MEDEDELINGEN LANDBOUWHOGESCHOOL
WAGENINGEN • NEDERLAND • 83-10 (1983)

ANALYSIS OF VARIANCE

NICOLAAS H. KUIPER

Department of Mathematics,
Wageningen Agricultural University, The Netherlands

(Received 9-XI-1983)

LANDBOUWHOGESCHOOL – WAGENINGEN – 1983

CIP-GEGEVENS

Kuiper, Nicolaas H.

Analysis of variance/Nicolaas H. Kuiper; [transl. from the Dutch by M. Keuls].
– Wageningen: Landbouwhogeschool. – (Mededelingen/Landbouwhogeschool Wageningen; 83–10 (1983))
Vert. van: Variantie analyse. – Leiden: Stenfert Kroese, 1952. – (Statistica Neerlandica; 6 (3)). – Met reg.
ISBN 90-6754-051-X
SISO 517.2 UDC 519.2
Trefw.: mathematische statistiek.

ANALYSIS OF VARIANCE[1]

by

NICOLAAS H. KUIPER

INTRODUCTION

At the time the dutch paper was written (1952) prof. Kuiper was a professor at Wageningen teaching Mathematics and Statistics. The mathematical basis of analysis of variance was vaguely perceived at that time. In his 1952 paper a system of notions and corresponding notations has been developed, which with some right could be called the Wageningen method. To be very short, it centres on the use of vectors, vector spaces and orthogonal (and skew) projections without the explicit use of matrices. The notation used has the advantage that it is easily generalized to multivariate situations (not dealt with in the paper) and to more complex experimental designs. Some additional references illustrate this point and connected topics.

ACKNOWLEDGEMENT

Thanks are due to prof. dr. L. C. A. Corsten for his encouraging support and effective help. For aid and assistence in the preparation of the paper mrs. Johanna D. Houwers, mrs. Renée M. Hoeffnagel and drs. Th. H. B. Hendriks are acknowledged. Mrs. ir. E. J. Kuiper suggested the translation.

The translator

[1] This paper aims at being a strict translation of the paper written in dutch language: N. H. KUIPER (1952) 'Variantie analyse'. Statistica Neerlandica 6 (3) 149–194.
Translated by M. Keuls, (1983); publication with the permission of the author and the Editorial Board of Statistica Neerlandica.

CONTENTS

1. INTRODUCTION[1]

In the analysis of observational results according to R. A. Fisher, well known as 'analysis of variance', attention is fixed on a positive number, the variance, to be computed for a set of observational results, and the possibility is examined to write this number as a sum of other positive numbers (also variances), to each of which a special meaning is attached. Each of these numbers is considered as the consequence of a distinct cause.

In the theory, that is developed in this paper, variances will also be mentioned, but they don't play an essential part in the theory. The centre of the theory will be in the *vectors*.

The purpose of using vectors in a theory aiming at the analysis of certain systems of observational results, and in connection with the design of experiments, is fourfold:

A. For a number of concepts, such as main effect, interaction, confounding, orthogonality (of an experimental design), clear definitions are often missing in the literature. A coherent set up of concepts is possible, and as it seems to us, is only possible, in terms of vectors.

B. Some parts of the analysis of variance can be developed only in an unclear and unconvincing way without the use of vectors, for instance the treatment of non-orthogonal designs (see the literature mentioned in chapter 11). By using vectors a transparent and convincing presentation is possible.

C. The theory allows the possibility of considering at a glance many more schemes of experimentation of great use in research practice. For instance it is not generally known, that there exist many more orthogonal experimental designs than those ordinarily met with in the literature (compare chapter 8 and 14).

D. Our notation is more simple than the notation of analysis of variance, generally used. The price one has to pay for these advantages lies in the need to study the theory of vectors and vector spaces. In chapter 2 we give some elements of this theory for those who are not acquainted with this subject.

2. VECTORS

We start by choosing a constant natural number N. Next we consider all rows of N real numbers $(x_1, \ldots, x_N)$. Such a row is called a *vector*. Vectors can be taken together in sets of two, three or an infinite number. The biggest of such sets is the set of *all* rows of N numbers (vectors). This set is called the N-dimensional vector space (E or also E_N).

[1] For critical remarks I am grateful to messrs. L. C. A. Corsten, G. Hamming, J. Hemelrijk, M. Keuls and C. A. G. Nass.

Two vectors a $= (a_1, \ldots, a_N)$ and b $= (b_1, \ldots, b_N)$ may be added. By the sum, denoted a + b, we mean the *vector*

$$a + b = (a_1 + b_1, \ldots, a_N + n_N).$$

It is easily understood, that for sums of vectors many properties hold, that also hold for ordinary sums (i.e. of numbers). These properties simplify computations. For instance

a + b is the same as	b + a
(a + b) + c is the same as	a + (b + c)
(a + b) + (c + (d + e)) is the same as	((a + c) + (b + d) + e.

It is generally true, that in sums of two or more terms the order and brackets may be chosen arbitrarily. In every case one gets the same answer.

A vector (a) can also be *multiplied by a number* (λ). A vector a multiplied by λ is by definition:

$$\lambda a = (\lambda a_1, \lambda a_2, \ldots, \lambda a_N).$$

Also here some easily retained rules hold, facilitating computations:

λ (a + b) is the same as λa + λb (check this)
($\lambda + \mu$) a is the same as λa + μa

The vector 0 which represents a row of N numbers zero: $(0, 0, \ldots, 0)$ is called *null vector*. It has properties in common with the number 0. Thus for example (a an arbitrary vector)

$$a + 0 = a \text{ and } a + (-1)a = (a_1, \ldots, a_N) + (-a_1, \ldots, a_N) = 0$$

The vector (-1) a or $-$a is called the opposite of a. As with ordinary algebra one can write a + ($-$b) more briefly as a $-$ b. All rules for the addition and subtraction also hold for vectors. Thus for example the solution of the equation in (the unknown vector) x:

$$2a + 3x = b \text{ equals } x = \tfrac{1}{3}b - \tfrac{2}{3}a.$$

The vectors in a three-dimensional vector space ($N = 3$) can be pictured in the (ordinary) space, in which we live. One chooses an orthogonal cross of coordinate axes and represents the vector (a_1, a_2, a_3) by the arrow that starts in the origin of the cross of axes and has its arrow point in the point with coordinates a_1, a_2 and a_3. The addition of vectors then is pictured by the well known *parallelogram construction*. The multiplication by a number is represented by a geometric multiplication with the origin as a centre of multiplication and the number as a multiplication factor. Historically, the more specific geometric concept of vectors depicted here arose first. By abstraction the more general concept, that we need, came into being.

A vector a is called (linearly) dependent on a vector b, if there exists a number λ, such that a $= \lambda$b. The vector (6, 9, $-$15, 6) for instance is dependent on (4,6, $-$10,4), for:

$$(6, 9, -15, 6) = \tfrac{3}{2}(4, 6, -10, 4);$$

a is called dependent on the vectors b, ..., c, if there exist numbers λ, ..., μ such that a = λb + ... μc. Thus (5, −6, 8, 1) is dependent on

$$(3, -2, 0, 7) \text{ and } (1, 0, -2, 5), \text{ for:}$$
$$(5, -6, 8, 1) = 3\,(3, -2, 0, 7) - 4\,(1, 0, -2, 5).$$

To some systems of vectors in the space E one gives the name linear vector space, or *subspace* or in brief *space*. All vectors λa, that are obtained from the constant vector a by multiplication with a real number, constitute a system that is called a 1-dimensional (sub-)space.

Given some (constant) vectors a, ..., b, the system of vectors λa + ... + μb, with λ, ..., μ variable real numbers, is a vector subspace of E. We will denote this space by a capital letter, for instance D. D is said to be spanned by the vectors a, ..., b. Notice that 0 belongs to every subspace. The set of vectors a, ..., b is called a *basis* of the space D. One sometimes can generate the same space using a different basis. For D one can find many bases. Every basis of D that consists of independent vectors (no one being dependent on the others), contains the same number of vectors. (This is the theorem of STEINITZ). This number is called the *dimension* of the space D.

In the books that deal with vectors, as a rule one pays little attention to those aspects which we will use; instead of this one treats determinants, linear equations and matrices. Some books in which one can read about vectors are: A. HEIJTING, Matrices en determinanten; A. C. AITKEN, Determinants and matrices; H. SCHWERDFEGER, Introduction to linear algebra and matrices; MCDUFFEE, Vectors and matrices; R. ZURMÜHL, Matrizen.

3. FUNCTIONS WITH AS A RANGE OF DEFINITION A FINITE NUMBER OF POINTS

The plots of a field trial $P_1, \ldots, P_N$ in an agricultural experiment can be characterized: by their subscript; or by, say, three numbers x_1 = dosage of nitrogen, x_2 = dosage of phosphate, x_3 = a number that indicates which plot with the dosages x_1, x_2 is meant; or in an analogous way by more numbers; or in many other ways. The result of an experiment may consist of a yield y_k ($k = 1, \ldots, N$) for each of the plots. These yields together may be conceived as a real function with the plots $P_1, \ldots, P_N$ as range of definition; for one understands by a function with this range of definition the assignment of a real number to each of the plots.

As our considerations also apply to many problems where the word field experiment will not be found, we will often replace the word experimental plot by 'point' (abstraction) and our interest will be focussed on real functions defined on N points. Such a function can be represented in a simple way, viz. by writing the values of the function in the order (numbering from 1 to N) of the points, i.e. $(y_1, y_2, \ldots, y_N)$.

It is clear, that the sum of two functions: $a = (a_1, a_2, \ldots, a_N)$ and $b = (b_1, b_2, \ldots, b_N)$ is obtained, by adding the function values i.e. the numbers at corresponding places: $a + b = (a_1 + b_1, a_2 + b_2, \ldots, a_N + b_N)$. Also if ω is a real number, $\omega a = \omega(a_1, a_2, \ldots, a_N) = (\omega a_1, \omega a_2, \ldots, \omega a_N)$.

By these properties of addition, and multiplication with a real number (ω), functions can be conceived as *vectors* and the set of all real functions on N points as a *vector space* E.

EXAMPLES OF FUNCTIONS AND FUNCTION-SPACES CONTAINED IN E

a) The function r assigns to each of N points the number 1: $(1, 1, \ldots, 1)$. The *constant functions* are of shape $(c, c, \ldots, c) = c.(1, 1, \ldots, 1)$. They form a 1-*dimensional subspace* of E with a *basis*: $r = (1, 1, \ldots, 1)$.

Definition 1.

This space is called the space of the general means.[1]

Suppose that one considers plots where a crop grows; all plots are treated in the same way and are subject to the same influences ('all influences are distributed uniformly over the plots'). If moreover accidental deviations do not occur, *then* the yield would be a constant function $(C, C, \ldots, C)$, hence a vector belonging to the space of the general means (See chapter 4 and 9).

b) *The elementary characteristic functions.* A function that assigns to one of the points the number 1, and to all other points the number 0, is called an elementary characteristic function. This function characterizes (= determines completely) the point concerned. There are N characteristic functions, viz. $(1, 0, \ldots, 0)$, $(0, 1, 0, \ldots, 0), \ldots, (0, 0, \ldots, 1)$.

Theorem: Every function can be written as a sum of multiples of the elementary characteristic functions, or in other words: *The elementary characteristic functions form a basis of the vector space of the functions of N points.* The vectors of this basis are independent, therefore the dimension of the space equals N.

Proof of the statement in italics: The arbitrarily chosen function $(y_1, y_2, \ldots, y_N)$ equals $y_1(1, 0, 0, \ldots, 0) + y_2(0, 1, 0, \ldots, 0) + \ldots + y_N(0, 0, 0, \ldots, 1)$.

Definition 2:

The number of degrees of freedom of a vector space is the dimension of that vector space. The number of degrees of freedom of the space of the general means is 1, that of the space E is N.

c) *The characteristic function of a class of points.* If $P_j, P_k, \ldots, P_n$ are a number of points (= a class of points), then the characteristic function of this class is the function, that is 1 on the points of this class and is 0 on the other points. The function r, and the elementary characteristic functions are examples.

d) *The subspace of functions, determined by a classification of* $P_1, \ldots, P_N$. Suppose that an assignment (classification) (named A) of the points $P_1, \ldots, P_N$ to m classes

[1] Translator's note: CORSTEN (1958, p. 26) introduced the symbol N for the space of the general means.

named $a_1, \ldots, a_m$) is given (see the scheme, where cells instead of points are drawn; this has the advantage that a function can be represented by writing in every cell the value of the function; the cells can for instance refer to agricultural plots with equal areas).

the function a_1

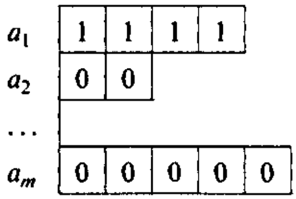


Fig. 1.

the function $z_1 \cdot \mathrm{a}_1 + \ldots + z_m \cdot \mathrm{a}_m$

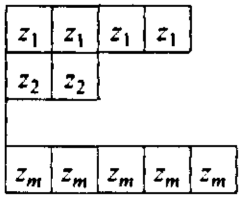


Fig. 2

The function represented in fig. 1 is the characteristic function of the class a_1. We call this function (= vector) a_1, and the naming of the other characteristic functions will be analogous. Next we consider all functions that are constant within each class of A. If the value of such a function at the points of the class a_k is equal to z_k ($k = 1, \ldots, m$), then it is clear that this function is a linear combination of the class characteristic functions of A, viz. $z_1 \cdot \mathrm{a}_1 + \ldots + z_m \mathrm{a}_m$. Because the m class characteristic functions are linearly independent, one has:

Theorem: The set of the functions that are constant within the m classes of a classification A form an m-dimensional vector space, with the class characteristic functions as a basis. This vector space is denoted by A.

A grouping of N plots into classes can be of practical relevance, when the plots within a class are subject to the same treatment. If random deviations are negligible then one will expect, that the yields of the plots within each class are equal, i.e. they are a function (= vector) in the space A corresponding to the classification A. If we indicate the treatment or influence that differs in the various classes, also by the letter A, then such a function (= vector) is called the 'crude' main effect of the influence A.

Definition 3:

The space A *is called space of crude main effects of the influence A.* The number of degrees of freedom of the crude main effect (here) is m.

To each of the classes of classification A one can assign a (different) value of a variable x_1. All ordinary functions $f(x_1)$ (with those values of x_1 as range of definition) can be taken as functions constant within the classes; thus in fact one gets the complete set i.e. the space A. The constant functions also belong to A, thus we can remark, that space A *contains the space of the general means, in particular* **r**.

Conversely, it may occur in practice that the points of a given set have a property that can be expressed quantitatively by a number, a value of the variable x_1. Those points for which x_1 has the same value, can be combined into a class;

thus a classification can be generated by the property (x_l). The 'ordinary' functions $f(x_l)$ determine the vector space of crude main effects corresponding to that classification.

In practice one often meets the situation that for N points two classifications (A and B) are given instead of one. In that case, there exists a third classification determined by A and B, viz. that is generated by grouping into one class those points that are at the same time contained in one class of A and in one of B. In the figures 3, 4, 5 this classification $A \times B$ is indicated by drawn lines.

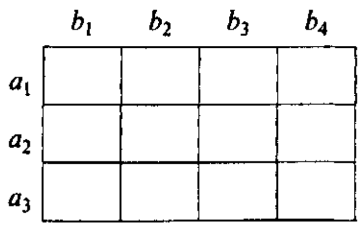


Fig. 3.

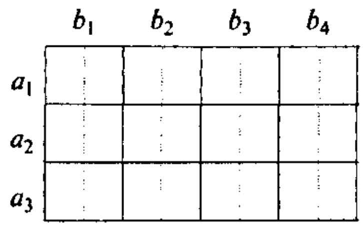


Fig. 4.

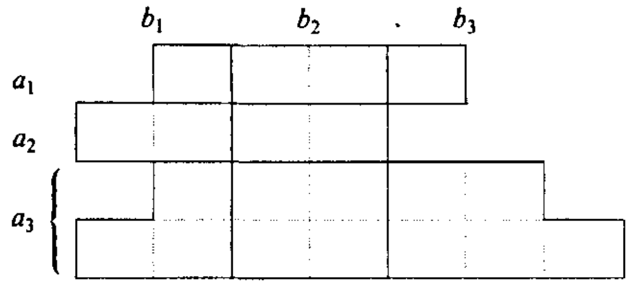


Fig. 5.

Theorem: The vector spaces A and B are contained in the vector space C = A × B determined by the two classifications; for each function that is constant within the classes of the classification A, is implicitly constant within the classes of the classification C.

Definition 4: Let A and B be classifications, based on two properties such that the vector spaces A and B are the crude main effects of the properties (influences) A and B, then the vector space A × B is called *the space of the crude interactions of the two properties A and B.* (Writing C for A × B, one could also consider space C as the space of the crude main effects of a property ('influence') C).

If with two classifications of N plots only combinations of the two corresponding influences A and B would be effective and if all other possible influences were the same on all plots, and there were no random deviations, then one could expect, that the yields within the classes $A \times B$ would be constant, i.e. they would form a function (= vector) in the space of crude interactions.

Just as before, one can also in this case assign to each of the classes of the classification $A(B)$ a (different) value of a variable x_1 (x_2). The functions of the space A can thus be written in the form $f(x_1)$ and A consists of all such functions.

The analogue holds for the functions of the shape $f(x_2)$ and vector space B. Also the functions of the space A × B have now a simple expression: They are all functions of shape $f(x_1, x_2)$. Thus again one sees, that the space A × B contains the spaces A and B.

The converse is also of interest: The points of a set have two properties, that are each quantitatively expressed by a value of the variables x_1 and x_2, respectively. Those points for which x_1 has the same value (x_2 has the same value; x_1 and x_2 have the same value) can be grouped into one class, and thus the classification A (resp. B and $A \times B$) is generated. The functions of shape $f(x_1)$, $f(x_2)$, $f(x_1, x_2)$ form the vector spaces corresponding to those classifications: the spaces of crude main effects and the space of crude interactions.

In addition to one or two classifications of N points, it may happen that more classifications are made, e.g. 5. If these classifications arise from variables x_1, $x_2, \ldots, x_5$, then some examples of interesting spaces of functions are:

1. The space of crude main effects of the second influence: the functions $f(x_2)$.
2. The space of crude interactions of the influences 2 and 3: the functions $f(x_2, x_3)$.
3. The space of crude interactions of the influences 2, 3 and 5: the functions $f(x_2, x_3, x_5)$.
4. The space of crude interactions: the functions $f(x_1, x_2, x_3, x_4, x_5)$. Each of these spaces contains the preceeding ones.

e. For a given problem one may need further different subspaces. Thus it may happen, that one has some explicity given functions: g, h and the function r, and that there is some reason to consider the space of functions that is spanned by these given functions, i.e. the set of functions $\lambda + \mu g + \nu h$; λ, μ, ν real variables.

Often a subspace is determined with respect to a property of the points, which has *to be measured*. For instance given a variable x_1, one can consider the linear functions $(\lambda + \mu x_1)$ or the quadratic functions $(\lambda + \mu x_1 + \nu x_1^2)$ or etc. in x_1 or in more variables.

Exercise (3.1)

Determine the dimensions of the spaces defined so far for the figures 3, 4 and 5.

Exercise (3.2)

When measuring property A for a number of points, it turns out that variable x_1 takes only three different values. Prove, that the space of crude main effects of the influence A, coincides with the space of the quadratic functions in x_1. What can be said about the space of fourth degree functions in x_1? Which dimension has the space of n^{th} degree functions in x_1 in this case? ($n = 0, 1, 2, \ldots$).

Exercise (3.3)

At a number of points the variable x_1 takes four different and the variable x_2 two different values. Both variables describe 'influences'. All combinations of the values of x_1 and x_2 are present. Which are the dimensions of the spaces of the crude main effects and of the space of the crude interactions? Which are the dimensions of the spaces of the functions of degree 1, 2, 3, 4, 5 in x_1; idem in x_2; idem in x_1 and x_2?

In this final paragraph we mention *the decomposition of a vector y into component vectors.* Suppose A, B, ..., H to be subspaces of E, that together span E, with d_A being the dimension of A etc., d the dimension of E, $d_A + \ldots + d_H = d$, the subspaces being linearly independent. One can choose then a basis for each subspace, and these bases together form a basis of the space E. An arbitrary vector y can be written as a sum of multiples of the basis vectors and this in only one way. These multiples can be taken together in an appropriate way, by which it appears that the given vector y can be written as a sum of component vectors each belonging to one of the subspaces. These component vectors, also called '(skew) projections', are indicated by y_{sA} etc., such that holds:

$$y = y_{sA} + y_{sB} \cdots + y_{sH}.$$

In the case that the projection is orthogonal, we drop the letter s. (Thus $y_{sA} = y_A$, if $y - y_{sA} \perp A$).

4. THE GAUSSION MULTIDIMENSIONAL PROBABILITY DISTRIBUTION

For a moment we return to agricultural terminology: There are given N plots. After an experiment one gets for each plot an observational result (yield) and one assumes it to be a sample from a normal distribution, thus (P = probability):

$$P(a_1 < \underline{y}_1 \leq b_1) = \int_{a_1}^{b_1} \left(\sigma\sqrt{2\pi}\right)^{-1} \exp \frac{-(y_1 - \check{y}_1)^2}{2\sigma^2} dy_1 \tag{1}$$

The yields on the N plots together are a function (vector), and similarly the expectations, denoted by y and $\check{y}$ respectively. One assumes, that the random variables $\underline{y}_1, \ldots, \underline{y}_N$ are independent, and it follows that probabilities as in (1) can be multiplied in order to obtain:

$$P(\underline{y} \in G, \text{i.e. here: } a_i < \underline{y}_i \leq b_i, i = 1, \ldots, N) =$$
$$= \int_G \left(\sigma\sqrt{2\pi}\right)^{-N} \exp \frac{-(y - \check{y})^2}{2\sigma^2} \times dy_1 dy_2 \ldots, dy_N \tag{2}$$

This determines the probability distribution of *the random vector* $\underline{y}$. The integrand is called the probability density of y.

Note that $(y - \check{y})^2$ is the inner product of the vector $y - \check{y}$ with itself, where the inner product of two vectors $a = (a_1, a_2, \ldots, a_N)$ and $b = (b_1, b_2, \ldots, b_N)$ is defined by $a_1 \cdot b_1 + a_2 \cdot b_2 + \ldots + a_N \cdot b_N = (a, b) = ab$.

(The inner product has the following properties:
$(a, b+c) = (a, b) + (a, c)$; $(a, b) = (b, a)$; $(\omega a, b) = (a, \omega b) = \omega\,(a, b) \cdot$)

The inner product opens the possibility to introduce some geometric terms, which for $N = 3$ with the mentioned presentation in *our* space appear to have the customary meaning. The square of the *length* of a vector y is defined by $(y, y) = y^2$; a *unit vector* is a vector of length 1; the cosine of the *angle* φ between two vectors a and b is $\cos \phi = ab/\sqrt{aa \cdot bb}$. Two vectors a, b are perpendicular to each other if $ab = 0$.

Suppose that $e_1, e_2, \ldots, e_N$ are mutually orthogonal unit vectors in E, and $z_1, z_2, \ldots, z_N$ are numbers such that

$$y = z_1 e_1 + z_2 e_2 + \ldots + z_N e_N,$$

then $(z_1, z_2, \ldots, z_N)$ are called coordinates of the vector y with respect to the coordinate system formed by $e_1, e_2, \ldots, e_N$. The inner product of two vectors, given by these new coordinates, has formally the same expression $(y, y') = y_1 y'_1 + \ldots + y_N y'_N = z_1 z'_1 + z_2 z'_2 + \ldots + z_N z'_N$, as in the old coordinates, as appears by substitution.

Integral calculus tells, that the integral (2) over an arbitrary region of the vector space is an expression in the variables z_k which is the same as that in the variables y_k. That the integrand remains formally unchanged, follows from the fact that the expression for the inner product remains the same. That the 'volume-element' $dy_1\, dy_2 \ldots dy_N$ remains formally unchanged may be understood as a property of determinants.

As (2) arose from a product of integrals like (1), so one can, at least if the region G lends itself to it, also write the new integral in $z_1, z_2, \ldots, z_N$ as a product of integrals of the form (1), but now in z_k. Anyhow the decomposition into factors holds for the integrand. From this it follows that the random variables $z_1, z_2, \ldots, z_N$ are stochastically independent. (The metric has been defined such, that stochastic independence of two linear vector functions implies, and follows from, orthogonality of the corresponding planes of constant function values.) Particularly the probability distribution of the random vector with new coordinates, say, $(z_1, z_2, z_3, z_4, 0, 0, \ldots, 0)$ has a probability density

$$(\sigma\sqrt{2\pi})^{-4} \exp. - \frac{(z_1 - \check{z}_1)^2 + (z_2 - \check{z}_2)^2 + (z_3 - \check{z}_3)^2 + (z_4 - \check{z}_4)^2}{2\sigma^2}$$

while the probability distribution of the exponent except for a factor -2 is that of chi-square with 4 degrees of freedom (= dimension).

Now let y be a random vector as dealt with in (2). Let A be a m-dimensional subspace of the vector space, and let y_A be the *orthogonal projection* of y on A. In the space A one may choose m mutually orthogonal unit vectors, which one could supplement to N such vectors in the vector space E. These unit vectors would determine a system of orthogonal coordinates, and in particular it would follow, that $(y_A - \check{y}_A)^2/\sigma^2$ is a random variable with a probability distribution as that of chi-square (χ^2) with m degrees of freedom. This random variable however is also of interest without the necessity of referring to unit vectors and coordinates. Finally as it appears, we can omit these and we obtain the

Theorem: Let A be a m-dimensional subspace of E, $\underline{y}$ a random vector with a central-symmetric normal probability distribution; σ the standard deviation of the random variable $\underline{y}_i$; $E(\underline{y}) = \check{y}$; $\check{y}_A$ and y_A, the projections of $\check{y}$ and y on A, respectively, then the random variable

$$\frac{(y_A - \check{y}_A)^2}{\sigma^2}$$ *has the probability distribution of chi-square ($\underline{\chi}^2$) with dimension m*

We shall now apply the metric in the vector space as introduced by the inner product, and return later to statistical considerations (chapter 7).

5. THE SPACE OF THE PURE MAIN EFFECTS OF AN INFLUENCE

Once we have introduced a metric, we know when two subspaces are completely orthogonal: each vector of one space is orthogonal to each vector of the other space.

Definition 5:

Let A be a classification, based on a property (influence) of the N points, A the vectorspace of crude main effects, then *the vectorspace of pure main effects* A* is the subspace of A that consists of all vectors of A, that are orthogonal to the space of the general means. Each vector in the space A* is called *a (pure) main effect.*

In chapter 3 we introduced the term 'crude main effect' by considering plot yields that would be obtained if only one influence would not affect all plots equally in form or degree. We do not wish, however, to accept as a component of a *pure main effect* of the influence A any increase in yield that has the same value on all plots and is exclusively due to a change of another influence, common to all plots. Thus we wish the definition of a *pure main effect* to be such, that the following two crude main effects determine the same pure main effect:

$$
\mathrm{z} = \begin{array}{|cc|c}
\hline
z_1 & z_1 & a_1 \\
\hline
z_2 & z_2 & a_2 \\
\hline
z_3 & z_3 & a_3 \\
\hline
\end{array}
\qquad
\begin{array}{|cc|}
\hline
z_1+4 & z_1+4 \\
\hline
z_2+4 & z_2+4 \\
\hline
z_3+4 & z_3+4 \\
\hline
\end{array}
$$

The space of crude main effects contains as two completely orthogonal subspaces: the space of the general means and the space of the pure main effects with as bases:

$$
\mathrm{r} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} +1 & +1 \\ -1 & -1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} +1 & +1 \\ 0 & 0 \\ -1 & -1 \end{pmatrix} \text{, respectively;}
$$

$\begin{pmatrix} p & p \\ q & q \\ r & r \end{pmatrix}$ is a pure main effect, if $2p + 2q + 2r = 0$.

Every crude main effect can be decomposed into component vectors in both these spaces, and the components are the *orthogonal* projections. As the orthogonal projection $\bar{\mathrm{z}}$ of the vector z onto the space of the general means is: $\omega \cdot \mathrm{r}$, and $\mathrm{z} - \omega \cdot \mathrm{r}$ is perpendicular to r, it follows that:

$$(\mathrm{z} - \omega \mathrm{r}, \mathrm{r}) = \mathrm{zr} - \omega \cdot 6 = 0$$

$$\omega = \frac{z_1 + z_1 + z_2 + z_2 + z_3 + z_3}{6} = \bar{z}.$$

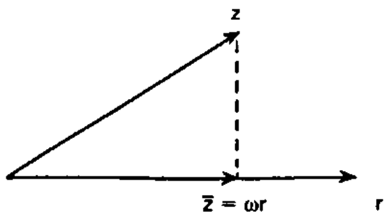


Fig. 6.

$\bar{\mathrm{z}}$ is the one component of z, the other, *the pure main effect* of z, *therefore* equals $\mathrm{z} - \bar{\mathrm{z}}$.

For the two schemes we find for the first and second component respectively:

$$
\begin{pmatrix} \bar{z} & \bar{z} \\ \bar{z} & \bar{z} \\ \bar{z} & \bar{z} \end{pmatrix}, \begin{pmatrix} \bar{z}+4 & \bar{z}+4 \\ \bar{z}+4 & \bar{z}+4 \\ \bar{z}+4 & \bar{z}+4 \end{pmatrix} \text{ and } \begin{pmatrix} z_1-\bar{z} & z_1-\bar{z} \\ z_2-\bar{z} & z_2-\bar{z} \\ z_3-\bar{z} & z_3-\bar{z} \end{pmatrix} \begin{pmatrix} z_1-\bar{z} & z_1-\bar{z} \\ z_2-\bar{z} & z_2-\bar{z} \\ z_3-\bar{z} & z_3-\bar{z} \end{pmatrix}
$$

The pure main effects of the given schemes are thus the same. A pure main effect is a vector from which one can derive only the 'differences' in effect, as a consequence of differences in the influence.

Another numerical example:

$$\begin{pmatrix} 23 & 23 & 23 \\ 13 & 13 & 13 \\ 22 & 22 & \end{pmatrix} = \begin{pmatrix} 19 & 19 & 19 \\ 19 & 19 & 19 \\ 19 & 19 & \end{pmatrix} + \begin{pmatrix} 4 & 4 & 4 \\ -6 & -6 & -6 \\ 3 & 3 & \end{pmatrix}$$

The pure main effect can be decomposed into the sum of two 'differences':

$$4\begin{pmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ 0 & 0 & \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 \\ -2 & -2 & -2 \\ 3 & 3 & \end{pmatrix}$$

Exercise:

Which are the dimensions of the spaces of the pure main effects for the schemes of chapter 3?

Note: The number of degrees of freedom of the space of the pure main effects of an influence A, generating a classification of N plots in m classes, equals $m - 1$.

6. THE SPACES OF THE PURE INTERACTIONS

Definition 6:

Let A and B be two classifications, based on two properties, then *the vector space of the pure interactions* is the subspace of the space of the crude interactions, consisting of all vectors orthogonal to the spaces of the crude main effects. Each vector in this space is called *a pure interaction of A and B*[1].

Examples

a) The space of the crude interactions (fig. 7) coincides with the space of all functions on the 4 points (= plots). We already know the spaces: (1) of the general means; (2) of the crude main effects A; (3) of the pure main effects A; (4) of the crude main effects B; (5) of the pure main effects B. The space of the pure interactions also is 1-dimensional with basis: (6).

Fig. 7.

[1]Translater's note: CORSTEN (1958, p. 48) introduced the notation (A × B)* for the space of the 'pure interactions', containing the vectors of type $y_{A \times B} - y_{A+B}$. One could be inclined to read $y_{(A \times B)*}$ as $y_{A \times B} - y_N$; however in cases where the productstructure in classification $A \times B$ is of no relevance, one may give a new name S to A × B and use $y_{S*} = y_S - y_N$ ($= y_{A \times B} - y_N$) without confusion.

$$\underset{(1)}{\begin{pmatrix} + & + \\ + & + \end{pmatrix}} \quad \overbrace{\begin{pmatrix} + & + \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ + & + \end{pmatrix}}^{(2)} \quad \underset{(3)}{\begin{pmatrix} + & + \\ - & - \end{pmatrix}} \quad \overbrace{\begin{pmatrix} + & 0 \\ + & 0 \end{pmatrix} \begin{pmatrix} 0 & + \\ 0 & + \end{pmatrix}}^{(4)}$$

$$\underset{(5)}{\begin{pmatrix} + & - \\ + & - \end{pmatrix}} \underset{(6)}{\begin{pmatrix} + & - \\ - & + \end{pmatrix}}$$

read: $+ = +1$, $- = -1$

The addition of the dimensions:

General means	1
A pure	1
B pure	1
$A \times B$ pure	1
total (= number of plots)	4

Example b)

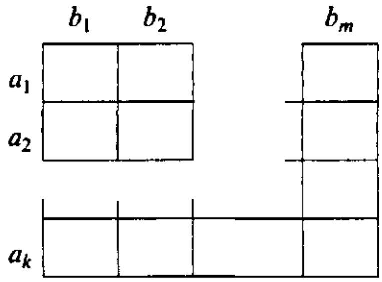


Fig. 8.

The spaces A* and B* are of dimension k − 1 and $m - 1$, respectively. The space of crude interactions is of dimension km. The space of pure interactions is of dimension $km - 1 - (k-1) - (m-1) = (k-1)(m-1)$; an independent basis is as follows:

$$\begin{pmatrix} + & - & 0 & \dots & 0 \\ - & + & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & & & & \dots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \begin{pmatrix} + & 0 & - & \dots & 0 \\ - & 0 & + & \dots & 0 \\ 0 & 0 & 0 & & 0 \\ \dots & & & & \dots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \dots, \begin{pmatrix} + & 0 & 0 & \dots & - \\ - & 0 & 0 & \dots & + \\ 0 & 0 & 0 & \dots & 0 \\ \dots & & & & \dots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix};$$

$$\begin{pmatrix} + & - & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ - & + & 0 & \dots & 0 \\ \dots & & & & \dots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \begin{pmatrix} + & 0 & - & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ - & 0 & + & \dots & 0 \\ \dots & & & & \dots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

, etc., number: $(k-1)(m-1)$

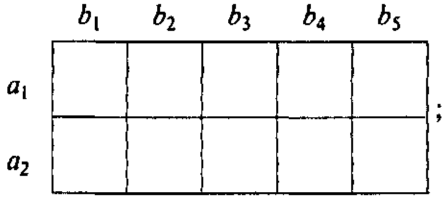
read: $+ = 1$ and $- = -1$

The addition of the dimensions:

General means	1
Pure main effect A	$k - 1$
Pure main effect B	$m - 1$
Pure interaction $A \times B$	$(k-1)(m-1)$
total	km

Exercise (6.1)

Consider in detail the classifications



;

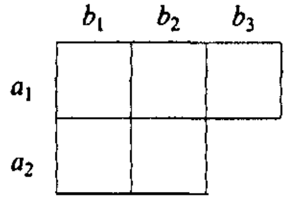


Fig. 9. Fig. 10.

Example c)

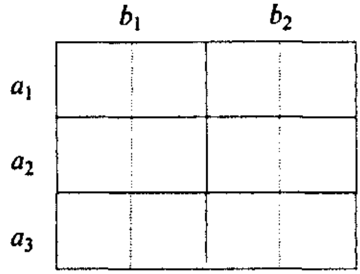


Fig. 11.

The important spaces: general means (1), A* (2), B* (3), pure interactions (4), having e.g. the independent bases

(1)	(2)		(3)	(4)	
+ + + +	+ + + +	+ + + +	+ + − −	+ + − −	+ + − −
+ + + +	− − − −	0 0 0 0	+ + − −	− − + +	0 0 0 0
+ + + +	0 0 0 0	− − − −	+ + − −	0 0 0 0	− − + +

Together they span the space of crude interactions (dimension 6). The space orthogonal to it is called here (see chapter 9) *the space* R *of pure error R*. Here R is of dimension $12 - 6 = 6$. A basis may be e.g.

+ − 0 0	0 0 + −	0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0
0 0 0 0	0 0 0 0	+ − 0 0	0 0 + −	0 0 0 0	0 0 0 0
0 0 0 0	0 0 0 0	0 0 0 0	0 0 0 0	+ − 0 0	0 0 + −

The addition of the degrees of freedom

General means	1
A pure	2
B pure	1
Interaction $A \times B$ pure	2
R (residual) pure error	6
total (= number of plots)	12

Exercise (6.2)

Determine the dimensions of the hitherto defined subspaces in each of the following eight cases, with bases and the addition of the dimensions for each of them.

(1)

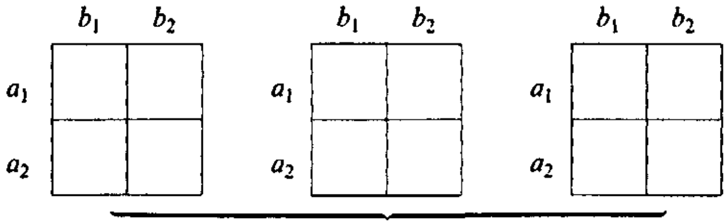


Fig. 12.

(2)

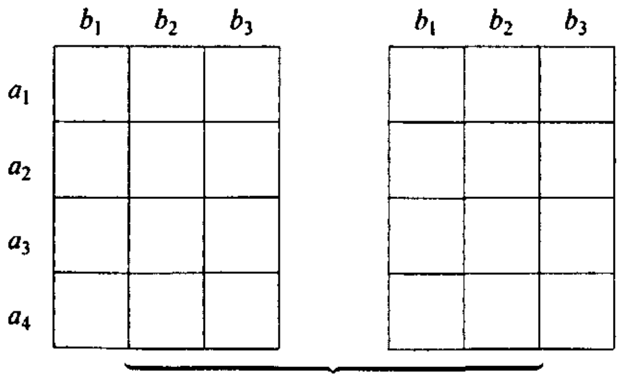


Fig. 13.

(3)

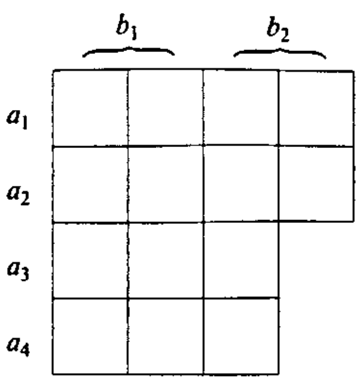


Fig. 14.

(4)

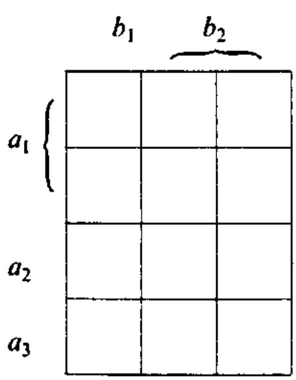


Fig. 15.

(5) The space of functions (with two influences) $f(x_1, x_2)$ defined at $(x_1; x_2) = (1;5), (3;5), (3;3), (3;6), (4;6), (4;3), (1;3)$.

(6) The space of functions (with three influences; define (and compute the dimension of) the space of the interactions of the three factors) $f(x_1, x_2, x_3)$ where the variables can take each the values 1, 2, 4 in all combinations.

(7) Define the spaces of the pure linear (quadratic, 3rd degree) main effects and interactions in the foregoing example.

(8) The space of functions $f(x_1, x_2, x_3)$ where the variables each take the values -1 and $+1$. Show that a basis of this space is given by the 8 functions $1, x_1, x_2, x_3, x_1 \cdot x_2, x_1 \cdot x_3, x_2 \cdot x_3, x_1 \cdot x_2 \cdot x_3$. These vectors also are bases of the spaces of the pure main effects and interactions!

Express *these* (5) ... (8) functions in terms of characteristic functions.

7. CONFOUNDING[1]

Definition 7.

Two influences, main effects or interactions are called *confounded*, if the corresponding vector spaces have a subspace of positive dimension in common. They are *completely confounded* if these vectorspaces coincide. *One vector space is called completely confounded with a second one*, if the former is contained in the latter vector space. (Complete and partial confounding, respectively.)

Examples

a)

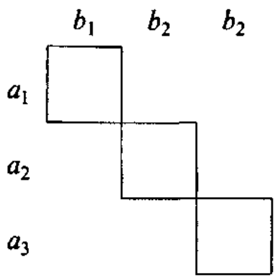


Fig. 16.

The spaces A and B coincide as well as the spaces A* and B*. The influences A and B are completely confounded.

b)

	c_1	c_2
a_1	b_1	b_2
a_2	b_2	b_1

Fig. 17.

$$\begin{pmatrix} + & - \\ - & + \end{pmatrix}$$

Fig. 17a.

[1] The dutch term 'strengelen' was suggested by dr. G. J. Vervelde.

There are given three classifications (A, B and C), where the classification C e.g. corresponds to an influence 'differences in fertility' ('blocks'). The space of the pure main effects B coincides with the space of the pure interactions of A and C (basis: fig. 17a). The (pure) interaction $A \times C$ and the (pure) influence B are (completely) confounded.

	b_1	b_2	b_3
a_1	1	2	3
a_2	3	1	2
a_3	2	3	1

Fig. 18.

c) Latin square: Three classifications: A, B and C (fig. 18). C is given by the figures in the plots. Application of the Latin square implies the assumption, that the interactions of the influences are negligible, because the following confounding applies. The pure main effect of C is completely contained in the space of pure interactions of A and B.

These 2- and 4-dimensional spaces have as bases:

$$\begin{pmatrix} + & - & 0 \\ 0 & + & - \\ - & 0 & + \end{pmatrix} = \begin{pmatrix} + & - & 0 \\ - & + & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ + & 0 & - \\ - & 0 & + \end{pmatrix} , \begin{pmatrix} + & 0 & - \\ - & + & 0 \\ 0 & - & + \end{pmatrix}$$

and

$$\begin{pmatrix} + & - & 0 \\ - & + & 0 \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} + & 0 & - \\ - & 0 & + \\ 0 & 0 & 0 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ + & - & 0 \\ - & + & 0 \end{pmatrix} , \begin{pmatrix} 0 & 0 & 0 \\ + & 0 & - \\ - & 0 & + \end{pmatrix},$$

respectively.

But also: The pure main effect A is completely confounded with (= contained in the space of) the pure interaction of B and C. The pure main effect B is completely confounded with the pure interaction of A and C.

d) (Compare Cox and Cochran (1) p. 156). Eight plots are grouped into two classes of four plots in four ways. A letter e.g. a indicates the plots that are in the second class, the other plots are in the first class of classification A etc. The fourth way (d) may refer to differences in fertility (= blocks). The space of the pure interactions of A, B and C has as basis:

ab	*ac*	*bc*	(1)

abcd	*ad*	*bd*	*cd*

+ + + +
– – – –

Fig. 19.

This vector is orthogonal to the spaces of the pure main effects A, B, C, but also orthogonal to the spaces of the pure interactions $A \times B$, $A \times C$ and $B \times C$:

$$A \text{ pure: } \begin{matrix} + & + & - & - \\ + & + & - & - \end{matrix}; \qquad A \times B \text{ pure: } \begin{matrix} + & - & - & + \\ + & - & - & + \end{matrix}$$

Conclusion: The pure main effect D and the pure interaction $A \times B \times C$ are completely confounded in this scheme. A similar confounding holds for e.g. $A * B$ and $C * D$, or A and $B * C * D$, etc., as may be verified.

8. ORTHOGONAL CLASSIFICATIONS

Definition 8.

Two classifications A and B are called orthogonal, if the spaces of the pure main effects A* and B* are (completely) orthogonal. If the scheme involves only these two classifications, then also *the scheme is said to be orthogonal.*

Where possible, orthogonal classifications are preferred to non-orthogonal ones. For if P, Q and R are spaces that together span E, and if moreover these spaces are mutually orthogonal, then the 'skew projections' which we would like to know, will turn out to be equal to the orthogonal projections of a vector, and the last can be computed in a simpler way (see chapter 9, 10). The 'main effect component' of a given vector can be computed directly through (orthogonal) projection. If however A and B are not orthogonal, then the computation of the main effect components gives technically greater difficulty, although an approximation can be given (See chapter 11).

Suppose that A and B are orthogonal classifications, i.e. the space A* is completely orthogonal to B*, or also A is completely orthogonal to B*. A basis for A may consist of the class-characteristic functions, the first of which will be: (1, 1, ..., 1, 0, 0, ..., 0, 0).

A basis of B* (that is also orthogonal to the space of the general means) can be obtained from the class characteristic functions of B. Suppose that the i-th class b_i of classification B consists of n_i elements (= points). A basis of the space B* is given by functions of the following type: On the elements of the i-th class the value of the function is $+1/n_i$, on the elements of the j-th class the value of the function is $-1/n_j$, on the other points the function is zero. The vectors of this basis are indeed orthogonal to the space of the general means. (Check this!). If such a vector is orthogonal to the first characteristic function of the classification A: (1, 1, 1, ..., 1, 0, ..., 0), and r_i (r_j) of the numbers $1/n_i$ $(-1/n_j)$ give a contribution $\neq 0$ in the inner product of the before mentioned basis vector of B* and the first class-characteristic function of A, then the inner product of the two vectors will be equal to zero and to $r_i \cdot 1/n_i - r_j \cdot 1/n_j$, and therefore:

$$r_i \cdot 1/n_i - r_j \cdot 1/n_j = 0.$$

Hence $r_1 : r_2 : r_3 : \ldots\ldots = n_1 : n_2 : n_3 : \ldots\ldots$, and the same holds for the other classes of A. *The elements of the classes of B occur in each class of A in the same proportion as in the whole scheme.* A classification that the generality of this necessary and sufficient condition suggests, is presented in fig. 20, where the area of each cell indicates the number of 'points' or plots in it. (Also see chapter 14 and fig. 24).

Other examples are given in chapter 3, fig. 3, 4; chapter 6 example *a, b, c*; chapter 7 example *c*.

Note. For use in agricultural experiments one will not take the scheme of fig. 20 as an experimental plan, but one will randomize the treatment combinations first.

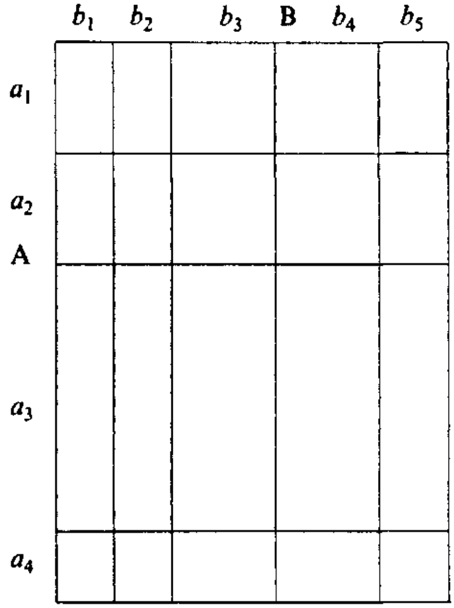


Fig. 20.

9. STATISTICAL CONSIDERATIONS

The first statistical problem

Referring to chapter 4 one can formulate the statistical problem in a simple manner. There is given an experimental result y, a vector in the N- dimensional space E. This y is an observation (realization) of a random vector $\underline{y}$ with a probability distribution as mentioned in chapter 4. It is further known, or it is supposed, that the expectation $\check{y} = E(\underline{y})$ is a vector in a linear subspace D. The subspace D has to be defined for each special problem. One asks for the most likely estimate S($\check{y}$) for $\check{y}$ that is the one for which the probability density at y is the biggest possible.

This maximum likelihood estimate (briefly denoted as M. L.-estimate) is found by minimizing $\{y - S(\check{y})\}^2$ (see chapter 4) and the solution is given by the orthogonal projection of y on D: y_D. (The proof of this geometric theorem will be omitted here.)

Theorem: Given the experimental result y *the maximum likelihood estimate of* $\check{y}$ *is* y_D.

As a rule the space D is spanned by a number of known linearly independent spaces U, V, ... W, such as pure main effect spaces. E.G. one assumes that $\check{y}$ can be composed by an addition of main effects. It is not so much y_D that is of interest, as well as the components of y_D in the spaces U etc. As said before, it is a great technical advantage, if U, V, ..., W are mutually orthogonal, since the mentioned component vectors then are orthogonal projections, which can be computed easily.

The second statistical problem

This concerns the demand for the most likely and sometimes for further still likely estimates for the standard deviation σ. Let R[1] be the space of all vectors in E orthogonal to D. The random vector $\underline{y}_R = \underline{y} - \underline{y}_D$ has the null vector as expectation, for $\check{y} = \check{y}_D$ according to the supposition at the start of this chapter, and therefore

$$E(\underline{y}_R) = E(\underline{y}) - E(\underline{y}_D) = \text{null vector}$$

where $E(\underline{y})$ means expectation of the random vector $\underline{y}$.

A subspace such as R, for which it is a priori known that $E(\underline{y}_R) = 0$, is called a zero-space or *a space of pure error*. The largest space of pure error is called *the space of pure error.*

The random variable $\underline{y}_R^2/\sigma^2$ has the distribution of chi-square $\underline{\chi}^2$ with a number of degrees of freedom, that equals the dimension of R: $d(R) = d(E) - d(D)$. Given an experimental result y, an unbiased estimate $S(\sigma^2)$ of σ^2 is the solution of the right hand side equation in $S(\sigma^2)$:

$$\chi^2 = \frac{y_R^2}{S(\sigma^2)} = \frac{y^2 - y_D^2}{S(\sigma^2)} = d(R)$$

Additional acceptable estimates of σ^2 can be determined in the usual way, applying the table of chi-square, since an estimate $S(\sigma^2)$ of σ^2 is called acceptable if the number

$$y_R^2/S(\sigma^2)$$

computed from it does not take a practically impossible value.

A value for $\chi^2_{d\,R}$ is called practically impossible, if it belongs to a predetemined 'critical region', having probability e.g. .01 or .05 for an experimental result.

[1] Translator's note: In the original T is used for the space of residuals. In the English text we have introduced consistently R of 'residuals' instead of T.

The acceptable values form a confidence-interval at significance level .01 or .05, respectively.

The third statistical problem

This problem, most often met in the analyses of field trials, can be formulated as follows:

In the space D, containing $\check{y}$, is contained a subspace V. One wishes to decide, whether for a given experimental result y the hypothesis that expectation vector $\check{y}$ lies in V, is acceptable. In case $\check{y}$ is in V, the random vector $\underline{y}_D - \underline{y}_V$ (in D, but orthogonal to V) would have the null-vector as its expectation. Further $(\underline{y}_D^2 - \underline{y}_V^2)/\sigma^2$ would have the chi-square-distribution with dimension $d(D) - d(V)$. From the experimental result y one might extract a second estimate of σ^2 by

$$\frac{(y_D - y_V)^2}{d(D) - d(V)}$$

The quotient of this last 'suspected' estimate and the former 'unsuspected' estimate (the two are stochastically independent) is the number

$$f = \frac{(y_D - y_V)^2}{d(D) - d(V)} : \frac{y_R^2}{d(R)}$$

Then f might be considered as a random sample from the F-distribution of Snedecor with $d(D) - d(V)$ and $d(R)$ degrees of freedom respectively. With the table of F we can see, whether the experimental result f may be considered practically possible. Therefore we look for the number F_0 that satisfies at dimensions $d(D) - d(V)$ and $d(R)$:

$$\text{Prob.} = P(\underline{F} \geqslant F_0) = .05 \text{ (or } .01).$$

The choice of the significance level (i.e. .05 or .01) is to be made by the experimenter. A value $F \geqslant F_0$ is considered practically impossible. (It has a probability of at most 1/20 of 1/100 resp.) If the experiment gives such a value for $\underline{f}$, then we conclude that the experimental result does not fit the hypothesis '$\check{y}$ lies in V' and we consider the hypothesis as non-acceptable and reject it.[1]

Note that computations are often simplified by the relation

$$(y_D - y_V)^2 = {y_D}^2 - {y_V}^2$$

which follows directly from the orthogonality: $(y_D - y_V, y_V) = 0$.

[1] For the power of the F-test, see H. B. MANN (7) p. 61.

10. TWO EXAMPLES

Example I

	b_1	b_2	b_3	b_4
a_1	y_{11}	y_{12}	y_{13}	y_{14}
a_2	y_{21}	y_{22}	y_{23}	y_{24}
a_3	y_{31}	y_{32}	y_{33}	y_{34}

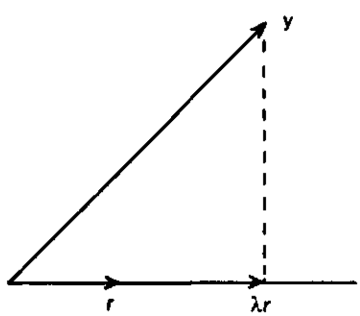
$= \mathrm{y}$

In an agricultural experiment with 12 plots two influences are investigated. One assumes that the effect of these influences on the yield can be written as the sum of an effect to be ascribed to the first influence and one ascribed to the second influence. In other words, one assumes that $\check{\mathrm{y}}$ is the sum of a general mean, a pure main effect in A and a pure main effect in B. In again other words: there is no interaction.

In this case D is the space spanned by A and B. These spaces and their bases have been illustrated in chapter 6, example *b*. We wish to know the vector $\mathrm{y_D}$, i.e. the best estimate for $\check{\mathrm{y}}$, and the skew (*here orthogonal*) components of $\mathrm{y_D}$ in the space of the general means (the vector $\bar{\mathrm{y}}$) and in the spaces of the pure main effects ($\mathrm{y_{A*}}$ and $\mathrm{y_{B*}}$). These are the M.L.-estimates of the general mean, the pure main effect *A* and the pure main effect *B* from this experiment.

The vector $\bar{\mathrm{y}}$. Let

$$\mathrm{r} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$



The required vector is a multiple $\lambda\mathrm{r}$ of r satisfying: $\mathrm{y} - \lambda\mathrm{r}$ is orthogonal to r. Hence

$$(\mathrm{y} - \lambda\mathrm{r}, \mathrm{r}) = (\mathrm{y}, \mathrm{r}) - \lambda(\mathrm{r}, \mathrm{r}) = \Sigma\, \mathrm{y}_{ij} - \lambda \cdot 12 = 0.$$

$$\lambda = \frac{\Sigma \mathrm{y}_{ij}}{12} = \bar{\mathrm{y}}$$

$$\bar{\mathrm{y}} = \begin{pmatrix} \bar{y} & \bar{y} & \bar{y} & \bar{y} \\ \bar{y} & \bar{y} & \bar{y} & \bar{y} \\ \bar{\mathrm{y}} & \bar{\mathrm{y}} & \bar{\mathrm{y}} & \bar{\mathrm{y}} \end{pmatrix}$$

The vector $\mathrm{y_{A*}}$ This vector can be computed most easily as the orthogonal component in the space A* of the vector $\mathrm{y_A}$. The other orthogonal component is $\bar{\mathrm{y}}$, therefore

$$\mathrm{y_{A*}} = \mathrm{y_A} - \bar{\mathrm{y}}.$$

The vector $\mathbf{y}_A$ In the space A we consider the following orthogonal (!) basis:

$$\begin{matrix}1&1&1&1\\0&0&0&0\\0&0&0&0\end{matrix} \qquad \begin{matrix}0&0&0&0\\1&1&1&1\\0&0&0&0\end{matrix} \qquad \begin{matrix}0&0&0&0\\0&0&0&0\\1&1&1&1\end{matrix}$$

The orthogonal components of y in the directions of these vectors are (as a computation, analogous to that for $\bar{\mathbf{y}}$, shows):

$$\begin{matrix}y_{1\cdot}&y_{1\cdot}&y_{1\cdot}&y_{1\cdot}\\0&0&0&0\\0&0&0&0\end{matrix} \qquad \begin{matrix}0&0&0&0\\y_{2\cdot}&y_{2\cdot}&y_{2\cdot}&y_{2\cdot}\\0&0&0&0\end{matrix} \qquad \begin{matrix}0&0&0&0\\0&0&0&0\\y_{3\cdot}&y_{3\cdot}&y_{3\cdot}&y_{3\cdot}\end{matrix}$$

with $y_{i.} = \sum_{j=1}^{4} y_{ij}/4$

The required $\mathbf{y}_A$ is the sum of the above three vectors and therefore consists of row means or more generally: means over the classes of the classification *A. The orthogonal projection* $\mathbf{y}_A$ *of* y *on* A *is thus a scheme that is obtained from* y *by replacing the numbers of* y *in each class of the classification A by the class mean.* Hence:

$$\mathbf{y}_A = \begin{pmatrix}y_{1\cdot}&y_{1\cdot}&y_{1\cdot}&y_{1\cdot}\\y_{2\cdot}&y_{2\cdot}&y_{2\cdot}&y_{2\cdot}\\y_{3\cdot}&y_{3\cdot}&y_{3\cdot}&y_{3\cdot}\end{pmatrix}$$

Analogously we find

$$\mathbf{y}_B = \begin{pmatrix}y_{\cdot1}&y_{\cdot2}&y_{\cdot3}&y_{\cdot4}\\y_{\cdot1}&y_{\cdot2}&y_{\cdot3}&y_{\cdot4}\\y_{\cdot1}&y_{\cdot2}&y_{\cdot3}&y_{\cdot4}\end{pmatrix};\ \mathbf{y}_B^* = \mathbf{y}_B - \bar{\mathbf{y}}$$

Thus the vector y can be decomposed into the following component vectors:

$$\mathbf{y} = \bar{\mathbf{y}} + (\mathbf{y}_A - \bar{\mathbf{y}}) + (\mathbf{y}_B - \bar{\mathbf{y}}) + \mathbf{y}_R \qquad (1)$$

The maximum likelihood estimate for $E(\underline{y})$ is

$$S(\check{\mathbf{y}}) = \bar{\mathbf{y}} + (\mathbf{y}_A - \bar{\mathbf{y}}) + (\mathbf{y}_B - \bar{\mathbf{y}})$$

$\mathbf{y}_A - \bar{\mathbf{y}}, \mathbf{y}_B - \bar{\mathbf{y}}$ are the M.L.-estimates of the pure main effects.

From (1) we can easily compute $\mathbf{y}_R$. Next it follows from (1) (Pythagoras):

$$\mathbf{y}^2 = \bar{\mathbf{y}}^2 + (\mathbf{y}_A - \bar{\mathbf{y}})^2 + (\mathbf{y}_B - \bar{\mathbf{y}})^2 + \mathbf{y}_R^2.$$

Since: $(\mathbf{y}_A - \bar{\mathbf{y}})^2 = \mathbf{y}_A^2 - \bar{\mathbf{y}}^2, (\mathbf{y}_B - \bar{\mathbf{y}})^2 = \mathbf{y}_B^2 - \bar{\mathbf{y}}^2.$

it follows $\mathbf{y}_R^2 = \mathbf{y}^2 - \mathbf{y}_A^2 - \mathbf{y}_B^2 + \bar{\mathbf{y}}^2.$

The terms at the right hand side can be computed directly from the scheme of the numbers y_{ij}. An estimate of σ^2 is given by $\mathbf{y}_R^2/6$ with $d(R) = 6$.

The addition of the dimensions:

General means	1
A pure	2
B pure	3
Residuals	6
	— +
Number of plots	12

It may be that one suspects that $\check{y}_{B*} = 0$ is quite well possible. Then one wishes to investigate if the hypothesis, that the influence *B* has no effect, is acceptable, i.e. whether the experimental results obtained fit that hypothesis! In that case one has $\check{y}_D = \check{y}_A$. The space V mentioned in chapter 9 coincides with A. The space in D that is orthogonal to A happens to be B*. Therefore the hypothesis $\check{y}_{B*} = 0$ will be tested, using the *F*-table with the dimensions 3 and 6, by considering the number f:

$$f = \frac{y_B^2 - \bar{y}^2}{3} : \frac{y_R^2}{6} = \frac{2\,(y_B^2 - \bar{y}^2)}{y^2 - y_A^2 - y_B^2 + \bar{y}^2}.$$

Exercise:

Do the computations for a numerical example (e.g. VAN UVEN [9]).

Example 2

	b_1	b_2		b_3		
a_1	y_{11}	y_{121}	y_{122}	y_{131}	y_{132}	
a_2	y_{211}	y_{221}	y_{222}	y_{231}	y_{232}	
	y_{212}	y_{223}	y_{224}	y_{233}	y_{234}	= y.
	y_{213}	y_{225}	y_{226}	y_{235}	y_{236}	

The addition of the dimensions:

1) General mean	1
3) Main effect *A*, pure	1
5) Main effect *B*, pure	2
7) Interaction $A \times B$, pure	2
8) Residuals, pure error	14
	—
number of plots	20

Besides the mentioned spaces, we may also consider the spaces of crude main effects *A* (2) and *B* (4) and the space of crude interactions $A \times B$ (6).

Bases of these spaces are (read: + = +1; − = −1)

(1) · (2) · (3)

$$\underbrace{\begin{pmatrix} + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \end{pmatrix}}_{(1)} ; \underbrace{\begin{pmatrix} + & + & + & + & + \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} ; \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ + & + & + & + & + \\ + & + & + & + & + \\ + & + & + & + & + \end{pmatrix}}_{(2)} ; \underbrace{\begin{pmatrix} 3 & 3 & 3 & 3 & 3 \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{pmatrix}}_{(3)};$$

$$4) \begin{pmatrix} + & 0 & 0 & 0 & 0 \\ + & 0 & 0 & 0 & 0 \\ + & 0 & 0 & 0 & 0 \\ + & 0 & 0 & 0 & 0 \end{pmatrix} ; \begin{pmatrix} 0 & + & + & 0 & 0 \\ 0 & + & + & 0 & 0 \\ 0 & + & + & 0 & 0 \\ 0 & + & + & 0 & 0 \end{pmatrix} ; \begin{pmatrix} 0 & 0 & 0 & + & + \\ 0 & 0 & 0 & + & + \\ 0 & 0 & 0 & + & + \\ 0 & 0 & 0 & + & + \end{pmatrix} ;$$

$$5) \begin{pmatrix} 2 & - & - & 0 & 0 \\ 2 & - & - & 0 & 0 \\ 2 & - & - & 0 & 0 \\ 2 & - & - & 0 & 0 \end{pmatrix} ; \begin{pmatrix} 0 & + & + & - & - \\ 0 & + & + & - & - \\ 0 & + & + & - & - \\ 0 & + & + & - & - \end{pmatrix} ;$$

$$6) \begin{pmatrix} + & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} ; \begin{pmatrix} 0 & + & + & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{etc.}$$

$$7) \begin{pmatrix} 6 & -3 & -3 & 0 & 0 \\ -2 & 1 & 1 & 0 & 0 \\ -2 & 1 & 1 & 0 & 0 \\ -2 & 1 & 1 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 3 & 3 & -3 & -3 \\ 0 & - & - & + & + \\ 0 & - & - & + & + \\ 0 & - & - & + & + \end{pmatrix}$$

$$8) \begin{pmatrix} 0 & + & - & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} ; \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ + & 0 & 0 & 0 & 0 \\ - & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} ; \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ + & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ - & 0 & 0 & 0 & 0 \end{pmatrix} ; \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & + & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & - & 0 & 0 \end{pmatrix}$$

etc.

Exercise:

Check the orthogonality, where required, of these vectors.

The component vectors

$$\bar{\mathbf{y}} = \begin{pmatrix} \bar{y} & \bar{y} & \bar{y} & \bar{y} & \bar{y} \\ \bar{y} & \bar{y} & \bar{y} & \bar{y} & \bar{y} \\ \bar{y} & \bar{y} & \bar{y} & \bar{y} & \bar{y} \\ \bar{y} & \bar{y} & \bar{y} & \bar{y} & \bar{y} \end{pmatrix} ; \; \bar{y} = \frac{1}{20} \Sigma\, y_{ij(k)}, \text{ summation over all plots.}$$

$$y_A{}^* = y_A - \bar{y}, y_A = \begin{pmatrix} y_{1\cdot} & y_{1\cdot} & y_{1\cdot} & y_{1\cdot} & y_{1\cdot} \\ y_{2\cdot} & y_{2\cdot} & y_{2\cdot} & y_{2\cdot} & y_{2\cdot} \\ y_{2\cdot} & y_{2\cdot} & y_{2\cdot} & y_{2\cdot} & y_{2\cdot} \\ y_{2\cdot} & y_{2\cdot} & y_{2\cdot} & y_{2\cdot} & y_{2\cdot} \end{pmatrix}.$$

where

$$y_{1.} = \frac{1}{5}\Sigma \text{ the numbers in the first row}$$

$$y_{2.} = \frac{1}{15}\Sigma \text{ the numbers of the other rows}$$

(Compare the general definition of y_A in the first example).

$y_{B*} = y_B - \bar{y}$; y_B analogous to y_A.
$y_{A \times B}$-pure $= y_{A \times B} - \bar{y} - (y_A - \bar{y}) - (y_B - \bar{y})$
where

$$y_{A \times B} = \begin{pmatrix} y_{11\cdot} & y_{12\cdot} & y_{12\cdot} & y_{13\cdot} & y_{13\cdot} \\ y_{21\cdot} & y_{22\cdot} & y_{22\cdot} & y_{23\cdot} & y_{23\cdot} \\ y_{21\cdot} & y_{22\cdot} & y_{22\cdot} & y_{23\cdot} & y_{23\cdot} \\ y_{21\cdot} & y_{22\cdot} & y_{22\cdot} & y_{23\cdot} & y_{23\cdot} \end{pmatrix}$$

and

$$y_{12\cdot} = \tfrac{1}{2}(y_{121} + y_{122}), y_{22\cdot} = \tfrac{1}{6}((y_{221} + \ldots + y_{226}), \text{etc.}$$

Finally we have: $y_R = y - y_{A \times B}$

The question now may arise, 'whether there is (in a numerical example) interaction'. This concerns the question whether it is likely that there would be no interaction (and that we then still would get the given experimental result). In different wording, the question is whether it is likely that the expectation vector of the pure interactions equals the null vector, i.e. that $\check{y}$ lies in the space spanned by the spaces A and B. This question (testing the null-hypothesis 'there is no interaction') is answered, by comparing for the given experimental result y, the number

$$f = \frac{(y_{A \times B} - y_A - y_B + \bar{y})^2}{2} : \frac{(y - y_{A \times B})^2}{14}$$

with the number in the *F*-table for the one-sided significance level .05 (or .01) at dimensions 2 and 14.

An other question whether it is acceptable that the influence *B* has no effect, i.e. $\check{y} = \check{y}_A$, is answered on the basis of the experimental result y by comparing the number

$$f = \frac{(y_{A \times B} - y_A)^2/4}{(y - y_{A \times B})^2/14}$$

with the table of *F* at dimensions 4 and 14.

Numerical example[1]

$$\mathbf{y} = \begin{pmatrix} 314 & 268 & 256 & 223 & 239 \\ 254 & 174 & 176 & 183 & 167 \\ 238 & 170 & 159 & 175 & 145 \\ 234 & 157 & 160 & 154 & 154 \end{pmatrix} ; \bar{\mathbf{y}} = \begin{pmatrix} 200 & 200 & 200 & 200 & 200 \\ 200 & 200 & 200 & 200 & 200 \\ 200 & 200 & 200 & 200 & 200 \\ 200 & 200 & 200 & 200 & 200 \end{pmatrix}$$

$$\mathbf{y}_A = \begin{pmatrix} 260 & 260 & 260 & 260 & 260 \\ 180 & 180 & 180 & 180 & 180 \\ 180 & 180 & 180 & 180 & 180 \\ 180 & 180 & 180 & 180 & 180 \end{pmatrix} ; \mathbf{y}_A - \bar{\mathbf{y}} = \begin{pmatrix} 60 & 60 & 60 & 60 & 60 \\ -20 & -20 & -20 & -20 & -20 \\ -20 & -20 & -20 & -20 & -20 \\ -20 & -20 & -20 & -20 & -20 \end{pmatrix}$$

$$\mathbf{y}_B = \begin{pmatrix} 260 & 190 & 190 & 180 & 180 \\ & & \text{etc.} & & \end{pmatrix} ; \mathbf{y}_B - \mathbf{y} = \begin{pmatrix} 60 & -10 & -10 & -20 & -20 \\ & & \text{etc.} & & \end{pmatrix}$$

$$\mathbf{y}_{A\times B} = \begin{pmatrix} 314 & 262 & 262 & 231 & 231 \\ 242 & 166 & 166 & 163 & 163 \\ 242 & 166 & 166 & 163 & 163 \\ 242 & 166 & 166 & 163 & 163 \end{pmatrix}$$

$$\mathbf{y}_{A\times B}\,\text{pure} = \mathbf{y}_{A\times B} - \mathbf{y}_A - \mathbf{y}_B + \bar{\mathbf{y}} = \begin{pmatrix} -6 & 12 & 12 & -9 & -9 \\ 2 & -4 & -4 & 3 & 3 \\ 2 & -4 & -4 & 3 & 3 \\ 2 & -4 & -4 & 3 & 3 \end{pmatrix}$$

$$\mathbf{y}_R = \mathbf{y} - \mathbf{y}_{A\times B} = \begin{pmatrix} 0 & 6 & -6 & -8 & 8 \\ 12 & 8 & 10 & 20 & 4 \\ -4 & 4 & -7 & 12 & -18 \\ -8 & -9 & -6 & -9 & -9 \end{pmatrix}$$

$(\mathbf{y}_{A\times B}\,\text{pure})^2 = 648, \dim = 2; \mathbf{y}_R^2 = 1816, \dim = 14.$

With the F-table we have to compare $f = \dfrac{648/2}{1816/14} = 2.50.$

At dimensions 2 and 14 and significance level .05 we find

$$P(\underline{F} > 3.74/2\,;14) = 0.05,$$

there is not sufficient indication to reject the null hypothesis 'there is no interaction'. 'No interaction' is an acceptable hypothesis.

Exercise:

Investigate under the priori condition 'there is no interaction', whether it is likely that the influence *A* (or *B*, respectively) has a zero pure main effect. What is under this a priori condition the space R? Is it necessary to start the computations all over again?

[1] For the construction and computation of some numerical examples I thank mr. L. C. A. Corsten.

11. NON-ORTHOGONAL SCHEMES

The given scheme y represents an observation from a central symmetric normal distribution of a random vector y in a 11-dimensional vector space (See chapter 4 and 9).

	b_1		b_2	b_3
a_1	y_{111}	y_{112}	y_{12}	y_{13}
a_2	y_{211}	y_{212}	y_{22}	y_{23}
a_3	y_{311}	y_{312}	y_{32}	

We suppose that apart from random error, the effect on the yield by the influences A and B are expressed exclusively by their main effects, and not by any interaction. In other words, the expectation vector $\check{y}$ is a vector in the space D spanned by the spaces A and B (additivity). The M.L.-estimate of $\check{y}$ for the experimental result y is the orthogonal projection y_D. The last one can be resolved in a unique way into components: 1° $\bar{y}$, a vector in the space of the general means, at the same time the orthogonal projection of y onto that space, 2° y_{sA^*}, a vector in the space of pure main effects A* and 3° y_{sB^*} in B*.

Thus we have: $y_D = \bar{y} + y_{sA^*} + y_{sB^*}$

and $y = y_D + y_R$

y_{sA^*} and y_{sB^*} are estimates of pure main effects.

Besides one can consider (= define) crude main effects, viz. $y_{sA} = \bar{y} + y_{sA^*}$; $y_{sB} = \bar{y} + y_{sB^*}$.

The addition of the dimensions:

general means	1
A pure	2
B pure	2
Residuals	6
number of plots	11

(If one would also consider an interaction $A \times B$, the space of the pure interactions would have dimension 3).

Computation of y_D and its component vectors mentioned does not present theoretical difficulties. By means of matrices a solution is easily formulated. However it is possible, as with orthogonal schemes, to compute good approximations to all estimates desired, by a technically simple and quick manner.

The iterative technique that will be discussed, has been devised and applied a.o. by STEVENS (8) and HAMMING (4) (see YATES (11) p. 138). A different important technique (by means of completing missing plots) will not be presented here (see (15)).

A problem in plane geometry

In order to provide an idea of the technique to be presented, we first consider a simple geometric problem. Two non-coinciding vector spaces A and B and the vector z are given in a (2-dimensional) plane.

One requires an expression for the component vectors z_{sA} and z_{sB} which contain only sums, differences and orthogonal projections of vectors. (Our preference for orthogonal projections naturally arises from the fact that orthogonal projections onto spaces that correspond to classifications, can be computed easily.)

Figure 21 suggests that z_{sB} is the limit of the sum of the vectors. (line segments) such as v_1, v_2 etc. An analogous remark can be made about z_{sA}. We now define:

$v_1 = z_B - z_{AB}$.

(z_{AB} is the orthogonal projection of z_A on B);

$u_1 = v_{1A}$, the orthogonal projection of v_1 on A;

$v_2 = u_{1B} = v_{1AB}$

etc.

$v_{i+1} = u_{iB} = v_{iAB} \qquad i = 1, 2, \ldots\ldots$

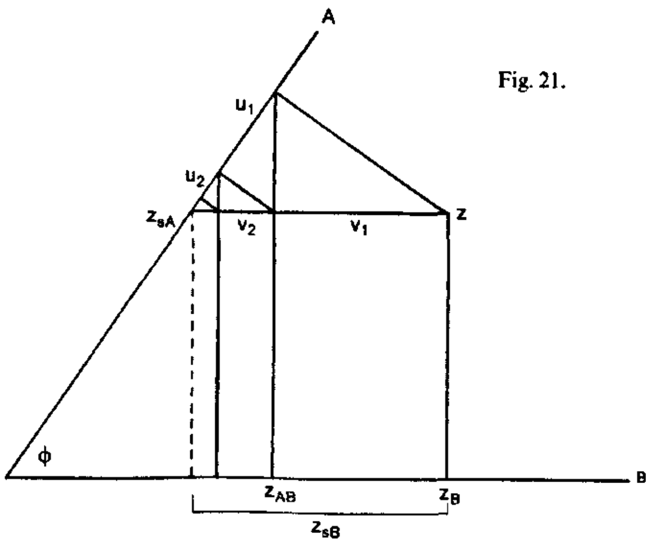


Fig. 21.

From the figure it appears that the lengths of the vectors v_1, u_1, v_2, u_2 .... in this order decrease: each 'new' element is cos ϕ (ϕ = the angle between A and B) times smaller than the preceeding one. By this observation one can prove that the limit vectors

$$z_{0B} = \lim_{n \to \infty} (v_1 + v_2 + v_3 + \ldots + v_n) \text{ and}$$

$$z_{0A} = \lim_{n \to \infty} (z_A - u_1 - u_2 - u_3 - \ldots - u_n)$$

exist. These vectors belong to B and A respectively.

Next we consider $w = z - z_{0A} - z_{0B}$:

$$w = z - z_A + u_1 + u_2 + \ldots - v_1 - v_2 - v_3 \ldots$$

One has:

$$w_A = z_A - z_A + u_1 + u_2 + \ldots - u_1 - u_2 - u_3 \ldots = 0.$$

(check this accurately)

$$w_B = z_B - z_{AB} + v_2 + v_3 + \ldots - (z_B - z_{AB}) - v_2 - v_3 \ldots = 0.$$

The orthogonal projections of w onto A and onto B are zero, therefore w itself is the null vector. Thus it has been proved, that:

$$w = z - z_{0A} - z_{0B} = 0, \text{ i.e. } z = z_{0A} + z_{0B},$$

therefore

$$z_{0A} = z_{sA} = z_A - \sum_{i=1}^{\infty} u_i$$

$$z_{0B} = z_{sB} = \sum_{i=1}^{\infty} v_i.$$

Now we return to our original problem, and take the letters A, B etc. in the original meaning. The difference with our geometrical example is first the fact that A and B do not span the space E. The space E is spanned by A, B and R. Moreover A and B have as intersection the space of the general means. We now define:

$$v_1 = y_B - y_{AB}.$$

This vector is in the space B, even in the space B*: for, because A and B contain the space of the general means, we have

$$v_1 = (y - y_A)_B = (y - y_A)_N + (y - y_A)_{B^*} = (y - y_A)_{B^*}$$

Next we define $u_1 = v_{1A}$; $v_2 = u_{1B}$; in general: $v_{i+1} = u_{iB} = v_{iAB}$.

The vectors u_i and v_i are in the spaces A* and B*, respectively, and the same holds for the vectors

$$y_{0B^*} = v_1 + v_2 + v_3 + \ldots \text{ ad inf.}$$

and

$$y_{0A^*} = y_A - \bar{y} - u_1 - u_2 - u_3 \ldots \text{ ad inf.}$$

The convergence of these infinite sums can be proved.

We will show that y_{0A^*} and y_{0B^*} are the skew projections y_{sA^*} and y_{sB^*}, and therefore are the best estimates of the required pure main effects. For this purpose we consider

$$w = y - y_{0A^*} - y_{0B^*} - \bar{y} = \quad (1)$$

$$w = y - y_A + u_1 + u_2 + u_3 + \ldots - v_1 - v_2 - v_3 \ldots.$$

As before we find directly: $w_A = w_B = 0$, and therefore w is orthogonal to the space A and B, i.e. to D. Thus w is a vector in the space R of residuals, and because the decomposition of y into component vectors, as determined by (1), is unique, we have

$$w = y_R, \; y = \bar{y} + y_{0A\bullet} + y_{0B\bullet} + y_R$$
$$y_{sA\bullet} = y_{0A\bullet}, \; y_{sB\bullet} = y_{0B\bullet}$$

Theorem:

$$y_{sA} = y_A - \sum_{i=1}^{\infty} u_i,$$

$$y_{sB\bullet} = \sum_{i=1}^{\infty} v_i.$$

N.B. If convergence is fast, a good approximation of the main effects is obtained from the first few terms of these series. For an *orthogonal design* the convergence is very fast:

$$y_{sA} = y_A, \; y_{sB\bullet} = v_1 = y_B - \bar{y}.$$

A numerical example.

	b_1		b_2	b_3		y_A (row-means)
a_1	314	327	304	285	4	307,5
$h = a_2$	329	326	305	306	4	316,5
a_3	269	271	264		3	268
	6		3	2		
y_B	306		291	295,5		$\bar{y} = 300$

The vector y_A is explicitly:

$$y_A = \begin{pmatrix} 307{,}5 & 307{,}5 & 307{,}5 & 307{,}5 \\ 316{,}5 & 316{,}5 & 316{,}5 & 316{,}5 \\ 268 & 268 & 268 & \end{pmatrix}$$

y_{AB} is obtained from it by replacing each number by the class average of the B-class to which it belongs. In an analoguous way one computes u_1 from v_1 and v_2 from u_1 etc.

	b_1	b_2	b_3		Y_A	u_1	u_2	u_3	u_4	u_5	
a_1	2	1	1	4	307.5	−1.37	−0.11	0.00	0.00	0.00	−1.48
a_2	2	1	1	4	316.5	−1.37	−0.11	0.00	0.00	0.00	−1.48
a_3	2	1		3	268	3,67	0.31	0.03	0.01	0.00	4.02
	6	3	2								
y_B	306	291	295.5								
y_{AB}	297.33	297.33	312								
v_1	8.67	−6.33	−16.5								
v_2	0.31	0.31	−1.37								
v_3	0.03	0.03	−0.11								
v_4	0.01	0.01	0.00								
v_5	0.00	0.00	0.00								
	9.02	−5.98	−17.98								

$y_{sB\bullet} = (9.02\ ;\ -5.98\ ;\ -17.98)$
$y_{sA\bullet} = (8.98\ ;\ 17.98\ ;\ -36.02)$

$$w = y - y_{sB\bullet} - y_{sA\bullet} - \bar{y} = \begin{pmatrix} -4 & 9 & 1 & -6 \\ 2 & -1 & -7 & 6 \\ -4 & -2 & 6 & \end{pmatrix}$$

As a check one may compute w_A and w_B. If no computational errors have been made, a good approximation to the null vector should be obtained in both cases. (Otherwise we repeat the same operation on w.) If: $w_A = w_B \approx 0$, then w is a good approximation to y_R, from which one can compute y_R^2.
The M.L.-estimate for σ^2 is

$$y_R^2/dR = y_R^2/6 = 280/6 = 46.67.$$

Additional acceptable estimates for σ^2 can be computed using the table of chi-square

We finally mention the third statistical problem for this case. E.g. we would like to test the hypothesis $\check{y}_{sB\bullet}$ = null vector.

We thus ask whether it is an acceptable hypothesis that the influence *B* is without any effect.

If this null hypothesis is true, then a second estimate of σ^2 will be (Note that y_A is the *orthogonal* projection):

$$\frac{(y_D - y_A)^2}{d(B^*)} = \frac{y_D^2 - y_A^2}{d(B^*)} = \frac{y^2 - y_R^2 - y_A^2}{d(B^*)}$$

The computation gives 1202/2 = 601.

Whether this null hypothesis is acceptable, may be decided by comparing the quotient f of the two estimates of σ^2 with the F-table at dimensions $dB^* = 2, dR = 6$.

$$f = \frac{y_D^2 - y_A^2}{d(B^*)} : \frac{y_R^2}{d(R)} = \frac{1202}{2} : \frac{280}{6} = 12{,}88$$

$$P(F > 10{,}92 / 2; 6) = 0{,}01$$

Hence the null hypothesis will be rejected. Influence *B* has effect.

On the speed of convergence

The vectors v_1, u_1, v_2, u_2, ... are each (except v_1) an orthogonal projection of a vector (the preceeding one) in the space B* (A*) onto the space A* (B*). The proportion of the lengths of two successive vectors in the sequence equals cos ϕ, where ϕ is the angle between the two vectors. The angle between an arbitrary vector in A* and a vector in B* has a minimum (if all pairs of vectors are considered). To this minimum ($\neq 0$, unless confounding occurs) corresponds a maximum of cos ϕ which will be represented by g. The $(n + 1)^{th}$ vector in the sequence v_1, u_1, v_2, ... has a length $\leqslant g^n \times$ length v_1. It follows that for small g convergence will be fast. The value of g for an orthogonal design equals $g = 0$.

For a non-orthogonal design that is obtained from an orthogonal scheme with m rows and n columns by omitting one plot, one has

$$g = \frac{1}{mn}$$

We remark that designs with more than two non-orthogonal classifications, can be analysed in an analogous way.

12. A FIELD TRIAL WITH LINEAR FERTILITY TREND

b_1	b_2	b_3	b_4
1	3	1	2
2	2	3	1
3	1	2	3

The given scheme represents a design of a trial field. The numbers 1, 2 and 3 refer to an influence *A* (three varieties of wheat). Based on previous experience one presumes a fertility trend increasing from left to right. Moreover it is supposed, that fertility increases linearly with distance. A fertility trend according to complete rows of the scheme is assumed to be absent. In order to avoid the confounding of any further fertility trends with varietal differences the varieties have been randomized within each column. The rearranged table of yields is as follows:

$$y = \begin{array}{c|c|c|c|c|} & b_1 & b_2 & b_3 & b_4 \\ \hline a_1 & y_{11} & y_{12} & y_{13} & y_{14} \\ \hline a_2 & y_{21} & y_{22} & y_{23} & y_{24} \\ \hline a_3 & y_{31} & y_{32} & y_{33} & y_{34} \\ \hline \end{array} = \begin{array}{|c|c|c|c|} \hline 479 & 491 & 512 & 510 \\ \hline 495 & 493 & 515 & 525 \\ \hline 484 & 492 & 497 & 507 \\ \hline \end{array}$$

It is also supposed that the influence of fertility trend and that of varietal differences have additive effects on yield.

If we consider the space B corresponding to classification *B*, then it may be noticed that the linear fertility trends belong to it. An example of a *pure* (= orthogonal to the space of the general means) linear trend is:

$$q = \begin{pmatrix} -3 & -1 & 1 & 3 \\ -3 & -1 & 1 & 3 \\ -3 & -1 & 1 & 3 \end{pmatrix} \quad ; \text{L is the space with basis q.}$$

It is now assumed that the expectation vector $\check{y}$, to be estimated, belongs to the space D, spanned by the spaces of (wheat varieties) main effects A, and L, the space with basis q. Again R is the space orthogonal to D, i.e. orthogonal to A and L.

As L is in B*, L is orthogonal to A and we thus have:

$$y_D = y_A + y_L; y_A = y_{A\bullet} + \bar{y}; y = y_D + y_R; \text{and also}$$
$$y^2 = y_A^2 + y_L^2 + y_R^2; y_{A\bullet}^2 = y_A^2 - \bar{y}^2.$$

The addition of the dimensions

General means	1
A pure	2
L	1
Residuals	8
	—
Number of plots	12

For $\bar{y}$ and y_A (row means) we find for the numerical example:

$$\bar{y} = \begin{pmatrix} 500 \text{ etc.} & & \\ & & \\ & & \end{pmatrix} \quad ; \quad y_A = \begin{pmatrix} 498 \text{ etc.} & & \\ 507 & & \\ 495 & & \end{pmatrix}$$

$$\bar{y}^2 = 12 \times 500^2 = 3{,}000{,}000; y_A^2 = (498^2 + 507^2 + 495^2).4 = 3{,}000{,}312.$$

The computation of y_L proceeds as follows: y_L is a vector of shape λq (λ a real number), such that $y - \lambda q$ is orthogonal to q: $(y - \lambda q, q) = (y, q) - \lambda(q, q) = 0$. Hence

$$\lambda = \frac{(y, q)}{(q, q)} =$$

$$= \frac{-3(y_{11}+y_{21}+y_{31}) - (y_{12}+y_{22}+y_{32}) + (y_{13}+y_{23}+y_{33}) + 3(y_{14}+y_{24}+y_{34})}{27 \quad + \quad 3 \quad + \quad 3 \quad + \quad 27}$$

$$= \frac{-3.1458 - 1476 + 1524 + 3.\,1542}{60} = \frac{300}{60} = 5;$$

$$y_L = \begin{pmatrix} -15 & -5 & 5 & 15 \\ -15 & -5 & 5 & 15 \\ -15 & -5 & 5 & 15 \end{pmatrix};$$

$$y_L{}^2 = \lambda^2 q^2 = 25.60 = 1500.$$

For the test of the null hypothesis: 'the varieties do not differ in productivity (equal yields)' we compare with the F-value at numbers of degrees of freedom $d(A^*) = 2$ and $d(R) = 8$, the number f:

$$f = \frac{y_{A*}{}^2}{2} : \frac{y_R{}^2}{8} = \frac{y_A{}^2 - \bar{y}^2}{2} : \frac{y^2 - y_A{}^2 - y_L{}^2}{8} = \frac{312}{2} : \frac{256}{8} = 4{,}875.$$

We find in the table: $P(\underline{F} > 4.46 \mid 2;8) = .05$.
The null hypothesis is rejected at a significance level .05; in other words: The varieties differ in their production.

Comparison with the usual procedure

The above procedure makes sense if we know beforehand that a linear fertility trend is a natural description of reality. Sometimes this hypothesis is plausible. The advantage in the method used is that the dimension of R is higher than if we define R as the space orthogonal to A and B (instead of A and L): d.f. equals 8 instead of 6. The 'unsuspected' estimate of σ^2 thus is more accurate (= more efficient = the standard deviation of the (random) estimate is smaller) and this implies, that if the varieties are really different, we have with the procedure given here, more chance of rejecting the null hypothesis than with the usual procedure at the same significance level (.05 or .01). In other words: The analysis (test procedure) used here has *more power*.

In the example this situation accidentally finds a clear illustration. The maximum likelihood estimates of σ^2 are

$$S_I = \frac{y^2 - y_A{}^2 - y_L{}^2}{8} = \frac{256}{8} = 32 \text{ against } S_{II} = \frac{y^2 - y_A{}^2 - y_B{}^2 + \bar{y}^2}{6} =$$

$$= \frac{196}{6} = 32.67.$$

The values of f are 4.88 and $(y_A^2 - \bar{y}^2)/2$: S_{II} = 4.77, respectively. The numbers of degrees of freedom are (2;8) and (2;6), respectively.
The F-table gives

$$P(\underline{F} > 4.46 \mid 2;8) = .05$$
$$P(\underline{F} > 5.14 \mid 2;6) = .05$$

Only in the first case the null hypothesis will be rejected at significance level .05.

Note that with non-orthogonal schemes of the above type, the theory of chapter 11 can be applied.

13. LATIN SQUARES

A Latin square is a scheme of n^2 plots, that in more than two (usually three[1]) ways can be grouped in n classes of n elements, such that every pair of classifications is orthogonal. In experiments the three classifications correspond to three influences, for which one assumes additivity in the effects on 'yield' or any other experimental result. In agricultural field trials sometimes two of the classifications represent fertility trends in two mutually perpendicular directions. In fig. 22 rows and columns correspond to classifications A and B;

	b_1	b_2	b_3	b_4	b_5
a_1	c_1	c_2	c_3	c_4	c_5
a_2	c_5	c_1	c_2	c_3	c_4
a_3	c_4	c_5	c_1	c_2	c_3
a_4	c_3	c_4	c_5	c_1	c_2
a_5	c_2	c_3	c_4	c_5	c_1

Fig. 22.

	c_1	c_2	c_3	c_4	c_5
a_1	b_1	b_2	b_3	b_4	b_5
a_2	b_2	b_3	b_4	b_5	b_1
a_3	b_3	b_4	b_5	b_1	b_2
a_4	b_4	b_5	b_1	b_2	b_3
a_5	b_5	b_1	b_2	b_3	b_4

Fig. 23

[1] The name Latin square refers to three orthogonal classifications. If four orthogonal classifications are involved, then the name Graeco-Latin square is used. Latin and Greek letters in addition to rows and columns are used to indicate classifications.

in fig. 23 the same classifications are represented, rearranged such that rows and columns represent classifications A and C. An experimental result may be as follows:

(We use the arrangement of fig. 22)

	b_1	b_2	b_3	b_4	b_5
a_1	$y_{11}{}^1$	$y_{12}{}^2$	$y_{13}{}^3$	$y_{14}{}^4$	$y_{15}{}^5$
a_2	$y_{21}{}^5$	$y_{22}{}^1$	$y_{23}{}^2$	$y_{24}{}^3$	$y_{25}{}^4$
$y = a_3$	$y_{31}{}^4$	$y_{32}{}^5$	$y_{33}{}^1$	$y_{34}{}^2$	$y_{35}{}^3$
a_4	$y_{41}{}^3$	$y_{42}{}^4$	$y_{43}{}^5$	$y_{44}{}^1$	$y_{45}{}^2$
a_5	$y_{51}{}^2$	$y_{52}{}^3$	$y_{53}{}^4$	$y_{54}{}^5$	$y_{55}{}^1$

The relevant spaces are:

	dimension	
General means	1	1
A pure	$n-1$	4
B pure	$n-1$	4
C pure	$n-1$	4
residuals	n^2-3n+2	12
Number of plots	n^2	25

In addition to the general assumption of chapter 9, we assume that the vector $\check{y}$ belongs to the space D spanned by spaces A, B and C. The best estimate of $\check{y}$ from the experimental result y is the orthogonal projection y_D. As A*, B*, C* are mutually orthogonal, the decomposition of y_D into component vectors is given by orthogonal projections:

$$y_D = \bar{y} + y_A{}^* + y_B{}^* + y_C{}^*$$
$$y_D = \bar{y} + (y_A - \bar{y}) + (y_B - \bar{y}) + (y_C - \bar{y})$$
and $$y = y_D + y_R$$

where y_{A^*} etc. are the best estimates of the pure main effects from the experimental result y.

According to the theorem of Pythagoras one has:

$$y^2 = y_D{}^2 + y_R{}^2 = \bar{y}^2 + (y_A - \bar{y})^2 + (y_B - \bar{y})^2 + (y_C - \bar{y})^2 + y_R{}^2,$$

and because of

$$(y_A - \bar{y})^2 = y_A{}^2 - \bar{y}^2 \text{ etc.}$$

$$y^2 = y_A{}^2 + y_B{}^2 + y_C{}^2 - 2\bar{y}^2 + y_R{}^2.$$

Thus y_R^2 may be computed.

If now A and B in an agricultural field trial refer to fertility trends, C to varieties, and y to production (yield), then the question might be posed whether the null hypothesis 'the varieties do not differ in productivity' is acceptable given the experimental result. At this null hypothesis $y_R^2/d(R) = y_R^2/(n^2 - 3n + 2)$ is an estimate of pure error and $y_{C\bullet}^2/dC^* = (y_C^2 - \bar{y}^2)/(n - 1)$ is a suspected estimate. Testing the null hypothesis requires comparing

$$f = \frac{(y_C^2 - \bar{y}^2)/(n-1)}{y_R^2/(n^2 - 3n + 2)}$$

with the F-table.

For numerical examples see the literature.

14. MULTIPLE COMPLETELY-ORTHOGONAL CLASSIFICATIONS (FACTORIAL DESIGNS)

A set of classifications $A, B, \ldots, C$ of N points is called completely orthogonal, if each (e.g. A) is orthogonal to the product classification of the other classifications ($B, \ldots, C$), i.e. to the classification that consists of classes that are intersections of classes of all other classifications. If the classifications refer to influences, as is usual in this paper, then the property can also be formulated as follows: The space of pure main effects of any influence, is orthogonal to all spaces of interactions and main effects of the other influences, i.e. orthogonal to the spaces of all crude interactions of the other influences.

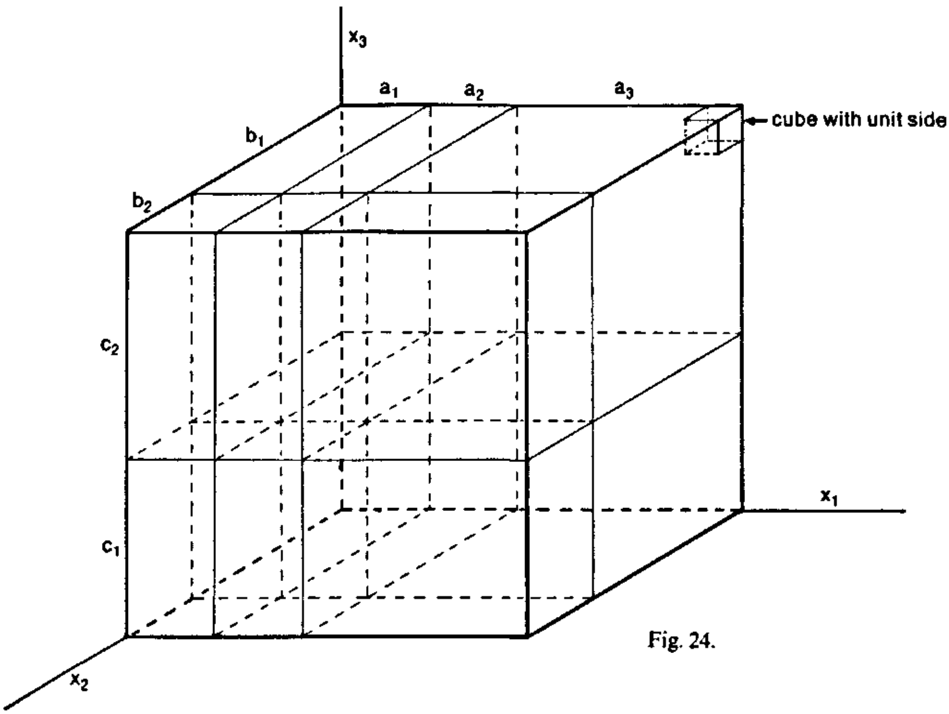


Fig. 24.

If the number of completely orthogonal classifications of N points is n, then it is possible to choose n variables $x_1, x_2, \ldots, x_n$ with range of definition: $x_i = 1, \ldots, p_i$ such that each of the $p_1, p_2, \ldots, p_n = N$ points will be characterized by the coordinates $(x_1, \ldots, x_n)$, where the i-th classification is characterized by the i-th coordinate: a class will consist of all points with a value in a given interval on the i-th line of real numbers as i-th coordinate. The general situation for $n = 3$ is suggested in fig. 24.

In the applications occurring in practice, the classes of a classification are often equally large: N/p_i (often with also $p_i = $ m for all i). The classification is then completely determined by these numbers p_i and one has a so-called $p_1 \times p_2 \times \ldots p_n$-scheme. Many examples are given in COX and COCHRAN (1) Ch. 5.

We will make a few remarks concerning a $p_1 \times p_2 \times p_3 \times p_4$-scheme for influences A, B, C, D and refer for examples and more details to the literature.

The addition of the dimensions

General means	1
A pure	$p_1 - 1$
B pure	$p_2 - 1$
etc.	
$A \times B$ pure	$p_1p_2 - p_1 - p_2 + 1 = (p_1 - 1)(p_2 - 1)$
etc.	
$A \times B \times C$	$(p_1 - 1)(p_2 - 1)(p_3 - 1)$
etc.	
$A \times B \times C \times D$	$p_1p_2p_3p_4 - \sum_{i<j<k} p_ip_jp_k + \sum_{i<j} p_ip_j - \Sigma p_i + 1$ +
number of plots	$p_1p_2p_3p_4$

Generally one will assume a priori in a case like this that some of the pure interactions do not exist (i.e. the relevant component of the expectation vector equals the null vector). The corresponding vector components of the vector y, representing an experimental result, will provide a reliable estimate of σ^2, against which one can test other suspected estimates of σ^2 by means of the F-table in the usual manner. Some components of y are:

$\bar{y}, y_A, y_{A^*}, y_{A\times B}$, with pure interaction com ponent:
$y_{A\times B} - y_A - y_B + \bar{y}$, etc.

If one assumes all interactions of 3[rd] and 4[th] order to vanish, then one has:

$$y = \bar{y} + (y_A - \bar{y}) + \text{etc.} + (y_{A\times B} - y_A - y_B + \bar{y}) + \text{etc.} + y_R$$

and because of mutual orthogonality of the component vectors:

$$y^2 = \bar{y}^2 + (y_A - \bar{y})^2 + \text{etc.} + (y_{A \times B} - y_A - y_B + \bar{y})^2 + \text{etc.} + y_R^2.$$

from which y_R^2 can be obtained.

We finally note, that if $p_I = 2$, the pure main effect space of the influence concerned is 1-dimensional. A component in such a space can be characterized by one number. If more pure main effect spaces are 1-dimensional, then so are also their pure interaction spaces. This is the case with $2 \times 2 \times 2 \times \ldots\ldots \times$ 2-designs (See COX and COCHRAN (1)).

SUMMARY

Introduction: Some theory of linear vector spaces can be applied at not too hard a mathematical level to some problems of analysis of variance. It is then possible to define some much used notions (main effect, interaction, confounding, orthogonality), and many experimental designs and their analysis get rather transparent, partly as a consequence of simple notations.

Chapter 3. A function $(y_1, \ldots, y_N)$ defined over N points (= e.g. experimental units in agricultural field trials) assigns a real number to any of these points. Such functions can be added and they can be multiplied with a real number: they can be considered as vectors and then form an N-dimensional vector space E. Subspaces are: the 1-dimensional space of constant functions (of general means); the space of functions that are constant within the classes of a classification of the N points (fig. 2). If the classification corresponds with some influence A (amount of phosphor added to the field; variety; fertility) then this space denoted by A is called the space of crude main-effects of influence A. An independent basis, and hence the dimension of A are determined. The number of degrees of freedom of a sub-space of E is defined to be its dimension. Also a space of crude interactions of two (or more) influences A and B is defined.

Chapter 4. The random variables $\underline{y}_1, \ldots \underline{y}_N$ have a normal distribution with expectation values $\check{y}_1, \ldots, \check{y}_N$ and all with the same variance σ^2. They are combined to a random N-dimensional vector $\underline{y}$ with expectation vector $\check{y}$.

It has sense to define a metric in the vector space, determined by a scalar product: $(a, b) = a_1b_1 + a_2b_2 + \ldots + a_Nb_N$. The length of a vector is $\sqrt{(a, a)}$. The angle between a and b is ϕ: $\cos\phi = (a, b) / \sqrt{(a, a) \cdot (b, b)}$. The orthogonal projection of a vector y on a linear space A is denoted by y_A. It follows that $(\underline{y} - \check{y})^2/\sigma^2$ and $(\underline{y}_A - \check{y}_A)^2/\sigma^2$ are random variables with distribution of χ^2 with N and d(A) (= dimension of A) degrees of freedom.

Chapter 5. The space A* of pure main effects of the influence A is the linear subspace of A perpendicular to the space of general means.

Chapter 6. The space of pure interactions of A and B is defined to be the subspace of the space A × B of crude interactions, perpendicular to the spaces A and B.

Chapter 7. Two influences or interactions are confounded, if their (pure) spaces meet in a space of dimension > 0. Confounding may be complete or partial.

Chapter 8. All orthogonal two-way classifications are determined.

Chapter 9. Statistical considerations are given. In particular the F-test is mentioned.

Chapter 10. Two examples of orthogonal two-way-classifications.

Chapter 11. How to deal with non-orthogonal classifications. An iteration technique to approximate maximum likelihood estimates for two main effects, given by Stevens and Hamming, is proved and understood in terms of vectors.

Chapter 12. An example with two effects, one of which is linear fertility, is given in detail.

Chapter 13. Latin squares.
Chapter 14. Factorial designs.

LITERATURE

1. COCHRAN, W. G. and G. M. COX (1950), *Experimental design*, New York, John Wiley.
2. CRAMÈR, H. (1946), *Mathematical methods of statistics*, Princeton, Princeton Univ. Press.
3. FISHER, R. A. (1946), *Statistical Methods for Research Workers*, 10th ed., Edinburgh, Oliver and Boyd.
4. HAMMING, G. (1952), *Cursus: De wiskundige achtergrond van variantieanalyse, correlatierekening en polyfactoranalyse*, Stencilled syllabus.
5. HOEL, P. G. (1947), *Introduction to Mathematical Statistics*, New York, John Wiley & Sons, Inc.
6. KENDALL, M. G. (1946), *The Advanced Theory of Statistics*, Vol. II, London, Charles Griffin & Cy, Ltd.
7. MANN, H. B. (1949), *Analysis and Design of Experiments*, New York, Dover Publications, Inc.
8. STEVENS, W. L. (1948), *Statistical Analysis of a non-orthogonal tri-factorial experiment*, Biom, (35) 346–367.
9. UVEN, M. J. VAN (1946), *Mathematical Treatment of the results of agricultural and other experiments*, 2d ed., Groningen, P. Noordhoff N.V.
10. WILKS, S. S. (1950), *Mathematical Statistics*, Princeton, Princeton Univ. Press.
11. YATES, F. (1949), *Sampling Methods for censuses and surveys*, London, Charles Griffin & Cy, Ltd.
12. YATES, F. (1937), *The Design and Analysis of Factorial Experiments*, Harpenden, England, Imp. Bureau of Soil Science.
13. KUIPER, N. H. (1953), *Vectoren en Variantieanalyse (elementair)*, Biom. Contact **5** (1) 22–35.
14. KUIPER, N. H. (1953), *On 4^3-factorial designs with confounding*, Neth. J. of Agr. Sci. **1**, 11–14.
15. KUIPER, N. H. and L. C. A. CORSTEN (1953), *Open plaatsen in variantieschema's*, Meded. van de Landbouwhogeschool **53** (2) 25–30.

ADDITIONAL REFERENCES ON ANALYSIS OF VARIANCE AND CONNECTED TOPICS

L. C. A. CORSTEN (1958). Vectors, a tool in statistical regression theory. Mededelingen van de Landbouwhogeschool, Wageningen, Nederland **58** (1) 1–92.

L. C. A. CORSTEN (1976). Canonical correlation in incomplete blocks. Essays in Probability and Statistics p. 125–154. S. Ikeda a.o. (ed.). Shinko Tsusho Co. Ltd. 1-7-1 Wakaba, Shinjukuku; Tokyo 160, Japan.

M. KEULS and F. GARRETSEN (1982). Statistical Analysis of growth curves in plant breeding. Euphytica **31** (1982) 51–64.

M. KEULS, W. J. KOOPS and J. T. N. M. THISSEN (1981). Hidden disconnectedness and analysis of variance. Genstat Newsletter, Issue no 7, March 1981, 10–13.

J. DE KROON and P. VAN DER LAAN. (1983) A generalization of Friedman's rank statistic. Statistica Neerlandica **37** (1) 1–14.

H. N. KUIPER (1960). Random variables and random vectors. Bull. de l'Institut Agronomique et des Stations de Recherches, Gembloux 344–355.

N. H. KUIPER (1961). Random variables rather than distribution functions, a different and useful approach in statistics. Mededelingen van de Landbouwhogeschool, Wageningen, Nederland **61**(12) 1–19.

N. H. KUIPER (1962). Linear Algebra and Geometry. Amsterdam, North-Holland Uitgeversmij.