VECTORS, A TOOL IN STATISTICAL REGRESSION THEORY

L. C. A. CORSTEN

VECTORS, A TOOL IN STATISTICAL REGRESSION THEORY

(MET EEN SAMENVATTING IN HET NEDERLANDS)

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD
VAN DOCTOR IN DE LANDBOUWKUNDE
OP GEZAG VAN DE RECTOR MAGNIFICUS IR. W. DE JONG,
HOOGLERAAR IN DE VEETEELTWETENSCHAP,
TE VERDEDIGEN TEGEN DE BEDENKINGEN
VAN EEN COMMISSIE UIT DE SENAAT
DER LANDBOUWHOGESCHOOL TE WAGENINGEN
OP VRIJDAG 10 JANUARI 1958 TE 16 UUR

DOOR

LEO CASPAR ANTOON CORSTEN

H. VEENMAN & ZONEN - WAGENINGEN - 1957

De inhoud van dit proefschrift verschijnt tevens in de Mededelingen van de Landbouwhogeschool te Wageningen/Nederland **58** (1) 1958 en als Mededeling van het Instituut voor Rassenonderzoek van Landbouwgewassen te Wageningen no 35.

Aan mijn Ouders
Voor mijn Vrouw

WOORD VOORAF

De goede gelegenheid, die de aanbieding van deze dissertatie biedt om openlijk allen te danken, die direct of indirect hebben medegewerkt aan mijn vorming en aan het tot stand komen van dit werk, grijp ik gaarne aan. Want aan het tot dankbaarheid stemmend bewustzijn, dat, hetgeen men maakt en kent, afgeleid wordt van God's gaven en van de vorming en hulp van anderen, kan men slechts zelden uiting geven. Zonder iemand van die „anderen" iets te kort te willen doen, zou ik enigen in het bijzonder willen noemen.

Hooggeleerde KUIPER, ofschoon ik reeds vele jaren contact met U heb gehad, word ik bij elke nieuwe ontmoeting getroffen door Uw scherpzinnigheid, gepaard aan een warme menselijkheid. Ik beschouw het daarom als een groot voorrecht, dat U zo een groot aandeel aan mijn vorming, en dat niet alleen in de wiskunde, hebt bijgedragen en nog steeds bijdraagt. Voor Uw stimulerende lessen, en voor de vele uren, gewijd aan de soms moeizame doorwerking van dit proefschrift, zeg ik U hartelijk dank. In mijn dank voor de gastvrijheid, die ik zo vaak in Uw huis mocht genieten, wil ik ook Mevrouw KUIPER betrekken.

Hooggeleerde VAN UVEN, aan Uw hand mocht ik de eerste stappen op het pad van de statistiek zetten. De geestdrift voor de wiskunde, die U toen en, ondanks een verdiende rust, nog vele malen daarna tentoonspreidde, vervult mij met grote bewondering.

Hooggeleerde HELLINGA, voor het feit dat Uw gewaardeerd enthousiasme voor de cultuurtechniek, niet voldoende voor mij is geweest om haar geheel trouw te blijven, ligt de verklaring bij mij. Ik dank U voor Uw voortgezette sympathieke belangstelling.

Hooggeleerde SCHUFFELEN, de vorm van Uw ingenieurscolleges was voor mij een voorbeeld in wetenschapsbeoefening.

Hooggeachte GROENEWOLT, de ruime interpretatie die U steeds geeft aan de beoefening van de statistiek ten behoeve van het rassenonderzoek en de daarmee corresponderende vrijheid en medewerking, die U mij daartoe hebt verleend, speciaal in de laatste en de komende maanden, stel ik op hoge prijs. In U dank ik ook alle medewerkers aan het Instituut voor Rassenonderzoek van Landbouwgewassen, wier hulp ik in gevarieerde vorm mocht ondervinden.

Waarde JUSTESEN, de vaak zeer uitvoerige gedachtenwisselingen over tal van onderwerpen uit de statistiek, waartoe U zich telkens weer, als was het vanzelfsprekend, bereid verklaarde, hebben mij zeer verrijkt; ik hoop mij aan Uw weldadige kritiek te kunnen blijven blootstellen.

Waarde KEULS, van Uw neiging tot generaliseren, in de gunstige zin des woords, zijn de sporen terug te vinden in dit proefschrift. Voor de goede invloed, die U op mijn statistisch denken gehad hebt, dank ik U hartelijk.

Het personeel van de Afdeling Wiskunde der Landbouwhogeschool en van het Centrum voor Landbouwwiskunde zeg ik voor de door mij als gast ondervonden belangstelling en hulp gaarne dank.

Voor Uw opofferingen en moeiten, Vader en Moeder, om mij tot de studie in staat te stellen, moge ik U, mede door de opdracht van dit proefschrift, bedanken. Juist daarom betreur ik het des te meer, Moeder, dat U het in akademische zin voltooien van mijn studie slechts van verre kunt meebeleven.

Mijn lieve vrouw, voor de steun, die jij mij, op de jou eigen wijze, bent geweest, kan ik niet dankbaar genoeg zijn.

CONTENTS

INTRODUCTION

An often met supposition about a set of random variables, each of which concerns the observation of the same, quantitatively expressible property of a set of individuals, is that the expectation of such a random variable is a function in the values of other, likewise quantitatively expressible, properties of the same individual. This function, common to the set of individuals, is called the regression function of the first property on the other properties. If this function is linear in a set of (unknown) parameters, the regression function is called linear. Estimation and testing concerning the parameters in linear regression functions is the subject of statistical regression theory.

This theory includes a great variety of subjects, such as analysis of variance, analysis of covariance, regression theory in a narrower sense (the mentioned function explicitly given), experimental designs, analysis of series of experiments etc. About all these subjects there exists an extensive literature. In the greater part of this literature the following facts are striking:

1. Several terms, intensively used, are ill-defined and handled more or less intuitively it seems; as examples we call: degrees of freedom, orthogonality, comparisons, adjustment, interaction, confounding, recovery of inter-block information etc.
2. Proofs and derivations are discouraging by their lengths, abundance of symbols and cumbersome computations. This is even true for rather simple situations.
3. Presumably as a consequence of both foregoing facts, *simple* methods are wanting in case of (in practice often inevitable) deviations from the customary particular situations and designs. In other words, the general cases are neglected, also because of their feared difficult analysis.
4. Testing is relatively overstressed in comparison with estimation, especially in the books on experimental design.

Although the representation of the set of observations of the mentioned random variables, as a point in an Euclidean (hyper)space (and similarly for the values of the other properties), has been used sometimes as an illustration of some results in this theory, KUIPER (11, 13) *founded* this theory on the notion of vector spaces.

By means of the notion of vector spaces many definitions, so far more or less vague, can be given in a sound and comprehensive way. The view proceeding from those definitions, by which the experimental result is considered as one entity that, for purposes of estimating and testing, must be decomposed in interesting components, is clarifying and facilitates the comprehension, also for the beginner in this field. The geometrical language is a great intuitive support: cumbersome computational processes can be overlooked and summarized by means of simple geometrical terms such as projection, orthogonality, perpendicular, length, angle, dimension, space. Further, properties, valid in ordinary solid geometry, can be applied by analogy. By means of this tool a simple and transparent notation arises. Proofs and derivations are considerably simplified by the introduced notions and notations. It brings the estimation, which in our view is the most important aspect of the analysis of statistical data, to the forefront; testing plays a secondary rôle. Also more general situations than those occurring in textbooks about the subject of regression analysis can be mastered without use of a deterrent and insurveyable arithmetic. It is true that, for instance, many non-orthogonal designs have the disadvantageous property that the variance of estimators is not determined simply, in general; but this is no reason to abandon these designs, the more as they are offered necessarily in this form in certain branches of research (e.g. varietal research), and as the customary designs are too restricted in their possibilities. Moreover, the study of general cases appears to lead to a bright insight in the particular situations usually treated.

The following study aims at the demonstration of the usefulness of this tool and at the presentation of new results. Although for other subjects in statistics this tool appeared to be also valuable (CORSTEN [4]) we have restricted ourselves to the subject of linear regression (in the wider sense of the word). Of course this study rests on the ideas and, partly, on the results of KUIPER. Now we give a brief outline of the construction and the origin of the study.

In chapter 1 we present the tool, linear algebra, which is of a non-statistical nature. We judged the insertion of this theory necessary for the following reason: although a great part of the exposition is not new, it is not simply within reach of the statistician; moreover, certain aspects of this theory must be considered in more detail than usually is done in abstract linear algebra, this in connection with the applications. In composing this chapter we are supported by the lectures of KUIPER and by the book of HALMOS (7). The general iterative method for decomposition of vectors is a new contribution; for special cases a (not proved) arithmetical method which amounts to the same, was invented by STEVENS (20) and HAMMING (8), (see also YATES [22]), while the special case of two classifications was derived by KUIPER (11) in terms of vectors.

In chapter 2 the notion of random vectors is introduced. This makes it possible to posit the regression problem in its most general form, to derive unbiased and most efficient estimators of the parameters, and to consider the appropriate tests, all in terms of vectors and, therefore, in a surprisingly simple way.

In chapter 3 we consider the application of the theory, developed in the first two chapters, in some special cases of increasing difficulty. In all these cases the random variables have covariance zero.

We mention: orthogonal polynomials, regression problems with one, two, or three classifications in general form (the customary analysis of variance is a part of this subject), and simultaneous occurrence of classifications and of explicitly given regression functions (the so-called analysis of covariance belongs to this domain). Particular attention is paid to the general iterative method for two classifications (compare KUIPER [11]); this also leads to a new insight in balanced, and partially balanced incomplete block designs and in two-dimensional lattices. For balanced blocks we owe this insight partly to KUIPER. We emphasize the importance of the mentioned general iterative method, also because adaptation for electronic equipment may be foreseen in the near future. Characteristic properties of the concerned vector spaces are established for some particular cases of three classifications, namely in the designs of PEARCE (15, 16), and for the case that one classification is orthogonal to the interaction of the other two. For the case that fertility in a trial is supposed to be a continuous function of the coordinates of the plots, an objective method of estimation is presented. For regression problems of the kind of covariance analysis, new more general solutions are derived. The missing plot technique has been considered in its most general form. All the generalizations, enunciated in this chapter, have proved to be very suitable and useful in practice.

In chapter 4 a new and very general consideration of regression problems is given for cases that some parameters in the regression functions are random variables. We mention the so-called recovery of inter-block information, designs of split-plot type, and the estimation from series of experiments. All these apparently different problems could be treated – by means of vectors again – in a uniform way. The derived iterative methods, which are related to those of chapter three, are simplified to partly known results in particular cases. The estimation of the variance of the random parameters, necessary in the performance of the derived methods, is the only subject in this study for which some computational work seemed inevitable.

CHAPTER 1

LINEAR ALGEBRA

1.1. VECTOR SPACES

1.1.1. *Definition of vector spaces*

A vector space E is a set of elements with the following properties:

(a) To every pair x and y of elements in E there corresponds an element z in E, called sum of x and y and denoted by $z = x + y$, such that addition is commutative i.e. $x + y = y + x$; addition is associative i.e. $x + (y + z) = (x + y) + z$; there exists in E one vector, 0, called the null vector, such that for every x in E: $x + 0 = x$; to every x in E corresponds a vector $-x$ such that $x + (-x) = 0$.

(b) To every pair, consisting of a real number λ and a vector x in E, there corresponds an element y in E, called product of λ and x and denoted by λx, such that multiplication is distributive with respect to vector addition i.e. $\lambda (x + y) = \lambda x + \lambda y$; is distributive with respect to addition of real numbers i.e. $(\lambda + \mu)x = \lambda x + \mu x$; is associative i.e. $\lambda (\mu x) = (\lambda \mu)x$; $0x = 0$ and $1x = x$.

It follows from this definition that every linear combination $\lambda_1 x_1 + \lambda_2 x_2 + \ldots$ of vectors $x_1, \ldots, x_n$ in E is a well defined element of E.

1.1.2. *Examples*

Consider the set of arrows from one point, called origin, in a plane or in an ordinary three-dimensional space. Let addition take place by the well-known parallelogram construction in mechanics, and let the product of $\lambda > 0$ and the arrow x be an arrow with the same direction as x, but with length λ times as large as that of x. Let an arrow with length zero be called 0, and an arrow with the same length as the arrow x, but in opposite direction, be called $-x$. Then the arrows satisfy the definition of vector space.

Another example of a vector space will be obtained by considering *n*-tuples of real numbers with the following properties. Let the elements x and y be $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ respectively. Further

$$\mathrm{x}+\mathrm{y} = (x_1+y_1, x_2+y_2, \ldots, x_n+y_n);\ \lambda\mathrm{x} = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n);$$
$$0 = (0, 0, \ldots, 0) \text{ and } -\mathrm{x} = (-x_1, -x_2, \ldots, -x_n).$$

One can verify that the set of all such n-tuples satisfies the definition of vector space, and therefore may be called a vector space.

1.1.3. *Dependence and independence*

A set of vectors $x_1, \ldots, x_n$ is called linearly dependent, if there is a set of numbers $\lambda_1, \ldots, \lambda_n$ not all equal zero such that $\lambda_1\mathrm{x}_1+\lambda_2\mathrm{x}_2+\ldots+\lambda_n\mathrm{x}_n = 0$. If on the other hand $\Sigma_i\lambda_i\mathrm{x}_i = 0$ implies that $\lambda_i = 0$ for $i = 1, \ldots, n$, the set is called linearly independent.

We remark that every set of vectors containing the null vector is dependent. If $\lambda_k \neq 0$ in $\Sigma_i^n\lambda_i\mathrm{x}_i = 0$, then $\mathrm{x}_k = \left(\frac{-\lambda_1}{\lambda_k}\right)\mathrm{x}_1+\left(\frac{-\lambda_2}{\lambda_k}\right)\mathrm{x}_2+\ldots+\left(\frac{-\lambda_n}{\lambda_k}\right)\mathrm{x}_n$.

Therefore a set of vectors is dependent, if and only if some vector of this set can be written as a linear combination of the others.

In the example of a vector space consisting of arrows in a three-dimensional space, three arrows are dependent if they belong to the same plane, and independent if they do not.

The vectors $\mathrm{a} = (1, 0, 0)$; $\mathrm{b} = (0, 1, 0)$; $\mathrm{c} = (0, 0, 1)$ and $\mathrm{d} = (1, 1, 1)$ are dependent because $\mathrm{a}+\mathrm{b}+\mathrm{c}-\mathrm{d} = 0$. Each can be written as a linear combination of the other three. Any three of them are independent.

1.1.4. *Basis*

A basis in a vector space E is a set of independent vectors in E such that every vector in E is a linear combination of this set.

In the following we shall confine ourselves to vector spaces which have a finite number of vectors in a basis.

Any two arrows with different directions in a plane form a basis for the space of arrows in that plane, and any three arrows not lying in a plane do for the space of arrows in a three-dimensional space.

Any n-tuple $(\lambda_1, \ldots, \lambda_n)$ can be written as $\lambda_1(1, 0, 0, \ldots, 0)+\lambda_2(0, 1, 0, \ldots, 0)+ +\ldots+\lambda_n(0, 0, \ldots, 0, 1) = \lambda_1\mathrm{e}_1+\ldots+\lambda_n\mathrm{e}_n$. The n vectors $\mathrm{e}_1, \ldots, \mathrm{e}_n$ are independent and thus form a basis of the vector space of n-tuples. As this basis will be used frequently, we call it the standard basis of this space.

We remark that the expression of an element x of E as a linear combination of a basis $\mathrm{x}_1, \ldots, \mathrm{x}_n$ is unique. For from $\mathrm{x} = \Sigma_i\lambda_i\mathrm{x}_i = \Sigma_i\mu_i\mathrm{x}_i$ it follows by subtraction that $\Sigma_i(\lambda_i-\mu_i)\,\mathrm{x}_i = 0$. Because of the independence of the basis, $\lambda_i = \mu_i$ for $i = 1, \ldots, n$. The numbers $\lambda_1, \ldots, \lambda_n$ are called coordinates of x (with respect to this basis).

1.1.5. *Dimension*

Theorem: Every basis of a vector space E contains the same number of elements, which number is called the dimension of E.

Proof: (Halmos [7]). Let $\mathrm{x}_1, \ldots, \mathrm{x}_m$ be a set of generators in E i.e. a set of vectors such that every vector in E is a linear combination of the vectors in that set. Let $\mathrm{y}_1, \ldots, \mathrm{y}_n$ be a set of independent vectors in E. In front of the vectors $\mathrm{x}_1, \ldots, \mathrm{x}_m$ we write y_n and obtain: $\mathrm{y}_n, \mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_m$. This set is dependent because y_n is a linear combination of $\mathrm{x}_1, \ldots, \mathrm{x}_m$. Going from the left to the right, we cancel the first vector which is a linear combination of the preceding vectors e.g. x_i. All vectors of E are linear combinations of the set $y_n, x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m$. In front of this set we write y_{n-1}, and, in a similar way as before, a vector x_k will be cancelled. We proceed in the same way, till the set $y_1, \ldots, y_n$ will be exhausted. We remark that, because of the independence of the $y_1, \ldots, y_n$, only vectors x_i will be cancelled by this procedure. One could ask oneself whether the x_i will not be exhausted before the set $y_1, \ldots, y_n$.

Then the remaining vectors of $y_1, \ldots, y_n$ would be linear combinations of the y_i already used, which is impossible in connection with the independence of the set $y_1, \ldots, y_n$. It follows that $m \geq n$.

If both sets are bases of E, their rôle in the argument above can be changed, and it follows that $n \geq m$. Thus $m = n$.

The definition of dimension yields the customary dimension for the geometrical vector spaces of arrows. The space of all *n*-tuples is *n*-dimensional.

1.1.6. *Subspaces*

A subset A, B, etc. of a vector space E is called a subspace of E, if it is a vector space with the same definition of addition and multiplication as in E.

An example of a subspace in the three-dimensional space of arrows is the set of all arrows in a plane which contains the origin. Another example is the subset, in the vector space of all *n*-tuples of numbers, of all those *n*-tuples of which the last m ($\leq n$) numbers are zero.

The set of all linear combinations of a set of vectors in E is a subspace too; for every linear combination of such linear combinations can be written as a linear combination of the first set of vectors, in other words, is an element of the subset. This property implies that the definition of vector space for E also holds for this subset. The set of generating vectors is said to span the subspace. Particularly a basis for E spans E. Similarly, the set of all linear combinations of vectors in the two subspaces A and B is said to be spanned by A and B; every vector in this subspace, denoted by A + B and called the join of A and B, can be considered as the sum of a vector in A and a vector in B.

A basis, and thus the dimension of a subspace A, of a *n*-dimensional space is found in the following way. If A consists of the null vector only, the dimension of A is *0*. If A contains a vector $x_1 \neq 0$, then x_1 spans a subspace A_1 in A. If $A = A_1$, then A has dimension one. If A contains a vector x_2 which is not in A_1, then x_1 and x_2 span a subspace A_2 in A etc. We can find at most n vectors x_i which span A and are independent, for every set of $n + 1$ vectors is dependent. The dimension of a subspace is thus at most n. Proceeding in the sketched way, one can choose a basis for E such that a part of this set forms a basis for A.

The intersection of two subspaces A and B i.e. the set of vectors which are both in A and in B, is a subspace; for if two vectors are in this intersection, in other words, are in both subspaces, then the same holds for every linear combination of this pair.

If two subspaces have the null vector only in common, they are called disjoint.

The decomposition of a vector z in the join A + B of two disjoint subspaces A and B to a sum $x + y$, with x in A and y in B, is unique. For from $z = x + y = x' + y'$ follows $x - x' = y - y'$ with $x - x'$ in A and $y - y'$ in B. Because A and B are disjoint $x - x' = y - y' = 0$ or $x = x'$ and $y = y'$. Let A be spanned by a basis $x_1, \ldots, x_n$ and B by a basis $y_1, \ldots, y_m$. Then these $m+n$ vectors span A+B. Because of the uniqueness of the decomposition $z = x+y$, it follows from $\lambda_1 x_1 + \ldots + \lambda_n x_n + \mu_1 y_1 + \ldots + \mu_m y_m = 0$ that $\Sigma_i \lambda_i x_i = 0$ and $\Sigma_j \mu_j y_j = 0$, so that λ_i and μ_i are all zero, and the $m+n$ vectors independent. The dimension of A+B is thus $m+n$.

If E has dimension n and the subspace A has dimension m, we can choose a subspace B with dimension $n-m$ such that A and B span E and are disjoint. For we can choose a basis for E such that m vectors of that set are in A. The subspace, spanned by the $n-m$ remaining vectors, is called B. From the definition of a basis follows that A and B are disjoint. We point out that such a residual space B can be chosen in various ways.

1.2. LINEAR TRANSFORMATIONS AND MATRICES

1.2.1. *Definition of linear transformations*

A linear transformation A in a vector space E assigns to every vector x in E a vector Ax in E (is a mapping of E into itself) such that, for x and y in E and λ and μ real numbers, $A(\lambda x + \mu y) = \lambda Ax + \mu Ay$. In particular $A0 = 0$. If, for all x, $Ax = 0$ or $Ax = x$, then A is represented by 0 or 1 respectively.

The sum $A+B$ of two linear transformations A and B is defined by: $(A+B)x = Ax + Bx$ for all x. The product of a real number λ and a linear transformation A is defined by: $(\lambda A)x = \lambda(Ax)$ for all x. The product $P = AB$ of two transformations A and B is defined by $Px = A(Bx)$ for all x. We remark that the order in this product is important. The given order means that x should be mapped by B, and that the result should be mapped by A. The product BA on the other hand means that x should be mapped by A, and that the result Ax should be mapped by B. By considering e.g. rotations in the three-dimensional space of arrows about the origin it will be seen that the result of AB in general is different from that of BA: linear transformations are in general not commutative.

From the definition we have the following rules concerning the calculus of transformations:

$$A0 = 0A = 0;\ A(B+C) = AB + AC;\ (A+B)C = AC + BC;$$
$$A1 = 1A = A;\ A(BC) = (AB)C.$$

As a consequence of the last property, not only AA may be written A^2, but also $AAA\ldots A$, which consists of m factors A, may be written A^m.

1.2.2. *Inverse of a transformation*

Theorem: The equation $Ax = y$ has a solution x for any y, if and only if $Ax = 0$ implies $x = 0$.

Proof: Suppose $Ax = 0$ implies $x = 0$. Let $x_1, \ldots, x_n$ be a basis for E. Then $Ax_1, \ldots, Ax_n$ is also a basis. For from $\Sigma_i \lambda_i A x_i = 0$ follows $A(\Sigma_i \lambda_i x_i) = 0$, so that $\Sigma_i \lambda_i x_i = 0$ and thus every λ_i is zero. The supposition implies that any y can be written as $\Sigma_i \lambda_i (Ax_i) = A(\Sigma_i \lambda_i x_i)$, so that $\Sigma_i \lambda_i x_i$ is a solution of $Ax = y$.

Conversely, if $Ax = y$ has a solution for any y, then, corresponding to the vectors of a basis $y_1, \ldots, y_n$, there can be found vectors x_i such that $Ax_i = y_i$. From $\Sigma_i \lambda_i x_i = 0$ follows $A \Sigma_i \lambda_i x_i = 0$, or $\Sigma_i \lambda_i A x_i = 0$ or $\Sigma_i \lambda_i y_i = 0$. Thus the vectors x_i form a basis, and any x can be written as $\Sigma_i \lambda_i x_i$. Hence $Ax = 0$ implies $A(\Sigma_i \lambda_i x_i) = 0$ or $\Sigma_i \lambda_i A x_i = 0$ or $\Sigma_i \lambda_i y_i = 0$; thus every λ_i is zero and $x = 0$.

The solution, if it exists, is unique. For if $x' \neq x$ or $x' - x \neq 0$, then $A(x' - x) \neq 0$ or $Ax \neq Ax'$. The assignment of a vector x to every y, obtained in this way, is a linear transformation (denoted by A^{-1}). Let $Ax_1 = y_1$ and $Ax_2 = y_2$; then $A(\lambda x_1 + \mu x_2) = \lambda y_1 + \mu y_2$; thus $A^{-1}(\lambda y_1 + \mu y_2) = \lambda x_1 + \mu x_2 = \lambda A^{-1} y_1 + \mu A^{-1} y_2$.

The linear transformation A^{-1} is called the inverse of A. If A^{-1} exists, A is called non-singular; if the inverse of A does not exist, A is singular. If A has an inverse, $AA^{-1} = A^{-1}A = 1$.

If $AB = 1$, A is the inverse of B (and thus B is the inverse of A) which may be seen as follows: $ABx = x$ for any x; therefore $x \neq 0$ implies $Bx \neq 0$, or, in other words, $Bx = 0$ implies $x = 0$; thus B has an inverse B^{-1}; multiplying both sides of $AB = 1$ on the right by B^{-1} yields $A = B^{-1}$.

1.2.3. *Matrices*

Let $x_1, \ldots, x_n$ be a basis for E. Let the linear transformation A be such that $Ax_j = \Sigma_i \alpha_{ij} x_i$, and let $x = \Sigma_j \lambda_j x_j$ be a vector in E. Then $Ax = \Sigma_j \lambda_j Ax_j = \Sigma_j \lambda_j (\Sigma_i \alpha_{ij} x_i) = \Sigma_i (\Sigma_j \alpha_{ij} \lambda_j)\, x_i$. The coordinates $\Sigma_j \alpha_{ij} \lambda_j$ of Ax are completely determined by a square array of numbers α_{ij}:

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1j} & \ldots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2j} & \ldots & \alpha_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ \alpha_{i1} & \alpha_{i2} & \ldots & \alpha_{ij} & \ldots & \alpha_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ \alpha_{n1} & \alpha_{n2} & \ldots & \alpha_{nj} & \ldots & \alpha_{nn} \end{bmatrix}$$

which is called the matrix of A in this coordinate system. The element α_{ij} is the i-th coordinate of Ax_j with respect to the basis $x_1, \ldots, x_n$.

It follows directly that the matrix of the transformation 0 consists of zeros only, and that in the matrix of the transformation 1 $\alpha_{ij} = 1$ for $i = j$, and $\alpha_{ij} = 0$ for $i \neq j$. If the transformations A and B have matrices with elements α_{ij} and β_{ij} respectively in a fixed coordinate system, then the elements of the matrix of the transformation $\lambda A + \mu B$ are $\gamma_{ij} = \lambda\alpha_{ij} + \mu\beta_{ij}$. In order to obtain the matrix of the product $C = AB$, we consider $Cx_j = A(Bx_j) = A(\Sigma_k \beta_{kj} x_k) = \Sigma_k \beta_{kj} Ax_k = \Sigma_k \beta_{kj} (\Sigma_i \alpha_{ik} x_i) = \Sigma_i (\Sigma_k \alpha_{ik} \beta_{kj}) x_i$. It follows that the element γ_{ij} of C is equal to $\Sigma_k \alpha_{ik} \beta_{kj}$, which is obtained from the i-th row of A and the j-th column of B.

Singularity of A i.e. $Ax = 0$ for some $x \neq 0$ implies $\Sigma_j \alpha_{ij} \lambda_j = 0$ for all i. If we consider the columns of the matrix of A as vectors in the space of n-tuples, this means that these columns are dependent. In order to obtain a necessary and sufficient condition of independence for the columns $a_1, \ldots, a_n$ of the matrix we consider the function

1.2.4. *Determinant*

The determinant is a real function of the columns $a_1, \ldots, a_n$ of a square matrix, $D(a_1, \ldots, a_n)$ such that:

$$D(a_1, \ldots, \lambda a_k, \ldots, a_n) = \lambda D(a_1, \ldots, a_k, \ldots, a_n);$$
$$D(a_1, \ldots, a_i + \lambda a_j, \ldots, a_n) = D(a_1, \ldots, a_i, \ldots, a_n);$$

and if we represent the columns of the matrix of the transformation 1 by $e_1, \ldots, e_n$: $D(e_1, \ldots, e_n) = 1$.

From the first condition it follows, by choosing $\lambda = 0$, that if $a_k = 0$ then $D = 0$. If the columns are dependent e.g. $a_1 = \Sigma_{j=2}^{n} \lambda_j a_j$, then $D(a_1, \ldots, a_n) = D(a_1 - \lambda_2 a_2, a_2, \ldots, a_n) = D(a_1 - \lambda_2 a_2 - \lambda_3 a_3, a_2, a_3, \ldots, a_n) = \ldots\ldots = D(0, a_2, a_3, \ldots, a_n) = 0$. In other words, singularity implies $D = 0$.

From $D(a_1, \ldots, a_n) = D(a_1 + a_2, a_2, a_3, \ldots, a_n) = D(a_1 + a_2, a_2 - a_1 - a_2, a_3, \ldots, a_n) = D(a_1 + a_2, -a_1, a_3, \ldots, a_n) = D(a_2, -a_1, a_3, \ldots, a_n) = -D(a_2, a_1, a_3, \ldots, a_n)$ it follows that exchanging two columns changes sign of D.

Further we will prove:

$$D(b_1 + b_2, a_2, \ldots, a_n) = D(b_1, a_2, \ldots, a_n) + D(b_2, a_2, \ldots, a_n).$$

If $a_2, \ldots, a_n$ are dependent the statement is trivial. If they are independent, the linear relation between $b_1, b_2, a_2, \ldots, a_n$ (for these are dependent) has not both coefficients of b_1 and b_2 equal zero, so that e.g. $b_2 = \lambda_1 b_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n$. Then: $D(b_2 + \lambda b_1, a_2, \ldots, a_n) = D(\lambda_1 b_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n + \lambda b_1, a_2, \ldots, a_n) = D(\lambda_1 b_1 + \lambda b_1, a_2, \ldots, a_n) = (\lambda_1 + \lambda) D(b_1, a_2, \ldots, a_n)$. Substituting $\lambda = 1$ yields: $D(b_1 + b_2, a_2, \ldots, a_n) = (\lambda_1 + 1)\, D(b_1, a_2, \ldots, a_n)$, while $\lambda = 0$ gives: $D(b_2, a_2, \ldots, a_n) = \lambda_1 D(b_1, a_2, \ldots, a_n)$. By subtraction the assertion follows.

When the columns $b_1, \ldots, b_n$ are dependent on $a_1, \ldots, a_n$ then, by repeated application of the foregoing property, $D(b_1, \ldots, b_n)$ can be reduced to a linear combination of determinants of matrices consisting of n columns of $a_1, \ldots, a_n$ (equality of such columns will occur). If $D(a_1, \ldots, a_n) = 0$ then all such determinants are zero and thus $D(b_1, \ldots, b_n) = 0$. If the set $a_1, \ldots, a_n$ is independent, then we may take the set $e_1, \ldots, e_n$ for $b_1, \ldots, b_n$. If $D(a_1, \ldots, a_n)$ were zero then also $D(e_1, \ldots, e_n)$ would be zero. While the last is not true, $D(a_1, \ldots, a_n)$ is not zero. We have: $a_1, \ldots, a_n$ *are dependent if and only if the determinant is zero.*

According to the method described in the last paragraph the determinant of the matrix of A can be reduced to a linear combination of determinants $D(e_{j_1}, e_{j_2}, \ldots, e_{j_n})$ with coefficients $\alpha_{j_1 1}, \alpha_{j_2 2}, \ldots, \alpha_{j_n n}$.

In connection with an application to follow, we observe that a function $L(a_1, \ldots, a_n)$, satisfying only the first two defining properties of determinants, is a linear combination of $L(e_{j_1}, \ldots, e_{j_n})$ with the same coefficients as in $D(a_1, \ldots, a_n)$. Hence for any such function there exists a constant C such that $L(a_1, \ldots, a_n) = C \cdot D(a_1, \ldots, a_n)$.

The application concerns the determinant of the product AB of two matrices. We observe: multiplication of a column of the matrix of B by λ implies multiplication of the corresponding column in the matrix of AB by λ, and thus multiplication of the determinant by λ; adding λ times a column of the matrix of B to another column implies a similar addition of corresponding columns in the matrix of AB, and therefore does not change the determinant; if $B = 1$, then $AB = A$, and the considered determinant is equal to the determinant of the matrix of A. Hence $D(AB) = L(B)$ with $C = D(A)$, so that $D(AB) = D(A) \cdot D(B)$.

In particular, if A has an inverse the product of the corresponding determinants is 1.

1.2.5. *Change of basis*

Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be two bases for E, and let the linear transformation A, which maps x_i onto y_i for every i, be given by $Ax_j = y_j = \Sigma_i \alpha_{ij} x_i$.

Let $\mathrm{z} = \Sigma_i \lambda_i \mathrm{x}_i = \Sigma_j \mu_j \mathrm{y}_j$. The relation between the coordinates λ_i and μ_j follows from: $\mathrm{z} = \Sigma_j \mu_j \mathrm{y}_j = \Sigma_j \mu_j A \mathrm{x}_j = \Sigma_j \mu_j (\Sigma_i \alpha_{ij} \mathrm{x}_i) = \Sigma_i (\Sigma_j \alpha_{ij} \mu_j) \mathrm{x}_i$, so that $\lambda_i = \Sigma_j \alpha_{ij} \mu_j$.

A is non singular, because $A\mathrm{z} = A\Sigma_i \lambda_i \mathrm{x}_i = \Sigma_i \lambda_i A \mathrm{x}_i = \Sigma_i \lambda_i \mathrm{y}_i = 0$ only if all $\lambda_i = 0$ i.e. if $\mathrm{z} = 0$. The transformation A^{-1} will be expressed by $\mathrm{x}_k = \Sigma_r \alpha^{-1}_{rk} \mathrm{y}_r$ in which α^{-1}_{rk} is an element of the matrix of A^{-1} with respect to the second basis. The product of the matrix of *A* with respect to the first basis and of A^{-1} with respect to the second basis is *1*; these matrices are called the inverse of each other.

Let *B* be a transformation of which the matrix with respect to the first basis consists of elements β_{ij} so that $B\mathrm{x}_j = \Sigma_i \beta_{ij} \mathrm{x}_i$. In order to obtain the matrix of *B* with respect to the second basis, we consider $B\mathrm{y}_j = B(A\mathrm{x}_j) = B\Sigma_i \alpha_{ij} \mathrm{x}_i = \Sigma_i \alpha_{ij} B \mathrm{x}_i = \Sigma_i \Sigma_k \alpha_{ij} \beta_{ki} \mathrm{x}_k = \Sigma_i \Sigma_k \Sigma_r \alpha_{ij} \beta_{ki} \alpha^{-1}_{rk} \mathrm{y}_r = \Sigma_r \{\Sigma_k \alpha^{-1}_{rk} (\Sigma_i \beta_{ki} \alpha_{ij})\} \mathrm{y}_r$. The required matrix is obtained by multiplying the given matrix of *B* on the right by the considered matrix of *A*, and on the left by the inverse of that matrix. It follows that the determinant of this product of matrices is equal to the determinant of the given matrix of *B*. Hence the determinant belonging to a transformation is independent of the basis.

1.2.6. *Proper vectors and proper values*

A vector $\mathrm{x} \neq 0$ is a proper vector and a number λ a proper value of the transformation *B* if $B\mathrm{x} = \lambda \mathrm{x}$. The proper values of *B* are those λ for which $(B - \lambda 1)\mathrm{x} = 0$ has a solution $\mathrm{x} \neq 0$, i.e. those λ for which the determinant of the transformation $B - \lambda 1$, $D(B - \lambda 1)$, is zero.

Instead of the matrix $B - \lambda 1$ of this transformation we may also use the matrix $A^{-1}(B - \lambda 1)A = A^{-1}BA - A^{-1}\lambda 1 A = A^{-1}BA - \lambda 1$ from which it follows that the proper values are independent of the basis. The function $D(B - \lambda 1)$ is a polynomial of degree *n* in λ with $(-1)^n$ as coefficient of λ^n. The number of proper values therefore is at most *n*. The multiplicity of a root λ of the equation $D(B - \lambda 1) = 0$ is called the algebraic multiplicity of that proper value.

The proper vectors that belong to a proper value form, together with the null vector, a subspace of E. The dimension of this subspace is called the geometric multiplicity of that proper value.

1.2.7. *Projection*

If E is spanned by the disjoint subspaces A and B, then there is a unique decomposition of z in E to $\mathrm{x} + \mathrm{y}$ with x in A and y in B. The transformation *P* which maps z onto $P\mathrm{z} = \mathrm{x}$ is called the projection on A along B. This transformation is linear; for, with x_i and y_i in A and B respectively, $P(\lambda \mathrm{z}_1 + \mu \mathrm{z}_2) = P(\lambda \mathrm{x}_1 + \lambda \mathrm{y}_1 + \mu \mathrm{x}_2 + \mu \mathrm{y}_2) = \lambda \mathrm{x}_1 + \mu \mathrm{x}_2 = \lambda P \mathrm{z}_1 + \mu P \mathrm{z}_2$. The transformation is singular in general: $P(\mathrm{x}_1 + \mathrm{y}_1) = P(\mathrm{x}_1 + \mathrm{y}_2)$ also for $\mathrm{y}_1 \neq \mathrm{y}_2$. Because x will be decomposed in $\mathrm{x} + 0$, we have $P\mathrm{x} = \mathrm{x}$ and thus $P^2\mathrm{z} = P(P\mathrm{z}) = P\mathrm{x} = \mathrm{x} = P\mathrm{z}$; in other words, $P = P^2$ for every z. Such a transformation is called idempotent.

If the first *k* vectors of a basis for E span the subspace A and the remaining vectors span B, then the projection *P* on A along B is expressed by a matrix with elements α_{ij} which are all zero except $\alpha_{jj} = 1$ for $j \leqq k$.

1.3. INNER PRODUCT; ORTHOGONALITY

1.3.1. *Definition of inner product*

Introduction of an inner product will enable us to introduce the notions length, distance and angle in vector spaces.

The inner product is a function which assigns to every pair of vectors x and y in a vector space E a real number, denoted by (x, y) or by xy, such that $(x, y) = (y, x)$; $(\lambda x_1 + \mu x_2, y) = \lambda(x_1, y) + \mu(x_2, y)$; $(x, x) \geqq 0$ and is zero only if $x = 0$.

(x, x) will also be denoted by x^2. The number $\sqrt{(x, x)}$ also denoted by $|x|$ is called the length of x. From the definition we have: $|\lambda x| = \sqrt{(\lambda x, \lambda x)} = \sqrt{\lambda(x, \lambda x)} = \sqrt{\lambda(\lambda x, x)} = \sqrt{\lambda^2(x, x)} = |\lambda| \cdot |x|$. Multiplication of a vector by the number λ makes the vector |λ| times longer. If x (or y) is 0, then $(x, y) = (0x, y) = 0(x, y) = 0$.

Let $x_1, \ldots, x_n$ be a basis for E, $\lambda_1, \ldots, \lambda_n$ the coordinates of x and $\mu_1, \ldots, \mu_n$ those of y. Then (x, y) can be reduced to $\Sigma_{ij}\lambda_i\mu_j(x_i, x_j)$. This form (which is a positive definite form in the variables λ_i if x = y) will be determined by the inner products of the basisvectors (x_i, x_j). Let the inner products (x_i, x_j) be the elements α_{ij} of a matrix which because of $\alpha_{ij} = \alpha_{ji}$ is called a symmetric matrix. Then this matrix together with the basis determines an inner product in E. The matrix itself will be named the metric (associated with this basis).

1.3.2. *Distance and angles*

Theorem: $(x, y)^2 \leqq x^2 \cdot y^2$.

Proof: For any λ: $(y - \lambda x)^2 = (y - \lambda x, y - \lambda x) \geqq 0$ or $y^2 - 2\lambda xy + \lambda^2 x^2 \geqq 0$. The quadratic form in λ thus has a discriminant $\leqq 0$, so that $(x, y)^2 \leqq x^2 \cdot y^2$. Equality occurs if and only if $y - \lambda x = 0$ i.e. if x and y are dependent.

The relation just proved (Schwarz's inequality) can also be expressed by: the number $\frac{(x, y)}{|x| \cdot |y|}$ is in absolute value at most *1*.

The distance of a pair of vectors x and y is defined as the length of the difference x–y. The name is justified by the following properties: $|y - x| = |x - y|$ i.e. the distance from x to y is as large as the distance from y to x; the distance is zero, if and only if x = y; otherwise it is positive; because $|(y+z) - (x+z)| = |y - x|$, the distance is said to be invariant under translations; the triangle inequality holds i.e. $|x - y| \leqq |x - z| + |z - y|$. The last property may be seen as follows. We have $|x+y|^2 = x^2 + 2(x, y) + y^2 \leqq |x|^2 + 2|x| \cdot |y| + |y|^2 = \{|x| + |y|\}^2$ or $|x+y| \leqq |x| + |y|$. Replacing x and y by x − z and z − y respectively gives the inequality.

The number $\frac{(x, y)}{|x| \cdot |y|}$ is named the cosine of the angle φ between x and y. This definition for the angle between vectors may be illustrated as follows: The sides of the triangle with lengths |x|, |y| and |x − y| respectively satisfy: $|x - y|^2 = (x - y)^2 = x^2 - 2xy + y^2 = |x|^2 + |y|^2 - 2\frac{x\,y}{|x| \cdot |y|} \cdot |x| \cdot |y|$. By substitution of the given definition for the angle φ between x and y we just obtain the ordinary rule for the cosine in a triangle. Remembering the addition of vectors in the space of arrows (of vectors y and x − y in this case) we observe that any pair of vectors x and y may be represented "congruently" in the space of arrows by two arrows with lengths $|x|$ and $|y|$ respectively and an angle φ between them. Similarly any set of three vectors can now be represented by arrows with the corresponding lengths and angles in an ordinary three-dimensional space.

1.3.3. *Orthogonality*

In accordance with the definition of angle two vectors are called orthogonal if their inner product is zero.

A set of vectors in E is called orthogonal if every vector of the set is orthogonal to every other vector of the set. If moreover every vector in the set has length *1*, the set is called orthonormal. (If a vector $x \neq 0$ is divided by its length $|x|$, a vector with length one, a so-called unit vector, is obtained).

An orthonormal set $x_1, \ldots, x_m$ is independent. For if $\Sigma_i\lambda_i x_i = 0$, then $(\Sigma_i\lambda_i x_i, x_j) = \Sigma_i\lambda_i(x_i, x_j) = \lambda_j = 0$ for all i. It follows that m is n at most.

We will prove that in a n-dimensional space there exists an orthonormal basis (of n vectors).

Proof: There exists a basis $x_1, \ldots, x_n$. This basis will be orthogonalized i.e. an orthonormal set $y_1, \ldots, y_n$ will be constructed such that every y_i is a linear combination of $x_1, \ldots, x_i$. This orthogonalization process will take place in n steps. In the first step $y_1 = x_1/|x|$. After completion of the r-th step so that $y_1, \ldots, y_r$ are available as linear combinations of $x_1, \ldots, x_r$, the $(r+1)$-th step is performed as follows. First we find a vector $z = x_{r+1} - (\lambda_1 y_1 + \lambda_2 y_2 + \ldots + \lambda_r y_r)$ orthogonal to $y_1, \ldots, y_r$. Because $(z, y_j) = (x_{r+1}, y_j) - \lambda_j$ for $j = 1, \ldots, r$, we must take $\lambda_j = (x_{r+1}, y_j)$. This z is not 0, for $x_1, \ldots, x_{r+1}$ are independent and the coefficient of x_{r+1} in z is not zero. Now $y_{r+1} = z/|z|$.

From the proof it follows that every x_r is a linear combination of $y_1, \ldots, y_r$.

In the space of arrows, where the inner product of two vectors is defined as the product of their lengths and the cosine of the angle between them, an orthonormal basis is formed by two (and three respectively) perpendicular arrows with length one. If the inner product in the space of n-tuples is defined by $(x, y) = \Sigma_i x_i y_i$ (as is often done), the standard basis $e_1, \ldots, e_n$ happens to be an orthonormal basis.

1.3.4. *Orthogonal polynomials*

An application of the orthogonalization process occurs, when the vector space of n-tuples (or a subspace of it) is considered as the set of n-tuples of function values for a function defined on a set of n real numbers. The function may for example be any real polynomial of degree $\leq m$ in one variable x. That the corresponding set of n-tuples of function values then is a vector space follows from the fact that, if such a n-tuple z_1 corresponds to the function $f_1(x)$ and z_2 to $f_2(x)$, then $\lambda_1 z_1 + \lambda_2 z_2$ contains the function values of the function $\lambda_1 f_1(x) + \lambda_2 f_2(x)$; if $f_1(x)$ and $f_2(x)$ are polynomials of degree $\leq m$, then the same is true for $\lambda_1 f_1(x) + \lambda_2 f_2(x)$. If the n values of x are different, a basis for the space of n-tuples is obtained from the functions $1, x, x^2, \ldots., x^{n-1}$. For, as $\lambda_0 + \lambda_1 x + \ldots + \lambda_{n-1} x^{n-1} = 0$ has at least n different roots, if and only if all coefficients are zero, the vector corresponding to $\lambda_0 + \lambda_1 x + \lambda_2 x^2 + \ldots + \lambda_{n-1} x^{n-1}$ is the null vector, if and only if every $\lambda_i = 0$, i.e. if the generating function is the polynomial *0*. Because the space of n-tuples has dimension n, it follows that a vector obtained from a polynomial of degree $\geqslant n$ also is generated by some polynomial of degree $\leq n - 1$.

For purpose of computation one often orthogonalizes the above-mentioned basis or if $m < n-1$ the first $m+1$ elements of this basis. This happens according to the process described in the foregoing section. Thus the j-th element of the orthonormal (or orthogonal) basis will be a linear combination of the first j elements of the first basis, which means that the corresponding function will be a polynomial of degree $j-1$. Such a set of functions are called orthogonal polynomials (associated with n values of x and a definition of inner product). Special cases of such functions and their tabulation will be considered in 3.1.4.

1.3.5. *Orthogonal subspaces and orthogonal projection*

Two subspaces A and B of E are called orthogonal if every vector of A is orthogonal to every vector in B. Orthogonal spaces are disjoint; for, if some z is in both, $z^2 = (z, z) = 0$ and thus $z = 0$. It is easily seen that A and B are orthogonal, if and only if a basis of one of them is orthogonal to the other space or to its basis. The set of vectors in E orthogonal to a subspace A clearly is a subspace too. It will be denoted by $A^\perp$. Let $x_1, \ldots, x_m$ be an orthonormal basis for A and let z be any vector in E. Then $x = \Sigma_{i=1}^m(z, x_i)x_i$ is in A, while $y = z - \Sigma_{i=1}^m(z, x_i)x_i$ is orthogonal to every x_i, thus to A, in other words y is in $A^\perp$. Thus z is decomposed in $x+y$ with x in A and y in $A^\perp$: A and $A^\perp$ span the n-dimensional space E. The $(n-m)$-dimensional residual space $A^\perp$ is called the orthogonal residual space of E, and is spanned by $n-m$ independent vectors orthogonal to A. It follows that $A^{\perp\perp}$ is the same as A.

The transformation which transforms z in x, the projection of z on A along $A^\perp$, appears to be determined by A only (as $A^\perp$ is determined by A). It is called the orthogonal projection on A. Similarly the assignment of $y = z - x$ to z is called the orthogonal projection on $A^\perp$. The orthogonal projection of z on A will be found as a linear combination, x, of the basis vectors of A such that $y = z - x$ is orthogonal to A i.e. to the basis vectors of A. It was this procedure which has been applied in the orthogonalization process (x_{r+1} is projected orthogonally on the space spanned by $y_1, \ldots, y_r$, while z is in the orthogonal residual space) and in the study of $A^\perp$ in fact. The vector y will often be called the perpendicular from z on A.

From $z = x+y$ and x orthogonal to y follows $z^2 = x^2+y^2$, the theorem of Pythagoras. Thus the length of x i.e. the length of the orthogonal projection of z is equal to the length of z at most. Equality holds only if $y = 0$ i.e. if z is in A.

Consider the distance of z to any vector u in A. Then we have $z - u = (z-x) + (x-u) = y+(x-u)$ and, because y is orthogonal both to x and to u, and thus to $x-u$: $(z-u)^2 = y^2+(x-u)^2$. It follows that the distance between z and u is at least as large as the length of y, the perpendicular from z on A. Equality holds only if $x = u$. In other words, x is the vector in A with the shortest distance to z, and the length of the perpendicular y is by definition the distance between z and A.

Let $A_1, \ldots, A_k$ be subspaces in E. Then the residual space of E, A_{k+1}, orthogonal to the space spanned by $A_1, \ldots, A_k$ is determined. If moreover $A_1, \ldots, A_k$ are orthogonal to each other, there is a unique decomposition of any vector z in E to $x_1 + \ldots + x_{k+1}$ with x_j in A_j for $j = 1, \ldots, k+1$. The uniqueness follows easily after choosing orthonormal bases in every A_j. Now every x_j is equal to the orthogonal projection of z on A_j, $P_j z$; for every $P_j z$ is a vector in A_j such that $z - P_j z$ is in the residual space of E orthogonal to A_j, in other words, is in the space spanned by the remaining subspaces.

Clearly z^2 is equal to $x_1^2 + x_2^2 + \ldots + x_{k+1}^2$ which is an extension of the Pythagorean theorem.

The sum of some $P_j z$ is also an orthogonal projection namely on the space spanned by the corresponding A_j, for the difference between z and this sum is orthogonal to that space.

With a slight change of wording we have: The difference between the orthogonal projection on the space spanned by A_1 and A_2 and the orthogonal projection on the space A_1, which is a subspace of the space $A_1 + A_2 = B$, is equal to the orthogonal projection on A_2, the residual space of B orthogonal to A_1.

If the considered orthogonal subspaces $A_1, \ldots$ are one-dimensional, in other words, form an orthogonal basis, then the corresponding orthogonal projections of z yield the coordinates of z in an orthonormal coordinate system. For the orthogonal projection of z on e.g. the space spanned by x is obtained as λx such that $z - \lambda x$ is orthogonal to x, or $(z, x) - \lambda(x, x) = 0$ or $\lambda = \frac{zx}{x^2}$. Thus the projection is $\frac{(z, x)}{(x, x)} x$, or (substituting $x' = x/|x|$) $(z, x')\, x'$. The coordinate of x' is equal to $(z, x') = \frac{(z, x)}{|x|}$.

1.4. CONVERGENCE OF VECTORS AND LINEAR TRANSFORMATIONS

1.4.1. *Convergence of vectors*

A sequence of vectors a_n ($n = 1, 2, \ldots$) in E is said to converge to a vector a in E if the corresponding sequence of distances $|a_n - a|$ converges to zero.

Let $x_1, \ldots, x_k$ be an orthonormal basis for E; let the sequence of vectors $a_n = \Sigma_{i=1}^{k} \alpha_{in} x_i$ and the vector $a = \Sigma_{i=1}^{k} \alpha_i x_i$ be such that α_{in} converges to α_i for every *i*. Then because of the triangle inequality:

$$|a_n - a| = |\Sigma_{i=1}^{k} (\alpha_{in} - \alpha_i) x_i| \leqq \Sigma_{i=1}^{k} |(\alpha_{in} - \alpha_i) x_i| = \Sigma_{i=1}^{k} |\alpha_{in} - \alpha_i|.$$

For any $\varepsilon > 0$ there exists N such that, for $n > N$, $|\alpha_{in} - \alpha_i| < \frac{\varepsilon}{k}$ for every *i*, and thus $|a_n - a| < \varepsilon$. In other words, a_n converges to a.

The analogon of Cauchy's characteristic of convergence for sequences of real numbers is also valid for sequences of vectors: if there is a sequence of vectors a_n such that for any $\varepsilon > 0$ there exists N so that, for all $n > m > N$, $|a_n - a_m| < \varepsilon$, then there exists a vector a to which a_n converges. To prove this write a_n and a_m like in the preceding paragraph. Then $a_n - a_m = \Sigma_{i=1}^{k} (\alpha_{in} - \alpha_{im}) x_i$. If $|a_n - a_m| < \varepsilon$ then also $|\alpha_{in} - \alpha_{im}|$, the length of the orthogonal projection of $a_n - a_m$ on x_i, is smaller than ε for every *i*. According to Cauchy's characteristic for real numbers the sequence of real numbers α_{in} has a limit α_i for every *i*. Thus a_n converges to $\Sigma_{i=1}^{k} \alpha_i x_i$.

1.4.2. *Bound of a linear transformation*

A linear transformation A is said to be bounded, if there exists a positive number K such that for every vector x in E: $|Ax| \leqq K|x|$, or (which is an equivalent relation) such that the length of the mapping of a vector with length 1 has an upper bound, K. The minimum of these upper bounds K is called the bound of the transformation A and will be denoted by $G(A)$. If A is an orthogonal projection e.g., the bound of A is equal to 1.

Theorem: Every linear transformation A is bounded.

Proof: Let $\mathrm{x}_1, \ldots, \mathrm{x}_k$ be an orthonormal basis for E; let M be the maximum of the set $|A\mathrm{x}_1|, \ldots, |A\mathrm{x}_k|$. Any vector x with length one is equal to $\Sigma_{i=1}^{k}(\mathrm{x}, \mathrm{x}_i)\mathrm{x}_i$ with $|(\mathrm{x}, \mathrm{x}_i)| \leqslant 1$. Applying the triangle inequality we have $|A\mathrm{x}| = |A\,\Sigma_{i=1}^{k}(\mathrm{x}, \mathrm{x}_i)\mathrm{x}_i| = |\Sigma_{i=1}^{k}(\mathrm{x}, \mathrm{x}_i)A\mathrm{x}_i| \leqq \Sigma_{i=1}^{k}|(\mathrm{x}, \mathrm{x}_i)A\mathrm{x}_i| = = \Sigma_{i=1}^{k}|(\mathrm{x}, \mathrm{x}_i)| \cdot |A\mathrm{x}_i| \leqq \Sigma_{i=1}^{k}|A\mathrm{x}_i| \leqq kM$. Thus kM is an upper bound as meant in the definition for a bounded transformation and A is bounded.

Theorem: Let the bound of A be $G(A)$. Then there exists a vector $\mathrm{x} \neq 0$ such that $|A\mathrm{x}| = G(A) \cdot |\mathrm{x}|$.

Proof: From the condition and the definition of a bound it follows that for any n ($n = 1, 2, \ldots$) there exists at least one vector x_n with length one such that $\left(1 - \frac{1}{n}\right)G(A) \leqq |A\mathrm{x}_n| \leqq G(A) \cdot |\mathrm{x}_n| = G(A)$. Let the coordinates of the vector x_n with respect to an orthonormal basis be $\alpha_n, \beta_n, \ldots$. The first coordinates α_n form an infinite sequence of real numbers in the interval from -1 to $+1$; this sequence has thus at least one limit point say α. It is possible to construct an infinite subsequence from the sequence α_n which converges to α, e.g. by choosing successively: α_{r_1} such that $|\alpha_{r_1} - \alpha| \leqq 1$; α_{r_2} with $r_2 > r_1$ such that $|\alpha_{r_2} - \alpha| \leqq \frac{1}{2}$; α_{r_3} with $r_3 > r_2$ such that $|\alpha_{r_3} - \alpha| \leqq \frac{1}{3}$; and so on.

Consider the corresponding subsequence of vectors $\mathrm{x}_{r_1}, \mathrm{x}_{r_2}, \ldots$. In this subsequence the sequence of second coordinates $\beta_{r_1}, \beta_{r_2}, \ldots$ may be treated analogously. In this way an infinite sub-subsequence of vectors can be constructed such that the sequence of their second coordinates converges to a limit β. The first coordinates will continue to have the limit α.

By continuation of this process one can obtain a subsequence of the vectors x_n such that the corresponding sequences of coordinates converge to $\alpha, \beta, \ldots$ respectively. According to the foregoing section this sequence of vectors converges to a vector x. Because the length of a vector is a continuous function of the (orthonormal) coordinates of the vector, it follows that the length of x is one. Further in the constructed subsequence, converging to x, we have that, for any $\varepsilon > 0$, there exists N such that for $n > N$: $G(A) \geqq |A\mathrm{x}| = |A\mathrm{x}_n - A(\mathrm{x}_n - \mathrm{x})| \geqq$ $\geqq |A\mathrm{x}_n| - |A(\mathrm{x}_n - \mathrm{x})| \geqq \left(1 - \frac{1}{n}\right) G(A) - \varepsilon \cdot G(A) \geqq (1 - 2\varepsilon)G(A)$, i.e. $G(A) \geqq |A\mathrm{x}| \geqq (1 - 2\varepsilon)G(A)$ for any $\varepsilon > 0$. Consequently $|A\mathrm{x}| = G(A)$, and the proof is complete.

For all x with length one we have $|(A+B)\mathrm{x}| \leqq |A\mathrm{x}| + |B\mathrm{x}| \leqq G(A) + G(B)$. Then the bound of $(A+B)$ obeys $G(A+B) \leqq G(A) + G(B)$.

Further $|AB\mathrm{x}| \leqq G(A) \cdot |B\mathrm{x}| \leqq G(A) \cdot G(B)$. Hence the bound of AB obeys $G(AB) \leqq G(A) \cdot G(B)$.

1.4.3. *Convergence of linear transformations*

A sequence of linear transformations A_n on E is said to converge to the linear transformation A on E, if the bound of the transformation $A_n - A$, $G(A_n - A)$, converges to zero.

It follows from the definition that for any vector x in E, $|A_n x - Ax| = |(A_n - A)x|$, which is at most $G(A_n - A) \cdot |x|$, converges to zero.

Now suppose, conversely, that $A_n x$ converges to Ax for every x in E. We will prove that A_n converges to A. Let $x_1, \ldots, x_k$ be an orthonormal basis for E. Then, for any $\varepsilon > 0$, there exists N such that for $n > N$ we have $|A_n x_i - Ax_i| < \frac{\varepsilon}{k}$ for $i = 1, \ldots, k$. Any unit vector $x = \Sigma_{i=1}^{k} (x, x_i)x_i$ satisfies $|(A_n - A)x| = |\Sigma_{i=1}^{k} (x, x_i)(A_n - A)x_i| \leqq \Sigma_{i=1}^{k} |(x, x_i)(A_n - A)x_i| \leqq \Sigma_{i=1}^{k} |(A_n - A)x_i| < \varepsilon$. Then also the least upper bound $G(A_n - A)$ of $(A_n - A)x$ is less than ε. Consequently $G(A_n - A)$ converges to zero, and A_n converges to A.

Here too we have an analogon of Cauchy's characteristic: if for any $\varepsilon > 0$ there exists $N(\varepsilon)$ such that $n > m > N$ implies $G(A_n - A_m) < \varepsilon$, then there exists a transformation A to which A_n converges.

The proof is as follows. From the assumption follows that, for any vector x and $\varepsilon' \geqslant \varepsilon \cdot |x|$, $\varepsilon > 0$, there exists N such that $n > m > N$ implies

$$|A_n x - A_m x| \leqq G(A_n - A_m) \cdot |x| \leqq \varepsilon \cdot |x| \leqq \varepsilon'.$$

According to Cauchy's characteristic for convergence of vectors, $A_n x$ converges to a limit named Ax. When we will have proved that the assignment of the vector Ax to x is a linear transformation, the proof that A_n has a limit namely A will be complete.

Let the limits of the sequences $A_n x$ and $A_n y$ be Ax and Ay respectively. Then for any $\varepsilon > 0$ there exists N such that for $n > N$ both $|A_n x - Ax| < \varepsilon$ and $|A_n y - Ay| < \varepsilon$. Then $|A_n(\lambda x + \mu y) - \lambda Ax - \mu Ay| = |\lambda A_n x - \lambda Ax + \mu A_n y - \mu Ay| \leqq |\lambda(A_n x - Ax)| + |\mu(A_n y - Ay)| < \{|\lambda| + |\mu|\}\varepsilon$. It follows that the limit of $A_n(\lambda x + \mu y)$ is equal to $\lambda Ax + \mu Ay$ or that $\lambda Ax + \mu Ay$ is assigned to $\lambda x + \mu y$ for any real λ and μ. Thus A is a linear transformation.

By means of the condition for convergence of A_n to A, that $A_n x$ should converge to Ax for every x, it is possible to prove that, if A_n converges to A, B_n to B and λ_n to λ, then $A_n + B_n$ converges to $A + B$, $\lambda_n A_n$ to λA and $A_n B_n$ to AB.

For $A_n + B_n$ the proof is as follows. For any fixed x and any $\varepsilon > 0$ there exists N such that for $n > N$ both $|A_n x - Ax| < \frac{1}{2}\varepsilon$ and $|B_n x - Bx| < \frac{1}{2}\varepsilon$. Then $|(A_n + B_n)x - (A + B)x| \leqq |A_n x - Ax| + |B_n x - Bx| < \varepsilon$ which means that the limit of $(A_n + B_n)x$ is $(A + B)x$.

For $\lambda_n A_n$: For any fixed x and any $\varepsilon > 0$ there exists N such that for $n > N$ both $|\lambda_n - \lambda| < \varepsilon$ and $|A_n x - Ax| < \varepsilon$. Then $|\lambda_n A_n x - \lambda Ax| \leqq |\lambda_n(A_n x - Ax)| + |(\lambda_n - \lambda)Ax| < |\lambda_n| \cdot \varepsilon + |Ax| \cdot \varepsilon \leqq \varepsilon(B + |Ax|)$ where B is an upper bound of the set $|\lambda_n|$. This means that the limit of $\lambda_n A_n x$ is λAx.

For $A_n B_n$: For any fixed x and any $\varepsilon > 0$ there exists N such that for $n > N$, $|B_n x - Bx| < \varepsilon$ and for $y = Bx$, $|A_n y - Ay| < \varepsilon$. Then $|A_n B_n x - ABx| \leqq |A_n B_n x - A_n Bx| + |A_n Bx - ABx| \leqq G(A_n) \cdot |B_n x - Bx| + |A_n(Bx) - A(Bx)| \leqq G \cdot \varepsilon + \varepsilon$, where G is an upper bound of the set of bounds of the sequence of transformations A_n. To prove the existence of G it will be shown that the sequence $G(A_n)$ has a limit namely $G(A)$. From $G(A + B) \leqq G(A) + G(B)$ we have $G(A_n) \leqq G(A_n - A) + G(A)$ or $G(A_n) - G(A) \leqq G(A_n - A)$ and $G(A) \leqq G(A - A_n) + G(A_n)$ or $G(A_n) - G(A) \geqslant - G(A - A_n)$. It follows that $|G(A_n) - G(A)| \leqq G(A_n - A)$ so that, if A_n converges to A, $G(A_n)$ converges to $G(A)$.

1.4.4. *A power series of transformations*

Consider the sequence $S_n = \Sigma_{p=0}^{n} A^p$, where A is a linear transformation, while A^0 represents the transformation 1. Let the bound of A obey $G(A) < 1$. It is clear that S_n is a linear transformation too. For $n > m$ we have $S_n - S_m =$ $= \Sigma_{p=m+1}^{n} A^p$ and according to the properties of a bound:

$$G(S_n - S_m) \leqq \Sigma_{p=m+1}^{n} G(A^p) \leqq \Sigma_{p=m+1}^{n} \{G(A)\}^p.$$

Because $G(A) < 1$ there exists, given $\varepsilon > 0$, N such that for $n > m > N$ $\Sigma_{p=m+1}^{n} \{G(A)\}^p < \varepsilon$. Thus S_n satisfies Cauchy's criterion for convergence of linear transformations and converges to a linear transformation S.

To calculate S we remark that the transformation $1 - A$ is non-singular. For $(1 - A)\mathrm{x} = 0$ and $\mathrm{x} \neq 0$ implies $A\mathrm{x} = \mathrm{x}$ or $|A\mathrm{x}| = |\mathrm{x}|$ on the one hand, and because $G(A) < 1$ also $|A\mathrm{x}| \leqq G(A) \cdot |\mathrm{x}| < |\mathrm{x}|$ on the other hand, a contradiction. Now $(1 - A)S_n = S_n - AS_n = 1 - A^{n+1}$, and $S_n = (1 - A)^{-1}(1 - A^{n+1})$. As $G(A^{n+1}) \leqq \{G(A)\}^{n+1}$ which converges to zero, A^{n+1} converges to 0.

Using the properties discussed in the foregoing section we may conclude that $1 - A^{n+1}$ converges to 1 and that S_n converges to $(1 - A)^{-1}$. Thus $S = (1 - A)^{-1}$.

1.5. EVALUATION OF PROJECTIONS

1.5.1. *General procedure*

In the applications we often meet the situation that E is spanned by the disjoint spaces A and B, both given by bases. The difficulty in evaluating the projection of a vector z on A along B, lies in the fact that usually z is given by the coordinates with respect to some basis of E, while the given basis vectors of A and B do not or do not all occur in that basis of E. If z is given as linear combination of the basis vectors of A and B the projection is trivial.

We wish to transform the general case to this trivial case. Let λ_i and μ_i be the coordinates of z (λ_i given and μ_i unknown respectively) with respect to the bases $\mathrm{x}_1, \ldots, \mathrm{x}_n$ and $\mathrm{y}_1, \ldots, \mathrm{y}_n$ respectively. Let $\mathrm{y}_1, \ldots, \mathrm{y}_k$ and $\mathrm{y}_{k+1}, \ldots, \mathrm{y}_n$ be bases of the disjoint spaces A and B respectively. Let y_j be given by $\mathrm{y}_j = \Sigma_{i=1}^{n} \alpha_{ij} \mathrm{x}_i$. The set of equations $\lambda_i = \Sigma_{j=1}^{n} \alpha_{ij} \mu_j$ (compare 1.2.5) is a system of n linear equations in $\mu_1, \ldots, \mu_n$. From the unique solution the coordinates of the projection of z on A with respect to $\mathrm{y}_1, \ldots, \mathrm{y}_k$ can be taken in order to evaluate this projection. The decomposition of z in components in more than two disjoint subspaces will be executed likewise.

Because solving n linear equations, especially if n is not very small, is very laborious in general, we look out for other simpler methods to decompose a vector in its components in subspaces of E. These methods involve the evaluations of orthogonal projections. If the metric is not too intricate orthogonal projections often can be calculated easily.

1.5.2. *Evaluation of orthogonal projections*

Let $\mathrm{x}_1, \ldots, \mathrm{x}_m$ be a basis of the subspace A of E. The orthogonal projection of a vector z in E on A will be a linear combination, $\lambda_1 \mathrm{x}_1 + \lambda_2 \mathrm{x}_2 + \ldots + \lambda_m \mathrm{x}_m$, of $\mathrm{x}_1, \ldots, \mathrm{x}_m$ such that $\mathrm{z} - \Sigma_{i=1}^{m} \lambda_i \mathrm{x}_i$ is orthogonal to A, or that its inner product with $\mathrm{x}_1, \ldots, \mathrm{x}_m$ is zero. This leads to a system of m linear equations in the unknowns λ_i of the form: $\lambda_1(\mathrm{x}_i, \mathrm{x}_1) + \lambda_2(\mathrm{x}_i, \mathrm{x}_2) + \ldots + \lambda_m(\mathrm{x}_i, \mathrm{x}_m) = (\mathrm{x}_i, \mathrm{z})$ for $i = 1, \ldots, m$. These equations are called *normal equations*.

It will be remarked that the matrix of coefficients is symmetric which, in addition to the fact that $m < n$, means a computational simplification in comparison with the general case discussed in the foregoing section. After the solution of the λ_i one can evaluate the orthogonal projection on A as $\Sigma_{i=1}^{m}\lambda_i x_i$ and the projection on $A^{\perp}$ as its difference with z.

If $x_1, \ldots, x_m$ is an orthogonal (orthonormal) basis of A, then the solution of the normal equations is simple: $\lambda_i = (x_i, z)/(x_i, x_i)$ or $\lambda_i = (x_i, z)$ respectively. This simple solution agrees with the remark at the end of 1.3.5: the orthogonal projection on the space spanned by the orthogonal basis $x_1, \ldots, x_m$ is equal to the sum of the orthogonal projections on the spaces spanned by $x_1, \ldots, x_m$ separately.

Therefore, if A is given by a non-orthogonal basis, one often prefers to orthogonalize this basis first, and to evaluate the orthogonal projection on A then, instead of solving the normal equations derived from the given basis. It is also for this reason that orthogonal polynomials have been introduced. We remark that the orthogonalized basis does not need to be orthonormal.

1.5.3. *An iterative method to obtain components in given spaces*

Let E be spanned by the not necessarily disjoint subspaces $A_1, \ldots, A_k$. We wish to decompose the vector z in E to $y_1 + \ldots + y_k$ such that y_i is in A_i for every *i*. Such an y_i is called an oblique component of y in A_i. We assume that each A_i is given by an orthogonal basis so that orthogonal projection is easy. We next present an iterative process to obtain oblique components of z, in which every step is an orthogonal projection. This process is valuable in the applications.

For convenience we denote $A_{sk+i} = A_i$, in which *s* is a positive integer. Consider the sequence of pairs of vectors: z_n, p_n defined by: z_1 is the orthogonal projection of z on A_1 and $p_1 = z - z_1$ is the corresponding perpendicular; …; z_r is the orthogonal projection of the perpendicular p_{r-1} on A_r and $p_r = p_{r-1} - z_r$ is the corresponding perpendicular.

Every step will yield a vector z_r in one of the spaces $A_1, \ldots, A_k$, while p_n is equal to $z - \Sigma_{r=1}^{n} z_r$.

We assert (which assertion will be proved) that the sequence of perpendiculars p_n converges to the null vector. This implies that the sequence of sums of orthogonal projections converges to z. These converging sums consist of contributions in every A_i, so that the desired components can be approximated with any required accuracy.

1.5.4. *Proof of the convergence*

Let the orthogonal projection of a vector z on $A_i^{\perp}$ be denoted by P_iz. Then the sequence of perpendiculars is: P_1z; P_2P_1z; $P_3P_2P_1$z; …; $P_kP_{k-1} \ldots P_2P_1$z; $P_1P_kP_{k-1} \ldots P_2P_1$z; $P_2P_1P_kP_{k-1} \ldots P_2P_1$z; … We wish to consider the linear transformation $D = P_{r+k} \ldots P_{r+2}P_{r+1}$ which is a mapping of $A_r^{\perp}$ into itself. Without loss of generality we may take $D = P_kP_{k-1} \ldots P_3P_2P_1$, by which to a vector in $A_k^{\perp}$ a vector in $A_k^{\perp}$ is assigned. Because *D* is a product of orthogonal projections the bound of *D* satisfies $G(D) \leq 1$. Our purpose is to prove: $G(D) < 1$. For if $G(D) < 1$, then the sequence of transformations $B_n = D^n$ converges to *0*, so that the sequence of perpendiculars in $A_k^{\perp}$ converges to the null vector and the same is true for the sequence of all perpendiculars. The proof is as follows:

Assume $G(D) = 1$. Then according to 1.4.2 there is a vector $x \neq 0$ in $A_k{}^{\perp}$ such that $|Dx| = |x|$. We know $|P_i z| \leqq |z|$ where equality holds only if $P_i z = z$. It follows that $P_1 x = x$; $P_2 P_1 x = x$; $P_3 P_2 P_1 x = x$ and so on. Consequently x is in all $A_i{}^{\perp}$ $(i = 1, \ldots, k)$, in other words, is orthogonal to a set of spaces that span E. But only the null vector is orthogonal to E which is in contradiction with $x \neq 0$. Hence $G(D) < 1$.

1.5.5. *Extension of the method*

Let $A_1, \ldots, A_k$ be subspaces of E which do not span E. Let z be a vector in E. When we perform the same sequence of operations with z now, as in the case where $A_1, \ldots, A_k$ span E, we will obtain again orthogonal projections on the spaces $A_1, \ldots, A_k$, and perpendiculars, each of which is equal to the difference between z and the sum of orthogonal projections determined in the preceding steps.

In order to find out the result of these operations we consider the (unique) decomposition of z in a component in the space spanned by $A_1, \ldots, A_k$ and a component in the residual space of E orthogonal to $A_1, \ldots, A_k$. The first component is in the same position now as z in the foregoing case: it will be decomposed in components in the subspaces $A_1, \ldots, A_k$, while the sequence of perpendiculars converges to the null vector. The second component will not give any contribution in $A_1, \ldots, A_k$ so that the corresponding sequence of perpendiculars has all elements equal to this component. It follows that the difference between z and the sum of the orthogonal projections on $A_1, \ldots, A_k$ converges to the perpendicular from z on the space spanned by $A_1, \ldots, A_k$.

So we have found a method to evaluate an orthogonal projection without an orthogonal basis and without solving normal equations. In addition we obtain (oblique) components of this projection in the subspaces that span the space on which z is projected orthogonally.

1.5.6. *Another description of the same method*

An elaboration of the method will show that the sequence of perpendiculars plays a little rôle in fact. For that purpose we consider the case $k = 3$ first. The subspaces A_1, A_2 and A_3 that span the space on which z is projected orthogonally are denoted now by A, B and C. Further we introduce the notation z_A and $(z_A)_B$ or z_{AB} for the orthogonal projection of z on A and the orthogonal projection of z_A on B respectively. In the successive rows we write the vectors to be projected on the left and the projections on the right. We obtain:

$$\begin{array}{ll}
z & u_1 = z_A \\
p_1 = z - u_1 & v_1 = (z - u_1)_B = z_B - z_{AB} \\
p_2 = z - u_1 - v_1 & w_1 = (z - u_1 - v_1)_C = z_C - z_{AC} - (v_1)_C \\
p_3 = z - u_1 - v_1 - w_1 & u_2 = z_A - (u_1)_A - (v_1 + w_1)_A = -(v_1 + w_1)_A \\
p_4 = p_1 - v_1 - w_1 - u_2 & v_2 = (p_1)_B - (v_1)_B - (w_1 + u_2)_B = -(w_1 + u_2)_B \\
p_5 = p_2 - w_1 - u_2 - v_2 & w_2 = (p_2)_C - (w_1)_C - (u_2 + v_2)_C = -(u_2 + v_2)_C \\
p_6 = p_3 - u_2 - v_2 - w_2 & u_3 = (p_3)_A - (u_2)_A - (v_2 + w_2)_A = -(v_2 + w_2)_A
\end{array}$$

and so on.

We observe that from the beginning of the second cycle of orthogonal projections every orthogonal projection is equal to the orthogonal projection of the sum of the $k-1$ foregoing orthogonal projections on the same space but with opposite sign. Only in the first cycle subtractions take place in fact. The left hand side, consisting of perpendiculars, can thus be left out of consideration.

In the case $k = 2$ and subspaces A and B the sequence of orthogonal projections is as follows: $u_1 = z_A$; $v_1 = z_B - z_{AB}$; $u_2 = -(v_1)_A$; $v_2 = -(u_2)_B$; $u_3 = -(v_2)_A$; $v_3 = -(u_3)_B$ and so on.

We remark that in this case after the first cycle all minus signs can be omitted. Then the vectors in B will keep the right sign, while the vectors to be found in A must be subtracted from z_A. It is obvious that the sum of these vectors to be subtracted from z_A is equal to the orthogonal projection of the component in B i.e. $\Sigma_{i=1}^{\infty} v_i$ on A.

1.6. SYMMETRIC TRANSFORMATIONS AND MATRICES

1.6.1. *Definitions*

Let A be a linear transformation on E. The transpose A^t of A with respect to an inner product is defined by the identity in x and y: $(Ax, y) = (x, A^t y)$. A^t is a linear transformation; for if $y = \lambda y_1 + \mu y_2$ then $(x, A^t y) = (Ax, y) = \lambda(Ax, y_1) + \mu(Ax, y_2) = \lambda(x, A^t y_1) + \mu(x, A^t y_2) = (x, \lambda A^t y_1 + \mu A^t y_2)$. From $(Ax, y) = (x, A^t y) = \{(A^t)^t x, y\}$ for every x and y follows: $(A^t)^t = A$.

Let the matrix of A associated with an orthonormal basis $x_1, \ldots, x_n$ consist of elements α_{ij}. Then, because $Ax_j = \Sigma_{k=1}^{n} \alpha_{kj} x_k$, $(Ax_j, x_i) = \Sigma_{k=1}^{n} \alpha_{kj} (x_k, x_i) = \alpha_{ij}$. If the matrix of A^t associated with the same basis consists of elements α^t_{ij}, then $\alpha^t_{ij} = (A^t x_j, x_i) = (x_j, Ax_i) = \alpha_{ji}$. The matrix of A^t where $\alpha^t_{ij} = \alpha_{ji}$ is called the transpose of the matrix of A.

If $A = A^t$, the transformation A (and also the matrix of A associated with an orthonormal basis) is said to be symmetric. A symmetric transformation A with $(x, Ax) \geqq 0$ for every x is called non-negative. If moreover $(x, Ax) = 0$ only if $x = 0$, then A is called positive definite. In that case we conclude from the Schwarz inequality $|(x, Ax)| \leqq |Ax| \cdot |x|$, that $Ax = 0$ implies $(x, Ax) = 0$ and $x = 0$, in other words, that A is non-singular.

Consider the inner product (x, y) of the vectors x and y. If A is positive definite with respect to that inner product then (Ax, y) is a second inner product of x and y. For (compare 1.3.1) : $(Ax, y) = (Ay, x)$; $(Ax, \lambda y_1 + \mu y_2) = \lambda(Ax, y_1) + \mu(Ax, y_2)$ and (Ax, x) is positive definite. Let the symbols for vectors x and y etc. stand for the n-tuples of their coordinates $(\lambda_1, \ldots, \lambda_n)$, $(\mu_1, \ldots, \mu_n)$ etc. with respect to the above-mentioned basis $x_1, \ldots, x_n$. Because this basis was named orthonormal, the first inner product of x and y is equal to $\lambda_1\mu_1 + \ldots + \lambda_n\mu_n$. Then the second inner product of the vectors x and y which, according to the foregoing, assigns the element α_{ij} from the matrix of A with respect to $x_1, \ldots, x_n$ to the pair x_i and x_j, and which thus is equal to $\Sigma_i \Sigma_j \alpha_{ij} \lambda_i \mu_j$, has the same value as the first inner product of the vectors Ax and y. In the sequel we will need that the projection, orthogonal with respect to the second inner product, of y on B, y_B, is that vector y_B in B, for which the projection, orthogonal with respect to the first inner product, of $A(y - y_B)$ on B is equal to the null vector.

1.6.2. *Proper values and vectors of symmetric transformations*

In the introduction of proper vectors and values we passed over the possibility of complex roots for λ. To study this possibility we admit for a moment n-tuples of complex numbers as coordinates of a vector with respect to an orthonormal basis. For solution of $x = (\lambda_1, \ldots, \lambda_n)$ in the equation $Ax = \lambda x$, in which A is represented by a matrix of real numbers α_{ij} with respect to an orthonormal basis and in which λ is a complex number, implies the solution of n linear equations of the form $\Sigma_j \alpha_{ij} \lambda_j = \lambda \lambda_i$, which will yield complex values for the λ_i in general. Then $(\mathrm{x}, \mathrm{y}) = \Sigma_{i=1}^{n} \lambda_i \mu_i$, and if A is symmetric $(A\mathrm{x}, \mathrm{y}) = \Sigma_i \Sigma_j \alpha_{ij} \lambda_i \mu_j = = (\mathrm{x}, A\mathrm{y})$. If every (complex) coordinate $a+bi$ of such a vector z is replaced by its complex conjugate $a-bi$ we obtain $\bar{\mathrm{z}}$.

Next we will prove that all proper values of a real symmetric transformation are real. First we observe that the complex conjugate of $(\bar{\mathrm{z}}, A\mathrm{z})$ equals (the complex conjugate of a product equals the product of the complex conjugates of the factors): $(\bar{\bar{\mathrm{z}}}, \overline{A\mathrm{z}}) = (\mathrm{z}, A\bar{\mathrm{z}}) = (\bar{\mathrm{z}}, A\mathrm{z})$. Hence $(\bar{\mathrm{z}}, A\mathrm{z})$ is real. If z is a proper vector that belongs to the proper value λ, then $(\bar{\mathrm{z}}, A\mathrm{z}) = (\bar{\mathrm{z}}, \lambda\mathrm{z}) = \lambda(\bar{\mathrm{z}}, \mathrm{z})$ and, as $(\bar{\mathrm{z}}, A\mathrm{z})$ and $(\bar{\mathrm{z}}, \mathrm{z})$ are real, so is λ. Complex coordinates may further be left out of consideration.

Consider two proper vectors x_1 and x_2 of the symmetric transformation A associated with the two different proper values λ_1 and λ_2. We have $0 = (A\mathrm{x}_1, \mathrm{x}_2) - (\mathrm{x}_1, A\mathrm{x}_2) = \lambda_1(\mathrm{x}_1, \mathrm{x}_2) - \lambda_2(\mathrm{x}_1, \mathrm{x}_2) = (\lambda_1 - \lambda_2)\ (\mathrm{x}_1, \mathrm{x}_2)$. Hence $(\mathrm{x}_1, \mathrm{x}_2) = 0$. Proper vectors of A belonging to different proper values are orthogonal. The same holds for two subspaces of proper vectors which belong to two different proper values.

It follows that the sum of geometric multiplicities of the proper values is n at most. On the other hand the space spanned by these orthogonal subspaces must be E.

To prove this we choose an orthonormal basis for E consisting of orthonormal bases of all these subspaces and, if necessary, completed with vectors orthogonal to these. The matrix of A with respect to this basis will contain in the column corresponding to a proper vector the associated proper value on the diagonal, and zeros elsewhere; because of the symmetry, the remaining columns will contain zeros in all the rows with an index equal to that of a column associated with a proper vector. Hence vectors in the space orthogonal to the subspaces of proper values are transformed by A in vectors in that same space. The linear transformation A considered within this space must have at least one proper value and an associated proper vector. But so we would find proper vectors of A not contained in the subspaces of proper vectors of A, which is a contradiction. Thus the subspaces of proper vectors span E.

Moreover the matrix of A corresponding to the chosen orthonormal basis contains the proper values each according to the geometric multiplicity in the diagonal and zeros elsewhere. From the determinant of $A - \lambda I$ it follows easily now, that the algebraic multiplicity is equal to the geometric multiplicity.

1.6.3. *Representation of symmetric transformations by means of projections*

Remembering our remark about the matrix of a projection (1.2.7) we may express the result at the end of the last section as follows: Let P_j be the orthogonal projection on the subspace of proper vectors associated with a proper value λ_j of A, so that $\Sigma_{j=1}^{p} P_j = I$ $(p \leq n)$; then A can be represented as $\Sigma_{j=1}^{p} \lambda_j P_j$ where the λ_j are distinct. This representation of A as a linear combination of orthogonal projections on orthogonal spaces, that together span E, is unique.

To prove this we assume that the symmetric transformation A can be written as $\Sigma_{j=1}^{p} \mu_j Q_j$ where the μ_j are distinct and the Q_j orthogonal projections on orthogonal spaces which together span E. Let x be a vector in the pointwise invariant space of Q_i i.e. $Q_i x = x$. Then $Q_j x = 0$ for $j \neq i$, and $Ax = \Sigma_j \mu_j Q_j x = \mu_i Q_i x = \mu_i x$. Thus every μ_j is a proper value of A, and the invariant space of Q_i is a set of proper vectors belonging to the proper value μ_i. Such a space must be the complete set of proper vectors associated with μ_i, because the invariant spaces of the Q_j span E.

Now the following properties are simple corollaries. A is singular if and only if some $\lambda_j = 0$ (this is also true for non-symmetric transformations as follows from the definition of proper values and vectors). A is non-negative only if all $\lambda_j \geqq 0$; for remembering $P_i^2 = P_i$ and $P_i P_j = 0$ for $i \neq j$, we find $(x, Ax) = (x, \Sigma_j \lambda_j P_j x) = [\Sigma_j (P_j x), \Sigma_j \lambda_j (P_j x)] = \Sigma_j \lambda_j (P_j x)^2$. A is positive definite only if all $\lambda_j > 0$. The inverse of A is equal to $\Sigma_j (\lambda_j)^{-1} P_j$ (with $\lambda_j \neq 0$ of course). If A is positive definite the same is true for A^{-1}.

If A is positive definite there is one positive definite transformation X such that $X^2 = A$; if $A = \Sigma_j \lambda_j P_j$, then this transformation X, which we call the square root of A, is $\Sigma_j \sqrt{\lambda_j} P_j$ where $\sqrt{\lambda_j}$ is the positive square root of λ_j. X is denoted by $\sqrt{A}$.

With reference to the end of 1.6.1 we observe that the second inner product (x, Ay) of the vectors x and y (determined by the second inner products α_{ij} of the basis vectors, which are elements of the symmetric matrix of the transformation A) is equal to the first inner product of the vectors x and Ay. Observe that Ay is a linear combination of orthogonal projections of y on some particular orthogonal spaces (with respect to the first inner product). We also have $(x, Ay) = (\sqrt{A}\, x, \sqrt{A}\, y)$ or in words: The second inner product of x and y is equal to the first inner product of $\sqrt{A}$ x and $\sqrt{A}$ y.

CHAPTER 2

STATISTICAL CONSIDERATIONS

2.1. RANDOM VECTORS

2.1.1. *Covariance matrix*

Let x be a vector in a n-dimensional space of n-tuples $(x_1, \ldots, x_n)$ with coordinates $x_1, \ldots, x_n$ with respect to the basis $e_1, \ldots, e_n$. Let the coordinates have a joint distribution function so that x is a random vector. Let Ex_i be the expectation (value) of x_i; let $\sigma^2(x_i) = E(x_i - Ex_i)^2$ be the variance of x_i and $\text{cov}\,(x_i, x_j) = E[(x_i - Ex_i)(x_j - Ex_j)]$ the covariance of x_i and x_j. We assume that the probability is not concentrated (with probability one) in any subset of E satisfying: $\Sigma_i w_i x_i = C$. Then $\Sigma_i w_i x_i$ has positive variance for any $(w_1, \ldots, w_n) \neq 0$. This variance, $\Sigma_{i,j}\, \text{cov}\,(x_i, x_j)\, w_i w_j$, is therefore a positive definite quadratic form in the variables $w_1, \ldots, w_n$, determined (compare 1.3.1) by the also positive definite covariance matrix V with elements $v_{ij} = \text{cov}\,(x_i, x_j)$.

Let A be a transformation with a matrix A of elements α_{ij} so that the coordinates of $y = Ax$ are $y_i = \Sigma_j \alpha_{ij} x_j$. Then $\text{cov}\,(y_i, y_j) = \text{cov}\,(\Sigma_k \alpha_{ik} x_k)(\Sigma_l \alpha_{jl} x_l) = \Sigma_{k,l} \alpha_{ik}\, \text{cov}\,(x_k, x_l) \alpha_{jl} = \Sigma_{k,l} \alpha_{ik} v_{kl} \alpha_{jl} = \Sigma_l (\Sigma_k \alpha_{ik} v_{kl}) \alpha^t{}_{lj}$. This can be summarized in saying that the covariance matrix of Ax is equal to AVA^t. Because

V is positive definite a possible choice for the transformation A is $\sqrt{V^{-1}}$. We find that the covariance matrix of $\sqrt{V^{-1}}\mathrm{x}$ is I.

In the following it will appear to be useful to introduce a metric I in case the covariance matrix is I. With reference to the end of the foregoing section we observe that a metric I for the coordinates of $\sqrt{V^{-1}}\mathrm{x}$ is equivalent to a metric $(\sqrt{V^{-1}})^2 = V^{-1}$ for the coordinates of x. In the case of a covariance matrix V of the coordinates of x we introduce therefore a metric V^{-1} for these coordinates. We then obtain a situation which is equivalent to the use of the metric I for the coordinates of $\sqrt{V^{-1}}\mathrm{x}$; these coordinates have covariance matrix I. Properties of orthogonal projections with metric I of random vectors with covariance matrix I also hold for orthogonal projections with metric V^{-1} of random vectors with covariance matrix V.

In the applications a slightly different metric will often be used, in case the covariance matrix of x is given in the form $\sigma^2 V$. We introduce a metric V^{-1}, which is equivalent to a metric I for the coordinates of $\sqrt{V^{-1}}\mathrm{x}$; these coordinates have the covariance matrix $\sqrt{V^{-1}}\cdot\sigma^2\cdot \mathrm{V}\cdot\sqrt{V^{-1}} = \sigma^2\cdot I$.

In the sequel we mainly study the special case $V = I$, as the general case is obtained immediately from it.

2.1.2. *Vectors with covariance matrix $\sigma^2\cdot I$ and metric I*

Let $\mathrm{x} = (x_1, \ldots, x_n)$ be a random vector with covariance matrix $\sigma^2\cdot I$ and in a space with metric given by $\mathrm{x}^2 = x_1^2+\ldots+x_n^2$. Let the orthogonal projection of x on the one-dimensional space spanned by the unit vector $\mathrm{a} = (a_1, \ldots, a_n)$ be $\lambda\mathrm{a}$. Then the coefficient $\lambda = (\mathrm{x}, \mathrm{a}) = (x_1a_1+\ldots+x_na_n)$ has variance $\Sigma_i a_i^2\sigma^2 = \sigma^2$. Further if the unit vector $\mathrm{b} = (b_1, \ldots b_n)$ is orthogonal to a, then the covariance of the coefficients of the projections, (x, a) and (x, b) respectively, is $\mathrm{cov}\,(x_1a_1+\ldots+x_na_n)\,(x_1b_1+\ldots+x_na_n) = \Sigma_i a_ib_i\sigma^2 = 0$.

If in particular $E\mathrm{x} = 0$, that is $Ex_i = 0$ for every i, then the square x_A^2 of the orthogonal projection of x on a m-dimensional subspace A, which is equal to the sum of the m squares of the orthogonal projections of x on m orthogonal one-dimensional subspaces of A, has the expectation value $m\sigma^2$.

If $E\mathrm{x} = (\mu_1, \ldots, \mu_n) = \mathrm{u}$, then the same applies to $\mathrm{x}-\mathrm{u}$. In particular $E\{(\mathrm{x}-\mathrm{u})_\mathrm{A}\}^2 = m\sigma^2$. Now $\mathrm{x}_\mathrm{A} = \mathrm{u}_\mathrm{A}+(\mathrm{x}-\mathrm{u})_\mathrm{A}$ and $\mathrm{x}_\mathrm{A}^2 = \mathrm{u}_\mathrm{A}^2+(\mathrm{x}-\mathrm{u})_\mathrm{A}^2+ +2\mathrm{u}_\mathrm{A}(\mathrm{x}-\mathrm{u})_\mathrm{A}$. Let B be a linear transformation; then $B\mathrm{y}$ is a vector of which the coordinates are linear functions of the coordinates of y, and thus $E(B\mathrm{y}) = = BE\mathrm{y}$; further (y, b), where b is some vector, is a linear function in the coordinates of y, so that $E(\mathrm{y}, \mathrm{b}) = (E\mathrm{y}, \mathrm{b})$. Thus $E\mathrm{y} = 0$ implies $E(B\mathrm{y}) = 0$ and $(E\mathrm{y}, \mathrm{b}) = 0$. It follows that $E(\mathrm{x}-\mathrm{u})_\mathrm{A} = 0$, and, from this, that $E\{\mathrm{u}_\mathrm{A}, (\mathrm{x}-\mathrm{u})_\mathrm{A}\} = 0$. So we have $E\mathrm{x}_\mathrm{A}^2 = \mathrm{u}_\mathrm{A}^2 + m\sigma^2$. In connection with the foregoing section we remark that the same is true, if the covariance matrix is $\sigma^2 V$ and the metric V^{-1} is used.

2.2. NORMAL DISTRIBUTION

2.2.1. *The case with covariance matrix $\sigma^2\cdot I$*

Let the coordinates $x_1, \ldots, x_n$ of $\mathrm{x} = (x_1, \ldots, x_n)$ have expectation 0 and covariance matrix $\sigma^2\cdot I$. It follows from the last section that the coordinates $z_1, \ldots, z_n$ of x with respect to any orthonormal basis of E have the covariance matrix $\sigma^2 \cdot I$ too. We consider the particular case in which the coordinates $x_1, \ldots, x_n$ have a normal $(0, \sigma)$ distribution and are mutually independent. The probability density of x is then the product of the probability densities of n normal $(0, \sigma)$ distributions i.e. $C \exp \left\{-\frac{1}{2\sigma^2}(x_1^2+\ldots+x_n^2)\right\} = C \exp \left(-\frac{\mathrm{x}^2}{2\sigma^2}\right)$. We observe that this density depends on the square of x only. The coordinates $z_1, \ldots, z_n$ with respect to another orthonormal basis satisfy $\mathrm{x}^2 = \Sigma_{i=1}^n z_i^2$. Because the functional determinant for change of variables is a constant, when the new variables are linear functions of the old ones (and equal to I here), the joint probability density for $z_1, \ldots, z_n$ is: $C \exp \left(-\frac{1}{2\sigma^2}\Sigma_i z_i^2\right) = C\, \Pi_i \exp \left(-\frac{z_i^2}{2\sigma^2}\right)$ from which follows that the z_i have a normal distribution too, and are independent.

Because σ^{-2} times the sum of squares of m independent variables with a normal $(0, \sigma)$ distribution has by definition a chi-square distribution with dimension m, it follows that σ^{-2} times the square of the orthogonal projection of x on A has a chi-square distribution with dimension m too. Similarly if B is a subspace with dimension m_1 orthogonal to A, then $\mathrm{x}_\mathrm{B}^2/\sigma^2$ has a m_1-dimensional chi-square distribution independent of the projection on A. It follows that $\frac{\mathrm{x}_\mathrm{A}^2/m}{\mathrm{x}_\mathrm{B}^2/m_1}$ has by definition a F-distribution with the dimensions m and m_1 as parameters. If $E\mathrm{x} = \mathrm{u}$ then the foregoing holds for $\mathrm{x} - \mathrm{u}$.

If $\mathrm{u}_\mathrm{B} = 0$, then it follows from the expectations of x_A^2 and x_B^2 which are equal to $\mathrm{u}_\mathrm{A}^2 + m\sigma^2$ and $m_1\sigma^2$ respectively, that an appropriate test criterion for the null hypothesis $\mathrm{u}_\mathrm{A} = 0$ will be formed by the quotient $\frac{\mathrm{x}_\mathrm{A}^2/m}{\mathrm{x}_\mathrm{B}^2/m_1}$ with a one-sided upper critical region in the F-distribution with dimensions m and m_1.

2.2.2. *The general case*

Let x be a random vector of which the coordinates have covariance matrix V. If the coordinates of $\mathrm{y} = \sqrt{V^{-1}}\mathrm{x}$ have the normal distribution given above (with covariance matrix I of course), then the coordinates of x are said to have a normal distribution too. The probability density of the coordinates of y is $C \exp (-\frac{1}{2}\mathrm{y}^2) = C \exp \{-\frac{1}{2}(\mathrm{x}, V^{-1}\mathrm{x})\}$, so that the probability density for the coordinates of x is $C' \exp\{-\frac{1}{2}(\mathrm{x}, V^{-1}\mathrm{x})\}$. C and C' are suitably chosen constants.

Conversely, if the coordinates of x have probability density $C \exp \{-\frac{1}{2}(\mathrm{x}, A\mathrm{x})\}$, where $(\mathrm{x}, A\mathrm{x})$ is a positive definite form with matrix A in the coordinates of x, then they have a normal distribution with covariance matrix A^{-1}. For, if $\mathrm{y} = \sqrt{A}\,\mathrm{x}$, the probability density can be written $C \exp (-\frac{1}{2}\mathrm{y}^2)$, so that the probability density for the coordinates of y is $C'' \exp (-\frac{1}{2}\mathrm{y}^2)$.

Because also in this respect the use of a metric A (the inverse of the covariance matrix) for the coordinates of x is equivalent to the use of the metric I for the coordinates of y, it follows immediately e.g. that the square (with respect to a metric A) of the orthogonal projection (with respect to a metric A) of x on a m-dimensional subspace has a m-dimensional chi-square distribution!

2.3. LINEAR REGRESSION

2.3.1. *The problem*

Consider n random variables $y_1, \ldots, y_n$ with expectation values $Ey_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik}$ for $i = 1, \ldots, n$.

The values of the y_i refer to observations of a property of individuals (persons, trial plots, countries, animals) to be studied. The observations are taken from n such individuals numbered $1, \ldots, n$.

The quantity x_{ij} is the value, for the i-th individual, of the j-th property from k also quantitatively expressible properties of the individuals. The x_{ij} are given, while the coefficients β_j, common to all individuals, are unknown.

The supposition about the n observations $y_1, \ldots, y_n$ can also be written as: $E\mathrm{y} = \beta_1 \mathrm{x}_1 + \ldots + \beta_k \mathrm{x}_k$, where $\mathrm{x}_1, \ldots, \mathrm{x}_k$ are vectors and y a random vector in the space E of n-tuples $(z_1, \ldots, z_n)$, such that numbers with the same index i in $y_i, x_{i1}, x_{i2}, \ldots, x_{ik}$ are corresponding coordinates (with respect to the basis $\mathrm{e}_1, \ldots, \mathrm{e}_n$).

The problem is to determine an unbiased and most efficient estimator of a linear function of the coefficients β_j, say $p_1\beta_1 + \ldots + p_k\beta_k$. Important examples of the functions to be estimated are the β_j themselves and the Ey_i. In connection with the condition of minimal variance of the estimator the covariance matrix of y is assumed to be a multiple of the known matrix V; it has thus the form $\sigma^2 V$, with σ not necessarily known.

2.3.2. *Unbiased linear estimators*

First we introduce the notion of linear functions on vector spaces. A linear function $f(\mathrm{z})$ on a vector space E assigns to every vector z in E a real number $f(\mathrm{z})$ such that, for z_1 and z_2 in E, $f(\lambda \mathrm{z}_1 + \mu \mathrm{z}_2) = \lambda f(\mathrm{z}_1) + \mu f(\mathrm{z}_2)$. In particular the value, assigned to the null vector, is equal to $f(0) = f(\mathrm{z} - \mathrm{z}) = f(\mathrm{z}) - f(\mathrm{z}) = 0$. Such a function is completely determined by the values assigned to the vectors of a basis of E.

We will look for an estimator which is a linear function f on E, so that the estimate will be $f(\mathrm{y})$. Even if we restrict ourselves to linear functions of the observation y (which may be written as $\Sigma_{i=1}^{n} c_i y_i$ if c_i is the value assigned to the i-th vector of the basis $\mathrm{e}_1, \ldots, \mathrm{e}_n$), it is possible to find an unbiased and most efficient estimator. Moreover such a function is multiplied by λ if y is multiplied by λ, in other words, it is not sensitive to the scale of y; finally in case every y_i is equal to the sum of e.g. two (possibly dependent) variables – for instance the yields of the underground and the overground parts of a crop such as turnips – and the observations of the parts have proportional covariance matrices, then the estimator of the whole will appear to be equal to the sum of the estimators belonging to the parts.

In order that $f(\mathrm{y})$ be an unbiased estimator, $Ef(\mathrm{y})$ must be equal to $p_1\beta_1 + \ldots + p_k\beta_k$ for any β_j. Now $Ef(\mathrm{y}) = f(E\mathrm{y}) = f(\Sigma_j \beta_j \mathrm{x}_j) = \Sigma_j \beta_j f(\mathrm{x}_j)$. Hence $f(\mathrm{x}_j)$ must be equal to p_j for all j.

In case the vectors $\mathrm{x}_1, \ldots, \mathrm{x}_k$ are independent, this restriction of f will not be contrary to the definition of a linear function; for then these vectors may be considered as a part of a basis for E, of which the function values determine f. If, however, the vectors x_j are dependent, say $\Sigma \alpha_j \mathrm{x}_j = 0$, then $f(0) = f(\Sigma \alpha_j \mathrm{x}_j) = \Sigma \alpha_j f(\mathrm{x}_j) = 0$, so that also $\Sigma \alpha_j p_j$ should be zero. Hence, if and only if the p_j satisfy all relations $\Sigma \alpha_j p_j = 0$ corresponding to all relations $\Sigma \alpha_j \mathrm{x}_j = 0$, then it is possible to find an unbiased estimator for $\Sigma p_j\beta_j$. Then $\Sigma p_j\beta_j$ is said to be estimable (Rao [19]).

Another formulation of this restriction of the p_j is found by writing $Ef(\mathrm{y}) = E\Sigma_i c_i y_i = \Sigma_i c_i E y_i$. Because $Ef(\mathrm{y})$ must be equal to $\Sigma_j p_j\beta_j$, it follows that *only linear combinations of the* $Ey_i = \Sigma_j x_{ij}\beta_j$ *are estimable.* If $\Sigma p_j\beta_j$ is estimable, then the value of any unbiased estimator f is completely determined for a vector in the subspace A spanned by $\mathrm{x}_1, \ldots, \mathrm{x}_k$: $f(\mathrm{x}_j) = p_j$.

Theorem: Let g be the linear function on A which is equal to the restriction to A of any unbiased estimator f of $\Sigma p_j\beta_j$, so that $g(\mathrm{x}_j) = p_j$ for all j. Then if P is any projection on A, gP is such an estimator f; conversely, if f is such an estimator, there exists a projection P on A such that $f = gP$.

Proof: Let P be any projection on A. Then gP is a linear function on E; for $gP(\lambda\mathrm{y}_1 + \mu\mathrm{y}_2) = g(\lambda P\mathrm{y}_1 + \mu P\mathrm{y}_2) = \lambda g(P\mathrm{y}_1) + \mu g(P\mathrm{y}_2)$. Moreover $gP\mathrm{x}_j = g(\mathrm{x}_j) = p_j$. Hence gP is as required. Conversely, let f be any linear function for which $f(\mathrm{x}_j) = p_j$, for all j. Because not all p_j are zero it is possible to choose a vector a_1 in the m-dimensional space A such that $f(\mathrm{a}_1) \neq 0$. Let $\mathrm{a}_1, \ldots, \mathrm{a}_n$ be a basis for E, with $\mathrm{a}_1, \ldots, \mathrm{a}_m$ in A. Now consider the set of vectors a_r $(r = 1, \ldots, m)$ and $\mathrm{b}_s = \mathrm{a}_s - \dfrac{f(\mathrm{a}_s)}{f(\mathrm{a}_1)}\mathrm{a}_1$ $(s = m+1, \ldots, n)$. Because dependence of this set would imply dependence of the first set, the second set is also a basis for E. Further we observe that $f(\mathrm{b}_s) = 0$ for all s. Let P be the projection on A along the space spanned by the vectors b_s. Then $f(\mathrm{y}) = g(P\mathrm{y})$ for every y in E. The proof is complete.

An unbiased estimate of an estimable $\Sigma p_j\beta_j$ is thus found with the help of a projection P of y on A, which yields $b_1\mathrm{x}_1 + \ldots + b_k\mathrm{x}_k$, and calculation of $g(P\mathrm{y}) = b_1p_1 + \ldots + b_kp_k$ i.e. substitution of b_j for β_j in the function to be estimated.

2.3.3. *Most efficient unbiased estimators*

A uniquely determined estimator f will be obtained with the help of the efficiency condition. Recalling that the covariance matrix of $\mathrm{y} = (y_1, \ldots, y_n)$ is $\sigma^2 \cdot V$, we introduce the metric V^{-1} with respect to the basis $\mathrm{e}_1, \ldots, \mathrm{e}_n$. Let $\mathrm{d}_1, \ldots, \mathrm{d}_n$ be an orthonormal basis such that $\mathrm{d}_1, \ldots, \mathrm{d}_m$ are in A. The coordinates of y with respect to this basis, coefficients of orthogonal projections (compare 2.1.2) of y, have the covariance matrix $\sigma^2 \cdot I$, so that the variance of the estimator $f(\mathrm{y})$ is equal to $\sigma^2\Sigma_{i=1}^{n}\{f(\mathrm{d}_i)\}^2$. The values $f(\mathrm{d}_i)$ being fixed by the condition of unbiasedness for $i = 1, \ldots, m$, the variance of $f(\mathrm{y})$ will be minimized by taking $f(\mathrm{d}_i) = 0$ for $i > m$. It follows immediately that the corresponding projection P, discussed in the last section, will then be orthogonal (with respect to the metric V^{-1}).

It is remarkable that this same orthogonal projection on A can be used for *any* estimable function. We observe: the most efficient unbiased estimate of the estimable function $\Sigma p_j\beta_j = \Sigma g(\mathrm{x}_j)\beta_j = g(\Sigma\beta_j\mathrm{x}_j) = g(E\mathrm{y})$ is $g(\mathrm{y}_\mathrm{A})$. The orthogonal projection y_A has the property that, for any linear function g on A, $g(\mathrm{y}_\mathrm{A})$ is the most efficient unbiased estimator of $g(E\mathrm{y})$. We also express this fact by saying: y_A is the most efficient unbiased estimator of $E\mathrm{y}$.

The numbers β_j as well as their estimates b_j in $\mathrm{y}_\mathrm{A} = b_1\mathrm{x}_1 + \ldots + b_k\mathrm{x}_k$ are called regression coefficients.

2.3.4. *Further remarks*

In contrast to the vector y_A the b_j are not necessarily unique, namely if $x_1, \ldots, x_k$ are dependent. In that case the normal equations in $b_1, \ldots, b_k$ have not a unique solution; for then the quadratic form, $(\lambda_1 x_1 + \ldots + \lambda_k x_k)^2$, in $\lambda_1, \ldots, \lambda_k$, of which the matrix is the matrix of coefficients of the normal equations, is non-negative (positive definite in case the x_j are independent). One way out of that difficulty is omission of dependent vectors x_j such that the remaining set is independent and spans A. However, when one still wants to maintain the whole set $b_1, \ldots, b_k$ (for purposes of computation and study of the covariance matrix of these regression coefficients), it is customary (KEMPTHORNE [9]) to fit the system of k normal equations in k variables into a system of $(k+m)$ independent equations in $(k+m)$ variables, such that the normal equations are obtained again by putting the new variables equal to zero in the first k equations.

We will not use normal equations in those cases, so that such method is not necessary. Abandoning the normal equations, however, means abandoning the covariance matrix of the b_j, which is the only drawback of the following exposition against other great advantages.

The square of the perpendicular from y on A, $(y - y_A)^2$, is equal to a quadratic form in the differences between corresponding coordinates of y and of the estimation of Ey (in the case of a covariance matrix $\sigma^2 \cdot I$ for y the quadratic form is the sum of squares of these differences). Because $(y - y_A)^2$ is the square of the shortest distance between y and A, the method of estimation of $E(y)$ is named the method of least squares or generally the method of the least value of a quadratic form.

In the particular case that y has a normal distribution, the estimator of Ey in A is such that $(y - Ey)^2$, which occurs in the exponent of the probability density of y, is minimal, in other words, such that the probability density of y is maximal. Hence it is the maximum likelihood estimator then.

Whether y is normally distributed or not, the expectation of $(y - y_A)^2$ is equal to $(n - m)\sigma^2$, so that an unbiased estimate of σ^2 is $(y - y_A)^2/(n - m)$. The calculation of $(y - y_A)^2$ is often simplified by the following: we know $(y - y_A)^2 = y^2 - y_A^2$ (Pythagoras); $(y - y_A)^2 = y(y - y_A) - y_A(y - y_A) =$ (the second term vanishes because of the orthogonality of y_A and $y - y_A) = y(y - y_A) = y^2 - yy_A$. So we have $y_A^2 = yy_A = y(b_1 x_1 + \ldots + b_k x_k) = b_1(y, x_1) + \ldots + b_k(y, x_k)$.

2.3.5. *Tests in linear regression*

Now and whenever tests are concerned in the following, we suppose that y is normally distributed. The tests will always have the same feature, namely a null hypothesis that one or more of the β_j are zero, against the alternative that they are not. The null hypothesis means that Ey is in a subspace B of A, with dimension, say, m_1. In that case the component of y_A, orthogonal to B, has the null vector as expectation. We consider the component of y_A orthogonal to B because its square is stochastically independent of y_B. According to 1.3.5, this component, the orthogonal projection of both y and y_A on the $(m - m_1)$-dimensional residual space of A orthogonal to B, is equal to $y_A - y_B$.

Remembering the end of 2.2.1 we find that $\dfrac{(y_A - y_B)^2/(m - m_1)}{(y - y_A)^2/(n - m)}$ is the test statistic, which has a $F(m - m_1; n - m)$ distribution under the null hypothesis.

Here we use the fact that $y - y_A$, which is orthogonal to A and thus to y_B and $y_A - y_B$, is independent of the last two components. $(y_A - y_B)^2$ will be determined either by direct orthogonal projection of y on the residual space of A orthogonal to B, or as $y_A^2 - y_B^2$, or as the difference of the squares of the perpendiculars $y - y_A$ and $y - y_B$ i.e. $y^2 - y_B^2 - (y^2 - y_A^2)$, or, as we will see, by means of quantities used in the performance of the orthogonal projection on A.

2.3.6. *Conditional observations*

The first problem is to obtain the best estimate of Ey from the observation of a random vector y with covariance matrix $\sigma^2 \cdot I$, while the Ey_i satisfy p linear conditions of the form: $a_{j1}Ey_1 + a_{j2}Ey_2 + \ldots + a_{jn}Ey_n = a_{jo}$; $(j = 1, \ldots, p)$.

Any of these conditions means that the difference between any pair of vectors satisfying the relation is orthogonal (with respect to the metric I) to the vector formed by the coefficients in that relation, say a_j. The relations, as a whole, imply that the difference between any pair of admitted vectors is orthogonal to the vectors $a_1, \ldots, a_p$, i.e. to the space $A^{\perp}$ spanned by these vectors. In other words, such a difference must belong to the space A with dimension, say, m and orthogonal to $A^{\perp}$. The set of admitted vectors for Ey will be found by adding some fixed vector x_o to A. In other words, the difference between Ey and x_o is in A. But now with this formulation the problem is reduced to the regression problem.

The best (from now on used for "unbiased and most efficient") estimate of $Ey - x_o$ is found by orthogonal projection of $y - x_o$ on A. This is done by choosing a vector in $A^{\perp}$: $\lambda_1 a_1 + \ldots + \lambda_p a_p$, such that $y - x_o - (\lambda_1 a_1 + \ldots + \lambda_p a_p)$ is in A; this is equivalent with the condition that the vector $y - (\lambda_1 a_1 + \ldots + \lambda_p a_p)$ is in the admitted set, i.e. satisfies the given conditions for Ey.

We find by substitution: $\lambda_1(a_1, a_j) + \ldots + \lambda_p(a_p, a_j) = (y, a_j) - a_{jo}$ for $j = 1, \ldots, p$. These equations for the so-called correlates λ_j can be solved, if $a_1, \ldots, a_p$ are independent. The square of the perpendicular $(\lambda_1 a_1 + \ldots + \lambda_p a_p)^2 = \Sigma_{j=1}^{p} \lambda_j\{\lambda_1(a_1, a_j) + \ldots + \lambda_p(a_p, a_j)\} = \Sigma_{j=1}^{p} \lambda_j\{(y, a_j) - a_{jo}\}$, divided by σ^2, has a $n - m = p$ dimensional chi-square distribution again.

The only point of difference between the first and the following problem is that the covariance matrix of y is $\sigma^2 \cdot V$ now. We then introduce the metric V^{-1}. Then the left hand sides of the conditions are inner products of the vectors Ey and Va_j. To obtain the best estimate of Ey we subtract a linear combination of the vectors Va_j from y such that the difference satisfies the conditions. The equations for λ_j are:

$\lambda_1(Va_1, a_j) + \ldots + \lambda_p(Va_p, a_j) = (y, a_j) - a_{jo}$ for $j = 1, \ldots, p$, where (Va_i, a_j) represents the inner product of a_i and a_j with metric V, and (y, a_j) the inner product of y and a_j with metric I. The square of the perpendicular is now $(\lambda_1 Va_1 + \ldots + \lambda_p Va_p)^2$ with metric V^{-1} which is equal to

$$\Sigma_{j=1}^{p} \lambda_j\{\lambda_1(Va_1, a_j) + \ldots + \lambda_p(Va_p, a_j)\}$$

with metric I or $\Sigma_{j=1}^{p} \lambda_j\{(y, a_j) - a_{jo}\}$ likewise with metric I.

The problem of linear regression and that of conditional observations are not essentially different and the one can be reduced to the other. The first problem (and method) will be preferred if $n - m > m$, the second if $n - m < m$. The translation of the one problem in the other is as follows. Suppose we have a linear regression problem. Then from some of the equations that express the Ey_i in the β_j, these β_j can be solved, and substituted in the remaining equations. In this way linear relations between the Ey_i are obtained. On the other hand, if the problem is in terms of conditional observations, then it is possible to express a part of all Ey_i in the remaining Ey_i, while the last are expressed in themselves. Consider now the remaining Ey_i, as the regression coefficients in the first problem.

CHAPTER 3

APPLICATION OF LINEAR REGRESSION WITH UNCORRELATED OBSERVATIONS

GENERAL REMARKS

In this chapter the random vector $\mathrm{y} = (y_1, \ldots, y_n)$ will have a diagonal covariance matrix $\sigma^2 \cdot D$, i.e. a matrix with elements d_{ij} for which $d_{ij} = 0$ for $i \neq j$. Then the appropriate metric D^{-1} will be diagonal too. The elements in the diagonal of D^{-1} are the reciprocals of the corresponding elements in D. These reciprocals (denoted by g_i or w_i) are called the weights of the coordinates. The inner product of the vectors $\mathrm{x} = (x_1, \ldots, x_n)$ and $\mathrm{y} = (y_1, \ldots, y_n)$ is then $\Sigma_{i=1}^{n} w_i x_i y_i$, a so-called weighted sum of products.

As many of the problems, to be dealt with, can easily be generalized from that with a covariance matrix $\sigma^2 \cdot I$ for y, to that with a diagonal one, by replacing the inner product $\Sigma_i x_i y_i$ by $\Sigma_i w_i x_i y_i$, we will generally assume a covariance matrix $\sigma^2 \cdot I$ for convenience, unless the reverse is declared explicitly. In each case it will be indicated whether the generalization, expressed in the last sentence, is possible or not.

3.1. SOME REGRESSION PROBLEMS WITH BASES EXPLICITLY GIVEN

3.1.1. *Level*

Let Ey_i be equal to a constant β for $i = 1, \ldots, n$, or in terms of vectors, let $E\mathrm{y}$ be equal to $\beta\mathrm{r}$, where $\mathrm{r} = (1, 1, \ldots, 1)$. The space N spanned by r will be called the space of levels.

To obtain the best estimate b of β, y will be projected orthogonally on this one-dimensional space N. We obtain (compare 1.3.5 and 1.5.2) $b\mathrm{r}$ with $b = \dfrac{\mathrm{yr}}{\mathrm{rr}} = \dfrac{y_1 + \ldots + y_n}{1 + \ldots + 1}$ or $\bar{y}$, the average of $y_1, \ldots, y_n$. The square of the perpendicular is $(\mathrm{y} - b\mathrm{r})^2 = \mathrm{y}^2 - (b\mathrm{r})^2 = \mathrm{y}^2 - b^2(\mathrm{r}, \mathrm{r}) = \mathrm{y}^2 - \dfrac{(\mathrm{y}, \mathrm{r})^2}{(\mathrm{r}, \mathrm{r})^2}(\mathrm{r}, \mathrm{r}) = \mathrm{y}^2 - \dfrac{(\mathrm{y}, \mathrm{r})^2}{(\mathrm{r}, \mathrm{r})} = \Sigma_i y_i^2 - \dfrac{(\Sigma_i y_i)^2}{n}$. Division of this square by the dimension of the subspace of E orthogonal to N namely $n - 1$ will give an unbiased estimate of σ^2.

To test the null hypothesis $\beta = 0$ we use the statistic $\dfrac{(\mathrm{br})^2}{(\mathrm{y} - \mathrm{br})^2/(n-1)} = \dfrac{(\Sigma y_i)^2/n}{\{\Sigma y_i^2 - (\Sigma y_i)^2/n\}/n(n-1)}$ which under the null hypothesis has a $F(1; n-1)$ distribution. It is the square of Student's *t*-statistic.

If the metric is diagonal, we obtain with the weights g_i: $b = \frac{(\mathrm{y}, \mathrm{r})}{(\mathrm{r}, \mathrm{r})} = \frac{g_1y_1 + \ldots + g_ny_n}{g_1 + \ldots\ldots + g_n}$ which is called the weighted mean of $y_1, \ldots, y_n$.

In that case $(\mathrm{y} - b\mathrm{r})^2 = \Sigma g_i y_i^2 - \frac{(\Sigma g_i y_i)^2}{\Sigma g_i}$.

3.1.2. *A linear function in one variable*

Let $Ey_i = \beta_o + \beta_1 x_i$; in words, the expectation of the interesting property of the individuals is a linear function of another property x, likewise quantitatively expressible. In terms of vectors $E\mathrm{y} = \beta_o \mathrm{r} + \beta_1 \mathrm{x}$.

To estimate β_o and β_1 we wish to project y on the two-dimensional space spanned by r and x. Because r and x in general are not orthogonal, we prefer (compare 1.5.2) to orthogonalize the basis first. For this purpose we replace x by its component x′ orthogonal to r. As we know that the orthogonal projection of x on N is $\frac{(\mathrm{xr})}{(\mathrm{rr})}\mathrm{r}$, we obtain $\mathrm{x}' = \mathrm{x} - \mathrm{x}_\mathrm{N} = \mathrm{x} - \frac{(\mathrm{xr})}{(\mathrm{rr})}\mathrm{r} = \mathrm{x} - \bar{x}\mathrm{r}$, which means that all the x_i are diminished with their average. The orthogonal projection of y on N is $\mathrm{y} - \mathrm{y}' = \mathrm{y}_\mathrm{N} = \frac{\mathrm{yr}}{\mathrm{rr}}\mathrm{r}$. The orthogonal projection of $\mathrm{y} = \mathrm{y}' + (\mathrm{y} - \mathrm{y}')$ on x′ is equal to that of y′ on x′, because $\mathrm{y} - \mathrm{y}' = \mathrm{y}_\mathrm{N}$ is orthogonal to x′. It is therefore $\frac{\mathrm{x}'\mathrm{y}'}{\mathrm{x}'\mathrm{x}'}\mathrm{x}'$.

The estimate of $E\mathrm{y}$ is thus $\mathrm{y}_\mathrm{N} + \frac{\mathrm{x}'\mathrm{y}'}{\mathrm{x}'\mathrm{x}'}\mathrm{x}' = (\mathrm{y}_\mathrm{N} - \frac{\mathrm{x}'\mathrm{y}'}{\mathrm{x}'\mathrm{x}'}\mathrm{x}_\mathrm{N}) + \frac{\mathrm{x}'\mathrm{y}'}{\mathrm{x}'\mathrm{x}'}\mathrm{x}$, from which follows that the estimate of β_1 is $b_1 = \frac{\mathrm{x}'\mathrm{y}'}{\mathrm{x}'\mathrm{x}'}$ and that of β_o is $b_o = \frac{\mathrm{yr}}{\mathrm{rr}} - b_1\frac{\mathrm{xr}}{\mathrm{rr}} = \bar{y} - b_1\bar{x}$. To evaluate b_1 we consider first $\mathrm{x}'\mathrm{x}' = (\mathrm{x} - \mathrm{x}_\mathrm{N})^2$ which according to the foregoing section is equal to $\mathrm{x}^2 - \mathrm{x}_\mathrm{N}^2$. Recalling that x_N is a vector which consists of n times the number $\bar{x}$, we find that $\mathrm{x}_\mathrm{N}^2 = \frac{(\Sigma x_i)^2}{n}$. In order to compute $\mathrm{x}'\mathrm{y}' = (\mathrm{x} - \mathrm{x}_\mathrm{N})(\mathrm{y} - \mathrm{y}_\mathrm{N})$ we use a lemma analogous to the Pythagorean theorem: if A and B are orthogonal spaces then $\mathrm{x}_\mathrm{A}\mathrm{y}_\mathrm{B} = \mathrm{x}_\mathrm{B}\mathrm{y}_\mathrm{A} = 0$, hence $(\mathrm{x}_\mathrm{A} + \mathrm{x}_\mathrm{B})(\mathrm{y}_\mathrm{A} + \mathrm{y}_\mathrm{B}) = \mathrm{x}_\mathrm{A}\mathrm{y}_\mathrm{A} + \mathrm{x}_\mathrm{B}\mathrm{y}_\mathrm{B}$.

We apply this to the case that A is N and B is the subspace orthogonal to N. We have $\mathrm{xy} = \mathrm{x}_\mathrm{N}\mathrm{y}_\mathrm{N} + (\mathrm{x} - \mathrm{x}_\mathrm{N})(\mathrm{y} - \mathrm{y}_\mathrm{N})$, so that $\mathrm{x}'\mathrm{y}' = \mathrm{xy} - \mathrm{x}_\mathrm{N}\mathrm{y}_\mathrm{N}$. Further

$$\mathrm{x}_\mathrm{N}\mathrm{y}_\mathrm{N} = n\left(\frac{\Sigma x_i}{n}\right)\left(\frac{\Sigma y_i}{n}\right) = \frac{(\Sigma x_i)(\Sigma y_i)}{n}$$

The vector y is decomposed in three orthogonal components namely y_N, $b_1\mathrm{x}'$ and $\mathrm{y}' - b_1\mathrm{x}' = \mathrm{y} - \mathrm{y}_\mathrm{N} - b_1\mathrm{x}'$. The square of the last component i.e. the perpendicular from y on the space, spanned by r and x, is equal to $\mathrm{y}^2 - \mathrm{y}_\mathrm{N}^2 - (b_1\mathrm{x}')^2$. Dividing this square by the dimension $n - 2$ yields an unbiased estimate of σ^2. Further the square of $b_1\mathrm{x}'$ will be the numerator of the statistic for the test of the null hypothesis that $E\mathrm{y}$ is in N, in other words, that $\beta_1 = 0$. This is the second reason why we prefer a (this) orthogonal basis: x′ is basis of the residual space, of the space spanned by r and x, orthogonal to r. Compare 2.3.5. The square of $b_1 x'$ is equal to $\frac{(x'y')^2}{(x'x')}$.

The generalization to the case of a diagonal covariance matrix presents no difficulties.

3.1.3. *Linear functions in several variables*

The regression problem with several variables has already been considered in fact in 2.3. Therefore we confine ourselves to some remarks.

In the supposition $Ey = \beta_1 x_1 + \ldots + \beta_k x_k$ each of the variables $x_1, \ldots, x_k$ represents the value of a quantitatively expressible property of the individual. Such a variable is often the function value of a given function φ in the values of one or more other such properties of the individual. In case Ey is supposed to be e.g. a polynomial of degree three in two variables x_1 and x_2 with unknown coefficients, $x_1^2 x_2$ is one of such variables. The vector of function values (in this case of the function $x_1^2 x_2$) for the n individuals is one of the basisvectors of the space in which E_J is supposed to be. The notion "function φ" should be understood in the broadest sense of the word: functions given not only by a formula in the values of the properties, but also by tables or graphs (obtained e.g. from previous experiments) may be considered. (KUIPER [14]).

Mostly the set of vectors x_j will contain the vector r, which corresponds to an unknown constant term in the supposition about Ey. Because the value of this constant is not of interest in general, the whole procedure of estimating and testing will, analogously to the last section, be performed in the residual space of E orthogonal to N (the space of contrasts). The determination of the projection of $y - y_N = y'$ on the space, spanned by the vectors $x_j - (x_j)_N = x_j'$, by normal equations requires quantities of the form $y'y'$, $y'x_j'$ and $x_j'x_j'$, of which the calculation has been discussed in the last section. The square of a projection will be obtained according to the end of 2.3.4. Testing some β_j requires the projection on a subspace i.e. solution of a new set of normal equations with coefficients, however, that also occur in the first set.

Another way of performing the projections will be by complete orthogonalization of the basis of the space A, spanned by the vectors x_j. This, however, is in general at least as cumbersome as the solution of normal equations.

Introduction of weights is no essential difficulty.

3.1.4. *Orthogonal polynomials*

For the particular case in which Ey is supposed to be a polynomial in one variable x, and the n values of x are equidistant, and the vector y has the covariance matrix $\sigma^2 \cdot I$, the result of the orthogonalization process, alluded to at the end of the last section, is fortunately given in tables (FISHER and YATES [6], PEARSON and HARTLEY [17], ANDERSON and HOUSEMAN [1], DELURY [5]). The functions u^j $(j = 0, 1, \ldots, k)$ are defined (compare 1.3.4) on $u = -n+1, -n+3, -n+5, \ldots, -1, 1, 3, \ldots, n-1$, when n is even, and on $u = -\frac{1}{2}(n-1), -\frac{1}{2}(n-3), \ldots, -2, -1, 0, 1, 2, \ldots, \frac{1}{2}(n-1)$ when n is odd. It follows that the set of function values of u^j is symmetrical if j is even, and asymmetrical if j is odd, so that the vectors of function values for j even are orthogonal to those for j odd.

These vectors of function values have been orthogonalized in the order of increasing j. In every step of this process either only even or only odd functions have been used. It follows that the orthogonalized vector is a symmetric or an antisymmetric set of numbers in case j is even or odd respectively. As a consequence only $\frac{1}{2}n$ (or $\frac{1}{2}n+\frac{1}{2}$) coordinates are recorded in fact. The length of the orthogonalized vector is not 1, but as small as possible and such that the coordinates are integers. The squares of these lengths are also noted in the table.

In order to find of which linear combination of the functions $u^j, u^{j-1}, \ldots, u, 1$ some recorded vector is the set of function values, – it is regrettable that this is not noted in the tables –, one may solve linear equations in the unknown coefficients, obtained by substitution of some u in that combination, and equating to the corresponding known function value. The fact that such a combination is either even or odd in u, and that the coefficient of the highest degree term is noted in the tables, simplifies this procedure.

Let the functions corresponding to the orthogonalized vectors $\mathrm{r}, \mathrm{v}_1, \mathrm{v}_2, \ldots$ be $1, u, au^2+b, cu^3+du$ etc. Let the average of the values x_i of the original variable x for the n individuals be $\bar{x}$ and the distance between two successive values of x_i be p. Then we take $u = \dfrac{x-\bar{x}}{p}$. Let the coefficients in the orthogonal projection be $b_0^* = \dfrac{\mathrm{yr}}{\mathrm{rr}}, b_1^* = \dfrac{\mathrm{yv}_1}{\mathrm{v}_1\mathrm{v}_1}, b_2^* = \dfrac{\mathrm{yv}_2}{\mathrm{v}_2\mathrm{v}_2}$ etc. The regression coefficients $b_0, b_1, b_2, \ldots$ of the polynomial in x will be obtained by reduction of:

$$b_0^* + b_1^*\left(\frac{x-\bar{x}}{p}\right) + b_2^*\left\{a\left(\frac{x-\bar{x}}{p}\right)^2 + b\right\} + b_3^*\left\{c\left(\frac{x-\bar{x}}{p}\right)^3 + d\left(\frac{x-\bar{x}}{p}\right)\right\} + \ldots$$

The square of the separate projection $\dfrac{(\mathrm{yv}_j)^2}{\mathrm{v}_j^2}$ serves as numerator of the statistic for the test that the regression coefficient β_j^* of v_j is zero, which is equivalent with the test that the coefficient β_j of x^j is zero, independently of whether the coefficients of the functions of *smaller* degree are zero or not. This consequence is the justification of the chosen order of orthogonalization.

3.2. REGRESSION PROBLEMS BASED ON A CLASSIFICATION

3.2.1. *Main effects*

Let the coordinates of the random vector y be grouped by virtue of some characteristic of the n individuals in a number, say k, of not necessarily equally large classes. Let it be supposed that the expectations of the coordinates within such a class are equal (but unknown). Then the set of vectors, to which Ey must belong, is a vector space. (Compare KUIPER [13]). Consider the k vectors which consist of ones in only one of the k classes and of zeros elsewhere. Every linear combination of these k vectors belongs to the set, and every element of the set can be written as a linear combination of these k vectors. For example:

$$\begin{bmatrix}\alpha\alpha\alpha\alpha \\ \beta\beta \\ \gamma\gamma\gamma \\ \delta\delta\delta\delta\delta\end{bmatrix} = \alpha\begin{bmatrix}1111 \\ 00 \\ 000 \\ 00000\end{bmatrix} + \beta\begin{bmatrix}0000 \\ 11 \\ 000 \\ 00000\end{bmatrix} + \gamma\begin{bmatrix}0000 \\ 00 \\ 111 \\ 00000\end{bmatrix} + \delta\begin{bmatrix}0000 \\ 00 \\ 000 \\ 11111\end{bmatrix}$$

Here the n-tuples are arranged according to convenient arrays. "Homologous" numbers are corresponding coordinates.

Simple inspection shows that the mentioned k vectors are independent. It follows that these vectors form a basis of the considered space, so that the dimension of that space is equal to the number of classes. Let the characteristic of classification be named A. Then also the classification will be named A. Further the k-dimensional subspace of vectors with equal coordinates within the classes is named A too and it is called the space of main effects (of A). The term effect of the characteristic should not be understood in the limited sense by which in everyday life the effect of a cause is meant, but it has the looser sense that different values of the expectation of the considered property y always go together with differences, qualitative or quantitative, in the characteristic A.

It is clear that the subspace N of levels is a subspace of A. The residual space of A orthogonal to N is called the space of pure main effects. It will be denoted by A* and it has dimension $k-1$. The reason of introducing this, is that the differences of the coordinates of vectors in A do interest us in fact. Two vectors in A of which corresponding coordinates differ by the same amount or, in other words, two vectors in A which differ by a vector in N are considered to represent the same effect of the classification A. The reason that the space A* of pure main effects, which together with N span A, is chosen orthogonal to N is that orthogonal projections are needed for tests (if Ey is in N, then $(E\mathrm{y})_{\mathrm{A}^*} = 0$. From the point of view of estimation any other choice of a $(k-1)$-dimensional subspace of A, which together with N spans A, is permissible.

3.2.2. *Estimation and test*

The best estimate of Ey is the orthogonal projection of y on A, $\mathrm{y_A}$. The considered basis of A happens to be orthogonal (also when the metric is diagonal and weights are used). The projection on A is thus equal to the sum of the projections on the one-dimensional spaces spanned by these basisvectors separately. Let such a basis vector be a. Then the coefficient in the projection is $\frac{\mathrm{ya}}{\mathrm{aa}}$. Here the numerator is equal to the sum of those coordinates of y, which are in the same class, as that in which a contains ones. The denominator is equal to the number of elements in that class. The coefficient is thus equal to the class average $\bar{y}_{i}$. In the mentioned example

$$\mathrm{y_A} = \begin{bmatrix} \bar{y}_{1.} & \bar{y}_{1.} & \bar{y}_{1.} & \bar{y}_{1.} & \\ \bar{y}_{2.} & \bar{y}_{2.} & & & \\ \bar{y}_{3.} & \bar{y}_{3.} & \bar{y}_{3.} & & \\ \bar{y}_{4.} & \bar{y}_{4.} & \bar{y}_{4.} & \bar{y}_{4.} & \bar{y}_{4.} \end{bmatrix}.$$

Further $\mathrm{y}_{\mathrm{A}}^2$ = (if the number of elements in class i is n_i)

$$= \Sigma_i n_i \bar{y}_{i.}^2 = \Sigma_i n_i \left(\frac{\Sigma_j y_{ij}}{n_i}\right)^2 = \Sigma_i \frac{(\Sigma_j y_{ij})^2}{n_i},$$

where y_{ij} is the j-th coordinate in class i. The orthogonal projection on A* cannot be performed easily directly in general. In our example the vectors

$$\begin{bmatrix} 5 & 5 & 5 & 5 & \\ 0 & 0 & & & \\ 0 & 0 & 0 & & \\ -4 & -4 & -4 & -4 & -4 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 & 0 & \\ 5 & 5 & & & \\ 0 & 0 & 0 & & \\ -2 & -2 & -2 & -2 & -2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 0 & 0 & 0 & \\ 0 & 0 & & & \\ 5 & 5 & 5 & & \\ -3 & -3 & -3 & -3 & -3 \end{bmatrix}$$

form a basis of this space indeed, but they are not orthogonal. However, as A* is orthogonal to N, $\mathrm{y}_{\mathrm{A}^*} = \mathrm{y_A} - \mathrm{y_N}$.

$$\mathbf{y}_{A^*} = \begin{bmatrix} \bar{y}_{1.}-\bar{y} & \bar{y}_{1.}-\bar{y} & \bar{y}_{1.}-\bar{y} & \bar{y}_{1.}-\bar{y} & \\ \bar{y}_{2.}-\bar{y} & \bar{y}_{2.}-\bar{y} & & & \\ \bar{y}_{3.}-\bar{y} & \bar{y}_{3.}-\bar{y} & \bar{y}_{3.}-\bar{y} & & \\ \bar{y}_{4.}-\bar{y} & \bar{y}_{4.}-\bar{y} & \bar{y}_{4.}-\bar{y} & \bar{y}_{4.}-\bar{y} & \bar{y}_{4.}-\bar{y} \end{bmatrix}$$

The vector $\mathbf{y}$ is thus decomposed in three orthogonal components $\mathbf{y}_N$, $\mathbf{y}_A-\mathbf{y}_N$ and $\mathbf{y}-\mathbf{y}_A$ in spaces with dimension 1, $k-1$, $n-k$ respectively. The square of $\mathbf{y}-\mathbf{y}_A$ which is equal to $\mathbf{y}^2-\mathbf{y}_A^2$, divided by the dimension $n-k$, is an unbiased estimate of σ^2, and is the denominator for the F-test criterion of the hypothesis that $E\mathbf{y}$ is in N. The numerator will be $\mathbf{y}_{A*}^2/(k-1)$ while $\mathbf{y}_{A*}^2=\mathbf{y}_A^2-\mathbf{y}_N^2$. The calculation of $\mathbf{y}_N^2$ is considered in 3.1.1.

If weights g_{ij} are used the coordinates of $\mathbf{y}_A$ are weighted class averages, and

$$\mathbf{y}_A^2 = \Sigma_i \frac{(\Sigma_j g_{ij} y_{ij})^2}{\Sigma_j g_{ij}}.$$

The remaining quantities are known from 3.1.1.

3.2.3. *Components of main effects*

Sometimes one is interested in the estimation and (or) testing of components of the main effects. These will be in subspaces of A*. The three mentioned basis vectors of A* e.g. span one-dimensional spaces representing the difference in effect of the first, the second and the third class respectively with respect to that of the fourth class.

The estimation of these effects is fairly simple: if

$$\mathbf{y}_{A^*} = \begin{bmatrix} \alpha & \alpha & \alpha & \alpha & \\ \beta & \beta & & & \\ \gamma & \gamma & \gamma & & \\ \delta & \delta & \delta & \delta & \delta \end{bmatrix},$$

then the first component is

$$\begin{bmatrix} \alpha & \alpha & \alpha & \alpha & \\ 0 & 0 & & & \\ 0 & 0 & 0 & & \\ -\frac{4}{5}\alpha & -\frac{4}{5}\alpha & -\frac{4}{5}\alpha & -\frac{4}{5}\alpha & -\frac{4}{5}\alpha \end{bmatrix} \quad \text{etc.}$$

The test of, say, the first component is not so simple, because the components are not orthogonal. Under the null hypothesis that there is no difference in effect between the first and the fourth class, $E\mathbf{y}$ is of the form

$$\begin{bmatrix} \alpha & \alpha & \alpha & \alpha & \\ \beta & \beta & & & \\ \gamma & \gamma & \gamma & & \\ \alpha & \alpha & \alpha & \alpha & \alpha \end{bmatrix}.$$

This means that the first and the fourth class have been united to one class. The orthogonal projection of $\mathbf{y}$ on the corresponding subspace of A contains class averages again, and differs with respect to $\mathbf{y}_A$ only in the former first and fourth class. The difference of the squares, necessary in the numerator of F, is thus equal to

$$\frac{(\Sigma_j y_{1j})^2}{n_1}+\frac{(\Sigma_j y_{4j})^2}{n_4}-\frac{(\Sigma_j y_{1j}+\Sigma_j y_{4j})^2}{n_1+n_4}.$$

The test is equivalent to Student's two-sided two sample test.

Another trio of components, which has the advantage to be orthogonal, is

$$\begin{bmatrix} 1 & 1 & 1 & 1 & \\ -2 & -2 & & & \\ 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 & 1 & 1 & \\ 1 & 1 & & & \\ -3 & -3 & -3 & & \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 5 & 5 & 5 & 5 & \\ 5 & 5 & & & \\ 5 & 5 & 5 & & \\ -9 & -9 & -9 & -9 & -9 \end{bmatrix}$$

They represent the differences in effect of the first and the second class, of the first and the second class together with respect to the third class, and of the first three classes together with respect to the fourth class respectively. The coefficient of the last vector in the orthogonal projection e.g. is equal to

$$\frac{5(\Sigma_j y_{1j} + \Sigma_j y_{2j} + \Sigma_j y_{3j}) - 9\,\Sigma_j y_{4j}}{9 \cdot 5^2 + 5 \cdot 9^2}$$

and the square of that projection is equal to:

$$\frac{[5\,(\Sigma_j y_{1j} + \Sigma_j y_{2j} + \Sigma_j y_{3j}) - 9\,\Sigma_j y_{4j}]^2}{9 \cdot 5^2 + 5 \cdot 9^2}$$

Another orthogonal trio of components is

$$\begin{bmatrix} 1 & 1 & 1 & 1 & \\ -2 & -2 & & & \\ 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 0 & 0 & \\ 0 & 0 & & & \\ 5 & 5 & 5 & & \\ -3 & -3 & -3 & -3 & -3 \end{bmatrix} \qquad \begin{bmatrix} 4 & 4 & 4 & 4 & \\ 4 & 4 & & & \\ -3 & -3 & -3 & & \\ -3 & -3 & -3 & -3 & -3 \end{bmatrix}$$

which represent the differences in effect of the first and the second class, of the third and the fourth class, and of the first and the second class together with respect to the third and the fourth class together.

When the classification is based on a quantitative characteristic x and the classes are equally large, then by means of orthogonal polynomials appropriate components may be chosen. Let the four classes have e.g. three elements each and let the values of x be equidistant, then we may choose the trio

$$\begin{bmatrix} -3 & -3 & -3 \\ -1 & -1 & -1 \\ 1 & 1 & 1 \\ 3 & 3 & 3 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \\ 1 & 1 & 1 \end{bmatrix} \qquad \begin{bmatrix} -1 & -1 & -1 \\ 3 & 3 & 3 \\ -3 & -3 & -3 \\ 1 & 1 & 1 \end{bmatrix}$$

which represent the spaces of linear, quadratic and cubic components of the effect of A. Sometimes it is useful to assume (if allowed) that Ey is in the space spanned by the linear and the quadratic component only. The dimension of the residual space orthogonal to A is raised by one then.

Introduction of weights requires the formation of other orthogonal components in general, while the first mentioned test can be generalized to this case along the same lines as in 3.2.2.

3.3. GENERAL REGRESSION PROBLEMS BASED ON TWO CLASSIFICATIONS

3.3.1. *Two spaces of main effects*

In addition to the classification A let the coordinates of y be grouped by virtue of a characteristic B in other classes at the same time, and for the present such that it is not possible to obtain one class of the one classification by uniting some classes of the other. Let e.g. in our example in 3.2.1 a second classification according to columns be introduced as follows:

	B		
A	×	×	× ×
	×	×	
	×	×	×
	× × ×	×	×

Corresponding to these two classifications we have two spaces of main effects A and B, and two spaces of pure main effects A* and B*. The spaces A and B have the space of levels in common. Now consider the basis vectors of A and B that consist of ones and zeros (compare 3.2.1). Because both classes of any pair of classes of the one classification (say *A*) contain some coordinates that, together, belong to some class of the other classification (say *B*), any linear combination of the basis vectors of A that is in B has all coefficients equal, and thus must be in N. It follows that N is the intersection of A and B, and that A* and B* are disjoint. If A and B have k_1 and k_2 classes respectively, the dimensions of A* and B* are $k_1 - 1$ and $k_2 - 1$.

The supposition that Ey is in the $(k_1 + k_2 - 1)$-dimensional space spanned by N, A* and B* implies that every Ey_i is considered as the sum of a general constant, a constant for the corresponding class of *A* and a constant for the corresponding class of *B*. In other words, the effects of *A* and *B* are considered as additive.

3.3.2. *Estimation of the main effects*

In order to obtain the best estimate of Ey and with that, the best estimate of the effects, y must be projected orthogonally on the space spanned by A and B. Because A* and B* in general are not orthogonal on the one hand, but orthogonal bases for A and B are known on which the orthogonal projection is technically simple, namely by taking averages, on the other hand, the iterative method of 1.5.6 with $k = 2$ will be used. Compare Yates (22), Hamming (8) and Kuiper (11).

The calculation of the following sequence of vectors is necessary: $u_1 = y_A$; $v_1 = y_B - y_{AB}$; $u_2 = (v_1)_A$; $v_2 = (u_2)_B$; $u_3 = (v_2)_A$; $v_3 = (u_3)_B$ and so on. The sum of the vectors v will yield a component in B, and u_1, diminished with the sum of the remaining vectors u, will give a component in A. These together form the orthogonal projection of y on the space A + B. Apart from the vector y the vectors will not be written in full but only one coordinate of a class of *A* or *B* will be noted. So we obtain the following computational scheme for an observation of the example in 3.3.1:

				number	sum	u_1	u_2	u_3	u_4	u_5	u_6	$y_{sA} = u_1 - \sum_{i=2}^{6} u_i$
	24.6	20.0	18.0 19.6	4	82.2	20.550	–0.498	–0.071	–0.010	–0.001	0.000	21.130
	24.1	30.9		2	55.0	27.500	0.812	0.094	0.012	0.002	0.000	26.580
	20.6	19.8	15.8	3	56.2	18.733	–0.061	–0.016	–0.002	0.000	0.000	18.812
	26.4 25.3 28.1	27.8	25.9	5	133.5	26.700	0.111	0.030	0.005	0.001	0.001	26.552
number	6	4	4	14								
sum	149.1	98.5	79.3		326.9							
y_B	24.850	24.625	19.825									
y_{AB}	24.480	23.371	21.633	check			0.004	0.006	0.003	0.005	0.005	
v_1	0.370	1.254	–1.808	0.004								
v_2	0.098	0.091	–0.236	0.008								
v_3	0.016	0.009	–0.032	0.004								
v_4	0.002	0.001	–0.004	0.000								
v_5	0.001	0.000	0.000	0.006								
$\longrightarrow$	0.487	1.355	–2.080									

$y_{sB*} = \sum_{i=1}^{5} v_i$

This will be self-explanatory for the greater part. The row and the column "numbers" are noted, because they occur as divisors, the sums of the classes and the general total as a check. The number (coordinate) of y_{AB} on the left, 24.480, has been found from $u_1 = y_A$ as $20.550 + 27.500 + 18.733 + 3 \times 26.700$ divided by 6, and the number on the right, 21.633, as$(2 \times 20.550) + 18.733 + 26.700$ divided by 4. Subtraction yields v_1.

The process of projection (averaging) will be continued, until all the coordinates in some vector are zero, or, as a consequence of rounding errors, remain small numbers. We will return to the latter case in the following section. Finally the required vectors in A and B can be calculated in the indicated way.

It will be remarked that $v_1 = y_B - y_{AB} = y_N + y_{B*} - (y_N + y_{A*})_B = y_{B*} - (y_{A*})_B$. The second term, an orthogonal projection of a vector orthogonal to N on a space that contains N, is orthogonal to N. The same holds for v_1, u_2, v_2 etc. This has two consequences: we have a check for every step, namely the inner product of v_1, u_2 etc. with the vector r must be zero; the values of this inner product have been noted in the row and the column marked with "check". Secondly the vector found in B, $\sum_{i=1}^{\infty} v_i$, is orthogonal to N, in other words is in B*; thus the vector found in A is equal to the sum of y_N and the unique component in A*. The latter can be found by subtraction of y_N (which is calculated by means of the general total of "sums" and "numbers") from the component in A. The unique components of y in A* and B* will be denoted by y_{sA*} and y_{sB*} where the letter *s* serves to distinguish from the orthogonal projections. In this section the letters A and B can be interchanged of course. Introduction of weights implies the use of weighted averages for every class and presents no further difficulties.

3.3.3. *Some computational remarks*

Let y′ be some approximation of the projection of y on a space, say A + B, and let y′ be in the space A + B. Then the required projection of y is equal to $y' + (y - y')_{A+B}$. The approximation of y_{A+B} may be given by means of vectors in A and B. This will be applied for instance in case the approximating vectors in A and B have been obtained by the procedure of the last section, but with rounding and possibly other errors. In order to obtain a more accurate result, y − y′ should be projected according to the same procedure and the projection added to the approximation. But, as follows from that procedure, only the class sums of y − y′ are necessary. In our example the residual sum for the second column e.g. is equal to $98.500 - (4 \times 1.355 + 21.130 + 26.580 + 18.812 + 26.552) = 0.006$. If such a great accuracy is wanted in our example, we will find 0.487; 1.357; –2.081 as coordinates of y_{sB^*}, and 21.130; 26.578; 18.813; 26.553 as coordinates of y_{sA}.

A similar procedure will be followed, when the process converges very slowly. It is our experience that after several steps the coordinates of the same class all have the same sign in two successive steps u or v and decrease. At that moment an approximation of one component, say in A, will be found, by the supposition that the successive coordinates *of the same class* decrease according to a geometric series with the proportion of the coordinates in the two successive steps as ratio. (In examples to follow this is exactly true.) An approximation of the component in B will be found by projecting the difference between y and the approximate component in A on B. This can be seen thus: $y = y_{sA} + y_{sB} + y_R$ with R orthogonal to A and B; then $(y - y_{sA})_B = (y_{sB} + y_R)_B = y_{sB}$. The subtraction will take place via the sums of the classes of *B* again. Next the method described in the foregoing paragraph will be applied.

The orthogonal projection y_{A+B} is such that $y - y_{A+B}$ is orthogonal to A and to B which means that the (possibly weighted) class sums in $y - y_{A+B}$ are zero. In the particular case that a class of one classification, say *A*, occurs in only one class of the other classification *B*, the coordinate of y_{A+B} in that class of *A* is thus equal to the (weighted) average of the coordinates of y in that class. In order to calculate the required projection y_{A+B}, this class can be left out of consideration first, so that vectors in the main effect spaces of the mutilated y will be obtained. The coordinates of the effect of *B* in the mutilated vector are used as the corresponding coordinates for the complete y. The same is done for the remaining classes of *A*. The coordinate of the so far omitted class of *A* in the effect of *A* is found by subtracting the coordinate of the effect of *B* for that class from the (already known) corresponding coordinate of y_{A+B}. This procedure is mentioned because it means speeding up the convergence compared with the iterative method for the complete y. We remark that, although the vectors y_{A+B} obtained by both methods are equal of course, both vectors in A and both vectors in B differ by the same multiple of the vector r. This is a consequence of the fact that the condition of orthogonality of N and A*, and of N and B* are different for the complete and for the mutilated y. The arbitrariness of the relation between N and the pure main effect is elucidated in this way.

Finally we discuss the situation that the projection y_{A+B} has been computed, and that a new class (with observations) of *B* is added to y afterwards. (Let e.g. *A* be a classification according to varieties and *B* according to trials so that the added class of *B* represents a new trial). In order to compute y_{A+B}

quickly then, new classes of *A* in the added class of *B* are left out of consideration first. The vector in A, that had been computed already, is considered as an approximation of the required component in A. An approximation of the component in B, according to the method of subtraction mentioned for the case of a slow convergence, will yield the coordinates of the component in B already computed (so that computation is not necessary) and a coordinate for the added class of *B*. Then the method of the first paragraph is applied to obtain the effect of *A* and *B* in the mutilated y. Finally the omitted classes of *A* will be treated as mentioned in the preceding paragraph. Because the residual sums will be considerably smaller than when one performs the iterative process from the beginning for the extended y, the described method is more advantageous.

3.3.4. *Another formulation of the method of estimation*

y_{A+B} must be found as the sum of two vectors in A and in B, y_{sA} and y_{sB}, such that $y - y_{sA} - y_{sB}$ is orthogonal to A and to B. The orthogonal projection of that difference on A and on B must thus be equal to the null vector. So we have: $y_A - y_{sA} - (y_{sB})_A = 0$ and $y_B - (y_{sA})_B - y_{sB} = 0$. Subtracting the orthogonal projection of the first expression for the null vector on B from the latter yields the equation: $y_{sB} - (y_{sB})_{AB} = y_B - y_{AB}$, while the first equation may be written as $y_{sA} = y_A - (y_{sB})_A$.

Let the product of the linear transformations in E consisting of the orthogonal projection on A followed by the orthogonal projection on B, $P_B P_A$, be denoted by Q. Then the equation for y_{sB} may be written $(1 - Q)y_{sB} = y_B - y_{AB}$. Now we consider the transformation Q within the space B*. This is possible because for any x in B* (that is orthogonal to N) Qx is in B and orthogonal to N and thus in B*. In the proof of 1.5.4 we used the theorem that a similar transformation D, which was a product of orthogonal projections too, could have a bound 1, only if there would exist a vector $x_o \neq 0$ such that $Dx_o = x_o$. In our case there should exist a vector in B* such that the orthogonal projection of that vector on A should leave that vector unchanged. Because A and B* are disjoint, this is impossible. It follows that the considered transformation Q has a bound smaller than one.

From 1.4.4 we see that the sequence $S_n = \Sigma_{p=o}^{n} Q^p$ converges to the inverse of $1 - Q$. While $y_B - y_{AB}$ according to 3.3.2 is in B* too, the equation $(1 - Q)y_{sB} = y_B - y_{AB}$, with Q a transformation in B*, is solved by $y_{sB} = \Sigma_{p=o}^{\infty} Q^p(y_B - y_{AB}) = v_1 + Qv_1 + Q^2v_1 + Q^3v_1 + \dots$. Further $y_{sA} =$ $= y_A - (y_{sB})_A$.

This solution is the same as that described in 3.3.2. It is given in connection with similar solutions in chapter 4.

3.3.5. *Testing main effects or components of main effects*

For testing the null hypothesis that e.g. there is no effect of *B* we need y^2_{A+B}, $y^2_R = y^2 - y^2_{A+B}$, and $(y_{A+B} - y_A)^2$. According to the end of 2.3.4 $y^2_{A+B} = yy_{A+B} = yy_{sA} + yy_{sB}$. In order to calculate yy_{sA} for every class of A, the corresponding coordinate of y_{sA} is multiplied by the corresponding class sum in y, and the products are added. In the computational scheme of 3.3.2 these quantities are available. From the description in words it follows that y^2_{A+B} also can be written as $y_A y_{sA} + y_B y_{sB}$. The calculation, however, by means of this identity is not accurate, because small errors in the components of the orthogonal projection y_{A+B} are inflated by the multiplication.

Therefore we calculate $(y_{A+B}-y_A)^2$ directly. Recalling that $y-y_{A+B}$ is orthogonal to A and to B, and $y_{A+B}-y_A$ orthogonal to A, we have:

$$(y_{A+B}-y_A)^2 = (y_{sB}+y_{sA}-y_A)(y_{A+B}-y_A) = y_{sB}(y_{A+B}-y_A) =$$
$$y_{sB}\{y-y_A-(y-y_{A+B})\} = y_{sB}(y-y_A)$$

which, on the analogy of the foregoing remark about yy_{sA}, is equal to $y_{sB}(y-y_A)_B = v_1y_{sB}$. This expression is equal to a sum of products. For each class of *B*, such a product must be calculated; it is equal to the coordinate in y_{sB} times the coordinate in v_1 times the number of coordinates in that class.

In order to simplify this calculation the computational scheme in 3.3.2 may be altered and simplified in the following way. The vectors y_B and y_{AB} will not be calculated. Instead of them the sums in the classes of *B* in $y-y_A$ will be determined and noted. They are obtained by subtraction of the class sums in $y_A = u_1$ from the corresponding sums in y. Division by the corresponding number of coordinates gives v_1. So v_1y_{sB} can be obtained as a sum of products consisting of two factors instead of three.

In this way we find at the same time the quantity necessary in the numerator of the test statistic (which must be divided by the dimension k_2-1), and the square of the perpendicular $y^2-(y_A^2+v_1y_{sB})$, which divided by the dimension $n-k_1-k_2+1$, occurs in the denominator of the *F*-test statistic. To test the effect of *A* we need $y_{A+B}^2-y_B^2$ which will be found as $y_A^2-y_B^2+v_1y_{sB}$.

When certain components of main effects must be estimated e.g. of A, then the estimate of the effect of *A*, y_{sA^*}, can be decomposed in the same way as y_A was in the case of one classification. But testing such components will be more difficult.

By way of example we consider again the classification of 3.3.1 and in particular *A*, also discussed in 3.2.3. Suppose we are interested in the difference of effect of the first three classes together and of the fourth class on the one hand, and in the mutual differences in effect between the first three classes on the other hand. Bases of the corresponding subspaces are the already mentioned vector c on the one hand, and (for instance) d_1 and d_2 on the other hand (fig. 1).

In order to test the null hypothesis that the first three classes have no differences in effect, y must be projected orthogonally on the space spanned by c and B. When one desires to use the method of 3.3.2, c can be replaced by the subspace of A, A′, spanned by b and d (see fig. 1).

Let now v_1 be $y_{A'}-y_{BA'}$ (changing of A′ and B). From the fact that v_1 is orthogonal to N, it follows that all vectors v_i are multiples of c. The linear transformation $Q = P_{A'}P_B$ is thus a multiplication by a real number less than one, say μ. It follows that $y_{sA'}$ is equal to $(1+\mu+\mu^2+\ldots)v_1 = \frac{1}{1-\mu}v_1$, so that v_1 and v_2 (the latter to determine μ) are sufficient to know the result of the iterative process. This is generally true, if one classification contains two classes only. This case is pointed out in connection with still following particular cases, but can be treated in a simpler way.

The vector c, obtained by orthogonalizing the vector d on N, can be orthogonalized on B, which gives the same result as orthogonalizing d on B. The orthogonal projection of d on B is d_B (see fig. 1). Subtracting this vector from

$$c=\left[\begin{array}{c|c|c} 5 & 5 & \begin{matrix}5\\5\end{matrix} \\ \hline 5 & 5 & \\ \hline -9 & -9 & -9 \\ \hline \begin{matrix}-9\\-9\\-9\end{matrix} & -9 & -9 \end{array}\right];\; d_1=\left[\begin{array}{c|c|c} 1 & 1 & \begin{matrix}1\\1\end{matrix} \\ \hline -2 & -2 & \\ \hline 0 & 0 & 0 \\ \hline \begin{matrix}0\\0\\0\end{matrix} & 0 & 0 \end{array}\right];\; d_2=\left[\begin{array}{c|c|c} 1 & 1 & \begin{matrix}1\\1\end{matrix} \\ \hline 1 & 1 & \\ \hline -3 & -3 & -3 \\ \hline \begin{matrix}0\\0\\0\end{matrix} & 0 & 0 \end{array}\right];\; b=\left[\begin{array}{c|c|c} 1 & 1 & \begin{matrix}1\\1\end{matrix} \\ \hline 1 & 1 & \\ \hline 1 & 1 & 1 \\ \hline \begin{matrix}0\\0\\0\end{matrix} & 0 & 0 \end{array}\right]$$

$$d=\left[\begin{array}{c|c|c} 0 & 0 & \begin{matrix}0\\0\end{matrix} \\ \hline 0 & 0 & \\ \hline 0 & 0 & 0 \\ \hline \begin{matrix}1\\1\\1\end{matrix} & 1 & 1 \end{array}\right];\; d_B=\left[\begin{array}{c|c|c} \frac{1}{2} & \frac{1}{4} & \begin{matrix}\frac{1}{4}\\\frac{1}{4}\end{matrix} \\ \hline \frac{1}{2} & \frac{1}{4} & \\ \hline \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \hline \begin{matrix}\frac{1}{2}\\\frac{1}{2}\\\frac{1}{2}\end{matrix} & \frac{1}{4} & \frac{1}{4} \end{array}\right];\; d'=\left[\begin{array}{c|c|c} -2 & -1 & \begin{matrix}-1\\-1\end{matrix} \\ \hline -2 & -1 & \\ \hline -2 & -1 & -1 \\ \hline \begin{matrix}2\\2\\2\end{matrix} & 3 & 3 \end{array}\right]$$

Figure 1

d and multiplying by 4 to get integers, we obtain the required vector d′ (see fig. 1). The sum of the squares of the orthogonal projections of y on B and on the space spanned by d′ is the square necessary in the numerator of the test statistic.

In order to test the null hypothesis that the effect of the fourth class does not differ from that of the first three classes together, we need the orthogonal projection of y on the space spanned by B and the residual space A″ of A orthogonal to A′. To calculate this projection we may use the iterative method, now with the spaces B and A″. The only question is how to perform orthogonal projections on A″. For the considered example an orthogonal basis of A″ has been given; use of this basis would imply for every orthogonal projection the calculation of two coefficients and summation of corresponding coordinates for three classes of *A*.

When the number of classes is larger, it will be simpler to calculate the projection on A″ as the difference between the projection on A and that on A′; such projections require only averaging. Application of this method in the considered example shows that every projection on A″ only pertains the first three classes: the coordinate for the fourth class will be zero. The coordinate of the other classes will be equal to the average of that class diminished with the average of the three classes together.

Introduction of weights gives no difficulties; averages will be replaced by weighted averages.

3.4. PARTICULAR CASES OF TWO CLASSIFICATIONS

3.4.1. *Orthogonal classifications*

Two classifications *A* and *B* are called orthogonal, if A* and B* are orthogonal. In that case the iterative method comes quickly to an end, because $u_2 = 0$. Further $v_1 = y_{B^*}$, and $y_{sA^*} = y_A - y_N$. Finally $v_1 y_{sB} = v_1^2 = y_{B^*}^2$, so that $y_{A+B}^2 = y_A^2 + y_{B^*}^2$. All these relations also follow directly from the orthogonality of A* and B*. So we find that the estimate of *E*y is equal to $y_N + y_{A^*} + y_{B^*} = y_A + y_B - y_N$ and the square of this estimate is

$$y_N^2 + y_A^2 - y_N^2 + y_B^2 - y_N^2 = y_A^2 + y_B^2 - y_N^2, \text{ so that } y_R^2 = y^2 - y_A^2 - y_B^2 + y_N^2.$$

Apart from the simplicity of the calculations we have the particularity that the estimate of an effect, say *A*, is the same whether the other effect *B* is supposed to be present or not.

To obtain a general condition for orthogonality, also in case of a diagonal metric, the sum of weights in class *j* of *B* is called $W(j)$. A vector with the coordinates $W(j_2)$ in class j_1 of *B*, and with coordinates $-W(j_1)$ in class j_2 of *B* and with zeros in the remaining classes is in B*. We choose a (any) class of *A* and also consider the corresponding basis vector of A consisting of ones and zeros. Let the sum of the weights of the coordinates that are in that class of *A* as well as in class *j* of *B* be r_j. Then the two mentioned vectors are orthogonal if and only if $r_{j_1} W(j_2) - r_{j_2} W(j_1) = 0$, or $r_{j_1} : r_{j_2} = W(j_1) : W(j_2)$. Because the numbers j_1 and j_2 have been chosen arbitrarily, it follows that the sums of weights, r_j, are proportional to the numbers $W(j)$. Moreover this is true for every class of *A*.

We have: Two classifications are orthogonal, if and only if the proportions between the sums of weights of the coordinates of all the classes of the one classification are equal within all classes of the other classification. This proportion will be equal to the proportion of the sums of weights of the complete classes of the first classification. Compare KUIPER (11).

3.4.2. *Balanced incomplete blocks*

In classifications with the name "incomplete blocks" *A* is often called a classification according to treatments, and *B* a classification according to blocks, which are more or less homogeneous groups of individuals (plots). The differences in effect of these blocks are taken into account, in order to eliminate the inevitable variation between them. Every block will be a natural unit such as animals with the same parents, estimates by one person, yields of adjacent plots etc. Especially when the number of classes of *A* is large, it is impossible, in general, to form so large homogeneous groups of individuals (plots, animals etc) that all classes of *A* can be included in such a group.

In balanced incomplete blocks any of the *t* classes of *A* consists of *r* coordinates; any of the *b* classes of *B* consists of *k* coordinates. Clearly $n = tr = kb$. Any class of *A* has with any class of *B* one or no coordinate in common. Any two classes of *A* are represented together in λ blocks. Considering the *r* classes of *B* in which class *i* of *A* is represented, we observe that these *r* classes of *B* contain $kr - r$ coordinates that belong to the $t - 1$ remaining classes of *A*.

Because every of these $t - 1$ classes is represented in a class of *B* together with class *i* equally frequently, namely λ times, we have: $\lambda = (kr - r)/(t - 1)$.

In order to observe the course of the iterative method for the orthogonal projection of y on $\mathrm{A}+\mathrm{B}$, we consider a vector a in A^* consisting of coordinates a_i, with $\Sigma_{i=1}^{t} a_i = 0$ because A^* is orthogonal to N. Applying the linear transformation $Q = P_\mathrm{A} P_\mathrm{B}$ on a we obtain $Q\mathrm{a}$. The coordinate of $Q\mathrm{a}$ in class i of A is equal to the average of r from the averages of the classes of B in a, namely of those classes of B in which class i of A is represented. Because every of the remaining $t-1$ classes of A is represented in those classes of B λ times, we find $(ra_i + \lambda\Sigma_{i' \neq i}\, a_{i'})/rk$, which, because of the relation between the a_i, is equal to $(r-\lambda)a_i/rk$. Because this is true for every i, it follows that Q is a multiplication by the number $\mu = (r-\lambda)/rk$. The factor μ is the only proper value of the transformation Q in A^*.

Analogously to the case of a classification in two classes we have

$\mathrm{y}_{s\mathrm{A}^*} = \left(1 - \frac{r-\lambda}{rk}\right)^{-1} (\mathrm{y}_\mathrm{A} - \mathrm{y}_\mathrm{BA})$ or substituting $\lambda = (kr-r)/(t-1)$:

$$\mathrm{y}_{s\mathrm{A}^*} = \frac{k(t-1)}{t(k-1)} (\mathrm{y}_\mathrm{A} - \mathrm{y}_\mathrm{BA}).$$

Clearly $\mathrm{v}_1 \mathrm{y}_{s\mathrm{A}}$, necessary for testing the effect of A and for computing y_R^2, is equal to $\frac{k(t-1)}{t(k-1)} (\mathrm{y}_\mathrm{A} - \mathrm{y}_\mathrm{BA})^2$.

Geometrically the property of this Q means that a in A^*, $P_\mathrm{B}\mathrm{a}$ and $P_\mathrm{A}P_\mathrm{B}\mathrm{a}$ can be represented by arrows in a plane, so that the angle between a and $P_\mathrm{B}\mathrm{a}$ is equal to that between $P_\mathrm{B}\mathrm{a}$ and $P_\mathrm{A}P_\mathrm{B}\mathrm{a}$. Denoting this angle by φ we remark that the length of $P_\mathrm{B}\mathrm{a}$ is $\cos\varphi$ times the length of a, and that the length of $P_\mathrm{A}P_\mathrm{B}\mathrm{a}$ is $\cos\varphi$ times the length of $P_\mathrm{B}\mathrm{a}$, so that $\mu = \cos^2\varphi$. We conclude: the balanced incomplete block design has the characteristic property that any vector in A^* forms the same angle φ with (its orthogonal projection on) B^* or B. Because $P_\mathrm{A}P_\mathrm{B}$ restricted to A^* is non-singular, there is a one-to-one correspondence between A^* and the set of orthogonal projections of the vectors in A^* on B. This set is a vector space $P_\mathrm{B}(\mathrm{A}^*)$, with the same dimension as A^* of course. Any vector in $P_\mathrm{B}(\mathrm{A}^*)$ forms the same angle φ with A^* (or A). The residual space of B orthogonal to $P_\mathrm{B}(\mathrm{A}^*)$ is also orthogonal to A^*; for, if a is in A^*, then the projection $P_\mathrm{B}\mathrm{a}$ and the perpendicular $\mathrm{a} - P_\mathrm{B}\mathrm{a}$ are orthogonal to that space, so that the same holds for their sum a.

It follows also geometrically that the square of the perpendicular, necessary for testing the effect of A, $\mathrm{y}_{\mathrm{A+B}} - \mathrm{y}_\mathrm{B} = \mathrm{y}_{s\mathrm{A}^*} + (\mathrm{y} - \mathrm{y}_{s\mathrm{A}^*})_\mathrm{B} - \mathrm{y}_\mathrm{B} = \mathrm{y}_{s\mathrm{A}^*} - (\mathrm{y}_{s\mathrm{A}^*})_\mathrm{B}$, i.e. the perpendicular from $\mathrm{y}_{s\mathrm{A}^*}$ on B, is equal to

$$\sin^2\varphi \cdot (\mathrm{y}_{s\mathrm{A}^*})^2 = (1-\mu) \left(\frac{1}{1-\mu}\mathrm{v}_1\right)^2 = \frac{1}{1-\mu}\mathrm{v}_1^2 .$$

Similarly the square of any perpendicular from a vector a in A^* on B is equal to $\frac{1}{1-\mu}\mathrm{a}_\mathrm{B}^2$.

Let a_1 and a_2 be two vectors in A^*, and b_1 and b_2 their orthogonal projections on B. Then the identity $\mu(\mathrm{a}_1 + \mathrm{a}_2)^2 = (\mathrm{b}_1 + \mathrm{b}_2)^2$ holds. Because $\mu\mathrm{a}_i^2 = \mathrm{b}_i^2$ $(i = 1, 2)$, we obtain $\mu\mathrm{a}_1\mathrm{a}_2 = \mathrm{b}_1\mathrm{b}_2$. Let φ_1 and φ_2 be the angle between a_1 and a_2, or between b_1 and b_2, respectively. Then the last identity may be written as

$$\frac{\mu a_1 a_2}{\sqrt{\mu}|a_1|\cdot\sqrt{\mu}|a_2|} = \frac{b_1 b_2}{|b_1|\cdot|b_2|} \text{ or } \cos\varphi_1 = \cos\varphi_2.$$

Hence the orthogonal projection of A* on B preserves the angles between vectors. In particular, the orthogonal projection of orthogonal vectors in A* on B will be orthogonal.

Let a_1 and a_2, and therefore b_1 and b_2, be orthogonal non-zero vectors. Then the inner product $a_1 b_2 = \{b_1 + (a_1 - b_1)\}b_2 = b_1 b_2 + (a_1 - b_1)b_2$ is zero, because $a_1 - b_1$ is a perpendicular on B. It follows that also the perpendiculars $a_1 - b_1$ and $a_2 - b_2$ are orthogonal. We conclude: the perpendiculars from any set of orthogonal basis vectors of A* on B span the residual space of A+B orthogonal to B. The sum of some of such perpendiculars is equal to the perpendicular from the sum of corresponding vectors in A* on B. The space spanned by B and some subspace A_1 of A* is also spanned by B and the set of perpendiculars from an orthogonal basis of A_1 on B. Let A_2 be the residual space of A* orthogonal to A_1. The orthogonal projection of y or of y_{A+B} on the residual space of A+B orthogonal to A_1+B is equal to the perpendicular on B from the orthogonal projection of y_{sA^*} on A_2. This projection of y is needed in the numerator of the statistic for the test of the hypothesis that the expectation of the component of y in A_2 vanishes. The square of the corresponding perpendicular is $\sin^2\varphi$ times the square of $(y_{sA^*})_{A_2}$ i.e.

$$\frac{1}{1-\mu}(v_1)^2_{A_2} = \frac{1}{1-\mu}(y_{A_2} - y_{BA_2})^2.$$

From the fact that the perpendiculars corresponding to orthogonal components of the pure effect of *A* to be tested are orthogonal, it follows that the set of corresponding squares are independent (in case of a normal distribution of y) and their sum is equal to $\frac{1}{1-\mu}v_1^2$. The square $v_1 y_{sA}$ can thus simply be decomposed in a sum of independent squares necessary for testing orthogonal components of the effect of *A*, namely by projection of the vector v_1 on the corresponding subspaces of A and multiplication of the squares of these projections by $\frac{1}{1-\mu}$. See also KRAMER and BRADLEY (10).

3.4.3. *Group divisible partially balanced incomplete blocks*

In partially balanced incomplete blocks with two associate classes there is, analogously to the balanced incomplete blocks, a classification *A* according to treatments and a classification *B* according to blocks. Any of the *t* classes of *A* consists of *r* coordinates; any of the *b* classes of *B* consists of *k* coordinates. Any class of *A* has with any class of *B* one or no coordinate in common. The pairs of classes of *A* satisfy a so-called relation of association. This relation is: any two classes are either first associates or second associates; every class has n_1 first associates and n_2 second associates (so that $n_1 + n_2 = t - 1$); any two first associates are represented together in λ_1 blocks and any two second associates are represented together in λ_2 blocks. If $\lambda_1 = \lambda_2$ we have balanced incomplete blocks again.

In order to observe the linear transformation $Q = P_A P_B$ we consider again a vector a in A* with coordinates a_i, with $\Sigma_{i=1}^{t} a_i = 0$. Applying *Q* we will find in class i of A, analogously to the foregoing case, $1/rk$ times the sum of the coordinates of a in those classes of B in which class i of A is represented. Because every of the remaining $t-1$ classes of A is represented in those classes of B, either λ_1 times if it is a first associate class of i, or λ_2 times if it is a second associate class of i, this sum is equal to $ra_i+\lambda_1 S_1+\lambda_2 S_2$; here S_1 represents the sum of coordinates a_i corresponding to the first associates of i, and S_2 the similar for second associates; $a_i+S_1+S_2=0$. So we find in class i:

$$(ra_i+\lambda_1 S_1+\lambda_2 S_2)/rk$$

which, because of the relation between the a_i, is equal to

$$\{(r-\lambda_2)\,a_i+(\lambda_1-\lambda_2)\,S_1\}/rk.$$

The classification A is named group divisible, if the classes of A can be divided in m groups of n such that any two classes in the same group are first associates (thus $n_1=n-1$) and any two classes in different groups are second associates (thus $n_2=nm-n$). Considering again the r classes of B in which a certain class of A is represented, we find the relation: $rk=r+\lambda_1 n_1+\lambda_2 n_2=$ $=r+\lambda_1(n-1)+\lambda_2(t-n)$. Or: $rk-\lambda_2 t=r-\lambda_2+(n-1)\ (\lambda_1-\lambda_2)$.

Uniting the classes which are first associates to new classes, we denote the corresponding space of main effects, which is a subspace of A*, by A_1. If a is in A_1, then $\{(r-\lambda_2)a_i+(\lambda_1-\lambda_2)S_1\}/rk$ is equal to $\{(r-\lambda_2)a_i+(\lambda_1-\lambda_2)\,(n-1)a_i\}/rk=$ $=\{(rk-\lambda_2 t)a_i\}/rk$. Q applied to any vector in A_1 is a multiplication by $\mu_1=$ $=(rk-\lambda_2 t)/rk$. If a is in the residual space A_2 of A* orthogonal to A_1 so that $a_i+S_1=0$, we find $\{(r-\lambda_2)a_i-(\lambda_1-\lambda_2)a_i\}/rk=\{(r-\lambda_1)a_i\}/rk$. Q applied to A_2 is a multiplication by $\mu_2=(r-\lambda_1)/rk$. The transformation Q in A* has two proper values μ_1 and μ_2 with A_1 and A_2 as associated spaces of proper vectors respectively. It follows immediately that

$$\begin{aligned}
y_{sA^*} &= \frac{1}{1-\mu_1}\,(y_A-y_{BA})_{A_1}+\frac{1}{1-\mu_2}\,(y_A-y_{BA})_{A_2}= \\
&=\frac{rk}{\lambda_2 t}\,(y_A-y_{BA})_{A_1}+\frac{rk}{rk-r+\lambda_1}\,(y_A-y_{BA})_{A_2}= \\
&=\frac{rk}{rk-r+\lambda_1}\left[y_A-y_{BA}+\frac{n(\lambda_1-\lambda_2)}{\lambda_2 t}\,(y_A-y_{BA})_{A_1}\right].
\end{aligned}$$

$$\begin{aligned}
\text{Further } v_1 y_{sA} &= \frac{rk}{\lambda_2 t}\,(y_A-y_{BA})^2_{A_1}+\frac{rk}{rk-r+\lambda_1}\,(y_A-y_{BA})^2_{A_2}= \\
&=\frac{rk}{rk-r+\lambda_1}\left[(y_A-y_{BA})^2+\frac{n(\lambda_1-\lambda_2)}{\lambda_2 t}\,(y_A-y_{BA})^2_{A_1}\right].
\end{aligned}$$

Analogously to the case of balanced incomplete block designs any vector in A_1 forms the same angle φ_1 with $P_B(A_1)$ and any vector in A_2 forms the same angle φ_2 with $P_B(A_2)$. Further $A_1+P_B(A_1)$ is orthogonal to $A_2+P_B(A_2)$ which is seen from the fact that $P_B(A_1)$ and $P_B(A_2)$ are spaces of proper vectors of $P_B P_A$ in B* associated with the proper values μ_1 and μ_2. It follows that any orthogonal component in A_1 or in A_2 will be estimated by multiplication by the appropriate factor $\dfrac{rk}{\lambda_2 t}$ or $\dfrac{rk}{rk-r+\lambda_1}$ of the orthogonal projection of v_1 on the corresponding subspace. For testing such a component we need the square of this projection of v_1 as well multiplied by the appropriate factor. The sum of such quantities (squares of perpendiculars on B) is equal to $v_1 y_{sA}$ again. Compare KRAMER and BRADLEY (10).

We remark that 283 of the 376 partially balanced designs compiled by BOSE e.a. (2) are group divisible.

3.4.4. *Latin square type partially balanced incomplete blocks*

20 designs in the just mentioned compilation are of the Latin square type. We consider the simplest type first. There are n^2 classes in A which can be represented by the lattice points of a square n by n.

Two classes are first associates, if they are represented in the same row or the same column of this square. Thus $n_1 = 2(n-1)$. Otherwise they are second associates. The orthogonal classifications according to rows and columns correspond to orthogonal subspace A_1 and A_2 of A^*.

If the vector a in A^* with coordinates a_i happens to be in A_1 or in A_2, then the coordinate of Qa in class i of A (compare foregoing section),

$$\{(r-\lambda_2)a_i + (\lambda_1-\lambda_2)S_1\}/rk,$$

is because $S_1 = (n-1)a_i + (0-a_i)$ equal to $\{(r-\lambda_2)a_i + (\lambda_1-\lambda_2)\,(n-2)a_i\}/rk$. Then Q is a multiplication by

$$\{(r-\lambda_2 + (n-2)\,(\lambda_1-\lambda_2)\}/rk = \{rk-\lambda_2 n^2 - n(\lambda_1-\lambda_2)\}/rk.$$

If the vector a is in the residual space A_3 of A^* orthogonal to A_1 and A_2, so that $S_1 = (0-a_i) + (0-a_i)$, $\{(r-\lambda_2)a_i + (\lambda_1-\lambda_2)S_1\}/rk$ is equal to $(r-2\lambda_1 + \lambda_2)a_i/rk$. It follows that

$$\begin{aligned} y_{sA^*} &= \frac{1}{1-\dfrac{rk-\lambda_2 n^2 - n(\lambda_1-\lambda_2)}{rk}}\,[(y_A - y_{BA})_{A_1} + (y_A - y_{BA})_{A_2}] + \\ &+ \frac{1}{1-\dfrac{r-2\lambda_1+\lambda_2}{rk}}\,(y_A - y_{BA})_{A_3} = \frac{rk}{rk-r+2\lambda_1-\lambda_2}\,(y_A - y_{BA}) + \\ &+ \left(\frac{rk}{\lambda_2 n^2 + n(\lambda_1-\lambda_2)} - \frac{rk}{rk-r+2\lambda_1-\lambda_2}\right)[(y_A - y_{BA})_{A_1} + (y_A - y_{BA})_{A_2}]\,. \end{aligned}$$

The expression for $v_1 y_{sA}$ and the investigation of orthogonal components of the effect of A will be analogous to the foregoing section.

An extension forms the case where in the square there is also a third classification in n classes of size n, orthogonal to the classifications according to rows and columns; now two classes in the same row or column or the same class of the third classification are first associates. To these three classifications correspond three orthogonal subspaces of A^* : A_1, A_2 and A_3. If the vector a in A^* is in one of these three subspaces, Q is a multiplication by

$$\{r-\lambda_2 + (n-3)\,(\lambda_1-\lambda_2)\}/rk \;=\; \{rk-\lambda_2 n^2 - 2n(\lambda_1-\lambda_2)\}/rk.$$

If a is orthogonal to these subspaces, then Q is a multiplication by

$$(r-3\lambda_1 + 2\lambda_2)/rk.$$

It follows that

$$\mathbf{y}_{s\mathrm{A}^*} = \frac{rk}{rk - r + 3\lambda_1 - 2\lambda_2}(\mathbf{y}_\mathrm{A} - \mathbf{y}_\mathrm{BA}) +$$
$$+ \left(\frac{rk}{\lambda_2 n^2 + 2n(\lambda_1 - \lambda_2)} - \frac{rk}{rk - r + 3\lambda_1 - 2\lambda_2}\right)\Sigma_{i=1}^{3}(\mathbf{y}_\mathrm{A} - \mathbf{y}_\mathrm{BA})_{\mathrm{A}_i}$$

The expression for $\mathbf{v}_1\mathbf{y}_{s\mathrm{A}}$ and the investigation of components in A is analogous to the preceding cases.

3.4.5. *Two-dimensional lattices*

In a two-dimensional lattice rp^2 coordinates are divided according to treatments (A) and blocks (B) again. There is a (sub)classification B_1 (of B) of the coordinates in r groups such that every group contains all p^2 treatments once. Each of these groups is divided in p blocks of p treatments such that these partitions of the treatments are orthogonal classifications of the p^2 treatments: every class of one of these classifications is represented once in every class of another classification. We observe that any vector in B$_2$, the residual space of B orthogonal to B$_1$, has a vanishing sum of coordinates in any class of B_1.

We consider the linear transformation $Q = P_\mathrm{B}P_\mathrm{A}$ on a vector b in B with coordinates b_i. If b is in B$_1$, then $Q = 0$, as B$_1$* is orthogonal to A. If b is in B$_2$, then the coordinate of Qb in a block will be $1/rp$ times the sum of the coordinates of b in those classes of A, which are represented in this block. If b_i is the coordinate of b in that block, the contribution to this sum from the block group, to which the considered block belongs, is equal to pb_i. The contribution from the remaining block groups is zero, as the p classes of A, represented in the considered block, occur in p different blocks of every group. Hence the required coordinate of Qb is $pb_i/rp = b_i/r$. As far as the considered vector is concerned, Q is a multiplication by $1/r$. Any vector in B$_2$ forms the same angle arc cos $r^{-\frac{1}{2}}$ with A. It follows that

$$\mathbf{y}_{s\mathrm{B}^*} = (\mathbf{y}_\mathrm{B} - \mathbf{y}_\mathrm{AB})_{\mathrm{B}_1} + \frac{1}{1 - \dfrac{1}{r}}(\mathbf{y}_\mathrm{B} - \mathbf{y}_\mathrm{AB})_{\mathrm{B}_2} =$$

$$= \frac{r}{r-1}(\mathbf{y}_\mathrm{B} - \mathbf{y}_\mathrm{AB}) - \frac{1}{r-1}(\mathbf{y}_{\mathrm{B}_1} - \mathbf{y}_{\mathrm{AB}_1}) = \frac{r}{r-1}(\mathbf{y}_\mathrm{B} - \mathbf{y}_\mathrm{AB}) - \frac{1}{r-1}(\mathbf{y}_{\mathrm{B}_1} - \mathbf{y}_\mathrm{N}).$$

Because the second term is orthogonal to A, $\mathbf{y}_{s\mathrm{A}^*} = \mathbf{y}_\mathrm{A} - \mathbf{y}_\mathrm{N} - \dfrac{r}{r-1}(\mathbf{y}_\mathrm{B} - \mathbf{y}_\mathrm{AB})_\mathrm{A}.$

For the calculation of $\mathbf{y}^2_{\mathrm{A+B}}$ we use $\mathbf{v}_1\mathbf{y}_{s\mathrm{B}} = \dfrac{r}{r-1}(\mathbf{y}_\mathrm{B} - \mathbf{y}_\mathrm{AB})^2 - \dfrac{1}{r-1}\{(\mathbf{y}_\mathrm{B} - \mathbf{y}_\mathrm{AB})_{\mathrm{B}_1}\}^2$ and $\mathbf{y}^2_\mathrm{A}$. Subtraction of $\mathbf{y}^2_\mathrm{B}$ from $\mathbf{y}^2_{\mathrm{A+B}}$ supplies the test statistic for the hypothesis that the treatments have no effect.

The particularities, established for incomplete blocks and lattices, are lost, when the covariance matrix of y is not $\sigma^2 \cdot 1$.

3.5. REGRESSION PROBLEMS WITH TWO MAIN EFFECTS AND INTERACTIONS

3.5.1. *Interaction*

By two classifications A and B of the coordinates of a vector y in E a third classification is determined, which arises by uniting those elements into one class which belong simultaneously to a certain class of A and to a certain class of B. In our example of 3.3.1 eleven classes are obtained in this way. (Compare KUIPER [11]).

Corresponding to such a classification the vector space of vectors, that have the same coordinates within a class, may be defined; a basis of this space is formed by vectors with coordinates one in one class and zeros elsewhere. This space, as well as the corresponding classification, is denoted by $\mathrm{A} \times \mathrm{B}$. It is called the space of interactions of A and B. It is easily seen that the spaces N, A, and B and thus also $\mathrm{A}+\mathrm{B}$ are subspaces of $\mathrm{A} \times \mathrm{B}$. The residual space of $\mathrm{A} \times \mathrm{B}$ orthogonal to the space $\mathrm{A}+\mathrm{B}$ is called the space of pure interactions and denoted by $(\mathrm{A} \times \mathrm{B})^*$. In our example it is 11-1-2-3 = 5 dimensional.

The supposition that $E\mathrm{y}$ is in the space $\mathrm{A} \times \mathrm{B}$ implies that to every combination of the qualitative or quantitative characters considered in the classifications A and B there corresponds an expectation, which cannot (in general) be described as the sum of effects of the characteristics A and B separately. The space $\mathrm{A} \times \mathrm{B}$ may also be considered as a space of main effects of one characteristic, which consists of combinations of the characteristics A and B.

3.5.2. *Estimation and test*

The estimate of $E\mathrm{y}$ is the projection of y on $\mathrm{A} \times \mathrm{B}$. Projecting is simple because $\mathrm{A} \times \mathrm{B}$ is a space of main effects: in every class of $A \times B$ the coordinates of y are replaced by their average. The square of the perpendicular is obtained in the same way as in 3.2.2; division by the dimension of the space orthogonal to $\mathrm{A} \times \mathrm{B}$ yields an unbiased estimate of σ^2, which also will be used as denominator in test statistics.

The estimation of the components of $E\mathrm{y}$ in N, A* and B* remains equal to what we found in the foregoing; for the orthogonal projection of y on $\mathrm{A}+\mathrm{B}$ is the same as the orthogonal projection of $\mathrm{y}_{\mathrm{A}\times\mathrm{B}}$ on $\mathrm{A}+\mathrm{B}$. The estimate of the component in $(\mathrm{A} \times \mathrm{B})^*$ is equal to $\mathrm{y}_{\mathrm{A}\times\mathrm{B}} - \mathrm{y}_{\mathrm{A}+\mathrm{B}}$.

The null hypothesis that $E\mathrm{y}$ is in $\mathrm{A}+\mathrm{B}$, against the alternative that $E\mathrm{y}$ is in $\mathrm{A} \times \mathrm{B}$, is equivalent to $E\mathrm{y}_{(\mathrm{A}\times\mathrm{B})^*} = 0$. In the test of this hypothesis $\mathrm{y}^2_{(\mathrm{A}\times\mathrm{B})^*} = \mathrm{y}^2_{\mathrm{A}\times\mathrm{B}} - \mathrm{y}^2_{\mathrm{A}+\mathrm{B}}$ will be used in the numerator of the test statistic. Rejecting this null hypothesis implies the presence of effects namely joint effects of A and B. When the null hypothesis is not rejected, a test on the presence of main effects may be performed along the same lines as in the foregoing. However, in order that the test will be independent of whether there is an interaction or not, the component in the denominator is the perpendicular on $\mathrm{A} \times \mathrm{B}$ and not on $\mathrm{A}+\mathrm{B}$.

Consider the following decomposition of a vector $E\mathrm{y}$ in its components in N, A*, B*, and $(\mathrm{A} \times \mathrm{B})^*$:

$$\begin{array}{cc} & B \\ A & \begin{bmatrix} 12 & 12 \\ 12 & 20 \end{bmatrix} \end{array} = \begin{bmatrix} 14 & 14 \\ 14 & 14 \end{bmatrix} + \begin{bmatrix} -2 & -2 \\ 2 & 2 \end{bmatrix} + \begin{bmatrix} -2 & 2 \\ -2 & 2 \end{bmatrix} + \begin{bmatrix} 2 & -2 \\ -2 & 2 \end{bmatrix}.$$

On the left hand side we observe that three combinations of characteristics yield the same effect. On the right hand side we find two main effects and an interaction. With this example in mind, we remark that estimation of pure interactions is full of sense, while testing the presence of interactions, or estimating the joint effects of two or more characteristics, but, that an isolated presentation of a formal pure interaction may be misleading for purposes of interpretation.

Similarly an isolated presentation of main effects is misleading, if there is an interaction between the corresponding characteristics. A related and self-evident remark is that any supposition about Ey which contains the interaction between A and B must contain the corresponding main effects. In particular, this remark applies also to the null hypothesis and the alternative in a test.

Often every class of $A \times B$ has only one coordinate so that $\mathrm{A} \times \mathrm{B}$ is identical with E. The residual space orthogonal to $\mathrm{A} \times \mathrm{B}$ then consists of the null vector only. In such a case (the incomplete blocks and the lattices form examples) one will suppose in general that Ey is in $\mathrm{A} + \mathrm{B}$; this supposition must be acceptable of course. The orthogonal projection of y on the space, formally indicated by $(\mathrm{A} \times \mathrm{B})^*$, will have expectation zero then. Another possibility is that $Ey_{(\mathrm{A}\times\mathrm{B})^*}$ is supposed to be in a special subspace of $(\mathrm{A} \times \mathrm{B})^*$. Examples will be considered in the following section.

3.5.3. *Components of interaction*

We here consider only orthogonal classifications with one coordinate in every class of $A \times B$. Let $\mathrm{x} = (\lambda_1, \ldots, \lambda_n)$ be a vector in the vector space X of n-tuples and $\mathrm{y} = (\mu_1, \ldots, \mu_m)$ a vector y in the vector space Y of m-tuples. To every pair of vectors x and y a vector $\mathrm{x} \bullet \mathrm{y}$, called *tensor product* of x and y, is assigned; this is a vector in the space of mn-tuples with coordinates $\lambda_i\mu_j$ $(i = 1, \ldots, n;\ \ j = 1, \ldots, m)$:

$$\mathrm{x} \bullet \mathrm{y} = \begin{bmatrix} \lambda_1\mu_1 & \lambda_1\mu_2 & \cdots\cdots & \lambda_1\mu_j & \cdots\cdots & \lambda_1\mu_m \\ \lambda_2\mu_1 & \lambda_2\mu_2 & \cdots\cdots & \lambda_2\mu_j & \cdots\cdots & \lambda_2\mu_m \\ \vdots & \vdots & & \vdots & & \vdots \\ \lambda_i\mu_1 & \lambda_i\mu_2 & \cdots\cdots & \lambda_i\mu_j & \cdots\cdots & \lambda_i\mu_m \\ \vdots & \vdots & & \vdots & & \vdots \\ \lambda_n\mu_1 & \lambda_n\mu_2 & \cdots\cdots & \lambda_n\mu_j & \cdots\cdots & \lambda_n\mu_m \end{bmatrix}$$

The space of all linear combinations of such products: $\Sigma_i\alpha_i(\mathrm{x}_i \bullet \mathrm{y}_i)$, is called the tensor product $\mathrm{X} \bullet \mathrm{Y}$ of the spaces X and Y.

The inner product (with respect to the metric I) of $(\lambda_1, \ldots, \lambda_n) \bullet (\mu_1, \ldots, \mu_m)$ and $(\lambda_1', \ldots, \lambda_n') \bullet (\mu_1', \ldots, \mu_m')$ i.e. of $\mathrm{x} \bullet \mathrm{y}$ and $\mathrm{x}' \bullet \mathrm{y}'$ is equal to

$$\begin{aligned} & \lambda_1\lambda_1'\mu_1\mu_1' + \lambda_1\lambda_1'\mu_2\mu_2' + \ldots + \lambda_1\lambda_1'\mu_m\mu_m' + \\ + \; & \lambda_2\lambda_2'\mu_1\mu_1' + \lambda_2\lambda_2'\mu_2\mu_2' + \ldots + \lambda_2\lambda_2'\mu_m\mu_m' + \\ & \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ + \; & \lambda_n\lambda_n'\mu_1\mu_1' + \lambda_n\lambda_n'\mu_2\mu_2' + \ldots + \lambda_n\lambda_n'\mu_m\mu_m' = \\ = \; & (\lambda_1\lambda_1' + \ldots + \lambda_n\lambda_n')\,(\mu_1\mu_1' + \ldots + \mu_m\mu_m'). \end{aligned}$$

So we have $(\mathrm{x} \bullet \mathrm{y}, \mathrm{x}' \bullet \mathrm{y}') = (\mathrm{x}, \mathrm{x}')\ (\mathrm{y}, \mathrm{y}')$. It follows that if x is orthogonal to x′, and y is orthogonal to y′ then $\mathrm{x} \bullet \mathrm{y}$ is orthogonal to $\mathrm{x}' \bullet \mathrm{y}'$. Hence the tensor products of the vectors of an orthogonal basis of X and those of an orthogonal basis of Y form an orthogonal basis for $\mathrm{X} \bullet \mathrm{Y}$.

Let $\mathrm{r_x}$ be the vector in X consisting of ones and $\mathrm{r_y}$ the vector in Y consisting of ones. In $\mathrm{X} \bullet \mathrm{Y}$ the subspaces $\mathrm{X} \bullet \mathrm{r_y}$ and $\mathrm{r_x} \bullet \mathrm{Y}$ are spaces of main effects

according to the columns and rows respectively in the given $m \times n$ array. Any main effect $x \bullet r_y$ or $r_x \bullet y$ will also be represented by x or y respectively. The tensor product of two pure main effects x and y (which are vectors orthogonal to r_x and r_y respectively) is a pure interaction, because with any main effect $x' \bullet r_y$ and $r_x \bullet y'$:
$(x \bullet y)(x' \bullet r_y) = (x \bullet y)(r_x \bullet y') = 0$; for e.g. $(x \bullet y)(x' \bullet r_y) = (x, x')(y, r_y) = 0$.

The tensor product $U \bullet V$ of the two subspaces, U in the space of pure main effects X* in X, and V in the space of pure main effects Y* in Y, is called the space of pure interactions of U and V.

As an example we consider the case in which U has the basis x = (1, 1, –1, –1), and V the basis y = (2, –1, –1). The tensor product $U \bullet V$ has the basis

$$x \bullet y = \begin{bmatrix} 2 & -1 & -1 \\ 2 & -1 & -1 \\ -2 & 1 & 1 \\ -2 & 1 & 1 \end{bmatrix}.$$

This is the basis of the pure interactions of the two mentioned main effects U and V.

Next let the vector x (or y) be the set of function values of a polynomial, defined on n (or m) equidistant real numbers x_1 (or x_2). If x represents the function $f(x_1)$ and y the function $g(x_2)$, then $x \bullet y$ represents the function $f(x_1) \cdot g(x_2)$, i.e. a polynomial in two variables, defined on a rectangular lattice of points in a Cartesian $(x_1; x_2)$ coordinate system. If $f(x_1)$ or $g(x_2)$ is the function 1, in other words, if one of the vectors x or y is r_x or r_y, then we obtain main effects again. The tensor products of a set of orthogonal vectors in X, representing orthogonal polynomials of degree $\leqq s$ in x_1, and a set of orthogonal vectors in Y, representing orthogonal polynomials of degree $\leqq t$ in x_2, form an orthogonal basis of the subspace of $X \bullet Y$, representing polynomials in two variables x_1 and x_2, which for every fixed x_2 are of degree $\leqq s$ in x_1, and for every fixed x_1 of degree $\leqq t$ in x_2.

Let, for example, x = (1, –1, –1, 1) represent the quadratic orthogonal component of pure main effects X*, and y = (1, 0, –1) the linear component of pure main effects Y*. Then the tensor product

$$x \bullet y = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \\ 1 & 0 & -1 \end{bmatrix}$$

is the basis of the pure interactions of the quadratic main effects U and the linear main effects V, and corresponds to the function $x_1^2 x_2 - 5x_2$. Bases of pure interactions can also be formed as tensor products of main effects based on a classification on the one hand, and of main effects based on polynomials on the other hand.

Another example is the tensor product of the best estimates of the pure main effects x and y (which stand for the corresponding vectors in A* and B*) of A and B. The subspace of pure interactions, spanned by this product, represents non-additivity of the effects of A and B (TUKEY [21]).

All such components can be tested in a way similar to that in 3.2.3. The supposition about Ey determines the residual space, while all the components are orthogonal; hence, the square of the projection of y on the corresponding spaces can be found easily. That the *F*-test is also valid in the last example (non-additivity), in which the space on which y must be projected has a random basis vector, follows by considering the conditional distribution of the *F*-test statistic, under the condition of fixed effects of *A* and *B*. Because this distribution is independent of the condition, this condition may be omitted.

3.6. GENERAL REGRESSION PROBLEMS BASED ON THREE CLASSIFICATIONS

3.6.1. *Definitions and hypotheses*

Let the coordinates of y be divided according to three classifications *A*, *B* and *C*. Such classifications can be represented in a spatial diagram best. By the three classifications a classification is determined, of which every class contains the coordinates which belong simultaneously to the same class of *A*, the same class of *B* and the same class of *C*. The subspace of vectors, with the same coordinates within every of these classes, is the space of second-order interactions of *A*, *B* and *C*, denoted by $\mathrm{A} \times \mathrm{B} \times \mathrm{C}$.

It contains N, A, B, and C, further $\mathrm{A} \times \mathrm{B}$, $\mathrm{A} \times \mathrm{C}$ and $\mathrm{B} \times \mathrm{C}$. The residual space of $\mathrm{A} \times \mathrm{B} \times \mathrm{C}$ orthogonal to the space spanned by $\mathrm{A} \times \mathrm{B}$, $\mathrm{B} \times \mathrm{C}$ and $\mathrm{A} \times \mathrm{C}$, and thus orthogonal to N, A*, B*, C*, $(\mathrm{A} \times \mathrm{B})^*$, $(\mathrm{A} \times \mathrm{C})^*$, and $(\mathrm{B} \times \mathrm{C})^*$ is the space of pure second-order interactions $(\mathrm{A} \times \mathrm{B} \times \mathrm{C})^*$. The supposition that *E*y is in $\mathrm{A} \times \mathrm{B} \times \mathrm{C}$ implies that, to every combination of the characteristics *A*, *B* and *C*, there corresponds an expectation which cannot always be described as the sum of the main effects and the ordinary interactions between *A* and *B*, *C* and *B*, and *A* and *C* only.

Generalizations to four or more classifications will not be discussed, because the treatment is analogous; interactions of third and higher order have to be introduced.

Every hypothesis about *E*y (and thus also the null hypothesis and the alternative in a test) is of the form: *E*y is in some subspace W of E. If some interaction is included in W, then all main effects and interactions of lower order, which pertain the same classifications as that interaction does, should also occur in W; if any of them is dropped, then interaction looses its sense. So we obtain the following types of admissible tests (and thus hypotheses), with the spaces that span W under the null hypothesis on the left, and those under the corresponding alternative on the right:

null hypothesis:	alternative hypothesis:
IV: N, A*, B*, C*, (A×B)*, (A×C)*, (B×C)*;	V: N, A*, B*, C*, (A×B)*, (A×C)*, (B×C)*, (A×B×C)*.
III: N, A*, B*, C*, (A×B)*, (A×C)*;	N, A*, B*, C*, (A×B)*, (A×C)*, (B×C)*.
II: N, A*, B*, C*, (A×B)*;	N, A*, B*, C*, (A×B)*, (A×C)*.
I: N, A*, B*, C*;	N, A*, B*, C*, (A×B)*.
N, A*, B*, (A×B)*;	N, A*, B*, C*, (A×B)*.
N, A*, B*;	N, A*, B*, C*.
N, A*, B*;	N, A*, B*, (A×B)*.
N, A*;	N, A*, B*.
N;	N, A*.

The following test concerning the effect of *C* is not reasonable: the alternative hypothesis is $\mathrm{W} = \mathrm{N} + \mathrm{A}^* + \mathrm{B}^* + \mathrm{C}^* + (\mathrm{A} \times \mathrm{B})^* + (\mathrm{A} \times \mathrm{C})^*$, and the null hypothesis is $\mathrm{W} = \mathrm{N} + \mathrm{A}^* + \mathrm{B}^* + (\mathrm{A} \times \mathrm{B})^* + (\mathrm{A} \times \mathrm{C})^*$. For W under the null hypothesis contains $(\mathrm{A} \times \mathrm{C})^*$ and not the main effect C*.

The last three of the mentioned admissible tests and the corresponding hypotheses need no further consideration, because they have been discussed in the foregoing; we need only remark that the denominator of the test statistic will be formed from the square of the perpendicular on the space, corresponding to the most extended hypothesis about Ey, taken in consideration in any special case.

3.6.2. *Performance of estimations and tests*

We consider some special cases of the set of hypotheses given in the previous section, namely those marked I, ..., V, in detail.

First we take *case I.* The estimate of Ey is the orthogonal projection of y on the space spanned by A, B and C then. Remembering that orthogonal projection on each of these spaces requires a very simple procedure (averaging within classes), we use the iterative method of 1.5.6 with $k = 3$. Compare YATES (22) and STEVENS (20). Analogously to the case of two classifications, the vectors u, v and w will not be written in full, but only one coordinate of every class will be noted. For the following example of y, in which the symbols a_i, b_i and c_i indicate the classes of A, B and C respectively:

	a_1			a_2			a_3			a_4			a_5			a_6		
	c_1	c_2	c_3	c_1	c_2	c_3	c_1	c_2	c_3	c_1	c_2	c_3	c_1	c_2	c_3	c_1	c_2	c_3
b_1	21.6	20.1	16.4	20.2			18.9		20.1	20.2	30.9		19.8	20.1	26.5	32.2	20.1	
b_2		26.7	17.3					14.3				22.8			15.9		11.0	
b_3					22.6	25.1			21.9		20.9						21.7	28.8
b_4					18.8	12.2		24.6				17.3		20.1				10.9

we obtain this computational scheme:

$u_1 = y_A$	20.42	19.78	19.96	22.42	20.48	20.92
u_2	0.17	–0.17	0.00	0.00	0.26	–0.26
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$\Sigma u = y_{sA}$	20.68	19.52	19.96	22.42	20.81	20.59

y_B	y_{AB}	v_1	v_2 ...	$\Sigma v = y_{sB}^*$
22.25	20.68	1.57	–0.04 ...	1.46
18.00	20.77	–2.77	–0.08 ...	–2.84
23.50	20.63	2.87	0.12 ...	3.08
17.32	20.56	–3.24	0.04 ...	–3.15

y_C	22.15	20.98	19.60
y_{AC}	20.66	20.66	20.66
$(v_1)_C$	1.57	–0.39	–0.39
w_1	–0.08	0.71	–0.67
w_2	0.04	–0.01	–0.01
⋮	⋮	⋮	⋮
$y_{sC}^* = \Sigma w$	0.02	0.68	–0.70

The steps y_A, $y_B - y_{AB}$, y_C and y_{AC} need no further explanation. $(v_1)_C$ is formed by averaging in v_1; the second coordinate e.g. is $3 \times 1.57 + 3 \times (-2.77) + 3 \times 2.87 + 3 \times (-3.24)$ divided by 12. $w_1 = y_C - y_{AC} - (v_1)_C$. The first coordinate of u_2 is found as $3 \times 1.57 + 2 \times (-2.77)$ from v_1, and $-0.08 + 2 \times 0.71 + 2 \times (-0.67)$ from w_1, the sum of which has been divided by 5, and the quotient noted with opposite sign: 0.17. The subsequent steps are found similarly. Analogously to the case of two classifications, v_1, w_1, u_2, v_2, w_2 etc. are orthogonal to N, which affords a check again. Further Σv and Σw are orthogonal to N, so that they are equal to y_{sB^*} and y_{sC^*} respectively, while $\Sigma u = y_N + y_{sA^*}$. Finally $y^2_{A+B+C} = yy_{sA} + yy_{sB} + yy_{sC} = y_A y_{sA} + y_B y_{sB} + y_C y_{sC}$, so that the calculation of this square is analogous to the first method discussed in 3.3.5.

In *case II* the hypothesis can be summarized in: $W = (A \times B) + C$. The estimate of *Ey* is the orthogonal projection of y on W. This is obtained by means of the iterative method for two classifications, namely $A \times B$ and C. If desired, the component in $A \times B$ can be decomposed in the components in N, A*, B*, and $(A \times B)^*$, by applying the iterative method for the two classifications *A* and *B* on $y_{sA \times B}$; the component in the space of pure interactions is found by subtraction. The last decomposition is not necessary for purposes of testing. In order to determine the square of the orthogonal projection, we need only the components in $A \times B$ and C.

In *case III* the hypothesis can be summarized in $W = (A \times B) + (A \times C)$. Estimation of *Ey* requires the iterative method for the two classifications $A \times B$ and $A \times C$. Arranging the classes of both $A \times B$ and $A \times C$ in groups according to the classes of *A* shows, that the array falls apart in separate arrays for every class of *A*. This is reflected in the technical performance of the iterative method: the method will be applied to every class of *A* separately, with respect to the classifications *B* and *C* in it. It is convenient that the subtraction which yields v_1, always takes place in the same subspace, either $A \times B$ or $A \times C$. In order to decompose the unique projection into further components, the method for two classifications will be applied to the (not unique) component in $A \times B$ with the classifications *A* and *B*, and to the (not unique) component in $A \times C$ with *A* and *C*. In one of them is y_N. Both yield a (not unique) component in A*, so that the required unique component in A, y_{sA^*}, is equal to the sum of these two not unique components.

In *case IV* the estimate of *Ey* is obtained by the iterative method for three classifications, namely $A \times B$, $A \times C$, and $B \times C$. If one is interested in the components of the projection in the separate subspaces mentioned in IV, the method for two classifications will then be applied to every of the three (not unique) components separately. One of them will contain y_N; the unique pure main effects will each be found as the sum of two not unique components.

In *case V* the estimate of *Ey* is the orthogonal projection on the space $A \times B \times C$, which can be obtained in a very simple way. The decomposition in the components in the subspaces mentioned in V takes place by the method discussed in the last paragraph; the second-order interaction will be obtained as a residual.

Because the squares of the projections corresponding to the null hypothesis and the alternative can be calculated in any case now, every admissible test can be performed (compare 2.3.5). For the denominator of the test statistic we refer to the end of the last section. We still need to know the dimensions of the subspaces. In our example we have 1 for N, 5 for A*, 3 for B*, 2 for C*, $9 - 1 - 3 - 2 = 3$ for $(B \times C)^*$, and $30 - 1 - 5 - 3 - 2 - 3 = 16$ for the residual space of E orthogonal to N, A*, B*, C*, and $(B \times C)^*$. We do not consider the remaining interaction spaces in this example, because, as a consequence of the distribution of the coordinates in comparatively many classes, they have vectors, and thus spaces, in common with each other and with the already considered spaces. This phenomenon is indicated by the term "confounding".

To the case that every class of $A\times B\times C$ contains one coordinate only, remarks similar to those at the end of 3.5.1 apply.

The considerations are also valid for any diagonal metric.

3.7. PARTICULAR REGRESSION PROBLEMS WITH THREE CLASSIFICATIONS

3.7.1. *Complete orthogonality*

Three classifications A, B and C are called completely orthogonal, if every classification is orthogonal to the classifications generated by the remaining classifications i.e. A* is orthogonal to B × C, B* is orthogonal to A × C, and C* is orthogonal to A × B. This definition implies 6 relations of orthogonality which are not true in the general case. In diagram:

$$\begin{tikzcd}[column sep=small]
(B\times C)^* \arrow[r, dash] & A^* \arrow[rr, dash] \arrow[dr, dash] & & B^* \arrow[r, dash] \arrow[dl, dash] & (A\times C)^* \\
 & & C^* \arrow[d, dash] & & \\
 & & (A\times B)^* & &
\end{tikzcd}$$

In this particular case e.g. also (A × C)* is orthogonal to (A × B)*. This may be seen as follows: Choose a class of $A\times B$, i.e. a class of A and a class of B. Any vector in (A × C)* will contain coordinates c_i corresponding to class i of C in the chosen class of A. Let the number of coordinates in class i of C and in the chosen class of A be n_i. Then, because (A × C)* is orthogonal to A, $\Sigma_i n_i c_i = 0$. Let the number of coordinates in class i of C, which are in the chosen class of $A\times B$, be equal to m_{ij}. Then, as the classes of C are represented in the classes of $A\times B$ in the same proportion as in the classes of A, we have also $\Sigma_i m_{ij} c_i = 0$. The inner product of any vector in (A × C)* with the basis vector of A × B, consisting of ones in the chosen class of $A\times B$ and of zeros elsewhere, is zero. This is true for every basis vector of A × B. Hence (A × C)* and A × B are orthogonal, and also the assertion is true.

Another sufficient condition for complete orthogonality is: A is orthogonal to B, and C is orthogonal to A × B. For let the number of coordinates in class i of A be l_i, in class j of B m_j, and in class k of C n_k, then the number of coordinates in class (i, j, k) of $A\times B\times C$ is equal to $\dfrac{l_i m_j n_k}{n^2}$, from which follows a proportional representation of the classes of B in those of $A\times C$, and similarly of the classes of A in those of $B\times C$.

Orthogonality of A, B and C is not sufficient for complete orthogonality, as may be seen by the following example, where the numbers are the numbers of coordinates in the classes of $A\times B\times C$:

	c_1			c_2	
	b_1	b_2		b_1	b_2
a_1	14	16	a_1	16	14
a_2	16	14	a_2	14	16

In the case of complete orthogonality the estimation of effects and the performance of tests is very simple again. We mention only

$$y_{(A\times B)^*} = y_{A\times B} - y_A - y_B + y_N; \qquad y^2_{(A\times B)^*} = y^2_{A\times B} - y^2_A - y^2_B + y^2_N$$

which are the formulas for two orthogonal classifications, and

$$y_{A\times B\times C} = y_N + y_{A^*} + y_{B^*} + y_{C^*} + y_{(A\times B)^*} + y_{(A\times C)^*} + y_{(B\times C)^*} + y_{(A\times B\times C)^*},$$

from which follows

$$y_{(A\times B\times C)^*} = y_{A\times B\times C} - y_{A\times B} - y_{A\times C} - y_{B\times C} + y_A + y_B + y_C - y_N, \text{ and}$$
$$y^2_{(A\times B\times C)^*} = y^2_{A\times B\times C} - y^2_{A\times B} - y^2_{A\times C} - y^2_{B\times C} + y^2_A + y^2_B + y^2_C - y^2_N.$$

With the definition of orthogonality, given in 3.4.1, the considerations can be extended to the case in which weights are present.

3.7.2. *Latin squares*

In the case of Latin squares there are n^2 coordinates in y and three orthogonal classifications, each in classes of n coordinates. In every class of one classification the classes of the two remaining classifications are represented with one coordinate. Such classifications have already been mentioned in connection with lattices and incomplete blocks. Two classifications are usually indicated by the terms "rows" and "columns".

The most extended hypothesis about Ey will concern three main effects only, because the spaces of pure interactions have vectors and thus spaces in common with the spaces of main effects: interactions and main effects are confounded. The estimate of Ey under the mentioned hypothesis is $y_N + y_{A^*} + y_{B^*} + y_{C^*} = y_A + y_B + y_C - 2y_N$, with the square $y^2_A + y^2_B + y^2_C - 2y^2_N$. These formulae also hold in every other case in which A, B, and C are orthogonal classifications.

By introduction of weights the particularity of Latin squares is lost.

3.7.3. *Two classifications orthogonal* (*Pearce*)

There are several designs with three classifications (compare PEARCE [15, 16]), where two of these classifications, say A and B, are orthogonal, and where only main effects are taken in consideration. It is true that the estimation of Ey can be found then by application of the iterative method for three classifications. But it is simpler to consider the two spaces A + B and C, and to apply the iterative method for obtaining orthogonal projections on a space spanned by *two* spaces. We obtain the following sequence: $u_1 = y_{A+B} = y_A + y_B - y_N$; the three components are not added but noted separately. $v_1 = (y - y_A - y_B + y_N)_C$; the contributions of the components will be added within every class of C before averaging. $u_2 = (v_1)_A + (v_1)_B$; the components are noted separately (the orthogonal projection of v_1 on N is 0 of course). $v_2 = \{(v_1)_A + (v_1)_B\}_C$; the contributions of the components will be determined class by class of C. And so on.

y_{sC^*} will be found as $\Sigma^s_{i=1} v_i$; $y_{sA}{}^*$ as $y_A - y_N - \Sigma^s_{i=1}(v_i)_A$; and $y_{sB}{}^*$ as $y_B - y_N - \Sigma^s_{i=1}(v_i)_B$.

An interesting example (PEARCE [16], figure XVIII) can be derived from a Latin square with p^2 coordinates ($p \geqq 3$). In the array p by p, one row (class of A) is deleted, and a column (class of B) is added such that an array $p - 1$ by $p + 1$ is obtained. In the added column all classes of C but for one are represented. This missing class is called class p of C. In column i ($i = 1, \ldots, p$) class i of C is not represented; in column $p + 1$, the added column, class p of C is not represented. In row j ($j = 1, \ldots, p - 1$) class j of C is represented twice. It is clear that the new classifications to rows (A) and columns (B) are orthogonal, but that C is not to A and to B. Orthogonal projection on A + B + C will be performed in the way just described. For this purpose we consider the transformation $Q = P_C P_{A+B}$ on a vector in C* with coordinates c_i in class i of C so that

$$p\,\Sigma_{i=1}^{p-1}\, c_i + (p-1)c_p = 0 \quad \text{or} \quad \Sigma_{i=1}^{p}\, c_i = \frac{c_p}{p}$$

P_A gives in row 1: $\quad \frac{1}{p+1}\,(2c_1 + c_2 + \ldots + c_p) = \frac{1}{p+1}\left(c_1 + \frac{c_p}{p}\right)$

$\vdots$

in row $p-1$: $\quad \frac{1}{p+1}\,(c_1 + \ldots + c_{p-2} + 2c_{p-1} + c_p) =$

$$= \frac{1}{p+1}\left(c_{p-1} + \frac{c_p}{p}\right).$$

P_B gives in column 1: $\quad \frac{1}{p-1}\,(c_2 + c_3 + \ldots + c_p) = \frac{1}{p-1}\left(\frac{c_p}{p} - c_1\right)$

$\vdots$

in column $p-1$: $\quad \frac{1}{p-1}\,(c_1 + c_2 + \ldots + c_{p-2} + c_p) = \frac{1}{p-1}\left(\frac{c_p}{p} - c_{p-1}\right)$

in column p: $\quad \frac{1}{p-1}\,(c_1 + c_2 + \ldots + c_{p-1}) = -\frac{c_p}{p}$

and in column $p+1$: $\quad \frac{1}{p-1}\,(c_1 + \ldots + c_{p-1}) = -\frac{c_p}{p}.$

$P_C P_{A+B}$ gives in class 1 of C:

$$\frac{1}{p}\left[\frac{2}{p+1}\left(c_1 + \frac{c_p}{p}\right) + \frac{1}{p+1}\left(c_2 + \frac{c_p}{p}\right) + \ldots + \frac{1}{p+1}\left(c_{p-1} + \frac{c_p}{p}\right) + \right.$$

$$\left. + \frac{1}{p-1}\left(\frac{c_p}{p} - c_2\right) + \frac{1}{p-1}\left(\frac{c_p}{p} - c_3\right) + \ldots + \frac{1}{p-1}\left(\frac{c_p}{p} - c_{p-1}\right) - \frac{2c_p}{p}\right] =$$

$$= \frac{2}{(p+1)(p-1)}\left(c_1 - \frac{c_p}{p^2}\right),$$

similarly in class 2 of C: $\quad \frac{2}{(p+1)\,(p-1)}\left(c_2 - \frac{c_p}{p^2}\right),$

$\vdots$

in class $p-1$: $\quad \frac{2}{(p+1)\,(p-1)}\left(c_{p-1} - \frac{c_p}{p^2}\right),$

and in class p of C:

$$\frac{1}{p-1}\left[\frac{1}{p+1}\left(c_1 + \frac{c_p}{p}\right) + \ldots + \frac{1}{p+1}\left(c_{p-1} + \frac{c_p}{p}\right) + \frac{1}{p-1}\left(\frac{c_p}{p} - c_1\right) + \ldots + \right.$$

$$\left. + \frac{1}{p-1}\left(\frac{c_p}{p} - c_{p-1}\right)\right] = \frac{2c_p}{p(p-1)}.$$

We consider two orthogonal subspaces of C, namely C_1 which corresponds to the classification obtained by uniting the first $p-1$ classes of C into one class, and the residual space C_2 of C orthogonal to C_1. A basis for the subspace of C_1 orthogonal to N will be formed by the vector for which $c_i = 1$ for

$i \neq p$, and $c_p = -p$. The transformation $Q = P_C P_{A+B}$ on that vector is a multiplication by $\frac{2}{p(p-1)}$. Further a vector in C_2 has $c_p = 0$. For vectors in C_2 Q is a multiplication by $\frac{2}{(p+1)(p-1)}$. Now, with this knowledge, the derivation of these proper values of Q in C^* would be very much simpler, if one would consider the transformation Q in the separate subspaces C_1 and C_2 from the beginning.

We find:

$$\mathrm{y}_{sC^*} = \frac{1}{1-\frac{2}{p(p-1)}}(\mathrm{v}_1)_{C_1} + \frac{1}{1-\frac{2}{p^2-1}}(\mathrm{v}_1)_{C_2} =$$

$$= \frac{p^2-1}{p^2-3}\,\mathrm{v}_1 - \left\{\frac{p^2-1}{p^2-3} - \frac{p(p-1)}{(p-2)(p+1)}\right\}(\mathrm{v}_1)_{C_1},$$

where $\mathrm{v}_1 = (\mathrm{y}-\mathrm{y}_A-\mathrm{y}_B+\mathrm{y}_N)_C$. The quantity $\mathrm{v}_1\mathrm{y}_{sC}$, necessary for testing and for calculating the square of the perpendicular, is equal to

$$\frac{p^2-1}{p^2-3}\,\mathrm{v}_1^2 - \left\{\frac{p^2-1}{p^2-3} - \frac{p(p-1)}{(p-2)(p+1)}\right\}(\mathrm{v}_1)^2_{C_1}.$$

Besides we need the quantity $\mathrm{y}_A^2 + \mathrm{y}_B^2 - \mathrm{y}_N^2$.

These considerations are only valid in case of a covariance matrix $\sigma^2 \cdot I$ for y. The design (PEARCE [16], figure XVII), in which a row and a column are added to a Latin square can be dealt with similarly, as well as other examples. One of the interesting subspaces of C^* is again the space of vectors with equal coordinates in all those classes of C which are equally large. Sometimes one finds that Q is a multiplication by a real number for all vectors in C^*. (e.g. PEARCE [16], figures XV and XVI).

3.7.4. *One classification orthogonal to the interaction of the other two*

The situation in which C is orthogonal to $A \times B$, but A not orthogonal to B, may also be somewhat simpler than the general case. Referring to the proof in 3.7.1, in which has been used only that the classes of C are represented in any class of $A \times B$ in the same proportion as in any class of A, we have: $(A \times C)^*$ is orthogonal to $A \times B$, i.e. to A^*, B^*, and $(A \times B)^*$. The same is true of course for $(B \times C)^*$. We observe that all pairs of subspaces, named pure main effects or interactions, are orthogonal except the pair A^* and B^*, and the pair $(A \times C)^*$ and $(B \times C)^*$.

Now we consider the estimate of $E\mathrm{y}$ in the four cases, indicated by I, …, IV in 3.6.1, and discussed for the general case in 3.6.2, again.

In *case I*, the best estimate y_{A+B+C} will be found as the sum of the orthogonal projection $\mathrm{y}_{C^*} = \mathrm{y}_C - \mathrm{y}_N$, and the orthogonal projection y_{A+B} which can be obtained iteratively. This is also true in any other case in which C is orthogonal to A and to B only. An example of the latter situation is formed by the Youden squares.

In *case II* the hypothesis may be: $W = N + A^* + B^* + C^* + (A \times B)^* = (A \times B) + C^*$. The orthogonal projection of y on W is equal to the sum of the orthogonal projections on $A \times B$ and on C^* separately. The hypothesis, however, that $E\mathrm{y}$ is in $N + A^* + B^* + C^* + (A \times C)^*$ will now be summarized in: $E\mathrm{y}$ is in $(A+B) + C^* + (A \times C)^*$. These three spaces are orthogonal. The orthogonal projection on A + B will be obtained iteratively again, the component in C* is equal to $y_C - y_N$, and the component in (A × C)* is, as *A* and *C* are orthogonal, equal to $y_{A\times C} - y_A - y_C + y_N$.

In *case III* the hypothesis may be: *E*y is in $N + A^* + B^* + C^* + (A\times B)^* + (A\times C)^* = (A\times B) + C^* + (A\times C)^*$. The three spaces in the right hand side member are orthogonal. The desired projection is equal to the sum of the separate orthogonal projections on these spaces. Another hypothesis may be: *E*y is in $N + A^* + B^* + C^* + (A\times C)^* + (B\times C)^*$. In the general case we observed that the projection on this space is found by separate applications of the iterative method for two classifications in every class of *C* with the classifications *A* and *B* (and the subtraction which yields v_1 always either in A or in B, say B). In this way we obtain in A × C: $u = y_{sA\times C}$. Because *A* and *C* are orthogonal, u can be decomposed in y_N and in components in A*, C* and (A × C)* by simple averaging. In particular we note $y_{sA} = u_A$. The second component $v = y_{sB\times C}$ will be orthogonal to N, while its orthogonal projection on C will be the null vector. Because *B* and *C* are orthogonal, v contains only components in B* and in (B × C)*. Hence $y_{sB^*} = v_B$ and $y_{s(B\times C)^*} = v - v_B$. The effect of *C* is completely in u.

In *case IV* the hypothesis: *E*y is in $N + A^* + B^* + C^* + (A\times B)^* + (A\times C)^* + (B\times C)^*$, is represented by: *E*y is in $(A\times C) + (B\times C) + (A\times B)^*$ because (A × B)* is orthogonal to A × C and to B × C. The corresponding orthogonal projection is thus found as the sum of the orthogonal projection on (A × C) + (B × C) on the one hand, and on (A × B)* on the other hand. The projection on $W = (A\times C) + (B\times C)$ has just been discussed in the last paragraph and is equal to $u + v$ with u in A × C and v in B × C. The projection on (A × B)* should be obtained as $y_{A\times B} - y_{A+B}$.

Now we assert that y_{A+B} needs not be calculated again, but is equal to the sum of $y_{sA} = u_A$ and $y_{sB^*} = v_B$ as found by the method of the foregoing paragraph.

Proof: When we have demonstrated that $y - u_A - v_B$ is orthogonal to A and to B the proof will be complete. Because A + B is a subspace of W, $y_{A+B} = (y_W)_{A+B}$. Now $(y - u_A - v_B)_A = (y_W - u_A - v_B)_A = (u + v - u_A - v_B)_A = v_A - v_{BA}$. We know $v = v_B + (v - v_B)$ with $v - v_B$ in (B × C)*. Because $v_{BA} - v_{BBA} = v_{BA} - v_{BA} = 0$ and $(v - v_B)_A - (v - v_B)_{BA} = 0 - 0 = 0$, $y - u_A - v_B$ is orthogonal to A. Similarly: $(y - u_A - v_B)_B = (u + v - u_A - v_B)_B = u_B - u_{AB}$. We know $u = u_A + (u - u_A)$ with $u - u_A$ in $C^* + (B\times C)^*$. Because $u_{AB} - u_{AAB} = u_{AB} - u_{AB} = 0$ and $(u - u_A)_B - (u - u_A)_{AB} = 0 - 0 = 0$, $y - u_A - v_B$ is also orthogonal to B.

Finally we remark that $y^2_{(A\times B)^*}$ will be obtained as $y^2_{A\times B} - y^2_A - v_1 y_{sB}$. Here $y_{sB} = v_B$, while the orthogonal projection of the first step in the mentioned iterative process on B, $\{y_{B\times C} - (y_{A\times C})_{B\times C}\}_B = y_B - (y_{A\times C})_B = y_B - y_{AB} = v_1$.

3.8. REGRESSION PROBLEMS WITH BASES GIVEN BOTH EXPLICITLY AND BY CLASSIFICATIONS

3.8.1. *One classification and orthogonal polynomials*

An example, in which at the same time occurred classifications and orthogonal polynomials, has been introduced in 3.5.3 already. There the corresponding basis of the space, in which *E*y was supposed to be, was orthogonal.

Now we wish to consider the case where Ey is in a space $A+B$ spanned by a space of main effects A on the one hand, and by the space B of function values of polynomials of degree k on the other hand. This situation occurs e.g. in a field trial of which the plots have been arranged in one strip. A corresponds to the "treatments" to be investigated. The contribution of the soil to the measured property y, often called fertility, is supposed to be a continuous function, e.g. a polynomial, of the serial number of the plot in the strip. If the covariance matrix is $\sigma^2 \cdot I$, the space B will be spanned by the (tabulated) orthogonal vectors of function values (orthogonal polynomials) $r, x_1, \ldots, x_k$. The vector r is in A, while the remaining vectors, in general, are not orthogonal to A*.

To estimate Ey the projection of y on the space spanned by A and by B* i.e. by $x_1, \ldots, x_k$ must be performed, and this is fairly simple by means of the iterative method for two spaces A and B.

If we choose $v_1 = (y - y_A)_B$, then we find here $\Sigma_{i=1}^{k} \{(yx_i - y_A x_i)/x_i^2\} x_i$ which, for computational facilities, will be replaced by $\Sigma_{i=1}^{k} \{(yx_i - y_A x_{iA})/x_i^2\} x_i$ Next $u_2 = \Sigma_{i=1}^{k} \{(yx_i - y_A x_{iA})/x_i^2\} x_{iA}$; this will be the first vector to be computed in fact (one coordinate in each class of A): After arranging the coordinates of the vectors $y, x_1, \ldots, x_k$ according to the classes of A the coordinates of the vectors x_{iA} are determined by averaging. Next the numbers $(yx_i - y_A x_{iA})/x_i^2$ are computed (the denominators are tabulated), and with these numbers as coefficients the coordinates of the linear combination of the vectors x_{iA} will be determined class by class of A.

Analogously, $u_3 = \Sigma_{i=1}^{k} \{u_2 x_{iA}/x_i^2\} x_{iA}$ the computation of which requires the coefficients $u_2 x_{iA}/x_i^2$ first; they are simple to compute because both u_2 and x_{iA} are in A; next linear combinations in each class of A are determined. Similarly $u_4 = \Sigma_{i=1}^{k} \{u_3 x_{iA}/x_i^2\} x_{iA}$, and so on. The orthogonality of every u to N affords a check on the computations.

The interesting component $y_{sA^*} = y_A - y_N - \Sigma_{j=2}^{\infty} u_j$, and

$$y_{sB^*} = \Sigma_{i=1}^{k} \{(yx_i - y_A x_{iA})/x_i^2 + \Sigma_{j=2}^{\infty} u_j x_{iA}/x_i^2\} x_i,$$

where the term in braces is the sum of all the coefficients found for x_{iA} in the vectors $u_2, u_3, \ldots$. In order to compute y_{A+B}^2 we use y_A^2 and

$$v_1 y_{sB^*} = \Sigma_{i=1}^{k} \{(yx_i - y_A x_{iA})/x_i^2\} \{(yx_i - y_A x_{iA})/x_i^2 + \Sigma_{j=2}^{\infty} u_j x_{iA}/x_i^2\} x_i^2 =$$
$$= \Sigma_{i=1}^{k} \{yx_i - y_A x_{iA}\} \{(yx_i - y_A x_{iA})/x_i^2 + \Sigma_{j=2}^{\infty} u_j x_{iA}/x_i^2\}$$

i.e. the sum of k products of already computed numbers. For testing the effect of A we finally need: $y_B^2 = y_N^2 + \Sigma_{i=1}^{k} (yx_i)^2/x_i^2$.

This method can simply be extended to the case of a polynomial in e.g. two variables defined on a rectangular lattice of points in a Cartesian coordinate system. We may consider, for instance, a field trial in which A corresponds to "treatments" again, and the plots are arranged in a rectangular array. By the method of 3.5.3 an orthogonal basis for effects of fertility can easily be obtained.

3.8.2. *One classification and one vector*

Let Ey be in the space spanned by A, the space of main effects corresponding to a classification A, and a vector x. This means that Ey_i is equal to the sum of a constant, which is the same for all y_i within every class of A, and β times x_i, where x_i is the value for the i-th individual of a quantitatively expressible property x of the n individuals. In general, A does not contain x. Let the space, spanned by A and x, be called $\mathrm{A+X}$. If A has k classes the dimension of $\mathrm{A+X}$ is $k+1$.

In order to obtain the best estimate of $E\mathrm{y}$, i.e. to project y orthogonally on this space, the vector x, after the example of 3.1.2 and 3.3.5, is replaced by $\mathrm{x_R = x - x_A}$, orthogonal to A. The orthogonal projection of y on $\mathrm{x_R}$ is $b_{\mathrm{R}}\mathrm{x_R}$, with $b_{\mathrm{R}} = \mathrm{yx_R/x_R^2}$ as estimate for β. Thus $\mathrm{y_{A+X} = y_A} + b_{\mathrm{R}}\mathrm{x_R}$. The square of this projection is equal to $\mathrm{y_A^2 + (yx_R)^2/x_R^2}$. The square of the perpendicular is $\mathrm{y^2 - y_A^2 - (yx_R)^2/x_R^2}$, which, divided by the dimension $n-k-1$, gives an unbiased estimate of σ^2. The F-test statistic for the null hypothesis that $\beta = 0$, or that $E\mathrm{y}$ is in A, has $(b_{\mathrm{R}}\mathrm{x_R})^2$ as numerator.

In general one does not wish to write the estimate $\mathrm{y_A} + b_{\mathrm{R}}\mathrm{x_R}$ as the sum of a multiple of x and an effect of A, but as the sum of a multiple of $\mathrm{x - x_N = x'}$ and an effect of A, so that we obtain: $\mathrm{y_A} - b_{\mathrm{R}}\mathrm{x_{A*}} + b_{\mathrm{R}}\mathrm{x'}$. The reason of this choice is that the characteristic for the classification A will be the object of the investigation in fact, while $\beta\mathrm{x'}$ will represent a disturbing, but mostly inevitable, effect on Ey from a property, independent of A and measured by x; this effect must be taken into account necessarily; the average of the measurements of this disturbing property is therefore taken as origin of the measurements. The component $b_{\mathrm{R}}\mathrm{x_{A*}}$, to be subtracted from $\mathrm{y_A}$, is named the correction for this disturbing property. The component in A, $\mathrm{y_A} - b_{\mathrm{R}}\mathrm{x_{A*}}$, is named the effect of A corrected for the disturbing property. Another advantage of the mentioned choice is that the average variance of the coordinates in the vector of corrections is minimal then.

The hypothesis that A has no effect implies that $E\mathrm{y}$ is in the space spanned by x and N. The corresponding orthogonal projection of y, or of $\mathrm{y_{A+X}}$, is known from 3.1.2, namely $\mathrm{y_N} + b\mathrm{x'}$, with $b = \mathrm{x'y'/x'x'}$. The orthogonal projection of y on the $(k-1)$-dimensional residual space of $\mathrm{A+X}$ orthogonal to x and N is $\mathrm{y_{A*}} + b_{\mathrm{R}}\mathrm{x_R} - b\mathrm{x'}$, while the square is $\mathrm{y_{A*}^2} + (b_{\mathrm{R}}\mathrm{x_R})^2 - (b\mathrm{x'})^2$. This square occurs in the numerator of the test statistic for the mentioned hypothesis.

Sometimes the investigation of the component in $\mathrm{A+X}$ will be more refined. For this purpose the orthogonal projection of $\mathrm{y_{A*}}$ on $\mathrm{x_{A*}}$ will be considered i.e. the linear dependence of the effect of A on the class means of x. This projection is equal to $b^*\mathrm{x_{A*}}$, with $b^* = \mathrm{x_{A*}y_{A*}/x_{A*}^2}$. So we have four orthogonal components in $\mathrm{y_{A+X}}$, namely $\mathrm{y_N}$, $b^*\mathrm{x_{A*}}$, $\mathrm{y_{A*}} - b^*\mathrm{x_{A*}}$, and $b_{\mathrm{R}}\mathrm{x_R}$, in spaces with dimension 1, 1, $k-2$ and 1 respectively.

The second and fourth vector span a two-dimensional space which contains $\mathrm{x_A^* + x_R = x'}$. It follows that the third component is orthogonal to $\mathrm{x'}$. The orthogonal projection of y on this two-dimensional space can be decomposed in two other orthogonal components, namely the orthogonal projection on $\mathrm{x'}$, and the component orthogonal to $\mathrm{x'}$. The last is equal to $b^*\mathrm{x_{A*}} + b_{\mathrm{R}}\mathrm{x_R} - b\mathrm{x'}$. Let $E\,b^* = Eb_{\mathrm{R}}$, which means that the linear dependence of the effect of A on the class means is the same as that of y on x. Then $E(b^*\mathrm{x_{A*}} + b_{\mathrm{R}}\mathrm{x_R} - b\mathrm{x'})$, which is orthogonal to $\mathrm{x'}$, is a multiple of $\mathrm{x'}$; this multiple must be the null vector. Hence the square of the component, $b^*\mathrm{x_{A*}} + b_{\mathrm{R}}\mathrm{x_R} - b\mathrm{x'}$, may serve as the numerator of the test statistic for the hypothesis: $Eb^* = Eb_{\mathrm{R}}$. The refinement thus consists of a decomposition of the component of the effect of A orthogonal to $\mathbf{x}'$, namely $\mathbf{y}_{A^*}+b_R\mathbf{x}_R-b\mathbf{x}'$, in the two orthogonal components $\mathbf{y}_{A^*}-b^*\mathbf{x}_{A^*}$ and $b^*\mathbf{x}_{A^*}+b_R\mathbf{x}_R-b\mathbf{x}'$.

The first component $\mathbf{y}_{A^*}-b^*\mathbf{x}_{A^*}$ represents the part of the effect of A that, although linear dependence of Ey on x has been taken into account, cannot be described only by linear dependence of the effect of A on the corresponding class means of x. If the component of Ey in this $(k-2)$-dimensional subspace is not zero, then there is an effect of A unmixed with linear dependence on x, of which it is doubtful sometimes whether it is an effect of A that happens to present itself as linear in x, or real dependence on x. Further, when the non-linear component of the effect of A is zero, one may test the null hypothesis that the expectations of b^* and b_R are equal. Under this null hypothesis there is no effect of A at all. Here we meet a second advantage of the refinement: when the real effect of A happens to be linear in x in substance, so that distinction between a real dependence of y on x and an effect of A that happens to be linear in x is necessary, then the last test will be more powerful than the over-all test on effect of A.

3.8.3. *Performance of tests and estimations*

The tests indicated in the last section will be performed as follows: One determines $\mathbf{y}_R^2$ as $\mathbf{y}^2-\mathbf{y}_A^2$. Further $(b_R\mathbf{x}_R)^2=(\mathbf{y}\mathbf{x}_R)^2/\mathbf{x}_R^2=(\mathbf{y}\mathbf{x}-\mathbf{y}\mathbf{x}_A)^2/(\mathbf{x}^2-\mathbf{x}_A^2)=$ $=(\mathbf{x}\mathbf{y}-\mathbf{x}_A\mathbf{y}_A)^2/(\mathbf{x}^2-\mathbf{x}_A^2)$ where $\mathbf{x}_A\mathbf{y}_A$ is found analogously to $\mathbf{x}_A^2$: for every class the product of corresponding sums of classes in $\mathbf{x}$ and $\mathbf{y}$ is divided by the number of coordinates in that class, and the quotients are added. Next the denominator for the test statistics and the unbiased estimate of σ^2 can be computed. In order to test the non-linear component of the effect of A (dimension $k-2$), we compute

$$(\mathbf{y}_{A^*}-b^*\mathbf{x}_{A^*})^2=\mathbf{y}_{A^*}^2-(b^*\mathbf{x}_{A^*})^2=\mathbf{y}_A^2-\mathbf{y}_N^2-(\mathbf{x}_{A^*}\mathbf{y}_{A^*})^2/\mathbf{x}_{A^*}^2=$$
$$=\mathbf{y}_A^2-\mathbf{y}_N^2-(\mathbf{x}_A\mathbf{y}_A-\mathbf{x}_N\mathbf{y}_N)^2/(\mathbf{x}_A^2-\mathbf{x}_N^2).$$

When the null hypothesis is rejected, there is an effect of A.

If the null hypothesis is true (which may be assumed when it is not rejected), the equality of the regression coefficients Eb^* and Eb_R can be tested by means of

$$(b^*\mathbf{x}_{A^*}+b_R\mathbf{x}_R-b\mathbf{x}')^2=(b^*\mathbf{x}_{A^*})^2+(b_R\mathbf{x}_R)^2-(b\mathbf{x}')^2=$$
$$(\mathbf{x}_A\mathbf{y}_A-\mathbf{x}_N\mathbf{y}_N)^2/(\mathbf{x}_A^2-\mathbf{x}_N^2)+(\mathbf{x}\mathbf{y}-\mathbf{x}_A\mathbf{y}_A)^2/(\mathbf{x}^2-\mathbf{x}_A^2)-(\mathbf{x}\mathbf{y}-\mathbf{x}_N\mathbf{y}_N)^2/(\mathbf{x}^2-\mathbf{x}_N^2).$$

When the null hypothesis is rejected, there is an effect of A.

If both null hypotheses are true (which may be assumed when neither is rejected), the hypothesis $\beta=0$ may be tested by means of

$$(b\mathbf{x}')^2=(\mathbf{x}\mathbf{y}-\mathbf{x}_N\mathbf{y}_N)^2/(\mathbf{x}^2-\mathbf{x}_N^2).$$

When the effect of A is not considered by its components but as a whole one makes use of the quantity

$$\mathbf{y}_A^2-\mathbf{y}_N^2+(\mathbf{x}\mathbf{y}-\mathbf{x}_A\mathbf{y}_A)^2/(\mathbf{x}^2-\mathbf{x}_A^2)-(\mathbf{x}\mathbf{y}-\mathbf{x}_N\mathbf{y}_N)^2/(\mathbf{x}^2-\mathbf{x}_N^2)$$

and the corresponding dimension $k-1$.

The effect of A and that of x may sometimes be investigated in the reverse order. Then $(b_R\mathbf{x}_R)^2=(\mathbf{x}\mathbf{y}-\mathbf{x}_A\mathbf{y}_A)^2/(\mathbf{x}^2-\mathbf{x}_A^2)$ will serve as numerator in the test statistic for the null hypothesis $\beta=0$. If this hypothesis is true, the effect of A will be tested by means of $\mathbf{y}_{A^*}^2$.

In accordance with the results of the tests the following estimates of $E\mathrm{y}$ may be possible:

$$(\mathrm{y}_{\mathrm{A}} - b_{\mathrm{R}}\mathrm{x}_{\mathrm{A}*}) + b_{\mathrm{R}}\mathrm{x}';$$

$$\mathrm{y}_{\mathrm{N}} + b^{*}\mathrm{x}_{\mathrm{A}*} + b_{\mathrm{R}}\mathrm{x}_{\mathrm{R}} = \mathrm{y}_{\mathrm{N}} + (b^{*} - b_{\mathrm{R}})\mathrm{x}_{\mathrm{A}*} + b_{\mathrm{R}}\mathrm{x}'; \quad \mathrm{y}_{\mathrm{N}} + b\mathrm{x}'; \quad \mathrm{y}_{\mathrm{A}}; \quad \mathrm{y}_{\mathrm{N}}.$$

3.8.4. *Interaction between A and x*

It is possible that the supposition about $E\mathrm{y}$, mentioned in 3.8.2, is too stringent, and that the linear dependence of Ey on x, described by a regression coefficient, is not the same in all classes of A. Let the classes of A be grouped in m sets. Each set determines a vector x_i $(i = 1, \ldots, m)$, which in that set of classes of A (e.g. one class) contains the same coordinates as x and zeros elsewhere; the sum of these vectors which are orthogonal, will be x. The supposition that $E\mathrm{y}$ is in the space, spanned by A and the m vectors x_i, attributes to each of the m sets of classes a regression coefficient.

In order to determine the orthogonal projection of y on this space we consider the vectors $\mathrm{x}_{i\mathrm{A}}$, i.e. the orthogonal projection of x_i on A, and $\mathrm{x}_{i\mathrm{R}} = \mathrm{x}_i - \mathrm{x}_{i\mathrm{A}}$. Because the projection on A is obtained by averaging within classes of A, the vectors $\mathrm{x}_{i\mathrm{R}}$, $(i = 1, \ldots, m)$, will have zeros in all except one of the sets of classes. Hence they are mutually orthogonal. The orthogonal projection of y is $\mathrm{y}_{\mathrm{A}} + \Sigma_{i=1}^{m} b_i \mathrm{x}_{i\mathrm{R}}$, with $b_i = \mathrm{y}\mathrm{x}_{i\mathrm{R}}/\mathrm{x}_{i\mathrm{R}}^2$.

The square of the perpendicular (in a space with dimension $n - k - m$) is

$$\mathrm{y}^2 - \mathrm{y}_{\mathrm{A}}^2 - \Sigma_{i=1}^{m} \{(\mathrm{y}\mathrm{x}_{i\mathrm{R}})^2/\mathrm{x}_{i\mathrm{R}}^2\} = \mathrm{y}^2 - \mathrm{y}_{\mathrm{A}}^2 - \Sigma_{i=1}^{m}\{(\mathrm{x}_i\mathrm{y} - \mathrm{x}_{i\mathrm{A}}\mathrm{y}_{\mathrm{A}})^2/(\mathrm{x}_i^2 - \mathrm{x}_{i\mathrm{A}}^2)\}.$$

This gives an unbiased estimate of σ^2, and the denominator for the test statistic of the null hypothesis that the Eb_i are equal; for, because $\Sigma_{i=1}^{m}\mathrm{x}_i = \mathrm{x}$, the space A+X is a subspace of the $(k+m)$-dimensional space in which $E\mathrm{y}$ is supposed to be. In the numerator the square of the difference

$$\mathrm{y}_{\mathrm{A}} + \Sigma_{i=1}^{m} b_i \mathrm{x}_{i\mathrm{R}} - (\mathrm{y}_{\mathrm{A}} + b_{\mathrm{R}}\mathrm{x}_{\mathrm{R}}) = \Sigma_{i=1}^{m} b_i \mathrm{x}_{i\mathrm{R}} - b_{\mathrm{R}}\mathrm{x}_{\mathrm{R}}$$

will occur. This square is

$$\Sigma_{i=1}^{m} (b_i\mathrm{x}_{i\mathrm{R}})^2 - (b_{\mathrm{R}}\mathrm{x}_{\mathrm{R}})^2 = \Sigma_{i=1}^{m} \{(\mathrm{x}_i\mathrm{y} - \mathrm{x}_{i\mathrm{A}}\mathrm{y}_{\mathrm{A}})^2/(\mathrm{x}_i^2 - \mathrm{x}_{i\mathrm{A}}^2)\} - (\mathrm{x}\mathrm{y} - \mathrm{x}_{\mathrm{A}}\mathrm{y}_{\mathrm{A}})^2/(\mathrm{x}^2 - \mathrm{x}_{\mathrm{A}}^2),$$

while the corresponding dimension is $m - 1$.

When the null hypothesis is rejected, so that there is "interaction" of A and x, and thus effect of both A and x, one may wish to obtain an impression of the influence of the effect of A on $E\mathrm{y}$, without disturbance of the property x. For that purpose, we introduce a vector $\mathrm{x}_{i\mathrm{N}}$ which is obtained from x_{N} in the same way as x_i has been formed from x. We have: $\mathrm{x}_{\mathrm{N}} = \Sigma_{i=1}^{m}\mathrm{x}_{i\mathrm{N}}$. Next we apply the following decomposition:

$$\begin{aligned}\mathrm{y}_{\mathrm{A}} + \Sigma_{i=1}^{m} b_i \mathrm{x}_{i\mathrm{R}} &= \mathrm{y}_{\mathrm{A}} - \Sigma_{i=1}^{m} b_i(\mathrm{x}_i - \mathrm{x}_{i\mathrm{N}} - \mathrm{x}_{i\mathrm{R}}) + \Sigma_{i=1}^{m} b_i(\mathrm{x}_i - \mathrm{x}_{i\mathrm{N}}) = \\ &= \{\mathrm{y}_{\mathrm{A}} - \Sigma_{i=1}^{m} b_i(\mathrm{x}_{i\mathrm{A}} - \mathrm{x}_{i\mathrm{N}})\} + \Sigma_{i=1}^{m} b_i(\mathrm{x}_i - \mathrm{x}_{i\mathrm{N}}).\end{aligned}$$

The first term in braces represents the effect of A "reduced" to one particular value of x, namely the average of x.

3.8.5. *Two classifications and one vector*

Suppose that $E\mathrm{y}$ is in a space A+B+X spanned by two spaces of main effects A and B, which in general are not orthogonal, and by a vector x, representing a disturbing property again. In a field trial, A may be the object of investigation, B the blocks in any incomplete block design, and x the number of plants, which is assumed to be independent of the effect of A.

In order to determine the projection of y on $A+B+X$ we use the component x_R of x orthogonal to $A+B$. By means of the iterative method x will be projected orthogonally on $A+B$, so that x_{sA} and x_{sB^*} are obtained. Next y must be projected on $A+B$ on the one hand, and on x_R on the other hand, so that the required projection is $y_{sA}+y_{sB^*}+b_R x_R$, with $b_R = yx_R/x_R^2$. For the evaluation of b_R we recall first that the denominator x_R^2 is equal to

$$x^2 - x_{A+B}^2 = x^2 - x_A^2 - v_1 x_{sB}, \text{ with } v_1 = (x - x_A)_B.$$

Further: $yx_R = y(x - x_{A+B}) = xy - yx_A - y(x_{A+B} - x_A)$.

Analogously to the reduction in 3.3.5:

$$y(x_{A+B} - x_A) = (y_{sA} + y_{sB^*})\,(x_{A+B} - x_A) = y_{sB^*}\,\{x - x_A - (x - x_{A+B})\} =$$
$$= y_{sB^*}\,(x - x_A) = v_1 y_{sB^*}.$$

Hence $yx_R = y_R x_R = xy - x_A y_A - y_{sB^*}(x - x_A)_B$. In the third term the symbols x and y may be interchanged. These expressions are preferred, because of their accuracy and computational facility, to expressions like

$$y(x - x_{sA} - x_{sB^*}) = xy - x_{sA} y_A - x_{sB^*} y_B.$$

The square of the perpendicular on $A+B+X$ is equal to

$$y^2 - y_{A+B}^2 - (b_R x_R)^2 = y^2 - y_A^2 - y_{sB^*}(y - y_A)_B - (xy - x_A y_A - v_1 y_{sB^*})^2/(x^2 - x_A^2 - v_1 x_{sB^*})$$

which yields an unbiased estimate of σ^2 and the denominator for test statistics. The last term in this expression is the numerator of the F-test statistic for the effect of x. If such an effect is assumed, $y_{A+B} + b_R x_R$ will often be written as $(y_{sA} - b_R x_{sA^*}) + (y_{sB^*} - b_R x_{sB^*}) + b_R x'$, in which the terms in brackets are the "corrected" effects of A and B.

In order to test the effect of A i.e. the null hypothesis that Ey is in $B+X$ we need the projection y_{B+X}. This can be found according to the method for one classification and one vector (3.8.2). The difference $y_{A+B+X}^2 - y_{B+X}^2$ will appear in the numerator of the test statistic.

The procedure to investigate whether the regression coefficients are equal in all classes of, say, A, will be as follows. After the example of the last section one forms vectors x_i from x which in a set of classes of A have the same coordinates as x and zeros elsewhere again. The space $A+B+X$ is in the space spanned by A, B and the vectors x_i. In order to determine the orthogonal projection of y on the latter space, we need the components of the vectors x_i orthogonal to $A+B$. Therefore every x_i will (iteratively) be projected on $A+B$, so that x_{isA} and x_{isB^*} are found. The components $x_{iR} = x_i - x_{isA} - x_{isB^*}$, in general, are not orthogonal. But, apart from the projection of y on $A+B$, we wish to know the orthogonal projection of y on the space spanned by the vectors x_{iR}; the coefficients of x_{iR} in these projections are the estimates of the regression coefficients. This will lead to a system of normal equations, with numbers x_{iR}^2 and $x_{iR}x_{jR}$ as coefficients of the unknowns, and numbers yx_{iR} as known terms. These numbers can easily be determined; for $x_{iR}^2 =$ $= x_i^2 - x_{iA}^2 - v_1 x_{isB^*}$ with $v_1 = v_1^{(i)} = (x_i - x_{iA})_B$; similarly $x_{iR}x_{jR} =$ $= x_i x_j - x_{iA} x_{jA} - v_1^{(i)} x_{jsB^*}$ and $yx_{iR} = x_i y - x_{iA} y_A - v_1^{(i)} y_{sB^*}$. We remark that

$x_i x_j = 0$ and that many terms in inner products, such as $x_{iA} x_{jA}$ and $x_i y$, are zero. The number of equations will be small in practice. Let the solutions be b_i. Then the square of the required projection is $y^2_{A+B} + \Sigma_i b_i (y x_{iR})$. Compare the end of 2.3.4. The denominator of the test statistic can be calculated now. In the numerator the difference between the last square and y^2_{A+B+X} appears.

When the null hypothesis of equality of the regression coefficients is rejected (interaction between *A* and *x*, then we write the orthogonal projection as $y_{A+B} + \Sigma_i b_i x_{iR}$. Let x_{iN} be a vector again, which is formed from x_N in the same way as x_i is formed from x, then the expression can be reduced to

$$\begin{aligned} & y_{A+B} - \Sigma_i b_i (x_i - x_{iN} - x_{iR}) + \Sigma_i b_i (x_i - x_{iN}) = \\ & = y_{A+B} - \Sigma_i b_i (x_{isA} + x_{isB^*} - x_{iN}) + \Sigma_i b_i (x_i - x_{iN}) = \\ & = \{y_{sA} - \Sigma_i b_i (x_{isA} - x_{iN})\} + \{y_{sB^*} - \Sigma_i b_i x_{isB^*}\} + \Sigma_i b_i (x_i - x_{iN}). \end{aligned}$$

The first and the second term are the "corrected" effects of *A* and *B*. It will be remarked that vectors x_{isA} and x_{isB} may have non-zero coordinates in classes in which the corresponding x_i and x_{iN} have compulsory zeros.

The methods discussed in the last four sections may be applied without difficulty to the case of a diagonal metric.

3.9. MISSING PLOTS

When y must be projected orthogonally on spaces, spanned by subspaces of the kind considered in this chapter, we meet sometimes the following situation: the technical performance of a similar projection would be considerably more simple, if the vectors in E would contain a few, say one, two or three, coordinates in addition; for then one of those particular situations would occur, in which the orthogonal projection is an explicit expression in terms of a finite number of simply workable orthogonal projections. It is true that the required projection can be found e.g. iteratively, but it may be possible to take advantage of the simplicity of the performance of the orthogonal projection in the mentioned particular situation.

For this purpose we consider the vector space E corresponding to the given problem, and the vector space E′ corresponding to the particular situation. Vectors in E′ contain as many coordinates more than vectors in E do, as the number of coordinates missing for "completeness", the number of "missing plots", amounts. Now we assume that an *independent* basis for the space D, in which *E*y is supposed to be, can be obtained by omission of the "missing plot" coordinates in a basis for the corresponding space D′ in E′. (The correspondence between D and D′ is determined by the supposition about the coordinates of *E*y). In that case the correspondence between vectors x′ in D′ and vectors x in D, in which x is obtained by omission of the "missing plot" coordinates in x′, is one-to-one. This is e.g. not true, if a whole block is omitted from a particular incomplete block design, or in case *Ey* is supposed to be a polynomial function of degree 3 in five equidistant values of *x*, from which two are deleted.

We will show (compare KUIPER and CORSTEN [12], CORSTEN [3]) that the following method leads to a simple solution: We construct from the given vector y a vector y′ in E′ such that homologous coordinates of y and y′ are equal, while the missing plots are filled up with unknown variables, and we further determine a vector x' in D' such that $(y' - x')^2$ is minimal. The vector x' depends on y and the unknown variables. Let x be the vector in D corresponding to x' in D' according to the foregoing paragraph. Then, if the metric is diagonal, $(y' - x')^2$ is equal to the sum of the square $(y - x)^2$ in E, and a non-negative contribution from the missing plot coordinates (a weighted sum of squares of differences). Whatever be the choice for x (or for x', which amounts to the same) the optimal values for the variables in y' will be such that the contribution from the missing plot coordinates to $(y' - x')^2$ is minimal, i.e. zero. Therefore, in minimizing $(y' - x')^2$ we have to take the unknown coordinates of y' equal to the (so far unknown) corresponding coordinates of x', so that $(y' - x')^2 = (y - x)^2$. In other words, the minimum of $(y' - x')^2$ is equal to the minimum of $(y - x)^2$. Because minimizing $(y - x)^2$ is our purpose (for then x is the required projection on D), our method will lead to the solution. The method consists in minimizing $(y' - x')^2$ i.e. orthogonal projection of y' on D'. The only question remains how to choose the missing plot coordinates in y'.

The coordinates of x', the orthogonal projection of y' on D', are simple expressions in the coordinates of y' i.e. in the known coordinates of y and the unknown missing plot coordinates of y'. These expressions follow from the expression for the orthogonal projection of y' on D'. According to the foregoing, each of the missing plot coordinates of x' is equal to the corresponding unknown coordinate in y'. So we obtain linear equations for these unknowns, which must be solved. After solution y' is projected on D'.

The method yields $x' = y'_{D'}$ from which $x = y_D$ follows. Further it yields $(y - x)^2 = (y' - x')^2$. However, it does not give automatically x^2 ($\neq x'^2$ in general). Further the components of x, say in a space of main effects, found by this method will, in general, differ from the same component, found by e.g. the iterative method, by some multiple of the vector r, owing to the different consequences from the definition of orthogonality in the different spaces E and E'.

Not all statistical procedures applied to y can be carried over by analogy to y'. Sometimes, however, wrong tests are performed in this way. Consider the null hypothesis that Ey is in C, a subspace of D, and let C' in E' correspond to C. Then (for convenience sake) one often uses in the numerator of the test statistic $y'^2_{D'} - y'^2_{C'}$ instead of $y^2_D - y^2_C$. Now the first quantity is equal to $(y'_{D'} - y'_{C'})^2$, which is a weighted sum of squares of coordinates. The contribution to this sum from the "not-missing plots" is the inner product in E, $(y_D - y_{sC})^2$, where y_{sC} is some vector in C. According to the property of orthogonal projections this quantity is at least as large as $(y_D - y_C)^2$. Hence $y'^2_{D'} - y'^2_{C'} \geqq (y_D - y_{sC})^2 \geqq (y_D - y_C)^2$. In words: the null hypothesis will be rejected wrongly more frequently than is indicated by the nominal level of significance. On the other hand: if the null hypothesis will not be rejected with the wrong test, it will certainly not with the right test.

Right tests require new projections of y. These may be evaluated by means of another application of the just described method, in case that is advantageous.

CHAPTER 4

SOME REGRESSION PROBLEMS WITH CORRELATED OBSERVATIONS

4.1. GENERAL CONSIDERATIONS

4.1.1. *Kind of the problems*

In this chapter we consider random vectors y with a non-diagonal covariance matrix, and we determine estimations of *E*y and its components. We restrict ourselves to regression problems in which the main effect spaces and the interaction spaces are given by classifications of the components of y.

In practice it will often occur that the individuals to which the observations within a class of y belong, form certain natural units which, as a group, may be considered as individuals of higher order. Such a unit, like a set of animals with the same parents, a block of adjacent plots, or a set of estimates by the same person, have been introduced already in 3.4.2 in connection with block designs. A classification of y may have the property that it can be considered as a classification of these units. It may also happen that, moreover, each unit is divided over the classes of one or more other classifications of y. Within the class of these situations we will choose a number of more or less general examples which have practical importance.

First example: y consists of measurements of differences in yield between two varieties. Each measurement has been obtained from a trial with these two varieties in a random place of the area, for which the expectation of this difference must be investigated. Because it will be expected that meteorologic conditions influence this difference, y will be divided according to the year of the trial. One might think that an appropriate supposition about *E*y would be that it is a vector in the space of year effects. In that case the best estimate of *E*y would be found by orthogonal projection of y on this space. This is right indeed, if one wishes to consider the level of this difference in *these* years in which the trials have been performed. But mostly it is the purpose of this kind of investigations to estimate the expectation of the considered difference over all years. For, although effects of the years are undeniable, the size of such an effect cannot be predicted, in the first place because the future meteorological conditions cannot predicted till now, secondly because the relation between the meteorological and connected conditions on the one hand, and the effect on the considered difference on the other hand are unknown. Therefore a random variable is introduced: the expectation of the difference in a random year. The expectation of this random variable is, under the sketched conditions, the best quantity for purposes of prediction; the estimation of this expectation is the problem in fact.

The supposition about y arises in two steps: to every year, unit of higher order, corresponds a random variable. All these random variables have the same expectation, covariance zero, and variance σ_1^2. Under the condition of a fixed year the coordinates of y in that year all have the same fixed expectation, covariance zero and variance σ_2^2. In other words, every coordinate of y is equal to the sum of a general constant (to be estimated), a random contribution of the year under consideration, and a random contribution of the trial under consideration in this year.

Second example: Consider an experiment with animals in which two treatments (to which two classifications A and B correspond) are investigated. Every class of $A \times B$ contains a set of animals of the same litter or in the same stable, while in different classes of $A \times B$ there are represented different stables or litters. It is possible to estimate and to test formally the main effects of A and B and their interaction according to the methods of the foregoing chapter, because there is more than one coordinate in every class of $A \times B$ (we suppose that a covariance matrix $\sigma^2 \cdot I$ for y is acceptable).

However, these formal components concern *these* litters or stables. If there are remarkable differences between some of these, they will be reflected as main effects and interaction. However, it will be in general the purpose of the experiment to investigate the effects of the treatments on *such* litters or stables, from the population of which these litters or stables are samples. For this purpose the considered estimates and tests are without value.

The appropriate supposition about y arises as follows: Consider some stable or litter in a fixed class of $A \times B$; the animals of this stable in that class of $A \times B$ have an expectation value. Another stable or litter in the same class of $A \times B$ will have another expectation. So we obtain a random variable in that class of $A \times B$, namely the expectation of a random stable or litter in that class. Now such a random variable is attached to every class of $A \times B$. We assume that each of these random variables is the sum of an effect of A, an effect of B, and a random variable with expectation zero and variance $\sigma_1{}^2$. Moreover, these variables are supposed to have covariance zero. In other words, the vector of random expectations for the classes of $A \times B$ has as expectation the sum of an effect of A and an effect of B. To every class of $A \times B$ a random variable is attached. These random variables all have expectation zero, covariance zero and variance σ_1^2. Finally, under the condition that the coordinates of y (animals) are in a certain class of $A \times B$, unit of higher order now, they are random variables with the same expectation (equal to the now fixed expectation for that class), with covariance zero, and with variance $\sigma_2{}^2$.

The investigation of the effects will, with this supposition, be different from that in case the expectations in every class of $A \times B$ are considered as constants. It will be observed that interaction of A and B is not included in the new hypothesis, because the classes of $A \times B$ contain only one unit of higher order. Otherwise the space of (random) expectations would have no residual space.

Third example: This is related to the above-mentioned block designs. As pointed out in 3.4.2 the classification according to blocks in block designs does not take place, because one is interested in the size of the block effect, but in order to take into account an inevitable variation. Only the effect of A, the treatments, is interesting. Analogously to the foregoing examples, it is often justified to construct the supposition about y thus: to each block, unit of higher order, a random variable is attached. These random variables have covariance zero, variance $\sigma_1{}^2$, and may or may not have the same expectation (in the last case each class of a classification, with as elements these units of higher order, has its own expectation). Further, every coordinate of y is, under the condition that it is in a certain unit, a random variable with variance $\sigma_2{}^2$; the expectation is equal to the sum of the (under the condition fixed) expectation for the class of B, and a contribution from the class of A to which the coordinate under consideration belongs. The covariance of these variables is zero. This supposition is equivalent to: Every coordinate of y is equal to the sum of: a constant determined by the class of A, secondly a constant determined by the class of the above-mentioned classification of the blocks, further a random contribution of the unit, the block, under consideration (all these contributions have expectation zero, covariance zero and variance $\sigma_1{}^2$), and finally a random contribution of the coordinate under consideration (under the condition that the expectation of the unit has its particular value, these contributions have expectation zero, covariance zero and variance σ_2^2).

Comparing the assumption, given here, with the assumption of a non-random block effect, we will find that the assumption of a random block effect gives a more efficient estimate of the effect of A, in general.

The mentioned problems and those to be treated, in which subjects of apparently different kind such as components of variance, analysis of series of experiments, recovery of inter-block information (RAO [18]), and split-plot experiments may be recognized, all have this in common: There is a random contribution common to all coordinates in a class of the classification B of y according to units of higher order; these contributions have expectation zero, covariance zero, and variance $\sigma_1{}^2$. Further the coordinates of y have, under the condition that they are in the units of higher order in which they are, the covariance matrix $\sigma_2{}^2 \cdot I$.

4.1.2. *General remarks about the estimation*

The covariance matrix of y, with respect to the standard basis $e_1, \ldots, e_n$ of the space of n-tuples, has a very unsuitable form in the situation sketched in the foregoing section. It is true that coordinates, belonging to different classes of the classification (say B) in units of higher order, have covariance zero, but if they belong to the same class, they will have a covariance $\sigma_1{}^2$. The inverse of the covariance matrix should be chosen as metric, in order to obtain the best estimate of Ey by means of orthogonal projection. It is not clever to use this metric directly.

According to the end of chapter 1, we know that the inner product of two vectors with some metric with respect to the basis $e_1, \ldots, e_n$ is equal to the inner product with metric I, after replacement of one of these vectors by a linear combination of its orthogonal (metric I) projections on some particular orthogonal (metric I) subspaces. If this metric is the inverse of the covariance matrix of the coordinates of y with respect to $e_1, \ldots, e_n$, then the covariance matrix of the coordinates of y with respect to another basis, orthonormal with respect to metric I, of which the basis vectors are either in or orthogonal to the above-mentioned subspaces, has the diagonal form. The elements in the diagonal of this matrix are the reciprocals of the coefficients in the mentioned linear combination.

Conversely, if a basis, orthonormal with respect to metric I, can be found such that the coordinates have a diagonal covariance matrix, the coefficients in the linear combination of orthogonal projections, occurring in the inner product, will be the reciprocals of the corresponding elements in that matrix.

Now we introduce a new *orthogonal* (metric I) basis of E: the basis vectors of the space B, consisting of ones in one class of B and of zeros elsewhere, completed with vectors orthogonal to B. Next we consider the corresponding *orthonormal* basis. The coordinates of y with respect to this basis have a covariance matrix which will turn out to be simple.

Let one class of B consist of k coordinates. Then the coordinate of the projection on the space spanned by the corresponding unit basis vector (compare the end of 1.3.5) is equal to $1/\sqrt{k}$ times the sum of the coordinates in that class; the variance of this coordinate is equal to $(k^2\sigma_1^2+k\sigma_2^2)/k = k\sigma_1^2+\sigma_2^2$. The covariance of two such coordinates is zero, because they have no random variable in common. In any basis vector, orthogonal to B, the sum of the coordinates in every class of B is zero. The inner product of y and such a basis vector will not contain a random variable common to any class of B, while the sum of the n squares of the coefficients of the other random variables in that inner product is equal to the square of the basis vector. It follows that the corresponding variance is σ_2^2. The covariance of two such coordinates of projections appears to be zero in the same way as in 2.1.2. Finally the covariance of the coordinate of the projection on a one-dimensional space in B and that on a one-dimensional space in $B^\perp$ is zero; in the expectation of the product of the two inner products the variables with variance σ_1^2, vanish, because they occur linearly, and the vectors of coefficients of the other random variables are orthogonal. Summarizing we have found a diagonal covariance matrix for these coordinates of y.

The appropriate inner product of two vectors x_1 and x_2 will be found thus: one of these vectors will be projected orthogonally (metric I) on each of the one-dimensional spaces spanned by the above-mentioned basis vectors of B, named B_i, and on $B^\perp$; each of these orthogonal components will be multiplied by the reciprocal of the corresponding variance (the numbers k may be different for different B_i, and be called k_i). The results are added; finally, the inner product (metric I) of this sum and the other vector will be determined.

The estimate of Ey in A, i.e. the appropriate orthogonal projection of y on A, will be that vector y_{sA} in A (the symbol s is used because it is not the customary inner product orthogonal projection), for which $y-y_{sA}$ is orthogonal to A according to the new metric. For this purpose $y-y_{sA}$ will be projected orthogonally (with metric I) on the spaces B_i and $B^\perp$; we form a linear combination of these components, with the reciprocals of the corresponding variances as coefficients. The orthogonality with respect to the new metric requires that the orthogonal projection of this linear combination on A, with respect to the ordinary metric I, is equal to the null vector. (Compare 1.6.1).

With the foregoing in mind, we will use the term "orthogonal" and the index, say, A in y_A, only with respect to the metric I from now on, unless indicated explicitly otherwise. Further B will always be the classification according to the units of higher order, to which random variables are attached. For the present, i.e. up to 4.4, σ_1^2 and σ_2^2 are supposed to be known.

4.2. A CONSTANT MAIN EFFECT AND A RANDOM MAIN EFFECT

4.2.1. *General method of estimation*

Let y be divided according to classifications A and B. Let the main effect of A be the object of investigation, while to every class of B a common random variable is attached, beside the random variables for every coordinate of y separately. One may think of any incomplete block design, where A corresponds to treatments and B to blocks. Let the number of coordinates in class i of B equal k_i. Put $(k_i\sigma_1^2+\sigma_2^2)^{-1} = w_i$ and $\sigma_2^{-2} = w$.

The best estimate of the effect of A will be found as y_{sA} such that the orthogonal projection of $\Sigma_i \{w_i(\mathrm{y}-\mathrm{y}_{s\mathrm{A}})_{\mathrm{B}_i}\} + w\{\mathrm{y}-\mathrm{y}_{s\mathrm{A}}-\Sigma_i(\mathrm{y}-\mathrm{y}_{s\mathrm{A}})_{\mathrm{B}_i}\}$ on A is the null vector. Consequently:

$$\Sigma_i\{w_i(\mathrm{y}-\mathrm{y}_{s\mathrm{A}})_{\mathrm{B}_i}\}_{\mathrm{A}} + w\{\mathrm{y}-\mathrm{y}_{s\mathrm{A}}-\Sigma_i(\mathrm{y}-\mathrm{y}_{s\mathrm{A}})_{\mathrm{B}_i}\}_{\mathrm{A}} = 0$$

or

$$(\Sigma_i \frac{w_i}{w}\mathrm{y}_{\mathrm{B}_i})_{\mathrm{A}} - \left[\Sigma_i\left\{\frac{w_i}{w}(\mathrm{y}_{s\mathrm{A}})_{\mathrm{B}_i}\right\}\right]_{\mathrm{A}} + \mathrm{y}_{\mathrm{A}} - \mathrm{y}_{s\mathrm{A}} - (\Sigma_i\mathrm{y}_{\mathrm{B}_i})_{\mathrm{A}} + \{\Sigma_i(\mathrm{y}_{s\mathrm{A}})_{\mathrm{B}_i}\}_{\mathrm{A}} = 0$$

or

$$\mathrm{y}_{s\mathrm{A}} - \left\{\Sigma_i\left(1-\frac{w_i}{w}\right)(\mathrm{y}_{s\mathrm{A}})_{\mathrm{B}_i}\right\}_{\mathrm{A}} = \mathrm{y}_{\mathrm{A}} - \left\{\Sigma_i\left(1-\frac{w_i}{w}\right)\mathrm{y}_{\mathrm{B}_i}\right\}_{\mathrm{A}}$$

We consider the following linear transformation by which to a vector x in B a vector $\mathrm{z} = \mathrm{x}_{g\mathrm{A}}$ in A is assigned: the coordinate of z in a class of *A*, containing *m* coordinates, is equal to $1/m$ times the linear combination of the coordinates of x in those classes *i* of *B*, which are represented in this class of *A*, with coefficients g_i. We remark that this transformation is the orthogonal projection on A, if the coefficients g_i are 1.

Choosing $g_i = w_i/w = k_i\sigma_1{}^2/(k_i\sigma_1{}^2+\sigma_2{}^2)$, we write the last equation as: $\mathrm{y}_{s\mathrm{A}} - \{(\mathrm{y}_{s\mathrm{A}})_{\mathrm{B}}\}_{g\mathrm{A}} = \mathrm{y}_{\mathrm{A}} - \{\mathrm{y}_{\mathrm{B}}\}_{g\mathrm{A}}$. Denoting the linear transformation, which consists of the orthogonal projection on B, P_{B}, followed by the transformation $g\mathrm{A}$, by the symbol Q, we have: $(1-Q)\mathrm{y}_{s\mathrm{A}} = \mathrm{y}_{\mathrm{A}} - Q\mathrm{y}$. This is very much similar to the resulting equation in 3.3.4.

The transformation Q of vectors in A may be paraphrased as follows: project orthogonally on B; multiply in the projection the coordinates in class *i* of *B* by g_i $(0 \leq g_i < 1)$; project the vector, obtained in this way, orthogonally on A.

The length of the vector in B, to be projected on A, is smaller than the length of the projection on B, so that the bound of Q satisfies $G(Q) < 1$. It follows that $\mathrm{y}_{s\mathrm{A}} = (I+Q+Q^2+\ldots)(\mathrm{y}_{\mathrm{A}} - \mathrm{y}_{\mathrm{B}\cdot g\mathrm{A}})$.

The difference with the iterative method for two classifications in 3.3 consists only in the insertion of an additional operation between the orthogonal projection on B and that on A, namely multiplication of the coordinates of the vector in B by the corresponding g_i.

4.2.2. *Special cases and computational remarks*

It will be clear that the iterative process will converge the faster the smaller the numbers g_i are. In particular, if these numbers are zero, which is the case if $\sigma_1 = 0$ (the random block variables are zero) we find at once: $\mathrm{y}_{s\mathrm{A}} = \mathrm{y}_{\mathrm{A}}$; the classification *B* will be neglected. On the other hand, if σ_1/σ_2 is large, then g_i is approximately 1, and the iterative process may be replaced by the process which is formally obtained for $g_i = 1$ i.e. the iterative process of 3.3. It follows that $\mathrm{y}_{\mathrm{A+B}}$ yields the most efficient estimate of the effect of *A*, if σ_2 is negligible in comparison with σ_1. When the effect of *A* is estimated from $\mathrm{y}_{\mathrm{A+B}}$ in case this extreme situation does not occur, the estimate is unbiased, but not most efficient, since the metric is not chosen in the appropriate way (compare 2.3.3). This conclusion proves our assertion at the end of the discussion of the third example in 4.1.1.

Equality of the numbers k_i implies equality of the numbers g_i: $g_i = g$. Then $Q = P_{\mathrm{A}}\cdot g\cdot P_{\mathrm{B}} = g\cdot P_{\mathrm{A}}P_{\mathrm{B}}$, which means a considerable simplification: the iterative method of 3.3 will be applied with the difference only that, instead of writing down the coordinates of the orthogonal projection of a vector in B on A, one notes these coordinates multiplied by g.

It will be useful to perform the transformation $(1-Q)^{-1}(P_{\mathrm{A}}-g\cdot P_{\mathrm{A}}P_{\mathrm{B}})$ on y_{N} and on $\mathrm{y}-\mathrm{y}_{\mathrm{N}}$ separately in this case, and to add the results, as will become clear in the following. The transformation $(P_{\mathrm{A}}-g\cdot P_{\mathrm{A}}P_{\mathrm{B}})$ of y_{N} yields $(1-g)\mathrm{y}_{\mathrm{N}}$, so that the whole transformation yields $(1+g+g^2+\ldots)(1-g)\mathrm{y}_{\mathrm{N}} = = \mathrm{y}_{\mathrm{N}}$. Further, application of the iterative method to $\mathrm{y}-\mathrm{y}_{\mathrm{N}} = \mathrm{y}'$ yields, as first component in A, $\mathrm{y}'_{\mathrm{A}}-g\cdot\mathrm{y}'_{\mathrm{BA}}$, a vector in A*. The same will be true for the following components in A. This fact can be used as a check in the computations. The first component will be obtained in fact as $\mathrm{y}_{\mathrm{A}}-\mathrm{y}_{\mathrm{N}}-g(\mathrm{y}_{\mathrm{B}}-\mathrm{y}_{\mathrm{N}})_{\mathrm{A}}$.

A second advantage of this separation is that the computations lead much faster to the results; for the coordinates of $\mathrm{y}-\mathrm{y}_{\mathrm{N}}$ are very much smaller, in general, than those of y_{N}.

A third advantage appears in the consideration of those incomplete block designs for which the transformation $P_{\mathrm{A}}P_{\mathrm{B}}$ of certain vectors in A* is a multiplication by a real number μ. The transformation $Q = g\cdot P_{\mathrm{A}}P_{\mathrm{B}}$ is then a multiplication by $g\mu$.

It follows directly that the estimate of $E\mathrm{y}$ for balanced incomplete blocks (compare 3.4.2) is, in this case, equal to

$$\mathrm{y}_{s\mathrm{A}} = \mathrm{y}_{\mathrm{N}} + (1-g\mu)^{-1}\{\mathrm{y}_{\mathrm{A}}-\mathrm{y}_{\mathrm{N}}-g(\mathrm{y}_{\mathrm{B}}-\mathrm{y}_{\mathrm{N}})_{\mathrm{A}}\}, \quad \text{with } \mu = (r-\lambda)/rk.$$

Similarly for group divisible partially balanced incomplete blocks (compare 3.4.3):

$$\mathrm{y}_{s\mathrm{A}}=\mathrm{y}_{\mathrm{N}}+(1-g\mu_1)^{-1}\{\mathrm{y}_{\mathrm{A}}-\mathrm{y}_{\mathrm{N}}-g(\mathrm{y}_{\mathrm{B}}-\mathrm{y}_{\mathrm{N}})_{\mathrm{A}}\}_{\mathrm{A}1}+(1-g\mu_2)^{-1}\{\mathrm{y}_{\mathrm{A}}-\mathrm{y}_{\mathrm{N}}-g(\mathrm{y}_{\mathrm{B}}-\mathrm{y}_{\mathrm{N}})_{\mathrm{A}}\}_{\mathrm{A}2}=$$
$$=\mathrm{y}_{\mathrm{N}}+(1-g\mu_1)^{-1}\{\mathrm{y}_{\mathrm{A}}-\mathrm{y}_{\mathrm{N}}-g(\mathrm{y}_{\mathrm{B}}-\mathrm{y}_{\mathrm{N}})_{\mathrm{A}}\}+\{(1-g\mu_1)^{-1}-(1-g\mu_2)^{-1}\}\{\mathrm{y}_{\mathrm{A}}-\mathrm{y}_{\mathrm{N}}-g(\mathrm{y}_{\mathrm{B}}-\mathrm{y}_{\mathrm{N}})_{\mathrm{A}}\}_{\mathrm{A}1}$$

with $\mu_1 = (rk-\lambda_2 t)/rk$ and $\mu_2 = (r-\lambda_1)/rk$.

Latin square type partially balanced incomplete blocks (3.4.4) yield completely analogous results.

Returning to the general case with unequal g_i, we remark that, unless the average of g_i over the elements in any class of A is the same, $Q\mathrm{y}_{\mathrm{N}}$ is not in N. Separation of y in y_{N} and $\mathrm{y}-\mathrm{y}_{\mathrm{N}}$, and performance of the transformation on the second component only, will be an improvement, also in the general case, if one is interested only in the differences between the coordinates in $\mathrm{y}_{s\mathrm{A}}$. Attention to this computational aspect may be justified by the illustrative remark, that in an example the performance of the calculations on y as a whole required 64 steps, and on $\mathrm{y}-\mathrm{y}_{\mathrm{N}}$ 2 steps!

If, however, $\mathrm{y}_{s\mathrm{A}}$ is wanted, and not only the differences between its coordinates, then $\mathrm{y}_{s\mathrm{N}}$ will be determined first, and after that the operations are performed on $\mathrm{y}-\mathrm{y}_{s\mathrm{N}}$. The equation for $\mathrm{y}_{s\mathrm{N}}$, which is a particular case of that for $\mathrm{y}_{s\mathrm{A}}$, will be: $\mathrm{y}_{s\mathrm{N}}-(\mathrm{y}_{s\mathrm{N}})_{g\mathrm{N}} = \mathrm{y}_{\mathrm{N}}-(\mathrm{y}_{\mathrm{B}})_{g\mathrm{N}}$. Let the coordinate of $\mathrm{y}_{s\mathrm{N}}$ be denoted by $\hat{\mu}$, the coordinate of y_{N} by $\bar{y}$, and the coordinate of y_{B} in class i by $\bar{y}_i$. Then we have:

$$\hat{\mu}(1-\Sigma_i k_i g_i/n) = \bar{y}-\Sigma_i k_i g_i \bar{y}_i/n, \quad \text{or} \quad \hat{\mu}\{\Sigma_i k_i(1-g_i)\} = \Sigma_i \bar{y}_i k_i(1-g_i), \quad \text{or}$$
$$\hat{\mu}\,\Sigma_i(\sigma_1{}^2+\sigma_2{}^2/k_i)^{-1} = \Sigma_i \bar{y}_i(\sigma_1{}^2+\sigma_2{}^2/k_i)^{-1}.$$

This is also the solution of the first problem discussed in 4.1.1.

Another particular case is that in which the classifications A and B are orthogonal in the customary sense of chapter 3. $\mathrm{y}_{s\mathrm{N}}$ is obtained as before. For the calculation of y_{sA} we consider y_N and $y - y_N$ separately. Because Qy_N is in N (the average of g_i over the elements in any class of A is the same), the contribution to y_{sA} from y_N is y_N.

Usually one has the situation that all g_i are equal. In that case the result of the transformation of $y - y_N$ will be $y_A - y_N$; for then $(y_B - y_N)_{gA} = 0$, and thus the contribution from $y - y_N$ is $(I + Q + \ldots)(y_A - y_N)$, which is $y_A - y_N$, because $(y_{A^*})_B = 0$. Hence, in this case $y_{sA} = y_A$.

4.2.3. *A design of split-plot type*

Let the classes of a classification B of y be units of higher order (e.g. blocks), with a random variable for every unit. Let C be a classification of these units. Let A be another classification of y. Let A, B, and C be the corresponding spaces. We suppose that Ey is in $A \times C$. In addition to the non-orthogonality this design deviates from the usual split-plot design in that a second classification of the units of higher order and the corresponding main effect (effect of replicates in orthogonal designs) is absent.

The estimation of the effects of A and C and their interaction will take place by projection of y on $A \times C$, orthogonal with respect to the appropriate metric. But this is the same problem as that in the previous sections: we need only replace A by $A \times C$.

In the application of the iterative method with the classifications B and $A \times C$ we may arrange y according to the classes of C (B is a subclassification of C). It follows that the method falls apart in separate applications of the method for every class of C. The vector in $A \times C$ obtained in this way can be decomposed in components in N, A*, C* and (A × C)* according to the methods of chapter 3.

It will be clear that the estimation is simplified, if in any class of C the units (blocks) have equal k_i, and thus equal g_i. If, moreover, in any class of C the classification according to A is orthogonal to the classification according to blocks, then the effect of $A \times C$ will be found by taking averages in every class of $A \times C$, as follows from the discussion in the foregoing section (whatever be σ_1 and σ_2). If, moreover, all classes of A are represented in every class of C, the decomposition of $y_{sA\times C}$ is also simple because of the orthogonality of A and C.

4.3. TWO CONSTANT MAIN EFFECTS, AND A RANDOM EFFECT

4.3.1. *The general case*

The classes of a classification B of y are units of higher order with a random variable with expectation zero and variance σ_1^2 for every unit. Let C be a classification of these units. Let A be another classification of y. We suppose that Ey is in $A + C$. The difference with the preceding problem is that interaction between A and C is not present.

A first example of this situation is that of a block design with treatments A, in which adjacent blocks (units) are taken together in classes, in order to take into account a great part of the variation between the blocks as a non-random block effect, while the effect of the blocks (units) within these classes is considered as a random variable. One hopes that the variance of this variable is much smaller than (as in 4.2.1) without the classification C. Further important examples will be considered later on (4.3.4 and 4.3.5).

The estimation of Ey will take place by determining y_{sA} in A and y_{sC} in

C such that the orthogonal projection of $y - y_{sA} - y_{sC}$, according to the appropriate metric, on A and on C is the null vector. With the same notation as in 4.2.1 we obtain the equations:

$$\Sigma_i\{w_i(y - y_{sA} - y_{sC})_{B_i}\}_A + w\{y - y_{sA} - y_{sC} - (y - y_{sA} - y_{sC})_B\}_A = 0 \quad (1)$$

$$\Sigma_i\{w_i(y - y_{sA} - y_{sC})_{B_i}\}_C + w\{y - y_{sA} - y_{sC} - (y - y_{sA} - y_{sC})_B\}_C = 0 \quad (2)$$

Because C is a subspace of B, so that $P_C P_B$ is equal to P_C, equation (2) is equivalent to: $\Sigma_i(w_i/w)\,(y - y_{sA} - y_{sC})_{B_i C} = 0$. We introduce the projection on C, which maps x as $z = x_{wC}$, with the property: the coordinate of z in a class of *C* is equal to the weighted average of the coordinates of x in that class of *C*, while the weight of every coordinate of x is equal to the number $w_i/w = 1 - g_i = g_i' = \sigma_2^2/(k_i\sigma_1^2 + \sigma_2^2)$ of the class of *B* to which it belongs.

In terms of this projection the last equation can be replaced by

$$(y - y_{sA} - y_{sC})_{wC} = 0 \qquad (2')$$

The left hand sides of (2′) and the former equation are not equal nor proportional, in general.

In passing it may be remarked that we have obtained the solution of the second example in 4.1.1 where A and C *both* are subspaces of B: for equation (1) is then $(y - y_{sA} - y_{sC})_{wA} = 0$, so that y_{sA} and y_{sC} can be found by the iterative method of 3.3 with A and C, attaching weights g_i' to the coordinates of y; this is equivalent to the application of that iterative method on a vector consisting of the class averages of *B* in y (every class counted once) attaching the weights $(\sigma_1^2 + \sigma_2^2/k_i)^{-1}$ to them.

Continuing with the present case we write the modified equation (2′) as: $y_{sC} = y_{wC} - (y_{sA})_{wC}$. Thus y_{sC} will be found easily, if y_{sA} is known. Let the transformation indicated by *g*A have the same meaning as in 4.2.1, and let the transformation indicated by *g*′A be a similar transformation, with the difference only that the numbers g_i are replaced by g_i'; the sum of these transformations is the orthogonal projection on A. Now equation (1) may be reduced as follows:

$$(y_B)_{g'A} - \{(y_{sA})_B\}_{g'A} - (y_{sC})_{g'A} + y_A - y_{sA} - (y_{sC})_A - y_{BA} + (y_{sA})_{BA} + (y_{sC})_A = 0.$$

Substituting from equation (2′): $(y_{sC})_{g'A} = (y_{wC})_{g'A} - \{(y_{sA})_{wC}\}_{g'A}$ we obtain by some reduction: $y_{sA} - [\{(y_{sA})_B\}_{gA} + \{(y_{sA})_{wC}\}_{g'A}] = y_A - [(y_B)_{gA} + (y_{wC})_{g'A}]$.

This can be summarized in $(1 - Q)y_{sA} = y_A - Qy$ where *Q* is the linear transformation with $Qz = (z_B)_{gA} + \{(z_B)_{wC}\}_{g'A}$.

Q can be described as follows: to every class *i* of *B* belong numbers

$$g_i = k_i\sigma_1^2/(k_i\sigma_1^2 + \sigma_2^2) \quad \text{and} \quad g_i' = \sigma_2^2/(k_i\sigma_1^2 + \sigma_2^2).$$

The first step in *Q* (on z) is the orthogonal projection on B (averaging within classes of *B*) and yields z_B. On the resulting vector z_B the projection *w*C is performed; the coordinate of a class of *C* will be found as the weighted mean of the coordinates of z_B in that class of *C* with weights g_i', or what amounts to the same, as the weighted mean of the means of the classes of *B* in that class of *C* (every class of *B* counted once) with weights $(\sigma_1^2 + \sigma_2^2/k_i)^{-1}$. So we obtain $(z_B)_{wC}$. Then follows a multiplication of the coordinate of z_B in class *i* of *B* by g_i, and similarly of $(z_B)_{wC}$ by g_i'. Addition of these products, class by class of *B*, yields a new vector, $z' = Rz$, in B. Orthogonal projection of z′ on A (averaging within classes of *A*) completes *Q*.

We conclude that, in comparison with the transformation $P_A P_B$, the transformation P_B has been extended with a projection on the subspace C of B, and with the formation of linear combinations of the coordinates of the results of these two projections in every class of B. Within every class i of B the coordinates of z_B and $(z_B)_{wC}$ are averaged with weights g_i and g_i' respectively.

We observe that Q transforms a vector in N in itself. Let $P_B z$ have coordinates b_i in class i of B. Then the vector z′ has as coordinate in class i of B:

$$c_i = g_i b_i + g_i' \frac{\Sigma g_i' b_i}{\Sigma g_i'},$$

where the summation pertains all the coordinates in the class of C, to which class i of B belongs. It follows that

$$\Sigma c_i = \Sigma g_i b_i + (\Sigma g_i') \frac{\Sigma g_i' b_i}{\Sigma g_i'} = \Sigma b_i.$$

Therefore, if $P_B z$ is orthogonal to N, the same holds for z′. Hence, if z is in A*, then Qz is in A* too.

The right hand member of the equation for y_{sA} may be written as

$$y_N + y_{A^*} - Q y_N - Q y' = y_{A^*} - Q y',$$

a vector in A*. If we know (which will be proved) that the bound of the transformation Q on vectors in A* is smaller than 1, then it follows that the solution of the equation is: $y_{sA} = (1 + Q + Q^2 + \dots)(y_A - Qy)$. This implies an iterative procedure similar to preceding cases. From the fact that $y_A - Qy$ and all the following components in A, obtained by this procedure, are orthogonal to N, we have a check on the computations. Because the projection wC will be performed on y and on all the components of which the sum is y_{sA}, all material for the computation of $y_{sC} = y_{wC} - (y_{sA})_{wC}$ comes also available.

4.3.2. *The bound of Q*

We will prove that the bound $G(Q)$ of the transformation Q of vectors in A* obeys $G(Q) < 1$. Let z be in A*, and $z' = Rz$ in B such that $Qz = P_A z' = P_A R z$. In case $|Rz| < |P_B z|$ we have $|Qz| = |P_A R z| \leqq |Rz| < |P_B z| \leqq |z|$; hence $|Qz| < |z|$. Further, we will find that, if $|Rz| = |P_B z|$, then $P_B z = P_C z$; this implies (from the definition of R): $Rz = P_C z$ and therefore $Q = P_A P_C$; then $|Qz| = |P_A P_C z| < |z|$, because A* and C* are disjoint (compare 3.3.4). From these considerations it follows that it is sufficient to prove that $|Rz| \leqq |P_B z|$ for all z in A*, with equality only if $Rz = P_C z$.

For this purpose we consider the relation between the coordinates c_i and b_i of $Rz = z'$ and $P_B z$, respectively, in class i of B given in the last section: $c_i = g_i b_i + g_i'(\Sigma g_i' b_i)/\Sigma g_i'$, and we prove that $\Sigma c_i^2 \leqq \Sigma b_i^2$ (where the summations pertain a class of C), with equality only if all b_i (within this class of C) are equal.

We have the inequality: $0 < g_i \leqq 1$. Then $b = \Sigma g_i' b_i' / \Sigma g_i'$ is not larger than the largest, and not smaller than the smallest of the numbers b_i; for if we denote the numbers b_i smaller than b by b_i', and those larger than b by b_i'', then in the identity $\Sigma' g_i'(b - b_i') = \Sigma'' g_i'(b_i'' - b)$ every term must be positive (the identity $0 = 0$ arises, if all b_i are equal).

From $c_i = g_i b_i + g_i' b = b_i - g_i'(b_i - b)$ it follows that every c_i arises by subtracting from b_i a part $(0 < g_i' \leqq 1)$ of its difference with b: every c_i is closer to b than the corresponding b_i. The total of the subtracted values amounts to

$\Sigma g_i'(b_i - b) = 0$, which has been used already in the preceding section. If $b = 0$, the assertion is thus trivial.

Now we assume $b > 0$ (for $b < 0$ an analogous argument may be given). First we observe that the c_i corresponding to $b_i \leqq -b$ will give a smaller contribution to Σc_i^2 than the corresponding b_i in Σb_i^2.

The c_i corresponding to the b_i in the interval $-b < b_i \leqq b$ may give a larger contribution to Σc_i^2 than the corresponding b_i to Σb_i^2. Let these b_i be denoted by the index j: b_j. Let the corresponding $-g_j'(b_j - b)$, which is positive or zero for every j, be denoted by d_j. Then the increase of the sum of squares is equal to $\Sigma_j(b_j + d_j)^2 - \Sigma_j b_j^2 = 2\Sigma_j b_j d_j + \Sigma_j d_j^2$ which, because $b_j \leqq b - d_j$ for every j, is at most $2b\Sigma_j d_j - \Sigma_j d_j^2$.

The c_i corresponding to $b_i > b$ will give a smaller contribution to Σc_i^2 than the corresponding b_i to Σb_i^2. Let these b_i be denoted by the index k, b_k, and let the corresponding $g_k'(b_k - b)$, which are positive, be denoted by d_k. Then the decrease in the sum of squares is equal to $\Sigma_k b_k^2 - \Sigma_k(b_k - d_k)^2 = 2\Sigma_k b_k d_k - \Sigma_k d_k^2$, which, because $b_k > b + d_k$ for every k, is larger than $2b\Sigma_k d_k + \Sigma_k d_k^2$.

Now we compare the upper bound of the increase, $2b\Sigma_j d_j - \Sigma_j d_j^2$, with the lower bound of the decrease, $2b\Sigma_k d_k + \Sigma_k d_k^2$. Because $\Sigma_j d_j$, the absolute value of the sum of the quantities subtracted from the b_j is at most equal to $\Sigma_k d_k$, the sum of the quantities subtracted from the b_k (the sum of the quantities subtracted from all the b_i was zero), the increase is smaller than the decrease, unless all d_k and thus all d_j are zero. Hence $\Sigma c_i^2 \leqslant \Sigma b_i^2$ with equality only if all b_i in class i of C are equal.

4.3.3. *Particular cases*

It follows from the transformation Q that, if σ_1/σ_2 is large, the orthogonal projection on B plays a preponderable rôle in the estimation of $\mathbf{y}_{sA}$. If σ_1 is small, the vector in C, which is very much similar to the orthogonal projection on C (g_i nearly 1) then, comes into prominence.

In the extreme case that $\sigma_1 = 0$, and thus $g_i = 0$ and $g_i' = 1$, the transformation wC is the orthogonal projection on C, and $Q = P_A P_C$. The projection on B will not be used at all. The equation for $\mathbf{y}_{sA}$ becomes completely the same as that in 3.3 for two classifications, now A and C. Further $\mathbf{y}_{sC}$ will be $\mathbf{y}_C - (\mathbf{y}_{sA})_C$ as in chapter 3.

In the other extreme case, approximated in case σ_1/σ_2 is large, $g_i = 1$ and $g_i' = 0$. Then the projection on C is absent in Q, so that $Q = P_A P_B$. We obtain the method of 3.3 with the classifications A and B in order to determine $\mathbf{y}_{sA}$. Further $\mathbf{y}_{sC}$ will be found as $(\mathbf{y} - \mathbf{y}_{sA})_{wC}$. If σ_1 is large, the numbers g_i' are proportional to $1/k_i$. The coordinate in a class of C will be calculated as follows: add the coordinates of a class of B and multiply the sum by $1/k_i$, i.e. take the class means of B; next add the class means of B in this class of C and divide by $\Sigma k_i(1/k_i)$, that is the number of classes of B in this class of C. The projection wC means thus averaging the class means of B within every class of C, every class of B counted once.

A condition that simplifies the calculations is equality of the numbers k_i. Then $Q = P_A(gP_B + g'P_C)$; the numbers g_i are replaced by g and the numbers g_i' by g'. A further simplification is obtained, if moreover all classes of A are proportionally represented in every class of C. This, for instance, is the case, if C represents a so-called classification according to replicates in an incomplete block design. Then P_C will only yield a contribution to y_{sA} in the first step. The first step yields $u_1 = y_A - (gy_B + g'y_C)_A$; the second $u_2 = g(u_1)_{BA}$; the third $u_3 = g(u_2)_{BA}$ and so on. We observe that, from the second step on, the operations are equal to those for an incomplete block design without the classification C (compare 4.2.2).

Let such a classification according to replicates be applied in the designs (compare 3.4.2, 3.4.3 and 3.4.4), for which we found explicit solutions for y_{sA} in 4.2.2. Then the estimate y_{sA} will be given by replacing the expression $y_A - y_N - g(y_B - y_N)_A$ in the solutions of 4.2.2 by $y_A - (gy_B + g'y_C)_A$, and omitting the first term y_N. This component will be represented in the component y_{sC}.

4.3.4. *Split-plot designs*

The classes of a classification B of y are units of higher order (e.g. blocks) with a random variable for every unit. Let C be a classification of these units according to a characteristic C (e.g. adjacent blocks form a class of C), and D another classification of these units according to a characteristic D (the main factor). Further there is another classification A in y (the split-plot factor). We suppose that Ey is in $C + (A \times D)$.

The problem to determine the best estimate of Ey can immediately be reduced to that of the preceding sections. We need only replace the space $A \times D$ of this problem by the space A in the foregoing problem. To every block (main plot) belong numbers g_i and g_i' which are necessary for the transformation Q. In this occur successively an orthogonal projection on B, a projection on the subspace C of B (weights g_i'), the formation of a new vector in B from these two projections, and finally an orthogonal projection on $A \times D$. In this way we find $y_{sA\times D}$, which can be decomposed according to chapter 3 in components in N, A*, D*, and $(A \times D)^*$, if desired. Further y_{sC} will be found as $(y - y_{sA\times D})_{wC}$.

Equality of the k_i is a simplification: the projection on C is now orthogonal and $Q = P_{A\times D}(gP_B + g'P_C)$, as in the last section. If, moreover, the classification according to A is orthogonal to the classification according to blocks in every class of D, then, from the second step on, all components of $y_{sA\times D}$ are in D (as follows from the orthogonal projection of a vector in B on $A \times D$). The third and following components in $A \times D$ are of the type: $u_3 = \{gu_2 + g'(u_2)_C\}_D$. Hence only projections on C and D occur. Moreover, the first step contains the whole main effect of A and the interaction of A and D. If, moreover, C is orthogonal to D, then the second term in the third and following components vanishes; the sum of the second and all following components is equal to $(1 - g)^{-1}u_2$. If, moreover, every class of D contains the same classes of A, we have orthogonality of A and D, of D and C, and hence of $A \times D$ and C. Now we find, with $y - y_N = y'$,

$$u_1 = y'_{A\times D} - (gy'_B + g'y'_C)_{A\times D} = y'_{A\times D} - gy'_D;$$
$$u_2 = g(y'_{A\times D} - gy'_D)_D = g(1 - g)y'_D; \quad \Sigma_{i=2}^{\infty} u_i = (1 - g)^{-1}u_2 = gy'_D.$$

Hence we will find $y_{sA\times D} = y'_{A\times D} = y_{A\times D} - y_N$. This very particular situation is the one usually indicated with the term split-plot designs.

4.3.5. *Series of experiments and similar problems*

Now we consider an extension of the first problem discussed in 4.1.1, in this sense that there are trials with more than two varieties (or other characteristics). We assume that every trial gives *one* measurement (the mean) of the yield of every variety in that trial. Different trials in the same year do not all contain the same set of varieties for many practical reasons, while trials in different years will not contain the same set of varieties by virtue of the nature of varietal research, which implies the introduction of new varieties. For purposes of prediction and advice the best estimate of the expectation of varietal differences is necessary.

Consider the classification of the yields y according to years on the one hand, and according to varieties on the other hand. It will be expected that, analogously to the case in 4.1.1, the expectations of the differences between varieties are not the same in different years, in other words, that there is interaction between varieties and years. Because, however, this interaction (which in 4.1.1 was a main effect of years) cannot be predicted, a random variable is introduced for every variety: the expectation of the yield of that variety in the considered area in a random year. The estimation of the expectations (over all years) of these random variables, but for a constant, will be our purpose.

Each of these random variables is equal to the sum of the varietal expectation and a random contribution from the year to this variety. The random contribution is supposed to be equal to the sum of a non-random contribution of the year common to all varieties, and a random variable, with expectation 0 and variance σ_1^2, both the same for all variables. Moreover, these variables are supposed to be uncorrelated. In other words, the vector of expectations for variety-year combinations has as expectation the sum of an effect of varieties and an effect of years; to every combination a random variable is attached. All these random variables have expectation zero, covariance zero and variance σ_1^2.

Every variety-year combination in y has been divided according to the trials in that year. This leads to the supposition that the coordinates, under the condition that they are in a fixed variety-year combination, have an expectation equal to the sum of the (now) fixed expectation, corresponding to this combination, and a contribution of the trial, common to all varieties in that trial. Further they have, under the same condition, covariance zero and variance σ_2^2.

The whole supposition may be summarized as follows: the coordinate of y belonging to some variety-trial combination is equal to the sum of a constant determined by the variety, a constant determined by the trial (the effect of years is included), a random variable with expectation zero common to the coordinates in the variety-year combination (these variables have covariance zero and variance σ_1^2) and a random variable for this particular coordinate (these variables have, under the condition that the expectation for the variety-year combination has its particular value, expectation zero, covariance zero and variance σ_2^2).

One might ask, why the effect of trials is not considered as a random variable too, because this effect (of trials and years) has a random character, and we are not interested in the effect of any particular year or trial field. The answer is in the foregoing sections: we observed, that when such an effect is assumed random and the associated variance is large, then the estimators can be approximated by the estimators under the assumption that random effects are constant effects. Because the effects of trials (and years) are very large in comparison with σ_1 and σ_2, we suppose them to be constants from the beginning.

It follows from our suppositions that the problem of the best estimate for varietal differences is a special case of the foregoing problems. The variety-year combinations are now the units of higher order (to be compared with blocks in the case of 4.3.1); they arise from the classification *B*. The classification according to varieties corresponds to the classification *C* of the classes of *B*. The classification according to trials corresponds to the classification *A*. We need the orthogonal projection on the space spanned by A and C. A difference with the preceding cases is that there the component in A is the most interesting, while here we aim at the component in C. The component in A is the sum of a component in D, the space of year effects, and a residual named trial effect within years. This is in contrast with the split-plot design, in which the effect corresponding to *A* consisted of two main effects and an interaction.

The solution of the estimation problem requires no difficulties now: to every variety-year combination (every class of $C \times D = B$) consisting of k_i coordinates belong numbers g_i and g_i'. The transformation Q consists in: orthogonal projection on B (varietal means for every year); projection on C (weighted means of the annual means for each variety with weights $(\sigma_1^2 + \sigma_2^2/k_i)^{-1}$); a combination of these two projections to a new vector in B by means of the numbers g_i and g_i'; finally orthogonal projection of the result in B on A.

In order to get a survey of the computation one may write the data for every year (arranged according to trial and variety) on different sheets, while there is a special sheet on which the vectors in C will be noted. It appears that the projections on B will be performed in every year (sheet) separately; when the projection of the result in B on C is performed, we pick one or none number from every sheet of the years, and the result is noted on the special sheet. By means of this last result and one of the other sheets every time, the required vector in B will be formed. The projection of this vector on A will be performed on every sheet separately. When the process has come to an end (all coordinates zero), the estimate y_{sC} will be found on the special sheet by subtracting from the first vector in C all the following vectors in C.

In the extreme case $\sigma_1 = 0$ i.e. no interaction between variety and year, Q will be (compare 4.3.3) $P_A P_C$. This means the application of the iterative method of 3.3 on the complete classification according to trials and to varieties, irrespective of the number of trials in which a variety is represented in a year. In the other extreme case of a very large interaction between variety and year, we obtain with reference to 4.3.3: In every year separately the iterative method of 3.3 will be applied in order to find y_{sA}, i.e. the effect of the trials within a year. Next $y_{sC} = (y - y_{sA})_{wC}$, which means that the coordinates of the effect of *C* for every year separately (found by the method of 3.3), also called the varietal means "adjusted for trial effects" for every year, are simply averaged, variety by variety, over the years, every year which contains that variety counted once.

Now we consider the particular case, in which any two trials within a year contain the same varieties (one coordinate for a variety-trial combination as before); the varieties in different years may be different. Then the transformation Q, which ends with an orthogonal projection of a vector in B on A, will yield the same coordinates within a year, in other words, yields a vector in D. The first step in Q is an orthogonal projection on B, which reduces for this vector to the transformation *1*. As $u_2 = Qu_1$ is also in D, $u_3 = Qu_2$ reduces to $Qu_2 = = (u_2)_{gD} + \{(u_2)_{wC}\}_{g'D}$, and so on. Because the effect of the trials is not interesting in the problem, we will simplify in replacing $u_1 = y_A - Qy$ by $y_{AB} - Qy$. Then the vectors u_2, u_3, ... obtained from this vector are equal to those obtained from the original u_1; for the orthogonal projection of $y_A - Qy$ on B is equal to $y_{AB} - Qy$. For the computations we observe $y_{AB} = y_D = y_{BD}$; hence the new $u_1 = y_{BD} - (y_B)_{gD} - (y_{wC})_{g'D} = (y_B - y_{wC})_{g'D}$. Now we work with the array of annual varietal means only. From the sum of the varietal means in year i the corresponding components in $y_{wC} = (y_B)_{wC}$ are subtracted, and the difference is multiplied by g_i'/m_i, where m_i is the number of varieties in that year. The resulting vector is u_1. Further $u_2 = (u_1)_{gD} + \{(u_1)_{wC}\}_{g'D}$, i.e. g_i'/m_i times the sum of the coordinates of $(u_1)_{wC}$, corresponding to varieties occurring in year i, is added to g_i times the coordinate of u_1 in year i. Similarly $u_3 = (u_2)_{gD} + \{(u_2)_{wC}\}_{g'D}$ and so on. Finally $y_{sC} = y_{wC} - \Sigma_{i=1}^{\infty} (u_i)_{wC}$.

We will show that this method is not the same as the following which sometimes has been proposed. To every annual varietal mean the weight $(\sigma_1^2 + \sigma_2^2/k_i)^{-1}$ is attached. Next the method of 3.3 is applied with the classifications C and D.

For we observe that, in the extreme case of a very large σ_1, the weights in different years will be equal, so that also then the iterative method of 3.3 would be applied. The right procedure, however, then consists only in simple averaging the annual means variety by variety. One might say that the proposed application of the method of 3.3 tends to a too great adjustment for year effects. If $\sigma_1 = 0$, the two methods are equivalent.

In case not only the trials within a year, but also the different years contain the same varieties, C is orthogonal to A and to D. Then y_{sC} is $y_{wC} - \Sigma_{i=1}^{\infty} (u_i)_{wC}$ again. But every $(u_i)_{wC}$ will be formed from a vector in D and thus is in N. Therefore, if the level of y_{sC} is not important, then y_{wC} yields all information required. If, moreover, the k_i are equal, $y_{wC} = y_C$.

Another particular situation in the case of the same varieties within a year (but not in different years) occurs, if the k_i for all years are equal. Then we find: $u_1 = g'(y_B - y_C)_D = g'(y_D - y_{CD})$; $u_2 = gu_1 + g'(u_1)_{CD}$;
$u_3 = gu_2 + g'(u_2)_{CD}$ and so on; $y_{sC} = y_C - \Sigma_{i=1}^{\infty} (u_i)_C$.

This method can be applied, if D corresponds to randomized block trials in one year in some area, A to the blocks (of which there are a constant number in every trial), C to varieties, and B to trial-variety combinations. The usual problem is now the estimation of the expectation of the varietal differences in that year over that area, taking into account a random interaction between trials and varieties.

The method of this section also applies in the following case. The classification C corresponds to trials in some year, A to varieties and D to more or less homogeneous groups of these varieties (e.g. with about the same sensitiveness to drought or to a disease, or of botanically very much related varieties). Further B corresponds to trial-group combinations. The supposition about the yields y is, that every coordinate is the sum of an effect of the corresponding variety, an effect of the corresponding trial (place), a random contribution common to the coordinates within a place-group combination (interaction between groups of varieties and places) and a random contribution for every coordinate separately. In the problem of estimation of Ey, assumed in A + C, the component in A will be the most interesting. The extreme $\sigma_1 = 0$ leads to the iterative method of 3.3 with the classifications A and C. The other extreme leads to separate applications of the iterative method of 3.3 with the subclassifications A and C in every class of D; the difference between groups of varieties will be estimated as zero, which is the best that may be said in this case indeed.

4.4. ESTIMATION OF THE VARIANCES

4.4.1. *General remark*

Up to now we have considered the quantities σ_1 and σ_2 as given. Mostly however, they must be taken, or rather be estimated from y itself. If we use the estimates for these quantities in the expressions of the preceding two sections, we replace constants by random variables. This procedure, however, will be applied, only if the variance of the estimate of σ_1^2 and σ_2^2 is sufficiently small. The effect, of using these random variables instead of constants, on the properties of the estimate of Ey or of its components will not be considered.

It will appear that the estimates of σ_1^2 and σ_2^2 will be obtained from the squares of orthogonal projections of y (or of vectors which arise by omitting coordinates from y) on orthogonal subspaces. We adopt the rule of thumb (compare KEMPTHORNE [9]) that the effect of using the estimates instead of the true values is negligible, if the dimension of these subspaces is 10 or more. If the dimension of the space, on which the orthogonal projection supplies the estimate for σ_1^2, is too small (the other will be large enough nearly always), then the random effects will be considered as constant effects.

4.4.2. *A constant main effect and a random main effect*

First we consider the case considered in 4.2.1. Let the coordinates of y be $y_{ij} = \alpha_j + \beta_i + \varepsilon_{ij}$, where α_j is the constant associated with class j of A, β_i the random variable (with expectation zero and variance σ_1^2) associated with class (block) i of B, and ε_{ij} the random variable (with expectation zero and variance σ_2^2) that belongs to the combination of the characteristic j of A and i of B. (We assume for the present that every class of A is represented at most once in any class of B). Let n_j be the number of coordinates in class j of A, so that $\Sigma_j n_j = n$. Let the number of classes of A be v and that of B be b.

Under the condition that the β_i have fixed values, Ey_{R}^2 (R is the residual space orthogonal to A and B) is

$$[n - 1 - (b - 1) - (v - 1)]\sigma_2^2 = (n - b - v + 1)\sigma_2^2.$$

Because the condition does not occur in this expression, the same is true unconditionally.

Further $Ey^2 = E\,\Sigma_{ij}(\alpha_j + \beta_i + \varepsilon_{ij})^2 =$ (because covariances are zero, all double products vanish) $= \Sigma_j n_j \alpha_j^2 + n\sigma_1^2 + n\sigma_2^2$. And

$$Ey_{\mathrm{A}}^2 = E\,\Sigma_j \{n_j^{-1}\,\Sigma_i(\alpha_j + \beta_i + \varepsilon_{ij})^2\},$$

where Σ_i pertains the coordinates in class j of A. Then

$$Ey_{\mathrm{A}}^2 = \Sigma_j\, n_j^{-1}\, E(n_j\alpha_j + \Sigma_i\beta_i + \Sigma_i\varepsilon_{ij})^2 =$$
$$\Sigma_j n_j^{-1}(n_j^2\alpha_j^2 + n_j\sigma_1^2 + n_j\sigma_2^2) = \Sigma_j n_j\alpha_j^2 + v\sigma_1^2 + v\sigma_2^2.$$

Hence $E(\mathrm{y}^2-\mathrm{y}_\mathrm{A}^2) = (n-v)\,\sigma_1^2 + (n-v)\sigma_2^2$. We observe that the numbers k_i in class i of B do not matter.

An unbiased estimate of σ_2^2 is $\mathrm{y}_\mathrm{R}^2/(n-v-b+1)$ and an unbiased estimate of σ_1^2 is $(\mathrm{y}^2-\mathrm{y}_\mathrm{A}^2)/(n-v)-\mathrm{y}_\mathrm{R}^2/(n-v-b+1)$.

We remark that, if in this and the following cases the expression for the estimator of σ_1^2 will yield a negative number, then the estimate will be taken equal to 0.

According to the rule of thumb (KEMPTHORNE [9]) given for special cases, $b-1$ must be larger than 10, in order that we will permit ourselves to use the random effect.

If classes of A are represented more than once in a class of B, then $E\mathrm{y}_\mathrm{A}^2$ will be different from the expression given above, because $\Sigma_i\beta_i$ for such a class j of A contains a multiple of some β_i. The expression must be calculated for every individual case.

A particular case is that where A = N (first example 4.1.1), so that every class of "A" is represented k_i times in class i of B. Then

$$E\mathrm{y}_\mathrm{A}^2 = E\mathrm{y}_\mathrm{N}^2 = En^{-1}\{\Sigma(\alpha+\beta_i+\varepsilon_{ij})\}^2 = n^{-1}(n^2\alpha^2+\Sigma_i k_i^2\sigma_1^2+n\sigma_2^2) =$$
$$= n\alpha^2+n^{-1}\Sigma_i k_i^2\sigma_1^2+\sigma_2^2.$$

Hence $E(\mathrm{y}^2-\mathrm{y}_\mathrm{N}^2) = (n-n^{-1}\Sigma_i k_i^2)\sigma_1^2 + (n-1)\sigma_2^2$. An unbiased estimate for σ_1^2 will be $(n-1)\,(n-n^{-1}\Sigma_i k_i^2)^{-1}\,\{(\mathrm{y}^2-\mathrm{y}_\mathrm{N}^2)/(n-1)-(\mathrm{y}^2-\mathrm{y}_\mathrm{B}^2)/(n-b)\}$. In the special case that all k_i are equal to k (hence $kb=n$), $n-n^{-1}\Sigma_i k_i^2$ is equal to $kb-k^2b/kb = k(b-1)$.

The design of split-plot type, discussed in 4.2.3, will be treated similarly. In order to estimate σ_2^2 the perpendicular on the space spanned by B and A × C is necessary. It is found by separate application of the method of 3.3 on each class of C, according to the subclassifications A and B. The estimation of σ_1^2 will take place by means of $\mathrm{y}^2-\mathrm{y}_{\mathrm{A}\times\mathrm{C}}^2$.

4.4.3. *Two constant main effects and a random effect*

The different cases discussed in 4.3 require separate consideration with respect to the estimation of the variances, especially of σ_1^2. First we consider the block designs with a classification C of the blocks and another classification A.

The estimation of σ_2^2 will take place by means of the perpendicular from y on the space A+B. For the estimation of σ_1^2 we take as many as possible coordinates from y, such that the classes of A are represented proportionally in every class of C (for the particular incomplete block designs, such as balanced or partially balanced incomplete block designs, this is true for the complete y). In the vector space corresponding to this mutilated y, we determine the square of the orthogonal projection on the corresponding space (A × C)*. When we know the expectation of this square, σ_1^2 can be estimated. In the calculation of this expectation the vector to be projected will be denoted by y, whether it is a mutilated y or not.

The supposition about the coordinates of y is $y_{ijk} = \gamma_i+\alpha_k+\beta_{ij}+\varepsilon_{ijk}$, where γ_i is the constant associated with class i ($i = 1, \ldots, r$) of C, α_k is the constant belonging to class k ($k = 1, \ldots, v$) of A, β_{ij} the random variable associated with block j in class i of C, also named block ij, and ε_{ijk} the random variable for the coordinate of class k of A in block ij. We assume that class k of A is represented at most once in block ij. Let the number of coordinates in block ij be k_{ij}. Then the number of coordinates in class i of C, k_i, is $\Sigma_j k_{ij}$, and $\Sigma_i k_i = n$. Let the number of coordinates, which are together in class k of A and in class i of C, be r_{ik}. Then $\Sigma_k r_{ik} = k_i$, and the number of coordinates in class k of A is $r_k = \Sigma_i r_{ik}$. Because of the orthogonality of A and C we have: $r_{ik}/k_i = r_k/n$ for all i. From the relation $\Sigma_i k_i \gamma_i = 0$ (C^* orthogonal to A) follows $\Sigma_i r_{ik} \gamma_i = 0$ for every k.

$$\begin{aligned}
\text{Now } Ey^2 &= \Sigma_{ijk} E(\gamma_i + \alpha_k + \beta_{ij} + \varepsilon_{ijk})^2 = \\
&= \Sigma_i k_i \gamma_i^2 + 2\Sigma_{ik} \gamma_i \alpha_k + \Sigma_k r_k \alpha_k^2 + n\sigma_1^2 + n\sigma_2^2 = \\
&= \Sigma_i k_i \gamma_i^2 + \Sigma_k r_k \alpha_k^2 + n\sigma_1^2 + n\sigma_2^2.
\end{aligned}$$

$$\begin{aligned}
Ey_{\mathrm{A}}^2 &= \Sigma_k \left[(1/r_k)\, E\,(\Sigma_{ij}\gamma_i + \Sigma_{ij}\alpha_k + \Sigma_{ij}\beta_{ij} + \Sigma_{ij}\varepsilon_{ijk})^2\right] = \\
&\qquad (\text{because } \Sigma_i r_{ik} \gamma_i = 0) \\
&= \Sigma_k\,(1/r_k)\,(r_k^2 \alpha_k^2 + r_k \sigma_1^2 + r_k \sigma_2^2) = \Sigma_k r_k \alpha_k^2 + v\sigma_1^2 + v\sigma_2^2.
\end{aligned}$$

$$\begin{aligned}
Ey_{\mathrm{C}}^2 &= \Sigma_i \left[(1/k_i)\, E\,(\Sigma_{jk}\gamma_i + \Sigma_{jk}\alpha_k + \Sigma_{jk}\beta_{ij} + \Sigma_{jk}\varepsilon_{ijk})^2\right] = \\
&= \Sigma_i[(1/k_i)\{k_i^2\gamma_i^2 + 2k_i\gamma_i\,\Sigma_k r_{ik}\alpha_k + (\Sigma_k r_{ik}\alpha_k)^2 + \Sigma_j k_{ij}^2\sigma_1^2 + k_i\sigma_2^2\}] = \\
&= \Sigma_i k_i \gamma_i^2 + \Sigma_i k_i \{\Sigma_k (r_{ik}/k_i)\alpha_k\}^2 + \Sigma_i\,(\Sigma_j k_{ij}^2/k_i)\sigma_1^2 + r\sigma_2^2 = \\
&= \Sigma_i k_i \gamma_i^2 + (1/n)\,(\Sigma_k r_k \alpha_k)^2 + \Sigma_i\,(\Sigma_j k_{ij}^2/k_i)\sigma_1^2 + r\sigma_2^2.
\end{aligned}$$

$$\begin{aligned}
Ey_{\mathrm{N}}^2 &= (1/n)\, E\,(\Sigma_{ijk}\gamma_i + \Sigma_{ijk}\alpha_k + \Sigma_{ijk}\beta_{ij} + \Sigma_{ijk}\varepsilon_{ijk})^2 = \\
&= (1/n)\,\{(\Sigma_k r_k \alpha_k)^2 + \Sigma_{ij} k_{ij}^2 \sigma_1^2 + n\sigma_2^2\} = \\
&= (1/n)\,(\Sigma_k r_k \alpha_k)^2 + (\Sigma_{ij} k_{ij}^2/n)\sigma_1^2 + \sigma_2^2.
\end{aligned}$$

It follows that $Ey_{(\mathrm{A}\times\mathrm{C})^*}^2 = Ey^2 - Ey_{\mathrm{A}}^2 - Ey_{\mathrm{C}}^2 + Ey_{\mathrm{N}}^2 =$

$$[n - v - \Sigma_i(\Sigma_j k_{ij}^2/k_i) + (\Sigma_{ij} k_{ij}^2/n]\sigma_1^2 + (n - v - r + 1)\sigma_2^2.$$

We observe that, if $\mathrm{C} = \mathrm{N}$, the expression for $E(y^2 - y_{\mathrm{A}}^2)$ in the preceding section will be found. In the particular case that all k_{ij} are equal, say k, the expression becomes: $[n - v - k\,(r - 1)]\sigma_1^2 + (n - v - r + 1)\sigma_2^2$. If, moreover, every class of C occurs once in every class of A, so that $n = vr$, then the expression will be: $(v - k)\,(r - 1)\sigma_1^2 + (v - 1)\,(r - 1)\sigma_2^2$.

Now we consider the case of which the series of experiments was an example. The estimation of σ_2^2 will take place by means of the perpendicular on $\mathrm{A}+\mathrm{B}$ again. This implies the application of the iterative method of 3.3 for every year (class of D) separately. The separate squares of perpendiculars will be added to find the required square.

In order to estimate σ_1^2 we lift out from y one or more parts, such that in every part at least two classes of D occur, and that in every class of A (every trial) the same classes of C (varieties) are represented. We will use the square of the projection on the corresponding space $(\mathrm{C}\times\mathrm{D})^*$. If there are more parts, then these squares and the corresponding expectations will be added in order to estimate σ_1^2. In the following calculation we call such a part y again.

The supposition about the coordinates of y is: $y_{ijk} = \gamma_i + \alpha_{jk} + \beta_{ij} + \varepsilon_{ijk}$ ($i = 1, \ldots, r$ and $j = 1, \ldots, m$), where γ_i is the constant associated with class i of C, α_{jk} the constant belonging to the k-th trial in class j of D (the

j-th year), β_{ij} the random variable associated with the combination of class i of C and class j of D (with variance σ_1^2) and ε_{ijk} the random variable for the coordinate of class i of C in class jk of A (with variance σ_2^2). Let the number of trials in year j be k_j, so that $r\Sigma_j k_j = n$. Because A* is taken orthogonal to C: $\Sigma_{jk}\alpha_{jk} = 0$.

$$\begin{aligned}
\text{Now } E\mathrm{y}_{\mathrm{N}}^2 &= (1/n)\, E\,(\Sigma_{ijk}\gamma_i + \Sigma_{ijk}\alpha_{jk} + \Sigma_{ijk}\beta_{ij} + \Sigma_{ijk}\varepsilon_{ijk})^2 = \\
&= (1/n)\,[(\Sigma_j k_j)^2\,(\Sigma_i\gamma_i^2) + r\Sigma_j k_j^2\sigma_1{}^2 + n\sigma_2{}^2].
\end{aligned}$$

$$\begin{aligned}
E\mathrm{y}_{\mathrm{C}}^2 &= (1/\Sigma_j k_j)\,[\Sigma_i E\,(\Sigma_{jk}\gamma_i + \Sigma_{jk}\alpha_{jk} + \Sigma_{jk}\beta_{ij} + \Sigma_{jk}\varepsilon_{ijk})^2] = \\
&= (1/\Sigma_j k_j)\,\Sigma_i\,[(\Sigma_j k_j)^2\gamma_i^2 + \Sigma_j k_j^2\sigma_1{}^2 + \Sigma_j k_j\sigma_2{}^2] = \\
&= (1/\Sigma_j k_j)\,[(\Sigma_j k_j)^2\,\Sigma_i\gamma_i^2 + r\Sigma_j k_j^2\sigma_1{}^2 + n\sigma_2{}^2].
\end{aligned}$$

$$\begin{aligned}
E\mathrm{y}_{\mathrm{D}}^2 &= \Sigma_j\,[(1/rk_j)\,E\,(\Sigma_{ik}\gamma_i + \Sigma_{ik}\alpha_{jk} + \Sigma_{ik}\beta_{ij} + \Sigma_{ik}\varepsilon_{ijk})^2] = \\
&= \Sigma_j(1/rk_j)[k_j^2\Sigma_i\gamma_i^2 + 2rk_j(\Sigma_i\gamma_i)(\Sigma_k\alpha_{jk}) + r^2\Sigma_k\alpha_{jk}{}^2 + rk_j^2\sigma_1{}^2 + rk_j\sigma_2{}^2] = \\
&= (1/r)\,(\Sigma_j k_j)\,(\Sigma_i\gamma_i^2) + r\Sigma_{jk}\,(\alpha_{jk}{}^2/k_j) + \Sigma_j k_j\sigma_1{}^2 + m\sigma_2{}^2.
\end{aligned}$$

$$\begin{aligned}
E\mathrm{y}_{\mathrm{C}\times\mathrm{D}}^2 &= \Sigma_{ij}\,[(1/k_j)\,E\,(\Sigma_k\gamma_i + \Sigma_k\alpha_{jk} + \Sigma_k\beta_{ij} + \Sigma_k\varepsilon_{ijk})^2] = \\
&= \Sigma_{ij}\,(1/k_j)\,[k_j^2\gamma_i^2 + 2k_j(\Sigma_i\gamma_i)\,(\Sigma_k\alpha_{jk}) + \Sigma_k\alpha_{jk}{}^2 + k_j^2\sigma_1{}^2 + k_j\sigma_2{}^2] = \\
&= (\Sigma_j k_j)\,(\Sigma_i\gamma_i^2) + r\Sigma_{jk}\,(\alpha_{jk}{}^2/k_j) + n\sigma_1{}^2 + rm\sigma_2{}^2.
\end{aligned}$$

Hence

$$\begin{aligned}
E\mathrm{y}_{(\mathrm{C}\times\mathrm{D})^*}^2 &= [n + (r\Sigma_j k_j{}^2/n) - (r\Sigma_j k_j^2/\Sigma_j k_j) - \Sigma_j k_j]\sigma_1{}^2 + (rm + 1 - m - r)\sigma_2{}^2 = \\
&= (r-1)\,[\Sigma_j k_j - (\Sigma_j k_j^2/\Sigma_j k_j]\sigma_1{}^2 + (r-1)\,(m-1)\sigma_2{}^2.
\end{aligned}$$

By means of this relation and the unbiased estimate of $\sigma_2{}^2$ we may find an unbiased estimate of $\sigma_1{}^2$. In case all k_j are equal, say k, then $E\mathrm{y}_{(\mathrm{C}\times\mathrm{D})^*}^2 = k\,(r-1)\,(m-1)\sigma_1{}^2 + (r-1)\,(m-1)\sigma_2{}^2$. Or $E\mathrm{y}_{(\mathrm{C}\times\mathrm{D})^*}^2/\dim\,(\mathrm{C}\times\mathrm{D})^* = k\sigma_1{}^2 + \sigma_2{}^2$.

For split-plot designs, as in 4.3.4, the estimation will proceed analogously. For the estimation of $\sigma_2{}^2$ the perpendicular on $(\mathrm{A}\times\mathrm{D}) + \mathrm{C} + \mathrm{B} = (\mathrm{A}\times\mathrm{D}) + \mathrm{B}$ is necessary. This implies the application of the iterative method of 3.3 on every class of D separately with the subclassifications A and B. In order to estimate $\sigma_1{}^2$, a part as large as possible is lifted out from y, such that the classification $A\times D$ (the combinations of main and split-plot factors) is orthogonal to the classification C. Then the just derived expression for $E\mathrm{y}_{(\mathrm{C}\times\mathrm{D})^*}^2$ will be used, where k_j is now the number of coordinates (plots) in (the residual of) a block which are together in a class of the main factor D and in a class of C.

If, for instance, there is one missing value in an ordinary orthogonal split-plot design, then, for the estimation of $\sigma_1{}^2$, the plots with the corresponding combination of main and split-plot factors remain out of consideration; in the square of the orthogonal projection of the residual on $(\mathrm{C}\times\mathrm{D})^*$ one of the k_j differs by one from the other k_j.

Finally we consider the case in which A and C are subspaces of B. Compare 4.1.1 and 4.3.1. For this purpose we consider first the classification of coordinates according to $A\times B$ with one coordinate in every class of $A\times B$, and A and B orthogonal.

Let the coordinate y_{ij} in class i of A and class j of B be equal to $\alpha_i + \beta_j + \varepsilon_{ij}$, where α_i is a constant for class i ($i = 1, \ldots, m$) of A, β_j a constant for class j

($j = 1, \ldots, k$) of B, and ε_{ij} a random variable with expectation 0 and variance $\sigma_{ij}{}^2$. The random variables have covariance zero. Because of the orthogonality of A and B we assume $\Sigma_{j=1}^{k}\beta_j = 0$. In order to determine $E\mathrm{y}^2_{(\mathrm{A}\times\mathrm{B})^*}$ we compute:

$$E\mathrm{y}^2 = E\,\Sigma_{ij}(\alpha_i+\beta_j+\varepsilon_{ij})^2 = k\Sigma_i\alpha_i{}^2+m\Sigma_j\beta_j^2+\Sigma_{ij}\sigma_{ij}{}^2;$$

$$E\mathrm{y}^2_{\mathrm{A}} = (1/k)\,\Sigma_i E[\Sigma_j(\alpha_i+\beta_j+\varepsilon_{ij})]^2 = (1/k)\,\Sigma_i E(k\alpha_i+\Sigma_j\varepsilon_{ij})^2 =$$
$$= k\Sigma_i\alpha_i^2+(1/k)\,\Sigma_{ij}\sigma_{ij}{}^2;$$

$$E\mathrm{y}^2_{\mathrm{B}} = (1/m)\,\Sigma_j E[\Sigma_i(\alpha_i+\beta_j+\varepsilon_{ij})]^2 = (1/m)\,\Sigma_j E(\Sigma_i\alpha_i+m\beta_j+\Sigma_i\varepsilon_{ij})^2 =$$
$$= (k/m)\,(\Sigma_i\alpha_i)^2 + m\Sigma_j\beta_j^2+(1/m)\Sigma_{ij}\sigma_{ij}{}^2;$$

$$E\mathrm{y}^2_{\mathrm{N}} = (1/km)\,E[\Sigma_{ij}(\alpha_i+\beta_j+\varepsilon_{ij})]^2 = (1/km)\,E(k\Sigma_i\alpha_i+\Sigma_{ij}\varepsilon_{ij})^2 =$$
$$= (k/m)\,(\Sigma_i\alpha_i)^2 + (1/km)\,\Sigma_{ij}\sigma_{ij}{}^2.$$

We find:

$$E\mathrm{y}^2_{(\mathrm{A}\times\mathrm{B})^*} = [1-(1/k)-(1/m)+(1/km)]\,\Sigma_{ij}\sigma_{ij}{}^2 = (k-1)\,(m-1)\,(\Sigma_{ij}\sigma_{ij}{}^2/km).$$

The formal estimation of σ^2 in case all $\sigma_{ij}{}^2$ were equal has an expectation value equal to the average of all $\sigma_{ij}{}^2$.

Returning to our problem, we estimate σ_2^2 with the help of the perpendicular on the space corresponding to the classification in units of higher order. Further, we consider the averages of these classes, which are supposed to satisfy a relation which agrees to that for the above-mentioned coordinates y_{ij}. For, let the number of coordinates in a unit of higher order be k_i, then the average has variance $\sigma_1^2 + \sigma_2^2/k_i$, while the expectation is the sum of the effects of the two classifications of the units. If we take from the set of averages a part, such that the two classifications become orthogonal, the way to estimate $\sigma_1{}^2$ is open now.

4.5. TESTS

4.5.1. *A constant main effect and a random effect*

First we consider the testing of the effect of A in the general regression problem discussed in 4.2.1 (incomplete blocks). Under the null hypothesis and under the condition of a fixed effect of B, the test statistic F, calculated according to chapter 3 with the help of the vectors $\mathrm{y}-\mathrm{y}_{\mathrm{A+B}}$ and $\mathrm{y}_{\mathrm{A+B}}-\mathrm{y}_{\mathrm{B}}$ both orthogonal to B, has a F-distribution. Because this distribution is independent of the condition, the same holds unconditionally. It follows that the test on A will be performed as in chapter 3.

Now we consider the design of split-plot type discussed in 4.2.3. Under the condition of fixed block effects, we perform first the test on the interaction $(\mathrm{A}\times\mathrm{C})^*$. This requires the perpendicular on $\mathrm{B}+(\mathrm{A}\times\mathrm{C})$, i.e. the iterative method of 3.3 for every separate class of C, according to the subclassifications A and B on the one hand, and the perpendicular on $\mathrm{A}+\mathrm{B}$ according to 3.3 on the other hand. Because both perpendiculars have expectations independent of the effect of B, the conditional test will be unconditional.

The best estimate of $E\mathrm{y}$ under the alternative hypothesis has been discussed in 4.2.3, while that under the null hypothesis ($E\mathrm{y}$ in $\mathrm{A}+\mathrm{C}$) will be found according to the method of 4.3.1.

The conditional test on effect of A requires the perpendiculars on $\mathrm{A}+\mathrm{B}$ and on B (both according to chapter 3) and appears to be unconditional again.

The best estimate of $E\mathrm{y}$ under the null hypothesis ($E\mathrm{y}$ in C) will be obtained by the method for the determination of $\mathrm{y}_{s\mathrm{N}}$ (end of 4.2.2) for every class of C separately.

For testing the main effect of C, with the null hypothesis that $E\mathrm{y}$ is in A, against the alternative that $E\mathrm{y}$ is in A + C, we cannot use a conditional test, as C is a subspace of B. Therefore we need (with the appropriate metric) the orthogonal projections of y on both subspaces. The projection (according to the new metric) on A takes place according to the method of 4.2.1, and that on A + C according to the method of 4.3.1. The square (with respect to the new metric) of the first projection $\mathrm{y}_{s\mathrm{A}}$ is equal to (compare 1.6.3)

$$\Sigma_i w_i \{(\mathrm{y}_{s\mathrm{A}})_{\mathrm{B}_i}\}^2 + w\{\mathrm{y}_{s\mathrm{A}} - (\mathrm{y}_{s\mathrm{A}})_\mathrm{B}\}^2 = \Sigma_i w_i \{(\mathrm{y}_{s\mathrm{A}})_{\mathrm{B}_i}\}^2 + w\mathrm{y}^2_{s\mathrm{A}} - w\{(\mathrm{y}_{s\mathrm{A}})_\mathrm{B}\}^2.$$

In case all k_i are equal and hence all w_i equal, say w', this is

$$(w' - w)\{(\mathrm{y}_{s\mathrm{A}})_\mathrm{B}\}^2 - w\mathrm{y}^2_{s\mathrm{A}}$$

Similarly the square (with respect to the new metric) of the second projection $\mathrm{y}_{s\mathrm{A}} + \mathrm{y}_{s\mathrm{C}}$ is: $\Sigma_i w_i \{(\mathrm{y}_{s\mathrm{A}} + \mathrm{y}_{s\mathrm{C}})_{\mathrm{B}_i}\}^2 + w\{\mathrm{y}_{s\mathrm{A}} + \mathrm{y}_{s\mathrm{C}} - (\mathrm{y}_{s\mathrm{A}} - \mathrm{y}_{s\mathrm{C}})_\mathrm{B}\}^2 =$

$$= \Sigma_i w_i \{(\mathrm{y}_{s\mathrm{A}})_{\mathrm{B}_i} + \mathrm{y}_{s\mathrm{C}}\}^2 + w[\mathrm{y}^2_{s\mathrm{A}} - \{(\mathrm{y}_{s\mathrm{A}})_\mathrm{B}\}^2].$$

The difference of these squares will occur in the numerator of F. In the denominator we need the square (with the appropriate metric) of the perpendicular on A × C, which can be found in a similar way.

For a special case, which we will study now, there exists a known method. We want to compare our technique with the known technique.

In the special case any of the k classes of A is represented once in every block. Let the number of blocks be b, so that $n = bk$, and let the number of classes of C be c. Then the estimate of $\sigma_2{}^2$, obtained from the perpendicular (with metric I) on $\mathrm{B} + (\mathrm{A} \times \mathrm{C})$ is $(\mathrm{y}^2 - \mathrm{y}^2_\mathrm{B} - \mathrm{y}^2_{\mathrm{A}\times\mathrm{C}} + \mathrm{y}^2_\mathrm{C})/\{bk - (b - c) - ck\}$ or $(\mathrm{y}^2 - \mathrm{y}^2_\mathrm{B} - \mathrm{y}^2_{\mathrm{A}\times\mathrm{C}} + \mathrm{y}^2_\mathrm{C})/(k-1)(b-c)$. In order to estimate $\sigma_1{}^2$, we need the perpendicular on $\mathrm{A} \times \mathrm{C}$: $\mathrm{y}^2 - \mathrm{y}^2_{\mathrm{A}\times\mathrm{C}}$ with expectation $k(b-c)(\sigma_1{}^2 + \sigma_2{}^2)$. The estimate of $\sigma_1{}^2$ will be

$$(\mathrm{y}^2 - \mathrm{y}^2_{\mathrm{A}\times\mathrm{C}})/\{k(b-c)\} - (\mathrm{y}^2 - \mathrm{y}^2_\mathrm{B} - \mathrm{y}^2_{\mathrm{A}\times\mathrm{C}} + \mathrm{y}^2_\mathrm{C})/\{k-1)(b-c)\}.$$

And the estimate of $k\sigma_1{}^2 + \sigma_2{}^2$ will be then: $(\mathrm{y}^2_\mathrm{B} - \mathrm{y}^2_\mathrm{C})/(b-c)$.

The orthogonal projection of y on A × C (with the appropriate metric) will according to 4.2.3 be $\mathrm{y}_{s\mathrm{A}\times\mathrm{C}} = \mathrm{y}_{\mathrm{A}\times\mathrm{C}}$ and the square (with that metric) of the corresponding perpendicular:

$$\begin{aligned} &w'\mathrm{y}^2_\mathrm{B} + w(\mathrm{y}^2 - \mathrm{y}^2_\mathrm{B}) - w'\{(\mathrm{y}_{\mathrm{A}\times\mathrm{C}})_\mathrm{B}\}^2 - w\{\mathrm{y}_{\mathrm{A}\times\mathrm{C}} - (\mathrm{y}_{\mathrm{A}\times\mathrm{C}})_\mathrm{B}\}^2 = \\ &= w'\mathrm{y}^2_\mathrm{B} + w\mathrm{y}^2 - w\mathrm{y}^2_\mathrm{B} - w'\mathrm{y}^2_\mathrm{C} - w\mathrm{y}^2_{\mathrm{A}\times\mathrm{C}} + w\mathrm{y}^2_\mathrm{C} = \\ &= w(\mathrm{y}^2 - \mathrm{y}^2_{\mathrm{A}\times\mathrm{C}} - \mathrm{y}^2_\mathrm{B} + \mathrm{y}^2_\mathrm{C}) + w'(\mathrm{y}^2_\mathrm{B} - \mathrm{y}^2_\mathrm{C}). \end{aligned}$$

This perpendicular is in a space with dimension $k(b-c)$.

The orthogonal (with the appropriate metric) projection on A + C will be $\mathrm{y}_\mathrm{C} + (\mathrm{y}_\mathrm{A} - \mathrm{y}_\mathrm{N})$, as follows from 4.3.3. Further the projection on A is y_A. The square (with this metric) of the difference $\mathrm{y}_\mathrm{C} - \mathrm{y}_\mathrm{N}$ will occur in the numerator of F. Hence the F-statistic will be

$$\frac{w'(\mathrm{y}^2_\mathrm{C} - \mathrm{y}^2_\mathrm{N})/(c-1)}{\{w(\mathrm{y}^2 - \mathrm{y}^2_{\mathrm{A}\times\mathrm{C}} - \mathrm{y}^2_\mathrm{B} + \mathrm{y}^2_\mathrm{C}) + w'(\mathrm{y}^2_\mathrm{B} - \mathrm{y}^2_\mathrm{C})\}/k(b-c)} =$$

$$= \frac{w'(\mathrm{y}_{\mathrm{C}}^2 - \mathrm{y}_{\mathrm{N}}^2)/(c-1)}{\{(k-1)(b-c)+(b-c)\}/k(b-c)} = \frac{(\mathrm{y}_{\mathrm{C}}^2 - \mathrm{y}_{\mathrm{N}}^2)/(c-1)}{(\mathrm{y}_{\mathrm{B}}^2 - \mathrm{y}_{\mathrm{C}}^2)/(b-c)}.$$

It is remarkable that we obtain the same value as would be found if, which is usual, the classification *A* would be left out of consideration, and if only the block sums would be considered as random variables. In our treatment, however, a *F*-distribution with dimensions $c-1$ and $k(b-c)$ must be used, while in the usual treatment the dimensions $c-1$ and $b-c$ would be used. This apparent greater power of our test, however, is merely the consequence of the fact that we have considered the estimates of σ_1^2 and σ_2^2 as constants, while they are random variables. It follows again that the effect of this wrong assumption will decrease, in case $b-c$ will be large. In order to escape this effect completely in a not special case, one should perform the usual test on a part of y, which part is of the kind just considered.

4.5.2. *Two constant main effects and a random effect*

The test on effect of *A* in the general case of 4.3.1 will be performed by a conditional test, under the condition of fixed block effects. This test appears to be unconditional again. It requires a projection on A + B and on B according to chapter 3; the classification *C* plays no rôle in this test.

The test on effect of *C* will be analogous to that on effect of *C* in designs of split-plot type considered in the last section. In this case, however, the denominator of *F* will be obtained from the square of the perpendicular on A + C (with respect to the new metric again). This square is equal to the difference of the square (new metric) of y, $\Sigma_i w_i \mathrm{y}_{\mathrm{B}_i}^2 + w\mathrm{y}^2 - w\mathrm{y}_{\mathrm{B}}^2$ and the square (new metric) of $(\mathrm{y}_{s\mathrm{A}} + \mathrm{y}_{s\mathrm{C}})$.

In case both A and C (compare 4.1.1 and 4.3.1) are subspaces of B, the square of the projection (new metric) of y on A + C is $\Sigma_i w_i \{(\mathrm{y}_{s\mathrm{A}} + \mathrm{y}_{s\mathrm{C}})_{\mathrm{B}_i}\}^2$. This is equal to the square of the projection of the vector of class averages of *B* with weights $(\sigma_1^2 + \sigma_2^2/k_i)^{-1}$ on the corresponding A + C. Similarly the square (with appropriate metric) of e.g. $\mathrm{y}_{s\mathrm{A}}$, the projection corresponding to the null hypothesis that *C* has no effect will be found. In this case the square (with appropriate metric) of y is equal to the sum of the (weighted) square of the vector of class averages and of $w(\mathrm{y}^2 - \mathrm{y}_{\mathrm{B}}^2)$.

If in the example of a series of experiments (4.3.5) $\sigma_1 \neq 0$, then there is effect of *C* (varieties) which is different in different years; testing the effect of varieties has thus no sense. In this example only the best estimation is of interest.

The last case to be discussed is the split-plot design (4.3.4), with a classification *C* of the blocks and a corresponding main effect, and with the interaction $A \times D$, where *D* is the main factor. The test on interaction of *A* and *D*, and that on main effect of *A*, are completely analogous to those in the split-plot design without the classification *C*. Conditional tests appear to be unconditional again. Projection on B + (A × D) requires the iterative method of 3.3 in every separate class of *D* according to the subclassifications *A* and *B*; projection on A + B (corresponding to the null hypothesis of no interaction) requires this iterative method with the classifications *A* and *B*; projection on B (corresponding to the null hypothesis: *E*y is in B + C) is quite simple.

The best estimates of $E\mathbf{y}$ under the different hypotheses are obtained as follows: Under the hypothesis that $E\mathbf{y}$ is in $\mathrm{C} + (\mathrm{A} \times \mathrm{D})$ we use the method of 4.3.4. Under the hypothesis that $E\mathbf{y}$ is in $\mathrm{C}+\mathrm{A}$ we use the method of 4.3.1. The estimate of $E\mathbf{y}$, under the hypothesis that $E\mathbf{y}$ is in C, will be obtained by the method for $\mathbf{y}_{s\mathrm{N}}$ (end of 4.2.2) for every class of *C* separately. For one case, namely that $E\mathbf{y}$ is in the space $\mathrm{A}+\mathrm{C}+\mathrm{D}$, we have no general solution, similar to the other solutions, available.

This absence of a general solution also complicates the test on the effect of *D* i.e. the test of the null hypothesis that $E\mathbf{y}$ is in $\mathrm{A}+\mathrm{C}$, against the alternative that $E\mathbf{y}$ is in $\mathrm{A}+\mathrm{C}+\mathrm{D}$.

In the particular case that the classes of *A* are represented proportionally, say once, in every class of *B* (which classes thus all have the same number of coordinates, say *k*) the estimate is a vector $\mathbf{y}_{s\mathrm{A}^*}+\mathbf{y}_{s\mathrm{C}}+\mathbf{y}_{s\mathrm{D}^*}$, such that
$w'(\mathbf{y}-\mathbf{y}_{s\mathrm{A}^*}-\mathbf{y}_{s\mathrm{C}}-\mathbf{y}_{s\mathrm{D}^*})_{\mathrm{B}} + w\{\mathbf{y}-\mathbf{y}_{s\mathrm{A}^*}-\mathbf{y}_{s\mathrm{C}}-\mathbf{y}_{s\mathrm{D}^*}-(\mathbf{y}-\mathbf{y}_{s\mathrm{A}^*}-\mathbf{y}_{s\mathrm{C}}-\mathbf{y}_{s\mathrm{D}^*})_{\mathrm{B}}\}$
is orthogonal to A*, C and D*. Because A* is orthogonal to B, and thus to C and D, we obtain the equations:

$$\begin{aligned}
w(\mathbf{y}-\mathbf{y}_{s\mathrm{A}^*}-\mathbf{y}_{s\mathrm{C}}-\mathbf{y}_{s\mathrm{D}^*})_{\mathrm{A}} &= 0 \quad \text{or} \quad (\mathbf{y}-\mathbf{y}_{s\mathrm{A}^*})_{\mathrm{A}^*} = 0\\
w'(\mathbf{y}-\mathbf{y}_{s\mathrm{A}^*}-\mathbf{y}_{s\mathrm{C}}-\mathbf{y}_{s\mathrm{D}^*})_{\mathrm{BC}} &= 0 \quad \text{or} \quad (\mathbf{y}-\mathbf{y}_{s\mathrm{C}}-\mathbf{y}_{s\mathrm{D}^*})_{\mathrm{C}} = 0\\
w'(\mathbf{y}-\mathbf{y}_{s\mathrm{A}^*}-\mathbf{y}_{s\mathrm{C}}-\mathbf{y}_{s\mathrm{D}^*})_{\mathrm{BD}} &= 0 \quad \text{or} \quad (\mathbf{y}-\mathbf{y}_{s\mathrm{C}}-\mathbf{y}_{s\mathrm{D}^*})_{\mathrm{D}} = 0.
\end{aligned}$$

It follows that $\mathbf{y}_{s\mathrm{A}^*} = \mathbf{y}_{\mathrm{A}}-\mathbf{y}_{\mathrm{N}}$ and that $\mathbf{y}_{s\mathrm{C}}$ and $\mathbf{y}_{s\mathrm{D}^*}$ are obtained by application of the iterative method of 3.3 with the classifications *C* and *D*. This implies that the classification *A* has no influence on this test (the orthogonal projection on $\mathrm{A}+\mathrm{C}$ in this special case has already been considered in the last section).

In the usual treatment of this problem for ordinary split-plot designs one leaves the classification *A* out of consideration and uses only the block sums. In the construction of the test statistic we will observe a similar difference in the dimensions of *F* between the usual and our technique, as in the last section.

If, however, *A* is not orthogonal to *B*, and the effect of *A* is significant, it is wrong to neglect the classification *A*. One way out of this difficulty is to perform the test (and the estimate) with the help of a part of $\mathbf{y}$ such that *A* is orthogonal to *B*. Another possibility for the test (not for the estimation), especially in case this part of $\mathbf{y}$ becomes too small, may be the following. We test (with the complete $\mathbf{y}$) the null hypothesis that $E\mathbf{y}$ is in $\mathrm{A}+\mathrm{C}$ against the alternative that $E\mathbf{y}$ is in $\mathrm{A}+(\mathrm{C}\times\mathrm{D})$. In this way, the main effect of *D* is investigated together with its interaction with *C*, which is supposed to be absent in fact. This procedure has the drawback that it leads to a decrease of the power of the test on effect of *D* only.

The projection (with appropriate metric) on $\mathrm{A}+\mathrm{C}$ requires the method of 4.3.1, while that on $\mathrm{A}+(\mathrm{C}\times\mathrm{D})$ will take place in a similar way, because $\mathrm{C}\times\mathrm{D}$ is in B. Mostly, the classification $C \times D$ will be the same as the classification *B* (e.g. in the case of missing plots in an ordinary split-plot design). But then the transformation *Q* is simply $P_{\mathrm{A}}P_{\mathrm{B}}$, so that the orthogonal projection will be obtained according to the method of 3.3 with the classifications *A* and *B*. The square (with appropriate metric) of the corresponding perpendicular is thus $w(\mathbf{y}-\mathbf{y}_{s\mathrm{A}}-\mathbf{y}_{s\mathrm{B}})^2$. The square (with appropriate metric) of the perpendicular on $\mathrm{A}+\mathrm{C}$ is (compare the foregoing section):

$$\Sigma_i w_i(\mathbf{y}_{\mathrm{B}_i})^2 + w\mathbf{y}^2 - w\mathbf{y}_{\mathrm{B}}^2 - \Sigma_i w_i\,\{(\mathbf{y}_{s\mathrm{A}})_{\mathrm{B}_i} + \mathbf{y}_{s\mathrm{C}}\}^2 - w\mathbf{y}_{s\mathrm{A}}^2 + w\{(\mathbf{y}_{s\mathrm{A}})_{\mathrm{B}}\}^2.$$

In the denominator of *F* we use $w\{\mathbf{y}-\mathbf{y}_{\mathrm{B}+(\mathrm{A}\times\mathrm{D})}\}^2$.

SUMMARY

In the introduction some objections against the usual definitions and notation, derivations and application of regression theory (to be understood in a broad sense) are enumerated. Therefore a sound and comprehensive foundation of this theory by linear algebra, as devised by KUIPER, is welcome. This study is an elaboration of that method, and the presentation of new results and insights obtained.

In chapter 1 linear algebra, as far as needed for the applications, is explained. In particular, more than usual attention is paid to the evaluation of projections. A new iterative method to obtain orthogonal projections is given.

In chapter 2 the general problems in regression theory are considered in terms of vectors. The most interesting conclusion is that unbiased estimates are obtained by projection, and that unbiased, most efficient estimates are obtained by orthogonal projection after choosing the appropriate metric.

In chapter 3 several regression problems for uncorrelated observations are considered in a general way. Beside some attention to the linear regression problem in a narrower sense (with e.g. orthogonal polynomials), the main interest is in regression problems, connected with classifications of the observations. Levels, main effects, interactions and components of effects are defined in terms of subspaces. A general definition of orthogonality of classifications, and the formation of components of interaction in orthogonal classifications by means of tensor products are given.

An iterative procedure, very useful for theoretical considerations as well as in practice, in order to solve the general regression problems with two and more classifications, is discussed. This procedure, together with the ideas developed in the previous chapters, leads to a surprising insight into the balanced incomplete block designs, the group divisible partially balanced incomplete block designs, the two-dimensional lattices, and some designs of Latin square type of PEARCE.

Another interesting case is that of one classification orthogonal to the interaction of the other two.

A general treatment of problems related to analysis of covariance is given. In particular, the case with two non-orthogonal classifications and a "concomitant variable" seems to be new. Further a simple iterative method is given in order to estimate and to test the effect of treatments in a trial field, where the fertility of the plots is supposed to be a polynomial in the coordinates of the centres of the plots.

The chapter ends with a general exposition of a missing plot technique.

In chapter 4 we consider regression problems, mainly estimation problems, in case of classifications of the observations, but with a random effect corresponding to one of the classifications. The so-called recovery of inter-block information in incomplete blocks forms an example.

A new iterative method for the best estimation in a very general situation of incomplete blocks is given. A design of split-plot type, also of a very general character, can be ranged under the same heading.

Similarly, a new iterative method for the best estimation in case of a random

(block) effect and two constant effects from two characteristics, of which at least one determines a classification in classes consisting of a number of whole blocks, is derived. A trivial particular case is that of replicates in an incomplete block design. But also the problem of estimation of a general effect of treatments from a series of trials, in which the interaction between years and treatments is considered as random, is solved for very general situations, under certain presuppositions. Further the customary split-plot design, however without the requirement of orthogonality, belongs to this type of problems.

For all the cases the estimation of the two variances, on which the derived methods are based, is presented. Finally some remarks are made about the tests of null hypotheses in all the considered cases.

SAMENVATTING

„Vectoren, een werktuig in de statistische regressietheorie" beoogt meer inzicht en, daardoor, vereenvoudiging te brengen in het terrein van de regressierekening. Tot dit terrein behoren vele onderwerpen, die met verschillende namen worden aangeduid, zoals variantie-analyse, covariantie-analyse, proefschema's, lineaire vereffening, vruchtbaarheidscorrecties, samenvatten van proeven enz. Door volgens de denkbeelden van Kuiper deze theorie te funderen op de lineaire algebra, waarin men verzamelingen van z.g. vectoren beschouwt, werd niet alleen genoemd doel bereikt, maar konden ook nieuwe resultaten en inzichten worden verworven.

In de inleiding wordt gewezen op enige ernstige bezwaren tegen de gebruikelijke fundering, die in vele opzichten vaag is en aanleiding geeft tot een staalkaart van onoverzichtelijke technieken en formules voor een toch nog vrij beperkt geheel.

Het niet-statistische hoofdstuk 1 is een beknopte, maar volledige uiteenzetting van die delen van de lineaire algebra, die voor de gewenste toepassing noodzakelijk zijn. Allereerst worden begrippen zoals vectoren, vectorruimten, basis, dimensie en deelruimten ingevoerd. Vervolgens behandelen wij lineaire transformaties in vectorruimten, en bijbehorende matrices en eigenwaarden. Met behulp van het inwendige product worden metrische eigenschappen, zoals lengten, afstanden, hoeken en loodrechtheid, ingevoerd, gevolgd door een bijzondere lineaire transformatie, de loodrechte projectie.

Na invoering van de begrippen convergentie van vectoren en van transformaties bespreken wij een belangrijke machtreeks van transformaties, analoog aan de gewone convergente meetkundige reeks.

In het gedeelte, waarin de uitvoering van projecties, o.a. met behulp van normaalvergelijkingen, ter sprake komt, wordt een nieuwe algemene iteratieve methode gegeven voor de bepaling van loodrechte projecties. Opmerkingen over symmetrische transformaties en matrices, dit in verband met de begrippen metriek en inwendig product, besluiten het hoofdstuk.

Vanaf hoofdstuk 2 zijn de te gebruiken vectoren rijtjes van n getallen, waarvoor bepaalde rekenregels zijn afgesproken. Na enige opmerkingen over de covariantie-matrix en de normale verdeling van een stochastische vector, en over de geschikte keuze van een metriek, volgen op eenvoudige wijze enige eigenschappen van loodrechte projecties van dergelijke vectoren, o.a. over de verwachting van het kwadraat van zo een projectie en de bijbehorende χ^2- en F-verdelingen. Vervolgens wordt de beste schatting in het algemene lineaire regressieprobleem afgeleid: zuivere schattingen ontstaan door projectie; zuivere, meest doeltreffende schattingen ontstaan door loodrechte projectie, na keuze van de passende metriek. Dan volgt een zeer algemene behandeling van de toetsing bij lineaire regressie. Tot slot geven wij de oplossing van het probleem der voorwaardelijke waarnemingen, dat in wezen hetzelfde blijkt te zijn als het lineaire regressieprobleem.

In hoofdstuk 3 worden de ontwikkelde begrippen toegepast in verscheidene bijzondere gevallen, waarbij de waarnemingsuitkomsten ongecorreleerd zijn. Wij onderscheiden „verklarende variabelen" die expliciet gegeven zijn, van op klasse-indelingen naar aanleiding van een of ander kenmerk berustende variabelen.

Van de eerste groep worden de lineaire functies in het algemeen, en de orthogonale polynomen in het bijzonder besproken.

Bij de tweede groep worden niveau's, hoofdeffecten en interacties gedefinieerd als vectoren in deelruimten. Voor het algemene geval van twee indelingen (twee hoofdeffectruimten) bespreken wij een uit de iteratieve methode van hoofdstuk 1 voortkomende methode van schatten, waarbij als vrijwel enige operatie het bepalen van gemiddelden optreedt. Als bijzondere gevallen verschijnen de orthogonale indelingen (met een zeer algemene definitie van orthogonaliteit), voorts de evenwichtige en sommige, gedeeltelijk evenwichtige, onverzadigde blokkenschema's en de twee-dimensionale roosterschema's. Bij de laatste drie groepen van schema's blijkt de iteratieve methode te leiden tot het gebruik van transformaties met slechts één of twee eigenwaarden met zeer bijzondere ruimten van eigenvectoren. Daaruit volgen merkwaardige eigenschappen van componenten van effecten, aan wier schatting en toetsing wij ook in het algemeen veel aandacht besteden.

Voor componenten van de interactie-ruimte bij een tweetal orthogonale indelingen blijkt het begrip tensor-product met vrucht te kunnen worden gebruikt.

Uitvoerig wordt ingegaan op schattingen en toetsingen in het algemene geval van drie indelingen. Ook hier speelt de iteratieve methode, afgeleid in hoofdstuk 1, een belangrijke rol. Als bijzonderheden treden op: volledige orthogonaliteit der drie indelingen, Latijnse vierkanten, schema's van PEARCE en het geval, dat één indeling orthogonaal is met de interactie van de andere twee. Bij de schema's van PEARCE zijn twee van de drie indelingen orthogonaal en blijkt de bijbehorende iteratieve methode eveneens te leiden tot een transformatie met twee eigenwaarden. Bij het laatst genoemde bijzondere geval treden merkwaardige eigenschappen en vereenvoudigingen op.

Tot de behandelde problemen, waarbij beide soorten van „verklarende variabelen" optreden, behoren die, welke ten dele met covariantie-analyse worden betiteld. Een nieuwe bijdrage daarin is de beschouwing van het algemene geval van twee indelingen en een „verklarende variabele". Voorts wordt een iteratieve methode gegeven, om in een proef in de vorm van een strook of van een rechthoekig rooster, onder de veronderstelling, dat de vruchtbaarheid een continue functie nl. een veelterm in de coördinaten der veldjesmiddelpunten is, schattingen en toetsingen met betrekking tot het te onderzoeken effect uit te voeren.

Het hoofdstuk eindigt met een algemene methode tot leemtevulling d.i. gebruik van een eenvoudig schema, als het gegeven schema enige gapingen vertoont ten opzichte van dat eenvoudiger schema.

In hoofdstuk 4 geven wij een algemene en nieuwe behandeling van regressieproblemen eveneens met klassenindelingen; sommige effecten worden echter zelf als stochastische variabelen opgevat. Na een inleiding omtrent de aard van deze problemen voeren wij op een, in hoofdstuk 1 voorbereide, wijze het passende inwendig product in.

De twee bij de veronderstellingen behorende varianties als gegeven beschouwend, leiden wij, met behulp van de machtreeks in hoofdstuk 1, een iteratieve methode af ter bepaling van de beste schatting voor het geval van onvolledige blokken met een stochastisch blokeffect (recovery of inter-block information). De reeds bekende oplossingen zijn bijzondere gevallen hiervan. Een zeer algemeen schema met het karakter van een split-plot schema blijkt tot hetzelfde onderwerp te behoren.

Eveneens geven wij een nieuwe iteratieve methode ter bepaling van de beste schatting voor het geval van onvolledige blokkenschema's, waarin de blokeffecten stochastisch zijn, maar waarin door een indeling van de blokken een constant hoofdeffect, naast dat voor de behandelingen, is toegevoegd. „Split-plot"-proeven in de gebruikelijke zin, maar zonder noodzakelijke orthogonaliteit der indelingen, zijn hiervan een bijzonder geval. Hetzelfde blijkt te gelden voor het probleem der schatting van gemiddeld te verwachten rasverschillen uit rassenproeven over verscheidene jaren, waarin de interactie tussen rassen en jaren als stochastisch wordt opgevat en waarin eveneens orthogonaliteit ontbreekt. Verder behoort hiertoe het geval van twee indelingen met meer dan een waarneming per combinatie van twee klassen en met een stochastische interactie.

Bij al de problemen van dit hoofdstuk wordt ingegaan op bijzondere situaties, zoals hetzelfde aantal veldjes in elk blok, orthogonaliteit van onderindelingen in de afzonderlijke klassen van een hoofdindeling, of een zeer kleine of een zeer grote verhouding van de genoemde twee varianties.

Hierna wordt de schatting van het tweetal benodigde varianties voor al de behandelde gevallen besproken en afgeleid. Enige opmerkingen over de uitvoering van toetsingen in al de beschouwde gevallen besluiten het hoofdstuk.

ACKNOWLEDGEMENT

The author is much obliged to the Institute for Research on Varieties of Field Crops for the benevolence which enabled him to complete this work as a thesis submitted to the Agricultural University of Wageningen in fulfilment of the requirements for the degree of Doctor of Agricultural science, and to insert results obtained in the course of the performance of his task. He wishes to express his thanks to Professor N. H. Kuiper for his encouragement, helpful suggestions and constructive criticism.

REFERENCES

1. Anderson, R. L. and E. E. Houseman. Tables of orthogonal polynomial values extended to $n = 104$. Research Bulletin no. 297, Iowa State College, Ames, Iowa, 1942.
2. Bose, R. C., W. H. Clatworthy and S. S. Shrikhande. Tables of partially balanced designs with two associate classes. North Carolina Agricultural Experiment Station, 1954.
3. Corsten, L. C. A. Een algemene methode tot leemtevulling. Statistica **16**, 141–151 (1957).
4. Corsten, L. C. A. Partition of experimental vectors connected with multinomial distributions. Biometrics **13**, (1957) (in press).
5. DeLury, D. B. Values and integrals of the orthogonal polynomials up to $n = 26$. University of Toronto Press, 1950.
6. Fisher, R. A. and F. Yates. Statistical Tables for Biological, Agricultural and Medical Research. Oliver and Boyd, London, 1953.
7. Halmos, P. R. Finite dimensional vector spaces. Princeton University Press, Princeton, 1948.
8. Hamming, G. Het samenvatten van rassenproeven en het toepassen van vruchtbaarheidscorrecties met niet-orthogonale methoden. Verslagen Landbouwkundige Onderzoekingen no. 54.19, 1949.
9. Kempthorne, O. The design and analysis of experiments. John Wiley and Sons, New York, 1952.
10. Kramer, C. Y. and R. A. Bradley. Examples of intra-block analysis for factorials in group divisible, partially balanced, incomplete block designs. Biometrics **13**, 197–224 (1957).
11. Kuiper, N. H. Variantie-analyse. Statistica **6**, 149–194 (1952).
12. Kuiper, N. H. en L. C. A. Corsten. Open plaatsen in variantieschema's. Mededelingen van de Landbouwhogeschool te Wageningen, **53**, 25–30 (1953).
13. Kuiper, N. H. Vectoren en variantie-analyse. Biometrisch Contact **5**, 22–35 (1953).
14. Kuiper, N. H. Een opmerking over het aanpassen van functies aan een groot aantal waarnemingsuitkomsten. Statistica **8**, 1–6 (1954).
15. Pearce, S. C. Some new designs of Latin square type. Journal Royal Statistical Society **14**, 101–106 (1952).
16. Pearce, S. C. Field experimentation with fruit-trees and other perennial plants. Commonwealth Agricultural Bureaux, 1953.
17. Pearson, E. S. and H. O. Hartley. Biometrika tables for statisticians. Volume I, Cambridge University Press, 1954.
18. Rao, C. R. General methods of analysis for incomplete block designs. Journal of the American Statistical Association **42**, 541–561 (1947).
19. Rao, C. R. Advanced statistical methods in biometric research. John Wiley and Sons, New York, 1952.
20. Stevens, W. L. Statistical analysis of a non-orthogonal tri-factorial experiment. Biometrika **35**, 346–367 (1948).
21. Tukey, J. W. One degree of freedom for non-additivity. Biometrics **5**, 232–242 (1949).
22. Yates, F. Sample methods for censuses and surveys. Charles Griffin and Co, London, 1949.