Aspects of the analysis of variance for classifactory data

Lakshmi Rangachari
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/7527

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

University Microfilms International
300 N. Zeeb Road
Ann Arbor, MI 48106
Rangachari, Lakshmi

ASPECTS OF THE ANALYSIS OF VARIANCE FOR CLASSIFICATORY DATA

Iowa State University

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

Ph.D. 1982
Aspects of the analysis of variance
for classificatory data

by

Lakshmi Rangachari

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:
Signature was redacted for privacy.
In Charge of Major Work
Signature was redacted for privacy.
For the Major Department
Signature was redacted for privacy.
For the Graduate College

Iowa State University
Ames, Iowa
1982
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. INTRODUCTION AND SUMMARY</td>
<td>1</td>
</tr>
<tr>
<td>2. DESCRIPTION OF ORDERINGS BY ORDERED POLYTOPIES</td>
<td>3</td>
</tr>
<tr>
<td>2.1. Introduction and Review of Finch's Paper</td>
<td>3</td>
</tr>
<tr>
<td>2.2. Characterizing Power of Ordered Polytomies for the Formal Situation $(\Delta_k, \Omega, \gamma)$</td>
<td>7</td>
</tr>
<tr>
<td>2.3. Descriptive Power of Ordered Polytomies for the Formal Situation $(\Delta_k, \Omega, \gamma)$</td>
<td>14</td>
</tr>
<tr>
<td>2.4. Characterizing and Descriptive Powers for the Formal Situation $(\Delta_k^*, \Omega, \gamma)$</td>
<td>22</td>
</tr>
<tr>
<td>2.5. Relation Between Finch's Gauge and Some Other Statistics</td>
<td>31</td>
</tr>
<tr>
<td>2.6. Use of Finch's Ideas</td>
<td>33</td>
</tr>
<tr>
<td>3. MEASURE OF EFFECTIVE VARIATION IN AN ARRAY</td>
<td>39</td>
</tr>
<tr>
<td>3.1. Background - Work Done by Wilk and Kemthorne, and Zyskind</td>
<td>39</td>
</tr>
<tr>
<td>3.2. Review of Cox's Paper</td>
<td>39</td>
</tr>
<tr>
<td>3.3. Derivation of the Components of Effective Variation for the General Case of $k$ Factors</td>
<td>47</td>
</tr>
<tr>
<td>3.4. Discussion</td>
<td>49</td>
</tr>
<tr>
<td>4. THE TWO-FACTOR MIXED LINEAR MODEL</td>
<td>65</td>
</tr>
<tr>
<td>4.1. Introduction</td>
<td>65</td>
</tr>
<tr>
<td>4.2. Derivations of the Mixed Effects Model by Various Authors</td>
<td>65</td>
</tr>
<tr>
<td>4.3. An Alternative Approach to Derivation of a Linear Model for the Mixed Factor Case</td>
<td>91</td>
</tr>
<tr>
<td>4.4. Conclusion</td>
<td>110</td>
</tr>
<tr>
<td>5. BIBLIOGRAPHY</td>
<td>112</td>
</tr>
<tr>
<td>6. ACKNOWLEDGEMENTS</td>
<td>115</td>
</tr>
</tbody>
</table>
1.

INTRODUCTION AND SUMMARY

The studies reported in this thesis, started with an attempt to obtain better understanding of a controversy with respect to a linear classificatory model and the associated analysis of variance, for which one factor of classification is fixed, with each level represented in the data, while the other factor is random, and the data contain only a sample of levels from a population of possible levels.

After preliminary study, it was judged appropriate to consider first, ideas of description of a data set, which have been formulated by Finch (1979). These are exposited and extended somewhat in Chapter 2. The ideas are useful in quantifying how well one description, δ, say from a class of possible descriptions Δ, describes a data set, ω, relative to the whole class of possible descriptions, Δ. The basic idea is to use a measure of badness of description, denoted by γ(δ,ω), which Finch calls the gauge. One then considers how γ(δ,ω) varies, as either δ varies over Δ, or ω varies over Ω, a collection of possible data sets.

It seemed then, that there is some relation of ideas of quality of description and of the ideas in the aforementioned mixed linear model controversy, to the ideas of Cox (1958), which turn out to be related to ideas on random sampling of a multidimensional array and to randomized experiments. These ideas and Cox's work are reviewed in Chapter 3, with development of a general formula for what Cox calls measures of effective variation for subsets of factors in a multi-dimensional array.

In Chapter 4, the background and various ideas put forward by various authors for the 2-factor model with one factor fixed and one-factor random
are described. The relationships between the different derivations of the mixed model, and also the connection with Cox's Σ described in Chapter 3, are discussed.
2. DESCRIPTION OF ORDERINGS BY ORDERED POLYTOPIES

2.1. Introduction and Review of Finch's Paper

This chapter examines an interesting idea of P. D. Finch (1979) on data description. In statistics, one is not only interested in analyzing data by testing various hypotheses, but also in describing the data. By trying to get a good description of the data and seeing what the data, on their own, have to tell, one can then form appropriate hypotheses to be tested. The results of these tests can then be used to make inferences about the underlying population. Thus, getting a good description of the data is very important.

The concept of description used is very general. A simple example is the description of a finite set of N scalar observations by an ordered dichotomy or polytomy. Another example is as follows: suppose we have a set of 2-vectors \( \{(y_i, x_i)/i=1,2, \ldots, n\} \). Then we may describe this set by the set \( \{(\hat{y}_i, x_i)/i=1,2, \ldots, n\} \), where \( \hat{y}_i \) is obtained by fitting a model \( y = f(x) \).

One may get any number of descriptions of a given data set. This gives rise to a number of questions such as, "What constitutes a good description?", "How can one compare two different descriptions of the same data set?"

So, the basic idea in this chapter is to examine the question, "How good is a particular description of a given data set?" For example, one might ask how well an ordered dichotomy describes a strong ordering.

In this paper, Finch defines a function which can be used to gauge
the effectiveness of a description. He makes the following definitions:

Let $\Omega$ be the set of objects to be described. Let $\Delta$ be a set of possible descriptions. Let $\gamma$ be a real-valued function on $\Delta \times \Omega$, such that $\gamma(\delta,\omega)$ gauges the extent to which $\delta$ falls short of a perfect description of $\omega$.

Then, $\gamma$ is non-negative and the smaller its value the better. Both $\Delta$ and $\Omega$ are assumed to be finite sets. Let $P(\delta_0||\omega_0) = \text{proportion} \{\delta/\delta \in \Delta, \gamma(\delta,\omega) \geq \gamma(\delta_0,\omega_0)\}$ and $Q(\omega_0||\delta_0) = \text{proportion} \{\omega/\omega \in \Omega, \gamma(\delta_0,\omega) \geq \gamma(\delta_0,\omega_0)\}$.

Then $P(\delta_0||\omega_0)$ and $Q(\omega_0||\delta_0)$ are called respectively, the descriptive power and the characterizing power of $\delta_0$ as a description of $\omega_0$.

The descriptive power shows how much better a description of $\omega_0$ is $\delta_0$, than the other descriptions in $\Delta$, while the characterizing power shows, how good a description of $\omega_0$ is $\delta_0$ relative to the other objects in $\Omega$. An ordered triple $(\Delta, \Omega, \gamma)$ is called a formal situation. These three quantities $\Delta, \Omega, \gamma$ can be chosen in different ways. Sometimes, they are decided on before the data are examined. Often, though, they are specified after the data are looked at.

Finch uses the basic ideas of a relation on a set $X$. He takes $\Delta$ to be the set of all binary relations on a finite set $X$ with $N$ elements. The relation $i$ denotes the identity relation on $X$. For the binary relation $\rho$ we have $x = [y/x, \rho].$ $\rho_x = [y/\rho_x].$ Distinct elements $x, y$ are $\rho$-tied when $x \rho = y \rho$ and $\rho x = \rho y$.

The zeta function of $\rho$ is the function $\zeta_\rho$ on $X \times X$ with $\zeta_\rho(x,y) = 1$ or $0$, depending on whether $(x,y)$ is or is not in $\rho$. The complement of $\rho$ is $\rho'$, its set complement in $X \times X$ and its converse is $\rho^*$, where $x \rho^* y$ implies $y \rho x$. An ordering of $X$ is a relation in $\Delta$ which is reflexive, transitive and such that $\omega \cap \delta = X \times X$. It is a strong ordering if it has.
no ties i.e. \( \omega \cap \delta = \emptyset \).

Only strong orderings were considered by Finch. Let \( \Omega \) be the set of all strong orderings on \( X \). Let \( D \subseteq X \) and \( D' \) be its set complement in \( X \). Then, the binary relation \( \delta_D = D' \times D \) is called the ordered dichotomy generated by \( D \). Also, Finch considered \( \Delta_n \), the set of all such dichotomies, and \( \Delta_n \) the subset of \( \Delta_n \) generated by subsets of size \( n \).

Finch defines a gauge \( \gamma \) which measures the extent to which \( \delta \) falls short of a perfect description of \( \omega \): \( \gamma(\delta, \omega) = \text{card}(\delta \cap \omega \cap \omega') + \text{card}(\delta' \cap \omega \cap \omega') \), where \( \text{card}(S) \) indicates cardinality of the set \( S \).

So, the gauge counts the number of pairs \( (x, y) \) with \( x \neq y \), which are either in \( \delta \) but not in \( \omega \) or in \( \omega \) but not in \( \delta \).

For the formal situation \( (\Delta, \Omega, \gamma) \), Finch obtains the descriptive power of \( \delta_\omega \) as the upper-tail determined by \( \gamma(\delta_\omega, \omega) \) on the binomial distribution with \( N(N-1) \) trials and success probability \( \frac{1}{2} \).

We wish to consider the variation of \( \gamma(\delta_\omega, \omega) \) as \( \omega \) ranges over \( \Omega \). We may then think of \( \omega \) as being a random member of \( \Omega \), and we can represent the variation of \( \gamma(\delta_\omega, \omega) \) as the variation of a random variable, in just the same way as we use mean and variance in the basic Chebychev formula to represent partially, a set of \( n \) numbers. So, in the ensuing we shall use \( M(\cdot) \) and \( \text{Var}(\cdot) \).

For the formal situation \( (\Delta, \Omega, \gamma) \), Finch obtains the characterizing power of \( \delta_\omega \) as the upper-tail determined by \( \gamma \), on a distribution that is approximately normal when \( N \) is moderately large. The mean of this distribution is \( M[\gamma(\delta_\omega, \cdot)] = \frac{1}{2} N(N-1) \), and the variance is

\[
\text{Var}[\gamma(\delta_\omega, \cdot)] = \frac{1}{2} \text{card}(\delta_\omega \cap \omega') - \text{card}(\delta_\omega \cap \omega') \sum_{x \in X} \left[ x(\delta_\omega \cap \omega') - \frac{x(\delta_\omega \cap \omega')}{x(\delta_\omega \cap \omega')} \right]^2
\]
where \( r_x(\alpha) = \text{card}(x \alpha) \) and \( \ell_x(\alpha) = \text{card}(\alpha x) \).

When \( N \) is sufficiently large, \( \gamma(\delta_0, \cdot) \) has an approximate normal distribution with mean \( M[\gamma(\delta_0, \cdot)] \) and \( \text{Var}[\gamma(\delta_0, \cdot)] \), as given above.

When \( \delta = \delta_D \) (the ordered dichotomy generated by the set \( D \)), then \( \text{Var}[\gamma(\delta_D, \cdot)] \) reduces to \( \text{Var}[\gamma(\delta_D, \cdot)] = \frac{1}{3} n(N-n)(N+1) \), where \( n = \text{card}(D) \) and \( N = \text{card}(X) \).

For the formal situation \((\Delta_x, \Omega, \gamma)\), Finch shows that the descriptive power of \( \delta_F \), as a description of \( \alpha_0 \), is the upper-tail determined by \( \gamma(\delta_F, \alpha_0) \) on a distribution that is approximately normal. The mean of this distribution is:

\[
M[\gamma(\cdot, \alpha_0)] = \frac{1}{2} N(N-1),
\]

and the variance is given by

\[
\text{Var}[\gamma(\cdot, \alpha_0)] = \frac{N(N-1)}{12}.
\]

For the formal situation \((\Delta_x, \Omega, \gamma)\) the computation of the characterizing power of \( \gamma(\delta_D, \cdot) \) does not change. Hence, \( \text{Var}[\gamma(\delta_D, \cdot)] \) is the same as before.

An ordered dichotomy has different descriptive powers according as we reference it to \( \Delta \), \( \Delta_x \) or \( \Delta_n \), even though the same gauge is used in all three cases.

In the last section of his paper, Finch discusses the description of one dichotomy by another. If one dichotomy is \( \delta_H \), and it is described by the other \( \delta_B \), the gauge then is \( \gamma(\delta_B, \delta_H) = \text{card}(\delta_B \setminus \delta_H) + \text{card}(\delta_H \setminus \delta_B) \).

He shows that for the formal situation \((\Delta_x, \Delta_x, \gamma)\), \( \gamma(\cdot, \delta_H) \) has, for large \( N \), approximately a normal distribution, and the descriptive power is given by the upper-tail determined by \( \gamma(\delta_B, \delta_H) \). The mean and variance for this distribution are:
\[ M[\gamma(\delta_H, \delta_H^*)] = \frac{1}{4}N(N-1) + 2n_Hn_H, \text{ and } \text{Var}[\gamma(\delta_H, \delta_H^*)] = \frac{1}{8}(N^2 + 2(N-1)n_Hn_H), \]

where \( n_H = \text{card}(H), N = \text{card}(X). \)

Similarly, \( \gamma(\delta_B, \delta_B^*) \) has approximately a normal distribution, with
\[ \text{mean } M[\gamma(\delta_B, \delta_B^*)] = \frac{1}{4}N(N-1) + 2n_Bn_B, \text{ and } \text{variance } \text{Var}[\gamma(\delta_B, \delta_B^*)] = \frac{1}{8}(N^2 + 2(N-1)n_Bn_B). \]

The characterizing power is given by the upper-tail determined by \( \gamma(\delta_B, \delta_H^*). \)

If instead, one considers \( (\Delta_B, \Delta_H, \gamma) \), then the descriptive power of \( \delta_B \) as a description of \( \delta_H \) is
\[ P(\gamma(\delta_B^*, \gamma) = \frac{n_H}{N} \text{ for } \delta_B^*, \gamma) / \binom{N}{b}, \]
and the characterizing power of \( \delta_B \) is given also by \( P(\gamma(\delta_B^*, \gamma) = \frac{n_H}{N} \text{ for } \delta_B^*, \gamma) / \binom{N}{b}. \)

Thus, the descriptive and characterizing powers are equal in \( (\Delta_B, \Delta_H, \gamma) \) and their common value is one minus the significance level of Fisher's exact test for 2x2 tables with fixed margins.

2.2. Characterizing Power of Ordered Polytomies for the Formal Situation \( (\Delta_K, \Omega, \gamma) \)

Consider the formal situation \( (\Delta_K, \Omega, \gamma) \). The \( \Delta_K \) is the set of all possible ordered \( k \)-tomies
\[ \delta_0 = D_1 < D_2 < D_3 < \ldots < D_{k-1} < D_k, \]

where \( D_i \subseteq X, i = 1, 2, \ldots, k, \) \( \bigcup_{i=1}^{k} D_i = X \),

\[ \bigcap_{i=1}^{k} D_i = \emptyset \text{ with card}(D_i) = n_i, i = 1, 2, \ldots, k. \] The cardinalities \( n_1, n_2, \ldots, n_k \) are fixed for \( \Delta_K \) and \( \sum_{i=1}^{k} n_i = N = \text{card}(X). \) The set \( \Omega \) is the set of all strong orderings on \( X. \) The function \( \gamma \), defined by Finch as
\[ \gamma(\delta_0, \omega_0) = \text{card} (\delta_0 \cap \omega_0' \cap \Omega') + \text{card} (\delta_0' \cap \omega_0' \cap \Omega') \]

can be given in an alternative form as
\[ \gamma(\delta_0, \omega_0) = \sum_{x,y} \left\{ \zeta_{\delta_0}(x,y) [1 - \zeta_{\omega_0}(x,y)] + [1 - \zeta_{\delta_0}(x,y)] \zeta_{\omega_0}(x,y) \right\}, \]

where
\[ \zeta_{\delta_0}(x,y) = 1 \quad \text{if} \ (x,y) \in \delta_0, \]
\[ = 0 \quad \text{otherwise}. \]

We have \( \zeta_{\delta_0}(x,y) = 1 \quad \text{iff} \ (x,y) \in \delta_0, \)
i.e., iff \( x \in D_i, \ y \in D_j \) with \( i < j, \ i, j = 1, 2, \ldots, k. \)

We shall first state and prove a lemma.

Lemma 2.1.: Consider the formal situation \( (\Delta, \Omega, \gamma) \), where \( \Delta \) is the set of all binary relations on the finite set \( X, \ \Omega \) is the set of all strong orderings and \( \gamma \) is the function given above. Then,
\[ \mathbb{E}[\gamma(\delta_0, \cdot)] = \frac{1}{2} N(N-1), \quad \text{and} \]
\[ \text{Var}[\gamma(\delta_0, \cdot)] = \left( \frac{1}{3} \right) \left[ \sum_x (A_x - A_x)^2 + \sum_x A_x \right], \]

where
\[ A_x = \sum_y a_{xy} = \sum_y \zeta_{\delta_0}(x,y), \]
and
\[ A_{\cdot x} = \sum_y a_{yx} = \sum_y \zeta_{\delta_0}(y,x). \]

Proof: We have
\[ \gamma(\delta_0, \omega) = \text{card} (\delta_0 \cap \omega \cap \Omega') + \text{card} (\delta_0' \cap \omega \cap \Omega') \]
\[ = \sum_{x,y} \left\{ \zeta_{\delta_0}(x,y) [1 - \zeta_{\omega}(x,y)] + [1 - \zeta_{\delta_0}(x,y)] \zeta_{\omega}(x,y) \right\}, \]
where
\[ \zeta_{o}(x,y) = 1 \quad \text{with relative frequency } \frac{1}{2}, \]
\[ = 0 \quad \text{with relative frequency } \frac{1}{2}. \]

So,
\[
E[\gamma(\delta_o, \cdot)] = \sum_{x,y} \left[ \frac{1}{2} \zeta_{o}(x,y) + \left[ 1 - \zeta_{o}(x,y) \right] \frac{1}{2} \right]
= \frac{1}{2} N(N-1)
= M[\gamma(\delta_o, \cdot)].
\]

\[
\text{Var}[\gamma(\delta_o, \cdot)] = \text{Var}\left[ \sum_{x,y} \zeta_{o}(x,y) + \sum_{x,y} \zeta_{o}(x,y) - 2 \sum_{x,y} \delta_{o}(x,y) \zeta_{o}(x,y) \right]
\]

Since \( \delta_o \) is fixed, \( \sum_{x,y} \zeta_{o}(x,y) \) is a constant.

Although \( \omega \) is a random variable and runs through \( \Omega \), the sum \( \sum_{x,y} \zeta_{o}(x,y) \) is a constant. Hence, \( \text{Var}[\gamma(\delta_o, \cdot)] = \text{Var}[2 \sum_{x,y} \zeta_{o}(x,y)] \),

where
\[ a_{xy} = \zeta_{o}(x,y). \] Then we have
\[
\text{Var}[\gamma(\delta_o, \cdot)] = 4 \sum_{x,y} a_{xy} \text{Var} \zeta_{o}(x,y) + \sum_{x,y} \sum_{z,t} a_{xy} a_{zt} \text{Cov}(\zeta_{o}(x,y), \zeta_{o}(z,t)).
\]

We must look at the following cases to find the covariances.

1) \( x = z, y = t \).
2) \( x = t, y = z \).
3) \( x = z, y \neq t \).
4) \( x = t, y \neq z \).
5) \( x \neq t, y = z \).
6) \( x \neq z, y = t \).
7) \( x \neq z, y \neq t \).

As an example, consider case 3. We wish to find \( \text{Cov}(\zeta_{o}(x,y), \zeta_{o}(x,t)) \).
We have the following situations:

<table>
<thead>
<tr>
<th>Situation</th>
<th>$\zeta_\omega(x,y)$</th>
<th>$\zeta_\omega(x,t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x &lt; y &lt; t$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$x &lt; t &lt; y$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$y &lt; x &lt; t$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$y &lt; t &lt; x$</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$t &lt; x &lt; y$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$t &lt; y &lt; x$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Each of these cases occurs with relative frequency (probability) of $1/6$.

So, $\text{Cov}(\zeta_\omega(x,y), \zeta_\omega(x,t)) = 2/6 - 1/2 \times 1/2$

$$= 1/3 - 1/4$$

$$= 1/12.$$

The covariances for the various cases are worked out similarly and we get the following:

<table>
<thead>
<tr>
<th>Case</th>
<th>Covariance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/4</td>
</tr>
<tr>
<td>2</td>
<td>-1/4</td>
</tr>
<tr>
<td>3</td>
<td>1/12</td>
</tr>
<tr>
<td>4</td>
<td>-1/12</td>
</tr>
<tr>
<td>5</td>
<td>-1/12</td>
</tr>
<tr>
<td>6</td>
<td>1/12</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
</tr>
</tbody>
</table>

A general formula for calculating covariance is, then,

$$\text{Cov}[\zeta_\omega(x,y), \zeta_\omega(z,t)]$$

$$= \frac{1}{12}(1-\delta_{xy})(1-\delta_{zt})[\delta_{xz} + \delta_{ty} + \delta_{tx} - \delta_{xy} \delta_{zt} + \delta_{ty} \delta_{tx} - \delta_{xy} \delta_{tx} - \delta_{zt} \delta_{tx}],$$

where

$$\delta_{ij} = 1 \text{ if } i = j,$$

$$= 0 \text{ otherwise.}$$
We have,

\[
\text{Var}[\gamma(\delta_o, \cdot)] = 4\sigma^2 \sum_{xyzt} (1 - \delta_{xy})a_{xy}(1 - \delta_{zt})a_{zt} \cdot \text{Cov}[\xi_o(x, y), \xi_o(z, t)]
\]

\[
= 4\sigma^2 \sum_{xyzt} (1 - \delta_{xy})(1 - \delta_{zt})a_{xy}a_{zt} \cdot \left( \frac{1}{12} \left( \delta_{xz} + \delta_{xy} + \delta_{zt} \right) \delta_{xy} \delta_{zt} - \delta_{xt} \delta_{yz} - \delta_{zt} \delta_{xy} \right)
\]

\[
= \left( \frac{1}{3} \right) \sum_{xyzt} (1 - \delta_{xy})(1 - \delta_{zt})a_{xy}a_{zt} + \sum_{xyzt} (1 - \delta_{xy})(1 - \delta_{zy})a_{xy}a_{zy}
\]

\[
+ \sum_{xyzt} (1 - \delta_{xy})(1 - \delta_{zt})a_{xy}a_{zt} - \sum_{xyzt} (1 - \delta_{xy})(1 - \delta_{zy})a_{xy}a_{zy}.
\]

Now, since z,t are dummy variables, we replace z by t. We simplify and write \(A_x = \sum_y a_{xy}, A_y = \sum_x a_{xy}\).

Then, the expression becomes:

\[
\text{Var}[\gamma(\delta_o, \cdot)] = \left( \frac{1}{3} \right) \left( \sum_x A_x^2 + \sum_y a_{xy}^2 - \sum_{xy} a_{xy}^2 yx - 2 \sum_x A_x, A_x \right)
\]

\[
= \left( \frac{1}{3} \right) \left( \sum_x (A_x - \bar{A}_x)^2 + \bar{A}_x, \bar{A}_x \right) .
\]

So,

\[
\text{Var}[\gamma(\delta_o, \cdot)] = \left( \frac{1}{3} \right) \left( \sum_x (A_x - \bar{A}_x)^2 + \bar{A}_x, \bar{A}_x \right) .
\]

The term \(\sum_{xy} a_{xy}a_{yx}\) will always equal zero. This is because we will be looking only at pairs \((x, y)\), such that \(x \neq y\), in which case,

\[
\zeta_o(x, y) \cdot \zeta_o(y, x) = 0 \quad \text{always}.
\]

So,

\[
\text{Var}[\gamma(\delta_o, \cdot)] = \left( \frac{1}{3} \right) \left( \sum_x (A_x - \bar{A}_x)^2 + \bar{A}_x, \bar{A}_x \right) .
\]

We show that equation (2.2.1) is the same as
\[
\text{Var}[\gamma(\delta_o, \cdot)] = \left(\frac{1}{3}\right) \left[ \text{card}(\delta_o \cap i') - \text{card}(\delta_o \cap \delta_o \cap i') \right] \\
+ \sum_{x \in X} \left[ r_x(\delta_o \cap i') - \ell_x(\delta_o \cap i') \right]^2.
\]

Now, \(\text{card}(\delta_o \cap i') = \frac{y}{xy} \zeta_{\delta_o}(x, y)\)

\[
= \sum_{xy} (1 - \delta_{xy}) \zeta_{\delta_o}(x, y) = \sum_{x, y} \zeta_{\delta_o}(x, y) - \sum_{xy} \zeta_{\delta_o}(x, y)
\]

\[
= \sum_{xy} a_{xy} - \sum_{x} a_{xx} = \sum_{x} A_{x} - \sum_{x} a_{xx}
\]

In a similar way, we show that \(\text{card}(\delta_o \cap \delta_o \cap i') = \sum_{xy} a_{xy} a_{xx}\),

\[
\tau_x(\delta_o \cap i') = \text{card}[x(\delta_o \cap i')] = \sum_{y} a_{xy} - \sum_{x} a_{xx} = A_{x} - \sum_{x} a_{xx},
\]

\[
\ell_x(\delta_o \cap i') = \text{card}[(\delta_o \cap i')x] = \sum_{y} a_{yx} - \sum_{x} a_{xx} = A_{x} - \sum_{x} a_{xx}
\]

and finally,

\[
\sum_{x \in X} \left[ r_x(\delta_o \cap i') - \ell_x(\delta_o \cap i') \right]^2 = \sum_{x} (A_{x} - A_{x}. - A_{x})^2.
\]

Hence, \(\text{Var}[\gamma(\delta_o, \cdot)] = \left(\frac{1}{3}\right) \left[ \text{card}(\delta_o \cap i') - \text{card}(\delta_o \cap \delta_o \cap i') \right] + \)

\[
\sum_{x \in X} \left[ r_x(\delta_o \cap i') - \ell_x(\delta_o \cap i') \right]^2
\]

\[
= \left(\frac{1}{3}\right) \left[ \sum_{x} A_{x} - \sum_{xy} a_{y}a_{xy} + \sum_{x} \left( A_{x} - A_{x}. - A_{x} \right)^2 \right]
\]

which is the same as (2.2.1).

Using Lemma 2.1, we prove Theorem 2.1.

Theorem 2.1: For the formal situation \((\Delta_{k}, \Omega, \gamma)\), the mean of \(\gamma(\delta_o, \cdot)\), as \(\omega\) runs through \(\Omega\) and \(\delta_o \in \Delta_k\), is

\[
M[\gamma(\delta_o, \cdot)] = \frac{1}{2} N(N-1),
\]

and the variance is

\[
\text{Var}[\gamma(\delta_o, \cdot)] = \left(\frac{1}{3}\right) \sum_{x} (\Sigma^x n_{1} n_{j})^2 - \Sigma^x n_{1} n_{j} n_{t}
\]

where \(\Sigma^x\) indicates that summation has been taken over all values of the suffixes which are in increasing order. For example, \(\Sigma^x n_{1} n_{j} n_{t} = \)
Proof: The formulas given in Lemma 2.2, hold also for the formal situation $(\Delta_k, \Omega, \gamma)$. Hence we have,

\[ M[\gamma(\delta_0, \cdot)] = \left(\frac{1}{2}\right) N(N-1) . \]

Now, \[ \text{Var}[\gamma(\delta_0, \cdot)] = \left(\frac{1}{3}\right) \sum_{x} (A_{x'} - A_x)^2 + \sum_{x} A_{x'} . \]

Suppose \( x \in D_1 \), then \( A_{x'} = \frac{k}{i=2} n_i \), and \( A_x = 0 \).

If \( x \in D_2 \), then \( A_{x'} = \frac{k}{i=2} n_i \), and \( A_x = n_1 \).

Continuing thus, if \( x \in D_r \), then \( A_{x'} = \sum_{i=r+1}^{k} n_i \), and \( A_x = \sum_{i=1}^{r-1} n_i \).

Finally, if \( x \in D_k \), \( A_{x'} = 0 \), \( A_x = \sum_{i=1}^{k-1} n_i \).

Also, \( \sum x A_{x'} = \sum n_i n_j \).

We have, \( \sum x (A_{x'} - A_x)^2 \)

\[ = n_1 (n_2 + n_3 + \ldots + n_k)^2 + n_2 (n_3 + n_4 + \ldots + n_k - n_1)^2 + n_3 (n_4 + \ldots + n_k - n_1 - n_2)^2 + \ldots + n_k (n_1 - n_2 - \ldots - n_{k-1})^2 \]

\[ = n_1 (S-n_1^2) + n_2 (S-n_2^2) + \ldots + n_k (S-n_k^2) + \text{(product terms)}, \]

where \( S = \sum_{i=1}^{k} n_i^2 \).

Now, consider the product terms. They will be of the form \( n_i n_j n_t \), with \( i < j < t \). For fixed \( i, j, t \), we will get \( 2 n_i n_j n_t \) from the term

\[ n_i (n_{i+1} + \ldots + n_j + \ldots + n_t + \ldots + n_k - n_1 - n_2 - \ldots - n_{i-1})^2 . \]

We will also get \( -2 n_i n_j n_t \) from the term

\[ n_j (n_{j+1} + \ldots + n_t + \ldots + n_k - n_1 - \ldots - n_i - \ldots - n_{j-1})^2 . \]
Lastly, we get $2n_t(n_t+\ldots+n_k-n_l-\ldots-n_i-\ldots-n_j-\ldots-n_{t-1})^2$.

Thus, we have

$$
\sum_x (A_x - \hat{A}_x)^2 = N.S. - \frac{k}{i=1} n_i^3 + Z_{ij} \hat{n}_i \hat{n}_j = \sum_{i=1}^k n_i^3 + \sum_{i=1}^k n_i^2 + \sum_{i=1}^k n_i \hat{n}_j
$$

$$
= n_1^2N + n_1n_3N + \ldots + n_1n_kN + n_2n_3N + \ldots + n_2n_kN + \ldots
$$

$$
+ n_{k-1}n_kN - n_1n_2n_3 - n_1n_2n_4 - \ldots - n_1n_2n_k - \ldots
$$

$$
- n_{k-2}n_{k-1}n_k
$$

$$
= N(\sum n_i) - \sum n_i \hat{n}_j
$$

Hence, $\text{Var}(\gamma(\delta_0, \nu)) = \left(\frac{1}{3}\right) \left[\frac{1}{X} \sum (A_x - \hat{A}_x)^2 + \sum A_x\right]$

$$
= \left(\frac{1}{3}\right) \left[\sum (n+1)(\sum n_i) - \sum n_i \hat{n}_j\right].
$$

If $N$ is sufficiently large, then $\gamma(\delta_0, \nu)$ has an approximate normal distribution, and the characterizing power is the upper-tail of this distribution, and is determined by $\gamma(\delta_0, \omega)$.

### 2.3. Descriptive Power of Ordered Polytomies for the Formal Situation $(\Delta K, \Omega, \gamma)$

Consider the formal situation $(\Delta K, \Omega, \gamma)$ described previously. Let $\omega_0 \in \Omega$ be a fixed strong ordering, and $\delta$ run through $\Delta K$.

$$
\gamma(\delta, \omega_0) = \sum_{x,y}^\# t_{\omega_0}(x, y) + \sum_{x,y}^\# t_{\delta}(x, y) - \sum_{x,y}^\# t_{\omega_0}(x, y) \cdot t_{\delta}(x, y).
$$
Now,

\[ \sum_{x,y} ^\# \zeta_\delta (x,y) = \frac{1}{4} N(N-1). \]

Since \( \Delta_k \) is the set of all ordered \( k \)-tomes of the form \( D_1 < D_2 < \ldots < D_k \) with cardinalities of \( D_1, D_2, \ldots, D_k \) fixed for \( \Delta_k \),

\[ \sum_{x,y} ^\# \zeta_\delta (x,y) = \text{card (}\delta\text{)} = \sum n_i n_j. \]

Now,

\[ \zeta_\delta (x,y) = 1 \quad \text{if } (x,y) \in \delta, \]

\[ = 0 \quad \text{otherwise}. \]

Consider the cases when \( \zeta_\delta (x,y) = 1 \). The function \( \zeta_\delta (x,y) = 1 \) if \( x \in D_1 \) and \( y \in D_2 \) or \( y \in D_3 \) or \ldots or \( y \in D_k \). So, for these cases

\[ E[\zeta_\delta (x,y) = 1] = \frac{n_1 n_2}{N(N-1)} + \frac{n_1 n_3}{N(N-1)} + \ldots + \frac{n_1 n_k}{N(N-1)}. \]

The function \( \zeta_\delta (x,y) = 1 \) if \( x \in D_2 \) and \( y \in D_3 \) or \( y \in D_4 \) or \ldots or \( y \in D_k \). So for these cases

\[ E[\zeta_\delta (x,y) = 1] = \frac{n_2 n_3}{N(N-1)} + \frac{n_2 n_4}{N(N-1)} + \ldots + \frac{n_2 n_k}{N(N-1)}. \]

Continuing thus, consider the case when \( \zeta_\delta (x,y) = 1 \) for \( x \in D_{k-1} \) and \( y \in D_k \). So, \( E[\zeta_\delta (x,y) = 1] = \frac{n_{k-1} n_k}{N(N-1)} \) for this case. Hence, \( E[\zeta_\delta (x,y) = 1] = \frac{[N(N-1)]^{-1}}{N(N-1)} \zeta_\delta (x,y) \) and \( E[\zeta_\delta (x,y)] = \frac{[N(N-1)]^{-1}}{N(N-1)} \zeta_\delta (x,y). \)

\[ E[\gamma(\ast, \omega_0)] = \frac{1}{4} N(N-1) + \sum n_j \omega_0 = \sum_{x,y} ^\# \zeta_\delta (x,y) \quad \text{for } \omega_0 = 0, \ldots, n-1. \]

\[ = \frac{1}{4} N(N-1) + \sum n_j \omega_0 \quad \text{for } \omega_0 = 0, \ldots, n-1. \]
\[= \frac{1}{2}N(N-1) + \sum_{i=1}^{N} n_i n_j - 2(N(N-1))^{-1} (\sum_{i=1}^{N} n_i) \cdot \frac{1}{2}N(N-1)\]

Since \(\sum_{x,y \neq (x,y)}^{N \times N} (x,y) = \frac{1}{2}N(N-1)\) is a constant, and \(\sum_{x,y \neq (x,y)}^{N \times N} \zeta_0(x,y) = \text{card} (\delta \cap i')\)

\[= \sum_{i=1}^{N} n_i n_j\] is a constant, both these quantities will not affect \(\text{Var} [y \gamma (y, \omega_0)]\). So,

\[\text{Var} [\delta(x, \omega_0)] = \text{Var} \left[ -2 \sum_{x,y \neq (x,y)}^{N \times N} \zeta_0(x,y) \cdot \zeta_0(x,y) \right] \]

\[= \frac{4\sum_{x,y \neq (x,y)}^{N \times N} \zeta_0(x,y)^2 \cdot \text{Var} (\zeta_0(x,y)) + \sum_{x,y \neq (x,y)}^{N \times N} \zeta_0(x,y) \cdot \zeta_0(z,t) \cdot \text{Cov} (\zeta_0(x,y), \zeta_0(z,t)) \cdot \text{Cov} (\zeta_0(x,y), \zeta_0(z,t)) \cdot (x,y) \neq (z,t) \]

Now, \(\text{Var} (\zeta_0(x,y)) = \left[ N(N-1) \right]^{-1} \sum_{i=1}^{N} n_i n_j - \left[ N(N-1) \right]^{-2} (\sum_{i=1}^{N} n_i)^2 \)

To find \(\text{Cov} (\zeta_0(x,y), \zeta_0(z,t))\), consider the following six cases:

Case 1: \(x \neq y \neq z \neq t\).
Case 2: \(x = z, y \neq t\).
Case 3: \(x = t, y = z\).
Case 4: \(x = t, y \neq z\).
Case 5: \(x \neq z, y = t\).
Case 6: \(x \neq t, y = z\).

We find, for each case, with what probability \(\zeta_0(x,y) \cdot \zeta_0(z,t) = 1\).

Consider the following notation: \(x|yz|t\) means that \(x \in D_i\), \(y\) and \(z\) are in class \(D_j\) with \(i < j\), and \(t \in D_s\) with \(j < s\). That is, when the letters \(x, y, z, t\) are separated by bars, it indicates, that they are in different classes. The letters are arranged according to the order of the classes to which they belong - the letters belonging to the lower classes appear
before the letters belonging to the higher classes.

Case 1: \( x \neq y \neq z \neq t \).

The product \( \zeta_\delta(x,y) \cdot \zeta_\delta(z,t) \) will equal one, only under certain conditions. For example, if \( x, y, z \in D_1 \) and \( t \in D_2 \), then \( \zeta_\delta(z,t) = 1 \), but \( \zeta_\delta(x,y) = 0 \). So, the product \( \zeta_\delta(x,y) \cdot \zeta_\delta(z,t) = 0 \).

Of all the possible arrangements of \( x, y, z, t \), in the classes \( D_1, D_2, \ldots, D_k \), we consider those that will give us \( \zeta_\delta(x,y) \cdot \zeta_\delta(z,t) = 1 \). A simple enumeration shows that, the only cases that give a non-zero product are the following:

1) \( x | z | y | t \)  
2) \( x | z | t | y \)  
3) \( z | x | y | t \)  
4) \( z | x | t | y \)  
5) \( x | y | z | t \)  
6) \( z | t | x | y \)  
7) \( x z | y | t \)  
8) \( x z | t | y \)  
9) \( z | x t | y \)  
10) \( x | z t | y \)  
11) \( x | z t | y \)  
12) \( z | x y t \)  
13) \( x z | y t \).

Hence, \( P[\zeta_\delta(x,y) \cdot \zeta_\delta(z,t) = 1] = E[\zeta_\delta(x,y) \cdot \zeta_\delta(z,t)] \)

\[ = \left[ N(N-1)(N-2)(N-3) \right]^{-1} \left[ \Sigma i \cdot n_i (n_i-1) n_j (n_j-1) + 2 \Sigma i \cdot n_i (n_i-1) n_j n_s \right. \]
\[ + 2 \Sigma i \cdot n_i n_j (n_j-1) n_s + 2 \Sigma i \cdot n_j n_j (n_j-1) n_s + 6 \Sigma i \cdot n_j n_j n_j n_j \] .

So, \( \text{Cov} [\zeta_\delta(x,y), \zeta_\delta(z,t)] = \left[ N(N-1)(N-2)(N-3) \right]^{-1} \cdot \]

\[ \left[ \Sigma i \cdot n_i (n_i-1) n_j (n_j-1) + 2 \Sigma i \cdot n_i (n_i-1) n_j n_s + 2 \Sigma i \cdot n_i n_j (n_j-1) n_s \right. \]
\[ + 2 \Sigma i \cdot n_i n_j n_j (n_j-1) + 6 \Sigma i \cdot n_j n_j n_j n_j \] - \[ N(N-1) \right]^{-2} \cdot (\Sigma i \cdot n_i n_j)^2 .

Case 2: \( x = z, y \neq t \).

Of all the possible arrangements of \( x, y, t \) in the classes \( D_1, \ldots, D_k \), the only ones that give \( \zeta_\delta(x,y) \cdot \zeta_\delta(x,t) = 1 \) are the following:
1) \(x|y|t\)  
2) \(x|y|t\)  
3) \(x|t|y\).

Hence, \(\mathbb{E}[(\zeta_0(x,y) \cdot \zeta_0(x,t)) = 1] = \mathbb{E}[(\zeta_0(x,y) \cdot \zeta_0(x,t))\]

\[= \left[\frac{N(N-1)(N-2)}{N(N-1)(N-2)}\right]^{-1} \left[\Sigma_{i,j} n_i n_j (n_j - 1) + 2\Sigma_{i,j} n_i n_j n_s\right].\]

So, \(\text{Cov} \{\zeta_0(x,y), \zeta_0(x,t)\} = \left[\frac{N(N-1)(N-2)}{N(N-1)(N-2)}\right]^{-1} \cdot \left[\Sigma_{i,j} n_i n_j (n_j - 1) + 2\Sigma_{i,j} n_i n_j n_s\right] - \left[\frac{N(N-1)(N-2)}{N(N-1)(N-2)}\right]^{-2} \cdot \left(\Sigma_{i,j} n_i n_j\right)^2.\]

Case 3: \(x = t, \ y = z\).

We cannot have \(\zeta_0(x,y) = 1\) and \(\zeta_0(y,x) = 1\). So, \(\mathbb{E}[(\zeta_0(x,y) \cdot \zeta_0(y,x)) = 1] = \mathbb{E}[(\zeta_0(x,y) \cdot \zeta_0(y,x))\]

\[= \frac{N(N-1)(N-2)}{N(N-1)(N-2)} \cdot \left[\Sigma_{i,j} n_i n_j (n_j - 1) + 2\Sigma_{i,j} n_i n_j n_s\right] - \left[\frac{N(N-1)(N-2)}{N(N-1)(N-2)}\right]^{-2} \cdot \left(\Sigma_{i,j} n_i n_j\right)^2.\]

Case 4: \(x = t, \ y \neq z\).

Of all the possible arrangements of \(x, y, z\) in the classes \(D_1, \ldots, D_k\), the following gives \(\zeta_0(x,y) \cdot \zeta_0(z,x) = 1:\)

1) \(z|x|y\)

Hence, \(\mathbb{E}[(\zeta_0(x,y) \cdot \zeta_0(z,x)) = 1] = \mathbb{E}[(\zeta_0(x,y) \cdot \zeta_0(z,x))\]

\[= \left[\frac{N(N-1)(N-2)}{N(N-1)(N-2)}\right]^{-1} \left[\Sigma_{i,j} n_i n_j (n_j - 1) + 2\Sigma_{i,j} n_i n_j n_s\right].\]

So,

\[
\text{Cov} \{\zeta_0(x,y), \zeta_0(z,x)\} = \left[\frac{N(N-1)(N-2)}{N(N-1)(N-2)}\right]^{-1} \cdot \left[\Sigma_{i,j} n_i n_j (n_j - 1) + 2\Sigma_{i,j} n_i n_j n_s\right]
- \left[\frac{N(N-1)(N-2)}{N(N-1)(N-2)}\right]^{-2} \cdot \left(\Sigma_{i,j} n_i n_j\right)^2.
\]

Case 5: \(x \neq z, \ y = t\).

Of all the possible arrangements of \(x, y, z\) in the classes \(D_1, \ldots, D_k\), the following give \(\zeta_0(x,y) \cdot \zeta_0(z,y) = 1:\)

1) \(xz|y\)  
2) \(x|z|y\)  
3) \(z|x|y\)
So,
\[
E[\xi_0(x,y) \cdot \xi_0(z,y)] = E[\xi_0(x,y) \cdot \xi_0(z,y)] = [N(N-1)(N-2)]^{-1} \cdot \left[ \sum i n_i(n_i-1)n_j + 2 \sum i n_i n_j n_s \right].
\]
Hence, Cov \{\xi_0(x,y), \xi_0(z,y)\} = \left[ [N(N-1)(N-2)]^{-1} \cdot \left[ \sum i n_i(n_i-1)n_j + 2 \sum i n_i n_j n_s \right] - [N(N-1)]^{-2} \cdot (\sum i n_i n_j)^2 \right].

Case 6: \(x \neq t, y = z\).

Of all the possible arrangements of \(x, y, t\) in the classes \(D_1, \ldots, D_k\), the one that gives \(\xi_0(x,y) \cdot \xi_0(y,t) = 1\), is:
1) \(x|y|t\).

So,
\[
E[\xi_0(x,y) \cdot \xi_0(y,t)] = E[\xi_0(x,y) \cdot \xi_0(y,t)] = [N(N-1)(N-2)]^{-1} \cdot [\sum i n_i n_j n_s].
\]
Hence, Cov \{\xi_0(x,y), \xi_0(y,t)\} = \left[ [N(N-1)(N-2)]^{-1} \cdot \left[ \sum i n_i n_j n_s \right] - [N(N-1)]^{-2} \cdot (\sum i n_i n_j)^2 \right].

All the covariance terms obtained are free of \(x, y, z, t\) and so can be taken outside the summation sign. Consider next,

\[
\sum_{x,y,z,t} \sum_{\omega_0} \xi_0(x,y) \cdot \xi_0(z,t) \text{ for the six cases.}
\]

Case 1: \(x \neq y \neq z \neq t\).

\[
\sum_{x,y,z,t} \sum_{\omega_0} \xi_0(x,y) \cdot \xi_0(z,t) = \binom{N}{4} \cdot 6 = \frac{N(N-1)(N-2)(N-3)}{4}.
\]
Case 2: $x = z, y \neq t$.
\[ \sum_{x, y} \sum_{z, t} \zeta_{\omega_0}(x, y) \cdot \zeta_{\omega_0}(x, t) = \binom{N}{3} \cdot 2 = \frac{N(N-1)(N-2)}{3} \]

Case 3: $x = t, y = z$.
\[ \sum_{x, y} \sum_{y, x} \zeta_{\omega_0}(x, y) \cdot \zeta_{\omega_0}(y, x) = 0, \text{ because we cannot have } \zeta_{\omega_0}(x, y) = 1 \]
and $\zeta_{\omega_0}(y, x) = 1$.

Case 4: $x = t, y \neq z$.
\[ \sum_{x, y} \sum_{z, t} \zeta_{\omega_0}(x, y) \cdot \zeta_{\omega_0}(z, x) = \binom{N}{3} \cdot 1 = \frac{N(N-1)(N-2)}{6} \]

Case 5: $x \neq z, y = t$.
\[ \sum_{x, y} \sum_{z, t} \zeta_{\omega_0}(x, y) \cdot \zeta_{\omega_0}(z, y) = \binom{N}{3} \cdot 2 = \frac{N(N-1)(N-2)}{3} \]

Case 6: $x \neq t, y = z$.
\[ \sum_{x, y} \sum_{z, y} \zeta_{\omega_0}(x, y) \cdot \zeta_{\omega_0}(y, t) = \binom{N}{3} \cdot 1 = \frac{N(N-1)(N-2)}{6} \]

So,
\[ \text{Var}[\chi(\omega_0)] = 4\sum_{x, y} \left[ \zeta_{\omega_0}(x, t) \right]^2 \text{Var}(\zeta_0(x, y)) \\
+ 4\sum_{x, y} \sum_{z, t} \zeta_{\omega_0}(x, y) \cdot \zeta_{\omega_0}(z, t) \cdot \text{Cov}[\zeta_0(x, y), \zeta_0(z, t)] \\
= 4\left[ \frac{N(N-1)}{2} \right] \cdot \left[ \sum_{i} n_{i} n_{j} - \binom{N}{2} \left( \sum_{i} n_{i} n_{j} \right)^2 \right] \cdot \frac{1}{4} \cdot N(N-1) \\
+ 4\left[ \frac{N(N-1)(N-2)(N-3)}{24} \right] \cdot \left[ \sum_{i} n_{i} n_{j} (n_{i} - 1)(n_{j} - 1) \right] \\
+ 2\sum_{i} n_{i} (n_{i} - 1)n_{j} n_{s} + 2\sum_{i} n_{i} n_{j} (n_{i} - 1)n_{s} \\
+ 2\sum_{i} n_{i} n_{j} n_{s} (n_{s} - 1) + 6\sum_{i} n_{i} n_{j} n_{s} n_{m} \\
- \frac{N(N-1)}{2} \left[ \sum_{i} n_{i} n_{j} \right]^2 \cdot \frac{N(N-1)(N-2)(N-3)}{4} \]
\[ + 4\left[\frac{N(N-1)(N-2)}{6}\right]^{-1}\left\{ \sum_{i=1}^{N} n_i (n_i-1) + 2\sum_{i=1}^{N} n_i n_j n_s \right\} \\
- \left[\frac{N(N-1)}{2}\right]^{-2} \cdot (\sum_{i=1}^{N} n_i)^2 \\
+ 4\left[\frac{N(N-1)(N-2)}{6}\right]^{-2} \cdot (\sum_{i=1}^{N} n_i)^2 \cdot 0 \\
+ 4\left[\frac{N(N-1)(N-2)}{6}\right]^{-1} \cdot \left\{ \sum_{i=1}^{N} n_i n_j n_s \right\} - \left[\frac{N(N-1)}{2}\right]^{-2} \cdot (\sum_{i=1}^{N} n_i)^2 \\
\cdot \frac{N(N-1)(N-2)}{6} \\
+ 4\left[\frac{N(N-1)(N-2)}{6}\right]^{-1} \cdot \left\{ \sum_{i=1}^{N} n_i (n_i-1)n_j + 2\sum_{i=1}^{N} n_i n_j n_s \right\} \\
- \left[\frac{N(N-1)}{2}\right]^{-2} \cdot (\sum_{i=1}^{N} n_i)^2 \cdot \frac{N(N-1)(N-2)}{6} \\
+ 4\left[\frac{N(N-1)(N-2)}{6}\right]^{-1} \cdot \left\{ \sum_{i=1}^{N} n_i n_j n_s \right\} - \left[\frac{N(N-1)}{2}\right]^{-2} \cdot (\sum_{i=1}^{N} n_i)^2 \\
\cdot \frac{N(N-1)(N-2)}{6} \]

The sum of all the negative terms is:
\[ \frac{N^2}{2} (N-1)^2 \cdot (-4) \left[\frac{N(N-1)}{2}\right]^{-2} \cdot (\sum_{i=1}^{N} n_i)^2 \]

= \(- \sum_{i=1}^{N} n_i (n_i)^2\)

The sum of all the positive terms is:
\[ \left(\frac{1}{3}\right) \sum_{i=1}^{N} n_i n_j n_k n_l + \sum_{i=1}^{N} n_i^2 n_j^2 + \sum_{i=1}^{N} n_i^2 n_j + \left(\frac{1}{3}\right) n_j n_k + \left(\frac{1}{3}\right) n_l n_j \]

\[ + 2\sum_{i=1}^{N} n_i^2 n_j n_s + 2\sum_{i=1}^{N} n_i n_j n_s + 2\sum_{i=1}^{N} n_i n_j n_s^2 \]

\[ + \left(\frac{2}{3}\right) \sum_{i=1}^{N} n_i n_j n_s + 6\sum_{i=1}^{N} n_i n_j n_s \cdot n_m \]

By adding and subtracting appropriate terms, and collecting together suitable terms, the above sum becomes:
\[ \sum_{i=1}^{N} n_i n_j n_s \]
\[ \left(\frac{1}{3}\right) \sum_{i=1}^{N} n_i (n_i)^2 \cdot N(N+1) \cdot \left(\sum_{i=1}^{N} n_i n_j \right) \]

\[ = \left(\sum_{i=1}^{N} n_i n_j \right)^2 + \left(\frac{1}{3}\right) N(N+1) \left(\sum_{i=1}^{N} n_i n_j \right) \]

Putting together the positive and negative terms of Var \([\gamma_{(\cdot, \omega_0)}]\),
we get \( \text{Var}[\gamma(\cdot, \omega_0)] = \left(\frac{1}{3}\right)[(N+1)(\sum^\omega n_i n_j) - \sum^\omega n_i n_j n_t] \).

So, we have:

Theorem 2.2: For the formal situation \((\Delta_k, \Omega, \gamma)\), the mean of \(\gamma(\cdot, \omega_0)\), as \(\delta\) runs through \(\Delta_k\), and \(\omega_0 \in \Omega\), is \(M[\gamma(\cdot, \omega_0)] = \frac{1}{2}N(N-1)\), and the variance is

\[
\text{Var} \{\gamma(\cdot, \omega_0)\} = \left(\frac{1}{3}\right)[(N+1)(\sum^\omega n_i n_j) - \sum^\omega n_i n_j n_t].
\]

If \(N\) is sufficiently large, then \(\gamma(\cdot, \omega_0)\) has an approximate normal distribution. The descriptive power is the upper-tail of this distribution and is determined by \(\gamma(\delta_0, \omega_0)\).

2.4. Characterizing and Descriptive Powers

for the Formal Situation \((\Delta^*_k, \Omega, \gamma)\)

Consider the formal situation \((\Delta^*_k, \Omega, \gamma)\), where \(\Delta^*_k\) is the set of all possible ordered \(k\)-tomies. Let \(\Omega\) and \(\gamma\) be as defined before.

\[
\gamma(\delta_0, \omega_0) = \sum_{x,y} [\zeta_{\delta_0}(x,y)[\zeta_{\omega_0}(x,y)] + [1 - \zeta_{\delta_0}(x,y)]\zeta_{\omega_0}(x,y)],
\]

where, \(\delta_0 \in \Delta^*_k\) and \(\omega_0 \in \Omega\). Let \(\delta_0\) be a fixed, ordered \(k\)-tomy in \(\Delta^*_k\) of the form \(D_1 < D_2 < \ldots < D_k\), with \(\text{card}(D_1) = n_1\), \(\text{card}(D_2) = n_2\), \ldots, \(\text{card}(D_k) = n_k\), so that \(\sum_{i=1}^{k} n_i = N = \text{card}(\Omega)\).

Then, as \(\omega\) runs through \(\Omega\), by Theorem 2.1, the mean \(M[\gamma(\delta_0, \cdot)] = \frac{1}{2}N(N-1)\) and the variance \(\text{Var} \{\gamma(\delta_0, \cdot)\} = \left(\frac{1}{3}\right)[(N+1)(\sum^\omega n_i n_j) - \sum^\omega n_i n_j n_t]\). If \(N\) is sufficiently large, then \(\gamma(\delta_0, \cdot)\) has approximately a normal distribution. The characterizing power is the upper-tail of this distribution, and is determined by \(\gamma(\delta_0, \omega_0)\).

For the same formal situation, consider the descriptive power.
\[
\gamma(\delta, \omega) = \sum_{x, y} \xi_\delta(x, y) + \sum_{x, y} \left[1 - 2\xi_\omega(x, y)\right] \cdot \xi_\delta(x, y) .
\]

Now, \(\sum_{x, y} \xi_\omega(x, y) = \frac{1}{2}N(N-1)\) is a constant.

\[
\zeta_\delta(x, y) = 1 \quad \text{if} \quad (x, y) \in \delta
\]
\[
= 0 \quad \text{otherwise}.
\]

In all, there are \(k + k(k-1) = k^2\) possible arrangements of \(x, y\) in the classes \(D_1, D_2, ..., D_k\). Consider those arrangements which will give \(\zeta_\delta(x, y) = 1\). There are \((k-1) + (k-2) + ... + 3 + 2 + 1 = \frac{k(k-1)}{2}\) arrangements of \(x, y\) in \(D_1, D_2, ..., D_k\) such that \(\zeta_\delta(x, y) = 1\).

So, \(F[\zeta_\delta(x, y) = 1] = E[\zeta_\delta(x, y)] = \frac{k(k-1)}{2} \cdot \frac{1}{k^2} = \frac{(k-1)}{2k}\), and

\[
\text{Var} \{\zeta_\delta(x, y)\} = \frac{(k-1)}{4k^2} - \frac{(k-1)^2}{4k^2} = \frac{(k-1)(k+1)}{4k^2} .
\]

Hence,

\[
E[\gamma(\delta, \omega)] = \sum_{x, y} \xi_\delta(x, y) + \sum_{x, y} \left[1 - 2\xi_\omega(x, y)\right] E[\zeta_\delta(x, y)]
\]
\[
= \frac{1}{2}N(N-1) + \frac{(k-1)}{2k} \sum_{x, y} \left[1 - 2\xi_\omega(x, y)\right]
\]
\[
= \frac{1}{2}N(N-1) + \frac{(k-1)}{2k} N(N-1) - \left(\frac{k-1}{2k}\right) \cdot 2 \cdot \frac{1}{2}N(N-1)
\]
\[
= \frac{1}{2}N(N-1)
\]

While finding \(\text{Var} \{\gamma(\delta, \omega)\}\), \(\sum_{x, y} \xi_\delta(x, y)\) need not be considered, as it is a constant. So, \(\text{Var} \{\gamma(\delta, \omega)\} = \)

\[
\text{Var} \left[\sum_{x, y} \left[1 - 2\xi_\omega(x, y)\right] \zeta_\delta(x, y)\right]
\]
\[
= \sum_{x, y} \left[1 - 2\xi_\omega(x, y)\right]^2 \text{Var} \{\zeta_\delta(x, y)\}
\]
The quantity \( [1-2\zeta_{(x,y)}(x,y)]^2 \) is always 1.

So,

\[
\sum_{x,y,z,t} [1-2\zeta_{(x,y)}(x,y)]^2 \text{Var} \{\zeta_{(x,y)}(x,y)\} = \sum_{x,y} \text{Var} \{\zeta_{(x,y)}(x,y)\}
\]

\[
= \frac{(k-1)(k+1)}{4k^2} \cdot N(N-1).
\]

To find \( \text{Cov} \{\zeta_{(x,y)}(x,y),\zeta_{(z,t)}(z,t)\} \), consider the following six cases:

Case 1: \( x \neq y \neq z \neq t \).

Case 2: \( x = z, y \neq t \).

Case 3: \( x = t, y = z \).

Case 4: \( x = t, y \neq z \).

Case 5: \( x \neq z, y = t \).

Case 6: \( x \neq t, y = z \).

Case 1: \( x \neq y \neq z \neq t \).

The total number of arrangements of \( x, y, z, t \) in \( D_1, D_2, \ldots, D_k \) is

\[
\left(\begin{array}{c}
k \\frac{k}{1} + \left(\begin{array}{c}
k \\frac{2}{2} \times \frac{4}{2} \right) \times 2 \times \left(\begin{array}{c}
k \\frac{2}{2} \times \frac{4}{3} \right) \times 3 \times \left(\begin{array}{c}
k \\frac{4}{4} \right) \times 4!
\end{array}\right).
\right.
\]

The term \( \left(\begin{array}{c}
k \\frac{k}{1} \right) \) gives the number of arrangements in which all four elements \( x, y, z, t \) belong to the same class. The term \( \left(\begin{array}{c}
k \\frac{2}{2} \times \frac{4}{3} \right) \) gives the number of arrangements in which two of the elements belong to one class, and the remaining two elements belong to another class. The term \( \left(\begin{array}{c}
k \\frac{4}{4} \right) \) gives the number of arrangements in which three elements belong to one class, and the last element belongs to another class. The term
\binom{k}{3} \cdot \binom{4}{2} \cdot 3! \text{ gives the number of arrangements in which, two elements belong to one class, and the remaining two elements belong to two other classes. Lastly, the term } \binom{k}{4} \cdot 4! \text{ gives the number of arrangements in which, each of } x, y, z, t \text{ belongs to a different class.}

Now, consider the arrangements that give } \zeta_0(x,y) \cdot \zeta_0(z,t) = 1. \text{ There are } \frac{k(k-1)}{2} \text{ arrangements such that } \zeta_0(x,y) = 1. \text{ Similarly, there are } \frac{k(k-1)}{2} \text{ arrangements such that } \zeta_0(z,t) = 1. \text{ Hence, there are } \left[\frac{k(k-1)}{2}\right]^2 = \frac{k^2(k-1)^2}{4} \text{ arrangements of } x, y, z, t \text{ such that } \zeta_0(x,y) \cdot \zeta_0(z,t) = 1. \text{ So,}

\[
P[\zeta_0(x,y) \cdot \zeta_0(z,t) = 1] = E[\zeta_0(x,y) \cdot \zeta_0(z,t)]
\]

\[
= \frac{k^2(k-1)^2}{4} \cdot \frac{1}{k^4} = \frac{(k-1)^2}{4k^2}.
\]

Then, \( \text{Cov} \{ \zeta_0(x,y), \zeta_0(z,t) \} = \frac{(k-1)^2}{4k^2} - \left[\frac{(k-1)}{2k}\right]^2 = 0. \)

So,

\[
\sum \sum \sum 1 - 2\zeta_0(x,y) \cdot 1 - 2\zeta_0(z,t) \sum \text{ Cov} \{ \zeta_0(x,y), \zeta_0(z,t) \} = 0.
\]

\text{Case 2: } x=z, \ y \neq t.

The total number of arrangements of } x, y, t \text{ in } D_1, \ldots, D_k, \text{ is } \binom{k}{1} \cdot 1 + 2 \cdot \binom{k}{2} \cdot \binom{3}{2} + \binom{k}{3} \cdot 3! = k^3. \text{ To obtain this, we use the same reasoning as in Case 1. The number of arrangements such that } \zeta_0(x,y) \cdot \zeta_0(x,t) = 1 \text{ is } \frac{k(k-1)(2k-1)}{6}. \text{ It is easy to see this. When } x \in D_1, \text{ there are } (k-1) \text{ places for } y \text{ and } (k-1) \text{ places for } t. \text{ So, there are } (k-1)^2 \text{ possible arrangements. When } x \in D_2, \text{ there are } (k-2) \text{ places for } y \text{ and } (k-2) \text{ places for } t. \text{ So, there are } (k-2)^2 \text{ arrangements. Continuing thus, the total number of arrangements such that } \zeta_0(x,y) \cdot \zeta_0(x,t) = 1 \text{ is } (k-1)^2 + (k-2)^2 + \ldots + (k-n)^2, \text{ where } n \text{ is the number of classes.} \)
\[(0) + (1-N)^{n_2} \cdot 2 - (1-N)^{n_3} \cdot 2 - (1-N)N =
\]
\[(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \xi^{n_1} + (x^r\xi)^{n_2} \xi^{n_1} - (x^r\xi)^{n_2} \xi^{n_1} - t \xi^{n_1} =
\]
\[\frac{C_4}{4} = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =
\]
\[I = (x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \] \quad \text{and} \quad I = (x^r\xi)^{n_2} \text{ so if we happen that case is}
\]
\[z = x, \quad I = I^I \cdot 2
\]
\[(Z-N)(1-N)N \cdot \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon = \]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
\[\{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \quad \text{and} \quad \epsilon = \{(x^r\xi)^{n_2} \cdot (x^r\xi)^{n_3} \} \epsilon =\]
Therefore,

\[ x_{xy} \left[ 1 - 2 \zeta_0(x, y) \right] \left[ 1 - 2 \zeta_0(y, x) \right] \text{Cov} \left[ \zeta_0(x, y), \zeta_0(y, x) \right] \]

\[ = \frac{(k-1)^2}{4k^2} \cdot N(N-1) \]

Case 4: \( x = t, y \neq z \).

The total number of arrangements of \( x, y, z \), in \( D_1, \ldots, D_k \) is \( k^3 \). Consider now the arrangements which give \( \zeta_0(x, y) \cdot \zeta_0(z, x) = 1 \). When \( x \in D_2 \), there is only 1 place for \( z \) but \( (k-2) \) places for \( y \). So \( 1 \cdot (k-2) \) arrangements are possible. When \( x \in D_3 \), there are 2 places for \( z \) and \( (k-3) \) places for \( y \). So, \( 2 \cdot (k-3) \) arrangements are possible. Continuing thus, when \( x \in D_{k-1} \), there are \( (k-2) \) places possible for \( z \) and only one for \( y \). Hence, \( (k-2) \cdot 1 \) arrangements are possible. So, in all, the number of arrangements such that \( \zeta_0(x, y) \cdot \zeta_0(z, x) = 1 \) is \( (k-2) \cdot 1 + (k-3) \cdot 2 + \ldots + 2 \cdot (k-3) + 1 \cdot (k-2) \)

\[ = (k-2) \cdot 1 + (k-3-1) \cdot 2 + (k-2-2) \cdot 3 + \ldots + (k-2 - k-3) (k-2) \]

\[ = (k-2) + 2(k-2) + 3(k-2) + \ldots + (k-2)(k-2) - 2 - 6 - 12 - \]

\[ (k-3)(k-2) \]

\[ = (k-2) \left[ 1 + 2 + \ldots + (k-2) \right] - 1 \cdot 2 - 2 \cdot 3 - 3 \cdot 4 - \ldots - \]

\[ (k-3)(k-2) \]

\[ = (k-2) \cdot \frac{(k-2)(k-1)}{2} - 1(1+1) - 2(2+1) - 3(3+1) - \ldots - \]

\[ (k-3)(k-3+1) \]

\[ = \frac{(k-2)^2(k-1)}{2} - \left[ 1^2 + 1 + 2^2 + 2 + 3^2 + 3 + \ldots + \right. \]

\[ (k-3)^2 + (k-3) \]
\[
\frac{(k-2)^2(k-1)}{2} - \left[1^2 + 2^2 + 3^2 + \ldots + (k-3)^2 + 1 + 2 + \ldots + (k-3)\right]
\]

\[
= \frac{(k-2)^2(k-1)}{2} - \frac{(k-3)(k-2)(2k-5)}{6} - \frac{(k-3)(k-2)}{2}
\]

\[
= \frac{k(k-1)(k-2)}{6}
\]

So, \(E[\zeta_0(x,y) \cdot \zeta_0(z,x) = 1] = E[\zeta_0(x,y) \cdot \zeta_0(z,x)] = \frac{k(k-1)(k-2)}{6} \cdot \frac{1}{k^3}
\]

\[
= \frac{(k-1)(k-2)}{6k^2}
\]

and

\[
\text{Cov} \{ \zeta_0(x,y), \zeta_0(z,x) \} = \frac{(k-1)(k-2)}{6k^2} - \frac{(k-1)^2}{4k^2} = -\frac{(k-1)(k+1)}{12k^2}.
\]

Also, \(\sum_{x \neq y \neq z} [1 - 2\zeta_{\omega_0}(x,y)] [1 - 2\zeta_{\omega_0}(z,x)]\)

\[
= \sum_{x \neq y \neq z} 1 - 2\sum_{x \neq y \neq z} \zeta_{\omega_0}(x,y) - 2\sum_{x \neq y \neq z} \zeta_{\omega_0}(z,x) + 4\sum_{x \neq y \neq z} \zeta_{\omega_0}(x,y) \cdot \zeta_{\omega_0}(z,x)
\]

\[
= N(N-1)(N-2) - 2\binom{N}{3} - 2\binom{N}{3} + 4\binom{N}{3}
\]

\[
= N(N-1)(N-2).
\]

So, \(\sum_{x \neq y \neq z} [1 - 2\zeta_{\omega_0}(x,y)] [1 - 2\zeta_{\omega_0}(z,x)] \cdot \text{Cov} \{ \zeta_0(x,y), \zeta_0(z,x) \}\)

\[
= -\frac{(k^2-1)}{12k^2} \cdot N(N-1)(N-2).
\]

Case 5: \(x \neq z, y = t\).

The total number of arrangements of \(x, y, z\) in \(D_1, D_2, \ldots, D_k\), is \(k^3\).

Consider the arrangements that give \(\zeta_0(x,y) \cdot \zeta_0(z,y) = 1\). If \(y \in D_2\), there is 1 place for \(x\) and 1 for \(z\). So, \(1^2 = 1\) arrangement is possible. If
y \in \mathcal{D}_2$, there are 2 places for x and 2 for z. So, $2^2$ arrangements are possible. When $y \in \mathcal{D}_4$, $3^2$ arrangements are possible. Continuing thus, when $y \in \mathcal{D}_k$, there are $(k-1)^2$ arrangements possible. So, in all, the number of arrangements of x, y, z such that $\zeta_\delta(x,y) \cdot \zeta_\delta(z,y) = 1$ is $1^2 + 2^2 + 3^2 + \ldots + (k-1)^2 = \frac{k(k-1)(2k-1)}{6}$. So $E[\zeta_\delta(x,y) \cdot \zeta_\delta(z,y)] = \frac{k(k-1)(2k-1)}{6k^2}$, and

$$\text{Cov} [\zeta_\delta(x,y), \zeta_\delta(z,y)] = \frac{(k-1)(2k-1)}{6k^2} - \frac{(k-1)^2}{4k^2}$$

$$= \frac{(k-1)(k+1)}{12k^2}.$$  

As before,

$$\sum_{x \neq y \neq z} \sum \sum [1 - 2\zeta_\omega(x,y)][1 - 2\zeta_\omega(z,y)] = N(N-1)(N-2) - 2 \cdot \binom{N}{3} - 2 \cdot \binom{N}{3} + 4 \cdot \binom{N}{3} \cdot 2 = \binom{5}{3} N(N-1)(N-2).$$

So,

$$\sum_{x \neq y \neq z} [1 - 2\zeta_\omega(x,y)][1 - 2\zeta_\omega(z,y)] \text{Cov} [\zeta_\delta(x,y), \zeta_\delta(z,y)]$$

$$= \frac{(k^2-1)}{12k^2} \binom{5}{3} N(N-1)(N-2).$$

Case 6: x \neq t, y = z.

The total number of arrangements of x, y, t in $D_1, D_2, \ldots, D_k$ is $k^3$.

Consider the arrangements when $\zeta_\delta(x,y) \cdot \zeta_\delta(y,t) = 1$. When $y \in \mathcal{D}_2$, there is 1 place for x and $(k-2)$ for t. When $y \in \mathcal{D}_3$, there are 2 places for x and $(k-3)$ for t and so on. This is the same as in Case 4. So, the total number of arrangements such that $\zeta_\delta(x,y) \cdot \zeta_\delta(y,t) = 1$ is $(k-2) \cdot 1 + (k-3) \cdot 2 + \ldots + 2 \cdot (k-3) + 1 \cdot (k-2) = \frac{k(k-1)(k-2)}{6}$. 

So,

\[ E[\tau_0(x,y) \cdot \tau_0(y,t)] = 1 \]

\[ = \frac{k(k-1)(k-2)}{6} \cdot \frac{1}{k^3} = \frac{(k-1)(k-2)}{6k^2} \]

and

\[ \text{Cov} \{ \tau_0(x,y), \tau_0(y,t) \} = \frac{(k-1)(k-2)}{6k^2} - \frac{(k-1)^2}{4k^2} \]

\[ = - \frac{(k-1)(k+1)}{12k^2} \]

Also,

\[ \sum_{x \neq y} \sum_{t} \left[ 1-2\tau_0(x,y) \right] \left[ 1-2\tau_0(y,t) \right] = N(N-1)(N-2) - 2 \left( \frac{N}{3} \right) - 2 \left( \frac{N}{3} \right) + 4 \left( \frac{N}{3} \right) = N(N-1)(N-2). \]

So,

\[ \sum_{x \neq y} \sum_{t} \left[ 1-2\tau_0(x,y) \right] \left[ 1-2\tau_0(y,t) \right] \text{Cov} \{ \tau_0(x,y), \tau_0(y,t) \} \]

\[ = - \frac{(k^2-1)}{12k^2} \cdot N(N-1)(N-2). \] Therefore, we have

\[ \text{Var} [\gamma(\cdot, \omega_0)] = \sum_{x,y} \left[ 1-2\tau_0(x,y) \right]^2 \text{Var} [\tau_0(x,y)] + \sum_{x,y} \sum_{z,t} \left[ 1-2\tau_0(x,y) \right] \text{Cov} \{ \tau_0(x,y), \tau_0(z,t) \} \]

\[ = \frac{(k^2-1)}{4k^2} N(N-1) + 0 + \frac{(k^2-1)}{12k^2} \cdot \frac{5}{3} N(N-1)(N-2) + \frac{(k-1)^2}{4k^2} \cdot N(N-1) - \frac{(k^2-1)}{12k^2} N(N-1)(N-2) + \frac{(k^2-1)}{12k^2} \cdot \frac{5}{3} N(N-1)(N-2) - \frac{(k^2-1)}{12k^2} N(N-1)(N-2) \]

So, we have:
Theorem 2.3: For the formal situation \((\Delta_k^*, \Omega, \gamma)\), the mean of \(\gamma(\cdot, \omega_0)\), as \(\delta\) runs through \(\Delta_k^*\), and \(\omega_0 \in \Omega\), is \(M[\gamma(\cdot, \omega_0)] = \frac{1}{2}N(N-1)\), and the variance is
\[
\text{Var} \{\gamma(\cdot, \omega_0)\} = \frac{1}{9k^2} N(N-1) \cdot \left( \frac{(k^2-1)N + (k-1)(5k-4)}{2} \right).
\]

If \(N\) is sufficiently large, then \(\gamma(\cdot, \omega_0)\) has an approximate normal distribution. The descriptive power is the upper-tail of this distribution, and is determined by \(\gamma(\delta_o, \omega_0)\).

2.5. Relation Between Finch's Gauge and Some Other Statistics

Finch gave the relation between the gauge \(\gamma\) and the Wilcoxon rank-sum statistic \(W_D\) for the formal situation \((\Delta_2, \Omega, \gamma)\).

\[
\gamma(\delta_D, \omega) = \frac{1}{2}N(N-1) + (N+1)n - 2W_D.
\]

The \(\delta_D\) is a dichotomy \(D^c < D\) with card \((D) = n\), where \(D^c\) is the complement of \(D\). There is also a relation between \(\gamma(\delta, \omega)\) and the Jonckheere statistic \(J\) for the general case \((\Delta_k, \Omega, \gamma)\). The Jonckheere statistic is used to test hypotheses of the form:

\[
\text{Ho: } \zeta_1 \text{'s are all equal.}
\]
\[
\text{Ha: } \zeta_1 < \zeta_2 < \ldots < \zeta_k.
\]

It is assumed that there are no ties. We have \(J = \sum W_{ij}\), where \(W_{ij}\) is the Mann-Whitney statistic.

Then \(E(J) = \frac{k}{4} \sum_{1 \leq i < j \leq 1} n_i n_j = \frac{N^2}{4} - \frac{\sum_{i=1}^{k} n_i^2}{4}\)

\(\text{Var}(J) = \frac{N^2(2N+3)}{2} - \frac{\sum_{i=1}^{k} n_i^2}{4}(2n_1 + 3)\).
For large $N$ we get an approximate normal distribution. The relationship of $J$ with Finch's $\gamma$ for $(\Delta_\xi, \Omega, \gamma)$ is given by:

$$\gamma(\delta_0, \omega) = \frac{1}{2}N(N-1) + \sum_{i<j} n_i n_j - 2J$$

The Jonckheere statistic is then equivalent to the Finch statistic, $\gamma(\delta_0, \omega)$ and the variance of the Jonckheere statistics is then one-fourth the variance of the Finch statistic, as $\omega$ varies over $\Omega$, the totality of all strong orderings.

It is easy to show that the above expressions given for $E(J)$ and $\text{Var}(J)$ are the same as the expressions for $E[\gamma(\delta_0, \omega)]$ and $\text{Var}[\gamma(\delta_0, \omega)]$.

Since we have:

$$E[\gamma(\delta_0, \omega)] = \frac{1}{2}N(N-1) + \sum_{i<j} n_i n_j - 2E(J)$$

$$\text{Var}[\gamma(\delta_0, \omega)] = 4 \text{Var}(J) = \frac{4}{32} \left[ 2N^3 + 3N^2 - 2 \sum_{i} n_i^3 - 3 \sum_{i} n_i^2 \right]$$

$$= \frac{1}{3} \left( \frac{1}{3} N N^3 + \frac{1}{2} N^2 - \frac{1}{3} \sum_{i} n_i^3 - \frac{1}{2} \sum_{i} n_i^2 \right)$$

$$= \frac{1}{3} \left[ \frac{1}{3} N \sum_{i} n_i^2 + \sum_{i<j} n_i n_j + \frac{1}{2} \sum_{i} n_i^2 + \sum_{i<j} n_i n_j - \frac{1}{3} \sum_{i} n_i^3 - \frac{1}{2} \sum_{i} n_i^2 \right]$$

$$= \frac{1}{3} \left[ \frac{1}{3} N \sum_{i} n_i^2 + \frac{2}{3} \sum_{i<j} n_i n_j + \sum_{i<j} n_i n_j - \frac{1}{3} \sum_{i} n_i^3 \right]$$
The expressions obtained here for $E[\gamma(\delta_0,')]$ and $\text{Var}[\gamma(\delta_0,')]$ are the same as those derived earlier.

2.6. Use of Finch's Ideas

The ideas of Finch can be applied to the general linear model setting. Consider the approximative univariate model:

$$y_i \equiv g_1(\beta), \quad \text{where } \beta' = [\beta_1, \ldots, \beta_p]$$

$g_1(\beta)$ is the hypothesized true value of $y_i$, or the form chosen to approximate $y_i$. Then the result of our model fitting is that we obtain $\{\hat{g}_1(\beta)\}$ where $\hat{g}_1(\beta)$ is the fitted value for $g_1(\beta)$. The result of such fitting is that we have a set of fitted values, $\{\hat{g}_1\}$. We can then ask how good a description of $\{y_i\}$ is $\{\hat{g}_1\}$. This question can be addressed by Finch's ideas, and in particular, we can examine how well the fitting recaptures the ordering in the data vector $y$. If a fitting process recaptures well the ordering of the data vector $y$, then we can take the view that the model and the fitting process are acceptable. So, the ideas can then be used to evaluate from a data-analysis viewpoint how well a particular model and mode of fitting represents the data vector. It can also be used to compare modes of fitting.

We shall illustrate these ideas by looking at simple linear model
situations as represented, generically, by the model: \( y = X \beta \). More particularly, we shall consider two rather simple classificatory data structures and associated linear models. There are three rather natural modes of fitting a linear model:

- \( \ell_1^{(n)} \): minimizing the sum of absolute deviations,
- \( \ell_2^{(n)} \): minimizing the sum of squares of the deviations,
- \( \ell_\infty^{(n)} \): minimizing the maximum of the absolute deviations.

We shall consider only \( \ell_1^{(n)} \) and \( \ell_2^{(n)} \) fittings.

Our first example is a set of data from Snedecor and Cochran (1967, page 263) for a one-way classification, with small alterations to avoid ties. We regret having to make such alterations. A deeper analysis in data with ties could be considered in a fairly simple way. With a total of, say, \( N \) observations, and with ties, our data is an ordered \( k \)-tomy with \( k \) equal to the number of distinct \( y \)-values. We would then consider how well the observed \( k \)-tomy is represented by a \( k' \)-tomy given by the fitting process. The problem of ties in the data set can be handled in this way, with minor complications that we shall not discuss.

Consider, then, the one-way classification, with data possessing no ties, as follows:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>41.8</td>
<td>33.0</td>
<td>38.5</td>
<td>43.7</td>
<td>34.2</td>
<td>32.6</td>
<td>36.2</td>
</tr>
<tr>
<td>2</td>
<td>38.9</td>
<td>37.5</td>
<td>35.9</td>
<td>38.8</td>
<td>38.6</td>
<td>38.4</td>
<td>33.4</td>
</tr>
<tr>
<td>3</td>
<td>36.1</td>
<td>33.1</td>
<td>33.9</td>
<td>36.3</td>
<td>40.2</td>
<td>34.8</td>
<td>37.9</td>
</tr>
</tbody>
</table>
These data consisting of 21 data points, can be fitted by $\ell_1^{(n)}$ or $\ell_2^{(n)}$. The latter is, of course, least squares fitting, which is trivially easy. We have applied these 2 modes of fitting to the data, using the IMSL Routine RLLAV for $\ell_1$ fitting, and the GLM Procedure from SAS '79 for the $\ell_2$ fitting. With $\ell_1$ fitting, the fitted values are the medians as shown below:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fitted value</td>
<td>38.9</td>
<td>33.1</td>
<td>35.9</td>
<td>38.8</td>
<td>38.6</td>
<td>34.8</td>
<td>36.2</td>
</tr>
</tbody>
</table>

Then, as a result of our fitting, we obtain the ordered 7-tomy Treatment 2 < Treatment 6 < Treatment 3 < Treatment 7 < Treatment 5 < Treatment 4 < Treatment 1, with each set having cardinality

3 i.e., $n_i = 3$, $i=1,2,3, ..., 7$.

The resulting gauge measure of this fitting is obtained as follows:

$$\gamma(\delta_o, \omega_o) = \sum_{x,y}^\delta \zeta_{\delta_o}(x,y) + \sum_{x,y}^\omega \zeta_{\omega_o}(x,y) - 2 \sum_{x,y}^\delta \zeta_{\delta_o}(x,y) \cdot \zeta_{\omega_o}(x,y).$$

$$\sum_{x,y}^\delta \zeta_{\delta_o}(x,y) = n_1(n_2 + n_3 + ... + n_7) + n_2(n_3 + ... + n_7) + ... + n_6n_7$$

$$= 189.$$

$$\sum_{x,y}^\omega \zeta_{\omega_o}(x,y) = \frac{1}{2}N(N-1) = 210.$$

$$\sum_{x,y}^\delta \zeta_{\delta_o}(x,y) \cdot \zeta_{\omega_o}(x,y) = 142.$$

Hence, $\gamma(\delta_o, \omega_o) = 189 + 210 - (2 \times 142) = 115.$

If we use $\ell_2$ fitting, we find $\gamma(\delta_o, \omega_o) = 111.$
We conclude that \( L_2 \) fitting gives a slightly better representation of the original strong ordering than does \( L_1 \) fitting.

Just how much better the one fitting is than the other can be assessed partially in the following way. Suppose, we consider a set of all possible 7-tomies, including, of course, that 7-tomy in which all the observations are described by a null polytomy, with just one class, corresponding to the data description that there are no group differences. In the notation of the earlier sections, that set would be \( A^*_7 \). The outcome, in this case, is as follows: For \( L_1 \) fitting

\[
\gamma(\cdot, a_0) \sim N(210, 1048.57), \text{ and the descriptive power is given by}
\]

\[
P[Z > \frac{115 - 210}{32} \cdot \frac{86}{38}] = 0.9983
\]

Also, \( \gamma(\delta_0, \cdot) \sim N(210, 1071) \), and the characterizing power is found to be 0.9981. For \( L_2 \) fitting, the descriptive power is 0.9989 and the characterizing power is 0.9987.

We now illustrate the ideas with a small 2-way data structure:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>112</td>
<td>90</td>
<td>123</td>
</tr>
<tr>
<td>Block</td>
<td>2</td>
<td>86</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>80</td>
<td>62</td>
</tr>
</tbody>
</table>

These data are from Snedecor and Cochran (1967, page 301).

We can consider representing these data by the model:

\[
y_{ij} \sim \mu + r_i + c_j, \text{ with } i \text{ indexing rows and } j \text{ indexing columns. Again, we can consider } L_1 \text{ and } L_2 \text{ fitting. The outcome is sur-}
\]
prising, perhaps, in that both modes of fitting recapture the original strong ordering in the data.

This will not happen, of course, in general. If our data set shows considerable interaction of row and column factors, the outcomes will be different. We shall then find, perhaps, that \( l_1 \) fitting recovers the ordering in the original data much better than \( l_2 \) fitting, or, of course, vice versa. We would have liked to consider other data sets to see what happens, but were prevented by time and computing considerations.

These deliberations are also related to ideas of transformation of data in the following way: Suppose we decide to use \( l_2 \) fitting. Then, if we use a monotonic transformation, as we obviously should, then we can consider the question: What transformation or class of transformations is best with respect to the fitting of an additive model? The monotonic transformation will not, of course, alter the original ordering of the data. This, we surmise, is strongly relevant to comparative experiments, with respect to analysis and interpretation. Often, we are concerned with obtaining an ordering of treatment effects, so we are concerned with the accuracy of description of the data by an ordered polytomy, in which certain treatment differences are null. Consider an experiment on 5 treatments, with \( l_2 \) fitting, and suppose that the ordering of treatment effects, indicated by the fitting, is \( t_2 < t_3 < t_1 < t_4 < t_5 \), but that \( t_3 \) and \( t_1 \) appear to be close. Then, we can consider the representation of the data by the linear model:

\[
y_{ij} = \mu + b_i + t_j \quad \text{with } t_1 = t_3.
\]

We can certainly fit this model, and we shall then obtain an ordering of fitted values. We can then examine the extent to which the ordering of
fitted values, with this approximative model, recovers or recaptures the ordering of the observed values. It is hoped to explore ideas of this type in the future.

The ideas have a considerably broader thrust than these small and easy examples suggest. Consider a data set \( \{(t, y_t) : t=1,2, \ldots, T\} \). Then, we may obtain a smooth "curve" \( \hat{y}_t \) by which to represent the data. We can then ask how well the resulting fit represents the actual ordering in the data set. Consider two modes of obtaining the set \( \{\hat{y}_t\} \), so that we have as competitors \( \{\hat{y}_t\} \) and \( \{\tilde{y}_t\} \). Then, we can consider the distance or gauge of the two descriptions, \( \{\hat{y}_t\} \) and \( \{\tilde{y}_t\} \) of the original data set \( \{y_t\} \). We may suggest that ideas of this type should have a very significant role in what is called "exploratory data analysis."

We see what are, numerically, very complicated procedures of fitting and smoothing of data sets, without any consideration of the utility of the procedures and examination of their descriptive and explanatory powers. It may be suggested that such considerations are important.

The critical point with respect to deliberations of the type mentioned, is that they are not at all based on stochastic models. They are oriented solely to data analysis by approximating models, fitted according to some optimum fit principle.

Finally, it is important to note that Finch has given developments of his ideas in Finch (1980).
3. MEASURE OF EFFECTIVE VARIATION IN AN ARRAY

3.1. Background - Work Done by Wilk and Kempthorne, and Zyskind

The important concept of randomization was introduced by R. A. Fisher, and it is used widely in comparative experiments. If randomization is used in an experiment, then the assumption of normality need not be made. In many situations, the general linear hypothesis theory leads to conclusions similar to those obtained when randomization tests are used. When normal theory is used for tests in a randomized experiment, the validity of these tests of significance comes from randomization theory.

When randomization is used in experiments, instead of making the normality assumption, one can obtain derived linear models. In such models, the random variables are the design random variables, and their properties can be derived under randomization.

M. B. Wilk (1955a), and M. B. Wilk and O. Kempthorne (1957), considered derived linear models for the completely randomized design, the randomized block design, the Latin square design and others. They gave the expected mean squares in terms of the variance components, as well as, in terms of certain linear combinations of the variance components. The linear combinations of the variance components were designated by \( \Sigma \) with suitable suffixes.

For the completely randomized design in the case of 2 factors with equal numbers and non-additivity of units and treatments, the following is the population model:

\[
x_{ijm} = \mu + a_i + b_j + (ab)_{ij} + p_m + q_{ijm} + \varepsilon_{ijm},
\]

where \( \mu \) = the overall mean, \( a_i \) = the main effect of the \( i \)-th level of factor A,
bj = the main effect of the j-th level of factor B,

(ab)_{ij} = the interaction effect of the i-th level of A and j-th level of B,

p_m = unit error,

q_{ijm} = interaction of treatment(ij) with unit m,

\epsilon_{ijm} = technical error.

The factor A has A levels of which a are randomly selected. The factor B has B levels of which b are selected at random. Each treatment is given to r units.

The statistical model is

\[ Y_{i^*j^*f} = \mu + \sum_i \alpha_i^{i^*} + \sum_j \beta_j^{j^*} b_j + \sum_i \sum_j \alpha_i^{i^*} \beta_j^{j^*} (ab)_{ij} + \sum_m \rho_m^{i^*j^*f} p_m + \sum_{i,j,k} \rho_{ijk}^{i^*j^*f} (q_{ijm} + \epsilon_{ijm}) , \]

where

\[ \alpha_i^{i^*} = 1 \text{ if } i^* \text{ corresponds to } i, \]
\[ = 0 \text{ otherwise.} \]

\[ \beta_j^{j^*} = 1 \text{ if } j^* \text{ corresponds to } j, \]
\[ = 0 \text{ otherwise.} \]

\[ \rho_m^{i^*j^*f} = 1 \text{ if } (i^*j^*f) \text{ corresponds to } m, \]
\[ = 0 \text{ otherwise.} \]

The following definitions are used:

\[ \sigma_a^2 = \frac{1}{A-1} \sum_i a_i^{2}, \quad \sigma_b^2 = \frac{1}{B-1} \sum_j b_j^{2}, \quad \sigma_p^2 = \frac{1}{p-1} \sum_m p_m^{2}, \quad \sigma^2 = E(\epsilon_{ijm}^2), \]

\[ \sigma_{ab}^2 = \frac{1}{(A-1)(B-1)} \sum_{i,j} (ab)_{ij}^{2}, \quad \sigma_q^2 = \frac{1}{AB(P-1)} \sum_m q_{ijm}^{2}, \]

\[ Q_{ap}^2 = \frac{1}{(A-1)(P-1)} \sum_m q_{i.m}^{2}, \quad Q_{bp}^2 = \frac{1}{(B-1)(P-1)} \sum_m q_{.jm}^{2}, \]
\[ Q_{abp}^2 = \frac{1}{(A-1)(B-1)(P-1)} \Sigma_{ijm} (q_{ijm} - q_{ij.m} - q_{i.jm} + q_{i.j.m})^2 \]

\[ \Sigma_a = \sigma_a^2 - \frac{1}{A} \sigma_{ab}^2 - \frac{1}{B} Q_{ap}^2 + \frac{1}{AB} Q_{abp}^2 \]

\[ \Sigma_b = \sigma_b^2 - \frac{1}{B} \sigma_{ab}^2 - \frac{1}{A} Q_{bp}^2 + \frac{1}{AB} Q_{abp}^2 \]

\[ \Sigma_{ab} = \sigma_{ab}^2 - \frac{1}{P} Q_{abp}^2 \]

\[ \Sigma_o = \sigma_o^2 + \sigma_p^2 + \sigma_q^2 \]

The following expected mean SS were obtained:

**Table 3.1. Expected mean sum of squares**

<table>
<thead>
<tr>
<th>MSS</th>
<th>E(MSS) in terms of variance components</th>
</tr>
</thead>
<tbody>
<tr>
<td>A*</td>
<td>( \sigma_a^2 + \sigma_b^2 + \sigma_{ab}^2 + \frac{r(b-b)}{B} (\sigma_{ab}^2 - \frac{1}{P} Q_{abp}^2) ) + ( \frac{ra(\sigma_{ab}^2 - \frac{1}{P} Q_{abp}^2)}{A} )</td>
</tr>
<tr>
<td>B*</td>
<td>( \sigma_p^2 + \sigma_q^2 + \sigma_{pq}^2 + \frac{r(A-a)}{B} (\sigma_{ab}^2 - \frac{1}{P} Q_{abp}^2) ) + ( \frac{ra(\sigma_{ab}^2 - \frac{1}{P} Q_{abp}^2)}{A} )</td>
</tr>
<tr>
<td>( r_{AB} )</td>
<td>( \sigma_p^2 + \sigma_q^2 + \sigma_{pq}^2 + \frac{r(\sigma_{ab}^2 - \frac{1}{P} Q_{abp}^2)}{A} )</td>
</tr>
<tr>
<td>R*</td>
<td>( \sigma_p^2 + \sigma_q^2 )</td>
</tr>
</tbody>
</table>

For the randomized block design with non-additivity of treatments with blocks or with units within blocks, the population model is:

\[ x_{ijk} = \mu + b_i + t_k + (bt)_{ik} + p_{ij} + n_{ijk} + \varepsilon_{ijk} \]

where \( \mu = \) overall mean,

\( b_i = \) effect of \( i\)-th block,
\( t_k \) = effect of \( k \)-th treatment,

\((bt)_{ik}\) = interaction of block \( i \) and treatment \( k \),

\( p_{ij} \) = within block additive unit error of the \( j \)-th experimental unit of block \( i \),

\( n_{ijk} \) = within block unit-treatment interaction,

\( \epsilon_{ijk} \) = technical error.

From a totality of \( T \) treatments, \( t \) are selected at random. From a total of \( B \) blocks, \( b \) are selected at random. From each selected block, \( p = rt \) experimental units are chosen at random, and treatments are applied within each block randomly, but so that each treatment appears \( r \) times in each block.

The statistical model here, using natural indicator variables, is:

\[
y_{ijk} = \mu + \sum_i^{B-1} a_i \beta_i^k \tau_k + \sum_k^{T-1} \alpha_i \beta_i^k (bt)_{ik} + \sum_j^{P-1} \rho_i \xi_i^j p_{ij} + \sum_{ijk}^{B(P-1)} \eta_{ijk} (n_{ijk} + \epsilon_{ijk}).
\]

The following definitions are used:

\[
\sigma_b^2 = \frac{1}{B-1} \Sigma_i b_i^2, \quad \sigma_t^2 = \frac{1}{T-1} \Sigma_k t_k^2, \quad \sigma_{bt}^2 = \frac{1}{(B-1)(T-1)} \Sigma_{ik} (bt)_{ik}^2,
\]

\[
\sigma_p^2 = \frac{1}{B(P-1)} \Sigma_{ij} p_{ij}^2, \quad \sigma_n^2 = \frac{1}{BT(P-1)} \Sigma_{ijk} n_{ijk}^2,
\]

\[
Q_{tp}^2 = \frac{1}{B(T-1)(P-1)} \Sigma_{ijk} n_{ijk}^2 = \frac{T}{T-1} \sigma_n^2, \quad \sigma^2 = E(\epsilon_{ijk}^2),
\]

\[
\Sigma_o = \sigma^2 + Q_{tp}^2 + \sigma_p^2 - \frac{1}{T} Q_{tp}^2, \quad \Sigma_{bt} = \sigma_{bt}^2 - \frac{1}{P} Q_{tp}^2,
\]

\[
\Sigma_t = \sigma_t^2 - \frac{1}{B} \sigma_{bt}^2, \quad \Sigma_b = \sigma_b^2 - \frac{1}{T} \sigma_{bt}^2 - \frac{1}{P} \sigma_p^2 + \frac{1}{TP} Q_{tp}^2.
\]

Table 3.2 below gives the expected mean SS:
Table 3.2. Expected mean sum of squares

<table>
<thead>
<tr>
<th>MSS</th>
<th>( E(MSS) ) in terms of variance components</th>
<th>( E(MSS) ) in terms of ( \Sigma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( B^* )</td>
<td>( rt[s_b^2 - \frac{1}{T} s_{bt}^2 + \frac{1}{P} s_{bt}^2 - \frac{1}{PT} s_{tp}^2] \Sigma_o + rE_{bt} + rE_b )</td>
<td>( E(MSS) ) in terms of ( \Sigma )</td>
</tr>
<tr>
<td>( I^* )</td>
<td>( rt[s_b^2 - \frac{1}{T} s_{bt}^2 + \frac{1}{P} s_{bt}^2 - \frac{1}{PT} s_{tp}^2] \Sigma_o + rE_{bt} + rE_t )</td>
<td>( E(MSS) ) in terms of ( \Sigma )</td>
</tr>
<tr>
<td>( I^* )</td>
<td>( rt[s_b^2 - \frac{1}{T} s_{bt}^2 + \frac{1}{P} s_{bt}^2 - \frac{1}{PT} s_{tp}^2] \Sigma_o + rE_{bt} )</td>
<td>( E(MSS) ) in terms of ( \Sigma )</td>
</tr>
<tr>
<td>( R^* )</td>
<td>( [s_p^2 - \frac{1}{T} s_{tp}^2] + q_{tp}^2 + \sigma^2 )</td>
<td>( E(MSS) ) in terms of ( \Sigma )</td>
</tr>
</tbody>
</table>

For the Latin square design with non-additivity of treatments with rows and columns, the population model is:

\[ x_{ijk} = \mu + r_i + c_j + t_k + (rc)_{ij} + (rt)_{ik} + (ct)_{jk} + (rct)_{ijk} + \varepsilon_{ijk}, \]

where \( \mu \) = overall mean, \( r_i \), \( c_j \), \( t_k \) are the effects of row \( i \), column \( j \),
treatment \( k \), respectively,

\( (rc)_{ij} \) = interaction of row \( i \) and column \( j \),

\( (rt)_{ik} \) = interaction of row \( i \) and treatment \( k \),

\( (ct)_{jk} \) = interaction of column \( j \) and treatment \( k \),

\( (rct)_{ijk} \) = interaction of row \( i \), column \( j \) and treatment \( k \),

\( \varepsilon_{ijk} \) = technical error.
From a total of $R$ rows, $t$ rows are selected at random. From a total of $C$ columns, $t$ columns are selected randomly. Then $t$ treatments are selected randomly from a total of $T$ treatments and are applied randomly to the $t^2$ units in a Latin square arrangement.

The statistical models are given by

$$Z_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

where $\alpha_i$, $\beta_j$, and $\epsilon_{ij}$ are random effects. The following definitions are used:

- $\delta_{ijk} = (R-1)(C-1)(T-1)$
- $\delta_{ij} = (R-1)(T-1)$
- $\delta_{i} = (C-1)(T-1)$
- $\delta_{t} = (R-1)(C-1)$
- $\delta_{rc} = (R-1)(C-1)$
- $\delta_{rt} = (R-1)(T-1)$
- $\delta_{ct} = (C-1)(T-1)$

The Table below gives the expected mean sum of squares:
Table 3.3. Expected mean sum of squares

<table>
<thead>
<tr>
<th>MSS</th>
<th>E(MSS) in terms of variance components</th>
<th>E(MSS) in terms of Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>R*</td>
<td>$\sigma^2 + \left(\frac{t}{C^T}\right)\sigma^2_{rc} + \left(\frac{t-C}{t}\right)\sigma^2_{ct} + \sigma^2_{t} + \sigma^2_{r}$</td>
<td>$\Sigma_o + t\Sigma_r$</td>
</tr>
<tr>
<td></td>
<td>$\frac{t-t}{T}\sigma^2_{rt} + t\sigma^2_{r}$</td>
<td></td>
</tr>
<tr>
<td>C*</td>
<td>$\sigma^2 + \left(\frac{t}{R^T}\right)\sigma^2_{rc} + \left(\frac{R-t}{R}\right)\sigma^2_{rc} + \sigma^2_{r}$</td>
<td>$\Sigma_o + t\Sigma_c$</td>
</tr>
<tr>
<td></td>
<td>$+(\frac{t-t}{T})\sigma^2_{ct} + \sigma^2_{rt} + t\sigma^2_{c}$</td>
<td></td>
</tr>
<tr>
<td>T*</td>
<td>$\sigma^2 + \left(\frac{t}{R^C}\right)\sigma^2_{rc} + \sigma^2_{rt} + \Sigma_o + t\Sigma_t$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$+(\frac{C-t}{C})\sigma^2_{ct} + \left(\frac{R-t}{R}\right)\sigma^2_{rt} + t\sigma^2_{t}$</td>
<td></td>
</tr>
<tr>
<td>D*</td>
<td>$\sigma^2 + \sigma^2_{rc} + \sigma^2_{ct} + \sigma^2_{rt} + \Sigma_o$</td>
<td></td>
</tr>
</tbody>
</table>

Zyskind (1962) has done considerable work on the Σ quantities. He discusses the ideas of relation and structure in experimental design and gives a general definition of the Σ quantities. The paper concerns mainly the simple and general Σ forms of the expected values of the squares of means of the observations. From this, it is easy to write down the expected mean squares in the analysis of variance in terms of the Σ, because in "balanced" cases, these expectations can be expressed uniquely as a linear function of expected values of squares of the typical observational means.

Zyskind starts by showing how a typical response can be expressed identically, as a sum of all its corresponding components, where the word components are used for linear combinations of the partial means. An admissible mean is one in which whenever a nested index appears, then
all the indices which nest it also appear. The indices of an admissible partial mean, which nest no other indices of that mean, are said to constitute the set of indices belonging to the rightmost bracket. He calls as the component of variation, of a given type of component, the quantity which is the sum of squares of values of components of the given type, over all the population ranges of the indices used to denote the component divided by the number of degrees of freedom for the type. These components of variation are denoted by $\sigma^2$ with suitable subscripts, bracketed into groups, corresponding to the subscripts of the types of components to which the $\sigma^2$'s refer. Then, the $\Sigma$ are defined as follows:

Definition: Consider a particular type of component and all $\sigma^2$'s of the following form (i) the set of subscripts of $\sigma^2$ includes the set of subscripts corresponding to the leading term of the component as a subset. (ii) the excess subscripts belong exclusively to the rightmost bracket of $\sigma^2$.

The linear combination of all such $\sigma^2$'s, where the coefficient of a particular $\sigma^2$ with $k$ excess subscripts, is

$$(-1)^k \frac{1}{\text{product of population ranges of the excess indices}}$$

is defined as the $\Sigma$ corresponding to the type of component under consideration. The subscript notation for the $\Sigma$ is to be same as for the type of component.

Zyskind, then used the derived linear model ideas of Wilk and Kempthorne to obtain the expectations of the mean sums of squares in analyses of variance for balanced samples from the structured population, and gives them in terms of the $\Sigma$'s.
3.2. Review of Cox's Paper

D. R. Cox (1958) also studied the $\Sigma$-quantities. His approach is different from that of Wilk and Kempthorne. Cox calls the $\Sigma$-quantities components of effective variation. For example, $\Sigma_c$ is the component of effective variation between treatments. It measures how much more variation there is between the treatment means than would be expected under random permutation. Cox obtains the expected values of the sums of squares in the analysis of variance, by considering random permutation of the levels of some factors within each combination of the levels of the remaining factors.

The one-way set-up is considered first. Suppose there are $K$ finite levels of a factor with $N$ units in each level. Let $x_{ij}$ denote the $j$-th unit in the $i$-th level, with $i = 1, 2, \ldots, K$, $j = 1, 2, \ldots, N$. Then, the components of variation between and within the levels are defined as

$$
\sigma_b^2 = \frac{1}{K-1} \sum_i (x_{i\cdot} - x_{\cdot\cdot})^2, \quad \sigma_\omega^2 = \frac{1}{K(N-1)} \sum_{ij} (x_{ij} - x_{i\cdot})^2,
$$

Replacing a suffix by a dot indicates that averaging has been done over that suffix, e.g., $x_{i\cdot} = \frac{1}{N} \sum_j x_{ij}$, and $x_{\cdot\cdot} = \frac{1}{KN} \sum_{ij} x_{ij}$.

Cox wants a measure of the variation between levels that reduces to $\sigma_b^2$ when there is no variation within each level, and has zero expectation when the whole set $x_{ij}$ is permuted at random into $K$ sets of $N$. Thus, this measure will describe how much more variation there is between the level means, than would be expected under random permutation.

The measure he obtains is $\Sigma_b = \sigma_b^2 - \frac{1}{N} \sigma_\omega^2$. Also, he writes $\Sigma_\omega = \sigma_\omega^2$. 
Next, Cox considered the two-way set-up. Suppose there are 2 factors, say, rows and columns. Let there be R rows and C columns, and let the value in the i-th row and j-th column be \( x_{ij} \). The following definitions are used:

\[
\sigma^2_{RC} = \frac{1}{(R-1)(C-1)} \sum_{ij} (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2,
\]

\[
\sigma^2_R = \frac{1}{R-1} \sum_i (\bar{x}_i - \bar{x})^2, \quad \sigma^2_C = \frac{1}{C-1} \sum_j (\bar{x}_j - \bar{x})^2.
\]

Again, dot indicates an average.

To find a measure of effective variation between rows, Cox wants a quantity that is not affected by arbitrary changes in columns, whose expectation is zero when the rows are permuted randomly within each column. The measure derived by Cox is:

\[
\Sigma_R = \sigma^2_R - \frac{1}{C} \sigma^2_{RC}.
\]

A similar measure of effective variation between columns can be obtained. It will be \( \Sigma_C = \sigma^2_C - \frac{1}{R} \sigma^2_{RC} \). Also, he writes \( \Sigma_{RC} = \sigma^2_{RC} \).

Lastly, Cox considered the Latin square design. The units are set out in a RxC array. Let \( x_{ijk} \) denote the response due to the k-th treatment being applied to the (ij)th unit. There are R rows, C columns and n treatments. The components of variation are defined as before, e.g.,

\[
\sigma^2_R = \frac{1}{(R-1)} \sum_i (\bar{x}_i - \bar{x})^2, \quad \sigma^2_{Rt} = \frac{1}{(R-1)(n-1)} \sum_{ik} (x_{ik} - \bar{x}_i - \bar{x}_k + \bar{x})^2, \quad \text{etc.}
\]

A measure of effective variation for rows x treatments must be unaffected by arbitrary changes in rows, treatments, columns, rows x columns, treatment x columns, i.e., in the effects or interactions that do not involve rows and treatments jointly. Also, the expectation under random permutation of the rows and treatments within each column must be zero. The measure found is \( \Sigma_{Rt} = \sigma^2_{Rt} - \frac{1}{C} \sigma^2_{RCt} \). Similarly, measures for
the other first-order interactions can be found.

Next, measures for the main effects are considered, e.g., main effect of treatments. This measure must not be affected by arbitrary changes in rows, columns, rows x columns. Also, expectation under random permutation of the treatments within each combination of the rows and columns must be zero. Two more conditions must be satisfied:

1) If the column classification is unnecessary, the 3-way set-up reduces to the 2-way set-up. Then, $x_{ijk} = x_{ik}$, and the measure should reduce to the measure in the 2-way set-up.

2) If the row classification is unnecessary, then as in the above condition, $x_{ijk} = x_{jk}$, and the 3-way set-up should reduce to the 2-way set-up, and the measure reduce to the measure in the 2-way set-up.

Using these conditions, Cox obtained

$$
\Sigma_t = \sigma_t^2 - \frac{1}{R} \sigma_{Rt}^2 - \frac{1}{C} \sigma_{Ct}^2 + \frac{1}{RC} \sigma_{RCt}^2.
$$

Similarly, components of effective variation can be obtained for $\Sigma_R$ and $\Sigma_C$.

3.3. Derivation of the Components of Effective Variation for the General Case of k Factors

We consider, now, a k-way set-up. Let there be k factors $F_1, F_2, \ldots, F_k$ with $L(1), L(2), \ldots, L(k)$ levels, respectively. Suppose we want a measure of effective variation for $F_{i_1} x F_{i_2} x \ldots x F_{i_s}$, $i_1 \neq i_2 \neq \ldots \neq i_s = 1, 2, \ldots, k$, with $s < k$. For convenience, we can take $F_{i_1} x F_{i_2} x \ldots x F_{i_s} = F_1 x F_2 x \ldots x F_s$. There is no loss of generality if we do this. Denote the measure we require by $Q$. Now, $Q$ must be such that it is unaffected by arbitrary changes in the effects that do not involve $F_1 x F_2 x \ldots x F_s$. 
With a subset $\alpha$, say, of the factors, and with the complement of $\alpha$ denoted by $\beta$, it must be the case that $Q$ is unaffected by arbitrary changes in effects and interactions associated with effects or interactions other than the ones involving all the factors of $\alpha$ with subsets of the factors of $\beta$. In other words, the $Q$ associated with $F_1x F_2x ... x F_s$ must be unaffected by arbitrary changes in effects and interactions of the whole system that do not involve all these $s$ factors.

The measure $Q$, is a measure of variation, and hence must be a symmetric, quadratic function of $\{X\}$, where $\{X\}$ is the conceptual population of observations. So, $Q$ is of the form:

$$Q = \alpha_0 C + \alpha_1 C_1 + \alpha_2 C_2 + \alpha_{12} C_{12} + \alpha_3 C_3 + ... + \alpha_{12}...k C_{12}...k'$$

where $C = [L(1)\cdot L(2)\cdots L(k)]^{-1}X^2\cdots = S$,

$$C_1 = [L(2)\cdot L(3)\cdots L(k)]^{-1} \sum_{i_1=1}^{L(1)} X_{i_1}^2 - C = S_1 - S,$$

$$C_2 = [L(1)\cdot L(3)\cdots L(k)]^{-1} \sum_{i_2=1}^{L(2)} X_{i_2}^2 - C = S_2 - S,$$

$$C_{12} = [L(3)\cdot L(4)\cdots L(k)]^{-1} \sum_{i_1=1}^{L(1)} \sum_{i_2=1}^{L(2)} X_{i_1}^2 - [L(1)\cdot L(3)\cdots L(k)]^{-1} \sum_{i_2=1}^{L(2)} X_{i_2}^2 \cdots + C$$

$$= S_{12} - S_1 - S_2 + S,$$

and so on.

The symbol $X$ in which some subscripts are replaced by dots represents the summation of elements of the original array over subscripts that have been replaced by dots.

Now consider an arbitrary change in any $F_i$, say $F_1$. Write the new observation as $y(i_1,i_2,...,i_k) = x(i_1,i_2,...,i_k) + z(i_1)$. 
A change in $F_1$ will affect only $C_1$, and none of the other sums of squares. Denote by $\tilde{C}_1(y), C_1(x), C_1(z)$ the sums of squares we get with $y, x, z$, respectively. If $Q$ is to be unchanged by arbitrary changes in $F$, then we must have $\tilde{C}_1(y) = C_1(x)$.

Now, $\tilde{C}_1(y) = \left[\sum_{i=2}^{k} L_i \right]^{-1} \sum_{i=1}^{k} x_i^2 - \left[\sum_{i=1}^{k} L_i \right]^{-1} y^2.$

$$= C_1(x) + 2\text{[Cross product terms]} + C_1(z). \quad (3.1)$$

The new observation produced by an arbitrary change in $F_1$, may also be expressed as $y'(i_1, i_2, \ldots, i_k) = x(i_1, i_2, \ldots, i_k) - z(i_1)$.

As noted before, a change in $F_1$ will only affect $C_1$ and none of the other sums of squares. Denote by $\tilde{C}_1(y'), C_1(x), C_1(z)$, the sums of squares obtained with $y', x, z$, respectively. Since $Q$ must be unaffected, we must have $\tilde{C}_1(y') = C_1(x)$.

Now $\tilde{C}_1(y') = C_1(x) - 2\text{[Cross product terms]} + C_1(z). \quad (3.2)$

Since it is the case that $\tilde{C}_1(y) = C_1(x) = \tilde{C}_1(y')$, if we multiply equations (3.1) and (3.2) by $\alpha_1$ and add, we must have

$$\alpha_1 [C_1(x) + C_1(z)] = \alpha_1 C_1(x) \quad (3.3)$$

The sum of squares $C_1(z)$ is positive. So, if equation (3.3) is to hold, we must have $\alpha_1 = 0$. Thus, $C_1$ does not enter the expression for $Q$.

Next, in general, consider an arbitrary change in $F_1x F_2x \ldots x F_t$ with $t < s$. This will affect only the sum of squares $C_{12 \ldots t}$ and no other sums of squares. The new observation obtained due to this change can be
denoted by \( y(i_1, i_2, \ldots, i_k) = x(i_1, i_2, \ldots, i_k) + z(i_1, i_2, \ldots, i_k) \),

where \( z(i_1, i_2, \ldots, i_t) \) is purely interactive in that, summation over the levels of any factor yields zero. As before, call as \( C_{12\ldots t}(y) \), \( C_{12\ldots t}(x) \), \( C_{12\ldots t}(z) \) the sums of squares obtained with \( y, x, z \), respectively.

If \( Q \) is to be unaffected, we must have \( C_{12\ldots t}(y) = C_{12\ldots t}(x) \).

Now, \( C_{12\ldots t}(y) = C_{12\ldots t}(x) + 2[\text{Cross product terms}] + C_{12\ldots t}(z) \) \hspace{1cm} (3.4)

Another new set of observations may also be expressed as \( y'(i_1, i_2, \ldots, i_k) = x(i_1, i_2, \ldots, i_t) - z(i_1, i_2, \ldots, i_t) \). Denote by \( C_{12\ldots t}(y') \), \( C_{12\ldots t}(x) \), \( C_{12\ldots t}(z) \), the sums of squares obtained with \( y', x, z \), respectively.

Since \( Q \) must remain unchanged, we must have \( C_{12\ldots t}(y') = C_{12\ldots t}(x) \).

Now \( C_{12\ldots t}(y') = C_{12\ldots t}(x) - 2[\text{Cross product terms}] + C_{12\ldots t}(z) \) \hspace{1cm} (3.5)

Since we must have \( C_{12\ldots t}(y') = C_{12\ldots t}(x) = C_{12\ldots t}(y) \), if we multiply equations (3.4) and (3.5) by \( \alpha_{12\ldots t} \) and add, we get

\[ \alpha_{12\ldots t}[C_{12\ldots t}(x) + C_{12\ldots t}(z)] = \alpha_{12\ldots t}C_{12\ldots t}(x) \] \hspace{1cm} (3.6)

The sum of squares \( C_{12\ldots t}(z) \) is non-zero. Hence, if equation (3.6) is to hold, we must have \( \alpha_{12\ldots t} = 0 \). Thus, \( C_{12\ldots t} \) does not enter the expression for \( Q \).

Thus, \( Q \) must be of the form:

\[ Q = \alpha_{12\ldots s}C_{12\ldots s} + \sum_{s=t_1}^{k} \alpha_{12\ldots st_1}C_{12\ldots st_1} + \sum_{s=t_1}^{k} \alpha_{12\ldots st_1 t_2}C_{12\ldots st_1 t_2} + \ldots + \alpha_{12\ldots k}C_{12\ldots k} \] \hspace{1cm} (3.7)

The measure \( Q \) must satisfy the following condition: The expectation
of \( Q \) must be zero under random permutation of the levels of \( F_1, F_2, \ldots, F_s \)
within each combination of the levels of \( F_{(s+1)}, F_{(s+2)}, \ldots, F_k \). Let \( \alpha \)
and \( \beta \) be two subsets of the factors whose levels are kept fixed.
Similarly, let \( \gamma_1 \) and \( \gamma_2 \) be two subsets of the factors, whose levels are
permuted. Then, let \( \alpha \cup \beta \cup \gamma_1 \cup \gamma_2 = \{ F_1, F_2, \ldots, F_k \} \), and \( \alpha, \beta, \gamma_1, \gamma_2 \) be
mutually exclusive.

Let \( X = \{ x(\alpha', \beta'; \gamma_1', \gamma_2') : I_{\alpha'}, I_{\beta'}, I_{\gamma_1'}, I_{\gamma_2'} \} \) represent the whole
population. Let \( x(\alpha', \beta'; \gamma_1', \gamma_2') \) be an individual element in the array \( X \).
This element is indexed by \( (I_{\alpha'}, I_{\beta'}, I_{\gamma_1'}, I_{\gamma_2'}) \), where \( I_{\alpha'} \) is some combination of
the levels of the factors in \( \alpha \). Similarly, \( I_{\beta'}, I_{\gamma_1'}, I_{\gamma_2'} \) are some combinations of the levels of the factors in \( \beta, \gamma_1, \gamma_2 \), respectively. Permute the
levels of the factors in \( \gamma_1 \) and \( \gamma_2 \) within each combination of the levels
of the factors in \( \alpha \) and \( \beta \). Then, let \( \tilde{x}(I_{\alpha'}, I_{\beta'}; I_{\gamma_1'}, I_{\gamma_2'}) \) represent a random
permutation of \( x(\alpha', \beta'; \gamma_1', \gamma_2') \).

Now, \( \tilde{x}(I_{\alpha'}, I_{\beta'}; I_{\gamma_1'}, I_{\gamma_2'}) = \sum_{I_{\gamma_1'}} \sum_{I_{\gamma_2'}} \delta(I_{\alpha'}, I_{\beta'}; I_{\gamma_1'}, I_{\gamma_2'}; I_{\gamma_1'}, I_{\gamma_2'}) \cdot x(I_{\alpha'}, I_{\beta'}; I_{\gamma_1'}, I_{\gamma_2'}) \),
where \( \delta(I_{\alpha'}, I_{\beta'}; I_{\gamma_1'}, I_{\gamma_2'}, I_{\gamma_1'}, I_{\gamma_2'}) \) is a random variable such that \( \delta(I_{\alpha'}, I_{\beta'}; I_{\gamma_1'}, I_{\gamma_2'}, I_{\gamma_1'}, I_{\gamma_2'}) = 1 \) when \( I_{\gamma_1} = I_{\gamma_1'} \) and \( I_{\gamma_2} = I_{\gamma_2'} \) for a given combination
of \( I_{\alpha'}, I_{\beta'} \). Under the random permutation, this occurs with probability

\[
\frac{1}{L(\gamma_1) \cdot L(\gamma_2)},
\]
and \( \delta(I_{\alpha'}, I_{\beta'}; I_{\gamma_1'}, I_{\gamma_2'}, I_{\gamma_1'}, I_{\gamma_2'}) = 0 \) otherwise.

Now \( L(\gamma_1) = \text{total number of levels in } \gamma_1 \)

\[ = \text{product of the levels of all factors in } \gamma_1. \]

Similarly, \( L(\gamma_2), L(\alpha), L(\beta) \) can be defined.
Hence, \[ E[X(I_\alpha, I_\beta; I_{Y_1}, I_{Y_2})] \]
\[ = E \left\{ \sum_{I_{Y_1}, I_{Y_2}} \delta(I_\alpha, I_\beta; I_{Y_1}, I_{Y_2}) \cdot x(I_\alpha, I_\beta; I_{Y_1}, I_{Y_2}) \right\} \]
\[ = \frac{1}{L(Y_1) \cdot L(Y_2)} \sum_{I_{Y_1}, I_{Y_2}} x(I_\alpha, I_\beta; I_{Y_1}, I_{Y_2}) \]
\[ = \frac{1}{L(Y_1) \cdot L(Y_2)} \sum_{I_{Y_1}, I_{Y_2}} x(I_\alpha, I_\beta) \]

The symbol + indicates that the index has been summed over.

Similarly, we get
\[ E[X(I^*_{\alpha}, I^*_{\beta}; I^*_{Y_1}, I^*_{Y_2})] \]
\[ = \frac{1}{L(Y_1) \cdot L(Y_2)} \sum_{I_{Y_1}, I_{Y_2}} x(I^*_{\alpha}, I^*_{\beta}; I_{Y_1}, I_{Y_2}) \]

and
\[ E[X(I^*_{\alpha}, I^*_{\beta}; I^*_{Y_1}, I^*_{Y_2}) \cdot X(I^*_{\alpha}, I^*_{\beta}; I^*_{Y_1}, I^*_{Y_2})] \]
\[ = \frac{1}{L(Y_1) \cdot L(Y_2)} \sum_{(I_{Y_1}, I_{Y_2}), \neq} x(I^*_{\alpha}, I^*_{\beta}; I^*_{Y_1}, I^*_{Y_2}) \cdot x(I^*_{\alpha}, I^*_{\beta}; I^*_{Y_1}, I^*_{Y_2}) \]

Now consider the Sum of Squares (SS) for \( \alpha \cup Y_1 \) after permutation. Denote it by \( S_{\alpha Y_1} \). Now, \[ S_{\alpha Y_1} = \frac{1}{L(Y_2) \cdot L(Y_2)} \sum_{I_{Y_1}, I_{Y_2}} [X(I^*_{\alpha}, I^*_{Y_1} +)]^2 \]

Now \[ X(I^*_{\alpha}, I^*_{Y_1} +) = \sum_{I_{Y_2}} X(I^*_{\alpha}, I^*_{\beta}; I^*_{Y_1}, I^*_{Y_2}) \]

\[ [X(I^*_{\alpha}, I^*_{Y_1} +)]^2 = [\sum_{I_{Y_2}} X(I^*_{\alpha}, I^*_{\beta}; I^*_{Y_1}, I^*_{Y_2})]^2 \]
\[ = \sum_{I_{Y_2}} [X(I^*_{\alpha}, I^*_{\beta}; I^*_{Y_1}, I^*_{Y_2})]^2 \]
\[
\sum_{(I_\beta', I_{Y_2'})} x(I_\alpha, I_\beta; I_{Y_1}, I_{Y_2}) \cdot x(I_\alpha, I_\beta; I_{Y_1}', I_{Y_2}')
\]

\[
= \sum_{I_\beta, I_{Y_2}} \left[ x(I_\alpha, I_\beta; I_{Y_1}, I_{Y_2}) \right]^2 + \\
\sum_{I_\beta, I_{Y_2}} \sum_{I_\beta', I_{Y_2'}} x(I_\alpha, I_\beta; I_{Y_1}, I_{Y_2}) \cdot x(I_\alpha, I_\beta; I_{Y_1}', I_{Y_2}') + \\
\sum_{I_\beta, I_{Y_2}} \sum_{all} x(I_\alpha, I_\beta; I_{Y_1}, I_{Y_2}) \cdot x(I_\alpha, I_\beta; I_{Y_1}', I_{Y_2}').
\]

So now, \( E[\left( x(I_\alpha', +; I_{Y_1}, +) \right)^2] \)

\[
= \sum_{I_\beta, I_{Y_2}} \sum_{I_\beta', I_{Y_2'}} \frac{1}{L(Y_1)L(Y_2)} \sum_{all} x(I_\alpha, I_\beta; I_{Y_1}, I_{Y_2}) \cdot x(I_\alpha, I_\beta; I_{Y_1}', I_{Y_2}').
\]

\[
+ \sum_{I_\beta, I_{Y_2}} \sum_{all} \frac{1}{L(Y_1)L(Y_2')} x(I_\alpha, I_\beta; +, +) \cdot \frac{1}{L(Y_1)L(Y_2)} x(I_\alpha, I_\beta; +, +).\]

We get the above expectation for the third term because, if \((I_\alpha, I_\beta) \neq (I_\alpha', I_\beta')\), the variables \(x(I_\alpha, I_\beta; I_{Y_1}, I_{Y_2})\) and \(x(I_\alpha, I_\beta; I_{Y_1}', I_{Y_2}')\) are independent.

So, \( E[\left( x(I_\alpha', +; I_{Y_1}, +) \right)^2] \)
Using this formula, we get

\[
E(\tilde{S}_{12\ldots s}) = \frac{1}{L(1)L(2)\ldots L(s)} \sum_{i_1, i_2, \ldots, i_k} x(i_1, i_2, \ldots, i_k)^2 - \frac{1}{L(1)L(2)\ldots L(s)} \sum_{i_1, i_2} x(i_1, i_2)^2 ,
\]

\[
i.e.,
E(\tilde{S}_{\alpha Y_1}) = \frac{[L(1) - 1]}{L(1)L(2)\ldots L(s)} \sum_{i_1, i_2} x(i_1, i_2)^2 - \frac{[L(1) - 1]}{L(1)L(2)\ldots L(s)} \sum_{i_1, i_2} x(i_1, i_2)^2 .
\]

Now, under random permutation,

\[
\tilde{S}_{12\ldots s} = \tilde{S}_{123\ldots s} - \tilde{S}_{12\ldots(s-1)} + \tilde{S}_{12\ldots(s-2)s} - \ldots + \tilde{S}_{12\ldots(s-2)} + \tilde{S}_{12\ldots(s-3)(s-1)} + \ldots ,
\]

where on the R.H.S we have an \( S \) associated with every subset of \( \{1,2,\ldots,s\} \), with sign equal to \( (-1)^k \), if \( k \) of the members of \( \{1,2,\ldots,s\} \) are omitted.

So, \( E(\tilde{S}_{12\ldots s}) \) and \( E(\tilde{S}_{\alpha Y_1}) \) follow from the above relations.
\[= \left[ \left( \frac{L(s+1)L(s+2)\ldots L(k)}{L(s+1)\ldots L(k)} \right)^{-1} s_{12}\ldots k - [L(s+1)\ldots L(k)]^{-1} s_{(s+1)\ldots k} + S \right] - \left( \frac{L(1)\ldots L(s-2)L(s-1)}{L(s+1)\ldots L(k)} \right) \left( \frac{L(1)\ldots L(s-1)}{L(s+1)\ldots L(k)} \right) S_{12}\ldots k - \left( \frac{L(1)\ldots L(s-1)}{L(s+1)\ldots L(k)} \right) S_{(s+1)\ldots k} + S \right] - \ldots \]

There are equal numbers of positive \( \tilde{S} \)'s and negative \( \tilde{S} \)'s in \( \tilde{c}_{12}\ldots s \).

When we take \( E(\tilde{S}) \) for any \( \tilde{S} \) in \( \tilde{c}_{12}\ldots s \), we find that the term \( S \) occurs in the expectation. Hence, in \( E(\tilde{c}_{12}\ldots s) \), there will be equal numbers of positive and negative values of the term \( S \), which will therefore vanish from \( E(\tilde{c}_{12}\ldots s) \).

So, \( E(\tilde{c}_{12}\ldots s) = (s_{12}\ldots k - s_{(s+1)\ldots k}) \left[ (L(s+1)L(s+2)\ldots L(k))^{-1} \right] - \left( L(1)\ldots L(s-1) L(1)\ldots L(s-1) - \ldots \right) \]

\[ = (s_{12}\ldots k - s_{(s+1)\ldots k}) \left[ (L(s+1)\ldots L(k))^{-1} \left[ (L(s+1)\ldots L(k))^{-1} \right] \right] \]

\[ \{L(1)L(2)\ldots L(s) - \sum_{i=1}^{s} \sum_{i_1 < i_2 < \ldots < i_p} L(i_1)L(i_2)\ldots L(i_p) = \sum_{i=1}^{s} L(i)\ldots L(i) \} \]

However, \( L(1)L(2)\ldots L(s) - \sum_{i=1}^{s} L(i)\ldots L(i) + \sum_{i=1}^{s} L(i)\ldots L(i) \) is the expansion of \( [L(1)-1][L(2)-1]\ldots[L(s)-1] \)

So, we have
Using the same procedure, we get

\[
E(C_{12...s}) = \frac{[L(1)-1][L(2)-1]...[L(s)-1]}{L(s+1)...L(k)L(1)...L(s)-1} (s_{12...k} - S(s+1)...k)
\]

Similarly, we can get the expectations of the other quantities in \(Q\).

So, \(E(Q) = \alpha_{12...s}E(C_{12...s}) + \sum_{t_1=s+1}^{k} \alpha_{12...st_1}E(C_{12...st_1}) + \ldots + \alpha_{12...k}E(C_{12...k})\)

\[
= \frac{[L(1)-1][L(s)-1]}{L(s+1)...L(k)L(1)...L(s)-1} (s_{12...k} - S(s+1)...k)\left[\alpha_{12...s} + \sum_{t_1=1}^{k} \sum_{t_1<t_2} [L(t_1)-1][L(t_2)-1] \alpha_{12...st_1t_2} + \ldots + \right.
\]

\[
\left. [L(s+1)-1][L(s+1)-1] \alpha_{12...s} \right]
\]

We must have \(E(Q) = 0\). This can happen only if \(\{\alpha_{12...s} + \sum_{t_1=1}^{k} [L(t_1)-1] \alpha_{12...st_1} + \ldots + [L(s+1)-1][L(s+1)-1] \alpha_{12...s}\} = 0\).

Now, suppose that the classifications by factors \(F(s+1), F(s+2), \ldots, F_k\) are unnecessary. Then \(x(i_1, i_2, \ldots, i_s, \ldots, i_k) = u(i_1, i_2, \ldots, i_k)\).

Let \(C^*\) denote the sum of squares with respect to \(u\). We want \(Q\) to be equal to \(\sigma^2_{12...s} = \left([L(1)-1][L(s)-1]\right)^{-1}C^*_{12...s}\) \hspace{1cm} (3.8)

But, with the above classifications unnecessary, the \(Q\) we derived (3.7), becomes \(Q = \alpha_{12...s}L(s+1)L(s+2)\ldots L(k)C^*_{12...s}\) \hspace{1cm} (3.9)
So, equating (3.8) and (3.9) we get

\[ \alpha_{12\ldots s} = \frac{1}{L(s+1)\ldots L(k) L(1)-1\ldots L(s)-1} \cdot \]

Now, suppose the classifications by factors \( F_{(s+2)}, \ldots, F_k \) are unnecessary.

Then \( x(i_1, i_2, \ldots, i_s, \ldots, i_k) = u(i_1, i_2, \ldots, i_s, i_{s+1}) \). Let \( C^\star \) denote the sum of squares with respect to \( u \). We want now

\[ Q = \alpha^2_{12\ldots s} L(s+2)\ldots L(k) C^\star_{12\ldots s} + \alpha^2_{12\ldots (s+1)} L(s+2)\ldots L(k) C^\star_{12\ldots (s+1)} \]

But, (3.7) reduces to

\[ Q = \alpha^2_{12\ldots s} L(s+1)\ldots L(k) C^\star_{12\ldots s} \]

So, equating (3.10) and (3.11) and using the fact that

\[ \alpha_{12\ldots s} = \frac{1}{L(s+1)\ldots L(k) L(1)-1\ldots L(s)-1} \cdot \]

we get

\[ \frac{1}{L(s+1)\ldots L(1)-1\ldots L(s)-1} C^\star_{12\ldots s} + \alpha_{12\ldots (s+1)} L(s+2)\ldots L(k) C^\star_{12\ldots (s+1)} \]

\[ = \frac{1}{L(s+1)\ldots L(1)-1\ldots L(s)-1} C^\star_{12\ldots s} \]

\[ \cdot [L(s+1)\ldots L(1)-1\ldots L(s+1)-1]^{-1} C^\star_{12\ldots (s+1)} \cdot \]

So, \( \alpha_{12\ldots (s+1)} = \frac{1}{L(s+1)L(s+2)\ldots L(k) L(1)-1\ldots L(s+1)-1} \cdot \]

We can continue thus, to obtain all the \( \alpha \)'s upto \( \alpha_{12\ldots (k-1)} \). To find \( \alpha_{12\ldots k} \) when we know the values of all the other \( \alpha \)'s, we use the equation \( E(Q) = 0 \). This equation will have only one unknown, \( \alpha_{12\ldots k} \).
Now \( C_{12 \ldots s} = L(s+1) \ldots L(k)[L(1)-1] \ldots [L(s)-1] \sigma^2_{12 \ldots s} \),

\( C_{12 \ldots (s+1)} = L(s+2) \ldots L(k)[L(k)-1] \ldots [L(s+1)-1] \sigma^2_{12 \ldots (s+1)}, \) etc.

So, the measure we are seeking, on substituting for the \( C \)’s, is

\[
Q = \sigma^2_{12 \ldots s} - \sum_{t_1=s+1}^{k} \frac{1}{L(t_1)} \sigma^2_{12 \ldots s} \cdot t_1 + \sum_{t_1<t_2} \frac{1}{L(t_1)L(t_2)} \sigma^2_{1 \ldots s} \cdot t_1 \cdot t_2 \ldots + \sum_{t_1=1}^{s+1} \frac{1}{L(s+1)L(s+2) \ldots L(k)} \sigma^2_{12 \ldots k}.
\]

\((-1)^s \frac{1}{L(s+1)L(s+2) \ldots L(k)} \sigma^2_{12 \ldots k} \cdot \)

Using Cox’s notation, we replace \( Q \) by \( \Sigma_{12 \ldots s} \). So we have

**Theorem 3.1:** Let \( F_1, F_2, \ldots, F_k \) be \( k \) factors with \( L(1), L(2), \ldots, L(k) \) levels, respectively. Then the Cox measure of effective variation for \( F_1 \times F_2 \times \ldots \times F_s \), \( i_1 \neq i_2 \neq \ldots \neq i_s = 1, 2, \ldots, k \), with \( s < k \), a subset \( \alpha \) of the factors, is given by

\[
\Sigma_{i_1, i_2 \ldots i_s} = \sigma^2_{i_1, i_2 \ldots i_s} - \sum_{t_1 \in \alpha} \frac{1}{L(t_1)} \sigma^2_{i_1, i_2 \ldots i_s} + \sum_{t_1<t_2} \frac{1}{L(t_1)L(t_2)} \sigma^2_{i_1, i_2 \ldots t_1 \cdot t_2} - \sum_{t_1<t_2<t_3} \frac{1}{L(t_1)L(t_2)L(t_3)} \sigma^2_{i_1, i_2, i_3 \ldots t_1 \cdot t_2 \cdot t_3} \ldots + \frac{1}{L(i_{s+1})L(i_{s+2}) \ldots L(i_k)} \sigma^2_{12 \ldots k},
\]

where the \( \sigma^2 \)’s are the components of variation.

**3.4 Discussion**

The ideas of Cox (1958) described above and extended to any arbitrary \( k \)-dimensional array of numbers obtained from a set of \( k \) classificatory factors, indicate a way in which the overall variation in the array can be described. One way is, of course, to give the ANOVA, and then take the resulting mean squares as representing the amount of variation associated.
The Cox alternative has a data-analytic basis. Consider, for example, a single factor \( F_1 \). Cox seeks a measure of variation associated with this factor, which will have zero expectation under random permutation of the levels of \( F_1 \) within each combination of the levels of the other factors. Also, if any of the other factors are without effect, the measure of variation should be that associated with \( F_1 \) in the lower-dimensional array with such factors omitted. The idea is then, that factor \( F_1 \) should be regarded as having no effect if the resulting "effective variation" associated with \( F_1 \) is zero. If the effective variation associated with levels of \( F_1 \) is not greater than what would be obtained on the average with random permutation of the levels of \( F_1 \), then the implication is drawn that we should regard this factor as not inducing "significant" variation. Here, the word "significant" does not have the usual meaning of significant in relation to some significance test associated with a stochastic model, but, significant with respect to the population of data sets that arise by random permutation of factor levels. So, for instance, with a \( R \times C \times T \) array, we should not regard \( T \) as contributing significant variation if

\[
\sum_T = \sigma_T^2 - \frac{1}{R} \sigma_{RT}^2 - \frac{1}{C} \sigma_{CT}^2 + \frac{1}{RC} \sigma_{RCT}^2
\]

is not greater than zero. Similarly, we should not regard, say, the interaction \( R \times T \) as contributing if

\[
\sum_{RT} = \sigma_{RT}^2 - \frac{1}{C} \sigma_{RCT}^2
\]

is not greater than zero, and so on.

Ideas supporting such measures of variation had been previously put forward from another viewpoint by Fairfield Smith (1955).

We note that if we have a purely random structure in which all
factors have a very large number of levels, the measures of effective variation reduce to the ordinary mean squares in the usual ANOVA.

In any finite situation, however, some of the effective measures of variation may be negative. This presents a curious problem with regard to tests of significance, in that, while the expected mean squares in an ANOVA for a sample from the large multi-dimensional array have a simple form, it is not clear how one should develop approximate tests of significance of whether $\Sigma > 0$, when the alternative is that it is $\leq 0$.

Attempts to understand why the measures of effective variation are those that arise in connection with randomized experiments, have not been successful.

We shall see in Chapter 4, that parts, at least, of controversies about mixed linear models, revolve around the questions of whether the $\Sigma$ quantities should be regarded as fundamental, and what the appropriate object of analysis of variance and of quantification of the amount of variation associated with different sources of variation is.

It is interesting to note that there is some relationship of the ideas of the present chapter to those of Finch described in Chapter 2. We tend to this view, because of the following: On the one hand, we have the description of the data by the usual model associated with the ANOVA. On the other hand, we have a family of descriptions, in which the levels of a factor are permuted at random. We would then take the view that including the factor as "significant" in the description of the data if the associated $\Sigma$ quantity is greater than zero, and otherwise we should regard the factor as "insignificant." In other words, if the
amount of effective variation is not greater than we would get on the average with random relabelling of factor levels, we should regard the factor as not contributing to the description of the data.

The question of why we should take the arrays formed by random permutation of levels of factors as a reference set in assessing "effective variation" is not decidable by any logical principle, just as in Finch's ideas, the class of possible descriptions has to be chosen. However, it does have an appeal to the data analyst who wishes to determine if the data suggest hypotheses. This should be the main thrust of data analysis, which we then see as a single, first step in the formulation and development of useful, predictive models.
4. THE TWO-FACTOR MIXED LINEAR MODEL

4.1. Introduction

The mixed-model controversy has been going on for a long time. The controversy is about the mean square to be used in the denominator of the F-statistic, while testing for the random effects in the mixed model. Many authors have written on this subject. Some of the writers, for example, Scheffé (1959), Wilk and Kempthorne (1955), say that the mean square for error should be used in the denominator of the F-statistic, while others like Searle (1971) and Nelder (1977) say that the mean square for Interaction should be used in the denominator of the F-statistic. In what follows, the formulations of the mixed-model as developed by the above-mentioned writers are given. An effort is made to see where they differed from each other, and what was the cause of these differences.

4.2. Derivations of the Mixed Effects Model by Various Authors

Scheffé (1959):

Scheffé gives a mixed model for the two-way layout. An example given by Scheffé, of a two-way layout in which one of the factors has fixed effects, and the other has random effects, is the example of machines and workers, where the workers are regarded as a random sample from a large population, the machines are not, the interest being in the individual performance of the machine. This would be the case if some of the machines in the experiment were of different makes.

Let A refer to machines, B to workers. It is assumed that the output of the j-th worker on the k-th day that he is assigned to the i-th
machine has the structure

$$y_{ijk} = m_{ij} + e_{ijk}, \quad (4.2.1)$$

where the "errors" \( \{e_{ijk}\} \) are independently and identically distributed with zero means and variance \( \sigma^2_e \), and independently of the "true" mean \( \{m_{ij}\} \).

The workers in the population are labelled by an index \( v \) with the population distribution \( P_v \). The "true" output of the worker labelled \( v \) on the \( i \)-th machine is denoted by \( m(i,v) \). Here \( v \) is a random quantity, corresponding to random selection of the worker according to \( P_v \), but \( i \) is not, referring to the particular machine labelled \( i \) in the experiment. The \( I \) random variables \( \{m(i,v)\} \) are the components of a vector random variable \( \mathbf{m} = \mathbf{m}(v) \) whose multivariate distribution is really the basic concept of the present model. The vector random variable

$$\mathbf{m} = \mathbf{m}(v) = (m(1,v), m(2,v), \ldots, m(I,v))' \quad (4.2.2)$$

is generated by the population of workers, the worker labeled \( v \) in the population carrying the value \( m(v) \) of the vector.

The vector of means \( \mathbb{E}(\mathbf{m}) \) for (4.2.2) will give the "true" means for the machines; i.e., the "true" mean for the \( i \)-th machine is defined to be

$$\mu_i = \mathbb{E}(m(i,v)) \quad (4.2.3)$$

where replacing \( v \) by a dot signifies that the expected value of \( m(i,v) \) has been taken w.r.t. \( P_v \). The general mean is defined as the arithmetic average of (4.2.3) over the \( I \) machines

$$\mu = \mu_\ast = \mathbb{E}(\ast\ast) \quad ,$$

where replacing \( i \) by a dot signifies that the arithmetic average has been taken over \( i \). The main effect of the \( i \)-th machine is defined as
The "true" mean for the worker labeled $v$ is defined as the average of his I "true" means on the I machines, namely, $m(\cdot, v)$. The main effect of the worker labeled $v$ in the population is defined to be

$$b(v) = m(\cdot, v) - m(\cdot, \cdot) .$$

The main effect of the worker labeled $v$, specific to the $i$-th machine is defined as $m(i, v) - m(i, \cdot)$, and hence the interaction of the $i$-th machine and the worker labeled $v$ in the population is defined as

$$c_i(v) = m(i, v) - m(i, \cdot) - m(\cdot, v) + m(\cdot, \cdot) .$$

Hence,

$$m(i, v) = \mu + \alpha_i + b(v) + c_i(v) .$$

From their definitions, the main effects and interactions in the population satisfy the following conditions:

$$\sum_i \alpha_i = 0, \quad E(b(v)) = 0, \quad \sum_i c_i(v) = 0 \quad \forall v, \quad E(c_i(v)) = 0, \quad \forall i .$$

The random effects $\{b(v), c_1(v), c_2(v), ..., c_I(v)\}$ are not independent; their variances and covariances are functions of the variance matrix of the variables $\{m(i, v)\}$. If $\sigma_{ii'} = \text{Cov}(m(i, v), m(i', v))$, then from the definitions of the random effects,

$$b(v) = I^{-1} \sum_i m(i, v) - \mu ,$$

$$\text{Var}(b(v)) = I^{-2} \sum_{ii'} \text{Cov}(m(i, v), m(i', v))$$

$$= I^{-2} \sum_{ii'} \sigma_{ii'} = \sigma_{\cdot \cdot} ,$$

$$\text{Cov}(c_i(v), c_{i'}(v)) = E[(m(i, v) - m(\cdot, v))(m(i', v) - m(\cdot, v))]$$

$$= \sigma_{ii'} - I^{-1} \sum_{i''} \sigma_{ii''} - I^{-1} \sum_{i''} \sigma_{i'i''} + I^{-2} \sum_{i''} \sum_{i'''} \sigma_{ii'''}$$

$$= \sigma_{ii'} - \sigma_{i'} - \sigma_{\cdot i'} + \sigma_{\cdot \cdot} .$$
Also, \( \sigma_{ii} = \sigma_{1'i'} \) and \( \sigma_{i'} = \sigma_{i} \) because of the symmetry of the matrix \( (\sigma_{ii'}) \).

Similarly,

\[
\text{Cov}(b(v), c_i(v)) = \sigma_{i} - \sigma_{..} .
\]

The following definitions are made:

\[
\begin{align*}
\sigma_A^2 &= (I-1)^{-1} \sum_i \alpha_i^2 , \\
\sigma_B^2 &= \text{Var}(b(v)) , \\
\sigma_{AB}^2 &= (I-1)^{-1} \sum_i \text{Var}(c_i(v)) .
\end{align*}
\]

In terms of the variance matrix of \( m(v) \), Scheffe gets,

\[
\begin{align*}
\sigma_B^2 &= \sigma_{..} , \\
\sigma_{AB}^2 &= (I-1)^{-1} \sum_i (\sigma_{ii} - \sigma_{..} ) .
\end{align*}
\]

Then, \( \sigma_B^2 = 0 \) iff \( b(v) = 0 \) \( \forall \ v \), i.e., iff the basic vector \( m(v) \) has a degenerate distribution satisfying \( \sum_i m_i(v) = \text{constant} = \mathbb{I} m \). Also, \( \sigma_{AB}^2 = 0 \) iff \( \text{Var}(c_i(v)) = 0 \) \( \forall \ i \), or \( m(i,v) = m(\cdot,v) + \alpha_i \), i.e., except for additive constants \( [\alpha_i] \), the r. v's \( m(i,v) \) are identical (not just identically distributed).

Further insight is obtained into the definitions by considering the highly symmetric case where the variance matrix of \( m(v) \) satisfies

\[
\sigma_{ii'} = \rho \sigma^2 \text{ if } i \neq i', \quad \sigma_{ii} = \sigma^2 .
\]

Then \( (4.2.10), (4.2.11), (4.2.12) \), give

\[
\begin{align*}
\sigma_B^2 &= \sigma_{1'i'}^{-1} (1 + \rho (I-1)) \\
\sigma_{AB}^2 &= \sigma^2 (1-\rho)
\end{align*}
\]

where

\[
-(I-1)^{-1} \leq \rho \leq 1 .
\]
Scheffé recommends that assumption (4.2.12) not be ordinarily made in applications where there usually exists no real symmetry corresponding to it.

If the J workers in the experiment are a random sample from \( P_v \) with labels \( \{v_1, v_2, \ldots, v_J\} \), then the "true" mean \( m_{ij} \) in (4.2.1) is \( m(i, v_j) \), and so the J vectors \( (m_{1j}, m_{2j}, \ldots, m_{Ij}) \) with I components or

\[
(m(1, v_j), m(2, v_j), \ldots, m(I, v_j))', \ j = 1, 2, \ldots, J,
\]

are independently distributed like (4.2.2). From (4.2.5),

\[
m_{ij} = \mu + \alpha_i + b_j + c_{ij},
\]

where \( b_j = b(v_j) \), and \( c_{ij} = c_i(v_j) \), and so the J vectors

\[
(b_j, c_{1j}, c_{2j}, \ldots, c_{Ij})',
\]

with I + 1 components are independently distributed like \( (b(v), c_1(v), c_2(v), \ldots, c_I(v))' \).

Suppose \( \sigma^2_{A/J} \) workers denoted the \( \sigma^2_A \) that would be defined for the I \( \times \) J layout of I machines and those J workers actually used in the experiment, with analogous definitions of \( \sigma^2_{B/J} \) workers and \( \sigma^2_{AB/J} \) workers.

These three \( \sigma^2 \)'s are then r.v.'s whose values depend on which set of J workers is sampled from the population of workers. Then \( \sigma^2_B = 0 = \sigma^2_{B/J} \) workers \( = 0 \), \( \forall \) set of J workers, and \( \sigma^2_{AB} = 0 \) has a similar implication, but \( \sigma^2_A = 0 \) does not.

If now the normality assumption is added, namely, that the vector r.v. \( m(v) \) has multivariate normal distribution, and that the \( \{e_{ijk}\} \) are normal, then the \( \Omega \) assumptions are expressed in 2 equivalent ways.
\[ y_{ijk} = m_{ij} + e_{ijk}, \] where the \( J \) vector random variables
\( (m_{ij}, m_{2j}, ..., m_{IJ})' \) are independently

\[ \Omega = N(\mu, \sigma_m^2) \] where \( \mu = (\mu_1, \mu_2, ..., \mu_J)' \),

and \( \sigma_m = (\sigma_{ij}) \) and are independent of the
\( \{e_{ijk}\} \) which are independently \( N(0,\sigma^2_e) \)

or

\[ y_{ijk} = \mu + \alpha_i + b_j + c_{ij} + e_{ijk}, \] where
\( \alpha = 0, c_{ij} = 0 \ \forall \ j, \) the \( \{b_j\}, \{c_{ij}\}, \)
\( \{e_{ijk}\} \) are jointly normal,
\( \{e_{ijk}\} \) are independently \( N(0,\sigma^2_e) \)

\[ \Omega = \text{independent of the} \ \{b_j\} \ \text{and} \ \{c_{ij}\} \text{ which have zero means and the following variances} \]
and covariances, defined in terms of an \( I \times I \) covariance matrix with elements \( \{\sigma_{ii}\} \)

\[ \text{Cov}(b_j, b_j') = \delta_{jj} \sigma_{..}, \]
\[ \text{Cov}(c_{ij}, c_{i'j'}) = \delta_{jj'}(\sigma_{ii} - \sigma_{i}, - \sigma_{i}, + \sigma_{..}), \] (4.2.13)
\[ \text{Cov}(b_j, c_{ij}) = \delta_{jj}(\sigma_{ii} - \sigma_{..}). \]

The only restriction on the \( \{\sigma_{ii}\} \) is that they be elements of a symmetric positive indefinite matrix.

If (4.2.13) is substituted into the four SS's,
\[ SS_A = JK \sum_i (\alpha_i - \alpha_{..} + e_{i..} - e_{..})^2, \] (4.2.14)
\[ SS_B = IK \sum_j (b_j - \bar{b} + e_{ij} - \bar{e} \ldots)^2, \]  
(4.2.15)

\[ SS_{AB} = K \sum_{ij} (c_{ij} - \bar{c} + e_{ij} - \bar{e}_{i} - \bar{e}_j + e \ldots)^2, \]  
(4.2.16)

\[ SS_e = \sum_{ijk} (e_{ijk} - \bar{e}_{ij})^2 \]  
(4.2.17)

Since \( c_{ij} = 0 \), \( c_{i} = 0 \), these four SS are pairwise independent except for \( SS_B \) and \( SS_{AB} \).

Table 4.2.1 gives the \( E(MS) \):

Table 4.2.1: Analysis of variance table

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>MS</th>
<th>( E(MS) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main effects of A (fixed)</td>
<td>(I-1)</td>
<td>( MS_A )</td>
<td>( \sigma^2_e + K \sigma^2_{AB} + JK \sigma^2_A )</td>
</tr>
<tr>
<td>Main effects of B (random)</td>
<td>(J-1)</td>
<td>( MS_B )</td>
<td>( \sigma^2_e + K \sigma^2_B )</td>
</tr>
<tr>
<td>A x B interactions</td>
<td>(I-1)(J-1)</td>
<td>( MS_{AB} )</td>
<td>( \sigma^2_e + K \sigma^2_{AB} )</td>
</tr>
<tr>
<td>Error</td>
<td>IJ(K-1)</td>
<td>( MS_e )</td>
<td>( \sigma^2_e )</td>
</tr>
</tbody>
</table>

Wilk and Kempthorne (1955):

Wilk and Kempthorne gave a development of a linear model. Suppose there are two factors \( A \) and \( B \) having \( A \) and \( B \) levels, respectively, which are to be studied w.r.t. a given population of \( P \) experimental units. Suppose that the experiment consists of selecting at random,

- \( a \) levels of \( A \),
- \( b \) levels of \( B \),
- \( p = rab \) experimental units,

and applying the selected \( ab \) treatment combinations at random to the selected units, so that each selected combination appears on \( r \) units.
The situation described is general because if \( A = a, B = b \), the "fixed model" is obtained; if \( A \gg a, B \gg b \), then the "random model" is obtained, and if \( A = a, B \gg b \), the "mixed model" is obtained. The symbol \( \gg \) denotes "much larger than."

To simplify the discussion it is assumed that technical errors are negligible. Then, a number is conceived which would be the response of a particular treatment combination when applied to a specified experimental unit. Thus, if

- \( i = 1, 2, \ldots, A \) denotes the levels of \( A \),
- \( j = 1, 2, \ldots, B \) denotes the levels of \( B \),
- \( k = 1, 2, \ldots, P \) denotes the experimental unit,

then a number \( Y_{ijk} \) is conceived which would be the true response if the \( k \)-th unit were subject to the treatment combination consisting of the \( i \)-th level of \( A \) and the \( j \)-th level of \( B \). Then, the set \( \{Y_{ijk}\} \) represents the total conceivable knowledge which might be experimentally acquired in this situation.

The following relationship is an algebraic identity:

\[
Y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_{ik} + n_{ijk},
\]

where \( \mu = Y_{...} \) is the conceptual overall mean response from all treatment combinations on all experimental units, \( a_i = (Y_{i..} - Y_{...}) \) is the difference between the mean response from all units when subjected to the \( i \)-th level of \( A \) in combination with each level of \( B \), and \( \mu \). The \( a_i \) is called the main effect of the \( i \)-th level of \( A \).

Then \( b_j = (Y_{.j} - Y_{...}) \) is the effect of the \( j \)-th level of \( B \), and \( (ab)_{ij} = (Y_{ij} - Y_{.j} - Y_{i..} + Y_{...}) \).
is the difference between the effect of the i-th level of A at the j-th level of B, and the effect of the i-th level of A averaged over all the levels of B. So, \((ab)_{ij}\) is the interaction of the i-th level of A and the j-th level of B. The \(e_k = (Y_{..k} - Y_{..})\) is the difference between the mean response from all treatment combinations on the k-th experimental unit, and \(\mu\). Thus, \(e_k\) measures the deviation of the k-th unit from the average of all units, w.r.t. the average of treatment combinations. The \(e_k\) is the additive error of the k-th unit.

The \(n_{ijk} = (Y_{ijk} - Y_{ij} - Y_{.k} + Y_{..})\) measures the difference between the deviation of the k-th experimental unit from the average w.r.t. treatment combination (ij) and \(e_k\). The \(n_{ijk}\) is the interactive error of the k-th unit and treatment (ij).

The above relationship is called the population model. It is based on definition. The definitions of the main effects of each factor depend in general on the levels of the other factor, and on the experimental units which are included in the experimental situation and design.

A further assumption is made (for simplification), that the \(n_{ijk}\) are all zero. This is not a trivial assumption. The population model then becomes

\[ Y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_k. \]

By definition,

\[ \sum_i a_i = \sum_j b_j = \sum_i (ab)_{ij} = \sum_j (ab)_{ij} = \sum_k e_k = 0. \]

Now, in the actual experiment, a randomly selected subset of the \(\{Y_{ijk}\}\) is observed. Let the indexes \(i^* = 1, 2, \ldots, a\), and \(j^* = 1, 2, \ldots, b\), denote the randomly selected levels of A and B, respectively, in order of
their selection. Thus, e.g., \( i^* = 1 \) corresponds to some value \( i = i_1 \).
The convention is made, however, that if \( A = a \), then \( i^* \) and \( i \) are the
same index and similarly for the case \( B = b \).

Let \( x_{i^*j^*f} \) denote the observation of the \( f \)-th replicate to which treatment combination \((i^*j^*)\) is applied, where \( f = 1, 2, \ldots, r \) for each value of \((i^*j^*)\). To each \((i^*j^*f)\), there corresponds a particular experimental unit, i.e., some value of the index \( k \).

Let \( \alpha_i^{i^*} = 1 \) if \( i^* \) corresponds to \( i \),
\[ = 0 \text{ otherwise,} \]
\( \beta_j^{j^*} = 1 \) if \( j^* \) corresponds to \( j \),
\[ = 0 \text{ otherwise,} \]
\( \rho_{k}^{i^*j^*f} = 1 \) if \((i^*j^*f)\) corresponds to \( k \),
\[ = 0 \text{ otherwise .} \]

The \( \alpha \)'s and \( \beta \)'s are dummy variables. The \( \rho \)'s are r.v.'s which specify how selected treatment combinations are assigned to experimental units. These are random variables because random methods of selection and allocation are employed. From the design of the experiment, certain distributional properties of these quantities are obtained, e.g.,
\[ \text{Prob}(\alpha_i^{i^*} = 1) = \frac{1}{A}, \]
\[ \text{Prob}(\alpha_i^{i^*} = 1, \alpha_{i'}^{i^*'} = 1) = \frac{1}{A(A-1)}, i \neq i', i^* \neq i^*'. \]
The \( \{\alpha_i^{i^*}\}, \{\beta_j^{j^*}\}, \{\rho_{k}^{i^*j^*f}\} \) are groupwise independent etc.

A model for \( x_{i^*j^*f} \) under the assumption all \( n_{ijk} = 0 \) is given by
\[ x_{i^*j^*f} = \mu + \sum_i \alpha_i^{i^*} a_i + \sum_j \beta_j^{j^*} b_j + \sum_{ij} \alpha_i^{i^*} \beta_j^{j^*} (ab)_{ij} + \sum_k \rho_{k}^{i^*j^*f} e_k. \]

This relationship is called the statistical model. The population model is derived from the experimental situation and the statistical
model is then obtained by imposing the conditions of the design of the experiment. The r.v.'s in the statistical model are the $\alpha_i^*, \beta_j^*, \gamma_k^*$ which take the values 0 or 1 with known probabilities. All the other quantities in the model are fixed, unknown parameters.

This development is general. For example, $A = a, B \gg b$, gives the "mixed model" situation with $A$ fixed and $B$ random. Then taking $i$ and $i^*$ to be identical,

$$\sum_i \alpha_i^* a_i = a_i^*, \sum_i \alpha_i^*(ab)_{ij} = (ab)_{ij}.$$

Hence, the mixed model becomes

$$x_i^* j^* f = x_{ij^*}^* f = \mu + a_i + \sum_j \beta_j^* b_j + \sum_j \beta_j^* (ab)_{ij} + \sum_k \gamma_k^* e_k.$$

The algebraic structure of the analysis of variance for this design is given in Table 4.2.2.

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>S.S.</th>
<th>M.S.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$(a-1)$</td>
<td>$A' = \sum_{i^<em>} (x_{i^</em>} - x_{..})^2$</td>
<td>$A^* = \frac{A'}{a-1}$</td>
</tr>
<tr>
<td>$B$</td>
<td>$(b-1)$</td>
<td>$B' = \sum_{j^<em>} (x_{..j^</em>} - x_{..})^2$</td>
<td>$B^* = \frac{B'}{b-1}$</td>
</tr>
</tbody>
</table>
| $A \times B$    | $(a-1)(b-1)$ | \begin{align*} I' &= \sum_{i^* j^*} (x_{i^* j^*} - x_{..} - x_{.j^*} - x_{i^*} + x_{..} )^2 \\
                        &\text{(a-l)(b-l)} \end{align*} | $I^* = \frac{I'}{(a-1)(b-1)}$ |
| Residual        | $ab(r-1)$ | $R' = \sum_{i^* j^* f} (x_{i^* j^* f} - x_{i^*} - x_{.j^*} - x_{..} )^2$ | $R^* = \frac{R'}{ab(r-1)}$ |
| Total           | $abr - 1$ | \begin{align*} G' &= \sum_{i^* j^* f} (x_{i^* j^* f} - x_{..} )^2 \\
                        &\text{(a-1)(b-1)(r-1)} \end{align*} | $G^* = \frac{G'}{(a-1)(b-1)(r-1)}$ |

Using the derived statistical model and some distributional properties of the $\{\alpha_i^*\}$, $\{\beta_j^*\}$ and $\{\gamma_k^*\}$, the expectations of the mean squares can be obtained. The results are given in Table 4.2.3. The
following definitions have been used:

\[ \sigma^2_a = \frac{1}{A-1} \sum_{i} a_i^2, \]
\[ \sigma^2_b = \frac{1}{B-1} \sum_{j} b_j^2, \]
\[ \sigma^2_{ab} = \frac{1}{(A-1)(B-1)} \sum_{ij} (ab)^2_{ij}, \]
\[ \sigma^2_e = \frac{1}{I-k} \sum_{k} e_k^2. \]

Table 4.2.3: Expectations of mean squares under assumptions of unit-treatment additivity

<table>
<thead>
<tr>
<th>Mean Square</th>
<th>Expectation of Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>A*</td>
<td>( \sigma^2_e + \frac{(B-b)}{B} \sigma^2_{ab} + rb \sigma^2_a )</td>
</tr>
<tr>
<td>B*</td>
<td>( \sigma^2_e + \frac{(A-a)}{A} \sigma^2_{ab} + rb \sigma^2_b )</td>
</tr>
<tr>
<td>I*</td>
<td>( \sigma^2_e + rb \sigma^2_{ab} )</td>
</tr>
<tr>
<td>R*</td>
<td>( \sigma^2_e )</td>
</tr>
</tbody>
</table>

Now, \( E_a^* \) is said to be the proper error term for A main effects, if \( E_a^* \) is a linear combination of the analysis of variance mean squares such that \( E(E_a^*) = E(A^*) - rb \sigma^2_a \) where \( E(E_a^*) \) is the expectation of \( E_a^* \).

The proper error terms, with unit-treatment additivity are given in Table 4.2.4.

Table 4.2.4. Proper error terms with unit-treatment additivity

<table>
<thead>
<tr>
<th>Category</th>
<th>Proper Error Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>A effects</td>
<td>( I^* - \frac{b}{B}(I^* - R^<em>) = \frac{B-b}{B} I^</em> + \frac{b}{B} R^* )</td>
</tr>
<tr>
<td>B effects</td>
<td>( I^* - \frac{a}{A}(I^* - R^<em>) = \frac{A-a}{A} I^</em> + \frac{a}{A} R^* )</td>
</tr>
<tr>
<td>A x B interactions</td>
<td>( R^* )</td>
</tr>
</tbody>
</table>
The special cases are:

(a) **Fixed Model:** \( A = a, B = b \)

The proper error term for each of \( A, B \) and \( A \times B \) is \( R^* \).

(b) **Random Model:** \( A \gg a, B \gg b \).

Now \( \frac{\bar{a}}{A}, \frac{\bar{b}}{B} \) approach zero and the proper error term for \( A \) and also for \( B \) is \( I^* \).

(c) **Mixed Model:** \( A \gg a, B = b \)

Here \( A \) is random and \( B \) is fixed. The proper error term for the random factor \( A \) is \( R^* \) and that for the fixed factor \( B \) is \( I^* \).

Results under general conditions are also given by Wilk and Kempthorne. Now the assumption that unit-treatment interactions are zero is not made, i.e., that \( n_{ijk} = 0 \) for all \( i, j, k \). Then, \( \sum_{ij} n_{ijk} = \sum_k n_{ijk} = 0 \).

The general statistical model is

\[
x_{ijk} = \mu + \sum_{i} \alpha_i a_i + \sum_{j} \beta_j b_j + \sum_{ij} \alpha_i \beta_j (ab)_{ij} + \sum_{ik} \rho_i e_{ik} + \sum_{j} \gamma_j \beta_j n_{ijk}.
\]

The analysis of variance is as given in Table 4.2.2. The expectations of the mean squares under these general conditions are given in Table 4.2.5. The following definitions are made:

\[
\begin{align*}
\sigma_n^2 &= \frac{1}{AB(P-1)} \sum_{ijk} n_{ijk}^2, \\
\sigma_{an}^2 &= \frac{1}{(A-1)(P-1)} \sum_i n_{i.k}^2, \\
\sigma_{bn}^2 &= \frac{1}{(B-1)(P-1)} \sum_j n_{.jk}^2, \\
\sigma_{abn}^2 &= \frac{1}{(A-1)(B-1)(P-1)} \sum_{ijk} (n_{ijk} - n_{i,k} - n_{.jk})^2.
\end{align*}
\]

It will be seen from Table 4.2.5, that in general, proper error terms do not exist. However, the "bias" is of order \( \frac{1}{P} \), where \( P \) is the
number of experimental units in the population. In many cases \( P \) will be large and hence the bias will be negligible.

Table 4.2.5. Expectations of mean squares (more general conditions)

<table>
<thead>
<tr>
<th>Mean Square</th>
<th>Expectation of Mean Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A^* )</td>
<td>( \sigma^2_n + \sigma^2_e + \frac{B-b}{B} \cdot r \left[ \sigma^2_{ab} - \frac{1}{p} \sigma^2_{abn} \right] + rb \left[ \sigma^2_a - \frac{1}{p} \sigma^2_{an} \right] )</td>
</tr>
<tr>
<td>( B^* )</td>
<td>( \sigma^2_n + \sigma^2_e + \frac{(A-a)}{A} \cdot r \left[ \sigma^2_{ab} - \frac{1}{p} \sigma^2_{abn} \right] + ra \left[ \sigma^2_b - \frac{1}{p} \sigma^2_{bn} \right] )</td>
</tr>
<tr>
<td>( I^* )</td>
<td>( \sigma^2_n + \sigma^2_e + r \left[ \sigma^2_{ab} - \frac{1}{p} \sigma^2_{abn} \right] )</td>
</tr>
<tr>
<td>( R^* )</td>
<td>( \sigma^2_n + \sigma^2_e )</td>
</tr>
</tbody>
</table>

There is, however, a defect in this work, when we consider the mixed model, that is, when one factor is fixed. It is concerned only with E.M.S's and does not consider the covariance structure of the observations except with random relabelling of the levels of the fixed factor. In the mixed model case, however, one will be interested in the properties of the observations without such random relabelling. This defect will be remedied later.

The use of the models given above is in a real sense quite incidental. They were for Wilk and Kempthorne a means of establishing EMS's with random sampling of the hypothesized population. The EMS's here do not depend on a specific model and could have been obtained purely from standard but somewhat complex sampling of the conceptual multi-dimensional array. Searle (1971):

In his book, Searle (1971) gathers together ideas of random and mixed models. The original formulations of fixed versus random models was given
by Eisenhart (1947). Contributions were made by many writers referenced in Searle's book. Specific discussions, other than by Searle, relating to the mixed model will be given later. In the following, the actual exposition of Searle (1971) is taken as a basis for presentation and discussion.

We shall try to summarize the ideas of Searle, using his mode of writing and description. Searle gives the two-way classification with interaction. He gives the equation of the linear model as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$  \hspace{1cm} (4.2.18)

where

- $y_{ijk}$ is the $k$-th observation in the $i$-th level of the $\alpha$ factor and the $j$-th level of the $\beta$ factor,
- $\mu$ is the general mean,
- $\alpha_i$ is the effect on $y_{ijk}$ due to the $i$-th level of the $\alpha$ factor,
- $\beta_j$ is the effect on $y_{ijk}$ due to the $j$-th level of the $\beta$ factor,
- $\gamma_{ij}$ is the interaction effect for the $i$-th level of the $\alpha$ factor and the $j$-th level of the $\beta$ factor,
- $e_{ijk}$ is the random error term peculiar to $y_{ijk}$ with $e \sim (0, \sigma^2 I)$.

For the non-empty cells,

$$y_{ij} = \sum_{k=1}^{n_{ij}} y_{ijk}$$

is total yield in cell (ij),

$$\bar{y}_{ij} = \frac{y_{ij}}{n_{ij}}$$

is the corresponding mean.

Here, replacing an index by a dot implies that summation over all values of that index has taken place, and placing a bar over the $y$ indicates the mean or average value.
At this stage, the only assumption made by Searle is that \( e \sim (0, \sigma^2 I) \).

Let \( R(\mu) = \text{SS due to fitting a mean } \mu \),

\[ R(\alpha | \mu) = \text{SS due to fitting } \alpha \text{ after } \mu, \]

\[ R(\beta | \mu, \alpha) = \text{SS due to fitting } \beta \text{ after fitting } \mu \text{ and } \alpha. \]

So,

\[ R(\mu) = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{i,j}^2 / n_{i,j}; \quad (4.2.19) \]

\[ R(\mu, \alpha) = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{i,j}^2 / n_{i,j}; \quad (4.2.20) \]

\[ R(\mu, \beta) = \sum_{i=1}^{a} \sum_{j=1}^{b} y_{i,j}^2 / n_{i,j}; \quad (4.2.21) \]

Model (4.2.18) involves the terms \( \mu, \alpha_i, \beta_j \) and \( \gamma_{ij} \). Hence, the SS for fitting it is

\[ R(\mu, \alpha, \beta, \gamma) = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} y_{i,j,k}^2 / n_{i,j,k}; \quad (4.2.22) \]

Searle says that formally, it might seem possible to define

\[ R(\beta | \mu, \alpha, \gamma) = R(\mu, \alpha, \beta, \gamma) - R(\mu, \alpha, \gamma). \]

However, before doing this one must look carefully at the meaning of the interaction \( \gamma \)-factor, because it will be found that \( R(\beta | \mu, \alpha, \gamma) \) is formally defined by the notation as identically equal to zero. For \( R(\mu, \alpha, \beta, \gamma) \), the model is (4.2.18), and \( R(\mu, \alpha, \beta, \gamma) = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{i,j} \gamma_{i,j}^2 \) as in (4.2.22). Similarly, in the context of the \( \alpha \)'s and \( \gamma \)'s of (4.2.18), the implied model for \( R(\mu, \alpha, \gamma) \) is \( y_{i,j,k} = \mu + \alpha_i + \gamma_{i,j} + e_{i,j,k} \). But this is exactly the model of the 2-way nested classification. Hence, the reduction in SS is

\[ R(\mu, \alpha, \gamma) = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{c} n_{i,j,k} \gamma_{i,j,k}^2; \quad (4.2.23) \]

and so \( R(\beta | \mu, \alpha, \gamma) = R(\mu, \alpha, \beta, \gamma) - R(\mu, \alpha, \gamma) \equiv 0 \). Similarly,
and so
\[ R(\alpha, \beta, \gamma) = 0 = R(\alpha, \beta, \gamma) . \]

Thus, the reduction SS due to fitting any model that contains the interaction \( \gamma \)-factor, is \( \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} \bar{y}_{ij}^2 \). More particularly, in (4.2.23) and (4.2.24) the reduction due to fitting any model which compared to (4.2.18) lacks either \( \alpha \) or \( \beta \) or both, is equal to \( R(\alpha, \beta, \gamma) = \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} \bar{y}_{ij}^2 \). Thus, \( \gamma \) cannot be fitted unless both \( \alpha \) and \( \beta \) are in the model.

Table 4.2.6 gives the analysis of variance for balanced data.

Table 4.2.6. Analysis of variance for a 2-way classification interaction model, with balanced data

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>d.f.</th>
<th>S.S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1</td>
<td>SSM = Ny^2</td>
</tr>
<tr>
<td>A-factor</td>
<td>a-1</td>
<td>SSA = ( bn \sum_{i=1}^{a} \bar{y}_{i..}^2 )</td>
</tr>
<tr>
<td>B-factor</td>
<td>b-1</td>
<td>SSB = ( an \sum_{j=1}^{b} \bar{y}_{.j}^2 )</td>
</tr>
<tr>
<td>AB interaction</td>
<td>(a-1)(b-1)</td>
<td>SSAB = ( n \sum_{i=1}^{a} \sum_{j=1}^{b} \bar{y}_{ij}^2 )</td>
</tr>
<tr>
<td>Residual Error</td>
<td>ab(n-1)</td>
<td>SSE = ( \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (y_{ijk} - \bar{y}<em>{ij..} - \bar{y}</em>{.j..} - \bar{y}_{...})^2 )</td>
</tr>
<tr>
<td>Total</td>
<td>N = abn</td>
<td>SST = ( \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} y_{ijk}^2 )</td>
</tr>
</tbody>
</table>

The model is
\[ y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} , \] (4.2.25)
with \( i = 1, \ldots, a \), \( j = 1, 2, \ldots, b \), \( k = 1, 2, \ldots, n \). To get the expected values of the SS, Searle writes \( \tilde{\alpha}_i = \frac{\alpha_i}{a} \) and \( \tilde{\beta}_j, \tilde{\gamma}_{ij} \) etc. in an analogous manner. From (4.2.25),

\[
\bar{y}_{ij} = \mu + \tilde{\alpha}_i + \tilde{\beta}_j + \tilde{\gamma}_{ij} + \epsilon_{ij} \, ,
\]

The values in (4.2.25) and (4.2.26) are substituted in the SS of Table 4.2.6.

Now, no matter what model is used the error terms are assumed to have zero means and variance \( \sigma_e^2 \), and to be independent of one another. Further, the expected value of the product of an error term with \( \mu \) an \( \alpha \), a \( \beta \) or a \( \gamma \) is zero.

Using this, the expected values of the mean SS are:

\[
\begin{align*}
E(\text{MSM}) & = E[N(\mu + \tilde{\alpha}_i + \tilde{\beta}_j + \tilde{\gamma}_{ij})^2] + \sigma_e^2, \\
E(\text{MSA}) & = \frac{bn}{a-1} \sum_{i=1}^a E(\alpha_i - \bar{\alpha}_i)^2 + \sigma_e^2, \\
E(\text{MSB}) & = \frac{an}{b-1} \sum_{j=1}^b E(\beta_j - \bar{\beta}_j)^2 + \sigma_e^2, \\
E(\text{MSAB}) & = \frac{n}{(a-1)(b-1)} \sum_{i=1}^a \sum_{j=1}^b E(\gamma_{ij} - \bar{\gamma}_{ij})^2 + \sigma_e^2, \\
E(\text{MSE}) & = \sigma_e^2.
\end{align*}
\]

These results hold whether the model is fixed, random or mixed.

To obtain his results for the mixed model, Searle supposes that \( \alpha \) effects are fixed and the \( \beta \)'s and \( \gamma \)'s are random. Then, he gets the expectations as shown in the table below.
Table 4.2.7. Expected mean squares of a 2-way classification interaction model, with balanced data (Searle)

<table>
<thead>
<tr>
<th>Mixed model: α's fixed, β's and γ's random</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(MSM) = abn(μ + \bar{α})^2 + anσ^2 + noσ^2 + σ^2</td>
</tr>
<tr>
<td>E(MSA) = \frac{bn}{a-1} \sum_{i=1}^{a} (α_i - \bar{α})^2 + nσ^2 + σ^2</td>
</tr>
<tr>
<td>E(MSB) = \frac{an}{σ^2} + nσ^2 + σ^2</td>
</tr>
<tr>
<td>E(MSAB) = nσ^2 + σ^2</td>
</tr>
<tr>
<td>E(MSE) = σ^2</td>
</tr>
</tbody>
</table>

Searle stresses that the expectations in Table 4.2.7 are arrived at without making any use of what he calls the "usual restrictions" on the elements of the model. He says, if, however, the restriction \( \sum_{i=1}^{a} α_i = 0 \) is taken as part of the mixed model, the E(MSA) of Table 4.2.7 reduces to

\[ E(MSA) = \frac{bn}{a-1} \sum_{i=1}^{a} α_i^2 + nσ^2 + σ^2. \]

An alternative mixed model that is often used is

\[ y_{ijk} = μ'' + α''_i + β''_j + γ''_ij + e_{ijk} \quad (4.2.27) \]

with the restriction

\[ \sum_{i=1}^{a} γ''_ij = γ''_i = 0, ∀j \quad (4.2.28) \]

Searle, it appears, interprets (4.2.28) to imply (4.2.29).

\[ \text{Cov}(γ''_{ij}, γ''_{i'j}) = c, ∀ i ≠ i' \text{ and } j. \quad (4.2.29) \]

Then from (4.2.28), \( \text{Var}(\sum_{i=1}^{a} γ''_{i}) = 0 \) gives

\[ c = -\frac{σ^2''}{a-1}, \quad (4.2.30) \]
and

\[ \text{Cov}(\gamma_{ij}, \gamma_{ij'}) = 0 \quad \forall \ i \text{ and } j \neq j'. \quad (4.2.31) \]

Searle then obtains the following:

\[ E(\text{MSM}) = N(\mu'' + \tilde{\alpha}'')^2 + \sum_{i=1}^{a} \sum_{j=1}^{b} \gamma''_{i}^2 + \sigma_e^2 \]

\[ E(\text{MSA}) = \frac{bn}{b-1} \left[ \sum_{i=1}^{a} (\alpha'' - \tilde{\alpha}'')^2 + \sum_{i=1}^{a} \gamma''_{i}^2 \right] + \sigma_e^2 \]

\[ E(\text{MSB}) = \frac{an}{a-1} \left[ \sum_{j=1}^{b} (\beta'' - \tilde{\beta}'')^2 \right] + \sigma_e^2 \]

\[ E(\text{MSAB}) = \frac{n}{(a-1)(b-1)} \sum_{i=1}^{a} \sum_{j=1}^{b} \gamma''_{ij}^2 + \sigma_e^2 \]

\[ E(\text{MSE}) = \sigma_e^2. \]

His expected values are shown in Table 4.2.8.

### Table 4.2.8. Expected mean squares of a 2-way classification interaction model with balanced data (Searle)

<table>
<thead>
<tr>
<th>Mixed model with restrictions on interactions: ( \gamma''_{ij} = 0 \quad \forall \ j )</th>
</tr>
</thead>
<tbody>
<tr>
<td>E(\text{MSM}) = N(\mu'' + \tilde{\alpha}'')^2 + \sum_{i=1}^{a} \gamma''_{i}^2 + \sigma_e^2</td>
</tr>
<tr>
<td>E(\text{MSA}) = \frac{bn}{b-1} \left[ \sum_{i=1}^{a} (\alpha'' - \tilde{\alpha}'')^2 + \sum_{i=1}^{a} \gamma''_{i}^2 \right] + \sigma_e^2</td>
</tr>
<tr>
<td>E(\text{MSB}) = \frac{an}{a-1} \left[ \sum_{j=1}^{b} (\beta'' - \tilde{\beta}'')^2 \right] + \sigma_e^2</td>
</tr>
<tr>
<td>E(\text{MSAB}) = \frac{n}{(a-1)(b-1)} \sum_{i=1}^{a} \sum_{j=1}^{b} \gamma''_{ij}^2 + \sigma_e^2</td>
</tr>
<tr>
<td>E(\text{MSE}) = \sigma_e^2</td>
</tr>
</tbody>
</table>

A relationship between Tables 4.2.7 and 4.2.8 is established. The model for Table 4.2.7 is given as

\[ y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}, \]
with $\alpha$ fixed, $\beta$ and $\gamma$ random. Searle rewrites it as

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk},$$

where

$$\mu'' = \mu, \alpha_i'' = \alpha_i, \beta_j'' = \beta_j + \gamma_{ij}$$

$$\gamma_{ij}'' = \gamma_{ij} - \gamma_{.j}. \quad (4.2.33)$$

Thus, he obtains model (4.2.27) corresponding to Table 4.2.8.

Nelder (1977):

Nelder gives what he calls a reformulation of linear models. Capital letters $A, B, C, \ldots$ are used to denote factors, associating indices $i, j, k, \ldots$ with the factors. Factor $A$ has $N_A$ levels in the population, and $n_A$ levels in a sample from that population, so $n_A < N_A$. A nested structure $A/B$ gives the model formula $A + A \cdot B$, where $A/B$ means $B$ within $A$. A crossed structure $A \times B$ gives $A + B + A \cdot B$, where $A \cdot B$ now is the interaction of $A$ and $B$. A simple term $B$ corresponds to a one-dimensional set of parameters $\beta_j$, and a compound term as $A \cdot B$ exemplifies a multi-dimensional set. Estimates of the parameters are given in Latin letters, e.g., $b_j$. The population elements are denoted by $x$ and the sample values by $y$.

Nelder samples from the levels of the factors. He calls it complete sampling if $n = N$, where $N$ is necessarily finite. The parameters in the linear model now, he says, correspond to what are usually called fixed effects. If a sub-set of size $n$ is selected non-randomly from a population of size $N'$, then the inferences are conditioned by the $n$ actually selected and $N'$ is reduced to $n$. It is incomplete sampling when $n < N$, i.e., a random sample of $n$ levels has been chosen from a population.
of N levels.

For simplicity, Nelder restricts himself to first and second order moments, and deals only with balanced structures having equal numbers in all the sub-classes. He distinguishes observational data from experimental data. The investigator may choose how to classify observational data, i.e., define the factors and their levels, but he has no control over the value of a factor level for a particular observational unit. When, however, experiments are done on a set of units, treatment factors and levels are defined by the investigator, who can now control the values of these factor levels for each unit.

Nelder first considers simple crossed structures. The structure is $A \times B$, where the levels of A index a row population and those of B a column population. Elements are defined for each intersection of a row and column.

If a population is finite, then the members are treated as equi-probable. An infinite population is assumed to have a probability distribution of the scale-and-location type, i.e., $f \left\{ \frac{x-u}{\sigma} \right\}$ with finite variance. The row and column populations may be finite or infinite. The scale parameter is assumed not to depend on the indexing. He defines $\sigma_A^2$ and $\sigma_B^2$ as the variance components of rows and columns, respectively, and $\sigma_{AB}^2$ as the corresponding interaction component. The population linear identity takes the form

$$x_{ij} = x_{..} + (x_{i..} - x_{..}) + (x_{.j..} - x_{..}) + (x_{ij..} - x_{i..} - x_{.j..} + x_{..}) \quad (4.2.34)$$

This expresses an element as the sum of the grand mean, the deviation of the i-th population mean from the grand mean, the deviation
of the j-th population mean from the grand mean and \((x_{ij} - x_i - x_j + x..)\).

He defines

\[
\begin{align*}
\sigma^2_A &= \frac{\sum_i (x_{i..} - x..)^2}{N_A - 1}, \\
\sigma^2_B &= \frac{\sum_j (x_{j..} - x..)^2}{N_B - 1}, \\
\sigma^2_{AB} &= \frac{\sum_{ij} (x_{ij} - x_{i..} - x_{j..} + x..)^2}{(N_A - 1)(N_B - 1)},
\end{align*}
\]

where \(N_A, N_B\) are finite.

If \(N_A, N_B\) are infinite, then averaging is replaced by integration over the corresponding suffix. By definition, \(x..\) is a constant, and so generates no variance component.

The population quadratic identity is derived by squaring the linear identity and taking expectations to give

\[
\begin{align*}
\text{Ex}^2_{ij} &= \text{Ex}^2_{ij} + \text{Ex}(x_{i..} - x..)^2 + \text{Ex}(x_{j..} - x..)^2 + \text{Ex}(x_{ij} - x_{i..} - x_{j..} + x..)^2, \\
&= \sigma^2_A (1 - \frac{1}{N_A}) + \sigma^2_B (1 - \frac{1}{N_B}) + \sigma^2_{AB} (1 - \frac{1}{N_A})(1 - \frac{1}{N_B}).
\end{align*}
\]

(4.2.35)

which for finite \(N_A\) and \(N_B\) can be written

\[
\text{Var}(x_{ij}) = \sigma^2_A (1 - \frac{1}{N_A}) + \sigma^2_B (1 - \frac{1}{N_B}) + \sigma^2_{AB} (1 - \frac{1}{N_A})(1 - \frac{1}{N_B}).
\]

(4.2.36)

Nelder considers the covariance structure next. There are now four distinct pairs of elements: identical, non-identical in same row, non-identical in same column, not in same row or column.

The four covariances are written as:

\[
\begin{align*}
\text{Cov}(x_{ij}, x_{ij}) &= \sigma^2, \\
\text{Cov}(x_{ij}, x_{i',j}) &= \rho_{12} \sigma^2, \\
\text{Cov}(x_{ij}, x_{i',j'}) &= \rho_{2} \sigma^2, \\
\text{Cov}(x_{ij}, x_{i',j'}) &= \rho_{1} \sigma^2.
\end{align*}
\]

(4.2.37)
Nelder then writes the population linear model in the form,

\[ x_{ij} = \mu + \alpha_i + \beta_j + (\alpha \beta)_{ij}, \]

i.e., \( x_{ij} \) is expressed in terms of \( \mu, \alpha_i, \beta_j, \gamma_{ij} \) which he says are random variables in a "certain formal sense." It is assumed that all the components on the RHS are uncorrelated. He claims that if to the components the following "variances" are assigned, then the covariance structure (4.2.37) is reproduced:

\[
\begin{align*}
\phi &= \text{Var} \mu = \rho \sigma^2 \\
\phi_A &= \text{Var} \alpha_i = (\rho_1 - \rho) \sigma^2 \\
\phi_B &= \text{Var} \beta_j = (\rho_2 - \rho) \sigma^2 \\
\phi_{AB} &= \text{Var}(\alpha \beta)_{ij} = (\rho_{12} - \rho_1 - \rho_2 + \rho) \sigma^2.
\end{align*}
\]

(4.2.38)

The quantities \( \phi \), he calls canonical components, and gets the relation between the \( \phi \)'s and the variance components and covariances. He refers to Fairfield Smith (1955) for use of the name "canonical component."

This is done for \( N_A, N_B \) finite. For infinite populations he gets results by letting \( N \to \infty \).

\[
\begin{align*}
\phi &= -\frac{\sigma_A^2}{N_A} - \frac{\sigma_B^2}{N_B} + \frac{\sigma_{AB}^2}{N_A N_B}, \\
\phi_A &= \sigma_A^2 - \frac{\sigma_{AB}^2}{N_A}, \\
\phi_B &= \sigma_B^2 - \frac{\sigma_{AB}^2}{N_B}, \\
\phi_{AB} &= \sigma_{AB}^2,
\end{align*}
\]
\[ \rho_{12}\sigma^2 = \phi + \phi_A + \phi_B + \phi_{AB}, \]
\[ \rho_2\sigma^2 = \phi + \phi_B, \]
\[ \rho_1\sigma^2 = \phi + \phi_A, \]
\[ \rho\sigma^2 = \phi. \]

He considers that the canonical components are the formal variances of the components of the population linear model, formal because \( \text{Var} \mu \) is not necessarily positive, and \( \text{Var} \alpha_i \) will be negative if \( \sigma_A^2 < \frac{\sigma_{AB}^2}{N_B} \), and so will \( \text{Var} \beta_j \) if \( \sigma_B^2 < \frac{\sigma_{AB}^2}{N_A} \).

The \( \phi \)'s coincide with the variance components if \( N_A \) and \( N_B \) both tend to infinity, then the standard infinite model with \( \text{Var} \mu = 0, \text{Var} \alpha_i = \sigma_A^2, \text{Var} \beta_j = \sigma_B^2, \text{Var} \gamma_{ij} = \sigma_{AB}^2 \), is recovered.

Thus, \( \text{Var} \gamma_{ij} \) is the interaction variance, \( \text{Var} \beta_j \) is the excess variance in the B margin over the interaction variance, \( \text{Var} \alpha_i \), similarly, for the A margin, while \( \text{Var} \mu + x^2 \) gives the excess of the square of the grand mean over the sum of the interaction variance and the two marginal excess variances.

Nelder claims that there is a difference between the linear identity and the linear model. The structures of the two are similar, but one must not equate the corresponding terms because, for example, \( x_{..} = \mu + \alpha_i + \beta_j + \gamma_{ij} \), and \( x_{..} \) is not necessarily \( \mu \).

By constraining the parameters, i.e., setting \( \sum \alpha_i = 0, \sum \beta_j = 0, \sum \gamma_{ij} = 0 = \sum \gamma_{ij} \), the corresponding terms in the linear identity and the linear model are similar. Without constraints, there are more
parameters than elements.

Next, Nelder gives the analysis of the samples. Suppose a sample of values $y_{ij}$ is generated from a $A \times B$ structure, $i = 1, 2, \ldots, n_A$, $j = 1, 2, \ldots, n_B$. If $N_A$ and $N_B$ are finite, then this amounts to choosing a sample of $n_A$ levels of $A$ and $n_B$ levels of $B$ at random without replacement. If either becomes infinite, it amounts to choosing a sample of $n$ from the appropriate distribution. It is noted that $n \leq N$, with equality implying that the sample is complete w.r.t. that factor. The linear and quadratic identities which arise due to the sample are:

$$ y_{ij} = y_{..} + (y_{i..} - y_{..}) + (y_{.j} - y_{..}) + (y_{ij} - y_{i..} - y_{.j} + y_{..}) , $$

$$ \sum_{ij} y_{ij}^2 = \sum_{ij} y_{..}^2 + \sum_{ij} (y_{i..} - y_{..})^2 + \sum_{ij} (y_{.j} - y_{..})^2 + \sum_{ij} (y_{ij} - y_{i..} - y_{.j} + y_{..})^2 , $$

where dots denote averaging over the sample, which now takes the place of expectations.

The sample quadratic identity gives the sample analysis of variance, and the expected mean squares are derived by substituting the population linear model, and using (4.2.38) to get the following table:

**Table 4.2.9. Expectations of mean squares of Nelder**

<table>
<thead>
<tr>
<th>Component</th>
<th>d.f.</th>
<th>EMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>1</td>
<td>$\phi_{AB} + n_B \phi_A + n_A \phi_B + n_A n_B (\phi + \mu^2)$</td>
</tr>
<tr>
<td>A</td>
<td>$n_A - 1$</td>
<td>$\phi_{AB} + n_B \phi_A$</td>
</tr>
<tr>
<td>B</td>
<td>$n_B - 1$</td>
<td>$\phi_{AB} + n_A \phi_B$</td>
</tr>
<tr>
<td>A•B</td>
<td>$(n_A - 1)(n_B - 1)$</td>
<td>$\phi_{AB}$</td>
</tr>
</tbody>
</table>
The EMS show the central role of the $\phi$'s. The separation of the $\phi$'s which are population quantities, from the multipliers which are sample quantities is noted; each EMS contains the $\phi$ with corresponding suffices, and also contributions from $\phi$'s of all terms to which that term is marginal. The forms of the EMS hold for finite or infinite populations and for complete or incomplete sampling.

The relation of this work to that of Wilk and Kempthorne will be discussed later.

4.3. An Alternative Approach to Derivation of a Linear Model for the Mixed Factor Case

We see that Searle specifies a linear model, and then to get the same EMS's as Scheffé, and Wilk and Kempthorne, he incorporates what he calls "usual restrictions" on the levels of the fixed factor.

However, we may proceed on a different route, on the lines of Wilk and Kempthorne, by deriving an appropriate linear model. The question at issue is the following. Suppose we write a linear model, assuming no technical errors, $y_{ij} = \mu + a_i + b_j + (ab)_{ij}$, what assumptions are reasonable with respect to the components of such a model?

A mode of examining this is given below. It is based partly on some unpublished notes of O. Kempthorne. As we proceed, we shall get forms for the EMS's that are the same as those of Scheffé (1959), and also those of Wilk and Kempthorne (1955) in the limiting case of $a = A$. It will be seen that part of this development is almost isomorphic to that of Scheffé. The point of the development is, however, to examine the appropriate linear model, which is a question beyond that of
The assumption is made that there exists a 2-way table of true values \( \{\tau_{ij}, i = 1,2,\ldots,v; j = 1,2,\ldots,p\} \). We refer to the rows as varieties and columns as places, because that is a very common situation in agronomy. The conceptual population of true values is sketched below:

Table 4.3.1. True values in a two-way layout

<table>
<thead>
<tr>
<th>Place (j)</th>
<th>1</th>
<th>2</th>
<th>\ldots</th>
<th>\ldots</th>
<th>P</th>
<th>Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ( \tau_{11} )</td>
<td>( \tau_{12} )</td>
<td>\ldots</td>
<td>( \tau_{1j} )</td>
<td>\ldots</td>
<td>( \tau_{1p} )</td>
<td>( \tau_1. )</td>
</tr>
<tr>
<td>2 ( \tau_{21} )</td>
<td>( \tau_{22} )</td>
<td>\ldots</td>
<td>( \tau_{2j} )</td>
<td>\ldots</td>
<td>( \tau_{2p} )</td>
<td>( \tau_2. )</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\ldots</td>
<td>\vdots</td>
<td>\ldots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>i ( \tau_{i1} )</td>
<td>( \tau_{i2} )</td>
<td>\ldots</td>
<td>( \tau_{ij} )</td>
<td>\ldots</td>
<td>( \tau_{ip} )</td>
<td>( \tau_i. )</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\ldots</td>
<td>\vdots</td>
<td>\ldots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>v ( \tau_{v1} )</td>
<td>( \tau_{v2} )</td>
<td>\ldots</td>
<td>( \tau_{vj} )</td>
<td>\ldots</td>
<td>( \tau_{vp} )</td>
<td>( \tau_v. )</td>
</tr>
</tbody>
</table>

| Means | \( \tau_1. \) | \( \tau_2. \) | \ldots | \( \tau_j. \) | \ldots | \( \tau_p \) | \( \tau_. \) |

We sample Table 4.3.1 by selecting \( p \) columns from the table at random with equal probability and without replacement. Denote the sample data by \( \{y_{iu}; i = 1,2,\ldots,v; u = 1,2,\ldots,p\} \) in which column \( u \) is the \( u \)-th selected column. We find the expectations under repetitions of the sampling.

Then,

\[ E(y_{iu}) = \tau_i. \]
To simplify notation, write
\[ \sum_j \tau_{ij}^2 = \left( \sum_j (\tau_{ij} - \tau_i) \right)^2 + \tau_i^2. \]
\[ = (p-1)\sigma_{ii} + \tau_i^2. \]
\[ \sum_j \tau_{ij} \tau_{i'j'} = (\sum_j \tau_{ij})^2 - \sum_j \tau_{ij}^2 \]
\[ = p(p-1)\tau_i^2 - (p-1)\sigma_{ii}. \]
\[ \sum_j \tau_{ij} \tau_{i'j'} = \sum_j (\tau_{ij} - \tau_i) (\tau_{i'j} - \tau_{i'}). \]
\[ = (p-1)\sigma_{ii}, + \tau_i \tau_{i'}. \]
\[ \sum_j \tau_{ij} \tau_{i'j'} = \sum_j \tau_{ij} (\tau_{i'j} - \tau_{i'}). \]
\[ = p(p-1)\tau_i \tau_{i'}. - (p-1)\sigma_{ii}. \]

So,
\[ E(y_{iu}) = \tau_i. \]
\[ E(y_{iu}^2) = \tau_i^2 + \frac{(p-1)}{p} \sigma_{ii}. \]
\[ E(y_{iu} y_{iu'}) = \tau_i^2 - \frac{1}{p} \sigma_{ii}, u \neq u'. \]
\[ E(y_{iu} y_{i'u'}) = \tau_i \tau_{i'}, + \frac{(p-1)}{p} \sigma_{ii}, , i \neq i'. \]
\[ E(y_{iu} y_{i'u'}) = \tau_i \tau_{i'}, - \frac{1}{p} \sigma_{ii}, , i \neq i', u \neq u'. \]

These expectations are similar to those obtained by Scheffe (1959). However,
ever, the method used to derive them is different.

Below is given the table of the sum of squares of the \( v \times p \) array of observations. Let \( C = vpy^2 \), where, now in this section, replacement of a subscript by a dot indicates that averaging has been done.

Table 4.3.2. The sample analysis of variance

<table>
<thead>
<tr>
<th>Source</th>
<th>S.S.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Varieties</td>
<td>( \sum_{i} \gamma_i^2 ) - C</td>
</tr>
<tr>
<td>Place</td>
<td>( \sum_{u} \gamma_u^2 ) - C</td>
</tr>
<tr>
<td>Varieties x Places</td>
<td>( \sum_{iu} (\gamma_{iu} - \gamma_i - \gamma_u + \gamma..)^2 )</td>
</tr>
<tr>
<td>Total</td>
<td>( \sum_{iu} \gamma_{iu}^2 ) - C</td>
</tr>
</tbody>
</table>

The expectations of these sum of squares can be worked out.

\[
E(C) = E\left[ \frac{1}{vp} \sum_{iu} \gamma_{iu}^2 \right]
= vpr^2 + \frac{1}{v} \left( \frac{p-1}{p} \right) \sigma_{i.\iota}^2
\]

Note that the last term is summed over all \( i \) and \( i' \).

\[
E(py_{i.}^2) = E\left[ \frac{1}{p} \sum_{u} \gamma_{iu}^2 \right]
= pr^2_{i.} + \left( \frac{p-1}{p} \right) \sigma_{i.}^2
\]

\[
E(vy_{..u}^2) = E\left[ \frac{1}{v} \sum_{i} \gamma_{iu}^2 \right]
= vr^2_{..u} + \frac{1}{v} \left( \frac{p-1}{p} \right) \sum_{i,\iota} \sigma_{i.\iota}^2
\]

Then we have,
\[
E(\text{Variety SS}) = p\Sigma_{i} \tau_{ii}^{2} - v\tau_{..}^{2} + \frac{(p-P)}{P} (\Sigma_{i} \sigma_{ii} - \frac{1}{v} \Sigma_{i,i',} \sigma_{ii'})
\]
\[
= p\Sigma(\tau_{ii}^{2} - \tau_{..}^{2}) + \frac{(p-P)}{P} (\Sigma_{i} \sigma_{ii} - \frac{1}{v} \Sigma_{i,i',} \sigma_{ii'})
\]
\[
E(\text{Places SS}) = \frac{(p-1)}{v} (\Sigma_{i,i',} \sigma_{ii'})
\]
\[
E(\text{Total SS}) = E(\Sigma_{i,i',} y_{ii}^{2} - C)
\]
\[
= p\Sigma(\tau_{ii}^{2} - \tau_{..}^{2}) + \frac{(p-1)p}{P} \Sigma_{i} \sigma_{ii} - \frac{1}{v} \frac{(p-P)}{P} \Sigma_{i,i',} \sigma_{ii'}
\]

Given below are tables with different forms of the expected sum of squares.

Table 4.3.3. Expectation of sum of squares - form 1

<table>
<thead>
<tr>
<th>Source</th>
<th>E(SS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Varieties</td>
<td>(\frac{(p-P)}{P} (\Sigma_{i} \sigma_{ii} - \frac{1}{v} \Sigma_{i,i',} \sigma_{ii'}) + \frac{(p-1)p}{P} \sigma^{2}_v)</td>
</tr>
<tr>
<td>Places</td>
<td>(\frac{(p-1)}{v} \Sigma_{i,i',} \sigma_{ii'})</td>
</tr>
<tr>
<td>Varieties x Places</td>
<td>(\frac{(p-1)}{P} \Sigma_{i} \sigma_{ii} - \frac{1}{v} \frac{(p-P)}{P} \Sigma_{i,i',} \sigma_{ii'}) + (\frac{(p-1)p}{P} \sigma^{2}_v)</td>
</tr>
<tr>
<td>Total</td>
<td>(\frac{(p-P)}{P} \Sigma_{i} \sigma_{ii} - \frac{1}{v} \frac{(p-P)}{P} \Sigma_{i,i',} \sigma_{ii'}) + (\frac{(p-1)p}{P} \sigma^{2}_v)</td>
</tr>
</tbody>
</table>

Table 4.3.4. Expectation of sum of squares - form 2

<table>
<thead>
<tr>
<th>Source</th>
<th>E(SS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Varieties</td>
<td>((v-1)\left[\frac{1}{(v-1)} \left{ \Sigma_{i} \sigma_{ii} - \frac{1}{v} \Sigma_{i,i'} \sigma_{ii'} \right}\right] + \frac{(v-1)p}{P} \sigma^{2}<em>v - \frac{1}{(v-1)} \frac{1}{P} \left[\Sigma</em>{i} \sigma_{ii} - \frac{1}{v} \Sigma_{i,i'} \sigma_{ii'} \right])</td>
</tr>
<tr>
<td>Places</td>
<td>((p-1)\left[\frac{1}{v} \Sigma_{j,j'} \sigma_{jj'}\right])</td>
</tr>
<tr>
<td>Varieties x Places</td>
<td>((v-1)(p-1)\left[\frac{1}{(v-1)} \left{ \Sigma_{i} \sigma_{ii} - \frac{1}{v} \Sigma_{i,i'} \sigma_{ii'} \right}\right])</td>
</tr>
</tbody>
</table>
This table is the same as that obtained by Scheffé.

Table 4.3.5. Expectation of sum of squares - form 3

<table>
<thead>
<tr>
<th>Source</th>
<th>E(S,S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Varieties</td>
<td>((v-1)[\sigma_v^2 + p(\sigma_v^2 - \frac{1}{p} \sigma_{vp}^2)])</td>
</tr>
<tr>
<td>Places</td>
<td>((p-1)\nu \sigma_p^2)</td>
</tr>
<tr>
<td>Varieties x Places</td>
<td>((v-1)(p-1)\sigma_{vp}^2)</td>
</tr>
</tbody>
</table>

where \(\sigma_{vp}^2 = \frac{1}{(v-1)} \left[ \sum_i \sigma_{ii} - \frac{1}{v} \sum_i \sigma_{ii} \right]\), \(\sigma_p^2 = \frac{1}{v} \sum_i \sigma_{ii}\),

and \(\sigma_v^2 = \frac{1}{v-1} \sum_i (\tau_{i.} - \tau_{..})^2\).

It is natural to make these definitions, because it can be shown that

\[
\frac{1}{v} \sum_i \sigma_{ii} = \frac{1}{v} \left( \sum_i \sigma_{ii} + \sum_i \sigma_{ii} \right) = v \frac{1}{(p-1)} \sum_j (\tau_{..} - \tau_{..})^2 = v \sigma_p^2,
\]

where \(\sigma_p^2\) is the variance amongst the place means.

Also, it can be shown that

\[
\frac{1}{(v-1)} \left[ \sum_i \sigma_{ii} - \frac{1}{v} \sum_i \sigma_{ii} \right] = \frac{1}{(v-1)(p-1)} \left\{ \sum_{ij} \tau_{ij}^2 - \nu \sum_j \tau_{..j}^2 - p \sum_i \tau_{..i}^2 + p \nu \tau_{..}^2 \right\} = \sigma_{vp}^2,
\]

where \(\sigma_{vp}^2\) is the usual definition for the component of variation for interaction in a complete two-way table.

The derivation of EMS's given above is obtained without the use of
a model. It assumes only that one has Table 4.3.1, with \( P \) some indefinitely large integer, and then that one has sampled by taking \( p \) of the \( P \) columns at random without replacement.

This development then does not use a model. However, the finite sampling approach can be used to obtain a model, on the basis of the sampling. The following describes how the process is put into effect. The idea is to develop, from a pure sampling point of view, properties of the terms in the model equation

\[
y_{iu} = \mu + v_i + \sum_j \delta^u_j p_j + \sum_j \delta^u_j (vp)_{ij}
\]

where

\[
y_{iu} = \mu + v_i + \beta_u + (v\beta)_{iu}.
\]

We now consider the nature of a linear model appropriate to this type of sampling, making no assumptions.

First we write, for the population,

\[
\tau_{..} = \mu,
\]

\[
\tau_i. = \mu + v_i, \quad \sum_i v_i = 0
\]

\[
\tau_.j = \mu + p_j, \quad \sum_j p_j = 0
\]

\[
\tau_{ij} = \mu + v_i + p_j + (vp)_{ij} \quad \text{with} \quad \sum_i (vp)_{ij} = 0, \sum_j (vp)_{ij} = 0.
\]

Also, let \( \delta_j^u = 1 \), if the \( u \)-th sample place is the \( j \)-th in the population; otherwise, \( = 0 \).

Then

\[
y_{iu} = \sum_j \delta^u_j \tau_{ij}
\]

\[
= \mu + v_i + \sum_j \delta^u_j p_j + \sum_j \delta^u_j (vp)_{ij}
\]

This can be written as

\[
y_{iu} = \mu + v_i + \beta_u + (v\beta)_{iu}
\]
and our task is to determine appropriate properties of the terms in the model.

Now, in the sampling framework, $\beta_u$ and $(v\beta)_{iu}$ are random variables.

Also, we find

$$E(\beta_u) = \frac{1}{F} \sum_j p_j = 0,$$
$$E(\beta_u^2) = \frac{1}{F} \sum_j p_j^2,$$
$$E(\beta_u \beta_u^*) = -\frac{1}{F(F-1)} \sum_j p_j^2,$$
$$E[(v\beta)_{iu}] = \frac{1}{F} \sum_j (vp)_{ij} = 0,$$
$$E[(v\beta)_{iu}]^2 = \frac{1}{F} \sum_j (vp)_{ij}^2,$$
$$E[(v\beta)_{iu}(v\beta)_{iu}] = \frac{1}{F} \sum_j (vp)_{ij}(vp)_{ij},$$
$$E[(v\beta)_{iu}(v\beta)_{iu}^*] = -\frac{1}{F(F-1)} \sum_j (vp)_{ij}^2,$$
$$E[\beta_u(v\beta)_{iu}] = \frac{1}{F} \sum_j p_j(vp)_{ij},$$
$$E[\beta_u(v\beta)_{iu}]^* = -\frac{1}{F(F-1)} \sum_j p_j(vp)_{ij}.$$

We consider $F$ to be indefinitely large. So, we can replace $\frac{1}{F}$ by 0.

It then seems reasonable to take the following as the first and second moments of the model terms.

$$E(\beta_u^2) = \frac{1}{F} \sum_j p_j^2 = \sigma_u^2, \text{ say,}$$
$$E(\beta_u \beta_u^*) = 0,$$
$$E[(v\beta)_{iu}] = \frac{1}{F} \sum_j (vp)_{ij} = \sigma_{iii}, \text{ say (I for interaction),}$$
$$E[(v\beta)_{iu}(v\beta)_{iu}] = 0,$$
$$E[(v\beta)_{iu}(v\beta)_{iu}^*] = \frac{1}{F} \sum_j (vp)_{ij}(vp)_{ij} = \sigma_{iii},$$
The reasonableness of these assumptions should be examined in any given set of data. The data of Yates and Cochran (1938) suggest that even these assumptions are not reasonable in some cases.

These formulas are interesting because there is no reason why \( \Sigma p_j(v_{ij})_j \) should be zero, or \( \Sigma (v_{ij})_j (v_{ij})_j \) should be zero. These formulas are considered important as there is no a priori reason why any of the above should be very small or zero.

These results are interesting because they do not lead to the conventional assumptions for the error components of the linear model

\[
y_{iu} = \mu + v_i + \beta_u + (v_{
\beta})_{iu},
\]

where \( \{\beta_u\} \) and \( \{(v_{
\beta})_{iu}\} \) are random. Hence, we have the following proposition. Under a sampling model, in the absence of technical errors, for \( P \) indefinitely large, and with homogeneity assumptions, a reasonable general linear model is

\[
y_{iu} = \mu + v_i + \beta_u + (v_{
\beta})_{iu},
\]

where \( \mu \) is a constant, \( \{v_i\} \) are constants with \( \Sigma v_i = 0 \),

\[
\beta_u \sim \text{IN}(0, \sigma^2_p),
\]

\[
(v_{
\beta})_{iu} \sim \text{N}(0, \sigma^2_{vp}),
\]

\[
\text{Cov}[(v_{
\beta})_{iu}, (v_{
\beta})_{iu}] = \rho_{ii} \sigma^2_v,
\]

\[
\text{Cov}[(\beta_u, (v_{
\beta})_{iu}] = \rho_{ii} \sigma_p \sigma_{vp}, -1 < v_i < 1.
\]
All other covariances are zero. This model will give the same expected sum of squares as in Table 4.3.5.

1) Our model is similar to that of Scheffé, but does not appear to be the same. We have the following relationships between the two models:

\[
\begin{align*}
\sigma_p^2 &= \sigma_B^2 = \sigma_{..}, \\
\sigma_{vp}^2 &= \sigma_{AB}^2 = \frac{1}{(I-1)} \sum (\sigma_{ii} - \sigma_{..}), \\
\rho_{ii'} &= \frac{(v-1)[\sigma_{ii'} - \sigma_i - \sigma_{..} + \sigma_{..}]}{\sum (\sigma_{ii} - \sigma_{..})}, \\
\nu_i &= \sqrt{\frac{\sigma_i^2 - \sigma_{..}}{\sigma_{..} (v-1) \sum (\sigma_{ii} - \sigma_{..})}}
\end{align*}
\]

We may obtain simpler symmetric models by supposing other homogeneity conditions. So next, we consider the relationship of this model to other simpler models given by other authors.

2) If in our model, we write \( \rho = \frac{-\lambda}{1-\lambda} \), for \( i \neq i' \), and put

\[
\nu_i = 0, \forall i, \text{ then we get the model given by Harville (1978).}
\]

\[
\begin{align*}
\beta_u &\sim N(0, \sigma_p^2), \\
(v \beta)_{iu} &\sim N(0, \sigma_{vp}^2), \\
\text{Cov}[(v \beta)_{iu}, (v \beta)_{i'u}] &= \frac{-\lambda}{1-\lambda} \sigma_{vp}^2,
\end{align*}
\]

with \( \beta_u \) and \( (v \beta)_{iu} \) uncorrelated.

A special case of this model, given by Graybill (1961), is obtained when we put \( \lambda = \frac{1}{t} \).

3) If we put \( \rho_{ii'} = 0, \forall i \neq i' \), \( \nu_i = 0, \forall i \), then we get the very
simple model that is often found in the literature. Here, all the random quantities are uncorrelated with each other. This model is given in Searle (1971), Mood (1950), and others.

Hocking (1973):

Hocking discusses the three different mixed models that are commonly used. He then gives the inter-relations between these models. The three models considered by him are:

Model 1: \( y_{ijk} = \mu + \alpha_i + b_j + c_{ij} + e_{ijk} \),

where

1. \( \mu \) and \( \alpha_i, i=1,2,\ldots,a \), are parameters with \( \sum \alpha_i = 0 \).
2. \( b_j, c_{ij} \) and \( e_{ijk} \) are random variables with zero means, and covariance structure described in terms of the axa positive definite, symmetric matrix \( \Sigma = (\sigma_{ij}) \), as given below:
   \[
   \begin{align*}
   \text{var}(e_{ijk}) &= \sigma^2, \\
   \text{var}(b_j) &= \sigma_{..}, \\
   \text{cov}(c_{ij}, c_{i'j}) &= \sigma_{ii} - \sigma_{i..} + \sigma_{..}, \\
   \text{cov}(b_j, c_{ij}) &= \sigma_{i..} - \sigma_{..}.
   \end{align*}
   \]
   All other covariances are zero.
3. \( c_{ij} = 0, j=1,\ldots,b \).
4. The variance components are defined as:
   \[
   \begin{align*}
   \sigma^2_B &= \frac{\sigma_{..}}{a}, \\
   \sigma^2_{AB} &= \frac{\sum_i (\sigma_{ii} - \sigma_{..})}{(a-1)}.
   \end{align*}
   \]
   This is Scheffé's model.

Model 2: \( y_{ijk} = \mu + \xi_i + \beta_j + (\xi \beta)_{ij} + e_{ijk} \).
where

1. $\mu$ and $\xi_i$, $i=1,...,a$, are parameters.

2. $\beta_j$, $(\xi\beta)_{ij}$, and $e_{ijk}$ are uncorrelated random variables with zero means and

   - $\text{var}(e_{ijk}) = \sigma^2$
   - $\text{var}(\beta_j) = \sigma^2_{\beta}$
   - $\text{var}((\xi\beta)_{ij}) = \sigma^2_{\xi\beta}$

3. The variance components are defined to be $\sigma^2_{\beta}$ and $\sigma^2_{\xi\beta}$.

Model 2 is a special case of Model 1 and Hocking gives the relation between them as:

1. $\mu + \alpha_i = \mu + \xi_i$

2. $\sigma_{ii} = \sigma^2_{\beta} + \sigma^2_{\xi\beta}$

3. $\sigma_{ii'} = \sigma^2_{\beta}$ ($i \neq i'$)

4. $\sigma^2_B = \sigma^2_{\beta} + \frac{1}{a} \sigma^2_{\xi\beta}$

5. $\sigma^2_{AB} = \sigma^2_{\xi\beta}$

6. $b_j = \beta_j + (\xi\beta)_{.,j}$

7. $c_{ij} = (\xi\beta)_{ij} - (\xi\beta)_{.,j}$

Model 3: $y_{ijk} = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + e_{ijk}$

where

1. $\mu$ and $\alpha_i$, $i=1,...,a$, are parameters.

2. $\gamma_j$, $(\alpha\gamma)_{ij}$, and $e_{ijk}$ are random variables with zero means and the following covariance structure:

   $\text{var}(e_{ijk}) = \sigma^2$
\[ \text{var}(\gamma_j) = \sigma^2_\gamma, \]
\[ \text{var}(\alpha\gamma)_{ij} = \frac{(a-1)}{a} \sigma^2_{\alpha\gamma}, \]
\[ \text{cov}(\alpha\gamma)_{ij}, (\alpha\gamma)_{i'j'} = -\frac{1}{a} \sigma^2_{\alpha\gamma}, \text{ if } i \neq i', \]
all other covariances are zero.

3. \((\alpha\gamma)_{ij} = 0.\)

4. The variance components are \(\sigma^2_\alpha\) and \(\sigma^2_{\alpha\gamma}\).

Model 3 is also a special case of Model 1 and Hocking gives the relations between them as:

1. \(b_j = \gamma_j, c_{ij} = (\alpha\gamma)_{ij}\)
2. \(\sigma_{ii} = \sigma^2_\gamma + \frac{(a-1)}{a} \sigma^2_{\alpha\gamma}\)
3. \(\sigma_{ii'} = \sigma^2_\gamma - \frac{1}{a} \sigma^2_{\alpha\gamma}\)
4. \(\sigma^2_B = \sigma^2_\gamma\)
5. \(\sigma^2_{AB} = \sigma^2_{\alpha\gamma}\)

Models 2 and 3 are shown to have the following relations:

1. \(\gamma_j = \beta_j + (\xi\beta)_{.j}\)
2. \((\alpha\gamma)_{ij} = (\xi\beta)_{ij} + (\xi\beta)_{.j}\)
3. \(\sigma^2_\gamma = \sigma^2_B + \frac{1}{a} \sigma^2_{\xi\beta}\)
4. \(\sigma^2_{\alpha\gamma} = \sigma^2_{\xi\beta}\)

Hocking discusses these three models for the two-way mixed, analysis of variance model. By reducing the description of the models to a statement of the first two moments on the observations, he obtains the relations given above. Then he leaves it to the researcher to choose the appropriate model for whatever he has in mind.
Discussion:

Scheffé (1959) begins with a true value and he defines the main effects and interaction in terms of the true values. He then writes the observed value as the sum of the true value and a random error. The restrictions on the parameters follow naturally. He adds the assumptions of normality and equality of cell variances. In the case of the random model, the random effects are all uncorrelated. In the mixed model, some of them are correlated. The normality assumption enables him to use the F-test.

So, for the fixed model, the error MS is used as the denominator for the F-statistic for all the factors. For the random model, the interaction MSS is used as the denominator of the F-statistic for all the factors. For the mixed model, the interaction MSS is used in the denominator to test the fixed factor and the error MSS for the random factor.

The Wilk and Kempthorne (1955) approach to the mixed model, is to consider it as the limiting case of a general sampling model. This gives the same rules of what MS are proper error terms, as those of Scheffé (1959). Again, as mentioned before, the covariance structure is considered only under random relabelling of the levels of the fixed factor. In the mixed model case, one is not interested in the random relabelling.

Wilk and Kempthorne gave Σ-quantities, which are the same as those developed later from a different viewpoint by Cox (1958).

According to Wilk and Kempthorne, all the parameters representing the effects and interactions are fixed unknown quantities. The random variables in the statistical model are the dummy variables introduced. They do not make any assumption of normality or independence of the random variables. From the design of the experiment, some distributional
properties of the dummy variables are obtained. Wilk and Kempthorne have just one derivation of the ANOVA table from the model. One can get the EMS for the different models by assuming one has infinite levels of the random factors or that one takes all levels (finite in number) in the case of the fixed factors.

Searle (1971) does not give a basis for his model as do Scheffé and Wilk and Kempthorne. He gives a model equation where the terms in the model are defined as effects (i.e., he gives the observation in terms of effects and a random error). He makes the assumption that all the random effects have zero means and are uncorrelated with each other. There are no restrictions on the parameters. When he does mention restrictions, he gives no reason why such restrictions may be imposed on the parameters. For fixed and random models, his ANOVA table is the same as Scheffé's and Wilk and Kempthorne. However, for the mixed model, for both the random factor and the fixed factor, he used the interaction MS in the denominator of the F-statistic. He also makes the normality assumption.

Like Wilk and Kempthorne, Nelder starts with an identity and has a population model and a sample one. According to him, the effects themselves are not as important as the canonical components. The parameters in the model are not random or fixed but what he calls random in a formal sense. They are all assumed to be uncorrelated.

He too has one derivation that leads to the ANOVA table. He too, like Wilk and Kempthorne gets three cases according to whether the number of levels in the population, is infinite or whether all the finite levels have been sampled. He does not use any restrictions on the parameters. His EMS are not in terms of the variance components but in terms of the
canonical components. He gets the same ANOVA table (i.e., EMS) for all three cases. The interaction is always to be used in the denominator of the F-statistic for any effect. Nelder does not make the normality assumption but assumes that the population has a scale-and-location type distribution. The canonical components obtained by Nelder are the same as the Σ-quantities obtained Wilk and Kempthorne and Cox, much earlier.

Nelder considers the effects A, B to be marginal to the interaction A·B. He considers that if A if fitted after A·B, then there is no information in A and hence, it is useless to test for A after elimination of A·B. Here, Nelder is attempting to say that with a model

\[ y = I + X_a a + X_b b + X_{ab} (ab) + e, \]

in which \(X_a, X_b,\) and \(X_{ab}\) are the incidence matrices for factors A, B and the interaction A·B, then \(C(X_{ab}) \subset C(X_a).\) Hence, after fitting for A·B, there is no sum of squares for the factor A. However, if conditions on interaction terms of the fixed model are introduced, then this is no longer true. He contends models in which A·B is postulated to exist whose marginal effects A or B are null, are of no practical interest.

On this point, argument can be made. There may, for example, be no variety effects, with the presence of variety place interactions. Also, we may be interested in whether there are place differences, and their magnitudes with all varieties grown at each place (conceptually, of course).

Scheffé says that if such a situation occurs then the correct conclusion is not that there are no differences, but that there are differences. However, when the effects of the levels of one factor are
averaged over the levels of the other, no difference of these averaged effects has been demonstrated, and this conforms to what we said in the previous paragraph.

Nelder equates fixed and random effects not with finite and infinite populations, respectively, but with complete and incomplete sampling and this framework includes randomization models. Nelder criticizes the nomenclature. The fixed effects model has actually a random quantity in it, namely, the error term. Also, in the random model, there is no distinction made between the random effects and the random error. He also complains about the constraints as he considers they have no basis and that they lead to unrealistic hypotheses. We sometimes get negative values for the variance components. He claims that they are not variances, but in fact, excesses of variance in the margins. Hence, excess can be negative. This is, of course, the potential negativity of the \( \Sigma \) quantities of Wilk and Kempthorne.

In the mixed model, Scheffé expressly mentions that one should not make the assumption

\[
\sigma_{ii'} = \rho \sigma^2 \quad \text{if } i \neq i' \\
= \sigma^2 \quad \text{if } i = i'
\]

because in most situations such symmetry does not exist. Nelder, however, makes this assumption about the covariance, but gives no justification.

Wilk and Kempthorne are of the opinion that the assumed linear model is not a causal or functional relationship, the normality assumption is nearly always false and the independence assumptions often bear no relationship to the physical situation.

The formulations of Scheffé (1959) are somewhat similar to that of
Wilk and Kempthorne (1955) and lead to the same EMS's. The reason for this is that each formulation starts with an observation being considered as the sum of a true value and an error. They define the effects and interactions in terms of the true values in a similar way. In both cases, the least squares estimates of these effects are obtained by replacing the true values in the definition by the sample values actually obtained. Hence, when expectation is taken of the MS, Scheffe and Wilk and Kempthorne get the same values. The restrictions $\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$ are consequences of the way the models are developed and are not arbitrary. In both formulations, the levels of the random factor used in the experiment are considered as a random sample from an infinite population of levels, while the levels of the fixed factor used in the experiment are considered to constitute all the possible levels (finite in number) of the factor.

Searle's results are different from those of Scheffe and Wilk and Kempthorne. The main cause for Searle getting different EMS is that Searle does not have the restrictions $\sum_j \alpha_i = 0$ etc. as an intrinsic part of the model, nor does he separately impose them. Hence, his EMS for the fixed and mixed models are different from those of Scheffe and Wilk and Kempthorne. In the case of the random model, they are the same because Searle makes the assumption that the effects have zero expectation.

The difference appears to be that Searle gives a model equation and then adds assumptions about the model terms that (presumably) he thought reasonable. If one thinks of the data set as arising by sampling a $v \times \infty$ table, then the first and second moment properties of model terms are derivable and given on page 98. We see that from this point of view,
the assumptions Searle uses on model terms, do not have a basis in sampling theory.

Nelder starts with a true value and gives an identity involving it. He does not define the terms in the identity as effects or interactions etc., as do Scheffe, and Wilk and Kempthorne.

The part where he differs from the latter three authors is in describing the covariance structure of the true values. Wilk and Kempthorne get the covariances from the randomization procedure. Scheffe gives a general variance-covariance matrix which is based on the example from which he derives the whole 2-way model. Nelder gives a simplified type of covariance structure. In his paper, he gives no justification for it. Scheffe is, in fact, against this simplification as he feels there is no justification for it in most cases. Nelder gives a model equation where the parameters are not directly equated with the terms of the identity. Scheffe, and Wilk and Kempthorne, on the contrary, equate the parameters with terms in the population identity and define them as effects, interactions, etc.

Nelder assumes that these parameters, which he considers random variables are uncorrelated. According to Wilk and Kempthorne, they are fixed unknown quantities. According to Scheffe, in the case of random effects, they are random variables, otherwise they are fixed, unknown quantities. Scheffe does not assume that the random effects are uncorrelated with each other.

The question of what assumptions on model terms are reasonable in terms of a reasonable sampling process is addressed earlier. From this mode of development, it is clear that the "usual" assumptions of zero covariance, apart from an obscure condition that sums of interactions
components are zero, is not reasonable.

4.4. Conclusion

The question of what statistical procedures to follow with the mixed 2-way linear model is unresolved, in our opinion. There can be, it seems, no question about the nature of expected values of mean squares, which conform to the structure given by Wilk and Kempthorne (1955). We may note, that this structure was hinted at in an unpublished note by Tukey written in 1949.

The question of what model to use, what homogeneity assumptions are appropriate and what assumptions about correlations should be made -- should presumably be addressed by data analysis and data plotting as Yates and Cochran did. However, development of formal tests of appropriateness of a covariance structure appears to be a very difficult problem, that we do not address.

Even if we adopt analysis of variance, the question of what tests are appropriate is quite unresolved. It is natural to form ratios of mean squares, but there is, it seems, no theoretical basis for this except under a model that is unrealistically simple.

The question of what assumptions are reasonable may be thought about in the light of the deliberations and data of Yates and Cochran (1938). They had a variety-place example. They plotted for each variety the means at the various places against the overall place means. If a simple linear model with uncorrelated components were appropriate, these plots should exhibit, approximately, parallel lines. This is not at all what was found.
It is our view that there is need for development of some technique of multivariate analysis. Each random place gives a vector of the variety means. So we have, for example, p vectors, each of v elements, and a proper statistical procedure would take this as a starting point. It is not at all clear that such a multivariate procedure would lead to the conventional analysis of variance. This is not meant, of course, to contravene the utility of the usual analysis of variance as a data analysis technique for forming good ideas of the amount of variability associated with the different sources of variation.

We regret that we have been unable to make any significant advance in this direction.


Pitman, E. J. G. 1938. Significance tests which may be applied to samples from any population. III. The analysis of variance test. Biometrika 29:322-335.


6. ACKNOWLEDGEMENTS

First and foremost, I'd like to thank Prof. O. Kempthorne for the tremendous amount of help and time he has given me. A simple thank you seems totally inadequate.

I wish to thank, also, all my friends in the United States, especially Dwaraka Rao, Aasha Kapur, Sharon Shepard and Frances Bradley for all the encouragement they gave me, and for listening patiently to my discomforts.

I owe a big debt of gratitude to Rebecca Shivvers for typing my thesis, and taking care of all the errands connected with it.

Lastly, I would like to thank my family for their sustained moral support at all times.