $K = 20$ the run cost $24.27 and there was 726 pages of output). This program was used to test several randomly chosen designs, where $k > 20$, and no counter examples were found. Therefore, while Conjecture 2 is to date unproved, the above indicates that it may be true.

The reader may have noticed that Conjectures 1 and 2 are concerned more with the permutations of the cell sizes than with the magnitudes of the cell sizes. It is helpful, in considering this idea, to introduce permutation matrices. A <u>permutation matrix</u> is an identity matrix with its rows or columns permuted. An $a \times b$ ECART design may be represented by an $ab \times 1$ vector of the cell sizes, Z, and permuting Z by an $ab \times ab$ permutation matrix, P_g, will produce a new vector, $Z_g = P_g Z$, associated with a design generated from the ECART design. Furthermore, for any design generated from the ECART design, there is some permutation matrix, P_g, where $Z_g = P_g Z$ is associated with the generated design. Let $(n_{11}, n_{12}, n_{13}, n_{21}, n_{22}, n_{23})' = Z_g$ be associated with a design generated from a 2×3 ECART design. For Z, associated with the ECART design, there is some permutation matrix, P, such that $Z_g = PZ$. For θ, the angle between $\hat{A}|J$ and $\hat{B}|J$ for this generated design, $\cos^2\theta$ may be written, in terms of the permutation matrix as, $\cos^2\theta = \dfrac{(a_1'PZ)(a_2'PZ)}{j_6'PZ}\left[\dfrac{1}{b_1'PZ}\left(\dfrac{e_1'PZ}{a_1'PZ} - \dfrac{e_4'PZ}{a_2'PZ}\right)^2 + \dfrac{1}{b_2'PZ}\left(\dfrac{e_2'PZ}{a_1'PZ} - \dfrac{e_5'PZ}{a_2'PZ}\right)^2 + \dfrac{1}{b_3'PZ}\left(\dfrac{e_3'PZ}{a_1'PZ} - \dfrac{e_6'PZ}{a_2'PZ}\right)^2\right]$, where $\{e_i\}_{i=1}^6$ is the natural basis for R^6, $a_1 = e_1 + e_2 + e_3$, $a_2 = e_4 + e_5 + e_6$, $b_1 = e_1 + e_4$, $b_2 = e_2 + e_5$, and $b_3 = e_3 + e_6$. Notice that in this formula, P appears to be the only variable. Thus it seems that the position of the cell sizes are more important than their magnitudes.

As an example, consider the following designs:

7	6	2
3	4	8

and

8	3	7
6	2	4

.

For the design on the left, an ECART design, $\theta = 64.40^\circ$. By rearranging the cell sizes of this ECART design the design on the right is obtained. For this new design $\theta = 86.56^\circ$. Thus, by only rearranging cell sizes, θ has been changed 22.16°.

CHAPTER IV

CONCLUSION

In this work it has been shown that for any 2×3 ECART design where $u=0$, or $v = 0$, or where $K = 1,2,\cdots,20$, the angle, θ_E, between $\hat{A}|J$ and $\hat{B}|J$ is less than or equal to the angle, θ_g, between $\hat{A}|J$ and $\hat{B}|J$ for any design generated from that ECART design. This same concept is conjectured and believed to be true, but not yet proved in general, for $K > 20$. It has also been shown that the angle between $\hat{A}|J$ and $\hat{B}|J$ is the same as the angle between $\hat{A}|\hat{B}$ and $\hat{B}|\hat{A}$ for any $a \times b$ design. Therefore, for any 2×3 ECART design with $u = 0$, or $v = 0$, or $K = 1,2,\cdots,20$, the angle, θ_E, between $\hat{A}|\hat{B}$ and $\hat{B}|\hat{A}$ is less than or equal to the angle, θ_g, between $\hat{A}|\hat{B}$ and $\hat{B}|\hat{A}$ for any design generated from that ECART design. Thus for a 2×3 ECART design with $u = 0$, or $v = 0$, or $K = 1,2,\cdots,20$, the angle, θ_E, between $\hat{A}|J$ and $\hat{B}|J$, ($\hat{A}|\hat{B}$ and $\hat{B}|\hat{A}$), may be used as a lower bound for the set of angles, between $\hat{A}|J$ and $\hat{B}|J$, ($\hat{A}|\hat{B}$ and $\hat{B}|\hat{A}$), for the designs generated from the ECART design. If $\hat{A}|J$ and $\hat{B}|J$, ($\hat{A}|\hat{B}$ and $\hat{B}|\hat{A}$), are "close enough" to orthogonal to express the results of the associated analysis as though $\hat{A}|J$ and $\hat{B}|J$, ($\hat{A}|\hat{B}$ and $\hat{B}|\hat{A}$), were orthogonal without being "far" from accurate, then the same may be done for any design generated from the ECART design.

BIBLIOGRAPHY

1. Burdick, D. S.; Herr, D. G.; O'Fallon, W. M.; O'Neill, B. V. "Exact Methods in Unbalanced, Two-way Analysis of Variance - A Geometric Approach." Comm. Statist., 3, 1974. pp.581-594.

2. Dixon, W. J.; Massey, F. J., Jr. Introduction to Statistical Analysis, 3rd edition. New York: McGraw Hill Book Company. 1969

3. Herr, D. G. "A Geometric Characterization of Connectedness in a Two-way Design." Biometrika, 1976. (to appear).

4. Herr, D. G. "Unbalanced Two-way ANOVA and SAS." Proceedings of first International SAS Users Meeting, 1976. (to appear).

5. Lancaster, Peter. Theory of Matrices. New York: Academic Press, Inc., 1969.

APPENDIX A

Theorems and Definitions

Consider two subspaces, U_1 and V_1, of euclidean k-space, R^k, and let $n = \min \{\dim(U_1), \dim(V_1)\}$. In this appendix it is assumed that none of the vectos mentioned are zero, unless otherwise stated. Let $u_1 \varepsilon U_1$ and $v_1 \varepsilon V_1$ be such that the angle, θ_1, between u_1 and v_1 is the smallest angle between any $u \varepsilon U_1$ and $v \varepsilon V_1$. Define $U_2 = \{u{:}u \quad U_1$ and $u \perp u_1\}$ and $V_2 = \{v{:}v \varepsilon V_1$ and $v \perp v_1\}$. Let $u_2 \varepsilon U_2$ and $v_2 \varepsilon V_2$ be such that the angle, θ_2, between u_2 and v_2 is the smallest angle between any $u \varepsilon U_2$ and $v \varepsilon V_2$. In the same way, let $U_i = \{u{:}u \varepsilon U_1$ and $u \perp u_1, u_2, \cdots, u_{i-1}\}$ and $V_i = \{v{:}v \varepsilon V_1$ and $v \perp v_1, v_2, \cdots, v_{i-1}\}$. Let $u_i \varepsilon U_i$ and $v_i \varepsilon V_i$ be such that the angle, θ_i, between u_i and v_i is the smallest angle between any $u \varepsilon U_i$ and $v \varepsilon V_i$. Notice that $\dim(U_i) = \dim(U_1) - i$ and $\dim(V_i) = \dim(V_1) - i$. This implies that U_i and V_i are defined for only $i = 1,2,\cdots,n$. Thus θ_i is defined only for $i = 1,2,\cdots, n$. Define the angle, θ, between U_1 and V_1 by $\theta = (\theta_1,\theta_2,\cdots,\theta_n)$.

For a 2×3 design, $\dim(\hat{A}|J) = 1$ and $\dim(\hat{B}|J) = 2$. Thus the angle, θ, between $\hat{A}|J$ and $\hat{B}|J$ has only one component, (i.e. $\theta = \theta_1$).

If X is an inner product space, with inner produce $<.,.>$, and $x, y \varepsilon X$, the angle, β, between x and y is defined by $\cos^2\beta = <x,y>^2/||x||^2 \cdot ||y||^2$. It is assumed that $0 \leq \beta \leq 90^o$. Here $||\cdot||$ is the length, or norm, defined by $||z||^2 = <z,z>$ for $z \varepsilon X$. For U and V, subspaces of X, $\dim(U) = 1$, the angle, θ, between U and V is defined by $\cos^2\theta_2 \max\{<x,y>^2/||x||^2 \cdot ||y||^2 : x \varepsilon U$ and $y \varepsilon V\}$.

Theorem A1. <u>If</u> U <u>and</u> V <u>are subspaces of euclidean k-space</u>, R^k, <u>and</u> $<\cdot,\cdot>$ <u>is the usual inner product (dot product), then</u> $\max\{<u,v>^2/||u||^2\cdot||v||^2 : u \in U \text{ and } v \in V\} = \max\{||P_U v||^2 : v \in V \text{ and } ||v||^2 = 1\}$.

Proof. Let $u \in U$ and $v \in V$. Now $u = \alpha P_U v + c$, where $c \in U$, $c \perp P_U v$, and $\alpha \in R$. Also $v = P_U v + z$ where $z \perp U$ and $v = \beta w$ where $w \in V, ||w||^2 = 1$, and $\beta \in R$. Thus $<u,v>^2/||u||^2\cdot||v||^2 =$

$<\alpha P_U v + c,\ P_U v + z>^2/||\alpha P_U v + c||^2\cdot||v||^2$

$$= \frac{(\alpha<P_U v, P_U v> + \alpha<P_U v, z> + <c, P_U v> + <c,z>)^2}{(|\alpha|^2||P_U v||^2 + ||c||^2)||v||^2}$$

$= (\alpha<P_U v, P_U v>)^2/(\alpha^2||P_U v||^2 + ||c||^2)||v||^2$

$= \alpha^2||P_U v||^4/(\alpha^2||P_U v||^2 + ||c||^2)||v||^2$

$\leq \alpha^2||P_U v||^4/\alpha^2||P_U v||^2\cdot||v||^2 = ||P_U v||^2/||v||^2 = ||P_U \beta w||^2/||\beta w||^2$

$= ||\beta P_U w||^2/||\beta w||^2 = \beta^2||P_U w||^2/\beta^2||w||^2 = ||P_U w||^2/||w||^2$

$= ||P_U w||^2$. Therefore, since u and v were arbitrary elements of U and V, respectively; and since when $u = \alpha P_U v$, $<u_1 v>^2/||u||^2\cdot||v||^2 = ||P_U w||^2$; $\max\{<u,v>^2/||u||^2\cdot||v||^2 : u \in U \text{ and } v \in V\}$

$= \max\{||P_U v||^2 : v \in V \text{ and } ||v|| = 1\}$.

Thus the angle, θ, between $\hat{A}|J$ and $\hat{B}|J$ for a 2 x 3 design may be defined by $\cos^2\theta = \max\{<x,y>^2/||x||^2\cdot||y||^2 : x \in \hat{A}|J \text{ and } y \in \hat{B}|J\} = \max\{||P_{\hat{B}|J}\, x||^2 : x \in \hat{A}|J \text{ and } ||x||^2 = 1\}$.

Theorem A 2. <u>The angle</u>, θ, <u>between</u> $\hat{A}|J$ <u>and</u> $\hat{B}|J$ <u>is given by</u>

$\cos^2\theta =$

$$\frac{n_{1.}n_{2.}}{n_{..}}\left(\frac{1}{n_{.1}}\left(\frac{n_{11}}{n_{1.}}-\frac{n_{21}}{n_{2.}}\right)^2+\frac{1}{n_{.2}}\left(\frac{n_{12}}{n_{1.}}-\frac{n_{22}}{n_{2.}}\right)^2+\left(\frac{n_{13}}{n_{1.}}-\frac{n_{23}}{n_{2.}}\right)^2\right). \qquad (1)$$

Proof. Since $\dim(\hat{A}|J) = 1$, all elements of $\hat{A}|J$ are constant multiples of a, the basis element of $\hat{A}|J$ defined on page 12. Thus, there is only one element of $\hat{A}|J$, u, such that $||u||^2 = 1$. $u = \frac{a}{||a||}$.

Now $\cos^2\theta = \max\{||P_{\hat{B}|J}x||^2 : x \in \hat{A}|J \text{ and } ||x||^2 = 1\}$. Since there is only one such $x \in \hat{A}|J$ $\cos^2\theta = ||P_{\hat{B}|J}u||^2 = ||P_{\hat{B}|J}a||^2/||a||^2$.

Next $P_{\hat{B}}a = P_{\hat{B}|J}a + P_J a$. But $a \in \hat{A}|J$ and thus $a \perp J$ so $P_{\hat{B}}a = P_{\hat{B}|J}a$. Therefore $\cos^2\theta = ||P_{\hat{B}}a||^2/||a||^2$. Now $||a||^2 = \langle a,a\rangle = 1/n_{1.} + 1/_{n2.} = (n_{1.} + n_{2.})/n_{1.}n_{2.} = n_{..}/n_{1.}n_{2.}$, so $\cos^2\theta = (n_{1.}n_{2.}/n_{..})||P_{\hat{B}}a||^2$.

Note here that b_1, b_2 and b_3, defined as basis elements of $\hat{B}$ in Chapter II, are mutually orthogonal, so $||P_{\hat{B}}a||^2 = ||P_{b_1}a||^2 + ||P_{b_2}a||^2 + ||P_{b_3}a||^2$. Finally, $||P_{b_i}a||^2 = \langle b_i,a\rangle/||b_i||^2 = ((n_{1i}/n_{1.}) - (n_{2i}/n_{2.}))^2/n_{.i}$, so $\cos^2\theta = (1)$.

Theorem A 3. <u>For a</u> 2 x 3 <u>design, the angle</u>, θ, <u>between</u> $\hat{A}|J$ <u>and</u> $\hat{B}|J$ <u>is the same as the angle</u>, γ, <u>between</u> $\hat{A}|\hat{B}$ <u>and</u> $\hat{B}|\hat{A}$

Proof. $\hat{A}|\hat{B} \oplus \hat{B}|J = (\hat{A} + \hat{B})|J = \hat{B}|\hat{A} \oplus \hat{A}|J$. Let $\{a_1\}$, $\{b_1, b_2\}$, $\{a_2\}$, $\{b_3, b_4\}$ be orthonormal bases for $\hat{A}|\hat{B}$, $\hat{B}|J$, $\hat{A}|J$, $\hat{B}|\hat{A}$ respectively. Let the matrices U_1, U_2, W_1 and W_2 be defined by $U_1 = [a_1]$, $U_2 = [b_1, b_2]$, $W_1 = [b_3, b_4]$, and $W_2 = [a_2]$. Thus $U_1'U_1 =$

$1 = W_2'W_2$ and $U_2'U_2 = I_2 = W_1'W_1$ (I_2 is the 2 x 2 identity matrix).

Now $\cos^2\gamma_1 = \max\{<u,w>^2/||u||^2 \cdot ||w||^2 : u \in \hat{A}|\hat{B}$ and $w \in \hat{B}|\hat{A}\}$ = $\max\{||P_{\hat{A}|\hat{B}}w||^2 : w \in \hat{B}|\hat{A}$ and $||w||^2 = 1\}$. Notice that for any $w \in \hat{B}|\hat{A}$, $\exists\, z \in R^2$ such that $W_1 z = w$. Consider $z \in R^2$ where $||z||^2 =$ 1. Then $1 = z'z = z'I_2 z = z'W_1'W_1 z = ||W_1 z||^2$ and $W_1 z \in B|A$. Thus $\cos^2\gamma_1 = \max\{||P_{\hat{A}|\hat{B}}W_1 z||^2 : z \in R^2$ and $||z||^2 = 1\}$. Notice that $1 = ||W_1 z||^2 = ||P_{\hat{A}|\hat{B}}W_1 z||^2 + ||P_{\hat{B}|J}W_1 z||^2$, so $\cos^2\gamma_1 = \max\{1 - ||P_{\hat{B}|J}W_1 z||^2 : z \in R^2$ and $||z||^2 = 1\} = 1 - \min\{||P_{\hat{B}|J}W_1 z||^2 : z \in R^2$ and $||z||^2 = 1\}$. Now $P_{\hat{B}|J}W_1 z = U_2(U_2'U_2)^{-1}U_2'W_1 z = U_2U_2'W_1 z$, so $||P_{\hat{B}|J}W_1 z||^2 = z'W_1'U_2U_2'U_2U_2'W_1 z = z'W_1'U_2U_2'W_1 z$. Thus $\cos^2\gamma_1 =$ $1 - \min\{z'W_1'U_2U_2'W_1 z : z \in R^2$ and $||z||^2 = 1\} = 1 - \min\{\lambda : \lambda$ is an eigen value of $(U_2'W_1)'(U_2'W_1)\}$ (Lancaster [5], p.109).

Similarly $\cos^2\theta_1 = \max\{<u,w>^2/||u||^2 \cdot ||w||^2 : u \in \hat{A}|J$ and $w \in \hat{B}|J\} = 1 - \min\{\mu : \mu$ is an eigen value of $(U_2'W_1)(U_2'W_1)'\}$.

Now, if M is a real matrix then $M'M$ is a real, symmetric matrix and has the same non-zero eigen values as MM'. When M' is square $M'M$ and MM' have all the same eigen values. Thus $\cos^2\gamma_1 =$ $1 - \min\{\lambda : \lambda$ is an eigen value of $(U_2'W_1)'(U_2'W_1)\} = 1 - \min\{\mu : \mu$ is an eigen value of $(U_2'W_1)(U_2'W_1)'\} = \cos^2\theta_1$. This implies that $\gamma_1 = \theta_1$ and since $\dim(\hat{A}|J)$ and $\dim(\hat{A}|\hat{B}) = 1$, $\lambda = \theta$.

Theorem A4. <u>For an</u> a x b <u>design, the angle,</u> θ, <u>between</u> $\hat{A}|J$ <u>and</u> $\hat{B}|J$ <u>is the same as the angle,</u> γ, <u>between</u> $\hat{A}|\hat{B}$ <u>and</u> $\hat{B}|\hat{A}$.

Proof. $\gamma_1 = \theta_1$ by the same proof as in Theorem A3. The proof that $\theta_i = \gamma_i$, $i < 1$ is similar.

APPENDIX B

PROGRAM FOR PROOF OF THEOREM 3

```
//ALLEN JOB ECS.UNCG.MA7329,KENDALL,T=(3,59),P=999,PRTY=0,
// FORMS=REAMS
// EXEC WATFIV
//SYSIN DD *
$JOB      ECS.UNCG.MA7968.C573/KENDALL,TIME=239,PAGES=999
      DIMENSION X(6),Y(6),L(2),FLAG(100),Z(6),DC(3),DR(2)
C DATA POINT IS BEING READ
    1 READ(1,2) K,LAST
    2 FORMAT(I5,74X,I1)
C DATA POINT UNDER CONSIDERATION, K, IS DENOTED
      WRITE(3,60) K
   60 FORMAT(1H1,50X,'K EQUALS ',I5)
      KA=K
    3 FORMAT(1H1)
      KK=K-1
C DO LOOP FOR IA BEGINS HERE AND ENDS AT 49
C IA RANGES FROM 1 TO K-1
      DO 49 IA=1,KK
C MIN(IA,K-IA) IS CALCULATED
      KKK=KA-IA
      IF(IA-KKK)51,51,52
   51 JB=IA
      GO TO 53
   52 JB=KKK
      GO TO 53
   53 CONTINUE
C DO LOOP FOR IB BEGINS HERE AND ENDS AT 49
C IB RANGES FROM 1 TO MIN(IA,K-IA)
      DO 49 IB=1,JB
C HERE K,IA,IB ARE WRITTEN TO SEPARATE THE
C ECART DESIGN ASSOCIATED WITH K, IA, IB AND ITS
C 60 INTERESTINGLY DIFFERENT DESIGNS FROM
C THE PRECEDING DESIGNS
      WRITE(3,4) KA,IA,IB
    4 FORMAT(1H1,'K = ',I3,2X,'A = ',I3,2X,'B = ',I3)
      WRITE(3,3)
      K=KA
      ICOUNT=1
      X(1)=K+IA
      X(2)=K+IB
      X(3)=K-IA-IB
      X(4)=K-IA
      X(5)=K-IB
      X(6)=K+IA+IB
```

```
C DO LOOP GENERATING THE INTERESTINGLY DIFFERENT
C DESIGNS FROM THE ECART DESIGN BEGINS HERE
C AND ENDS AT 25
      DO 10 J=2,5
      JJ=J+1
      DO 10 K=JJ,6
      Z(1)=X(1)
      Z(2)=X(J)
      Z(3)=X(K)
      L(1)=J
      L(2)=K
      II=4
      DO 20 I=2,6
      IF (L(1).EQ.I) GO TO 20
      IF (L(2).EQ.I) GO TO 20
      Y(II)=X(I)
      II=II+1
   20 CONTINUE
      DO 25 N1=1,3
      DO 25 N2=1,3
      IF (N1.EQ.N2) GO TO 25
      N3=6-N1-N2
      Y(1)=Z(N1)
      Y(2)=Z(N2)
      Y(3)=Z(N3)
C ROW AND COLUMN TOTALS CALCULATED
      AN1D=Y(1)+Y(2)+Y(3)
      AN2D=Y(4)+Y(5)+Y(6)
      AND1=Y(1)+Y(4)
      AND2=Y(2)+Y(5)
      AND3=Y(3)+Y(6)
C SQUARED COSINE CALCULATED FOR EACH DESIGN
      ACOS2=(((Y(1)/AN1D)-(Y(4)/AN2D))**2)/AND1
      ACOS2=ACOS2+(((Y(2)/AN1D)-(Y(5)/AN2D))**2)/AND2
      ACOS2=ACOS2+(((Y(3)/AN1D)-(Y(6)/AN2D))**2)/AND3
      ACOS2=ACOS2*AN1D*AN2D/(6*KA)
C ANGLE CALCULATED FROM THE SQUARED COSINE
      ANGLE=SQRT(ACOS2)
      ANGLE=ARCOS(ANGLE)
      ANGLE=ANGLE*(180.0/3.141593)
      FLAG(ICOUNT)=ACOS2
C DESIGN PRINTED ALONG WITH SQUARED COSINE AND
C ANGLE
      WRITE(3,41) Y(1),Y(2),Y(3),ANGLE
   41 FORMAT(1H0,3(F3.0,2X),2X,'ANGLE=',F10.5)
      WRITE(3,42) Y(4),Y(5),Y(6),ACOS2
   42 FORMAT(1H0,3(F3.0,2X),2X,'COSIN SQ =',F10.5)
```

```
C SQUARED COSINE FOR DESIGN IS SUBTRACTED FROM
C SQUARED COSINE FOR ASSOCIATED ECART DESIGN, IF
C DIFFERENCE IS NEGATIVE A WARNING IS PRINTED
      IF (FLAG(1)-FLAG(ICOUNT)) 43,25,25
   43 WRITE(3,44)
   44 FORMAT(1H+,110X,'HELP')
      ICOUNT=ICOUNT+1
   25 CONTINUE
   10 CONTINUE
   49 CONTINUE
C RETURN FOR A NEW DATA POINT, IF NO MORE DATA
C IS THERE THE PROGRAM IS TERMINATED
      IF (LAST) 50,1,50
   50 STOP
      END
$DATA
   15
   16
//
```

The following is sample out-put from the above program where $K = 21$, $u = 13$, and $v = 7$. The first two lines is an essentially different design and each pair of lines afterwards is a new essentially different design.

```
 8.  28.  34.     ANGLE=  74.21489
 1.  14.  41.     COSIN SQ =   0.07400
34.  28.  14.     ANGLE=  49.86916
 1.   8.  41.     COSIN SQ =   0.41543
34.  14.  28.     ANGLE=  60.17793
 1.   8.  41.     COSIN SQ =   0.24731
```