LEAST SQUARE ASPECTS OF ANALYSIS OF VARIANCE

by

R. C. Bose

Institute of Statistics
Mimeo. Series 9
For distribution to classes.

CHAPTER I

1. The expectation of a random variable x, which takes the values $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$ is defined as $E(x) = p_1x_1 + p_2x_2 + \ldots + p_nx_n$. When x is a continuous variate in the range (a,b), then

$$E(x) = \int_a^b xp(x)dx,$$

where p(x) is the frequency density. It is easy to see that

(i) $E(cx) = cE(x)$

(ii) $E(x + y) = E(x) + E(y)$

(iii) $E(c_1y_1 + c_2y_2 + \ldots + c_ny_n) = c_1E(y_1) + c_2E(y_2) + \ldots + c_nE(y_n)$.

If $E(x) = m$, then V(x) is defined as $E(x-m)^2$. Hence $V(x) = E(x^2) - m^2$, or $E(x^2) = V(x) + \{E(x)\}^2$

2. Let

$$E(x) = m_1, \qquad E(y) = m_2$$
$$V(x) = \sigma_1^2, \qquad V(y) = \sigma_2^2$$

then the correlation between x and y is defined by the equation

$$E\{(x_1 - m_1)(x_2 - m_2)\} = \rho\sigma_1\sigma_2 = \mathrm{Cov}(x_1, x_2)$$

This gives us

$$E(x_1x_2) - m_1m_2 = \rho\sigma_1\sigma_2 = \mathrm{Cov}\ (x_1, x_2)$$

When x_1 and x_2 are not correlated, $\rho = 0$, and in this case

$$E(x_1x_2) = m_1m_2 = E(x_1)E(x_2)$$

3. Let us now calculate

$V(c_1x_1 + c_2x_2 + \ldots + c_nx_n)$, given that $V(x_1) = \sigma_1^2, \ldots, V(x_n) = \sigma_n^2$,

and correlation between x_i and $x_j = \rho_{ij}$.

Let $E(x_1) = m_1, \ldots, E(x_n) = m_n$, then

$$V(c_1x_1+c_2x_2+\ldots+c_nx_n)=E\left\{(c_1x_1+c_2x_2+\ldots+c_nx_n)-(c_1m_1+c_2m_2+\ldots c_nm_n)\right\}^2$$

$$=c_1^2E(x_1-m_1)^2+c_2^2E(x_2-m_2)^2+\ldots+c_n^2E(x_n-m_n)^2$$

$$+2c_1c_2E(x_1-m_1)(x_2-m_2)+\ldots$$

$$=c_1^2\ \sigma_1{}^2+c_2^2\ \sigma_2{}^2+\ldots+2c_1c_2\ \rho_{12}\ \sigma_1\ \sigma_2+\ldots$$

$$=\sum_{i,j=1}^{n} c_ic_j\ \rho_{ij}\ \sigma_i\ \sigma_j$$

In partiuular if x_1, x_2, ..., x_n are independent variates then

$$V(c_1x_1+c_2x_2+\ldots+c_nx_n) = c_1^2\ \sigma_1^2+c_2^2\ \sigma_2^2+\ldots+c_n^2\ \sigma_n^2.$$

If further x_1, x_2, ..., x_n have the same variance σ^2, then

$$V(c_1x_1+c_2x_2+\ldots+c_nx_n) = (c_1^2+c_2^2+\ldots+c_n^2)\ \sigma^2$$

4. Again let us calculate the covariance of

$$y = c_1x_1+c_2x_2+\ldots+c_nx_n,$$

and $$y' = c_1'x_1+c'_2x_2+\ldots+c_n'x_n$$

then proceeding as before

$$\mathrm{Cov}(y,y') = \mathrm{Cov}\ \left\{c_1x_1+c_2x_2+\ldots+c_nx_n,\ \ c_1'x_1+c_2'x_2+\ldots+c_n'x_n\right\}$$

$$= c_1c_1'\ \sigma_1^2+\ldots+c_nc_n'\ \sigma_n^2+(c_1c_2'+c_2c'_1)\ \rho_{12}\ \sigma_1\ \sigma_2+\ldots$$

$$= \sum_{i,j=1}^{n} c_ic_j'\ \rho_{ij}\ \sigma_i\ \sigma_j$$

In particular, if x_1, x_2, ..., x_n are independent, then

$$\mathrm{Cov}(c_1x_1+c_2x_2+\ldots+c_nx_n,\ \ c_1'x_1+c_2'x_2+\ldots+c_n'x_n)$$

$$= c_1c_1'\ \sigma_1^2+c_2c_2'\ \sigma_2^2+\ldots+c_nc_n'\ \sigma_n^2$$

and if further the variances of the x's are equal, then the covariance is

$$(c_1c_1'+c_2c_2'+\ldots+c_nc_n')\,\sigma^2$$

5. Dependence, independence and orthogonality of linear functions.

In order to deal effectively with problems of linear estimation and tests of linear hypotheses, we shall find it convenient to introduce the notions of dependence and independence, and orthogonality of linear functions and vectors.

Consider the linear functions

$$Y_1 = 2y_1+3y_2+y_3$$

$$Y_2 = y_1+y_2+y_3$$

$$Y_3 = 4y_1+5y_2+3y_3$$

Given $Y_1 = 5$, $Y_2 = 7$, what is the value of Y_3? The clever student will notice that $Y_3 = Y_1+2Y_2$ so that its value must be 19, without going into the question of the actual values of y_1, y_2, y_3. If, however, given $Y_1 = 5$, $Y_2 = 7$, I want you to find the value of

$$Y' = 4y_1+5y_2+4y_3$$

then it is obvious that the value of Y' cannot be calculated simply by knowing the values of Y_1 and Y_2, but depends on the actual values of y_1, y_2, y_3. We, therefore, say that Y_3 is dependent on Y_1 and Y_2, but Y' is not dependent on Y_1 and Y_2.

In general, if

$$Y_1 = a_{11}y_1+a_{21}y_2+\ldots+a_{n1}y_n$$

$$Y_2 = a_{12}y_1+a_{22}y_2+\ldots+a_{n2}y_n$$

$$\ldots \qquad \ldots \qquad \ldots$$

$$Y_k = a_{1k}y_1+a_{2k}y_2+\ldots+a_{nk}y_n$$

then $Y = \ell_1y_1+\ell_2y_2+\ldots+\ell_ny_n$ will be said to be dependent on

$Y_1, Y_2, \ldots, Y_k$ if we can find constants, $b_1, b_2, \ldots, b_k$ such that

$$Y = b_1Y_1+b_2Y_2+\ldots+b_kY_k \quad \text{(identically)}$$

The null function

$$oy_1+oy_2+\ldots+oy_k$$

is always dependent on $Y_1, Y_2, \ldots, Y_k$ since we may take $b_1, b_2, \ldots, b_k$ to be zero.

The linear functions

$$Y_1, Y_2, \ldots, Y_r$$

may be said to be independent if none depends on the rest. The necessary and sufficient condition for this is that it is impossible to find $b_1, b_2, \ldots, b_r$ (not all zero) such that

$$b_1Y_1+b_2Y_2+\ldots+b_rY_r = 0 \qquad \text{(identically)}$$

Consider the two linear functions

$$Y = c_1y_1+c_2y_2+\ldots+c_ny_n$$

$$Y' = c_1'y_1+c_2'y_2+\ldots+c_n'y_n$$

of independent random variates $y_1, y_2, \ldots, y_n$ with a common variance σ^2. Then

$$\text{Cov}(Y,Y') = (c_1c_1'+c_2c_2'+\ldots+c_nc_n')\ \sigma^2$$

If the covariance is to vanish,

$$c_1c_1'+c_2c_2'+\ldots+c_nc_n' = 0$$

In this case Y and Y' are said to be orthogonal to each other.

6. <u>Dependence and independence of vectors.</u>

Any n-plet of numbers, e.g.

$$(a_1, a_2, \ldots, a_n) \text{ or } (\ell_1, \ell_2, \ldots, \ell_n)$$

is called a vector and is usually denoted by a Greek symbol. Thus we may write

$$\alpha = (a_1, a_2, \ldots, a_n), \quad \lambda = (\ell_1, \ell_2, \ldots, \ell_n)$$

The addition of vectors is defined by

$$\alpha + \lambda = (a_1 + \ell_1, a_2 + \ell_2, \ldots, a_n + \ell_n)$$

$$-\lambda = (-\ell_1, -\ell_2, \ldots, -\ell_n)$$

$$\alpha - \lambda = (a_1 - \ell_1, a_2 - \ell_2, \ldots, a_n - \ell_n)$$

Multiplication by a number c is defined by

$$c\,\alpha = (ca_1, ca_2, \ldots, ca_n)$$

The relation $\alpha = \lambda$ would mean,

$$a_1 = \ell_1, a_2 = \ell_2, \ldots, a_n = \ell_n$$

Thus if

$$\alpha_1 = (a_{11}, a_{21}, \ldots, a_{n1})$$

$$\alpha_2 = (a_{12}, a_{22}, \ldots, a_{n2})$$

$$\ldots \quad \ldots \quad \ldots \quad \ldots$$

$$\alpha_k = (a_{1k}, a_{2k}, \ldots, a_{nk})$$

then the relation

$$b_1\alpha_1 + b_2\alpha_2 + \ldots + b_m\alpha_m = \lambda$$

would mean that

$$b_1a_{11} + b_2a_{12} + \ldots + b_ma_{1m} = \ell_1$$

$$b_1a_{21} + b_2a_{22} + \ldots + b_ma_{2m} = \ell_2$$

etc.

Vectors are very closely connected with linear functions and linear equations. Thus, if there is a linear function

$$a_1y_1+a_2y_2+\dots+a_my_m$$

then we say that its coefficient vector is α .

Coefficient Vector	Linear Function
$\alpha = (a_1, a_2, \dots, a_n)$	$Y_1 = a_1y_1+a_2y_2+ \dots +a_ny_n$
$\beta = (b_1, b_2, \dots, b_n)$	$Y_2 = b_1y_1+b_2y_2+ \dots +b_ny_n$
Null vector $0 = (0, 0, \dots, 0)$	Null function $0 = 0y_1+0y_2+\dots+0y_n$
$\alpha+\beta$	Sum $= Y_1+Y_2=(a_1+b_1)y_1+(a_2+b_2)y_2+\dots+(a_n+b_n)y_n$
$c\alpha$	$cY_1 = ca_1y_1+ca_2y_2+\dots+ca_ny_n$

Thus if operations of addition or multiplication with numbers are performed on linear functions, the coefficient vectors also undergo the same operation.

Coefficient Vector	Linear Function
α_1	Y_1
α_2	Y_2
.	.
.	.
.	.
α_k	Y_k
$c_1\alpha_1+c_2\alpha_2+\dots+c_k\alpha_k$	$c_1Y_1+c_2Y_2+\dots+c_kY_k$

In defining the dependence and independence of vectors, we keep to this correspondence. Thus, if

Coefficient Vector	Linear Function
$\alpha_1 = (a_{11}, a_{21}, \dots, a_{n1})$	$Y_1 = a_{11}y_1+a_{21}y_2+\dots+a_{n1}y_n$
$\alpha_2 = (a_{12}, a_{22}, \dots, a_{n2})$	$Y_2 = a_{12}y_1+a_{22}y_2+\dots+a_{n2}y_2$
$\alpha_k = (a_{1k}, a_{2k}, \dots, a_{nk})$	$Y_k = a_{1k}y_1+a_{2k}y_2+\dots+a_{nk}y_k$
and	
$\lambda = (\ell_1, \ell_2, \dots, \ell_k)$	$Y = \ell_1y_1+\ell_2y_2+\dots+\ell_ky_k$

then λ is said to be dependent on $\alpha_1, \alpha_2, \ldots, \alpha_k$ provided we can find constants $b_1, b_2, \ldots, b_k$ such that

$$\lambda = b_1 \alpha_1 + b_2 \alpha_2 + \ldots + b_k \alpha_k.$$

You would notice that in this case the linear function Y with coefficient vector λ is also dependent on the linear functions $Y_1, Y_2, \ldots, Y_k$ with coefficient vectors $\alpha_1, \alpha_2, \ldots, \alpha_k$.

The vectors $\alpha_1, \alpha_2, \ldots, \alpha_r$ may be said to be independent if none depends on the rest. The necessary and sufficient condition for this is that it is impossible to find $b_1, b_2, \ldots, b_r$ (not all zeros), such that

$$b_1 \alpha_1 + b_2 \alpha_2 + \ldots + b_r \alpha_r = 0 \quad \text{(null vector)}$$

The set of all vectors dependent on $\alpha_1, \alpha_2, \ldots, \alpha_k$, is called <u>vector space</u> generated by $\alpha_1, \alpha_2, \ldots, \alpha_k$.

If the generating set is independent it may be called the basis of the vector space.

The set of all linear functions dependent on $Y_1, Y_2, \ldots, Y_k$ constitutes the <u>linear set</u> generated by $Y_1, Y_2, \ldots, Y_k$.

If the generating set is independent it may be called the basis of the linear set.

I will now give you some of the well known results relating to vector spaces and linear sets, the proofs for most of which are pretty obvious.

1. The vector space generated by a set of vectors $\alpha_1, \alpha_2, \ldots, \alpha_k$ remains unchanged if

   (a) α_i is replaced by $c_i \alpha_i$ where $c_i \neq 0$

   (b) α_i is replaced by $\alpha_i + \alpha_j$ where $(i \neq j)$

Combining (a) and (b) we may replace α_i by

$c_1 \alpha_1 + \ldots + c_i \alpha_i + \ldots + c_k \alpha_k$
if $c_i \neq 0$

   (c) If the null vector happens to appear in the generating set it is dropped.

1. the linear set generated by a set of linear functions $Y_1, Y_2, \ldots, Y_r$ remains unchanged if

   (a) Y_i is replaced by $c_i Y_i$ where $c_i \neq 0$

   (b) Y_i is replaced by $Y_i + Y_j$ where $(i \neq j)$

Combining (a) and (b) we may replace Y_i by

$c_1 Y_1 + c_2 Y_2 + \ldots + c_k Y_k$
if $c_i \neq 0$.

   (c) If the null function happens to be in the set, it is dropped.

2. For any vector space V there exists a number r such that it is possible to choose r independent vectors in V, but not more. Any r independent vectors form a basis of V. This number is called the rank of the vector space.

3. If we consider vectors with n coordinates, then there cannot exist more than n independent vectors. Hence every vector space must have rank $\leq$ n. The vector space consisting of all vectors with n coordinates has the rank n.

2. For any linear set there exists a number r, such that it is possible to choose r independent linear functions in the set, but not more. The linear set can be generated by any r linear functions belonging to the set. r is said to be the number of <u>degrees of</u> freedom carried by the functions of the set.

3. If we consider linear functions of n variates $y_1, y_2, \ldots, y_n$, there cannot exist more than n independent linear functions. Hence the degrees of freedom carried by any linear set $\leq$ n. The linear set of all linear functions of $y_1, y_2, \ldots, y_n$ has n degrees of freedom.

The notion of the rank of a vector space, or the degrees of freedom belonging to a linear set is also connected with the rank of a matrix. Thus, if

the rank of the vector space generated by

$$\alpha_1 = (a_{11}, a_{21}, \ldots, a_{n1})$$
$$\alpha_2 = (a_{12}, a_{22}, \ldots, a_{n2})$$
$$\ldots \quad \ldots \quad \ldots \quad \ldots$$
$$\alpha_k = (a_{1k}, a_{2k}, \ldots, a_{nk})$$

the number of degrees of freedom carried by the linear set generated by

$$Y_1 = a_{11}y_1 + a_{21}y_2 + \ldots + a_{n1}y_1$$
$$Y_2 = a_{12}y_1 + a_{22}y_2 + \ldots + a_{n2}y_2$$
$$\ldots \quad \ldots \quad \ldots \quad \ldots$$
$$Y_k = a_{1k}y_1 + a_{2k}y_2 + \ldots + a_{nk}y_k$$

is r, then r is also the rank of the matrix,

$$\begin{pmatrix} a_{11} & a_{21} & \ldots & a_{n1} \\ a_{12} & a_{22} & \ldots & a_{n2} \\ \ldots & \ldots & \ldots & \ldots \\ a_{1k} & a_{2k} & \ldots & a_{nk} \end{pmatrix}$$

i.e. r is the order of the largest non-vanishing partial determinant.

We may summarize this as follows:

Number of degrees of freedom carried by a set of linear functions

= Rank of the vector space of the coefficient vectors.

= Rank of the matrix of the coefficients.

7. Orthogonality of Vectors.

Consider n independent random variates $y_1, y_2, \ldots, y_n$ with a common variance σ^2, then we have seen that if

$$Y_1 = c_1 y_1 + c_2 y_2 + \ldots + c_n y_n$$

$$Y_2 = c_1' y_1 + c_2' y_2 + \ldots + c_n' y_n$$

then

$$\text{Cov}\,(Y_1, Y_2) = (c_1 c_1' + c_2 c_2' + \ldots + c_n c_n')\,\sigma^2$$

Here the coefficient vectors are

$$\gamma = (c_1, c_2, \ldots, c_n)$$
$$\gamma' = (c_1', c_2', \ldots, c_n')$$

We define the scalar product of γ and γ' as

$$(\gamma . \gamma') = c_1 c_1' + c_2 c'_2 + \ldots + c_n c'_n$$

For $(\gamma . \gamma) = c_1^2 + c_2^2 + \ldots + c_n^2$ we sometimes use the notation γ^2. Note that the sum of the two vectors is a vector, but their scalar product is a pure number. We can now write:

$$\text{Cov}(Y_1, Y_2) = (\gamma . \gamma')\,\sigma^2$$

$$V(Y_1) = (\gamma . \gamma)\,\sigma^2 = \gamma^2\,\sigma^2$$

If Y_1 and Y_2 are uncorrelated, i. e. when

$$c_1 c_1' + c_2 c_2' + \ldots + c_n c_n' = (\gamma . \gamma') = 0$$

we have already called Y_1 and Y_2 orthogonal. In this case we also call γ and γ' orthogonal. Thus the condition for the orthogonality of two vecors is that the sum of the products of the corresponding coefficients vanishes.

The following theorem about orthogonal vectors and orthogonal linear functions may be stated:

1. If the vector β is orthogonal to each of the vectors $\alpha_1, \alpha_2, \ldots, \alpha_n$, then β is orthogonal to all vectors dependent on $\alpha_1, \alpha_2, \ldots \alpha_n$ i.e. to all vectors of the vecotr space V generated by $\alpha_1, \alpha_2, \ldots, \alpha_n$. In this case we say that β is orthogonal to V.

2. Given a vector space V of rank r (consisting of vectors with n coordinates) then all vectors orthogonal to V constitute a vector space V' of rank (n-r). Thus

Rank V + Rank V' = n

V' is said to be the complete orthogonal space to V. Likewise, V is also the complete orthogonal space to V'.

1. If the linear function Y is orthogonal to (uncorrelated with) $Y_1, Y_2, \ldots, Y_m$, then it is orthogonal to (uncorrelated with) all linear functions depending on $Y_1, Y_2, \ldots, Y_m$, i. e. to the functions of the linear set generated by $Y_1, Y_2, \ldots, Y_m$.

2. Given a set of linear functions with r degrees of freedom, all linear functions orthogonal to the linear functions of the set, form a linear set with n-r degrees of freedom (considering linear functions with n variates.)

The random variates $y_1, y_2, \ldots, y_n$ may themselves be regarded as constituting the vector

$$\eta = (y_1, y_2, \ldots, y_n)$$

and the linear function $Y = c_1y_1 + c_2y_2 + \ldots + c_ny_n$ may be written as $(\gamma \cdot \eta)$ where $\gamma = (c_1, c_2, \ldots, c_n)$. With this notation

$$V(\gamma \cdot \eta) = (\gamma \cdot \gamma)\,\sigma^2 = \gamma^2\sigma^2$$

$$\mathrm{Cov}\left\{(\gamma \cdot \eta), (\gamma' \cdot \eta)\right\} = (\gamma \cdot \gamma')\,\sigma^2$$

8. Homogeneous equations.

Consider now a system of homogeneous linear equations to be solved. We shall first take a simple example.

Equations			Coefficient vectors
$0y_1 + 4y_2 + 8y_3 - 16y_4 - 12y_5 = 0$	0	0	(0, 4, 8, -16, -12) = α_1
$3y_1 + 2y_2 - 5y_3 - 17y_4 - 3y_5 = 0$	-3	-3	(3, 2, -5, -17, -3)= α_2
$4y_1 + 2y_2 - 8y_3 - 20y_4 - 2y_5 = 0$	-4	-3	(4, 2, -8, -20, -2) = α_3
$y_1 - 3y_2 - 7y_3 + 13y_4 + 8y_5 = 0$	-1	-1	(1,-3, -7, 13, 8) = α_4
$0y_1 - y_2 - 2y_3 + 4y_4 + 3y_5 = 0$	0	0	(0, -1, -2, 4, 3)
$3y_1 + 0y_2 - 9y_3 - 9y_4 + 3y_5 = 0$	-3	-3	(3, 0, -9, -9, 3)
$4y_1 + 0y_2 - 12y_3 - 12y_4 + 4y_5 = 0$	-4	-3	(4, 0, -12, -12, 4)
$y_1 + 0y_2 - y_3 - y_4 - y_5 = 0$	-1	-1	(1, 0, -1, 1, -1)
$0y_1 - y_2 - 2y_3 + 4y_4 + 3y_5 = 0$	0	0	(0, -1, -2, 4, 3)
$-y_1 + 0y_2 + 3y_3 + 3y_4 + y_5 = 0$	1	-1	(-1, 0, 3, 3, -1)
$0y_1 + 0y_2 + 0y_3 + 0y_4 + 0y_5 = 0$	0	1	(0, 0, 0, 0, 0)
$0y_1 + 0y_2 + 2y_3 + 4y_4 - 2y_5 = 0$	0	0	(0, 0, 2, 4, -2)
$0y_1 - y_2 + 0y_3 + 8y_4 + y_5 = 0$	0		(0, -1, 0, 8, 1) = β_1
$-y_1 + 0y_2 + 0y_3 - 3y_4 + 2y_5 = 0$	1		(-1, 0, 0, -3, 2) = β_2
$0y_1 + 0y_2 + 0y_3 + 0y_4 + 0y_5 = 0$	0		(0, 0, 0, 0, 0)
$0y_1 + 0y_2 - y_3 - 2y_4 + y_5 = 0$	0		(0, 0, -1, -2, 1) = β_3

$$y_1 = -3y_4 + 2y_5$$

$$y_2 = 8y_4 + y_5$$

$$y_3 = -2y_4 + y_5$$

Hence the general solution of the equation is

$$(-3\ell + 2m,\ 8\ell + m,\ -2\ell + m,\ \ell,\ m)$$

or the vector space generated by

$$\gamma_1 = (-3,\ 8,\ -2,\ 1,\ 0)$$
$$\gamma_2 = (\ 2,\ 1,\ \ 1,\ 0,\ 1)$$

Right hand side of the picture shows us how to obtain the basis of a vector space V, and to determine the rank. The vectors ultimately left are independent. The connection between the rank of vector spaces and matrices also become obvious.

The left hand side teaches us how to solve a system of linear homogeneous equations, and obtain the basis of the vector space completely orthogonal to V. The relation

$$\text{Rank } V + \text{Rank } V' = n$$

is also exemplified.

This may be expressed by saying that the rank of the vector space of solutions and the rank of the vector space of coefficient vectors, add up to n.

9. <u>Non-homogeneous Equations</u>

Next let us consider non-homogeneous equations. Suppose we consider the same equations, but the right hand sides are now 0, -3, -4, -1. Then the result is ultimately obtained as

$$y_1 = 3y_4 + 2y_5 - 1$$
$$y_2 = 8y_4 + \ y_5$$
$$y_3 = 2y_4 + \ y_5$$

The general solution is now given by

$$(-3\ell + 2m + 1,\ \ 8\ell + m,\ \ -2\ell + m,\ \ell,\ m) \text{ or } \ell\gamma_1 + m\gamma_2 + \gamma_3$$

where

$$\gamma_3 = (-1,\ 0,\ 0,\ 0,\ 0)$$

For the general non-homogeneous linear equations,

$$a_{11}y_1 + a_{21}y_2 + \dots + a_{n1}y_n = a_{o1}$$

$$a_{12}y_1 + a_{22}y_2 + \dots + a_{n2}y_n = a_{o2}$$

$$\dots \quad \dots \quad \dots \quad \dots$$

$$a_{1m}y_1 + a_{2m}y_2 + \dots + a_{nm}y_n = a_{om}$$

It is clear that if

augmented

$$A = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1m} & a_{2m} & \dots & a_{nm} \end{pmatrix} \quad \text{and } \bar{A} = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} & a_{01} \\ a_{12} & a_{22} & \dots & a_{n2} & a_{02} \\ \dots & \dots & \dots & \dots & \dots \\ a_{1m} & a_{2m} & \dots & a_{nm} & a_{om} \end{pmatrix}$$

then Rank $\bar{A} \geqslant$ Rank A.

If, however, the system is to be solvable, whenever a null vector appears on the right hand side, a zero must appear on the left hand side, otherwise there will be inconsistency. (Make clear by considering the example when the right hand side numbers are 0, -3, -3, -1.) Hence the necessary and sufficient condition for the solvability of the system is that

$$\text{Rank } A = \text{Rank } \bar{A}$$

or the rank of the vector space of the coefficients vector of the homogeneous portion does not increase by the adjunction of a new coordinate corresponding to the non-homogeneous portion, to each vector.

10. Projections

The length of the vector $\alpha = (a_1, a_2, \dots, a_n)$ is defined to be

$$\sqrt{a_1^2 + a_2^2 + \dots + a_n^2}$$

Then the square of the length = $(\alpha . \alpha) = \alpha^2$. If we confine the coordinates to real numbers only, then it is seen that the length cannot vanish unless the vector is null. Since the vanishing of the length is the condition of self-orthogonality, we may say that a vector cannot be self-orthogonal unless it is null.

A vector with unit length is usually called a unit vector. A vector can always be converted into a unit vector by suitable multiplication with a constant, e. g., if

$$\alpha = (a_1, a_2, \ldots, a_n)$$

then $c\,\alpha = (ca_1, ca_2, \ldots, ca_n)$ is a unit vector if we take

$$c = \frac{1}{\sqrt{(a_1^2 + a_2^2 + \ldots + a_n^2)}}$$

I shall now give you a few theorems on the orthogonality of vectors, and their connection with independence. It must be remembered that we are dealing with real constants as coefficients.

(1) If $\alpha_1, \alpha_2, \ldots, \alpha_n$ are mutually orthogonal non-null vectors, they form an independent set.

Cor. There cannot exist more than n mutually orthogonal vectors (with n coordinates).

(2) If $\alpha_1, \alpha_2, \ldots, \alpha_m$ and $\beta_1, \beta_2, \ldots, \beta_{m'}$ are two sets of independent vectors such that any α_i is orthogonal to any β_j, then the set $\alpha_1, \ldots, \alpha_m, \beta_1, \ldots, \beta_{m'}$, is an independent set.

(1) If $Y_1, Y_2, \ldots, Y_n$ are mutually orthogonal non-null linear functions, they form an independent set.

Cor. There cannot exist more than n mutually orthogonal linear functions with n variates.

(2) If $Y_1, Y_2, \ldots, Y_m$ and $Y'_1, Y'_2, \ldots, Y'_{m'}$, are two sets of independent linear functions such that any Y_i is orthogonal to any Y'_j, then the set $Y_1, Y_2, \ldots, Y_m$, $Y'_1, Y'_2, \ldots, Y'_{m'}$, is an independent set.

Proof of (1). If possible let

$$c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n = 0$$

$$\therefore\ \alpha_i\cdot(c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n) = 0$$

$$\therefore\ c_i\alpha_i^2 = 0$$

$$\therefore\ c_i = 0 \qquad (i = 1, 2, \ldots, n)$$

which shows that there cannot exist a relation $c_1\alpha_1 + c_2\alpha_2 + \ldots + c_n\alpha_n = 0$ in which the c's are not all zero.

Proof of (2). If possible let

$$c_1\alpha_1 + c_2\alpha_2 + \ldots + c_m\alpha_m + d_1\beta_1 + d_2\beta_2 + \ldots + d_{m'}\beta_{m'} = 0$$

or putting

$$\lambda = c_1\alpha_1 + c_2\alpha_2 + \ldots + c_m\alpha_m, \quad \mu = d_1\beta_1 + d_2\beta_2 + \ldots + d_{m'}\beta_{m'},$$

$$\lambda + \mu = 0$$

But $\quad \lambda\cdot\mu = 0$

$$\therefore\ (\lambda + \mu)^2 = \lambda^2 + \mu^2 = 0$$

This implies : $\lambda = 0$, $\mu = 0$ or $c_1 = c_2 = \ldots = c_m = 0$, $d_1 = d_2 = \ldots = d_m = 0$
Hence the result.

(3) If α is a non-null vector and γ is any vector, then we can uniquely express γ in the form

$$\gamma = \beta_1 + \beta_2$$

where β_2 is orthogonal to α, and β_1 is dependent on α, i.e. $\beta_1 = c\alpha$.

The vectors β_1 and β_2 are said to be the components of γ along and orthogonal to α. β_1 may be said to be the projection of γ on α.

(3) If Y_o is a non-null linear function and Y is any linear function, we can always express Y uniquely in the form

$$Y = Y_1 + Y_2$$

where Y_2 is orthogonal to Y_o and Y_1 is dependent on Y_o. (Thus Y is uniquely decomposed into two parts, one dependent on Y_o and perfectly correlated with it, and one orthogonal to Y_o and therefore completely uncorrelated with it).

(4) If $\beta_1, \beta_2, \ldots, \beta_m$ are any system of mutually orthogonal non-null vectors, then any vector γ can be uniquely expressed in the form

$$\gamma = \beta + \beta_{m+1}$$

where β depends on $\beta_1, \beta_2, \ldots, \beta_m$ and β_{m+1} is orthogonal to them.

(4) If $Y_1, Y_2, \ldots, Y_m$ are any system of mutually orthogonal non-null linear functions, then any linear function Y can be uniquely expressed in the form

$$Y = Y' + Y_{m+1}$$

where Y' depends on $Y_1, Y_2, \ldots, Y_m$ and Y_{m+1} is orthogonal to them.

Proof of (3). Suppose $\gamma = \beta_1 + \beta_2$ where β_1, β_2 satisfy the above properties then, $\beta_1 = c\,\alpha$. Thus

$$\gamma = c\alpha + \beta_2$$

$$\therefore\ (\alpha\cdot\gamma) = c\,\alpha^2 \quad \text{or} \quad c = \frac{(\alpha\cdot\gamma)}{\alpha^2}$$

$$\therefore\ \beta_1 = \frac{(\alpha\cdot\gamma)}{\alpha^2}\alpha\ , \quad \beta_2 = \gamma - \frac{(\alpha\cdot\gamma)}{\alpha^2}\alpha$$

Conversely if β_1 and β_2 are as above, they satisfy the required properties. Hence the result.

Proof of (4). Suppose $\gamma = \beta + \beta_{m+1}$, where β and β_{m+1} satisfy the above properties, then

$$\beta = c_1\beta_1 + c_2\beta_2 + \dots + c_m\beta_m$$

$$\therefore \gamma = c_1\beta_1 + c_2\beta_2 + \dots + c_m\beta_m + \beta_{m+1}$$

$$\therefore (\beta_i\gamma) = c_i\beta_i^2, \text{ or } \quad c_i = \frac{(\beta_i \cdot \gamma)}{\beta_i^2}, \quad i = 1, 2, \dots, m$$

$$\therefore \beta = \frac{(\beta_1 \cdot \gamma)}{\beta_1^2}\beta_1 + \frac{(\beta_2 \cdot \gamma)}{\beta_2^2}\beta_2 + \dots + \frac{(\beta_m \cdot \gamma)}{\beta_m^2}\beta_m,$$

$$\beta_{m+1} = \gamma - \frac{(\beta_1 \cdot \gamma)}{\beta_1^2}\beta_1 \dots - \frac{(\beta_m \cdot \gamma)}{\beta_m^2}\beta_m$$

Conversely if β and β_{m+1} are as above, they satisfy the required properties. Hence the result.

(5) Given a vector space V of rank r, we can always choose r mutually orthogonal vectors $\beta_1, \beta_2, \dots, \beta_r$ forming a basis of V.

If $\alpha_1, \alpha_2, \dots, \alpha_r$ is a basis of V, then this choice can always be made in such a way that β_i depends only on the first i vectors of the basis $\alpha_1, \alpha_2, \dots, \alpha_i$.

Cor. We can always find n mutually orthogonal vectors (with n-coordinates).

(5) Given a linear set of functions with r degrees of freedom, we can always choose r mutually orthogonal linear functions $Y_1', Y_2', \dots, Y_r'$ which generate the set.

If $Y_1, Y_2, \dots, Y_r$ are independent linear functions generating the set, then the choice can always be made in such a way that Y_i' depends only on $Y_1, Y_2, \dots, Y_i$.

Cor. We can always find n mutually orthogonal linear functions with n variates.

Proof of (5). Let a basis of V be $\alpha_1, \alpha_2, \dots, \alpha_r$. Now we can express α_2 in the form

$$\alpha_2 = \beta_1 + \beta_2$$

where β_1 depends on α_1 and β_2 is orthogonal to α_1(cf Theorem 3). Since $\beta_1 = c\,\alpha$, so β_2 depends on α_1 and α_2. Also β_2 is orthogonal to β_1. We replace α_1 by β_1 and α_2 by β_2. Now we can express

$$\alpha_3 = c_1\beta_1 + c_2\beta_2 + \beta_3$$

where $\beta_1, \beta_2, \beta_3$ are mutually orthogonal. β_3 depends on $\alpha_1, \alpha_2, \alpha_3$. We replace α_3 by β_3.

Continuing this process we get r mutually orthogonal vectors $\beta_1, \beta_2, \ldots, \beta_r$ lying in V and forming a basis of V. Clearly β_i depends on $\alpha_1, \alpha_2, \ldots, \alpha_i$.

(6) Given a vector space V of rank r, and any vector γ, then we can uniquely express γ in the form

$$\gamma = \alpha + \beta$$

where α lies in V, and β is orthogonal to V.

The vectors α and β are called the components of γ lying in and orthogonal to V. α may be called the projection of γ on V. Clearly β is the projection of γ on V', the space completely orthogonal to V.

Cor.

$$\gamma^2 = \alpha^2 + \beta^2$$

(6) Given a linear set with r d.f., then any linear function Y can be uniquely expressed in the form $Y = Y_1 + Y_2$ where Y_1 belongs to the set, and Y_2 is orthogonal to the functions of the set.

Y_1 and Y_2 may be called the components of Y lying in and orthogonal to the set.

Cor.

$$V(Y) = V(Y_1) + V(Y_2)$$

Proof of (6). From Theorem 5, we can find a basis $\beta_1, \beta_2, \ldots, \beta_r$ of V such that $\beta_1, \beta_2, \ldots, \beta_r$ are orthogonal. The result follows from Theorem 4.

(7) Let V_1 be a sub-space of the vector space V. Let α be the projection of γ on V, and α_1 the projection of α on V_1. Then α_1 is also the projection of γ on V_1.

(7) Let the linear set L_1 be a subset of the linear set L, and let Y_o be the component of the linear function Y lying in L and let Y_1 be the component of Y_o lying in L_1, then Y_1 is the component of Y lying in L_1.

Proof of (7).

$\gamma = \alpha + \beta$ where β is orthogonal to V, and thus to V_1

$\alpha = \alpha_1 + \beta_1$ where β_1 is orthogonal to V_1

$\therefore \gamma = \alpha_1 + (\beta + \beta_1)$ where $\beta + \beta_1$ is orthogonal to V_1.

Hence α_1 must be the projection of γ on V_1.

CHAPTER II

1. Consider n independent random variates or observables

$$y_1,\ y_2,\ \ldots,\ y_n \tag{1.1}$$

with a common variance σ^2, whose expectations are linear functions with known coefficients of m unknown parameters $p_1, p_2, \ldots, p_m$. Thus

$$\left.\begin{array}{l} E(y_1) = a_{11}p_1 + a_{12}p_2 + \ldots + a_{1m}p_m \\ E(y_2) = a_{21}p_1 + a_{22}p_2 + \ldots + a_{2m}p_m \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ E(y_n) = a_{n1}p_1 + a_{n2}p_2 + \ldots + a_{nm}p_m \end{array}\right\} \tag{1.2}$$

A linear function

$$Y = c_1y_1 + c_2y_2 + \ldots + c_ny_n \tag{1.3}$$

of $y_1, y_2, \ldots, y_n$ will be called an unbiased linear estimate of the function

$$\Pi = \ell_1p_1 + \ell_2p_2 + \ldots + \ell_mp_m \tag{1.4}$$

of the parameters, if

$$E(Y) = \Pi$$

independently of the parameters. Now

$$\begin{aligned} E(Y) &= c_1E(y_1) + c_2E(y_2) + \ldots + c_nE(y_n) \\ &= (c_1a_{11} + c_2a_{21} + \ldots + c_na_{n1})p_1 \\ &\quad + (c_1a_{12} + c_2a_{22} + \ldots + c_na_{n2})p_2 \\ &\quad + \ \ldots \quad \ldots \quad \ldots \quad \ldots \\ &\quad + (c_1a_{1m} + c_2a_{2m} + \ldots + c_na_{nm})p_m \end{aligned} \tag{1.5}$$

Thus a necessary and sufficient condition for Y to be an unbiased linear estimate of Π is

$$\begin{array}{l} c_1a_{11} + c_2a_{21} + \ldots + c_na_{n1} = \ell_1 \\ c_1a_{12} + c_2a_{22} + \ldots + c_na_{n2} = \ell_2 \\ \ldots \quad \ldots \quad \ldots \quad \ldots \\ c_1a_{1m} + c_2a_{2m} + \ldots + c_na_{nm} = \ell_m \end{array} \tag{1.6}$$

A linear function Π of the parameters, is said to be estimable if there exists a linear function Y of the variates, which is an unbiased linear estimate of Π. In this case there must exist $c_1, c_2, \ldots, c_n$ satisfying (1.6). Hence we get:

Theorem (1). The parametric function $\Pi = \ell_1 p_1 + \ell_2 p_2 + \ldots + \ell_m p_m$ is estimable if and only if the matrices

$$A = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{pmatrix} \text{ and } \bar{A} = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} & \ell_1 \\ a_{12} & a_{22} & \cdots & a_{n2} & \ell_2 \\ \cdots & \cdots & \cdots & \cdots & \\ a_{1m} & a_{2m} & \cdots & a_{nm} & \ell_n \end{pmatrix} \quad (1.7)$$

have the same rank.

Corollary. If the rank of A is m, then every parametric function is estimable.

Proof: Rank $\bar{A} \geq$ Rank A, but Rank $\bar{A}$ cannot exceed m since it has m rows. Hence Rank $\bar{A}$ = Rank A.

The column vectors in the equation of expectation may be denoted by

$$\alpha_1 = (a_{11}, a_{21}, \ldots, a_{m1}), \quad \alpha_2 = (a_{12}, a_{22}, \ldots, a_{m2}), \ldots, \quad \alpha_m = (a_{1m}, a_{2m}, \ldots, a_{nm})$$

and we may denote the observables by the vector

$$\eta = (y_1, y_2, \ldots, y_n)$$

which may be called the observation vector. The equation of expectation can then be simply written as

$$E(\eta) = p_1 \alpha_1 + p_2 \alpha_2 + \ldots + p_m \alpha_m \quad (1.8)$$

The linear function Y can then be written

$Y = (\gamma \cdot \eta)$ where $\gamma = (c_1, c_2, \ldots, c_n)$ is the coefficient vector

$$E(Y) = E(\gamma \cdot \eta) = p_1(\gamma \cdot \alpha_1) + p_2(\gamma \cdot \alpha_2) + \ldots + p_m(\gamma \cdot \alpha_m) \quad (1.9)$$

which is the result (1.5) in a compact form.

2. When $\Pi = \ell_1 p_1 + \ell_2 p_2 + \ldots + \ell_m p_m$ is estimable, there will exist in general an infinity of solutions for (1.6), so that an infinity of unbiased linear estimates of Π is possible. Out of these we have to pick out the one whose variance is the least. This may be called the best unbiased linear estimate. Before proceeding to this, we shall establish the notions of error and estimation spaces.

A linear function $Y = c_1y_1 + c_2y_2 + \dots + c_ny_n$ may be said to belong to 'error' if $E(Y) = 0$ independently of the parameters.

Hence for Y to belong to error

$$(\gamma \cdot \alpha_1) = 0, \quad (\gamma \cdot \alpha_2) = 0, \ \dots, \quad (\gamma \cdot \alpha_m) = 0 \qquad (2.1)$$

Thus the coefficient vector γ of Y lies in the vector space V' completely orthogonal to the vector space V generated by the vectors $\alpha_1, \alpha_2, \dots, \alpha_m$. We may call V' the error space. The space V is called the 'estimation' space for a reason which will presently appear.

Theorem (2). If $\Pi = \ell_1 p_1 + \ell_2 p_2 + \dots + \ell_m p_m$ is any estimable parametric function, then there exists a unique linear function Y_o whose coefficient vector $\gamma_o = (c_{10}, \dots, c_{no})$ lies in the estimation space and for which

$$E(Y_o) = \Pi$$

This function Y_o is the best estimate of Π

Since Π is estimable, there exists a linear function

$$Y = (c_1 y_1 + c_2 y_2 + \dots + c_n y_n) = (\gamma \cdot \mathcal{Y})$$

such that $E(Y) = \Pi$. Now let $Y_o = (c_{10}, c_{20}, \dots, c_{no})$ and $\gamma' = (c_1', c_2', \dots, c_n')$ be the components of γ along and orthogonal to V. Then $\gamma = \gamma_o + \gamma'$

$$Y = (c_1y_1 + c_2y_2 + \dots c_ny_n) = (c_{10}y_1 + c_{20}y_2 + \dots + c_{no}y_n) + (c_1' y_1 + c_2' y_2 + \dots c_n' y_n)$$

$\therefore \ \Pi = E(Y) = E(Y_o)$ since $E(Y') = 0$ as it belongs to error.

This shows that there exists a linear function Y_o whose expectation is Π, and whose coefficient vector lies in the estimation space. If possible let there exist another such function Y_o', with coefficient vector $\gamma_o' = (c_{10}', c_{20}', \dots c_{no}')$.

Then the expectation of the linear function

$$(c_{10} - c_{10}') y_1 + (c_{20} - c_{20}') y_2 + \dots + (c_{no} - c_{no}') y_n$$

with coefficient vector $\gamma_o - \gamma_o'$ is zero. Hence $\gamma_o - \gamma_o'$ belongs to error and is orthogonal to V. But it lies in V. This is impossible unless $\gamma_o - \gamma_o' = 0$ or $\gamma_o = \gamma_o'$ i. e. $Y_o = Y_o'$. This proves the uniqueness of Y_o.

Also, $V(Y) = \sigma^2 Y^2 = \sigma^2(\gamma_o^2 + \gamma'^2) = V(Y_o) + V(Y')$

$$\therefore V(Y_o) \leq V(Y)$$

the equality holding when and only when $\gamma' = 0$, i. e. when Y coincides with Y_o. This completes our proof.

Corollary. Between the estimable parametric functions, and their best estimates there is a (1, 1) correspondence such that if $Y_1, Y_2, \ldots, Y_k$ are the best estimates of $\Pi_1, \Pi_2, \ldots, \Pi_k$, then $Y = b_1Y_1 + b_2Y_2 + \ldots + b_kY_k$ is the best estimate of $\Pi = b_1\Pi_1 + b_2\Pi_2 + \ldots + b_k\Pi_k$.

Proof: Clearly, $E(Y) = \Pi$, and since the coefficient vectors of $Y_1, \ldots, Y_k$ lie in the estimation spaces, the same is true for the coefficient vector of Y.

3. The previous theorem may be put in a slightly different form:

If $Y_1 = (\alpha_1 \cdot \eta)$, $Y_2 = (\alpha_2 \cdot \eta)$, ..., $Y_m = (\alpha_m \cdot \eta)$

and $\Pi = \ell_1 p_1 + \ell_2 p_2 + \ldots + \ell_m p_m$ is an estimable parametric function, there exists one and only one linear function of the form

$$Y_o = q_1Y_1 + q_2Y_2 + \ldots + q_nY_n \tag{3.1}$$

for which $E(Y_o) = \Pi$. This linear function is the best estimate of Π.

To actually determine the best estimate we have to find the q's. Now

$$E(Y_o) = q_1E(\alpha_1 \cdot \eta) + q_2E(\alpha_2 \cdot \eta) + \ldots + q_mE(\alpha_m \cdot \eta) = \ell_1p_1 + \ell_2p_2 + \ldots + \ell_mp_m$$

Hence using (1.9) we have

$$\begin{aligned}
q_1(\alpha_1 \cdot \alpha_1) + q_2(\alpha_2 \cdot \alpha_1) + \ldots + q_m(\alpha_m \cdot \alpha_1) &= \ell_1 \\
q_1(\alpha_1 \cdot \alpha_2) + q_2(\alpha_2 \cdot \alpha_2) + \ldots + q_m(\alpha_m \cdot \alpha_2) &= \ell_2 \\
\ldots \qquad \ldots \qquad \ldots \qquad \ldots & \\
q_1(\alpha_1 \cdot \alpha_m) + q_2(\alpha_2 \cdot \alpha_m) + \ldots + q_m(\alpha_m \cdot \alpha_m) &= \ell_m
\end{aligned} \tag{3.2}$$

Solving (3.2) for the q's and substituting in (3.1), we get the best estimate.

We shall now prove the following fundamental theorem:

Theorem: If $\Pi = \ell_1 p_1 + \ell_2 p_2 + \dots + \ell_m p_m$ is an estimable parametric function, then its best estimate is obtained by substituting for the p's in Π, any solution of the normal equations

$$\begin{array}{l} p_1(\alpha_1\cdot\alpha_1) + p_2(\alpha_1\cdot\alpha_2) + \dots + p_m(\alpha_1\cdot\alpha_m) = (\alpha_1\cdot\eta) = Y_1 \\ p_1(\alpha_2\cdot\alpha_1) + p_2(\alpha_2\cdot\alpha_2) + \dots + p_m(\alpha_2\cdot\alpha_m) = (\alpha_2\cdot\eta) = Y_2 \\ \vdots \\ p_1(\alpha_m\cdot\alpha_1) + p_2(\alpha_m\cdot\alpha_2) + \dots + p_m(\alpha_m\cdot\alpha_m) = (\alpha_m\cdot\eta) = Y_m \end{array} \tag{3.3}$$

Let $\hat{p}_1, \hat{p}_2, \dots, \hat{p}_m$ be any solution of (3.3). Substitute these in (3.3), multiply by $q_1, q_2, \dots, q_m$ and add. Using (3.2) we get

$$\ell_1\hat{p}_1 + \ell_2\hat{p}_2 + \dots + \ell_m\hat{p}_m = q_1(\alpha_1\cdot\eta) + q_2(\alpha_2\cdot\eta) + \dots + q_m(\alpha_m\cdot\eta) = \text{best estimate.}$$

Corollary: It should be noted that the lefthand side in (3.3) is simply the expectation of the right hand side. Since any estimating function is of the form (3.1), it follows that any estimable parametric function must be a linear combination of the parametric functions occuring on the left side of (3.3).

4. Markoff's Theorem

If $\Pi = \ell_1 p_1 + \ell_2 p_2 + \dots + \ell_m p_m$ is estimable, then its best estimate is obtained by substituting for the p's those values which minimize the sum of squares of the deviations of the observations from expectation, i.e. the sum of squares,

$$S_e^2 = (y_1 - a_{11}p_1 - a_{12}p_2 - \dots - a_{1m}p_m)^2 + (y_2 - a_{21}p_1 - a_{22}p_2 - \dots - a_{2m}p_m)^2 + \dots + (y_n - a_{n1}p_1 - a_{n2}p_2 - \dots - a_{nm}p_m)^2$$

$$\frac{1}{2}\frac{\partial S_e^2}{\partial p_i} = -a_{1i}(y_1 - a_{11}p_1 - a_{12}p_2 - \dots - a_{1m}p_m) - a_{2i}(y_2 - a_{21}p_1 - a_{22}p_2 - \dots - a_{2m}p_m) - \dots - a_{ni}(y_n - a_{n1}p_1 - a_{n2}p_2 - \dots - a_{nm}p_m)$$

Hence equating the partial differential coefficients with respect to $p_1, p_2, \dots, p_m$ to zero we get

$$(\alpha_i \cdot \alpha_1)p_1 + (\alpha_i \cdot \alpha_2)p_2 + \ldots + (\alpha_i \cdot \alpha_m)p_m = (\alpha_i \cdot \eta)$$

$$i = 1, 2, \ldots, m$$

which are identical with the normal equation (3.3) already deduced. Hence the theorem.

5. Variance of the Best Estimate

We have seen that the best estimate depends on the linear function $Y_i = (\alpha_i \cdot \eta)$ in all cases. We therefore start with writing down the variances and covariances of these.

$$V(\alpha_i \cdot \eta) = (\alpha_i \cdot \alpha_i)\sigma^2 = \alpha_i^2 \sigma^2$$

$$\mathrm{Cov}\{(\alpha_i \cdot \eta), (\alpha_j \cdot \eta)\} = (\alpha_i \cdot \alpha_j)\sigma^2$$

Best estimate of $\pi = \ell_1 p_1 + \ell_2 p_2 + \ldots + \ell_m p_m$ is

$$Y_o = q_1(\alpha_1 \cdot \eta) + q_2(\alpha_2 \cdot \eta) + \ldots + q_m(\alpha_m \cdot \eta)$$

where the q's satisfy (3.2).

$$V(Y_o) = \left[q_1^2(\alpha_1 \cdot \alpha_1) + \ldots + q_m^2(\alpha_m \cdot \alpha_m) + 2q_1q_2(\alpha_1 \cdot \alpha_2) + \text{etc.}\right]\sigma^2 =$$

$$\sigma^2 \Sigma q_i q_j (\alpha_i \cdot \alpha_j)$$

$$= \sigma^2 \Big[q_1\{q_1(\alpha_1 \cdot \alpha_1) + q_2(\alpha_1 \cdot \alpha_2) + \ldots + q_m(\alpha_1 \cdot \alpha_m)\}$$

$$+ q_2\{q_1(\alpha_2 \cdot \alpha_1) + q_2(\alpha_2 \cdot \alpha_2) + \ldots + q_m(\alpha_2 \cdot \alpha_m)\}$$

$$+ \quad \ldots\ldots \qquad \ldots\ldots$$

$$+ q_m\{q_1(\alpha_m \cdot \alpha_1) + q_2(\alpha_m \cdot \alpha_2) + \ldots + q_m(\alpha_m \cdot \alpha_m)\}\Big]$$

$$= (\ell_1 q_1 + \ell_2 q_2 + \ldots + \ell_m q_m)\sigma^2 \text{ where the q's satisfy (3.2).}$$

Suppose a solution of the normal equations (3.3) is

$$\begin{aligned} p_1 &= C_{11}Y_1 + C_{12}Y_2 + \ldots + C_{1m}Y_m \\ p_2 &= C_{21}Y_1 + C_{22}Y_2 + \ldots + C_{2m}Y_m \\ &\ldots \quad \ldots\ldots \quad \ldots\ldots \\ p_m &= C_{m1}Y_1 + C_{m2}Y_2 + \ldots + C_{mm}Y_m \end{aligned}$$

then a solution of (3.2) would be

$$q_1 = C_{11}\ell_1 + C_{12}\ell_2 + \dots + C_{1m}\ell_m$$
$$q_2 = C_{21}\ell_1 + C_{22}\ell_2 + \dots + C_{2m}\ell_n$$
$$\dots \qquad \dots\dots \qquad \dots\dots$$
$$q_n = C_{m1}\ell_1 + C_{m2}\ell_2 + \dots + C_{mm}\ell_n$$

Then $V(Y_o) = \sigma^2\Sigma\,\ell_i\,\ell_j C_{ij}$.
We can therefore express $V(Y_o)$ in these forms

$$V(Y_o) = \sigma^2\Sigma q_i q_j u_{ij} = \sigma^2(q_1\ell_1 + q_2\ell_2 + \dots + q_n\ell_m) =$$
$$= \sigma^2\Sigma\,\ell_i\,\ell_j C_{ij} \tag{5.2}$$

To obtain the coefficients C_{ij} Fisher has suggested the following procedure: $(C_{1i}, C_{2i}, \dots, C_{mi})$ is a solution of the auxiliary equations obtained from the normal equations by putting the right-hand side zero except in the i-th equation where it is put 1.

6. Example

Let $y_1, y_2, \dots, y_n$ be the values of the dependent variate corresponding to the dependent variate $x_1, x_2, \dots, x_n$. If we want to find the linear regression we may take

$$y_i = a + b(x_i - \bar{x}) + \varepsilon_i$$

where $\bar{x} = (x_1 + x_2 + \dots + x_n)/n$, and ε_i is a random variate with mean zero. Then the equations of expectation are

$$E(y_1) = a + b(x_1 - \bar{x})$$
$$E(y_2) = a + b(x_2 - \bar{x})$$
$$\dots \qquad \dots \qquad \dots$$
$$E(y_n) = a + b(x_n - \bar{x})$$

Then the estimation space is generated by

$$\alpha_1 = (1, 1, \dots, 1), \quad \alpha_2 = (x_1-\bar{x}, x_2-\bar{x}, \dots, x_n-\bar{x})$$

$$(\alpha_1\cdot\alpha_1) = n, \quad \alpha_1\cdot\alpha_2 = 0, \ (\alpha_2\cdot\alpha_2) = \Sigma(x_i-\bar{x})^2,$$

$$Y_1 = (\alpha_1\cdot\eta) = n\bar{y}, \ Y_2 = (\alpha_2\cdot\eta) = \Sigma y_i(x_i-\bar{x})$$

The normal equations are

$$(\alpha_1 \cdot \alpha_1)a + (\alpha_1 \cdot \alpha_2)b = Y_1 \quad \text{or } a = \frac{1}{n} Y_1 \qquad C_{11} = \frac{1}{n}, \quad C_{12} = 0$$

$$(\alpha_1 \cdot \alpha_2)a + (\alpha_2 \cdot \alpha_2)b = Y_2 \qquad b = \frac{1}{\Sigma(x_i - \bar{x})^2} Y_2 \quad C_{21} = 0, \quad C_{22} = \frac{1}{\Sigma(x_i-\bar{x})^2}$$

Hence

$$\hat{a} = \bar{y}, \qquad \hat{b} = \frac{\Sigma y_i (x_i - \bar{x})}{\Sigma(x_i - \bar{x})^2}$$

$$V(\hat{a}) = \frac{\sigma^2}{n}, \qquad V(\hat{b}) = \frac{\sigma^2}{\Sigma(x_i - \bar{x})^2}$$

7. The sum of squares belonging to a single degree of freedom

The quantity

$$S^2 = \frac{(c_1 y_1 + c_2 y_2 + \dots + c_n y_n)^2}{c_1^2 + c_2^2 + \dots + c_n^2} = \frac{Y^2}{c_1^2 + c_2^2 + \dots + c_n^2} = \frac{(\gamma \cdot \eta)^2}{\gamma^2}$$

is called the sum of squares, corresponding to the single degree of freedom carried by the linear function

$$Y = c_1 y_1 + c_2 y_2 + \dots + c_n y_n = (\gamma \cdot \eta)$$

Now

$$E(Y^2) = V(Y) + \left\{ E(Y) \right\}^2 = (c_1^2 + c_2^2 + \dots + c_n^2)\, \sigma^2 + \left\{ E(Y) \right\}^2$$

$$\therefore E(S^2) = \sigma^2 + \frac{\{E(Y)\}^2}{c_1^2 + c_2^2 + \dots + c_n^2} = \sigma^2 + S_n^2$$

where S_n^2 is obtained from S^2 by substituting for the observations their expectations. Hence S_n^2 is an essentially positive quantity.

If Y belongs to error, we say that the degree of freedom carried by it belongs to error. In that case

$$E(S^2) = \sigma^2$$

Thus the expectation of the s.s. corresponding to any linear function which caries a d.f. belonging to error is always σ^2.

Let α be the projection of the observation vector η on the coefficient vecotr γ of Y. We shall show that s.s. corresponding to Y, is the squaro of the length of α, i. e. α^2.

$\eta = \alpha + \beta$ where $\alpha = c\gamma$, and β is orthogonal to γ.

$$\therefore\ (\gamma\cdot\eta) = (\gamma\cdot\alpha) = c\gamma^2$$

$$\therefore\ c = \frac{(\gamma\cdot\eta)}{\gamma^2}$$

$$\therefore\ \alpha = \frac{(\gamma\cdot\eta)}{\gamma^2}\cdot\gamma$$

$$\therefore\ \alpha^2 = \frac{\gamma\cdot\eta}{\gamma^2}$$

8. <u>The sum of squares corresponding to a set of k degrees of freedom</u>

Consider a linear set of functions with <u>k</u> d.f. We can always find <u>k</u> mutually orthogonal linear functions, $Y_1, \ldots, Y_k$ belonging to the set. If $S_1^2, \ldots, S_k^2$ are the corresponding s.s. then the s. s. corresponding to the <u>k</u> d.f. carried by the functions of the set is defined as $$S^2 = S_1^2 + S_2^2 + \ldots + S_k^2$$

It will in general be possible to choose $Y_1, \ldots, Y_k$ in an infinity of ways. For our definition to be unambiguous, we must show that it is independent of any particular choice.

Let $\gamma_1, \ldots, \gamma_k$ be the coeff. vectors of $Y_1, \ldots, Y_k$, and let α be the projection of η on the vector space Σ generated by $\gamma_1, \ldots \gamma_k$. Let $$\alpha = c_1\gamma_1 + c_2\gamma_2 + \ldots + c_k\gamma_k$$
then $c_i\gamma_i$ is also the projection of η on γ_i. We thus have

$$\alpha = \alpha_1 + \ldots + \alpha_k \qquad \alpha^2 = \alpha_1^2 + \ldots + \alpha_k^2 = S_1^2 + \ldots + S_k^2 = S^2$$

Hence the s.s. defined before is the square of the projection of η on the vector space Σ and is therefore unique.

Cor. 1. $$E(S^2) = k\sigma^2 + S_m^2$$
where S_m^2 is the quantity obtained from S^2, by replacing the observations by their expectations. In particular if the <u>k</u> d.f. belong to error, $E(S^2) = k\sigma^2$.

$s^2 = S^2/k$ is defined as the mean square for the <u>k</u> d.f. in question.

$$E(s^2) = \sigma^2 + S^2/k = \sigma^2 + s_m^2$$

Corollary 2. It is clear that if there are two linear sets carrying k_1 and k_2 d.f., such that the functions of the first set are orthogonal to the functions of the second set, and if S_1^2 and S_2^2 are the s.s. corresponding to those k_1 and k_2 d.f. then the s.s. corresponding to the $k_1 + k_2$ degrees of freedom carried by the two sets taken together is given by $S^2 = S_1^2 + S_2^2$.

In general, if k d.f. are partitioned into mutually orthogonal sets of $k_1, k_2, \ldots, k_k$ d.f., the sum of squares S^2, belonging to the k d.f. can also be partitioned into the corresponding orthogonal components $S_1^2, S_2^2, \ldots, S_k^2$.

The sum of squares corresponding to n degrees of freedom carried by all the linear functions of the observations is $\gamma^2 = y_1^2 + y_2^2 + \ldots + y_n^2$.

We have seen that there is a (1.1) correspondence botweem estimable parametric functions and their best estimates, such that if k of the parametric functions are independent then k of the best estimates are independent. Thus if k d.f. are carried by a linear set of parametric functions, then their best estimates will carry k d.f. These k d.f. are spoken of as either belonging to the parametric functions or the estimates, but the corresponding s.s. is always calculated from the best estimates.

9. Analytical formula for the sum of squares belonging to any number of d.f.

Let $Y_1, Y_2, \ldots, Y_k$ be k linear functions given by

$$\begin{aligned} Y_1 &= C_{11}y_1 + C_{21}y_2 + \ldots + C_{n1}y_n \\ Y_2 &= C_{12}y_1 + C_{22}y_2 + \ldots + C_{n2}y_n \\ &\ldots \qquad \ldots \qquad \ldots \qquad \ldots \\ Y_k &= C_{1k}y_1 + C_{2k}y_2 + \ldots + C_{nk}y_n \end{aligned} \tag{9.1}$$

not necessarily orthogonal to one another. To find the s.s. corresponding to these, we have to find the square of the projection of γ, on the vector space generated by $\gamma_1, \gamma_2, \ldots, \gamma_k$. Let this projection be

$$\alpha = t_1\gamma_1 + t_2\gamma_2 + \ldots t_k\gamma_k$$

then $\gamma = t_1\gamma_1 + t_2\gamma_2 + \ldots + t_k\gamma_k + \beta$ where β is orthogonal to $\gamma_1, \gamma_2, \ldots \gamma_k$.
Hence the t's are determined by

$$\begin{aligned} t_1(\gamma_1\cdot\gamma_1) + t_2(\gamma_1\cdot\gamma_2) + \ldots + t_k(\gamma_1\cdot\gamma_k) &= (\gamma_1\cdot\gamma) \\ t_1(\gamma_2\cdot\gamma_1) + t_2(\gamma_2\cdot\gamma_2) + \ldots + t_k(\gamma_k\cdot\gamma_k) &= (\gamma_2\cdot\gamma) \\ \ldots \qquad \ldots \qquad \ldots \qquad &\ldots \\ t_1(\gamma_k\cdot\gamma_1) + t_2(\gamma_k\cdot\gamma_2) + \ldots + t_k(\gamma_k\cdot\gamma_k) &= \gamma_k\cdot\gamma) \end{aligned} \tag{9.2}$$

The required sum of squares is

$$S^2 = (t_1\gamma_1 + t_2\gamma_2 + \dots + t_k\gamma_k)$$

$$= t_1(\gamma_1 \cdot \eta) + t_2(\gamma_2 \cdot \eta) + \dots + t_k(\gamma_k \cdot \eta) \qquad (9.3)$$

where $t_1, t_2, \dots, t_k$ satisfy (9.2).

It should be noticed that even if $Y_1, Y_2, \dots, Y_k$ are not independent the formula (9.3) for the sum of squares is valid.

Cor. 1. To find the s.s. due to all the estimates, we have to take the linear function $(\alpha_1 \ \eta), (\alpha_2 \cdot \eta), \dots (\alpha_m \cdot \eta)$. The equations (9.2) now become the normal equations. Hence the s.s. due to all estimating functions, (or all estimable parametric functions) is

$$S_o^2 = \hat{p}_1(\alpha_1 \cdot \eta) + \hat{p}_2(\alpha_2 \cdot \eta) + \dots + \hat{p}_m(\alpha_m \cdot \eta)$$

Cor. 2. Let n_o be the rank of the estimation space, then the rank of the error space is $n_e = n - n_o$. Thus n_o d.f. belong to the estimates and n_e to the error. These two sets are mutually orthogonal and if S_e^2 is the s.s. due to error, then

$$S_o^2 + S_e^2 = \eta^2 = y_1^2 + y_2^2 + \dots + y_n^2$$

$$\text{or} \quad S_e^2 = \eta^2 - \hat{p}_1(\alpha_1 \cdot \eta) - \hat{p}_2(\alpha_2 \cdot \eta) - \dots - \hat{p}_m(\alpha_m \cdot \eta)$$

$$E(S_e^2) = n_e \sigma^2$$

Hence σ^2 is estimated by

$$s_e^2 = \frac{S_e^2}{n_e}$$

The Markoff set up would give for S_e^2 the minimum value of

$$\Sigma(y_i - a_{i1}p_1 - a_{i2}p_2 - \dots - a_{im}p_m)^2$$

which as we have seen is

$$\Sigma(y_i - a_{i1}\hat{p}_1 - a_{i2}\hat{p}_2 - \dots - a_{im}\hat{p}_m)^2$$

$$= \eta^2 - 2\hat{p}_1(\alpha_1 \cdot \eta) - 2\hat{p}_2(\alpha_2 \cdot \eta) - \dots - 2\hat{p}_m(\alpha_m \cdot \eta)$$

$$+ \Sigma_{i,j}\, \hat{p}_i\hat{p}_j(\alpha_i \cdot \alpha_j)$$

$$= \eta^2 - \hat{p}_1(\alpha_1 \cdot \eta) - \hat{p}_2(\alpha_2 \cdot \eta) - \dots - \hat{p}_m(\alpha_m \cdot \eta)$$

on using the normal equations.

Hence S_e^2 is also given by the sum of the squares of the deviations of the observations from their graduated values.

10. <u>The generalized t-test</u>

So far we had not assumed anything about the nature of the universe of the y's except that their expectations were given by (1.4), and that they had a common variance σ^2. In what follows it will be assumed that the y's are normally distributed variates. Suppose we want to test whether the estimable parametric function

$$\Pi = \ell_1 p_1 + \ell_2 p_2 + \dots + \ell_m p_m \qquad (10.1)$$

is significantly different from an assigned value 'a'.

Let Y_o be the best estimate of Π. Then the coefficient vector $\gamma_o = (C_{10}, \dots, C_{m0})$ of Y_o vector is in the estimation space. If n_o is the rank of the estimation space, we choose another n_o-1 vectors $\gamma_1, \dots, \gamma_{n_o-1}$ of the estimation space, and mutually orthogonal to one another, and orthogonal to γ_o. Let $Y_1, \dots, Y_{n_o-1}$ be the corresponding linear functions. Also if n_e is the rank of the error space $(n_o+n_e=n)$, we can choose $\gamma_1', \dots, \gamma_{n_e}'$ mutually orthogonal unit vectors in the error space and let $Y_1', \dots, Y_{n_e}'$ be the corresponding linear functions. Then $Y_o, Y_1, \dots, Y_{n_o-1}, Y_1', \dots Y_{n_e}'$ will be independently and normally distributed. The mean of Y_o is 'a' under the hypothesis, and its variance is $\gamma_o^2\sigma^2 = \sigma_o^2$. The means of $Y_1, \dots, Y_{n_o-1}$ are unknown. Let them be $M_1, \dots, M_{n_o-1}$, whereas $Y_1', \dots, Y_{n_e}'$ are error functions and have therefore zero means. Since the coeff. vectors of $Y_1, \dots, Y_{n_o-1}, Y_1', \dots, Y_{n_e}'$ are of unit length, their variances are σ^2. Hence the joint distribution of the Y's is given by

$$\text{const} \times e^{-(Y_o - a)^2/2\sigma_o^2}\, dY_o$$

$$\times\, e^{-\sum_{i=1}^{n_o-1}((Y_i - M_i)^2/2\sigma^2)}\, dY_1, dY_2, \dots, dY_{n_o-1}$$

$$\times\, e^{-\sum_{i=1}^{n_e}(Y_i'^2/2\sigma^2)}\, dY_1', dY_2', \dots, dY_{n_e}' \qquad (10.2)$$

We can at first integrate out for $Y_1, \dots, Y_{n_o-1}$. Next we note that each of Y_i'/σ is a normal variate with mean zero and unit variance. Hence

$$\chi^2 = \sum_{j=1}^{n_e} Y_j'^2/\sigma^2 = S_e^2/\sigma^2 \tag{10.3}$$

obeys the χ^2 distribution with n_e d.f. Hence the joint distribution of Y_o and χ^2 is

$$\text{const} \times e^{-(Y_o - a)^2/2\sigma_o^2}(\chi^2)^{(n_e - 2)/2} e^{-\chi^2/2} dY_o d\chi^2 \tag{10.4}$$

The estimate of σ^2 is

$$s_e^2 = S_e^2/n_e$$

Hence an estimate of the variance of Y_o is

$$s_{eY_o}^2 = s_e^2(c_{10}^2 + c_{20}^2 + \dots + c_{n0}^2)$$

Let us take

$$t = (Y_o - a)/(s_e\sqrt{c_{10}^2 + c_{20}^2 + \dots + c_{n0}^2}) \tag{10.5}$$

$$= \frac{\text{Best estimate of } \Pi - \text{Value of } \Pi \text{ under the hypothesis}}{\text{Sq. root of the estimate of the variance of the best estimate}}$$

then from (10.4), it can be proved in the usual manner that $\underline{t}$ obeys Student's t-distribution with n_e d.f. For let

$$(Y_o - a)/\sigma_o = R \sin \phi, \quad \chi = R \cos \phi$$

then the joint distribution of R and ϕ is

$$\text{const} \times e^{-R^2/2} R^{n_e} \cos^{n_e-1}\phi \, d\phi \, dR$$

Integrating out for R, we get

$$\text{const} \times C_o \cos^{n_e-1}\phi \, d\phi, \text{ where } C_o = \frac{\Gamma\left(\frac{n_e-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n_e}{2}\right)}$$

Now

$$t = \frac{Y_o - a}{s_e\sqrt{c_{10}^2 + \dots + c_{n0}^2}} = \frac{(Y_o - a)/\sigma_o}{s_e/\sigma} = \frac{(Y_o - a)/\sigma_o}{\chi/\sqrt{n_e}} = \sqrt{n_e}\tan\phi$$

Hence the sampling distribution of $\underline{t}$ is

$$C\, dt \,/\, (1 + t^2/n_e)^{\frac{n_e+1}{2}}$$

where

$$C = \frac{\Gamma\left(\frac{n_e+1}{2}\right)}{\sqrt{n_e\pi}\,\Gamma\left(\frac{n_e}{2}\right)}$$

This is the well known t-distribution with n_e, d.f. Find the 5 o/o and the 1 o/o values for this distribution with n_e, d.f. If the hypothesis is true, then the 5 o/o value will be exceeded by the observed t given by (10.5), only in 5 o/o cases. When the observed t exceeds this value, we may, therefore, think the observed value to be too large to have occurred by chance, and reject the hypothesis. We say in this case that t is significant on the 5 o/o level. In so doing we shall, however, be rejecting a true hypothesis in 5 o/o cases. This is expressed by saying that we shall commit a mistake of the first kind in 5 o/o cases. If we do not want to commit a mistake of the first kind in so many cases, we may work at the 1 o/o level, and reject the hypothesis only where the observed t exceeds the 1 o/o value of t. In this case we shall be on safer ground so far as the unwarranted rejection of a true hypothesis is concerned, and shall commit a mistake of the first kind in only 1 o/o cases. But there is another side of the picture. When the hypothesis is not true, the observed value of t will exceed the 5 o/o value of t in many more cases, then it will exceed the 1 o/o value. Hence, working on the 5 o/o level, we shall be rejecting the hypothesis when false, in many more cases than when working on the 1 o/o level. We shall, therefore, be committing an error of the second kind (non-rejection of a false hypothesis) in a lesser number of cases. Thus, in substituting the 1 o/o level, what is gained in the first kind of error is lost on the second kind. This is made clear by the following considerations.

If the hypothesis to be tested is not true, and as a matter of fact

$$\pi = \ell_1 p_1 + \ell_2 p_2 + \dots + \ell_m p_n = a'$$

then $E(Y_o) = a'$, and the joint distribution of Y_o and X^2 will be

$$Ce^{-(Y_o-a')^2/2\sigma_o^2} (X^2)^{\frac{n_e-2}{2}} dY_o dX^2$$

where

$$C = \frac{1}{\sqrt{2\pi}\ \sigma_o} \cdot \frac{1}{\sqrt{2\pi}\ \sigma_o} \cdot \frac{1}{2^{n_o/2}\,\Gamma(\frac{n_o}{2})}$$

Putting

$$\Delta = \frac{a' - a}{o},\ \frac{Y_o - a}{o} = R \sin \phi,\ X = R \cos \phi$$

we now get for the distribution of R and ϕ

$$\frac{1}{\sqrt{\pi}\, 2^{\frac{n_e-1}{2}}\,\Gamma(\frac{n_e}{2})} e^{-\frac{1}{2}(R^2 - 2R\Delta \sin\phi + \Delta^2)} R^{n_e} \cos^{n_e-1}\phi\, dR\, d\phi$$

We have to integrate out for R which varies from 0 to ∞ . Now

$$I_n(x) = \frac{1}{\sqrt{2\pi}\,\Gamma(n+1)} \int_0^\infty t^n e^{-\frac{1}{2}(t+x)^2} dx$$

where I_n is the function introduced by Fisher in the British Association Tables Vol. I and defined by

$$I_o(x) = \frac{1}{\sqrt{2\pi}} \int_o^x e^{-\frac{1}{2}t^2} dt, \quad I_n(x) = \int_x^\infty I_{n-1}(x)dx$$

(The function $Hh_n(x) = \sqrt{2\pi}\, I_n(x)$ is tabulated in the British Association Tables Vol. I)

Now integrating out for R we get as the distribution of ϕ

$$\frac{\Gamma(n_e+1)}{2^{\frac{n_e-2}{2}}\,\Gamma(\frac{n_e}{2})} e^{-\frac{1}{2}\Delta^2\cos^2\phi} \cos^{n_e-1}\phi\, I_n(-\Delta\sin\phi)\, d\phi$$

Finally since $t = \sqrt{n_e}\tan\phi$ the distribution of t comes out as

$$\frac{\Gamma(n_e+1)}{2^{(n_e-2)/2}\sqrt{n_e}\,\Gamma(\frac{n_e}{2})} e^{-\frac{1}{2}\frac{n_e\Delta^2}{n_e+t^2}} I_n\left(-\frac{\Delta t}{\sqrt{t^2+n_e}}\right) \frac{dt}{(1+\frac{t^2}{n_e})^{\frac{n_e+1}{2}}}$$

Remembering that

$$\sqrt{\pi}\,\Gamma(n_e + 1) = 2^{n_e}\,\Gamma\left(\frac{n_e+1}{2}\right)\Gamma\left(\frac{n_e+2}{2}\right);$$

we can write this distribution as

$$f(n_e,\ t)\ \phi(n_e,\ t,\ \Delta)\ dt$$

where

$$f(n_e,\ t)\ dt = \frac{\Gamma\left(\frac{n_e+1}{2}\right)}{\sqrt{n_e\pi}\ \Gamma\left(\frac{n_e}{2}\right)}\ \frac{dt}{\left(1+\frac{t^2}{n_e}\right)^{\frac{n_e+1}{2}}}$$

is the distribution of t on the null hypothesis on n_e d.f. and

$$\phi(n_e,\ t,\ \Delta) = 2^{\frac{n_e+2}{2}}\ \Gamma\left(\frac{n_e+2}{2}\right) e^{-\frac{1}{2}\frac{n_e\Delta^2}{n_e+t^2}}\ I_n\left(-\frac{\Delta t}{t^2+n_e}\right).$$

Now

$$I_{n_e}(0) = \frac{1}{2^{\frac{n_e+2}{2}}\ \Gamma\left(\frac{n_e+2}{2}\right)}$$

So that when $\Delta = 0$, $\phi(n_e,\ t,\ \Delta) = 1$, and the distribution of t reduces to the familiar distribution on the null hypothesis with n_e d.f.

If we want to test our hypothesis on the α o/o level, then we shall not reject the hypothesis if the observed t lies between $-t_\alpha$ and $+t_\alpha$ where t_α is the table value of t (null hypothesis) at the α o/o level with n_e d.f. Hence when the hypothesis is wrong the chance of our not rejecting it is given by

$$P_2 = \int_{-t_\alpha}^{+t_\alpha} f(n_e,\ t)\ \phi(n_e,\ t,\ \Delta)\ dt$$

which is therefore the magnitude of the second kind of error. When α decreases t_α increases, and hence P_2 increases. The quantity $1 - P_2$ is the power of the test namely the probability with which the test enables us to reject a wrong hypothesis. Of course the power depends on both n_e and Δ, and for a fixed α can be shown to be a monotonic increasing function of both.

11. <u>Example.(i).</u>

Let $x_1, x_2, \ldots, x_n$ be a random sample of n observations from a normal population with mean m and variance σ^2. We want to test the hypothesis $m = a$. Now

$$\begin{aligned} E(x_1) &= m \\ E(x_2) &= m \\ \ldots &= \ldots \\ E(x_n) &= m \end{aligned}$$

The estimation space is generated by the single vector $\alpha = (1, 1, \ldots, 1)$, hence the rank of the estimation space is 1, and n-1 d.f. belong to error. The normal equations are

$$(\alpha \cdot \alpha)m = (\alpha \cdot \xi) \quad \text{where} \quad \xi = (x_1, x_2, \ldots, x_n)$$

$$\text{or } nm = (x_1 + x_2 + \ldots + x_n)$$

Thus m is estimated by $\bar{x} = (x_1 + x_2 + \ldots + x_n)$. The graduated value of each observation is $\bar{x}$. Hence

$$S_e^2 = \Sigma(x_i - \bar{x})^2$$

and the estimate of ^{2} is

$$s_e^2 = \frac{\Sigma(x_i - \bar{x})^2}{n-1} = \text{mean sq. due to error.}$$

Since $V(\bar{x}) = \frac{{}^2}{n}$ its estimate is $\frac{s_e^2}{n}$. Thus to test our hypothesis we have to take

$$t = \frac{\bar{x} - a}{s_e/\sqrt{n}}$$

and working on the α o/o level we reject the hypothesis if the observed t exceeds in absolute value the α o/o value with d.f. n-1 found from the tables.

Ex. (ii). Let $x_{11}, x_{12}, \ldots, x_{1n_1}$ and $x_{21}, x_{22}, \ldots, x_{2n_2}$, be two random samples of sizes n_1 and n_2 from normal population with variance σ^2 and means m_1 and m_2. It is required to test the hypothesis $m_1 = m_2$.

$$E(x_{11}) = m_1$$
$$E(x_{12}) = m_2$$
$$\dots$$
$$E(x_{1n_1}) = m_1$$
$$E(x_{21}) = \qquad m_2$$
$$E(x_{22}) = \qquad m_2$$
$$\dots$$
$$E(x_{2n_2}) = \qquad m_2$$

The estimation space is generated by the vectors

$$\alpha_1 = (1, 1, \dots, 1, \quad 0, 0, \dots, 0), \qquad \alpha_2 = (0,0,\dots,0, 1,1,\dots,1)$$

So that the estimation space has rank 2, and $n_1 + n_2 - 2$ d.f. belong to error. The observation vector is

$$\xi = (x_{11}, x_{12}, \dots, x_{1n_1}, x_{21}, x_{22}, \dots, x_{2n_2})$$

The normal equations are

$$(\alpha_1 \cdot \alpha_1)m_1 + (\alpha_1 \cdot \alpha_2)m_2 = (\alpha_1 \cdot \xi)$$
$$(\alpha_2 \cdot \alpha_1)\, m_1 + (\alpha_2 \cdot \alpha_2)m_2 = (\alpha_2 \cdot \xi)$$
$$m_1 = \bar{x}_1 \qquad m_2 = \bar{x}_2$$

where $\bar{x}_1$, $\bar{x}_2$ are the means of the two samples,

$$S_e^2 = \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 + \sum_{j=1}^{n_1} (x_{2j} - \bar{x}_2)^2$$

and an estimate of σ^2 is

$$s_e^2 = \frac{S_e^2}{n_1 + n_2 - 2}$$

The best estimate of $m_1 - m_2$ is $\bar{x}_1 - \bar{x}_2$ and

$$V(\bar{x}_1 - \bar{x}_2) = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

Hence the t-statistic which we have to use is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_e\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Working on the α o/o level, we reject the hypothesis if the observed value of t, exceeds the α o/o value of t with d.f. $n_1 + n_2 - 2$.

Ex. (iii). In the example considered in paragraph 6, let it be required to test if the regression coefficient b can be considered to have a particular value B. Here

$$S_e^2 = \sum_{i=1}^{n} \left\{ y_i - \bar{y} - \hat{b}\,(x_i - \bar{x}) \right\}^2$$

Hence an estimate of σ^2 is given by

$$s_e^2 = \frac{\sum_{i=1}^{n} \left(y_i - \bar{y} - \hat{b}(x_i - \bar{x}) \right)^2}{n - 2}$$

Hence the t-statistic required is

$$t = \frac{(b - B)\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}{s_e}$$

and we have to test for its significance on d.f. n-2.

12. The generalized z-test.

To test whether k independent estimable parametric functions can be simultaneously regarded as significantly zero, i.e. to test the null hypothesis,

$$\begin{aligned} \pi_1 &= \ell_{11}p_1 + \ell_{21}p_2 + \dots + \ell_{m1}p_m = 0 \\ \pi_2 &= \ell_{12}p_1 + \ell_{22}p_2 + \dots + \ell_{m2}p_m = 0 \\ \dots & \quad \dots \quad \dots \quad \dots \quad \dots \\ \pi_k &= \ell_{1k}p_1 + \ell_{2k}p_2 + \dots + \ell_{mk}p_m = 0 \end{aligned} \qquad (12.1)$$

Let $Y_1, Y_2, \dots, Y_k$ be the best estimates of $\pi_1, \pi_2, \dots, \pi_k$. Then their coefficient vectors $\gamma_1, \gamma_2, \dots, \gamma_k$ must lie in the estimation space and be independent. In the vector space of rank k generated by $\gamma_1, \gamma_2, \dots, \gamma_k$ we may choose k mutually orthogonal vectors of unit length $\gamma_{1o}, \gamma_{2o}, \dots, \gamma_{ko}$ and let $Y_{10}, Y_{20}, \dots, Y_{k0}$ be the corresponding linear functions. It is clear that if $\pi_1, \pi_2, \dots, \pi_k$, the expectations of $Y_1, Y_2, \dots, Y_k$ are zero, then also the expectations of $Y_{10}, Y_{20}, \dots, Y_{k0}$ are zero and conversely.

Also S_k^2, the s.s. due to $Y_1, Y_2, \dots, Y_k$ is the same as the s.s. due to $Y_{10}, Y_{20}, \dots, Y_{k0}$. Thus

$$S_k^2 = Y_{10}^2 + Y_{20}^2 + \dots + Y_{k0}^2$$

Let $\gamma_{k+1}, \gamma_{k+2}, \ldots, \gamma_{n_o}$ be n_o-k mutually orthogonal vectors of unit length in the estimation space, so that they are at the same time orthogonal to $\gamma_{10}, \gamma_{20}, \ldots, \gamma_{k0}$, and let $Y_{k+1}, Y_{k+2}, \ldots, Y_{n_o}$ be the corresponding linear functions. Let their means be $M_{k+1}, M_{k+2}, \ldots, M_{n_o}$. Finally, let $\gamma'_1, \gamma'_2, \ldots, \gamma'_{n_e}$ be as before mutually orthogonal vecotrs of unit length in the error space, and let $Y'_1, Y'_2, \ldots, Y'_{n_e}$ be the corresponding linear functions.

The joint distribution of $Y_{10}, Y_{20}, \ldots, Y_{k0}, Y_{n_o}, Y'_1, \ldots Y'_{n_e}$ can be written

$$\text{const} \times e^{-\sum_{i=1}^{k} Y_{io}^2/2\sigma^2} \, d\,Y_{10} \ldots d\,Y_{k0}$$

$$\times\, e^{-\sum_{i=k+1}^{n_o} (Y_i - M_i)^2/2\sigma^2} \, d\,Y_{k+1} \ldots dY_{n_o}$$

$$\times\, e^{-\sum_{j=1}^{n_e} Y_j'^2/2\sigma^2} \, dY'_1 \ldots d\,Y'_{n_e}$$

We can at first integrate out for $Y_{k+1}, Y_{k+2}, \ldots, Y_{n_o}$. Next we note that Y_{i_o}/σ and Y'_j/σ are normal variates with zero mean and unit variance. Hence if we put

$$\chi_1^2 = \sum_{i=1}^{k} \frac{Y_{io}^2}{\sigma^2} = \frac{S_k^2}{\sigma^2} \tag{12.2}$$

$$\chi_2^2 = \sum_{j=1}^{n_o} \frac{Y_j'^2}{\sigma^2} = \frac{S_e^2}{\sigma^2} \tag{12.3}$$

then χ_1^2 and χ_2^2 obey the χ^2 distribution with k and n_e d.f. respectively. Hence their joint distribution is

$$\text{const} \times (\chi_1^2)^{\frac{k-2}{2}} (\chi_2^2)^{\frac{n_e-2}{2}} e^{-\frac{1}{2}(\chi_1^2+\chi_2^2)} \, d\,\chi_1^2 \, d\,\chi_2^2$$

Let $s_k^2 = \frac{S_k^2}{k}$ = mean square due to hypothesis

$s_e^2 = \frac{S_e^2}{n_e}$ = mean square due to error

Let us take

$$F = \frac{s_k^2}{s_e^2}, \quad z = \frac{1}{2} \log F = \frac{1}{2} \log \frac{s_k^2}{s_e^2} \tag{12.4}$$

Putting $\chi_1 = R \sin \theta$, $\chi_2 = R \cos \theta$, we get the joint distribution of $\underline{R}$ and θ as

$$\text{const} \times R^{n_e+k-1} e^{-\frac{1}{2}R^2} \sin^{k-1}\theta \cos^{n_e-1}\theta \, dR \, d\theta$$

whence integrating out for $\underline{R}$, we have as the distribution of θ

$$\text{const} \times \sin^{k-1}\theta \cos^{n_e-1}\theta \, d\theta$$

Now

$$F = \frac{s_k^2}{s_e^2} = \frac{S_k^2 / k}{S_e^2 / n_e} = \frac{\chi_1^2 / k}{\chi_2^2 / n_e} = \frac{n_e}{k} \tan^2\theta$$

which gives as the distribution of $\underline{F}$

$$\text{const} \times \frac{F^{\frac{k-2}{2}} \, dF}{\left(1 + \frac{kF}{n_e}\right)^{\frac{n_e+k}{2}}} \tag{12.5}$$

This as we know is the $\underline{F}$ distribution with $\underline{k}$, n_e d.f.

If

$$z = \frac{1}{2} \log F = \frac{1}{2} \log \frac{s_k^2}{s_e^2},$$

then

$$F = e^{2z}, \quad dF = 2e^{2z} \, dz$$

whence the distribution of $\underline{z}$ is

$$\text{const} \quad \frac{e^{kz} \, dz}{\left(1 + \frac{k}{n_e} e^{2z}\right)^{\frac{n_e+k}{2}}} \tag{12.6}$$

Observe carefully the structure of $F = \frac{s_k^2}{s_e^2}$. The expectation of the denominator s_e^2 is σ^2 independently of any hypothesis. On the other hand $E(s_k^2) = \sigma^2$ if the hypothesis is true, but exceeds σ^2 if the hypothesis is wrong. In fact, if $E(Y_{io}) = M_{io}$ $(i=1,2,.,k)$, then for the hypothesis to be true it is necessary and sufficient that $M_{10} = M_{20} = \dots = M_{k0} = 0$. When the hypothesis is wrong,

$$E(s_k^2) = \sigma^2 + \frac{M_{10}^2 + M_{20}^2 + \dots + M_{k0}^2}{k} = \sigma^2 + \Delta^2 \qquad (12.7)$$

If we work at the α o/o level, we shall reject the hypothesis if the observed F or z given by (12.4) exceeds the corresponding α o/o value on d.f. k, n_e. This will happen in α o/o cases when the hypothesis is true. So that we shall be commiting an error of the first kind in α o/o cases. We can obtain a better control over this by diminishing α. This will however increase the α o/o value of F or z, and when the hypothesis is wrong, will lead to its rejection in fewer cases. Thus a reduction of the first kind of error will involve an increase of the second kind of error. The usual conventional levels are 5 o/o and 1 o/o. With given α it is clear that rejection of a wrong hypothesis will be in a larger and larger percentage of cases, as Δ^2 increases, so that Δ^2 may be taken as a measure of the departure of the actual state of affairs from the hypothesis.

13. Ex. i. To distinguish between group means. Let there be k samples

Sample		Sample means
1	$x_{11}\ x_{12} \dots x_{1n_1}$	$\bar{x}_1$
2	$x_{21}\ x_{22} \dots x_{2n_2}$	$\bar{x}_2$
...	... ...	...
k	$x_{k1}\ x_{k2} \dots x_{kn_k}$	$\bar{x}_k$

of sizes $n_1, n_2, \dots, n_k$, supposed to have come from normal populations with a common variance. We wish to test whether the means of these populations can be regarded as identical.

Suppose the means of the k populations to be $m_1, m_2, \ldots, m_k$, and the common variance to be σ^2. Then we have

$$\begin{array}{ll} E(x_{1j}) = m_1 + 0m_2 + \ldots + 0m_k & j = 1, 2, \ldots, n_1 \\ E(x_{2j}) = 0m_1 + m_2 + \ldots + 0m_k & j = 1, 2, \ldots, n_2 \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots & \\ E(x_{kj}) = 0m_1 + 0m_2 + \ldots + m_k & j = 1, 2, \ldots, n_k \end{array}$$

The number of variates is now

$$n_1 + n_2 + \ldots + n_k = N$$

The estimation space is generated by the column vectors $\alpha_1, \alpha_2, \ldots, \alpha_k$ where for α_u, the first $n_1 + n_2 + \ldots + n_{u-1}$ coordinates are zero, the next n_u unity and the rest zero. Clearly, $\alpha_1, \ldots, \alpha_k$ are orthogonal and hence independent. The rank of the estimation space is k, and N-k d.f. belong to error. The observation vector is

$$\xi = (x_{11}, \ldots, x_{1n_1}; x_{21}, \ldots, x_{2n_2}; \ldots; x_{k1}, \ldots, x_{kn_k})$$

$$(\alpha_1 \cdot \alpha_1)m_1 = (\alpha_1 \cdot \xi), \ (\alpha_2 \cdot \alpha_2)m_2 = (\alpha_2 \cdot \xi), \ \ldots, \ (\alpha_k \cdot \alpha_k)m_k = (\alpha_k \cdot \xi)$$

or

$$n_1 m_1 = n_1 \bar{x}_1, \ n_2 m_2 = n_2 \bar{x}_2, \ \ldots, \ n_k m_k = n_k \bar{x}_k \tag{13.1}$$

Thus $m_1, m_2, \ldots, m_k$ are estimated by $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$. The s.s. due to the estimates is

$$S_k^2 = n_1 \bar{x}_1^2 + n_2 \bar{x}_2^2 + \ldots + n_k \bar{x}_k^2 \tag{13.2}$$

and the s.s. due to error is

$$S_e^2 = \Sigma x_{ij}^2 - (n_1 \bar{x}_1^2 + n_2 \bar{x}_2^2 + \ldots + n_k \bar{x}_k^2)$$

$$= \sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1)^2 + \sum_{j=1}^{n_2} (x_{2j} - \bar{x}_2)^2 + \ldots + \sum_{j=1}^{n_k} (x_{kj} - \bar{x}_k)^2 \tag{13.3}$$

Now we want to test the hypothesis

$$m_1 = m_2 = \ldots m_k \tag{13.4}$$

Any linear function of the m's shall be called a contrast between the m's if the sum of the coefficients is zero. Thus

$$c_1 m_1 + c_2 m_2 + \ldots + c_k m_k$$

is a contrast if $\Sigma c = 0$. Evidently then our hypothesis is that all the contrasts vanish, and we have to find the s.s. due to the k-1 d.f. carried by the contrasts.

The contrast (13.3) is estimated by

$$c_1\bar{x}_1 + c_2\bar{x}_2 + \dots + c_m\bar{x}_m$$

and it is readily seen that this is orthogonal to $\bar{x}$, representing the general mean,

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \dots + n_k\bar{x}_k}{N} = \frac{\Sigma x_{ij}}{N}$$

The s.s. due to $\bar{x}$ is

$$N\bar{x}^2 \tag{13.5}$$

Subtracting it from (13.2), the s.s. due to the estimates, we get s.s. due to all the contrasts. Hence the s.s. due to the contrasts is

$$S_{k-1}^2 = n_1\bar{x}_1^2 + n_2\bar{x}_2^2 + \dots + n_k\bar{x}_k^2 - N\bar{x}^2$$

$$= n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \dots + n_k(\bar{x}_k - \bar{x})^2$$

Hence the mean square due to the contrasts between the group means is

$$s_{k-1}^2 = \frac{S_{k-1}^2}{k-1} = \frac{\sum_j n_j(\bar{x}_j - \bar{x})^2}{k - 1}$$

and the mean square due to error is

$$s_e^2 = \frac{S_e^2}{N-k} = \frac{\sum_{i,j} (x_{ij} - \bar{x}_j)^2}{N - k}$$

To test the hypothesis we use the F-statistic

$$F = \frac{s_{k-1}^2}{s_e^2}$$

and working on the α o/o level, reject the hypothesis if the observed value of F exceeds the α o/o value of F on d.f. k-1, N-k.

The result may be presented in form of an analysis of variance table,

Due to	d.f.	sum of squares	mean square
Hypothesis (between groups)	k-1	$S^2_{k-1} = \sum_j n_j(\bar{x}_j - \bar{x})^2$	$s^2_{k-1} = S^2_{k-1}/k-1$
Error (within groups)	N-k	$S^2_e = \sum_{i,j} (x_{ij} - \bar{x}_j)^2$	$s^2_e = S^2_e/N-k$
Total	N-1	$\sum (x_{ij} - \bar{x})^2$	

$$F = \frac{s^2_{k-1}}{s^2_e}$$

In the usual presentation of the analysis of variance the degree of freedom due to the grand mean is always omitted, in the guise of the correction for the mean. Adding the s.s., $n\bar{x}^2$ corresponding to this degree of freedom we get the total sum of squares as $\sum_{i,j} x^2_{ij}$ and the number of degrees of freedom as <u>N</u>.

14. Ex. (ii). <u>Two way classification</u>

Suppose there are <u>mn</u> individual readings subject to two way classification, viz.,

$$\begin{array}{cccc|c} x_{11} & x_{12} & \cdots & x_{1n} & \bar{x}_{1\cdot} \\ x_{21} & x_{22} & \cdots & x_{2n} & \bar{x}_{2\cdot} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} & \bar{x}_{m\cdot} \\ \hline \bar{x}_{\cdot 1} & \bar{x}_{\cdot 2} & & \bar{x}_{\cdot n} & \bar{x} \end{array}$$

The reading x_{ij} belonging to the i-th row and the j-th column belongs to the i-th A-class, and the j-th B-class. The variates x_{ij} are supposed to be normal with common variance σ^2, and the mean of x_{ij} is supposed to be $p_i + q_j$, the portion p_i being supposed contributed by the i-th A-class, and the portion q_j by the j-th B-class.

It is required to test the hypothesis

$$p_1 = p_2 = \cdots = p_m$$

independently of any hypothesis regarding the $q_1, q_2, \dots, q_n$ (i.e., we wish to test whether the A-class means can be regarded as identical). Now

$$E(x_{ij}) = p_i + q_j \tag{14.2}$$

The estimation space is generated by the vectors

$$\alpha_1, \alpha_2, \dots, \alpha_m, \alpha'_1, \alpha'_2, \dots, \alpha'_n$$

where the coordinates of α_i are

$$\begin{pmatrix} 0, & 0, & \dots, & 0 \\ 0, & 0, & \dots, & 0 \\ \dots & \dots & \dots & \dots \\ 1, & 1, & \dots, & 1 \\ \dots & \dots & \dots & \dots \\ 0, & 0, & \dots, & 0 \end{pmatrix}$$

the unities being only in the i-th row of the above scheme.

Similarly the coordinates of α'_j are

$$\begin{pmatrix} 0, & 0, & \dots, & 1, & \dots, & 0 \\ 0, & 0, & \dots, & 1, & \dots, & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0, & 0, & \dots, & 1, & \dots, & 0 \end{pmatrix}$$

the unities being only in the j-th column of the above scheme. Clearly,

$$\alpha_1 + \alpha_2 + \dots + \alpha_m - \alpha'_1 - \alpha'_2 - \dots - \alpha'_n = 0$$

and it is readily seen that this is the only connecting relation between them. Since if

$$c_1\alpha_1 + c_2\alpha_2 + \dots + c_m\alpha_m + c'_1\alpha'_1 + c'_2\alpha'_2 + \dots + c'_n\alpha'_n = 0$$

we get

$$c_i + c'_j = 0, \quad i = 1, 2, \dots, m; \ j = 1, 2, \dots, n$$

from which it follows

$$c_1 = c_2 = \dots = c_m = -c'_1 = -c'_2 = \dots = -c'_n$$

Thus exactly m+n-1 of the vectors $\alpha_1, \alpha_2, \dots, \alpha_m, \alpha'_1, \alpha'_2, \dots, \alpha'_n$ are independent, so that the estimation space has rank m+n-1. Consequently the number of d.f. belonging to error is $mn-m-n+1 = (m-1)(n-1)$.

Setting

$$\bar{p} = \frac{p_1 + p_2 + \dots + p_m}{m}, \quad \bar{q} = \frac{q_1 + q_2 + \dots + q_n}{n} \tag{14.3}$$

the normal equations can be written as

$$\begin{array}{ll} p_1 + \bar{q} = \bar{x}_{1.} & \bar{p} + q_1 = \bar{x}_{.1} \\ p_2 + \bar{q} = \bar{x}_{2.} & \bar{p} + q_2 = \bar{x}_{.2} \\ \cdots & \cdots \\ p_m + \bar{q} = \bar{x}_{m.} & \bar{p} + q_n = \bar{x}_{.n} \end{array} \tag{14.4}$$

Thus any contrast between the p's, viz.,

$$c_1p_1 + c_2p_2 + \ldots + c_mp_m, \quad \sum_{i=1}^{m} c_i = 0 \tag{14.5}$$

is estimated by

$$c_1\bar{x}_1 + c_2\bar{x}_2 + \ldots + c_m\bar{x}_m \tag{14.6}$$

and since there are (m-1) independent contrasts, there are (m-1) such linear functions.

The s.s. due to the m linear functions

$$\bar{x}_{1.}, \bar{x}_{2.}, \ldots, \bar{x}_{m.} \tag{14.7}$$

which estimate

$$p_1 + \bar{q}, \; p_2 + \bar{q}, \; \ldots, \; p_m + \bar{q}$$

is clearly

$$n(\bar{x}_{1.}^2 + \bar{x}_{2.}^2 + \ldots + \bar{x}_{m.}^2)$$

It is also readily seen that the functions (14.6) are all orthogonal to the grand mean

$$\bar{x} = \frac{1}{mn} \Sigma x_{ij} = \frac{1}{m} (\bar{x}_{1.} + \bar{x}_{2.} + \ldots + \bar{x}_{m.})$$

for which the sum of squares is $mn\bar{x}^2$.

Hence the m d.f. carried by the linear functions (14.7) can be split up into two orthogonal sets, viz., the m-1 d.f. carried by the functions (14.6) which estimate the contrasts between the p's, and the one d.f. carried by the grand mean $\bar{x}$. Thus S^2_{m-1}, the s.s. due to the contrasts between the p's, is given by

$$S^2_{m-1} = n(\bar{x}_{1.}^2 + \bar{x}_{2.}^2 + \ldots + \bar{x}_{m.}^2) - mn\bar{x}^2$$

$$= n \sum_i (\bar{x}_{i.} - \bar{x})^2$$

Likewise the s.s. due to the n-1 independent contrasts between the q's is

$$m \sum_{j=1}^{n} (\bar{x}_{.j} - \bar{x})^2$$

To find the s.s. due to error two ways are open to us.

Method I. The graduated value of x_{ij} is the best estimate of $p_i + q_j$. This from the normal equation is seen to be

$$\bar{x}_{i\cdot} + \bar{x}_{\cdot j} - \bar{x}$$

Hence the s.s. due to error is

$$S_e^2 = \sum_{i,j} (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})^2$$

Method II. It is easy to see that the linear functions estimating contrasts between q's are orthogonal to $\bar{x}$, as well as to linear functions estimating contrasts between the p's. Hence the total s.s. due to the estimates is

$$n \sum_i (\bar{x}_{i\cdot} - \bar{x})^2 + m \sum_j (\bar{x}_{\cdot j} - \bar{x})^2 + mn\bar{x}^2 = n \sum_i \bar{x}_{i\cdot}^2 + m \sum_j \bar{x}^2_{\cdot j} - mn\bar{x}^2$$

Hence the s.s. due to error is

$$S_e^2 = \sum_{i,j} x_{ij}^2 - n \sum_i \bar{x}_{i\cdot}^2 - m \sum_j \bar{x}^2_{\cdot j} + mn\bar{x}^2$$

$$= \sum_{i,j} (x_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{x})^2$$

To test our hypothesis we have therefore to put

$$s_{m-1}^2 = \frac{S_{m-1}^2}{m-1} = \text{mean sq. due to hypothesis}$$

$$s_e^2 = \frac{S_e^2}{(m-1)(n-1)} = \text{mean sq. due to error}$$

$$F = \frac{s_{m-1}^2}{s_e^2} \quad \text{or} \quad z = \frac{1}{2} \log \frac{s_{m-1}^2}{s_e^2}$$

Then working on the α o/o level, we reject the hypothesis, if the observed F (or z) exceeds the α o/o value of F (or z), on d.f. $m-1$, $(m-1)(n-1)$.

The results may be presented in the form of an analysis of variance table.

Due to	d.f.	s.s.	mean square
Hypothesis (Contrasts between the p's)	m-1	$S^2_{m-1} = n \sum_i (\bar{x}_i - \bar{x})^2$	$s^2_{m-1} = S^2_{m-1}/(m-1)$
Contrasts between the q's	n-1	$S^2_{n-1} = m \sum_j (\bar{x}_{\cdot j} - \bar{x})^2$	$s^2_{n-1} = S^2_{n-1}/(n-1)$
Error	(m-1)(n-1)	$S^2_e = \sum_{i,j} (x_{ij} - \bar{x}_i - \bar{x}_{\cdot j} + \bar{x})^2$	$s^2_e = S^2_e/(m-1)(n-1)$
Total	mn-1	$\sum_{i,j} (x_{ij} - \bar{x})^2$	

$$F = s^2_{m-1}/s^2_e$$

To complete the mn degrees of freedom, we have to add the degree of freedom for the grand mean, to which corresponds a sum of squares $mn\bar{x}^2$. The completed s.s. is then $\sum_{i,j} x^2_{ij}$.

15. Conditional error

The calculation of the s.s. due to the hypothesis is in some cases simplified by using the concept of conditional error. Suppose we want to test the hypothesis

$$\Pi_1 = \ell_{11}p_1 + \ell_{21}p_2 + \ldots + \ell_{m1}p_m = 0$$
$$\Pi_2 = \ell_{12}p_1 + \ell_{22}p_2 + \ldots + \ell_{m2}p_m = 0$$
$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad .$$
$$\Pi_k = \ell_{1k}p_1 + \ell_{2k}p_2 + \ldots + \ell_{mk}p_m = 0$$

Then any linear function $\underline{Y}$ of the observations may be said to belong to error conditionally if $E(Y) = 0$, in virtue of the hypothesis. Linear functions belonging to error also belong to error conditionally, but other linear functions, namely those whose expectations are of the form $b_1\Pi_1 + b_2\Pi_2 + \ldots + b_k\Pi_k$ conditionally belong to error. In the equations of expectation, if we make a change of parameters, and substitute from the hypothesis, then we can reduce the number of parameters to m-k. If we now calculate the s.s. due to error, then this will exceed the s.s. due to ordinary error, the difference being exactly equal to the s.s. due to the hypothesis.

s.s. due to hypothesis =

= s.s. due to conditional error - s.s. due to error

= s.s. due to parameters - s.s. due to conditional parameters

We may illustrate by using the example in para. 13. The hypothesis to be tested is

$$m_1 = m_2 = \ldots = m_k$$

which is equivalent to the vanishing of the (k-1) independent contrasts. We can therefore reduce the number of parameters to one only, by putting

$$m_1 = m_2 = \ldots = m_k = m$$

The equations of expectation now become

$$\begin{array}{lll} E(x_{1j}) = m & j = 1, 2, \ldots, n_1 \\ E(x_{2j}) = m & j = 1, 2, \ldots, n_2 \\ \ldots \quad \cdot & \cdot \qquad \ldots \\ E(x_{kj}) = m & j = 1, 2, \ldots, n_k \end{array}$$

The estimate of $\underline{m}$ is $\bar{x}$, and so the conditional error is

$$S_e'^2 = \Sigma x_{ij}^2 - N\bar{x}^2$$

Hence from (13.3), the s.s. due to the hypothesis is

$$S_{k-1}^2 = n_1\bar{x}_1^2 + n_2\bar{x}_2^2 + \ldots + n_k\bar{x}_k^2 - N\bar{x}^2$$

which is the result we otherwise derived.

CHAPTER III

1. Up to this time we have been considering a case when the observations $y_1, y_2, \ldots, y_n$ have a given variance σ^2. When however we design an experiment, we have some choice, and by using what has been called "local control" we may succeed in reducing σ^2, and improving the efficiency of our experiment. This will be first illustrated by a simple example:

Suppose it is required to test whether two drugs, A and B, are of equal value in producing sleep, or whether one of them can be regarded significantly better. We could conduct the experiment in one of two ways.

Method I. We choose 2n individuals at random from the universe of individuals for which we want our results to be true, and to $\underline{n}$ of these administer the drug A, and to the other $\underline{n}$ administer the drug B, and note the number of hours of sleep produced.

Suppose the values are as follows:

Drug A	x_1	x_2	...	x_n
Drug B	x'_1	x'_2	...	x'_n

If the mean effect, of the drug A is $\underline{m}$, and the mean effect of the drug B is $\underline{m'}$, then we have to test the hypothesis $m = m'$. Here

$$x_i = m + \varepsilon_i \ , \quad x'_i = m' + \varepsilon'_i \ , \quad i = 1, 2, \ldots, n$$

Where ε_i, ε'_i are random variables with mean zero and variance, say σ^2. Hence

$$E(x_i) = m, \quad E(x'_i) = m', \quad V(x_i) = V(x'_i) = \sigma^2, \ i = 1, \ldots, n.$$

To test our hypothesis, we have to use the t-statistic

$$t = \frac{\bar{x} - \bar{x}'}{s_e\sqrt{2/n}}$$

where

$$s_e^2 = \frac{\Sigma(x_i - \bar{x})^2 + \Sigma(x'_i - \bar{x}')^2}{2n - 2}$$

and reject the hypothesis on the $\alpha\%$ level if the observed value $\underline{t}$ exceeds the $\alpha\%$ value of $\underline{t}$ with d.f. 2n - 2.

Method II. We pick out <u>n</u> individuals at random, and to each administer both drugs A and B, (with a suitable interval to eliminate any carry over effects), and note the results.

We now have the following scheme:

	Individuals			
	1	2	...	n
Drug A	y_1	y_2	...	y_n
Drug B	y'_1	y'_2	...	y'_n

The variability in the responses to the two drugs may be supposed to be compounded of two parts; the individual response, and the variability due to the other residual factors. Thus,

$$y_i = m + p_i + \delta_i, \; y'_i = m' + p_i + \delta'_i, \; (i = 1, 2, \ldots, n)$$

where p_i is the personal response of the i-th individuals, and δ_i, δ'_i are random variables with mean zero and variance say $\sigma_1^{\,2}$. We may write

$$Z_i = y_i - y'_i = (m - m') + (\delta_i - \delta'_i)$$

and $E(Z_i) = E(y_i - y'_i) = m - m'$; $V(Z_i) = V(y_i - y'_i) = 2\sigma_1^{\,2}$.

To test our hypothesis, we now have to use the t-statistic

$$t = \frac{\overline{Z}}{s'_e/\sqrt{n}} = \frac{\overline{y} - \overline{y}'}{s'_e/\sqrt{n}}$$

where

$$s'^{\,2}_e = \frac{\Sigma(Z_i - \overline{Z})^2}{n - 1} = \frac{\Sigma(y_i - y'_i - \overline{\overline{y} - \overline{y}'})^2}{n - 1}$$

and reject the hypothesis on the α % level if the observed value of <u>t</u> exceeds the α % value of <u>t</u> with $n - 1$ d.f.

Now let us compare the two methods. If the hypothesis is true, then of course both tests would lead to the rejection of the true hypothesis in α % cases, so that there is nothing to choose between them. What happens when the hypothesis is wrong? The expectation of the numerator is in both cases the same, viz., $m - m'$. The expectation of the square of the denominator is in the first case $2\sigma^2/n$ whereas, in the second case it is $2\sigma_1^2/n$. If we denote the variance in personal response by $V(p)$, we have $$\sigma^2 = V(p) + \sigma_1^2 \quad .$$

If V(p) is at appreciable, then σ^2 exceeds σ_1^2. Thus $\underline{t}$ in the second method will in the long run deviate from zero more, than in the first method. On the other hand the deviation of $\underline{t}$ from zero will be considered significant only if $\underline{t}$ exceeds the α% value on $n - 1$ d.f., whereas, in the first method, it is considered significant if $\underline{t}$ exceeds the α% value on $2n - 2$ d.f. For example, if $n = 10$, then the 1% and 5% values of $\underline{t}$ are as follows

d.f.	5%	1%
9	2.262	3.250
18	2.101	2.878

The α% value on $n - 1$ d.f. is always larger than the corresponding value on $2n - 2$ d.f., though the difference tends to decrease as $\underline{n}$ increases. Hence, the larger deviation of $\underline{t}$ from zero in the second method is compensated by its being required to exceed a larger α% value, for significance. The physical reason for this is that in the second method our estimate of error is relatively uncertain, being based only on $n - 1$ d.f. as against $2n - 2$ d.f. in the first method, and for the same margin of safety as regards the first kind of error, the deviation of $\underline{t}$ from zero should be relatively larger in order to be considered significant.

It is therefore clear that if the variance of the personal response is appreciable enough to compensate for the loss of $n - 1$ d.f. in the estimation of error, the second method will give a better result. The elimination of personal response in the second method is an example of local control. It is a general principle in the design of experiments, that if by instituting suitable local control we can eliminate significant causes of variation, then we shall usually gain in precision. We must, however, be careful not to leave too few degrees of freedom for the estimation of error by over elaborating the experiment.

As another example, suppose there are $\underline{m}$ treatments in an agricultural experiment, and we want to test whether all the treatments are equally efficacious so far as the yield is concerned, or whether they are significantly different. Let mn plots of land be available for the experiment. In this case we could again use two methods.

First Method. In this case each treatment is applied to n randomly selected plots from all the plots available. We can apply the method of para. 13. Taking all the yields from the same treatment as forming one group, we have simply to test for the equality of the group means. If the yields for the i-th treatment are

$$x_{i1}, x_{i2}, \ldots, x_{in}$$

then the analysis of variance is

due to	d.f.	sum of squares	mean square
Treatment	$m - 1$	$S^2_{m-1} = \sum_i (\bar{x}_i - \bar{x})^2$	$s^2_{m-1} = S^2_{m-1}/(m - 1)$
Error	$mn - m$	$S^2_e = \sum_{i,j} (x_{ij} - \bar{x}_i)^2$	$s^2_e = S^2_e / m(n - 1)$
Total	$mn - 1$	$\sum_{i,j} (x_{ij} - \bar{x})^2$	

$$F = \frac{s^2_{m-1}}{s^2_e}$$

The expectation σ^2 of the mean square in the denominator of F includes, besides other residual factors, the variance due to the different fertility of the different plots.

Second Method. In this case the land is divided into n compact blocks, each consisting of m plots. Within any block each of the m treatments is applied to one plot chosen at random. This is known as the randomized block design. Let y_{ij} be the yield of the i-th treatment, for the j-th block. To distinguish between the treatment means, we can apply the method of para. 14 and exhibit the analysis of variance as

due to	d.f.	sum of squares	mean square
Treatments	m-1	$S'^2_{m-1}=n\sum_i(\bar{y}_{i.}-\bar{y})^2$	$s'^2_{m-1}=S'^2_{m-1}/(m-1)$
Blocks	n-1	$S'^2_{n-1}=m\sum_j(\bar{y}_{.j}-\bar{y})^2$	$s'^2_{n-1}=S'^2_{n-1}/(n-1)$
Error	(m-1)(n-1)	$S'^2_e=\sum_{i,j}(x_{ij}-\bar{x}_{i.}-\bar{x}_{.j}+\bar{x})^2$	$s'^2_e=S'^2_e/(m-1)(n-1)$
Total	mn-1	$\sum_{i,j}(x_{ij}-\bar{x})^2$	

$$F = \frac{s'^2_{m-1}}{s'^2_e}\text{, d.f., m-1, mn-1}$$

To compare the two methods we observe that in the second method we have eliminated from the error differences of fertility between blocks. This may be exhibited by writing

$$x_{ij} = t_i + \varepsilon_{ij},\ y_{ij} = t_i + b_j + \delta_{ij}$$

where t_i is the effect of i-th treatment, and b_j the mean effect of the j-th block. We shall denote by $\bar{t}$ the mean effect of all the treatments. Thus

$$\bar{t} = (t_1 + t_2 + \dots + t_m)/m$$

$$\sigma^2 = E(s_e^2) = V(\varepsilon);\ \sigma'^2 = E(s_e'^2) = V(\delta)$$

But since $V(\varepsilon) = V(b) + V(\delta)$, where $V(b)$ denotes the component of the variance due to the differences between the blocks, $\sigma^2 \geq \sigma'^2$. Thus, if plots within the same block tend to be relatively more alike in their response to treatments, than plots in different blocks, then σ^2 will appreciably exceed σ'^2. Now the ratio of the expectation of the numerator and denominator of F

$$1 + \frac{n\Sigma(t_i - \bar{t})^2}{(m-1)\sigma^2} \text{ and } 1 + \frac{n\Sigma(t_i - \bar{t})^2}{(m-1)\sigma'^2}$$

in the two methods, when the hypothesis of equality of the treatment effects is false. Thus, <u>F</u> in the second method will deviate more from unity than in the first method. On the other hand, for significance on the $\alpha\%$ level, F in the second method must exceed the $\alpha\%$ value of F on d.f. $m - 1, (m - 1)(n - 1)$. This value will be somewhat larger than the $\alpha\%$ value of F on d.f. $m - 1$, $mn - m$, which has to be exceeded by the observed F in the first method. For example, note 1% and 5% values of F shown below for the case $m = 5$.

d.f.	5%	1%
4,20	2.87	4.43
4,25	2.76	4.18

Hence, as in the previous example it is clear that if the variance between blocks is appreciable enough to compensate for the loss of $n - 1$ d.f. in the estimation of error, the second method will give a better result. As a matter of fact, the randomized block design has been found to be of great value and is now in almost universal use.

2. The randomized block design discussed above is efficient only when the number of treatments is not very large, for the variability in response of plots within the same block, always goes to swell the error. When the block size is large, the design tends to become inefficient. To overcome this, "incomplete block designs" have been introduced. I shall first discuss the general theory of analysis of incomplete block designs before coming to concrete 'incomplete block designs'.

Let us consider N observations on heterogeneous material to which u treatments (whose expected effects are given by the parameters $t_1, t_2, \ldots, t_u$) have to be applied in order to test their relative efficacy. It would be advantageous to divide the material into b relatively homogeneous parts, before applying the treatments. Thus in a field experiment we would divide the whole land on which the experiment is to be made into b compact blocks. Each block is then divided into a number of plots to which treatments are applied, and the yield (or whatever other effect we are interested in) is observed. In a biological experiment the animals on whom the experiment is to be performed may be divided into b relatively homogeneous groups.

Each such group corresponds to a block, and the animals within a group to a plot.

We have thus in the general case u treatments and b blocks. Let the number of plots in the j-th block be k_j, and let the number of replications of the i-th treatment be r_i. Clearly

$$r_1 + r_2 + \dots + r_u = k_1 + k_2 + \dots + k_b = N \qquad (2.1)$$

In the usual experiments, the number of plots within a block is kept constant, so that $k_1 = k_2 = \dots = k_b$, but due to accidental circumstances such as a missing plot, these numbers may become unequal. Hence it is advisable so far as the general theory is concerned to keep open the possibility of their being different. Suppose the i-th treatment has been applied to n_{ij} plots in the j-th block. Then n_{ij} is either 1 or 0, according as the j-th treatment has or has not been applied to some plot in the j-th block. (The same treatment is never applied to more than one plot, in the same block). The numbers n_{ij} can be represented in the matrix form

$$\begin{array}{lccccc} & & & & & \text{Total} \\ & n_{11} & n_{12} & \dots & n_{1b} & r_1 \\ & n_{21} & n_{22} & \dots & n_{2b} & r_2 \\ & \dots & \dots & \dots & \dots & .. \\ & n_{u1} & n_{u2} & \dots & n_{ub} & r_u \\ \hline \text{Total} & k_1 & k_2 & \dots & k_b & N \end{array} \qquad (2.2)$$

If $n_{ij} = 1$, then the yield of the i-th treatment in the j-th block may be denoted by y_{ij}. We then have

$$E(y_{ij}) = t_i + b_j$$

the part t_i being due to the i-th treatment, and the part b_j to the j-th block.

The plots being assumed to come in a certain order, the equations of expectation can now be written in full as

$$E(y_{ij}) = 0t_1 + .. + t_i + .. + 0t_u + 0b_1 + .. + b_j + .. + 0b_b \qquad (2.3)$$

Let η be the observation vector whose coordinates are the yields of the N plots; and let $\tau_1, \tau_2, \dots, \tau_u, \beta_1, \beta_2, \dots, \beta_b$, be the column vectors, corresponding to the parameters $t_1, t_2, \ldots, t_u$, $b_1, b_2, \ldots, b_b$. It is readily seen that

$$(\eta \cdot \tau_i) = T_i \; ; \; (\eta \cdot \beta_j) = B_j \tag{2.4}$$

where T_i is the total yield of the i-th treatment, and B_j is the total yield of the j-th block. Since any best estimate of a linear function of the parameters, must be compounded of $(\eta \cdot \tau_i)$ and $(\eta \cdot \beta_j)$, $(i = 1, 2, \ldots, u)$, $(j = 1, 2, \ldots, b)$ we at once get the following result:

Theorem 1. Corresponding to any estimable function Π, of the treatment and block effects $t_1, t_2, \ldots, t_u, b_1, b_2, \ldots, b_b$ there exists a unique linear function Y,

$$Y = q_1T_1 + q_2T_2 + \ldots + q_uT_u + q_1'B_1 + q_2'B_2 + \ldots + q_b'B_b \tag{2.5}$$

of the yield and block totals for which $E(Y) = \Pi$. This linear function $\underline{Y}$ is the best estimate of Π.

3. We are however usually more interested in those parametric functions which do not contain the block effects. Let us find out the general form of the best estimates of functions of the treatment effects only. Now

$$E(T_i) = n_{i1}(t_i + b_1) + n_{i2}(t_i + b_2) + \ldots + n_{ib}(t_i + b_b)$$
$$= r_it_i + (n_{i1}b_1 + n_{i2}b_2 + \ldots + n_{ib}b_b) \tag{3.1}$$

$$E(B_j) = n_{1j}(t_1 + b_j) + n_{2j}(t_2 + b_j) + \ldots + n_{uj}(t_u + b_j)$$
$$= k_jb_j + (n_{1j}t_1 + n_{2j}t_2 + \ldots + n_{uj}t_u) \tag{3.2}$$

Hence the coefficient of b_j in

$$E(Y) = E(q_1T_1 + q_2T_2 + \ldots + q_uT_u + q_1'B_1 + q_2'B_2 + \ldots + q_b'B_b)$$

is given by

$$k_jq_j' + q_1n_{1j} + q_2n_{2j} + \ldots + q_un_{uj} \tag{3.3}$$

Hence if E(Y) is to be free from block effects, we must have

$$q_j' = -\frac{1}{k_j}(n_{1j}q_1 + n_{2j}q_2 + \ldots + n_{uj}q_u) \tag{3.4}$$

Substituting in $\underline{Y}$ we find that it must be of the form

$$q_1Q_1 + q_2Q_2 + \ldots + q_uQ_u \tag{3.5}$$

x
n J

where
$$Q_i = T_i - \frac{n_{i1}B_1}{k_1} - \frac{n_{i2}B_2}{k_2} - \dots - \frac{n_{ib}B_b}{k_b}$$
Conversely it is easy to verify that $E(Q_i)$ does not contain any block effects. The quantities Q_i are called the adjusted yield because Q_i is obtained by deducting from T_i, the average yield of the blocks in which the i-th treatment occurs.

We can now state the following theorem:

Theorem 2. Corresponding to any estimable parametric function
$$\ell_1 t_1 + \ell_2 t_2 + \dots + \ell_u t_u$$
of the treatment effects only, there exists a unique linear function of the adjusted yields, viz:
$$q_1 Q_1 + q_2 Q_2 + \dots + q_u Q_u$$
such that
$$E(q_1 Q_1 + q_2 Q_2 + \dots + q_u Q_u) = \ell_1 t_1 + \ell_2 t_2 + \dots + \ell_u t_u$$
This linear function ΣqQ is the best estimate of the parametric function $\Sigma \ell t$.

4. Let us now find $E(Q_i)$, $V(Q_i)$ and $Cov(Q_i, Q_{i'})$.
$$E(Q_i) = E(T_i) - \frac{n_{i1}}{k_1}E(B_1) - \frac{n_{i2}}{k_2}E(B_2) - \dots - \frac{n_{ib}}{k_b}E(B_b)$$
Substituting from (3.1) and (3.2), we get
$$E(Q_i) = C_{i1}t_1 + C_{i2}t_2 + \dots + C_{iu}t_u \tag{4.1}$$
where
$$E(C_{ii'}) = -\left(\frac{n_{i1}n_{i'1}}{k_1} + \frac{n_{i2}n_{i'2}}{k_2} + \dots + \frac{n_{ib}n_{i'u}}{k_b}\right) \tag{4.2}$$
$$C_{ii} = r_i - \left(\frac{n_{i1}^2}{k_1} + \frac{n_{i2}^2}{k_2} + \dots + \frac{n_{ib}^2}{k_b}\right) \tag{4.3}$$

When as is usually the case in field experiment designs, the block size is constant and equal to $\underline{k}$, we have the simple relations
$$C_{ii'} = -\frac{\lambda_{ii'}}{k},\ i \neq i' \tag{4.4}$$
$$C_{ii} = r_i(1 - \frac{1}{k}) \tag{4.5}$$
where $\lambda_{ii'}$ is the number of blocks in which the treatments $\underline{i}$ and $\underline{i'}$ occur together.

Again
$$V(T_i) = r_i\sigma^2,\ Cov(T_i,\ T_{i'}) = 0$$
$$V(B_j) = k_j\sigma^2,\ Cov(B_j,\ B_{j'}) = 0 \qquad (4.6)$$
$$Cov(T_i,\ B_j) = n_{ij}\sigma^2$$

Hence
$$V(Q_i) = V(T_i - \frac{n_{i1}}{k_1}B_1 - \frac{n_{i2}}{k_2}B_2 - \dots - \frac{n_{ib}}{k_b}B_b)$$
$$= V(T_i) + \sum_j \frac{n_{ij}^2}{k_j^2}V(B_j) - 2\sum_j\left(\frac{n_{ij}}{k_j}Cov(T_i,\ B_j)\right)$$
$$= \sigma^2\left(r_i + \sum_j \frac{n_{ij}^2}{k_j} - 2\sum_j \frac{n_{ij}^2}{k_j}\right)$$
$$= C_{ii}\sigma^2 \qquad (4.7)$$

$Cov(Q_i,\ Q_{i'}) =$
$$= Cov(T_i - \frac{n_{i1}}{k_1}B_1 - \frac{n_{i2}}{k_2}B_2 - \dots - \frac{n_{ib}}{k_b}B_b,\ T_{i'} - \frac{n_{i'1}}{k_1}B_1 - \frac{n_{i'2}}{k_2}B_2 - \dots - \frac{n_{i'b}}{k_b}B_b)$$
$$= -\sum_j \frac{n_{i'j}}{k_j}Cov(T_i,\ B_{j'}) - \sum_j \frac{n_{ij}}{k_j}Cov(T_{i'},\ B_j) + \sum \frac{n_{ij}n_{i'j}}{k_j^2}V(B_j)$$
$$= -\sigma^2\sum_j \frac{n_{ij}n_{i'j}}{k_j}$$
$$= C_{ii'}\sigma^2 \qquad (4.8)$$

Corollary.
$$Cov(B_j,\ Q_i) = Cov(B_j,\ T_i) - \frac{n_{ij}}{k_j}V(B_j)$$
$$= n_{ij}\sigma^2 - \frac{n_{ij}}{k_j}k_j\sigma^2$$
$$= 0 \qquad (4.9)$$

This shows that any block total is orthogonal to any adjusted yield.

5. The constituents of the matrix $((C_{ii'}))$ occur both in $E(Q_i)$ and $V(Q_i)$. This matrix plays a very important part in our theory.

(i) The matrix is symmetrical, i.e., $C_{ii'} = C_{i'i}$

(ii) Each row and column adds up to zero.

(iii) It follows from (ii) that the determinant of the matrix vanishes.

(iv) We shall see later that the rank of the matrix is equal to the number of independent estimable linear functions of the treatment effects.

Since any estimate of a linear function of treatment effects is of the form

$$q_1 Q_1 + q_2 Q_2 + \dots + q_u Q_u$$

it follows that any estimable linear function of the treatment effects must be of the form

$$\begin{aligned} L &= q_1 E(Q_1) + q_2 E(Q_2) + \dots + q_u E(Q_u) \\ &= q_1 (C_{11} t_1 + C_{12} t_2 + \dots + C_{1u} t_u) \\ &+ q_2 (C_{21} t_1 + C_{22} t_2 + \dots + C_{2u} t_u) \\ &+ \quad \dots \quad \dots \quad \dots \quad \dots \\ &+ q_u (C_{u1} t_1 + C_{u2} t_2 + \dots + C_{uu} t_u) \end{aligned}$$

Hence the sum of the coefficients of $t_1, t_2, \dots, t_u$ in $\underline{L}$, must vanish in consequence of (ii).

A linear function of the treatment effects, in which the sum of the coefficients vanishes, may be called a treatment contrast. Thus we get:

If any linear function of the treatment effects is estimable it must be a contrast. In particular the sum of treatment effects

$$t_1 + t_2 + \dots + t_u$$

is non-estimable.

To answer the question whether every treatment contrast is estimable, we have to bring in the concept of connectedness.

A treatment and block may be said to be associated if the treatment is contained in the block. Two treatments, two blocks, or a block and a treatment are said to be connected if it is possible to pass from one to the other by means of a chain consisting alternately of blocks and treatments such that any two members of the chain are associated. Thus in a design if i_o and i_n are connected treatments we must have a chain

$$i_o, j_1, i_1, j_2, \dots j_{n-1}, i_{n-1}, j_n, i_n \qquad (5.1)$$

such that the block j_p is associated to the treatments i_{p-1}, i_p for $p = 1, 2, \dots, n$.

A design is said to be a connected design if every block and treatment of the design is connected to every other. Likewise, a portion of the design may be said to be a connected portion if every block and treatment in the portion is connected to every other. Any general design must break up into a number of connected parts, such that a block or treatment belonging to one part, is unconnected with a block or a treatment belonging to the other part.

6. Let us first study the properties of connected designs. All the usual designs for varietal trials, viz., randomized blocks, balanced incomplete blocks, lattices (including rectangular lattices) and the more general class of designs known as partially balanced incomplete block designs, are connected designs in the sense defined above.

We shall first show that for a connected design any treatment contrast is estimable.

Since a general treatment contrast

$$\ell_1 t_1 + \ell_2 t_2 + \dots + \ell_u t_u, \ \Sigma \ell_i = 0$$

can be written in the form

$$\ell_1(t_1 - t_u) + \ell_2(t_2 - t_u) + \dots + \ell_{n-1}(t_{n-1} - t_u)$$

it is sufficient for us to show that $t_{i_o} - t_{i_n}$ is estimable, where i_o and i_n are any two treatments. Since the design is connected there exists a chain (5.1), showing that i_{p-1} and i_p both occur in the block j_p. Hence $t_{i_{p-1}} - t_{i_p}$ is estimated by the difference of the yields of the plots in this block to which i_{p-1} and i_p have been applied. Since

$$t_{i_o} - t_{i_n} = (t_{i_o} - t_{i_1}) + \dots + (t_{i_{p-1}} - t_{i_p}) + \dots + (t_{i_{n-1}} - t_{i_n})$$

it is clear that $t_{i_o} - t_{i_n}$ is estimable.

The linear function $E(Q_1)$, $E(Q_2)$, ..., $E(Q_u)$ of the treatment effects are not independent since their sum vanishes. But at least $u - 1$ of them are independent since every treatment contrast being estimable is of the form

$$q_1E(Q_1) + q_2E(Q_2) + \dots + q_uE(Q_u)$$

and there are $u - 1$ independent treatment contrasts. This shows that in the case of a connected design the rank of the matrix $((C_{ii'}))$ is $u - 1$. It also follows that of the linear function $Q_1, Q_2, \dots, Q_u$ of the yields, just $u - 1$ are independent. Of course the relation

$$Q_1 + Q_2 + \dots + Q_u = 0$$

is easy to verify.

Since the linear function Q_i is orthogonal to B_j, and since B_j and $B_{j'}$ $(j \neq j')$ are orthogonal to each other, it is clear that in the case of a connected design, the estimation space is at least of the rank $b + u - 1$, as it contains the vectors corresponding to $Q_1, Q_2, \dots, Q_u, B_1, B_2, \dots, B_b$. On the other hand, the identity

$$B_1 + B_2 + \dots + B_b = T_1 + T_2 + \dots + T_u$$

shows that the vectors $\beta_1, \beta_2, \dots, \beta_b, \tau_1, \tau_2, \dots, \tau_u$ generating the estimation space cannot be all independent, as we must have

$$\beta_1 + \beta_2 + \dots + \beta_b = \tau_1 + \tau_2 + \dots + \tau_u$$

Accordingly the rank of the estimation space is exactly $b + u - 1$. Thus $b + u - 1$ d.f. belong to the estimates, and $N - b - u + 1$ degrees of freedom belong to error.

7. Let us now turn to the problem of estimation of any contrast between treatment effects. For this purpose we can enunciate the following theorem.

<u>Theorem</u>. The best estimate of the contrast

$$L = \ell_1t_1 + \ell_2t_2 + \dots + \ell_ut_u \tag{7.1}$$

is obtained by substituting in <u>L</u>, any set of values of <u>t</u> obtained by solving the following system of normal equations:

$$\begin{aligned} c_{11}t_1 + c_{12}t_2 + \dots + c_{1u}t_u &= Q_1 \\ c_{21}t_1 + c_{22}t_2 + \dots + c_{2u}t_u &= Q_2 \\ \dots \quad \dots \quad \dots \quad \dots \quad & \dots \\ c_{u1}t_1 + c_{u2}t_2 + \dots + c_{uu}t_u &= Q_u \end{aligned}$$

Proof: If $\Sigma q Q$ is the best estimate of L, we have

$$q_1 E(Q_1) + q_2 E(Q_2) + \dots + q_u E(Q_u) = \ell_1 t_1 + \ell_2 t_2 + \dots + \ell_u t_u$$

which gives rise to the following equations for determining the q's.

$$\begin{array}{l} C_{11}q_1 + C_{21}q_2 + \dots + C_{u1}q_u = \ell_1 \\ C_{12}q_1 + C_{22}q_2 + \dots + C_{u2}q_u = \ell_2 \\ \dots \quad \dots \quad \dots \quad \dots \quad .. \\ C_{1u}q_1 + C_{2u}q_2 + \dots + C_{uu}q_u = \ell_u \end{array}$$

Now if $\hat{t}_1, \hat{t}_2, \dots, \hat{t}_u$ is any particular solution of (7.2), then substituting these values in (7.2), multiplying these equations by $q_1, q_2, \dots, q_u$ respectively, adding and using (7.3), we find that

$$\ell_1 \hat{t}_1 + \ell_2 \hat{t}_2 + \dots + \ell_u \hat{t}_u = q_1 Q_1 + q_2 Q_2 + \dots + q_u Q_u$$
$$= \text{best estimate}$$

Corollary 1. Only u-1 of the equations (7.2) are independent. Since the general solution of the homogeneous equations corresponding to (7.2) is

$$(\theta, \theta, \dots, \theta)$$

So the general solution of (7.2) is

$$t_1 = \hat{t}_1 + \theta, \quad t_2 = \hat{t}_2 + \theta, \dots, t_u = \hat{t}_u + \theta \qquad (7.4)$$

Thus (as is to be expected), the difference between any two t's and in general the contrast

$$\ell_1 t_1 + \ell_2 t_2 + \dots + \ell_u t_u, \quad \Sigma \ell = 0$$

is uniquely determined. In order to render the solution of the normal equations unique, we may if we like take the equations (7.2) together with some arbitrary restraint

$$C_1 t_1 + C_2 t_2 + \dots + C_u t_u = 0 \qquad (7.5)$$

where $C_1 + C_2 + \dots + C_u \neq 0$. The unique solution thus obtained is bound to satisfy the normal equations, and on substitution in $\Sigma \ell t$ will give its best estimate. The reason why in the restraining condition (7.5) we must not take $\Sigma C = 0$ is that in this case $(C_1, C_2, \dots, C_u)$, will depend on the vectors $(C_{i1}, C_{i2}, .., C_{iu})$, and so the imposition of this condition will not lead to a unique solution of the normal equations.

Corollary 2. Let us find the variance of the best estimate of $\Sigma \ell t$. We have seen that the best estimate will be ΣqQ, where the Q's are given by (7.3). Now

$$
\begin{aligned}
V(q_1Q_1 + q_2Q_2 + \dots + q_uQ_u) &= \sum_{i,i'} \mathrm{Cov}(Q_i, Q_{i'}) \\
&= \sigma^2 \sum_{i,i'} q_i q_{i'} c_{ii'} \\
&= \{ q_1(c_{11}q_1 + c_{12}q_2 + \dots + c_{1u}q_u) \\
&\quad + q_2(c_{21}q_1 + c_{22}q_2 + \dots + c_{2u}q_u) \\
&\quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\
&\quad + q_u(c_{u1}q_1 + c_{u2}q_2 + \dots + c_{uu}q_u)\} \sigma^2 \\
&= (q_1\ell_1 + q_2\ell_2 + \dots + q_u\ell_u)\sigma^2 \qquad (7.6)
\end{aligned}
$$

On comparing (7.2) and (7.3) it appears that if the normal equations lead to

$$
\begin{aligned}
t_1 &= c_{11}Q_1 + c_{12}Q_2 + \dots + c_{1u}Q_u \\
t_2 &= c_{21}Q_1 + c_{22}Q_2 + \dots + c_{2u}Q_u \\
&\dots \quad \dots \quad \dots \quad \dots \quad \dots \\
t_u &= c_{u1}Q_1 + c_{u2}Q_2 + \dots + c_{uu}Q_u
\end{aligned}
\qquad (7.7)
$$

then $q_1, q_2, \dots, q_u$ may be obtained from $t_1, t_2, \dots, t_u$ in (7.7), on replacing Q's by ℓ's. Thus

$$q_i = \ell_1 c_{i1} + \ell_2 c_{i2} + \dots + \ell_u c_{iu}$$

Substituting in (7.6) we see that the variance of the best estimate is

$$\sigma^2 \sum_{i,i'} \ell_i \ell_{i'} c_{ii'} \qquad (7.8)$$

Thus the variance of the best estimate of $\Sigma \ell t$, involves only the coefficients in the algebraic solution of the normal equations. In particular the variance of the estimate of $t_i - t_{i'}$ is

$$(c_{ii} - c_{ii'} - c_{i'i} + c_{i'i'})\sigma^2 \qquad (7.9)$$

8. Continuing our study of the connected design, let us now turn to the problem of partitioning the total sum of squares into its various constituents, and to the estimation of the error σ^2.

The s.s. due to all the observations is

$$S'^2 = \Sigma y_{ij}^2 \qquad (8.1)$$

If G is the grand total of all the observations, then G estimates

$$r_1t_1 + r_2t_2 + \dots + r_ut_u + k_1b_1 + k_2b_2 + \dots + k_bb_b$$

The s.s. due to

$$G = \Sigma y_{ij} \tag{8.2}$$

is then

$$G^2/N$$

which is known as the 'correction for the mean'. Subtracting it from S'^2, we get the quantity

$$S^2 = S'^2 - \frac{G^2}{N}$$

which in the language of the analysis of variance is called the total sum of squares. It corresponds to the N-1 d.f. belonging to the contrasts between the observations. S^2 is the sum of squares of the deviations of the individual observations from the general mean, and we can write $S^2 = \text{dev}^2 y$.

The total sum of squares can be partitioned into three orthogonal components.

(i) The s.s. due to the contrasts between the block totals, which carry b-1 d.f.

(ii) The s.s. belonging to the adjusted yields, which carry u-1 d.f.

(iii) The s.s. due to error, carrying N-b-u+1 d.f.

(i) The block totals $B_1, B_2, \dots, B_b$ are orthogonal to one another, the s.s. due to B_j being B_j^2/k_j. Now

$$B_1 + B_2 + \dots + B_b = G$$

Hence from the sum of squares

$$\sum_{j=1}^{b} B_j^2/k_j$$

due to the b degrees of freedom belonging to the B_j's, we have to subtract the s.s. due to G, in order to obtain the s.s. due to the b-1 d.f. carried by the contrasts between the block totals. This s.s. is therefore

$$S_b'^2 = \frac{B_1^2}{k_1} + \frac{B_2^2}{k_2} + \dots + \frac{B_b^2}{k_b} - \frac{G^2}{N}$$

In the language of the analysis of variance, the above is called the s.s. due to the blocks (ignoring treatments).

We can write $S_b'^2$ in the form

$$S_b'^2 = \sum_j k_j (\frac{B_j}{k_j} - \frac{G}{N})^2$$

Thus $S_b'^2$ is the sum of squares of the deviations of the block averages from the general average, weighted by the number of plots in the block. For the designs in current use, the block size is constant and we get

$$S_b'^2 = \frac{1}{k} \sum_j (B_j - \frac{G}{b})^2 = \frac{\text{dev}^2 B}{k}$$

(ii) Let us next find the s.s. due to the adjusted yields, Q_1, Q_2, ..., Q_u, i.e., the s.s. due to the estimates of the treatment contrasts.

Let γ_i be the vector corresponding to Q_i. Then

$$\sigma^2(\gamma_i \cdot \gamma_{i'}) = \text{Cov}(Q_i, Q_{i'}) = \sigma^2 C_{i,i'}$$

or $$(\gamma_i \cdot \gamma_{i'}) = C_{i,i'}, \quad i,i' = 1, 2, \ldots, u$$

If δ is the projection of the observation vector on the vector space generated by the γ's then $\eta = \delta' + \delta_o$, where

$$\delta_o = d_1\gamma_1 + d_2\gamma_2 + \ldots + d_u\gamma_u$$

and δ' is orthogonal to $\gamma_1, \gamma_2, \ldots, \gamma_u$. Hence

$$Q_i = (\eta \cdot \gamma_i) = (\delta_o \cdot \gamma_i) = d_1(\gamma_1 \cdot \gamma_i) + d_2(\gamma_2 \cdot \gamma_i) + \ldots + d_u(\gamma_u \cdot \gamma_i)$$

or $$C_{i1}d_1 + C_{i2}d_2 + \ldots + C_{iu}d_u = Q_i, \quad i = 1, 2, \ldots, u$$

Thus $d_1, d_2, \ldots, d_u$ satisfy the normal equations, and we may take them to be $\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_u$. Hence

$$\delta_o = \hat{t}_1\gamma_1 + \hat{t}_2\gamma_2 + \ldots + \hat{t}_u\gamma_u$$

Consequently the s.s. due to the adjusted yields is

$$\begin{aligned} S_t^2 &= (\hat{t}_1\gamma_1 + \hat{t}_2\gamma_2 + \ldots + \hat{t}_u\gamma_u)^2 \\ &= \sum_i \hat{t}_i \{\hat{t}_1(\gamma_1\ \gamma_i) + \hat{t}_2(\gamma_2\ \gamma_i) + \ldots + \hat{t}_u(\gamma_u\ \gamma_i)\} \\ &= \sum_i \hat{t}_i(C_{i1}\hat{t}_1 + C_{i2}\hat{t}_2 + \ldots + C_{iu}\hat{t}_u) \\ &= \hat{t}_1Q_1 + \hat{t}_2Q_2 + \ldots + \hat{t}_uQ_u \end{aligned}$$

(iii) Finally, the s.s. due to error is obtained by subtracting from the total sum of squares, the s.s. corresponding to the b-1 d.f. carried by the block contrasts, and the s.s. corresponding to the u-1 d.f. carried by the adjusted yields. Thus the s.s. due to error is

$$S_e^2 = S^2 - S_b'^2 - S_t^2$$

The estimate of the per lot variance σ^2, based on the N-b-u+1 d.f. belonging to error is

$$s = \frac{S_e^2}{N - b - u + 1}$$

The general scheme of the analysis of variance for any connected design can therefore be given as

Due to	d.f.	s.s.	mean square
Treatments Eliminating Blocks	u - 1	$S_t^2 = \sum_i t_i Q_i$	$s_t^2 = S_t^2/(u-1)$
Blocks ignoring treatments	b - 1	$S_b^2 = \Sigma\, k_i \left(\frac{B_i}{k_i} - \frac{G}{N}\right)^2$ $= (\text{dev}^2 B)/k$ when the block size is constant	
Error	N-b-u+1	S_e^2 (by subtraction)	$s_e^2 = S_e^2/(N-b-u+1)$
Total	N - 1	$\text{dev}^2 y$	

To test for the hypothesis that the treatments are significantly differentiated, we have to use the F-statistic

$$F = s_t^2/s_e^2$$

with u-1, N-b-u+1 d.f.

Suppose this result comes out significant. We can then proceed to test whether two treatments _i_ and _i'_ are significantly different. Now the estimate of $t_i - t_{i'}$ is $\hat{t}_i - \hat{t}_{i'}$, where

$$V(\hat{t}_i - \hat{t}_{i'}) = (c_{ii} - c_{ii'} - c_{i'i} + c_{i'i'})\sigma^2$$

Hence the required difference is tested by using the t-statistic

$$t = (\hat{t}_i - \hat{t}_{i'})/(s_e\sqrt{c_{ii} - c_{ii'} - c_{i'i} + c_{i'i'}})$$

with N-b-u+1 d.f.

9. Ex. (i) Balanced Incomplete Block Designs

In a balanced incomplete block design, there are <u>u</u> treatments, arranged in <u>b</u> blocks, such that each block has <u>k</u> plots, and each treatment is replicated <u>r</u> times. Further, each pair of treatments must occur together in exactly λ blocks. It is easy to see that

$$bk = ur, \quad \lambda(u - 1) = r(k - 1)$$

Also, Fisher has proved the inequality

$$b \geqq u \quad \text{or} \quad k \leqq r$$

The actual designs are listed in Fisher and Yates tables when $r \leqq 10$.

Let us consider the problem of analysis of these designs. Now we have

$$C_{ii} = r(1 - 1/k), \quad C_{ii'} = -\lambda/k, \quad (i \neq i').$$

Hence the normal equations are

$$-\frac{\lambda}{k}t_1 - \frac{\lambda}{k}t_2 - \ldots + r(1-\frac{1}{k})t_i - \ldots - \frac{\lambda}{k}t_u = Q_i, \quad (i = 1, 2, \ldots, u)$$

Taking these together with the restriction

$$t_1 + t_2 + \ldots + t_u = 0$$

we get

$$(r - \frac{r}{k} + \frac{\lambda}{k})\hat{t}_i = Q_i, \quad (i = 1, 2, \ldots, u)$$

Now

$$r - \frac{r}{k} + \frac{\lambda}{k} = \frac{r(k-1) + \lambda}{k} = \frac{\lambda(u-1) + \lambda}{k} = \frac{u\lambda}{k}$$

$$\therefore \hat{t}_i = \frac{k}{u\lambda} Q_i = \frac{kr}{u\lambda}\frac{Q_i}{r} = \frac{1}{E}\frac{Q_i}{r}$$

where

$$E = \frac{u\lambda}{kr} = \frac{u(k-1)}{k(u-1)} = \frac{1 - \frac{1}{k}}{1 - \frac{1}{u}} < 1$$

is defined as the efficiency factor for a reason which will presently appear.

The contrast between the i-th and i'-th treatment is estimated by

$$\hat{t}_i - \hat{t}_{i'} = \frac{1}{E}\frac{Q_i - Q_{i'}}{r}$$

$$V(\hat{t}_i - \hat{t}_{i'}) = \sigma^2(C_{ii} - C_{ii'} - C_{i'i} + C_{i'i'}) = \frac{2\sigma^2}{rE}$$

In ordinary randomized block design, the corresponding variance, for the same number of replications would be $2\sigma^2/r$. Hence if there is no reduction in the per plot variance due to the reduction of the block size, the variance of $\hat{t}_i - \hat{t}_{i'}$ is increased in the ration $1/E$ or the information (which is defined as the reciprocal of the variance) is decreased in the ratio E. Hence, E is called the efficiency factor. Of course this theoretical loss of efficiency will in general be more than offset by the reduction in the error variance per plot. The analysis of variance is

Due to	d.f.	s.s.	mean square
Treatments Eliminating blocks	$u - 1$	$S_t^2 = dev^2 Q/rE$	$s_t^2 = S_t^2/(u-1)$
Blocks ignoring treatments	$b - 1$	$S_b'^2 = dev^2 B/k$	
Error	$N-b-u+1$ where $N = bk = ur$	S_e^2 (by subtraction)	$s_e^2 = S_e^2/(N-b-u+1)$
Total	$N - 1$	$dev^2 y$	

Hence to test whether the treatments are significantly differentiated we have to use the F-statistic

$$F = s_t^2/s_e^2$$

with $u-1$, $N-b-u+1$ d.f.

To test for the significance of the difference between any two treatments, we have to use the t-statistic

$$t = (Q_i - Q_{i'})/s_e\sqrt{2rE}$$

with $N-b-u+1$ d.f.

10. The Lattice Design

Consider a kxk two-dimensional square. The k^2 cells correspond to u $(=k^2)$ treatments. We can form k blocks by taking sets of k treatments occuring in the same row. Similarly k blocks may be obtained by taking sets of treatments in the same column. For example from the 5x5 square

1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

we get the two sets of blocks

Set I	Set II
(1, 2, 3, 4, 5)	(1, 6, 11, 16, 21)
(6, 7, 8, 9, 10)	(2, 7, 12, 17, 22)
(11, 12, 13, 14, 15)	(3, 8, 13, 18, 23)
(16, 17, 18, 19, 20)	(4, 9, 14, 19, 24)
(21, 22, 23, 24, 25)	(5, 10, 15, 20, 25)

Each set of blocks gives one complete replication. The two replications are orthogonal in the sense that the treatments in any block of one replication are distributed one each among the blocks of the second replication. We can get another orthogonal replication by taking a 5x5 Latin square, and taking those varieties which correspond to the letters of the Latin square. Thus, if we take the Latin square (L_1), shown below

C	E	D	B	A
A	C	B	E	D
E	B	A	D	C
B	D	C	A	E
D	A	E	C	B

then the blocks of the third replication are given by

Set III

(5, 6, 13, 19, 22)
(4, 8, 12, 16, 25)
(1, 7, 15, 18, 24)
(3, 10, 14, 17, 21)
(2, 9, 11, 20, 23)

If there exists a Latin square orthogonal to (L_2), then another replication may be taken corresponding to this Latin square. It is known that when k is a prime or a power of a prime, there always exists a set of k-1 mutually orthogonal Latin squares. The practically useful values of k are $k \leq 9$, so that only in the case k=6, there does not exist a Latin square orthogonal to a given one. When k=6, we can go up to three replications only. Of course, we may stop at any number of replications, but in case k is a prime or a prime power, and we go on to the maximum possible number of orthogonal replications, viz., k+1, then every pair will occur once, and we get a 'balanced lattice', which is identical with the balanced incomplete design with parameters

$$u = k^2, \quad b = k(k + 1), \quad r = k + 1, \quad k = k, \quad \lambda = 1$$

A lattice design with m orthogonal replications may be called an m-ple lattice. Of course, we may extend the number of blocks by repeating the whole design, say n times, so that we get mn replications, but in general it will be preferable to add new orthogonal replications, instead of duplicating or triplicating the old ones. But in the case k=6, not more than three orthogonal replications are available, and if it is desired to have more replications, we may duplicate or triplicate each replication getting 6 or 9 replications in all.

These remarks may be illustrated by taking the case k=5. In this case the complete set of orthogonal Latin squares, consists of four squares. Thus, three other squares orthogonal to (L_1) and mutually orthogonal to one another are

(L_2)

E	B	A	D	C
A	C	B	E	D
D	A	E	C	B
C	E	D	B	A
B	D	C	A	E

(L_3)

B	D	C	A	E
A	C	B	E	D
C	E	D	B	A
D	A	E	C	B
E	B	A	D	C

(L_4)

D	A	E	C	B
A	C	B	E	D
B	D	C	A	E
E	B	A	D	C
C	E	D	B	A

and the corresponding sets of blocks are

Set IV	Set V	Set VI
(3, 6, 12, 20, 24)	(4, 6, 15, 17, 23)	(2, 6, 14, 18, 25)
(2, 8, 15, 19, 21)	(1, 8, 14, 20, 22)	(5, 8, 11, 17, 24)
(5, 7, 14, 16, 23)	(3, 7, 11, 19, 25)	(4, 7, 13, 20, 21)
(4, 10, 11, 18, 22)	(2, 10, 13, 16, 24)	(1, 10, 12, 19, 23)
(1, 9, 13, 17, 25)	(5, 9, 12, 18, 21)	(3, 9, 15, 16, 22)

If we take all the six replications possible, we shall get a balanced lattice, but we can stop at any stage. Thus, by taking only the first four sets, we would get four replications (quadruple lattice with 25 treatments).

Let us now consider the analysis for a lattice design with k^2 treatments, laid out in <u>mk</u> blocks, consisting of <u>m</u> orthogonal replications. We then have

$$u = k^2, \quad b = km, \quad r = m.$$

Also, $\lambda_{ii'} = 1$, if the treatments <u>i</u> and <u>i'</u> correspond to the same row, same column, or to the same letter of one of the Latin squares. Otherwise, $\lambda_{ii'} = 0$. Then

$$c_{ii} = m(1 - 1/k), \quad c_{ii'} = -\lambda_{ii'}/k$$

Denote by $S_r(t_i)$, $S_c(t_i)$, $S_j(t_i)$, respectively the sum of the effects of the treatments, in the same row, in the same column, corresponding to the same letter of the j-th Latin square. Similar meanings are given to $S_r(Q_i)$, $S_c(Q_i)$, $S_j(Q_i)$, with reference to the adjusted yields. Thus with example (k=5) considered before

$$S_r(t_1) = t_1 + t_2 + t_3 + t_4 + t_5, \quad S_r(Q_1) = Q_1 + Q_2 + Q_3 + Q_4 + Q_5$$

$$S_c(t_1) = t_1 + t_6 + t_{11} + t_{16} + t_{21}, \quad S_c(Q_1) = Q_1 + Q_6 + Q_{11} + Q_{16} + Q_{21}$$

$$S_1(t_1) = t_1 + t_7 + t_{15} + t_{18} + t_{24}, \quad S_1(Q_1) = Q_1 + Q_7 + Q_{15} + Q_{18} + Q_{24}$$

It should be noted that $S_r(t_1)$ could also be denoted by $S_r(t_2)$, $S_r(t_3)$, $S_r(t_4)$, or $S_r(t_5)$; and similarly for the other symbols. Now the i-th normal equation is

$$m(1 - \frac{1}{k}) + \frac{m}{k} t_i - \frac{1}{k} S_r(t_i) + S_c(t_i) + S_1(t_i) + \,..\, + S_{m-2}(t_i) = Q_i$$

or

$$mt_i - \frac{1}{k} S_r(t_i) + S_c(t_i) + S_1(t_i) + \ldots + S_{m-2}(t_i) = Q_i$$

Also, let us impose the restriction,

$$t_1 + t_2 + \dots + t_u = 0$$

Summing up over the row containing the i-th treatment, we get

$$mS_r(t_i) - S_r(t_i) = S_r(Q_i)$$

$$S_r(t_i) = S_r(Q_i)/(m-1)$$

Similarly

$$S_c(t_i) = S_c(Q_i)/(m-1)$$

Also, if we sum up over the treatments, which correspond to the same letter as the i-th treatment in the j-th Latin square, we have

$$S_j(t_i) = S_j(Q_i)/(m-1)$$

Hence finally we get

$$\hat{t}_i = \frac{1}{m} Q_i + \frac{1}{m(m-1)k} S_r(Q_i) + S_c(Q_i) + S_1(Q_i) + \dots + S_{m-2}(Q\ )$$

From this we get

$$c_{ii} = \frac{1}{m} + \frac{1}{k(m-1)}$$

For $i \neq i'$ we get

$$c_{ii'} = \frac{1}{m(m-1)k} \text{ or } 0$$

according as the treatments <u>i</u> and <u>i'</u> occur in a common block or do not occur in a common block. We may thus write

$$c_{ii'} = \frac{\lambda_{ii'}}{m(m-1)k}$$

Now $\qquad V(\hat{t}_i - \hat{t}_{i'}) = (c_{ii} - c_{ii'} - c_{i'i} + c_{i'i'})\sigma^2$

Hence if <u>i</u> and <u>i'</u> occur together in the same block

$$V(\hat{t}_i - \hat{t}_{i'}) = 2\sigma^2 \frac{1}{m} + \frac{1}{k(m-1)} - \frac{1}{m(m-1)k}$$

$$= \frac{2\sigma^2}{m} (1 - \frac{1}{k})$$

If <u>i</u> and <u>i'</u> do not occur together in the same block

$$V(\hat{t}_i - \hat{t}_{i'}) = 2\sigma^2 \frac{1}{m} + \frac{1}{k(m-1)}$$

$$= \frac{2\sigma^2}{m} 1 + \frac{m}{(m-1)k}$$

Notice that in the second case the variance will be slightly larger. Tests of significance can easily be carried out by applying the general formula.

N.B. If each of $\underline{m}$ orthogonal replications in an m-ple lattice on k^2 treatments is laid out $\underline{n}$ times, then we have

$$u = k^2, \quad b = mnk, \quad r = mn$$

Also, $\lambda_{ii'}$ is $\underline{n}$ if the i-th and i'-th treatment correspond to the same row, column, or the same letter of one of the Latin squares; otherwise it is 0. Thus in effect C_{ii} and $C_{ii'}$ are multiplied by $\underline{n}$, so that the new normal equations can be obtained from the old on replacing Q_i by Q_i/n for $i = 1, 2, \ldots, u$. Hence our estimates are given by

$$\hat{t}_i = \frac{1}{mn}Q_i + \frac{1}{mn(m-1)k}\left\{S_r(Q_i) + S_c(Q_i) + S_1(Q_i) + \ldots + S_{m-2}(Q_i)\right\}$$

Also we now have

$$C_{ii} = \frac{1}{n}\left\{\frac{1}{m} + \frac{1}{k(m-1)}\right\}, \quad C_{ii'} = \frac{1}{mn(m-1)k} \text{ or } 0$$

$$V(\hat{t}_i - \hat{t}_{i'}) = \frac{2\sigma^2}{mn}\left(1 + \frac{1}{k}\right), \text{ when } \underline{i} \text{ and } \underline{i'} \text{ occur together in } \underline{n} \text{ blocks}$$

$$= \frac{2\sigma^2}{mn}\left(1 + \frac{m}{(m-1)k}\right), \text{ when } \underline{i} \text{ and } \underline{i'} \text{ do not occur together in } \underline{n} \text{ blocks}$$

11. The Partailly Balanced Incomplete Block Designs

The Partially Balanced Designs are a general class of designs, which include as a special case, both the Balanced Incomplete Block designs and the lattice designs. Here, as in the case of lattice designs, two treatments are not always compared with the same accuracy.

The conditions for a partially balanced design are:

(i) Every treatment is replicated $\underline{r}$ times, and each block contains $\underline{k}$ plots.

(ii) With respect to any given treatment, the rest fall into $\underline{m}$ groups of $n_1, n_2, \ldots, n_m$ each, such that any treatment of the i-th group occurs with the given treatment λ_i times, the numbers λ_i and n_i being independent of the treatment with which we start. The treatments of the i-th group may be called i-associates of the given treatment.

(iii) If α is i-associate of β, then β is an i-associate of α, and the number of treatments which are at the same time j-associates of α, and k-associates of β, is p^i_{jk} and is independent of α and β.

If <u>u</u> be the total number of treatments, and <u>b</u> the number of blocks, then the following conditions are satisfied by the parameters occuring in the design.

$$
\left.\begin{aligned}
bk &= ur \\
u - 1 &= n_1 + n_2 + \dots + n_m \\
r(k - 1) &= n_1 \;{}_1 + n_2 \;{}_2 + \dots + n_m \;{}_m \\
p^i_{i1} + p^i_{i2} + \dots + p^i_{im} &= n_i - 1 \\
p^i_{j1} + p^i_{j2} + \dots + p^i_{jm} &= n_j, \quad (i \neq j) \\
n_i p^i_{jk} &= n_j p^j_{ik}
\end{aligned}\right\} \qquad (11.1)
$$

The parameters $u, b, r, k, \lambda_1, \lambda_2, \dots, \lambda_m, n_1, n_2, \dots, n_m$ are called parameters of the first kind, and the p^i_{jk}'s are called parameters of the second kind.

It shall now be shown that if a design satisfies the conditions of a partially balanced incomplete block design, then the analysis can be carried through by solving sets of m+1 linear equations. For the practically useful designs, <u>m</u> would be 2 or at most 3 (when m=1, we get a balanced incomplete block design).

Let us denote by $S_i(t)$ the sum of the effects of the i-associates of the treatment <u>t</u>, with a similar convention for $Q_i(t)$. Then our normal equations are

$$r(k-1)t - \lambda_1 S_1(t) - \lambda_2 S_2(t) - \dots - \lambda_m S_m(t) = kQ \qquad (11.2)$$

Summing up over i-associates we get

$$
\begin{aligned}
kQ_i(t) = r(k-1)S_i(t) &- \lambda_1 \left\{ p^1_{i1} S_1(t) + p^2_{i1} S_2(t) + \dots + p^m_{i1} S_m(t) \right\} \\
&- \lambda_2 \left\{ p^1_{i2} S_1(t) + p^2_{i2} S_2(t) + \dots + p^m_{i2} S_m(t) \right\} \\
&\qquad \dots \qquad\quad \dots \qquad\quad \dots \qquad\quad \dots \\
&- \lambda_m \left\{ p^1_{im} S_1(t) + p^2_{im} S_2(t) + \dots + p^m_{im} S_m(t) \right\} \\
&- \lambda_i n_i t
\end{aligned} \qquad (11.3)
$$

Now impose the restriction

$$t_1 + t_2 + \dots + t_u = 0$$

which may be written as

$$t + S_1(t) + S_2(t) + \dots + S_m(t) = 0$$

Then (11.3) can be written as

$$
\begin{aligned}
kQ_i(t) = &\left\{\lambda_i n_i - \lambda_1 p_{i1}^1 - \lambda_2 p_{i2}^1 - \dots - \lambda_m p_{im}^1\right\} S_1(t) \\
&+ \left\{\lambda_i n_i - \lambda_1 p_{i1}^2 - \lambda_2 p_{i2}^2 - \dots - \lambda_m p_{im}^2\right\} S_2(t) \\
&\quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\
&+ \left\{r(k-1) + \lambda_i n_i - \lambda_1 p_{i1}^i - \lambda_2 p_{i2}^i - \dots - \lambda_m p_{im}^i\right\} S_i(t) \\
&\quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\
&+ \left\{\lambda_i n_i - \lambda_1 p_{i1}^m - \lambda_2 p_{i2}^m - \dots - \lambda_m p_{im}^m\right\} S_m(t)
\end{aligned}
\tag{11.4}
$$

Let us set

$$
a_{i\ell} = \frac{1}{k}\left\{\lambda_i n_i - \lambda_1 p_{i1}^\ell - \lambda_2 p_{i2}^\ell - \dots - \lambda_m p_{im}^\ell\right\}, \quad i \neq \ell
$$

$$
a_{ii} = \frac{1}{k}\left\{r(k-1) + \lambda_i n_i - \lambda_1 p_{i1}^i - \lambda_2 p_{i2}^i - \dots - \lambda_m p_{im}^i\right\} \tag{11.5}
$$

Then we have

$$
\begin{aligned}
a_{11}S_1(t) + a_{12}S_2(t) + \dots + a_{1m}S_m(t) &= Q_1 \\
a_{21}S_1(t) + a_{22}S_2(t) + \dots + a_{2m}S_m(t) &= Q_2 \\
\dots \quad \dots \quad \dots \quad \dots \quad & \quad \dots \\
a_{m1}S_1(t) + a_{m2}S_2(t) + \dots + a_{mm}S_m(t) &= Q_m
\end{aligned}
\tag{11.6}
$$

Solving these we can express $S_1(t)$, $S_2(t)$, ..., $S_m(t)$ as

$$
\begin{aligned}
S_1(t) &= A_{11}Q_1(t) + A_{21}Q_2(t) + \dots + A_{m1}Q_m(t) \\
\dots \quad & \quad \dots \quad \dots \quad \dots \quad \dots \\
S_m(t) &= A_{1m}Q_1(t) + A_{2m}Q_2(t) + \dots + A_{mm}Q_m(t)
\end{aligned}
\tag{11.7}
$$

and finally

$$
\begin{aligned}
r(k-1)t = kQ &+ \left\{\lambda_1 A_{11} + \lambda_2 A_{12} + \dots + \lambda_m A_{1m}\right\} Q_1 \\
&+ \quad \dots \quad \dots \quad \dots \quad \dots \\
&+ \left\{\lambda_1 A_{m1} + \lambda_2 A_{m2} + \dots + \lambda_m A_{mm}\right\} Q_m
\end{aligned}
\tag{11.8}
$$

Thus our main task in solving the normal equations is to evaluate the constituents of $((a_{i\ell}))$, and then solve (11.6).

It appears from (11.8) that if the treatment <u>i'</u> is the j-th associate of the treatment <u>i</u>, then

$$
C_{ii} = k/r(k-1)
$$

$$
C_{ii'} = \left\{\lambda_1 A_{j1} + \lambda_2 A_{j2} + \dots + \lambda_m A_{jm}\right\}/r(k-1)
$$

$$
V(\hat{t}_i - \hat{t}_{i'}) = \frac{2\sigma^2}{r(k-1)}\left\{k - \lambda_1 A_{j1} - \lambda_2 A_{j2} - \dots - \lambda_m A_{jm}\right\}
$$

Tests of significance can be carried out in the usual way.

12. Two way elimination of heterogeneity.

In this case the blocks are classified into two ways, so that each experimental unit is a part of one block of either system. Thus, in the case of a field experiment the design is laid out in the form of a square or a rectangle, with 'rows' and 'columns'. We shall use the names 'rows' and 'columns' even in the general case. Let there be $\underline{k}$ rows and $\underline{k'}$ columns. Then there are besides the treatment effects

$$t_1, t_2, \ldots, t_u$$

the $\underline{k}$ parameters corresponding to the row effects, viz.,

$$b_1, b_2, \ldots, b_k$$

and the $\underline{k'}$ parameters corresponding to the column effects

$$b_1', b_2', \ldots, b_{k'}'$$

There are now $\underline{kk'}$ plots. Let the yield in the plot j, j', i.e., the plot occuring in the j-th row and j'-th column, be denoted by $y_{jj'}$. If the i-th treatment occurs in this plot, then the equations of expectation are

$$E(y_{jj'}) = t_i + b_j + b_{j'}'$$

Let $\beta_1, \ldots, \beta_k, \beta_1', \ldots, \beta_{k'}', \tau_1, \ldots, \tau_u$ be the coefficients of $b_1, \ldots, b_k, b_1', \ldots, b_{k'}', t_1, \ldots, t_u$ when the equations of expectation are written out in full. If η is the vector of the observed yields, clearly

$$(\eta\cdot\beta_j) = B_j, \quad (\eta\cdot\beta_{j'}') = B_{j'}', \quad (\eta\cdot\tau_i) = T_i$$

where B_j is the total yield of the j-th row, $B_{j'}'$ is the total yield of the j'-th column, and T_i is the total yield of the i-th treatment. Hence from the general theory we get the following theorem:

Theorem: Corresponding to any estimable parametric function Π of the row, column and treatment effects, there exists a unique linear function Y_o of the row, column and treatment totals, such that

$$E(Y_o) = \Pi$$

This linear function Y_o is the best estimate of Π.

Ex. Before proceeding to the general theory let us apply the above to a Latin square design. Such a design is formed by taking a $\underline{k \times k}$ Latin square, and assigning $\underline{k}$ treatments corresponding to the letters. Of course, there must be proper randomization.

Here $k'=k$, as there are k rows and k columns. Also, $u=k$. Let

B	D	C	A	E
A	C	B	E	D
C	E	D	B	A
D	A	E	C	B
E	B	A	D	C

$$S(b) = b_1 + b_2 + \dots + b_k$$

$$S(b') = b_1' + b_2' + \dots + b_k'$$

$$S(t) = t_1 + t_2 + \dots + t_k$$

Now

$$E(T_i) = kt_i + S(b) + S(b')$$

Also if

$$G = T_1 + \dots + T_k = B_1 + \dots + B_k = B_1' + \dots + B_k' \qquad (12.1)$$

is the grand total, then

$$E(G) = k\left\{S(t) + S(b) + S(b')\right\}$$

$$\therefore\quad E\left\{\frac{1}{k}\left(T_i - \frac{G}{k}\right)\right\} = \left\{t_i - \frac{S(t)}{k}\right\}, \quad i = 1, 2, \dots, k$$

If we multiply those equations by $\ell_1, \ell_2, \dots, \ell_k$, where $\Sigma\ell = 0$, and add, we at once see that

$$E\left\{(\ell_1T_1 + \ell_2T_2 + \dots + \ell_kT_k)/k\right\} = \ell_1t_1 + \ell_2t_2 + \dots + \ell_kt_k \qquad (12.2)$$

Thus the treatment contrast $\ell_1t_1 + \dots + \ell_kt_k$ is estimated by

$$(\ell_1T_1 + \ell_2T_2 + \dots + \ell_kT_k)/k$$

In particular $t_i - t_{i'}$ is estimated by

$$\frac{1}{k}\,(T_i - T_{i'})$$

Similar estimates can be obtained for row contrasts in terms of row totals, and the column contrasts in terms of column totals.

We have hitherto accounted for $3(k-1)$ degrees of freedom, viz., those belonging to the contrasts between the T's, the contrasts between the B's, and the contrasts between the B''s. Due to the relation (12.1), there is only one other independent linear function of the T, B and B''s, viz., G. Hence the estimation space is of rank $3k-2$ and $k^2-3k+2 = (k-1)(k-2)$ degrees of freedom belonging to error.

Clearly $T_1, \dots, T_k$ are orthogonal. The s.s. belonging to the treatment contrasts is therefore

$$\frac{1}{k}(T_1^2 + T_2^2 + \dots + T_k^2 - \frac{G^2}{k}) = \frac{1}{k}\,\mathrm{dev}^2 T$$

with similar expressions for the s.s. due to row contrasts, and the column contrasts. The s.s. due to G is of course G^2/k^2, which is the correction for the mean. The s.s. due to error is therefore

$$S_e^2 = \Sigma\, y_{jj'}^2 - \frac{G^2}{k^2} - \frac{\mathrm{dev}^2 T}{k} - \frac{\mathrm{dev}^2 B}{k} - \frac{\mathrm{dev}^2 B'}{k}$$

$$= \mathrm{dev}^2 y - \frac{\mathrm{dev}^2 T}{k} - \frac{\mathrm{dev}^2 B}{k} - \frac{\mathrm{dev}^2 B'}{k}$$

Hence we have the following analysis of variance

Due to	d.f.	s.s.	Mean square
Treatment Contrasts	$k - 1$	$S_t^2 = \mathrm{dev}^2 T/k$	$s_t^2 = S_t^2/(k-1)$
Row Contrasts	$k - 1$	$S_b^2 = \mathrm{dev}^2 B/k$	$s_b^2 = S_b^2/(k-1)$
Column Contrasts	$k - 1$	$S_{b'}^2 = \mathrm{dev}^2 B'/k$	$s_{b'}^2 = S_{b'}^2/(k-1)$
Error	k^2-3k+2	S_e^2 (by subtraction)	$s_e^2 = S_e^2/(k-1)(k-2)$
Total	k^2-1	$\mathrm{dev}^2 y$	

An estimate of the per-plot error variance σ^2 is given by

$$s_e^2 = S_e^2/(k^2-3k+2)$$

To test whether the treatments are significantly differentiated, we use the F-statistic

$$F = s_t^2/s_e^2$$

with k-1, (k-1)(k-2) degrees of freedom.

To test for the significance of the difference between the i-th and i'-th treatments we use the t-statistic

$$t = (T_i - T_{i'})/(s_e\sqrt{2k})$$

with (k-1)(k-2) degrees of freedom.

13. Normal equations for two way elimination in the general case.

$$E(T_i) = r_i t_i + \sum_{j=1}^{k} n_{ij} b_j + \sum_{j'=1}^{k'} n'_{ij'} b'_{j'}$$

where $n_{ij} = 1$ or 0 according as the i-th treatment does or does not occur in the j-th row. Likewise $n'_{ij'} = 1$ or 0 according as the i-th treatment does or does not occur in the i'-th column. Now

$$E(B_j) = k'b_j + b'_1 + b'_2 + \dots + b'_{k'} + \sum_{i=1}^{u} n_{ij}t_i$$

$$E(B'_{j'}) = kb'_{j'} + b_1 + b_2 + \dots + b_k + \sum_{i=1}^{u} n'_{ij'}t_i$$

The adjusted yield Q_i is now defined to be

$$Q_i = T_i - \frac{1}{k'}\sum_{j=1}^{k} n_{ij}B_j - \frac{1}{k}\sum_{j'=1}^{k'} n'_{ij'}B'_{j'} + \frac{r_i G}{kk'}$$

A little calculation shows that $E(Q_i)$ is given by

$$E(Q_i) = r_i t_i - \frac{1}{k'}(\lambda_{i1}t_1 + \dots + \lambda_{iu}t_u) - \frac{1}{k}(\lambda'_{i1}t_1 + \dots + \lambda'_{iu}t_u)$$
$$+ \frac{1}{kk'}(r_1t_1 + \dots + r_ut_u)$$

where $\lambda_{ii'}$ is the number of rows in which the i-th and i'-th treatments come together, and $\lambda'_{ii'}$ is the number of columns in which the i-th and i'-th treatments come together, and $\lambda_{ii} = \lambda'_{ii} = r_i$ $\quad \therefore \quad E(Q_i) = C_{i1}t_1 + \dots + C_{iu}t_u$

where

$$C_{ii'} = \left(\frac{r_i r_{i'}}{kk'} - \frac{\lambda_{ii'}}{k'} - \frac{\lambda'_{ii'}}{k}\right), \quad i \neq i'$$

$$C_{ii} = r_i\left(1 - \frac{1}{k} - \frac{1}{k'} + \frac{r_i}{kk'}\right)$$

These are the coefficients which take the place of the coefficients in the theory of one way elimination.

By calculating the expectation of an arbitrary linear function $\underline{Y}$ of $T_1, \dots, T_u, B_1, \dots, B_k, B'_1, \dots, B'_{k'}$, we can show that in order that $\underline{Y}$ may be an estimate of pure functions of treatment effects. $\underline{Y}$ must be of the form

$$q_1Q_1 + q_2Q_2 + \dots + q_uQ_u$$

Also, a little calculation shows that

$$V(Q_i) = C_{ii}\sigma^2$$

$$Cov(Q_i, Q_{i'}) = C_{ii'}\sigma^2$$

The idea of connectedness will now be slightly generalized. In finding whether two treatments are connected, we have to allow connection both through rows and through columns. If any two treatments are connected, then the design may be said to be a connected design. For such a design any treatment contrast is estimable, and the estimate is obtained by substituting in the contrast solution of the normal equations

$$C_{i1}t_1 + C_{i2}t_2 + \dots + C_{iu}t_u = Q_i$$

The s.s. due to the u-1 d.f. belonging to the adjusted yields can be proved to be

$$\hat{t}_1 Q_1 + \hat{t}_2 Q_2 + \dots + \hat{t}_u Q_u$$

It is readily seen that the row and column contrasts are orthogonal to the Q's. Hence we get the following scheme for the analysis of variance

Due to	d.f.	s.s.	mean square
Treatments eliminating rows & cols.	$u - 1$	$S_t^2 = \Sigma\hat{t}Q$	$s_t^2 = S_t^2/(u-1)$
Row cont. (ignoring treatments)	$k - 1$	$dev^2 B/k$	
Col. cont. (ignoring treatments)	$k' - 1$	$dev^2 B'/k'$	
Error	$kk'-u-k-k'+2$	S_e^2 (by subtraction)	$s_e^2 = \dfrac{S_e^2}{kk'-u-k-k'+2}$
Total	$kk' - 1$	$dev^2 y$	

14. Youden's Squares.

Sometimes in a design the position within the block is important. The classical example is due to Youden, in his studies on the tobacco mosaic virus. He found that the response to treatments also depends on the position of the leaf on the plant. If the number of leaves is sufficient so that every treatment can be applied to one leaf, then we get an ordinary Latin square, in which the trees are columns and the leaves belonging to the same position constitute the rows. But if the number of treatments is larger than the number of leaf positions available, then we must have incomplete columns. Youden used a design in which the columns constituted a balanced incomplete block design, whereas the rows were complete. These designs are known as Youden's squares.

For example, consider the following, in which there are eleven treatments but rows are incomplete, each containing only five treatments.

3	4	5	6	7	8	9	10	11	1	2
9	10	11	1	2	3	4	5	6	7	8
5	6	7	8	9	10	11	1	2	3	4
4	5	6	7	8	9	10	11	1	2	3
1	2	3	4	5	6	7	8	9	10	11

Of course, here the design is given in a schematic form and we must randomize rows and columns. Also, the actual physical treatments have to be assigned numbers 1 to 11 in a random order. It will be seen that every treatment is replicated 5 times, and any pair of treatments occur in the same column (i.e., on the same tree) twice.

In the general case, if there are u treatments, we have b columns, each with k treatments, so that each treatment is replicated r times, and each pair of treatments occurs in the same column λ times. Since the rows are complete, each contains u treatments.

Thus we must have $b = u$, and consequently $r = k$. Hence the balanced incomplete block design formed by the columns is a symmetrical balanced incomplete block design. It can be shown that every symmetrical balanced incomplete design can be converted into a Youden's square, i.e., the blocks can be so arranged that in a particular position within the block, each treatment just occurs once.

Let us now consider the problem of analysis. Here

$$k = k, \quad k' = u, \quad r_i = r, \quad \lambda_{ii'} = r, \quad \lambda'_{ii'} = \lambda$$

Hence

$$C_{ii'} = \left(\frac{r_i r'_i}{kk'} - \frac{\lambda_{ii'}}{k'} - \frac{\lambda'_{ii'}}{u}\right)$$

$$= \left(\frac{r^2}{ku} - \frac{r}{u} - \frac{}{k}\right)$$

$$= -\lambda/k \quad \text{since} \quad r = k$$

$$c_{ii} = r_i(1 - \frac{1}{k} - \frac{1}{k'} + \frac{r_i}{kk'})$$

$$= r(1 - \frac{1}{k})$$

These coefficients are the same as in the case of an ordinary balanced incomplete block design, and we find as before that $t_i - t_{i'}$ is estimated by

$$\hat{t}_i - \hat{t}_{i'} = (Q_i - Q_i')/rE, \text{ where } E = u\lambda/rk$$

and that

$$V(\hat{t}_i - \hat{t}_{i'}) = 2\sigma^2/rE$$

The only difference is in the estimation of σ^2. The number of d.f. belonging to error is

$$ur - 2u - r + 2$$

The s.s. due to error is

$$\text{dev}^2 y - \Sigma\, \hat{t}Q - \frac{\text{dev}^2 B}{u} - \frac{\text{dev}^2 B'}{k}$$

Thus there is a loss of r-1 d.f. in the estimation of error. this error is offset by the reduction in σ^2 (the per unit variance). If this reduction is large enough, i.e., the positions within the block are sufficiently differentiated, then we shall gain by using a Youden's square in place of an ordinary balanced incomplete block design.

15. Missing variates

Consider the problem of linear estimation with variates $y_1, \ldots, y_n; x_1, \ldots, x_k$, with a common variance σ^2, and expectations given by

$$\begin{aligned} E(y_1) &= a_{11}p_1 + a_{12}p_2 + \cdots + a_{1m}p_m \\ E(y_2) &= a_{21}p_1 + a_{22}p_2 + \cdots + a_{2m}p_m \\ &\cdots \\ E(y_n) &= a_{n1}p_1 + a_{n2}p_2 + \cdots + a_{nm}p_m \end{aligned} \qquad (15.1)\text{A}$$

$$\begin{aligned} E(x_1) &= b_{11}p_1 + b_{12}p_2 + \cdots + b_{1m}p_m \\ &\cdots \\ E(x_k) &= b_{k1}p_1 + b_{k2}p_2 + \cdots + b_{km}p_m \end{aligned} \qquad (15.1)\text{B}$$

Let $\alpha_j = (a_{1j}, a_{2j}, \ldots, a_{nj}, 0, 0, \ldots, 0)$

$\beta_j = (0, 0, \ldots, 0, b_{1j}, b_{2j}, \ldots, b_{kj})$

$$\gamma_j = (a_{1j}, a_{2j}, \dots, a_{nj}, b_{1j}, b_{2j}, \dots, b_{kj}) = \alpha_j + \beta_j$$
$$\eta = (y_1, y_2, \dots, y_n, 0, 0, \dots, 0)$$
$$\xi = (0, 0, \dots, 0, x_1, x_2, \dots, x_k)$$
$$f = (y_1, y_2, \dots, y_n, x_1, x_2, \dots, x_k) = \eta +$$

The normal equations are

$$\begin{aligned}
(\gamma_1 \cdot \gamma_1)p_1 + (\gamma_1 \cdot \gamma_2)p_2 + \dots + (\gamma_1 \cdot \gamma_m)p_m &= (\gamma_1 \cdot f) \\
(\gamma_2 \cdot \gamma_1)p_1 + (\gamma_2 \cdot \gamma_2)p_2 + \dots + (\gamma_2 \cdot \gamma_m)p_m &= (\gamma_2 \cdot f) \\
\dots \qquad \dots \qquad \dots \qquad \dots &\qquad \dots \\
(\gamma_m \cdot \gamma_1)p_1 + (\gamma_m \cdot \gamma_2)p_2 + \dots + (\gamma_m \cdot \gamma_m)p_m &= (\gamma_m \cdot f)
\end{aligned} \tag{15.2}$$

Now suppose that the variates $\underline{x}$ are missing. We make the analysis as is they were present, and minimize the sum of squares due to error, which is

$$S_e^2 = f^2 - (\gamma_1 \cdot f)p_1 - (\gamma_2 \cdot f)p_2 - \dots - (\gamma_m \cdot f)p_m$$

Where $p_1, \dots, p_m$ are solutions of the normal equations

$$\frac{dS_e^2}{dx_i} = 2x_i - (\gamma_1 \cdot f)\frac{dp_1}{dx_i} - \dots - (\gamma_m \cdot f)\frac{dp_m}{dx_i} - p_1 b_{i1} - \dots - p_m b_{im} \tag{15.3}$$

Differentiating (15.2) we get

$$(\gamma_1 \cdot \gamma_1)\frac{dp_1}{dx_i} + \dots + (\gamma_1 \cdot \gamma_m)\frac{dp_m}{dx_i} = b_{i1}, \; (i = 1, 2, \dots, m) \tag{15.4}$$

Multiplying the equation (15.4) by $p_1, \dots, p_m$, and adding, we find on using (15.2), that

$$(\gamma_1 \cdot f)\frac{dp_1}{dx_i} + \dots + (\gamma_m \cdot f)\frac{dp_m}{dx_i} = p_1 b_{i1} + \dots + p_m b_{im}$$

Hence

$$\frac{1}{2}\frac{dS_e^2}{dx_i} = x_i - p_1 b_{i1} - p_2 b_{i2} - \dots - p_m b_{im}$$

Thus to minimize S_e^2 we must put

$$x_i = p_1 b_{i1} + p_2 b_{i2} + \dots + p_m b_{im}, \; (i = 1, 2, \dots, k) \tag{15.5}$$

The normal equations (15.2) can be written as

$$\begin{array}{l}
[(\alpha_1\cdot\alpha_1)+(\beta_1\cdot\beta_1)]\,p_1 + \dots + [(\alpha_1\cdot\alpha_m)+(\beta_1\cdot\beta_m)]\,p_m = (\alpha_1\cdot\eta)+(\beta_1\cdot\xi) \\
[(\alpha_2\cdot\alpha_1)+(\beta_2\cdot\beta_1)]\,p_1 + \dots + [(\alpha_2\cdot\alpha_m)+(\beta_2\cdot\beta_m)]\,p_m = (\alpha_2\cdot\eta)+(\beta_2\cdot\xi) \\
\quad\dots \qquad \dots \qquad \dots \qquad \dots \\
[(\alpha_m\cdot\alpha_1)+(\beta_m\cdot\beta_2)]\,p_1 + \dots + [(\alpha_m\cdot\alpha_m)+(\beta_m\cdot\beta_m)]\,p_m = (\alpha_m\cdot\eta)+(\beta_m\cdot\xi)
\end{array} \quad (15.6)$$

But from (15.5), if $x_1, \dots, x_k$ are the values of x which minimize S_e^2, then

$$\begin{array}{l}
(\beta_1\cdot\xi) = p_1(\beta_1\cdot\beta_1) + p_2(\beta_1\cdot\beta_2) + \dots + p_m(\beta_1\cdot\beta_m) \\
(\beta_2\cdot\xi) = p_1(\beta_2\cdot\beta_1) + p_2(\beta_2\cdot\beta_2) + \dots + p_m(\beta_2\cdot\beta_m) \\
\quad\dots \qquad \dots \qquad \dots \qquad \dots \qquad \dots \\
(\beta_m\cdot\xi) = p_1(\beta_m\cdot\beta_1) + p_2(\beta_m\cdot\beta_2) + \dots + p_m(\beta_m\cdot\beta_m)
\end{array} \quad (15.7)$$

Hence

$$\begin{array}{l}
(\alpha_1\cdot\alpha_1)p_1 + (\alpha_1\cdot\alpha_2)p_2 + \dots + (\alpha_1\cdot\alpha_m)p_m = (\alpha_1\cdot\eta) \\
(\alpha_2\cdot\alpha_1)p_1 + (\alpha_2\cdot\alpha_2)p_2 + \dots + (\alpha_2\cdot\alpha_m)p_m = (\alpha_m\cdot\eta) \\
\quad\dots \qquad \dots \qquad \dots \qquad \dots \qquad \dots \\
(\alpha_m\cdot\alpha_1)p_1 + (\alpha_m\cdot\alpha_2)p_2 + \dots + (\alpha_m\cdot\alpha_m)p_m = (\alpha_m\cdot\eta)
\end{array} \quad (15.8)$$

This shows that if those values of x are taken which minimize S_e^2, then the solutions of the normal equations would be the same as if we had formed them by neglecting the missing variates.

Hence we get Fisher's principle: If the variates $x_1, x_2, \dots, x_k$ are missing, we can proceed as if they are present, providing that we put for them values which would minimize the sum of squares. The procedure would not change the estimates at all.

16. Let us now consider what would be the effect on tests of significance. We can write

$$S_e^2 = \sum_{i=1}^{n} (y_i - a_{i1}p_1 - \dots - a_{im}p_m)^2 + \sum_{i=1}^{k} (x_i - b_{i1}p_1 - \dots - b_{im}p_m)^2$$

where $p_1, \dots, p_m$ are solutions of (15.2). But if values of x_i minimizing S_e^2 be taken, then (15.5) shows that

$$S_e^2 = \Sigma\,(y_i - a_{i1}p_1 - a_{i2}p_2 - \dots - a_{im}p_m)^2$$

which is the actual sum of squares due to error when the variates are missing. Hence

The sum of squares due to error in the case of missing variates can be obtained by proceeding as if the variates were present, and taking those values of them which minimize this sum of squares.

If now we had to test any hypothesis, then the

(s.s. due to hypothesis) = (s.s. due to conditional error) - (s.s. due to error)

But our previous reasoning is valid for conditional error. Hence the actual s.s. due to conditional error can be obtained by proceeding as if the missing variates were present and minimizing the s.s. due to conditional error. Of course, the values of x_1, $x_2, \ldots, x_k$ which minimize the s.s. due to conditional error will not be the same as those which minimized the s.s. due to error. Hence, to get the s.s. due to hypothesis we have the following rules:

(actual s.s. due to hypothesis) = Min{(s.s. due to hypothesis) + (s.s. due to error)} - Min{(s.s. due to error)}

the s.s. on the right hand side being calculated as if the variates are present.

17. Missing Plot in a Randomized Block Experiment.

Suppose in a randomized block experiment one plot is missing. Let the missing plot occur in the j-th block, and suppose that the i-th treatment had been administered to it. If we write down the sum of squares due to error as if there had been no missing plot, then

$$S_e^2 = \Sigma\, y^2 - \frac{T_1^2 + \ldots + T_m^2}{n} - \frac{B_1^2 + \ldots + B_n^2}{m} + \frac{G^2}{mn} \qquad (17.1)$$

where there are $\underline{m}$ treatments and $\underline{n}$ blocks

$$\frac{dS_e^2}{dy_{ij}} = 2y_{ij} - \frac{2T_i}{n} - \frac{2B_i}{m} + \frac{2G}{mn}$$

Now the T_i, B_j, and $\underline{G}$ as given above also contain y_{ij}. Hence if we denote by T_i', B_j', G' the actual yield total, the actual block total and actual grand total, then we may write

$$y_{ij}\left(1 - \frac{1}{m} - \frac{1}{n} + \frac{1}{mn}\right) = \frac{T_i'}{n} + \frac{B_j'}{m} - \frac{G'}{mn}$$

$$\therefore\quad y_{ij} = \frac{mT_i' + nB_j' - G'}{(m-1)(n-1)} = \alpha \qquad (17.2)$$

To obtain the actual s.s. due to error, we have to put this value of y_{ij} in (17.1).

Again the conditional error is obtained by adding the s.s. due to error and the s.s. due to treatments. Thus the conditional error is

$$S_e'^2 = \Sigma\, y^2 - \frac{B_1^2 + B_2^2 + \ldots + B_n^2}{m} \tag{17.3}$$

$$\therefore \quad \frac{dS_e'^2}{dy_{ij}} = 2y_{ij} - \frac{2B_j}{m}$$

We then have to put

$$y_{ij}(1 - \frac{1}{m}) = \frac{B_j'}{m}$$

$$\text{or} \quad y_{ij} = \frac{B_j'}{m-1} = \beta \tag{17.4}$$

in (17.2), to get the actual conditional error, i.e., the actual sum of squares due to treatments and to error. If we have already carried out the analysis of variance, by using the value (17.2) for y_{ij} we would have over estimated the s.s. due to treatments, by an amount which is equal to the difference of the values of (17.3), when putting in it y_{ij} as given by (17.2) and y_{ij} as given by (17.4). This is called the bias in the s.s. due to treatments. Now

$$S_e'^2 = y_{ij}^2 - \frac{(B_j' + y_{ij})^2}{m} + \text{other terms not containing } y$$

$$= \frac{m-1}{m} y_{ij}^2 - \frac{2B_j'}{m} y_{ij}$$

Hence bias is given by

$$\frac{m-1}{m}(\alpha^2 - \beta^2) - \frac{2B_j'}{m}(\alpha - \beta) = \frac{(\alpha-\beta)}{m}\left\{(m-1)(\alpha+\beta) - 2\beta_j'\right\} = \frac{m-1}{m}(\alpha - \beta)^2$$

Thus we get the following rule: Estimate the missing value by the formula

$$y_{ij} = (mT_i' + nB_j' - G')/\left[(m-1)(n-1)\right]$$

Carry out the analysis of variance, with this value, but correct the s.s. due to the treatments by subtracting the amount

$$\frac{m}{m-1}\left\{y_{ij} - \frac{B_j'}{m-1}\right\}^2$$

but remember that only $(m-1)(n-1) - 1$ d.f. belong to error.

If i_1 and i_2 are two treatments for which the blocks are complete, then $t_{i_1} - t_{i_2}$ is estimated by

$$(T_{i_1} - T_{i_2})/n$$

and the variance of this comparison is clearly

$$2\sigma^2/n$$

It is of interest to compare this with the variance of a contrast between the i-th treatment for which a plot is missing and any other treatment say the i_o-th. In this case the estimate will be

$$(T_i - T_{i_o})/n$$

which may be written as

$$\frac{1}{n}\left\{\frac{mT_i' + nB_j' - G}{(m-1)(n-1)} + T_i' - T_{i_o}\right\}$$

Some algebraical calculation shows that the variance of this estimate is

$$\frac{\sigma^2}{n}\left\{2 + \frac{m}{(m-1)(n-1)}\right\}$$

Hence the contrast between these two treatments is measured by somewhat greater inaccuracy.