UNIVERSITY OF CAPE TOWN

DEPARTMENT OF MATHEMATICAL STATISTICS

APPLICATIONS OF ANALYSIS OF VARIANCE IN WOOL MARKETING

by

J.J.J. Du Plessis

A thesis prepared under the supervision of Professor C.G. Troskie in partial fulfilment of the requirements for the degree of Master of Science in Mathematical Statistics.

TO FAAN and DITA

A C K N O W L E D G E M E N T S

I am heavily indebted to certain individuals who have influenced the writing of this thesis. I should like to mark my special debt to PROFESSOR TROSKIE for his guidance and encouragement which made this thesis the reality that it is today and for introducing me to the practical application of statistics.

I also wish to thank MR.H.IRELAND, technical liason officer of the S.A. Wool Commission for his assistance in interpreting some of the results and arriving at sound and meaningful conclusions. My indebtness goes also to MR.A.VAN WYK for his efficiency in proofreading and to the S.A. Wool Commission by whom I am employed for being the source of the data used and which forms the basis of this thesis.

It is pleasure to acknowledge the patience and skill with which my wife DITA typed this thesis. This task necessitated numerous hours of night work and the loss of many weekends. I also owe great debt to her for her help and encouragement and her patience during the many evenings I worked on this thesis. Had it not been for her, I probably would not have been able to attend university.

Finally, I must also place on record the excellent co-operation of MRS.M.I.(Tib) COUSINS for the various formalities which she took to task after I have left the university of Cape Town.

Port Elizabeth, South Africa

March 1971 J. du P.

P R E F A C E

Analysis of variance could be described as a statistical technique for analysing measurements depending on several kinds of effects operating simultaneously so as to decide which kinds of effects are important and to estimate the effects. Although probably not susceptible of a very precise definition, it in general consists of a body of tests of hypotheses and methods of estimation using statistics which are linear combinations of sums of squares of linear functions of the observed values. Having been developed mainly in connection with problems of agricultural experimentation, the application thereof in the South African Wool Trade seems non existent. I hope that this thesis will illustrate some of the very useful applications, especially to the extent where the rejection of all (or some) of the hypotheses under consideration is in itself as significant as the acceptance thereof would have been.

As one of the aims of science is to describe and predict events in the world in which we live, one way in which this can be accomplished is by finding a formula or equation that relates quantities in the real world. Although functional relationships are assumed to hold in many fields of science, such as physics, there are many scientific areas such as biology, economics etc., where relationships are more obscure. For example, the price of wool on a given sale cannot be pre-

dicted accurately. Many of the factors effecting this price are known, but the equation relating these quantities is obscure. It could be known that type of wool X_1, yield X_2, demand X_3, supply X_4, and many other factors influence the price Y. Although all the factors that effect this price are not known and although the relationship is not known, it is useful nevertheless to assume that there exists a finite number of factors $X_1, \ldots, X_n$ and a function g such that the price Y can be exactly determined by

$$Y = g(X_1, \ldots, X_n)$$

Based on this assumption, amongst other things, various linear mathematical models have been developed to aid the prediction of real world events. By a linear model we shall mean an equation that involves random variables, mathematical variables, and parameters that is linear in the parameters and in the random variables. Following Scheffé (21), these models can best be described as follows: Suppose we have n observations or measurements. It is assumed that these observations are values taken on by n random variables $Y_1, Y_2, \ldots, Y_n$ which are constituted of linear combinations of p unknown quantities $\beta_1, \beta_2, \ldots, \beta_p$ plus errors $e_1, e_2, \ldots, e_n$,

$$Y_i = x_{1i}\beta_1 + x_{2i}\beta_2 + \ldots + x_{pi}\beta_p + e_i \qquad (i = 1, 2, \ldots, n) \tag{1}$$

where the $\{x_{ji}\}$ are known constant coefficients. The $\{\beta_j\}$ are more or less idealized formulations of some aspects of

interest to the investigator in the phenomena underlying the observations, while the $\{e_i\}$ are unobservable random variables about which we assume having zero expected values.

A more precise definition is now appropriate. The analysis of variance is a body of statistical methods of analyzing measurements of the structure (1) where the coefficients $\{x_{ji}\}$ are integers, either 0 or 1. In the analysis of variance the $\{x_{ji}\}$ are the values of "counter variables" which refer to the presence of absence of the effects $\{\beta_j\}$ in the conditions under which the observations are taken. If the $\{x_{ji}\}$ are values taken on by continious variables, called "concomitant variables", we have a case of regression analysis.

If there are $\{x_{ji}\}$ of both kinds, we have an analysis of covariance.

Although the boundaries between the three kinds are not very sharp or universally agreed to, the following destinction can be made. In the analysis of variance all factors are treated qualitatively, in regression analysis all factors are quantitative and treated quantitatively whereas in the analysis of covariance some factors are present that are treated qualitatively and some that are treated quantitatively.

The nature of the unknown effects $\{\beta_j\}$ needs further speci-

fication. They may be unknown constants which we then call parameters, or unobservable random variables subject to further assumptions about their distribution involving other unknown parameters. A model in which all the $\{\beta_j\}$ are unknown constants is called a fixed-effect model. It often happens that one of the $\{\beta_j\}$ is a constant which occurs with every observation with coefficient 1 so that for this j, $x_{ji} = 1$ for all i. Such a constant is called an additive constant (usually a "general mean" in some sense). A model in which all the $\{\beta_j\}$ are random variables, except possible for one which is an additive constant, is called a random-effects model. Intermediate cases where at least one β_j is a random variable and at least one is a constant (not an additive constant) are called mixed models. These clasifications are due to Scheffé (21).

More specifically it is the purpose of this thesis to apply some of the various special cases of the analysis of variance model to wool marketing endevouring in this way to elliminate some of the components of variance in this complex structure of price formation. In particular it is the purpose of this thesis to establish the effects of lot size of wool for selected types when offered at auction on the price of wool and to compare the results with a similar study undertaken in Australia by Whan (53). Due to the "observational" nature of the data, most of the models will be restricted to the fixed

effect type.

For some of the models, basic assumptions will be tested and multiple comparisons made where applicable. To make this thesis as self-contained as possible, the general theory underlying the analysis of variance is included as Chapter I.

As this thesis is being written further provisional research results are coming to hand, but being provisional, it would be premature to report on them at this stage, though, to some extent, these provisional conclusions might colour some of the discussions which follows. A lot more work in this field is required and will be done, and for this reason many of the conclusions could be regarded as a report of work in progress.

J. du P.

C O N T E N T S

CHAPTER I. GENERAL THEORY

(1.1)	Introduction and Notation	I.1
(1.2)	Least-Squares Estimates and Normal Equations	I.2
(1.3)	Estimable Functions and the Gauss-Markoff theorem	I.5
(1.4)	The Canonical Form of the Underlying Assumptions Ω	I.12
(1.5)	Distribution of Estimates Ψ under Ω	I.15
(1.6)	Test of Hypothesis H derived from the Likelihood Ratio	I.17
(1.7)	Canonical form of Ω and H. Distribution of F under Ω	I.20

CHAPTER II. THE ONE-WAY LAYOUT: EQUAL CELL NUMBERS

(2.1)	Method of Analysis	II.1
(2.2)	Description of Data	II.4
(2.3)	Results	II.7
(2.4)	Further Analysis (Multiple Comparisons)	II.10
(2.5)	Conclusions	II.18

CHAPTER III. THE ONE-WAY LAYOUT: UNEQUAL CELL NUMBERS

(3.1)	Method of Analysis	III.1
(3.2)	Description of Data	III.3
(3.3)	Results	III.5
(3.4)	Further Analysis (Multiple Comparisons)	III.8
(3.5)	Conclusions	III.9

CHAPTER IV. THE TWO-WAY LAYOUT: ONE OBSERVATION PER CELL NO: INTERACTION

(4.1)	Method of Analysis	IV.1
(4.2)	Description of Data	IV.6
(4.3)	Results	IV.7
(4.4)	Further Analysis (Multiple Comparisons)	IV.10
(4.5)	Conclusions	IV.13

CHAPTER V. THE TWO-WAY LAYOUT: ONE OBSERVATION PER CELL: INTERACTION

(5.1)	Method of Analysis	V.1
(5.2)	Description of Data	V.7
(5.3)	Results	V.7
(5.4)	Further Analysis (Multiple Comparisons)	V.8
(5.5)	Conclusions	V.8

CHAPTER VI. THE TWO-WAY LAYOUT: EQUAL CELL NUMBERS WITH INTERACTION

(6.1) Method of Analysis VI.1
(6.2) Description of Data VI.2
(6.3) Results VI.9
(6.4) Further Analysis (Multiple Comparisons) VI.11
(6.5) Conclusions VI.12

CHAPTER VII. THE TWO-WAY LAYOUT: UNEQUAL CELL NUMBERS AND A TEST FOR INTERACTION

(7.1) Method of Analysis VII.1
(7.1.1) Test for Interaction VII.4
(7.1.2) Inferences about main effects assuming additivity VII.9
(7.1.3) Test for Main effect under Ω VII.12
(7.2) Description of Data VII.12
(7.3) Results VII.15
(7.4) Conclusions VII.31

CHAPTER VIII. GENERAL

APPENDIX I A.I

Chapter I

General Theory

(1.1) Introduction and Notation.

The development of the general theory of this chapter is greatly facilitated by the use of vector and matrix algebra. The structure (1) of chapter I can be written as

$$Y = X' \beta + e \tag{1.1}$$

where

$$Y^{nx1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \beta^{px1} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \quad e^{nx1} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix},$$

$$X^{pxn} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & & & \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix}$$

and X' is the transpose of X and the superscripts rxs on a matrix denotes that the matrix has r rows and s columns.

A minimal assumption which is always made on the random variables $\{e_i\}$ is that their expected values are zero:

$$E(e) = 0 \tag{1.2}$$

Further we shall always assume that

$$E(ee') = \sigma^2 I \tag{1.3}$$

which is equivalent to saying that the $\{e_i\}$ are uncorrelated and have equal variance σ^2.

(1.2) Least-Squares Estimates and Normal Equations.

We let Ω be the set of underlying assumptions

$$\Omega : Y^{nx1} = X' \beta^{px1} + e^{nx1}$$

$$E(e) = 0$$

$$E(ee') = \sigma^2 I \text{ or more briefly}$$

$$\Omega : E(Y) = X' \beta, \; \mathcal{Y}_y = \sigma^2 I. \qquad (1.4)$$

To estimate β , we let $b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}$

be any quantity which we might consider as an estimate for β and with the natural requirements that b must be some function of the observations and must yield estimates of y close to the observed values.

We define the quantity

$$S(Y,b) = \sum_{i=1}^{n} (y_i \text{ observed} - y_i \text{ estimate})^2$$

$$= (Y-X'b)'(Y-X'b)$$

$$= ||Y - X'b||^2$$

where $||U||$ is the length of vector U.

If b estimates β, then $S(Y,b)$ can be interpreted as:

(1) A measure of how well the model with β estimated by b fits the observations.

(2) $\sum_{i=1}^{n} \hat{e}_i^2$ where $\hat{e}_i$ denotes the estimate of the error e_i in the observation y_i.

Clearly the smaller $S(Y,b)$, the better the fit. The value of b that minimizes $S(Y,b)$ is called the least squares estimate of β, $\hat{\beta}$ say.

DEFN. A set of functions of Y

$\hat{\beta}_1 = \hat{\beta}_1(Y)$, $\hat{\beta}_2 = \hat{\beta}_2(Y)$, ..., $\hat{\beta}_p = \hat{\beta}_p(Y)$

such that the values $b_j = \hat{\beta}_j$ $(j = 1, 2, ..., p)$ minimize $S(Y,b)$, is called a set of least-squares estimates (L.S.E.) of the $\{\beta_j\}$.

To find a vector b that minimizes $S(Y,b)$ we require

$$\partial S(Y,b)/\partial b_t = 0 \quad (t = 1, ..., p)$$

Since

$$\partial S(Y,b) = (y - X'b)'(y - X'b)$$

we have

$$\partial S(Y,b)/\partial b_t = 2XY - 2XX'b = 0$$

Hence $XX'b = XY$ is a set of linear equations in b and are called the NORMAL EQUATIONS (N.E.). With $S=XX'$ we have,

$$Sb = XY \tag{1.5}$$

THM. I Least-squares estimates always exist and are the solutions to the normal equations.

PROOF: We shall first show that the value of b which minimizes $S(Y,b)$ is a function of Y only.

Let: $\rho(X) = r$

Vr = space spanned by the columns of X'

$(\xi_1, \ldots, \xi_p)$ = columns of X'

$Z = X'b$ where we think of varying b

Now $S(Y,b) = ||Y-X'b||^2$

$= ||Y-Z||^2$ which by (7) appendix I, has a minimum when Z is the projection of Y on Vr.

Let $\hat{\eta}$ = projection of Y on Vr

Then

$S(Y,b)$ is a minimum when $Z=\hat{\eta}$

Since

$\hat{\eta} \in Vr$, there exist $\{b_1,\ldots, b_p\}$ such that

$\hat{\eta} = b_1 \xi_1 + \ldots + b_p\xi_p$

Now $\hat{\eta}$ is unique but $\{b_1,\ldots, b_p\}$ is not necessarily unique since $\{\xi_j\}$ may be linearly dependent.

Hence $\hat{\eta}$ is a function of Y only, and not of unknown parameters, $\{b_j\}$ may also be taken as functions of Y only. Further they minimize $S(Y,b)$ so that $\{b_1,\ldots,b_p\}$ are L.S.E. Thus we have shown that L.S.E. always exist.

We shall now show that if $\{b_j\}$ are functions of Y only, then they are L.S.E. if and only if they satisfy the N.E. Let

$X'b = \hat{\eta}$

Then each of the following statements hold if and only if each successive one is true:

$X'b = \hat{\eta}$

$Y-X'b \perp Vr$

$$Y-X'b \perp \xi_j \quad (j=1, \ldots, p) \qquad *$$

$$\xi_j'(Y-X'b) = 0 \quad (j=1, \ldots, p)$$

$$X(Y-X'b) = 0$$

$$XX'b = XY$$

where * follows from (4) appendix I.

Q.E.D.

NOTATION: We shall use $\hat{\beta}$ for the value of b that minimizes S(Y,b). Hence $\hat{\beta}$ is a solution to the NE and $\hat{\beta}$ is also a set of L.S.E.

S_Ω will denote the minimum value of S(Y,b) and is called the error sum of squares.

(1.3) Estimable functions and the Gaus-Markoff Theorem.

DEFN. A parametric function is defined to be a linear function of the unknown parameters $\{\beta_1, \ldots, \beta_p\}$ with known constant coefficients

$$\psi = \sum_{j=1}^{p} c_j \beta_j = c'\beta.$$

DEFN. A parametric function ψ is said to be estimable if it has an unbiased linear estimate, in other words, ψ is estimable if there exists a vector a such that $E(a'Y) = \psi$, identically in β.

THM.II $\psi = c'\beta$ is estimable if and only if c' is a linear combination of the rows of X', i.e., if and only if there exists a vector a^{nx1} such that $c'=a'X'$.

PROOF: $\psi=c'\beta$ is estimable if there exists a vector a^{nx1} such that

$$E(a'Y) = \psi$$

But

$$E(a'Y) = a'E(Y) = a'X'\beta$$

The condition $a'X'\beta=c'\beta$ is satisfied identically in β if and only if $a'X=c'$

Q.E.D.

LEMMA I Let $\psi=c'\beta$ be estimable and Vr the space spanned by the columns of X'. Then:

(1) There exists a unique linear unbiased estimate of ψ, say $a^{*'}Y$, with $a^*\varepsilon Vr$

(2) If $a'Y$ is any unbiased linear estimate, then a^* is the projection of a on Vr.

PROOF: We shall prove (2) first:

Since ψ is estimable there exists an a^{nx1} for which $Ea'Y=\psi$.

Let $a=a^*+b$ where $a^*\varepsilon Vr$ and $b \perp Vr$. Then

$$\psi = E(a'Y) = E\left[(a^*+b)' \right]$$
$$= E(a^{*'}Y) + E(b'Y)$$
$$= E(a^{*'}Y)$$

since $E(b'Y) = b'X'\beta=0$ $(b \perp Vr)$.

Thus $a^{*'}Y$ is an unbiased estimate of ψ with $a^*\varepsilon Vr$.

(1) Uniqueness

Suppose $\alpha'Y$ is also an unbiased estimate of ψ with $\alpha\varepsilon Vr$.

Thus we have

$$0 = E(a^{*'}Y) - E(\alpha'Y)$$

$= (a^*-\alpha)'X'\beta$ identically in β.

Thus

$$(a^*-\alpha)'X'\beta=0$$

Hence either $(a^*-\alpha) \perp Vr$,

or $(a^*-\alpha)=0$

But a^* and α are both in Vr so that

$$(a^*-\alpha)=0$$

and $a^*=\alpha$

Thus $a^{*\prime}Y$ is unique.

THM.III (Gauss-Markoff Theorem-G.M.T.):

Under the assumptions Ω: $E(Y)= X'\beta$, $\mathcal{V}_y = \sigma^2 I$

(1) Every estimable function $\psi=c'\beta$ has a unique unbiased linear estimate ψ which has minimum variance in the class of all unbiased linear estimates.

(2) The estimate ψ can be obtained from $\psi=c'\beta$ by replacing β by any set of L.S. estimates.

PROOF: (1) Let $a^{*\prime}Y$ be the unbiased linear estimate of ψ with $a^* \varepsilon Vr$ (by the above lemma this estimate exists and is unique). Let $a'Y$ be any unbiased linear estimate of ψ, then by the same lemma, a^* is the projection of a on Vr, and

$$||a||^2 = ||a^*||^2 + ||a-a^*||^2$$

$$\begin{aligned} \text{Var}\,(a'Y) = a'\,\mathcal{V}_y a &= \sigma^2\,||a||^2 \\ &= \sigma^2\,||a^*||^2 + \sigma^2\,||a-a^*||^2 \\ &= \text{Var}||a^{*\prime}Y|| + \sigma^2\,||a-a^*||^2 \end{aligned}$$

Hence Var $(a'Y) \geq$ Var $(a^{*\prime}Y)$ with equality only if $a=a^*$. Thus $a^{*\prime}Y$ is the unique unbiased linear estimate of ψ with minimum variance.

(2) It remains to prove that $a^{*\prime}Y=c'\hat{\beta}$ i.e., ψ can be obtained by using L.S.E.

Let $\hat{\eta}=X'\beta$ where $\hat{\eta}$ is the projection of Y on Vr. Then since $(Y-\hat{\eta}) \perp$ Vr we have $a^{*\prime}(Y-\hat{\eta})=0$ because $a^* \varepsilon$ Vr. Hence $a^{*\prime}Y=a^{*\prime}\hat{\eta}$.

Now $c'=a^{*\prime}X'$ since $c'\beta \equiv E(a^{*\prime}Y) \equiv a^{*\prime}X'\beta$ identically in β. Choose β as $\hat{\beta}$ and we have

$$a^{*\prime}Y=a^{*\prime}\hat{\eta}=a^{*\prime}X'\hat{\beta}=c'\hat{\beta}$$

Thus $\hat{\psi}=a^{*\prime}Y=c'\hat{\beta}$ where $\hat{\beta}$ is any set of L.S.E.

Q.E.D.

DEFN. The unique unbiased linear estimate of ψ with minimum variance will be called the least square estimate of ψ.

COROLLARY. If $\{\psi_1, \ldots\ldots \psi_q\}$ are estimable functions, every linear combination

$\psi = \sum_{i=1}^{q} h_i \psi_i$ is estimable and its L.S. estimate is $\hat{\psi} = \sum_{i=1}^{q} h_i \hat{\psi}_i$ where $\hat{\psi}_i$ is the L.S. estimate of ψ_i.

PROOF: Since $\sum_{i=1}^{q} h_i \hat{\psi}_i$ is an unbiased linear estimate of ψ, ψ is estimable. Suppose

$$\psi_i = \sum_{j=1}^{p} c_{ij} \beta_j .$$

Then

$$\psi = \sum_{i=1}^{q} h_i \left(\sum_{j=1}^{p} c_{ij}\beta_j\right)$$

$$= \sum_{j=1}^{p} \left(\sum_{i=1}^{q} h_i c_{ij}\right)\beta_j$$

By G.M.T.

$$\hat{\psi}_i = \sum_{j=1}^{p} c_{ij}\hat{\beta}_j$$

$$\hat{\psi} = \Sigma_j(\Sigma_i h_i c_{ij})\hat{\beta}_j$$

where the $\{\beta_j\}$ are any set of L.S.E.

Hence

$$\hat{\psi} = \Sigma_i h_i \hat{\psi}_i$$

REMARKS: (1) So far we have treated the case where $Y \sim (X'\beta, \sigma^2 I)$ i.e., $\{y_i\}$ are independent and all have equal variances. If however the $\{y_i\}$ are correlated and we know the correlations of all pairs of observations and the ratio of their variances, we have

$Y \sim (X'\beta, \theta B)$, $|B| \neq 0$, $\rho(X') = r$.

This case may be reduced to the case where $\mathcal{V}_y = \sigma^2 I$ by using (9) of appendix I. If we let $\tilde{Y} = P'Y$, then $E(\tilde{Y}) = P'E(Y) = P'X'\beta = \tilde{X}'\beta$ where $\tilde{X}' = P'X'$; so rank $\tilde{X}'$ = rank X' = r and

$\mathcal{V}_{\tilde{y}} = P'\mathcal{V}_y P = \theta P'BP = \sigma^2 I$ where

$\sigma^2 = \theta$. We can thus write

$\Omega : E(Y) = X'\beta$; $\mathcal{V}_y = \theta B$, $\rho(X') = r$

as

Ω : $E(\tilde{Y}) = \tilde{X}'\beta$; $\mathcal{Z}_{\tilde{y}} = \sigma^2 I$, $\rho(\tilde{X}') = r$

which is the same as the case considered before.

(2) Case where $\hat{\beta}$ is unique.

The case where the pxn matrix X is of rank p is called the case of full rank because usually $p < n$. If $\rho(X) = p$, then (1.5) has a unique solution (and only then). From (8) appendix I we have that $\rho(S) = \rho(X)$ so that S is nonsingular. Thus S^{-1} exists and the solution is uniquely given by

$\hat{\beta} = S^{-1}XY$

Further S^{-1} is symmetric since S is so that

$$\mathcal{Z}_{\hat{\beta}} = (S^{-1}X)\mathcal{Z}_y(S^{-1}X)'$$
$$= \sigma^2 S^{-1}XX'S^{-1}$$
$$= \sigma^2 S^{-1}.$$

(3) Case where $\hat{\beta}$ is not unique.

If $\rho(X) < p$, then the L.S. estimate $\hat{\beta}$ is not unique since it consists of any set $\{b_1,\ldots,b_p\}$ satisfying $b_1\xi_1 + \ldots + b_p\xi_p = \hat{\eta}$ where ξ_j is the jth column of X' and $\hat{\eta}$ is the projection of Y on Vr, the space spanned by the $\{\xi_i\}$. A similar indeterminacy affects the parameters $\{\beta_1,\ldots, \beta_p\}$ through the relation $\beta_1\xi_1 + \ldots + \beta_p\xi_p = \eta$

in the sense that different sets of $\{\beta_j\}$ will give the same η and hence the same vector $Y = \eta + e$. To elliminate these indeterminacies two courses are open:

(i) Consider a "reduced" problem with only r parameters $\{\beta_j\}$. This can be achieved by choosing r linearly independent vectors from the set $\{\xi_1, \ldots, \xi_p\}$ as a basis for Vr and keeping only the r corresponding $\{\beta_j\}$. This gives a new nxr matrix of coefficients instead of the old X'. The resulting "reduced" problem is a case of full rank.

(ii) Put suitable side conditions on the p parameters $\{\beta_j\}$ and there estimates. We would achieve the same result as in (i) if we agreed that for the p-r parameters $\{\beta_j\}$ discarded we always take $\beta_j = 0$ and $\hat{\beta}_j = 0$. In most analysis of variance situations it is convenient to add linear restrictions of a more general form than this to produce the desired uniqueness. The $\{\beta_j\}$ are therefore subjected to $t(t \geq p-r)$ linear restrictions $H'\beta=0$, where H' is a txp matrix of known constants. The restrictions make the $\{\beta_j\}$ unique in the sense that for every possible set $\{\beta_j\}$ in the original problem there will

exist a unique set $\{\tilde{\beta}_j\}$ satisfying

$$X'\beta = X'\tilde{\beta} \text{ and } H'\tilde{\beta} = 0. \qquad (1.6)$$

The first of these conditions says the $\{\tilde{\beta}_j\}$ give the same $\eta = X'\beta$ as the $\{\beta_j\}$. The two conditions (1.6) will then make the $\{\tilde{\beta}_j\}$ uniquely determined functions of the $\{\beta_j\}$. That these are estimable functions in the original problem so that every parametric function $c'\tilde{\beta}$ in the new problem is an estimable function in the old problem and that there is then a unique set of L.S. estimates $\{\hat{\beta}_j\}$ which satisfy the side conditions $H'\hat{\beta} = 0$, is proved in (21) chapter I.

(1.4) <u>The Canonical form of the Underlying Assumptions Ω.</u>

Let Vn be the space of the observation vector Y^{nx1} and $\{\rho_1,\ldots, \rho_n\}$ be an orthonormal basis for Y where $\rho_i=(\delta_{i1},\ldots,\delta_{in})'$ so that

$$Y = \sum_{i=1}^{n} y_i\rho_i.$$

Let Vr be the space spanned by the columns of X'.

Let $\{\alpha_1,\alpha_2,\ldots, \alpha_r\}$ be an orthonormal basis for Vr and complete it to an orthonormal basis for Vn:

$$\{\alpha_1,\ldots, \alpha_r, \alpha_{r+1},\ldots, \alpha_n\}.$$

Let $Y = \sum_{i=1}^{n} z_i \alpha_i$ where $\{z_i\}$ are the coordinates of Y relative to the new basis and hence $z_i = \alpha_i' Y$. Thus $Z = PY$ where $P^{n \times n}$ is an orthogonal matrix whose i th row is α_i'.

Let $\xi_i = E(z_i)$, so $\xi_i = E(\alpha_i' Y) = \alpha_i' \eta$ $(\eta = X'\beta)$. But $\eta \varepsilon Vr$ so that $\alpha_i \eta = 0$ for $i>r$, $\alpha_i \perp Vr$.

Hence

$$\xi_i = 0 \text{ for } i>r.$$

If we let $\mathcal{Z}_z$ be the covariance matrix of the transformed "observations" we have

$$\begin{aligned} \mathcal{Z}_z &= P \mathcal{Z}_y P' \\ &= \sigma^2 PP' \\ &= \sigma^2 I \quad (P \text{ is orthogonal}). \end{aligned}$$

It is thus evident that the Ω assumptions $Y = X'\beta$; $\mathcal{Z}_y = \sigma^2 I$ can always be transformed by a suitable orthogonal transformation to the canonical form

$$\Omega: \begin{cases} Z = (z_1, \ldots, z_n)'. \\ E(z_i) = \xi_i \quad (i=1, \ldots, r) \\ E(z_i) = 0 \quad (i=r+1, \ldots, n) \\ \mathcal{Z}_z = \sigma^2 I. \end{cases} \tag{1.7}$$

The canonical form is very useful for derivation of distribution theory: The error sum of squares S_Ω may be written $S_\Omega = ||Y-\hat{\eta}||^2$ where $\hat{\eta}$ is the projection of Y on Vr. But $Y = \sum_1^n z_i \alpha_i$ and $\hat{\eta} = \sum_1^r z_i \alpha_i$ where $\{\alpha_1, \ldots, \alpha_n\}$ is the basis of the canonical form. Hence we have that

$$S_\Omega = \left|\left| \sum_1^n z_i \alpha_i - \sum_1^r z_i \alpha_i \right|\right|^2$$

$$= \left|\left| \sum_{r+1}^{n} z_i \alpha_i \right|\right|^2$$

$$= \sum_{r+1}^{n} z_i^2 \quad \text{since } \alpha_i \text{ are O.N. set.}$$

For $i>r$, $E(z_i) = 0$ which implies $\text{Var}(z_i) = E(z_i) = \sigma^2$

so that $E(S_\Omega) = E\left[\sum_{r+1}^{n} z_i^2\right]$

$$= \sum_{r+1}^{n} E\left[z_i^2\right]$$

$$= (n-r)\ \sigma^2.$$

Defining $s^2 = S_\Omega/(n-r)$ we have that $E(s^2)=\sigma^2$ so that s^2 is an unbiased estimate of σ^2. s^2 is called the mean square for error (MS_e) and is said to have (n-r) degrees of freedom (d.o.f.). The cononical variables $\{z_1,\ldots, z_n\}$ are linear forms in the observations $\{y_i\}$ and they define two orthogonal spaces of linear forms, namely the space spanned by $\{z_1,\ldots, z_r\}$, is called the estimation space, and that spanned by $\{z_{r+1},\ldots, z_n\}$, called the error space. Since $z_i = \alpha_i'Y$, the $\{z_1,\ldots, z_n\}$ constitute an orthonormal basis for the n-dimensional space so that the two spaces are orthogonal. The error sum of squares S_Ω involves only the set $\{z_{r+1},\ldots, z_n\}$. It is easily shown that if ψ is an estimable function and $\hat{\psi}$ is its L.S.E., then the linear form $\hat{\psi}$ is a linear combination of $\{z_1,\ldots, z_r\}$, i.e., ψ is in the estimation space. We will need this result in the next theorem to show that under Ω, $\hat{\Psi}$ and S_Ω/σ^2 are statistically independent.

(1.5) Distribution of estimates $\hat{\Psi}$ under Ω.

If in addition to the underlying assumptions Ω we assume that the observations $\{y_i\}$ have a joint normal distribution, the Ω-assumptions become

$$\Omega: \; Y \sim N(X'\beta, \sigma^2 I), \; \rho(X') = r \qquad (1.8)$$

which permits the derivation of confidence intervals and the test of hypotheses about the parameter values.

Let $\psi_1, \ldots, \psi_q$ be a set of q estimable functions and let $\hat{\psi}_1, \ldots, \hat{\psi}_q$ be their L.S.E. We have that

$$\psi_i = \sum_{j=1}^{p} c_{ij} \beta_j \qquad (i=1, \ldots, q)$$

$$\hat{\psi}_i = \sum_{j=1}^{n} a_{ij} y_j \qquad (i=1, \ldots, q).$$

$\hat{\psi}_i$ can be found by substitution of a solution $\{\hat{\beta}_j\}$ of the N.E., or by finding some linear combination of the observations whose expectation is ψ. In vector notation

$$\psi^{qxl} = C^{qxp} \beta^{pxl}$$

$$\hat{\psi}^{qxl} = A^{qxn} Y^{nxl}$$

and $\mathrm{Cov}(\hat{\psi}) = \Sigma_{\hat{\psi}} = \mathrm{Cov}(AY) = A\Sigma_y A' = \sigma^2 AA'$.

An unbiased estimate of σ^2 is

$$s^2 = S_\Omega/(n-r).$$

We now derive the joint distribution of $\{\hat{\psi}_i\}$ and the error sum of squares S_Ω.

THM. IV Under Ω, (1) $\hat{\Psi} \sim N(\Psi, \Sigma_{\hat{\psi}})$ and (2) $S_\Omega/\sigma^2 \sim \chi^2_{n-r}$.

(3) $\hat{\Psi}$ and S_Ω/σ^2 are statistically independent.

PROOF: (1)From (1) appendix I we have that any linear transformation of a multivariate normal Y remains multivariate normal. More precisely, if $Y \sim N(X'\beta,\sigma^2 I)$ and $\hat{\Psi} = AY$, then $\hat{\Psi} \sim N(AX'\beta,\sigma^2 AA')$. Further, $E(\hat{\Psi}) = E(AY) = AX'\beta$. But $\hat{\Psi}$ is an estimate of an estimable function and hence $E(AY) = AX'\beta$ must hold identically in β by the Gauss-Markoff theorem. Hence $AX = C$ and therefore $E(\hat{\Psi}) = C\beta = \Psi$.

(3) To show $\hat{\Psi}$ is independent of S_Ω/σ^2 it suffices to show that $\hat{\Psi}$ is independent of S_Ω. Consider the canonical form $Z = PY$ where $P'P = I$ so that $Z \sim N(\xi,\sigma^2 I)$. Then

$$E(z_i) = \begin{cases} \xi_i & i=1,\ldots, r \\ 0 & i=r+1,\ldots, n \end{cases}$$

$$\mathcal{V}_{z_i} = \sigma^2 I$$

At the end of section (1.4) we found that $\hat{\Psi}$ is a function of the set $\{z_1,\ldots, z_r\}$, and S_Ω only of the set $\{z_{r+1},\ldots, z_n\}$. Since the two sets are statistically independent, so are $\hat{\Psi}$ and S_Ω.

(2) We now have to show $S_\Omega/\sigma^2 \sim \chi^2_{n-r}$. We have that

$$Z \sim N(\xi,\sigma^2 I) \qquad E(z_i) = \begin{cases} \xi_i & i=1,\ldots, r \\ 0 & i=r+1,\ldots, n \end{cases}$$

Hence

$$z_i \overset{\text{independently}}{\sim} N(0,\sigma^2) \qquad i=r+1,\ldots, n$$

Hence

$$S_\Omega/\sigma^2 = \sum_{i=r+1}^{n} z_i^2/\sigma^2 \text{ where } z_i/\sigma \overset{\text{independently}}{\sim} N(0,1).$$

Hence

S_Ω/σ^2 is the sum of the squares of (n-r) indepen-

dent $N(0,1)$ variates so that

$$S_\Omega/\sigma^2 \sim \chi^2_{n-r}$$

Q.E.D.

(1.6) Test of Hypothesis H derived from the Likelihood Ratio. The Statistic F.

Let Ω be our set of underlying assumptions and H the hypothesis about the parameters of the density function of Y.

We introduce the symbols

$$\omega = \Omega \cap H$$

meaning that the set of assumptions obtained by imposing the assumptions of the hypothesis H in addition to the assumptions Ω, i.e. ω is the set of assumptions obtained by combining

$$\Omega : Y \sim N(X'\beta, \sigma^2)$$

and

$$H : \psi_1 = \psi_2 = ... = \psi_q = 0.$$

Let $p(Y)$ be the density function of the observations Y. The likelihood-ratio statistic λ is defined by

$$\lambda = \max_{\omega} p(Y) / \max_{\Omega} p(Y).$$

We note that $0 \leq \lambda \leq 1$ since any value of $p(Y)$ possible under ω is also possible under Ω. The likelihood-ratio test in rejecting H if $\lambda < \lambda_o$ where the constant λ_o is chosen to give the desired level of significance.

Another form of the Ω- and ω- assumptions which is useful in geometrical arguments is

$\Omega : Y \sim N(\eta, \sigma^2 I)$, $\eta \varepsilon Vr$, a given r-dimensional sub-

space of Vn.

H : $\eta \varepsilon Vr-q$, a given (r-q)-dimensional subspace of Vr.

To calculate the likelihood-ratio statistic for testing H under Ω we need the joint density function of the observations,

$$p(Y) = p_1(y_1)p_2(y_2) \ldots\ldots p_n(y_n)$$

where

$$p_i(y_i) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(y_i - \sum_{j=1}^{p} x_{ji}\beta_j)^2/\sigma^2\right].$$

The joint density is

$$p(Y) = (2\pi\sigma^2)^{-n/2} \exp\left[-\tfrac{1}{2}S(Y,\beta)/\sigma^2\right] \qquad (1.9)$$

where $S(Y,\beta) = ||Y-X'\beta||^2$ is the sum of squares minimized in the least squares theory. We shall find the maximum of p(Y) under Ω and ω simultaneously by writing

$$\Omega = \Omega_1 \qquad Vr = V(1)$$

$$\omega = \Omega_2 \qquad Vr-q = V(2).$$

The problem is now to find the maximum of (1.9) under Ω_i (i=1,2), or the maximum of

$$p(Y) = (2\pi\sigma^2)^{-n/2} \exp\left[-\tfrac{1}{2}\sigma^{-2}||Y-\eta||^2\right] \qquad (1.10)$$

for $0<\sigma^2<\infty$, $\eta \varepsilon V(i)$.

For fixed σ^2 the maximum of p(Y) in (1.10) occurs when $||Y-\eta||^2$ is a minimum i.e., when η is the projection of Y on V(i) (which will be denoted by $\eta_{\Omega i}$). The maximum of (1.10) can be found by first calculating the maximum for fixed σ^2 and varying $\eta \varepsilon V(i)$, and then maximizing this for varying σ^2. Thus, the maximum of (1.10) for fixed σ^2 is

$$(2\pi\sigma^2)^{-n/2} \exp\left[-\tfrac{1}{2}\sigma^{-2}||Y-\eta_{\Omega i}||^2\right].$$

The question now is for what value of σ^2 does this maximum occur. If $p(Y)$ has a maximum, then $\log p(Y)$ also has a maximum. For a maximum we must have:

$$\partial\left[p(Y)\right]/\partial\sigma^2 = 0 \; ; \; \partial^2\left[p(Y)\right]/\partial^2\sigma^2 < 0$$

Thus, letting $||Y-\eta_{\Omega i}||^2 = S_{\Omega i}$ we have that

$$\log p(Y) = -(n/2)\log(2\pi) - (n/2)\log\sigma^2 - \tfrac{1}{2}(S_{\Omega i}|\sigma^2)$$

$$\partial(\log p(Y))/\partial\sigma^2 = -n/(2\sigma^2) + S_{\Omega i}/(2\sigma^4) = 0 .$$

Hence

$$n/\sigma^2 = S_{\Omega i}/\sigma^4$$

or

$$\sigma^2 = S_{\Omega i}/n .$$

$$\partial^2(\log p(Y))/\partial^2\sigma^2 = n/(2\sigma^4) - S_{\Omega i}/\sigma^6$$

which is negative for $\sigma^2 = S_{\Omega i}/n$.

Hence

$$\underset{\Omega i}{\text{Max}}\; p(Y) = (2\pi S_{\Omega i}/n)^{-n/2} \exp\left[\frac{n}{-2}\right]$$

so that

$$\lambda = \left[\frac{(1/n)\; 2\pi\; S_{\Omega 2}}{(1/n)\; 2\pi\; S_{\Omega 1}}\right]^{-n/2}$$

$$= (S_\omega/S_\Omega)^{-n/2}$$

where

$$S_\Omega = ||Y-\hat{\eta}||^2 \quad ; \quad S_\omega = ||Y-\hat{\eta}_\omega||^2$$

and $\hat{\eta}$ the projection of Y on Vr ;

$\hat{\eta}_\omega$ the projection of Y on Vr-q .

In practice we use the statistic

$$F = \frac{n-r}{q}\;\frac{S_\omega - S_\Omega}{S_\Omega} \tag{1.11}$$

instead of λ but the test is the same as the λ test since

$$F = F(\lambda) = (n-r)(\lambda^{-2/n}-1)/q$$

is a single valued decreasing function of λ everywhere in $0 \leq \lambda \leq 1$. Hence, if we define $F_o = F_o(\lambda_o)$, $\lambda < \lambda_o$ if and only if $F > F_o$. Thus the λ-test is equivalent to rejecting H if and only if $F > F_o$, where the constant F_o is to be determined to yield the desired significance level. Hence the likelihood ratio test of H is

Reject H : $\psi_1 = \psi_2 = \ldots = \psi_q$ if

$$\frac{(n-r)}{q}\left[\frac{S_\omega - S_\Omega}{S_\Omega}\right] > F_{\alpha,q,n-r}. \qquad (1.12)$$

(1.7) Canonical form of Ω and H. Distribution of F under Ω.

DEFN. A basis $\{\alpha_1, \ldots, \alpha_r\}$ for $Vr \subset Vn$ is called orthonormal if the r vectors α_i are pairwise orthogonal and have unit norm.

ω, Ω and H determine three vector spaces $Vr-q \subset Vr \subset Vn$. By the Schmidt orthogolisation process ((10) appendix I) we can find an orthonormal basis (O.N) for these spaces. By (11) appendix I we can choose a set of r-q vectors $\{\alpha_i\}$ as orthonormal basis for Vr-q and augment this basis to form an orthonormal basis for Vr which can then again be augmented to form an orthonormal basis for Vn. We then have

$$\alpha_1, \ldots, \alpha_{q+1}, \ldots, \alpha_r, \alpha_{r+1}, \ldots, \alpha_n$$

where $\alpha_{q+1}, \ldots, \alpha_r$ is an orthonormal basis for Vr-q

$\alpha_1, \ldots, \alpha_r$ is an orthonormal basis for Vr

$\alpha_1, \ldots, \alpha_n$ is an orthonormal basis for Vn.

Let $(z_1, \ldots, z_n)$ be coordinates of Y relative to the basis $\{\alpha_i\}$. Then $z_i = \alpha_i' Y$, $Z^{nx1} = PY$.

Then $z_i = \alpha_i' Y$

$Z^{nx1} = PY$ P orthogonal matrix with i th row α_i'

$\xi_i = E(z_i)$

$\xi^{nx1} = E(PY) = P(\eta)$.

Under Ω we have that $Z \sim N(\xi, \sigma^2 I)$.

Under Ω : $\xi_i = 0 \quad i>r$

Under ω : $\xi_i = 0 \quad i \leq q$

Then we have the following canonical form.

Ω : $\{z_i\}$ are statistically independent,

$z_i \sim N(\xi_i, \sigma^2) \quad i=1, \ldots, n,$

$\xi_{r+1}, \ldots, \xi_n = 0$

H : $\xi_1 = \xi_2 = \ldots\ldots = \xi_q = 0$.

Now $\hat{\eta}$ and $\hat{\eta}_\omega$ are the respective projections of Y on Vr and Vr-q so that

$$\hat{\eta} = \sum_{1}^{r} z_i \alpha_i \; ; \; \hat{\eta}_\omega = \sum_{q+1}^{r} z_i \alpha_i \text{ so that}$$

$$S_\Omega = ||Y-\hat{\eta}||^2 = ||\sum_{i=r+1}^{n} z_i \alpha_i||^2 = \sum_{i=r+1}^{n} z_i^2$$

$$S_\omega = ||Y-\hat{\eta}_\omega||^2 = ||\sum_{i=1}^{q} z_i \alpha_i + \sum_{i=r+1}^{n} z_i \alpha_i||^2$$

$$= \sum_{i=1}^{q} z_i^2 + \sum_{i=r+1}^{n} z_i^2 .$$

Hence $S_\omega - S_\Omega = \sum_{i=1}^{q} z_i^2$.

Since the two sets $\{z_1, \ldots, z_q\}$ and $\{z_{r+1}, \ldots, z_n\}$ are statistically independent, $S_\omega - S_\Omega$ and S_Ω are statistically independent. Also $E(z_i) = 0$ $i>r$ and $z_i \sim N(\xi_i, \sigma^2)$ $i>r$ so that $(S_\omega - S_\Omega)/\sigma^2 = \sum_{i=1}^{q} (z_i/\sigma)^2$ is the sum of q normal variables. Hence

$$(S_\omega - S_\Omega)/\sigma^2 \sim \chi'^2_q(\delta) \text{ with noncentrality parameter}$$

$$\delta = \left(\sum_{i=1}^{q} \xi_i^2/\sigma^2\right)^{\frac{1}{2}}$$

according to (12) appendix I.

By (13) appendix I

$$F = \frac{n-r}{q} \frac{S_\omega - S_\Omega}{S_\Omega} \sim F'_{q,n-r;\,\delta}.$$

In particular, under $\omega = \Omega \cap H$ where H: $\xi_1 = \ldots = \xi_q = 0$ this reduces to the central F-distribution so that the F-test under Ω at α level of significance is:

Reject H if $F > F_{\alpha,q,n-r}$.

A P P E N D I X I

(1) A vector X is any ordered n-tuple of real numbers which are usually written in a row as $(x_1, x_2, \ldots, x_n)$.

(2) For any given n, the set of all vectors is denoted by Vn and called the n-dimensional vector space.

(3) Let $Vr \subset Vn$, and $X \in Vn$. Then X is said to be orthogonal to Vr (we write $X \perp Vr$) if and only if X is orthogonal to every vector in Vr (Vr is an r-dimensional vector space contained in Vn).

(4) If $\{\alpha_1, \alpha_2, \ldots, \alpha_s\}$ spans $Vr \subset Vn$, then a vector $X \in Vn$ is orthogonal to Vr if and only if X is orthogonal to each α_i $(i=1, 2, \ldots, s)$.

PROOF: If $X \perp Vr$, then $X \perp \alpha_i$ by (3).

If $X \perp \alpha_i$ for $i=1, 2, \ldots, s$, and if $Y \in Vr$, we can write $Y = \sum_{i=1}^{s} b_i \alpha_i$.

Then $X'Y = \sum_{i=1}^{s} b_i X'\alpha_i = 0$. Thus $X \perp Y$ for all $Y \in Vr$.

Q.E.D.

(5) If $Vr \subset Vn$ and $Y \in Vn$, then there exist vectors X and Z where $X \in Vr$ and $Z \perp Vr$ and such that $X = X+Z$. This decomposition is unique. A detailed proof of this

statement can be found in (21).

(6) Given a vector $Y \varepsilon Vn$, the vector $X \varepsilon Vr$ such that $(Y-X) \perp Vr$ is called the projection of Y on Vr.

(7) Given a fixed $Vr \subset Vn$, a fixed vector $Y \varepsilon Vn$ and a variable vector $X \subset Vr$, then $||Y-X||$ has a minimum value which is attained if and only if X is the projection of Y on Vr.

PROOF: Let X^* be the projection of Y on Vr. X is a variable vector. If we write

$$(Y-X) = (Y-X^*) + (X^*-X) \quad \text{then}$$

$$||Y-X||^2 = (Y-X^*)'(Y-X^*) + (X^*-X)'(X^*-X) + (Y-X^*)'(X^*-X) + (X^*-X)'(Y-X).$$

But $(Y-X^*) \perp Vr$, while X, X^* and hence $X-X^*$ are all in Vr.

Hence

$$||Y-X||^2 = ||Y-X^*||^2 + ||X-X^*||^2.$$

If X varies in Vr, the first of the two terms on the right is fixed while the second is variable with value ≥ 0, and $= 0$ if and only if $X = X^*$.

Q.E.D.

(8) For any (real) A, the matrix AA' is symmetric, positive definite, and of the same rank as A. A proof of this statement is found in (21) appendix II.

(9) For every symmetric B^{nxn} there exists a nonsingular P^{nxn} such that P'BP = I. A proof of this statement is found in (21) appendix II.

(10) Schmidt Orthogonalisation process: Given an arbitrary basis $\{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ for Vr, there exists an orthonormal basis $\{\gamma_1, \ldots, \gamma_r\}$ for Vr such that each γ_i is a linear combination of $\alpha_1, \ldots, \alpha_i$. A proof of this statement is found in (21) appendix I.

(11) If $\{\alpha_1, \ldots, \alpha_r\}$ is an orthonormal basis for $Vr \subset Vn$ it is always possible to extend it to an orthonormal basis $\{\alpha_1, \ldots, \alpha_r, \alpha_{r+1}, \ldots, \alpha_n\}$ for Vn. A proof of this statement is found in (21) appendix I.

(12) If $x_1, \ldots, x_r$ are undependently distributed and x_i is $N(\mu_i, 1)$, then the random variable $U = \sum_1^r x_i^2$ is called a non-central chi-square variable with r d.o.f. (degrees of freedom). We call $\delta = (\sum_1^r \mu_i^2)^{\frac{1}{2}}$ the noncentrality parameter of the distribution and

use the symbol $\chi'^2_r(\delta)$ for a noncentral chi-square variable with r d.o.f. and noncentrality parameter δ. In particular, when $\delta = 0$ we have the central chi-square χ^2_r.

(13) If U_1 is an independent noncentral chi-square variable with v_1 d.o.f. and noncentrality parameter δ, and U_2 an independent cetral chi-square variable with v_2 d.o.f., then $(U_1/v_1)/(U_2/v_2)$ is called a noncentral F-distribution with v_1 and v_2 d.o.f. and noncentrality parameter δ and is written $F'_{v_1,v_2;\delta}$. In particular when $\delta = 0$ this reduces to a central F-distribution.

Chapter II

The One-Way Layout: Equal Cell Numbers.

(2.1) Method of Analysis (the model).

This is the simplest case in which the analysis of variance is applied. The one-way layout enables us to compare the means of several univariate normal distributions and is a generalization of the two-sample t-test with k>2. Let the means be $\beta_1,\ldots,\beta_I$. We will assume that:

(1) The I populations are normal with equal variance σ^2.

(2) Independent random samples, each of size J are drawn from the respective populations. If we denote the sample from the i th population by $y_{i1},\ldots, y_{iJ}$, then our underlying assumptions are

$$\Omega : y_{ij} = \beta_i + e_{ij} \quad (i=1,\ldots, I;\ j=1,\ldots, J).$$
$$\{e_{ij}\} \text{ are independently } N(0,\sigma^2)$$

$$H : \beta_1 = \beta_2 = \ldots = \beta_I.$$

In the general theory of chapter I we have that n = IJ, r = I i.e., the rank of the nxI matrix X' is I so that all parametric functions are estimable. The sum of squares to be minimized under Ω is

$$S(Y,\beta) = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - \beta_i)^2 \qquad (2.1)$$

so that the N.E. under Ω becomes

$$\frac{\partial S(Y,\beta)}{\partial \beta_t} = -2 \sum_{j=1}^{J} (y_{tj} - \beta_t) = 0 \quad t=1,\ldots, I.$$

Hence

$$\hat{\beta}_t = \left(\sum_{j=1}^{J} y_{tj}\right)/J$$

NOTATION: Replacing a subscript by a dot means that the arithmetic average of the quantities to which the subscript is attached has been taken over all possible values of the subscript.

With the convenient dot notation we have

$$\hat{\beta}_i = y_{i\cdot} \quad (i=1,\ldots, I).$$

Under ω, minimizing (2.1) is the same as minimizing

$$S(Y,\beta_\omega) = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - \beta)^2.$$

Solving the one N.E.

$$\frac{\partial S(Y,\beta_\omega)}{\partial \beta} = -2 \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - \beta) = 0$$

which gives

$$\hat{\beta}_\omega = \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij}/n \quad \text{where } n = IJ$$

$$= y_{\cdot\cdot} \quad .$$

Recalling that $E(Y) = \eta$ and that $\hat{\eta}$ and $\hat{\eta}_\omega$ are the L.S. estimates under Ω and ω respectively, and $E(y_{ij}) = \eta_{ij}$, we have that

$$\hat{\eta}_{ij} = \hat{\beta}_i = y_{i\cdot} \quad .$$

Similarly under ω

$$\hat{\eta}_{ij,\omega} = y_{\cdot\cdot} \quad .$$

The numerator and denominator sums of squares (SS) of the test-statistic F is obtained from the identities

$$SS_H = ||\hat{\eta} - \hat{\eta}_\omega||^2; \quad SS_e = ||Y - \hat{\eta}||^2$$

which become

$$SS_H = \sum_{i,j} (\hat{\eta}_{ij} - \eta_{ij,\omega})^2 = \sum_{i,j} (y_{i\cdot} - y_{\cdot\cdot})^2 = J \sum_{i} (y_{i\cdot} - y_{\cdot\cdot})^2$$

$$SS_e = \sum_{i,j} (y_{ij} - \hat{\eta}_{ij})^2 = \sum_{i,j} (y_{ij} - y_{i.})^2.$$

Also we have

$$SS_e + SS_H = ||Y - \hat{\eta}_\omega||^2 = \sum_{i,j} (y_{ij} - y_{..})^2$$

which is called the total SS about the grand mean and is denoted by SS_{tot}. SS_H is called the SS between groups, and SS_e the SS within groups. The hypothesis H: $\beta_1 = \ldots = \beta_I$ can be specified by equating to zero I-1 linearly independent estimable functions H: $\beta_2 - \beta_1 = 0, \ldots, \beta_I - \beta_1 = 0$ so that d.o.f. for SS_H is $q = I-1$. Since r=I the number of d.o.f. for SS_e is n-I so that the F-statistic is MS_H/MS_e where $MS_H = SS_H/(I-1)$; $MS_e = SS_e/(n-I)$. Hence the F-test of H consists in rejecting H at α level of significance if and only if $F > F_{\alpha,I-1,n-I}$. It is convenient to summarize these results in an analysis of variance table as follows

Table 2.1

Analysis of Variance of One-Way Layout

will equal cell numbers.

Source	SS	dof.	MS	F
Between groups	$SS_H = J \sum_i (y_{i.} - y_{..})^2$	I-1	$SS_H/(I-1)$	MS_H/MS_e
Within groups	$SS_e = \sum_i \sum_j (y_{ij} - y_{i.})^2$	n-I	$SS_e/(n-I)$	
Total	$SS_{tot} = \sum_i \sum_j (y_{ij} - y_{..})^2$	n-1		

REMARKS: (1) For calculation purposes the following identities are used:

$$SS_H = J \sum_{i}^{I} y_{i.}^2 - ny_{..}^2 \qquad (2.2)$$

$$SS_{tot} = \sum_{i=1}^{I} \sum_{j=1}^{J} y_{ij}^2 - ny_{..}^2 \qquad (2.3)$$

(2) An alternative formulation of the above model is

$$y_{ij} = \begin{cases} \mu + \alpha_i + e_{ij} & i=1,\ldots, I;\ j=1,\ldots, J \\ \{e_{ij}\} \text{ are independent } N(0,\sigma^2) \end{cases}$$

$H : \alpha_1 = \alpha_2 = \ldots = \alpha_I = 0$

In this model the rank of X' is less than I. Imposing the side condition $\sum_i \hat{\alpha}_i = 0$ on the solution of the N.E. makes the solution unique and hence every parametric function is estimable.

(2.2) Description of Data.

It would seem appropriate to say something about wool marketing in general and explain some of the concepts which are standard wool trade terminology.

The auction is a public sale by open outcry at which wool is sold to the highest of successive bidders. Wool is presented for sale at auction in South Africa at the four ports Port Elizabeth, East London, Cape Town and Durban. The selling season in South Africa extends from September to May and four sales, one in each of the four ports, are held per week. The special feature of the sale by auction is that it displays the product publicly to all potential buyers, with opportunity of providing maximum competition. The receiving, storing, handling and displaying of wool for auction is handled by the wool selling brokers.

There is a large range of wool types sold in South Africa and because of differences in their technical properties (e.g. fineness and processing performance) different prices are paid for them at auction. The type number given to wool indicates the quality number, style, staple length, vegetable matter fault and colour of that wool. A complete list of the approximate 600 South African wool types can be found in (49). Various species of seed and burr, sand and dirt can be collected in the fleece of sheep during a woolgrowing period. The 'yield' of any wool is the proportion of clean wool per quantity under consideration and is expressed as a percentage of the total mass after allowing for wool grease. Yields vary considerably and ranges from 30%-70% are quite common. Hence, the clean mass of a bale of wool weighing 150 kilograms with an estimated yield of 50% is 75 kilograms. If the realized grease price of this wool was 70 cents per kilogram, the clean price of the same wool would be 70/.5 = 140 cents/kg. Grease prices, discounted for yields, are declared at auction. The higher the estimated yield, the higher the corresponding grease price.

Indeed an auction sale is based on the existence of differences between different buyers estimates of value on the same lots. Potential buyers in a wool auction are invited to estimate values for each sale lot. The traditional method of classifying and appraising wool at auctions is the result of many years of trial and error in the market. Most of the standards used by

appraisers and buyers are based on subjective associations between fleece or staple characteristics and fibre properties. For example, the clean yield of a sale lot of greasy wool is estimated from the appearance of fleeces in the lot and by association with the performance of similar wools previously purchased or appraised. The development of laboratory methods of wool classification, either as a supplement or as an alternative to the traditional methods of appraisal, pose important questions for the wool industry.

In South Africa it is the task of the South African Wool Commission (SAWC) to protect farmers against abnormal low wool prices by its system of reserve prices for all wool types beyond which no wool is sold to the trade. In the event of a lot not attaining the SAWC's predetermined reserve price for that lot, the SAWC declares it's reserve price for the lot at auction. Should no further bids be made, the lot is sold to the Commission. It is therefore clear that the technical efficiency of the traditional methods of yield estimation is as important to the Commission as it is to the trade. It is for this reason that specific lots are laboratory tested for yield and compared with the estimates made by the Wool Commission appraisers so as to keep them in line.

The yield data used in this chapter are the estimates for percentage yield made by the Wool Commission appraisers in the various ports during the 1970/71 season together with the

laboratory tested results. Two types, type 48 and type 53, were selected and sixteen estimates of yield for each type were taken from each of the four ports. The tested yield for each estimate completes the data. The following is a description of the two types used in this study.

Type 48: Good topmaking merino fleece with a staple lenght of 2 3/4 to 3 1/4 inches and a 64's spinning count.

Type 53: Good topmaking merino fleece with a staple lenght of 2 1/4 to 2 1/2 inches and a 64's spinning count.

Two seperate one-way analysis of variances, one for each of the two types, were made with the estimated and tested results for all ports forming the I populations, I being 8 in this case. It was felt that the estimates in the four ports should not differ significantly and that the assumption of equal variances was quite tenable so that, if the appraisers are equally skilled in the four ports, the one-way layout should not give a significant F-statistic.

(2.3) Results.

For ease of notation the following abbreviations will be used: CT. - Cape Town; DBN - Durban; EL. - East London; PE. - Port Elizabeth; E-estimated yield; T - tested yield. Hence we can write PE(T) for tested yields at PE etc. The following are summaries of the I sample means and variances, followed by the analysis of variance (ANOVA) tables for the two types respectively:

TYPE 48:

	$\hat{\beta}_i = y_{i.}$	s_i^2
CT(E)	64.75	3.80
CT(T)	65.69	5.03
DBN(E)	65.69	6.23
DBN(T)	63.75	8.20
EL(E)	62.56	5.33
EL(T)	62.44	8.26
PE(E)	63.38	5.98
PE(T)	64.75	13.40

$I = 8$; $J = 16$; $n = IJ = 128$; $y_{..} = 64.13$

Table 2.2

One-Way ANOVA for % Yield estimates:

Type 48 season 1970-1971.

Source	SS	dof	MS	F
Between yields	186.50	7	26.64	3.79**
Within yields	843.50	120	7.03	
Total	1030.00	127		

* - significant at 5%

** - significant at 1%

Hence H: $\beta_1 = \beta_2 = \ldots = \beta_I$ is rejected

TYPE 53:

	$y_{i.}$	s_i^2
CT(E)	63.50	5.47
CT(T)	65.63	8.12
DBN(E)	63.75	5.40
DBN(T)	62.44	7.73
EL(E)	60.31	4.76
EL(T)	59.63	7.72
PE(E)	62.00	7.20
PE(T)	62.56	10.26

$I = 8$; $J = 16$; $n = IJ = 128$; $y_{..} = 62.48$

Table 2.3

One-Way ANOVA for % Yield estimates:

Type 53 season 1970-1971.

Source	SS	dof	MS	F
Between yields	410.12	7	58.59	8.28**
Within yields	849.81	120	7.08	
Total	1259.93	127		

* - significant at 5%

** - significant at 1%

H : $\beta_1 = \beta_2 = \ldots = \beta_I$ is rejected

(2.4) Further Analysis (Multiple Comparisons)

The fact that the H hypothesis have been rejected in either case at the 1% level suggests that further inferences about the means are desirable. This brings us in the domain of multiple comparisons. The general principles of multiple comparisons were forged into their current structure between 1947 and 1955 by three principal investigators: Duncan, Scheffé and Tukey. Whereas the S-method is due to Scheffé and utilizes the F-distribution, the T-method is due to Tukey and utilizes the distribution of $q_{k,v}$, the Studentized range. We will give a short description of either method. Either method will tell us which parameters are responsible for the rejection of H.

DEFN. A contrast among the parameters $\beta_1,\ldots,\beta_I$ is a linear function of the β_i, $\sum_{i=1}^{I} c_i\beta_i$, with known constant coefficients subject to the condition $\sum_{i=1}^{I} c_i = 0$.

DEFN. A set L of estimable functions $\{\psi\}$ is called a q-dimensionalspace of estimable functions if there exists q linearly independent estimable functions $\{\psi_1,\ldots,\psi_q\}$ such that every ψ in L is of the form $\psi = \sum_{i=1}^{q} h_i\psi_i$, where $h_1,\ldots,h_q$ are known constant coefficients i.e. L is the set of all linear combinations of $\psi_1,\ldots,\psi_q$.

The Scheffé technique will, under Ω, give a probability of $1-\alpha$ that simultaneously for all $\psi \varepsilon L$,

$$\hat{\psi} - S\hat{\sigma}_{\hat{\psi}} \leq \psi \leq \hat{\psi} + S\hat{\sigma}_{\hat{\psi}} \tag{2.4}$$

where $\hat{\sigma}_{\hat{\psi}}^{\,2}$ = the variance of $\hat{\psi}$

$$S = (q\, F_{\alpha,q,n-r})^{\frac{1}{2}} . \tag{2.5}$$

Under certain conditions simultaneous confidence statements about contrasts among a set of parameters $\{\beta_j\}$ in terms of unbiased estimates $\{\hat{\beta}_j\}$ and s^2 can be made by the T-method. One of the restrictions is that the $\{\hat{\beta}_j\}$ have equal variances which implies equal sample sizes. Under the assumptions

$$\Omega : \left[\begin{array}{l} \text{The } \{\hat{\beta}_i\} \text{ are statistically independent} \\ \text{and } \hat{\beta}_i \text{ is } N(\beta_i, a^2\sigma^2),\ i=1,\ldots,k, \text{ where} \\ a \text{ is a known pos. constant. } s^2 \text{ is an} \\ \text{independent estimate of } \sigma^2 \text{ with } v \text{ d.o.f.} \\ \text{i.e. } vs^2/\sigma^2 \sim \chi^2_v, \text{ independent of the} \\ \{\hat{\beta}_i\} \end{array}\right. \tag{2.6}$$

the probability is $1-\alpha$ that the values of all contrasts $\psi = \sum_{i=1}^{k} c_i\beta_i$ $(\sum_i c_i = 0)$ simultaneously satisfy

$$\hat{\psi} - Ts(\tfrac{1}{2}\sum_{i=1}^{k} |c_i|) \leq \psi \leq \psi + Ts(\tfrac{1}{2}\sum_{i=1}^{k} |c_i|) \tag{2.7}$$

where $\hat{\psi} = \sum_{i=1}^{k} c_i\hat{\beta}_i$; $T = aq_{\alpha,k,v}$ and $q_{\alpha,k,v}$ is the upper α point of the Studentized range $q_{k,v}$.

REMARKS: (1) Detailed proofs of (2.4) and 2.6 can be found in (21) and (15).

(2) In the one-way layout a contrast $\psi = \sum_i c_i\beta_i$

admits the unbiased estimate $\hat{\psi}=\sum_i c_i\hat{\beta}_i = \sum_i c_i y_{i\cdot}$.
The variance of $\hat{\psi}$ is

$$\hat{\sigma}_{\hat{\psi}}^2 = \sum_i c_i^2 \text{ var } (y_{i\cdot}) = \sigma \sum_i (c_i^2/J)$$

which is estimated by

$$\hat{\sigma}_{\hat{\psi}}^2 = s^2 \sum_i (c_i^2/J) \text{ where } s^2 = \text{MSe.} \qquad (2.8)$$

(3) For a given space L of estimable functions and confidence coefficient $1-\alpha$, the L.S. estimate $\hat{\psi}$ of an estimable function $\hat{\psi}\varepsilon L$ will be said to be significantly different from zero according to the S-criterion if the interval (2.4) does not cover $\psi = 0$ i.e., if $|\hat{\psi}|>S\hat{\sigma}_{\hat{\psi}}$.

(4) The T-method is of limited applicability since it is available only for the case of equal variances of the $\{\hat{\beta}_i\}$. Since the T-method was originally designed to give intervals for the differences $\{\beta_i-\beta_{i'}\}$, we might expect the T-method to give shorter intervals for these differences.

(5) For pairwise comparisons $(\frac{1}{2}\sum_{i=1}^{k} |c_i|)$ in (2.7) is equal to one.

APPLICATION. (at 5%)

For pairwise comparisons it is convenient to rearange the true means in the order of the observed means so that $\hat{\beta}_{1i}\leq \hat{\beta}_{2i}\leq\ldots\leq\hat{\beta}_{Ii}$. $(i=1,\ldots, I)$ references the random position of the original estimates $\hat{\beta}_i$ and the first subscript indicates the order of magnitude of the $\hat{\beta}_i$. Hence we have:

Type 48 - S Method:

62.44 62.56 63.38 63.75 64.75 64.75 65.69 65.69

$\hat{\beta}_{16}$ $\hat{\beta}_{25}$ $\hat{\beta}_{37}$ $\hat{\beta}_{44}$ $\hat{\beta}_{58}$ $\hat{\beta}_{61}$ $\hat{\beta}_{72}$ $\hat{\beta}_{83}$

(1) Pairwise comparisons. We have from (2.8):

$$\hat{\sigma}^2_{\hat{\psi}} = (S^2/J) \left[\Sigma c_i^2 \right]$$

$$= (7.03/16)2$$

$$= .8788 \rightarrow$$

and from (2.5):

$$S = (qF_{\alpha,q,n-r})^{\frac{1}{2}} \quad (q = I-1)$$

$$= 3.82 \ (5\%)$$

so that

$$S\,\hat{\sigma}_{\hat{\psi}} = 3.58 \rightarrow$$

From remark (3) we have that only pairwise differences with absolute value greater than 3.58 will be significantly different by the Scheffé technique at the 5% level so that we have to conclude no pairwise differences by this method.

(2) For the contrast average of E against average T we have:

$$\hat{\psi} = 1/4(\hat{\beta}_2+\hat{\beta}_4+\hat{\beta}_6+\hat{\beta}_8) - 1/4(\hat{\beta}_1+\hat{\beta}_3+\hat{\beta}_5+\hat{\beta}_7)$$

$$\hat{\psi} = .75 \rightarrow$$

$$S\hat{\sigma}_{\hat{\psi}} = 3.82 \ (.4687) = 1.79 \rightarrow$$

Hence no difference between average E and average T by the S-method.

Type 48 - T Method:

(1) For pairwise comparisons we have from (2.7) that

$$Ts = .25(4.36)\ 2.6515$$

$$= 2.89 \rightarrow$$

so that all pairwise differences with absolute value greater than 2.89 will be significantly different by the T-method. Thus we have:

$$\beta_{83} > \beta_{25} \; ; \; \beta_{16}$$

$$\beta_{72} > \beta_{25} \; ; \; \beta_{16}$$

all significant at the 5% level.

(2) For the contrast average E against average T we have:

$\hat{\psi} = .75$

$$Ts \; \tfrac{1}{2} \sum_{i=1}^{k} |c_i| = (.25)\ 4.36\ (2.6515)(.5)2$$

$$= 2.89 \; \rightarrow$$

REMARK: Note that the S-method gives an interval of $\pm$ 3.58 for pairwise comparisons while the T-method gives an interval of only $\pm$ 2.89 which agrees with our previous remark no (4).

Type 53 - S Method:

Reorganizing the $\{\hat{\beta}_i\}$ we have:

59.63	60.31	62.00	62.44	62.56	63.50	63.75	65.63
$\hat{\beta}_{16}$	$\hat{\beta}_{25}$	$\hat{\beta}_{37}$	$\hat{\beta}_{44}$	$\hat{\beta}_{58}$	$\hat{\beta}_{61}$	$\hat{\beta}_{73}$	$\hat{\beta}_{82}$

(1) Pairwise Comparisons. From (2.8) we have:

$\hat{\sigma}^2_{\hat{\psi}} = .885$; S = 3.82 so that

$S\hat{\sigma}_{\hat{\psi}} = 3.59 \; \rightarrow$

Hence the following differences are significant at the 5% level.

$$\beta_{82} > \beta_{37}; \; \beta_{25}; \; \beta_{16}$$

$$\beta_{73} > \beta_{16}$$

$$\beta_{61} > \beta_{16}$$

(2) Average E against Average T.

$$\hat{\psi} = .70$$

$$S\hat{\sigma}_{\hat{\psi}} = (3.82)(.4704) = 1.80 \rightarrow$$

Hence no difference by S-method.

(3) EL(T) against average rest (T): $\hat{\psi} = \hat{\beta}_6 - 1/3(\hat{\beta}_2 + \hat{\beta}_4 + \hat{\beta}_8)$

$$\hat{\psi} = 3.91$$

$$S\hat{\sigma}_{\hat{\psi}} = 3.82(.7672) = 2.92 \rightarrow$$

This interval does not include zero so that we can conclude at the 5% level that there is a difference between EL(T) and the average of the T in the other three ports.

(4) EL(E) against average E of other ports:

$$\hat{\psi} = \hat{\beta}_5 - 1/3(\hat{\beta}_1 + \hat{\beta}_3 + \hat{\beta}_7)$$

$$= 2.77$$

$$S\hat{\sigma}_{\hat{\psi}} = 3.82(.7672)$$

$$= 2.92 \rightarrow$$

Hence no difference between EL(E) and the average (E) of the other three ports.

(5) CT(T) against average T of other ports:

$$\hat{\psi} = 4.09$$

$$S\hat{\sigma}_{\hat{\psi}} = 2.92$$

Hence significant difference at 5%.

Type 53 - T Method:

(1) Pairwise comparisons: From (2.7) we have

$$Ts = .25(4.36)2.66$$

$$= 2.90$$

Hence

$$\beta_{82} > \beta_{58};\ \beta_{44};\ \beta_{37};\ \beta_{25};\ \beta_{16}$$

$$\beta_{73} > \beta_{25};\ \beta_{16}$$

$$\beta_{61} > \beta_{25};\ \beta_{16}$$

$$\beta_{58} > \beta_{16}$$

(2) Average E against average T:

$$Ts\left(\tfrac{1}{2}\sum_{i=1}^{k} |c_i|\right) = .25(4.36)2.66$$

$$= 2.90$$

$$\hat{\psi} = .70$$

Hence no difference at 5%.

(3) EL(E) against average rest (E):

$$\hat{\psi} = 2.77$$

$$Ts\ \tfrac{1}{2}\sum_{1}^{k} |c_i| = 2.90$$

Hence no difference at 5%.

(4) EL(T) against average T of other ports:

$$\hat{\psi} = 3.91$$

$$Ts\ \tfrac{1}{2}\Sigma|c_i| = 2.90$$

This interval does not include zero so that there is a significant difference at α=5%.

Before summarizing it is worthwhile to note that in both the samples the variance for PE(T) seems high compared to the others so that a question hangs over the assumption of equal variances. Although inequality of variances has little effect on the F-test in a balanced design, we will nevertheless test

the equality of variances with a test which was derived by Bartlett (2):

With only two mean squares, the usual F-test is applicable. With a priori reason to anticipate inequality of variance the alternative in this case is a two sided one: $\sigma_1^2 \neq \sigma_2^2$. Hence the F-criterion is $F = s_1^2/s_2^2$ where s_1^2 is the larger mean square. The distribution of F when the null hypothesis is true was worked out by Fisher (6) early in the 1920's. With more than two independent estimates of variance Bartlett provided a test: If there are a estimates s_i^2, each with the same number of degrees of freedom f, the test criterion is

$$M = \log_e 10f \ (a \log \bar{s}^2 - \sum_i \log s_i^2) \qquad (2.9)$$

where

$$\bar{s}^2 = (\Sigma \ s_i^2)/a.$$

Under the null hypothesis that each s_i^2 is an estimate of the same σ^2, the quantity M/C is distributed approximately as χ^2 with (a-1) degrees of freedom, where $C = 1+(a+1)/(3\ af)$. When the degrees of freedom differ as in samples of unequal sizes, the computation of χ^2 is more tedious. Then

$$M = \log_e 10 \left[(\Sigma f_i) \log \bar{s}^2 - \Sigma f_i \log s_i^2 \right] \qquad (2.10)$$

where

$$\bar{s}^2 = (\Sigma s_i^2)/a \ ; \quad C = 1 + \frac{1}{3(a-1)} \left[\Sigma \frac{1}{f_i} - \frac{1}{\Sigma f_i} \right]$$

χ^2 = M/C with (a-1) degrees of freedom.

We will now apply (2.9) to our two samples.

		Type 48	Type 53
$\bar{s}^2$	:	7.03	7.08
$\log \bar{s}^2$	:	0.8470	0.8500
$\sum_i \log s_i^2$	:	6.5372	6.704
a	:	8.0	8.0
f	:	16.0	16.0
M	:	8.80	3.54
C	:	1.026	1.026
M/C	:	8.58	3.45
χ^2_7 (.05)	:	14.07	14.07

The values M/C of 8.58 and 3.45 respectively are clearly not significant so that the assumptions of equality of variances have not been violated.

(2.5) Conclusions.

In both cases the F-test rejected the null hypothesis of no differences in yields at the 1% level although the F-test was more significant in the case of type 53. From descriptions of the two types, it is clear that the only difference between the two lies in their staple length, the type 53 being of shorter staple length. What seems important at this stage is that the length classification of wool might influence the appraisement of yield and one would be inclined to reason that the shorter the staple length, the higher the variance in the estimated yields. Discussions with the technical staff of the SAWC

revealed that such a tendency could exist because of the smaller variety of wool from the fleece that qualify for inclusion in the longer length categories. Further research in this respect is required to establish the exact relationship between staple length and yield estimates.

The rejection of the F-tests suggested multiple comparisons. For the pairwise comparisons we choose to summarize the significant differences of the T-method as follows:

Type 48:

DBN(E) > EL(E); EL(T)

CT(T) > EL(E); EL(T)

Type 53:

CT(T) > PE(T); DBN(T); PE(E); EL(E); EL(T)

DBN(E) > EL(E); EL(T)

CT(E) > EL(E); EL(T)

PE(T) > EL(T)

The pairwise comparisons give no evidence of real differences in estimated and tested yields within a given port considered separately for each type so that, based on these two samples, it can be concluded that appraisers are equally skillfull in all the ports. However, the fact that CT(T) is consistently higher than the tested results in other ports but that CT(E) does not quite follow the same pattern, suggests the inclination to underestimate in Cape Town. Likewise, the fact that DBN(E),in both cases, are significantly higher than EL(E) while the same trend is not found in the tested results could, to a lesser extend, indicate

that Durban is inclined to overestimate. Comparing individual means does not seem to contradict these assertions.

The significance of the contrast CT(T) against the average of the other tested yields for type 53 shows that the true yields for this type is higher in the districts sending wool to Cape Town than those sending to the other ports.

It was also pleasing to note that averaged over all the ports, there was no overall difference between estimated and tested yields.

A most interesting fact was highlighted by the data. Comparing individual variances in the two samples show that variances for tested results are considerably higher than the corresponding variances for estimated yields in all cases.

Although the reason for this occurrence is not so obvious the following explanation seems well motivated.

Assuming that appraisers subconsciously associate a given type with a certain basic yield, the incorrect determination of type could account for greater variance in the estimated yields.

The spinning count (fineness) of wool refers to fibre tickness and is subjectively determined by the size of the crimp. The smaller the crimp, the finer the wool and vice versa.

In the good topmaking group a change in spinning count from a 64 to a 60/64 raises the type by one grade while a change in the quality from a 64 to a 64/70 lowers the type one grade.

It is obvious that an incorrect appraisement of spinning count will lead to an incorrect type classification and it's associated yield. We can thus conclude that it is to the combined analysis of tested yields and fibre thickness that future studies must be directed for a full understanding of the variation and consistency of estimated yields.

Chapter III

The One-Way Layout : Unequal Cell Numbers.

(3.1) Method of Analysis.

More often than not the one-way layout is unbalanced. Basically the model is the same as that treated in chapter II except that the assumption of equal cell numbers has been dropped. It is clear that the model in chapter one is a special case of the unbalanced design. We will now assume that:

(1) The I populations are normal with equal variance σ^2.

(2) Independent random samples of sizes $J_1, J_2, \ldots, J_I$ are drawn from the respective populations. Our Ω assumptions are:

$$\Omega : y_{ij} = \beta_i + e_{ij} \quad (i=1,\ldots, I;\ j=1,\ldots, J_i)$$

$$\{e_{ij}\} \text{ are independently } N(o,\sigma^2)$$

$$H : \beta_1 = \beta_2 = \ldots = \beta_I$$

with $n=\Sigma J_i$ and $r=I$ in the general theory of chapter I. The sum of squares to be minimized under Ω is.

$$S(Y,\beta) = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (y_{ij} - \beta_i)^2 \qquad (3.1)$$

so that the N.E. under Ω becomes

$$\frac{\partial S(Y,\beta)}{\partial \beta_t} = -2 \sum_{j=1}^{J_i} (y_{tj} - \beta_t) = 0 \quad (i=1,\ldots, I)$$

Hence

$$\hat{\beta}_t = (\sum_{j=1}^{J_i} y_{tj})/J_i \text{ so that}$$

$$\hat{\beta}_i = y_{i.} \text{ as before.}$$

Under ω we have to minimize

$$S(Y,\beta_\omega) = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (y_{ij} - \beta)^2$$

where β denotes the common (known) value of $\beta_1, \ldots, \beta_I$.

Solving for the only one N.E. we have

$$\hat{\beta}_\omega = \sum_{i=1}^{I} \sum_{j=1}^{J_i} y_{ij}/n \quad \text{where } n = \sum_{i=1}^{I} J_i.$$

$$= \bar{y}.$$

REMARK: We write $\bar{y}$ instead of the $y_{..}$ to denote the weighted average of the $\{y_{i.}\}$ instead of the unweighted average used in chapter I.

Further, we still have that

$$\hat{\eta}_{ij} = \hat{\beta}_i = y_{i.}$$

$$\hat{\eta}_{ij,\omega} = \bar{y}.$$

The numerator and denominator SS's of the test statistic F now becomes

$$SS_H = \sum_i \sum_j (y_{i.} - \bar{y})^2 = \sum_i J_i (y_{i.} - \bar{y})^2$$

$$= \sum_i J_i \, y_{i.}^2 - n\,\bar{y}^2.$$

$$SS_e = \sum_i \sum_j (y_{ij} - y_{i.})^2$$

Further $SS_{tot} = \sum_{i=1}^{I} \sum_{j=1}^{J_i} (y_{ij} - \bar{y})^2$

$$= \sum_{i=1}^{I} \sum_{j=1}^{J_i} y_{ij}^2 - n\bar{y}^2.$$

Bearing in mind that $n = \sum_{i=1}^{I} J_i$ the F statistic remains MS_H/MS_e with $q = I-1$ and $n-r = n-I$ degrees of freedom for the numerator and denominator respectively. The analysis of variance can be

summarized as follows:

Table 3.1

One-Way ANOVA with unequal cell numbers.

Source	SS	dof	MS	F
Between groups	$SS_H=\Sigma_i J_i(y_{i\cdot}-\bar{y})^2$	I-1	$SS_H/(I-1)$	MS_H/MS_e
Within groups	$SS_e=\Sigma_i \Sigma_j (y_{ij}-y_{i\cdot})^2$	n-I	$SS_E/(n-1)$	
Total	$SS_{tot}=\Sigma\Sigma(y_{ij}-\bar{y})^2$	n-1		

(3.2) Description of Data.

The price of wool, like the price of other comodities, is determined by the forces of supply and demand, a consideration which, incidentally, is often used to justify non intervention in the free marketing of wool. However, as we look deeper at the meaning of supply and demand, we find it necessary to be much more precise, especially if these concepts are to be used in any attempt to unravel and measure the various forces at work in causing wool prices to fluctuate. Firstly, supply and demand will have different meanings according to the time period involved. For example, are we talking about the equality of supply and demand over the course of one days wool sale, or one month's sales, or over the whole selling season, or even longer periods. Commodity prices generally follow a trend.

A trend is a persistent change occuring over a period of time. Superimposed on a trend are price movements, which in the short term often tend to obscure the direction of the trend. These deviations from a trend are fluctuations. Price fluctuations occur for particular types of wool and in the margin between types. They take place over different periods of time as for example within a sale; between sales within a week; between sales within a season; or over a number of seasons.

It is the purpose of this chapter to investigate the price fluctuations for a given type within a sale. A sale day is normally devided into three sessions. In particular it is the purpose of this chapter to investigate whether a significant price difference exsists between the three sessions of a sale day, and if it does, to try and establish a trend.

The price data used in the analysis was obtained from the SAWC appraisers in Port Elizabeth which recorded the grease prices of type 48 according to whether it was sold during the first, second or third session. The data refer to type 48 sold in Port Elizabeth during the 1969-1970 season and was recorded for a number of sales with reasonable representation in each session. Prices are still cents per pound and have been converted to clean prices to exclude yield differences.

REMARK: The data can at this stage be combined in a two-way layout with unequal cell numbers (chapter VII). The one-way layouts were however already analysed

weekly as the data became available, and for this reason only, kept as such.

(3.3) Results.

We refer to successive sales as catalogues and write cat 1 to refer to the first sale etc. Means, variances and ANOVA tables are given below for a number of catalogues. The sample sizes are also recorded. Significance of the F-statistic will be indicated by * for 5% and ** for 1%. All prices are clean prices in cents per lb.

Catalogue 1:

	Session 1	Session 2	Session 3
$y_{i.}$	69.01	66.58	67.42
s_i^2	2.18	3.35	3.77
J_i	10	9	10

Table 3.2

One-Way ANOVA For Within Sale Price Variation:

Type 48 Cat 1 1969/70.

Source	SS	dof	MS	F
Between Sessions	29.18	2	14.59	4.72*
Within Sessions	80.40	26	3.09	
Total	109.58	28		

Catalogue 2:

	Session 1	Session 2	Session 3
$y_{i\cdot}$	66.16	66.52	65.42
s_i^2	3.19	3.66	2.98
J_i	25	21	9

Table 3.3

One-Way ANOVA For Within Sale Price Variation:

Type 48 Cat 2 1969/70.

Source	SS	dof	MS	F
Between Sessions	7.53	2	3.77	1.13
Within Sessions	173.65	52	3.34	
Total	181.18	54		

Catalogue 4:

	Session 1	Session 2	Session 3
$y_{i\cdot}$	66.00	65.82	65.12
s_i^2	4.11	4.23	5.10
J_i	17	9	9

Table 3.4

One-Way ANOVA For Within Sale Price Variation:

Type 48 Cat 4 1969/70.

Source	SS	dof	MS	F
Between Sessions	4.68	2	2.34	.53
Within Sessions	140.40	32	4.39	
Total	145.08	34		

Catalogue 5:

	Session 1	Session 2	Session 3
$y_{i.}$	64.79	65.24	64.31
s_i^2	1.91	4.92	5.15
J_i	9	6	10

Table 3.5

One-Way ANOVA For Within Sale Price Variation:

Type 48 Cat 5 1969/70.

Source	SS	dof	MS	F
Between Sessions	3.35	2	1.68	.43
Within Sessions	86.17	22	3.92	
Total	89.52	24		

Catalogue 8:

	Session 1	Session 2	Session 3
$y_{i.}$	64.04	65.70	63.12
s_i^2	4.24	2.20	2.97
J_i	5	4	7

Table 3.6

One-Way ANOVA For Within Sale Price Variation:

Type 48 Cat 8 1969/70.

Source	SS	dof	MS	F
Between Sessions	16.92	2	8.46	2.66
Within Sessions	41.38	13	3.18	
Total	58.30	15		

(3.4) Further Analysis.

The following is a summary of the F-statistics obtained from the various one-way layouts:

	F-Statistic
Cat 1	4.72*
Cat 2	1.13
Cat 4	.53
Cat 5	.43
Cat 8	2.66

Only during the first catalogue was there a difference between the three selling periods. This was however only significant at the 5% level. The fact that the null hypothesis of no difference between the three sessions of a day was accepted for four catalogues indicate that such a difference does not exist for the type in question. To test the assumption of equality of variance we can use Bartlett's test for samples of unequal sizes.

	M	C	M/C
Cat 1	4.7256	1.0515	4.49
Cat 2	.1667	1.0329	.16
Cat 4	.1418	1.0469	.14
Cat 5	2.145	1.065	2.01
Cat 8	.3749	1.1124	.34

With $\chi_2^2(.05) = 5.99$ we can conclude that the intra period variances were the same in all cases.

(3.5) Conclusions.

Based on this data it seems that there is no relationship between the time of sale within a sale day and the realized price for a type 48. The F-statistic being less than one in two cases indicate a higher variation in prices within sale periods than between sale periods. This means that other factors not provided for in the model mostly contribute to the variation in prices during the sale. It is well known that the price paid for "identical" lots sold at one auction sale can vary over the sale period. This variation arises because factors other than the technical characteristics of wool determine the price paid at auction. These factors could be:

(1) Variations in demand throughout a sale period. These variations occur when buyers fill orders and then withdraw from the sale, or if buyers employ a strategy that requires a temporary withdrawal from the market. In other cases a buyer may withdraw from bidding simply to balance his purchases against orders.

(2) Differences in price limits between buyers. These differences reflects variations in the competitive position of individual buyers. In a commodity market variations in price limits for specified grades will arise from differences in the efficiency of the buyer; in processing or retailing the commodity, differences in the stocks held and the level of orders. In a large market involving day to day transactions with effective methods of communica-

ting market intelligence it is unlikely that price limits would vary over a large range. Variations in the price limits will however also arise from differences in the time available for individual buyers to fill orders.

(3) Errors in specification. These frequently occur when the composite properties of a grade are estimated subjectively. Such a situation is common in auction markets for primary products such as livestock, cotton, tobacco or wool. If all buyers in such a market have identical price limits, then the successful bidder will be the buyer whose estimate of value contains the largest possible error. The difference between the buyers' estimate of value and the price paid will vary according to the number of buyers in the market and the distribution of valuations.

These constraints were placed on a simulated model of an auction market by R.B. Whan and R.A. Richardson (61) from which they found that there are on the average eleven bidders usually bidding at wool auction sales in Australia. The model also indicated that an auction held with less than four bidders does not provide enough competition to force buyers to pay their predetermined valuation.

Chapter IV

The Two-way Layout : One Observation Per Cell

No Interaction

(4.1) Method of Analysis (The model).

The two-way layout where there is only one observation per cell is a case frequently occuring in practice. In order to get exact tests and confidence intervals concerning the main effects it is generally necessary with the fixed-effects model to assume that there are no interactions, an assumption which we make in this chapter. The two-way layout with one observation per cell where we do not make the assumption of'no interactions' is treated in the next chapter where a test of the hypothesis of no interaction will be presented. Henceforth the parameterization introduced in the second remark of section 2.1 will be used.

In a two-factor experiment we assume that factor A can operate on I levels and factor B on J levels. The observations can be arranged in an IxJ table with y_{ij} denoting the observation on the "i,j treatment combination" where factor A is at the i th level and B at the j th, i=1,..., I; j=1,..., J. Without the assumption of ' no interaction ' the observation y_{ij} can be considered to consist of the following:

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

where

(1) μ is the average of the factors taken over all their levels;

(2) α_i is the main effect of the i th level of factor A in the sense that it is the excess of the mean for the i th level

of A over the general mean μ irrespective of the level of B,

(3) β_j is the main effect of the j th level of factor B in the sense that it is the excess of the mean for the j th level of B over the general mean μ irrespective of the level of A.

(4) γ_{ij} is the interaction effect of the i th level of A with the j th level of B, i.e., that portion of their combined effect after we have corrected for the main effect of factor A, the main effect of factor B and the total average effect and originates because of this particular level combination of the factors.

DEFN. A case of no interaction is called a case of additivity of effects i.e., $\gamma_{ij} = 0$ $i=1,\ldots, I$; $j=1,\ldots, J$.

If we denote the "true" mean of the i,j cell by η_{ij}, we have under the assumption of additivity that $\eta_{ij} = \mu + \alpha_i + \beta_j$ where $\alpha_\cdot = \beta_\cdot = 0$ so that our Ω assumptions now become:

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

$$\Omega : \alpha_\cdot = \beta_\cdot = 0,$$

$\{e_{ij}\}$ are independently $N(0,\sigma^2)$

H_A : all $\alpha_i = 0$; H_B : all $\beta_j = 0$.

H_A states that the means $\{\mu+\alpha_i\}$ for the different levels of A are all equal. Likewise for H_B. The rank of X' of the general assumption $\eta=E(Y) = X'\beta$ is equal to I+J-1 since the vector η is determined by I+J+1 parameters $\{\mu,\alpha_i, \beta_j\}$ subject to the two linearly independent side conditions $\Sigma_i \alpha_i = 0$ and $\Sigma_j \beta_j = 0$.

The SS to be minimized under Ω is

$$S(Y,\beta) = \sum_{ij}\sum (y_{ij} - \mu - \alpha_i - \beta_j)^2.$$

Equating to zero

$$\frac{\partial S(Y,\beta)}{\partial \mu} = -2 \sum_{ij}\sum (y_{ij} - \mu - \alpha_i - \beta_j),$$

and using $\alpha_. = \beta_. = 0$, we find

$$\hat{\mu} = (\Sigma y_{ij})/IJ = y_{..} \qquad (4.1)$$

Also, equating to zero

$$\frac{\partial S(Y,\beta)}{\partial \alpha_i} = -2 \sum_{j} (y_{ij} - \mu - \alpha_i - \beta_j)$$

we have $\hat{\mu} + \hat{\alpha}_i = y_{i.}$ so that

$$\hat{\alpha}_i = y_{i.} - y_{..} \qquad (4.2)$$

Similarly

$$\hat{\beta}_j = y_{.j} - y_{..} \qquad (4.3)$$

The error $SS = S_\Omega = SS_e$

$$SS_e = \sum_{ij}\sum (y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2$$

$$= \sum_{ij}\sum (y_{ij} - y_{i.} - y_{.j} + y_{..})^2. \qquad (4.4)$$

The degrees of freedom for SS_e is:

$$n-r = IJ - (I+J-1)$$

$$= (I-1)(J-1)$$

Under $\omega = \Omega \cap H_A$; H_A all $\alpha_i = 0$, we must minimize

$$S = \sum_{ij}\sum (y_{ij} - \mu - \beta_j)^2.$$

On equating to zero the partial derivatives we find that the LS estimates $\hat{\mu}_\omega$ and $\hat{\beta}_{j,\omega}$ have the same values as under Ω:

$\partial S/\partial \mu$ same as before

$$\partial S/\partial \beta_v = -2 \sum_{i} (y_{ij} - \mu - \beta_v) = 0$$

i.e. $$I\beta_v = \sum_i y_{iv} - I\, y_{..} \qquad \text{since } \hat{\mu} = y_{..}$$

Hence $\hat{\beta}_v = y_{.j} - y_{..}$

We will denote the SS_H for testing H_A : all $\alpha_i = 0$ by SS_A.

Hence

$$SS_A = ||\hat{\eta} - \hat{\eta}_\omega||^2$$

$$= \sum_{ij} (\hat{\mu}+\hat{\alpha}_i+\hat{\beta}_j - \hat{\mu}_\omega - \hat{\alpha}_{i,\omega} - \hat{\beta}_{j,\omega})^2$$

$$= \sum_{ij} \hat{\alpha}_i^2 \quad \text{since } \hat{\alpha}_{i,\omega} = 0,\ \hat{\beta}_{j,\omega} = \hat{\beta}_j,\ \hat{\mu}_\omega = \hat{\mu}$$

$$= J \sum_i \hat{\alpha}_i^2$$

$$= J \sum_i (y_{i.} - y_{..})^2$$

$$= J \sum_i y_{i.}^2 - IJ\, y_{..}^2 \qquad (4.5)$$

Since H_A states that I-1 linearly independent estimable functions are zero, the degrees of freedom for SS_A is (I-1). The F-test of H_A at significance level α consists in rejecting H_A if

$$MS_A/MS_e > F_{\alpha,(I-1),(I-1)(J-1)}$$

where $MS_A = SS_A/(I-1)$ and $MS_e = SS_e/(I-1)(J-1)$. We can similarly derive a test for H_B which is: reject H_B at level α if

$$MS_B/MS_e > F_{\alpha,\ J-1,(I-1)(J-1)}$$

where

$$MS_B = SS_B/(J-1) \text{ and}$$

$$SS_B = I \sum_j (y_{.j} - y_{..})^2$$

$$= I \sum_j y_{.j}^2 - IJy_{..}^2 \qquad (4.6)$$

with J-1 degrees of freedom.

The total SS is calculated as

$$SS_{tot} = \sum_{ij} (y_{ij} - y_{..})^2$$

$$= \sum_{ij}\sum y_{ij}^2 - IJ\, y_{..}^2 \qquad (4.7)$$

The analysis of variance can now be summarized in the usual ANOVA table.

Table 4.1

Two-way ANOVA With One Observation Per Cell

('No Interaction' Assumed)

Source	SS	dof	MS	F
Rows	$J \sum_i (y_{i.}-y_{..})^2$	(I-1)	$SS_A/(I-1)$	MS_A/MS_e
Columns	$I \sum_j (y_{.j}-y_{..})^2$	(J-1)	$SS_B/(J-1)$	MS_B/MS_e
Residual	$(\Sigma\Sigma y_{ij}-y_{i.}-y_{.j}+y_{..})^2$	(I-1)(J-1)	$SS_e/(I-1)(J-1)$	
Total	$\Sigma\Sigma(y_{ij}-y_{..})^2$	IJ-1		

REMARKS: (1) To construct the table like Table 4.1, SS_A is usually calculated from (4.5), SS_B from (4.6) and SS_{tot} from (4.7). SS_e is then found by subtraction since

$$SS_e = SS_{tot} - SS_A - SS_B.$$

(2) For a more thorough scrutiny of the data it is good practice to construct an IxJ table in which the i,j entry is

$$\hat{\gamma}_{ij} = y_{ij} - y_{i.} - y_{.j} + y_{..}$$

If there is a relatively large interaction $\hat{\gamma}_{ij}$, it may suggest that the Ω assumptions have somehow been violated.

(3) Usually a further column, the expected mean square, E(MS), is added to the ANOVA table and is calculated by replacing y_{ij} by $E(y_{ij})$ under Ω and adding σ^2 to the result:

$$E(MS_A) = \sigma^2 + \frac{J}{(I-1)} \sum_i \left[E(y_{i.} - y_{..})\right]^2$$

$$= \sigma^2 + J \sum_i \left[E(\hat{\alpha}_i)\right]^2/(I-1)$$

$$= \sigma^2 + J \sum \alpha_i^2/(I-1)$$

$$= \sigma^2 + J \sigma_A^2$$

where σ_A^2 is NOT a variance but a convenient notation for $(\Sigma\alpha_i^2)/(I-1)$

Similarly $E(MS_B) = \sigma^2 + I \sigma_B^2$ so that H_A and H_B may be expressed as $H_A : \sigma_A^2 = 0$ and $H_B : \sigma_B^2 = 0$.

(4.2) Description of Data.

The reserve price scheme of the South African Wool Commission necessitates the appraisement of type and yield for every lot of wool offered for sale in South Africa. These in conjunction with a barome of reserve prices for all types, determine the reserve prices on the various lots offered. The reserve prices are not made known to the trade.

The average of the differences between market price and reserve price for a given type and sale, expressed as a percentage, measures to some extent the degree of protection given to the type. A combined weighted average of these differences for all types, expressed as the total percentage difference between

market and reserve prices, gives a measure of the closeness of market and reserve price levels. This measure is a useful guideline for determining future reserve price levels since there is every reason to believe that there exists a consistent relationship between the total percentage variation between market and reserve price and the total percentage of all wool bought by the SAWC.

In this chapter we analize the total percentage difference between market and reserve prices for the various ports over 18 sales. The two-way analysis consists of I=4 ports and J=18 sales. The sales are ordered from 1 to 18 over the selling season so that this model allows an estimate of the effect of time of sale on the percentage price difference. Further, with this model it is possible to make some estimate of the effect of the various ports on this percentage difference. Market and reserve price data for all types offered during 18 sales held in the four South African selling centres during the (current) 1970/1971 season were converted to total percentage differences between market and reserve prices.

(4.3) Results.

A summary of the results are given below. $\beta_1,\ldots,\beta_{18}$ refer to the 18 consecutive sales while $\alpha_1,\alpha_2,\alpha_3,\alpha_4$ refer to the four ports as follows:

α_1 = Cape Town (CT)

α_2 = Durban (DBN)

α_3 = East London (EL)

x_4 = Port Elizabeth (PE)

Sale	$y_{.j}$	$\hat{\beta}_j$	Port	$y_{i.}$	$\hat{\alpha}_i$
1: 15/9/70	3.21	-2.18	CT	5.98	0.59
2: 22/9/70	4.66	-0.73	DBN	7.32	1.93
3: 29/9/70	6.66	1.27	EL	4.31	-1.08
4: 6/10/70	5.16	-0.23	PE	3.95	-1.44
5: 13/10/70	5.09	-0.30			
6: 20/10/70	3.63	-1.76			
7: 27/10/70	5.99	0.60	$\hat{\mu} = y_{..} = 5.39$		
8: 3/11/70	11.91	6.52			
9: 10/11/70	10.45	5.06			
10: 17/11/70	8.76	3.37			
11: 22/11/70	5.65	0.26			
12: 1/12/70	3.56	-1.83			
13: 8/12/70	3.23	-2.16			
14: 19/1/71	3.34	-2.05			
15: 11/2/71	3.72	-1.67			
16: 29/2/71	4.34	-1.05			
17: 9/3/71	4.16	-1.23			
18: 23/3/71	3.53	-1.86			

where $\hat{\mu}$, $\hat{\alpha}_i$, $\hat{\beta}_j$, were calculated from (4.1), (4.2), (4.3) respectively.

Table 4.2

Two-Way ANOVA For Total % Difference

Between Market And Reserve Prices Season 1970/1971.

Source	SS	dof	MS	F
Ports	131.58	3	43.86	6.37**
Sales	444.59	17	26.15	3.80**
Residual	351.29	51	6.89	
Total	927.46	71		

From the analysis set out in the table the total variation in percentage differences between market and reserve prices has been resolved into components owing to ports and sales. The remaining residual variation represents that portion of variation that cannot be accounted for by the present model.

Since we made the additivity assumption we have to look at the estimates $\hat{\gamma}_{ij}$ for relatively large interactions. Port names will be abbreviated and sales will consecutively be numbered 1 to 18.

Summary of interactions: $\hat{\gamma}_{ij} = y_{ij} - y_{i.} - y_{.j} + y_{..}$

Sale	CT	DBN	EL	PE
1	3.18	-2.56	-0.56	-0.06
2	3.94	-2.84	-0.65	-0.44
3	3.59	-3.27	0.82	-1.14
4	1.52	-1.44	0.95	-1.03
5	-0.84	0.42	0.90	-0.49
6	-0.95	1.50	-0.15	-0.39
7	-2.20	4.61	-2.17	-0.24
8	-1.51	5.80	-2.86	-1.43
9	-3.99	9.47	-2.60	-2.87
10	0.36	0.12	2.34	-2.82
11	-0.39	1.77	-0.61	-0.76
12	-0.52	-0.79	0.28	1.03
13	-0.37	-1.13	0.26	1.23
14	0.52	-2.73	0.40	1.79
15	-0.08	-1.75	0.39	1.43
16	-1.28	-2.17	1.34	2.12
17	-0.54	-2.70	0.90	2.35
18	-0.46	-2.30	1.00	1.76

(4.4) Further Analysis (Multiple Comparisons).

Although it does not seem from the estimates $\hat{\gamma}_{ij}$ that all interactions are zero, we nevertheless make the assumption and continue with multiple comparisons between row and column contrasts.

(a) Row contrasts : S-Method.

Let ψ be any linear function of the $\{\alpha_i\}$, $\psi = \sum_i c_i \alpha_i$.

The set L of all such ψ is the same as the set of all contrasts among the true means $\{A_i = \mu+\alpha_i = \eta_{i\cdot}\}$ since $\sum_i c_i \alpha_i = \sum_i c_i' \eta_{i\cdot}$, where $c_i' = c_i - c.$ and hence $\sum_i c_i' = 0$. Conversely, if $\sum_i c_i = 0$, $\sum_i c_i'\eta_{i\cdot} = \sum_i c_i'\alpha_i$.

Hence the LS estimate of ψ is

$$\hat{\psi} = \sum_i c_i \hat{\alpha}_i .$$

Using (2.4) and (2.5) of chapter 2 where $q = I-1$, $n-r = (I-1)(J-1)$ an interval for the contrast ψ is given by

$$\psi \in \hat{\psi} \pm \left[(I-1)\, F_{\alpha,(I-1),(I-1)(J-1)}\right]^{\frac{1}{2}} \hat{\sigma}_{\hat{\psi}}$$

For pairwise comparisons we will conclude ψ to be significantly different from zero if $|\hat{\psi}| > S\hat{\sigma}_{\hat{\psi}}$ where S is defined by (2.5).

Rearanging we have

$\hat{\alpha}_4$	$\hat{\alpha}_3$	$\hat{\alpha}_1$	$\hat{\alpha}_2$
-1.44	-1.08	.59	1.93

and

$$S\hat{\sigma}_{\hat{\psi}} = 2.6249 \sqrt{2/18} \sqrt{3 . F_{3,51(.05)}}$$
$$= 2.54 \rightarrow$$

Thus only:

DBN > EL; PE

by the S-method.

(b) Row Contrasts: T-Method.

For pairwise comparisons of ψ we have from (2.7) chapter 2 that an interval for the contrast ψ is given by

$$\psi \in \hat{\psi} \pm (1/\sqrt{J})\, q_{I,(I-1)(J-1)}\, S$$

and from which we can conclude ψ to be significantly different from zero if $|\hat{\psi}| > Ts$ where T is defined by 2.7. Thus

$$Ts = (1/\sqrt{18})\, q_{4,51}(.05)(2.6249)$$
$$= 2.33 \rightarrow$$

Thus only

DBN > EL; PE

by the T-method. Although the significant comparisons by the T-method is the same as that of the S-method, we note that the T-method gives an interval which is shorter by .21 on either side.

(c) Column Contrasts: S-Method.

For pairwise comparisons we have

$$S\hat{\sigma}_{\hat{\psi}} = 2.6249\ \sqrt{2/4}\ \sqrt{17.F_{17,51}(.05)}$$
$$= 10.27 \rightarrow$$

so that the S-method indicates no significant pairwise differences.

To test whether the average for November is significantly different from the rest, we have

$$\hat{\psi} = 1/4(\hat{\beta}_8+\hat{\beta}_9+\hat{\beta}_{10}+\hat{\beta}_{11}) - 1/4(\hat{\beta}_1+\ldots+\hat{\beta}_7+\hat{\beta}_{12}+\hat{\beta}_{13}+\ldots+\hat{\beta}_{18})$$
$$= 4.88 \rightarrow$$

with

$$S\hat{\sigma}_{\hat{\psi}} = 2.6249\ (5.5317)\ (.558)$$
$$= 8.10$$

which indicates no significance.

(d) Column Contrasts: T-Method.

For pairwise comparisons we have

$$Ts = (.5)\ q_{18,51}(.05)\ 2.6249$$
$$= 6.84$$

which indicates

$$\beta_8 > \beta_1;\ \beta_2;\ \beta_6;\ \beta_{12};\ \beta_{13};\ \beta_{14};\ \beta_{15};\ \beta_{16};\ \beta_{17};\ \beta_{18}.$$
$$\beta_9 > \beta_1;\ \beta_{12};\ \beta_{13};\ \beta_{14};\ \beta_{18}.$$

The T-method also indicates no significance between the average November variations and that of the rest of the season for which data was analysed.

(4.5) Conclusions.

On average there was a difference of approximately 5% between market and reserve price levels for the period under consideration. The high activity of the SAWC during the current (1970/71) season indicates that the SAWC is required to buy a large percentage of the total wool offered at auction in South Africa. Purchases for the season up to and including March 1971 amounted to 199,546 bales (26.3%) of the total offering while the SAWC had to bid on 68.4% of all wools offered. Unfortunately comparative figures for previous seasons during which the SAWC was required to be less active than at present, is not available.

The fact that the percentage variation in Durban was significantly higher than that of Port Elizabeth and East London can be ascribed to either one of two reasons (or a combination of both).

Underestimation:

Although we found in chapter II that Durban, compared to East London, is inclined to overestimate types 48 and 53, the opposite is true for the clip as a whole. Conservative appraisement of the bulk of wool sold at Durban results in lower reserve prices and a higher percentage difference from the market price.

Improvement in the market.

For some reason market prices are higher in Durban than in the other ports. It is not expected that this would be the case for all types offered in Durban but it is possible that the composition of the clip is such that the bulk of the wool offered from this area is in good demand

The contrast $\psi = \alpha_2 - 1/3(\alpha_1+\alpha_3+\alpha_4)$ indicated that the percentage variation in Durban was significantly higher than the average of the other three ports.

It is interesting to note that the highest percentage differences attained occurred during the month of November during which there was a general improvement in the market. This is illustrated by the model through the estimates $\hat{\beta}_j$ which attained their highest values during this period. Negative values of the $\hat{\beta}_j$ before and after November illustrate the lower trend in the market for these periods.

From the summary of interactions it does not quite seem that the interactions are zero. A test for interactions in the two-way layout with one observation will be presented in the next chapter

where we will test the same data for the validity of the additivity assumption.

Chapter V

The Two-Way Layout: One Observation Per Cell With Interaction

(5.1) Method of Analysis (The Model).

A test for interactions in the two-way layout with one observation per cell was presented by Tukey (52) by partitioning one degree of freedom out of the usual old error SS for non-additivity. Tukey also generalized this procedure to other designs (51).

We first consider the general assumptions

$$\Omega: \begin{cases} y_{ij} = \mu+\alpha_i+\beta_j+\gamma_{ij}+e_{ij} \\ \{e_{ij}\} \text{ are independent } N(0,\sigma^2) \\ \alpha_.=\beta_.=\gamma_{i.}=\gamma_{.j}=0 \text{ for all } i,j. \end{cases}$$

where we would like to test the hypothesis

$$H : \text{all } \gamma_{ij}=0.$$

Because we have more parameters than observations we will have to derive a test of H under Ω' which imposes some restrictions on the $\{\gamma_{ij}\}$. We will assume that the interaction γ_{ij} for a cell is a function of the main effect α_i and β_j for that cell:

$$\gamma_{ij} = G\alpha_i\beta_j \qquad (5.1)$$

The Ω' assumptions are then

$$\Omega': \begin{cases} y_{ij} = \mu+\alpha_i+\beta_j+G\alpha_i\beta_j+e_{ij} \\ \{e_{ij}\} \text{ are independent } N(0,\sigma^2) \\ \alpha_.=\beta_.=0 \end{cases}$$

To derive LS estimates of the parameters $\mu, \{\alpha_i\}$, $\{\beta_j\}$ and G under Ω' we pretend for the moment that the $\{\alpha_i\}$ and $\{\beta_j\}$ are

known. Under this fiction a LS estimate $\tilde{G}$ of G is obtained by minimizing

$$S = \sum_{ij}\sum (y_{ij} - \mu - \alpha_i - \beta_j - G\alpha_i\beta_j)^2.$$

Equating to zero

$$\delta S/\delta G = -2 \sum_{ij}\sum \alpha_i\beta_j(y_{ij} - \mu - \alpha_i - \beta_j - G\alpha_i\beta_j)$$

gives

$$\tilde{G} = \frac{\sum_{ij}\sum \alpha_i\beta_j y_{ij}}{\sum_i \alpha_i^2 \sum_j \beta_j^2}$$

since

$$\sum_{ij}\sum \alpha_i\beta_j = \sum_{ij}\sum \alpha_i^2\beta_j = \sum_{ij}\sum \alpha_i\beta_j^2 = 0.$$

Let $\tilde{\gamma}_{ij} = \tilde{G}\,\alpha_i\beta_j$. Then

$$\sum_{ij}\sum \tilde{\gamma}_{ij}^2 = \tilde{G}^2 \sum_i \alpha_i^2 \sum_j \beta_j^2$$

$$= \frac{(\sum_{ij}\sum \alpha_i\beta_j y_{ij})^2}{\sum_i \alpha_i^2 \sum_j \beta_j^2}$$

can be considered to be a "SS for interactions". By replacing the assumed known $\{\alpha_i\}$ and $\{\beta_j\}$ by their LS estimates under Ω, we obtain

$$SS_G = \frac{(\sum_{ij}\sum \hat{\alpha}_i\hat{\beta}_j y_{ij})^2}{\sum_i \hat{\alpha}_i^2 \sum_j \hat{\beta}_j^2}. \qquad (5.2)$$

Provided we can find the distribution of SS_G, SS_G can be used to test H by rejecting H for "large" SS_G. In order to derive the distribution of SS_G under $\omega = H \cap \Omega$, we need the following theorem.

THM. I Let L be a t-dimensional space of linear functions of $\{y_1, \ldots, y_n\}$ and let $f_i = \sum_{j=1}^{n} b_{ij}y_j$ $(i=1, \ldots, m)$ be m $(m<t)$ linear functions belonging to L. Provided the $\{f_i\}$ are orthogonal we have under Ω that $f_1^2/c_1, \ldots, f_m^2/c_m$ and $SS_L - \sum_{i=1}^{m} f_i^2/c_i$ are m+1 statistically independent quantities which, after division by σ^2, all have noncentral chi-square distributions with 1, 1,..., 1 and t-m degrees of freedom respectively.

PROOF: Let $z_i = f_i/\sqrt{c_i}$, $i=1, \ldots, m$. Extend the orthonormal basis $\{z_1, \ldots, z_m\}$ to an orthonormal basis $\{z_1, \ldots, z_m, z_{m+1}, \ldots, z_t\}$ for L. Then from

$$f_1^2/c_1 = z_1^2, \ldots, f_m^2/c_m = z_m^2, \text{ and}$$

$$SS_L - \sum_{i=1}^{m} f_i^2/c_i = \sum_{i=1}^{t} z_i^2 - \sum_{i=1}^{m} z_i^2$$

$$= \sum_{i=m+1}^{t} z_i^2$$

the theorem follows.

Q.E.D.

The distribution of SS_G under $\omega = H \cap \Omega$ is derived in the following theorem.

THM. II Under $\omega = H \cap \Omega$, SS_G/σ^2 and SS_{res}/σ^2 are statistically independent and have chi-square distributions with one and IJ-I-J degrees of freedom respectively where SS_G is defined by (5.2), $SS_{int} = \sum\sum_{ij} (y_{ij} - y_{i\cdot} - y_{\cdot j} + y_{\cdot\cdot})^2$, $SS_{res} = SS_{int} - SS_G$ and $\hat{\alpha}_i$ and $\hat{\beta}_j$ are given by

$y_{i.} - y_{..}$ and $y_{.j} - y_{..}$ respectively.

PROOF: Letting $\hat{\gamma}_{ij} = y_{ij} - y_{i.} - y_{.j} + y_{..}$, SS_G can be written

$$SS_G = \frac{(\Sigma_i \Sigma_j \hat{\alpha}_i \hat{\beta}_j \hat{\gamma}_{ij})^2}{\Sigma \hat{\alpha}_i^2 \; \Sigma \hat{\beta}_j^2}$$

The number of linearly independent restrictions imposed by H : all $\gamma_{ij} = 0$ is $(I-1)(J-1)$. Let L be the $(I-1)(J-1)$-dimensional space spanned by the linear forms $\{\hat{\gamma}_{ij}\}$ and consider the linear forms

$$f = \sum_{ij}\sum a_i b_j \hat{\gamma}_{ij} = \sum_{ij}\sum a_i b_j y_{ij} \qquad (5.3)$$

where $\{a_i, b_j\}$ are constants subject to $\Sigma_i a_i = \Sigma_j b_j = 0$ and $\Sigma a_i^2 > 0$ and $\Sigma b_j^2 > 0$. From the previous theorem we have that

$$f^2/\sigma^2(\sum_i a_i^2 \Sigma b_j^2) \text{ and } \{SS_{int} - f^2/(\sum_i a_i^2 \Sigma_j b_j^2)\}/\sigma^2$$

are statistically independent and have chi-square distributions with one and IJ-I-J degrees of freedom respectively. Further, since $E(\gamma_{ij}) = 0$ both chi-square distributions are central. The decomposition of the error space and the estimation space into orthogonal spaces in chapter I can be extended by a method of nested ω's to generate all the orthogonal spaces of the linear forms involved. The following spaces spanned by the four sets of linear forms in the observations $\{y_{ij}\}$ are mutually orthogonal:

Space	Spanned by	Dimension
L_α	$\hat{\alpha}_1, \ldots, \hat{\alpha}_I$	I-1
L_β	$\hat{\beta}_1, \ldots, \hat{\beta}_J$	J-1
L_μ	$\hat{\mu}$	1
L_e	$\{\hat{\gamma}_{ij} = y_{ij} - y_{i\cdot} - y_{\cdot j} + y_{\cdot\cdot}\}$	(I-1)(J-1)

The set $\{\hat{\alpha}_i\}$, $\{\hat{\beta}_j\}$ and $\{\hat{\gamma}_{ij}\}$ are thus statistically independent so that the conditional distribution of the $\{\hat{\gamma}_{ij}\}$, given the $\{\hat{\alpha}_i\}$ and $\{\hat{\beta}_j\}$, is identical with the unconditional distribution of the $\{\hat{\gamma}_{ij}\}$. Let $\{\hat{\alpha}_i\}$ and $\{\hat{\beta}_j\}$ be considered fixed and substitute $a_i = \hat{\alpha}_i$, $b_j = \hat{\beta}_j$ in (5.3). The random variables $\{\hat{\alpha}_2, \ldots, \hat{\alpha}_I, \hat{\beta}_2, \ldots, \hat{\beta}_J\}$ have a joint density and hence the probability that they are all zero is zero. Since $\Sigma_i \hat{\alpha}_i^2 > 0$ and $\Sigma_j \hat{\beta}_j^2$ with probability one, the joint conditional distribution under ω of SS_G/σ^2 and SS_{res}/σ^2, given the $\{\hat{\alpha}_i\}$ and $\{\hat{\beta}_j\}$ is that of two statistically independent chi-square variables with one and IJ-I-J degrees of freedom respectively. This however does not depend on the fixed values of the $\{\hat{\alpha}_i\}$ and the $\{\hat{\beta}_j\}$ so that the unconditional distribution is the same as the conditional which proves the theorem.

Q.E.D.

In the two-way layout with one observation per cell the usual error sum of squares can be partitioned by the preceding theorem into two components SS_G and SS_{res}. A test for interaction is done with the statistic

$$(IJ-I-J)\ SS_G/SS_{res}$$

which has under ω a central F-distribution with 1 and IJ-I-J degrees of freedom.

The test for interactions can be derived from the ANOVA table for no interactions in the following way.

Table 5.1

Two-Way ANOVA With One Observation Per Cell With Interactions.

Source	SS	dof	MS	F
Rows	SS_A	I-1	$SS_A/(I-1)$	
Columns	SS_B	J-1	$SS_B/(J-1)$	
Interactions	SS_G	1	SS_G	$(IJ-I-J)SS_G/SS_{res}$
Residual	SS_{res}	IJ-I-J	$SS_{res}/(IJ-I-J)$	
Total	SS_{tot}	IJ-1		

where

$$SS_A = J \sum_i (y_{i.}-y_{..})^2$$

$$SS_B = I \sum_j (y_{.j}-y_{..})^2$$

$$SS_G = (\sum_{ij} \hat{\alpha}_i \hat{\beta}_j y_{ij})^2 / (\sum_i \hat{\alpha}_i^2 \sum_j \hat{\beta}_j^2)$$

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

$$SS_{res} = SS_{int} - SS_G$$

The test for interactions is:

Reject H : all $\gamma_{ij} = 0$ if

$$(IJ-I-J)\ SS_G/SS_{res} > F_{\alpha,1,IJ-I-J}$$

(5.2) Description Of Data.

The data of this chapter is that analysed in chapter IV where we estimated the effect of time of sale and port on the percentage variation between market and reserve prices. The test derived in section 5.1 will be employed to test the assumption of additivity made in the previous chapter.

(5.3) Results.

From the estimates $\{\hat{\alpha}_i\}$, $\{\hat{\beta}_j\}$ in chapter IV we have

$$\sum_i \hat{\alpha}_i^2 = 7.31$$

$$\sum_j \hat{\beta}_j^2 = 111.12$$

$$\sum_i \hat{\alpha}_i^2 \sum_j \hat{\beta}_j^2 = 812.61$$

$$\sum\sum_{ij} \hat{\alpha}_i \hat{\beta}_j y_{ij} = 307.76$$

$$\left(\sum\sum_{ij} \hat{\alpha}_i \hat{\beta}_j y_{ij}\right)^2 = 94715.91$$

Hence $SS_G = 94715.91/812.61$

$= 116.56$ →

$SS_{res} = SS_{int} - SS_G$

$= 351.29 - 116.56$

$= 234.73$

Also IJ-I-J = 50.

Table 5.2

Two-Way ANOVA For % Variation Between Market And Reserve Prices : 1970/71.

Source	SS	dof	MS	F
Ports	131.58	3	43.86	
Sales	444.59	17	26.15	
Interactions	116.56	1	116.56	24.85**
Residual	234.73	50	4.69	
Total	927.46	71		

(5.4) Further Analysis.

Multiple comparisons for row and column contrasts were made in chapter IV. Since H : all $\gamma_{ij}=0$ is rejected by the F-test we do not try to explore the interactions further statistically.

(5.5) Conclusions.

The interpretation of an ANOVA model is always much simpler if no interaction is present. We will however demonstrate that if the relevant interactions are wrongly assumed to be zero, the expected value of the pooled mean square is inflated by the extra component. This will have the effect of widening the confidence intervals so that, if either variance-ratio is found to be significant, the corresponding effect is real.

We shall now show that in a two-way layout with only one observation per cell in the i,j combination, that if the interaction is assumed to be zero but is really not zero, it vitiates the

test of significance in the sense that the test will not reject the hypothesis as often as the level of significance indicates.

Under Ω the two way layout without interaction was defined by the model

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \quad i=1,\ldots, I;\ j=1,\ldots, J.$$

The error sum of squares divided by σ^2, that is SS_e/σ^2, is distributed as central chi-square $\chi^2_{(I-1)(J-1)}$. If however the model contains a two-factor interaction term and is given by

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ij} \quad i=1,\ldots, I;\ j=1,\ldots, J$$

then $\frac{SS_e}{\sigma^2}$ is distributed as non-central chi-square $\chi'^2_{(I-1)(J-1),\delta}$ where

$$\delta = \Sigma\Sigma\{(\alpha\beta)_{ij}\}^2/2\sigma^2.$$

Hence the quantity

$$u = MS_A/MS_e$$

used to test the hypothesis $\alpha_1 = \alpha_2 = \ldots = \alpha_I$ is not distributed as a central F even if the hypothesis is true. If no interaction is present $E(MS_e) = \sigma^2$. If interaction is present,

$$E(MS_e) = \sigma^2 + \Sigma\Sigma\{(\alpha\beta)_{ij}\}^2/(I-1)(J-1)$$

so that the error term is increased on the average.

We can thus conclude that the presence of interaction makes the rejection of main effects more difficult. The significance of the variance ratios for ports and sales therefore indicate that these effects definitely exist.

Although significant main effects by the F-ratio still indicate true effects, the interpretation of the analysis must somehow be

adjusted. With the assumption of additivity we conclude that

(main effect for Durban) - (main effect for PE) > 0

which implies that percentage differences in Durban are higher than that of Port Elizabeth by the same amount in all sales of the experiment. Without the assumption of additivity our conclusion is that, averaged over the J sales of the experiment, percentage differences in Durban are higher than that of Port Elizabeth; however, it might happen that in a particular sale in the experiment the percentage difference in Port Elizabeth is higher than that of Durban.

Chapter VI

The Two-Way Layout : Equal Cell Numbers With Interaction.

(6.1) Method of Analysis (The Model).

We will now generalize the two-way layout with interaction to a model with K observations in the i,j cell. A model with equal number of observations per cell is called a complete layout in contrast to the incomplete layout which has unequal numbers of observations per cell. A two-way layout with unequal cell numbers and which occurs frequently in practice will be treated in chapter VII. We shall assume about K that K>1.

Let y_{ijk} denote the k th observation in the i,j cell. Then

$$\Omega : \begin{cases} y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \\ \{e_{ijk}\} \text{ are independent } N(0,1) \\ k = 1, \ldots, K \\ \sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0 \end{cases}$$

Under Ω we must minimize

$$S = \sum_{ijk}\sum\sum (y_{ijk} - \mu - \alpha_i - \beta_j - \gamma_{ij})^2 \qquad (6.1)$$

Let $E(y_{ij}) = \eta_{ij}$. Then

$$\eta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij};$$

so that S can be written as

$$S = \sum_{ijk}\sum\sum (y_{ijk} - \eta_{ij})^2. \qquad (6.2)$$

By equating to zero the partial derivatives with respect to η_{ij}

we have the LS. estimates

$$\hat{\eta}_{ij} = y_{ij\cdot}$$

so that the error SS, the minimum of (6.2), is

$$SS_e = \sum_{ijk}\sum\sum (y_{ijk} - y_{ij\cdot})^2$$

with n-r = IJK-IJ = IJ(K-1) degrees of freedom.

Because $\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0$, the general mean, main effects and interactions are uniquely determined by the $\{\eta_{ij}\}$, i.e.,

$$\mu=\eta_{\cdot\cdot};\ \alpha_i=\eta_{i\cdot}-\eta_{\cdot\cdot};\ \beta_j=\eta_{\cdot j}-\eta_{\cdot\cdot};\ \gamma_{ij}=\eta_{ij}-\eta_{i\cdot}-\eta_{\cdot j}+\eta_{\cdot\cdot}.$$

It is easily seen that $\rho(X') = IJ$ so that all linear functions of the parameters $\{\eta_{ij}\}$ are estimable. By the Gauss-Markoff theorem the L.S. estimates under Ω can thus be obtained by replacing the $\hat{\eta}_{ij}$ by $y_{ij\cdot}$.

We thus have the L.S. estimates

$$\hat{\mu} = y_{\cdot\cdot\cdot}$$

$$\hat{\alpha}_i = y_{i\cdot\cdot}$$

$$\hat{\beta}_j = y_{\cdot j\cdot}$$

$$\hat{\gamma}_{ij} = y_{ij\cdot} - y_{i\cdot\cdot} - y_{\cdot j\cdot} + y_{\cdot\cdot\cdot}$$

The hypotheses of interest are

$$H_A : \text{all } \alpha_i = 0$$

$$H_B : \text{all } \beta_j = 0$$

$$H_{AB} : \text{all } \gamma_{ij} = 0$$

If we substitute the identity

$$y_{ijk}-\mu-\alpha_i-\beta_j-\gamma_{ij} = (y_{ijk}-\hat{\mu}-\hat{\alpha}_i-\hat{\beta}_j-\hat{\gamma}_{ij}) + (\hat{\mu}-\mu) + (\hat{\alpha}_i-\alpha_i) + (\hat{\beta}_j-\beta_j) + (\hat{\gamma}_{ij}-\gamma_{ij})$$

in (6.1) and preserve the parentheses on squaring and summing over i, j, k, we have that

$$S=SS_e+IJK(\hat{\mu}-\mu)^2+JK\Sigma_i(\hat{\alpha}_i-\alpha_i)^2+IK\Sigma_j(\hat{\beta}_j-\beta_j)^2+K\Sigma\Sigma_{ij}(\hat{\gamma}_{ij}-\gamma_{ij})^2 \tag{6.3}$$

It is clear from this expression that, except for the parameters which are zero under H_A, H_B and H_{AB} respectively, the L.S. estimates under these hypotheses are the same as under Ω. Under H_A (6.3) becomes

$$S=SS_e+IJK(\hat{\mu}-\mu)^2+JK\Sigma_i\hat{\alpha}_i^2+IK\Sigma_j(\hat{\beta}_j-\beta_j)^2+K\Sigma\Sigma_{ij}(\hat{\gamma}_{ij}-\gamma_{ij})^2$$

which is minimized by the values $\mu=\hat{\mu}$, $\beta_j=\hat{\beta}_j$ and $\gamma_{ij}=\hat{\gamma}_{ij}$. Hence the minimum value of H_A is

$$S_{\omega_A} = SS_e + JK\Sigma\hat{\alpha}_i^2.$$

For testing H_A the numerator SS of F is thus

$$SS_A = JK\Sigma_i\hat{\alpha}_i^2.$$

and similarly for testing H_B and H_{AB}

$$SS_B = IK\Sigma_j\hat{\beta}_j^2.$$

$$SS_{AB} = K\Sigma\Sigma_{ij}\hat{\gamma}_{ij}^2.$$

while the denominator is SS_e in each case.

The number of linearly independent side conditions imposed by H_A and H_B are (I-1) and (J-1) respectively so that the d.o.f. for SS_A is (I-1) and the d.o.f. for SS_B is (J-1). The number of estimable restrictions imposed by H_{AB} is IJ. However, in an IxJ table of the $\{\gamma_{ij}\}$ we see that, if all $\gamma_{ij}=0$ in the sub-table

by deleting the last row and column, then $\gamma_{ij}=0$ in the whole table since row and column sums must zero respectively. This suggests that H_{AB} imposes $(I-1)(J-1)$ linearly independent restrictions so that the d.o.f. for SS_{AB} is $(I-1)(J-1)$.

If we calculate the expected values of MS_A, MS_B and MS_{AB} (follow the procedure outlined in remark three section (4.1)), the resulting formulas suggest that we introduce the notations

$$\sigma_A^2 = \sum_i \alpha_i^2/(I-1)$$

$$\sigma_B^2 = \sum_j \beta_j^2/(J-1)$$

$$\sigma_{AB}^2 = \sum\sum_{ij} \gamma_{ij}^2/(I-1)(J-1)$$

where again the symbols σ_A^2, σ_B^2 and σ_{AB}^2 are not variances but convenient abbreviations for the functions of parameters.

The results are summarized in the following ANOVA table.

Table 6.1

Two-Way ANOVA With K Observations Per Cell.

Source	SS	dof	F
Rows	$JK\sum_i (y_{i..}-y_{...})^2$	$(I-1)$	MS_A/MS_e
Columns	$IK\sum_j (y_{.j.}-y_{...})^2$	$(J-1)$	MS_B/MS_e
Interactions	$K\sum\sum_{ij} (y_{ij.}-y_{i..}-y_{.j.}+y_{...})^2$	$(I-1)(J-1)$	MS_{AB}/MS_e
Error	$\sum\sum\sum_{ijk} (y_{ijk}-y_{ij.})^2$	$IJ(K-1)$	
Total	$\sum\sum\sum (y_{ijk}-y_{...})^2$	$IJK-1$	

where MS_A, MS_B, MS_{AB} and MS_e are obtained by deviding SS_A, SS_B, SS_{AB} and SS_e by their respective d.o.f.,i.e.

(I-1), (J-1), (I-1)(J-1) and IJ(K-1).

REMARKS: (1) If we let $C=IJKy_{...}^2$, the SS for main effects can be calculated as

$$SS_A = JK\sum_i y_{i..}^2 - C$$

$$SS_B = IK\sum_j y_{.j.}^2 - C.$$

(2) If we calculate a SS for "cell means about the grand mean"

$$SS_{cells} = K\sum\sum_{ij} y_{ij.}^2 - C.$$

Then the SS for interaction is found by subtraction

$$SS_{AB} = SS_{cells} - SS_A - SS_B.$$

(3) The total SS about the grand mean is calculated as usual from

$$SS_{tot} = \sum\sum\sum y_{ijk}^2 - C.$$

(4) The error SS can now be found by subtraction

$$SS_e = SS_{tot} - SS_A - SS_B - SS_{AB}$$

$$= SS_{tot} - SS_{cells}.$$

(5) Because of rounding errors it is more accurate to work with sums rather than means so that we will use the following "+" notation:

Replacing a subscript by a "+" will indicate summation over that subscript. Hence

$$y_{+++} = \sum_{ijk}\sum\sum y_{ijk}; \quad y_{...} = y_{+++}/IJK.$$

$$y_{i++} = \sum_{jk}\sum\sum y_{ijk}; \quad y_{i..} = y_{i++}/JK.$$

$$y_{+j+} = \sum_{ik}\sum\sum y_{ijk}; \quad y_{.j.} = y_{+j+}/IK.$$

With this very convenient notation the various SS can be expressed in terms of sums of observations rather than means.

Let $C = IJKy_{...}^{2} = (y_{+++})^{2}/IJK.$

Then

$$SS_{A} = \sum_{i}(y_{i++})^{2}/IK - C$$

$$SS_{B} = \sum_{j}(y_{+j+})^{2}/JK - C$$

$$SS_{cells} = \sum_{ij}\sum(y_{ij+})^{2}/K - C$$

$$SS_{tot} = \sum_{ijk}\sum\sum y_{ijk}^{2} - C$$

(6.2) Description of Data.

The South African Woolclip can be categorized into the following six broad classes:

Merino wool : Merino wool, spinning count 60^{s} and over, and overstrong (extra strong Merino wool, spinning count average (bulk 58^{s}), mean classes of wool consisting of all white wool, which are by nature free from kemp fibres or hair and is derived from sheep showing the typical characteristics of the Merino in their wool.

Karakul wool : Karakul wool means a class of wool consisting of all wool derived from sheep showing the typical characteristics of the Karakul in their wool.

Crossbred wool : Crossbred wool means a class of wool consisting of all white wool which is free from kemp fibres or hair, has a spinning count under 60^{s}, and is derived from crossbred woolled sheep and from woolled breeds other than the Merino.

Coarse and coloured wool : Coarse white wool and coarse and coloured wool mean classes of wool which include all wool containing by nature kemp fibres or hair and/or coloured wool or hair fibres.

Native wool : Wools produced in the native territories i.e., Transkei, Ciskei, Basutoland and Pondoland.

Dead wool : Wools taken from sheep that have died from natural and/or other causes but not slaughtered.

The significance of percentage variations between market and reserve prices between the four ports Cape Town, Durban, East London and Port Elizabeth suggest a further brakedown of these percentage differences in the main wool classes. Because of uneven distribution and lack of offering in all ports, Karakul

and Native wools have been excluded from the analysis.

The data again consists of percentage differences between market and reserve prices during the beginning of the 1970/71 season. Whereas in chapter IV and V we considered the percentage variation for the wool clip as a whole, we now make a breakdown of the clip into the four classes - Merino, Coarse and Coloured, Crossbred and Dead wool and consider the percentage variation for each class in the various ports. Due to the fact that only auctions in which all ports and classes were represented were taken into consideration, sales have not been included as a factor in the analysis. The four levels for row effects represent the four ports in the now familiar sequence Cape Town, Durban East London and Port Elizabeth. Four column effects represent the classes of wool:merino, crossbred, coarse and coloured and dead wool. Thirteen "observations" were calculated for each port and class combination by grouping all types within a given class for a given sale and port. Weighted reserve and market prices were computed for each cell combination from which percentage differences were then calculated. Hence we have

$$I=4,\ J=4,\ K=13 \text{ and } \rho(X') = 16.$$

We will use the usual abbreviations for the four ports and the following for the various classes of wool.

Cr.Bred = Cross/Bred Wools.

C & C = Coarse and Coloured Wools.

D/Wool = Dead Wools.

(6.3) Results.

A summary of cell means bordered by row totals and means, and column totals and means are given in the following table.

Table 6.2

	Merino	Cr.Bred	C & C	D/Wool	y_{i++}	$y_{i..}$
CT	6.06	13.92	14.38	47.43	1063.19	20.45
DBN	6.40	11.32	40.35	21.95	1040.17	20.00
EL	4.03	10.49	20.54	16.92	675.71	12.99
PE	3.16	13.27	14.55	29.68	788.56	15.17
y_{+j+}	255.31	636.99	1167.64	1507.69		
$y_{.j.}$	4.91	12.25	22.46	28.99		
				y_{+++}	3567.63	
				$y_{...}$	17.15	

REMARK: (1) The row means are not comparable with that obtained in the previous two chapters for the following two reasons. (a) The same sales were not taken into consideration, (b) the row means in table (6.2) are unweighted means of the four classes of wool whereas in chapter IV all classes (types) were combined to form weighted reserve and market prices from which percentage differences were calculated.

Row and column effects are given by the L.S. estimates

$$\hat{\alpha}_1 = y_{1..} - y_{...} = 3.30$$
$$\hat{\alpha}_2 = y_{2..} - y_{...} = 2.85$$
$$\hat{\alpha}_3 = y_{3..} - y_{...} = -4.16$$
$$\hat{\alpha}_4 = y_{4..} - y_{...} = -1.98$$

$$\hat{\beta}_1 = y_{.1.} - y_{...} = -12.24$$
$$\hat{\beta}_2 = y_{.2.} - y_{...} = -4.90$$
$$\hat{\beta}_3 = y_{.3.} - y_{...} = +5.31$$
$$\hat{\beta}_4 = y_{.4.} - y_{...} = +11.84$$

while interactions, given by $\hat{\gamma}_{ij} = y_{ij.} - y_{i..} - y_{.j.} + y_{..}$, are summarized in the following table:

Table 6.3

	Merino	Cr.Bred	C & C	D/Wool
CT	-2.15	-1.63	-11.38	+15.14
DBN	-1.30	-3.78	+15.04	- 9.89
EL	3.28	+2.40	+ 2.24	- 7.91
PE	.23	+3.00	- 5.93	+2.67

Further we have

$$\sum_{ijk}\sum\sum (y_{ijk})^2 = 125,392.91$$

$$\sum_{ij}\sum (y_{ij+})^2/K = 92,020.94$$

$$\sum_{i} (y_{i++})^2/JK = 62,821.84$$

$$\sum_{j} (y_{+j+})^2/IK = 78,989.46$$

$(y_{+++})^2/IJK$ = 61,192.23 = C

and $IJ(K-1)$ = 192

The ANOVA table is given as table 6.4.

Table 6.4

Two-Way ANOVA Of % Differences Between Market and Reserve Prices For Main Wool Classes : 1970/71.

Source	SS	dof	MS	F
Ports	1,629.61	3	543.21	3.13*
Wool class	17,797.23	3	5932.41	34.13**
Interaction	11,401.87	9	1266.87	7.29**
Error	33,371.97	192	173.81	
Total	64,200.68	207		

(6.4) Further Analysis (Multiple Comparisons).

We are mainly interested in pairwise column contrasts and will derive the T and S intervals. For the S-method we have

$$(S\hat{\sigma}_{\hat{\psi}})^2 = (J-1)\, F_{(J-1),IJ(K-1)}(.05)\, s^2(\Sigma c_i^2/JK)$$

$$= 3\, F_{3,192}(.05)\ \ 173.81\ \ (2/52)$$

$$= 52.74$$

$$S\hat{\sigma}_{\hat{\psi}} = 7.26$$

so that by the S-method, the following pairwise row contrasts are significantly different from zero.

D/Wool > Cr. Bred ; Merino

C & C > Cr. Bred ; Merino

Cr. Bred > Merino

By the T method we have

$$Ts = (q_{J,IJ(K-1)}(.05)\sqrt{173.81})/\sqrt{IK}$$

$$= 3.63\ (13.18)/7.21$$

$$= 6.64$$

from which it is obvious that the same column contrasts are significantly different from zero as given by the S-method. Note however that the T-method gives shorter intervals.

For row contrasts it is immediately clear that the following contrasts are different from zero by the T-method.

CT > EL

DBN > EL

REMARK: Since we reject H_{AB} we could explore the interactions further statistically, as can be done with any F-test in fixed effect models, by the S-method. The q for the S-method applied to the whole space of interactions spanned by the $\{\gamma_{ij}\}$ would be the number of degrees of freedom for SS_{AB}, namely $(I-1)(J-1)$. The T-method would not apply since the covariances of the $\{\hat{\gamma}_{ij}\}$ are not equal.

(6.5) Conclusions.

We have stated before that the tests for main effects are valid regardless of the true values of the interactions. We have not

assumed additivity in this chapter so that conclusions about positive contrasts must be that, averaged over the columns (rows) in the experiment, the contrasts hold. It might however happen that in a particular location in the experiment, a contrast does not hold.

From the row contrasts it is evident that, averaged over the four wool classes, variations between market and reserve prices are higher in Cape Town and Durban than in East London. The data does not support the claim that it is higher than in Port Elizabeth as well.

High positive interactions for Dead Wool in Cape Town and coarse and coloured wools in Durban indicate that special forces are at work in these two ports for the two classes of wool. In the absence of any sound reason for a better market for these wools in the two ports, one would be inclined to reason that undervaluation by commission appraisers of these wools exist in the two ports. For Durban however, this assertion proved to be unfounded.

Technical advice revealed that a highly competitive market for coarse and coloured wools exists in Durban. This can be ascribed to the proportionately smaller quantity of coarse and coloured wools offered in Durban as compared to Port Elizabeth and which forms the essence of the wool washeries established in these centres.

In the light of conclusions drawn at the end of chapter I, and no facts to prove the contrary, we can conclude that commission appraisement in Cape Town, especially for Dead wool, is lower than that of the trade and lower than commission appraisement in the other ports.

Comparing column effects and the highly significant F-statistic, the following is evident. All classes of wool do not enjoy the same amount of support by the SAWC through its reserve price scheme. Highest support goes to Merino, followed by Crossbred, Coarse and Coloured and the least to Dead wool. This seems justifiable because, proportionately, this is the sequence of representation in the South African wool clip.

Chapter VII

The Two-Way Layout : Unequal Cell Numbers (and a test for interactions).

(7.1) Method of analysis (The Model).

In many non-experimental studies the investigator classifies his sample according to the factors or variables of interest, exercising no control over the way in which the numbers fall. With a one-way classification, the handling of the "unequal numbers" case was discussed in Chapter III. In this chapter we present a method for analysing a two-way classification. A two-way classification with unequal numbers in subclasses are treated in detail by Graybill(9) although he does not derive a test for interactions. A test of the hypothesis.

$$H_{AB} : E(y_{ijk}) = \mu + \alpha_i + \beta_j$$

is given by Scheffé (21) while an example worked in detail is given by Snedecor and Cochran (25). With unequal sub-class numbers the expression for the expected values of the mean squares in terms of components of variance are complicated. Wilk and Kempthorne (65) have developed formulas for two -and three-factor arrangements. In order to derive the theory underlying the model we need further results and notation.

Suppose that we have two factors A and B that vary in an experiment and that A can take on any of I levels and B any of J levels.

DEFN. Let $\{w_j\}$ be an arbitrary set of numbers with $w_j \geq 0$ and $\Sigma_j w_j = 1$. The mean for the i th level of A is

defined to be

$A_i = \Sigma_j w_j \eta_{ij}$.

where η_{ij} is the "true mean" of the i,j cell.

DEFN. If we let $\{v_{ij}\}$ be an arbitrary set of numbers with $v_{ij} \geq 0$ and $\Sigma_i v_{ij} = 1$, we define the mean for the j th level of B as

$B_j = \Sigma_i v_{ij} \eta_{ij}$.

DEFN. The general mean, denoted by μ, is defined to be

$\mu = \Sigma_j w_j B_j = \Sigma_i v_i A_i = \underset{ij}{\Sigma\Sigma} v_i w_j \eta_{ij}$.

DEFN. The main effect of the i th level of A is

$\alpha_i = A_i - \mu$ where we note that

$\Sigma_i v_i \alpha_i = 0$.

DEFN. The main effect of the j th level of B is

$\beta_j = B_j - \mu$ where $\underset{j}{\Sigma} w_j \beta_j = 0$

DEFN. The interaction of the i th level of A with the j th level of B can now be defined as

$\gamma_{ij} = \eta_{ij} - A_i - B_j + \mu$

where $\underset{i}{\Sigma} v_i \gamma_{ij} = \underset{j}{\Sigma} w_j \gamma_{ij} = 0$ for all i and j.

Hence we have

$\eta_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$.

THM. I If the interactions $\{\gamma_{ij}\}$ are all zero for some system of weights $\{v_i\}$ and $\{w_j\}$, then they are zero for every system of weights. Hence every contrast in the main effects $\{\alpha_i\}$ and $\{\beta_j\}$ has a value that does not depend on the system of weights $\{v_i\}$ and $\{w_j\}$, and the same is

true for contrasts in the means $\{A_i\}$ and $\{B_j\}$ for the levels of A and B respectively.

PROOF: A proof of this theorem can be found in (21) chapter IV.

Although the tests for main effects and the associated multiple comparisons by the S-method remain relatively simple in the case of unequal cell numbers, the test for interactions is more difficult to compute, requiring the solution of m linear equations in m unknowns, where m is one less than the minimum of I and J. The number of observations in cell i,j will be denoted by n_{ij}. In chapter VI all n_{ij} were equal to K, and the general mean, main effects and interactions were defined with equal weights $\{v_i\}$ and $\{w_j\}$. If y_{ijk} denotes the k th observation in the i,j cell, and D denotes the set of all pairs $\{(i,j)\}$ which label the non-empty cells, our Ω assumptions become

$$\Omega : \begin{cases} y_{ijk} = \eta_{ij} + e_{ijk} \\ \{e_{ijk}\} \text{ are independently } N(0,\sigma^2) \\ k = 1,\ldots, n_{ij} \; ; \; (i,j) \; \varepsilon \; D \end{cases}$$

The hypotheses of chief interest concern the main effects and interactions. Under Ω we have to minimize

$$S = \sum_{(i,j)\varepsilon D} \sum_{k=1}^{n_{ij}} (y_{ijk} - \eta_{ij})^2 \qquad (7.1)$$

so that we obtain the LSE

$$\hat{\eta} = y_{ij\cdot} \text{ for } (i,j)\varepsilon D \qquad (7.2)$$

Since the NE have the unique solutions (7.2) the rank r of the matrix X' is equal to p where p is the number of nonempty cells. The error SS, the minimum of (7.1) is thus

$$S_{\Omega} = SS_e = \sum_{(i,j)\in D} \sum_{k=1}^{n_{ij}} (y_{ijk} - y_{ij\cdot})^2 \qquad (7.3)$$

with n-p d.o.f., where n is the number of observations.

We will again use our own "plus" notation in addition to the normal "dot" notation.

$$N_{i+} = \sum_{j=1}^{J} n_{ij}$$

$$N_{+j} = \sum_{i=1}^{I} n_{ij}$$

$$N_{++} = \sum_{ij} n_{ij}$$

$$y_{ij+} = \sum_{k=0}^{n_{ij}} y_{ijk}$$

$$y_{i++} = \sum_{j=1}^{J} \sum_{k=0}^{n_{ij}} y_{ijk}$$

$$y_{+j+} = \sum_{i=1}^{I} \sum_{k=0}^{n_{ij}} y_{ijk}$$

$$y_{+++} = \sum\sum\sum_{ijk} y_{ijk}$$

(7.1.1) Test for Interactions.

The hypothesis H_{AB} of additivity is

$$H_{AB} : E(y_{ijk}) = \mu + \alpha_i + \beta_j.$$

From theorem I, the question whether H_{AB} is true or false does not depend on the weights $\{v_i\}$ and $\{w_j\}$ used to define the general mean μ and the main effects $\{\alpha_i\}$ and $\{\beta_j\}$. Let $\omega = \Omega \cap H_{AB}$. The effect of choosing the weights $\{v_i\}$ and $\{w_j\}$ under ω is only

to impose the side conditions $\Sigma_i v_i \alpha_i = \Sigma_j w_j \beta_j = 0$. We may derive the N.E. under ω first and choose convenient side conditions later. The pattern of nonempty cells in the two-way layout must satisfy a certain condition in order that the parameters μ, $\{\alpha_i\}$, $\{\beta_j\}$ be estimable under ω and the side conditions. To examine the N.E. we shall make the following assumption.

Assumption 7.1: The n_{ij} are values such that $\alpha_i - \alpha_j$ is estimable for every $i \neq j = 1,2,\ldots, I$ and $\beta_i - \beta_j$ is estimable for every $i \neq j = 1,2,\ldots, J$. If the n_{ij} satisfy this condition, we have under ω that there are exactly I+J-1 linearly independent estimable functions. A proof of this statement can be found in (9) chapter 13.

Under ω we must minimize

$$\Sigma\Sigma\Sigma_{ijk}(y_{ijk} - \mu - \alpha_i - \beta_j)^2.$$

Equating to zero the partial derivatives with respect to μ, α_i and β_j, we get the N.E. Throughout this section the L.S.E. will be that obtained under ω so that we will omit the subscripts ω unless they are required to prevent confusion.

$$\mu : N_{++}\hat{\mu} + \Sigma_i N_{i+}\hat{\alpha}_i + \Sigma_j N_{+j}\hat{\beta}_j = y_{+++} \qquad (7.4)$$

$$\alpha_i : N_{i+}\hat{\mu} + N_{i+}\hat{\alpha}_i + \Sigma_j n_{ij}\hat{\beta}_j = y_{i++} \quad i=1,\ldots, I.$$

$$\beta_j : N_{+j}\hat{\mu} + \Sigma_i n_{ij}\hat{\alpha}_i + N_{+j}\hat{\beta}_j = y_{+j+} \quad j=1,\ldots, J.$$

The L.S.E. under ω will not be unique until suitable side conditions are chosen, but the various SS's involving them will never-

theless be unique.

We shall solve the N.E. for contrasts of the α_s. We can use the β_j equation and obtain

$$\hat{\mu}+\hat{\beta}_j = y_{.j.} - (\Sigma_s n_{sj}\hat{\alpha}_s)/N_{+j} \qquad j=1,\ldots,J. \quad (7.5)$$

The α_i equation is

$$\Sigma_j n_{ij}(\hat{\mu}+\hat{\beta}_j) + N_{i+}\hat{\alpha}_i = y_{i++} \qquad i=1,\ldots,I. \quad (7.6)$$

Substituting $\hat{\mu}+\hat{\beta}_j$ from (7.5) into (7.6) gives

$$\Sigma_j n_{ij}(y_{.j.} - \Sigma_s n_{sj}\hat{\alpha}_s/N_{+j}) + N_{i+}\hat{\alpha}_i = y_{i++} \quad i=1,\ldots, I. \quad (7.7)$$

This gives

$$N_{i+}\hat{\alpha}_i - \sum_{s=1}^{I}\sum_{j=1}^{J}(n_{ij}n_{sj}\hat{\alpha}_s)/N_{+j} = y_{i++} - \sum_{j=1}^{J} n_{ij}y_{.j.}$$

$$i=1,\ldots,I. \quad (7.8)$$

If we isolate the quantity involving $\hat{\alpha}_i$ from the second term on the left hand side of (7.8) we get

$$\left[N_{i+} - \sum_{j=1}^{J}\frac{n_{ij}^2}{N_{+j}}\right]\hat{\alpha}_i - \sum_{\substack{s=1\\ s\neq i}}^{I}\frac{n_{ij}n_{sj}}{N_{+j}}\hat{\alpha}_s = q_i \quad i=1,\ldots, I. \quad (7.9)$$

where q_i is given by the right hand side of (7.8). The system (7.9) represents I equations in I unknowns $\hat{\alpha}_s$. They can be represented in matrix form as

$$A\hat{\alpha} = q \quad (7.10)$$

where the IxI matrix $A = (a_{is})$ has elements as follows:

Diagonal elements:

$$a_{ii} = N_{i+} - \sum_{j=1}^{J}\frac{n_{ij}^2}{N_{+j}} \qquad i=1,2,\ldots, I. \quad (7.11)$$

Off-diagonal elements:

$$a_{ir} = -\sum_{j=1}^{J} \frac{n_{ij}n_{rj}}{N_{+j}} \qquad i \neq r = 1,\ldots,I.$$

The Ixl vector $\hat{\alpha}$ has elements $\hat{\alpha}_i$, and q has elements q_i given by

$$q_i = y_{i++} - \sum_{j=1}^{J} n_{ij}y_{\cdot j\cdot} \qquad (7.12)$$

We remark that the elimination of the $\{\hat{\beta}_j\}$ resulted also in the elimination of $\hat{\mu}$. We could also have eliminated the $\{\hat{\alpha}_i\}$ instead of the $\{\hat{\beta}_j\}$ to obtain a set of equations similar to (7.10) i.e.,

$$B\,\hat{\beta} = r \qquad (7.13)$$

where B is a JxJ matrix, $\hat{\beta}$ a Jxl vector and r having elements

$$r_i = y_{+j+} - \sum_{i=1}^{I} n_{ij}y_{i\cdot\cdot}$$

THM.II The rank of the matrix A in (7.10) is (I-1).

PROOF: By assumption (7.1) and the statement following it, there are I-1 linearly independent estimable functions of the α_i. These must come from equation (7.10) so that the rank of A must at least be (I-1).

If the equations of (7.9) are added together, i.e. summed over i=1 to i=I, the result is zero, which shows that there is at least one dependent equation in the system. Therefore, A is IxI with one linearly dependent row; hence it's rank must be at most I-1.

Therefore, it's rank is exactly I-1.

Q.E.D.

Since rank A is (I-1) there are an infinite number of vectors $\hat{\alpha}$ that will satisfy $A\hat{\alpha} = q$. To solve the system any nonestimable side condition can be used. Using the condition $\Sigma \hat{\alpha}_i = 0$ means that we must solve a system of I+1 equations for I+1 unknowns. The work can be reduced by using other nonestimable functions. If we use the side condition $\alpha_I = 0$, this immediately reduces the system to I-1 equations by crossing out the I th equation in (7.10) and setting $\hat{\alpha}_I = 0$.

We note that S_ω can be expressed as

$$S_\omega = \Sigma\Sigma\Sigma_{ijk} y^2_{ijk} - \Sigma\Sigma\Sigma_{ijk} y_{ijk}\hat{\mu} - \Sigma_i y_{i++}\hat{\alpha}_i - \Sigma_j y_{+j+}\hat{\beta}_j . \qquad (7.14)$$

We can eliminate $\{\hat{\beta}_j\}$ from (7.14) by substitution from (7.5) to obtain

$$S_\omega = \Sigma\Sigma\Sigma_{ijk} y^2_{ijk} - \Sigma_i N_{i+}\hat{\alpha}_i - \Sigma_j (y_{+j+})^2/N_{+j} \qquad (7.15)$$

Similarly we can eliminate the $\{\hat{\alpha}_i\}$ to obtain

$$S_\omega = \Sigma\Sigma\Sigma_{ijk} y^2_{ijk} - \Sigma_j N_{+j}\hat{\beta}_j - \Sigma_i (y_{i++})^2/N_{i+} \qquad (7.16)$$

from which we see that $\hat{\mu}$ drops out in both cases.

If we choose the side condition $\hat{\alpha}_I = 0$ with the system of equations (7.10) i.e., $A\hat{\alpha} = q$, we can calculate S_ω by solving the I-1 linear equations in I-1 unknowns $\{\hat{\alpha}_i\}$ by deleting the I th row and column from the matrix A.

Alternatively, if we choose the side condition $\hat{\beta}_J = 0$ with the system of equations (7.13), i.e., $B\hat{\beta} = r$, we can calculate S_ω by solving the J-1 linear equations in the J-1 unknowns $\{\hat{\beta}_j\}$ by deleting the J th row and column from the matrix B. If $I \neq J$,

we choose the method with the fewer number of unknowns.

The numerator of the statistic F for testing H_{AB} is $S_\omega - S_\Omega$ where S_Ω is given by (7.3). From the general theory of chapter I we know that ω constrains the vector of means η to an (r-q)-dimensional subspace. In the present case the parameters in η under ω are those of a two-way layout with additivity, so that ω constrains η to a space of dimension I+J-1. Hence r-q = I+J-1 so that q = p-I-J+1 since r = p. The statistic is thus

$$\frac{n - p}{p-I-J+1} \qquad \frac{S_\omega - S_\Omega}{S_\Omega} \tag{7.18}$$

which under H_{AB} has the F-distribution with p-I-J+1 and n-p degrees of freedom.

REMARKS: (1) If we have no empty cells in the layout, p=IJ so that

$n-p = n-IJ = N_{++} - IJ.$

and

$p-I-J+1 = IJ-I-J+1 = (I-1)(J-1)$

(2) With empty cells present, deduct 1 d.o.f. for each empty cell.

(3) An approximate analysis to replace the tedious exact calculations in the case of unequal cell numbers is given by Scheffé (21), page 362.

(7.1.2) <u>Inferences about main effects, assuming additivity.</u>

The tests for and estimation of main effects depend on whether we incorporate the hypothesis of additivity H_{AB} into the under-

lying assumption. If we do this, the assumptions become the ω defined above. Suppose we wish to test

$$H_A : \text{all } \alpha_i = 0 \quad \text{under } \omega$$

the hypothesis

$$\omega_1 = \omega \cap H_A$$

is the same as the Ω for the one-way layout with J classes and N_{+j} observations in the j th class so that S_ω is the error SS for the one-way layout, i.e.,

$$\begin{aligned} S_{\omega_1} &= \Sigma\Sigma\Sigma_{ijk}(y_{ijk} - y_{\cdot j\cdot})^2 \\ &= \Sigma\Sigma\Sigma_{ijk} y^2_{ijk} - \Sigma_j (y_{+j+})^2 / N_{+j} \\ &= \Sigma\Sigma\Sigma_{ijk} y^2_{ijk} - \Sigma_j N_{+j} y^2_{\cdot j\cdot} \end{aligned} \tag{7.19}$$

To obtain the d.o.f. for the F-statistic for testing H_A under ω, we reason as follows:

We already have that ω constrains η to a space of dimension I+J-1, which corresponds to the rank r of the general theory, where Ω corresponds to the present ω, so that n-r of the general theory becomes here n-I-J+1. The hypothesis H_A imposes I-1 linearly independent estimable restrictions. Hence the statistic

$$\frac{n-I-J+1}{I-1} \qquad \frac{S_{\omega_1} - S_\omega}{S_\omega}$$

has under ω_1 the F-distribution with I-1 and n-I-J+1 d.o.f. The sum of squares for rows, adjusted for columns, can be expressed as Row SS (adjusted) = $\Sigma_i \hat{\alpha}_i q_i$

where $\hat{\alpha}_i$ is any solution to $A\hat{\alpha} = q$. Similarly, the sum of squares for column, adjusted for rows, can be expressed as

$$\text{Column SS (adjusted)} = \sum_j \hat{\beta}_j r_i$$

where $\hat{\beta}_j$ is any solution to $B\hat{\beta} = r$.

If we calculate the SS between sub-classes and the unadjusted SS for rows and columns, these being, respectively

$$\sum_{ij}\sum y^2_{ij+}/n_{ij} - C \; ; \; \sum_i y^2_{i++}/N_{i+} - C \; ; \; \sum_j y^2_{+j+}/N_{+j} - C \; ;$$

where $C = y^2_{+++}/N_{++}$, we can partition the SS between sub-classes into either of the following partitions:

Row SS (unadjusted) + Column SS (adjusted) + Interaction SS

or into

Row SS (adjusted) + Column SS (unadjusted) + Interaction SS.

We can now summarize the results in the following ANOVA table.

Table 7.1

Two-Way ANOVA With Unequal Numbers In The Subclasses.

Source	SS(1)	d.o.f.(2)
Rows(adj)	$\sum_i \hat{\alpha}_i q_i$	(I-1)
Columns(unadj)	$\sum_j y^2_{+j+}/N_{+j} - N_{++}y^2_{...}$	(J-1)
Interactions	$\sum\sum_{ij} n_{ij}y_{ij\cdot}^2 - \sum_j N_{+j}y_{\cdot j\cdot}^2 - \sum_i \hat{\alpha}_i q_i$	(I-1)(J-1)
Error	$\sum\sum\sum_{ijk}(y_{ijk}-y_{ij\cdot})^2$	$N_{++}-(IJ)$
Total	$\sum\sum\sum_{ijk}(y_{ijk}-y_{...})^2$	$N_{++} - 1$

REMARK: Note that the mean squares for the F-test of the main effects must be the adjusted mean squares. An adjusted SS for columns can be found by subtraction:
$SScol(adj) = SS_{cells} - SS_I - SSrow(unadj.)$
and vice versa if we had eliminated the $\{\hat{\alpha}_i\}$ instead of the $\{\hat{\beta}_j\}$.

(7.1.3) Tests for main effects under Ω.

Computations are much simpler if we wish to estimate main effects without assuming interactions. In this case the definition of the main effects depends on the system of weights. It is now necessary to assume that there are no empty cells, else the main effects are not estimable under Ω. Then $p = IJ$. The numerator SS for the F-test for testing H_A, i.e. SS_A is given by Scheffé (21) while the statistic M_{SA}/MS_e of the F-test of H_A under Ω has I-1 and n-IJ d.o.f.

(7.2) Description of Data.

Wool is presented for sale in South Africa in lots of varying size. A "lot" consists of a number of bales of similar type wool and is classified according to it's size as big or star lot, big lots containing three or more bales and star lots one or two bales. Big and star lots are usually sold in separate sale-rooms. As a result there are often two groups of buyers, one concerned with valuing and buying big lots, the other concentrating on star lots.

Some firms allot the responsibility for valuing star lots to

junior buyers. Others, usually the smaller ones, do not appear to operate in the star lot room at all, because of the greater cost involved. This may result in a smaller number of buyers bidding on star lots than on big lots.

There is some evidence (53), (55), (41) that prices paid for star lots in Australia are lower than those paid for big lots of equivalent types of wool sold at the same sale in the same season. However, in order for it to be profitable for a grower to convert small lots into large lots, the price for large lots must exceed that for star lots by more than the cost of lot conversion plus any discount on rehandled wools.

It has also been demonstrated in Australia (41) that the unit marketing costs associated with both the buying and selling of wool at auction vary inverse with the number of bales in a sale lot. Some of the factors leading to higher buying costs per bale for small lots could be: the longer inspection time required to value small lots; higher invoicing costs; increased sampling and the fact that small lots take longer to buy. It is for such reasons that star lots are claimed to be more expensive to market than big lots.

Wool-buying is a highly competitive business in which profit margins are relatively low so that the size of these margins depends to some extend on the proportion of small lots a buyer has to purchase as these involve higher costs. For this reason buyers have pressed for an increase in the average size

of sale lots and this matter is an important feature of current discussions of wool marketing reform. It is conceivable that the inverse relationship which apparently exists between unit buying costs and size of lots could lead to wool prices varying with the size. These considerations suggests the proposition that a price discount on star lots can exist owing to the possibility of buyers avoiding the higher costs of purchasing this wool operating on big lots only.

Whan (53) illustrated that big lots of selected types sold in Sydney during the 1963-64 Australian season demanded a higher price than star lots. Available data for lots of the Australian types 62 and 62A however suggested that the discount for star lots may have decreased over the period 1957-58 to 1963-64 which could have resulted because of a reduction in the proportion of star lots offered at auction. The average premium paid for big lots in Australia for the types and periods under consideration were found to be just over 2 cents per lb. clean which represents a greasy price margin for big lots of 1.2 cents per lb. calculated at an assumed average yield of 60%. It was also found that the magnitude of the price differences varied from year to year and from selling centre to selling centre.

Because the composition of the South African and Australian wool clips differ considerably, as well as the fact that there is a higher proportion of star lots in South Africa than in Australia, raises the question whether price dis-

counts for star lots exist in South Africa and whether they are of the same magnitude as that found in Australia if they exist. Such an examination is the purpose of this chapter.

(7.3) Results:

Price data and lot size details were captured for a number of types sold at the various auction centres in South Africa during the 1969/70 season. A description of the types used in the analysis aregiven in table no. 7.2, combined for all ports.

Table 7.2

Wool type	Description	No. of Lots Star	Big	Total Lots
48	Good topmaking 12 months 64's	1851	4259	6110
53	Good topmaking 10/12 months 64's	3126	3895	7021
58	Good topmaking 9/11 months 64's	2399	2450	4849

where the relation of growth period to staple length in inches are as follows:

12 months - $2\frac{3}{4}$ to $3\frac{1}{4}$ inches

10/12 months - $2\frac{1}{4}$ to $2\frac{1}{2}$ inches

9/11 months - 2 to $2\frac{1}{4}$ inches.

Prices paid at auction for greasy wool were converted to clean wool prices using appraisers assessments for type and yield so as to eliminate effects of changes in yield on the price of wool. Errors of judgement by appraisers will almost certainly occur, but it is a reasonable assumption that such errors will be of random incidence, not influencing average values,

and that they will be consistent, affecting big lots as well as star lots. We also assume that the classing of big lots and star lots of the same type is of a similar standard of uniformity. If the classing of star lots is generally more irregular than that of big lots, shortcomings in classing could be a factor contributing to a reduction in the value of star lots.

Data for types 48, 53 and 58 were analysed and unbiased estimates of price differences between big and star lots for each type, combined for all ports, were obtained for the 1969/70 season. Inequalities in the number of observations per cell and the effect of wool price trends can produce biased estimates of price differences. The results quoted have been adjusted to remove bias due to these factors. The formula used to obtain unbiased estimates of price differences is

$$\frac{\sum_i w_i d_i}{\sum_i d_i} \quad \text{where } w_i = \frac{n_{i1}\, n_{i2}}{n_{i1} + n_{i2}}$$

and d_i is the uncorrected price difference between big and star lots for sale "i". Nineteen sales had adequate representation of the types under consideration and were used in the analysis. The analyses of variance for the price data relating to all lots of the three types, combined for the four ports over the 1969/70 season, are given in tables 7.3, 7.4 and 7.5.
In the analyses set out in the tables the total variation in

prices has been resolved into components owing to lot size, sales and interaction. The remaining or residual variation represents that portion of variation that cannot be accounted for by the present model. The components of variation due to classification are given in the sums of squares column. The average effect of one change in classification (i.e. one sale) is given in the mean square column and finally the significance of the sales and lot size is measured by the size of the ratio of the mean square classification and the mean square for residual variation. Since mean squares for F-tests of the row and column effects must be the adjusted mean squares, the adjusted SS for both rows and columns will be given in the tables. Note however that because of this breakdown SS for rows, columns and interactions do not add up to the SS between subclasses.

Table 7.3

Analysis of Variance for clean prices paid for type 48, all ports, season 1969/70 (cents per lb).

Source	SS	dof	MS	F
Lots	242.05	1	242.05	37.24**
Sales	9970.50	18	553.92	85.22**
Interaction	146.45	18	8.14	125
Residual	39348.76	6072	6.50	
Total		6109		

Between SS = 10,321.19

Total SS = 49,669.76

Table 7.4

Analysis of Variance for clean prices paid for type 53, all ports; season 1969/70 (cents/lb).

Source	SS	dof	MS	F
Lots	86.89	1	86.89	11.54**
Sales	12,767.78	18	709.32	94.20**
Interaction	149.45	18	8.30	1.10
Residual	52,613.35	6983	7.53	
Total		7020		

Between SS = 13,072.30

Total SS = 65,685.65

Table 7.5

Analysis of Variance for clean prices paid for type 58, all ports; season 1969/70 (cents/lb).

Source	SS	dof	MS	F
Sales	93.44	1	93.44	10.44**
Lots	10,995.75	18	610.88	68.25**
Interaction	113.91	18	6.33	.71
Residual	43,075.26	4811	8.95	
Total		4848		

Between SS = 11,220.42

Total SS = 54,277.68

An examination of the ratios for the present data shows that the most significant factor affecting price variation is the

time of selling; the second important factor is lot size while interaction is not significant! A very important fact was highlighted by the data: the significant effect for lot size resulted from higher prices obtained for star lots. Unbiased estimates of discounts for big lots for the three types are given in table 7.6.

Table 7.6

Price Differences between Big and Star lots for three selected types sold in South Africa, 1969/70.

Wool type	Ave.Price Discount for Big lots cents/lb clean	Ratio:No Star lots to		Variance Ratio	
		Big lots	Tot. lots	Lot Size	Interaction
48	0.43	0.43	0.30	37.24**	1.25
53	0.23	0.80	0.45	11.54**	1.10
58	0.30	0.98	0.49	10.44**	.71

Because of the close relationship between the three types there is hardly any difference between the types in discount for big lots. The discount, however, seems marginal. Since interaction is not significant, the unbiased estimates of prices differences are reliable indications of the discounts actually obtained.

The analysis was carried a step further to test whether the data showed any price differences between (i) one-bale lots and lots of size two or more bales; (ii) one-, two- and three-bale lots and lots of size four or more bales (iii) lots of one to four bales and lots of five bales or more. The following table 7.7

summarizes the results. Unbiased estimates of price differences are in cents per pound clean. Small lots will refer to the lots with the lesser number of bales while large lots will refer to the bigger lots.

Table 7.7

Price Differences between lots of various sizes for three selected types, all ports, 1969/70.

Lot Size	Wool type	No. Lots Small	No. Lots Large	Ave. Price discount for Big lots	Variance Ratio Lot Size	Variance Ratio Interaction
1 vs 2	48	867	5,243	0.52	30.51**	.96
or	53	1,663	5,358	0.21	7.31**	1.17
More	58	1,315	3,534	0.13	1.90	.54
Star	48	1,851	4,259	0.43	37.24**	1.25
vs	53	3,126	3,895	0.23	11.54**	1.10
Big	58	2,399	2,450	0.30	10.44**	.71
1-3 vs	48	2,774	3,336	0.44	45.88**	1.05
4 or	53	4,369	2,652	0.31	20.72**	1.15
More	58	3,194	1,655	0.42	20.95**	1.03
1-4	48	3,511	2,599	0.48	52.81**	.96
vs 5 or	53	5,186	1,835	0.44	35.20**	1.15
More	58	3,688	1,161	0.49	23.38**	1.39

The writer has examined the available data for a continuous relationship between size of lots and the prices paid for them. Both linear and logarithmic relationships have been tested and in every case the correlation coefficients between

these two variables is close to zero. Similar results have been obtained by Whan (53).

So far the data used in the analyses were combined for the four ports. The composition of the South African wool clip is however such that there is no a priori reason to assume that the relationship between price and lot size is the same in all South African ports. It was hence decided to make a further breakdown of the existing data by port.
Tables 7.8 to 7.19 are the analyses of variance for differences in the four ports between big and star lots for the three selected types.

Table 7.8

ANOVA for clean prices paid for type 48 : Cape Town, 1969/70.

Source	SS	d.o.f.	MS	F
Lots	34.64	1	34.64	7.47**
Sales	1,526.82	18	84.82	18.28**
Interaction	81.85	18	4.55	.98
Error	4,413.69	952	4.64	
Total		989		

Between SS = 1,615.16

Total SS = 6,028.85

Table 7.9

ANOVA for clean prices paid for type 48 : Durban 1969/70.

Source	SS	d.o.f.	MS	F
Lots	49.16	1	49.16	6.85**
Sales	3,344.04	18	185.78	25.87**
Interaction	272.69	18	15.15	2.11**
Error	8,525.62	1187	7.18	
Total		1224		

Between SS = 3,681.40

Total SS = 12,207.02

Table 7.10

ANOVA for clean prices paid for type 48 : East London 1969/70.

Source	SS	d.o.f.	MS	F
Lots	118.68	1	118.68	18.09**
Sales	6,176.59	18	343.14	52.31**
Interaction	257.83	18	14.32	2.18**
Error	16,821.92	2565	6.56	
Total		2603		

Between SS = 6,553.01

Total SS = 23,374.93

Table 7.11

ANOVA for clean prices paid for type 48 : Port Elizabeth 1969/70.

Source	SS	d.o.f.	MS	F
Lots	20.07	1	20.07	4.20*
Sales	1,844.12	18	102.45	21.43**
Interaction	93.77	18	5.21	1.09
Error	5,994.66	1254	4.78	
Total		1292		

Between SS = 1,961.40

Total SS = 7,956.06

Table 7.12

ANOVA for clean prices paid for type 53 : Cape Town 1969/70.

Source	SS	d.o.f.	MS	F
Lots	17.54	1	17.54	2.96
Sales	2,106.82	18	117.05	19.74**
Interaction	80.30	18	4.46	.75
Error	6,477.84	1092	5.93	
Total		1129		

Between SS = 2,191.58

Total SS = 8,669.42

Table 7.13

ANOVA for clean prices paid for type 53 : Durban 1969/70.

Source	SS	d.o.f.	MS	F
Lots	23.64	1	23.64	3.28
Sales	5,938.36	18	330.46	45.83**
Interaction	224.28	18	12.46	1.73
Error	12,885.20	1788	7.21	
Total		1825		

Between SS = 6,347.19

Total SS = 19,232.39

Table 7.14

ANOVA for clean prices paid for type 53 : East London 1969/70.

Source	SS	d.o.f.	MS	F
Lots	14.83	1	14.83	1.93
Sales	6,069.86	18	337.21	43.85**
Interaction	155.25	18	8.63	1.12
Error	19,086.22	2482	7.69	
Total		2519		

Between SS = 6,248.39

Total SS = 25,335.22

Table 7.15

ANOVA for clean prices paid for type 53 : Port Elizabeth 1969/70.

Source	SS	d.o.f.	MS	F
Lots	8.72	1	8.72	1.55
Sales	2,879.10	18	159.95	28.46**
Interaction	45.10	18	2.51	.45
Error	8,466.79	1507	5.62	
Total		1544		

Between SS = 2,878.61

Total SS = 11,399.22

Table 7.16

ANOVA for clean prices paid for type 58 : Cape Town 1969/70.

Source	SS	d.o.f.	MS	F
Lots	24.22	1	24.22	2.17
Sales	2,027.28	18	112.62	15.62**
Interaction	170.94	18	9.50	1.32
Error	5,898.81	818	7.21	
Total		855		

Between SS = 2,212.61

Total SS = 8,111.42

Table 7.17

ANOVA for clean prices paid for type 58 : Durban 1969/70.

Source	SS	d.o.f.	MS	F
Lots	15.95	1	15.95	1.73
Sales	2,043.27	18	113.52	12.29**
Interaction	230.05	18	12.78	1.38
Error	7,612.37	824	9.24	
Total		861		

Between SS = 2,291.78

Total SS = 9,904.15

Table 7.18

ANOVA for clean prices paid for type 58 : East London 1969/70.

Source	SS	d.o.f.	MS	F
Lots	1.04	1	1.04	.13
Sales	4,484.29	18	249.13	30.34**
Interaction	165.01	18	20.10	2.45**
Error	11,080.88	1350	8.21	
Total		1387		

Between SS = 4,650.24

Total SS = 15,731.12

Table 7.19

ANOVA for clean prices paid for type 58 : Port Elizabeth 1969/70.

Source	SS	d.o.f.	MS	F
Lots	7.12	1	7.12	.99
Sales	4,941.57	18	274.53	38.24**
Interaction	126.87	18	7.05	.98
Error	12,243.63	1705	7.18	
Total		1742		

Between SS = 5,110.54

Total SS = 17,354.17

Table 7.20

Port	Wool type	No. Lots Star	No. Lots Big	Price diff. for big lots.	$\hat{\sigma}^2$	F-Ratio Lots	F-Ratio Interaction
all	48	1851	4259	0.43	6.50	37.24**	1.25
CT	48	341	649	0.40	4.64	7.47**	.98
DBN	48	383	842	0.44	7.18	6.85**	2.11**
EL	48	846	1757	0.46	6.56	18.09**	2.18**
PE	48	281	1011	0.30	4.78	4.20**	1.09
all	53	3126	3895	0.23	7.53	11.54**	1.10
CT	53	586	544	0.25	5.93	2.96	.75
DBN	53	757	1069	0.23	7.21	3.28	1.73
EL	53	1221	1299	0.15	7.69	1.93	1.12
PE	53	562	983	0.16	5.62	1.55	.45
all	58	2399	2450	0.30	8.95	10.44**	.71
CT	58	499	357	0.35	7.21	2.17	1.32
DBN	58	474	388	-0.28	9.24	1.73	1.38
EL	58	723	665	0.06	8.21	.13	2.45**
PE	58	703	1040	0.13	7.18	.99	.98

Linear and logarithmic relationships have been tested but the correlation coefficient between size and lots and the price paid for them were found to be non-significant. The analysis was again carried a step further to test whether the data showed any price differences between one-bale lots and lots of size two or more bales. Although significant discounts for type 48 were found to exist in the various ports, highly significant in-

teractions in each case made the estimates of price differences meaningless. Although it was not the purpose of this chapter to discriminate between prices obtained in the various ports for the selected types, such differences did exist and is illustrated by the significant F-ratios in table 7.21. Individual one-way analyses of variance were calculated for each of the 19 "sale" weeks while no distinction was made between lots of varying size. In each case the four levels of the one-way analysis refer to the four South African selling centres. Observations were the clean prices obtained for the three selected types in the various ports for sales held during the same week. Percentages of total variance explained by the model are given for each week.

Table 7.21

Clean price differences amongst South African Ports : 1969/70.

Week ending	Type 48		Type 53		Type 58	
	% Var	F-Ratio	% Var	F-Ratio	% Var	F-Ratio
12/9/69	10.07	12.76**	16.18	29.49**	10.23	16.52**
26/9	15.99	24.00**	12.33	25.34**	11.79	19.12**
3/10	7.45	10.63**	8.99	15.34**	7.98	9.45**
17/10	2.55	2.81*	9.04	14.38**	2.68	2.78**
24/10	2.38	2.61	3.26	4.70**	5.84	6.04**
31/10	9.80	13.03**	17.17	27.69**	17.63	19.88**
7/11	5.83	7.46**	16.00	25.63**	5.86	5.33**
14/11	4.82	5.64**	9.43	13.50**	7.36	6.70**
21/11	12.45	17.49**	10.21	16.30**	23.11	28.86**
28/11	9.58	12.70**	8.68	12.71**	15.41	15.05**
5/12	14.41	18.24**	5.53	7.63**	10.36	10.90**
12/12	11.34	13.44**	12.50	19.39**	21.42	26.81**
23/1/70	5.62	12.93**	5.71	8.25**	11.73	9.08**
6/2	7.59	18.14**	6.68	8.67**	25.23	20.81**
20/2	.32	.49	3.17	3.38*	7.86	5.17**
27/2	3.63	3.89*	4.35	3.73*	22.69	15.27**
6/3	4.74	3.53*	10.89	7.99**	26.90	19.73**
20/3	1.45	1.14	4.67	2.89*	14.98	6.22**
10/4	2.21	.94	32.74	16.41**	28.09	10.93**

Further analysis in this respect is at present undertaken. Provisional results however reveal that clean prices in Durban

are consistently higher than that obtained in the other ports. On average the premium for Durban wools were about 2 cents clean per pound and in some cases as high as 4 cents per pound clean.

(7.4) Conclusions:

The most significant conclusion to be made is that there exists no discount for star lots compared to big lots for the selected types in South Africa as was found in Australia. On the contrary, a marginal discount for big lots was highlighted by the data when combined for all ports. If it is assumed that an average yield of 60% applies to the types in question, an average clean price margin of .40 cents per pound represents a greasy price margin for star lots of approximately .25 cents per pound. This, for all practical purposes, is not significant so that we must conclude that there are no differences in clean prices paid for big and star lots for the selected types in South Africa. It is also evident from table 7.20 that, however marginal, the unbiased estimates of price discounts for big lots increased with a shift in length classification: the longer the fibre, the higher the price discount.

A breakdown of the data with respect to the various ports indicated that F-ratios were only significant for type 48. Price discounts for the other two types were non-existent in all four ports.

In three of the analyses significant interactions were present. The estimates of price differences in these cases are therefore

not reliable indications of the actual price differences. These interactions indicate that the size and direction of the price differences between big and star lots depend on some other factor not considered in the analysis that is associated with the time at which the lots are sold. An inspection of the actual price differences revealed no consistent preference for lots of different sizes over the selling season. It seems probable that district of origin can produce an interaction between the effects of time of selling and lot size on the price difference between big and star lots. There may be a tendency for big or star lots to be associated with particular districts in accordance with differences in the nature or size of woolgrowers' enterprises. The ratio of big lots to star lots at sales would thus be related to times of shearing in different districts. Changes in the ratio of star lots to big lots of a particular type of wool at the sales may also be associated with seasonal changes in demand for that type. For example, if the main concentration of say type 48 was in the star lot room at a time when orders for this wool were heavy, the buyers wanting type 48 would presumably give their main attention to the star lots.

The author is of the opinion that buyers, because they operate in South Africa with it's very high proportion of star lots, have already discounted for this fact in their overall price levels so that no further discrimination is made in the sale room.

Any considerations to eliminate one and two bale lots from the South African auction floor (as recommended by the Committee of Enquiry on Wool Marketing in South Africa and presented to His Honourable, Senator D.C.H. Uys, Minister of Agriculture) should therefore only be based on a possible saving to brokers and buyers in the cost of offering star lots. The wool farmer can only benefit from this practice indirectly if this saving is reflected in higher prices for his wool and a reduction in the costs which he has to pay for the services rendered by his broker. It would appear that the solution to problems associated with star lots is to be sought by adjusting selling practices so as to reduce the costs of brokers and buyers, without raising the costs borne by growers. The full benefit of such a cost saving can best be passed on to the farmer if the proposed new marketing system of wool whereby the South African Wool Commission buys in the whole clip at predetermined prices, is implemented.

Chapter VIII

General

It should be evident at this stage that analysis of variance can have wide applicability in Wool Marketing. In this thesis we have treated only a few applications, but further work is still in progress. Although we have limited curselves to one- and two-way layouts, it is clear that higher-way layouts can be applied just as effectively. A higher-way layout which is at present being investigated by the author is the following:

$$P_{ijklmno} = \frac{G_{ijklmo}}{Y_n}$$

$$= \mu+T_i+L_j+C_k+D_l+S_m+P_o+W_{ijklmo}$$

where

$P_{ijklmno}$ = the calculated clean price paid

$G_{ijklmno}$ = the greasy price paid at auction

Y_n = the yield estimated for the particular lot by an SAWC appraiser; n=1,...,N lots.

T_i = effect of type of wool; i=48,53,58,...,etc.

L_j = effect of lot size; j= big or star lots.

C_k = effect of wool preparation or classing; k = grower's brand, bulk classed or inter lotted.

D_l = effect of district or origin of wool; l = Karoo, Grassveld etc.

S_m = effect of time of sale; m=1,..., 35 sales.

P_o = effect of port at which wool is sold; o = Cape Town, Durban, etc.

$W_{ijklmno}$ = residual or error term

μ = a general mean.

This will lead to a six-way layout which will most probably have unequal numbers of observations per cell, unless a random effects model is chosen to give the required balance.

Finally, it is also worthwhile to mention that the percentage variation between market and reserve prices and the percentage of wool bought by the SAWC were tested for a continuous relationship. High correlations were obtained for the various models tested and although results are still provisional, the best fit was obtained by fitting and exponential equation.

For a full treatment of the assumptions made in the Analysis of Variance technique, and the departures therefrom, the reader is refered to Scheffé (21). The following is just a brief summary of the various assumptions.

The distribution of the F-test was derived on the basis of four assumptions about errors:

1) $E(e_i) = 0$

2) $Var(e_i) = \sigma^2$

3) $Cov(e_i, e_j) = 0 \qquad i \neq j$

4) $e_i \sim N(0, \sigma^2)$

(1) : Violation of assumption one can be avoided by arranging the experiment (if possible) to avoid bias by applying treatments at random.

(2) : Inequality of Variance.

This has little effect on the F-test if the same number of observations are made on each mean. With different

cell numbers, inequality of variance has considerable effect. Bartlett's test for inequality of variance is useful but extremely sensitive to non-normality so that it has to be used with caution. When variances are definitely unequal, a suitable transformation of the data can be used to stabilize them.

(3) : Correlation between errors.

This is the most serious departure and should be avoided. Unfortunately this is not always possible in Economical applications.

(4) : Non-Normality.

Non-normality has little effect on the F-test provided the destribution of the $\{e_i\}$ is not too skew and has well behaved tails. Randomization can help to approximate normality.

B I B L I O G R A P H Y

BOOKS:

(1) ANDERSON, T.W., "Introduction to Multivariate Statistical Analysis", John Wiley & Sons, New York, 1958.

(2) BARTLETT, M.S., Jour. Royal Statist. Soc. Suppl., 4:137 (1937).

(3) BRUNK, H.D., "An Introduction to Mathematical Statistics", Blaisdell Publishing Company, New York.

(5) DIXON, W.J. and MASSEY, F.J.(Jr.), "Introduction to Statistical Analysis", Mc Graw-Hill Book Company, New York, 1951.

(6) FISHER, R.A., Proc. Int. Math. Conf., Toronto, 805(1924)

(7) GOLDBERGER, A.S., "Econometric Theory", John Wiley & Sons, New York, 1964.

(9) GRAYBILL, F.A., "An Introduction to Linear Statistical Models", Volume I, Mc Graw-Hill Book Company, New York, 1961.

(11) JOHNSTON, J., "Econometric Methods", Mc Graw-Hill Book Company, New York, 1963.

(13) MC CARTHY, E., "Wool Disposals 1945-52. The Joint Organization", Hobbs, Southampton, 1967.

(15) MILLER, R.C.(Jr.), "Simultaneous Statistical Inference", Mc Graw-Hill Book Company, New York, 1966.

(17) NATIONAL WOOL GROWERS ASSOCIATION OF SOUTH AFRICA, "Classing Standards", Craft Press, Pretoria, 1965.

(19) PEARSON, E.S. and HARTLEY, H.O., "Biometrika Tables for Statisticians", Volume I, Cambridge University Press, London, 1954.

(21) SCHEFFE, H., "The Analysis of Variance", John Wiley & Sons, New York, 1959.

(23) SIEGEL,S., "Nonparametric Statistics for the Behavioral Science", Mc Graw-Hill Book Company, New York, 1956.

(25) SNEDECOR, G.W. and COCHRAN, W.G., "Statistical Methods", Iowa State College Press, Ames, Iowa, 1967.

PAPERS:

(27) AUSTRALIAN WOOL BOARD, "Report on Wool Marketing", Vol part 1, Melbourne, October 1967.

(29) AUSTRALIAN WOOL BOARD, "Report on Wool Marketing", Vol. part 2, Melbourne, October 1967.

(31) AUSTRALIAN WOOL BOARD, "Report on Wool Marketing", Vol. II, Melbourne, October 1967.

(33) AUSTRALIAN WOOL BOARD, "Report on Wool Marketing", Vol. III, Melbourne, October 1967.

(35) BARTLETT, M.S., Jour. Royal Statist. Soc. Suppl., 4:137 (1937).

(37) BUREAU OF ECONOMIC RESEARCH, STELLENBOSCH, "The Marketing of Wool in South Africa", July 1969.

(39) FISHER, R.A., Proc. Int. Math. Corf. Toronto, 805 (1924).

(41) FOURLINNIE J.P. and WHAN R.B., "The Influence of the size of sale lots of Wool on Wool Buyers costs", Q Rev of Agric Econ., Vol. XX, no. 3, July 1967.

(43) JENKINS, E.L., "An Assessment of Costs and Capital of a Reserve price Scheme for Australian Wool", Wool Economic Research Report, No. 7, December 1964.

(45) PHILPOTT, B.P., "Analysis of trends and Fluctuations in Wool Prices", Wool Marketing Study Group Submission Paper, no. 3, Lincoln College, April 1966.

(47) PHILPOTT, B.P., "Price Formation in the Raw Wool Market", Agricultural Economics Research Unit Discussion Paper no. 8, Lincoln College, N.Z., February 1969.

(49) SOUTH AFRICAN WOOL COMMISSION, "Type List" Amended July 1965.

(50) STEVENS, W.L., "Statistical Analysis of a Non-orthogonal Tri-Factorial Experiment", Biometrika, Vol. 35 (1948), p 346.

(51) TUKEY, J.W., Biometrics, Vol. II (1955), pp 111-113.

(52 TUKEY, J.W., "One degree of freedom for non-additivity", Biometrics, Vol. 5 (1949), pp 232-242.

(53) WHAN, R.B., "Differences in the Prices paid at Auction for Big and Star lots of wool sold in Sydney", Q Rev. of Agric. Econ. Vol. XIX, No. 4, October 1966.

(55) WHAN, R.B., "Factors Influencing the Price discount for Star lots of Wool", Quarterly Review of Agricultural Economics", Vol. XXI, No. 3, July 1968.

(57) WHAN, R.B., "Price Variation within Wool Auction Sales", Quarterly Review of Agricultural Economics, Vol. XXII, No. 2, April 1969.

(59) WHAN, R.B. and FOURLINNIE, J.P., "The consistency of a Buyer's Estimates of Yield of Greasy Wool", Quarterly Review of Agricultural Economics, Vol. XXI, No. 1, January 1968.

(61) WHAN, R.B. and RICHARDSON, R.A., "A Simulated Study of an Auction Market", Aust. J. Agric. Econ. 13:2, 91, 1969.

(63) WHAN, R.B. and WEBSTER, J.D., "The effect of Vegetable Fault on clean price paid for wool", Quarterly Review of Agricultural Economics, Vol. XXI, No. 2, April 1968.

(65) WILK, M.B. and KEMPTHORNE, O. WADC Technical Report 55-244, Vol. II, Office of Technical Services, U.S. Dept. of Commerce, Washington, D.C. (1956).

(67) YATES, F., J. Amer. Stat. Ass., 29:51 (1934).