# STATISTICAL INFERENCE ON VARIANCE COMPONENTS

Promotor: dr. P. van der Laan, hoogleraar in de Wiskundige Statistiek.

L. R. Verdooren

# STATISTICAL INFERENCE ON VARIANCE COMPONENTS

Proefschrift
ter verkrijging van de graad van
doctor in de landbouwwetenschappen,
op gezag van de rector magnificus,
dr. C.C. Oosterlee,
in het openbaar te verdedigen op
woensdag 3 februari 1988,
des namiddags te vier uur in de aula
van de Landbouwuniversiteit te Wageningen.

ISN 266302.

Aan Pieneke

Voor Moes

NN08201, 1196

STELLINGEN

1. In de schattingstheorie moet van een schatter geëist worden dat deze behoort tot de klasse van toegelaten (Eng.: "permissible") schatters.

   (Dit proefschrift.)

2. Bij het rassenonderzoek van cultuurgewassen zijn voorspellingen van rascontrasten van meer nut dan de schattingen ervan.

   (Dit proefschrift.)

3. Het gebruik van Satterthwaite's benadering bij het bepalen van de kans op negatieve schattingen van genetische varianties, zoals dat gebeurt in Bridges en Knapp (1987), is af te raden.

   (Bridges Jr., W.C. and Knapp, S.J. (1987). Probabilities of negative estimates of genetic variances. Theor. Appl. Genet. 74, 269-274.)

4. Bij de benadering van de variantie van de schatters voor de erfelijkheidsgraad $h^2$, wordt in vele handboeken, o.a. Becker (1975), de covariantie van gemiddelde kwadraatsommen (Eng."Mean Squares") gelijk aan nul gesteld. Voor ongebalanceerde schema's is dit niet juist.

   (Becker, W.A. (1975). Manual of quantitative genetics, 3rd ed. Washington State University, Washington.)

5. Het klakkeloos gebruiken van een schatter voor de genetische correlatie tussen identieke kenmerken, die gemeten zijn in verschillende milieus binnen eenzelfde niveau van een fokprogramma, kan leiden tot uitkomsten die groter zijn dan 1.

   (Merks, J.W.M. (1986). Genotype x Environment interactions in pig breeding programmes. I. Central test. Livestock Production Science 14, 365-381.)

6. Bij het streven naar hogere gewasopbrengsten in ontwikkelingslanden liggen de beperkingen heden ten dage niet zozeer bij de mogelijkheden tot toelevering van verbeterde cultivars, doch vooral in de noodzaak om bij grote

aantallen kleine producenten verbeteringen van teelttechnieken en van bodemvruchtbaarheid te realiseren.

7. In de plantenveredeling kan men bij het selecteren uit een aantal lijnen in het algemeen beter de procedure van Gupta gebruiken dan de tot nu toe gebruikelijke "multiple comparisons" procedures.

   (Gupta, S.S.(1965). On some multiple decision (selection and ranking) rules. Technometrics 7, 225-245.)

8. In het buitenland, en vooral in kustvlaktegebieden, neemt als gevolg van intensieve irrigatie de noodzaak tot drainage toe. Nu in Nederland het drainage uitvoeringsprogramma nagenoeg voltooid is, kunnen Nederlandse experts met hun uitrusting in dergelijke gebieden worden ingezet, dit tot voordeel van beide partijen.

9. Bij vele Nederlandse projecten in het kader van ontwikkelingssamenwerking vindt toegepast onderzoek plaats. Het is dan ook uiterst wenselijk dat statistische begeleiding een vast onderdeel wordt van dergelijke projecten, zoals dat al gebruikelijk is bij de projecten van de Britse "Overseas Development Administration".

10. Om de herkenbaarheid in het buitenland van afgestudeerden van de Landbouwuniversiteit te verbeteren, zou het gewenst zijn na het eerste trimester doctoraal B een meetpunt in te stellen. Een succesvolle afsluiting van dit gedeelte van de studie kan dan gehonoreerd worden met een "B.Sc."-diploma. De daarop volgende vijf trimesters van de studie leiden dan tot een "M.Sc."-diploma. Bovendien kunnen abituriënten van het Hoger Agrarisch Onderwijs dan instromen in verwante "M.Sc."-opleidingen.

11. De emancipatie van de vrouw kan wellicht bevorderd worden door in de veefokkerij de kruisingen niet als ♂ x ♀ te noteren maar, zoals in de plantenveredeling gebruikelijk is, als ♀ x ♂.

Proefschrift van L.R. Verdooren
Statistical Inference on Variance Components
Wageningen, 3 februari 1988

VOORWOORD

Bij de publicatie van dit proefschrift wil ik mijn diepe erkentelijkheid uitspreken aan mijn promotor Professor dr. P. van der Laan. Uw grote belangstelling voor mijn onderzoek en de vriendschappelijke drang om dit onderzoek af te ronden, heeft geleid tot dit proefschrift. Ik wil U nogmaals bedanken voor alle aandacht die U steeds voor mij heeft gehad.

Ik wil aan al mijn leermeesters dank betuigen. In het bijzonder wil ik twee hoogleraren noemen:

(i) Professor dr. N.H. Kuiper.

Uw heldere en stimulerende colleges in de Wiskunde en Wiskundige Statistiek hebben mij tot de liefde voor de Statistiek gebracht. De tijd waarin ik op Uw afdeling Wiskunde mocht werken is beslissend geweest voor mijn vorming.

(ii) Professor dr.ir. F. Hellinga.

Uw meesterlijke aanpak hoe theorie toegepast moet worden in de praktijk probeer ik nog steeds na te volgen. U heeft mij, toen ik nog student in de Cultuurtechniek was, in contact gebracht met het wetenschappelijk onderzoek, waarin ik nog altijd verkeer.

Om onderzoek te doen is de werkomgeving heel belangrijk. Ik wil al mijn collega's bedanken voor de prettige sfeer waarin wij naast het onderwijs nog onderzoek doen. Toch wil ik een aantal speciaal noemen. Professor dr.ir. L.C.A. Corsten ben ik zeer erkentelijk voor zijn grote bijdrage aan de verdieping van mijn statistische kennis. Dr. A.C. van Eijnsbergen dank ik voor de prettige samenwerking en in het bijzonder voor de tijd die wij samen besteedden aan Seely's theorie omtrent het schatten van variantiecomponenten. Dr.ir. M.A.J. van Montfort heeft met zijn praktische kijk op de toepassing van de statistiek mij altijd gestimuleerd. Drs. A. Otten dank ik voor zijn bereidheid om

altijd naar mijn wilde theorieën te luisteren en dan met zijn scherpzinnige commentaar te komen.

Zonder de secretariële ondersteuning zou er niets van mijn onderzoek tot stand zijn gekomen. Ik wil daarom de dames bedanken die mijn slordige en soms onleesbare manuscripten tot persklare artikelen omgetoverd hebben: Petra Hendriks, Heleen van Holland, Ditsie Massop en niet te vergeten onze vroegere medewerksters Renée Hoeffnagel en Jodien Houwers.

De mooie omslag van dit proefschrift is ontworpen door Annette van der Laan. Haar wil ik hierbij danken voor alle tijd die ze besteed heeft om dit fraaie ontwerp te maken.

Mijn ouders dank ik dat zij mij hebben laten studeren. Na het overlijden van Pappie heeft Moes mij altijd aangespoord om een proefschrift te schrijven. Dit werk is er nu voor Moes.

Tenslotte wil ik mijn echtgenote Pieneke bedanken voor al de steun die zij mij heeft gegeven en nog geeft. Een loopbaan als onderzoeker doet een aanslag op het gezinsleven. Dat het gezin dit op kan brengen is te danken aan mijn lieve dochters Karin en Angela en zeker aan jou Pieneke. Daarom is dit proefschrift ook aan jou opgedragen.

Ontwerp omslag: Annette van der Laan

CONTENTS

Page

CHAPTER 1

# 1. OUTLINE OF AUTHOR'S RESEARCH ON THE INFERENCE OF VARIANCE COMPONENTS

The historical development of variance components estimation is described in chapter 2.

Just as with the fixed effects model, the description of the models for variance components (random models and mixed models) was from 1947 onwards in terms of equations with subscripts. Sometimes the whole set of observations was described with matrices. The derivation of the interesting quantities, such as Sum of Squares (SS) in the Analysis of Variance Table (ANOVA-table), Expected Mean Squares (EMS) and distribution of ratio of Sums of Squares, was often done by manipulating expressions with sums and squares of sums, where several summation symbols ($\Sigma$ 's) with different indices were needed. These clumsy notations obscured the insight into the derivation of the results.

For the fixed effects model there were already some persons who made systematic use of vectors and vector spaces to describe it, see Kuiper (1952), Corsten (1958) and Scheffé (1959). In chapter 3 the representation of variance components models with vectors and vector spaces is described. This approach for random and mixed models has been adopted also in the lectures on Applied Statistics at the Agricultural University in Wageningen, since 1972.

For the balanced nested designs exact tests about ratio of variance components and the calculation of their power, are well-known. For the unbalanced nested designs an exact test sometimes exists, but the calculation of the power was troublesome, or an exact test was unknown.

In a seemingly unrelated problem field, that of economics, there is the problem to determine serial correlations in series

of data. The often used test statistic is that of Durbin and Watson. Another procedure is the use of the residuals with the best linear unbiased estimators with a scalar covariance matrix (BLUS). For the distribution function of these test statistics, the distribution function of a ratio of quadratic forms in normally distributed random variables is needed. For the calculation of this distribution function a computer program was available since 1969. Statisticians not working in economics were unaware of this. The author encountered this problem also in his investigation on exact tests for variance ratios in unbalanced three-stage nested designs. He recognized the similarity and used the program to calculate the distribution function of such a ratio of quadratic forms. These results of determining exact tests about ratio of variance components in unbalanced nested two and three-stage nested designs have been presented at the "8th International Biometric Conference" at Constanta, Romania, in 1974 (see section 4.2).

The related problem of an exact confidence interval for the ratio of variance components has also been tackled (see section 4.2). Because this confidence interval was very useful in practice, this part has again been presented at the second International Symposium of "Computational Statistics (COMPSTAT)" at West-Berlin in 1976 (see section 4.3).

In the meantime many approximate and incorrect tests about ratio of variance components have been published. An overview and an extension of the work about exact tests and confidence intervals for ratio of variance components have been presented at the International Conference on the "Analysis of the Unbalanced Mixed Model" at Gainesville, Florida, U.S.A., in 1987 (see section 7.2).

Using the possibility to compute the exact distribution of ratios of quadratic forms in normally distributed variables, the probability

of negative outcomes of the ANOVA estimator for the balanced one-way classification has been computed (see chapter 6).

The development of estimators of variance components was concentrated more and more on best quadratic unbiased estimators (Best = Minimum Variance). The author introduced the concept of a permissible estimator to find another and "better" estimator at the "VI-th International Conference on Mathematical Statistics" at Wisla, Poland, in 1978. The unified least squares procedure has been used to find a non-negative, closest to unbiased, estimator for variance components (see section 5.2).

These ideas about the unified least squares procedure for estimating variance components and an overview over permissible estimators for variance components were also presented at the International Conference on the "Analysis of the Unbalanced Mixed Model" at Gainesville, Florida, U.S.A., in 1987 (see section 8.2).

The use of variance components has extensively been applied in Animal Breeding, especially using the BLUP-procedure. In Plant Breeding no use has yet been made of variance components to get the best linear unbiased estimator for varietal contrasts and the best linear unbiased predictor for varietal contrasts. At the second International Seminar on "Statistical Methods in Variety Testing" at Slupia Wielka, Poland, in 1985, the author presented this application (see section 9.2).

## REFERENCES

Corsten, L.C.A. (1958). Vectors, a tool in statistical regression theory. Mededelingen van de Landbouwhogeschool 58 (1), Wageningen. Mededelingen van het Instituut voor Rassenonderzoek van Landbouwgewassen, no. 35, Wageningen.

Kuiper, N.H. (1952). Variantie-analyse. Statistica 6, 149-194. An english translation by M. Keuls appeared in 1983: Analysis of Variance. Mededelingen Landbouwhogeschool, 83-10, Wageningen.

Scheffé, H. (1959). The Analysis of Variance. John Wiley and Sons, New York.

CHAPTER 2

# 2. HISTORICAL DEVELOPMENT OF VARIANCE COMPONENTS ESTIMATION

## 2.1. Introduction

For the history of models with random effects we rely on Scheffé (1956), Anderson (1979), Khuri and Sahai (1985) and Searle (1987). The historical development reflects the recognition of the importance of variance component models in practice.

The struggle for good estimators of variance components is sketched. Our personal choice of a good estimator for a variance component will be given in chapter 5.

## 2.2. The first period from 1861 to 1947

Estimation of fixed effects in linear models, in which the covariance matrix of the errors is known up to a scalar factor, essentially began with Legendre (1806) and Gauss (1809). Independently from each other they invented the theory of least squares in books on astronomical problems, so that it is with these problems that we must associate the origin of the fixed-effects models (see also Plackett (1972) and Stigler (1981)).

What is even more interesting is that the first appearance of variance components is also in astronomy books by Airy (1861) and by Chauvenet (1863). A one-way random model was formulated by Airy (1861, especially part IV) and there is also a provision for unbalanced data, unequal numbers of telescopic observations from night to night on the same phenomenon of interest. Suppose that on I nights [Airy uses days (!)] *observations are made with a telescope on the same phenomenon,* $J_i$ observations on the i-th night. Airy assumes the following structure for the j-th observation on the i-th night:

$$y_{ij} = \mu + c_i + e_{ij} \qquad (2.1)$$

where $\mu$ is the general mean or "true" value, and the $\{c_i\}$ and $\{e_{ij}\}$ are random effects with the following meanings: He calls $c_i$ the "constant error", meaning it is constant on the i-th night; (we would call it the i-th night effect); it is caused by the "atmospheric and personal circumstances" peculiar to the i-th night. The $\{e_{ij}\}$ for fixed i we would call the errors about the (conditional) mean $\mu + c_i$ on the i-th night. It is implied by Airy's discussion that he assumes all the $e_{ij}$ independently and identically distributed, similarly for the $c_i$, that the $\{e_{ij}\}$ are independent of the $\{c_i\}$, and that all have zero means. Let us denote the variances of the $\{e_{ij}\}$ and the $\{c_i\}$ by $\sigma_e^2$ and $\sigma_c^2$.
To decide about his equivalent of the hypothesis $\sigma_c^2 = 0$, Airy compares, as we would, a between-nights measure of variability with a within-nights measure, but he uses different measures than we would.

The second use of a random effects model appears to be Chauvenet (1863, Vol. II, Articles 163 and 164). Although he did not write model equations like (2.1) with all the subscripts, nevertheless he implied a one-way classification random model in which, using today's notation, he derived the variance of $\bar{y}_{..} = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}/(IJ)$ as $\mathrm{var}(\bar{y}_{..}) = (\sigma_c^2+\sigma_e^2/J)/I$ from (2.1). Chauvenet suggests that there is in his example little practical advantage in having J greater than 5, and refers to Bessel (1820, p. 166) for this idea. Chauvenet's reference to Bessel on this specific point (J = 5) is incorrect, but the page he cites does contain a formula for the probable error of a sum of independent random variables which could be the basis for such a conclusion. Probably Bessel made the remark elsewhere. (Probable error is an older measure of sampling variability and is equal to $0.6745\sigma$. The reason for this measure is that for a random variable $\underline{x}$, which is normally distributed with mean $\mu$ and variance $\sigma^2$, $N(\mu,\sigma^2)$, $P(-0.6745\sigma < \underline{x} - \mu < +0.6745\sigma) = 0.5$)

Chauvenet's contribution was not so much in his implicit use of a linear model, but was rather his demonstration of the utility of an a priori knowledge of the relative magnitudes of components of variance in the design of experiments.

The more modern beginning of variance components is Fisher's (1918) paper on quantitative genetics in which he initiated (i) the use of the terms "variance" and "analysis of the variance"; (ii) an implicit employment of variance components models; and (iii) definitive ascription of percentages of a total variance to constituent causes, e.g. that dominance deviations accounted for 21 percent of the total variance in human stature wheras, prior to the study, dominance deviations were ascribed to be environmental effects. (The other two basic papers on quantitative genetics are from Wright (1921) and Haldane (1932)).

In his book: Statistical Methods for Research Workers, section 40, Fisher (1925) made yet another major contribution to variance component models. He initiated there what has come to be known as the analysis of variance (ANOVA) method of estimation: equate sums of squares from an analysis of variance to their expected values and thereby obtain a set of equations that are linear in the variance components to be estimated. This idea arose from using an analysis of variance to estimate an intra-class correlation from balanced data pertaining to a one-way random model. In to-day's notation for the one-way classification random model like (2.1), with balanced data, Fisher used

$$E(SSE) = E\sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij} - \bar{y}_{i.})^2 = I(J-1)\sigma_e^2$$

and

$$E(SSA) = E\sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{y}_{i.} - \bar{y}_{..})^2 = E\sum_{i=1}^{I}J(\bar{y}_{i.} - \bar{y}_{..})^2 = (I-1)(J\sigma_c^2+\sigma_e^2).$$

From these the estimation equations are taken as SSE = $I(J-1)\hat{\sigma}_e^2$ and

SSA = $(I-1)(J\hat{\sigma}_c^2 + \hat{\sigma}_e^2)$ and so $\hat{\sigma}_e^2$ = SSE/[I(J-1)] = MSE and

$\hat{\sigma}_c^2$ = (MSA - MSE)/J where MSA = SSA/(I-1).

Although Fisher's book extended the analysis of variance to the one-way model with unbalanced data (different $J_i$'s) and to the two-way crossed classification model with interaction, his book did not treat the estimation of variance components.

The analysis of variance method for unbiased estimation of variance components from balanced as well as unbalanced one-way random models was later elucidated by Tippett (1931, Table XXIV) who also, for the first time, extended the method to handle the two-way crossed classification random model without interaction, and displayed in the second edition of his book (in 1937) some explicit estimators. Tippett (1931) also considered the problem of selecting the optimal sampling designs for particular experimental situations for a one-way random model, as Chauvenet (1863) had done, and perhaps Bessel (1820).

Yates and Zacopanay (1935) later dealt comprehensively with higher order sampling designs in several experiments. Neyman, Iwaszkiewicz and Kolodziejczyk (1935) considered the efficiency of randomized blocks and Latin Square designs, and in doing so made extensive use of linear models (including mixed models).

Although Fisher (1935) used the term "components of variation" in an acrimonious review of Neyman et al. (1935), and they themselves had used the phrase "error components", the first apparent use of "components of variance" is by Daniels (1939). He also derived the sampling variances of the variance component estimators for complex models, such as three-way crossed classification random models containing two-factor and three-factor interaction effects.

The analysis of unbalanced data is much more complex. Cochran (1939) used the analysis of variance in selecting an optimal design in which the need for a method for estimating variance components arose. The analysis of variance (ANOVA) method for unbiased estimating of variance components in the one-way random model with unbalanced data has been given by Winsor and Clarke (1940). Yet, although Fisher (1925) has the idea of taking expected values, he had not there specifically formulated it using the E Operator (E = Expected value) as do Daniels (1939), and Winsor and Clarke (1940). The method was subsequently extended to k-stage nested classifications (k > 2) by Ganguli (1941), and to the two-way crossed classification random model with interaction by Crump (1946, 1947).

Both Ganguli and Crump noted that the ANOVA method sometimes produced negative estimates and they suggested that such estimates be replaced by zero.

While a mixed model is implied by Fisher's (1935, Sec. 65) discussion of varietal trials in a randomly selected set of locations, and by Yates' (1935) analysis of the split-plot design, the first explicit clear description of a mixed model is in a paper on mental tests by Jackson (1939), where the score $y_{ij}$ of the j-th individual on the i-th trial of a test is assumed to have the structure $y_{ij} = \mu + \alpha_i + b_j + e_{ij}$, where the trial effects $\{\alpha_i\}$ are treated as fixed and the "individual" effects $\{b_j\}$ are treated as random. The $\{b_j\}$ have to be independently and identically distributed with zero means. The errors $\{e_{ij}\}$ are assumed to be independently distributed with zero means and equal variances and also independent of the $\{b_j\}$.

The systematic description of fixed effects models as "Model I", random effects models as "Model II" and mixed effects models as "Model III" is due to Eisenhart (1947). This is a landmark paper.

Finally it can be remarked that the use of variance components in

animal breeding has been applied from 1923 onwards by J.L. Lush who in 1937 wrote his well-known book "Animal breeding plans". The theory of the selection index was developed by Smith (1936) and Hazel (1943). Estimating variance components is an integral aspect of animal and plant breeding for two reasons: (i) to identify sources of variation, principally genetic variation; and (ii) as an adjunct to the prediction of breeding values of candidates for selection.

Use of components of variance in animal breeding probably started with simple between SS and within SS in one-way ANOVA, to get estimates of between-group variation and as a way to compute correlations and regressions when the same attributes were not measured on each individual. An example of the latter is when an estimate of repeatability is wanted. Lush and Jones (1923) published the concept of repeatability. Dickerson (1942) estimated components of variance in order to design experiments for testing inbred lines of swine. For further interesting historical facts in animal breeding, we refer to Freeman (1973, 1979).

Some early applications are contained in the papers of Lush, Hetzer and Culbertson (1934), Bywaters (1937), Stonaker and Lush (1942), Hetzer, Dickerson and Zeller (1944), Knapp and Nordskog (1946) and Anderson (1947), to name just a few.

## 2.3. The second period from 1947 to 1987

Impetus for the fledgling method of variance component estimation came from a wide variety of practical problems, especially genetics. We already mentioned in animal breeding Lush (1937); another pioneer there is Lerner (1950). In plant breeding pioneers were Comstock and Robinson (1948), Mather (1949), Hayman (1954a, b), Jinks (1954), Griffing (1956), Finlay and Wilkinson (1963).

The first comprehensive discussion in book form was presented by Anderson and Bancroft (1952), with its four chapters on variance components.

The above mentioned papers dealt with balanced data or with unbalanced data in nested designs with random effects. Because it directed attention to the need for alternative methods of estimating, which would facilitate the handling of mixed models with unbalanced data, the study of Henderson (1948) represented a major contribution to variance component theory. For estimating variance components from unbalanced data the landmark paper is undoubtedly Henderson (1953). This contains three methods 1, 2 and 3 for unbiased estimation of variance components. In that paper the ANOVA method of estimation, based on equating analysis of variance sums of squares to their expected values, was extended for unbalanced data, to equating a wide variety of quadratic forms (not all of them sums of squares) to their expected values.

Henderson's Method 1 may only be applied to random effects models. For each effect an "analogous sum of squares" of the effect is calculated, with the same recipe as for the balanced case. (For a crossed classification the "analogous sums of squares" for interactions are not always non-negative definite!).
The expected values of these "analogous sum of squares" are determined. The unbiased estimator for the variance components follows from equating the expected values of the "analogous sum of squares" to their outcomes. Notice that Henderson's Method 1 gives the correct unbiased estimators for nested designs.

Henderson's Method 2 may only be applied to mixed effects models, which do not include either interaction effects between fixed and random factors or nesting of fixed and random factors within each other. The basic strategy of Henderson's Method 2 is to estimate the

fixed effects in the model using least squares (computing as though the random effects were fixed), correct the data vector in accordance with these estimates, and then apply Henderson's Method 1 to the adjusted data vector. Accordingly, the computational simplicity of Method 1 is retained.

Henderson's Method 3 utilizes reductions in sums of squares resulting from fitting the linear model and sub-models thereof using the least squares procedures of fitting constants. The random effects are considered as constant effects for the least squares fitting procedure. For mixed models the reductions start with fitting the fixed effects and then the random effects. Equations leading to unbiased estimators of the variance components are then obtained by equating each such reduction to its expected value under the full model.

The difficulty with Henderson's Method 3 was that even for the easiest crossed classification with two random factors A and B and unbalanced data, there was no unique estimator! The estimator depends on whether the reduction in sums of squares was calculated in the order of A and then B corrected for A, or in the order of B and then A corrected for B. Nevertheless this Method 3 has been used extensively in practice, because there soon were computer programs to handle it (e.g. Harvey's program (1960)).

Then followed a period (from 1953 to 1970) of trying to evaluate those methods mostly through deriving expressions, under normality assumptions, for sampling variances of the resulting estimators. These results are compiled in Searle's (1971) book. In every case the results are, of course, quadratic functions of the unknown variance components; but the coefficients of the squares and products of those components are such hopelessly intractable functions of the numbers of observations in the cells of the data that it is impossible to make analytic comparisons

either between different estimation methods, or between the effects of different degrees of data unbalancedness on any one method of estimation. This absence of tractable criteria, on which judgement can be based as to which application of the ANOVA method has optimal features, thus became very frustrating.

For balanced data an optimal criterion exists. Graybill (1954), Graybill and Wortham (1956) and Graybill and Hultquist (1961) had established the properties of minimum variance for the unbiased quadratic ANOVA estimators.

But for unbalanced data the frustration of "no optimality criterion" persisted. In contrast to the optimal properties of the ANOVA estimators for balanced data, the best estimators in the unbalanced cases of random models and mixed models are not known. Read (1961) showed that unbiased quadratic estimators with uniformly minimum variance do not exist if the design for the random model is not balanced.

For unbalanced data Henderson's Method 3 was used, and among the several pieces of work which have dealt with this method only that of Harville (1967) seems to have been concerned with consistency of the equations leading to the estimators and to the existence of unbiased (quadratic) estimators under various conditions. Harville (1967), however, only treats a completely random two-way classification model with interaction. One other result which deals with existence of unbiased quadratic estimators in a completely random model is given by Graybill and Hultquist (1961).

Seely (1970a, b; 1971; 1972; 1977) and Seely and Zyskind (1971) introduced a least squares approach for estimating variance components. Seely's method has the advantage of giving necessary and sufficient conditions for the existence of unbiased quadratic estimators,

which Henderson's Method 3 does not always supply. For a description of Seely's method we refer to chapter 5, section 5.2 and chapter 8, section 8.2.

In the meantime Rao (1970; 1971a, b; 1972) introduced a method equivalent to Seely's method for deriving estimators for variance components. Rao's minimum norm quadratic unbiased estimator (MINQUE) has become well-known.

For joint normally distributed data LaMotte (1973) derived the minimum variance quadratic unbiased estimator (MIVQUE). This estimator is the same as Rao's MINQUE in the case of joint normally distributed data. For a description of Rao's MINQUE see chapter 5, section 5.2 and chapter 8, section 8.2.

When we restrict ourselves to joint normally distributed data, the method of maximum likelihood to get estimators for the parameters has been applied for the unbalanced mixed model by Hartley and Rao (1967). Another modification of this method was the restricted maximum likelihood (REML) estimation, initiated for balanced data by Anderson and Bancroft (1952) and Thompson (1962), and extended by Patterson and Thompson (1971, 1975) to block designs and thence to unbalanced data generally. For a description of this REML method see chapter 5, section 5.2.

Further desirable developments and recommendations in variance components estimation will be made in section 8.3.

## REFERENCES

Airy, G.B. (1861). On the algebraical and numerical theory of errors of observations and the combination of observations. Macmillan and Co., Ltd., Cambridge and London.

Anderson, R.D. (1979). On the history of variance component estimation. Proceedings of a conference in honor of C.R. Henderson, July 16-17, 1979. Variance components and animal breeding. Ed. Vleck, L.D. van and Searle, S.R.; Cornell University, Ithaca, N.Y., 19-42.

Anderson, R.L. (1947). Use of variance components in the analysis of hog prices in two markets. Journ. Amer. Statist. Assoc. 42, 612-634

Anderson, R.L. and Bancroft, T.A. (1952). Statistical Theory in Research. Mc Graw-Hill, New York.

Bessel, I. (1820). Beschreibung des auf der Königsberger Sternwarte aufgestellten Reichenbachschen Meridiankreises, dessen Anwendung und Genauigkeit, imgleichen der Repoldschen Uhr. Astronomisches Jahrbuch für das Jahr 1823. Berlin.

Bywaters, J.H. (1937). The hereditary and environmental portions of the variance in weaning weights of Poland-China pigs. Genetics 22, 457-468.

Chauvenet, W. (1863). A manual of spherical and practical astronomy, Vol. I and II. J.B. Lippincott Company, Philadelphia.

Cochran, W.G. (1939). The use of the analysis of variance in enumeration by sampling. Journ. Amer. Statist. Ass. 34, 492-510.

Comstock, R.E. and Robinson, H.F. (1948). The components of genetic variation in populations of biparental progenies and their use in estimating the average degree of dominance. Biometrics 4, 254-266.

Crump, S.L. (1946). The estimation of variance components in analysis of variance. Biometrics Bulletin 2, 7-11.

Crump, S.L. (1947). The estimation of components of variance in multiple classifications. Ph.D. Thesis, Iowa State University, Ames, Iowa.

Daniels, H.E. (1939). The estimation of components of variance. J. Roy. Stat. Soc., Suppl. 6, 186-197.

Dickerson, G.E. (1942). Experimental design for testing inbred lines of swine. Journ. Anim. Sci. 1, 326-341.

Eisenhart, C. (1947). The assumptions underlying the analysis of variance. Biometrics 3, 1-21.

Finlay, K.W. and Wilkinson, G.N. (1963). The analysis of adaptation in a plant breeding program. Austr. Journ. Agr. Res. 14, 742-754.

Fisher, R.A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. Trans. Royal Society Edinburgh 52, 399-433.

Fisher, R.A. (1925). Statistical Methods for Research Workers, 1st Edition, Oliver and Boyd Ltd., Edinburgh.

Fisher, R.A. (1935). The design of experiments. Oliver and Boyd Ltd, Edinburgh.

Freeman, A.E. (1973). Genetic statistics in animal breeding. Proceedings of Animal Breeding and Genetics Symposium in Honor of Dr. J.L. Lush. American Society of Animal Science and the American Dairy Science Association, Champaign, Illinois, 1-9.

Freeman, A.E. (1979). Components of variance: Their history, use and problems in animal breeding. Proceedings of a conference in honor of C.R. Henderson, July 16-17, 1979. Variance components and animal breeding. Ed. Vleck, L.D. van and Searle, S.R. Cornell University, Ithaca, N.Y., 43-57.

Ganguli, M. (1941). A note on nested sampling. Sankhyā 5, 449-452.

Gauss, C.F. (1809). Theoria motus corporum coelestrium in sectionibus conicis solem ambientium. Perthes and Besser, Hamburg. (Translation reprinted in 1963 as Theory of the motions of the heavenly bodies moving about the sun in copic sections. Dover, New York.)

Graybill, F.A. (1954). On quadratic estimates of variance components. Ann. Math. Statist. 25, 367-372.

Graybill, F.A. and Wortham, A.W. (1956). A note on uniformly best unbiased estimators for variance components. Journ. Amer. Statist. Assoc. 51, 266-268.

Graybill, F.A. and Hultquist, R.A. (1961). Theorems concerning Eisenhart's model II. Ann. Math. Statist. 32, 261-269.

Griffing, B. (1956). Concept of general and specific combining ability in relation to diallel crossing systems. Austr. Journ. Biol. Sci 9, 463-493.

Haldane, J.B.S. (1932). The causes of evolution. Longmans, Green and Co., London.

Hartley, H.O. and Rao, J.N.K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. Biometrika 54, 93-108.

Harvey, W.R. (1960). Least squares analysis of data with unequal subclass numbers. United States Department of Agriculture. AR2 20-8.

Harville, D.A. (1967). Estimability of variance components for the two-way classification with interaction. Ann. Math. Statist. 38, 1508-1519.

Hayman, B.I. (1954a). The analysis of variance in diallel tables. Biometrics 10, 235-244.

Hayman, B.I. (1954b). The theory and analysis of diallel crosses. Genetics 39, 789-809.

Hazel, L.N. (1943). The genetic basis for constructing selection indexes. Genetics 28, 476-490.

Henderson, C.R. (1948). Estimation of general, specific and maternal combining abilities in crosses among inbred lines of swine. Ph.D. Thesis, Iowa State University, Ames, Iowa.

Henderson, C.R. (1953). Estimation of variance and covariance components. Biometrics 9, 226-252.

Hetzer, H.O., Dickerson, G.E. and Zeller, J.H. (1944). Heritability of type in Poland China swine as evaluated by scoring. Journ. Anim. Sci. 3, 390-398.

Jackson, R.W.B. (1939). The reliability of mental tests. British J. Psych. 29, 267-287.

Jinks, J.L. (1954). The analysis of continuous variation in a diallel cross of Nicotiana rustica varieties. Genetics 39, 767-788.

Khuri, A.I. and Sahai, H. (1985). Variance components analysis: A selective literature survey. Intern. Statist. Review 53, 279-300.

Knapp, B.Jr. and Nordskog, A.W. (1946). Heritability of growth and efficiency of beef cattle. Journ. Anim. Sci. 5, 62-70.

LaMotte, L.R. (1973). Quadratic estimation of variance components. Biometrics 29, 311-330.

Legendre, A.M. (1806). Nouvelles méthodes pour la détermination des orbites des comètes; avec un supplément contenant divers perfectionnements de ces méthodes et leur application aux deux comètes de 1805. Courcier, Paris.

Lerner, I.M. (1950). Population genetics and animal improvement. Cambridge University Press, Cambridge.

Lush, J.L. (1937). Animal Breeding Plans, 1st edition. Iowa State College Press, Ames, Iowa.

Lush, J.L. and Jones, J.M. (1923). The influence of individuality, age and season upon weights of fleeces produced by range sheep. Texas Agricultural Experiment Station Bulletin 311.

Lush, J.L., Hetzer, H.O. and Culbertson, C.C. (1934). Factors affecting birth weights of swine. Genetics 19, 329-343.

Mather, K. (1949). Biometrical genetics. Methuen & Co., London.

Neyman, J., Iwaszkiewicz, K. and Kolodziejczyk, St. (1935). Statistical problems in agricultural experimentation. J. Royal Stat. Soc. Suppl. 2, 107-154.

Patterson, H.D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545-554.

Patterson, H.D. and Thompson, R. (1975). Maximum likelihood estimation of components of variance. Proceedings of the 8th International Biometric Conference. Ed. Corsten, L.C.A. and Postelnicu, J., Constanta, 197-207.

Plackett, R.L. (1972). Studies in history of probability and statistics XXIX. The discovery of the method of least squares. Biometrika 59, 239-251.

Rao, C.R. (1970). Estimation of heteroscedastic variances in linear models. Journ. Amer. Statist. Assoc. 65, 161-172.

Rao, C.R. (1971a). Estimation of variance and covariance components - MINQUE theory. Journ. Multivar. Anal. 1, 257-275.

Rao, C.R. (1971b). Minimum variance quadratic unbiased estimation of variance components. Journ. Multivar. Anal. 1, 445-456.

Rao, C.R. (1972). Estimation of variance and covariance components in linear models. Journ. Amer. Statist. Assoc. 67, 112-115.

Read, R.R. (1961). On quadratic estimates of the interclass variance for unbalanced designs. Journ. Roy. Statist. Soc. B 23, 493-497.

Scheffé, H. (1956). Alternative models for the analysis of variance. Ann. Math. Statist. 27, 251-271.

Searle, S.R. (1971). Linear Models, John Wiley and Sons, Inc., New York.

Searle, S.R. (1987). Mixed models and unbalanced data: Wherefrom, wherat and whereto? To appear in Commun. Statist. Theor. -Meth. A16, ?.

Seely, J. (1970a). Linear spaces and unbiased estimation. Ann. Math. Statist. 41, 1725-1734.

Seely, J. (1970b). Linear spaces and unbiased estimation - application to the mixed linear model. Ann. Math. Statist. 41, 1735-1748.

Seely, J. (1971). Quadratic subspaces and completeness. Ann. Math. Statist. 42, 710-721.

Seely, J. (1972). Completeness for a family of multivariate normal distributions. Ann. Math. Statist. 43, 1644-1647.

Seely, J. (1977). Minimal sufficient statistics and completeness for multivariate normal families. Sankya A39, 170-185.

Seely, J. and Zyskind, G. (1971). Linear spaces and minimum variance unbiased estimation. Ann. Math. Statist. 42, 691-703.

Smith, H. Fairfield (1936). A discriminant function for plant selection. Ann. Eugen. 7, 240-250.

Stigler, S.M. (1981). Gauss and the invention of least squares. Ann. Statist. 9, 465-474.

Stonaker, H.H. and Lush, J.L. (1942). Heritability of conformation in Poland-China swine as evaluated by scoring. Journ. Anim. Sci. 1, 99-105.

Thompson, W.A., Jr. (1962). The problem of negative estimates of variance components. Ann. Math. Statist. 33, 273-289.

Tippett, L.H.C. (1931). The Methods of Statistics, 1st Edition. Williams and Norgate Ltd., London. (2nd Edition (1937)).

Winsor, C.P., and Clarke, G.L. (1940). A statistical study of variation in the catch of plankton nets. Sears Foundation J. Marine Res. 3, 1-34.

Wright, S. (1921). Systems of mating. Genetics 6, 111-178.

Yates, F. (1935). Complex experiments. J. Roy. Stat. Soc., Suppl. 2, 181-247.

Yates, F. and Zacopanay, I. (1935). The estimation of the efficiency of sampling, with special reference to sampling for yield in cereal experiments. J. Agric. Sci., Camb. 25, 545-577.

CHAPTER 3

# 3. REPRESENTATION OF VARIANCE COMPONENT MODELS

## 3.1. Introduction

For the balanced scheme the representation with vectors and vector spaces is described in the following article, given in section 3.2.

In this article it was stated in the Introduction that for the more complicated mixed models with correlated random effects, the theory presented there was to be extended in a subsequent publication. The author later used the ideas developed here, together with the general mixed model equation $\underline{y} = X\beta + Z\underline{u} + \underline{e}$, where the stochastically independent random vectors $\underline{u}$ and $\underline{e}$ have zero mean vectors ($E\underline{u} = 0$ and $E\underline{e} = 0$), and dispersion matrices $D(\underline{u}) = E(\underline{u}\underline{u}')$ and $D(\underline{e}) = E(\underline{e}\underline{e}')$, respectively. The correlated effects have been included in $D(\underline{u})$ and $D(\underline{e})$. Hence the expectation vector $E\underline{y} = X\beta$, and the dispersion matrix of $\underline{y}$ is $D(\underline{y}) = ZD(\underline{u})Z' + D(\underline{e})$.
This representation has been used throughout in the following chapters and, especially for correlated random effects, in section 9.2.

The projection of the observation vector $\underline{y}$ on a vector space A is denoted as $P_A\underline{y}$ rather than $\underline{y}_A$. Here $P_A$ stands for the matrix of orthogonal projection on the vector space A.

## 3.2. Representation of ANOVA models with vectors and vector spaces

Published in:

Statistica Neerlandica 23 (1969) no. 1, pp. 53 - 65.

Also published as:

Mededelingen van het Instituut voor Rassenonderzoek van Landbouwgewassen, no. 57-1969, Wageningen.

# Representation of ANOVA models with vectors and vector spaces

by L. R. VERDOOREN *

Summary The use of vectors and vector spaces for the representation of the fixed-effects models for the analysis of variance (ANOVA) is well-known. This representation gives a clear understanding of the estimation and hypothesis testing problems.
A similar representation can be used for the random-effects models and some mixed models. As a result the distribution and the expectation of the mean squares in the ANOVA table can easily be derived.

## 1 Introduction

The representation of the fixed-effects models for the analysis of variance (ANOVA) with vectors and vector spaces was developed in a systematic manner by KUIPER (1952) and CORSTEN (1958). Furthermore the distribution of the sums of squares in the ANOVA table can easily be found by interpreting such a sum of squares as the square of the orthogonal projection of a normal random vector on a subspace, compare KUIPER (1959, 1960). To recall the procedure a few examples are given in section 2.

The vector representation can also be applied to the random-effects ANOVA models to elucidate there consequences. This was pointed out by KUIPER in his 1962 lectures on mathematical statistics at the Landbouwhogeschool (Agricultural University) of Wageningen. A few examples are given to illustrate this technique in section 3.

For some mixed-effects ANOVA models the theory can readily be applied as shown in section 4. For the more complicated mixed models with correlated random-effects the theory presented here should be extended, and this will be done in a subsequent publication.

## 2 Fixed-effects ANOVA models

### 2.1 *The one-way classification*

The one-way layout refers to the comparison of the means $\beta_i$ of the populations $i$ ($i = 1, \ldots, I$) which are normal, each with the same variance $\sigma^2$. From each population there are $n_i$ independent random samples $y_{ij}$ ($j = 1, \ldots, n_i$). The sample scheme is then:

| Populations or Classification $A$: | |
|---|---|
| 1 | $y_{11}, \ldots, y_{1n_1}$ |
| 2 | |
| . | |
| . | |
| $I$ | $y_{I1}, \ldots, y_{In_I}$ |

* Institute for Research on Varieties of Field Crops (IVRO), Wageningen.

The model is:

$$\underline{y}_{ij} = \beta_i + \underline{e}_{ij}$$

whereby $\{\underline{e}_{ij} \simeq \sigma\underline{\chi}\}$ are mutually stochastically independent ($i = 1, \ldots, I$; $j = 1, \ldots, n_i$).

Note that the standard normal random variable $\underline{\chi}$ is defined by

$$P(\underline{\chi} < x) = \int_{-\infty}^{x} (2\pi)^{-\frac{1}{2}} e^{-t^2/2} dt;$$

two random variables $\underline{x}$ and $\underline{y}$ are called isomorous ($\underline{x} \simeq \underline{y}$) when they have the same distribution function.

The sample scheme is looked upon as a vector $y$ in a real vector space $R^n$ of dimension $n$ with $n = \sum_{i=1}^{I} n_i$. Corresponding to the one-way classification $A$ we define a vector subspace $A \subset R^n$, consisting of vectors with equal coordinates within each class of $A$. A basis of this vector subspace of dimension $I$ is given by the class-characteristic vectors $s_1, \ldots, s_I$, whereby $s_i$ has coordinates equal to one merely in the class $i$ of $A$ and zeros elsewhere. Hence the expectation vector $\mathscr{E}\underline{y} \in A$. Under the null hypothesis $H_0 : \beta_1 = \beta_2 = \ldots = \beta_I$ the vector $\mathscr{E}\underline{y} \in L$, the subspace of levels with the basisvector $s_0$ consisting merely of ones. Obviously: $s_0 = s_1 + \ldots + s_I$, $L \subset A$.

Let the orthogonal complement of $L$ in $A$ be $A^*$, then under $H_0$ $\mathscr{E}\underline{y} \in L$ and the orthogonal projection of $\mathscr{E}\underline{y}$ on $A^*$, $(\mathscr{E}\underline{y})_{A^*}$, is the null vector. Let the orthogonal complement of $A$ in $R^n$ be $R$, the space of residuals or the error space, then the dimension of $R$, $\dim R = r = n - I$. Also $(\mathscr{E}\underline{y})_R = 0$ whether $H_0$ is true or not.

The vector representation of this model is then $\underline{y} = \mathscr{E}\underline{y} + \sigma\underline{u}$; $\mathscr{E}\underline{y} \in A$, with $\underline{u} \simeq \underline{\chi}_n$, the normal random vector with mutually independent coordinates $\underline{u}_i \simeq \underline{\chi}$ ($i = 1, \ldots, n$).

Note that sometimes the model is given in the form $\underline{y}_{ij} = \lambda + \alpha_i + \underline{e}_{ij}$ with the side condition $\sum_{i=1}^{I} n_i\alpha_i = 0$. The vector $\mathscr{E}\underline{y} = \mu \in A$ can be written as $\mu = \mu_L + \mu_{A^*}$ with $\mu_L = \lambda s_0$ and $\mu_{A^*} = \sum_{i=1}^{I} \alpha_i s_i$.

For the ANOVA table we decompose $y$ into orthogonal components $y = y_L + y_{A^*} + y_R$ and thus $y^2 = y_L^2 + y_{A^*}^2 + y_R^2$ with the corresponding dimension splitting $n = 1 + (I-1) + (n-I)$.

The Gauss-Markov estimator of $\mathscr{E}\underline{y} = \mu$ is $\underline{y}_A$.

The expectation of $\underline{y}_{A^*}^2$ is found as follows:

$$\underline{y}_{A^*} = \mu_{A^*} + \sigma\underline{u}_{A^*},$$

$$(\underline{y}_{A^*} - \mu_{A^*})^2 = \sigma^2\underline{u}_{A^*}^2 \simeq \sigma^2\underline{\chi}_p^2$$

with $p = \dim A^* = I - 1$.

Therefore

$$\mathscr{E}(\underline{y}_{A^*}-\mu_{A^*})^2 = \mathscr{E}\sigma^2\underline{\chi}_p^2 = p\sigma^2$$

and

$$\begin{aligned}\mathscr{E}\underline{y}_{A^*}^2 &= \mathscr{E}\left[(\underline{y}_{A^*}-\mu_{A^*})+\mu_{A^*}\right]^2 = \\ &= \mathscr{E}(\underline{y}_{A^*}-\mu_{A^*})^2+2\mathscr{E}(\underline{y}_{A^*}-\mu_{A^*})\mu_{A^*}+\mathscr{E}\mu_{A^*}^2 \\ &= \mu_{A^*}^2+p\sigma^2.\end{aligned}$$

The distribution of $\underline{y}_{A^*}^2$ is derived as

$$\begin{aligned}\underline{y}_{A^*} &= \mu_{A^*}+\sigma\underline{u}_{A^*}, \\ \underline{y}_{A^*}^2 &= (\mu_{A^*}+\sigma\underline{u}_{A^*})^2 = \sigma^2(\sigma^{-1}\mu_{A^*}+\underline{u}_{A^*})^2 \simeq \\ &\simeq \sigma^2[(\sigma^{-2}\mu_{A^*}^2)^{\frac{1}{2}}+\underline{\chi}]^2+\sigma^2\underline{\chi}_{p-1}^2 \simeq \sigma^2\underline{\chi}_p^{2,\gamma}\end{aligned}$$

a non-central chi-square variable with non-centrality parameter

$$\gamma = (\sigma^{-2}\mu_{A^*}^2)^{\frac{1}{2}} = \sigma^{-1}(\mu_{A^*}^2)^{\frac{1}{2}}$$

and $p$ degrees of freedom.

Furthermore $\underline{y}_R = \sigma\underline{u}_R$ and $\underline{y}_R^2 \simeq \sigma^2\underline{\chi}_r^2$ with $r = \dim R = n-I$. Since $A^*$ is orthogonal to $R$, $\underline{y}_{A^*}^2$ is stochastically independent of $\underline{y}_R^2$ and thus the well-known test statistic for $H_0$: $\beta_1 = \ldots = \beta_I$ or "$\mu_{A^*}$ equals the null vector" is:

$$\frac{\underline{y}_{A^*}^2/p}{\underline{y}_R^2/r} \simeq \frac{\sigma^2\underline{\chi}_p^{2,\gamma}/p}{\sigma^2\underline{\chi}_r^2/r} \simeq \underline{F}_r^{p,\gamma},$$

a non-central $\underline{F}$ variable with $p$ and $r$ degrees of freedom and non-centrality parameter $\gamma = \sigma^{-1}(\mu_{A^*}^2)^{\frac{1}{2}}$. Under $H_0$ is $\gamma = 0$ and the test-statistic is a central $\underline{F}_r^p$ variable.

### 2.2 *The two-way classification (additive model)*

The additive model for the connected two-way layout with a classification $A$ of $I$ classes and a classification $B$ with $J$ classes and $n_{ij}$ observations per subclass $A_i \cap B_j$ is:

$$\underline{y}_{ijk} = \alpha_i'+\beta_j'+\underline{e}_{ijk}$$

or

$$\underline{y}_{ijk} = \lambda+\alpha_i+\beta_j+\underline{e}_{ijk}$$

with the arbitrary side conditions

$$\textstyle\sum_i n_{ij}\alpha_i = 0, \quad \sum_j n_{ij}\beta_j = 0 \quad (i = 1,\ldots,I; \quad j = 1,\ldots,J; \quad k = 1,\ldots,n_{ij})$$

whereby $\{\underline{e}_{ijk} \simeq \sigma\underline{\chi}\}$ are mutually stochastically independent.

Let the vector spaces $A$ and $B$ correspond to the classifications $A$ and $B$ respectively. The space $L$ is the intersection of $A$ and $B$ and for any connected two-way layout the dimension of $L$ is 1. The residual space $R$ is the orthogonal complement of $D=A+B$ in $R^n$, $n=\sum_i\sum_j n_{ij}$.

The vector representation is: $\underline{y} = \mathscr{E}\underline{y}+\sigma\underline{u}$ with $\mathscr{E}\underline{y} \in D = A+B$ and $\underline{u} \simeq \underline{\chi}_n$. The Gauss-Markov estimator of $\mathscr{E}\underline{y}$ is $\underline{y}_D = \underline{y}_{A+B}$.

Under e.g. $H_{0_1}$: $\alpha_1 = \ldots = \alpha_I = 0$ the vector $\mathscr{E}\underline{y} \in D_0 = B$. Let $D^*$ be the orthogonal complement of $D_0$ in $D$. The test statistic for $H_{0_1}$ is then

$$\frac{\underline{y}^2_{D^*}/d^*}{\underline{y}^2_R/r} \simeq \underline{F}_r^{d^*,\gamma}$$

with $d^* = \dim D^*$, $\gamma = \sigma^{-1}(\mu^2_{D^*})^{\frac{1}{2}}$ and $r = \dim R$.

Because $y$ can be decomposed as $y = y_D+y_R$ with $D$ and $R$ orthogonal, $y^2 = y_D^2+y_R^2 = y^2_{A+B}+y^2_R$.

For an orthogonal design $A^*$ is orthogonal to $B^*$ (this is the case if and only if $n_{ij} = (\sum_i n_{ij})(\sum_j n_{ij})/n$), thus $y^2_{A+B} = y^2_L+y^2_{A^*}+y^2_{B^*}$ and $y^2_{D^*} = y^2_D-y^2_{D_0} = y^2_{A^*}$.

For a non-orthogonal design let $y_{A\perp B}$ be the component of $y$ in $A+B$ orthogonal to $B$, then $y^2_{A+B} = y^2_{A\perp B}+y^2_B = y^2_L+y^2_{A\perp B}+y^2_{B^*}$ and $y^2_{D^*} = y^2_{A\perp B}$.

The vector $y_{A\perp B}$ can be found by an iterative procedure see CORSTEN (1958).

In the ANOVA table the various sums of squares are given according to the decomposition $y^2 = y^2_L+y^2_{A\perp B}+y^2_{B^*}+y^2_R$ with the corresponding dimension splitting $n = 1+(I-1)+(J-1)+(n+1-I-J)$.

Analogously one can test $H_{0_2}$ : $\beta_1 = \ldots = \beta_J = 0$.

2.3 *The two-way classification (interaction model)*

The model is in this case

$$\underline{y}_{ijk} = \lambda+\alpha_i+\beta_j+\gamma_{ij}+\underline{e}_{ijk}$$

with the arbitrary side conditions

$$\sum_i n_{ij}\alpha_i = 0,\ \sum_j n_{ij}\beta_j = 0,\ \sum_i n_{ij}\gamma_{ij} = 0,\ \sum_j n_{ij}\gamma_{ij} = 0$$

$$(i = 1, \ldots, I;\ j = 1, \ldots, J;\ k = 1, \ldots, n_{ij})$$

whereby $\{\underline{e}_{ijk} \simeq \sigma\underline{\chi}\}$ are mutually stochastically independent.

We define the product-classification $A \times B$ as the classification according to the non-empty subclasses $A_i \cap B_j$. A vector in the corresponding vector space $A \times B$ has equal coordinates in each subclass $A_i \cap B_j$.

The vector representation for the interaction model is then:

$$\underline{y} = \mathscr{E}\underline{y}+\sigma\underline{u}$$

with

$$\mathscr{E}\underline{y} \in D = A \times B, \ \underline{u} \simeq \sigma\underline{\chi}_n \quad \left(n = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij}\right)$$

Under $H_0$: "$\gamma_{ij} = 0, \ i = 1, \ldots, I; \ j = 1, \ldots, J$" the vector $\mathscr{E}\underline{y} \in D_0 = A + B$. The same procedure for testing $H_0$ can be applied as described in section 2.2. For details see VERDOOREN (1967).

## 3 Random-effects ANOVA models

### 3.1 *The one-way classification*

Let the classification $A$ consist of $I$ rows, with $J$ experimental units per row. In the random-effects models we assume that there is a joint action of two random influences. Firstly there is a random variable $\underline{a}_i$ which attains a value in each row $i$. These $\underline{a}_i$ are isomorous and mutually stochastically independent with $\underline{a}_i \simeq \sigma_a\underline{\chi}$. Secondly there is another random variable $\underline{e}_{ij}$ which attains a value on each experimental unit in a row. These $\underline{e}_{ij}$ are isomorous and mutually stochastically independent with $\underline{e}_{ij} \simeq \lambda + \sigma\underline{\chi}$ and moreover independent of $\underline{a}_i$. The model is thus

$$\underline{y}_{ij} = \lambda + \underline{a}_i + \sigma\underline{u}_{ij}$$

with the $\{\underline{a}_i \simeq \sigma_a\underline{\chi}\}$ and $\{\underline{u}_{ij} \simeq \underline{\chi}\}$ completely stochastically independent, ($i = 1, \ldots, I; \ j = 1, \ldots, J$).

With the same notations as in section 2.1 the first vector representation is

$$\underline{y} = \lambda s_0 + \sum_{i=1}^{I} \underline{a}_i s_i + \sigma\underline{u}$$

with $\underline{u} \simeq \underline{\chi}_n, \ n = IJ$.

Let $\underline{v} \simeq \underline{\chi}_n$ and independent of $\underline{u}$. Observe now on the one side that the orthogonal projection of $\underline{v}$ on $A$ is

$$\underline{v}_A = \sum_{i=1}^{I} \underline{v}_{i.} s_i \text{ with } \underline{v}_{i.} = \sum_{j=1}^{J} \underline{v}_{ij}/J \text{ and } \underline{v}_{i.} \simeq J^{-\frac{1}{2}}\underline{\chi}.$$

On the other hand $\underline{a}_i \simeq \sigma_a\underline{\chi}$ and we can write therefore $\underline{a}_i \simeq \sqrt{J}\sigma_a\underline{v}_{i.}$, hence

$$\sum_{i=1}^{I} \underline{a}_i s_i \simeq \sigma_a\sqrt{J}\,\underline{v}_A.$$

The vector representation of the random model is then

$$\underline{y} \simeq \lambda s_0 + \sigma_a\sqrt{J}\,\underline{v}_A + \sigma\underline{u}$$

or

$$\underline{y} \simeq \mathscr{E}\underline{y} + \sigma_a\sqrt{J}\,\underline{v}_A + \sigma\underline{u}$$

with $\mathscr{E}\underline{y} \in L$.

Because $R^n = A + R$ with $A$ and $R$ orthogonal we can write $\underline{u} = \underline{u}_A + \underline{u}_R$ with $\underline{u}_A$ independent of $\underline{u}_R$ and

$$\underline{y} \simeq \lambda s_0 + \sigma_a \sqrt{J}\, \underline{v}_A + \sigma \underline{u}_A + \sigma \underline{u}_R.$$

Since for each component of the vectors $\underline{u}$ and $\underline{v}$ we have

$$\sigma_a \sqrt{J}\, \underline{v}_{ij} + \sigma \underline{u}_{ij} \simeq (\sigma_a^2 J + \sigma^2)^{\frac{1}{2}} \underline{\chi},$$

because $\underline{v}_{ij} \simeq \underline{\chi}$, $\underline{u}_{ij} \simeq \underline{\chi}$ and $\underline{v}_{ij}$ independent of $\underline{u}_{ij}$, is

$$\sigma_a \sqrt{J}\, \underline{v}_A + \sigma \underline{u}_A = (\sigma_a \sqrt{J}\, \underline{v} + \sigma \underline{u})_A \simeq (\sigma_a^2 J + \sigma^2)^{\frac{1}{2}} [\underline{\chi}_n]_A \simeq (\sigma_a^2 J + \sigma^2)^{\frac{1}{2}} \underline{v}_A$$

while $\underline{v} \simeq \underline{\chi}_n$.

Hence

$$\underline{y} \simeq \lambda s_0 + (\sigma_a^2 J + \sigma^2)^{\frac{1}{2}} \underline{v}_A + \sigma \underline{u}_R.$$

Furthermore $A = L + A^*$ with $L$ orthogonal to $A^*$, hence $\underline{v}_A = \underline{v}_L + \underline{v}_{A^*}$ with

$$\underline{v}_L \simeq n^{-\frac{1}{2}} \underline{\chi} s_0.$$

Finally we obtain

$$\underline{y} \simeq [\lambda + n^{-\frac{1}{2}} (\sigma_a^2 J + \sigma^2)^{\frac{1}{2}} \underline{\chi}] s_0 + (\sigma_a^2 J + \sigma^2)^{\frac{1}{2}} \underline{v}_{A^*} + \sigma \underline{u}_R.$$

We can thus decompose $\underline{y}$ into the following independent components $\underline{y} = \underline{y}_L + \underline{y}_{A^*} + \underline{y}_R$ with

$$\underline{y}_L = \bar{\underline{y}} s_0 \simeq [\lambda + n^{-\frac{1}{2}} (\sigma_a^2 J + \sigma^2)^{\frac{1}{2}} \underline{\chi}] s_0, \quad \bar{\underline{y}} = \sum_i \sum_j \underline{y}_{ij}/n;$$

$$\underline{y}_{A^*} \simeq (\sigma_a^2 J + \sigma^2)^{\frac{1}{2}} \underline{v}_{A^*}$$

and

$$\underline{y}_R \simeq \sigma \underline{u}_R.$$

In the ANOVA table we calculate the sums of squares $y_L^2$, $y_{A^*}^2$ and $y_R^2$ with the dimensions 1, $(I-1)$ and $(n-I)$ respectively.

The distribution of $\underline{y}_{A^*}^2$ is

$$(\sigma_a^2 J + \sigma^2) \underline{v}_{A^*}^2 \simeq (\sigma_a^2 J + \sigma^2) \underline{\chi}_p^2$$

with $p = \dim A^* = I - 1$. The expectation of the mean square $\underline{y}_{A^*}^2/p$ is thus $\sigma_a^2 J + \sigma^2$.

Stochastically independent of $\underline{y}_{A^*}^2$ is $\underline{y}_R^2 \simeq \sigma^2 \underline{\chi}_r^2$ with $r = \dim R = n - I$. To test the hypothesis $H_0$: $\sigma_a = 0$ we use the test statistic

$$\frac{\underline{y}^2_{A^*}/p}{\underline{y}^2_R/r} \simeq \frac{\sigma_a^2 J+\sigma^2}{\sigma^2}\underline{F}^p_r$$

and under $H_0$ this is isomorous with $\underline{F}^p_r$.

3.2 *The two-way classification*

For the two-way layout with a classification $A$ of $I$ classes and a classification $B$ with $J$ classes and an equal number $m$ of observations per subclass $A_i \cap B_j$, the random-effects model is

$$\underline{y}_{ijk} = \lambda+\underline{a}_i+\underline{b}_j+\underline{e}_{ijk},$$

whereby all random variables

$$\{\underline{a}_i \simeq \sigma_a\underline{\chi}\},\ \{\underline{b}_j \simeq \sigma_b\underline{\chi}\} \quad \text{and} \quad \{\underline{e}_{ijk} \simeq \sigma\underline{\chi}\}$$

are mutually stochastically independent, $(i = 1, \ldots, I;\ j = 1, \ldots, J;\ k = 1, \ldots, m)$.

With the same notations as in section 2.2 we can write

$$\underline{y} = \lambda s_0+\sum_{i=1}^{I}\underline{a}_i s_i+\sum_{j=1}^{J}\underline{b}_j t_j+\sigma\underline{u}$$

whereby the class-characteristic vectors $s_1, \ldots, s_I$ and $t_1, \ldots, t_J$ form a basis of $A$ and $B$ respectively.

Let $\underline{u}$, $\underline{v}$ and $\underline{w}$ be independent normal random vectors $\underline{\chi}_n$ $(n = IJm)$, then

$$\underline{y} \simeq \lambda s_0+\sigma_a\sqrt{mJ}\,\underline{v}_A+\sigma_b\sqrt{mI}\,\underline{w}_B+\sigma\underline{u},$$

$$\underline{y} \simeq \lambda s_0+\sigma_a\sqrt{mJ}(\underline{v}_L+\underline{v}_{A^*})+\sigma_b\sqrt{mI}(\underline{w}_L+\underline{w}_{B^*})+\sigma(\underline{u}_L+\underline{u}_{A^*}+\underline{u}_{B^*}+\underline{u}_R)$$

and

$$\underline{y} \simeq [\lambda+n^{-\frac{1}{2}}(\sigma_a^2 mJ+\sigma_b^2 mI+\sigma^2)^{\frac{1}{2}}\underline{\chi}]s_0+(\sigma_a^2 mJ+\sigma^2)^{\frac{1}{2}}\underline{v}_{A^*}+(\sigma_b^2 mI+\sigma^2)^{\frac{1}{2}}\underline{w}_{B^*}+\sigma\underline{u}_R.$$

Hence

$$\underline{y}_{A^*} = (\sigma_a^2 mJ+\sigma^2)^{\frac{1}{2}}\underline{v}_{A^*} \qquad \text{and}$$

$$\underline{y}^2_{A^*} \simeq (\sigma_a^2 mJ+\sigma^2)\underline{\chi}^2_p \qquad \text{with} \quad p = \dim A^* = I-1,$$

$$\underline{y}^2_{B^*} \simeq (\sigma_b^2 mI+\sigma^2)\underline{\chi}^2_q \qquad \text{with} \quad q = \dim B^* = J-1,$$

$$\underline{y}^2_R \simeq \sigma^2\underline{\chi}^2_r \qquad \text{with} \quad r = \dim R = IJm+1-I-J,$$

whereby $\underline{y}^2_{A^*}$, $\underline{y}^2_{B^*}$ and $\underline{y}^2_R$ are mutually stochastically independent since $A^*$, $B^*$ and $R$ are mutually orthogonal.

### 3.3 *Hierarchal classifications*

A twofold nested classification with $m$ observations per class has the following random model:

$$\underline{y}_{ijk} = \lambda + \underline{a}_i + \underline{b}_{ij} + \underline{e}_{ijk}$$

whereby all random variables

$$\{\underline{a}_i \simeq \sigma_a \underline{\chi}\}, \quad \{\underline{b}_{ij} \simeq \sigma_b \underline{\chi}\} \quad \text{and} \quad \{\underline{e}_{ijk} \simeq \sigma \underline{\chi}\}$$

are mutually stochastically independent, $(i = 1, \ldots, I;\ j = 1, \ldots, J;\ k = 1, \ldots, m)$.

Let $\underline{u}$, $\underline{v}$ and $\underline{w}$ be independent normal random vectors $\underline{\chi}_n$ with $n = IJm$. If the vector space $B$ corresponds to the classification $B$ and the vector space $A$ to the classification $A$ then $A \subset B$. Decompose $B$ into mutually orthogonal subspaces as $B = L + A^* + B^*$, whereby $B^*$ is the orthogonal complement of $A$ in $B$. Furthermore let the vector space $R$ be the orthogonal complement of $B$ in $R^n$.

The vector representation of this model is then:

$$\underline{y} \simeq \lambda s_0 + \sigma_a \sqrt{mJ}\, \underline{v}_A + \sigma_b \sqrt{m}\, \underline{w}_B + \sigma \underline{u}$$

or

$$\underline{y} \simeq \lambda s_0 + \sigma_a \sqrt{mJ}(\underline{v}_L + \underline{v}_{A^*}) + \sigma_b \sqrt{m}(\underline{w}_L + \underline{w}_{A^*} + \underline{w}_{B^*}) + \sigma(\underline{u}_L + \underline{u}_{A^*} + \underline{u}_{B^*} + \underline{u}_R)$$

and

$$\underline{y} \simeq [\lambda + n^{-\frac{1}{2}}(\sigma_a^2 mJ + \sigma_b^2 m + \sigma^2)^{\frac{1}{2}} \underline{\chi}] s_0 + (\sigma_a^2 mJ + \sigma_b^2 m + \sigma^2)^{\frac{1}{2}} \underline{v}_{A^*} + (\sigma_b^2 m + \sigma^2)^{\frac{1}{2}} \underline{w}_{B^*} + \sigma \underline{u}_R.$$

Thus

$$\underline{y}_{A^*}^2 \simeq (\sigma_a^2 mJ + \sigma_b^2 m + \sigma^2)\underline{\chi}_p^2 \quad \text{with} \quad p = \dim A^* = I - 1,$$

$$\underline{y}_{B^*}^2 \simeq (\sigma_b^2 m + \sigma^2)\underline{\chi}_q^2 \quad \text{with} \quad q = \dim B^* = I(J-1),$$

$$\underline{y}_R^2 \simeq \sigma^2 \underline{\chi}_r^2 \quad \text{with} \quad r = \dim R = IJ(m-1),$$

whereby $\underline{y}_{A^*}^2$, $\underline{y}_{B^*}^2$ and $\underline{y}_R^2$ are mutually stochastically independent.

## 4 Mixed-effects ANOVA models

For mixed-effects models with uncorrelated random-effects the exposed theory can readily be applied. A few examples of hierarchal classifications will be presented here. In the more complicated mixed models there appear correlated random-effects. The theory presented here should then be extended and this will be done in a subsequent publication.

4.1 *Mixed-effects hierarchal classifications*

Suppose a firm produces a material in different plants. If $\alpha_i$ denotes the effect of the $i$-th plant, then this is a fixed effect since the plants do not change in a replication of the experiment. The material is processed in batches and let $\underline{b}_{ij}$ be the batch effect. Suppose that a sample of size $m$ is taken from each of $J$ batches and let $\underline{e}_{ijk}$ be the effect of the $k$-th unit taken from the $ij$-th batch. The observations $\underline{y}_{ijk}$ on these units have the structure

$$\underline{y}_{ijk} = \lambda+\alpha_i+\underline{b}_{ij}+\underline{e}_{ijk},\ \sum_i \alpha_i = 0,\ \{\underline{b}_{ij} \simeq \sigma_b \underline{\chi}\} \text{ and } \{\underline{e}_{ijk} \simeq \sigma \underline{\chi}\}$$

are mutually stochastically independent ($i = 1, \ldots, I;\ j = 1, \ldots, J;\ k = 1, \ldots, m$).

With the same notations as in section 3.3 the vector representation is

$$\underline{y} \simeq \mathscr{E}\underline{y}+\sigma_b\sqrt{m}\underline{v}_B+\sigma\underline{u}$$

with $\mathscr{E}\underline{y} \in A$ or

$$\underline{y} \simeq \lambda s_0+\sum_{i=1}^{I} \alpha_i s_i+\sigma_b\sqrt{m}(\underline{v}_L+\underline{v}_{A^*}+\underline{v}_{B^*})+\sigma(\underline{u}_L+\underline{u}_{A^*}+\underline{u}_{B^*}+\underline{u}_R).$$

Thus

$$\underline{y} \simeq \lambda s_0+(\sigma_b^2 m+\sigma^2)^{\frac{1}{2}}\underline{v}_L+\sum_{i=1}^{I} \alpha_i s_i+(\sigma_b^2 m+\sigma^2)^{\frac{1}{2}}\underline{v}_{A^*}+(\sigma_b^2 m+\sigma^2)^{\frac{1}{2}}\underline{v}_{B^*}+\sigma\underline{u}_R.$$

We find

$$\underline{y}_{A^*} \simeq \sum_{i=1}^{I} \alpha_i s_i+(\sigma_b^2 m+\sigma^2)^{\frac{1}{2}}\underline{v}_{A^*}$$

since $s_0$, $L$, $B^*$ and $R$ are orthogonal to $A^*$ and these vector components vanish by projection on $A^*$.

Hence

$$\underline{y}^2_{A^*} \simeq (\sigma_b^2 m+\sigma^2)\underline{\chi}^{2,\gamma}_p \qquad \text{with} \quad p = \dim A^* = I-1$$

and

$$\gamma = (\sigma_b^2 m+\sigma^2)^{-\frac{1}{2}}[(\sum_i \alpha_i s_i)^2]^{\frac{1}{2}} = (\sigma_b^2 m+\sigma^2)^{-\frac{1}{2}}(mJ\sum_i \alpha_i^2)^{\frac{1}{2}};$$

$$\underline{y}_{B^*} \simeq (\sigma_b^2 m+\sigma^2)^{\frac{1}{2}}\underline{v}_{B^*}$$

and

$$\underline{y}^2_{B^*} \simeq (\sigma_b^2 m+\sigma^2)\underline{\chi}^2_q$$

with $q = \dim B^* = I(J-1)$;

$\underline{y}_R \simeq \sigma\underline{u}_R$ and $\underline{y}^2_R \simeq \sigma^2\underline{\chi}^2_r$ with $r = \dim R = IJ(m-1)$, whereby $\underline{y}^2_{A^*}$, $\underline{y}^2_{B^*}$ and $\underline{y}^2_R$ are mutually stochastically independent.

The test on $H_{0_1}$: $\alpha_1 = \ldots = \alpha_I = 0$ is equivalent to $H_{0_1}$: $\gamma = 0$. The test statistic is thus

$$\frac{\underline{y}^2_{A^*}/p}{\underline{y}^2_{B^*}/q} \simeq \underline{F}_q^{p,\gamma},$$

and under $H_{0_1}$ this is a central $\underline{F}_q^p$ variable. For the test on $H_{0_2}$: $\sigma_b = 0$ we must use

$$\frac{\underline{y}^2_{B^*}/q}{\underline{y}^2_R/r} \simeq \frac{(\sigma_b^2 m+\sigma^2)}{\sigma^2}\underline{F}_r^q.$$

4.2 *A split-plot design*

In many agricultural or industrial experiments treatments of a factor $A$ on $I$ levels are given to whole plots in $m$ blocks or replications. Other treatments of a factor $B$ on $J$ levels are given to the split-plots in the whole plots. The model for such an experiment is:

$$\underline{y}_{ijk} = \lambda+\alpha_i+\underline{\varrho}_k+\sigma_1\underline{u}_{ik}+\beta_j+\gamma_{ij}+\sigma_2\underline{v}_{ijk}$$

with

$$\sum_i \alpha_i = 0, \sum_k \underline{\varrho}_k = 0, \sum_j \beta_j = 0, \sum_i \gamma_{ij} = 0, \sum_j \gamma_{ij} = 0,$$

$$\{\underline{u}_{ik} \simeq \underline{\chi}\} \quad \text{and} \quad \{\underline{v}_{ijk} \simeq \underline{\chi}\}$$

are mutually stochastically independent ($i = 1, \ldots, I$; $j = 1, \ldots, J$; $k = 1, \ldots, m$).

Let $A$ and $B$ be the vector spaces corresponding to the classifications of treatments $A$ and $B$ respectively, $A \times B$ be the vector space corresponding to the product-classification of $A$ and $B$, $C$ be the vector space corresponding to the replications and $P$ be the vector space corresponding to the subdivision of whole plots. The vector representation is then:

$$\underline{y} \simeq \mathscr{E}\underline{y}+\sigma_1\sqrt{J}\,\underline{u}_{AC}+\sigma_2\underline{v}$$

with $\mathscr{E}\underline{y} = \mu \in A \times B + C$; $\underline{u}$ and $\underline{v}$ are independent normal random vectors $\underline{\chi}_n$ ($n = IJm$).

Proceeding as before we derive

$$(\underline{y}_{A^*}-\mu_{A^*})^2 \simeq (\sigma_1^2 J+\sigma_2^2)\underline{\chi}_p^2$$

with $p = \dim A^* = I-1$ or

$$\underline{y}^2_{A^*} \simeq (\sigma_1^2 J+\sigma_2^2)\underline{\chi}_p^{2,\gamma}$$

with $\gamma = (\sigma_1^2 J+\sigma_2^2)^{-\frac{1}{2}}(\mu^2_{A^*})^{\frac{1}{2}}$.

Let $R_1$ be the orthogonal complement of $A+C$ in $P$ then

$$\underline{y}^2_{R_1} \simeq (\sigma_1^2 J+\sigma_2^2)\underline{\chi}^2_{r_1}$$

with $r_1 = \dim R_1 = (I-1)(m-1)$.

Further

$$(\underline{y}_{B^*}-\mu_{B^*})^2 \simeq \sigma_2^2 \underline{\chi}_q^2$$

with $q = \dim B^*$.

Let $(A \times B)^*$ be the orthogonal complement of $A+B$ in $A \times B$ and $R_2$ be the orthogonal complement of $A \times B+C$ in $R^n$. Then

$$(\underline{y}_{(A \times B)^*}-\mu_{(A \times B)^*})^2 \simeq \sigma_2^2 \underline{\chi}_s^2$$

with $s = \dim (A \times B)^* = (I-1)(J-1)$, and

$$\underline{y}_{R_2}^2 \simeq \sigma_2^2 \underline{\chi}_{r_2}^2$$

with $r_2 = \dim R_2 = (m-1)I(J-1)$.

To test $H_{0_1}$: $\alpha_1 = \ldots = \alpha_I = 0$ we use the test statistic

$$\frac{\underline{y}_{A^*}^2/p}{\underline{y}_{R_1}^2/r_1} \simeq \underline{F}_{r_1}^{p,\gamma}$$

with $\gamma = (\sigma_1^2 J+\sigma_2^2)^{-\frac{1}{2}}(\mu_{A^*}^2)^{\frac{1}{2}}$

and to test $H_{0_2}$: "There is no interaction" we use the test statistic

$$\frac{\underline{y}_{(A \times B)^*}^2/s}{\underline{y}_{R_2}^2/r_2} \simeq \underline{F}_{r_2}^{s,\delta}$$

with $\delta = \sigma_2^{-1}(\mu_{(A \times B)^*}^2)^{\frac{1}{2}}$.

## 5 Final remarks

(i) For unbalanced designs the random-effects models give difficulties in hypotheses testing. The vector representation however gives an easy derivation for the expectation of the mean squares. Also the fact that the mean square of effects is not distributed as chi-square can easily be derived.

As an example we take the one-way classification with unequal observations per row. The random model is

$$\underline{y}_{ij} = \lambda+\underline{a}_i+\underline{e}_{ij}$$

with $\{\underline{a}_i \simeq \sigma_a \underline{\chi}\}$ and $\{\underline{e}_{ij} \simeq \sigma \underline{\chi}\}$ mutually stochastically independent ($i = 1, \ldots, I$; $j = 1, \ldots, n_i$).

The vector model is

$$\underline{y} = \lambda s_0+\sum_{i=1}^{I} \underline{a}_i s_i+\sigma \underline{u}$$

with $\underline{u} \simeq \underline{\chi}_n$ $(n = \sum_i n_i)$.

Let $A_i$ be the space of dimension one with basis $s_i$, then $A$ can be decomposed in mutual orthogonal subspaces as $A = A_1 + \ldots + A_I$.

Hence

$$\underset{\sim}{y} \simeq \lambda s_0 + \sum_{i=1}^{I} \sigma_a \sqrt{n_i}\, \underset{\sim}{v}_{A_i} + \sigma\left(\sum_{i=1}^{I} \underset{\sim}{u}_{A_i} + \underset{\sim}{u}_R\right)$$

or

$$\underset{\sim}{y} \simeq \lambda s_0 + \sum_{i=1}^{I} (\sigma_a^2 n_i + \sigma^2)^{\frac{1}{2}} \underset{\sim}{v}_{A_i} + \sigma \underset{\sim}{u}_R.$$

When we want to calculate $\mathcal{E}\, \underset{\sim}{y}^2_{A^*}$ we need $\mathcal{E}\, \underset{\sim}{y}^2_A$ and $\mathcal{E}\, \underset{\sim}{y}^2_L$ since $\mathcal{E}\, \underset{\sim}{y}^2_{A^*} = \mathcal{E}(\underset{\sim}{y}^2_A - \underset{\sim}{y}^2_L)$.

But $\underset{\sim}{y}^2_L$ is equal to

$$(\underset{\sim}{y} s_0)^2/(s_0 s_0) = (\underset{\sim}{y} s_0)^2/n \text{ and } \underset{\sim}{y}^2_A = [\lambda s_0 + \sum_{i=1}^{I} (\sigma_a^2 n_i + \sigma^2)^{\frac{1}{2}} \underset{\sim}{v}_{A_i}]^2.$$

Since

$$\underset{\sim}{y} s_0 \simeq \lambda n + \sum_{i=1}^{I} (\sigma_a^2 n_i + \sigma^2)^{\frac{1}{2}} \underset{\sim}{v}_{A_i} s_0$$

and

$$\underset{\sim}{v}_{A_i} s_0 = \underset{\sim}{v}_{i.} s_i (s_1 + \ldots + s_I) = s_i^2 \underset{\sim}{v}_{i.} \simeq n_i (n_i)^{-\frac{1}{2}} \underset{\sim}{\chi} \simeq n_i^{\frac{1}{2}} \underset{\sim}{\chi}$$

we find

$$\underset{\sim}{y} s_0 \simeq \lambda n + \left[\sum_{i=1}^{I} n_i (\sigma_a^2 n_i + \sigma^2)\right]^{\frac{1}{2}} \underset{\sim}{\chi}.$$

Hence

$$\mathcal{E}\, \underset{\sim}{y}^2_L = n\lambda^2 + n^{-1} \sum_{i=1}^{I} n_i (\sigma_a^2 n_i + \sigma^2).$$

Since

$$\underset{\sim}{v}_{A_i} \simeq n_i^{-\frac{1}{2}} \underset{\sim}{\chi} s_i, \quad \mathcal{E}\, \underset{\sim}{v}^2_{A_i} = 1$$

15

$$\mathcal{E}\, \underset{\sim}{y}^2_A = n\lambda^2 + \sum_{i=1}^{I} (\sigma_a^2 n_i + \sigma^2).$$

Thus

$$\mathcal{E}\, \underset{\sim}{y}^2_{A^*} = \mathcal{E}\, \underset{\sim}{y}^2_A - \mathcal{E}\, \underset{\sim}{y}^2_L = (I-1)\sigma^2 + (n - \sum_i n_i^2/n)\sigma_a^2$$

and

$$\mathcal{E}\, \underset{\sim}{y}^2_{A^*}/(I-1) = \sigma^2 + (n^2 - \sum n_i^2)\, [n(I-1)]^{-1}\, \sigma_a^2.$$

The distribution of $\underset{\sim}{y}^2_A$ is a linear combination of independant non-central $\chi^2_1$ – variables with unequal coefficients and thus $\underset{\sim}{y}^2_{A^*}$ is not distributed as a constant times a chi-square variable.

(ii) The vector representation can be used to calculate the expectation of mean squares in the ANOVA tables even when we have not normal random vectors but only random vectors with uncorrelated components, each with the same variance. The crucial point is that if $\underset{\sim}{u}_{ij} \simeq \underset{\sim}{v}_{ij}$ and $\underset{\sim}{u}_{ij}$ is uncorrelated with $\underset{\sim}{v}_{ij}$, then the variance

of a linear combination of $\underset{\cdot}{u}_{ij}$ and $\underset{\cdot}{v}_{ij}$ equals the sum of the square of the coefficients times the variance of the components. For the expectation of mean squares only this fact is necessary.

(iii) The author wishes to acknowledge the helpful comments and suggestions during the preparation of the manuscript of Professor Dr. Ir. L. C. A. CORSTEN.

## 6 References

[1] CORSTEN, L. C. A., Vectors, a tool in statistical regression theory, Mededelingen van de Landbouwhogeschool **58** (1), Mededelingen van het Instituut voor Rassenonderzoek van Landbouwgewassen no. **35**, Wageningen (1958).
[2] KUIPER, N. H., Variantie-analyse, Statistica **6**, pp. 149–194 (1952).
[3] KUIPER, N. H., Een symbool voor isomoor, Statistica Neerlandica **13**, pp. 295–317 (1959).
[4] KUIPER, N. H., Random variables and random vectors, Bulletin de l'Institut Agronomique et des Stations de Recherches de Gembloux, Hors série Vol. 1, pp. 344–355 (1960).
[5] VERDOOREN, L. R., De betekenis van en het onderzoek naar interactie, Landbouwkundig Tijdschrift **79**, pp. 159–171 (1967).

CHAPTER 4

# 4. EXACT TESTS AND CONFIDENCE INTERVALS FOR VARIANCE-RATIOS IN NESTED DESIGNS

## PART I: ONSET

### 4.1. Introduction

Let us consider a three-stage nested design. From a population I samples are taken at random. From the i-th sample, $J_i$ subsamples are taken at random. From the j-th subsample, drawn from the i-th sample, $n_{ij}$ subsubsamples are drawn at random. Let $x_{k(ij)}$ be the observation made on the k-th such subsubsample (drawn from the j-th subsample, which has been drawn from the i-th sample).

A model with random effects is the following:

$$\underline{x}_{k(ij)} = \lambda + \underline{a}_i + \underline{b}_{j(i)} + \underline{e}_{k(ji)} \qquad (4.1)$$

where $\lambda$ is a general mean; the random variables $\{\underline{a}_i\}$ are independent and identically distributed (i.i.d.), where $\underline{a}_i$ are $N(0,\sigma_a^2)$; the random variables $\underline{b}_{j(i)}$ are i.i.d. and $N(0,\sigma_b^2)$; the random variables $\underline{e}_{k(ji)}$ are i.i.d. and $N(0,\sigma_e^2)$. The random variables $\{\underline{a}_i\}$, $\{b_{j(i)}\}$ and $\{\underline{e}_{k(ji)}\}$ are mutually stochastically independent.

For a rigorous derivation of a two-stage nested design (e.g. half-sib design), see chapter 6. This derivation can easily be extended to a three-stage nested design (e.g. full-sib design).

In animal breeding such a full-sib design has already been used before 1945 by Lush and his coworkers. In the first stage I sires have been drawn at random from a population, in the second stage dams are drawn at random and each sire has been mated to a different group of $J_i$ dams. From the offspring $n_{ij}$ animals are drawn at random. Interest lies in estimating variance components and heritabilities.

In plant breeding propaganda has been made by Comstock and

Robinson (1948; 1952). From then on this design was called North Carolina mating design 1.

The ANOVA-table reads (where $J_0 = \sum_{i=1}^{I} J_i$):

| Source of variation | df | SS | MS | E(MS) |
|---|---|---|---|---|
| Correction term | 1 | SSL | | |
| Between samples | I-1 | SSA | MSA | $\sigma_e^2+K_2\sigma_b^2+K_1\sigma_a^2$ |
| Within samples, between subsamples | $J_0$-I | SSB | MSB | $\sigma_e^2 + K_3\sigma_b^2$ |
| Within subsamples = Residual, or Error | $n_{..}-J_0$ | SSR | MSR | $\sigma_e^2$ |
| Uncorrected total | $n_{..}$ | SST | | |

$$SSL = [\sum_{i=1}^{I}\sum_{j=1}^{J_i}\sum_{k=1}^{n_{ij}} x_{k(ij)}]^2/n_{..},$$

where $n_{..} = \sum_{i=1}^{I}\sum_{j=1}^{J_i} n_{ij}$;

$$SSA = \sum_{i=1}^{I}[\sum_{j=1}^{J_i}\sum_{k=1}^{n_{ij}} x_{k(ij)}]^2/n_{i.} - SSL,$$

where $n_{i.} = \sum_{k=1}^{J_i} n_{ij}$;

$$SSB = \sum_{i=1}^{I}\sum_{j=1}^{J_i}[\sum_{k=1}^{n_{ij}} x_{k(ij)}]^2/n_{ij} - SSA - SSL;$$

$$SST = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\sum_{k=1}^{n_{ij}} x_{k(ij)}^2;$$

$$SSR = SST - SSL - SSA - SSB.$$

The column of the expected values of the Mean Squares, E(MS), in the ANOVA-table is such that $E(\underline{MSR}) = \sigma_e^2$; $E(\underline{MSB}) = \sigma_e^2 + K_3\sigma_b^2$ with

$$K_3 = (J_0-I)^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J_i} n_{ij}^2(n_{ij}^{-1} - n_{i.}^{-1});\ E(MSA) = \sigma_e^2 + K_2\sigma_b^2 + K_1\sigma_a^2,$$

where $K_2 = (I-1)^{-1} \sum_{i=1}^{I} \sum_{j=1}^{J_i} n_{ij}^2 (n_{i.}^{-1} - n_{..}^{-1})$ and

$$K_1 = (I-1)^{-1} \sum_{i=1}^{I} n_{i.}^2 (n_{i.}^{-1} - n_{..}^{-1}).$$

For the (totally) balanced case: $J_i = J$ for all i and $n_{ij} = m$ for all i and j, the E(MS) column reads:

$$E(\underline{MSB}) = \sigma_e^2 + m\sigma_b^2 \text{ and}$$

$$E(\underline{MSA}) = \sigma_e^2 + m\sigma_b^2 + Jm\sigma_a^2.$$

Furthermore in the balanced case, MSR, MSB and MSA are mutually stochastically independent; MSB is distributed as $(\sigma_e^2+m\sigma_b^2)[I(J-1)]^{-1}$ times a chi-square variable with I(J-1) degrees of freedom, $\chi^2(I(J-1))$; MSA is distributed as $(\sigma_e^2 + m\sigma_b^2 + Jm\sigma_a^2)(I-1)^{-1}$ times a chi-square variable with (I-1) degrees of freedom, $\chi^2(I-1)$, and MSR is distributed as $\sigma_e^2[IJ(m-1)]^{-1}$ times a $\chi^2(IJ(m-1))$ variable.

To test the null-hypothesis $H_{02}$: $\sigma_b^2 = 0$ against the alternative hypothesis $H_{12}$: $\sigma_b^2 > 0$, or equivalently $H_{02}$: $\sigma_b^2/\sigma_e^2 = 0$ against $H_{12}$: $\sigma_b^2/\sigma_e^2 > 0$; the test statistic is MSB/MSR, which has, under $H_{02}$, an F-distribution with I(J-1) and IJ(m-1) degrees of freedom.

To test the null-hypothesis $H_{01}$: $\sigma_a^2 = 0$ against the alternative hypothesis $H_{11}$: $\sigma_a^2 > 0$, or equivalently $H_{01}$: $\sigma_a^2/\sigma_e^2 = 0$ against $H_{11}$: $\sigma_a^2/\sigma_e^2 > 0$, the test statistic is MSA/MSB, which has, under $H_{01}$, an F-distribution with (I-1) and I(J-1) degrees of freedom.

For a derivation of the above mentioned facts, see section 7.2.

For the unbalanced case we have $E(\underline{MSR}) = \sigma_e^2$; $E(\underline{MSB}) = \sigma_e^2 + K_3\sigma_b^2$ and $E(\underline{MSA}) = \sigma_e^2 + K_2\sigma_b^2 + K_1\sigma_a^2$, where $K_3$ is seldom equal to $K_2$. Under $H_{02}$: $\sigma_b^2 = 0$ we see that $E(\underline{MSB}) = E(\underline{MSR})$. Hence as the test statistic MSB/MSR should be taken, which indeed under $H_{02}$ has an F-distribution with $(J_0-I)$ and $(n_{..}-J_0)$ degrees of freedom.

Under $H_{01}$: $\sigma_a^2 = 0$ we see that $E(\underline{MSA}) = \sigma_e^2 + K_2\sigma_b^2$ and

$E(\underline{MSB}) = \sigma_e^2 + K_3\sigma_b^2$.

To find a test statistic, an approximation of Satterthwaite (1946) is often used. Let $\underline{MSh}$, h = 1, ..., H, be mutually stochastically independent Mean Squares from an ANOVA table; each $\underline{MSh}$ has $E(\underline{MSh})=\tau_h^2$ and is distributed as $\tau_h^2(\nu_h)^{-1}$ times a chi-square variable with $\nu_h$ degrees of freedom, $\chi^2(\nu_h)$.

A linear combination $\sum_{h=1}^{H} c_h(\underline{MSh})$ will be approximated by $C\nu^{-1}\chi^2(\nu)$ such that the expectation and variance are equal, hence

$$E[\sum_{h=1}^{H} c_h(\underline{MSh})] = E[C\nu^{-1}\chi^2(\nu)] \tag{4.2}$$

and

$$\text{var}[\sum_{h=1}^{H} c_h(\underline{MSh})] = \text{var}[C\nu^{-1}\chi^2(\nu)] \tag{4.3}$$

From (4.1) it follows that

$$\sum_{h=1}^{H} c_h\tau_h^2 = C \tag{4.4}$$

and from (4.2) that

$$\sum_{h=1}^{H} c_h^2\,\tau_h^4\,2\,\nu_h^{-1} = C^2\,2\,\nu^{-1} \tag{4.5}$$

From (4.4) and (4.5) Satterthwaite found

$$\nu = [\sum_{h=1}^{H} c_h\tau_h^2]^2/[\sum_{h=1}^{H} c_h^2\tau_h^4/\nu_h] \tag{4.6}$$

As an estimate for $\tau_h^2$ he used MSh, hence the degrees of freedom $\nu$ will be estimated as

$$\hat{\nu} = [\sum_{h=1}^{H} c_h(\text{MSh})]^2/[\sum_{h=1}^{H} c_h^2(\text{MSh})^2/\nu_h] \tag{4.7}$$

Tietjen and Moore (1968) used Satterthwaite's procedure to construct an approximate F-test to test $H_{01}$: $\sigma_a^2 = 0$ against $H_{11}$: $\sigma_a^2 > 0$.

They used as test statistic MSA/MSC where

MSC = $(K_2/K_3)$MSB + $(1-K_2/K_3)$MSR, hence under $H_{01}$: $\sigma_a^2 = 0$, E(MSA) = E(MSC). They then approximated the distribution of MSC as E(MSC).$\nu^{-1}$ times a chi-square variable with $\nu$ degrees of freedom, where $\nu$ is estimated by (4.7). They also provided, upon request, a computer program to perform this approximate F-test for all unbalanced nested designs!

But Kruskal (1968) pointed out that, for the unbalanced case of a completely nested design, MSA and MSB are not independent.

How to find an exact test for $H_{01}$: $\sigma_a^2 = 0$ against $H_1$: $\sigma_a^2 > 0$ is described in section 4.2.

An extended version is given in section 7.2.

Also an exact confidence interval for the ratio of variance components is derived in sections 4.2 and 4.3.

REFERENCES


Comstock, R.E. and Robinson, H.F. (1948). The components of genetic variance in populations of biparental progenies and their use in estimating the average degree of dominance. Biometrics 4, 254-266.

Comstock, R.E. and Robinson, H.F. (1952). Estimation of average dominance of genes. In Heterosis, ed. J.W. Gowen, Iowa State College Press, Ames, Iowa, 494-516.

Kruskal, W.H. (1968). Correction to Tietjen and Moore's note (in Biometrics 24, 423-429). Biometrics 24, 1025-1026.

Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. Biometrics 2, 110-114.

Tietjen, G.L. and Moore, R.H. (1968). On testing significance of components of variance in the unbalanced nested analysis of variance. Biometrics 24, 423-429.

## 4.2. Exact tests about variance ratios in unbalanced two- and three-stage nested designs

*Contributed paper at the 8th International Biometric Conference*, Constanta, Romania, August 25-30, 1974.

Delivered as hand-out to interested participants of the 8th International Biometric Conference.

Contributed paper 8th International Biometric Conference, Constanta, Romania, August 25 - 30, 1974.

# Exact tests about variance ratios in unbalanced two- and three-stage nested designs.

L.R. Verdooren
Agricultural University, Wageningen, The Netherlands.

Summary

The first part deals with the problem of testing the hypothesis $\Delta = 0$ against $\Delta > 0$ for nested designs, where $\Delta$ is the ratio of variance components. Even for the simplest random models, such as those for nested designs, exact tests of variance ratios and their power function are troublesome in the unbalanced case. For testing one needs the distribution of a linear combination of independent chi-square variables (with positive and negative coefficients). Such distributions are not tabulated. However, with an algorithm, suggested by Imhof (1961), the exact distribution can easily be calculated by means of a computer. Our numerical results with this procedure agree with that of Spjøtvoll (1968) for the case he was able to handle, viz. the unbalanced one-way lay-out with three classes.

It is shown that for partially balanced three-stage nested designs (balanced per second stage), the sum of squares "Between" and "Within Samples" of the second stage are stochastically independent under the normality assumption.

The second part deals with the construction of an exact confidence interval for $\Delta$, concerning the last two stages of an unbalanced nested design. Wald (1940) indicated an iterative procedure to obtain an exact confidence interval for $\Delta$ in the case of the unbalanced one-way lay-out (two-stage nested design), but his procedure was never used in practice. With a computer this procedure can be applied easily. A similar problem for unbalanced three-stage nested designs can be handled by a generalization of Wald's procedure.

Some examples are presented.

## 1. Notation

### 1.1. Two-stage nested designs

The model reads:

$$x_{j(i)} = \lambda + a_i + e_{j(i)},\ j = 1,\ldots,\ n_i \quad \text{for} \quad i = 1,\ \ldots,\ I,$$

where $\lambda$ is a general mean, and independently and identically distributed (i.i.d.) $a_i \sim N(0,\sigma_a^2)$, i.i.d. $e_{j(i)} \sim N(0,\sigma_e^2)$. The random variables $\{a_i\}$ and $\{e_{j(i)}\}$ are mutually statistically independent.
The design is said to be balanced if all $n_i$ are the same; otherwise it is unbalanced. The ANOVA-table reads:

| Source | df | SS | MS | Variance ratio |
|---|---|---|---|---|
| Correction term (level) | 1 | x'Lx | | |
| Between samples | I-1 | x'Ax | MSA | $f$ = MSA/MSR |
| Within samples = Residual | $n_.$-I | x'Rx | MSR | |
| Uncorrected total | $n_.$ | x'x | | |

Remark that a sum of squares (SS) is a quadratic form in the observations.

To test the null-hypothesis $H_0$: $\Delta = 0$ against the alternative $H_1$: $\Delta > 0$ ($\Delta = \sigma_a^2/\sigma_e^2$) the usual test statistic is the variance ratio $f$ = MSA/MSR.

## 1.2. Three-stage nested designs

The model reads:

$$x_{k(ij)} = \lambda + a_i + b_{j(i)} + e_{k(ji)}, \quad k = 1, \ldots, n_{ij} \text{ for } j = 1, \ldots, J_i \text{ and } i = 1, \ldots, I$$

where $\lambda$ is a general mean, and i.i.d. $a_i \sim N(0,\sigma_a^2)$, $b_{j(i)} \sim N(0,\sigma_b^2)$, $e_{k(ij)} \sim N(0,\sigma_e^2)$ respectively. The random variables $\{a_i\}$, $\{b_{j(i)}\}$ and $\{e_{k(ji)}\}$ are mutually independent.

The design is called (totally) balanced if $J_i = J$ for all i and $n_{ij} = m$ for all i and j. In the case where $n_{ij} = m_i$ for all j, the design is called partially balanced. In the case where for at least one i, $n_{ij} \neq n_{ij'}$, the design is called unbalanced.

The ANOVA-table reads (where $J_0 = \sum_{i=1}^{I} J_i$):

| Source | df | SS | MS | Variance ratio |
|---|---|---|---|---|
| Correction term (level) | 1 | x'Lx | | |
| Between samples | I-1 | x'Ax | MSA | $f_1$ = MSA/MSB |
| Within samples, between subsamples | $J_0$-I | x'Bx | MSB | $f_2$ = MSB/MSR |
| Within subsamples = Residual | $n_{..}-J_0$ | x'Rx | MSR | |
| Uncorrected total | $n_{..}$ | x'x | | |

To test the null-hypothesis $H_{02}$: $\Delta_2 = 0$ against $H_{12}$: $\Delta_2 > 0$ ($\Delta_2 = \sigma_b/\sigma_e$) the usual test statistic is the variance ratio $f_2$ = MSB/MSR and for the test of $H_{01}$ : $\Delta_1 = 0$ against $H_{11}$: $\Delta_1 > 0$ ($\Delta_1 = \sigma_a^2/\sigma_e^2$) the usual test statistic is the variance ratio: $f_1$ = MSA/MSB.

## 2. Distribution of a variance ratio (normal variables)

### 2.1. General theory

A variance ratio is the ratio of two sums of squares. This is a special case of the ratio of two quadratic forms x'Ax and x'Bx in normal variables, A and B being symmetric. Knowing the distribution of a quadratic form in normal variables we can derive the distribution of a ratio of two such quadratic forms.

Let $x = (x_1,\ldots,x_n)'$ be a normal random vector with mean vector $\mu = (\mu_1,\ldots,\mu_n)'$, non-singular covariance matrix $V = E(x-\mu)(x-\mu)'$, and consider the quadratic form $Q = x'Ax$, with rank $A = k$. It can be shown that Q is distributed as $\sum_{j=1}^{m} \lambda_j \chi^2_{h_j}(\delta_j^2)$ where $\lambda_j$ are the distinct non-zero characteristic roots of AV, the $h_j$ the respective multiplicities with $\sum_{j=1}^{m} h_j = k$, the $\delta_j$ certain linear combinations of $\mu_1, \ldots, \mu_n$ and the $\chi^2_{h_j}(\delta_j^2)$ are independent $\chi^2$-variables with $h_j$ degrees of freedom and non-centrality parameter $\delta_j^2$.

Imhof (1961) discovered that the distribution function of Q can be obtained quite easily by straightforward numerical integration of the characteristic function of Q.

A computer program in FORTRAN IV was given by Koerts and Abrahamse (1969), pp.159-160. On a CDC-3200 computer we applied this program to the examples given by Imhof (1961) and our results agreed with his results. With this program fast numerical calculation of the distribution function of a linear combination of independent (non-) central chi-square variables with positive and negative coefficients is possible.

For the case of an ANOVA variance ratio $f$ = MSA/MSB = $cx'Ax/x'Bx$, with

$c = p/k$, where $k$ = rank A and $p$ = rank B, the sum of squares $x'Ax$ and $x'Bx$ are independent and such that $A\mu = B\mu = 0$ while the eigenvalues of $AV$ and $BV$ are non-negative. Hence the distribution of $x'Ax$ is $\sum_{i=1}^{k} \lambda_i S_i^2$ where $S_i^2$ is a $\chi_1^2$-variable and the $S_i^2$ are independent, and $\lambda_i$ are the positive eigenvalues of $AV$, $i = 1,\ldots,k$ = rank A. Analogously the distribution of $x'Bx$ is $\sum_{j=1}^{p} \lambda_j' S_j^2$ with $p$ = rank B. Because $Pr(x'Bx = 0) = Pr(\sum_{j=1}^{p} \lambda_j' S_j^2 = 0) = 0$, the distribution of $f$ is as follows:

$$Pr(f \leq t') = Pr(\frac{x'Ax}{x'Bx} \leq t) = Pr(\sum_{i=1}^{k} \lambda_i S_i^2 - \sum_{j=1}^{p} t\lambda_j' S_j^2 \leq 0)$$ and with Imhof's algorithm we can calculate the distribution for each $t > 0$, $t = kt'/p$.

## 2.2. Application to nested designs

Let V be the variance-covariance matrix of the vector of observations. For the unbalanced two-stage nested designs, the sum of squares $x'Ax$ and $x'Rx$ are stochastically independent, because $AVR = 0$. For the test on $\Delta = \sigma_a^2/\sigma_e^2 = 0$ the distribution of $f$ = MSA/MSR is needed. Now $Pr(f \leq t') = Pr(x'Ax/x'Rx) \leq t) = Pr(\sum_{i=1}^{I-1} \lambda_i S_i^2 - t\chi^2_{n_\cdot - I} \leq 0)$, where the random part consists of a linear combination of independent chi-square variables, $S_i^2$ is a $\chi_1^2$-variable, $t = (I-1)t'/(n_\cdot - I)$ and $\lambda_i$ are the positive eigenvalues of $AV/\sigma_e^2$. Under $H_0$: $\Delta = 0$ the distribution of $f$ is an F-distribution with $I-1$ and $n_\cdot - I$ degrees of freedom.

For the balanced three-stage nested designs the sums of squares $x'Ax$ and $x'Rx$ are stochastically independent and also $x'Bx$ and $x'Rx$ are independent.

We have proved the theorem: In a three-stage nested design the sums of squares x'Ax and x'Bx are stochastically independent if and only if for each $i = 1,\ldots,I$: $n_{ij} = n_{ij'}$ $(j, j' = 1,\ldots,J_i)$. Thus only for the partially balanced designs these sums of squares are independent. For the unbalanced designs x'Ax and x'Bx are dependent.

For the test on $\Delta_2 = \sigma_b^2/\sigma_e^2 = 0$ the distribution of $f_2$ = MSB/MSR is needed.

Now $Pr(f_2 \leq t_2') = Pr(x'Bx/x'Rx \leq t_2) = Pr(\sum_{i=1}^{J_0-I} \lambda_i S_i^2 - t_2 \chi^2_{n_{..}-J_0} \leq 0)$, where the random part consists of a linear combination of independent chi-square variables with $S_i^2$ a $\chi_1^2$-variable, $t_2 = (J_0 - I)t_2'/(n_{..} - J_0)$, and $\lambda_i$ are the positive eigenvalues of $BV/\sigma_e^2$.

For the test on $\Delta_1 = \sigma_a^2/\sigma_e^2 = 0$ the distribution of $f_1$ = MSA/MSB is found as

$Pr(f_1 \leq t_1') = Pr(x'Ax/x'Bx \leq t_1) = Pr(x'(A-t_1B)x \leq 0) = Pr(\sum_{i=1}^{k} \lambda_i S_i^2 \leq 0)$ where $t_1 = (I-1)t_1' / (J_0-I)$, $S_i^2$ are independent $\chi_1^2$-variables and $\lambda_i$ are the non-zero eigenvalues of $(A-t_1B)V/\sigma_e^2$. However, even under $H_0$: $\Delta_1 = 0$ the distribution of $f_1$ depends on the value of $\Delta_2$.

Remark 1: The procedure of Satterthwaite (1946) to approximate the distribution of the variance ratio by an F-distribution, can not be used when the numerator and denominator are dependent. Also for the unbalanced three-stage nested design the distribution of $f_1$ can not be approximated with Satterthwaite's procedure. This has been overlooked by Tietjen and Moore (1968).

Remark 2: The problem of testing the hypothesis $\Delta \leq \Delta_0$ against $\Delta > \Delta_0$, where $\Delta = \sigma_a^2/\sigma_e^2$, for the unbalanced two-stage nested design has been treated by Spjøtvoll (1967). Spjøtvoll derives the most powerful invariant test for $H_0$:

$\Delta = \Delta_0$ against an alternative $\Delta = \Delta_1 > \Delta_0$.
His test statistic reads:

$$W(\Delta_0,\Delta_1) = \frac{\sum_{i=1}^{I} g_{i0}(\bar{x}_i - \bar{\bar{x}}_0)^2 - \sum_{i=1}^{I} g_{i1}(\bar{x}_i - \bar{\bar{x}}_1)^2}{\sum_{i=1}^{I} g_{i0}(\bar{x}_i - \bar{\bar{x}}_0)^2 + \sum_{i=1}^{I}\sum_{j=1}^{n_i} (x_{j(i)} - \bar{x}_i)^2}$$

where $g_{ik} = n_i/(\Delta_k n_i + 1)$, $\bar{\bar{x}}_k = \sum_{i=1}^{I} g_{ik}\bar{x}_i \,/\, \sum_{i=1}^{I} g_{ik}$ for $k = 0, 1$ and
$\bar{x}_i = \sum_{j=1}^{n_i} x_{j(i)}/n_i$.

For the case $\Delta \to \infty$ (test against large alternatives $\Delta_1$) he used the test statistic:

$$T(\Delta_0) = \frac{\sum_{i=1}^{I} g_{i0}(\bar{x}_i - \bar{\bar{x}}_0)^2}{\sum_{i=1}^{I}\sum_{j=1}^{n_i} (x_{j(i)} - \bar{x}_i)^2}$$

which is the one used by Scheffé (1959) to test the hypothesis $\Delta_0 = 0$.

The distributions of W and T consist of linear combinations of independent $\chi^2$-variables with positive and negative coefficients.

Spjøtvoll (1968) gives several examples of the exact power function calculations of the test statistics W and T for the hypothesis $\Delta = 0$ against $\Delta = 0.1$ at the 1% level for the special case he was able to handle, viz. the unbalanced one-way lay-out with three classes. We have checked these examples and our results agree with his results. However, we are now able to calculate the exact power function of unbalanced one-way lay-out with more than three classes.

Conclusion: The problem of exact tests concerning variance ratios and their power functions, for the unbalanced two- and three-stage nested designs, has been solved in practice. Copies of the program are available from the author upon request.

## 3. An exact confidence interval for Δ

### 3.1. Two-stage nested designs

Wald (1940) and also Spjøtvoll (1967) indicated a procedure for an exact confidence interval for $\Delta = \sigma_a^2/\sigma_e^2$ with confidence coefficient $1 - \alpha$.

For the computation we need the expression of $T(\Delta_0)$, see remark 2 in section 2.2. The lower confidence limit $\Delta_L$ of $\Delta$ is given by the root of the equation in $\Delta$, $(n_{.}-I)T/(I-1) = f_2$, where $f_2$ is the upper $\alpha/2$-point of the F-distribution with I-1 and $n_{.}-I$ degrees of freedom. The upper confidence coefficient $\Delta_U$ of $\Delta$ is given by the root of the equation in $\Delta$: $(n_{.}-I)T/(I-1) = f_1$, where $f_1$ is the lower $\alpha/2$-point of the F-distribution with I-1 and $n_{.}-I$ degrees of freedom.

The construction of this exact confidence interval for $\Delta$ can be solved only iteratively, and perhaps for that reason it was not used in practice. With a computer, however, this procedure can readily be applied as we did with a FORTRAN IV program.

### 3.2. Three-stage nested designs

A generalization to the unbalanced three-stage nested design will be given here. Again Wald's procedure can be used for an exact confidence interval of the variance ratio $\Delta_2 = \sigma_b^2/\sigma_e^2$ for the last two stages.

Let $f_1$ and $f_2$ be the lower and upper $\alpha/2$-point of the F-distribution with $J_0 - I$ and $n_{..} - J_0$ degrees of freedom.

$$\text{Define } T(\Delta_0) = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J_i} g_{ij}(\bar{x}_{ij} - \bar{\bar{x}}_{i.})^2}{\sum_{i=1}^{I}\sum_{j=1}^{J_i}\sum_{k=1}^{n_{ij}} (x_{k(ij)} - \bar{x}_{ij})^2} \quad \text{with } g_{ij} = n_{ij}/(\Delta_0 n_{ij}+1);$$

$\bar{x}_{ij} = \sum_{k=1}^{n_{ij}} x_{k(ij)}/n_{ij}$ and $\bar{\bar{x}}_{i.} = \sum_{j=1}^{J_i} g_{ij}\bar{x}_{ij} \Big/ \sum_{j=1}^{J_i} g_{ij}$. The lower confidence limit $\Delta_L$ of $\Delta$ is given by the root of the equation in $\Delta$: $(n_{..} - J_0)T/(J_0 - I) = f_2$ and the upper confidence limit $\Delta_U$ of $\Delta$ is given by the root of the equation in $\Delta$: $(n_{..}-J_0)T/(J_0-I) = f_1$.

The generalization to the unbalanced k-stage nested design in order to obtain an exact confidence interval for the variance ratio $\Delta$ of the last two stages, is straightforward.

Conclusion: The problem of exact confidence intervals for $\Delta$, concerning the last two stages of an unbalanced nested design, has been solved in practice. Copies of the program are available from the author upon request.

## 4. Numerical examples

### 4.1. A two-stage nested design

The first sample has been taken from Snedecor and Cochran (1967), example 10.18.1, p. 290. In research on artificial insemination of cows, a series of semen samples from bulls is sent out and tested for its ability to produce conception. The percentages of conceptions from samples for six bulls are shown below.

| Bull (i) | $x_{j(i)}$ | $n_i$ | $\bar{x}_i$ |
|---|---|---|---|
| 1 | 46, 31, 37, 62, 30 | 5 | 41.20 |
| 2 | 70, 59 | 2 | 64.50 |
| 3 | 52, 44, 57, 40, 67, 64, 70 | 7 | 56.29 |
| 4 | 47, 21, 70, 46, 14 | 5 | 39.60 |
| 5 | 42, 64, 50, 69, 77, 81, 87 | 7 | 67.14 |
| 6 | 35, 68, 59, 38, 57, 76, 57, 29, 60 | 9 | 53.22 |
| Total | 1876 | 35 | |

ANOVA-table

| Source | df | SS | MS | E(MS) | Variance ratio |
|---|---|---|---|---|---|
| Correction term | 1 | 100553.60 | | | |
| Between bulls | 5 | 3322.06 | 664.41 | $\sigma_e^2 + 5.669\sigma_a^2$ | 2.68 |
| Within bulls | 29 | 7200.34 | 248.29 | $\sigma_e^2$ | |
| Uncorrected total | 35 | 111076.00 | | | |

$\hat{\sigma}_e^2 = 248.29$; $\hat{\sigma}_a^2 = 73.40$.

Test the hypothesis $H_0$: $\Delta = 0$ against $\Delta = \Delta_1 > 0$ with $\Delta = \sigma_a^2/\sigma_e^2$.

Let $\gamma_1$ be the critical level for the W test statistic, $\gamma_1 = \Pr(W(0,\Delta_1) > W)$, and $\beta_1$ be the corresponding power when $\gamma_1$ is taken as the significance level of the test. Let $\gamma_0$ be the critical level for the T test statistic, $\gamma_0 = \Pr(T(0) > T)$, and $\beta_0$ be the corresponding power when $\gamma_0$ is taken as the significance level of the test. Because for this example $\Pr(F(5,29) > 2.68) = 0.042$, $\gamma_0$ is taken as 0.042 throughout the calculation of $\beta_0$.

| $\Delta_1$ | $\gamma_1$ | $\beta_1$ | $\beta_0$ |
|---|---|---|---|
| 0 | .042 | .042 | .042 = $\gamma_0$ |
| .02 | .052 | .078 | .061 |
| .04 | .050 | .102 | .084 |
| .06 | .048 | .129 | .109 |
| .08 | .047 | .156 | .136 |
| .1 | .045 | .184 | .165 |
| .2 | .042 | .324 | .309 |
| .4 | .040 | .545 | .539 |
| .6 | .040 | .685 | .684 |
| .8 | .040 | .775 | .775 |
| 1.0 | .040 | .833 | .834 |
| 1.2 | .040 | .872 | .873 |
| 1.4 | .040 | .900 | .901 |
| 1.6 | .040 | .920 | .921 |
| 1.8 | .040 | .935 | .936 |
| 2 | .040 | .946 | .947 |
| 3 | .041 | .976 | .976 |
| 4 | .041 | .987 | .987 |
| 5 | .041 | .992 | .992 |

The exact confidence interval $\Delta_L < \Delta < \Delta_U$, for $\Delta$ with confidence coefficient $1-\alpha$ is:

| $1-\alpha$ | $\Delta_L$ | $\Delta_U$ |
|---|---|---|
| .70 | .09 | 1.13 |
| .80 | .05 | 1.46 |
| .90 | .01 | 2.16 |
| .95 | .0*) | 3.08 |
| .99 | .0*) | 5.00 |

*) On the interval $0 \leq \Delta < \infty$ there is no solution for the equation in $\Delta$:

$$\sum_{i=1}^{I} g_i(\bar{x}_i - \bar{\bar{x}})^2 = \frac{I-1}{n-I} f_2 \sum_{i+}^{I} \sum_{j}^{n_i} (x_{j(i)} - \bar{x}_i)^2$$

When we allow negative values of $\Delta$ with the restriction that the weights $g_i = n_i/(\Delta n_i + 1)$ remain positive, we find for $\Delta_L$: -0.02 for $\alpha = 0.05$ and -0.06 for $\alpha = 0.01$.

4.2. A three-stage nested design

The second example consists of fictitious data for a three-stage nested design.

| | $A_1$ | | $A_2$ | | $A_3$ | | |
|---|---|---|---|---|---|---|---|
| | $B_{1(1)}$ | $B_{2(1)}$ | $B_{1(2)}$ | $B_{2(2)}$ | $B_{1(3)}$ | $B_{2(3)}$ | $B_{3(3)}$ |
| | 7 | 7 | 4 | 6 | 4 | 5 | 4 |
| | 5 | 6 | 4 | 5 | 6 | 3 | 4 |
| | 7 | 6 | | 7 | 5 | 5 | 3 |
| | | 7 | | 6 | 5 | | |
| Total | 19 | 26 | 8 | 24 | 20 | 13 | 11 |
| $n_{ij}$ | 3 | 4 | 2 | 4 | 4 | 3 | 3 |
| $\bar{x}_{ij}$ | 6.33 | 6.5 | 4.0 | 6.0 | 5.0 | 4.33 | 3.67 |
| $n_i$ | 7 | | 6 | | 10 | | |
| $\bar{x}_{i.}$ | 6.43 | | 5.33 | | 4.4 | | |

ANOVA-table

| Source | df | SS | MS | E(MS) |
|---|---|---|---|---|
| Correction term | 1 | 636.565 | | |
| Between $A_i$ | 2 | 16.987 | 8.494 | $\sigma_e^2+3.435\sigma_b^2+7.478\sigma_a^2$ |
| Between $B_{j(i)}$ | 4 | 8.448 | 2.112 | $\sigma_e^2+3.174\sigma_b^2$ |
| Within $B_{j(i)}$ = Residual | 16 | 11.000 | 0.688 | $\sigma_e^2$ |
| Uncorrected total | 23 | 673 | | |

$\hat{\sigma}_e^2 = 0.688$; $\hat{\sigma}_b^2 = 0.449$; $\hat{\sigma}_a^2 = 0.838$.

Test on $H_{02}$: $\Delta_2 = 0$ against $H_{12}$: $\Delta_2 > 0$ with $\Delta_2 = \sigma_b^2/\sigma_e^2$, $f_2$ = MSB/MSR = 2.112/0.688 = 3.072; the critical level $\gamma_2 = Pr(f_2 > 3.072 \mid \Delta_2 = 0) = 0.047$. The power of this test for $\Delta_2$ with significance level $\gamma_2 = 0.047$ is $\beta_2(\Delta_2) = Pr(f_2 > 3.072 \mid \Delta_2)$. Some values for $\beta_2(\Delta_2)$ are given below.

| $\Delta_2$ | $\beta_2(\Delta_2)$ |
|---|---|
| 0 | .047 = $\gamma_2$ |
| 0.1 | .100 |
| 0.5 | .353 |
| 1 | .580 |
| 1.5 | .712 |
| 2 | .792 |
| 5 | .944 |
| 10 | .983 |
| 1000 | 1.000 |

Test on $H_{01}$: $\Delta_1 = 0$ against $H_{11}$: $\Delta_1 > 0$ with $\Delta_1 = \sigma_a^2/\sigma_e^2$, given the value of $\Delta_2$,

requires $f_1$ = MSA/MSB = 8.494/2.112 = 4.022. The critical level $\gamma_1(\Delta_2)$ = $Pr(f_1 > 4.022 \mid \Delta_1 = 0, \Delta_2)$, and the power of this test for a certain value $\Delta_2$ with significance level $\gamma_1(\Delta_2)$ is $\beta_1(\Delta_1|\Delta_2) = Pr(f_1 > 4.022 \mid \Delta_1, \Delta_2)$. Some values for $\gamma_1(\Delta_2) = \beta_1(\Delta_1 = 0, \Delta_2)$ and $\beta_1(\Delta_1, \Delta_2)$ are given below.

| $\Delta_1$ \ $\Delta_2$ | 0 | .1 | .5 | 1 | 1.5 | 2 | 5 | 10 | 1000 |
|---|---|---|---|---|---|---|---|---|---|
| 0 | .111 | .114 | .117 | .119 | .119 | .119 | .120 | .120 | .120 |
| .1 | .217 | .195 | .160 | .145 | .138 | .135 | .127 | .124 | .120 |
| .5 | .439 | .432 | .307 | .244 | .212 | .193 | .153 | .137 | .121 |
| 1 | .653 | .591 | .439 | .345 | .293 | .260 | .185 | .154 | .121 |
| 1.5 | .737 | .681 | .530 | .424 | .360 | .318 | .215 | .171 | .121 |
| 2 | .788 | .739 | .597 | .487 | .417 | .369 | .244 | .187 | .121 |
| 5 | .903 | .876 | .783 | .693 | .624 | .569 | .386 | .276 | .122 |

Using Satterthwaite's procedure (which is incorrect, because MSA and MSB are dependent and MSA and MSB are not a multiple of a $\chi^2$-variable, in this example), we need a linear combination MSC with E(MSC) = E(MSA). From the E(MS) column in the ANOVA-table we find MSC = (1-3.435/3.174)MSR + (3.435/3.174)MSB = -0.057 + 2.286 = 2.229, with corresponding
df = $(2.229)^2/\{(-0.057)^2/16 + (2.286)^2/4\}$ = 3.80. Now Pr(MSA/MSC > 8.494/2.229) $\approx$ Pr(F(2,3.8) > 3.81) = 0.124; Pr(F(2,4) > 3.81) = 0.118.
An exact confidence interval $\Delta_L < \Delta_2 < \Delta_U$ for $\Delta_2$ with confidence coefficient $1-\alpha$ is:

| $1-\alpha$ | $\Delta_L$ | $\Delta_U$ |
|---|---|---|
| .70 | .194 | 2.80 |
| .80 | .107 | 3.73 |
| .90 | .007 | 5.82 |
| .95 | 0*) | 8.76 |
| .99 | 0*) | 21.15 |

*) On the interval $0 \leq \Delta_2 < \infty$ there is no solution for the equation in $\Delta$:

$$\sum_i \sum_j g_{ij}(\bar{x}_{ij} - \bar{\bar{x}}_{i.})^2 = \frac{J_0 - I}{n_{..} - J_0} f_2 \sum_i \sum_j \sum_k (x_{k(ij)} - \bar{x}_{ij})^2.$$

When we allow negative values of $\Delta_2$, with the restriction that the weights $g_{ij} = n_{ij}/(\Delta_2 n_{ij} + 1)$ remain positive, we find for $\Delta_L$: -0.059 for $\alpha = 0.05$ and -0.151 for $\alpha = 0.01$ respectively.

References

IMHOF, J.P. (1961). Computing the distribution of quadratic forms in normal variables. Biometrika 48, 419-426. Correction, Biometrika 49 (1962), 284.

KOERTS, J. and ABRAHAMSE, A.P.J. (1969). On the theory and application of the general linear model. Rotterdam University Press, Rotterdam.

SATTERTHWAITE, F.E. (1946). An approximate distribution of estimates of variance components. Biometrics 2, 110-114.

SCHEFFÉ, H. (1959). The analysis of variance. John Wiley and Sons, New York.

SNEDECOR, G.W., and COCHRAN, W.G. (1967). Statistical Methods (6th ed.). Iowa State University Press, Ames, Iowa.

SPJØTVOLL, E. (1967). Optimum invariant tests in unbalanced variance component models. Ann. Math. Stat. 38, 422-428.

SPJØTVOLL, E. (1968). Most powerful tests for some non-exponential families. Ann. Math. Stat. 39, 772-784.

TIETJEN, G.L. and MOORE, R.H. (1968). On testing significance of components of variance in the unbalanced nested analysis of variance. Biometrics 24, 423-429. Correction by W.H. Kruskal, Biometrics 24, 1025-1026.

WALD, A. (1940). A note on the analysis of variance with unequal class frequencies. Ann. Math. Stat. 11, 96-100.

Address
L.R. Verdooren
Department of Mathematics
Agricultural University
De Dreijen 8
Wageningen, NL 6140
The Netherlands.

## 4.3. Exact confidence interval for the variance ratio in nested designs

Paper presented at the second International Symposium of Computational Statistics (COMPSTAT) at West-Berlin, in 1976.

*Published in:*

COMPSTAT 1976.

Proceedings in Computational Statistics, 2nd Symposium held in Berlin (West) 1976.

Edited by Johannes Gordesch and Peter Naeve, Physica-Verlag, Wien, pages 221-228.

# Exact Confidence Interval for the Variance Ratio in Nested Designs

**L. R. Verdooren, Wageningen**

Summary

This paper deals with the construction of an exact confidence interval for the ratio $\Delta$ of variance components, concerning the last two stages of an unbalanced nested design. Wald (1940) indicated an iterative procedure to obtain an exact confidence interval for $\Delta$ in the case of the unbalanced one-way lay-out (two stage nested design), but the application of his procedure had to wait for computing facilities. With a computer this procedure can be applied easily. A similar problem for unbalanced three stage nested designs can be handled by a generalization of Wald's procedure.

Some examples are presented.

KEYWORDS: Variance components; variance ratio; nested designs

## 1. Two stage nested designs

We are concerned with data which are sampled from a population in two stages. In the first stage random samples $A_i$ ($i=1,\ldots,I$) are drawn from the population. In the second stage we draw from each sample $A_i$ a random subsample of size $n_i$ ($i=1,\ldots,I$). The observation on each element $j$ of the subsample taken from sample $A_i$ will be denoted as $x_{j.i}$.

The model reads:

$$x_{j.i} = \lambda + a_i + e_{j.i},\ i = 1,\ldots,I,\ j = 1,\ldots,n_i$$

where $\lambda$ is a general mean, identically and independently distributed (i.i.d.) $a_i = N(0,\sigma_a^2)$, i.i.d. $e_{j.i}$ $N(0,\sigma_e^2)$. The random variables $\{a_i\}$ and $\{e_{j.i}\}$ are mutually stochastically independent. A design is said to be balanced if all $n_i$ are the same; otherwise it is unbalanced. The ANOVA-table reads (where $n = \Sigma_{i=1}^{I} n_i$):

| Source | df | SS | MS | |
|---|---|---|---|---|
| Correction term (level) | 1 | x'Lx | | |
| Between samples | I - 1 | x'Ax | MSA | $\mathcal{F}$ = MSA/MSR |
| Within samples = Residual | n - I | x'Rx | MSR | |
| Total | n | x'x | | |

Note that the sum of squares (SS) is a quadratic form in the vector x of observations. Often the interest of inference lies in the ratio of the variance components $\sigma_a^2$ and $\sigma_e^2$; $\Delta = \sigma_a^2/\sigma_e^2$. To test the null-hypothesis $H_0$: $\Delta = 0$ against the alternative $H_1$: $\Delta > 0$, the usual test statistic is $\mathcal{F} = MSA/MSR$.

Let V be the variance-covariance matrix of the vector x of observations.

For the unbalanced two stage nested designs, the sum of squares x'Ax and x'Rx are stochastically independent, while AVR = 0. At $H_0$ $\mathcal{F}$ has an F distribution with I - 1 and n - I degrees of freedom.

The problem of testing the hypothesis $\Delta \leq \Delta_0$ against $\Delta > \Delta_0$ for the unbalanced two stage nested design has been treated by Spjøtvoll (1967). Spjøtvoll derives the most powerful invariant test for $H_0$: $\Delta = \Delta_0$ against an alternative $\Delta = \Delta_1 > \Delta_0$.
His test statistic reads:

$$W(\Delta_0,\Delta_1) = \frac{\Sigma_{i=1}^{I} g_{i0}(\bar{x}_i - \tilde{x}_0)^2 - \Sigma_{i=1}^{I} g_{i1}(\bar{x}_i - \tilde{x}_1)^2}{\Sigma_{i=1}^{I} g_{i0}(\bar{x}_i - \tilde{x}_0)^2 + \Sigma_{i=1}^{I}\Sigma_{j=1}^{n_i}(x_{j.i} - \bar{x}_i)^2}$$

where $g_{ik} = n_i/(\Delta_k n_i + 1)$, $\tilde{x}_k = \Sigma_{i=1}^{I} g_{ik}\bar{x}_i/\Sigma_{i=1}^{I} g_{ik}$ for k = 0, 1 and $\bar{x}_i = \Sigma_{j=1}^{n_i} x_{j.i}/n_i$.

For the case $\Delta_1 \to \infty$ (test against large alternatives $\Delta$) the limiting form of the test statistic W becomes, say, W'. Rejecting $H_0$ when W' > constant is equivalent with rejecting $H_0$ when $T(\Delta_0)$ > constant, where

$T(\Delta_0) = \dfrac{\Sigma_{i=1}^{I} g_{i0}(\bar{x}_i - \tilde{x}_0)^2}{\Sigma_{i=1}^{I}\Sigma_{j=1}^{n_i}(x_{j.i} - \bar{x}_i)^2}$. This statistic $T(\Delta_0)$ was also used by Scheffé (1959) to test the hypothesis $\Delta_0 = 0$. When $\Delta$ is the true variance ratio, $(n - I)T(\Delta)/(I - 1)$ has an F-distribution with (I - 1) and (n - I) degrees of freedom.

Spjøtvoll (1968) gives several examples of the exact power function of the tests based on W and T when testing the hypothesis $\Delta = 0$ against $\Delta = 0.1$ at the 1% level for the special case of the unbalanced one-way lay-out with three classes. For the calculation of the exact power function of unbalanced one-way lay-out with other values of $\Delta$ and with more than three classes see Verdooren (1974).

Wald (1940) and also Spjøtvoll (1967) indicated a procedure for an exact confidence interval for $\Delta = \sigma_a^2/\sigma_e^2$ with prescribed confidence. For the computation we need the expression $T(\Delta)$. An exact $(1 - \alpha)$-confidence interval for the variance ratio consists of all the values of $\Delta$ which are not rejected with the test statistic T at the significance level $\alpha$.

The lower confidence limit $\Delta_L$ of $\Delta$ is given by the root of the equation in $\Delta$, $(n - I)T(\Delta)/(I - 1) = f_2$, where $f_2$ is the upper $\alpha/2$-point of the F-distribution with $I - 1$ and $n - I$ degrees of freedom. The upper confidence limit $\Delta_U$ of $\Delta$ is given by the root of the equation in $\Delta$: $(n - I)T(\Delta)/(I - 1) = f_1$, where $f_1$ is the lower $\alpha/2$-point of the F-distribution with $I - 1$ and $n - I$ degrees of freedom.

The calculation of this exact confidence interval for $\Delta$ can be done only iteratively, and perhaps therefore it was not used in practice. With a computer, however, this procedure can readily be applied. Copies of the FORTRAN IV program list are available from the author upon request.

2. Three stage nested designs.

Now we are concerned with data which are sampled from a population in three stages. In the first stage random samples $A_i$ $(i = 1, \ldots, I)$ are drawn. In the second stage from a sample $A_i$ are drawn subsamples $B_{j.i}$ $(j = 1, \ldots, J_i)$. In the third stage from $B_{j.i}$ a sub-subsample of $n_{ij}$ elements is drawn. Let the observation made on the element k of the subsample $B_{j.i}$ from $A_i$ be denoted by $x_{k.ij}$.

The model reads:

$$x_{k.ij} = \lambda + a_i + b_{j.i} + e_{k.ij} \qquad \begin{matrix} i = 1,\ldots,I \\ j = 1,\ldots,J_i \\ k = 1,\ldots,n_{ij} \end{matrix}$$

where $\lambda$ is a general mean, and i.i.d.

$a_i = N(0,\sigma_a^2)$, i.i.d. $b_{j.i} = N(0,\sigma_b^2)$, i.i.d. $e_{k.ij} = N(0,\sigma_e^2)$ respectively. The random variables $\{a_i\}$, $\{b_{j.i}\}$ and $\{e_{k.ij}\}$ are mutually independent. The design is called balanced if $J_i = J$ for all i and $n_{ij} = m$ for all i and j. In the case where $n_{ij} = m_i$ for all i, a design is called partially balanced. In the case where $n_{ij} \neq n_{ij'}$ for at least one i, the design is called unbalanced.

The ANOVA-table reads (where $J_0 = \Sigma_{i=1}^{I} J_i$ and $n = \sum_{i=1}^{I} \sum_{j=1}^{J_i} n_{ij}$):

| Source | df | SS | MS | |
|---|---|---|---|---|
| Correction term (level) | 1 | x'Lx | | |
| Between samples | I - 1 | x'Ax | MSA | |
| Within samples, between subsamples | $J_0$ - I | x'Bx | MSB | $\tilde{\mathcal{F}} = MSB/MSR$ |
| Within subsamples = Residual | n - $J_0$ | x'Rx | MSR | |
| Total | n | x'x | | |

Often the interest of inference lies in the ratio of the variance components $\sigma_b^2$ and $\sigma_e^2$: $\Delta_2 = \sigma_b^2/\sigma_e^2$.

In order to test the null hypothesis $H_{02} : \Delta_2 = 0$ against $H_{12} : \Delta_2 > 0$, the usual test statistic is $\tilde{\mathcal{F}} = MSB/MSR$.

Note that for the unbalanced three stage nested designs the sum of squares x'Ax and x'Rx and also x'Bx and x'Rx are stochastically independent. For exact tests about the variance ratio $\sigma_a^2/\sigma_e^2$ see Verdooren (1974). The complication arises from the theorem: In a three stage nested design the sum of squares x'Ax and x'Bx are stochastically independent if and only if for each $i = 1,\ldots,I$: $n_{ij} = n_{ij'}$, $(j,j'=1,\ldots,J_i)$. Thus only for the partially balanced designs these sums of squares are independent. For the unbalanced designs x'Ax and x'Bx are dependent, hence MSA/MSB cannot be approximated by an F-distribution.

At $H_0$: $\Delta_2 = 0$ $\tilde{\mathcal{F}}$ has an F-distribution with $J_0$-I and n-$J_0$ degrees of freedom.

A generalization of Wald's procedure can be used for the construction of an exact (1-α)-confidence interval of the variance ratio $\Delta_2 = \sigma_b^2/\sigma_e^2$ for the last two stages.

Define $T(\Delta) = \dfrac{\sum_{i=1}^{I}\sum_{j=1}^{J_i} g_{ij}(\bar{x}_{ij}-\tilde{x}_i)^2}{\sum_{i=1}^{I}\sum_{j=1}^{J_i}\sum_{k=1}^{n_{ij}}(x_{k.ij}-\bar{x}_{ij})^2}$ with $g_{ij} = n_{ij}/(\Delta\, n_{ij}+1)$;

$\bar{x}_{ij} = \sum_{k=1}^{n_{ij}} x_{k.ij}/n_{ij}$ and $\tilde{x}_i = \sum_{j=1}^{J_i} g_{ij}\bar{x}_{ij}/\sum_{j=1}^{J_i} g_{ij}$. When Δ is the true variance ratio, $(n-J_0)T(\Delta)/(J_0-I)$ has an F-distribution with $J_0$-I and n-$J_0$ degrees of freedom. Let $f_1$ and $f_2$ be the lower and upper α/2-point of the F-distribution with $J_0$-I and n-$J_0$ degrees of freedom.

The lower confidence limit $\Delta_L$ of $\Delta$ is given by the root of the equation in $\Delta$ : $(n-J_0)T(\Delta)/(J_0-I) = f_2$ and the upper confidence limit $\Delta_U$ of $\Delta$ is given by the root of the equation in $\Delta$ : $(n-J_0)T(\Delta)/(J_0-I) = f_1$.

The generalization to the unbalanced k-stage nested design in order to obtain an exact confidence interval for the variance ratio $\Delta$ of the last two stages, is straightforward.

Copies of the program are available from the author upon request.

## 3. Numerical examples

### 3.1. A two stage nested design

The first example has been taken from Snedecor and Cochran (1967), example 10.18.1, p. 290. In research on artificial insemination of cows, series of semen samples from bulls are sent out and tested for their ability to produce conceptions. The percentages of conceptions from samples for six bulls are shown below.

| Bull (i) | $x_{j.i}$ | $n_i$ | $\bar{x}_i$ |
|---|---|---|---|
| 1 | 46,31,37,62,30 | 5 | 41.20 |
| 2 | 70,59 | 2 | 64.50 |
| 3 | 52,44,57,40,67,64,70 | 7 | 56.29 |
| 4 | 47,21,70,46,14 | 5 | 39.60 |
| 5 | 42,64,50,69,77,81,87 | 7 | 67.14 |
| 6 | 35,68,59,38,57,76,57,29,60 | 9 | 53.22 |
| Total | 1876 | 35 | |

ANOVA-table

| Source | df | SS | MS | E(MS) |
|---|---|---|---|---|
| Correction term | 1 | 100553.60 | | |
| Between bulls | 5 | 3322.06 | 664.41 | $\sigma_e^2+5.669\sigma_a^2$ |
| Within bulls | 29 | 7200.34 | 248.29 | $\sigma_e^2$ |
| Total | 35 | 111076.00 | | |

$\hat{\sigma}_e^2 = 248.29$; $\hat{\sigma}_a^2 = 73.40$.

Test on $H_0$: $\Delta = 0$ against $\Delta = \Delta_1 > 0$ with $\Delta = \sigma_a^2/\sigma_e^2$, $F = MSA/MSR = 2.68$.

Because $Pr(F_{29}^5 > 2.68) = 0.042$, $H_0$ is rejected at significance level $\alpha=0.05$.

The exact confidence interval $\Delta_L < \Delta < \Delta_U$ for $\Delta$ for some confidence coefficient $1 - \alpha$ is:

| $1-\alpha$ | $\Delta_L$ | $\Delta_U$ |
|---|---|---|
| .70 | .093 | 1.13 |
| .80 | .055 | 1.46 |
| .90 | .010 | 2.16 |
| .95 | 0*) | 3.08 |
| .99 | 0*) | 6.50 |
| .995 | 0*) | 8.79 |
| .999 | 0*) | 17.36 |

*) On the interval $0 \leq \Delta < \infty$ there was no solution for the equation in $\Delta$:

$\Sigma_{i=1}^{I} g_i(\bar{x}_i-\bar{\bar{x}})^2 = \frac{I-1}{n-I} f_2 \Sigma_i^I \Sigma_j^{n_i}(x_{j.i}-\bar{x}_i)^2$, namely $\frac{n-I}{I-1} T(\Delta) = f_2$.

When we allow negative values of $\Delta$ with the restriction that the weights $g_i = n_i/(\Delta n_i+1)$ remain positive, we find for $\Delta_L$ : -0.021 for $\alpha = 0.05$, -0.065 for $\alpha = 0.01$, -0.077 for $\alpha = 0.005$ and -0.096 for $\alpha = 0.001$ respectively.

3.2. *A three stage nested design*

The second example consists of fictitious data for a three stage nested design.

| | $A_1$ | | $A_2$ | | $A_3$ | | |
|---|---|---|---|---|---|---|---|
| | $B_{1.1}$ | $B_{2.1}$ | $B_{1.2}$ | $B_{2.2}$ | $B_{1.3}$ | $B_{2.3}$ | $B_{3.3}$ |
| | 7 | 7 | 4 | 6 | 4 | 5 | 4 |
| | 5 | 6 | 4 | 5 | 6 | 3 | 4 |
| | 7 | 6 | | 7 | 5 | 5 | 3 |
| | | 7 | | 6 | 5 | | |
| Total | 19 | 26 | 8 | 24 | 20 | 13 | 11 |
| $n_{ij}$ | 3 | 4 | 2 | 4 | 4 | 3 | 3 |
| $\bar{x}_{ij}$ | 6.33 | 6.5 | 4.0 | 6.0 | 5.0 | 4.33 | 3.67 |
| $n_i$ | 7 | | 6 | | 10 | | |
| $\bar{x}_i$ | 6.43 | | 5.33 | | 4.4 | | |

$n_i = \Sigma_{j=1}^{J_i} n_{ij}$ and $\bar{x}_i = \Sigma_{j=1}^{J_i} \Sigma_{k=1}^{n_{ij}} x_{k.ij}/n_i$ .

ANOVA-table

| Source | df | SS | MS | E(MS) |
|---|---|---|---|---|
| Correction term | 1 | 636.565 | | |
| Between $A_i$ | 2 | 16.987 | 8.494 | $\sigma_e^2+3.435\sigma_b^2+7.478\sigma_a^2$ |
| Between $B_{j.i}$ | 4 | 8.448 | 2.112 | $\sigma_e^2+3.174\sigma_b^2$ |
| Within $B_{j.i}$ = Residual | 16 | 11.000 | 0.688 | $\sigma_e^2$ |
| Total | 23 | 673 | | |

$\hat{\sigma}_e^2 = 0.688$; $\hat{\sigma}_b^2 = 0.449$; $\hat{\sigma}_a^2 = 0.838$.

Test on $H_{02}$: $\Delta_2 = 0$ against $H_{12}$: $\Delta_2 > 0$ with $\Delta_2 = \sigma_b^2/\sigma_e^2$, $\widetilde{F^*}$ = MSB/MSR = 2.112/0.688 = 3.072. Because $Pr(F_{16}^4 > 3.072) = 0.047$, $H_0$ is rejected at significance level $\alpha = 0.05$.

An exact confidence interval $\Delta_L < \Delta_2 < \Delta_U$ for $\Delta_2$ for some confidence coefficient $1-\alpha$ is:

| $1-\alpha$ | $\Delta_L$ | $\Delta_U$ |
|---|---|---|
| .70 | .194 | 2.80 |
| .80 | .107 | 3.73 |
| .90 | .007 | 5.82 |
| .95 | 0*) | 8.76 |
| .99 | 0*) | 21.15 |

*) On the interval $0 \leq \Delta_2 < \infty$ there was no solution for the equation in $\Delta$:

$$\Sigma_i\Sigma_j\, g_{ij}(\bar{x}_{ij}-\tilde{x}_i)^2 = \frac{J_0-I}{n-J_0}\, f_2\Sigma_i\Sigma_j\Sigma_k(x_{k.ij}-\bar{x}_{ij})^2, \text{ namely } \frac{n-J_0}{J_0-I}\, T(\Delta) = f_2.$$

When we allow negative values of $\Delta_2$, with the restriction that the weights $g_{ij} = n_{ij}/(\Delta_2 n_{ij}+1)$ remain positive, we find for $\Delta_L$: -0.059 for $\alpha = 0.05$, -0.151 for $\alpha = 0.01$ respectively.

References

Scheffé, H. The analysis of variance.
John Wiley and Sons, New York, 1959

Snedecor, G.W. and Cochran, W.G.,Statistical Methods (6th ed.).
Iowa State University Press, Ames, Iowa, U.S.A., 1967

Spjøtvoll, E. Optimum invariant tests in unbalanced variance component models,
Ann. Math. Stat. 38, 422-428, 1967

Spjøtvoll, E., Most powerful tests for some non-exponential families.
Ann. Math. Stat. 39, 772-784, 1968

Verdooren, L.R., Exact tests about variance ratios in unbalanced two and three stages nested designs.
Contributed paper, 8th International Biometric Conference, Constanţa, Romania, 1974

Wald, A., A note on the analysis of variance with unequal class frequencies,
Ann. Math. Stat. 11, 96-100, 1940

CHAPTER 5

# 5. ESTIMATION OF VARIANCE COMPONENTS

## PART I: ONSET

### 5.1. Introduction

As described in chapter 2, the search for estimators of variance components till 1975 was concentrated on unbiased estimators with some optimality criterion. For several designs (nested or crossed) quadratic forms in the observations, $\underline{y}' Q \underline{y}$, were sought to get an unbiased estimator of variance components. A nice feature of unbiased estimators $\hat{\underline{\sigma}}_i^2 = \underline{y}'Q_i\underline{y}$ for the variance components $\sigma_i^2$ $(i = 1,\ldots,k)$ respectively, is that a linear combination of variance components $\sum_{i=1}^{k} \lambda_i\sigma_i^2$ will be unbiasedly estimated by $\sum_{i=1}^{k} \lambda_i\hat{\underline{\sigma}}_i^2$.

As optimality criterion for an estimator of a variance component $\sigma_i^2$, which consists of a quadratic form in the observations, an unbiased estimator was used which has minimum variance, the so called Best Quadratic Unbiased Estimator (BQUE) or Minimum Variance Unbiased Estimator (MIVQUE).

A critique could however be made. Only the error variance component $\sigma_k^2$ would be unbiasedly estimated by a quadratic form $\underline{y}'Q_k\underline{y}$, such that the estimates were always non-negative; otherwise stated $Q_k$ is a non-negative definite matrix. The other variance components $\sigma_i^2$ $(i = 1,\ldots,k-1)$ were unbiasedly estimated by quadratic functions $\underline{y}'Q_i\underline{y}$ such that the outcomes were not always non-negative; otherwise stated $Q_i$ is <u>not</u> a non-negative definite matrix. Because a variance component $\sigma_i^2$ is by definition non-negative, it is amazing to see how much research has been done to find yet another type of unbiased estimator. It is the very requirement of unbiasedness that leads to

variance component estimators which give (some quite often) negative estimates, which makes no sense. Of course an unbiased non-negative estimator for $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ is a desirable one, if such an estimator exists.

To focus attention on estimators which do make sense, the author introduced the concept of a permissible estimator. The necessary condition for an estimator must be that it is a permissible estimator, meaning an estimator which gives estimates that, with probability one, lie in the parameter space. Hence a permissible variance component estimator is a non-negative estimator.

In section 5.2 this basic idea of estimation is elucidated. (In chapter 6 the probabilities are given that the Best Quadratic Unbiased Estimator for the one-way balanced random model gives negative outcomes).

Furthermore, the several types of estimators of variance components in use around 1980 are described in section 5.2. Also a modification of a least squares estimator for a variance component is derived which is non-negative. Such a closest-to-unbiased non-negative estimator can be found by solving a quadratic programming problem.

5.2. On estimation of variance components

Paper presented at the VI-th International Conference on Mathematical Statistics at Wisla, Poland, in 1978.

The content of this paper was also delivered as an invited lecture at the Fourth International Summer School on Problems of Model Choice and Parameter Estimation in Regression Analysis at Mühlhausen/Thuringia (GDR), in 1979.

Published in: Statistica Neerlandica 34 (1980) no. 2, pp. 83 - 106.

# On estimation of variance components*

by L.R. VERDOOREN**

**Summary** In this survey paper the estimation of variance components is given. The least squares approach in variance component estimation is a unifying principle which includes the analysis of variance estimators and the MINQUE. When normality is assumed the maximum likelihood estimators can be used. Many variance component estimators are not permissible because they are not non-negative. The development of non-negative variance component estimators is indicated.

## 1 Introduction

Let us first recall some basic ideas of estimation theory in section 1.1. Some criticism can be made on the usual estimation procedures. In section 1.2 we will focus our attention on the minimum variance unbiased estimation of a variance and variance components.

### *1.1 BASIC IDEAS OF ESTIMATION*

The probabilistic model implies that the outcome $y$ is observed of a random variable $\underline{y}$ which assumes values in the "sample space" $\mathcal{Y}$ according to a probability distribution $P$ which is known to belong to a given class $\mathcal{P} = \{P_\theta ; \theta \in \Theta\}$ of probability distributions. If $P = P_\theta$ then $\theta$ is the true value of the underlying unknown parameter. We shall be interested in the estimation of a real valued parameter $g$: $\Theta \to \mathbb{R}$, or equivalently of the corresponding true value $\gamma = g(\theta)$. We need an estimator $d$: $\mathcal{Y} \to \mathbb{R}$ with the interpretation that $d(y)$ will be used as an estimate for $\gamma = g(\theta)$ if the outcome $y \in \mathcal{Y}$ of $\underline{y}$ is obtained.

Estimators will be compared by means of their mean squared error (or risk function if squared error loss is introduced):

$$\mathrm{MSE}_\theta(d) = \mathrm{E}_\theta\,(d(\underline{y}) - g(\theta))^2 = \mathrm{var}_\theta\, d(\underline{y}) + (\mathrm{bias}_\theta\,(d))^2$$

where

$$\mathrm{bias}_\theta(d) = \mathrm{E}_\theta\, d(\underline{y}) - g(\theta).$$

An estimator is said to be admissible if no estimator exists with uniformly (in $\theta$) smaller MSE. Or, more precisely, $d$ is admissible if no $d'$ exist with $\mathrm{MSE}_\theta\,(d') \leqslant \mathrm{MSE}_\theta\,(d)$ for all $\theta \in \Theta$ and $\mathrm{MSE}_\theta\,(d') < \mathrm{MSE}_\theta\,(d)$ for at least one $\theta$.

In practice estimators are often constructed by requiring unbiasedness: the estimator $d$ for $\gamma = g(\theta)$ is said to be unbiased if $\mathrm{bias}_\theta\,(d) = 0$ for all $\theta \in \Theta$. Since estimators with uniformly minimum mean squared error rarely exist, a reasonable procedure is to restrict the class of estimators and to look for estimators with uniformly minimum MSE within that class. One such class which is used (too) often, is that of unbiased estimators.

---

* An earlier version of this paper was presented at the VI-th International Conference on Mathematical Statistics ar Wisla, Poland, 7-13 December 1978.

** Department of Mathematics, Agricultural University, Wageningen.

An estimator within that class with minimum MSE is also an estimator with uniformly minimum variance among unbiased estimators (UMVUE).

Sometimes there is a sufficient statistic $t(\underline{y})$ for $g(\theta)$, i.e. if the partition of the sample space generated by $t$ is sufficient. (A partition $\mathscr{A}$ of the sample space is said to be sufficient for $\theta$ (more precisely, for the family $\{P_\theta; \theta \in \Theta\}$) if for any $A \in \mathscr{A}$ and $B \in \mathscr{B}$ holds that $P_\theta(\underline{y} \in B | \underline{y} \in A)$ does not depend on $\theta$). Hence for the sufficient statistic $t(\underline{y})$ the conditional distribution $P_\theta(\underline{y} \in B | t(y) = t)$ on the sample space, for given $t$, does not depend on $\theta$. The relevance of sufficiency to the question of minimum variance unbiased estimators is given by the RAO-BLACKWELL theorem:

Let $\{P_\theta; \theta \in \Theta\}$ be a family of distributions on a sample space $\mathscr{Y}$ and suppose that $\tilde{g}(\underline{y})$ is an unbiased estimator of a real-valued function $g$ of $\theta$. Then if $t(\underline{y})$ is a sufficient statistic for $\theta$, $E_\theta\{\tilde{g}(\underline{y})|t\}$ is also an unbiased estimator of $g(\theta)$ and it has variance uniformly no larger than that of $\tilde{g}(\underline{y})$.

Before this theorem can be used to establish the existence of MVUE's we have to demand the existence of a sufficient statistic $t$ with an additional property. Suppose that we are estimating $g(\theta)$ and that there exists a sufficient statistic $t$ with the property that there is a unique function of $t$, $\hat{g}(t)$ say, which is an unbiased estimator of $\theta$. Then $\hat{g}(t)$ is an MVUE of $g(\theta)$. So another question arises. How can we ascertain whether a sufficient statistic has this additional property? Sometimes we can do so by using the notion of completeness of a family of distributions. Let $\{P_\theta; \theta \in \Theta\}$ be a family of distributions on the sample space $\mathscr{Y}$. This family is said to be complete if $E_\theta\{f(\underline{y})\} = 0$ for all $\theta \in \Theta$ implies $f(y) = 0$ almost everywhere.

We can express now the usefulness of these notions in the following special case of the general LEHMANN-SCHEFFÉ theorem:

Let $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n$ be a random sample from a distribution that has a probability density function $f(y, \theta), \theta \in \Theta$; let $t(\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n)$ be a sufficient statistic for $g(\theta)$, and let the family $\{h(t, \theta): \theta \in \Theta\}$ of probability density functions be complete. If there is a continuous function of $\underline{t}$ which is an unbiased statistic for $g(\theta)$, then this function of $\underline{t}$ is the unique MVUE for $g(\theta)$.

Or loosely formulated: if a complete sufficient statistic exists, then every function of it is a uniformly MVUE of its expected value.

A well known example to illustrate these ideas is the following.

Let $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n$ be a random sample from the normal distribution $N(\mu, \sigma^2)$. A complete sufficient statistic for $(\mu, \sigma^2)$ is $\left(\overline{\underline{y}}, \sum_{i=1}^{n} (\underline{y}_i - \overline{\underline{y}})^2\right)$ where $\overline{\underline{y}} = \sum_{i=1}^{n} \underline{y}_i/n$. The reason for using $\underline{s}^2 = (n-1)^{-1} \sum_{i=1}^{n} (\underline{y}_i - \overline{\underline{y}})^2$ as an estimator for $\sigma^2$ is that it is UMVUE by the Lehmann-Scheffé theorem. However the above-mentioned estimator $\underline{s}^2$ for $\sigma^2$ can be improved in the class of quadratic functions of the observations by using the (biased) estimator $(n+1)^{-1} \sum_{i=1}^{n} (\underline{y}_i - \overline{\underline{y}})^2$. So in the class of "quadratic estimators" (quadratic functions of the observations), the estimator $\underline{s}^2$ is inadmissible, but in the class of unbiased quadratic estimators it is admissible.

In our view it is also worthwhile to introduce the concept of permissible estimation. Let $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n$ be a random sample of a random variable $\underline{y}$ with probability distribution $P \in \mathcal{P} = \{P_\theta : \theta \in \Theta\}$. An estimator $\underline{t} = t(\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n)$ for $\theta$ is said to be permissible if $\underline{t} \in \Theta$ holds with probability one.

In some situations $\Gamma = g(\Theta) = \{\gamma; \gamma = g(\theta)$ for some $\theta \in \Theta\}$ is an interval of the form $[a, b]$ or $[a, \infty)$ or $(-\infty, b]$ or some other bounded set. If the estimator $d: \mathcal{Y} \to \mathbb{R}$ assumes values outside $\Gamma$ with positive probability, (hence $d$ is impermissible), then it can obviously be improved by truncating. This is seen immediately as follows.

Let $d: \mathcal{Y} \to \mathbb{R}$ be any estimator for $\gamma = g(\theta)$ (it needs not be unbiased) such that $\Gamma = g(\Theta) = [a, b]$ and $P_\theta\{d(\underline{y}) \notin \Gamma\} > 0$ for at least one $\theta \in \Theta$. Let $d'$ be the corresponding truncated estimator:

$$d'(y) = \begin{cases} a & \text{if } d(y) < a \\ d(y) & \text{if } a \leqslant d(y) \leqslant b \\ b & \text{if } d(y) > b \end{cases}$$

then for any $\theta \in \Theta$ we have $g(\theta) \in \Gamma$ and hence $(d(y) - g(\theta))^2 \geqslant (d'(y) - g(\theta))^2$ for all $y \in \mathcal{Y}$. Hence $\mathrm{MSE}_\theta(d) \geqslant \mathrm{MSE}_\theta(d')$ for all $\theta \in \Theta$. For $\theta$ with $P_\theta\{d(\underline{y}) \notin \Gamma\} > 0$ we obviously have strict inequality and hence $d$ is inadmissible.

*Remark*

The theorem can be extended to the case where $\underline{y}$ is a random vector and $\Theta$ is a closed convex set. Truncation here means that an estimate which falls outside this set $\Theta$ is replaced by the boundary point of $\Theta$ nearest to it.

We conclude this section by mentioning some textbooks examples. Note that uncritical use of the restriction of unbiasedness may lead to anomalies.

*Example* (i) KENDALL and STUART [19]. If the zero frequency of a Poisson distribution with parameter $\theta$ cannot be observed, the distribution is called a Truncated Poisson distribution. From a single observation $\underline{x}$ the only unbiased estimator of the probability of non-zero observations $1 - e^{-\theta}$ takes the values 0 when $x$ is odd and 2 when $x$ is even.

*Example* (ii) FERGUSON [7, page 135]. Let $\underline{x}_1, \ldots, \underline{x}_n$ be a random sample from the normal distribution $N(\theta, 1)$. The statistic $\underline{t} = \sum_{j=1}^{n} \underline{x}_j$ is a complete sufficient statistic for $\theta$. Furthermore $\bar{\underline{x}} = \underline{t}/n$ is an unbiased estimator of $\theta$. Hence $\bar{\underline{x}}$ is a minimum variance unbiased estimator (MVUE) of $\theta$. On the other hand $E\bar{\underline{x}}^2 = \theta^2 + 1/n$, so that $\bar{\underline{x}}^2 - 1/n$ is MVUE of $\theta^2$. But unbiasedness itself is not an optimum property. The MVUE of $\theta^2$ is impermissible for it occasionally gives negative values for a parameter known to be positive. The (biased) estimator $\max(0, \bar{\underline{x}}^2 - 1/n)$ has smaller mean squared error for all values of $\theta$ hence the MVUE is inadmissible.

*Example* (iii) FERGUSON [7, pag. 136]. From a Poisson distribution with parameter $\theta$, we wish to estimate $e^{-2\theta}$ as a function of $\underline{x}$. The only unbiased estimator of $e^{-2\theta}$ is

$t(\underline{x}) = (-1)^{\underline{x}}$. This ridiculous estimator is a MVUE, as the class of unbiased estimators has only one element. WASAN [49] looks for an unbiased estimator of $e^{-3\theta}$. The MVUE $(-2)^{\underline{x}}$ is impermissible.

*1.2 MVUE OF A VARIANCE AND VARIANCE COMPONENTS*

Let us consider the usual linear regression model $\underline{y} = X\beta + \underline{e}$, where $X$ is an $n \times p$-matrix of rank $p$; $y$ and $e \in \mathbb{R}^n$, the parameter vector $\beta \in \mathbb{R}^p$; $E(\underline{e}) = 0$, var $\underline{e} = \sigma^2 I$ and $\underline{e}$ has a multinormal distribution $N(0, \sigma^2 I)$. The MVUE for $\sigma^2$ is given by the customary unbiased estimator $\underline{s}^2$ of $\sigma^2$ with $n$-$p$ degrees of freedom.

$$\underline{s}^2 = \{\underline{y}'\underline{y} - \underline{y}'X(X'X)^{-1}X'\underline{y}\} / (n - p)$$

The minimum mean squared error estimator for $\sigma^2$ is the biased estimator $(n - p) \times \underline{s}^2/(n - p + 2)$. For a proof along the lines of section 1.1 see CORSTEN [6].

Another example to show that unbiased estimation is inadequate follows now. Consider the balanced one-way random effects model: Let $\underline{y}_{ij} = \mu + \underline{a}_i + \underline{e}_{ij}$, $i = 1, \ldots, n_1$; $j = 1, \ldots, n_2$. The random variables $\underline{a}_i$ are $N(0, \sigma_1^2)$ and are identically and independently distributed (i.i.d.); the random variables $\underline{e}_{ij}$ are $N(0, \sigma_2^2)$ and also i.i.d.; the $\underline{a}_i$ and $\underline{e}_{ij}$ are independent. From the model it follows that $\text{var}(\underline{y}_{ij}) = \sigma_1^2 + \sigma_2^2$ and these variances $\sigma_1^2$ and $\sigma_2^2$ are called the variance components.

The analysis of variance table reads:

| Source of variation | df | SS | MS | E(MS) |
|---|---|---|---|---|
| General mean level | 1 | SSL | MSL | $\sigma_2^2 + n_2\sigma_1^2 + n_1 n_2 \mu^2$ |
| Between groups | $n_1 - 1$ | SSA | MSA | $\sigma_2^2 + n_2\sigma_1^2$ |
| Within groups, or error | $n_1(n_2 - 1)$ | SSE | MSE | $\sigma_2^2$ |
| Total | $n_1 n_2$ | SST | | |

$$\text{SST} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} y_{ij}^2, \quad \text{SSL} = (\Sigma_i \Sigma_j y_{ij})^2 / (n_1 n_2),$$

$$\text{SSA} = \sum_{i=1}^{n_1} (\sum_{j=1}^{n_2} y_{ij})^2 / n_2 - \text{SSL}; \quad \text{SSE} = \text{SST-SSL-SSA}.$$

An unbiased estimator for $\mu$ is $\underline{\bar{y}} = \sum_i \sum_j \underline{y}_{ij}/(n_1 n_2)$; an unbiased estimator for $\sigma_2^2$ is $\underline{\hat{\sigma}}_2^2 = \underline{\text{MSE}}$, and an unbiased estimator for $\sigma_1^2$ is $\underline{\hat{\sigma}}_1^2 = (\underline{\text{MSA}}\text{-}\underline{\text{MSE}})/n_2$. The estimators $\underline{\hat{\sigma}}_2^2$ and $\underline{\hat{\sigma}}_1^2$ are called the analysis of variance estimators. The statistics $\underline{\bar{y}}$, $\underline{\text{SSA}}$ and $\underline{\text{SSE}}$ are a set of independent minimal sufficient statistics and it can be shown that they are complete. Hence $\underline{\bar{y}}$, $\underline{\hat{\sigma}}_2^2$ and $\underline{\hat{\sigma}}_1^2$ are uniformly minimum variance unbiased estimators for $\mu$, $\sigma_2^2$ and $\sigma_1^2$ respectively. (GRAYBILL and WORTHAM [9] and GRAYBILL [10, page 342].

Remark that $\underline{\hat{\sigma}}_2^2$ is a permissible estimator for $\sigma_2^2$. A minimum MSE estimator for $\sigma_2^2$ is given by $SSE/(n_1(n_2 - 1) + 2)$.

However $\underline{\hat{\sigma}}_1^2$ is an impermissible estimator since there is a positive probability that $\underline{\hat{\sigma}}_1^2$ takes negative values under the model assumed:

$$P(\underline{\hat{\sigma}}_1^2 \leqslant 0) = P((\underline{MSA} - \underline{MSE})/n_2 \leqslant 0) = P(\underline{MSA}/\underline{MSE} \leqslant 1) =$$

$$= P(\underline{F}_q^p \leqslant (1 + n_2\gamma)^{-1})$$

where $\gamma = \sigma_1^2/\sigma_2^2$, $p = n_1 - 1$ and $q = n_1(n_2 - 1)$. In the following table, values of $100\ P(\underline{F}_q^p \leqslant (1 + n_2\gamma)^{-1})$ are given for several values of $n_2\gamma$.

| | $p$ | | | | | | $p$ | | | | | | $p$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $q$ | 1 | 5 | 10 | 30 | 50 | 100 | 1 | 5 | 10 | 30 | 50 | 100 | 1 | 5 | 10 | 30 | 50 | 100 | $q$ |
| 1 | 50 | 36 | 34 | 33 | 32 | 32 | 44 | 28 | 25 | 23 | 23 | 23 | 33 | 14 | 11 | 09 | 09 | 09 | 1 |
| 5 | 64 | 50 | 47 | 43 | 43 | 42 | 55 | 33 | 27 | 22 | 21 | 20 | 41 | 13 | 07 | 03 | 02 | 01 | 5 |
| 10 | 66 | 53 | 50 | 47 | 46 | 45 | 57 | 34 | 27 | 19 | 17 | 15 | 42 | 12 | 05 | 01 | 00 | 00 | 10 |
| 30 | 67 | 57 | 53 | 50 | 49 | 48 | 58 | 35 | 25 | 14 | 10 | 07 | 43 | 11 | 04 | 00 | 00 | 00 | 30 |
| 50 | 68 | 57 | 54 | 51 | 50 | 49 | 58 | 35 | 25 | 12 | 08 | 04 | 43 | 11 | 03 | 00 | 00 | 00 | 50 |
| 100 | 68 | 58 | 55 | 52 | 51 | 50 | 58 | 35 | 25 | 10 | 05 | 02 | 43 | 11 | 03 | 00 | 00 | 00 | 100 |
| | $n_2\gamma = 0$ | | | | | | $n_2\gamma = 0.5$ | | | | | | $n_2\gamma = 2$ | | | | | | |

The biased estimator $\max(0, \underline{\hat{\sigma}}_1^2)$ has smaller MSE than $\underline{\hat{\sigma}}_1^2$ for all values of $\sigma_1^2$.

In reading the literature on the estimation of variance components, one often gets the impression that unbiasedness is a desirable property. We have already seen that concept of unbiasedness is only a mathematical tool. Perhaps one of the reasons for the many publications concerning unbiased estimators of variance components is that, in the restricted class of quadratic functions of the observations, one often gets one solution for the uniformly minimum variance quadratic unbiased estimator (UMVQUE) or for the best quadratic unbiased estimator (BQUE). Some statisticians seem to argue that a sequence of averages of estimators is consistent if and only if the individual estimators are unbiased. But a research worker mostly does not lean on an average of several estimates; he is more interested in individually permissible estimators and he will be happy if these estimators are admissible too.

It is not true that a negative estimate of a variance component gives strong evidence that the model does not apply. Under validity of the assumptions of the model there may be a positive (and sometimes considerable) probability that such unbiased estimators attain negative values, see VERDOOREN [48].

If the frequency of negative values of unbiased variance components estimators is much larger than expected under the model we may ask ourselves whether the model is correct. Of course the testing of the adequacy of the model is another question and it does not depend only on the variance component estimators.

## 2 Unbiased variance component estimators

### *2.1 ANALYSIS OF VARIANCE ESTIMATORS*

Let us consider in detail the variance components model:

$$\underline{y} = X\beta + \sum_{i=1}^{k} U_i \underline{a}_i = X\beta + U\underline{a},$$

with $U = (U_1 : U_2 : \ldots : U_k)$ and $\underline{a}' = (\underline{a}_1' : \underline{a}_2' : \ldots : \underline{a}_k')$, where $\underline{y}$ is an $n \times 1$-vector of observations, $X$ is an $n \times p$-design matrix of known constants and of full rank ($p < n$), $\beta$ is a $p \times 1$-vector of fixed effects parameters, $U_i$ a known $n \times m_i$-matrix, $\underline{a}_i$ a $m_i \times 1$ vector of random effects such that

(1) $E(\underline{a}_i) = 0, (i = 1, 2, ..., k)$

(2) the $\underline{a}_i$'s are uncorrelated: $E(\underline{a}_i \underline{a}_j') = 0, i \neq j$

(3) $E(\underline{a}_i \underline{a}_i') = \sigma_i^2 \, I_{m_i}$

The column space of (X, U) is supposed to be $\mathbb{R}^n$. The assumptions imply VAR $(\underline{y}) = \Sigma_i \sigma_i^2 U_i U_i' = \Sigma_i \sigma_i^2 V_i$ where $V_i = U_i U_i'$.

Let $P_0 = X(X'X)^{-1}X'$ be the orthogonal projection operator on the column space of $X$. Let $P_i$ for $i = 1, 2, \ldots, k$ be the orthogonal projection operators on the column space of $(X, U_1, U_2, \ldots, U_i)$ for $i = 1, \ldots, k$ respectively, hence $P_k = I_n$. Let $N_j = P_j - P_{j-1}$ for $j = 1, 2, \ldots, k\text{-}1$ and $N_k = P_k - P_{k-1}$. The projection operator $N_j$ denotes the orthogonal projection on the orthogonal complement of the columnspace of $(X, U_1, \ldots, U_{j-1})$, (the columnspace of $X$ in case $j = 1$), in the columnspace of $(X, U_1, \ldots, U_j)$ for $j = 1, \ldots, k\text{-}1$ and $N_k = P_k - P_{k-1} = I_n - P_{k-1}$ is the orthogonal projection on the orthogonal complement of $(X, U_1, \ldots, U_{k-1})$. HENDERSON's [16] Method III can be described as follows.

(1) Calculate $y'N_j y$ for $j = 1, 2, \ldots, k$ and

(2) Solve the system of linear equations of sums of squares

$$y'N_j y = \sum_{i=j}^{k} \hat{\sigma}_i^2 \operatorname{tr}(N_j V_i) \qquad (\text{for } j = 1, \ldots, k).$$

This procedure is the hierarchical set up of the Analysis of Variance table. HENDERSON's method III is based on the following obvious decomposition of $\underline{y}$ into orthogonal complements:

$$\underline{y} = P_0 \underline{y} + \sum_{j=1}^{k} N_j \underline{y}$$

with the consequence that

$$\underline{y}'\underline{y} = \underline{y}'P_0\underline{y} + \sum_{j=1}^{k} \underline{y}'N_j\underline{y}$$

where $P_0\underline{y}$ can be used for estimating $\beta$, $\underline{y}'N_k\underline{y}$ for estimating the residual variance $\sigma_k^2$, while $\underline{y}'N_j\underline{y}$ for $j = 1, \ldots, k\text{-}1$ can be used to give unbiased estimators for the variance components $\sigma_1^2, \ldots, \sigma_{k-1}^2$. Because $N_jX = 0, N_jU_i = 0$ for $i < j$,

$$E(\underline{y}'N_j\underline{y}) = \mathrm{tr}(N_jV) = \sum_{i=j}^{k} \sigma_i^2 \mathrm{tr}(N_jV_i)$$

for $j = 1, \ldots, k$, the estimators $\underline{\hat{\sigma}}_i^2$ of $\sigma_i^2$ are unbiased. This estimation method depends however on the order of the $U_j$ in the definition of the projection operators $P_i$. A further disadvantage of this method is that there is a positive probability for negative values of $\underline{\hat{\sigma}}_i^2 = \underline{y}'A_i\underline{y}$ when $A_i$ is not non-negative definite. Thus all estimators $\hat{\sigma}_i^2$ $(i = 1, \ldots, k\text{-}1)$ are impermissible.

*Example 1*

Let us consider the balanced one-way random effects model of section 1.2 with $n_1 = 3$ classes and $n_2 = 2$ observations per class, hence $n = n_1n_2 = 6$, and $y' = (y_{11}, \ldots, y_{1\,n_2}$ $y_{21}, \ldots, y_{n_1n_2})$. Written in the model $\underline{y} = X\beta + \sum_{i=1}^{k} U_i\underline{a}_i$ with $k = 2$, we have $X$ is a 6x1-matrix with $X' = (1,1,1,1,1,1)$, $\beta$ is equal to $\mu$; $U_1$ is the 6x3-matrix given below, $U_2 = I_6$, $P_0$ and $P_1$ are given below, $P_2 = I_6$, $N_1 = P_1 - P_0$ and $N_2 = I_6 - P_1$ are given below.

$$U_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad P_0 = \tfrac{1}{6} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}, P_1 = \tfrac{1}{2} \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}, N_1 = \tfrac{1}{6} \begin{pmatrix} 2 & 2 & -1 & -1 & -1 & -1 \\ 2 & 2 & -1 & -1 & -1 & -1 \\ -1 & -1 & 2 & 2 & -1 & -1 \\ -1 & -1 & 2 & 2 & -1 & -1 \\ -1 & -1 & -1 & -1 & 2 & 2 \\ -1 & -1 & -1 & -1 & 2 & 2 \end{pmatrix}$$

$$N_2 = \tfrac{1}{2} \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$$

In the analysis of variance table the degrees of freedom are the ranks of $P_0$, $N_1$ and $N_2$ respectively and the sum of squares are SSL = $y'P_0y$, SSA = $y'N_1y$, SSE = $y'N_2y$ and SST = $y'y$. VAR($\underline{y}$) = $\sigma_1^2U_1U_1' + \sigma_2^2U_2U_2' = \sigma_1^2V_1 + \sigma_2^2V_2$ where $V_1$ is given below and $V_2 = I_6$.

$$V_1 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

GRAYBILL [8] showed that for the random effects model with balanced experimental design (i.e. equal numbers in the subclasses) the analysis of variance (ANOVA) estimators have uniformly minimum variance in the class of unbiased quadratic estimators, regardless the probability distributions of the components. READ [39] showed that unbiased quadratic estimators with uniformly minimum variance do not exist if the design for the random effects model is not balanced. KLEFFE [20] proves a similar result. GRAYBILL and WORTHAM [9] showed that for a balanced design with normally distributed random effects the ANOVA estimators are functions of independent minimal sufficient statistics, which are also complete, by which these estimators are uniformly minimum variance unbiased estimators.

Despite these properties of the ANOVA estimators there are some drawbacks.

*Example 2*

Let us consider the following example of a mixed model with two missing observations $y_{22}$ and $y_{23}$.

| | $B_1$ | $B_2$ | $B_3$ |
|---|---|---|---|
| $A_1$ | $y_{11}$ | $y_{12}$ | $y_{13}$ |
| $A_2$ | $y_{21}$ | | |

The model is $\underline{y}_{ij} = \alpha_i + \underline{b}_j + \underline{e}_{ij}$ $(i = 1, 2; j = 1, \ldots, 3)$ where the effects $\alpha_i$ are fixed, while the effects $\underline{b}_j$ and $\underline{e}_{ij}$ are random:

$E\underline{b}_j = 0$, $E\underline{e}_{ij} = 0$, $\text{var}\,\underline{b}_j = \sigma_1^2$, $\text{var}\,\underline{e}_{ij} = \sigma_2^2$, $\text{cov}(\underline{b}_j, \underline{b}_{j'}) = 0$ for $j \neq j'$,

$\text{cov}(\underline{e}_{ij}, \underline{e}_{i'j'}) = 0$ for $i \neq i'$ or $j \neq j'$; $\text{cov}(\underline{b}_j, \underline{e}_{ij'}) = 0$ for all $i$, $j$ and $j'$.

The vector model with $y' = (y_{11}, y_{12}, y_{13}, y_{21})$ is

$$\underline{y} = X\beta + U_1\underline{a}_1 + U_2\underline{a}_2$$

with $\beta = (\alpha_1, \alpha_2)'$; $\underline{a}_1 = (\underline{b}_1, \underline{b}_2, \underline{b}_3)'$ and $\underline{a}_2 = (\underline{e}_{11}, \underline{e}_{12}, \underline{e}_{13}, \underline{e}_{21})'$, $X$ and $U_1$ are given below, $U_2 = I_4$, $\text{VAR}(\underline{y}) = \sigma_1^2 U_1 U_1' + \sigma_2^2 U_2 U_2' = \sigma_1^2 V_1 + \sigma_2^2 V_2$ where $V_1$ is given below and $V_2 = I_4$.

$$X = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} ; U_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} ; V_1 = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

The analysis of variance table reads:

| Source of variation | *df* | SS | |
|---|---|---|---|
| General mean level | 1 | $y'P_{00}y$ | } $y'P_0y$ |
| Fixed A effects | 1 | $y'N_0y$ | |
| Random B effects | 2 | $y'N_1y$ | |
| Error | 0 | $y'N_2y$ | |
| Total | 4 | $y'y$ | |

The orthogonal projectors $P_0$, $P_{00}$ and $N_1$ are given below while $N_0 = P_0 - P_{00}$, $P_1 = P_2 = I_4$ and $N_2 = 0$.

$$P_0 = \tfrac{1}{3}\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} , P_{00} = \tfrac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} , N_1 = \tfrac{1}{3}\begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Since no unique solution for the variance components from the EMS-column exists, ANOVA estimation fails here. In the following section 3 we discuss the least squares estimators by SEELY which provide a solution.

### *2.2 MINQUE ESTIMATORS*

In this section we introduce first the MINQUE according to the its originator RAO. In section 4 we will show that MINQUE is a generalized least squares estimator.

RAO [34, 35, 36, 37] introduced the minimum norm quadratic unbiased estimator (MINQUE) $\underline{y}'A\underline{y}$, $A$ symmetric, for a linear combination $q'\theta = \sum_{i=1}^{k} q_i\sigma_i^2$. For unbiasedness we must have

$$E(\underline{y}'A\underline{y}) = \mathrm{tr}(AV) + \beta'X'AX\beta = \sum_{i=1}^{k} \sigma_i^2 \,\mathrm{tr}(AV_i) + \beta'X'AX\beta,$$

so that necessary and sufficient conditions for $\underline{y}'A\underline{y}$ to be unbiased for $q'\theta$ are that

$$\left.\begin{aligned} &X'AX = 0 \\ &\mathrm{tr}(AV_i) = q_i \quad i = 1, \ldots, k \end{aligned}\right\} \tag{1}$$

If we want an estimator which is invariant under translation of $\beta$ we consider $\beta_d = \beta - \beta_0$,

and put $\underline{y}_d = \underline{y} - X\beta_0$. Then the invariant estimator satisfies $\underline{y}_d' A \underline{y}_d = \underline{y}' A \underline{y}$ for all $\beta_0$, which is equivalent to $AX = 0$. Thus invariance implies the first condition of unbiasedness. The conditions for an unbiased invariant estimator for $q'\theta$ are:

$$\left.\begin{aligned} &AX = 0 \\ &\mathrm{tr}(AV_i) = q_i, \quad i = 1, \ldots, k \end{aligned}\right\} \tag{2}$$

Let $\alpha_1^2, \ldots, \alpha_k^2$ be a priori (or approximate) values of $\sigma_1^2, \ldots, \sigma_k^2$. (If no a priori information is available, then each $\alpha_i^2$ may be taken to be unity.) Then we can rewrite

$$\underline{y} = X\beta + W_1\underline{b}_1 + \ldots + W_k\underline{b}_k, \quad W_i = \alpha_i U_i, \quad \underline{a}_i = \alpha_i \underline{b}_i \quad \text{and}$$

$$\mathrm{VAR}(\underline{y}) = \gamma_1^2 T_1 + \ldots + \gamma_k^2 T_k, \quad T_i = W_i W_i' = \alpha_i^2 V_i,$$

where $\gamma_i^2 = \sigma_i^2/\alpha_i^2$, the rescaled variance components to be estimated. If the hypothetical variables $\underline{b}_i$ were known, then a natural estimator of $\gamma_i^2$ would be $\underline{b}_i'\underline{b}_i/m_i$ and hence

$$\widehat{\underline{q'\theta}} = \sum_{i=1}^{k} q_i \alpha_i^2 \underline{b}_i'\underline{b}_i/m_i = \underline{b}'\Delta\underline{b} \text{ (say)},$$

where $\Delta$ is a suitably defined diagonal matrix and $\underline{b}' = (\underline{b}_1' : \ldots : \underline{b}_k')$. But the proposed estimator is $\underline{y}'A\underline{y} = \underline{b}'W'AW\underline{b}$ since $X'AX = 0$ (unbiased) or $AX = 0$ (invariant), where $W = (W_1 : \ldots : W_k)$. The difference between $\underline{b}'W'AW\underline{b}$ and $\underline{b}'\Delta\underline{b}$ is $\underline{b}'(W'AW - \Delta)\underline{b}$ which can be made small, in some sense, by minimizing $\| W'AW - \Delta \|$, where the norm $\| \cdot \|$ is suitably chosen. We shall use the Euclidean norm, i.e. $\| B \|^2 = \mathrm{tr}\, BB'$ = the sum of squares of all the elements of $B$.

For the MINQUE (without invariance) the problem reduces to that of determining a symmetric $A$ such that $\mathrm{tr}(ATAT) + 2\mathrm{tr}(AXX'AT)$ is a minimum subject to the conditions (1), where

$$T = \sum_{i=1}^{k} T_i = \sum_{j=1}^{k} \alpha_i^2 V_i.$$

For the invariant MINQUE the problem reduces to that of determining a symmetric $A$ such that $\mathrm{tr}(ATAT)$ is a minimum subject to the conditions (2). PRINGLE [30] gives an explicit solution for $A$ in the MINQUE (without invariance) $\underline{y}'A\underline{y}$, i.e.

$$A = \sum_{j=1}^{k} \lambda_j A_j, \quad \text{where}$$

$$A_j = (T + XX')^{-1}(V_j - QT^{-1}V_jT^{-1}Q)(T + XX')^{-1},$$

$$Q = X(X'T^{-1}X)^{-1}X',$$

and the $\lambda_j$ are determined from

$$\sum_{j=1}^{k} \lambda_j \mathrm{tr}(A_j V_i) = q_i, \quad i = 1, \ldots, k.$$

RAO [35, 37] gives an explicit solution for $A$ in the invariant MINQUE:

$$A = \sum_{i=1}^{k} \lambda_i R T_i R,$$

where

$$R = T^{-1} - T^{-1} X (X' T^{-1} X)^{-1} X' T^{-1}$$

and $\lambda_i$ satisfies the equations

$$\sum_{j=1}^{k} \lambda_j \mathrm{tr}(R T_j R T_i) = \alpha_i^2 q_i, \quad i = 1, \ldots, k.$$

When $\underline{y}$ has an $n$-variate normal distribution, $\mathrm{var}(\underline{y}' A \underline{y}) = 2\mathrm{tr}(AVAV) + 4\beta' X' AVAX\beta = 2\mathrm{tr}(AVAV)$ since $AX = 0$. Thus if the true variance components are $\alpha_1^2, \ldots, \alpha_k^2$, $\mathrm{tr}(ATAT)$ is proportional to $\mathrm{var}(\underline{y}' A \underline{y})$ and the invariant MINQUE is the locally minimum variance unbiased quadratic estimator under normal distribution for $\underline{y}$. See also KLEFFE [21].

The question arises if MINQUEs are permissible estimators for $q'\theta$. LAMOTTE [23] proves that in variance components models the only individual variance component which can be estimated unbiasedly by a non-negative quadratic function $\underline{y}' A \underline{y}$ is $\sigma_k^2$, and even $\sigma_k^2$ is so estimable only if all $V_i$, $i = 1, \ldots, k-1$ are singular. Further only for $q$ with all $q_i$ non-negative, $q'\theta$ may be estimated unbiasedly by a non-negative quadratic $\underline{y}' A \underline{y}$. For the balanced one-way layout of example (iv), $q_1\sigma_1^2 + q_2\sigma_2^2$ has an unbiased estimator $\underline{y}' A \underline{y}$ with $A$ non-negative (definite) if and only if $q_2 \geq n_2^{-1} q_1 \geq 0$. Clearly $\sigma_2^2$ and $\sigma_2^2 + n_2\sigma_1^2$ can be estimated unbiasedly by a non-negative quadratic, but $\sigma_1^2$ cannot.

## 3 Least squares estimators

For a non-singular linear model $\underline{y} = X\beta + \underline{e}$, where $y, e \in \mathbb{R}^n$, $X$ is an $nxp$-matrix of rank $p$, the parameter vector $\beta \in \mathbb{R}^p$, $\mathrm{E}(\underline{e}) = 0$, $\mathrm{VAR}(\underline{e}) = \sigma^2 V$ and $V$ positive definite, it is known that the least squares estimator for $\beta$, $\hat{\underline{\beta}} = (X'X)^{-1} X' \underline{y}$, is unbiased or, if $V$ is known, the best linear unbiased estimator or Gauss-Markov estimator is $\hat{\underline{\beta}} = (X' V^{-1} X)^{-1} X' V^{-1} \underline{y}$.

Focussing our attention on variance component estimation we have the linear model given above with

$$\sigma^2 V = \sigma^2 (\gamma_1 V_1 + \ldots + \gamma_{k-1} V_{k-1} + I_n),$$

with variance components $\sigma_i^2 = \sigma^2 \gamma_i$, $i = 1, \ldots, k-1$ and $\sigma^2 = \sigma_k^2$.

The least squares approach to estimate variance components or linear combinations of variance components has been used by SEELY [40, 41]. Instead of quadratic functions of $\underline{y}$, SEELY utilizes linear functions of $\underline{Z} = \underline{y}\underline{y}'$ in $\mathcal{A}$, a vector space of symmetric $nxn$-

matrices, with inner product of $A,B \in \mathcal{A}$ defined by $(A,B) = \mathrm{tr}(AB)$. The linear estimator for variance components is $(A,\underline{Z})$. The expected value equals.

$$E(\underline{Z}) = \sum_{i=1}^{p} \sum_{j=1}^{p} \beta_i \beta_j B_{ij} + \sum_{i=1}^{k} \sigma_i^2 V_i, \quad \text{with } i \leqslant j$$

where $B_{ii} = x_i x_i'$ and $B_{ij} = x_i x_j' + x_j x_i'$ for $i < j$ is symmetric and $x_j$ is the $j$-th column of $X$, i.e.

$$E(\underline{Z}) = X \Psi X' + \sum_{i=1}^{k} \sigma_i^2 V_i,$$

with $\Psi = \beta\beta'$. Hence $E(\underline{Z})$ belongs to $\{\text{span } \mathcal{B}_0 + \text{span } \mathcal{B}_1\}$ with $\mathcal{B}_0 = \{B_{ij}; 1 \leqslant i \leqslant j \leqslant p\}$ and $\mathcal{B}_1 = \{V_1, V_2, \ldots, V_k\}$. To get rid of the nuisance parameter $\Psi$ we use the linear operator $\mathbb{I} - \mathbb{P}$ where $\mathbb{I}$ is the identity mapping of $\mathcal{A}$ to $\mathcal{A}$, $\mathbb{P}$ is the projection operator on $\mathcal{B}_0$. In this case $\mathbb{P}(A) = PAP$, where $P$ is the orthogonal projection on the column space of $X$, i.e. $P = X(X'X)^{-1}X'$, and

$$(\mathbb{I} - \mathbb{P})(A) = A - PAP.$$

The unbiased estimation of $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ is possible if the equations

$$\begin{bmatrix} (V_1,(\mathbb{I}-\mathbb{P})V_1) & \ldots & (V_1,(\mathbb{I}-\mathbb{P})V_k) \\ \vdots & & \vdots \\ (V_k,(\mathbb{I}-\mathbb{P})V_1) & \ldots & (V_k,(\mathbb{I}-\mathbb{P})V_k) \end{bmatrix} \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_k \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{bmatrix}$$

or $(\mathrm{tr}(V_i(V_j - PV_jP))\,\rho = \lambda$ have a solution. The unbiased estimator for $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ is then given by

$$\sum_{i=1}^{k} \rho_i((\mathbb{I} - \mathbb{P})V_i, \underline{Z}).$$

SEELY uses the operator $\pi(A) = \frac{1}{2}(\{(I - P)A + A(I - P)\}$ to get rid of $\Psi$ and therefore we must solve the normal equation

$$\begin{bmatrix} \mathrm{tr}(V_1(I-P)V_1) & \ldots & \mathrm{tr}(V_1(I-P)V_k) \\ \vdots & & \vdots \\ \mathrm{tr}(V_k(I-P)V_1) & \ldots & \mathrm{tr}(V_k(I-P)V_k) \end{bmatrix} \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_k \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{bmatrix},$$

and the unbiased estimator for $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ is given by $\sum_{i=1}^{k} \rho_i(\underline{y}'(I - P)V_i\underline{y})$.

Note that these estimators are completely analogous to the estimators in the linear

model $\underline{y} = X\beta + \underline{e}$, where $X = (X_1, X_2)$ with $X_1$ and $X_2$ of full column rank and $\beta = (\beta_1', \beta_2')'$. If we wish to get rid of the nuisance parameters $\beta_1$ to estimate $\beta_2$ unbiasedly, we may write $E\underline{y} = X\beta$ in the form

$$E\underline{y} = X_1\beta_1 + X_2\beta_2 = X_1\beta_1 + PX_2\beta_2 + (I - P)X_2\beta_2 =$$
$$= X_1\alpha + (I - P)X_2\beta_2$$

where $P$ is the orthogonal projector on the column space of $X_1$, $P = X_1(X_1'X_1)^{-1}X_1'$. A linear combination $(\lambda,\beta_2)$ of parameters in $\beta_2$ is identifiable if there is a $\rho$ such that $X_2'(I - P)X_2\rho = \lambda$. From the normal equations we estimate $\beta_2$ as the solution of

$$X_2'(I - P)X_2\,\hat{\beta}_2 = X_2'(I - P)\underline{y}$$

and the linear unbiased estimator for $(\lambda,\beta_2) = \Sigma\lambda_i\beta_{2i}$ is given by

$$((I - P)X_2\rho,\underline{y}) = \Sigma\rho_i((I - P)X_{2i},\underline{y})$$

where $X_{2i}$ are the columns of $X_2$.

*Example*

Let us derive the least squares unbiased estimator for $\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2$ in example 2 of section 2.1. The orthogonal projector $P = X(X'X)^{-1}X'$ and the matrices $(I - P)V_i = V_i - PV_iP$ are given by

$$P = \tfrac{1}{3}\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}, \quad V_1 - PV_1P = \tfrac{1}{3}\begin{pmatrix} 2 & -1 & -1 & 2 \\ -1 & 2 & -1 & -1 \\ -1 & -1 & 2 & -1 \\ 2 & -1 & -1 & 0 \end{pmatrix} \quad V_2 - PV_2P = \tfrac{1}{3}\begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & -1 & 0 \\ -1 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Since $(V_i,(I - P)V_j) = \mathrm{tr}V_i(V_j - PV_jP)$, we must solve the following set of equations in $\rho_1$ and $\rho_2$:

$$\begin{pmatrix} 10/3 & 2 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \Rightarrow \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} = \tfrac{1}{4}\begin{pmatrix} 3 & -3 \\ -3 & 5 \end{pmatrix}\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}$$

Because

$$((I - P)V_i\underline{y},\underline{y}') = \underline{y}'(V_i - PV_iP)\underline{y}$$

the unbiased estimator for $\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2$ is

$$\rho_1 \underline{y}'(V_1 - PV_1P)\underline{y} + \rho_2 \underline{y}'(V_2 - PV_2P)\underline{y} =$$

$$\rho_1 \{\tfrac{2}{3}(\underline{y}_{11}^2 + \underline{y}_{12}^2 + \underline{y}_{13}^2 - \underline{y}_{11}\underline{y}_{12} - \underline{y}_{11}\underline{y}_{13} + 2\underline{y}_{11}\underline{y}_{21} - \underline{y}_{12}\underline{y}_{13} - \underline{y}_{12}\underline{y}_{21} - \underline{y}_{13}\underline{y}_{21})\} +$$

$$+ \rho_2 \{\tfrac{2}{3}(\underline{y}_{11}^2 + \underline{y}_{12}^2 + \underline{y}_{13}^2 - \underline{y}_{11}\underline{y}_{12} - \underline{y}_{11}\underline{y}_{13} - \underline{y}_{12}\underline{y}_{13})\}.$$

The unbiased estimator for $\sigma_1^2$ is

$$\hat{\underline{\sigma}}_1^2 = \tfrac{1}{2}(2\underline{y}_{11}\underline{y}_{21} - \underline{y}_{12}\underline{y}_{21} - \underline{y}_{13}\underline{y}_{21})$$

and the unbiased estimator for $\sigma_2^2$ is:

$$\hat{\underline{\sigma}}_2^2 = \tfrac{1}{6}(2\underline{y}_{11}^2 + 2\underline{y}_{12}^2 + 2\underline{y}_{13}^2 - 2\underline{y}_{11}\underline{y}_{12} - 2\underline{y}_{11}\underline{y}_{13} - 6\underline{y}_{11}\underline{y}_{21} - 2\underline{y}_{12}\underline{y}_{13} + 3\underline{y}_{12}\underline{y}_{21} + 3\underline{y}_{13}\underline{y}_{21}).$$

## 4 MINQUE, a generalised least squares estimator

We give now the derivation of MINQUE as a generalised least squares estimator. If we have the idea that the variance-covariance matrix of $\underline{y}$

$$V = \sum_{i=1}^{k} \sigma_i^2 V_i$$

is of the form

$$T = \sum_{i=1}^{k} \alpha_i^2 V_i,$$

or, in other words, if $\alpha_i^2$ is the a priori (or approximate) value of $\sigma_i^2$, then we may transform $\underline{y}$ into $\tilde{\underline{y}} = T^{-\frac{1}{2}}\underline{y}$ with $E\tilde{\underline{y}} = T^{-\frac{1}{2}}X\beta$ and VAR $\tilde{\underline{y}} = T^{-\frac{1}{2}}VT^{-\frac{1}{2}} =$

$$\sum_{i=1}^{k} \sigma_i^2 T^{-\frac{1}{2}} V_i T^{-\frac{1}{2}} = \sum_{i=1}^{k} \sigma_i^2 \tilde{V}_i.$$

When $T$ is given, the unbiased estimator for $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ is

$$\sum_{i=1}^{k} \rho_i((I - \tilde{P})\tilde{V}_i, \tilde{\underline{Z}})$$

where $\tilde{\underline{Z}} = \tilde{\underline{y}}\tilde{\underline{y}}'$, $(I - \tilde{P})(A) = A - \tilde{P}A\tilde{P}$ with $\tilde{P} = \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$ and $\rho$ is the solution of the matrix equation

$$\{(\tilde{V}_i, (I - \tilde{P})\tilde{V}_j)\}\rho = \lambda.$$

Using the properties of the trace operator $\mathrm{tr}(ABC) = \mathrm{tr}(BCA)$ we find that the unbiased estimator for

$$\sum_{i=1}^{k} \lambda_i \sigma_i^2$$

can be written as $\underline{y}'A_*\underline{y}$ where

$$A_* = \sum_{i=1}^{k} \rho_i T^{-1}(V_i - P_T V_i P_T')T^{-1},$$

in which $P_T = X(X'T^{-1}X)^{-1}X'T^{-1}$ and $\rho_1, \ldots, \rho_k$ satisfy

$$\sum_{i=1}^{k} \rho_i \operatorname{tr} T^{-1}(V_i - P_T V_i P_T')T^{-1}V_j = \lambda_j \quad \text{for } j = 1, \ldots, k.$$

This is the estimator which RAO [34, 35, 36, 37] introduced as the minimum norm quadratic unbiased estimator (MIN QUE) if we minimize tr(ATAT') only instead of tr(ATAT) +2 tr(AXX'AT).

If we want an estimator for $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ which is invariant under translation of $\beta$ we consider $\beta_d = \beta - \beta_0$, and put $\underline{y}_d = \underline{y} - X\beta_0$. Then the invariant estimator satisfies

$$\underline{y}_d' A \underline{y}_d = \underline{y}' A \underline{y}$$

for all $\beta_0$, which is equivalent to $AX = 0$. From the derivation of the unbiased estimator $\underline{y}'A\underline{y}$ without invariance

$$\mathrm{E}(\underline{y}'A\underline{y}) = \operatorname{tr} AV + \beta' X'AX\beta = \sum_{i=1}^{k} \sigma_i^2 \operatorname{tr}(AV_i) + \beta' X'AX\beta,$$

it follows that necessary and sufficient conditions for $\underline{y}'A\underline{y}$ to be unbiased for $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ are

$$\left.\begin{aligned} &X'AX = 0 \\ &\operatorname{tr}(AV_i) = \lambda_i, \quad i = 1, \ldots, k \end{aligned}\right\}$$

We thus see that invariance implies the first condition of unbiasedness. The conditions for an unbiased invariant estimator $\underline{y}'A\underline{y}$ for $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ are:

$$\left.\begin{aligned} &AX = 0 \\ &\operatorname{tr}(AV_i) = \lambda_i, \quad i = 1, \ldots, k \end{aligned}\right\}$$

SEELY [44] shows that $Q'\underline{y}$ is a maximal invariant statistic from $\mathbb{R}^n$ to $\mathbb{R}^q$, where $Q$ is an $n \times q$-matrix with $q = n - \text{rank } X$, column space of $Q$ = null space of $X'$ and $Q'Q = I_q$, i.e. $QQ' = I - P$, where $P$ is the orthogonal projector on the column space of $X$. Now $\mathrm{E}(Q'\underline{y}) = Q'X\beta = 0$ and $\mathrm{VAR}(Q'\underline{y}) = Q'VQ = \sum_{i=1}^{k} \sigma_i^2 Q'V_iQ$.

The set of normal equations for the least-squares estimator of $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ is now given by

$$\sum_{i=1}^{k} \rho_i (Q'V_iQ, Q'\underline{y}\underline{y}'Q)$$

where the $\rho_i$'s satisfy the matrix equation

$$\{(Q'V_iQ, Q'V_jQ)\}\rho = \lambda.$$

Using the properties of trace operators we find that the invariant unbiased estimator for $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ can be written as

$$\underset{\sim}{y}'A_*\underset{\sim}{y}$$

where

$$A_* = \sum_{i=1}^{k} \rho_i (I - P)V_i(I - P)$$

and $\rho_1, \ldots, \rho_k$ satisfy

$$\sum_{i=1}^{k} \rho_i \operatorname{tr}(I - P)V_i(I - P)V_j = \lambda_j \quad \text{for } j = 1, \ldots, k.$$

If we have the idea that $V$ is of the form $T$ we first transform $\underset{\sim}{y}$ to

$$\tilde{\underset{\sim}{y}} = T^{-\frac{1}{2}}\underset{\sim}{y} \quad \text{with} \quad E\tilde{\underset{\sim}{y}} = T^{-\frac{1}{2}}X\beta = \tilde{X}\beta$$

and

$$\operatorname{VAR}(\tilde{\underset{\sim}{y}}) = \sum_{i=1}^{k} \sigma_i^2 T^{-\frac{1}{2}} V_i T^{-\frac{1}{2}} = \sum_{i=1}^{k} \sigma_i^2 \tilde{V}_i.$$

The invariant statistic now is $\tilde{Q}\tilde{\underset{\sim}{y}}$ where $\tilde{Q}\tilde{Q}' = I - \tilde{P}$ and $\tilde{P} = \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$. The invariant unbiased estimator for $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ is then given by

$$\sum_{i=1}^{k} \rho_i (\tilde{Q}'\tilde{V}_i\tilde{Q}, \tilde{Q}'\tilde{\underset{\sim}{y}}\tilde{\underset{\sim}{y}}'\tilde{Q})$$

where the $\rho_i$'s satisfy the matrix equations

$$\{(\tilde{Q}'\tilde{V}_i\tilde{Q}, \tilde{Q}'\tilde{V}_j\tilde{Q})\}\rho = \lambda.$$

The invariant unbiased estimator for $\sum_{i=1}^{k} \lambda_i \sigma_i^2$, given $T = \sum_{i=1}^{k} \alpha_i^2 V_i$ can then be written as $\underset{\sim}{y}'A_*\underset{\sim}{y}$ where

$$A_* = \sum_{i=1}^{k} \rho_i (I - P_T')T^{-1}V_iT^{-1}(I - P_T).$$

in which $\rho_1, \ldots, \rho_k$ satisfy

$$\sum_{i=1}^{k} \rho_i \operatorname{tr}(I - P_T')T^{-1}V_iT^{-1}(I - P_T)V_j = \lambda_j \quad \text{for} \quad j = 1, \ldots, k,$$

and in which

$$P_T = X(X'T^{-1}X)^{-1}X'T^{-1}.$$

This is RAO's invariant MINQUE estimator. When $\underset{\sim}{y}$ has an $n$-variate normal distribution,

$\text{var}(\underline{y}'A\underline{y}) = 2\text{tr}(AVAV) + 4\beta'X'AVAX\beta = 2\text{tr}(AVAV)$ since $AX = 0$. Thus if the true variance components are $\alpha_1^2, \ldots, \alpha_k^2$, $\text{tr}(ATAT)$ is proportional to $\text{var}(\underline{y}'A\underline{y})$ and the invariant MINQUE is the locally minimum variance unbiased quadratic estimator under normal distribution for $\underline{y}$. See also KLEFFE [21].

Finally we remark that SEELY [43] proves that when $\underline{y}$ is normally distributed, for the class of unbiased invariant estimators (thus $AX = 0$), the best quadratic unbiased estimator for $\sum_{i=1}^{k} \lambda_i\sigma_i^2$ is a function of $Q'\underline{y}$ if and only if

$$\mathscr{B} = \text{span}\{Q'V_1Q, \ldots, Q'V_kQ\}$$

is a quadratic subspace of all real symmetric $q \times q$ matrices $\mathscr{A}$. A subspace $\mathscr{B}$ of $\mathscr{A}$ with the property that $B \in \mathscr{B}$ implies that $B^2 \in \mathscr{B}$ is said to be a quadratic subspace of $\mathscr{A}$. SEELY [43, 44, 45] used the notion of quadratic subspaces also in the derivation of completeness of certain statistics for a family of multivariate normal distributions. For a survey of invariant methods for estimating variance components, where the restriction of non-negativity of the quadratic estimator has not been taken into account, see KLEFFE [22].

## 5 Non-negative estimators

(i) At the end of section 2 the question arises of MINQUE's are permissible estimators for $q'\theta$, where $\theta = (\sigma_1^2, \ldots, \sigma_k^2)'$. The answer was that many variance components were estimated by MINQUE which were not non-negative estimators. RAO [36] indicates a modification of MINQUE which would provide non-negative estimators, but the resulting estimators would generally be neither quadratic nor unbiased. BROWN [2] describes an iterative feedback procedure using residuals which ensures non-negative estimation of variance components. See also RAO and CHAUBEY [38].

(ii) For a linear model $\underline{y} = X\beta + \underline{e}$, where $y, e \in \mathbb{R}^n$, $X$ is an $n \times p$-matrix of rank $p$, the parameter vector $\beta \in \mathbb{R}^p$, $E(\underline{e}) = 0$, $VAR(\underline{e}) = \sigma^2 V$ and $V$ positive definite, it is known that the least squares estimator for $\beta$, $\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y}$, is unbiased or, if $V$ is known, the best linear unbiased estimator or Gauss-Markov estimator is $\hat{\underline{\beta}} = (X'V^{-1}X)^{-1}X'V^{-1}\underline{y}$.

Sometimes order restrictions are imposed on the parameters $\beta_i$ in $\beta = (\beta_1, \ldots, \beta_p)'$. The estimation of $\beta_i$ must e.g. be order preserving or isotonic. For permissible estimators of $\beta$ see BARLOW et al. [1].

In the case of variance component estimation we have the linear model given above with

$$\sigma^2 V = \sigma^2(\gamma_1 V_1 + \ldots + \gamma_{k-1} V_{k-1} + I_n),$$

with variance components $\sigma_i^2 = \sigma^2\gamma_i$, $i = 1, \ldots, k-1$ and $\sigma^2 = \sigma_k^2$.

Isotonic variance components estimators are given for the balanced case by THOMPSON [46] and THOMPSON and MOORE [47]. See also BARLOW et al. [1].

(iii) In the linear model $\underline{y} = X\beta + \underline{e}$ sometimes $\beta$ belongs to the positive orthant $\mathbb{R}_+^p$ of $\mathbb{R}^p$. A permissible estimator $\tilde{\underline{\beta}}$ of $\beta$, better than the one found by truncation (i.e. replacing

negative estimates of $\beta_i$ by zero), is provided by the solution of the quadratic programming problem:

$$\text{minimize } \| y - Xb \|, \qquad b_i \geqslant 0, \quad (i = 1, \ldots, p),$$

or, if $V$ is known,

$$\text{minimize } \| V^{-\frac{1}{2}} y - V^{-\frac{1}{2}} Xb \| \quad b_i \geqslant 0, \quad (i = 1, \ldots, p).$$

If there are inequality constraints for some $\beta_i$, $\beta$ must belong to a convex subset of $\mathbb{R}^p$ and we again have a quadratic programming problem. See e.g. JUDGE and TAKAYAMA [18], MANTEL [25], WATERMAN [50] and LIEW [24].

In the case of the variance components model we discussed in section 3 the least squares approach of SEELY to estimate the variance components $\sigma_1^2, \ldots, \sigma_k^2$. In his procedure SEELY has not taken into consideration the non-negativity of $\sigma_1^2, \ldots, \sigma_k^2$. Imposing this restriction on the parameters means that we may again use the quadratic programming approach in order to ensure that the estimators for $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ be non-negative.

The author developed independently of others the procedure to find non-negative variance components estimators just as in the case of a postive $\beta$ in the linear model $\underline{y} = X\beta + \underline{e}$. For solving the related quadratic programming problem we use the algorithm of WATERMAN [50]. Consider $Q'\underline{y}$, which is a maximal invariant statistic from $\mathbb{R}^n$ to $\mathbb{R}^q$, where $Q$ is an $n \times q$ matrix with $q = n - \text{rank } X$, column space of $Q$ = null space of $X'$ and $Q'Q = I_q$, i.e. $QQ' = I - P$, where $P = X(X'X)^{-1}X'$ is the orthogonal projector on the column space of $X$. Let $\underline{Z} = (Q'\underline{y})(Q'\underline{y})'$, then $\mathrm{E}\underline{Z} = \sum_{i=1}^{k} \sigma_i^2 Q'V_iQ$. Let the index set $I$ range over all $2^k$ subsets of $\{1, 2, \ldots, k\}$. For each such set we minimize $\| Z - \widehat{\mathrm{E}(Z)} \|$ and obtain the unrestricted solution. If all the estimates of the parameters of $\sigma_i^2$ belonging to a certain set $J \in I$ are non-negative, we obtain $\| Z - \{\widehat{\mathrm{E}(Z)} \mid \text{corresponding to } J\} \|$. The minimum of these norms solves the problem for we then set all those parameters $\sigma_i^2$ equal to zero which do not belong to the minimal set of $\widehat{\mathrm{E}(Z)}$.

PUKELSHEIM [32, 33] investigated the unbiased estimation of variance components by means of non-negative quadratics, using convex programming.

Another interesting development is the use of restricted generalized inverse operators such as given by HARTUNG [12]. He considers minimum bias estimators (as introduced by CHIPMAN [3] for estimating the mean value parameter $\beta$), which are invariant under the group of translations of $\beta$, where the minimum is taken over the appropriate class of positive semi-definite matrices after a reduction by invariance. These estimators always exist. They are of course non-negative and are unbiased if non-negative unbiased quadratic estimation of $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ is given. If these estimates are not unique, the one with minimum norm has to be taken. The minimum norm minimum bias invariant positive semi-definite estimator has been characterized by introducing a non-linear cone-restricted

pseudo inverse, for which HARTUNG gave a representation, allowing computation without consideration of the boundary.

## 6 Maximum likelihood estimators

Another estimator which is always permissible is the proper maximum likelihood estimator (MLE), which yields a point in the parameter set $\Theta$ at which the likelihood function attains an absolute maximum. It should be noted that in many situations the estimators obtained by merely solving the likelihood equations, i.e. setting first derivatives of the likelihood equal to zero, will be impermissible. Solving likelihood equations while taking into account the restrictions imposed by the parameter set $\Theta$ will give the proper MLE.

For the general variance components model with normal vector $\underline{y}$ the likelihood is given by

$$l = (2\pi)^{-n/2}(\det(V))^{-\frac{1}{2}}\exp\{-(y - X\beta)'V^{-1}(y - X\beta)/2\}$$

or

$$\log l = L(\theta,\beta,y) = -(n/2)\log 2\pi - \tfrac{1}{2}\log\det(V) - (y - X\beta)'V^{-1}(y - X\beta)/2,$$

defined for $\theta = (\sigma_1^2, \ldots, \sigma_k^2)'$ and $\beta$ such that $\sigma_k^2 > 0$, $\sigma_i^2 \geqslant 0$, $i = 1, \ldots, k-1$. We write $V$ as

$$V = \sigma_k^2 H = \sigma_k^2(\gamma_1 V_1 + \ldots + \gamma_{k-1}V_{k-1} + I_n)$$

where $\gamma_i = \sigma_i^2/\sigma_k^2$, $i = 1, \ldots, k-1$. Hence

$$L(\theta,\beta,y) = -(n/2)\log 2\pi - (n/2)\log\sigma_k^2 - \tfrac{1}{2}\log\det(H) +$$
$$-(y - X\beta)'H^{-1}(y - X\beta)/(2\sigma_k^2).$$

The likelihood equations are

$$\frac{\partial L}{\partial \beta} = \sigma_k^{-2}\{X'H^{-1}y - (X'H^{-1}X)\beta\} = 0$$

$$\frac{\partial L}{\partial \sigma_k^2} = -\frac{n}{2\sigma_k^2} + \frac{1}{2\sigma_k^4}(y - X\beta)'H^{-1}(y - X\beta) = 0$$

$$\frac{\partial L}{\partial \gamma_i} = -\tfrac{1}{2}\mathrm{tr}\left(H^{-1}\frac{\partial H}{\partial \gamma_i}\right) - \frac{1}{2\sigma_k^2}(y - X\beta)'\frac{\partial H^{-1}}{\partial \gamma_i}(y - X\beta)$$

$$= -\tfrac{1}{2}\mathrm{tr}(H^{-1}V_i) + \frac{1}{2\sigma_k^2}(y - X\beta)'H^{-1}V_iH^{-1}(y - X\beta) = 0$$

for $i = 1, \ldots, k-1$.

We obtain

$$\hat{\beta} = (X'H^{-1}X)^{-1}(X'H^{-1}y)$$

and

$$\hat{\sigma}_k^2 = y'R'H^{-1}Ry/n$$

where

$$R = I - X(X'H^{-1}X)^{-1}X'H^{-1}.$$

Inserting this in $\partial L/\partial\gamma_i = 0$ we get $k-1$ equations for the $k-1$ values of $\gamma_i$:

$$y'R'H^{-1}V_iH^{-1}Ry/\hat{\sigma}_k^2 = \mathrm{tr}(H^{-1}V_i) \quad (i = 1, \ldots, k-1)$$

or, if we put $V_i = U_iU_i'$,

$$y'R'H^{-1}U_iU_i'H^{-1}Ry/\hat{\sigma}_k^2 = \mathrm{tr}(U_iH^{-1}U_i').$$

For the balanced one-way layout the ML estimators are given in explicit form by HERBACH [17]:

$$\left.\begin{aligned}\hat{\sigma}_2^2 &= \mathrm{MSE}\\ \hat{\sigma}_1^2 &= \{(1-n_1^{-1})\mathrm{MSA} - \mathrm{MSE}\}/n_2\end{aligned}\right\} \quad \text{if } (1-n_1^{-1})\mathrm{MSA} \geqslant \mathrm{MSE}$$

$$\left.\begin{aligned}\hat{\sigma}_2^2 &= \{(n_1-1)\mathrm{MSA} + n_1(n_2-1)\mathrm{MSE}\}/(n_1n_2)\\ \hat{\sigma}_1^2 &= 0\end{aligned}\right\} \quad \text{otherwise.}$$

HARTLEY and RAO [13] gave an iterative procedure for solving the estimation equations above under the restriction $\gamma_i \geqslant 0$, $i = 1, \ldots, k-1$. For a simplification of the procedure see HEMMERLE and HARTLEY [15].

If one believes in unbiased estimators, then a criticism of the ML approach for estimating $\theta$ is that it does not take into account the loss in degrees of freedom due to estimating $\beta$. In the particular case $\underline{y} = X\beta + \underline{e}$, i.e. with $k = 1$, $V = \sigma_1^2 I_n$ and $\sigma_1^2 > 0$, the ML estimator for the single variance component $\sigma_1^2$ is

$$\underline{\hat{\sigma}}_1^2 = (\underline{y} - X\hat{\beta})'(\underline{y} - X\hat{\beta})/n$$

where

$$\hat{\beta} = X(X'X)^{-1}X'\underline{y}, \quad \mathrm{E}(\underline{\hat{\sigma}}_1^2) = \sigma_1^2(n-p)/n,$$

hence the estimator $\underline{\hat{\sigma}}_1^2$ is biased downwards.

The restricted maximum likelihood (REML) or modified maximum likelihood approach of PATTERSON and THOMPSON [28, 29] maximises the likelihood, not of the complete vector $\underline{y}$, but of all error contrasts. Any linear combination $u'\underline{y}$ of the observations such that $\mathrm{E}(u'\underline{y}) = 0$, i.e. such that $u'X = 0$ with $u$ independent of $\theta$ or $\beta$ is an error contrast. Each such vector belongs to the column space of the singular matrix $S = I - X(X'X)^{-1}X'$.

Since $SX$ is null, $S\underline{y}$ is singular $N(0,\sigma_k^2 SHS)$. The maximum possible number of linearly independent error contrasts is $n-p$. A particular set of $n-p$ linearly independent error contrasts is given by $T\underline{y}$, where $T$ is an $(n-p)\times n$-matrix whose rows are any $n-p$ linearly independent rows of $S$, and $TX=0$. Now the transformation

$$\underline{z} = \begin{pmatrix} T \\ X'H^{-1} \end{pmatrix} \underline{y} \simeq N \left( \begin{pmatrix} 0 \\ X'H^{-1}X\beta \end{pmatrix}, \begin{pmatrix} TH\,T'\sigma_k^2 & 0 \\ 0 & X'H^{-1}X\sigma_k^2 \end{pmatrix} \right),$$

is non-singular, because both $X'$ and $T$ have full rank, while the rows of $T$ are linearly independent of those of $X'$. The log likelihoods of $T\underline{y}$ and $X'H^{-1}\underline{y}$ are respectively:

$$L_1 = -\tfrac{1}{2}(n-p)\log(2\pi) - \tfrac{1}{2}(n-p)\log\sigma_k^2 - \tfrac{1}{2}\log\det(THT') +$$

$$- \frac{1}{2\sigma_k^2}\, y'T'(THT')^{-1}Ty$$

and

$$L_2 = -\tfrac{1}{2}p\log(2\pi) - \tfrac{1}{2}p\log\sigma_k^2 - \tfrac{1}{2}\log\det(X'H^{-1}X) +$$

$$- \frac{1}{2\sigma_k^2}\,(y-X\beta)'H^{-1}X(X'H^{-1}X)^{-1}X'H^{-1}(y-X\beta).$$

Differentiation of $L_1$ gives:

$$\frac{\partial L_1}{\partial \sigma_k^2} = -\frac{1}{2\sigma_k^2}\,(n-p) + \frac{1}{2\sigma_k^4}\, y'T'(THT')^{-1}Ty$$

and

$$\frac{\partial L_1}{\partial \gamma_i} = -\tfrac{1}{2}\mathrm{tr}(U_i'T'(THT')^{-1}TU_i) + \frac{1}{2\sigma_k^2}\, y'T'(THT')^{-1}TU_iU_i'T'(THT')^{-1}Ty$$

for $i = 1, \ldots, k-1$.

Setting these derivatives equal to zero we get the restricted maximum likelihood (REML) estimators. An iterative procedure, starting with initial values for $\gamma = (\gamma_1, \ldots, \gamma_{k-1})$, requires solving $\hat{\sigma}_k^2 = y'T'(THT')^{-1}Ty/(n-p)$ and calculating new $\gamma_i$-values which bring $\partial L_1/\partial\gamma_i$ closer to zero. For the numerical procedure see CORBEIL and SEARLE [4].

PATTERSON and THOMPSON [29] mentioned that the MINQUE procedure is equivalent to a single iteration of REML. Further, for balanced variance components models the REML estimators are the usual analysis of variance estimators. They used these properties as a justification of their procedure. But these properties imply that the REML procedure of PATTERSON and THOMPSON gives impermissible estimators. The authors do not use a proper maximum likelihood procedure because they only set the likelihood derivatives $\partial L_1/\partial\sigma_k^2$ and $\partial L_1/\partial\gamma_i$ $i = 1, \ldots, k-1$ equal to zero, without taking into consideration the constraints $\gamma_i \geqslant 0$ $(i = 1, \ldots, k-1)$.

Also in the comparisons of variance components estimators of CORBEIL and SEARLE [4] they used the improper REML-procedure ignoring these non-negativity constraints. Therefore their results are questionable.

For a thorough discussion of maximum likelihood approaches to variance component estimation see HARVILLE [14], in which it is argued that the correct REML procedure makes sense. From the comment by J. N. K. RAO on HARVILLE's article we mention the warning concerning the use of ML-procedures, bacause none of the proposed algorithms guarantees a solution which is indeed ML.

The behaviour of the likelihood as a function of the variance components appears to be complex; even for the simple unbalanced one-way layout, the likelihood equation may have multiple roots, or the ML-estimate may be a boundary point rather than a root.

Another procedure to get maximum likelihood estimates for variance components is given by MILLER [27]. The asymptotic properties of these maximum likelihood estimates are given by MILLER [26].


## Acknowledgement

The author wants to thank Prof. Dr. W. Schaafsma for many helpful suggestions to improve the original draft.


## References


[1] BARLOW, R. E., D. J. BARTHOLOMEW, J. M. BREMNER, H. D. BRUNK (1972), Statistical inference under order restrictions; the theory and applications of isotonic regression. Wiley & Sons, New York.

[2] BROWN, K. G. (1978), Estimation of variance components using residuals, Journal of the American Statistical Association 73, 141–146.

[3] CHIPMAN, J. S. (1964), On least squares with insufficient observations, Journal of the American Statistical Association 59, 1078–1111.

[4] CORBEIL, R. R. and S. R. SEARLE (1976), Restricted maximum likelihood (REML) estimation of variance components in the mixed model, Technometrics 18, 31–38.

[5] CORBEIL, R. R. and S. R. SEARLE (1976), A comparison of variance component estimators, Biometrics 32, 779–791.

[6] CORSTEN, L. C. A. (1964), A different solution of a problem posed by THEIL and SCHWEITZER, Statistica Neerlandica 18, 15–18. Correction, Statistica Neerlandica 18, 307.

[7] FERGUSON, T. S. (1967), Mathematical statistics, a decision theoretic approach. Academic Press, New York.

[8] GRAYBILL, F. A. (1954), On quadratic estimates of variance components, Annals of Mathematical Statistics 25, 367–372.

[9] GRAYBILL, F. A. and A. W. WORTHAM (1956), A note on uniformly best unbiased estimators for variance components, Journal of the American Statistical Association 51, 266–268.

[10] GRAYBILL, F. A. (1961), An introduction to linear statistical models, I, McGraw-Hill, New York.

[11] HARTUNG, J. (1978), On non-negative estimation of variance components, Contributed paper, presented at the 11-th European Meeting of Statisticians, Oslo, August 14–18.

[12] HARTUNG, J. (1978), Non-negative minimum biased invariant estimation in variance components models, Submitted for publication.

[13] HARTLEY, H. O. and J. N. K. RAO (1967), Maximum likelihood estimation for the mixed analysis of variance model, Biometrika 54, 93–108.

[14] HARVILLE, D. A. (1977), Maximum likelihood approaches to variance component estimation and to related problems, (with comment by J. N. K. RAO), Journal of the American Statistical Association 72, 320–340.
[15] HEMMERLE, W. J. and H. O. HARTLEY (1973). Computing maximum likelihood estimates for the mixed A.O.V. model using the W transformation, Technometrics 15, 819–831.
[16] HENDERSON, C. R. (1953), Estimation of variance and covariance components, Biometrics 9, 226–252.
[17] HERBACH, L. H. (1959), Properties of model II-type analysis of variance tests: A. Optimum nature of the F-test for model II in the balanced case, Annals of Mathematical Statistics 30, 939–959.
[18] JUDGE, G. G. and T. TAKAYAMA (1966). Inequality restrictions in regression analysis, Journal of the American Statistical Association 61, 166–181.
[19] KENDALL, M. G. and A. STUART (1973). The advanced theory of statistics, Vol. 2, Griffin, London.
[20] KLEFFE, J. (1975), Best quadratic unbiased estimators for variance components in the balanced two-way classification model, Mathematische Operationsforschung und Statistik 6, 189–196.
[21] KLEFFE, J. (1976), A note on MINQUE for normal models, Mathematische Operationsforschung und Statistik 7, 707–714.
[22] KLEFFE, J. (1977), Invariant methods for estimating variance components in mixed linear models, Mathematische Operationsforschung und Statistik, series Statistics, 8, 233–250.
[23] LAMOTTE, L. R. (1973). On non-negative quadratic unbiased estimation of variance components, Journal of the American Statistical Association 68, 728–730.
[24] LIEW, C. K. (1976), Inequality constrained least-squares estimation, Journal of the American Statistical Association 71, 746–751.
[25] MANTEL, N. (1969), Restricted least squares regression and convex quadratic programming, Technometrics 11, 763–773.
[26] MILLER, J. J. (1977), Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance, Annals of Statistics 5, 746–762.
[27] MILLER, J. J. (1979), Maximum likelihood estimation of variance components – a Monte Carlo study, Journal of Statistical Computation and Simulation 8, 175–190.
[28] PATTERSON, H. D. and R. THOMPSON (1971), Recovery of inter-block information when block sizes are unequal, Biometrika 58, 545–554.
[29] PATTERSON, H. D. and R. THOMPSON (1975), Maximum likelihood estimation of components of variance, Proceedings of the 8th International Biometric Conference, Constanta, ed. L. C. A. CORSTEN and T. POSTELNICU, 197–207.
[30] PRINGLE, R. M. (1974), Some results on the estimation of variance components by MINQUE, Journal of the American Statistical Association 69, 987–989.
[31] PUKELSHEIM, F. (1977), Punktschätzungen in linearen statistischen Modellen, Dissertation Freiburg im Breisgau.
[32] PUKELSHEIM, F. (1977), Linear models and convex programs: unbiased nonnegative estimation in variance component models, Technical report no. 104, Department of Statistics, Stanford University, Stanford California.
[33] PUKELSHEIM, F. (1978), Examples for unbiased nonnegative estimation in variance component models, Technical report no. 113, Department of Statistics, Stanford University, Stanford California.
[34] RAO, C. R. (1970), Estimation of heteroscedastic variances in linear models, Journal of the American Statistical Association 65, 161–172.
[35] RAO, C. R. (1971), Estimation of variance and covariance components-MINQUE theory, Journal of multivariate analysis 1, 257–275.
[36] RAO, C. R. (1972), Estimation of variance and covariance components in linear models, Journal of the American Statistical Association 67, 112–115.
[37] RAO, C. R. (1973), Linear statistical inference and its applications; second edition, Wiley & Sons, New York.
[38] RAO, P. S. R. S. and Y. P. CHAUBEY (1978), Three modifications of the principle of the MINQUE, Communications in Statistics, Theory and Methods, A7, 767–778.

[39] READ, R. R. (1961), On quadratic estimates of the interclass variance for unbalanced designs, Journal of the Royal Statistical Society B23, 493–497.
[40] SEELY, J. (1970), Linear spaces and unbiased estimation, Annals of Mathematical Statistics 41, 1725–1734.
[41] SEELY, J. (1970), Linear spaces and unbiased estimation – applications to the mixed linear model, Annals of Mathematical Statistics 41, 1735–1748.
[42] SEELY, J. and G. ZYSKIND (1971), Linear spaces and minimum variance unbiased estimation, Annals of Mathematical Statistics 42, 691–703.
[43] SEELY, J. (1971), Quadratic subspaces and completeness, Annals of Mathematical Statistics 42, 710–721.
[44] SEELY, J. (1972), Completeness for a family of multivariate normal distributions, Annals of Mathematical Statistics 43, 1644–1647.
[45] SEELY, J. (1977), Minimal sufficient statistics and completeness for multivariate normal families. Sankhyā A, 39, 170–185.
[46] THOMPSON, W. A. (1962), The problem of negative estimates of variance components, Annals of Mathematical Statistics 33, 273–289.
[47] THOMPSON, W. A. and J. R. MOORE (1963), Non-negative estimates of variance components, Technometrics 5, 441–450.
[48] VERDOOREN, L. R. (1980), How large is the probability for the estimate of a variance component to be negative? Technical Note 80–02, Department of Mathematics, Statistics Division, Agricultural University Wageningen.
[49] WASAN, M. T. (1970). Parametric estimation, McGraw-Hill, New York.
[50] WATERMAN, M. S. (1974), A restricted least squares problem, Technometrics 16, 135–136.

CHAPTER 6

# 6. HOW LARGE IS THE PROBABILITY FOR THE ESTIMATE OF A VARIANCE COMPONENT TO BE NEGATIVE?

Published in:

Biometrical Journal 24 (1982) no. 4, pp. 339 - 360.

Biom. J. vol. 24 · 1982 · no. 4 · pp 339–360

Department of Mathematics, Agricultural University Wageningen,
The Netherlands

# How Large is the Probability for the Estimate of a Variance Component to be Negative?

L. R. VERDOOREN

Abstract

For a balanced one-way classification, where the normally distributed observations obey a random model $y_{ij} = \mu + \underline{b}_i + \underline{e}_{ij}$ with two variance components var $(\underline{b}_i) = \sigma_1^2$ and var $(\underline{e}_{ij}) = \sigma_2^2$, the probability is given that the analysis of variance estimate of $\sigma_1^2$ will be negative. This probability depends on $\sigma_1^2/\sigma_2^2$ and the degrees of freedom in the ANOVA table. Tables for this probability are given. If the normally distributed observations obey an intra-class correlation model, the probability that the Mean Square between groups is smaller than the Mean Square within groups can also be evaluated from the given tables.

*Key words:* Variance components, probability of negative estimates, balanced random model, balanced intra-class correlation model.

## 1. Introduction

Often there occur situations in statistical practice which can be described in the form of a linear model with random effects. The main interest is directed towards the magnitude of the variances of the several sources of variation. The use of the well-known analysis of variance (ANOVA) estimators for the variance components sometimes produces negative estimates. A first idea, and a good one, which comes to mind is to check the calculations. When no errors are found in the computation, a second idea is to consider (check) more carefully the assumptions implied by the model.

In § 2 the assumptions of the simplest variance component model will be generated from the sampling procedure. We shall consider a balanced two-stage sampling procedure or a balanced two-stage nested design.

In § 3 the ANOVA estimator for the variance components for such a design will be described. When the model assumptions fit the practical situation, then there will often be doubt about the normality of the observations.

In § 4 it is shown that this doubt is not always justified, since there is a reasonably large probability for negative outcomes of the ANOVA estimator for a variance component.

In § 5 the intra-class correlation model will be described. This model can often be applied in situations where primarily a variance component model seemed appropriate.

When all the assumptions of the variance component model are fullfilled then it follows that the ANOVA estimator is not a good estimation procedure for variance components. A permissible estimator for a non-negative parameter as a variance component should be non-negative. An example of such a non-negative estimator is proposed in § 6.

## 2. The Construction of a Variance Component Model

A variance component model is shown to be generated by the sampling procedure of the observations. Let us first consider the following example. From a large bulk of natural phosphatic fertilizers one wishes to determine the average content $\mu$ of $P_2O_5$. One uses a balanced two-stage sampling procedure. In the first stage $n_1$ samples (for example bales) are drawn at random from the bulk. In the second stage $n_2$ subsamples are drawn from each sample. Chemical analysis of $P_2O_5$ is performed at each of the $n_1 n_2$ subsamples. The bulk of natural phosphatic fertilizer will not be completely homogeneous, hence the average content of $P_2O_5$ of samples will vary. The average content of the random samples from the bulk can be seen as random drawings from a probability distribution with variance $\sigma_1^2$. The chemical determination will not give completely reproducible analysis, hence the determinations of subsamples from samples will vary. The $P_2O_5$ content found from the analysis can be seen as random drawings from a probability distribution with variance $\sigma_2^2$. The goal of this investigation is to estimate the interesting parameters $\mu$, $\sigma_1^2$ and $\sigma_2^2$.

For a formal derivation of a random model for such a practical situation we consider a large population (= bulk). Note that in practice small random samples without replacement are drawn from a large population. This type of sampling can be considered, without any harm, as sampling with replacement. In a *balanced* two-stage nested design the same number $n_2$ of sub-samples is drawn in the second stage from each sample.

An element chosen at random from the population according to a nested sampling scheme in two stages has a composite subscript $(\underline{i}, \underline{j})$, where $i$ refers to the first sampling step, $\underline{j}$ refers to the second sampling step. (We denote a random variable $\underline{x}$ by underlining the symbol $x$; an observed value of this random variable is denoted by the symbol $x$ without a bar below).

Sampling starting from a specified element $i$ in the first stage is denoted by $(i, \underline{j})$. Conditional sampling of $\underline{j}$ given that $\underline{i} = i$ is identical to sampling $(i, \underline{j})$. The observation at the element obtained at the last sampling stage is denoted by $y_{i\underline{j}}$.

The random variable $y_{i\underline{j}}$ has expectation $\mathrm{E}(y_{i\underline{j}} \mid \underline{i} = i) = \mathrm{E}_{\underline{j}} y_{i\underline{j}} = \mu_i$, where ex-

pectation has been taken over the second sampling stage, given the result of the first sampling stage. Define the random variable $e_{ij}=y_{ij}-\mu_i$. The expectation of $e_{ij}$ is zero by definition. Now assume that the variance of $e_{ij}$ is the same for all $i$. It will be denoted as $\mathrm{var}_j\,(e_{ij})=\sigma_2^2$.

Next define $\mathrm{E}_i\,(\mu_i)=\mu$ and the random variable $e_i=\mu_i-\mu$ with expectation zero and variance $\sigma_1^2$.

It follows that

$$y_{ij}=\mu_i+e_{ij}=\mu+e_i+e_{ij}$$

where $\mu_i=\mu+e_i$ .

From $\mathrm{E}_{i,j}\,(e_ie_{ij})=\mathrm{E}_i\,[\mathrm{E}_j(e_ie_{ij})]=\mathrm{E}_i\,[e_i\mathrm{E}_j(e_{ij})]=\mathrm{E}_i\,(e_i.0)=0$

it follows that $e_i$ and $e_{ij}$ are uncorrelated. Note that this is a consequence of the accurate description of the sampling procedure and not an a priori assumption. It also follows from the sampling scheme that $e_{ij}$ and $e_{ij'}$, $(j\neq j')$ are independent. So we have not only that $\mathrm{var}_i\,(\mu_i)=\sigma_1^2$ according to the definition, but also $\mathrm{var}_{i,j}\,(y_{ij})=\sigma_1^2+\sigma_2^2$. The two variances $\sigma_1^2$ and $\sigma_2^2$ are called the variance components of $\mathrm{var}_{i,j}\,(y_{ij})$.

So far we considered sampling of one element only from the population. An additional sampling in the first stage will be identical by definition while in the second stage it will be independent of the previous sampling result. Denote two resulting observations by $y_{ij}$ and $y_{i'j'}$. In the following we mean by $i=i'$ that both observations have the first sampling stage in common, and analogously for $j=j'$.

In the expressions:

$$y_{ij}=\mu+e_i+e_{ij}$$

and

$$y_{i'j'}=\mu+e_{i'}+e_{i'j'}$$

any two terms with the same number of subscripts are either identical or uncorrelated. If $(i,j)=(i',j')$ then

$$e_i=e_{i'},\ e_{ij}=e_{i'j'}.$$

If $i=i'$ but $j\neq j'$ then only $e_i=e_{i'}$ are identical while $e_{ij}$ and $e_{i'j'}$ are uncorrelated. Finally, if $i\neq i'$ then the two pairs $(e_i, e_{i'})$ and $(e_{ij}, e_{i'j'})$ are uncorrelated.

The following expressions for covariances are now immediate:

$$\mathrm{cov}\,(y_{ij}, y_{i'j'})=\delta_{ii'}\,[\sigma_1^2+\delta_{jj'}\sigma_2^2]$$

$$\mathrm{cov}\,(y_{ij}, y_{i'j'})=\delta_{ii'}\delta_{jj'}\sigma_2^2\,,$$

where the Kronecker delta notation is used

$$\delta_{ii'}=1 \quad \text{if} \quad i=i' \quad \text{and} \quad \delta_{ii'}=0 \quad \text{if} \quad i\neq i'\,.$$

The usual description of the model is:

$$y_{ij}=\mu+b_i+c_{ij} \quad (i=1,\ldots,n_1;\ j=1,\ldots,n_2),$$

where the subscripts $i$ and $j$ refer to the samples rather than to the sampling procedure. With this relabelling $\underline{b}_i$ plays the role of $\underline{e}_i$ and $\underline{c}_{ij}$ that of $\underline{e}_{ij}$.
We summarize as follows:

$$\underline{y}_{ij} = \mu + \underline{b}_i + \underline{c}_{ij}$$

(1) with $\mathrm{E}\underline{b}_i = 0$ and $\mathrm{cov}\,(\underline{b}_i, \underline{b}_{i'}) = \delta_{ii'}\,\sigma_1^2$ for $i, i' = 1, \ldots, n_1$;
$\mathrm{E}\underline{c}_{ij} = 0$, $\mathrm{cov}\,(\underline{c}_{ij}, \underline{c}_{i'j'}) = \delta_{ii'}\delta_{jj'}\,\sigma_2^2$ for $i, i' = 1, \ldots, n_1$;
$j, j' = 1, \ldots, n_2$ and $\mathrm{cov}\,(\underline{b}_i, \underline{c}_{i'j'}) = 0$.

Let $y$ be the vector of all observations, $y = (y_{11}, \ldots, y_{1n_2}, y_{21}, \ldots, y_{2n_2}, \ldots, y_{n_1n_2})'$, then $y \in \boldsymbol{R}^n$ with $n = n_1 n_2$, $b = (b_1, \ldots, b_{n_1})' \in \boldsymbol{R}^{n_1}$, $c = (c_{11}, \ldots, c_{1n_2}, \ldots, c_{n_1n_2})' \in \boldsymbol{R}^n$.

The model can now be written as

$$\underline{y} = \mathbf{1}_n \mu + U_1 \underline{b} + U_2 \underline{c}$$

where $\mathbf{1}_n$ is a $n \times 1$-columnvector of ones, $U_1$ is an $n \times n_1$ designmatrix related to the first sampling stage, hence the $i^{\text{th}}$-column of $U_1$ consists of 1 according to observations of the $i$-th sample and 0 elsewhere; $U_2$ is the $n \times n$ identity matrix.

The covariance matrix of $y$ can be written as $V = \sigma_1^2 U_1 U_1' + \sigma_2^2 U_2 U_2' = \sigma_1^2 V_1 + \sigma_2^2 V_2$ with $V_1 = I_{n_1} \times J_{n_2}$ and $V_2 = I_n$, $n = n_1 n_2$, where the KRONECKER product $A \times B$ of the $(p \times q)$-matrix $A = \{a_{ij}\}$ and the $(r \times s)$-matrix $B = \{b_{kl}\}$ is the partitioned $(pr \times qs)$-matrix $\{a_{ij}B\}$. (For the properties and use of the KRONECKER product $A \times B$ see the Appendix). $I_{n_1}$ is the $n_1 \times n_1$ unit matrix $\{\delta_{ij}\}$, and $J_{n_2}$ is the $n_2 \times n_2$-matrix of ones, $\{1\}$.

In the described model (1) there are no further assumptions about the probability distribution of the random effects. Often an additional assumption is suitable for practice. This assumption is that of normally distributed random effects. Note that uncorrelated random effects in the case of joint normally distributed effects is equivalent to stochastic independence of the effects. In this case we formulate model (2) as:

$$\underline{y} = \mathbf{1}_n \mu + U_1 \underline{b} + U_2 \underline{c}$$

with

(2) $\underline{b} = N(0, \sigma_1^2 I_{n_1})$, $\underline{c} = N(0, \sigma_2^2 I_n)$, $n = n_1 n_2$,

and $\underline{b}$ and $\underline{c}$ are stochastically independent.

Hence $\underline{y}$ has a multivariate normal distribution $N(\mathbf{1}_n \mu, V)$.

## 3. Estimation of the Variance Components

The best estimator for the expectation $\mathrm{E}\underline{y}_{ij} = \mu$ is $\bar{\underline{y}} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \underline{y}_{ij}/(n_1 n_2)$. For model (1) this estimator $\bar{\underline{y}}$ has minimum variance among all unbiased linear estimators. For model (2) this estimator $\bar{\underline{y}}$ has minimum variance among all unbiased estimators. This follows from the fact that $\bar{\underline{y}}$ is a minimal sufficient complete statistic of the

normally distributed observations with respect to $\mu$. For the estimation of the variance components the analysis of variance table can be calculated:

| Source of variation | degrees of freedom df | Sum of Squares SS | Mean Squares MS | Expected Mean Squares E(MS) |
|---|---|---|---|---|
| Between samples | $a=n_1-1$ | SSB | MSB | $\sigma_2^2+n_2\sigma_1^2$ |
| Within samples | $e=n_1(n_2-1)$ | SSW | MSW | $\sigma_2^2$ |
| Corrected Total | $n_1n_2-1$ | SST | | |

where $SST=\sum_i\sum_j y_{ij}^2-C$ with $C=(\sum_i\sum_j y_{ij})^2/(n_1n_2)$; $SSB=\sum_i(\sum_j y_{ij})^2/n_2-C$; $SSW=SST-SSB$; $MSB=SSB/a$; $MSW=SSW/e$ with $a=n_1-1$ and $e=n_1(n_2-1)$.

Well known is $\underline{EMSB}=\sigma_2^2+n_2\sigma_1^2$ and $\underline{EMSW}=\sigma_2^2$.
Unbiased estimators for $\sigma_2^2$ and $\sigma_1^2$ are $\hat{\underline{\sigma}}_2^2=\underline{MSW}$ and $\hat{\underline{\sigma}}_1^2=(\underline{MSB}-\underline{MSW})/n_2$, respectively. These estimators are the so-called analysis of variance (ANOVA) estimators.

Under model (1) these estimators have minimal variance in the class of unbiased quadratic functions of the observations, GRAYBILL (1954); only for balanced designs. READ (1961). Under model (2) these estimators have minimal variance in the class of unbiased estimators. This follows from the fact that for a balanced design with normally distributed random effects the ANOVA estimators are functions of independent minimal sufficient complete statistics, with respect to the parameters $\mu$, $\sigma_1^2$ and $\sigma_2^2$ (GRAYBILL and WORTHAM (1956)).

In practical applications the problem arises of negative outcomes for $\hat{\underline{\sigma}}_1^2$. (Remark that $\hat{\underline{\sigma}}_2^2$ can be written as $\sum_i\sum_j(\underline{y}_{ij}-\bar{\underline{y}}_i)^2/(n_1(n_2-1))$ with $\bar{\underline{y}}_i=\sum_j\underline{y}_{ij}/n_2$, hence $\hat{\underline{\sigma}}_2^2$ can never be negative).

In that case one will often questions the validity of the model. But in § 4 the probability will be calculated for negative outcomes of $\hat{\underline{\sigma}}_1^2$ in the case of normally distributed observations (model (2)).

## 4. Probability of Negative Outcomes of $\hat{\underline{\sigma}}_1^2$

Under the assumptions of model (2) $\underline{SSW}$ is distributed as $\sigma_2^2\underline{\chi}_e^2$ with $e=n_1(n_2-1)$, and $\underline{SSB}$ as $(\sigma_2^2+n_2\sigma_1^2)\,\underline{\chi}_a^2$ with $a=n_1-1$. Furthermore $\underline{SSW}$ and $\underline{SSB}$ are stochastically independent.

Let $\underline{MSW}=\underline{SSW}/e$ and $\underline{MSB}=\underline{SSB}/a$. The probability for a negative outcome of $\hat{\underline{\sigma}}_1^2$ is $P(\hat{\underline{\sigma}}_1^2<0)=P((\underline{MSB}-\underline{MSW})/n_2<0)=P(\underline{MSB}<\underline{MSW})=P(\underline{MSB}/\underline{MSW}<1)$ since $P(\underline{MSW}=0)=0$. However $\underline{MSB}/\underline{MSW}$ is distributed as

$$\frac{(\sigma_2^2+n_2\sigma_1^2)\,\underline{\chi}_a^2/a}{\sigma_2^2\underline{\chi}_e^2/e}=(1+n_2\gamma)\,\underline{F}(a,e)$$

where $\gamma = \sigma_1^2/\sigma_2^2$ and $\underline{F}(a, e)$ has the $F$-distribution with $a$ and $e$ degrees of freedom respectively.

Hence $P(\hat{\underline{\sigma}}_1^2 < 0) = P((1+n_2\gamma)\,\underline{F}(a, e) < 1) = P\left(\underline{F}(a, e) < \frac{1}{1+n_2\gamma}\right)$. Another objection against the assumption of normality may be that we may find a certain negative outcome $-D$ $(D>0)$ whereas the probability is negligible for outcomes of $\hat{\underline{\sigma}}_1^2$ equal to or smaller than $-D$. To investigate this objection in more detail we express $D$ in units of $\sigma_2^2$, i.e. $D = \delta\sigma_2^2$ with $\delta > 0$. Then

$$\begin{aligned} P(\hat{\underline{\sigma}}_1^2 \leqq -\delta\sigma_2^2) &= P((\underline{\mathrm{MSB}} - \underline{\mathrm{MSW}})/n_2 \leqq -\delta\sigma_2^2) \\ &= P((\sigma_2^2 + n_2\sigma_1^2)\,\underline{\chi}_a^2/(an_2) - \sigma_2^2\underline{\chi}_e^2/(en_2) \leqq -\delta\sigma_2^2) \\ &= P\left(\frac{1+n_2\gamma}{an_2}\,\underline{\chi}_a^2 - \frac{1}{en_2}\,\underline{\chi}_e^2 \leqq -\delta\right) \quad \text{with} \quad \gamma = \sigma_1^2/\sigma_2^2\,. \end{aligned}$$

This probability can be evaluated by numerical inversion of the characteristic function of a linear combination of independent chi-square variables, see IMHOF (1961). The special case $\delta = 0$ has been handled before. In Table 1 this probability multiplied by 1000 is presented for several combinations of $n_1$, $n_2$, $\gamma$ and $\delta$. If for a certain combination of $\gamma$ and $\delta$ this probability is less than 0.001 it has been omitted.

It should be noted that sometimes there are available separate estimators for $\sigma_2^2$ and $\sigma_2^2 + m\sigma_1^2$ from different experiments. (For example in plant or animal breeding experiments). From one experiment one will obtain an unbiased estimator $\underline{\mathrm{SSW}}/q$ for $\sigma_2^2$ with $q$ degrees of freedom and from another experiment an unbiased estimator $\underline{\mathrm{SSB}}/p$ for $\sigma_2^2 + m\sigma_1^2$, which is independent of $\underline{\mathrm{SSW}}/q$. Now we find $P(\hat{\underline{\sigma}}_1^2 \leqq -\delta\sigma_2^2) = P((1+m\gamma)\,\underline{\chi}_p^2/p - \underline{\chi}_q^2/q \leqq -\delta m)$ where $\gamma = \sigma_1^2/\sigma_2^2$.

In Table 2 this probability multiplied by 1000 is presented for several combinations of the degrees of freedom $p$ and $q$ and several values of $\gamma$ and $\delta$. If for a certain combination of $\gamma$ and $\delta$ this probability is less than 0.001 it has been omitted.

The use of these tables is hence to find the probability of falsely rejecting the assumption of normality. To use a value $\delta$ one must know the value of $\sigma_2^2$. An idea about the value of $\sigma_2^2$ can be obtained from an $(1-\alpha)$-confidence interval for $\sigma_2^2$:

$$\underline{\mathrm{SSW}}\,/U < \sigma_2^2 < \underline{\mathrm{SSW}}/L$$

with

$$P(\underline{\chi}_e^2 < L) = \alpha/2 \quad \text{and} \quad P(\underline{\chi}_e^2 < U) = 1 - \alpha/2\,.$$

**Table 1**

For given $\gamma = \sigma_1^2/\sigma_2^2$ and stochastically independent $\underline{\chi}_a^2$ and $\underline{\chi}_e^2$ the probability multiplied by 1000 is given for a negative estimate of the variance component $\sigma_1^2$:

$$1000 \times P(\hat{\underline{\sigma}}_1^2 < -\delta\sigma_2^2) = 1000 \times P\left(\frac{1+n_2\gamma}{an_2}\,\underline{\chi}_a^2 - \frac{1}{en_2}\,\underline{\chi}_e^2 \leqq -\delta\right), \quad \delta > 0,$$

with degrees of freedom $a = n_1 - 1$, $e = n_1 (n_2 - 1)$ for

$$n_1 = 2, 5, 10, 30, 50, 70 \quad \text{and} \quad n_2 = 2, 3, 4, 5, 10, 15 .$$

This probability has been computed for $\gamma = 0.0, 0.1, 0.5, 1.0, 2.0, 10.0$ and for $\delta = 0.0, 0.01, 0.05, 0.1\ (0.1)\ 1.0, 1.5, 2.0$.

If for a certain combination of $\gamma$ and $\delta$ this probability is less than 0.001 it is not given.

$n_1 = 2, n_2 = 2; a = 1, e = 2$

| $\gamma$ \ $\delta$ | 0.00 | 0.01 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.5 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 578 | 568 | 522 | 473 | 387 | 317 | 259 | 212 | 174 | 142 | 116 | 95 | 78 | 29 | 11 |
| 0.1 | 542 | 532 | 491 | 444 | 363 | 298 | 244 | 200 | 163 | 134 | 110 | 90 | 73 | 27 | 10 |
| 0.5 | 447 | 438 | 405 | 366 | 300 | 246 | 201 | 164 | 146 | 122 | 103 | 87 | 60 | 22 | 8 |
| 1.0 | 377 | 370 | 342 | 309 | 253 | 208 | 170 | 139 | 114 | 93 | 76 | 62 | 51 | 19 | 7 |
| 2.0 | 301 | 296 | 273 | 247 | 202 | 165 | 136 | 111 | 91 | 74 | 61 | 50 | 41 | 15 | 6 |
| 10.0 | 152 | 149 | 138 | 125 | 102 | 84 | 69 | 56 | 46 | 38 | 31 | 25 | 21 | 8 | 3 |

$n_1 = 2, n_2 = 3; a = 1, e = 4$

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 625 | 614 | 563 | 491 | 350 | 237 | 154 | 98 | 61 | 38 | 23 | 14 | 8 | 1 | |
| 0.1 | 569 | 559 | 512 | 446 | 316 | 214 | 139 | 88 | 55 | 34 | 20 | 12 | 7 | 0 | |
| 0.5 | 438 | 429 | 392 | 341 | 241 | 162 | 105 | 67 | 42 | 26 | 16 | 9 | 6 | 0 | |
| 1.0 | 356 | 349 | 318 | 276 | 195 | 131 | 85 | 54 | 34 | 21 | 12 | 8 | 4 | 0 | |
| 2.0 | 275 | 269 | 245 | 213 | 150 | 101 | 66 | 41 | 26 | 16 | 10 | 6 | 3 | 0 | |
| 10.0 | 133 | 130 | 119 | 103 | 73 | 48 | 32 | 20 | 12 | 8 | 5 | 3 | 2 | 0 | |

$n_1 = 2\ \ n_2 = 4; a = 1, e = 6$

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 644 | 631 | 568 | 472 | 274 | 138 | 62 | 27 | 11 | 4 | 2 | 1 |
| 0.1 | 570 | 557 | 500 | 413 | 239 | 120 | 54 | 23 | 9 | 4 | 1 | 1 |
| 0.5 | 415 | 405 | 361 | 297 | 171 | 85 | 38 | 16 | 6 | 2 | 1 | 0 |
| 1.0 | 329 | 321 | 286 | 235 | 135 | 67 | 30 | 13 | 5 | 2 | 1 | 0 |
| 2.0 | 250 | 243 | 216 | 177 | 102 | 50 | 23 | 9 | 4 | 2 | 0 | 0 |
| 10.0 | 119 | 116 | 102 | 84 | 48 | 24 | 11 | 5 | 2 | 1 | 0 | 0 |

$n_1 = 2, n_2 = 5; a = 1, e = 8$

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 654 | 638 | 562 | 436 | 190 | 61 | 17 | 4 | 1 |
| 0.1 | 563 | 548 | 480 | 370 | 160 | 52 | 14 | 3 | 1 |
| 0.5 | 393 | 382 | 331 | 254 | 108 | 35 | 9 | 2 | 1 |
| 1.0 | 306 | 298 | 257 | 197 | 84 | 27 | 7 | 2 | 0 |
| 2.0 | 230 | 223 | 192 | 147 | 62 | 20 | 5 | 1 | 0 |
| 10.0 | 108 | 105 | 90 | 69 | 29 | 9 | 2 | 1 | 0 |

$n_1 = 2, n_2 = 10; a = 1, e = 18$

| | | | | | |
|---|---|---|---|---|---|
| 0.0 | 669 | 642 | 477 | 167 | 2 |
| 0.1 | 511 | 487 | 354 | 122 | 2 |
| 0.5 | 312 | 296 | 211 | 72 | 1 |
| 1.0 | 234 | 221 | 157 | 53 | 1 |
| 2.0 | 170 | 161 | 114 | 39 | 0 |
| 10.0 | 78 | 74 | 52 | 18 | 0 |

$n_1=2, n_2=15; a=1, e=28$

| $\gamma$ \ $\delta$ | 0.00 | 0.01 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.5 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 674 | 632 | 328 | 12 | | | | | | | | | | | |
| 0.1 | 468 | 434 | 216 | 8 | | | | | | | | | | | |
| 0.5 | 266 | 245 | 119 | 4 | | | | | | | | | | | |
| 1.0 | 196 | 180 | 87 | 3 | | | | | | | | | | | |
| 2.0 | 141 | 130 | 63 | 2 | | | | | | | | | | | |
| 10.0 | 64 | 59 | 28 | 1 | | | | | | | | | | | |

$n_1=5, n_2=2; a=4, e=5$

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 514 | 504 | 463 | 413 | 318 | 237 | 173 | 124 | 88 | 61 | 42 | 28 | 19 | 2 |
| 0.1 | 442 | 433 | 396 | 352 | 269 | 200 | 145 | 103 | 73 | 51 | 35 | 23 | 16 | 2 |
| 0.5 | 261 | 255 | 231 | 203 | 153 | 112 | 80 | 56 | 40 | 28 | 19 | 13 | 8 | 1 |
| 1.0 | 155 | 151 | 136 | 119 | 88 | 64 | 46 | 32 | 22 | 16 | 11 | 7 | 5 | 1 |
| 2.0 | 72 | 70 | 63 | 55 | 40 | 29 | 20 | 14 | 10 | 7 | 5 | 3 | 2 | 0 |
| 10.0 | 6 | 6 | 5 | 4 | 3 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |

$n_1=5, n_2=3; a=4, e=10$

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 548 | 532 | 466 | 380 | 224 | 113 | 51 | 21 | 8 | 3 | 1 |
| 0.1 | 431 | 416 | 359 | 288 | 166 | 82 | 37 | 15 | 6 | 2 | 1 |
| 0.5 | 195 | 188 | 157 | 123 | 67 | 32 | 14 | 6 | 2 | 1 | 0 |
| 1.0 | 97 | 92 | 77 | 59 | 32 | 15 | 6 | 3 | 1 | 0 | 0 |
| 2.0 | 38 | 36 | 30 | 23 | 12 | 6 | 2 | 1 | 0 | 0 | 0 |
| 10.0 | 2 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |

$n_1=5, n_2=4; a=4, e=15$

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0.0 | 562 | 540 | 448 | 327 | 128 | 34 | 7 | 1 |
| 0.1 | 405 | 386 | 311 | 220 | 82 | 21 | 4 | 1 |
| 0.5 | 149 | 140 | 108 | 72 | 25 | 6 | 1 | 0 |
| 1.0 | 66 | 62 | 46 | 31 | 10 | 2 | 0 | 0 |
| 2.0 | 23 | 22 | 16 | 10 | 4 | 1 | 0 | 0 |
| 10.0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |

$n_1=5, n_2=5; a=4, e=20$

| | | | | | | |
|---|---|---|---|---|---|---|
| 0.0 | 569 | 542 | 423 | 268 | 57 | 5 |
| 0.1 | 378 | 355 | 264 | 159 | 31 | 3 |
| 0.5 | 116 | 108 | 75 | 42 | 8 | 1 |
| 1.0 | 47 | 43 | 30 | 16 | 3 | 0 |
| 2.0 | 16 | 14 | 10 | 5 | 1 | 0 |
| 10.0 | 1 | 1 | 0 | 0 | 0 | 0 |

$n_1=5, n_2=10; a=4, e=45$

| | | | | |
|---|---|---|---|---|
| 0.0 | 582 | 526 | 265 | 31 |
| 0.1 | 264 | 228 | 97 | 10 |
| 0.5 | 46 | 38 | 14 | 1 |
| 1.0 | 15 | 12 | 4 | 0 |
| 2.0 | 4 | 4 | 1 | 0 |

$n_1=5,\ n_2=15;\ a=4,\ e=70$

| $\gamma$ \ $\delta$ | 0.00 | 0.01 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.5 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 587 | 500 | 107 | | | | | | | | | | | | |
| 0.1 | 192 | 150 | 23 | | | | | | | | | | | | |
| 0.5 | 24 | 18 | 2 | | | | | | | | | | | | |
| 1.0 | 7 | 5 | 1 | | | | | | | | | | | | |
| 2.0 | 2 | 2 | 0 | | | | | | | | | | | | |

$n_1=10,\ n_2=2;\ a=9,\ e=10$

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 505 | 491 | 438 | 373 | 257 | 165 | 100 | 58 | 32 | 17 | 9 | 5 | 2 |
| 0.1 | 397 | 385 | 339 | 285 | 191 | 120 | 72 | 41 | 23 | 12 | 6 | 3 | 2 |
| 0.5 | 156 | 150 | 128 | 104 | 66 | 39 | 23 | 13 | 7 | 4 | 2 | 1 | 0 |
| 1.0 | 57 | 54 | 46 | 36 | 22 | 13 | 7 | 4 | 2 | 1 | 0 | 0 | 0 |
| 2.0 | 12 | 11 | 9 | 7 | 4 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |

$n_1=10,\ n_2=3;\ a=9,\ e=20$

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0.0 | 529 | 507 | 416 | 306 | 132 | 42 | 10 | 2 |
| 0.1 | 354 | 336 | 265 | 186 | 74 | 22 | 5 | 1 |
| 0.5 | 80 | 74 | 54 | 34 | 12 | 3 | 1 | 0 |
| 1.0 | 19 | 17 | 12 | 8 | 2 | 1 | 0 | 0 |
| 2.0 | 2 | 2 | 2 | 1 | 0 | 0 | 0 | 0 |

$n_1=10,\ n_2=4;\ a=9,\ e=30$

| | | | | | | |
|---|---|---|---|---|---|---|
| 0.0 | 539 | 508 | 380 | 230 | 47 | 4 |
| 0.1 | 308 | 285 | 197 | 108 | 19 | 2 |
| 0.5 | 43 | 39 | 24 | 11 | 2 | 0 |
| 1.0 | 8 | 7 | 4 | 2 | 0 | 0 |
| 2.0 | 1 | 1 | 0 | 0 | 0 | 0 |

$n_1=10,\ n_2=5;\ a=9,\ e=40$

| | | | | | |
|---|---|---|---|---|---|
| 0.0 | 544 | 504 | 340 | 159 | 10 |
| 0.1 | 266 | 239 | 141 | 57 | 3 |
| 0.5 | 25 | 22 | 11 | 4 | 0 |
| 1.0 | 4 | 3 | 1 | 0 | 0 |

$n_1=10,\ n_2=10;\ a=9,\ e=90$

| | | | | |
|---|---|---|---|---|
| 0.0 | 554 | 470 | 142 | 3 |
| 0.1 | 129 | 97 | 18 | 0 |
| 0.5 | 3 | 2 | 0 | 0 |

$n_1=10,\ n_2=15;\ a=9,\ e=140$

| | | | |
|---|---|---|---|
| 0.0 | 557 | 428 | 25 |
| 0.1 | 67 | 41 | 1 |
| 0.5 | 1 | 0 | 0 |

$n_1=30,\ n_2=2;\ a=29,\ e=30$

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 501 | 479 | 391 | 290 | 135 | 51 | 16 | 4 | 1 |
| 0.1 | 313 | 295 | 229 | 159 | 66 | 23 | 6 | 2 | 0 |
| 0.5 | 33 | 30 | 20 | 12 | 4 | 1 | 0 | 0 | 0 |
| 1.0 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |

| $\gamma$ \ $\delta$ | 0.00 | 0.01 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.5 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n_1=30, n_2=3; a=29, e=60$ | | | | | | | | | | | | | | |
| 0.0 | 515 | 477 | 327 | 172 | 24 | 1 | | | | | | | | | |
| 0.1 | 222 | 198 | 115 | 50 | 5 | 0 | | | | | | | | | |
| 0.5 | 4 | 4 | 2 | 0 | 0 | 0 | | | | | | | | | |
| | $n_1=30, n_2=4; a=29, e=90$ | | | | | | | | | | | | | | |
| 0.0 | 521 | 467 | 260 | 83 | 2 | | | | | | | | | | |
| 0.1 | 153 | 127 | 51 | 11 | 0 | | | | | | | | | | |
| 0.5 | 1 | 0 | 0 | 0 | 0 | | | | | | | | | | |
| | $n_1=30, n_2=5; a=29, e=120$ | | | | | | | | | | | | | | |
| 0.0 | 524 | 454 | 198 | 32 | | | | | | | | | | | |
| 0.1 | 103 | 79 | 20 | ? | | | | | | | | | | | |
| | $n_1=30, n_2=10; a=29, e=270$ | | | | | | | | | | | | | | |
| 0.0 | 530 | 382 | 20 | | | | | | | | | | | | |
| 0.1 | 14 | 6 | 0 | | | | | | | | | | | | |
| | $n_1=30, n_2=15; a=29, e=420$ | | | | | | | | | | | | | | |
| 0.0 | 532 | 308 | | | | | | | | | | | | | |
| 0.1 | 2 | 0 | | | | | | | | | | | | | |
| | $n_1=50, n_2=2; a=49, e=50$ | | | | | | | | | | | | | | |
| 0.0 | 500 | 472 | 361 | 238 | 79 | 18 | 3 | | | | | | | | |
| 0.1 | 262 | 241 | 168 | 98 | 26 | 5 | 1 | | | | | | | | |
| 0.5 | 8 | 7 | 4 | 2 | 0 | 0 | 0 | | | | | | | | |
| | $n_1=50, n_2=3; a=49, e=100$ | | | | | | | | | | | | | | |
| 0.0 | 511 | 462 | 275 | 108 | 5 | | | | | | | | | | |
| 0.1 | 135 | 130 | 58 | 16 | 0 | | | | | | | | | | |
| | $n_1=50, n_2=4; a=49, e=150$ | | | | | | | | | | | | | | |
| 0.0 | 516 | 446 | 196 | 35 | | | | | | | | | | | |
| 0.1 | 87 | 65 | 16 | 1 | | | | | | | | | | | |
| | $n_1=50, n_2=5; a=49, e=200$ | | | | | | | | | | | | | | |
| 0.0 | 518 | 429 | 130 | 8 | | | | | | | | | | | |
| 0.1 | 47 | 31 | 4 | 0 | | | | | | | | | | | |
| | $n_1=50, n_2=10; a=49, e=450$ | | | | | | | | | | | | | | |
| 0.0 | 523 | 335 | 4 | | | | | | | | | | | | |
| 0.1 | 2 | 0 | 0 | | | | | | | | | | | | |
| | $n_1=50, n_2=15; a=49, e=700$ | | | | | | | | | | | | | | |
| 0.0 | 524 | 245 | | | | | | | | | | | | | |

| $\gamma$ \ $\delta$ | $n_1=70, n_2=2; a=69, e=70$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.00 | 0.01 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.5 | 2.0 |
| 0.0 | 500 | 467 | 337 | 200 | 48 | 7 | 1 | | | | | | | | |
| 0.1 | 225 | 202 | 127 | 63 | 11 | 1 | 0 | | | | | | | | |
| 0.5 | 2 | 2 | 1 | 0 | 0 | 0 | 0 | | | | | | | | |

| | $n_1=70, n_2=3; a=69, e=140$ | | | | |
|---|---|---|---|---|---|
| 0.0 | 509 | 451 | 237 | 70 | 1 |
| 0.1 | 112 | 89 | 30 | 5 | 0 |

| | $n_1=70, n_2=4; a=69, e=210$ | | | |
|---|---|---|---|---|
| 0.0 | 513 | 431 | 152 | 16 |
| 0.1 | 52 | 35 | 5 | 0 |

| | $n_1=70, n_2=5; a=69, e=280$ | | | |
|---|---|---|---|---|
| 0.0 | 515 | 410 | 88 | 2 |
| 0.1 | 23 | 13 | 1 | 0 |

| | $n_1=70, n_2=10; a=69, e=630$ | | |
|---|---|---|---|
| 0.0 | 519 | 299 | 1 |

| | $n_1=70, n_2=15; a=69, e=980$ | |
|---|---|---|
| 0.0 | 520 | 200 |

**Table 2**

For given $\gamma=\sigma_1^2/\sigma_2^2$ and stochastically independent $\underset{\sim}{\chi}_p^2$ and $\underset{\sim}{\chi}_q^2$ the probability multiplied by 1000 is given for a negative estimate of the variance component $\sigma_1^2$:

$$1000\times P\,(\hat{\underset{\sim}{\sigma}}_1^2 \leqq -\delta\sigma_2^2)=1000\times P\left(\frac{1+m\gamma}{p}\,\underset{\sim}{\chi}_p^2-\frac{1}{q}\,\underset{\sim}{\chi}_q^2 \leqq -\delta m\right)$$

with degrees of freedom
$p$ and $q=1$, 5, 10, 30, 50, 100 and $m\gamma=0.0$, 0.1, 0.5, 1.0, 2.0, 10.0 and $\delta>0$, $m\delta=0.00$, 0.01, 0.05, 0.1 (0.1) 1.0, 1.5, 2.0. If for a certain combination of $m\gamma$ and $m\delta$ this probability is less than 0.001 it is not given.

| $my$ \ $m\delta$ | $p=1, q=1$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.00 | 0.01 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.5 | 2.0 |
| 0.0 | 500 | 490 | 462 | 435 | 391 | 356 | 326 | 300 | 277 | 256 | 237 | 220 | 205 | 145 | 104 |
| 0.1 | 485 | 475 | 448 | 422 | 382 | 348 | 320 | 291 | 269 | 249 | 231 | 214 | 199 | 141 | 102 |
| 0.5 | 436 | 428 | 404 | 380 | 343 | 312 | 286 | 264 | 244 | 226 | 209 | 194 | 181 | 141 | 92 |
| 1.0 | 392 | 385 | 363 | 343 | 309 | 282 | 259 | 238 | 220 | 204 | 189 | 176 | 164 | 116 | 84 |
| 2.0 | 333 | 328 | 309 | 292 | 264 | 241 | 221 | 204 | 188 | 175 | 162 | 151 | 140 | 100 | 72 |
| 10.0 | 186 | 184 | 173 | 164 | 148 | 136 | 125 | 115 | 106 | 99 | 92 | 85 | 80 | 57 | 41 |

| | $p=1, q=5$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 636 | 633 | 618 | 599 | 556 | 509 | 459 | 408 | 360 | 314 | 273 | 236 | 202 | 86 | 34 |
| 0.1 | 615 | 612 | 598 | 579 | 537 | 491 | 443 | 394 | 347 | 303 | 263 | 227 | 194 | 83 | 33 |
| 0.5 | 548 | 545 | 532 | 515 | 477 | 435 | 392 | 348 | 306 | 267 | 232 | 200 | 171 | 73 | 29 |
| 1.0 | 488 | 486 | 474 | 458 | 424 | 386 | 348 | 309 | 272 | 237 | 205 | 176 | 151 | 64 | 26 |
| 2.0 | 418 | 416 | 398 | 384 | 355 | 324 | 291 | 258 | 227 | 198 | 171 | 147 | 126 | 54 | 21 |
| 10.0 | 224 | 223 | 217 | 209 | 193 | 176 | 158 | 140 | 123 | 107 | 92 | 79 | 68 | 29 | 11 |

| | $p=1, q=10$ | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 660 | 657 | 645 | 629 | 593 | 551 | 502 | 448 | 391 | 334 | 279 | 228 | 184 | 50 | 11 |
| 0.1 | 638 | 635 | 623 | 608 | 572 | 531 | 484 | 432 | 377 | 321 | 268 | 219 | 176 | 48 | 10 |
| 0.5 | 567 | 565 | 554 | 539 | 506 | 469 | 426 | 379 | 330 | 281 | 235 | 192 | 154 | 42 | 9 |
| 1.0 | 505 | 502 | 492 | 479 | 449 | 415 | 376 | 335 | 291 | 248 | 206 | 168 | 135 | 37 | 8 |
| 2.0 | 432 | 429 | 420 | 409 | 383 | 354 | 322 | 287 | 250 | 214 | 180 | 140 | 112 | 31 | 6 |
| 10.0 | 231 | 230 | 225 | 218 | 203 | 187 | 169 | 150 | 130 | 110 | 92 | 75 | 60 | 16 | 3 |

| | $p=1, q=30$ | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 675 | 672 | 662 | 648 | 618 | 583 | 543 | 495 | 436 | 367 | 292 | 217 | 151 | 10 |
| 0.1 | 652 | 650 | 639 | 625 | 596 | 562 | 522 | 476 | 419 | 352 | 280 | 208 | 145 | 10 |
| 0.5 | 579 | 577 | 567 | 554 | 526 | 494 | 458 | 416 | 365 | 306 | 243 | 180 | 134 | 8 |
| 1.0 | 515 | 513 | 503 | 491 | 465 | 436 | 404 | 366 | 321 | 268 | 212 | 158 | 110 | 7 |
| 2.0 | 432 | 430 | 422 | 411 | 388 | 364 | 336 | 303 | 265 | 231 | 184 | 139 | 98 | 6 |
| 10.0 | 235 | 233 | 229 | 223 | 210 | 196 | 180 | 162 | 141 | 118 | 93 | 69 | 48 | 3 |

| | $p=1, q=50$ | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 678 | 675 | 665 | 652 | 622 | 589 | 551 | 506 | 451 | 383 | 301 | 214 | 136 | 3 |
| 0.1 | 655 | 653 | 642 | 629 | 600 | 567 | 530 | 486 | 433 | 367 | 288 | 205 | 130 | 3 |
| 0.5 | 582 | 580 | 570 | 557 | 529 | 499 | 465 | 425 | 377 | 319 | 250 | 177 | 113 | 2 |
| 1.0 | 517 | 515 | 506 | 494 | 468 | 440 | 409 | 373 | 331 | 279 | 218 | 155 | 98 | 2 |
| 2.0 | 434 | 432 | 424 | 413 | 391 | 366 | 340 | 309 | 273 | 230 | 180 | 127 | 81 | 2 |
| 10.0 | 236 | 235 | 230 | 224 | 211 | 197 | 182 | 165 | 145 | 122 | 95 | 67 | 42 | 1 |

| | $p=1, q=100$ | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 680 | 678 | 668 | 654 | 626 | 593 | 556 | 514 | 463 | 400 | 316 | 215 | 117 | 14 |
| 0.1 | 657 | 655 | 645 | 632 | 603 | 571 | 535 | 493 | 444 | 383 | 303 | 205 | 112 | 0 |
| 0.5 | 584 | 582 | 572 | 559 | 532 | 502 | 469 | 431 | 387 | 332 | 262 | 177 | 97 | 0 |
| 1.0 | 519 | 517 | 508 | 496 | 471 | 443 | 413 | 378 | 339 | 290 | 228 | 154 | 84 | 0 |
| 2.0 | 435 | 433 | 425 | 415 | 393 | 369 | 342 | 313 | 280 | 239 | 188 | 127 | 69 | 0 |
| 10.0 | 236 | 235 | 230 | 224 | 212 | 198 | 183 | 167 | 149 | 127 | 99 | 67 | 36 | 0 |

| $m\gamma$ \ $m\delta$ | 0.00 | 0.01 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.5 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $p=5, q=1$ | | | | | | | | | | | | | | |
| 0.0 | 364 | 360 | 347 | 332 | 306 | 283 | 263 | 244 | 226 | 210 | 196 | 183 | 171 | 123 | 89 |
| 0.1 | 343 | 340 | 328 | 314 | 289 | 268 | 249 | 231 | 215 | 200 | 186 | 174 | 162 | 117 | 85 |
| 0.5 | 276 | 273 | 264 | 253 | 234 | 217 | 202 | 189 | 176 | 164 | 153 | 143 | 134 | 97 | 71 |
| 1.0 | 217 | 215 | 208 | 200 | 185 | 172 | 160 | 150 | 140 | 131 | 123 | 115 | 108 | 78 | 57 |
| 2.0 | 144 | 143 | 139 | 134 | 124 | 115 | 108 | 101 | 94 | 88 | 83 | 78 | 74 | 54 | 39 |
| 10.0 | 21 | 21 | 20 | 20 | 18 | 17 | 16 | 15 | 14 | 13 | 13 | 12 | 11 | 8 | 6 |
| | $p=5, q=5$ | | | | | | | | | | | | | | |
| 0.0 | 500 | 495 | 474 | 448 | 396 | 348 | 302 | 261 | 224 | 191 | 163 | 138 | 116 | 46 | 18 |
| 0.1 | 460 | 455 | 435 | 410 | 362 | 317 | 275 | 237 | 203 | 173 | 147 | 124 | 105 | 42 | 16 |
| 0.5 | 334 | 330 | 314 | 295 | 259 | 225 | 194 | 166 | 141 | 120 | 101 | 85 | 72 | 29 | 10 |
| 1.0 | 233 | 230 | 218 | 204 | 178 | 154 | 132 | 112 | 95 | 80 | 68 | 57 | 48 | 19 | 7 |
| 2.0 | 127 | 125 | 119 | 111 | 96 | 82 | 70 | 59 | 50 | 42 | 35 | 29 | 24 | 10 | 4 |
| 10.0 | 10 | 10 | 9 | 9 | 7 | 6 | 5 | 4 | 4 | 3 | 2 | 2 | 2 | 1 | 0 |
| | $p=5, q=10$ | | | | | | | | | | | | | | |
| 0.0 | 535 | 529 | 506 | 476 | 417 | 359 | 303 | 251 | 205 | 164 | 130 | 101 | 78 | 18 | 4 |
| 0.1 | 488 | 482 | 460 | 432 | 376 | 322 | 271 | 224 | 182 | 145 | 114 | 89 | 68 | 16 | 3 |
| 0.5 | 342 | 338 | 320 | 298 | 256 | 216 | 179 | 146 | 118 | 93 | 73 | 56 | 43 | 10 | 2 |
| 1.0 | 230 | 226 | 214 | 198 | 168 | 140 | 115 | 93 | 74 | 58 | 45 | 35 | 26 | 6 | 1 |
| 2.0 | 118 | 116 | 109 | 100 | 84 | 69 | 56 | 45 | 36 | 28 | 21 | 16 | 12 | 3 | 0 |
| 10.0 | 8 | 8 | 8 | 7 | 6 | 4 | 4 | 3 | 2 | 2 | 1 | 1 | 1 | 0 | 0 |
| | $p=5, q=30$ | | | | | | | | | | | | | | |
| 0.0 | 565 | 559 | 535 | 503 | 437 | 369 | 301 | 235 | 176 | 125 | 84 | 54 | 33 | 1 | |
| 0.1 | 512 | 506 | 482 | 452 | 389 | 326 | 263 | 204 | 152 | 107 | 72 | 46 | 28 | 1 | |
| 0.5 | 348 | 344 | 324 | 300 | 252 | 206 | 162 | 123 | 90 | 62 | 41 | 26 | 15 | 1 | |
| 1.0 | 226 | 223 | 208 | 191 | 157 | 126 | 98 | 73 | 52 | 36 | 23 | 14 | 9 | 0 | |
| 2.0 | 111 | 109 | 101 | 92 | 74 | 58 | 44 | 33 | 23 | 16 | 10 | 6 | 4 | 0 | |
| 10.0 | 7 | 7 | 6 | 6 | 4 | 3 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | |
| | $p=5, q=50$ | | | | | | | | | | | | | | |
| 0.0 | 572 | 566 | 542 | 510 | 442 | 372 | 300 | 231 | 167 | 112 | 69 | 39 | 20 | | |
| 0.1 | 517 | 511 | 487 | 456 | 392 | 326 | 261 | 199 | 142 | 95 | 58 | 33 | 17 | | |
| 0.5 | 350 | 345 | 325 | 300 | 251 | 203 | 158 | 117 | 82 | 53 | 32 | 18 | 9 | | |
| 1.0 | 225 | 222 | 207 | 189 | 155 | 123 | 94 | 68 | 47 | 30 | 18 | 10 | 5 | | |
| 2.0 | 110 | 108 | 100 | 90 | 72 | 56 | 42 | 30 | 20 | 13 | 7 | 4 | 2 | | |
| 10.0 | 7 | 6 | 6 | 5 | 4 | 3 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | | |
| | $p=5, q=100$ | | | | | | | | | | | | | | |
| 0.0 | 578 | 572 | 547 | 515 | 446 | 374 | 300 | 227 | 159 | 100 | 55 | 26 | 10 | | |
| 0.1 | 522 | 516 | 491 | 460 | 394 | 327 | 260 | 194 | 135 | 84 | 46 | 21 | 8 | | |
| 0.5 | 350 | 345 | 325 | 300 | 250 | 223 | 155 | 112 | 76 | 46 | 24 | 11 | 4 | | |
| 1.0 | 224 | 221 | 206 | 188 | 153 | 120 | 90 | 64 | 42 | 25 | 13 | 6 | 2 | | |
| 2.0 | 108 | 106 | 98 | 89 | 70 | 54 | 40 | 28 | 18 | 10 | 5 | 2 | 1 | | |
| 10.0 | 6 | 6 | 6 | 5 | 4 | 3 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | | |

$p=10, q=1$

| $m\gamma$ \ $m\delta$ | 0.00 | 0.01 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.5 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 340 | 338 | 327 | 314 | 291 | 270 | 250 | 233 | 216 | 202 | 188 | 176 | 165 | 118 | 86 |
| 0.1 | 318 | 316 | 306 | 294 | 273 | 254 | 236 | 219 | 204 | 190 | 178 | 166 | 156 | 112 | 82 |
| 0.5 | 248 | 246 | 239 | 230 | 215 | 200 | 187 | 174 | 162 | 152 | 142 | 133 | 124 | 91 | 66 |
| 1.0 | 187 | 186 | 181 | 174 | 163 | 152 | 143 | 133 | 125 | 117 | 109 | 102 | 96 | 70 | 52 |
| 2.0 | 113 | 113 | 110 | 106 | 100 | 93 | 88 | 82 | 78 | 73 | 68 | 64 | 60 | 44 | 33 |
| 10.0 | 8 | 8 | 8 | 7 | 7 | 6 | 6 | 6 | 6 | 5 | 5 | 5 | 4 | 3 | 2 |

$p=10, q=5$

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 465 | 460 | 437 | 410 | 357 | 310 | 266 | 228 | 194 | 164 | 139 | 117 | 98 | 38 | 14 |
| 0.1 | 418 | 413 | 392 | 367 | 319 | 275 | 236 | 201 | 171 | 145 | 122 | 102 | 85 | 33 | 12 |
| 0.5 | 274 | 270 | 255 | 237 | 203 | 174 | 148 | 125 | 106 | 89 | 74 | 62 | 52 | 20 | 7 |
| 1.0 | 164 | 162 | 152 | 141 | 120 | 102 | 86 | 72 | 60 | 51 | 42 | 35 | 29 | 11 | 4 |
| 2.0 | 66 | 64 | 60 | 55 | 47 | 39 | 33 | 27 | 23 | 19 | 16 | 13 | 11 | 4 | 1 |
| 10.0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

$p=10, q=10$

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 500 | 493 | 466 | 422 | 367 | 305 | 250 | 201 | 160 | 125 | 96 | 74 | 56 | 12 | 2 |
| 0.1 | 442 | 435 | 410 | 378 | 318 | 263 | 214 | 171 | 135 | 105 | 81 | 62 | 46 | 10 | 2 |
| 0.5 | 267 | 262 | 244 | 222 | 182 | 147 | 117 | 92 | 72 | 55 | 42 | 31 | 23 | 5 | 1 |
| 1.0 | 145 | 142 | 131 | 118 | 95 | 76 | 59 | 46 | 35 | 27 | 20 | 15 | 11 | 2 | 0 |
| 2.0 | 49 | 48 | 44 | 39 | 31 | 24 | 18 | 14 | 10 | 8 | 6 | 4 | 3 | 1 | 0 |

$p=10, q=30$

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 535 | 527 | 494 | 453 | 370 | 291 | 219 | 157 | 107 | 69 | 43 | 25 | 14 |
| 0.1 | 463 | 455 | 424 | 386 | 310 | 240 | 178 | 126 | 84 | 54 | 33 | 19 | 11 |
| 0.5 | 254 | 249 | 228 | 202 | 155 | 114 | 81 | 55 | 36 | 22 | 13 | 7 | 4 |
| 1.0 | 124 | 120 | 108 | 94 | 70 | 50 | 34 | 22 | 14 | 8 | 5 | 3 | 2 |
| 2.0 | 35 | 34 | 30 | 26 | 18 | 12 | 8 | 5 | 3 | 2 | 1 | 1 | 0 |

$p=10, q=50$

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 544 | 535 | 502 | 458 | 371 | 286 | 208 | 141 | 89 | 52 | 28 | 14 | 6 |
| 0.1 | 468 | 460 | 428 | 387 | 307 | 232 | 166 | 110 | 69 | 39 | 21 | 10 | 4 |
| 0.5 | 251 | 245 | 223 | 196 | 147 | 105 | 71 | 45 | 26 | 15 | 7 | 4 | 2 |
| 1.0 | 118 | 115 | 103 | 88 | 63 | 43 | 28 | 17 | 10 | 5 | 3 | 1 | 0 |
| 2.0 | 32 | 31 | 27 | 23 | 16 | 10 | 6 | 4 | 2 | 1 | 0 | 0 | 0 |

$p=10, q=100$

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 551 | 543 | 508 | 463 | 371 | 281 | 197 | 127 | 73 | 36 | 16 | 6 | 2 |
| 0.1 | 472 | 464 | 430 | 388 | 304 | 225 | 154 | 97 | 54 | 27 | 11 | 4 | 1 |
| 0.5 | 247 | 241 | 218 | 191 | 140 | 97 | 62 | 37 | 19 | 9 | 4 | 1 | 0 |
| 1.0 | 114 | 110 | 98 | 83 | 58 | 38 | 23 | 13 | 7 | 3 | 1 | 0 | 0 |
| 2.0 | 30 | 29 | 25 | 21 | 14 | 9 | 5 | 3 | 1 | 1 | 0 | 0 | 0 |

$p=30, q=1$

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 325 | 323 | 313 | 301 | 279 | 260 | 241 | 225 | 210 | 196 | 183 | 171 | 160 | 114 | 84 |
| 0.1 | 303 | 300 | 291 | 281 | 261 | 242 | 226 | 210 | 196 | 183 | 171 | 160 | 150 | 109 | 80 |
| 0.5 | 230 | 229 | 222 | 214 | 200 | 187 | 174 | 163 | 152 | 143 | 134 | 125 | 118 | 86 | 63 |
| 1.0 | 168 | 166 | 162 | 157 | 147 | 137 | 128 | 120 | 113 | 106 | 99 | 93 | 88 | 64 | 48 |
| 2.0 | 94 | 93 | 91 | 88 | 82 | 78 | 73 | 68 | 64 | 60 | 57 | 54 | 50 | 38 | 28 |
| 10.0 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 |

| $m\gamma$ \ $m\delta$ | 0.00 | 0.01 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.5 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $p=30, q=5$ | | | | | | | | | | | | | | |
| 0.0 | 435 | 429 | 405 | 377 | 325 | 279 | 237 | 201 | 170 | 143 | 120 | 100 | 83 | 32 | 12 |
| 0.1 | 381 | 376 | 354 | 329 | 282 | 241 | 204 | 173 | 145 | 122 | 102 | 85 | 70 | 27 | 10 |
| 0.5 | 219 | 216 | 202 | 186 | 158 | 133 | 111 | 93 | 78 | 64 | 54 | 44 | 36 | 14 | 5 |
| 1.0 | 107 | 106 | 98 | 90 | 75 | 63 | 52 | 43 | 36 | 29 | 24 | 20 | 16 | 6 | 2 |
| 2.0 | 26 | 25 | 24 | 21 | 18 | 14 | 12 | 10 | 8 | 7 | 5 | 4 | 4 | 1 | 0 |
| | $p=30, q=10$ | | | | | | | | | | | | | | |
| 0.0 | 465 | 458 | 426 | 388 | 318 | 256 | 203 | 158 | 122 | 93 | 70 | 52 | 38 | 8 | 1 |
| 0.1 | 394 | 386 | 358 | 324 | 262 | 208 | 164 | 126 | 97 | 73 | 55 | 40 | 30 | 6 | 1 |
| 0.5 | 188 | 184 | 167 | 148 | 116 | 89 | 68 | 51 | 38 | 28 | 20 | 15 | 11 | 2 | 0 |
| 1.0 | 70 | 68 | 61 | 53 | 40 | 30 | 22 | 16 | 12 | 9 | 6 | 4 | 3 | 0 | 0 |
| 2.0 | 10 | 9 | 8 | 7 | 5 | 4 | 3 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| | $p=30, q=30$ | | | | | | | | | | | | | | |
| 0.0 | 500 | 489 | 444 | 390 | 288 | 201 | 133 | 84 | 50 | 28 | 15 | 8 | 4 | | |
| 0.1 | 398 | 388 | 347 | 299 | 214 | 145 | 93 | 57 | 33 | 18 | 10 | 5 | 2 | | |
| 0.5 | 136 | 131 | 112 | 92 | 59 | 36 | 21 | 12 | 6 | 3 | 2 | 1 | 0 | | |
| 1.0 | 31 | 30 | 25 | 19 | 11 | 6 | 4 | 2 | 1 | 0 | 0 | 0 | 0 | | |
| 2.0 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| | $p=30, q=50$ | | | | | | | | | | | | | | |
| 0.0 | 511 | 498 | 448 | 387 | 272 | 176 | 105 | 57 | 29 | 13 | 6 | 2 | 1 | | |
| 0.1 | 397 | 385 | 340 | 287 | 192 | 119 | 68 | 36 | 17 | 8 | 3 | 1 | 0 | | |
| 0.5 | 119 | 114 | 94 | 74 | 43 | 23 | 12 | 6 | 2 | 1 | 0 | 0 | 0 | | |
| 1.0 | 22 | 21 | 17 | 12 | 6 | 3 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | | |
| 2.0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| | $p=30, q=100$ | | | | | | | | | | | | | | |
| 0.0 | 521 | 507 | 452 | 383 | 255 | 150 | 77 | 34 | 13 | 4 | 1 | | | | |
| 0.1 | 395 | 382 | 332 | 273 | 170 | 94 | 46 | 19 | 7 | 2 | 0 | | | | |
| 0.5 | 102 | 97 | 78 | 59 | 30 | 14 | 6 | 2 | 1 | 0 | 0 | | | | |
| 1.0 | 16 | 15 | 12 | 8 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | | | | |
| | $p=50, q=1$ | | | | | | | | | | | | | | |
| 0.0 | 322 | 320 | 310 | 298 | 277 | 257 | 240 | 223 | 208 | 194 | 182 | 170 | 159 | 115 | 84 |
| 0.1 | 294 | 292 | 284 | 273 | 258 | 240 | 224 | 209 | 195 | 182 | 170 | 159 | 149 | 108 | 79 |
| 0.5 | 226 | 225 | 219 | 211 | 197 | 184 | 172 | 161 | 150 | 140 | 131 | 124 | 116 | 85 | 62 |
| 1.0 | 164 | 162 | 158 | 153 | 143 | 134 | 126 | 118 | 110 | 104 | 97 | 91 | 86 | 63 | 47 |
| 2.0 | 89 | 89 | 87 | 84 | 79 | 74 | 70 | 66 | 62 | 58 | 55 | 52 | 48 | 36 | 27 |
| 10.0 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | $p=50, q=5$ | | | | | | | | | | | | | | |
| 0.0 | 428 | 422 | 398 | 370 | 318 | 272 | 231 | 195 | 165 | 138 | 116 | 96 | 80 | 31 | 11 |
| 0.1 | 372 | 367 | 346 | 320 | 274 | 233 | 197 | 166 | 140 | 117 | 98 | 81 | 67 | 26 | 9 |
| 0.5 | 207 | 203 | 190 | 175 | 147 | 124 | 103 | 86 | 72 | 59 | 49 | 40 | 33 | 12 | 4 |
| 1.0 | 95 | 93 | 87 | 79 | 66 | 55 | 45 | 37 | 31 | 25 | 21 | 17 | 14 | 5 | 2 |
| 2.0 | 19 | 19 | 17 | 16 | 13 | 10 | 9 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |

| $m\gamma$ \ $m\delta$ | 0.00 | 0.01 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.5 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $p=50, q=10$ | | | | | | | | | | | | | | |
| 0.0 | 456 | 448 | 416 | 377 | 306 | 244 | 192 | 148 | 113 | 86 | 64 | 48 | 35 | 7 | 1 |
| 0.1 | 381 | 373 | 344 | 310 | 248 | 195 | 152 | 116 | 88 | 66 | 49 | 36 | 26 | 5 | 1 |
| 0.5 | 167 | 163 | 148 | 130 | 100 | 76 | 57 | 42 | 31 | 23 | 16 | 12 | 8 | 2 | 0 |
| 1.0 | 53 | 52 | 46 | 40 | 30 | 22 | 16 | 12 | 8 | 6 | 4 | 3 | 2 | 0 | 0 |
| 2.0 | 5 | 5 | 4 | 4 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| | $p=50, q=30$ | | | | | | | | | | | | | | |
| 0.0 | 489 | 477 | 428 | 368 | 261 | 175 | 111 | 66 | 38 | 20 | 11 | 5 | 2 | | |
| 0.1 | 375 | 364 | 321 | 270 | 184 | 118 | 72 | 42 | 23 | 12 | 6 | 3 | 1 | | |
| 0.5 | 101 | 96 | 80 | 63 | 38 | 22 | 12 | 6 | 3 | 2 | 1 | 0 | 0 | | |
| 1.0 | 15 | 14 | 11 | 8 | 4 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | | |
| | $p=50, q=50$ | | | | | | | | | | | | | | |
| 0.0 | 500 | 486 | 429 | 360 | 237 | 142 | 78 | 39 | 18 | 7 | 3 | 1 | | | |
| 0.1 | 369 | 356 | 306 | 249 | 154 | 87 | 45 | 21 | 9 | 4 | 1 | 0 | | | |
| 0.5 | 78 | 73 | 58 | 43 | 22 | 10 | 5 | 2 | 1 | 0 | 0 | 0 | | | |
| 1.0 | 8 | 7 | 6 | 4 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | | | |
| | $p=50, q=100$ | | | | | | | | | | | | | | |
| 0.0 | 511 | 494 | 429 | 348 | 208 | 106 | 46 | 17 | 5 | 1 | | | | | |
| 0.1 | 360 | 345 | 289 | 224 | 122 | 57 | 23 | 8 | 2 | 0 | | | | | |
| 0.5 | 57 | 53 | 40 | 27 | 11 | 4 | 1 | 0 | 0 | 0 | | | | | |
| 1.0 | 4 | 4 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | | | | | |
| | $p=100, q=1$ | | | | | | | | | | | | | | |
| 0.0 | 320 | 317 | 308 | 296 | 275 | 256 | 238 | 222 | 207 | 193 | 181 | 169 | 158 | 114 | 84 |
| 0.1 | 292 | 290 | 281 | 271 | 252 | 235 | 218 | 204 | 190 | 178 | 169 | 158 | 148 | 107 | 79 |
| 0.5 | 224 | 222 | 216 | 208 | 195 | 182 | 170 | 159 | 149 | 140 | 131 | 123 | 115 | 84 | 62 |
| 1.0 | 160 | 159 | 155 | 150 | 141 | 132 | 124 | 116 | 109 | 102 | 96 | 90 | 85 | 62 | 46 |
| 2.0 | 86 | 86 | 84 | 81 | 76 | 72 | 68 | 64 | 60 | 56 | 53 | 50 | 47 | 35 | 26 |
| 10.0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| | $p=100, q=5$ | | | | | | | | | | | | | | |
| 0.0 | 422 | 416 | 392 | 364 | 312 | 266 | 226 | 191 | 160 | 134 | 112 | 94 | 78 | 30 | 11 |
| 0.1 | 365 | 360 | 339 | 313 | 267 | 227 | 192 | 161 | 135 | 113 | 94 | 78 | 65 | 24 | 9 |
| 0.5 | 196 | 193 | 181 | 166 | 139 | 116 | 97 | 81 | 67 | 55 | 46 | 38 | 31 | 11 | 4 |
| 1.0 | 85 | 84 | 78 | 71 | 59 | 48 | 40 | 33 | 27 | 22 | 18 | 15 | 12 | 4 | 2 |
| 2.0 | 14 | 14 | 13 | 12 | 10 | 8 | 6 | 5 | 4 | 3 | 3 | 2 | 2 | 1 | 0 |
| | $p=100, q=10$ | | | | | | | | | | | | | | |
| 0.0 | 449 | 440 | 407 | 368 | 296 | 234 | 182 | 140 | 107 | 80 | 60 | 44 | 32 | 6 | 1 |
| 0.1 | 370 | 362 | 333 | 298 | 236 | 184 | 142 | 108 | 81 | 61 | 45 | 33 | 24 | 4 | 1 |
| 0.5 | 150 | 147 | 132 | 115 | 87 | 65 | 48 | 36 | 26 | 19 | 14 | 10 | 7 | 1 | 0 |
| 1.0 | 41 | 40 | 35 | 30 | 22 | 16 | 11 | 8 | 6 | 4 | 3 | 2 | 1 | 0 | 0 |
| 2.0 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

| $m\delta$ / $my$ | 0.00 | 0.01 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0 | 1.5 | 2.0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $p=100,\ q=30$ | | | | | | | | | | | | | | |
| 0.0 | 479 | 465 | 412 | 348 | 237 | 152 | 92 | 53 | 29 | 15 | 7 | 4 | 2 | | |
| 0.1 | 353 | 341 | 295 | 243 | 157 | 96 | 56 | 31 | 16 | 8 | 4 | 2 | 1 | | |
| 0.5 | 70 | 67 | 54 | 41 | 22 | 12 | 6 | 3 | 1 | 1 | 0 | 0 | 0 | | |
| 1.0 | 6 | 5 | 4 | 3 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | |
| | $p=100,\ q=50$ | | | | | | | | | | | | | | |
| 0.0 | 489 | 473 | 408 | 332 | 203 | 112 | 55 | 25 | 10 | 4 | 1 | | | | |
| 0.1 | 338 | 324 | 270 | 211 | 118 | 60 | 28 | 12 | 5 | 2 | 1 | | | | |
| 0.5 | 44 | 41 | 30 | 21 | 9 | 4 | 1 | 0 | 0 | 0 | 0 | | | | |
| 1.0 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| | $p=100,\ q=100$ | | | | | | | | | | | | | | |
| 0.0 | 500 | 480 | 401 | 307 | 158 | 66 | 23 | 7 | 2 | | | | | | |
| 0.1 | 317 | 300 | 237 | 170 | 76 | 28 | 9 | 2 | 0 | | | | | | |
| 0.5 | 22 | 20 | 13 | 8 | 2 | 0 | 0 | 0 | 0 | | | | | | |

## 5. The Intra-Class Correlation Model

In § 2 we have seen that the observations within the same sample are correlated, $\operatorname{cov}(\underline{y}_{ij}, \underline{y}_{ij'}) = \sigma_1^2$ for $j \neq j'$; hence the correlation coefficient $\varrho(\underline{y}_{ij}, \underline{y}_{ij'}) = \sigma_1^2/(\sigma_1^2 + \sigma_2^2)$. The observations from several samples are uncorrelated. In our variance component model the correlation coefficient $\varrho$ is non-negative by definition.

In several applications however, especially in plant breeding or animal breeding, one wants a model where the observations within the same sample may be negatively correlated. It is frequently employed in studies about twins. The following intra-class correlation model has been proposed by R. A. FISHER long before the variance component model.

From a population $n_1$ samples (groups) are drawn at random. From each group $n_2$ elements are drawn. Let $y_{ij}$ be the observation at the $j$-th element drawn from the $i$-th group. Then the model (3) is now:

$$\underline{y}_{ij} = \mu + \underline{e}_{ij} \qquad \begin{matrix} i = 1, \ldots, n_1 \\ j = 1, \ldots, n_2; \end{matrix}$$

where

(3) $$\mathrm{E}\underline{e}_{ij} = 0, \ \operatorname{var}(\underline{e}_{ij}) = \sigma^2$$

$$\operatorname{cov}(\underline{e}_{ij}, \underline{e}_{i'j'}) = \delta_{ii'}\, \varrho\sigma^2 \quad \text{for} \quad j \neq j' .$$

The parameter $\varrho$ is called the intra-class correlation coefficient.

Let

$$y' = (y_{11}, y_{12}, \ldots, y_{1n_2}, \quad y_{21}, \ldots, y_{n_1 n_2}) \in \boldsymbol{R}^n$$

and

$$e' = (e_{11}, e_{12}, \ldots, e_{1n_2}, \quad e_{21}, \ldots, e_{n_1 n_2}) \in \boldsymbol{R}^n$$

with $n=n_1n_2$; $\boldsymbol{I}_n$ a column vector with coordinates of merely ones, then we have in vectorform: $\underline{y}=\mu\boldsymbol{I}_n+\underline{\varepsilon}$ with non-singular covariance matrix of $\underline{y}$:

$$V=\sigma^2 I_{n_1}\times\{(1-\varrho)\, I_{n_2}+\varrho J_{n_2}\}\,.$$

From the requirement of positive-definiteness of $V$ it follows that $-1/(n_2-1)<\varrho<1$. The proof is as follows. The positive definite matrix $V$ has only positive eigenvalues. The eigenvalues of $V$ are found as the product of the eigenvalues of $\sigma^2 I_{n_1}$ and those of $\{(1-\varrho)\, I_{n_2}+\varrho J_{n_2}\}$. The latter are to be found from $|(1-\varrho)\, I_{n_2}+\varrho J_{n_2}-\lambda I_{n_2}|=0$ or

$$(1-\lambda+(n_2-1)\,\varrho)\,(1-\varrho-\lambda)^{n_2-1}=0\,.$$

Hence those eigenvalues are $\lambda_1=1+(n_2-1)\,\varrho$ and $\lambda_2=(1-\varrho)$ with multiplicity $n_2-1$. The eigenvalues of $V$ are therefore $\sigma^2(1+(n_2-1)\,\varrho)$ with multiplicity $n_1$ and $\sigma^2(1-\varrho)$ with multiplicity $n_1(n_2-1)$. Since they must be positive it follows that $-1/(n_2-1)<\varrho<1$.

If we adopt the assumption of normally distributed observations, then we have the model:

$$\underline{y}_{ij}=\mu+\underline{e}_{ij}$$

$\underline{e}_{ij}$ are identical $N(0,\sigma^2)$ for

$$(4)\qquad i=1,\ldots,n_1;\ j=1,\ldots,n_2;$$
$$\operatorname{cov}(\underline{y}_{ij},\underline{y}_{i'j'})=\delta_{ii'}\,\varrho\sigma^2 \quad \text{for} \quad j\neq j'\,.$$

The estimator of $\mu$, given by $\bar{\underline{y}}=\sum_i\sum_j \underline{y}_{ij}/(n_1n_2)$ is for model (3) the minimum variance unbiased linear estimator. For model (4) this estimator $\bar{\underline{y}}$ has minimum variance among all unbiased estimators. This follows from the fact that $\bar{\underline{y}}$ is a minimal sufficient complete statistic of the normally distributed observations with respect to $\mu$.

For the estimation of the relevant parameters we use the same analysis of variance table as in § 3 but the expected mean squares are now

$$\mathrm{E}\,(\underline{\mathrm{MSB}})=\sigma^2\,(1+(n_2-1)\,\varrho)$$

and

$$\mathrm{E}\,(\underline{\mathrm{MSW}})=\sigma^2\,(1-\varrho)$$

respectively.

Now $\underline{\mathrm{SSW}}$ is distributed as $\sigma^2(1-\varrho)\,\underline{\chi}^2_e$ and $\underline{\mathrm{SSB}}$ as $\sigma^2(1+(n_2-1)\,\varrho)\,\underline{\chi}^2_a$.

From $\mathrm{E}\,(\underline{\mathrm{MSB}})=\sigma^2(1+(n_2-1)\,\varrho)$ and $\mathrm{E}\,(\underline{\mathrm{MSW}})=\sigma^2(1-\varrho)$ it follows that $\mathrm{E}\,(\underline{\mathrm{MSB}})<\mathrm{E}\,(\underline{\mathrm{MSW}})$, if and only if $\varrho<0$.

The probability $P\,(\underline{\mathrm{MSB}}<\underline{\mathrm{MSW}})$ is equal to $P\,(\underline{\mathrm{MSB}}/\underline{\mathrm{MSW}}<1)$ since $P(\underline{\mathrm{MSW}}=0)=0$, and $\underline{\mathrm{MSB}}/\underline{\mathrm{MSW}}$ is distributed as $(1+n_2\varrho/(1-\varrho))\,\underline{F}(a,e)$, hence $P(\underline{\mathrm{MSB}}<\underline{\mathrm{MSW}})=P(\underline{F}(a,e)<1/(1+n_2\varrho/(1-\varrho))$. Entering Table 1 with $\gamma=\varrho/(1-\varrho)$ we may read this probability numerically. Note that $P(\underline{\mathrm{MSB}}<\underline{\mathrm{MSW}})\geqslant P(\underline{F}(a,e)<$

$<1$), if $-1/(n_2-1)<\varrho<0$, with equality if $\varrho=0$. We can also read from Table 1 the probability $P(\underline{\text{MSB}}-\underline{\text{MSW}}<-\delta n_2\sigma^2(1-\varrho))=$

$$P\left(\frac{1+n_2\gamma}{n_2 a}\,\underline{\chi}_a^2-\frac{1}{n_2 e}\,\underline{\chi}_e^2\leqq-\delta\right),\quad\text{with}\quad\delta>0\,.$$

From the analysis of variance table it follows that ($\underline{\text{MSB}}-\underline{\text{MSW}}$) is an unbiased estimator for $n_2\varrho\sigma^2$, and ($\underline{\text{MSB}}+(n_2-1)\,\underline{\text{MSW}}$) for $n_2\sigma^2$. SNEDECOR and COCHRAN (1969) propose as an estimator for $\varrho$:

$$\underline{\hat{\varrho}}_1=(\underline{\text{MSB}}-\underline{\text{MSW}})/(\underline{\text{MSB}}+(n_2-1)\,\underline{\text{MSW}})\,.$$

Note that $\underline{\hat{\varrho}}_1$ is not unbiased.

Another estimator for $\varrho$ follows from the $n_2(n_2-1)$ ordered pairs of observations which can be formed in each group. The correlation coefficient between these pairs gives an estimate for $\varrho$. When we take the sum over all groups we get another, but also biased, estimator for $\varrho$:

$$\underline{\hat{\varrho}}_2=\frac{(n_1-1)\,\underline{\text{MSB}}-n_1\underline{\text{MSW}}}{(n_1-1)\,\underline{\text{MSB}}+n_1(n_2-1)\,\underline{\text{MSW}}}$$

Since $\underline{\hat{\varrho}}_2$ differs only slightly from $\underline{\hat{\varrho}}_1$, $\underline{\hat{\varrho}}_1$ is the usual estimator in practice.

## 6. A Non-Negative Estimation for $\sigma_1^2$

Finally we want to remark that even if all assumptions of model (1) or (2) are satisfied, the analysis of variance estimator is not a desirable estimator for $\sigma_1^2$. The variance component $\sigma_1^2$ is non-negative and one would prefer an estimator which can not give negative outcomes. The problem of negative outcomes of $\underline{\hat{\sigma}}_1^2$ stems from the demand of unbiasedness. In the theory of estimation, the estimators $\underline{t}$ of a parameter $\vartheta$ can be compared by the second moment with respect to $\vartheta$.

$$\mathrm{E}\,(\underline{t}-\vartheta)^2=\mathrm{Var}\,(\underline{t})+(\mathrm{E}\underline{t}-\vartheta)^2\,.$$

If an estimator $\underline{t}$ is unbiased, then $\mathrm{E}\underline{t}=\vartheta$ and $\mathrm{E}\,(\underline{t}-\vartheta)^2=\mathrm{Var}\,(\underline{t})$. But there are many examples of estimators known, which have a smaller second moment than the minimum variance unbiased estimator. For example, the estimator $\underline{\text{SSW}}/e$ for $\sigma_2^2$ in model (2) has within the class of unbiased estimators minimal variance. Another estimator with smallest second moment with respect to $\sigma_2^2$ is $\underline{\text{SSW}}/(e+2)$. (For model (1) we can only say that $\underline{\text{SSW}}/(e+2)$ has smallest second moment within the class of all quadratic functions of the observations).

For the variance component $\sigma_1^2$ an estimator with smaller second moment with respect to $\sigma_1^2$ than the analysis of variance estimator $\underline{\hat{\sigma}}_1^2$ is the truncated estimator of $\underline{\hat{\sigma}}_1^2$, that is the estimator $\begin{cases}(\underline{\text{MSB}}-\underline{\text{MSW}})/n_2 & \text{if } \text{MSB}\geqq\text{MSW}\\ 0 & \text{if } \text{MSB}<\text{MSW}\,.\end{cases}$

This truncated estimator is a biased, but non-negative estimator for $\sigma_1^2$, which has smaller second moment than $\hat{\sigma}_1^2$.

Another non-negative estimator for $\sigma_1^2$, which is also biased, is given by the maximum likelihood estimator. This estimator can only be calculated if the probability distributions of the random components in the model are completely specified, such as for example in model (2). The truncated $\hat{\sigma}_1^2$-estimator can be used even in model (1). For another non-negative estimator see VERDOOREN (1980).

## Appendix: Kronecker Product

Let $A=(a_{ij})$ and $B=(b_{ij})$ be $(m_1 \times m_2)$ and $(n_1 \times n_2)$ matrices, respectively. Then the KRONECKER product $A \times B=(a_{ij}B)$ is an $(m_1 n_1 \times m_2 n_2)$ matrix expressible as a partitioned matrix with $a_{ij}B$ as the $(i,j)$-th partition, $i=1, \ldots, m_1$ and $j=1, \ldots, m_2$.

The following results are consequences of the definition (see e.g. NEUDECKER (1968) and RAO (1973)):

a) $O \times A = A \times O = O$ ($O$ = zero matrix)
b) $(A_1+A_2) \times B = (A_1 \times B)+(A_2 \times B)$
c) $A \times (B_1+B_2) = (A \times B_1)+(A \times B_2)$
d) $aA \times bB = abA \times B$
e) $(A_1A_2) \times (B_1B_2) = (A_1 \times B_1)\,(A_2 \times B_2)$, if the matrix products $A_1A_2$ and $B_1B_2$ exist.
f) $(A \times B)' = A' \times B'$

In case $A$ and $B$ are square of order $(m \times m)$ and $(n \times n)$ respectively the following results hold:

(i) $(A \times B)^{-1} = A^{-1} \times B^{-1}$, if the inverses exist.
(ii) $(A \times B)\,(A^{-1} \times B^{-1}) = I_m \times I_n = I_{mn}$, if the inverses exist.
(iii) Let $A$ have eigenvalues $\alpha_i$, eigencolumnvectors $u_i$ and eigenrowvectors $v_i'(i=1, \ldots, m)$:
Let $B$ have eigenvalues $\beta_j$, eigencolumnvectors $w_j$ and eigenrowvectors $z_j'(j=1, \ldots, n)$;
Then $A \times B$ has eigenvalues $\alpha_i\beta_j$, eigencolumnvectors $u_i \times w_j$ and eigenrowvectors $v_i' \times z_j'(i=1, \ldots, m;\ j=1, \ldots, n)$.

*Proof of (iii)*

Use result e):

$$(A \times B)\,(u_i \times w_j) = (Au_i) \times (Bw_j) = (\alpha_i u_i) \times (\beta_j w_j) = \alpha_i\beta_j (u_i \times w_j)\,.$$
$$(v_i' \times z_j')\,(A \times B) = (v_i'A) \times (z_j'B) = (\alpha_i v_i') \times (\beta_j z_j') = \alpha_i\beta_j\,(v_i' \times z_j')\,.$$

(iv) $|A \times B| = |A|^n |B|^m$

*Proof of (iv)*

$A \times B$ has eigenvalues $\alpha_i \beta_j$ $(i = 1, \ldots, m; j = 1, \ldots, n)$. So

$$A \times B = \prod_{i=1}^{m} \prod_{j=1}^{n} \alpha_i \beta_j = \prod_{i=1}^{m} \left( \alpha_i^n \prod_{j=1}^{n} \beta_j \right) = \left( \prod_{i=1}^{m} \alpha_i^n \right) \left( \prod_{j=1}^{n} \beta_j^m \right) = |A|^n |B|^m .$$

(v) $\operatorname{tr}(A \times B) = (\operatorname{tr} A)(\operatorname{tr} B)$

*Proof of (v)*

$$\operatorname{tr}(A \times B) = \sum_i \sum_j \alpha_i \beta_j = \left( \sum_i \alpha_i \right) \left( \sum_j \beta_j \right) = (\operatorname{tr} A)(\operatorname{tr} B)$$

(vi) $\varrho(A \times B) = \varrho(A) \cdot \varrho(B)$ where $\varrho(A)$ is the rank of $A$.

*Proof of (vi)*

$\varrho(A)$ = number of non-zero eigenvalues $\alpha_i$ of $A$; $\varrho(B)$ = number of non-zero eigenvalues $\beta_j$ of $B$. Hence the number of non-zero eigenvalues $\alpha_i \beta_j$ of $A \times B$ is $\varrho(A) \cdot \varrho(B)$.

Summary

Consider a balanced two-stage sampling scheme where the observations obey a random model. Under the assumption of normally distributed observations, the probability is given that the estimate of a variance component will be negative.

Zusammenfassung

Es wurde eine balanzierte zwei-stufige Stichprobenerhebung betrachtet, wobei die Beobachtungen mittels eines stochastischen Modells beschrieben werden können. Vorausgesetzt, die Beobachtungen seien normal verteilt, so wird die Wahrscheinlichkeit gegeben, daß die Schätzung einer Varianzkomponente negativ ist.

References

GRAYBILL, F. A., 1954: On quadratic estimates of variance components, Ann. Math. Statist. **25**, 367–372.

GRAYBILL, F. A., WORTHAM, A. W., 1956: A note on uniformly best unbiased estimators for variance components. J. Amer. Statist. Assoc. **51**, 266–268.

IMHOF, J. P., 1961: Computing the distribution of quadratic forms in normal variables. Biometrika **48**, 419–426, Correction, Biometrika **49**, (1962), 284.

NEUDECKER, H., 1968: The KRONECKER matrix product and some of its applications in econometrics. Statistica Neerlandica **22**, 69–82.

RAO, C. R., 1973: Linear statistical inference and its applications, 2nd edition, John Wiley and Sons, New York.

READ, R. R., 1961: On quadratic estimates of the interclass variance for unbalanced designs, J. Roy. Statist. Soc., series B, **23**, 493–497.

SNEDECOR, G. W., COCHRAN, W. G., 1969: Statistical Methods, 6th ed., The Iowa State College Press, Ames, Iowa.

VERDOOREN, L. R., 1980: On estimation of variance components, Statistica Neerlandica 34, 83–106.

Manuscript received: 18. 12. 1980

Author's address:

L. R. VERDOOREN
Department of Mathematics
Agricultural University
De Dreijen 8, 6703 BC Wageningen
the Netherlands

CHAPTER 7

# 7. EXACT TESTS AND CONFIDENCE INTERVALS FOR VARIANCE RATIOS IN NESTED DESIGNS

PART II: CONTINUATION

## 7.1. Introduction

In chapter 5 the onset of the author's research on exact tests in unbalanced two- and three-stage nested designs was given. A systematic derivation of the distribution of sums of squares in the Analysis of Variance table is given in section 7.2. For this derivation a fruitful use had been made of vectors, vector spaces and orthogonal projection of a vector on a vector space. The notions presented here are extensions of the concepts introduced in chapter 3.

Furthermore the author introduces, in section 7.2, a new exact test concerning the ratios of variance components in three-stage nested designs. For the three-stage nested design there are two interesting ratios of variance components, $\rho_1 = \sigma_A^2/\sigma_e^2$ and $\rho_2 = \sigma_B^2/\sigma_e^2$, where $\sigma_A^2$ and $\sigma_B^2$ are the variance components in the first and second stages respectivily, and $\sigma_e^2$ is the error variance. The new test on $H_{01}$: $\rho_1 = \rho_{10}$ against $H_{11}$: $\rho_1 > \rho_{10}$ depends on $\rho_2$ and the author's approach is to test $\rho_1$ after fixing $\rho_2$. In practice the researcher already has a priori knowledge about $\rho_2$; a range of plausible $\rho_2$ has already been estimated from previous experiments or can be found in the literature. The situation is the same in animal breeding, where the Best Linear Unbiased Prediction (BLUP) of breeding values depends on plausible values for ratios of variance components. These ratios are estimated (sometimes not so well) from previous experiments and are regularly updated.

In section 7.3 an example is given in which the performance of the new exact test is demonstrated.

## 7.2. Exact tests and confidence intervals for ratio of variance components in unbalanced two- and three-stage nested designs

Paper presented at the International Conference on the "Analysis of the Unbalanced Mixed Model" at Gainesville, Florida, U.S.A., in 1987.

Submitted for publication to:

Communications in Statistics, Part A: Theory and methods.

Submitted April, 1987;

Revised October, 1987;

Accepted November, 1987.

# EXACT TESTS AND CONFIDENCE INTERVALS FOR RATIO OF VARIANCE COMPONENTS IN UNBALANCED TWO- AND THREE-STAGE NESTED DESIGNS

L.R. Verdooren

Department of Mathematics
Agricultural University Wageningen
The Netherlands

Key Words and Phrases: projections on vector spaces; exact distribution of ratio of SS; Wald's procedure.

## ABSTRACT

For the unbalanced two-stage nested design an exact size $\alpha$ test for $H_0$: $\rho \leq \rho_0$ against $H_1$: $\rho > \rho_0$ exists, where $\rho = \sigma_B^2/\sigma_e^2$ with $\sigma_B^2$ the variance component in the first sampling stage and $\sigma_e^2$ the error variance in the second stage. Using Wald's procedure one can construct an exact $(1-\alpha)$-confidence interval for the ratio $\rho$ of the variance components.

For the three-stage nested design there are two interesting ratios of variance components $\rho_1 = \sigma_A^2/\sigma_e^2$ and $\rho_2 = \sigma_B^2/\sigma_e^2$, where $\sigma_A^2$ and $\sigma_B^2$ are the variance components in the first stage and

second stage respectively, and $\sigma_e^2$ is the error variance. The construction of exact size $\alpha$ tests and $(1-\alpha)$-confidence intervals for $\rho_2$ is no problem using the same procedure as for the two-stage nested design. In the balanced three-stage nested design an exact size $\alpha$ test is well known for $H_0$: $\rho_1 = 0$ against $H_1$: $\rho_1 > 0$. A test for $H_0$: $\rho_1 \leq \rho_{10}$ against $H_1$: $\rho_1 > \rho_{10}$ exists but depends also on the parameter $\rho_2$. In the unbalanced three-stage nested design even the test for $H_0$: $\rho_1 = 0$ against $H_1$: $\rho_1 > 0$ depends also on $\rho_2$. Also an exact $(1-\alpha)$-confidence interval for $\rho_1$ irrespective of $\rho_2$ is not possible. Many approximate solutions were proposed in the past. One way to overcome this difficulty is to construct simultaneous confidence regions for $\rho_1$ and $\rho_2$. In this paper a different solution will be proposed, which produces exact tests and confidence intervals for $\rho_1$ given a certain value of $\rho_2$. The experimenter who is interested in a test or confidence interval for $\rho_1$ decides beforehand in which range of $\rho_2$ he is interested in his experiment to make decisions about $\rho_1$. Exact size $\alpha$ tests and $(1-\alpha)$-confidence intervals for $\rho_1$ given a value of $\rho_2$ are presented here.

## 1. INTRODUCTION

### 1.1. Model and notations

For a sampling procedure in two stages we consider a large population. In the first sampling stage b samples are drawn at random from the population. In the second sampling stage $n_i$ subsamples are drawn at random from the i-th sample.
In a balanced two-stage nested design the same number m of subsamples is drawn in the second stage from each sample, hence $n_i = m$ for $i = 1,2,\ldots,b$. Let the observation made on the k-th subsample from the i-th sample be denoted by $y_{ik}$, $k = 1,2,\ldots,n_i$ for $i = 1,2,\ldots,b$. A linear model for the random variable $y_{ik}$ is then as follows:

$$y_{ik} = \beta + u_i + e_{ik},\ k = 1,2,\ldots,n_i;\ i = 1,2,\ldots,b. \qquad (1)$$

The $u_i$'s and the $e_{ij}$'s have zero means and are statistically independent with $\text{var}(u_i) = \sigma_B^2$ and $\text{var}(e_{ik}) = \sigma_e^2$. In the sequel we make the assumption of normally distributed effects. Note that uncorrelated random effects in the case of joint normally distributed effects are equivalent to stochastic independence of the effects. A two-stage nested design is also called a one-way random model.

Let y be the column vector of all observations

$\underset{\sim}{y} = (y_{11},\ldots,y_{1n_1}, y_{21},\ldots,y_{2n_2},\ldots,y_{b1},\ldots,y_{bn_b})'$, then $\underset{\sim}{y} \in R^{n.}$

with $n. = \sum_{i=1}^{b} n_i$, $u = (u_1,\ldots,u_b)' \in R^b$, $\underset{\sim}{e} = (e_{11},\ldots,e_{bn_b})' \in R^{n.}$.

The model can now be written as

$$\underset{\sim}{y} = X\beta + \underset{\sim}{Zu} + \underset{\sim}{e} \qquad (2)$$

where X is a n.x1-column vector of ones, $\underset{\sim}{1}_{n.}$, Z is an n.xb design matrix related to the first sampling stage. Let $\underset{\sim}{1}_{n_i}$ be a column vector of $n_i$ ones then Z can be described as $\text{diag}(\underset{\sim}{1}_{n_1}, \underset{\sim}{1}_{n_2},\ldots,\underset{\sim}{1}_{n_b})$, where diag means a diagonal block matrix. Let $I_b$ be the bxb identity matrix then $\underset{\sim}{u}$ has a multivariate normal distribution $N(\underset{\sim}{0},\sigma_B^2 I_b)$ and $\underset{\sim}{e}$ is $N(\underset{\sim}{0},\sigma_e^2 I_{n.})$. The dispersion matrix of $\underset{\sim}{y}$ (or covariance matrix of $\underset{\sim}{y}$), $D(\underset{\sim}{y}) = E[(\underset{\sim}{y}-E\underset{\sim}{y})(\underset{\sim}{y}-E\underset{\sim}{y})']$ where $E\underset{\sim}{y} = \underset{\sim}{1}_{n.}\beta$ can be written as

$$D(\underset{\sim}{y}) = E[(\underset{\sim}{Zu}+\underset{\sim}{e})(\underset{\sim}{Zu}+\underset{\sim}{e})'] = \sigma_B^2 ZZ' + \sigma_e^2 I_{n.} = \sigma_e^2(\rho ZZ' + I_{n.}) = \sigma_e^2 V \qquad (3)$$

with $\rho = \sigma_B^2/\sigma_e^2$, the ratio of the variance components. Hence $\underset{\sim}{y}$ has a multivariate normal distribution $N(1_{n..}\beta, \sigma_e^2 V)$ with V depending on $\rho$.

Analogously we derive a model for a three-stage nested sampling procedure by taking a random sample of size a in the first sampling stage, $b_h$ random subsamples from the h-th sample and $n_{hi}$ random subsubsamples in the third sampling stage from the i-th subsample of the h-th sample. Let $y_{hik}$ be the observation of the k-th subsubsample of the h-th subsample of the i-th sample ($k = 1,2,\ldots,n_{hi}$; $i = 1,2,\ldots,b_h$; $h = 1,2,\ldots,a$), then the linear model is

$$y_{hik} = \beta + u_{1h} + u_{2hi} + e_{hik} \tag{4}$$

with $Eu_{1h} = Eu_{2hi} = Ee_{hik} = 0$, $var(u_{1h}) = \sigma_A^2$, $var(u_{2hi}) = \sigma_B^2$, $var(e_{hik}) = \sigma_e^2$, and uncorrelated random effects $u_{1h}$, $u_{2hi}$, $e_{hik}$.

We again assume normally distributed random effects.
Let $\underset{\sim}{y}$ be the vector of observations in ordered form with last index running first. We have

$$\underset{\sim}{y} = X\beta + Z_1\underset{\sim}{u}_1 + Z_2\underset{\sim}{u}_2 + \underset{\sim}{e} \tag{5}$$

where $X = 1_{n..}$, $n_{..} = \sum_{h=1}^{a}\sum_{i=1}^{b_h} n_{hi}$; $Z_1 = \mathrm{diag}(1_{n_{1.}},\ldots,1_{n_{a.}})$ of order $n_{..}xa$, $n_{h.} = \sum_{i=1}^{b_h} n_{hi}$; $Z_2 = \mathrm{diag}(1_{n_{11}},\ldots,1_{n_{a1}})$ of order $n_{..}xb_.$,

$b_{\cdot} = \sum_{h=1}^{a} b_h$. Now $\underset{\sim}{u}_1$ is $N(\underset{\sim}{0},\sigma_A^2 I_a)$, $\underset{\sim}{u}_2$ is $N(\underset{\sim}{0},\sigma_B^2 I_{b_{\cdot}})$, $\underset{\sim}{e}$ is $N(0,\sigma_e^2 I_{n..})$ and $\underset{\sim}{u}_1$, $\underset{\sim}{u}_2$ and $\underset{\sim}{e}$ are independent. Hence $\underset{\sim}{y}$ has a multivariate normal distribution $N(\underset{\sim}{1}_{n..}\beta,\sigma_e^2 V)$ with dispersion matrix

$$D(\underset{\sim}{y}) = \sigma_e^2 V = \sigma_e^2(\rho_1 Z_1 Z_1' + \rho_2 Z_2 Z_2' + I_{n..}) \qquad (6)$$

and with $\rho_1 = \sigma_A^2/\sigma_e^2$, $\rho_2 = \sigma_B^2/\sigma_e^2$.

A three-stage nested design is called balanced if $b_h = m_1$ for $h = 1,2,\ldots,a$ and $n_{hi} = m_2$ for $h = 1,2,\ldots,a$ and $i = 1,2,\ldots,b_h$. The design is called partially balanced if $n_{hi} = m_h$ for all $i$, but $b_h$ is not the same for all $h = 1,2,\ldots,a$.

When we look at the model (2) for a two-stage nested design and at model (5) for a three-stage nested design we see that these models are the simplest examples of a general mixed model

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + Z_1\underset{\sim}{u}_1 + \ldots + Z_k\underset{\sim}{u}_k + \underset{\sim}{e}$$

with $D(\underset{\sim}{y}) = \sigma_1^2 Z_1 Z_1' + \ldots + \sigma_k^2 Z_k Z_k' + \sigma_e^2 I_{n..\,..}$ where $\sigma_i^2$ is the variance component belonging to the random effect vector $\underset{\sim}{u}_i$.

## 1.2. Sum of Squares in the ANOVA-table

For the description of tests and confidence intervals for the ratio of variance components we need an analysis of variance table. We will describe the calculation of the sum of squares as the squares of projections of the observation vector $\underset{\sim}{y}$ on certain

subspaces.

For the two-stage nested design we have the observation vector $\underset{\sim}{y} \in R^{n.}$ with $n. = \sum_{h=1}^{b} n_h$. Let L be the column space of $\underset{\sim}{1}_{n.}$ and B the column space of Z. The dimension of the subspace L, dim L, equals the rank of $\underset{\sim}{1}_{n.}$ hence dim L = 1; also dim B = b, the rank of Z. Notice that $L \subset B$ and define the subspace $B^*$ as the intersection $B \cap L^{\perp}$, the orthogonal complement of L in B, with dim $B^*$ = b - 1. Let C be the orthogonal complement of B in $R^{n.}$ with dim C = n.- b. The space $R^{n.}$ is now decomposed into 3 orthogonal subspaces L, $B^*$ and C. Let $P_L$ be the orthogonal projector of $\underset{\sim}{y}$ on B, then $P_B = X(X'X)^{-1}X' = n^{-1}\underset{\sim}{1}_{n.}\underset{\sim}{1}'_{n.} = n_.^{-1}J_{n.}$, where $J_{n.}$ is the n.xn. matrix with elements 1. Let $P_B$ be the orthogonal projector of $\underset{\sim}{y}$ on B, then $P_B = Z(Z'Z)^{-1}Z' = \text{diag}(B_1, B_2, \ldots, B_b)$ with $B_i = n_i^{-1}J_{n_i}$. The orthogonal projector of $\underset{\sim}{y}$ on $B^*$ is now $P_B - P_L$.

The orthogonal projector of $\underset{\sim}{y}$ on C is $P_C = I_{n.} - P_B = \text{diag}(I_{n_1}-B_1, I_{n_2}-B_2, \ldots, I_{n_b}-B_b)$. Remark that a projection matrix P is symmetric (P = P') and idempotent ($P^2=P$). For a first-stage class, $B_i$ is also a projection matrix on the space with base $\underset{\sim}{1}_{n_i}$. We can decompose the vector $\underset{\sim}{y} \in R^{n.}$ into three orthogonal components $P_L\underset{\sim}{y}$, $P_{B^*}\underset{\sim}{y}$ and $P_C\underset{\sim}{y}$ and hence (Pythagorean theorem) we have for the squared lengths of $\underset{\sim}{y}$ and its projections:

$$\underset{\sim}{y}'\underset{\sim}{y} = (P_L\underset{\sim}{y})'(P_L\underset{\sim}{y}) + (P_{B^*}\underset{\sim}{y})'(P_{B^*}\underset{\sim}{y}) + (P_C\underset{\sim}{y})'(P_C\underset{\sim}{y})$$

$$= \underset{\sim}{y}'P_L\underset{\sim}{y} + \underset{\sim}{y}'P_{B^*}\underset{\sim}{y} + \underset{\sim}{y}'P_C\underset{\sim}{y}.$$

In the Analysis of Variance table (ANOVA table) this orthogonal decomposition is used. A Sum of Squares (SS) is the square of the length of a projection on a subspace and the degrees of freedom (df) associated with such SS is the dimension of the subspace.

The ANOVA-table for a two-stage nested design is

| Source of variation | sub-space | dim df | Sum of Squares SS | Mean Squares MS |
|---|---|---|---|---|
| Between samples | $B^*$ | $b-1$ | $\underset{\sim}{y}'P_{B^*}\underset{\sim}{y} = SSB$ | $MSB = SSB/(b-1)$ |
| Within samples= Error | $C$ | $n_. - b$ | $\underset{\sim}{y}'P_C\underset{\sim}{y} = SSE$ | $MSE = SSE/(n_.-b)$ |
| Corrected total | $L^\perp$ | $n_.-1$ | $\underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{y}'P_L\underset{\sim}{y} = SST$ | |

Calculations: $\underset{\sim}{y}'\underset{\sim}{y} = \sum_{i=1}^{b}\sum_{k=1}^{n_i} y_{ik}^2;$

Correction term = $CT = \underset{\sim}{y}'P_L\underset{\sim}{y} = (\sum_{i=1}^{b}\sum_{k=1}^{n_i} y_{ik})^2/n_.;$

$SSB = \underset{\sim}{y}'P_{B^*}\underset{\sim}{y} = \underset{\sim}{y}'P_B\underset{\sim}{y} - \underset{\sim}{y}'P_L\underset{\sim}{y} = \sum_{i=1}^{b}(\sum_{k=1}^{n_i} y_{ik})^2/n_i - CT;$

$$SSE = \underset{\sim}{y}'P_C\underset{\sim}{y} = (\underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{y}'P_L\underset{\sim}{y}) - \underset{\sim}{y}'P_{B^*}\underset{\sim}{y} = \underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{y}'P_B\underset{\sim}{y}.$$

Analogously for the three-stage nested design we have the observation vector $\underset{\sim}{y} \in R^{n..}$ with $n.. = \sum_{h=1}^{a}\sum_{i=1}^{b_h} n_{hi}$, with space $L$ = column space of $1_{n..}$, $A$ = column space of $Z_1$, $B$ = column space of $Z_2$.

Notice that $L \subset A \subset B$ and define $A^* = A \cap L^{\perp}$, $B^* = B \cap A^{\perp}$ hence $\dim L = 1$, $\dim A^* = a-1$, $\dim B^* = b. - a$ with $b. = \sum_{h=1}^{a} b_h$. Let $C$ be the orthogonal complement of $B$ in $R^{n..}$, hence $\dim C = n.. - b.$.

The matrices $P_L$, $P_{A^*}$, $P_{B^*}$ and $P_C$ are the orthogonal projectors on the spaces $L$, $A^*$, $B^*$ and $C$ respectively.

Because $L$, $A^*$, $B^*$ and $C$ are orthogonal the following decomposition holds:

$$\underset{\sim}{y}'\underset{\sim}{y} = \underset{\sim}{y}'P_L\underset{\sim}{y} + \underset{\sim}{y}'P_{A^*}\underset{\sim}{y} + \underset{\sim}{y}'P_{B^*}\underset{\sim}{y} + \underset{\sim}{y}'P_C\underset{\sim}{y}.$$

Furthermore $P_L = \underset{\sim}{1}_{n..}(\underset{\sim}{1}_{n..}'\underset{\sim}{1}_{n..})^{-1}\underset{\sim}{1}_{n..}' = n..^{-1}J_{n..}$;

$$P_A = Z_1(Z_1'Z_1)^{-1}Z_1' = \mathrm{diag}(A_1, A_2, \ldots, A_a)$$

where $A_h = n_{h.}^{-1}J_{n_{h.}}$ is symmetric and idempotent; $P_{A^*} = P_A - P_L$;

$$P_B = Z_2(Z_2'Z_2)^{-1}Z_2' = \mathrm{diag}(B_{11}, \ldots, B_{1b_1}, \ldots, B_{a1}, \ldots, B_{ab_a})$$

where $B_{hi} = n_{hi}^{-1}J_{n_{hi}}$ is symmetric and idempotent;

$$P_{B^*} = P_B - P_A;\ P_C = I_{n..} - P_B = \mathrm{diag}(I_{11} - B_{11}, \ldots, I_{ab_a} - B_{ab_a}).$$

The ANOVA table for a three-stage nested design is

| Source of variation | sub-space | dim df | Sum of Squares SS | Mean Square MS |
|---|---|---|---|---|
| Between first-stage samples | $A^*$ | $a-1$ | $\underset{\sim}{y}'P_{A^*}\underset{\sim}{y} = SSA$ | MSA |
| Between second-stage within first-stage samples | $B^*$ | $b_.-a$ | $\underset{\sim}{y}'P_{B^*}\underset{\sim}{y} = SSB$ | MSB |
| Error | C | $n_{..}-b_.$ | $\underset{\sim}{y}'P_C\underset{\sim}{y} = SSE$ | MSE |
| Corrected total | $L^{\perp}$ | $n_{..}-1$ | $\underset{\sim}{y}'\underset{\sim}{y}-\underset{\sim}{y}'P_L\underset{\sim}{y} = SST$ | |

Calculations: $\underset{\sim}{y}'\underset{\sim}{y} = \sum_{h=1}^{a}\sum_{i=1}^{b_h}\sum_{k=1}^{n_{ih}} y_{hik}^2;$

$$CT = \underset{\sim}{y}'P_L\underset{\sim}{y} = \Big(\sum_{h=1}^{a}\sum_{i=1}^{b_h}\sum_{k=1}^{n_{ih}} y_{hik}\Big)^2/n_{..};$$

$$SSA = \underset{\sim}{y}'P_{A^*}\underset{\sim}{y} = \underset{\sim}{y}'P_A\underset{\sim}{y} - \underset{\sim}{y}'P_L\underset{\sim}{y} = \sum_{h=1}^{a}\Big(\sum_i\sum_k y_{hik}\Big)^2/n_{h.} - CT;$$

$$SSB = \underset{\sim}{y}'P_{B^*}\underset{\sim}{y} = \underset{\sim}{y}'P_B\underset{\sim}{y} - \underset{\sim}{y}'P_A\underset{\sim}{y} = \sum_h\sum_i\Big(\sum_k y_{hik}\Big)^2/n_{hi} - \underset{\sim}{y}'P_A\underset{\sim}{y};$$

$$SSE = \underset{\sim}{y}'P_C\underset{\sim}{y} = (\underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{y}'P_L\underset{\sim}{y}) - \underset{\sim}{y}'P_{A^*}\underset{\sim}{y} - \underset{\sim}{y}'P_{B^*}\underset{\sim}{y} = \underset{\sim}{y}'\underset{\sim}{y} - \underset{\sim}{y}'P_B\underset{\sim}{y}.$$

## 1.3. Distribution of Sum of Squares

For the derivation of the distribution of the test statistics we need the distribution of Sum of Squares in the ANOVA table. A

short review of the basic facts will be given now. For more details we refer to Searle (1971), Rao (1973) or Graybill (1976). Let $\underset{\sim}{y}$ be $N(\underset{\sim}{0},\sigma^2 I_n)$ and let the matrix A be a non-negative matrix, then the quadratic form $\underset{\sim}{y}'A\underset{\sim}{y}$ can be written as $\sum_{i=1}^{k} \lambda_i \sigma^2 z_i^2$ with $z_i$ independent N(0,1) random variables and $\lambda_i$ the k positive eigenvalues of A. Hence $\underset{\sim}{y}'A\underset{\sim}{y} = \sum_{i=1}^{k} \lambda_i \sigma^2 S_i^2$ where $S_i^2$ are independent $\chi^2$ random variables with 1 degree of freedom, $\chi^2(1)$. If $P_A$ is a projection matrix ($P_A = P_A'$, $P_A^2 = P_A$) on a subspace A of $R^n$ with dim A = k, then $P_A$ has k eigenvalues equal to 1 and n-k eigenvalues equal to 0. The quadratic form $\underset{\sim}{y}'P_A\underset{\sim}{y}$ can now be written as $\sigma^2 \sum_{i=1}^{k} S_i^2$ where $S_i^2$ are independent $\chi^2(1)$ random variables, which is a multiple of a $\chi^2$-variable with k df. Let now $\underset{\sim}{y}$ be $N(X\beta,\sigma^2 V)$, $\beta \in R^p$, where V is a nxn positive definite matrix, and let rank X be $k \leq p$, then $\underset{\sim}{y} = X\underset{\sim}{\beta} + \sigma V^{\frac{1}{2}}\underset{\sim}{x}$ with $\underset{\sim}{x}$ is $N(\underset{\sim}{0},I_n)$. ($V^{\frac{1}{2}}$ is a non-singular symmetric matrix such that $(V^{\frac{1}{2}})(V^{\frac{1}{2}}) = V$; if $HH' = H'H = I_n$ and $H'VH = \Lambda$, $\Lambda = \mathrm{diag}(\lambda_1,\lambda_2,\ldots,\lambda_n)$ with $\lambda_i > 0$, then $V^{\frac{1}{2}} = H\Lambda^{\frac{1}{2}}H'$ with $\Lambda^{\frac{1}{2}} = \mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n})$).

Let A be the orthogonal complement of the column space of X then dim A = n-k. Let $P_A$ be the orthogonal projector on A, then $P_A X = \underset{\sim}{0}$ and $P_A = I_n - X(X'X)^{-}X'$, where $(X'X)^{-}$ is a generalized inverse of X'X such that $(X'X)(X'X)^{-}(X'X) = X'X$. Now let us con-

sider the quadratic from

$$\underset{\sim}{y}'P_A\underset{\sim}{y} = (X\underset{\sim}{\beta} + \sigma V^{\frac{1}{2}}\underset{\sim}{x})'P_A(X\underset{\sim}{\beta} + \sigma V^{\frac{1}{2}}\underset{\sim}{x}) = \underset{\sim}{x}'\sigma^2V^{\frac{1}{2}}P_AV^{\frac{1}{2}}\underset{\sim}{x} = \sum_{i=1}^{n-k} \lambda_i\sigma^2S_i^2,$$

where $S_i^2$ are independent $\chi^2(1)$ variables and $\lambda_i$ are the non-zero eigenvalues of $V^{\frac{1}{2}}P_AV^{\frac{1}{2}}$. The set of positive eigenvalues of $V^{\frac{1}{2}}P_AV^{\frac{1}{2}}$ is the same as that of $VP_A$ or $P_AV$.

In the ANOVA table for nested designs a Sum of Squares is of the form $\underset{\sim}{y}'P_D\underset{\sim}{y}$ with $P_D$ a projection matrix such that $P_DX = 0$, hence $\underset{\sim}{y}'P_D\underset{\sim}{y}$ is distributed as a linear combination of d independent chi-square variables with one degree of freedom, where d is the rank of $P_D$ or dim D = d.

We also consider general quadratic forms $\underset{\sim}{y}'Q\underset{\sim}{y}$ where $Q\underset{\sim}{X} = \underset{\sim}{0}$ with Q symmetric but not necessarily semi-positive definite. The distribution of $\underset{\sim}{y}'Q\underset{\sim}{y}$ is that of $\sum_{i=1}^{q} \lambda_i\sigma^2S_i^2$ where $\lambda_i$ are the non-zero (positive and negative) eigenvalues of $V^{\frac{1}{2}}QV^{\frac{1}{2}}$ or equivalently of VQ or QV, $S_i^2$ are independent $\chi^2(1)$ random variables.

## 1.4. Eigenvalues of the dispersion matrix $D(\underset{\sim}{y})$

The dispersion matrix $D(\underset{\sim}{y})$ for the two-stage nested design is given in (3):

$D(\underset{\sim}{y}) = \sigma_e^2(\rho ZZ' + I_n) = \sigma_e^2V$. But $V = \text{diag}(V_1, V_2, \ldots, V_b)$ where

$$V_i = \rho J_{n_i} + I_{n_i} = \rho n_i B_i + B_i + I_{n_i} - B_i = (1+\rho n_i)B_i + (I_{n_i} - B_i)$$

where $B_i$ and $I_{n_i}-B_i$ are idempotent matrices with $B_i = n_i^{-1}J_{n_i}$ $(i=1,\ldots,b)$; and $B_i(I_{n_i}-B_i) = 0$, the zero-matrix.
Hence $V_i$ has one eigenvalue $(1+\rho n_i)$ for eigenvector $1_{n_i}$ and $n_i-1$ eigenvalues 1 with eigenvectors belonging to the orthogonal complement of the space spanned by $1_{n_i}$. Hence V has in total b eigenvalues $1+\rho n_1,\ldots,1+\rho n_b$, and $\sum_{i=1}^{b}(n_i-1)$ eigenvalues 1. We have now written V in the spectral form as:

$$V = \mathrm{diag}(V_1,V_2,\ldots,V_b) \text{ with } V_i = (1+\rho n_i)B_i + 1(I_{n_i}-B_i) \qquad (7)$$

hence $V_i^{\frac{1}{2}} = (1+\rho n_i)^{\frac{1}{2}}B_i + 1^{\frac{1}{2}}(I_{n_i}-B_i)$

and $V_i^{-1} = (1+\rho n_i)^{-1}B_i + 1^{-1}(I_{n_i}-B_i)$.

Finally $V^{\frac{1}{2}} = \mathrm{diag}(V_1^{\frac{1}{2}}, \ldots, V_b^{\frac{1}{2}})$ and $V^{-1} = \mathrm{diag}(V_1^{-1}, \ldots, V_b^{-1})$.

The dispersion matrix $D(\underset{\sim}{y})$ for the three-stage nested design is given in (6): $D(\underset{\sim}{y}) = \sigma_e^2(\rho_1 Z_1 Z_1' + \rho_2 Z_2 Z_2' + I_{n..}) = \sigma_e^2 V$.

But $V = \mathrm{diag}(V_1,V_2,\ldots,V_a)$ where

$$V_h = \rho_1 J_{n_{h.}} + \rho_2 \mathrm{diag}(J_{n_{h1}},\ldots,J_{n_{hb_h}}) + I_{n_{h.}} \qquad (8)$$

$$= \rho_1 n_{h.} A_h + \rho_2 \mathrm{diag}(n_{h1}B_{h1},\ldots,n_{hb_h}B_{hb_h}) + I_{n_{h.}}$$

with symmetric and idempotent matrices $A_h = n_{h.}^{-1}J_{n_{h.}}$,

$B_{hi} = n_{hi}^{-1}J_{n_{hi}}$, $(i = 1,2,\ldots,b;\ h = 1,2,\ldots,a)$.

For the balanced three-stage nested design $n_{hi} = m_2$, $b_i = m_1$, $n_{h.} = m_1m_2$ and $V_h = \rho_1 m_1 m_2 A + \rho_2 m_2 \text{diag}(B, \ldots, B) + I_{m_1m_2}$

where $A = (m_1m_2)^{-1}J_{m_1m_2}$, $B = m_2^{-1}J_{m_2}$, hence

$$V_h = (1+m_2\rho_2+m_1m_2\rho_1)A + (1+m_2\rho_2)[\text{diag}(B,\ldots,B) - A] + [I_{m_1m_2} - \text{diag}(B,\ldots,B)].$$

The three matrices in $V_h$ are symmetric and idempotent and the product of each two is zero. The coefficients of these three matrices are the eigenvalues of $V_h$ where the eigenvalue $(1 + m_2\rho_2 + m_1m_2\rho_1)$ has multiplicity 1, $(1+m_2\rho_2)$ has multiplicity $m_1-1$ and 1 has multiplicity $m_1(m_2-1)$. The matrix V has the same eigenvalues repeated a times. Hence in the balanced case we can easily find $V_h^{-1}$ or $V_h^{\frac{1}{2}}$ and thus $V^{-1}$ or $V^{\frac{1}{2}}$.

This nice structure of $D(\underset{\sim}{y})$ in the nested design has also been used by e.g. Fuller and Battese (1973). See also LaMotte (1972) for a nice representation of $V^{-1}$.

## 2. TWO-STAGE NESTED DESIGNS

### 2.1. Balanced case

In the model (1) we have the situation

$$y_{ik} = \beta + u_i + e_{ik} \quad \text{with } i = 1, 2, \ldots, b;\ k = 1,2,\ldots,m$$

for each i. With the tools of chapter 1 we can readily derive

$$SSB = \underset{\sim}{y}'P_{B*}\underset{\sim}{y} \quad \text{and} \quad P_{B*}V = (P_B - P_L)\,\text{diag}(V_1, V_2, \ldots, V_b)$$

with $P_B = \text{diag}(B_1, \ldots, B_b) = \text{diag}(B, B, \ldots, B)$, $B = m^{-1}J_m$;

$V_i = (1+\rho m)B + (I_m - B) = W$ for $i = 1,2,\ldots,b$; $P_L = (bm)^{-1}J_{bm}$.

In the balanced case we can use also the direct (or Kronecker) product x of two matrices. Properties of direct products can be found in Neudecker (1968), Rao (1973) or Khuri (1982). In our case $P_B = I_b \times B$, $V = I_b \times W$ thus $(P_B - P_L)V = (1+\rho m)(P_B - P_L) = (1+\rho m)P_{B^*}$. Hence $SSB = \underset{\sim}{y}'P_{B^*}\underset{\sim}{y}$ is distributed as $\sigma_e^2(1+\rho m)\chi^2(b-1)$. In the same way we derive

$SSE = \underset{\sim}{y}'P_C\underset{\sim}{y}$ is distributed as $\sigma_e^2\chi^2(b(m-1))$. Furthermore SSB and SSE are stochastically independent because

$P_{B^*}D(\underset{\sim}{y})P_C = \sigma_e^2 P_{B^*}VP_C = 0$ (see Searle (1971)).

Hence MSB/MSE has a $(1+\rho m)$F-distribution with b-1 and b(m-1) df.

For the test of $H_0$: $\rho = 0$ against $H_1$: $\rho > 0$ (with $\rho = \sigma_B^2/\sigma_e^2$) we calculate the test statistic MSB/MSE. We reject $H_0$ at significance level $\alpha$ if $MSB/MSE > F(1-\alpha; b-1, b(m-1))$, the $(1-\alpha)$-distribution point of $F(b-1, b(m-1))$. For the test of $H_0$: $\rho \leq \rho_0$ against $H_1$: $\rho > \rho_0$ we must calculate the test statistic $(1+\rho_0 m)^{-1}MSB/MSE$, which has under $H_0$ again an $F(b-1,b(m-1))$-distribution. This test is UMP invariant with respect to a group of transformations generated by translations and scale transformations. Also the test is UMP unbiased of size $\alpha$ (see Lehmann (1959) or Herbach (1959)); for mixed models containing random effects which are ordered hierarchically see Spjøtvoll (1966) and Roebruck (1982).

For the sequel we notice that this test statistic $(1+\rho_0 m)^{-1}MSB/MSE$ is equivalent to $MSB(\rho_0)/MSE(\rho_0)$ where

$MSB(\rho_0) = \underset{\sim}{y}'P_{B^*}[V(\rho_0)]^{-1}P_{B^*}\underset{\sim}{y}/(b-1)$ and

$MSE(\rho_0) = \underset{\sim}{y}'P_C[V(\rho_0)]^{-1}P_C\underset{\sim}{y}/(b(m-1))$ and $\sigma_e^2V(\rho_0)$ is the dispersion matrix of $\underset{\sim}{y}$ with $\rho = \rho_0$ or $V(\rho_0) = I_b \times [(1+\rho_0 m)B + (I_m - B)]$. It follows readily that

$$\underset{\sim}{y}'P_{B^*}[V(\rho_0)]^{-1}P_{B^*}\underset{\sim}{y} = \underset{\sim}{y}'P_{B^*}\underset{\sim}{y}/(1+\rho_0 m)$$

and

$$\underset{\sim}{y}'P_C[V(\rho_0)]^{-1}P_C\underset{\sim}{y} = \underset{\sim}{y}'P_C\underset{\sim}{y}.$$

An exact $(1-\alpha)$-confidence interval for $\rho$ now follows from $(1+\rho m)^{-1}(MSB/MSE)$ which has an $F(b-1,b(m-1))$-distribution. Let $F_1 = F(\alpha/2;\ b-1,\ b(m-1))$ and $F_2 = F(1-\alpha/2;\ b-1,\ b(m-1))$ be the $\alpha/2$ and $1-\alpha/2$ points of the distribution function of $F(b-1,b(m-1))$ respectively. Then a $(1-\alpha)$-confidence interval for $\rho$, which is uniformly most accurate invariant is given by:

$$[F_2^{-1}(MSB/MSE) - 1]/m < \rho < [F_1^{-1}(MSB/MSE) - 1]/m\ .$$

## 2.2. Unbalanced case

Recall model (1) which is

$$y_{ik} = \beta + u_i + e_{ik} \quad \text{with } k = 1, 2, \ldots, n_i \text{ for } i = 1, 2, \ldots, b.$$

Lehmann (1959) already mentioned that for the unbalanced case there is no UMP invariant test. Spjøtvoll (1967) found that for the problem of testing $H_0$: $\rho \leq \rho_0$ against $H_1$: $\rho = \rho_1$ $(\rho > \rho_0)$ the most powerful similar invariant test depends upon the value of $\rho_1$. His test statistic is

$$W(\rho_0,\rho_1) = \frac{\underset{\sim}{y}'P_{B*}[V(\rho_0)]^{-1}P_{B*}\underset{\sim}{y} - \underset{\sim}{y}'P_{B*}[V(\rho_1)]^{-1}P_{B*}\underset{\sim}{y}}{\underset{\sim}{y}'P_{B*}[V(\rho_0)]^{-1}P_{B*}\underset{\sim}{y} + \underset{\sim}{y}'P_C\underset{\sim}{y}}$$

where $V(\rho_0)$ and $V(\rho_1)$ are the matrices $\rho ZZ' + I_{n.}$ with $\rho = \rho_0$ and $\rho = \rho_1$ respectively. This test has the property of maximizing the minimal power over the set of alternatives with $\rho \geq \rho_1$. To get a size $\alpha$ test a constant c must be determined such that $P_0(W(\rho_0,\rho_1) > c) = \alpha$. For the case under $H_1$: $\rho_1 \to \infty$ (test against large $\rho_1$) the sum of squares is $\underset{\sim}{y}'P_{B*}[V(\rho_1)]^{-1}P_{B*}\underset{\sim}{y} \to 0$. To reject the $H_0$ when $W(\rho_0,\infty)$ is large is the same as to reject $H_0$ when

$$T(\rho_0) = \underset{\sim}{y}'P_{B*}[V(\rho_0)]^{-1}P_{B*}\underset{\sim}{y}/\underset{\sim}{y}'P_C\underset{\sim}{y} > \text{constant } c' \ .$$

Using the results of section 1.4 it can easily be derived that

$$W(\rho_0,\rho_1) = \frac{\sum_{i=1}^{b} g_{i0}(\bar{y}_{i.}-\bar{\bar{y}}_0)^2 - \sum_{i=1}^{b} g_{i1}(\bar{y}_{i.}-\bar{\bar{y}}_1)^2}{\sum_{i=1}^{b} g_{i0}(\bar{y}_{i.}-\bar{\bar{y}}_0)^2 + \sum_{i=1}^{b}\sum_{k=1}^{n_i}(y_{ik}-\bar{y}_{i.})^2}$$

where $g_{ij} = n_i/(\rho_j n_i+1)$, $\bar{\bar{y}}_j = \sum_{i=1}^{b} g_{ij}\bar{y}_{i.} / \sum_{i=1}^{b} g_{ij}$ for $j = 0$ and 1, and

$\bar{y}_{i.} = \sum_{k=1}^{n_i} y_{ik}/n_i$; hence $T(\rho_0) = \sum_{i=1}^{b} g_{i0}(\bar{y}_{i.}-\bar{\bar{y}}_0)^2 / \sum_{i=1}^{b}\sum_{k=1}^{n_i}(y_{ik}-\bar{y}_{i.})^2$.

To calculate $P(W(\rho_0,\rho_1) > c)$ we note that $W(\rho_0,\rho_1) > c$ is equivalent to

$$\underset{\sim}{y}'[P_{B*}((1-c)[V(\rho_0)]^{-1} - [V(\rho_1^2)]^{-1})P_{B*} - cP_C]\underset{\sim}{y} > 0$$

or $\underset{\sim}{y}'Q\underset{\sim}{y} > 0 = \underset{\sim}{y}'Q\underset{\sim}{y}/\sigma_e^2 > 0.$

The distribution of $\underset{\sim}{y}'Q\underset{\sim}{y}/\sigma_e^2$ is that of a linear combination of independent $\chi^2(1)$ variables with positive and negative coefficients, where the coefficients are the eigenvalues of $QV(\rho_0)$ under $H_0$ (for calculating the P-value) and of $QV(\rho_1)$ under $H_1$ (for calculating the power function). The distribution of $T(\rho_0)$ at $\rho = \rho_0$ is easily found. The numerator $\underset{\sim}{y}'P_{B*}[V(\rho_0)]^{-1}P_{B*}\underset{\sim}{y}$ is a linear combination of independent $\sigma_e^2\ \chi^2(1)$ variables with coefficients the eigenvalues of $P_{B*}[V(\rho_0)]^{-1}P_{B*}V(\rho_0)$, which turns out to be $P_{B*}$; hence the numerator is a $\sigma_e^2\ \chi^2(b-1)$ variable. Furthermore the numerator and denominator of $T(\rho_0)$ are independent under $H_0$ because their covariance is zero, which follows from

$$P_{B*}[V(\rho_0)^{-1}]P_{B*}V(\rho_0)P_C = P_{B*}P_C = 0.$$

Hence $(n_{.}-b)T(\rho_0)/(b-1)$ has under $H_0$ an $F(b-1,n_{.}-b)$-distribution. Note that

$$\underset{\sim}{y}'P_C[V(\rho_0)]^{-1}P_C\underset{\sim}{y} = \underset{\sim}{y}'P_C\underset{\sim}{y} \text{ and is a } \sigma_e^2\chi^2(n_{.}-b) \text{ variable.}$$

We have again as test statistic $MSB(\rho_0)/MSE(\rho_0)$ as in the balanced case. For the power calculations at $\rho = \rho_1$: $P(T(\rho_0) > c) =$

$$P(\underset{\sim}{y}'[P_{B*}[V(\rho_0)]^{-1}P_{B*} - cP_C]\underset{\sim}{y} > 0) = P(\underset{\sim}{y}'Q\underset{\sim}{y} > 0) = P(\underset{\sim}{y}'Q\underset{\sim}{y}/\sigma_e^2 > 0),$$

where $\underset{\sim}{y}'\underset{\sim}{Q}\underset{\sim}{y}/\sigma_e^2$ has the distribution of $\sum_{i=1}^{q} \lambda_i S_i^2$, where $S_i^2$ are independent $\chi^2(1)$ variables and $\lambda_i$ are the positive and negative eigenvalues of $QV(\rho_1)$ and $q$ = rank Q.

Spjøtvoll (1968) gives several examples of the exact power function calculations of the test statistics W and T for the hypothesis $\rho = 0$ against $\rho = 0.1$ at the 1% level for the special case he was able to tackle, viz. the unbalanced two-stage design (one-way lay-out) with b = 3 classes only.

Verdooren (1974) shows how the exact P-values and power function of unbalanced one-way lay-out in general (thus also for more than three classes) can be calculated. The calculation of $P(\underset{\sim}{y}'\underset{\sim}{Q}\underset{\sim}{y}/\sigma_e^2 > c)$ can be handled with the procedure of Imhof (1961). He shows that the distribution function of $\underset{\sim}{y}'\underset{\sim}{Q}\underset{\sim}{y}/\sigma_e^2$ can be obtained quite easily by straightforward numerical integration of the characteristic function of $\underset{\sim}{y}'\underset{\sim}{Q}\underset{\sim}{y}/\sigma_e^2$. A computer program in FORTRAN was given by Koerts and Abrahamse (1969). With this program fast numerical calculation of the distribution function of a linear combination of independent (non-)central chi-square variables with positive and negative coefficients is possible. Another good program in ALGOL is given by Davies (1980) and Farebrother (1980), (1984a). The Koerts and Abrahamse program can in most cases be used, see Farebrother (1984b). For examples of calculations see Verdooren (1974) and Verdooren (1982).

Wald (1940, 1947) and also Spjøtvoll (1967) use $T(\rho)$ for constructing an exact $(1-\alpha)$-confidence interval for $\rho$. The lower confidence limit $\rho_L$ of $\rho$ is given by the root of the following equation in $\rho$,

$$(n_. - b)T/(b-1) = F(1-\alpha/2;\ b-1,\ n_. - b)$$

and the upper confidence limit $\rho_U$ of $\rho$ is given by the root of the equation in $\rho$:

$(n_{.}-b)T/(b-1) = F(\alpha/2;\ b-1,\ n_{.}-b)$.

The construction of this exact $(1-\alpha)$-confidence interval for $\rho$ can only be solved iteratively, which can easily be programmed, see Verdooren (1976). For an approximate $(1-\alpha)$-confidence interval for $\rho$ based on an approximation of $T(\rho)$ see Thomas and Hultquist (1978). At the present time however, the calculation for an exact $(1-\alpha)$-confidence interval is so easy, that we would use the procedure of Thomas and Hultquist only as a starting point for the iterative calculation of an exact $(1-\alpha)$-confidence interval.

The upper bound $\rho_U$ can be negative which is contrary to the model assumptions. This happens when $(n_{.}-b)T(0)/(b-1)$ is smaller than $F(\alpha/2;\ b-1, n_{.}-b)$. Seely and El-Bassiouni (1983) prove that the probability that $\rho_U$ is negative is at most $\alpha/2$ and goes to zero as $\rho$ gets large. To avoid the possibility of negative endpoints, one can use an adjusted interval suggested by Thompson (1955). This confidence interval for $\rho$ is $[\max(0,\rho_L),\ \max(0,\rho_U)]$ which has confidence level $1-\alpha$ for $\rho > 0$ and $1-\alpha/2$ for $\rho = 0$.

A locally most powerful test for $H_0$: $\rho = \rho_0$ against $H_1$: $\rho = \rho_0 + \Delta$, where $\Delta$ is small, is given by Mostafa (1967). His power calculations are not exact but based on approximations. With Imhof's procedure exact results can be given.

Recently, Seely and El-Bassiouni (1983) were able to obtain Wald's (1947) exact test via reductions in sums of squares for the random effects adjusted for the fixed effects in a general mixed model. They also gave necessary and sufficient conditions under Wald's test, which can be used in mixed models. Harville and Fenech (1985) outlined the computational aspects of Seely and El-Bassiouni's method in the context of a specific animal breeding application.

A computational remark can however be made. By calculating Wald's exact test via reductions in sums of squares, our method given

here by calculating the sum of squares as squares of projections, gives a more rapid and lucidly interpreted result. Seely et al. (1983) and Harville et al. (1985) use a linear transformation of $\underset{\sim}{y}$ which is not a projection operator. Their linear transformations can however be easily transformed to projection operators.

## 3. THREE-STAGE NESTED DESIGN

### 3.1. Balanced case

In the model (4) we have the situation

$$y_{hik} = \beta + u_{1h} + u_{2hi} + e_{hik}$$

with $h = 1,2,\ldots,a$; $i = 1,2,\ldots,m_1$ for each h and $k = 1,2,\ldots,m_2$ for each h and i. With the tools of chapter 1 we readily derive the distribution of $SSA = \underset{\sim}{y}'P_{A^*}\underset{\sim}{y}$, $SSB = \underset{\sim}{y}'P_{B^*}\underset{\sim}{y}$ and $SSE = \underset{\sim}{y}'P_C\underset{\sim}{y}$.

In the balanced case we have seen in 1.4: $P_A = I_a \times A$, $P_{A^*} = P_A - P_L$, $P_B = I_{am_1} \times B$, $P_{B^*} = P_B - P_A$; $P_C = I_{am_1m_2} - P_B$, where A and B are symmetric and idempotent matrices with $A = (m_1m_2)^{-1}J_{m_1m_2}$ and $B = m_2^{-1}J_{m_2}$. The dispersion matrix $D(\underset{\sim}{y}) = \sigma_e^2 V$ with $V = I_a \times W$ with $W = \rho_1 J_{m_1m_2} + \rho_2 I_{m_1} \times J_{m_2} + I_{m_1m_2}$.

For the distribution of $SSA = \underset{\sim}{y}'P_{A^*}\underset{\sim}{y}$ we need the eigenvalues of

$$P_{A^*}V = (m_1m_2\rho_1 + m_2\rho_2 + 1)P_{A^*}.$$

Because $P_{A^*}$ is a projection operator on $A^*$ it has rank

$P_{A^*} = \dim A^* = a-1$. So a-1 eigenvalues are 1 and the other eigenvalues are 0. SSA has therefore the distribution of

$\sigma_e^2(m_1m_2\rho_1+m_2\rho_2+1)\chi^2(a-1)$.

Analogously we derive the eigenvalues of $P_{B^*}V = (m_2\rho_2+1)P_{B^*}$. Because $P_{B^*}$ is a projection operator on $B^*$ it has rank $P_{B^*} = \dim B^* = a(m_1-1)$. So $a(m_1-1)$ eigenvalues are 1 and the other eigenvalues are 0. SSB has therefore the distribution of $\sigma_e^2(m_2\rho_2+1)\chi^2(a(m_1-1))$. Also $P_CV = P_C$. Hence SSE has a

$\sigma_e^2\chi^2(am_1(m_2-1))$-distribution because rank $P_C = \dim C = am_1(m_2-1)$.

Because $P_{A^*}VP_{B^*} = (m_1m_2\rho_1+m_2\rho_2+1)P_{A^*}P_{B^*} = 0$, SSA and SSB are independent. Also $P_{A^*}VP_C = (m_1m_2\rho_1+m_2\rho_2+1)P_{A^*}P_C = 0$ and $P_{B^*}VP_C = (m_2\rho_2+1)P_{B^*}P_C = 0$ hence SSA, SSB ar[illegible]utually independent. Hence MSB/MSE has a $(1+m_2\rho_2)$F-distribution with $a(m_1-1)$ and $am_1(m_2-1)$ degrees of freedom respectively.

For the test of $H_0:\rho_2 = 0$ against $H_1:\rho_2 > 0$ we reject $H_0$ at significance level $\alpha$ if $MSB/MSE > F(1-\alpha; a(m_1-1), am_1(m_2-1))$. This was the special case $\rho_{20} = 0$; now we consider the general case.

For the test of $H_0$: $\rho_2 \leq \rho_{20}$ against $H_1$: $\rho_2 > \rho_{20}$ the test statistic $(1+m_2\rho_{20})^{-1}MSB/MSE$ is used, which has under $H_0$ again an $F(a(m_1-1), am_1(m_2-1))$-distribution. The argument for deriving this test is completely analogous to the argument used in section 2.1, which shows that we have a UMP unbiased size $\alpha$ test. Notice again that we can use the information about $\rho_2$ under $H_0$ to calculate

$$\underset{\sim}{y}'P_{B^*}[V(\rho_{20})]^{-1}P_{B^*}\underset{\sim}{y} = (1+m_2\rho_{20})^{-1}\underset{\sim}{y}'P_{B^*}\underset{\sim}{y}$$

and

$$\underset{\sim}{y}'P_C[V(\rho_{20})]^{-1}P_C\underset{\sim}{y}' = \underset{\sim}{y}'P_C\underset{\sim}{y}.$$

The test statistic is hence equivalent to $MSB(\rho_{20})/MSE(\rho_{20})$.

An exact $(1-\alpha)$-confidence interval for $\rho_2$ now follows from $(1+m_2\rho_2)^{-1}(MSB/MSE)$, which has an $F(a(m_1-1), am_1(m_2-1)$-distribution, just as in section 2.1.

For the test of $H_0$: $\rho_1 = 0$ against $H_1$: $\rho_1 > 0$ we use the fact that MSA/MSB has the distribution of

$$[(1+m_2\rho_2+m_1m_2\rho_1)/(1+m_2\rho_2)]F(a-1,\ a(m_1-1)),$$

which has under $H_0$ an $F(a-1,a(m_1-1))$ distribution.

For the test of $H_0$: $\rho_1 \leq \rho_{10}$ against $H_1$: $\rho_1 > \rho_{10}$ the test statistic MSA/MSB has the distribution of

$$[1+m_1m_2\rho_{10}/(1+m_2\rho_2)]F(a-1,a(m_1-1)) \text{ under } H_0.$$

This distribution depends on the nuisance parameter $\rho_2$!

Let us use the information about the dispersion matrix $D(\underset{\sim}{y})$ for $\rho_1 = \rho_{10}$,

$$D(\underset{\sim}{y}|\rho_{10}) = \sigma_e^2[\rho_{10}Z_1Z_1' + \rho_2Z_2Z_2' + I_{am_1m_2}] = \sigma_e^2V(\rho_{10}).$$

We calculate $SSA(\rho_{10}) = \underset{\sim}{y}'P_{A*}[V(\rho_{10})]^{-1}P_{A*}\underset{\sim}{y}$ which is equal to

$$(1+m_2\rho_2+m_1m_2\rho_{10})^{-1}\underset{\sim}{y}'P_{A*}\underset{\sim}{y}.$$

Also $SSB(\rho_{10}) = \underset{\sim}{y}'P_{B*}[V(\rho_{10})]^{-1}P_{B*}\underset{\sim}{y} = (1+m_2\rho_2)^{-1}\underset{\sim}{y}'P_{B*}\underset{\sim}{y}$.

If we use as a test statistic $MSA(\rho_{10})/MSB(\rho_{10})$ we get $[1+m_1m_2\rho_{10}/(1+m_2\rho_2)]^{-1}MSA/MSB$ and this has an $F(a-1,a(m_1-1))$-distribution. But the test statistic depends on the nuisance parameter $\rho_2$.

Also the construction of an exact $(1-\alpha)$-confidence interval for $\rho_1$ depends on the value of $\rho_2$.

Several proposals are made to solve this problem from which we mention some interesting ones. Broemeling (1969) gives a

simultaneous confidence region for $\rho_1$ and $\rho_2$ from which an upper bound and lower bound respectively can be found for $\rho_1$ and $\rho_2$ separately. By projecting on the $\rho_1$ axis an upper bound and lower bound respectively for $\rho_1$ can be found. These bounds are conservative. How to find the exact confidence coefficients associated with Broemeling's confidence regions for $(\rho_1,\rho_2)$ is demonstrated by Sahai and Anderson (1973) in terms of the upper tail of the inverted Dirichlet distribution. A computer program for calculating the upper tail of the inverted Dirichlet distribution is given by Yassaee (1976). An extension of the results of Sahai and Anderson (1973) is given by Yassaee (1981).

Broemeling (1978) proposed a confidence interval for $\rho_1$ based on a two-sided version of the simultaneous confidence region of $(\rho_1,\rho_2)$. Unfortunately this proposal was wrong, as Tong (1979) proved.

Khuri (1981) presented a technique for the construction of simultaneous confidence intervals for the values of all continuous functions of the variance components for a general balanced random model. For the balanced three-stage nested design an approximate $(1-\alpha)$-confidence interval for $\rho_1$ is given by Graybill and Wang (1979) and Wang and Graybill (1981).

Having mentioned the proposals in the literature we now propose a new procedure. Let us consider again the test of $H_0$: $\rho_1 \leq \rho_{10}$ against $H_1$: $\rho_1 > \rho_{10}$, with test statistic

$$MSA(\rho_{10})/MSB(\rho_{10}) = [1+m_1m_2\rho_{10}/(1+m_2\rho_2)]^{-1}MSA/MSB = f.$$

This test statistic has an $F(a-1,a(m_1-1))$-distribution for each $\rho_2$. In an experiment we get an outcome f. Let us now calculate the P-value or critical level $P(F(a-1,a(m_1-1)) > f)$ for each possible $\rho_2$. In practice we can calculate this for several $\rho_2$-values such as $\rho_2$ = 0, 0.1, 0.5, 1, 1.5, 2, 4,8 etc. For a certain $\rho_{10}$ we now have a set of P-values for the range of $\rho_2$ values. The experimenter often knows which range of $\rho_2$ is of

interest to him. For each value of $\rho_2$ in this range we can decide at significance level $\alpha$ whether $H_0$: $\rho_1 \leq \rho_{10}$ must be rejected in favour of $H_1$: $\rho_1 > \rho_{10}$. If the experimenter has no previous knowledge about $\rho_2$, an idea about the interesting range can be suggested from a confidence interval of $\rho_2$. (A Bayesian approach would assume an a priori distribution for $\rho_2$ and derive the a posteriori set of not rejected $\rho_1$'s).

The idea of calculating the P-values under the $H_0$: $\rho_1 = \rho_{10}$ for several given values of $\rho_2$ can be used more fruitfully. If we use our knowledge about $\rho_2$ we see that, irrespective of the value of $\rho_{10}$, $\underset{\sim}{y}'P_{B*}[V(\rho_{10})]^{-1}P_{B*}\underset{\sim}{y} = (1+m_2\rho_2)^{-1}\underset{\sim}{y}'P_{B*}\underset{\sim}{y} = (1+m_2\rho_2)^{-1}$ SSB has a $\sigma_e^2\chi^2(a(m_1-1))$-distribution. But also $\underset{\sim}{y}'P_C[V(\rho_{10})]^{-1}P_C\underset{\sim}{y} = \underset{\sim}{y}'P_C\underset{\sim}{y}$ = SSE has a $\sigma_e^2\chi^2(am_1(m_2-1))$-distribution. Furthermore SSB and SSE are independent, hence $(1+m_2\rho_2)^{-1}$SSB + SSE has a $\sigma_e^2\chi^2$-distribution with $a(m_1-1) + am_1(m_2-1) = a(m_1m_2-1)$ degrees of freedom. Let us define SSD = $[(1+m_2\rho_2)^{-1}$SSB + SSE] and MSD = SSD/d with $d = a(m_1m_2-1)$ then $MSA(\rho_{10})/MSD = [1 + m_2\rho_2 + m_1m_2\rho_{10}]^{-1}MSA/MSD$ has an F(a-1,d)-distribution (because SSA is independent of SSB and SSE, SSA is independent of SSD).
By calculating for a range of $\rho_2$ values the P-values for a certain $H_0$ with value $\rho_{10}$, we can decide for which interesting $\rho_2$ we reject $H_0$: $\rho_1 \leq \rho_{10}$ in favour of $H_1$: $\rho_1 > \rho_{10}$.

Also an exact $(1-\alpha)$-confidence interval for $\rho_1$ can now be given if we use a given value of $\rho_2$. For each $\rho_2$ we calculate MSA/MSD. Because $[1+m_2\rho_2+m_1m_2\rho_1]^{-1}$MSA/MSD for a given value of $\rho_2$ has an F(a-1,d)-distribution, we get an exact $(1-\alpha)$-confidence interval for $1 + m_2\rho_2 + m_1m_2\rho_1$:

$$(MSA/MSD)/F(1-\alpha/2;\ a-1,d) < 1+m_2\rho_2+m_1m_2\rho_1 < (MSA/MSD)/F(\alpha/2;\ a-1,d)$$

and hence an exact $(1-\alpha)$-confidence interval for $\rho_1$ given the value of $\rho_2$, is

$$L < \rho_1 < U \text{ with } L = ([(MSA/MSD)/F(1-\alpha/2;\ a-1,d)] - 1 - m_2\rho_2)/(m_1m_2)$$

and $U = ([(MSA/MSD)/F(\alpha/2;\ a-1,d)] - 1 - m_2\rho_2)/(m_1m_2)$.

A conservative confidence interval for $\rho_1$ is obtained by choosing

$$\text{lower bound} = \min_{\rho_2}(L),\quad \text{upper bound} = \max_{\rho_2}(U),$$

with $\rho_2$ varying in the range given by the experimenter.

When a priori information about $\rho_2$ is not available a Bonferroni type approach could be used. First a $(1-\frac{1}{2}\alpha)$-confidence interval is derived for $\rho_2$, based on MSB/MSE. Second, the procedure above is carried out with this interval as range for $\rho_2$, and with L and U determined (for given $\rho_2$) with confidence coefficient $1-\frac{1}{2}\alpha$ (instead of $1-\alpha$).

## 3.2. Unbalanced case

Recall model (4) which is

$$y_{hik} = \beta + u_{1h} + u_{2hi} + e_{hik}$$

with $h = 1,2,\ldots,a$; $i = 1,2,\ldots,b_h$ and $k = 1,2,\ldots,n_{hi}$.
It can easily be shown that $SSE = \underset{\sim}{y}'P_C\underset{\sim}{y}$ has a $\sigma_e^2\chi^2(n_{..}-b_.)$-distribution with $n_{..} = \sum_{h=1}^{a}\sum_{i=1}^{b_h} n_{hi}$ and $b_. = \sum_{h=1}^{a} b_h$. $SSB = \underset{\sim}{y}'P_{B^*}\underset{\sim}{y}$

has a distribution of $\sigma_e^2 \sum_{i=1}^{b_.-a} \lambda_i S_i^2$ where $\lambda_i$ are the eigenvalues of $P_{B^*}V$ and $S_i^2$ are independent $\chi^2(1)$ variables.

The test of $H_0$: $\rho_2 = 0$ against $\rho_2 > 0$ can be based on

$\underset{\sim}{y}'P_{B^*}\underset{\sim}{y}$ = SSB and $\underset{\sim}{y}'P_C\underset{\sim}{y}$ = SSE. Under $H_0$ SSB has a $\sigma_e^2\chi^2(b_.-a)$-distribution. Furthermore SSB and SSE are independent while $P_{B^*}VP_C = P_{B^*}P_C = 0$. Hence MSB/MSE has under $H_0$ an $F(b_.-a, n_{..}-b_.)$-distribution.

For the test of $H_0$: $\rho_2 \leq \rho_{20}$ against $\rho_2 > \rho_{20}$ we use an analogous extension of the test statistic $T(\rho)$ of the unbalanced two-stage nested design.

Define $T(\rho_{20}) = SSB(\rho_{20})/SSE$

$$= \sum_{h=1}^{a}\sum_{i=1}^{b_h} g_{hi}(\bar{y}_{hi.}-\bar{\bar{y}}_{h..})^2 \Big/ \sum_{h=1}^{a}\sum_{i=1}^{b_h}\sum_{k=1}^{n_{hi}} (y_{hik}-\bar{y}_{hi.})^2$$

with $g_{hi} = n_{hi}/(\rho_{20}n_{hi}+1)$, $\bar{y}_{hi.} = \sum_{k=1}^{n_{hi}} y_{hik}/n_{hi}$ and

$$\bar{\bar{y}}_{h..} = \sum_{i=1}^{b_h} g_{hi}\bar{y}_{hi.} \Big/ \sum_{i=1}^{b_h} g_{hi}.$$

We can directly apply Wald's procedure to construct an exact $(1-\alpha)$-confidence interval for $\rho_2$ using $(n_{..}-b_.)T(\rho_2)/(b_.-a)$ which has again an $F(b_.-a, n_{..}-b_.)$-distribution. See Verdooren (1976).

We now consider the test of $H_0$: $\rho_1 = 0$ against $H_1$: $\rho_1 > 0$. We can easily calculate the expected Mean Squares (EMS), using the fact that the expected value of a quadratic form $\underset{\sim}{y}'Q\underset{\sim}{y}$ with $E\underset{\sim}{y} = X\beta$ and $D(\underset{\sim}{y}) = \sigma_e^2 V$,

$$E(\underset{\sim}{y}'Q\underset{\sim}{y}) = \text{tr } Q\sigma_e^2 V + \beta'X'QX\beta,$$

with V given in (8) and where tr stands for trace of the square matrix.

For a sum of squares in the ANOVA table Q is equal to $P_{A^*}$, $P_{B^*}$ and $P_C$ respectively, hence $QX = 0$.

Furthermore we have

$$E(\underset{\sim}{y}'P_C\underset{\sim}{y}) = \sigma_e^2 tr(P_C V) = \sigma_e^2(n_{..}-b_.) \text{ and}$$

$$E(\underset{\sim}{y}'P_{B^*}\underset{\sim}{y}) = \sigma_e^2 tr(P_{B^*}V) = (b_.-a)\sigma_e^2[1+k_3\rho_2],$$

$$\text{with } k_3 = \sum_{h=1}^{a}\sum_{i=1}^{b_h} n_{hi}^2(n_{hi}^{-1} - n_{h.}^{-1}).$$

$$E(\underset{\sim}{y}'P_{A^*}\underset{\sim}{y}) = \sigma_e^2 tr(P_{A^*}V) = (a-1)\sigma_e^2[1+k_2\rho_2+k_1\rho_1]$$

$$\text{with } k_2 = \sum_{h=1}^{a}\sum_{i=1}^{b_h} n_{hi}^2(n_{h.}^{-1} - n_{..}^{-1}) \text{ and } k_1 = \sum_{h=1}^{a} n_{h.}^2(n_{h.}^{-1} - n_{..}^{-1}).$$

In the past Tietjen and Moore (1968) advocated a quasi-F test by using a linear combination $a_1$MSB + $(1-a_1)$MSE such that the expectation of this combination was equal to E(MSA) under $H_0$: $\rho_1 = 0$. The idea for the quasi-F test was the application of the procedure of Satterthwaite (1946). The idea of Satterthwaite was to approximate a linear combination of independent chi-square variables $\sum_{i=1}^{m} a_i\chi^2(g_i)$ by $c\chi^2(g)$, such that the expectation and variance of these two random quantities are equal. Kruskal (1968) pointed out that for such an unbalanced nested design MSA and MSB are not independent ($P_{A^*}VP_{B^*}$ is not equal to a null matrix). It can easily be shown that MSA and MSB are independent if the three-stage nested design is partially balanced ($n_{hi} = m_h$ for all i) or has last stage uniformity. Note that last stage uniformity requires $n_{h_i} = m$ for all h and i. Even for partially balanced designs SSA and SSB are not always distributed as a chi-square variable. And an expression as a linear combination of independent $\chi^2(1)$ variables is still needed.

Next Tietjen (1974) proposed the conventional F-test based on MSA/MSB for testing $H_0$: $\rho_1 = 0$, instead of Satterthwaite's procedure which was earlier proposed by Tietjen and Moore in 1968. The P-value has been approximated by $P(F(a-1,b_{\cdot}-a) > MSA/MSB)$. Tietjen (1974) based his proposal on several simulations, but this is only a crude approximate test.

Cummings and Gaylor (1974) also give crude approximate P-values for MSA/MSD with MSD = cMSB + (1-c)MSE if $c = k_2/k_3 < 1$, such that under $H_0$: $\rho_1 = 0$, E(MSA) = E(MSD), and for(dMSA + (1-d)MSE)/MSB if $d = k_3/k_2 < 1$, such that under $H_0$: $\rho_1 = 0$, E(dMSA + (1-d)MSE) = E(MSB). Tan and Cheng (1984) compare Tietjen's test statistic, Cummings and Gaylor's test statistic and their own test statistic $(k_3MSA+k_2MSE)/(k_2MSB+k_3MSE)$, such that under $H_0$: $\rho_1 = 0$, $E(k_3MSA+k_2MSE) = E(k_2MSB+k_3MSE)$. Using a better approximation for the distribution of the test statistics, based on Laguerre polynomial expansions, they conclude that their test statistic is a preferable one.

However Verdooren (1974) has shown how an exact P-value can be calculated. Let the actual outcome of MSA/MSB be f, then the exact P-value is

$$P_0(MSA/MSB > f, \text{ given that } \rho_1 = 0) =$$

$$P_0((\underset{\sim}{y}'P_{A^*}\underset{\sim}{y}/\underset{\sim}{y}'P_{B^*}\underset{\sim}{y}) > (a-1)f/(b_{\cdot}-a)) =$$

$$P_0(\underset{\sim}{y}'[P_{A^*}-[(a-1)f/(b_{\cdot}-a)]P_{B^*}]\underset{\sim}{y} > 0) = P_0(\underset{\sim}{y}'Q\underset{\sim}{y} > 0) = P_0(\underset{\sim}{y}'Q\underset{\sim}{y}/\sigma_e^2 > 0).$$

The distribution under $H_0$ of $\underset{\sim}{y}'Q\underset{\sim}{y}/\sigma_e^2$ is equal to that of $\sum_{i=1}^{q} \lambda_i S_i^2$,

where q = rank Q, and $\lambda_i$ are the non-zero eigenvalues of QV and $S_i^2$

are independent $\chi^2(1)$ variables. This probability can easily be determined using a computer program such as that given by Koerts and Abrahamse (1969), see section 2.2, and an example is calculated by Verdooren (1974). In that way exact P-values and power calculations of the other test statistics can be found.

Now we consider the test of $H_0$: $\rho_1 \leq \rho_{10}$ against $H_1$: $\rho_1 > \rho_{10}$. For a three-stage nested design Wald's procedure, as indicated by Seely and El-Bassiouni (1983), does not work because rank $(X,Z_1,Z_2)$-rank$(X,Z_1) = 0$, for the column space of $Z_1$ lies in the column space of $Z_2$ (otherwise stated $A \subset B$).

Neither can a suggestion by Pincus (1977) be applied in this situation because we can not find a subspace of dimension larger than zero which is orthogonal to the range of X and $(Z_2Z_2')$, which could give information about $\sigma_A^2$.

We therefore propose our approach of section 3.1. Calculate for a given $\rho_2$:

$$\underset{\sim}{y}'P_{A^*}[V(\rho_{10})]^{-1}P_{A^*}\underset{\sim}{y} = SSA(\rho_{10}),\quad \underset{\sim}{y}'P_{B^*}[V(\rho_{10})]^{-1}P_{B^*}\underset{\sim}{y} = SSB(\rho_{10})$$

and use as a test statistic: $MSA(\rho_{10})/MSB(\rho_{10})$.
For each given value of $\rho_2$ we can calculate the exact P-value given an outcome of an experiment of $MSA(\rho_{10})/MSB(\rho_{10}) = f$. This P-value is $P(MSA(\rho_{10})/MSB(\rho_{10}) > f)$. Using the computer program we can evaluate

$$P(\underset{\sim}{y}'Q\underset{\sim}{y}/\sigma_e^2 > 0) = P(\sum_{i=1}^{q} \lambda_i S_i^2 > 0)$$

where $Q = P_{A^*}[V(\rho_{10})]^{-1}P_{A^*} - [f(a-1)/(b_.-a)]P_{B^*}[V(\rho_{10})]^{-1}P_{B^*}$, q is the rank of Q and $\lambda_i$ are the positive or negative eigenvalues of $QV(\rho_{10})$, and $S_i^2$ are independent $\chi^2(1)$ variables. We can calculate the P-value according to our $H_0$: $\rho_1 = \rho_{10}$ and we can decide whether $H_0$ must be rejected in favour of $H_1$: $\rho_1 > \rho_{10}$.

A better use of the information from a given value of $\rho_2$ can be found in improving the estimator for $\sigma_e^2$. The distribution of

$\underset{\sim}{y}'P_{B^*}[V(\rho_{10})]^{-1}P_{B^*}\underset{\sim}{y} = \underset{\sim}{y}'Q\underset{\sim}{y}$ is that of $\sigma_e \sum_{i=1}^{q} \lambda_i S_i^2$ where $S_i^2$ are independent $\chi^2(1)$ variables and $\lambda_i$ are the positive eigenvalues of $Q[V(\rho_{10})]$, or equivalently $[V(\rho_{10})]^{\frac{1}{2}}Q[V(\rho_{10})]^{\frac{1}{2}}$, and $q = b_{\cdot}-a$. Let $\{h_i\}$ be a complete set of orthonormal eigenvectors of $[V(\rho_{10})]^{-\frac{1}{2}}Q[V(\rho_{10})]^{-\frac{1}{2}} = M(B^*)$ corresponding to the positive eigenvalues of $\{\lambda_i\}$. Note that $\rho_{10}$ appears in $V(\rho_{10})$ but not in $M(B^*)$ and consequently not in $\{h_i\}$ and $\{\lambda_i\}$.

Now $\sum_{i=1}^{q} \underset{\sim}{y}'[V(\rho_{10})]^{-\frac{1}{2}}h_i h_i'[V(\rho_{10})]^{-\frac{1}{2}}\underset{\sim}{y}/\lambda_i = SS\tilde{B}$ has a $\sigma_e^2\chi^2(q)$-distribution for each given value of $\rho_2$. But $SSE=\underset{\sim}{y}'P_C\underset{\sim}{y}=\underset{\sim}{y}'P_C[V(\rho_{10})]^{-1}P_C\underset{\sim}{y}$ has, irrespective of $\rho_2$ and $\rho_{10}$, a $\sigma_e^2\chi^2(n_{\cdot\cdot}-b_{\cdot})$-distribution. Furthermore SSE and $SS\tilde{B}$ are independent, hence $SSD = SS\tilde{B} + SSE$ has a $\sigma_e^2\chi^2(n_{\cdot\cdot}-a)$-distribution. For each given value of $\rho_2$ we can use $MSA(\rho_{10})/MSD$ to test $H_0$: $\rho_1 \leq \rho_{10}$.

Also we can now use $MSA(\rho_{10})/MSD$ to get, after some iterations, a $(1-\alpha)$-confidence interval for $\rho_1$ for each given value of $\rho_2$. Of course this procedure must be handled by computer. But in principle we have for each interesting value of $\rho_2$ an answer for the test of $H_0$: $\rho_1 \leq \rho_{10}$ against $H_1$: $\rho_1 > \rho_{10}$.


## ACKNOWLEDGMENTS

The author would like to express his gratitude to mr. A. Otten for his helpful remarks, and to the referee for many improvements of this article.

## BIBLIOGRAPHY


Broemeling, L.D. (1969). Confidence regions for variance ratios of random models. J. Amer. Statist. Assoc. 64, 660-664.

Broemeling, L.D. (1978). Simultaneous inferences for variance ratios of some mixed linear models. Commun. Statist.-Theor. Meth. A7, 297-305.

Cummings, W.B. and Gaylor, D.W. (1974). Variance component testing in unbalanced nested designs. J. Amer. Statist. Assoc. 69, 765-771.

Davies, R.B. (1980). Algrorithm AS155: The distribution of a linear combination of $\chi^2$ random variables. Applied Statistics 29, 323-333.

Farebrother, R.W. (1980). Algorithm AS153. Pan's procedure for the tail probabilities of the Durbin-Watson statistic. Applied Statistics 29, 224-227 and 30(1981), 189.

Farebrother, R.W. (1984a). The distribution of a linear combination of central $\chi^2$ random variables. A remark on AS153: Pan's procedure for the tail probabilities of the Durbin-Watson statistic. Applied Statistics 33, 363-366.

Farebrother, R.W. (1984b). A remark on algorithms AS106, AS153 and AS155. The distribution of a linear combination of $\chi^2$ random variables. Applied Statistics 33, 366-369.

Fuller, W.A. and Battese, G.E. (1973). Transformations for estimation of linear models with nested-error structure. J. Amer. Statist. Assoc. 68, 626-632.

Graybill, F.A. (1976). Theory and Application of the Linear Model. North Scituate: Duxbury Press.

Graybill, F.A. and Wang, C.M. (1979) Confidence intervals for proportions of variability in two-factor nested variance components models. J. Amer. Statist. Assoc. 74, 368-374.

Harville, D.A. and Fenech, A.P. (1985). Confidence intervals for a variance ratio, or for heritability, in an unbalanced mixed linear model. Biometrics 41, 137-152.

Herbach, L.H. (1959). Properties of model II type analysis of variance tests, A. Optimum nature of the F-test for model II in the balanced case. Ann. Math. Statist. 30, 939-959.

Imhof, J.P. (1961). Computing the distribution of quadratic forms in normal variables. Biometrika 48, 419-426. Correction, Biometrika 49 (1962), 284.

Khuri, A.I. (1981). Simultaneous confidence intervals for functions of variance components in random models. J. Amer. Statist. Assoc. 76, 878-885.

Khuri, A.I. (1982). Direct products: a powerful tool for the analysis of balanced data. Commun. Statist. Theor.-Meth. A11, 2903-2920.

Koerts, J. and Abrahamse, A.P.J. (1969). On the theory and application of the general linear model. Rotterdam: Rotterdam University Press.

Kruskal, W.H. (1968). Note to Tietjen and Moore (1968) article. Biometrics 24, 1025-1026.

LaMotte, L.R. (1972). Notes on the covariance matrix of a random nested ANOVA model. Ann. Math. Statist. 43, 659-662.

Lehmann, E.L. (1959). Testing statistical hypotheses. New York: John Wiley and Sons.

Mostafa, M.G. (1967). Note on testing hypotheses in an unbalanced random effects model. Biometrika 54, 659-662.

Neudecker, H. (1968). The Kronecker matrix product and some of its applications in econometrics. Statistica Neerlandica 22, 69-82.

Pincus, R. (1977). On tests in variance components model. Math. Operationsforsch. Statist., Series Statistics 8, 251-255.

Rao, C.R. (1973). Linear statistical inference and its applications, second edition. New York: John Wiley and Sons, Inc.

Roebruck, P. (1982). On the optimality of F-test on variance components in mixed linear models. Biometrical Journal 24, 811-815.

Sahai, H. and Anderson, R.L. (1973). Confidence regions for variance ratios of random models for balanced data. J. Amer. Statist. Assoc. 68, 951-952.

Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. Biometrics Bulletin 2, 110-114.

Searle, S.R. (1971). Linear Models. New York: John Wiley and Sons, Inc.

Seely, J.F. and El-Bassiouni, Y. (1983). Applying Wald's variance component test. Ann. Statist. 11, 197-201.

SpjØtvoll, E. (1966). A mixed model in the analysis of variance. Optimal properties. Skandinavisk Aktuarietidskrift 49, 1-38.

SpjØtvoll, E. (1967). Optimum invariant tests in unbalanced variance component models. Ann. Math. Statist. 38, 422-428.

SpjØtvoll, E. (1968). Most powerful tests for some non-exponential families. Ann. Math. Statist. 39, 772-784.

Tan, W.Y. and Cheng, S.S. (1984). On testing variance components in three-stages unbalanced nested random effects models. Sankhyã B 46, 188-200.

Thomas, J.D. and Hultquist, R.A. (1978). Interval estimation for the unbalanced case of the one-way random effects model. Ann. Statist. 6, 582-587.

Thompson, W.A. (1955). On the ratio of variances in the mixed incomplete block model. Ann. Math. Statist. 26, 721-733.

Tietjen, G.L. (1974). Exact and approximate tests for unbalanced random effects designs. Biometrics 30, 573-581.

Tietjen, G.L. and Moore, R.H. (1968). On testing significance of components of variance in the unbalanced nested analysis of variance. Biometrics 24, 423-429.

Tong, L.Y. (1979). Counterexamples to a result of Broemeling on simultaneous inferences for variance ratios of some mixed linear models. Commun. Statist.-Theor. Meth. A8, 1197-1204.

Verdooren, L.R. (1974). Exact tests about variance ratios in unbalanced two and three stage nested designs. Contributed paper 8th International Biometric Conference, Constanta, Romania, August 25-30, 1974.

Verdooren, L.R. (1976). Exact confidence interval for the variance ratio in nested designs. Compstat 1976, ed. J. Gordesch and P. Naeve, Wien: Physica-Verlag, 221-228.

Verdooren, L.R. (1982). How large is the probability for the estimate of a variance component to be negative? Biometrical Journal 24, 339-360.

Wald, A. (1940). A note on the analysis of variance with unequal class frequencies. Ann. Math. Statist. 11, 96-100.

Wald, A. (1947). A note on regression analysis. Ann. Math. Statist. 18, 586-589.

Wang, C.M. and Graybill, F.A. (1981). Confidence intervals on a ratio of variances in the two-factor nested components of variance models. Commun. Statist.-Theor. Meth. A10, 1357-1368.

Yassaee, H. (1976). Probability integral of inverted Dirichlet distribution and its applications. Compstat 1976, ed. J. Gordesch and P. Naeve, Wien: Physica Verlag, 64-71.

Yassaee, H. (1981). On integrals of Dirichlet distributions and their applications. Commun. Statist.-Theor. Meth. A10, 897-906.

## 7.3. An example of the performance of the new exact test

We give an example of the performance of the exact tests about the variance ratios $\rho_2 = \sigma_B^2/\sigma_e^2$ and $\rho_1 = \sigma_A^2/\sigma_e^2$ for the three-stage nested design, as described in section 7.2 (subsection 3). The data are taken from Gill (1981) in which age-adjusted milk production records (305 days) are given, obtained in the same year and herd from cows whose sires and dams were considered randomly representative of a large population. Production records (in kg) are shown in table 7.3.1.

Table 7.3.1. Milk Production Records, kg (305 days)

| Sire | Dam | Production of full-sib daughters | | |
|---|---|---|---|---|
| 1 | 1 | 4379 | 6560 | |
| | 2 | 5560 | 7733 | 7198 |
| | 3 | 4637 | 5639 | 8072 |
| | 4 | 5726 | 5576 | |
| | 5 | 4968 | 4574 | |
| 2 | 6 | 5355 | 7057 | 7052 |
| | 7 | 4605 | 4180 | |
| | 8 | 4393 | 4530 | |
| 3 | 9 | 5195 | | |
| | 10 | 6137 | 4748 | 7351 |
| | 11 | 6253 | | |
| | 12 | 5553 | 6026 | 6666 |
| 4 | 13 | 6268 | 7575 | 7024 |
| | 14 | 7112 | | |
| | 15 | 5840 | 7316 | 6382 |
| | 16 | 6246 | 5595 | |
| | 17 | 5400 | 6440 | |
| | 18 | 7301 | 6615 | |
| | 19 | 5453 | | |
| | 20 | 7374 | 6693 | 6592 |

The statistical model for this example may be written

$$\underline{y}_{hik} = \mu + \underline{s}_h + \underline{d}_{hi} + \underline{e}_{hik} \qquad (7.3.1.)$$

$(k = 1,2,\ldots,n_{hi}$ for $i = 1,2,\ldots,b_h$ and $h = 1,2,3,4)$,

where $\mu$ is the general population mean, $\underline{s}_h$, $\underline{d}_{hi}$ and $\underline{e}_{hik}$ symbolize the

random effects of sires, dams, and progeny together with environment, respectively.

The analysis of variance table is given in table 7.3.2.; for the recipe of calculation see section 4.1.

Table 7.3.2. Analysis of Variance table for the data of Table 7.3.1.

| Source of Variation | df | SS | MS | E(MS) |
|---|---|---|---|---|
| Between sires | 3 | 8298165.5 | 2766055.2 (=MSS) | $\sigma_e^2 + K_2\sigma_D^2 + K_1\sigma_S^2$ |
| Between dams, within sires | 16 | 18089233.5 | 1130577.1 (=MSD) | $\sigma_e^2 + K_3\sigma_D^2$ |
| Error | 24 | 20639926.0 | 859996.9 (=MSE) | $\sigma_e^2$ |
| Corrected total | 43 | 47027325.0 | | |

The values are $K_3 = 2.135$; $K_2 = 2.462$ and $K_1 = 10.530$.

The ANOVA estimates for the variance components are:

$$\hat{\sigma}_e^2 = 859996.9;$$

$$\hat{\sigma}_D^2 = (1130577.1 - 859996.9)/2.135 = 126735.5;$$

$$\hat{\sigma}_S^2 = (2766055.2 - 2.462 \times 126735.5)/10.530 = 151380.4.$$

Outcomes of the biased estimators for $\rho_2 = \sigma_S^2/\sigma_e^2$ and $\rho_1 = \sigma_D^2/\sigma_e^2$ are $\hat{\rho}_2 = \hat{\sigma}_S^2/\hat{\sigma}_e^2 = 0.176$ and $\hat{\rho}_1 = \hat{\sigma}_D^2/\hat{\sigma}_e^2 = 0.147$ respectively.

For the test of $H_{02}$: $\rho_2 = 0$ against $H_{12}$: $\rho_2 > 0$, the test statistic f = MSD/MSE is used. The outcome of f in this experiment is 1.315; the P-value or critical level $\gamma = P(f > 1.315 \mid \rho_2 = 0) = 0.265$. The

power of this test for $\rho_2$ with significance level $\gamma = 0.265$ is

$\beta(\rho_2) = P(f > 1.315 \mid \rho_2)$.

Some values for $\beta(\rho_2)$ are given below.

| $\rho_2$ | $\beta(\rho_2)$ |
|---|---|
| 0 | $0.265 = \gamma$ |
| 0.1 | 0.419 |
| 0.2 | 0.558 |
| 0.5 | 0.821 |
| 1 | 0.960 |
| 2 | 0.997 |
| 2.5 | 0.999 |
| 3 | 1.000 |

For the test of $H_{02}$: $\rho_2 = \rho_{20}$ against $H_{12}$: $\rho_2 > \rho_{20}$, the test statistic $f = T(\rho_{20}) \times 24/16$ is used with $T(\rho_{20}) = SSB(\rho_{20})/SSE =$

$$\sum_{h=1}^{4} \sum_{i=1}^{b_h} g_{hi}(\bar{y}_{hi.} - \bar{\bar{y}}_{h..})^2 / \sum_{h=1}^{4} \sum_{i=1}^{b_h} \sum_{k=1}^{n_{hi}} (y_{hik} - \bar{y}_{hi.})^2$$

where

$$g_{hi} = n_{hi}/(\rho_{20}\, n_{hi} + 1); \quad \bar{y}_{hi.} = \sum_{k=1}^{n_{hi}} y_{hik}/n_{hi}$$

and

$$\bar{\bar{y}}_{hi.} = \sum_{i=1}^{b_h} g_{hi}\, \bar{y}_{hi.} / \sum_{i=1}^{b_h} g_{hi}.$$

For some values of $\rho_{20}$ and $\rho_2$ are given

the P-value or critical level

$$\gamma(\rho_{20}) = P(f(\rho_2) > f \mid \rho_2 = \rho_{20})$$

and some power values

$$\beta(\rho_2) = P(f(\rho_2) > f \mid \rho_2 > \rho_{20}).$$

$\rho_2 \geq \rho_{20}$ (Probabilities x 1000)

| $\rho_{20}$ | 0 | 0.1 | 0.2 | 0.5 | 1 | 2 | 2.5 | 3 |
|---|---|---|---|---|---|---|---|---|
| 0 | 265 | 419 | 558 | 821 | 960 | 997 | 999 | 1000 |
| 0.1 | | 430 | 567 | 826 | 961 | 997 | 999 | 1000 |
| 0.2 | | | 574 | 830 | 962 | 997 | 999 | 1000 |
| 0.5 | | | | 836 | 964 | 997 | 999 | 1000 |
| 1 | | | | | 965 | 997 | 999 | 1000 |
| 2 | | | | | | 997 | 999 | 1000 |
| 2.5 | | | | | | | 999 | 1000 |
| 3 | | | | | | | | 1000 |

For the test of $H_{01}$: $\rho_1 = \rho_{10}$ against $H_{11}$: $\rho_1 > \rho_{10}$ given the value of $\rho_2$, the test statistic $f = MSA(\rho_{10})/MSB(\rho_{10})$ is used where

$$MSA(\rho_{10}) = y'P_{A^*} [V(\rho_{10})]^{-1} P_{A^*}y/3$$

and

$$MSB(\rho_{10}) = y'P_{B^*} [V(\rho_{10})]^{-1} P_{B^*}y/16.$$

For some values of $\rho_{10}$, $\rho_1$ and $\rho_2$ are given the P-value or critical level

$$\gamma(\rho_{10} \mid \rho_2) = P(f(\rho_{10}) > f \mid \rho_{10}, \rho_2)$$

and some power values

$$\beta(\rho_1 \mid \rho_2) = P(f(\rho_1) > f \mid \rho_1 > \rho_{10}, \rho_2).$$

$\underline{\rho_2 = 0}$

$\rho_1 \geq \rho_{10}$ (Probabilities x 1000)

| $\rho_{10}$ | 0 | 0.1 | 0.2 | 0.5 | 1 | 2 | 2.5 | 3 |
|---|---|---|---|---|---|---|---|---|
| 0 | 101 | 343 | 514 | 756 | 882 | 950 | 963 | 971 |
| 0.1 | | 355 | 525 | 764 | 887 | 952 | 965 | 972 |
| 0.2 | | | 525 | 764 | 887 | 952 | 965 | 972 |
| 0.5 | | | | 762 | 886 | 952 | 964 | 972 |
| 1 | | | | | 885 | 952 | 964 | 972 |
| 2 | | | | | | 952 | 964 | 972 |
| 2.5 | | | | | | | 964 | 972 |
| 3 | | | | | | | | 972 |

$\rho_2 = 0.1$

$\rho_1 \geq \rho_{10}$ (Probabilities x 1000)

| $\rho_{10}$ | 0 | 0.1 | 0.2 | 0.5 | 1 | 2 | 2.5 | 3 |
|---|---|---|---|---|---|---|---|---|
| 0 | 104 | 304 | 460 | 706 | 851 | 935 | 951 | 962 |
| 0.1 | | 316 | 472 | 716 | 857 | 938 | 953 | 963 |
| 0.2 | | | 472 | 716 | 957 | 938 | 954 | 964 |
| 0.5 | | | | 714 | 856 | 937 | 953 | 963 |
| 1 | | | | | 855 | 937 | 953 | 963 |
| 2 | | | | | | 937 | 953 | 963 |
| 2.5 | | | | | | | 953 | 963 |
| 3 | | | | | | | | 963 |

$\rho_2 = 0.2$

$\rho_1 \geq \rho_{10}$ (Probabilities x 1000)

| $\rho_{10}$ | 0 | 0.1 | 0.2 | 0.5 | 1 | 2 | 2.5 | 3 |
|---|---|---|---|---|---|---|---|---|
| 0 | 106 | 278 | 420 | 664 | 822 | 920 | 939 | 952 |
| 0.1 | | 290 | 431 | 674 | 829 | 923 | 942 | 954 |
| 0.2 | | | 432 | 674 | 829 | 923 | 942 | 954 |
| 0.5 | | | | 673 | 828 | 922 | 942 | 954 |
| 1 | | | | | 827 | 922 | 941 | 954 |
| 2 | | | | | | 922 | 941 | 954 |
| 2.5 | | | | | | | 941 | 954 |
| 3 | | | | | | | | 954 |

$\rho_2 = 0.5$

$\rho_1 \geq \rho_{10}$ (Probabilities x 1000)

| $\rho_{10}$ | 0 | 0.1 | 0.2 | 0.5 | 1 | 2 | 2.5 | 3 |
|---|---|---|---|---|---|---|---|---|
| 0 | 113 | 233 | 342 | 566 | 745 | 875 | 904 | 923 |
| 0.1 | | 243 | 352 | 576 | 753 | 880 | 907 | 926 |
| 0.2 | | | 355 | 578 | 754 | 880 | 908 | 926 |
| 0.5 | | | | 577 | 753 | 880 | 908 | 926 |
| 1 | | | | | 752 | 879 | 907 | 926 |
| 2 | | | | | | 878 | 907 | 925 |
| 2.5 | | | | | | | 906 | 925 |
| 3 | | | | | | | | 925 |

$\rho_2 = 1.0$

$\rho_1 \geq \rho_{10}$ (Probabilities x 1000)

| $\rho_{10}$ | 0 | 0.1 | 0.2 | 0.5 | 1 | 2 | 2.5 | 3 |
|---|---|---|---|---|---|---|---|---|
| 0 | 119 | 200 | 278 | 463 | 646 | 807 | 847 | 875 |
| 0.1 | | 208 | 286 | 471 | 653 | 812 | 851 | 879 |
| 0.2 | | | 289 | 474 | 655 | 814 | 853 | 880 |
| 0.5 | | | | 475 | 656 | 814 | 853 | 880 |
| 1 | | | | | 654 | 813 | 852 | 879 |
| 2 | | | | | | 812 | 851 | 879 |
| 2.5 | | | | | | | 851 | 878 |
| 3 | | | | | | | | 878 |

$\rho_2 = 2.0$

$\rho_1 \geq \rho_{10}$ (Probabilities x 1000)

| $\rho_{10}$ | 0 | 0.1 | 0.2 | 0.5 | 1 | 2 | 2.5 | 3 |
|---|---|---|---|---|---|---|---|---|
| 0 | 126 | 175 | 223 | 354 | 515 | 695 | 748 | 788 |
| 0.1 | | 181 | 230 | 361 | 521 | 701 | 753 | 792 |
| 0.2 | | | 233 | 364 | 524 | 703 | 755 | 794 |
| 0.5 | | | | 367 | 527 | 705 | 757 | 795 |
| 1 | | | | | 526 | 705 | 756 | 795 |
| 2 | | | | | | 703 | 755 | 793 |
| 2.5 | | | | | | | 754 | 793 |
| 3 | | | | | | | | 792 |

$\rho_2 = 2.5$

$\rho_1 \geq \rho_{10}$ (Probabilities x 1000)

| $\rho_{10}$ | 0 | 0.1 | 0.2 | 0.5 | 1 | 2 | 2.5 | 3 |
|---|---|---|---|---|---|---|---|---|
| 0 | 128 | 169 | 210 | 323 | 470 | 651 | 706 | 749 |
| 0.1 | | 174 | 215 | 329 | 476 | 656 | 711 | 753 |
| 0.2 | | | 218 | 332 | 479 | 658 | 714 | 756 |
| 0.5 | | | | 336 | 483 | 661 | 716 | 758 |
| 1 | | | | | 483 | 661 | 716 | 756 |
| 2 | | | | | | 659 | 714 | 756 |
| 2.5 | | | | | | | 713 | 755 |
| 3 | | | | | | | | 755 |

## REFERENCE

Gill, J.L. (1981). Design and analysis of experiments in the animal and medical sciences, vol. 1. The Iowa State University Press, Ames, Iowa, 204 - 207.

CHAPTER 8

# 8. ESTIMATION OF VARIANCE COMPONENTS

## PART II: CONTINUATION

### 8.1. Introduction

In chapter 6 the onset of the author's research on estimation of variance components was given.

Because many statisticians did not easily grasp the rationale of Rao's minimum norm derivation in the MINQUE, the unified least squares approach in the derivation of estimators for variance components is given in section 8.2.

Also recent developments in permissible (hence non-negative) estimators for variance components are presented.

In section 8.3 the connection between the calculation of Maximum Likelihood (ML) estimators, Restricted Maximum Likelihood (REML) estimators and Iterative Minimum Norm Quadratic Unbiased Estimators (I-MINQUE) is elucidated.

Furthermore some recommendations are made for the procedure of estimating variance components.

Finally, no discussion of the Bayesian approach in variance component estimation will be found here. Previous knowledge about variance components must of course be incorporated. This can be effected by using prior weights for ratios of variance components in the dispersion matrix, which is used in a non-negative modification of a least squares estimator (or MINQUE) for the variance components.

## 8.2. Least squares estimators and non-negative estimators of variance components

Paper presented at the International Conference on the "Analysis of the Unbalanced Mixed Model" at Gainesville, Florida, U.S.A., in 1987.

Submitted for publication to:
Communications in Statistics, Part A: Theory and Methods.

Submitted April, 1987;

Revised October, 1987;

Accepted November, 1987.

# LEAST SQUARES ESTIMATORS AND NON-NEGATIVE ESTIMATORS OF VARIANCE COMPONENTS

L.R. Verdooren

Department of Mathematics
Agricultural University Wageningen
The Netherlands

Key Words and Phrases: permissible estimators; MINQUE; MIVQUE; ML; REML; quadratic or convex programming.

ABSTRACT

Using the least squares method we have a unified procedure for the derivation of estimators for variance components in the linear model of ZZ', where Z is the residual vector from a simple least squares fit for $X\beta$. These least squares estimators are unbiased but not always non-negative. The invariant and unbiased least squares estimators for variance components are the MINQUE estimators. For multivariate normally distributed variables the MINQUE is the same as the MIVQUE.

Non-negative estimators of variance components are permissible estimators. Taking into account the constraints of non-negativity of the variance components, a quadratic programming procedure is suggested.

For multivariate normally distributed variables the ML method or REML method can be used. In practice many computer programs do

not take into account the non-negativity constraints on the parameter space. Henderson's iterative procedure using the mixed model equations guarantees a non-negative solution for REML.

## 1. INTRODUCTION

The estimation of variance components for the analysis of variance models has a long history. A good expository review has been given by Khuri and Sahai (1985). Several times one can find remarks on negative estimates produced by certain estimators. Rao (1977) and Kleffe (1980) reviewed the estimation procedures based on the MINQUE-theory. However, leading principles in these papers are properties such as unbiasedness and minimum variance of the unbiased estimator. These properties are often not so relevant for the estimation procedures for variance components. We feel that the leading principle must be permissible estimation and, for the usually adopted quadratic loss function, the minimum mean squared error.

We examine the least squares method for estimating variance components. This method gives for many situations a unified procedure, which can also be adapted to meet the non-negativity constraints for variance components.

We introduce in section 2 the concept of permissible estimation. In section 3 the least squares estimators are discussed in the linear model and in the dispersion-mean model. In section 4 a procedure has been given which can be adopted to produce non-negative variance component estimators. In section 5 some remarks on non-negative ML and REML estimators for normally distributed variables are made.

## 2. PERMISSIBLE ESTIMATION

At the VI-th International Conference on Mathematical

Statistics at Wisła, Poland (7-13 December 1978), the concept of permissible estimation has been introduced, see Verdooren (1980). In our view this concept must be the first requirement for estimating variance components.

Consider a random variable $\underline{y}$ which assumes values in the sample space Y according to a probability distribution P which is known to belong to a given class $\{P_\theta;\ \theta \in \Theta\}$ of probability distributions, and let y denote the observed value. If $P = P_\theta$, $\theta$ is the true value of the underlying unknown parameter. We shall be interested in the estimation of a real valued parameter $g: \Theta \to R$, or equivalently of the corresponding true value $\gamma = g(\theta)$. We need an estimator $d: Y \to R$ with the interpretation that $d(y)$ will be used as an estimate for $\gamma = g(\theta)$ if the outcome $y \in Y$ of $\underline{y}$ is observed.

Estimators will be compared by means of their mean squared error (MSE) (or risk function if squared error loss is considered):

$$MSE_\theta(d) = E_\theta(d(\underline{y}) - g(\theta))^2 = var_\theta d(\underline{y}) + (bias_\theta(d))^2$$

where $bias_\theta(d) = E_\theta d(\underline{y}) - g(\theta)$.

An estimator is said to be <u>admissible</u> if no estimator exists with uniformly (in $\theta$) smaller MSE. Or, more precisely, d is admissible if no d' exists with $MSE_\theta(d') \leq MSE_\theta(d)$ for all $\theta \in \Theta$ and $MSE_\theta(d') < MSE_\theta(d)$ for at least one $\theta$.

In our view it is worthwhile to introduce the concept of <u>permissible estimation</u> as a necessary condition. Let $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n$ be a random sample of a random variable $\underline{y}$ with probability distribution $P \in \{P_\theta;\ \theta \in \Theta\}$. An estimator $\underline{t} = t(\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n)$ for $\theta$ is said to be permissible if $\underline{t} \in \Theta$ holds with probability one. In some situations $\Gamma = \{g(\theta)\} = \{\gamma;\ \gamma = g(\theta)$ for some $\theta \in \Theta\}$ is an interval of the form $[a,b]$ or $[a,\infty)$ or $(-\infty,b]$ or some other bounded set. If the estimator $d: Y \to R$ takes on values outside $\Gamma$ with positive probability, (and hence is not permissible), then it

can be improved by truncating. This is seen immediately as follows.

Let $d: Y \to R$ be any estimator for $\gamma = g(\theta)$ (it need not be unbiased) such that $\Gamma = \{g(\theta)\} = [a,b]$ and $P_\theta\{d(\underline{y}) \notin \Gamma\} > 0$ for at least one $\theta \in \Theta$. Let $d'$ be the corresponding truncated estimator:

$$d'(y) = \begin{cases} a & \text{if } d(y) < a \\ d(y) & \text{if } a \leq d(y) \leq b \\ b & \text{if } d(y) > b. \end{cases}$$

Now for any $\theta \in \Theta$ we have $g(\theta) \in \Gamma$ and hence $(d(y) - g(\theta))^2 \geq (d'(y) - g(\theta))^2$ for all $y \in Y$. Hence $MSE_\theta(d) \geq MSE_\theta(d')$ for all $\theta \in \Theta$. For $\theta$ with $P_\theta\{d(\underline{y}) \notin \Gamma\} > 0$, we obviously have strict inequality and hence d is inadmissible.

Note that this result can be extended to the case where $\underline{y}$ is a random vector and $\Theta$ is a closed convex set. Truncation here means that an estimate which falls outside this set $\Theta$ is replaced by the boundary point of $\Theta$ nearest to it.

Examples of not permissible estimators are probability estimators with possible values outside [0,1]; and estimators for a parameter known to be positive, which may take occasionally negative values.

For the sake of illustration, consider the balanced one-way random effects model:

$$\underline{y}_{ij} = \mu + \underline{a}_i + \underline{e}_{ij}, \; i = (1,2,\ldots, m); \; j = (1,2,\ldots, n).$$

The random variables $\underline{a}_i$ are $N(0,\sigma_1^2)$ and are identically and independently distributed (i.i.d); the random variables $\underline{e}_{ij}$ are $N(0,\sigma_2^2)$ and also i.i.d.; the $\underline{a}_i$ and $\underline{e}_{ij}$ are assumed to be independent. By calculating

$$\underline{CT} = \Big(\sum_{i=1}^{m}\sum_{j=1}^{n} \underline{y}_{ij}\Big)^2/(mn);\quad \underline{SSA} = \sum_{i=1}^{m}\Big(\sum_{j=1}^{n} \underline{y}_{ij}\Big)^2/n - \underline{CT},$$

$$\underline{SSR} = \sum_{i=1}^{m}\sum_{j=1}^{n} \underline{y}_{ij}^2 - \underline{CT} - \underline{SSA};\quad \underline{MSA} = \underline{SSA}/(m-1),\ \underline{MSR} = \underline{SSR}/(m(n-1))$$

we have an unbiased estimator $\underline{\hat{\sigma}}_2^2 = \underline{MSR}$ for $\sigma_2^2$, and an unbiased estimator for $\sigma_1^2$ is

$$\underline{\hat{\sigma}}_1^2 = (\underline{MSA}-\underline{MSR})/n.$$

Notice that

$$\hat{\sigma}_2^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} (\underline{y}_{ij} - \bar{\underline{y}}_{i.})^2/(m(n-1))$$

is a permissible estimator for $\sigma_2^2$. However $\underline{\hat{\sigma}}_1^2$ is not a permissible estimator since there is a positive probability that $\underline{\hat{\sigma}}_1^2$ takes negative values under the model assumed! We have

$$P(\underline{\hat{\sigma}}_1^2 \leq 0) = P((\underline{MSA} - \underline{MSR})/n \leq 0) = P(\underline{MSA}/\underline{MSR} \leq 1)$$

$$= P(\underline{F}(m-1),m(n-1)) \leq (1 + n\gamma)^{-1}) \quad \text{where} \quad \gamma = \sigma_1^2/\sigma_2^2.$$

Verdooren (1982) gives extensive tables for this probability for several values of m, n and $\gamma$. From this table it can be seen that negative estimates can occur frequently, in certain cases 25 % of the time.

It is possible to use the biased estimator $\max(0,\underline{\hat{\sigma}}_1^2)$, which

has smaller MSE than $\underline{\theta}_1^2$ for all values of $(\sigma_1^2,\sigma_2^2)$.

Another example is given by the estimators in the linear regression model with unequal variances. See Rao and Subrahmaniam (1971), Hartley and Jayatillake (1973), Chaubey and Rao (1976) and Kleffe (1984).

## 3. LEAST SQUARES ESTIMATORS

### 3.1. The linear model

Let us consider the linear model

$$\underline{y} = X\beta + \underline{e},$$

where $y,e \in R^n$, X is an nxp-matrix of full rank p, the parameter vector $\beta \in R^p$; $E(\underline{e}) = 0$, dispersion matrix $D(\underline{e}) = D(\underline{y}) = \sigma^2 I_n$. It is known that the simple or ordinary least squares estimator for $\beta$, $\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{y}$, is unbiased. Each linear combination $c'\beta$ (or written as inner product of c and $\beta$: $(c,\beta)$) has as an unbiased estimator

$$c'\hat{\underline{\beta}} = (c,\hat{\underline{\beta}}).$$

Let X be partitioned as $(X_1,X_2)$ with $X_1$ and $X_2$ of full column rank $p_1$ and $p_2$ respectively ($p_1 + p_2 = p$) and $\beta = (\beta_1',\beta_2')'$. If we wish to remove the nuisance parameters $\beta_1$ while estimating $\beta_2$ unbiasedly, we may write the expectation vector $E\underline{y} = X\beta$ in the form

$$E\underline{y} = X_1\beta_1 + X_2\beta_2 = X_1\beta_1 + P_1X_2\beta_2 + (I_n-P_1)X_2\beta_2 = X_1\alpha + M_1X_2\beta_2$$

where $P_1$ is the orthogonal projector on the column space of $X_1$,

$P_1 = X_1(X_1'X_1)^{-1}X_1'$ and $M_1 = I_n - P_1$, the orthogonal projector on the *null space of* $X_1$.

A linear combination $\lambda'\beta_2 = (\lambda,\beta_2)$ of parameters is identifiable if and only if there is a $\rho$ such that

$$X_2'M_1X_2\rho = \lambda.$$

In this case of full rank, any linear combination of the parameters is identifiable. From the normal equations we estimate $\beta_2$ from

$$X_2'M_1X_2\hat{\underline{\beta}}_2 = X_2'M_1\underline{y},$$

and the linear unbiased estimator for

$$(\lambda,\beta_2) = \lambda'\beta_2 = \sum_{i=1}^{p_2} \lambda_i\beta_{2i}$$

is given by

$$(M_1X_2\rho,\underline{y}) = \sum_{i=1}^{p_2} \rho_i(M_1x_{2i},\underline{y}),$$

where $x_{2i}$, $i = (1,2,\ldots,p_2)$ are the columns of $X_2$.

If the linear model $\underline{y} = X\beta + \underline{e}$ has dispersion matrix

$$D(\underline{e}) = D(\underline{y}) = \sigma^2V$$

and V is symmetric and positive definite, it is known that the ordinary least squares estimator for $\beta$, $\hat{\underline{\beta}} = X(X'X)^{-1}X'\underline{y}$, is unbiased. Let V be known, $V = V^{\frac{1}{2}}V^{\frac{1}{2}}$ with $V^{\frac{1}{2}}$ symmetric. Now we transform the model by multiplying with $V^{-\frac{1}{2}}$, that is,

$$V^{-\frac{1}{2}}\underline{y} = V^{-\frac{1}{2}}X\beta + V^{-\frac{1}{2}}\underline{e}.$$

or $\tilde{\underline{y}} = \tilde{X}\beta + \tilde{\underline{e}}$ with $D(\tilde{\underline{e}}) = D(\tilde{\underline{y}}) = \sigma^2 I_n$.

In the transformed linear model, we apply the ordinary least squares estimator for $\beta$, which gives

$$\tilde{\underline{\beta}} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{\underline{y}}.$$

Substituting from the above transformation, the generalized least squares estimator is

$$\tilde{\underline{\beta}} = (X'V^{-1}X)^{-1}X'V^{-1}\underline{y},$$

which is the best linear unbiased estimator or Gauss-Markov estimator for $\beta$.

The ordinary least squares estimator will be the same as the generalized least squares if and only if the range of VX is contained in the range of X and rank of X = p.

## 3.2. Variance Components model

Let us consider the variance components model:

$$\underline{y} = X\beta + U_1\underline{a}_1 + \ldots + U_k\underline{a}_k,$$

where $\underline{y}$ is an nx1-vector of observations, X is an nxp-design matrix of known constants and of full rank ($p < n$), $\beta$ is px1-vector of fixed effects parameters, $U_i$ is a known $nxm_i$-matrix, $\underline{a}_i$ is an $m_i$x1-vector of random effects such that

(1) $E(\underline{a}_i) = 0$,
(2) $D(\underline{a}_i) = \sigma_i^2 I_{m_i}$, $i = (1,2,\ldots,k)$.

Further, $U_k = I_n$ and $\text{var}(\underline{a}_k) = \sigma_k^2 > 0$. The column space of $(X, U_1, U_2, \ldots, U_k)$ is $R^n$. These assumptions imply that

$$D(\underline{y}) = \sum_{i=1}^{k} \sigma_i^2 U_i U_i' = \sigma_1^2 V_1 + \sigma_2^2 V_2 + \ldots + \sigma_{k-1}^2 V_{k-1} + \sigma_k^2 I_n$$

$$= \sigma^2(\gamma_1 V_1 + \gamma_2 V_2 + \ldots + \gamma_{k-1} V_{k-1} + I_n) = \sigma^2 V$$

where the variance components $\sigma_i^2 = \sigma^2 \gamma_i$ for $i = (1,2,\ldots,k-1)$, $\sigma_k^2 = \sigma^2$, and $D(\underline{y})$ is positive definite.

The least squares approach to estimate variance components or linear combinations of variance components has been used by Seely (1970 a,b) and Seely and Zyskind (1971). Instead of quadratic functions of $\underline{y}$, Seely utilizes linear functions of $\underline{Z} = \underline{y}\underline{y}'$ in S, a vector space of symmetric nxn-matrices, with inner product of A, B $\in$ S defined by $(A,B) = \text{tr } A'B = \text{tr } AB$, where tr A means trace of A. The linear estimator for variance components is $(A,\underline{Z})$. The expected value of $\underline{Z}$ is:

$$E(\underline{Z}) = \sum_{\substack{i=1 \\ i \leq j}}^{p} \sum_{j=1}^{p} \beta_i \beta_j B_{ij} + \sum_{i=1}^{k} \sigma_i^2 V_i$$

where $B_{ii} = x_i x_i'$ and $B_{ij} = x_i x_j' + x_j x_i'$ for $i < j$ are symmetric and $x_j$ is the j-th column of X. Alternatively,

$$E(\underline{Z}) = X\Psi X' + \sum_{i=1}^{k} \sigma_i^2 V_i \quad \text{with } \Psi = \beta\beta'.$$

Hence $E(\underline{Z})$ belongs to $\{\text{span } B_0 + \text{span } B_1\}$ with

$$B_0 = \{B_{ij};\ 1 \leq i \leq j \leq p\} \text{ and } B_1 = \{V_1, V_2, \ldots, V_k\}.$$

This model has a form similar to the one in section 3.1.

The disperson matrix of $\underline{Z}$ consists of the product of the

second and fourth moments of $\underline{y}$. Let vec A be the column vector obtained from writing the rows of the matrix A one behind the other and transposing the resulting row vector. Then the inner product of A and B is the ordinary inner product of vec A and vec B: (A,B) = (vec A)'.(vec B) = tr(A'B).

We have a nuisance parameter $\Psi$ and parameters $\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2$ with the column vectors vec $V_1$, vec $V_2$, ..., vec $V_k$, respectively. We apply the ordinary least squares method to this model after eliminating the nuisance parameter $\Psi$. To remove the nuisance parameter $\Psi$ we use the linear operator $I - \Pi$ where I is the identity mapping of S to S, $\Pi$ is the projection operator on $B_0$. In this case $\Pi(A) = PAP$, where P is the orthogonal projection on the column space of X, i.e.

$$P = X(X'X)^{-1}X', \text{ and } (I-\Pi)(A) = A - PAP.$$

The unbiased estimation of $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ is possible if the equations

$$\begin{bmatrix} (V_1,(I-\Pi)(V_1)), & \ldots, & (V_1,(I-\Pi)(V_k)) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ (V_k,(I-\Pi)(V_1)), & \ldots, & (V_k,(I-\Pi)(V_k)) \end{bmatrix} \begin{bmatrix} \rho_1 \\ \cdot \\ \cdot \\ \cdot \\ \rho_k \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \cdot \\ \cdot \\ \cdot \\ \lambda_k \end{bmatrix}$$

or $[\mathrm{tr}\{V_i(V_j - PV_jP)\}]\rho = \lambda$ have a solution.

The unbiased estimator of $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ is then given by

$$\sum_{i=1}^{k} \rho_i((I-\Pi)(V_i),\underline{Z}) = \sum_{i=1}^{k} \rho_i(\underline{y}'(V_i-PV_iP)\underline{y}).$$

Seely (1970b) uses the operator

$$\pi(A) = \tfrac{1}{2}\{(I-P)A + A(I-P)\}$$

to remove $\Psi$ and obtained the normal equations

$$\begin{bmatrix} \mathrm{tr}(V_1(I-P)V_1) & \dots & \mathrm{tr}(V_1(I-P)V_k) \\ \vdots & & \vdots \\ \mathrm{tr}(V_k(I-P)V_1) & \dots & \mathrm{tr}(V_k(I-P)V_k) \end{bmatrix} \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_k \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{bmatrix}.$$

The unbiased estimator of $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ is given by

$$\sum_{i=1}^{k} \rho_i (\underline{y}'(I-P)V_i\underline{y}) .$$

In Rao (1973) one finds that the conditions for an invariant and unbiased estimator $\underline{y}'A\underline{y}$ of $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ are:

$$\left.\begin{aligned} &AX = 0 \\ &\mathrm{tr}(AV_i) = \lambda_i \quad i = (1,2, \dots, k) \end{aligned}\right] .$$

Seely (1972) shows that $Q'\underline{y}$ is a maximal invariant statistic from $R^n$ to $R^q$, where Q is a nxq-matrix with q = n-rank X, column space of Q = null space of X' and $Q'Q = I_q$, i.e. $QQ' = I-P = M$, where P is the orthogonal projector on the column space of X. Now we use $QQ'\underline{y} = M\underline{y}$ or the simple residuals as our invariant statistic. We have, since $M = M'$,

$$E(M\underline{y}) = MX\beta = 0 \text{ and } D(M\underline{y}) = MD(\underline{y})M = \sum_{i=1}^{k} \sigma_i^2 MV_iM.$$

The set of normal equations for the least squares estimator of

$\sum_{i=1}^{k} \lambda_i \sigma_i^2$ gives a solution

$$\sum_{i=1}^{k} \rho_i (MV_iM, M\underline{y}\,\underline{y}'M)$$

where the $\rho_i$'s satisfy the matrix equation $\{(MV_iM, MV_jM)\}\rho = \lambda$.
Using the properties of trace operators we find that the invariant

and unbiased estimator of $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ can be written as $\underline{y}'A_*\underline{y}$ where

$A_* = \sum_{i=1}^{k} \rho_i MV_iM$ and $\rho_1, \ldots, \rho_k$ satisfy $\sum_{i=1}^{k} \rho_i tr(MV_iMV_j) = \lambda_j$ for

$j = (1, 2, .., k)$.

If we know that $D(\underline{y})$ is of the form $T = \sum_{i=1}^{k} \alpha_i^2 V_i$, or in other

words, if $\alpha_i^2$ is the a priori (or approximate) value of $\sigma_i^2$,
we first transform $\underline{y}$ into $\tilde{\underline{y}} = T^{-\frac{1}{2}}\underline{y}$ with $E\tilde{\underline{y}} = T^{-\frac{1}{2}}X\beta = \tilde{X}\beta$

and $D(\tilde{\underline{y}}) = \sum_{i=1}^{k} \sigma_i^2 T^{-\frac{1}{2}} V_i T^{-\frac{1}{2}} = \sum_{i=1}^{k} \sigma_i^2 \tilde{V}_i$.

The invariant statistic is now $\tilde{M}\tilde{\underline{y}}$ with $\tilde{M} = I - \tilde{P}$ and $\tilde{P} = \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$.

The invariant unbiased estimator of $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ is then given by

$\sum_{i=1}^{k} \rho_i (\tilde{MV_iM}, \tilde{M\underline{y}\underline{y}'M})$ where the $\rho_i$'s satisfy the matrix equations

$$\{\tilde{MV_iM}, \tilde{MV_jM})\}\rho = \lambda.$$

The invariant and unbiased estimator of $\sum_{i=1}^{k} \lambda_i \sigma_i^2$, given $T = \sum_{i=1}^{k} \alpha_i^2 V_i$, can then be written as $\underline{y}'A_*\underline{y}$ where

$$A_* = \sum_{i=1}^{k} \rho_i (I-P_T')T^{-1}V_iT^{-1}(I-P_T).$$

The coefficients $\rho_1, \ldots, \rho_k$ satisfy

$$\sum_{i=1}^{k} \rho_i \, tr(I-P_T')T^{-1}V_iT^{-1}(I-P_T)V_j = \lambda_j \quad \text{for } j = (1,2,\ldots,k),$$

where

$$P_T = X(X'T^{-1}X)^{-1}X'T^{-1}.$$

This is the invariant MINQUE (Minimum Norm Quadratic Unbiased Estimator) of Rao (1970, 1971, 1972, 1973). We have derived here these estimators using the well known least squares procedure. The rationale of Rao's minimum norm is not so easily understood by several statisticians.

When $\underline{y}$ has a n-variate normal distribution, since $Ax = 0$,

$$var(\underline{y}'A\underline{y}) = 2tr(AD(\underline{y})AD(\underline{y})) + 4\beta'X'AD(\underline{y})AX\beta = 2tr(AD(\underline{y})AD(\underline{y})).$$

Thus, if the true variance components are $\alpha_1^2, \ldots, \alpha_k^2$, tr(ATAT)

is proportional to var($\underline{y}'A\underline{y}$) and the invariant MINQUE is the locally minimum variance unbiased quadratic estimator under multivariate normal distribution of $\underline{y}$.

Finally we remark that Seely (1971) shows that when $\underline{y}$ has a multivariate normal distribution, the best quadratic invariant and unbiased estimator of $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ is a function of $M\underline{y}$ if and only if $B = \text{span}\{MV_1M, \ldots, MV_kM\}$ is a quadratic subspace of all real symmetric matrices S. A subspace B of S with the property that $B_1 \in B$ implies $B_1^2 \in B$ is said to be a quadratic subspace of S, that is, the elements of B form a Jordan algebra with operator $\frac{1}{2}[(A,B)+(B,A)]$. Notice that in this case the ordinary least squares estimator of $\sigma_i^2$ is equal to the generalized least squares estimator.

Seely (1972, 1977) used the notion of quadratic subspaces also in the derivation of completeness of certain statistics for a family of multivariate normal distributions.

The approach of looking at the quadratic estimators $\underline{y}'A\underline{y}$ as linear functions of $\underline{y}\underline{y}'$ has been exploited by Pukelsheim (1976, 1977). That the mean and dispersion share the same structure (dispersion-mean correspondence) has been elaborated by Pukelsheim (1977). See also Drygas (1977), Pukelsheim (1979) and Anderson, Henderson, Pukelsheim and Searle (1984) for overviews of this topic.

If $\underline{y}$ has a multivariate normal distribution the least squares approach gives the MINQUE which is the same as the MIVQUE (Minimum Variance Quadratic Unbiased Estimator). See Swallow and Searle (1978) for the comparison of MIVQUE and ANOVA estimators.

## 4. NON-NEGATIVE ESTIMATION

The question now arises whether the least squares solutions of the variance components in the dispersion-mean correspondence model approach, are permissible estimators. For balanced designs these estimators (thus also the MINQUE estimator) are the same as the ANOVA estimators. We saw already in section 2 that the ANOVA estimator is not permissible. For unbalanced designs with two or more classifications, there are no unique ANOVA estimators. For the one-way model with unequal variances several estimators for the mean and the variances are examined by Rao, Kaplan and Cochran(1981). Rao and Chaubey (1978) showed that for the model

$$\underline{y}_{ij} = \mu_i + \underline{e}_{ij},\ j = (1,2,\ldots,n_i) \text{ for } i = (1,2,\ldots,k),$$

with $E(\underline{e}_{ij}) = 0$, $var(\underline{e}_{ij}) = \sigma_i^2$ and independent $\underline{e}_{ij}$'s, non-negativeness and unbiasedness result in $\underline{s}_i^2 = \sum_{j=1}^{n} (\underline{y}_{ij}-\bar{\underline{y}}_{i.})^2/(n_i-1)$ as the estimator for $\sigma_i^2$.

LaMotte (1973) shows that for the variance components models the only individual variance component which can be estimated unbiasedly by a non-negative quadratic function $\underline{y}'A\underline{y}$ is $\sigma_k^2$, the last variance component; $\sigma_k^2$ can have a non-negative estimator only if all $V_i$, $i = (1,2,\ldots,k-1)$ are singular. (Note $V_k = I_n$ is non-singular). Further $\sum_{i=1}^{k} \lambda_i \sigma_i^2$ with all $\lambda_i$ non-negative, may be estimated unbiasedly by a non-negative quadratic $\underline{y}'A\underline{y}$. For the balanced

one-way lay-out of the example in section 2, $\lambda_1\sigma_1^2 + \lambda_2\sigma_2^2$ has an unbiased estimator $\underline{y}'A\underline{y}$ with A symmetric and non-negative definite if and only if $\lambda_2 \geq n^{-1}\lambda_1 \geq 0$. Clearly $\sigma_2^2 = E(\underline{MSR})$ and $\sigma_2^2 + n\sigma_1^2 = E(\underline{MSA})$ can be estimated by a non-negative quadratic.

We shall now describe a procedure for obtaining non-negative estimators from the least squares residuals. Consider the model

$$\underline{y} = X\beta + \sum_{i=1}^{k} U_i\underline{a}_i,$$

with the positive definite dispersion matrix

$D(\underline{y}) = \sigma_k^2(\gamma_1 V_1 + \ldots + \gamma_{k-1}V_{k-1} + I_n)$, (see the top of section 3.2). Consider the residuals $M\underline{y}$, $M = I - P$ with $P = X(X'X)^{-1}X'$ the orthogonal projector on the column space of X. $M\underline{y}$ is a maximal invariant statistic in a subspace of dimension $q = n - \text{rank } X$. We now have the residual model

$$M\underline{y} \text{ with } E(M\underline{y}) = 0 \text{ and } D(M\underline{y}) = \sum_{i=1}^{k} \sigma_i^2 MV_iM.$$

The natural parameter set for $(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)'$ is

$$G_M = \{(s_1^2, s_2^2, \ldots, s_k^2)' \in R^k \mid \sum_{i=1}^{k} s_i^2 MV_iM \in S\},$$

where $\sum_{i=1}^{k} s_i^2 MV_iM$ is positive definite and S is the set of all real symmetric non-negative definite nxn-matrices.

Pukelsheim (1981 a,b) characterized non-negative estimability

of a form $\Sigma \lambda_i^2 \sigma_i$ by means of the natural parameter set in a residual model. This leads to the surprising alternative that in the presence of a quadratic subspace condition either the standard unbiased estimators $\underline{s}_i^2$, $i = (1,2,\ldots,k)$, provide an unbiased non-negative definite quadratic estimator, $\Sigma \lambda_i \underline{s}_i^2$, or no such estimator exists. This is proved by Mathew (1984a).

Mathew (1984b), Peddada (1984) and Massam and Muller (1985) extend Pukelsheim's results, but the checking of conditions of non-negative estimability is not easy if there is no quadratic subspace in the natural parameter set. In the linear model

$$\underline{y} = X\beta + \underline{e}$$

with $E\underline{e} = 0$, $D(\underline{e}) = D(\underline{y}) = \sigma^2 V$, $V$ positive definite, sometimes $\beta$ belongs to the positive orthant $R_+^p$ of $R^p$. A permissible estimator $\hat{\underline{\beta}}$ of $\beta$, better than the one found by truncation (i.e. replacing negative estimates of $\beta_i$ by zero), is provided by the solution of the quadratic programming problem:

$$\text{minimize } \|y - Xb\|^2, \; b_i \geq 0, \; i = (1,2, \ldots, p)$$

or, if $V$ is known,

$$\text{minimize } \|V^{-\frac{1}{2}}y - V^{-\frac{1}{2}}Xb\|^2, \; b_i \geq 0, \; i = (1,2, \ldots, p).$$

If there are inequality constraints for some $\beta_i$, $\beta$ must belong to a convex subset of $R^p$ and we again have a quadratic or convex programming problem. See e.g. Judge and Takayama (1966), Mantel (1969), Waterman (1974), Liew (1976), and Bremner (1982).

In the case of the variance components model, we discussed in section 3 the least squares approach of Seely (or the MINQUE of Rao) to estimate the variance components $\sigma_1^2,\ldots,\sigma_k^2$. In his procedure Seely had not taken into consideration the non-negativity of $\sigma_1^2,\ldots,\sigma_k^2$. Imposing this restriction on the parameters means that we may use again the quadratic programming approach in order to assure that the estimators of $\sum_{i=1}^{k} \lambda_i\sigma_i^2$ are non-negative.

A solution of the related quadratic programming problem is found using the algorithm of Waterman (1974). Consider all possible $2^k$ subsets of $\{1, 2, \ldots, k\}$. For each such set, say J, we minimize $\|\underline{Z} - \hat{E}(\underline{Z})\|$ and obtain the unrestricted solution. More precisely we project Z orthogonally on the space generated by $\{MV_iM;\ i \in J\}$. If all the estimates of the parameters with index in J are non-negative, we retain these estimates and the corresponding norm $\|Z - \{\hat{E}(\underline{Z})$ corresponding to $J\}\|$. The retained subset J which gives the minimum of these norms solves the original problem; those parameters $\sigma_i^2$ for which $i \notin J$ are estimated by zero. These non-negative estimators are closest to the least squares estimators.

Brown (1978) also started with the residual model and also used a quadratic programming technique to find a non-negative solution. But now he uses these estimated variance components to get a generalized least squares solution for $\beta$. By this he gets an updated $M\underline{y}$ and covariance matrix of $\underline{Z} = (M\underline{y})(M\underline{y})'$. Using these $D(\underline{Z})$ for the projection on the parameter space with as inner product of 2 elements A, B $\in$ S: $tr(A[D(y)]^{-1}B)$ he stops after a certain number of iterations or after the convergence criterion has been met. Rich and Brown (1979) reported some experimental evi-

dence for the usefulness of this procedure.

For the special case with two variance components, non-negative unbiased estimators are given by Baksalary and Molińska (1984) and non-negative, admissible invariant estimators by Gnot, Kleffe and Zmyślony (1985).

Another type of a non-negative estimating procedure is given by Rao and Chaubey (1978). They minimize the norm of the invariant MINQUE without restrictions of unbiasedness, and called this estimator MINQE.

For estimating $\sum_{i=1}^{k} p_i\sigma_i^2$ their estimator is $\underline{y}'A\underline{y}$; with $\alpha_i^2$ are a priori values of $\sigma_i^2$, hence $D(\underline{y})$ is the form $T = \sum_{i=1}^{k} \alpha_i^2 V_i$ and $A = \sum_{i=1}^{k} (p_i/n_i)\alpha_i^4(I-P_T)'T^{-1}V_iT^{-1}(I-P_T)$ where $P_T = X(X'T^{-1}X)^{-1}X'T^{-1}$.

If $\underline{z} = T^{-1}(I-P_T)\underline{y}$ then the MINQUE for $\sigma_i^2$ is given by

$\hat{\sigma}_i^2 = \alpha_i^4\, \underline{z}'V_i\underline{z}/n_i$ and is non-negative.

Chaubey (1983) proposed a non-negative estimator closest to MINQUE. Let the estimator $\sigma_i^2$ (for some i) be the MINQUE $\underline{y}'A\underline{y}$, where A is not necessarily non-negative. Write A in its spectral form $A = \sum_{i=1}^{n} \lambda_i e_i e_i'$ where $\lambda_i$ are the eigenvalues of A and $e_i$ is the orthonormal eigenvector belonging to $\lambda_i$.

Skip the negative eigenvalues $\lambda_i$ (say n-r ones) and define $B'B = \sum_{j=1}^{r} \lambda_i e_i e_i'$. The euclidean norm $\|A-B'B\|$ is minimal, or otherwise stated the A matrix is closest approximated by a non-negative matrix B'B. The estimator $\underline{y}'B'B\underline{y}$ is a non-negative esti-

mator for $\sigma_i^2$ and can be seen as a truncated MINQUE, since it is obtained by taking the spectral decomposition of the matrix A. The estimator $\underline{y}'B'B\underline{y}$ remains within the class of invariant quadratic estimators but is not always unbiased any more.

For a discussion of the properties of these non-negative estimators see Rao (1977), Rao, Kaplan and Cochran (1981), Rao and Sylvestre (1984) and Kleffe (1980). The comparison of these non-negative estimators is given by Chaubey (1984).

For the balanced case Thompson (1962) and Thompson and Moore (1963) derive non-negative estimators such that linear combinations of variance components (e.g. $\sigma_1^2 \leq \sigma_1^2 + \sigma_2^2$) are order preserving. Another type of non-negative estimator for variance components has been proposed by Hartung (1981), but his method is not order preserving for estimators of linear combinations of variance components! Therefore we do not discuss this estimator.

Kleffe and Rao (1986) discussed the existence of asymptotically unbiased non-negative estimators. For finite n we always have a positive probability to get negative estimates. For infinite n this probability is zero.

Simple conditions can be found from the expressions of the estimators in several cases. Secondly, adjustments to non-negativeness result in the bias for the unbiased estimators like the ANOVA and the MINQUE. The variances of these adjusted biased estimators may not become small unless certain conditions are satisfied. Rao (1987) examined for two models the conditions for the biased and unbiased estimators of the variance components to become small in MSE.

## 5. MAXIMUM LIKELIHOOD ESTIMATORS

Another estimator which is always permissible is the proper maximum likelihood estimator (MLE), which yields a point in the

parameter set $\theta$ at which the likelihood function attains an absolute maximum. It should be noted that in many situations the estimators obtained by merely solving the likelihood equations, i.e. setting first derivatives of the likelihood equal to zero, will not be permissible. Solving likelihood equations while taking into account the restrictions imposed by the parameter set $\theta$ will give the proper MLE.

For the variance components model with multivariate normally distributed $\underline{y}$, Hartley and Rao (1967) give an iterative procedure for solving the likelihood equations under the restriction of non-negative variance components.

The restricted maximum likelihood (REML) or modified maximum likelihood approach of Patterson and Thompson (1971, 1975) maximizes the likelihood, not of the complete vector $\underline{y}$, but of all error contrasts.

For a discussion of maximum likelihood approaches to variance component estimation see Harville (1977). J.N.K. Rao warned in his comment on Harville's article against the use of ML-procedures, because none of the proposed algorithms in that article guarantees a solution which is indeed ML.

The behaviour of the likelihood as a function of the variance components appears to be complex; even for the simple unbalanced one-way lay-out, the likelihood equation may have multiple roots, or the ML-estimate may be a boundary point rather than a root.

A procedure which guarantees a non-negative solution is the iterative process of Henderson (1984) using the mixed model equations to get a REML solution. The comparison of ANOVA, MINQUE, REML and ML estimators for variance components in the unbalanced one-way classification has been done by Swallow and Monahan (1984). Conerly and Webster (1987) extended the comparison between the MSE of the MINQE and the estimators considered by Swallow and Monahan (1984).

ACKNOWLEDGEMENTS

The author would like to express his gratitude to mr. A. Otten for his helpful remarks, and to the referee for many improvements of this article.

BIBLIOGRAPHY

Anderson, R.D., Henderson, H.V., Pukelsheim, F. and Searle, S.R. (1984). Best estimation of variance components from balanced data, with arbitrary kurtosis. Math. Operationsforsch. Statist., Series Statistics. 15, 163-176.

Baksalary, J.K. and Molińska, A. (1984). Nonnegative unbiased estimability of linear combinations of two variance components. J. Statist. Planning and Inference 10, 1-8.

Bremner, J.M. (1982). An algorithm for non-negative least squares and projections onto cones. Compstat 1982, ed. H. Caussinus, P. Ettinger, R. Tomassone. Vienna: Physica-Verlag, 155-160.

Brown, K.G. (1978). Estimation of variance components using residuals. J. Amer. Statist. Assoc. 73, 141-146.

Chaubey, Y.P. (1983). A non-negative estimator of variance component closest to MINQUE. Sankhyā A 45, 201-211.

Chaubey, Y.P. (1984). On the comparison of some non-negative estimators of variance components for two models. Commun. Statist. - Simula. Computa. 13, 619-633.

Chaubey, Y.P. and Rao, P.S.R.S. (1976). Efficiencies of five estimators for the parameters of two linear models with unequal variances. Sankhyā B 38, 364-370.

Conerly, M.D. and Webster, J.T. (1987). MINQE for the one-way classification. Technometrics 29, 229-236.

Drygas, H. (1977). Best quadratic unbiased estimation in variance-covariance component models. Math. Operationsforsch. Statist., Series Statistics. 8, 211-231.

Gnot, S., Kleffe, J. and Zmyślony, R. (1985). Nonnegativity of admissible invariant quadratic estimates in mixed linear models with two variance components. J. Statist. Planning and Inference 12, 249-258.

Hartley, H.O. and Rao, J.N.K. (1967). Maximum likelihood estimation for the mixed analysis of variance model. Biometrika 54, 93-108.

Hartley, H.O. and Jayatillake, K.S.E. (1973). Estimation for linear models with unequal variances. J. Amer. Statist. Assoc. 68, 189-192.

Hartung, J. (1981). Nonnegative minimum biased invariant estimation in variance component models. Ann. Statist. 9, 278-292.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems, (with comment by J.N.K. Rao). J. Amer. Statist. Ass. 72, 320-340.

Henderson, C.R. (1984). Applications of linear models in animal breeding. Guelph: University of Guelph.

Judge, G.G. and Takayama, T. (1966). Inequality restrictions in regression analysis. J. Amer. Stat. Assoc. 61, 166-181.

Khuri, A.I. and Sahai, H. (1985). Variance components analysis: A selective literature survey. International Statistical Review 53, 279-300.

Kleffe, J. (1980). On recent progress of MINQUE theory - Nonnegative estimation, consistency, asymptotic normality and explicit formulae. Math. Operationsforsch. Statist., Series Statistics. 11, 563-588.

Kleffe, J. (1984). On variance components estimation in heteroscedastic models, Computational Statistics Quaterly 1, 179-204.

Kleffe, J. and Rao, J.N.K. (1986). The existence of asymptotically unbiased nonnegative quadratic estimates of variance components in ANOVA models. J. Amer. Stat. Ass. 81, 692-698.

LaMotte, L.R. (1973). On non-negative quadratic unbiased estimation of variance components. J. Amer. Statist. Assoc.68, 728-730.

Liew, C.K. (1976). Inequality constrained least-squares estimation. J. Amer. Statist. Assoc. 71, 746-751.

Mantel, N. (1969). Restricted least squares regression and convex quadratic programming. Technometrics 11, 763-773.

Massam, H. and Muller, J. (1985). The nonnegative MINQUE estimate. J. Multivar. Anal. 16, 253-259.

Mathew, T. (1984a). On nonnegative quadratic unbiased estimability of variance components. Ann. Statist. 12, 1566-1569.

Mathew, T. (1984b). On the characterization of nonnegatively estimable linear combinations of variance components. Technical report no. 85-06, Center for Multivariate Analysis. Pittsburgh: University of Pittsburgh.

Patterson, H.D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika 58, 545-554.

Patterson, H.D. and Thompson, R. (1975). Maximum likelihood estimation of components of variance. Proceedings of the 8th International Biometric Conference, Constanta, ed. Corsten, L.C.A. and Postelnicu, T., 197-207.

Peddada, S.D. (1984). Some results on non-negative estimation of variance components. Technical Report no. 84-12. Center for Multivariate Analysis. Pittsburgh: University of Pittsburgh.

Pukelsheim, F. (1976). Estimating variance components in linear models. J. Multivar. Anal. 6, 626-629.

Pukelsheim, F. (1977). On Hsu's model in regression analysis. Math. Operationsforsch. Statist., Series Statistics 8, 323-331.

Pukelsheim, F. (1977). Punktschätzungen in linearen statistischen Modellen. Dissertation zur Erlangung des Doktorgrades der Mathematischen Fakultät der Albert-Ludwigs-Universität. Freiburg im Breisgau.

Pukelsheim, F. (1979). Classes of linear models. Proceedings of a conference in honor of C.R. Henderson, July 16-17, 1979. Variance components and animal breeding. ed. Van Vleck, L.D. and Searle, S.R.; Ithaca, N.Y.: Cornell University.

Pukelsheim, F. (1981a). On the existence of unbiased nonnegative estimates of variance covariance components. Ann. Statist. 9, 293-299.

Pukelsheim, F. (1981b). Linear models and convex geometry: Aspects of non-negative variance estimation. Math. Operationsforsch. Statist., Series Statistics 12, 271-286.

Rao, C.R. (1970). Estimation of heteroscedastic variances in linear models. J. Amer. Statist. Ass. 65, 161-172.

Rao, C.R. (1971). Estimation of variance and covariance components-MINQUE theory. J. Multivar. Anal. 1, 257-275.

Rao, C.R. (1972). Estimation of variance and covariance components in linear models. J. Amer. Statist. Assoc. 67, 112-115.

Rao, C.R. (1973). Linear statistical inference and its applications; second edition. New York: John Wiley & Sons.

Rao, J.N.K. and Subrahmaniam, K. (1971). Combining independent estimators and estimation in linear regression with unequal variances. Biometrics 27, 971-990.

Rao, P.S.R.S. (1977). Theory of the MINQUE - a review. Sankhya B 39, 201-210.

Rao, P.S.R.S. (1987). Estimation of variance components; small and large properties. To appear in Commun. Statist.-Theor. Meth. A 16, ?.

Rao, P.S.R.S. and Chaubey, Y.P. (1978). Three modifications of the principle of the MINQUE. Commun. Statist.-Theor. Meth. A7, 767-778.

Rao, P.S.R.S., Kaplan, J. and Cochran, W.G. (1981). Estimators for the one-way random effects model with unequal error variances. J. Amer. Statist. Assoc. 76, 89-97.

Rao, P.S.R.S. and Sylvestre, E.A. (1984). ANOVA and MINQUE type of estimators for the one-way random effects model. Commun. Statist.-Theor. Meth. A13, 1667-1673.

Rich, D.K. and Brown, K.G. (1979). Estimation of variance components using residuals: some empirical evidence. Proceedings of a conference in honor C.R. Henderson, July 16-17; 1979. Variance components and animal breeding. ed. Van Vleck, L.D. and Searle, S.R.; Ithaca, N.Y.: Cornell University, 139-154.

Seely, J. (1970a). Linear spaces and unbiased estimation. Ann. Math. Statist. 41, 1725-1734.

Seely, J. (1970b). Linear spaces and unbiased estimation-application to the mixed linear model. Ann. Math. Statist. 41, 1735-1748.

Seely, J. (1971). Quadratic subspaces and completeness. Ann Math. Statist. 42, 710-721.

Seely, J. (1972). Completeness for a family of multivariate normal distributions. Ann. Math. Statist. 43, 1644-1647.

Seely, J. (1977). Minimal sufficient statistics and completeness for multivariate normal families. Sankyā A. 39, 170-185.

Seely, J. and Zyskind, G. (1971). Linear spaces and minimum variance unbiased estimation. Ann. Math. Statist. 42, 691-703.

Swallow, W.H. and Searle, S.R. (1978). Minimum variance quadratic unbiased estimation (MIVQUE) of variance components. Technometrics 20, 265-272.

Swallow, W.H. and Monahan, F. (1984). Monte Carlo comparison of ANOVA, MIVQUE, REML, and ML estimators of variance components. Technometrics 26, 47-57.

Thompson, W.A. (1962). The problem of negative estimates of variance components. Ann. Math. Statist. 33, 273-289.

Thompson, W.A. and Moore, J.R. (1963). Non-negative estimates of variance components. Technometrics 5, 441-449.

Verdooren, L.R. (1980). On estimation of variance components. Statistica Neerlandica 34, 83-106.

Verdooren, L.R. (1982). How large is the probability for the estimate of a variance component to be negative? Biometrical Journal 24, 339-360.

Waterman, M.S. (1974). A restricted least squares problem. Technometrics 16, 135-136.

8.3. Which estimation procedure to be used?

For the variance components estimators we gave, in section 5.2 subsection 2.1, the mixed model

$$\underline{y} = X\beta + \sum_{i=1}^{k} U_i \underline{a}_i \qquad (8.3.1)$$

where $\underline{a}_k$ are the error effects with $D(\underline{a}_k) = \sigma_k^2 V_k$ with $V_k = I_n$.
Let v stand for the column vector of variance components:

$$v' = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2) = \sigma_k^2(\gamma_1, \gamma_2, \ldots, 1) \qquad (8.3.2)$$

To estimate v, at the present time four main options are available: ANOVA, Least Squares or MINQUE, ML and REML.

For balanced designs ANOVA and MINQUE are the same. For unbalanced designs the absence of an optimality criterion for ANOVA estimators is a serious deficiency. With more than two variance components, Henderson's Method III estimators give no unique estimator either. Therefore our recommendation is to abandon the ANOVA method of estimating variance components for unbalanced data.

MINQUE, ML and REML are all to be preferred over ANOVA because they have built-in optimality properties. There are also computational connections between these three estimators.

Let us first consider the invariant MINQUE as described in section 5.2, subsection 4, (§ 5.2.4). The dispersion matrix of $\underline{y}$ is:

$$D(\underline{y}) = V = \sum_{i=1}^{k} \sigma_i^2 V_i \qquad (8.3.3)$$

where $V_i = U_i U_i'$. If the a priori (or approximate) values of $\sigma_i^2$ are $\alpha_i^2$, then the variance components are estimated by calculating with the approximate dispersion matrix

$$T = \sum_{i=1}^{k} \alpha_i^2 V_i .$$

As described in § 5.2.4 the invariant unbiased estimator for

$\sum_{i=1}^{k} \lambda_i \sigma_i^2 = \lambda' v$, given $T = \sum_{i=1}^{k} \alpha_i^2 V_i$, is

$$\sum_{i=1}^{k} \rho_i \underline{y}'(I-P_T')T^{-1}V_iT^{-1}(I-P_T)\underline{y}$$

in which $\rho_1, \ldots, \rho_k$ satisfy

$$\sum_{i=1}^{k} \rho_i \mathrm{tr}[(I-P_T')T^{-1} V_iT^{-1}(I-P_T)V_j] = \lambda_j$$

for $j = 1, 2, \ldots, k$, and in which $P_T = X(X'T^{-1}X)^{-1}X'T^{-1}$.

We use the notation

$\{a_{ij}\}_{i,j=1}^{i,j=k}$ for a square matrix of order k and $\{b_i\}_{i=1}^{i=k}$ for a column vector of order k. Let $\omega$ be the column vector of a priori values $\alpha_i^2$, which are used in place of $\sigma_i^2$, $\omega = (\alpha_1^2, \alpha_2^2, \ldots, \alpha_k^2)'$.

Let $P_\omega = T^{-1}(I-P_T)$, then $P_\omega = P_\omega'$.

Let $\rho = (\rho_1, \rho_2, \ldots, \rho_k)'$, then the estimator of $\lambda' v$ is

$$\lambda'\underline{\hat{v}} = \rho'\{ \underline{y}' P_\omega V_i P_\omega \underline{y}\}_{i=1}^{i=k} \qquad (8.3.4)$$

and $\lambda'\underline{\hat{v}}$ exists if

$$\{\mathrm{tr}(P_\omega V_i P_\omega V_j)\}_{i,j=1}^{i,j=k} = \lambda . \qquad (8.3.5)$$

Inserting (8.3.5) into (8.3.4) means that the estimation of the variance components vector v must satisfy the equations in matrix form

$$\{\mathrm{tr}(P_\omega V_i P_\omega V_j)\}_{i,j=1}^{i,j=k} \underline{\hat{v}} = \{\underline{y}' P_\omega V_i P_\omega \underline{y}\}_{i=1}^{i=k}$$

or, using the property of the trace operator tr(ABC) = tr(BCA) and

$V_i = U_i U_i'$,

$$\{tr(U_i U_i' P_\omega U_j U_j' P_\omega)\}_{i,j=1}^{i,j=k} \underline{\theta} = \{\underline{y}' P_\omega U_i U_i' P_\omega \underline{y}\}_{i=1}^{i=k} . \qquad (8.3.6)$$

A very natural extension is to use as a new $\omega$ the outcomes $\hat{\theta}$ from (8.3.6) and solve the equation (8.3.6) with this new $\omega$.

One iterates till the solution is sufficiently accurate. This is the Iterative MINQUE or I-MINQUE for the variance components.

The ML estimator for the variance components is derived in section 5.2, subsection 6, (§ 5.2.6). This estimator can only be used if we assume that the observations $\underline{y}$ have a multivariate normal distribution. This restriction is not needed for the MINQUE or I-MINQUE.

The dispersion matrix $D(\underline{y}) = V$ of (8.3.4) is also described as

$$D(\underline{y}) = \sigma_k^2 H = \sigma_k^2 \sum_{i=1}^{k} \gamma_i V_i \qquad (8.3.7)$$

with $\gamma_k = 1$.

The ML estimators follow after iterating as the solutions of

$$\hat{\beta} = (X'\hat{H}^{-1}X)^{-1}(X'\hat{H}^{-1}y) \qquad (8.3.8)$$

and

$$\hat{\sigma}_k^2 = y'\hat{R}'\hat{H}^{-1}\hat{R}\, y/n \qquad (8.3.9)$$

where $\hat{R} = I - X(X'\hat{H}^{-1}X)^{-1}X'\hat{H}^{-1}$,

and

$$tr(\hat{H}^{-1}V_i) = y'\, \hat{R}'\, \hat{H}^{-1}\, V_i\, \hat{H}^{-1}\, \hat{R}\, y/\hat{\sigma}_k^2 \qquad (8.3.10)$$

for $i = 1,2, \ldots, k-1$.

Let us define

$$\hat{P} = \hat{H}^{-1}\, \hat{R}/\hat{\sigma}_k^2 = \hat{V}^{-1} - \hat{V}^{-1}X(X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}. \qquad (8.3.11)$$

Because $\mathrm{tr}(\hat{H}^{-1}\hat{V}_i) = \mathrm{tr}(\hat{H}^{-1}\hat{V}_i\hat{H}^{-1}\hat{H}) = \mathrm{tr}(\hat{H}^{-1}U_iU_i'\hat{H}^{-1}\sum_{j=1}^{k}\hat{\gamma}_jU_jU_j') =$
$\sum_{j=1}^{k}\mathrm{tr}(\hat{H}^{-1}U_iU_i'\hat{H}^{-1}U_jU_j')\hat{\sigma}_j^2/\hat{\sigma}_k^2$,

we can write equations (8.3.9) and (8.3.10) in matrix form, using

$\hat{V}^{-1} = \hat{H}^{-1}/\hat{\sigma}_k^2$, as:

$$\{\mathrm{tr}(U_iU_i'\hat{V}^{-1}U_jU_j'\hat{V}^{-1})\}_{i,j=1}^{i,j=k}\ \hat{\underline{v}} = \{\underline{y}'\hat{P}U_iU_i'\hat{P}\underline{y}\}_{i=1}^{i=k}\ . \qquad (8.3.12)$$

The equations (8.3.8) and (8.3.12) are iteratively solved for $\hat{v}$. Notice that for each iteration cycle $\hat{H}$ and $\hat{P}$ must be adapted. A ML estimator for the variance components is given by the solution of the set of equations (8.3.12) with the restriction that all $\hat{\sigma}_i^2 \geqslant 0$.

The REML estimator for the variance components is derived in section 5.2, subsection 6, (§ 5.2.6). This estimator is found as the solution of $\tilde{\sigma}_k^2 = y'T'(T\tilde{H}T')^{-1}Ty/(n-p)$ (8.3.13)

and

$$\mathrm{tr}(U_i'T'(T\tilde{H}T')^{-1}TU_i) = y'T'(T\tilde{H}T')^{-1}TU_iU_i'T'(T\tilde{H}T')^{-1}Ty/\tilde{\sigma}_k^2 \qquad (8.3.14)$$

for i = 1, 2, ..., k-1,

where T is an (n-p)xn matrix of which the rows are any n-p linear independent rows of $S = I-X(X'X)^{-1}X'$, and $TX = 0$. Because S can be written as QQ', where Q is a nx(n-p) matrix with $Q'Q = I_{n-p}$, it follows that $T' = QF$, where F is a non-singular matrix.

Now $T'(T\tilde{H}T')^{-1}T = Q\ F(F'Q'\tilde{H}QF)^{-1}\ F'Q' = QF[F'(Q'\tilde{H}Q)F]^{-1}\ F'Q' =$
$QFF^{-1}(Q'\tilde{H}Q)^{-1}(F')^{-1}\ F'Q' = Q(Q'\tilde{H}Q)^{-1}Q'$.

Furthermore the dispersion matrix of $S\underline{y}$ is $\sigma_k^2$ SHS, where SHS is singular.

Let the Moore-Penrose form of a generalised inverse of an non-null matrix B be $B^+$, so that $B^+$ is the unique solution of

(i) $B\,B^{+}\,B = B$ ,

(ii) $B^{+}\,B\,B^{+} = B^{+}$ ,

(iii) $B\,B^{+} = (B\,B^{+})'$ ,

(iv) $B^{+}\,B = (B^{+}\,B)'$.

It can be shown that the Moore-Penrose inverse of $S\tilde{H}S$ is $(S\tilde{H}S)^{+} =$

$Q(Q'\tilde{H}Q)^{-1}Q'$, but also $(S\tilde{H}S)^{+} = \tilde{H}^{-1}-\tilde{H}^{-1}X(X'\tilde{H}^{-1}X)^{-1}X'\tilde{H}^{-1} = \tilde{P}\tilde{\sigma}_k^2$.

We have now proved that $T'(T\tilde{H}T')^{-1}\,T = \tilde{P}\tilde{\sigma}_k^2$, and the equations (8.3.13) and (8.3.14) can be written in matrix-form, using $\tilde{V}^{-1} = \tilde{H}^{-1}/\tilde{\sigma}_k^2$, as:

$$\{tr(U_iU_i'\tilde{P}U_jU_j'\tilde{P})\}_{i,j=1}^{i,j=k}\ \underline{\tilde{v}} = \{\underline{y}'\tilde{P}U_iU_i'\tilde{P}\underline{y}\}_{i=1}^{i=k} \qquad (8.3.15)$$

and solved by iteration for $\tilde{v}$.

A REML estimator for the variance components is given by the solution of the set of equations (8.3.15) with the restriction that all $\tilde{\sigma}_i^2 \geq 0$.

We now see that the REML equations (8.3.15) differ from the ML equations (8.3.12) only in having the $\tilde{P}$-matrix where the ML equations have the $V^{-1}$-matrix in the left-hand side.

Further the MINQUE equations (8.3.6) are exactly the same as the REML equations (8.3.15), except that the $\tilde{P}$-matrix in REML is replaced by $P_\omega$ for MINQUE. Thus a MINQUE = a first iterate of REML. This was first observed by Hocking and Kutner (1975). If the I-MINQUE gives positive estimates for the variance components it is the same as the REML. Only for normally distributed data do we have a REML; if there is no good foundation for the normality assumption, then using the REML procedure and calling the estimate I-MINQUE would be acceptable . Brown (1976) has shown that I-MINQUE has a limiting distribution that is normal.

But a MINQUE is not a permissible estimator; MINQUE can produce negative estimates. Therefore a non-negative adapted MINQUE must be used. To get rid of the pre-assigned vector $\omega$ of a priori values of

variance components, a non-negative I-MINQUE would be a preferable estimator.

If the data are multivariate normally distributed, the choice is between ML or REML. Both estimators can only be found by iterativatily solving the equations (8.3.12) for ML or (8.3.15) for REML, and this raises a number of problems that are in the realm of numerical analysis. Searle (1987) posed the following questions.
Does the choice of a starting value affect the attained value at convergence? Does that attained value always correspond to a global maximum of the likelihood that is being maximized, or does it sometimes correspond to a local maximum?

Since at each successive round of the iteration a numerical matrix is being used for V, how does one ensure that it is always positive-definite? What are the consequences if an updated V is not positive-definite? If, after some iteration, the numerical value $\hat{\sigma}_j^2$ to be given to $\sigma_j^2$ is negative, what action is to be taken? Were that negative $\hat{\sigma}_j^2$ the result from the last round of iteration then it would, according to the maximum likelihood principles, be changed to zero. The model would then be altered correspondingly, and the remaining variance components re-estimated. But suppose $\hat{\sigma}_j^2 < 0$ occurs before convergence; and suppose it is changed to zero and the model altered, and iteration continues using that altered model. Is this a good procedure if, as a result of some numerical peculiarity of those data, continuing with that unchanged negative $\hat{\sigma}_j^2$ would have, at a subsequent round of iteration, led to a positive $\hat{\sigma}_j^2$ ? Perhaps changing $\hat{\sigma}_j^2$ to zero and altering the model is the wrong thing to do.

Up to now no available computing package takes care of these difficulties. The research worker believes uncritically in the result given by the computer package. A test with different starting values is usually not made, because it is time consuming. However, if one gets the

same solution some confidence in this result would seem to be justified.

So far we have sketched the difficulties and dangers inherent in using ML or REML procedures, but we have not answered the difficult question: ML or REML? One favoured characteristic of REML is that with balanced data the REML equations reduce to the same equations as are used by ANOVA estimators. We have seen in section 2.3 that ANOVA estimators for balanced designs have the attractive minimum variance property. But ANOVA estimators are not permissible, they can lead to negative estimates. The solutions of the REML equations (8.3.15) are not necessarily REML estimators; they only are if they are positive!

Also REML gives no unbiased estimators even for balanced designs. It is true that the expected value of the right-hand side of the REML equations in (8.3.15) can be written in the same form as the left-hand side of those equations, but this does not imply unbiasedness. The non-negativity constraint of REML has to be taken into account. For unbalanced designs, after iteration, neither ML nor REML is unbiased.

One of the merits of ML over REML is that the ML procedure includes providing an ML estimator for the fixed effects, namely $\hat{E}(\underline{y}) = X(X'\hat{V}^{-1}X)^{-1} X' \hat{V}^{-1}\underline{y}$, where the ML solutions for the variance components are inserted into $\hat{V}$. Kackar and Harville (1981, 1984) have shown that this ML estimator for $E(\underline{y})$ is unbiased, and that its sampling variance can be calculated, and hence we asymptotically have normality.

The REML method provides no such estimator, although intuitively one would be inclined to use $X(X'\tilde{V}^{-1}X)^{-1} X'\tilde{V}^{-1}\underline{y}$ as an estimator for $E(\underline{y})$ as well, where the REML solutions for the variance components are inserted into $\tilde{V}$.

It is difficult to be anyting but inconclusive about which of ML

and REML is the preferred method. ML has the merit of simultaneously providing estimators for both the fixed effects and the variance components, and that is appealing. On the other hand, REML has the attraction of providing variance components estimators that are unaffected by the fixed effects. As already discussed before, for data where we do not want to assume normality, a REML procedure as given in most computer packages, gives I-MINQUE results. One must be aware, however, that, in the case of negative results, one should repeat the calculations searching along the boundary of the parameter space (see section 8.2, subsection 4).

## REFERENCES


Brown, K.G. (1976). Asymptotic behaviour of MINQUE-type estimators of variance components. Ann. of Statist. 4, 746 - 754.

Hocking, R.R. and Kutner, M.H. (1975). Some analytical and numerical comparisons of estimators for the mixed A.O.V. model. Biometrics 31, 19 - 28.

Kackar, R.N. and Harville, D.A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. Commun. Statist. Theor.- Meth. A 10, 1249 - 1261.

Kackar, R.N. and Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. Journ. Amer. Statist. Assoc. 79, 853 - 861.

Searle, S.R. (1987). Mixed models and unbalanced data: Wherefrom, whereat and whereto? To appear in Commun. Statist. Theor.- Meth. A 16, ?.

CHAPTER 9

# 9. PREDICTION OF RANDOM EFFECTS

## 9.1. Introduction

In the mixed model the estimation of fixed effects and prediction of the random effects are two other aspects that demand attention besides the estimation of variance components. If the dispersion matrix $D(\underline{y}) = V$ is known, the generalized least squares estimator of $E(\underline{y}) = X\beta$ is $X(X'V^{-1}X)^{-1}X'V^{-1}\underline{y}$.

At the end of section 8.3 we have discussed the estimation of $E(\underline{y})$. We now proceed to discuss the prediction of random effects. Any random effect that occurs in the data is not actually a random variable, but is a realization of a random variable. Nevertheless, it is usually unobservable, e.g. the genetic make-up of the dairy cow, whose annual milk yield has been recorded. Whilst we cannot measure these realizations we can think of predicting random effects. This has many applications in practice. The most far-reaching, on a world-wide scale, is in dairy cow breeding, where bulls used in artificial insemination programs each have many offspring, half of whom have milk production records. These records can be used to calculate predictions of the bull's genetic merits. This leads to a ranking of bulls, and only the best few get continuing use in artificial insemination. It is this procedure that has, over the last 30 years, drastically increased the per-cow milk production in many countries around the world. The same principles have also been used in other contexts, such as meat, wool and egg production. In section 9.2 an application in variety testing of field crops is described.

Let us consider again the mixed model, which is described in section 5.2, subsection 2.1:

$$\underline{y} = X\beta + \sum_{i=1}^{k} U_i \underline{a}_i \qquad (9.1.1.)$$

where $\underline{a}_k$ are the error effects, with $D(\underline{a}_k) = \sigma_k^2 V_k$ with $V_k = I_n$.

Let $\underline{U} = (\underline{a}_1', \underline{a}_2', \ldots, \underline{a}_k')'$ then the dispersion matrix $D(\underline{a})$ is $D = \text{diag}(\sigma_1^2 I_{m_1}, \ldots, \sigma_k^2 I_{m_k})$, a block diagonal matrix, with the blocks being the matrices $\sigma_i^2 I_{m_i}$ for $i = 1, 2, \ldots, k$.

Furthermore, the dispersion matrix

$$D(\underline{y}) = \sum_{i=1}^{k} U_i U_i' \sigma_i^2 = V \qquad (9.1.2.)$$

and the covariance matrix of $\underline{y}$ and $\underline{a}$ is

$$C = \text{cov}(\underline{y}, \underline{a}') = \sum_{i=1}^{k} U_i \sigma_i^2.$$

So the feature of interest is prediction of $\underline{a}$, where $\underline{y}$ and $\underline{a}$ are jointly distributed:

$$\begin{bmatrix} \underline{y} \\ \underline{a} \end{bmatrix} \quad \text{with } E \begin{bmatrix} \underline{y} \\ \underline{a} \end{bmatrix} = \begin{bmatrix} X\beta \\ 0 \end{bmatrix}$$

and disperion matrix $D\begin{bmatrix} \underline{y} \\ \underline{a} \end{bmatrix} = \begin{bmatrix} V & C \\ C' & D \end{bmatrix}$. (9.1.3.)

Cochran (1951) or Rao (1973, section 4g.1) describe the best (i.e. minimum mean square) predictor for $\underline{a}$ as

$$\tilde{\underline{a}} = E(\underline{a} \mid \underline{y}). \qquad (9.1.4.)$$

The best linear predictor, i.e. linear in elements of $\underline{y}$, is

$$\tilde{\underline{a}}_L = E(\underline{a}) + C' V^{-1} (\underline{y} - X\beta). \qquad (9.1.5.)$$

These results $\tilde{\underline{a}}$ and $\tilde{\underline{a}}_L$ hold for any distribution (satisfying the usual regularity conditions) having finite first and second moments. Furthermore, when that distribution is the multivariate normal distribution, (9.1.4.) and (9.1.5.) give the same result, i.e. under normality

$$\tilde{\underline{a}} = \tilde{\underline{a}}_L = E(\underline{a} \mid \underline{y}) = E(\underline{a}) + C'V^{-1}(\underline{y} - X\beta). \qquad (9.1.6.)$$

An extension of (9.1.5.) is given by Henderson (1963) for the mixed model. He shows that the best linear predictor of

$$\underline{w} = T'X\beta + \underline{a}$$

is

$$\tilde{\underline{w}} = T'X\hat{\beta} + E(\underline{a}) + C'V^{-1}(\underline{y} - X\hat{\beta}) \qquad (9.1.7.)$$

where

$$X\hat{\beta} = X(X'V^{-1}X)X'V^{-1}\underline{y}.$$

This is what animal breeders refer to as Best Linear Unbiased Prediction (BLUP). Again, it requires values of $\sigma_i^2$ for practical application. A comprehensive compilation of the theory for BLUP and its application in animal breeding can be found in Henderson (1984).

## REFERENCES


Cochran, W.G. (1951). Improvement by means of selection. Proceedings second Berkeley Symposium, Math. Statist. and Prob., 449-470.

Henderson, C.R.(1963). Selection index and expected genetic advance. In: Statistical Genetics and Plant Breeding. Eds. Hanson, W.D. and Robinson, H.F., National Academy of Sciences, National Research Council, Washington. Publication 982, 141-163.

Henderson, C.R. (1984). Applications of linear models in animal breeding. University of Guelph.

Rao, C.R. (1973). Linear statistical inference and its applications, second edition. John Wiley & Sons, New York.

## 9.2. The best estimation of varietal contrasts according to the model from combined varietal trials

Paper presented at the second International Seminar on "Statistical Methods in Variety Testing" at Slupia Wielka, Poland, in 1985.

Published in:

Biuletyn Oceny Odmian 12(1987) no. 1-2, pp 135-149.

BIULETYN OCENY ODMIAN
Tom XII Zeszyt 1-2 (17-18) 1987

L. R. VERDOOREN

Department of Mathematics, Agricultural University,
Wageningen, The Netherlands

# THE BEST ESTIMATION OF VARIETAL CONTRASTS ACCORDING TO THE MODEL FROM COMBINED VARIETAL TRIALS

## 1. INTRODUCTION

One of the main objects of varietal trials is to estimate the differences in yielding capacity of varieties. The assumptions to derive a good estimation procedure are relatively simple and the observations do not need to be normally distributed. Although the principles of the best linear unbiased estimators (BLUE) for varietal contrasts are described in several mathematical statistical handbooks (e.g. Rao (1973), Searle (1971)), there seems to be no extensive use in the field of variety research. Our colleagues of animal breeding make much more use of these tools. Also the method of best linear unbiased prediction (BLUP) of varietal contrasts has not been used in variety research.

In section 2 we give a review of the best linear unbiased estimator in linear models, demonstrated on a little numerical example. In section 3 the linear model of combined varietal trials is described. According to this model the BLUE can be applied. In section 4 the method of best linear prediction is described. Often in varietal research the prediction of varietal contrasts is the relevant question and not the estimation.

## 2. BEST LINEAR UNBIASED ESTIMATORS (BLUE)

### 2.1. THE MODEL $(\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{I}_n)$

Let us start with the following simple linear model. Consider uncorrelated observations* $(\underline{y}_1, \ldots, \underline{y}_n)$ such that

$$\left.\begin{aligned} E(\underline{y}_i) &= x_{i1}\beta_1 + \ldots + x_{im}\beta_m \\ V(\underline{y}_i) &= \sigma^2 \end{aligned}\right\} \quad i = 1, 2, \ldots, n, \tag{1}$$

where $(\beta_1, \ldots, \beta_m)$ and $\sigma^2$ are unknown parameters and $(x_{ij}) = \mathbf{X}$ is a $n \times m$-matrix of known coefficients with $n > m$. If $\underline{\mathbf{y}} = (\underline{y}_1, \ldots, \underline{y}_n)'$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)'$ stand for column vectors of the variables $\underline{y}_i$ and the parameters $\beta_j$, this fixed effects model (1) can be written in

* Random variables and vectors (r. v.) are denoted by a letter which is underscored; outcomes of r. v. are denoted by the same letter which is not underscored.

matrix notation:

$$\underline{\mathbf{y}}=\mathbf{X}\boldsymbol{\beta}+\underline{\mathbf{e}}, \quad E(\underline{\mathbf{e}})=\mathbf{0}, \quad D(\underline{\mathbf{e}})=\sigma^2\mathbf{I}_n \tag{2}$$

where $D$ stands for the dispersion-matrix (matrix of variances and covariances) and $\mathbf{I}_n$ for the unit matrix of order $n$; hence $E(\underline{\mathbf{y}})=\mathbf{X}\boldsymbol{\beta}$ and $D(\underline{\mathbf{y}})=\sigma^2\mathbf{I}_n$. We denote this linear model as $(\underline{\mathbf{y}}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_n)$. The problem is that of estimating the unknown parameters $\beta_j$ on the basis of the observations $\underline{\mathbf{y}}$. Using the method of least squares for estimating $\beta_j$ we must minimize

$$\sum_{i=1}^{n}(\underline{y}_i-x_{i1}\underline{b}_1-\ldots-x_{im}\underline{b}_m)^2=(\underline{\mathbf{y}}-\mathbf{X}\underline{\mathbf{b}})'(\underline{\mathbf{y}}-\mathbf{X}\underline{\mathbf{b}}),$$

this is the ordinary least squares (OLS) method and the solution is found from the so-called normal equations

$$\mathbf{X}'\mathbf{X}\underline{\mathbf{b}}=\mathbf{X}'\underline{\mathbf{y}}. \tag{3}$$

Often the linear model is defined with too much parameters $\boldsymbol{\beta}$, which is called over-parametrization. This means that the normal equations (3) may have more solutions $\underline{\mathbf{b}}$. An unbiased estimator as $\mathbf{p}'\underline{\mathbf{b}}$ for $\mathbf{p}'\boldsymbol{\beta}=p_1\beta_1+\ldots+p_m\beta_m$ needs not to be unique. The concept of an identifiable linear function of parameters $\mathbf{p}'\boldsymbol{\beta}$ is now needed. Now $E(\underline{\mathbf{y}})=\mathbf{X}\boldsymbol{\beta}=\mathbf{X}\boldsymbol{\beta}^*$ implies $\mathbf{p}'\boldsymbol{\beta}=\mathbf{p}'\boldsymbol{\beta}^*$. Therefore $\mathbf{p}'\boldsymbol{\beta}$ is identifiable if $\mathbf{p}'\underline{\mathbf{b}}$ is unbiased and is unique for all solutions $\underline{\mathbf{b}}$ of the normal equations (3) or equivalently $\mathbf{p}\in C(\mathbf{X}') \Leftrightarrow \mathbf{p}\in C(\mathbf{X}'\mathbf{X})$ or, equivalently there exists a linear function of $\mathbf{y}$ with expectation $\mathbf{p}'\boldsymbol{\beta}$, [$C(\mathbf{A})$ is the column-space of the matrix $\mathbf{A}$].

Note that it is not sufficient to solve (3) and if for two solutions $\underline{\mathbf{b}}_1$ and $\underline{\mathbf{b}}_2$, $\mathbf{p}'\underline{\mathbf{b}}_1=\mathbf{p}'\underline{\mathbf{b}}_2$, to say that $\mathbf{p}'\boldsymbol{\beta}$ is identifiable (which is sometimes done in practice) because it must be true for all solutions of (3). From the uniqueness of $\mathbf{p}'\underline{\mathbf{b}}$ for all solutions of the normal equations (3) in most books $\mathbf{p}'\boldsymbol{\beta}$ is called an estimable linear function of the parameters instead of an identifiable linear function. The Gauss-Markov theorem is now as follows: If $\mathbf{p}'\boldsymbol{\beta}$ is estimable, then $\mathbf{p}'\underline{\mathbf{b}}$ has minimum variance in the class of linear unbiased estimators of $\mathbf{p}'\boldsymbol{\beta}$. The estimator $\mathbf{p}'\underline{\mathbf{b}}$ is called the best linear unbiased estimator (BLUE) for $\mathbf{p}'\boldsymbol{\beta}$.

Till now we did not solve the normal equation (3). If the rank of the $n\times m$-matrix, $R(\mathbf{X})=m$, then $R(\mathbf{X}'\mathbf{X})=m$ and hence the normal equation has an unique solution $\underline{\mathbf{b}}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\underline{\mathbf{y}}$ and naturally $\mathbf{p}'\underline{\mathbf{b}}$ is unique for any given $\mathbf{p}$. Hence all linear parametric functions are estimable, if and only if $R(\mathbf{X})=m$. If rank $R(\mathbf{X})<m$, then some parametric functions do not admit unbiased estimators. For such non-estimable or non-identifiable parametric functions nothing can be inferred.

Let $(\mathbf{X}'\mathbf{X})^-$ be a generalised inverse of $\mathbf{X}'\mathbf{X}$, such that $(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^-(\mathbf{X}'\mathbf{X})=\mathbf{X}'\mathbf{X}$. It can than be shown that $\underline{\mathbf{b}}=(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\underline{\mathbf{y}}$ is a solution of the normal equations. There are two equivalent criterions to conclude that $\mathbf{p}'\boldsymbol{\beta}$ is estimable.

(i) A necessary and sufficient condition that $\mathbf{p}'\boldsymbol{\beta}$ is estimable is that

$$\mathbf{p}'[\mathbf{I}-(\mathbf{X}'\mathbf{X})^-(\mathbf{X}'\mathbf{X})]=\mathbf{0}$$

or

$$\mathbf{p}'=\mathbf{p}'(\mathbf{X}'\mathbf{X})^-(\mathbf{X}'\mathbf{X}).$$

(ii) Let $\mathbf{q}_i$ be basis vectors of the kernel or null space of $\mathbf{X}$, or $\mathbf{X}\mathbf{q}_i=\mathbf{0}$, $i=1, \ldots, s$, where $s=m-R(\mathbf{X})$. Let $\mathbf{Q}'$ be the matrix $(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_s)$ or $\mathbf{X}\mathbf{Q}'=\mathbf{0}$. A necessary and sufficient condition that $\mathbf{p}'\boldsymbol{\beta}$ is estimable is that $\mathbf{p}'\mathbf{Q}'=\mathbf{0}'$ or $\mathbf{Q}\mathbf{p}=\mathbf{0}$.

Because the null space X is often already needed for the construction of a generalised inverse of $\mathbf{X}'\mathbf{X}$ criterion (ii) is easier to use to check the estimability of $\mathbf{p}'\boldsymbol{\beta}$. A generalised inverse of $\mathbf{X}'\mathbf{X}$ is found as $(\mathbf{X}'\mathbf{X}+c\mathbf{Q}'\mathbf{Q})^{-1}$ with $c\neq 0$.

Note that an easy check for estimability in experimental designs is that if $\mathbf{p}'\boldsymbol{\beta}$ can be expressed as a linear combination of expected cell means it is estimable.

Let $\mathbf{p}'\underline{\mathbf{b}}$ and $\mathbf{r}'\underline{\mathbf{b}}$ be the least squares estimators of the estimable functions $\mathbf{p}'\boldsymbol{\beta}$ and $\mathbf{r}'\boldsymbol{\beta}$. Let $(\mathbf{X}'\mathbf{X})^-$ be any generalized inverse of $\mathbf{X}'\mathbf{X}$, then

$$V(\mathbf{p}'\underline{\mathbf{b}})=\sigma^2\mathbf{p}'(\mathbf{X}'\mathbf{X})^-\mathbf{p}, \quad \operatorname{cov}(\mathbf{p}'\underline{\mathbf{b}}, \mathbf{r}'\underline{\mathbf{b}})=\sigma^2\mathbf{p}'(\mathbf{X}'\mathbf{X})^-\mathbf{r}, \tag{4}$$

so $\sigma^2(\mathbf{X}'\mathbf{X})^-$ can be formally considered as the dispersion matrix of $\underline{\mathbf{b}}$ as long as formulas (4) are applied to estimable parametric functions.

When $\mathbf{X}'\mathbf{X}$ is non-singular we have the well-known fact $D(\underline{\mathbf{b}})=\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Note that the calculation of $(\mathbf{X}'\mathbf{X})^-$ is not always necessary to find the variance of $\mathbf{p}'\underline{\mathbf{b}}$. Suppose from the normal equations $\mathbf{X}'\mathbf{X}\underline{\mathbf{b}}=\mathbf{X}'\underline{\mathbf{y}}$ we find that on multiplying by a vector $\mathbf{c}$, $\mathbf{c}'(\mathbf{X}'\mathbf{X})\underline{\mathbf{b}}=\mathbf{c}'(\mathbf{X}'\underline{\mathbf{y}})$ reduces to $\mathbf{p}'\underline{\mathbf{b}}=\mathbf{c}'(\mathbf{X}'\underline{\mathbf{y}})$. Then $\mathbf{c}'(\mathbf{X}'\underline{\mathbf{y}})$ is the BLUE of $\mathbf{p}'\boldsymbol{\beta}$ and $V(\mathbf{p}'\underline{\mathbf{b}})=V(\mathbf{c}'\mathbf{X}'\underline{\mathbf{y}})=\sigma^2\mathbf{c}'(\mathbf{X}'\mathbf{X})\mathbf{c}=\sigma^2\mathbf{c}'\mathbf{p}$. Similarly if $\mathbf{c}'(\mathbf{X}'\mathbf{X})=\mathbf{p}'$ and $\mathbf{d}'(\mathbf{X}'\mathbf{X})=\mathbf{r}'$, then $\mathbf{c}'(\mathbf{X}'\underline{\mathbf{y}})$ and $\mathbf{d}'(\mathbf{X}'\underline{\mathbf{y}})$ are the BLUE of $\mathbf{p}'\boldsymbol{\beta}$ and $\mathbf{r}'\boldsymbol{\beta}$ and $\operatorname{cov}(\mathbf{c}'(\mathbf{X}'\underline{\mathbf{y}}), \mathbf{d}'(\mathbf{X}'\underline{\mathbf{y}}))=\sigma^2\mathbf{c}'\mathbf{r}=\sigma^2\mathbf{d}'\mathbf{p}$.

Example 1. In a randomized (complete) block design two varieties $A_1$ and $A_2$ are tested in 3 blocks $B_1$, $B_2$, $B_3$, each block consist of 2 plots. Unfortunately a tractor destroyed the plot of $A_2$ in block $B_3$. The observations $y_{ij}$ are as follows:

| | $B_1$ | $B_2$ | $B_3$ |
|---|---|---|---|
| $A1$ | $y_{11}=10$ | $y_{12}=11$ | $y_{13}=13$ |
| $A2$ | $y_{21}=10$ | $y_{22}=14$ | * |

One assumes that the fixed effects of factor $A$ and $B$ are additive, hence

$$E(\underline{y}_{ij})=\alpha_i+\beta_j \quad (i=1,2;\ j=1,2,3).$$

We can consider the observations to be uncorrelated with the same variance $\sigma^2$. Hence we have the model $(\underline{\mathbf{y}}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}_5)$ with $\underline{\mathbf{y}}=(\underline{y}_{11}, \underline{y}_{12}, \underline{y}_{13}, \underline{y}_{21}, \underline{y}_{22})'$ and

$$E(\underline{\mathbf{y}})=\mathbf{X}\boldsymbol{\beta}=\begin{bmatrix}1&0&1&0&0\\1&0&0&1&0\\1&0&0&0&1\\0&1&1&0&0\\0&1&0&1&0\end{bmatrix}\begin{bmatrix}\alpha_1\\\alpha_2\\\beta_1\\\beta_2\\\beta_3\end{bmatrix}.$$

The normal equations $\mathbf{X}'\mathbf{X}\mathbf{b}=\mathbf{X}'\mathbf{y}$ are

$$\begin{bmatrix}3&0&1&1&1\\0&2&1&1&0\\1&1&2&0&0\\1&1&0&2&0\\1&0&0&0&1\end{bmatrix}\begin{bmatrix}a_1\\a_2\\b_1\\b_2\\b_3\end{bmatrix}=\begin{bmatrix}34\\24\\20\\25\\13\end{bmatrix}.$$

Note that for many varieties in incomplete blocks a solution of $\mathbf{X}'\mathbf{X}\mathbf{b}=\mathbf{X}'\mathbf{y}$ is often found

by an iterative procedure. Remark that the columns $x_i$ of X are linearly dependent: $x_1+x_2=$ $=x_3+x_4+x_5$ or $x_1+x_2-x_3-x_4-x_5=0$ hence $\mathbf{Xq}=\mathbf{0}$ with $\mathbf{q}=(1, 1, -1, -1, -1)'$, where $\mathbf{q}$ is the basis of the null space of X. The matrix $\mathbf{Q}'=(\mathbf{q})$. The rank of X is $R(\mathbf{X})=$ $=5-1=4$ hence $\mathbf{X}'\mathbf{X}$ is singular.

A generalized inverse $(\mathbf{X}'\mathbf{X})^-=(\mathbf{X}'\mathbf{X}+c\mathbf{Q}'\mathbf{Q})^{-1}$ with $c\neq 0$. When we take $c=1$ we get a

$$(\mathbf{X}'\mathbf{X})^- = \begin{bmatrix} 0.28 & -0.12 & 0.02 & 0.02 & -0.08 \\ -0.12 & 0.48 & -0.08 & -0.08 & 0.32 \\ 0.02 & -0.08 & 0.43 & -0.07 & -0.22 \\ 0.02 & -0.08 & -0.07 & 0.43 & -0.22 \\ -0.08 & 0.32 & -0.22 & -0.22 & 0.88 \end{bmatrix}.$$

With this generalized inverse we find a solution of the normal equations as $\begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} =$

$$=(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y}=\begin{bmatrix} 6.5 \\ 8 \\ 2.75 \\ 5.25 \\ 6.5 \end{bmatrix}.$$

Remark the varietal contrast $\alpha_2-\alpha_1=\mathbf{p}'\boldsymbol{\beta}=(-1, 1, 0, 0, 0)\boldsymbol{\beta}$ is estimable because $\mathbf{p}'\mathbf{q}=0$; $(\alpha_1+\alpha_2)/2=\mathbf{p}'\boldsymbol{\beta}=(\frac{1}{2}, \frac{1}{2}, 0, 0, 0)\boldsymbol{\beta}$ is not estimable because $\mathbf{p}'\mathbf{q}=1\neq 0$; $\alpha_2+(\beta_1+\beta_2+\beta_3)/3=$ $=\mathbf{p}'\boldsymbol{\beta}=(0, 1, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})\boldsymbol{\beta}$ is estimable because $\mathbf{p}'\mathbf{q}=0$. $E(\underline{y}_{23})=\alpha_2+\beta_3$ is estimable with estimate $a_2+b_3=8+6.5=14.5$.

Another generalised inverse of $\mathbf{X}'\mathbf{X}$ can be found by stating one of the five parameters (of which only four parameters are independent) equal to e.g. 0. A solution is found by stating e.g. $a_1=0$. The relevant part of $\mathbf{X}'\mathbf{X}$ is now $\begin{bmatrix} 2 & 1 & 1 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ and this has an inverse $\frac{1}{4}\begin{bmatrix} 4 & -2 & -2 & 0 \\ -2 & 3 & 1 & 0 \\ -2 & 1 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$. A generalised inverse of $\mathbf{X}'\mathbf{X}$ is then $\frac{1}{4}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & -2 & -2 & 0 \\ 0 & -2 & 3 & 1 & 0 \\ 0 & -2 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$. The solution of the normal equations are now $\begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1.5 \\ 9.25 \\ 11.75 \\ 13 \end{bmatrix}$. The BLUE for $\alpha_2-\alpha_1$ is in both cases $a_2-a_1$ with result $1.5=(8-6.5)=(1.5-0)$. The BLUE for $\alpha_1+(\beta_1+\beta_2+\beta_3)/3$ is 11.333 and for $\alpha_2+(\beta_1+\beta_2+\beta_3)/3$ is 12.833. Var $(a_2-a_1)=\sigma^2(0.28+0.48-2(-0.12))=$ $=\sigma^2(0+4-2(0))/4=1\sigma^2$.

2.2. THE MODEL ($\underline{y}$, $\mathbf{X}\boldsymbol{\beta}$, $\mathbf{V}$)

We consider now the model where the $n$ observations $\underline{y}$ can have unequal variances and correlated observations are possible. The linear model is

$$\underline{y}=\mathbf{X}\boldsymbol{\beta}+\underline{e}, \quad \text{with } E(\underline{e})=\mathbf{0} \text{ and } D(\underline{e})=\mathbf{V}=\sigma^2\mathbf{W}, \ \mathbf{V} \text{ is positive definite.}$$

This linear model is denoted as ($\underline{y}$, $\mathbf{X}\boldsymbol{\beta}$, $\mathbf{V}$). By considering $\underline{z}=\mathbf{W}^{-\frac{1}{2}}\underline{y}$ we have

$$E(\underline{z})=\mathbf{W}^{-\frac{1}{2}}\mathbf{X}\boldsymbol{\beta}=\mathbf{U}\boldsymbol{\beta}$$

$$D(\underline{z})=\sigma^2\mathbf{I}_n.$$

The transformed model is $\underline{z}=\mathbf{U}\boldsymbol{\beta}+\underline{f}$ with $E(\underline{f})=\mathbf{0}$ and $D(\underline{f})=\sigma^2(\mathbf{I})$. We can apply now the results of section 2.1 on ($\underline{z}$, $\mathbf{U}\boldsymbol{\beta}$, $\sigma^2\mathbf{I}_n$). The least squares estimator for $\boldsymbol{\beta}$ is found as the solution of the normal equations $\mathbf{U}'\mathbf{U}\underline{b}=\mathbf{U}'\underline{z}$ or

$$\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}\underline{b}=\mathbf{X}'\mathbf{W}^{-1}\underline{y}. \tag{5}$$

The BLUE for estimable parametric functions $\mathbf{p}'\boldsymbol{\beta}$ are given by $\mathbf{p}'\underline{b}$ where $\underline{b}$ is a solution of (5) and such an estimator is called the generalised least squares (GLS) estimator of $\mathbf{p}'\boldsymbol{\beta}$. The variance of an estimable $\mathbf{p}'\underline{b}$ is

$$V(\mathbf{p}'\underline{b})=\sigma^2\mathbf{p}'(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-}\mathbf{p}.$$

Remark: If we take in the model ($\underline{y}$, $\mathbf{X}\boldsymbol{\beta}$, $\mathbf{V}$) as an estimator $\underline{b}$ the solution of $\mathbf{X}'\mathbf{X}\underline{b}=\mathbf{X}'\underline{y}$ then we have for an estimable function $\mathbf{p}'\boldsymbol{\beta}$ a linear unbiased estimator $\mathbf{p}'\underline{b}$, but this estimator has in general no minimum variance in the class of linear unbiased estimators of $\mathbf{p}'\boldsymbol{\beta}$. Sometimes are OLS and GLS estimators for $\mathbf{p}'\boldsymbol{\beta}$. If and only if the column space $C(\mathbf{VX})\subset C(\mathbf{X})$, or $\mathbf{X}$ is invariant under $\mathbf{V}$, then OLS is GLS. See also Baksalary and Kala (1981, 1983).

Example 2. Let us take as an example 2 varieties $A1$ and $A2$ which are tested on 3 random locations (places or sites) $L1$, $L2$, $L3$ drawn at random from a certain region. Due to an accident, the yield of variety $A2$ in place $B3$ was lost. Let the observations be the same as in example 1.

| | $L\,1$ | $L\,2$ | $L\,3$ |
|---|---|---|---|
| $A1$ | $y_{11}=10$ | $y_{12}=11$ | $y_{13}=13$ |
| $A2$ | $y_{21}=10$ | $y_{22}=14$ | * |

From previous experiments in this region, which has the same climatic conditions everywhere, one can state that there is no interaction of variety and places. The statistical model for this mixed model is

$$\underline{y}_{ij}=\alpha_i+\underline{u}_j+\underline{e}_{ij} \quad (i=1,2;\ j=1,2,3)$$

with $E(\underline{u}_j)=0$, $\operatorname{cov}(\underline{u}_j, \underline{u}_{j'})=\delta_{jj'}\sigma_L^2$; $E(\underline{e}_{ij})=0$, $\operatorname{cov}(\underline{e}_{ij}, \underline{e}_{i'j'})=\delta_{ii'}\delta_{jj'}\sigma_e^2$; $\operatorname{cov}(\underline{u}_j, \underline{e}_{ij})=0$ for all $i$ and $j$; where $\delta_{ii'}=1$ if $i=i'$ and $\delta_{ii'}=0$ if $i\neq i'$.

In matrix notation is

$$\underline{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\underline{u}+\underline{e}$$

with $D(\underline{u})=\sigma_L^2\mathbf{I}_3$; $D(\underline{e})=\sigma_e^2\mathbf{I}_5$, $\operatorname{cov}(\underline{u},\underline{e})=\mathbf{0}$, $\mathbf{X}\boldsymbol{\beta}=\begin{bmatrix}1&0\\1&0\\1&0\\0&1\\0&1\end{bmatrix}\begin{bmatrix}\alpha_1\\\alpha_2\end{bmatrix}$; $\mathbf{Z}\underline{u}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\1&0&0\\0&1&0\end{bmatrix}\begin{bmatrix}\underline{u}_1\\\underline{u}_2\\\underline{u}_3\end{bmatrix}$.

The expectation vector $E(\underline{y})=\mathbf{X}\boldsymbol{\beta}$ and the dispersion matrix of $\underline{y}$ is

$$E[(\underline{y}-\mathbf{X}\boldsymbol{\beta})(\underline{y}-\mathbf{X}\boldsymbol{\beta})']=\mathbf{Z}D(\underline{u})\mathbf{Z}'+D(\underline{e})=\sigma_L^2\mathbf{Z}\mathbf{I}_3\mathbf{Z}'+\sigma_e^2\mathbf{I}_5.$$

Hence $D(\underline{y})=\mathbf{V}=\sigma_e^2\mathbf{W}=\sigma_e^2[\gamma\mathbf{Z}\mathbf{Z}'+\mathbf{I}_5]$ with $\gamma=\sigma_L^2/\sigma_e^2$ and

$$\mathbf{W}=\gamma\begin{bmatrix}1&0&0&1&0\\0&1&0&0&1\\0&0&1&0&0\\1&0&0&1&0\\0&1&0&0&1\end{bmatrix}+\begin{bmatrix}1&0&0&0&0\\0&1&0&0&0\\0&0&1&0&0\\0&0&0&1&0\\0&0&0&0&1\end{bmatrix}.$$

Let us assume that from previous experiments is known $\gamma=\sigma_L^2/\sigma_e^2=3$.

$$\mathbf{W}=\begin{bmatrix}4&0&0&3&0\\0&4&0&0&3\\0&0&4&0&0\\3&0&0&4&0\\0&3&0&0&4\end{bmatrix}\quad\text{and}\quad\mathbf{W}^{-1}=\frac{1}{196}\begin{bmatrix}112&0&0&-84&0\\0&112&0&0&-84\\0&0&49&0&0\\-84&0&0&112&0\\0&-84&0&0&112\end{bmatrix}.$$

The normal equations are

$$(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})\begin{bmatrix}a_1\\a_2\end{bmatrix}=\mathbf{X}'\mathbf{W}^{-1}\underline{y}\to\frac{1}{196}\begin{bmatrix}273&-168\\-168&224\end{bmatrix}\begin{bmatrix}a_1\\a_2\end{bmatrix}=\frac{1}{196}\begin{bmatrix}973\\924\end{bmatrix},$$

$$\begin{bmatrix}a_1\\a_2\end{bmatrix}=\frac{196}{32928}\begin{bmatrix}224&168\\168&273\end{bmatrix}\frac{1}{196}\begin{bmatrix}973\\924\end{bmatrix}=\frac{1}{32928}\begin{bmatrix}373184\\415716\end{bmatrix}=\begin{bmatrix}11.333\\12.625\end{bmatrix}.$$

In the general mixed model situation

$$\underline{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\underline{u}+\underline{e},\quad\text{with } E(\underline{u})=\mathbf{0},\ E(\underline{e})=\mathbf{0};\ D(\underline{u})=\sigma_e^2\mathbf{G},\ D(\underline{e})=\sigma_e^2\mathbf{R}$$

$$\text{and } \operatorname{cov}(\underline{u},\underline{e})=\mathbf{0},$$

we have $D(\underline{y})=\mathbf{V}=(\mathbf{Z}\mathbf{G}\mathbf{Z}'+\mathbf{R})\sigma_e^2=\sigma_e^2\mathbf{W}$.

The generalised least squares solution $\underline{b}$ for $\boldsymbol{\beta}$ follows from the normal equations

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})\underline{b}=\mathbf{X}'\mathbf{V}^{-1}\underline{y}\quad\text{or equivalently}\quad(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})\underline{b}=\mathbf{X}'\mathbf{W}^{-1}\underline{y}. \qquad (6)$$

In solving these equations one needs $\mathbf{W}^{-1}$ where $\mathbf{W}$ is a $n\times n$-matrix. The difficulty with this method is that $\mathbf{W}$ is often a matrix so large that its inversion is very costly. An alternative

method is given by Henderson et al. (1959). The idea is that in the fixed model situation

$$\underline{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{u}+\underline{e}=(\mathbf{X}\ \mathbf{Z})\begin{bmatrix}\boldsymbol{\beta}\\ \mathbf{u}\end{bmatrix}+\underline{e}\quad \text{with } D(\underline{e})=\mathbf{R}\sigma_e^2=D(\underline{y}),$$

the normal equations are

$$(\mathbf{X}\ \mathbf{Z})'\mathbf{R}^{-1}(\mathbf{X}\ \mathbf{Z})\begin{bmatrix}\mathbf{b}\\ \hat{\mathbf{u}}\end{bmatrix}=(\mathbf{X}\ \mathbf{Z})'\mathbf{R}^{-1}\mathbf{y}$$

or

$$\begin{bmatrix}\mathbf{X}'\\ \mathbf{Z}'\end{bmatrix}\mathbf{R}^{-1}(\mathbf{X}\ \mathbf{Z})\begin{bmatrix}\mathbf{b}\\ \hat{\mathbf{u}}\end{bmatrix}=\begin{bmatrix}\mathbf{X}'\\ \mathbf{Z}'\end{bmatrix}\mathbf{R}^{-1}\mathbf{y}$$

$$\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\end{bmatrix}\begin{bmatrix}\mathbf{b}\\ \hat{\mathbf{u}}\end{bmatrix}=\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}\end{bmatrix}.$$

In the mixed model situation we have the same patterned matrix at the left-hand-side with the addition of the inverse of $D(\underline{\mathbf{u}})\sigma_e^{-2}=\mathbf{G}$ to $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}$. The generalised least squares estimate $\mathbf{b}$ of $\boldsymbol{\beta}$ is then found as a solution of

$$\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{b}\\ \hat{\mathbf{u}}\end{bmatrix}=\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{y}\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}\end{bmatrix}. \tag{7}$$

These equations (7) are called Henderson's Mixed Model Equations (MME).

Because often $\mathbf{R}$ is a diagonal matrix the only difficulty is the inverse of $\mathbf{G}$ which is of a much smaller order than $\mathbf{W}$. Another computational trick is the absorbing of e.g. the $\hat{\mathbf{u}}$ equations in the $\mathbf{b}$ equations in (7). Let the MME in (7) be written as

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B}' & \mathbf{C}\end{bmatrix}\begin{bmatrix}\mathbf{b}\\ \hat{\mathbf{u}}\end{bmatrix}=\begin{bmatrix}\mathbf{r}_1\\ \mathbf{r}_2\end{bmatrix}$$

then the absorbed equations are

$$[\mathbf{A}-\mathbf{B}\mathbf{C}^{-1}\mathbf{B}']\mathbf{b}=\mathbf{r}_1-\mathbf{B}\mathbf{C}^{-1}\mathbf{r}_2. \tag{8}$$

The proof that a solution of MME (7) is a solution of the GLS equations (6) is based on the identity

$$\mathbf{W}^{-1}=(\mathbf{Z}\mathbf{G}\mathbf{Z}'+\mathbf{R})^{-1}=\mathbf{R}^{-1}-\mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}.$$

From the second line of the MME (7) it follows

$$\hat{\mathbf{u}}=(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1})^{-1}[\mathbf{Z}'\mathbf{R}^{-1}\mathbf{y}-\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\mathbf{b}].$$

Inserting $\hat{\mathbf{u}}$ in the first line of the MME (7) gives

$$\mathbf{X}'[\mathbf{R}^{-1}-\mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}]\mathbf{X}\mathbf{b}=$$

$$=\mathbf{X}'[\mathbf{R}^{-1}-\mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}]\mathbf{y}$$

and hence

$$\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}\mathbf{b}=\mathbf{X}'\mathbf{W}^{-1}\mathbf{y}.$$

Example 2 (continuation). The Mixed Model Equations are with $\mathbf{R}=\mathbf{I}_5=\mathbf{R}^{-1}$, $\mathbf{G}=3\mathbf{I}_3$ hence $\mathbf{G}^{-1}=\frac{1}{3}\mathbf{I}_3$

$$\left[\begin{array}{c|c} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \hline \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z}+\frac{1}{3}\mathbf{I}_3 \end{array}\right]\left[\begin{array}{c} \mathbf{a} \\ \hline \hat{\mathbf{u}} \end{array}\right]=\left[\begin{array}{c} \mathbf{X}'\mathbf{y} \\ \hline \mathbf{Z}'\mathbf{y} \end{array}\right]$$

$$\left[\begin{array}{cc:ccc} 3 & 0 & 1 & 1 & 1 \\ 0 & 2 & 1 & 1 & 0 \\ \hdashline 1 & 1 & \frac{7}{3} & 0 & 0 \\ 1 & 1 & 0 & \frac{7}{3} & 0 \\ 1 & 0 & 0 & 0 & \frac{4}{3} \end{array}\right]\left[\begin{array}{c} a_1 \\ a_2 \\ \hdashline \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \end{array}\right]=\left[\begin{array}{c} 34 \\ 24 \\ \hdashline 20 \\ 25 \\ 13 \end{array}\right].$$

Absorbing of the $\hat{u}_j$ equations in the $a_i$ equations gives

$$\left[\begin{array}{cc} (3-\frac{3}{7}-\frac{3}{7}-\frac{3}{4}) & (0-\frac{3}{7}-\frac{3}{7}) \\ (0-\frac{3}{7}-\frac{3}{7}) & (2-\frac{3}{7}-\frac{3}{7}) \end{array}\right]\left[\begin{array}{c} a_1 \\ a_2 \end{array}\right]=\left[\begin{array}{c} 34-\frac{3}{7}20-\frac{3}{7}25-\frac{3}{4}13 \\ 24-\frac{3}{7}20-\frac{3}{7}25 \end{array}\right]$$

$$\frac{1}{28}\left[\begin{array}{cc} 39 & -24 \\ -24 & 32 \end{array}\right]\left[\begin{array}{c} a_1 \\ a_2 \end{array}\right]=\frac{1}{28}\left[\begin{array}{c} 139 \\ 132 \end{array}\right]$$

$$\left[\begin{array}{c} a_1 \\ a_2 \end{array}\right]=\frac{1}{672}\left[\begin{array}{cc} 32 & 24 \\ 24 & 39 \end{array}\right]\left[\begin{array}{c} 139 \\ 132 \end{array}\right]=\frac{1}{672}\left[\begin{array}{c} 7616 \\ 8484 \end{array}\right]=\left[\begin{array}{c} 11.333 \\ 12.625 \end{array}\right].$$

## 3. A MODEL FOR COMBINED VARIETAL TRIALS

Let us now discuss the testing of $I$ varieties ($A_1, A_2, \ldots, A_I$) in a region. Let us take a random sample of $J$ places (or sites) $P_1, P_2, \ldots, P_J$ from this region. In each location we use a completely randomized design with $H$ plots for each variety. Let $\underline{y}_{ijh}$ be the yield of variety $A_i$ on the $h$-th plot in place $P_j$. A combined analysis is based on the cross-classification of varieties and places.

| | $P_1$ | $\ldots P_j \ldots$ | $P_J$ |
|---|---|---|---|
| $A_1$ | $y_{111}, \ldots, y_{11H}$ | | |
| $\vdots$ | | | |
| $A_i$ | | | |
| $\vdots$ | | | |
| $A_I$ | | | |

The mixed model is

$$\underline{y}_{ijh}=\mu+\alpha_i+\underline{p}_j+(\underline{ap})_{ij}+\underline{e}_{ijh} \tag{9}$$

where $\mu$ is a general mean, $\alpha_i$ is the fixed effect of variety $A_i$ ($i=1, \ldots, I$) and to get rid of the overparametrization we have chosen the $\alpha_i$'s such that $\sum_{i=1}^{I}\alpha_i=0$; $\underline{p}_j$ is the random effect of place $P_j$ ($j=1, \ldots, J$) with $E(\underline{p}_j)=0$, $V(\underline{p}_j)=\sigma_P^2$ and $\mathrm{cov}(\underline{p}_j, \underline{p}_{j'})=0$ if $j\neq j'$; $(\underline{ap})_{ij}$ is the random interaction effect of variety $A_i$ at place $P_j$ with $E[(\underline{ap})_{ij}]=0$, $V[(\underline{ap}_{ij}]=\sigma_{AP}^2$ for all $i$ and $j$; $\underline{e}_{ijh}$ is the residual random error or plot to plot variation with $E(\underline{e}_{ijh})=0$, $V(\underline{e}_{ijh})=\sigma_e^2$ for

all $i, j$ and $h$; the $\underline{e}_{ijh}$'s are uncorrelated; furthermore $\mathrm{cov}(\underline{p}_j, (\underline{ap})_{ij})=0$, $\mathrm{cov}(\underline{p}_j, \underline{e}_{ijh})=0$, $\mathrm{cov}((\underline{ap})_{ij}, \underline{e}_{ijh})=0$.

If we calculate the analysis of variance (ANOVA) table, one will find in the literature two different Expected Mean Squares columns (i) and (ii):

| Source of Variation | D.F. | M.S. | E.M.S. (j) | E.M.S. (ii) |
|---|---|---|---|---|
| Varieties (fixed) | $I-1$ | MSA | $\sigma_e^2+H\sigma_{AP}^2+JH\kappa_A^2$ | $\sigma_e^2+H\sigma_{AP}^2+JH\kappa_A^2$ |
| Places (random) | $J-1$ | MSP | $\sigma_e^2+H\sigma_{AP}^2+IH\sigma_P^2$ | $\sigma_e^2+IH\sigma_P^2$ |
| Interaction | $(I-1)(J-1)$ | MSAP | $\sigma_e^2+H\sigma_{AP}^2$ | $\sigma_e^2+H\sigma_{AP}^2$ |
| Error | $IJ(H-1)$ | MSE | $\sigma_e^2$ | $\sigma_e^2$ |
| Corrected Total | $IJH-1$ | | | |

with $\kappa_A^2=\sum_{i=1}^{I}(\alpha_i-\bar{\alpha})^2/(I-1)=\sum_{i=1}^{I}\alpha_i^2/(I-1)$ while $\bar{\alpha}=\sum_{i=1}^{I}\alpha_i/I=0$.

At a first glance the difference in the E. M. S. column is in $E(\underline{\mathrm{MSP}})$ where the term $H\sigma_{AP}^2$ is present in (i) and absent in (ii). In the statistical handbooks we find EMS (i) for example in Steel and Torrie (1980) and Kirk (1968) and one finds EMS (ii) for example in Scheffé (1959), Snedecor and Cochran (1980) and Winer (1971). Inclusion of $H\sigma_{AP}^2$ from $E(\underline{\mathrm{MSP}})$ in (i) arise when the random effects $(\underline{ap})_{ij}$ in (9) for the mixed model, are defined uncorrelated. Exclusion of $H\sigma_{AP}^2$ from $E(\underline{\mathrm{MSP}})$ in (ii) can arise when $(\underline{ap})_{ij}$ in (9) are defined such that $\sum_{i=1}^{I}(\underline{ap})_{ij}=0$ for all $j$. In this case there is a correlation between $(\underline{ap})_{ij}$ and $(\underline{ap})_{i'j}$ $(i\neq i'; j=1, \ldots, J)$. Reasonable is to assume that $\mathrm{cov}[(\underline{ap})_{ij}, (\underline{ap})_{i'j}]=\omega$ for all $j$ and $i\neq i'$. Now is

$$V[\sum_{i=1}^{I}(\underline{ap})_{ij}]=V(0)=0$$

but otherwise

$$V[\sum_{i=1}^{I}(\underline{ap})_{ij}]=\sum_{i=1}^{I}\sigma_{AP}^2+\sum_{\substack{i=1\\ i\neq i'}}^{I}\sum_{i'=1}^{I}\omega=I\sigma_{AP}^2+I(I-1)\omega$$

and hence

$$\omega=-\sigma_{AP}^2/(I-1).$$

The limitation $\sum_{i=1}^{I}(\underline{ap})_{ij}=0$ for all $j$ is quite reasonable when we adopt the mixed model derivation of Scheffé (1959). Searle (1971) shows that the mixed model (i) can be redefined to get mixed model (ii).

Redefine

$$\underline{p}_j^*=\underline{p}_j+\sum_{i=1}^{I}(\underline{ap})_{ij}/I \quad \text{and} \quad (\underline{ap})_{ij}^*=(\underline{ap})_{ij}-\sum_{i=1}^{I}(\underline{ap})_{ij}/I.$$

Now

$$V(\underline{p}_j^*)=\sigma_p^2+\sigma_{AP}^2/I=\sigma_p^{*2},$$

$$V[(\underline{ap})_{ij}^*]=\sigma_{AP}^2+\sigma_{AP}^2/I-2\sigma_{AP}^2/I=\sigma_{AP}^2(I-1)/I=\sigma_{AP}^{*2},$$

$$\mathrm{cov}[(\underline{ap})_{ij}^*, (\underline{ap})_{i'j}^*]=-\sigma_{AP}^{*2}/(I-1) \quad \text{for} \quad i\neq i'.$$

The variance components in EMS (ii) are $\sigma^2_{AP}(ii)=\sigma^{*2}_{AP}I/(I-1)=\sigma^2_{AP}$,

$$\sigma^2_P(ii)=\sigma^{*2}_P=\sigma^2_P+\sigma^2_{AP}/I;$$

so $\sigma^2_P$ has in EMS$(i)$ and EMS$(ii)$ a different meaning! Furthermore there is no correlation between $(\underline{ap})_{ij}$ and $(\underline{ap})_{i'j}$ in mixed model (i) but there is a correlation in mixed model (ii).

In matrix notation the mixed model can be written as

$$\underline{\mathbf{y}}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}_1\underline{\mathbf{u}}_1+\mathbf{Z}_2\underline{\mathbf{u}}_2+\underline{\mathbf{e}} \quad \text{with} \quad \boldsymbol{\beta}=(\mu,\alpha_1,\alpha_2,\ldots,\alpha_I)'$$

$$\underline{\mathbf{u}}_1=(\underline{p}_1,\underline{p}_2,\ldots,\underline{p}_J)$$

$$\text{and} \quad \underline{\mathbf{u}}_2=((\underline{ap})_{11},(\underline{ap})_{12},\ldots,(\underline{ap})_{IJ})'$$

where X, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are the design matrices; the dispersion matrix of $\underline{\mathbf{u}}_1$ is $D(\underline{\mathbf{u}}_1)=\sigma^2_P\,\mathbf{I}_J$, $D(\underline{\mathbf{e}})=\sigma^2_e\,\mathbf{I}_n$ with $n=IJH$ and for mixed model (i)

$$D(\underline{\mathbf{u}}_2)=\sigma^2_{AP}\,\mathbf{I}_{IJ}$$

and for mixed model (ii)

$$D(\underline{\mathbf{u}}_2)=\sigma^2_{AP}/(I-1)\,[I\times\mathbf{I}_I-\mathbf{J}_I]\otimes\mathbf{I}_J=[\sigma^2_{AP}/(I-1)]\,\Omega,$$

where $\mathbf{J}_I$ is $I\times I$-matrix with all elements equal to 1 and the Kronecker product $\mathbf{A}\otimes\mathbf{B}$ is $(a_{ij}\mathbf{B})$. The dispersion matrix of $\underline{\mathbf{y}}$ is for model (i):

$$D(\underline{\mathbf{y}})=V(i)=\mathbf{Z}_1(\sigma^2_P\,\mathbf{I}_J)\,\mathbf{Z}_1'+\mathbf{Z}_2(\sigma^2_{AP}\,\mathbf{I}_{IJ})\,\mathbf{Z}_2'+\sigma^2_e\,\mathbf{I}_n=$$

$$=\sigma^2_e\,[\lambda_1\,\mathbf{Z}_1\,\mathbf{Z}_1'+\lambda_2\,\mathbf{Z}_2\,\mathbf{Z}_2'+\mathbf{I}_n]=\sigma^2_e\,\mathbf{W}(i) \quad \text{with} \quad \lambda_1=\sigma^2_P/\sigma^2_e;\ \lambda_2=\sigma^2_{AP}/\sigma^2_e.$$

The dispersion matrix of $\underline{\mathbf{y}}$ is for model (ii):

$$D(\underline{\mathbf{y}})=V(ii)=\mathbf{Z}_1(\sigma^2_P\,\mathbf{I}_J)\,\mathbf{Z}_1'+\mathbf{Z}_2(\sigma^2_{AP}/(I-1)\,\Omega)\,\mathbf{Z}_2'+\sigma^2_e\,\mathbf{I}_n=$$

$$=\sigma^2_e\,[\lambda_1\,\mathbf{Z}_1\,\mathbf{Z}_1'+\lambda_2\,\mathbf{Z}_2\Omega\mathbf{Z}_2'+\mathbf{I}_n]=\sigma^2_e\,\mathbf{W}(ii)$$

$$\text{with} \quad \lambda_1=\sigma^2_P/\sigma^2_e;\ \lambda_2=\sigma^2_{AP}/[\sigma^2_e(I-1)].$$

For the estimation of varietal contrasts we can now use the BLUE such as already have been discussed in section 2. It turns out that for a balanced cross-classification the ordinary least squares estimator (OLS) for $\mathbf{p}'\boldsymbol{\beta}$ is the same as the generalised least squares estimators (GLS). This feature has to do with the fact that the column space of X is invariant with respect to the dispersion matrix V. But in the unbalanced case the OLS and the GLS estimators give different outcomes. See for conditions that OLS is GLS Baksalary and Kala (1981, 1983).

A more realistic example is the testing of $I$ varieties on $J$ random places, but in each location there is a need for blocking, for example there is a fertility gradient in the experimental field which can be accounted for by rectangular blocks with the long side perpendicular on the gradient, or the experimental field consits of homogeneous parts of different fertility. In each location we use a randomized block design with $H$ complete blocks of $I$ plots (such complete blocks are often called replications). Let $\underline{y}_{ijh}$ be the yield

of variety $A_i$ at place $P_j$ in the $h$-th block $B_{hj}$ of place $P_j$, then the mixed model is

$$\underline{y}_{ijh}=\mu+\alpha_i+\underline{p}_j+\beta_{h(j)}+(\underline{ap})_{ij}+\underline{e}_{ijh}. \tag{10}$$

For the fixed block effects $\beta_{h(j)}$ we choose the condition $\sum_{h=1}^{H}\beta_{h(j)}=0$ for each $j=1, \ldots, J$ to get rid of the overparametrization. The other effects are defined as in (9). The ANOVA-table gives then according to model (i) – no limitation on $(\underline{ap})_{ij}$ and model (ii) – $\sum_{i=1}^{I}(\underline{ap})_{ij}=0$ for each $j$:

| Source of Variation | D.F. | M.S. | E.M.S. (i) | E.M.S. (ii) |
|---|---|---|---|---|
| Varieties | $I-1$ | MSA | $\sigma_e^2+H\sigma_{AP}^2+JH\kappa_A^2$ | $\sigma_e^2+H\sigma_{AP}^2+JH\kappa_A^2$ |
| Places | $J-1$ | MSP | $\sigma_e^2+H\sigma_{AP}^2+IH\sigma_P^2$ | $\sigma_e^2+IH\sigma_P^2$ |
| Blocks within places | $J(H-1)$ | MSB | $\sigma_e^2+H\sigma_{AP}^2+I\kappa_B^2$ | $\sigma_e^2+I\kappa_B^2$ |
| Interaction $A\times P$ | $(I-1)(J-1)$ | MSI | $\sigma_e^2+H\sigma_{AP}^2$ | $\sigma_e^2+H\sigma_{AP}^2$ |
| Error | $(I-1)J(H-1)$ | MSE | $\sigma_e^2$ | $\sigma_e^2$ |
| Corrected Total | $IJH-1$ | | | |

with $\kappa_A^2=\sum_{i=1}^{I}\alpha_i^2/(I-1)$ and $\kappa_B^2=\sum_{j}^{J}\sum_{h}^{H}\beta_{h(j)}^2/[J(H-1)]$.

In the balanced case with a completely randomized design [mixed model (9)] or a randomized (complete) block design [mixed model (10)] we can also analyze the two-way table of varietal means

$$\bar{\underline{y}}_{ij.}=\sum_{h=1}^{H}\underline{y}_{ijh}/H.$$

The mixed model for $\bar{\underline{y}}_{ij\bullet}$ is than for both cases (i) and (ii):

$$\bar{\underline{y}}_{ij.}=\mu+\alpha_i+\underline{p}_j+(\underline{ap})_{ij}+\bar{\underline{e}}_{ij.}, \tag{11}$$

with $V(\bar{\underline{e}}_{ij.})=\sigma_e^2/H$ but $D[(\underline{ap})_{ij}]$ is different in model (i) or (ii).

If we have a different number of replications $r_j$ per place $P_j$, $V(\bar{\underline{e}}_{ij.})=\sigma_e^2/r_j$ and $D(\bar{\underline{e}})$ is a diagonal matrix with diagonal elements $\sigma_e^2/r_j$. In the case of a randomized block design with missing values at a certain place or with an incomplete block design the best estimators for $\mu+\alpha_i$ is not $\bar{\underline{y}}_{ij}$. more and the BLUE of $\mu+\alpha_i$ per place has a dispersion-matrix which is not diagonal any more! Also the use of $\alpha$-designs of Patterson et al. (1978) in varietal testing has the impact of a more complicated structure for the dispersion-matrix of the BLUE of $\mu+\alpha_i$. Till now in practice one neglects this and uses for the analysis of a two-way table of varieties and places with the estimates of $\mu+\alpha_i$ per place an error dispersion-matrix which is a constant times the identity matrix. With the possibility now to use the computer it is wise to use the generalised least squares estimators with the correct dispersion matrix.

From the generalised least squares method with use of the mixed model normal equations (7) one can see that if the ratio of the variance components $\sigma_P^2/\sigma_e^2$ and $\sigma_{AP}^2/\sigma_e^2$ are very large the generalised least squares method for a mixed model is about the same as the ordinary

least squares method for a fixed model where the effect of places are considered to be fixed. This means that if we take places at random from a country with different climatic and/or soil regions, we can expect a very large $\sigma_P^2/\sigma_e^2$. Often it is better to divide the country in strata (=regions) according to the more homogeneous climatic or soil conditions and the ratio of $\sigma_P^2/\sigma_e^2$ is then not so large.

One can expect in our developed countries in Europe that the differences of yield that we may expect to put forward in varietal field experiments are about 5 percent of the mean yield. Just for such small differences in yielding capacity it is worthwile to use the more elaborated generalised least square method to find BLUE for varietal contrast!

Over-year summaries can be obtained by assembling the BLUE of varietal means for each season in a varieties $x$ years table. The same remarks we have made before can be used for analyzing such tables according to the correct model. Prior experiments give ideas for $\sigma_P^2/\sigma_e^2$ and $\sigma_{\lambda P}^2/\sigma_e^2$.

## 4. BEST LINEAR UNBIASED PREDICTION

A mixed linear model arises in combined variety trials and in many genetic applications and can be represented as we have seen in section 2 as

$$\underline{y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\underline{u}+\underline{e} \tag{12}$$

where $\underline{y}$ is an $n\times 1$-observation vector. $\mathbf{X}$ is a known $n\times m$-matrix, $\boldsymbol{\beta}$ is an unknown fixed $m\times 1$-vector and $\mathbf{Z}$ is a known $n\times k$-matrix; $\underline{u}$ and $\underline{e}$ are nonobservable random vectors with $E(\underline{u})=0$ and $E(\underline{e})=0$ and dispersion-matrix

$$D\begin{bmatrix}\underline{u}\\ \underline{e}\end{bmatrix}=\begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}\sigma_e^2,$$

where $\sigma_e^2$ is a scalar, possibly unknown, and $\mathbf{G}$ and $\mathbf{R}$ are both nonsingular. Now, given a sample vector $\underline{y}$, we wish now to do often

1. Predict $\underline{u}$ or some linear function of $\underline{u}$;
2. Predict linear functions of $\boldsymbol{\beta}$ and $\underline{u}$ jointly.

The varietal advisory service and the practical breeder wish to do these predictions. C. R. Henderson has done much work to apply prediction in animal breeding, see e.g. Henderson (1973, 1975). It can be shown that if $\boldsymbol{\beta}$, $\mathbf{G}$ and $\mathbf{R}$ are known, the best linear predictor (BLP) of $\mathbf{p}'\boldsymbol{\beta}+\mathbf{q}'\underline{u}$, is,

$$\mathbf{p}'\boldsymbol{\beta}+\mathbf{q}'\operatorname{cov}(\underline{u},\underline{y}')\mathbf{V}^{-1}(\underline{y}-\mathbf{X}\boldsymbol{\beta}), \quad \text{where } \operatorname{cov}(\underline{u},\underline{y}')=\mathbf{G}\mathbf{Z}'\sigma_e^2, \tag{13}$$

$$D(\underline{y})=\mathbf{V}=\sigma_e^2[\mathbf{Z}\mathbf{G}\mathbf{Z}'+\mathbf{R}]=\sigma_e^2\mathbf{W}.$$

In the multivariate normal case, (13) is of course $E(\mathbf{p}'\boldsymbol{\beta}+\mathbf{q}'\underline{u}|\underline{y})$ and consequently is the best predictor. Whether or not normality is implied, (13) is the usual selection index used in animal and plant breeding. Now if $\boldsymbol{\beta}$ is unknown, as is usually the case, the method (13) cannot be used. A modification involving unbiasedness can be employed, however. By unbiased it is mean $E(\text{predictor})=E(\mathbf{p}'\boldsymbol{\beta}+\mathbf{q}'\underline{u})=\mathbf{p}'\boldsymbol{\beta}$. It can be shown (e.g. Henderson

(1973)) that the best linear unbiased predictor (BLUP) of $\mathbf{p}'\boldsymbol{\beta}+\mathbf{q}'\underline{\mathbf{u}}$, where $\mathbf{p}'\boldsymbol{\beta}$ is estimable, is:

$$\mathbf{p}'\underline{\mathbf{b}}+\mathbf{q}'\mathbf{G}\mathbf{Z}'\mathbf{W}^{-1}(\underline{\mathbf{y}}-\mathbf{X}\underline{\mathbf{b}})=\mathbf{p}'\underline{\mathbf{b}}+\mathbf{q}'\hat{\underline{\mathbf{u}}}, \tag{14}$$

with $\hat{\underline{\mathbf{u}}}=\mathbf{G}\mathbf{Z}'\mathbf{W}^{-1}(\underline{\mathbf{y}}-\mathbf{X}\underline{\mathbf{b}})$ and where $\underline{\mathbf{b}}$ is any solution of the generalized least squares equations

$$\mathbf{X}'\mathbf{W}^{-1}\mathbf{X}\underline{\mathbf{b}}=\mathbf{X}'\mathbf{W}^{-1}\underline{\mathbf{y}}. \tag{15}$$

In animal breeding programs, such as in the national sire evaluation programs the order of $\underline{\mathbf{y}}$ may be in the millions, and $\mathbf{V}$ is not diagonal. Therefore (14) and (15) are not computationally feasible. However, $\underline{\mathbf{b}}$ and $\hat{\underline{\mathbf{u}}}$ can be obtained from the solution of the mixed model equations, as we already seen in section 2.2.:

$$\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\end{bmatrix}\begin{bmatrix}\underline{\mathbf{b}}\\ \hat{\underline{\mathbf{u}}}\end{bmatrix}=\begin{bmatrix}\mathbf{X}'\mathbf{R}^{-1}\underline{\mathbf{y}}\\ \mathbf{Z}'\mathbf{R}^{-1}\underline{\mathbf{y}}\end{bmatrix}, \tag{16}$$

which can be solved either directly or by iterative methods when $\mathbf{R}$ is a diagonal or a patterned matrix. From the inverse of the coefficient matrix $V(\underline{\mathbf{b}})$ and $V(\hat{\underline{\mathbf{u}}}-\underline{\mathbf{u}})$ can be obtained.

In addition to unbiasedness and minimum prediction error variance, $\hat{\underline{\mathbf{u}}}$ from (16) has the following properties (Henderson 1973):

(i) it maximizes the probability that elements of $\underline{\mathbf{u}}$ are correctly ranked pairwise, in a multivariate setting;

(ii) under normality, it is the maximum likelihood estimator of $E(\underline{\mathbf{u}}|\underline{\mathbf{y}})$;

(iii) *it is linearly invariant, i.e. BLUP* $(\mathbf{q}'\underline{\mathbf{u}})=\mathbf{q}'\hat{\underline{\mathbf{u}}}$, *where* $\mathbf{q}$ *is a vector of constants.*

In a Bayesian setting with normal prior and normal likelihood, $\hat{\underline{\mathbf{u}}}$ is the median of the posterior distribution. Hence, any function of $\underline{\mathbf{u}}$ is estimated with minimum absolute error loss by the same function of $\hat{\underline{\mathbf{u}}}$. In varietal testing literature the notion of BLUP is seldom (or never) mentioned. In my opinion many questions in varietal research is the prediction and not the estimation. Use of equation (14) can be fruitfully adopted for the prediction of the yielding capacity of varieties.

Example 3. Let us take as an example 2 varieties $A1$ and $A2$ which are tested on 3 random places $P1$, $P2$, $P3$ drawn at random from a certain region. The experimental design used at each place was a randomized block design with four blocks. Due to a misunderstanding in place $P3$ variety $A2$ was not sown but another variety. The varietal means $\bar{y}_{ij.}$ per place were

| | $P1$ | $P2$ | $P3$ |
|---|---|---|---|
| $A1$ | $\bar{y}_{11.}=10$ | $\bar{y}_{12.}=11$ | $\bar{y}_{13.}=13$ |
| $A2$ | $\bar{y}_{21.}=10$ | $\bar{y}_{22.}=14$ | * |

The mixed model for these varietal means is

$$\bar{\underline{\mathbf{y}}}_{ij.}=\alpha_i+\underline{p}_j+(\underline{ap})_{ij}+\bar{\underline{e}}_{ij.} \tag{17}$$

where we assume model (i) of section 3, uncorrelated random effects $(\underline{ap})_{ij}$ in (9). We can

redefine the model as

$$\bar{y}_{ij.}=\alpha_i+p_j+f_{ij} \tag{18}$$

with $E(p_j)=0$, Var $(p_j)=\sigma_P^2$ and uncorrelated $p_j$ and $E(f_{ij})=0$, Var $(f_{ij})=\sigma_{AP}^2+\sigma_e^2/4=\sigma_f^2$ with uncorrelated $f_{ij}$ and also $p_j$ and $f_{ij}$ are uncorrelated. Let us further assume that from previous experiments we know $\gamma=\sigma_P^2/\sigma_f^2=3$. We wish to predict the varietal mean of $A2$ at place $P3$. Let

$$\underline{y}=(\bar{y}_{11.}, \bar{y}_{12.}, \bar{y}_{13.}, \bar{y}_{21.}, \bar{y}_{22.})'$$

then

$$E\underline{y}=\mathbf{X}\beta+\mathbf{Z}\underline{U}+\underline{f}$$

with $\underline{u}=(p_1, p_2, p_3)'$, and $\mathbf{X}$, $\beta$ and $\mathbf{Z}$ as given in Example 2, section 2.2. The dispersion matrix $D\begin{bmatrix}\underline{u}\\ \underline{f}\end{bmatrix}=\begin{bmatrix}\mathbf{G} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}\end{bmatrix}\sigma_f^2$ where $G=\gamma\mathbf{I}_3$ and $\mathbf{R}=\mathbf{I}_5$.

The best linear unbiased predictor of $\mathbf{p}'\beta+\mathbf{q}'\underline{u}$ with $\mathbf{p}'=(0, 1)$ and $\mathbf{q}'=(0, 0, 1)$ is according to (14)

$$\mathbf{p}'\underline{b}+\mathbf{q}'\mathbf{G}\mathbf{Z}'\mathbf{W}^{-1}(\underline{y}-\mathbf{X}\underline{b})$$

with

$$\mathbf{W}=\mathbf{Z}\gamma\mathbf{I}_3\mathbf{Z}'+\mathbf{I}_5=3\mathbf{Z}\mathbf{Z}'+\mathbf{I}_5.$$

The inverse $\mathbf{W}^{-1}$ is given in example 2 and the solution of (15) of $\mathbf{b}$ is found there as

$$\mathbf{b}=\begin{bmatrix}a_1\\ a_2\end{bmatrix}=\begin{bmatrix}11.333\\ 12.625\end{bmatrix}.$$

Inserting these results we get as BLUP for $\mathbf{p}'\beta+\mathbf{q}'\underline{u}$:

$$(0, 1)\begin{bmatrix}11.333\\ \\ 12.625\end{bmatrix}+(0, 0, 1)\begin{bmatrix}-1.69629\\ 0.44657\\ 1.25025\end{bmatrix}=13.875.$$

## REFERENCES

Baksalary J. K., Kala R., (1981). Simple least squares estimation versus best linear unbiased prediction. Journal of Statistical Planning and Inference 5, 147 - 151.

Baksalary J. K., Kala R., (1983). On equalities between BLUEs, WLSEs, and SLSEs. The Canadian Journal of Statistics, 11, 119 - 123.

Henderson C. R., Kempthorne O., Searle S. R., Von Krosigk C. M., (1959). The estimation of environmental and genetic trends from records subject to culling. Biometrics 15, 192 - 218.

Henderson C. R., (1973). Sire evaluation and genetic trends. In proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr. Jay. L. Lush, 10 - 41. Champaign, Illinois: American Society of Animal Science and American Dairy Science Association.

Henderson C. R., (1975). Best linear unbiased estimation and prediction under a selection model. Biometrics 31, 423 - 447.

Kirk R. E., (1968). Experimental design: Procedures for the Behavioural Science. Brooks/Cole, Belmont, California.

Patterson H. D., Williams E. R., Hunter E. A., (1978). Block designs for variety trials. Journal of Agricultural Science, Cambridge, 90, 395 - 400.

Rao C. R., (1973). Linear statistical inference and its applications, 2nd ed. John Wiley and Sons, New York.

Scheffé H., (1959). The analysis of variance. John Wiley and Sons, New York.

Searle S. R., (1971). Linear models. John Wiley and Sons, New York.

Searle S. R., (1971). Topics in variance component estimation. Biometrics 27, 1 - 76.

Snedecor G. W., Cochran W. G., (1980). Statistical methods, 7th ed. The Iowa State University Press, Ames, Iowa.

Steel R. G. D., Torrie J. H., (1980). Principles and procedures of Statistics, a biometrical approach, 2nd ed. McGraw-Hill, New York.

Winer B. J., (1971). Statistical principles in experimental design, 2nd ed. McGraw-Hill.

## NAJLEPSZA ESTYMACJA KONTRASTÓW OBIEKTOWYCH STOSOWANA DO ZŁOŻONYCH MODELI DOŚWIADCZEŃ ODMIANOWYCH

Jednym z głównych powodów prowadzenia doświadczeń odmianowych jest ocena możliwości plonowania odmian. Założenia stojące u podstaw poprawnych procedur etymacyjnych są względnie proste i nie wymagają normalności rozkładów. Chociaż zasady najlepszej liniowej nieobciążonej estymacji BLUE dla kontrastów obiektowych są opisane w wielu podręcznikach statystyki matematycznej (np. Rao 1973 lub Searle'a 1971), to wydaje się, że nie są one rozpowszechnione w badaniach odmianowych. Częściej można je spotkać w badaniach hodowlanych zwierząt. Podobnie, metoda najlepszej liniowej nieobciążonej predykcji (BLUP) kontrastów nie jest stosowana w badaniach odmianowych.

W pracy zaprezentowano zasady najlepszej liniowej nieobciążonej estymacji w modelach liniowych oraz zademonstrowano prosty przykład numeryczny. Ponadto, przedstawiono model liniowy złożonego doświadczenia odmianowego, w którym pokazano estymację BLU. Wreszcie, opisano metodę najlepszej liniowej predykcji, która w badaniach odmianowych jest częściej stosowana niż postępowanie estymacyjne.

## SUMMARY

In several sciences but especially in animal and plant breeding, the general mixed model with fixed and random effects plays a great role. Statistical inference on variance components means tests of hypotheses about variance components, constructing confidence intervals for them, estimating them, and using the variance components to get best estimates for fixed effects as well as to predict random effects.

Many problems in the statistical inference of variance components already arise in even the most simple mixed model, describing nested designs; they are already present in the balanced nested designs, but become more pronounced in the unbalanced case. To find a guideline for solving the problems for the general mixed model, the study of the nested designs is worthwile.

In chapter 1 the outline of the author's research on statistical inference of variance components is given.

In chapter 2 the historical development of variance components estimation is described. More than 125 years ago the notion of variance components appears explicitly in astronomy. The expansion began with the development of quantitative genetics in 1918 after the first World War, but the tremendous increase in research in variance components dates from after the second World War.

Just as in the fixed effects model, the notions of vectors, vector spaces and projections of vectors on vector spaces, can be fruitfully used for the mixed effects model. This approach is given in chapter 3.

For balanced nested designs exact tests about ratios of variance components and the calculation of their power, are well-known. For the unbalanced three-stage nested designs an exact test exists about the variance component belonging to the second stage, but for the first stage variance component no exact test was available in practice before

1974. In chapter 4 the onset of the author's research on the testing problem is described. It was possible to calculate the critical level or P-value for the exact test.

The estimation of variance components was directed towards finding, for several designs, the best quadratic unbiased estimators for the variance components. In chapter 5 the onset of author's research on the estimation problem is described. On introduction of the concept of a permissible estimator it becomes clear which approaches to estimate variance components are unsatisfactory. To use permissible estimators for variance components means to use estimators which give non-negative estimates. A solution for a non-negative estimator has been found using the least squares approach as a unified procedure in variance components estimation.

To demonstrate the danger of the use of the best quadratic unbiased estimator in the simplest random model, the balanced one-way classification, the probabilities for negative estimates were calculated in chapter 6.

In chapter 7 a new exact test about ratios of variance components in the unbalanced three-stage nested design is given.

The description of the least squares method as a unified procedure for the estimation of variance components and for the derivation of permissible (non-negative) estimators of variance components is given in chapter 8. The use of the main estimation procedures such as iterated Least Squares (or I-MINQUE), Maximum Likelihood (ML) and Restricted Maximum Likelihood (REML) estimators, is also discussed there.

Finally in chapter 9 the use of variance components in predicting the random effects is discussed. Best Linear Unbiased Prediction (BLUP), which is extensively used in animal breeding, has not so far

been applied on the same scale in plant breeding. How to use it is the subject of the last section 9.2.

SAMENVATTING

In diverse wetenschapsgebieden, maar voornamelijk in de veefokkerij en de plantenveredeling, speelt het algemene gemengde model met vaste en stochastische effecten een grote rol. Statistische gevolgtrekkingen maken omtrent variantiecomponenten betekent: het toetsen van hypothesen over variantiecomponenten, het construeren van betrouwbaarheidsintervallen ervoor, het schatten ervan, en het gebruik ervan, zowel om de beste schattingen voor de vaste effecten te verkrijgen, alsook om de stochastische effecten te kunnen voorspellen.

Veel problemen die zich voordoen bij het maken van statistische gevolgtrekkingen omtrent variantiecomponenten, komen reeds voor in het eenvoudigste gemengde model, dat de hiërarchische of geneste schema's beschrijft. Deze problemen zijn versluierd aanwezig bij de evenwichtige geneste schema's, maar komen zeer sterk naar voren bij de onevenwichtige schema's. Om een richtsnoer te vinden voor het oplossen van de problemen bij het algemene gemengde model, is de studie van hiërarchische schema's de moeite waard.

In hoofdstuk 1 wordt een overzicht gegeven van het onderzoek van de auteur met betrekking tot het maken van statistische gevolgtrekkingen omtrent variantiecomponenten.

In hoofdstuk 2 wordt de historische ontwikkeling van het schatten van variantiecomponenten beschreven. Reeds meer dan 125 jaar geleden kwam het begrip variantiecomponenten expliciet voor in de sterrenkunde. De verdere ontwikkeling begon met die van de kwantitatieve genetica in 1918, na de eerste wereldoorlog, maar de enorme uitbreiding in het onderzoek naar variantiecomponenten dateert van na de tweede wereldoorlog.

Net als bij het model met vaste effecten, kunnen de begrippen vector, vectorruimte en projectie van vectoren op vectorruimten met vrucht gebruikt worden bij het model met gemengde effecten. Deze aanpak wordt in hoofdstuk 3 beschreven.

Voor evenwichtige hiërarchische bemonsteringsschema's zijn exacte toetsen omtrent verhoudingen van variantiecomponenten algemeen bekend, evenals de berekening van het onderscheidingsvermogen. Voor het niet-evenwichtige hiërarchische bemonsteringsschema in drie trappen bestaat er een exacte toets op de variantiecomponent die bij de tweede trap behoort. Maar voor de variantiecomponent die bij de eerste trap behoort was het voor 1969 practisch niet mogelijk om de kansverdeling van de toetsingsgrootheid te berekenen. In hoofdstuk 4 wordt het begin beschreven van het door de auteur verrichte onderzoek aangaande dit toetsingsprobleem. Het bleek mogelijk om de overschrijdingskans van de exacte toets te berekenen.

Het schatten van variantiecomponenten was erop gericht om, voor verschillende schema's, de beste zuivere kwadratische schatters te vinden. In hoofdstuk 5 komt aan de orde het begin van het door de schrijver verrichte onderzoek betreffende het schattingsprobleem. Na invoering van het begrip van een toegestane ('permissible') schatter ziet men duidelijk of een bepaalde aanpak om variantiecomponenten te schatten tot onbevredigende resultaten leidt. Een toegestane schatter voor variantiecomponenten is een schatter die niet-negatieve uitkomsten levert. Hoe een niet-negatieve schatter door de schrijver gevonden is, door het toepassen van het kleinste kwadraten principe als unificerende schattingsprocedure, wordt ook beschreven in hoofdstuk 5.

In hoofdstuk 6 wordt geïllustreerd hoe gevaarlijk het kan zijn om zelfs in het eenvoudigste model met enkel stochastische effecten, nameljk het evenwichtige bemonsteringsschema met één classificatie,

de beste zuivere kwadratische schatter te gebruiken. Hiertoe worden de kansen berekend dat zo'n schatter negatieve uitkomsten levert.

In hoofdstuk 7 wordt een nieuwe exacte toets beschreven omtrent verhoudingen van variantiecomponenten in het niet-evenwichtige drietraps geneste bemonsteringsschema.

Hoofdstuk 8 bevat de beschrijving van de methode van de kleinste kwadraten als unificerende procedure voor het schatten van variantiecomponenten. Verder wordt de methode van kleinste kwadraten gebruikt om toegestane (niet-negatieve) schatters te vinden.

Dit hoofdstuk besluit met een bespreking van het gebruik van de voornaamste schattingsprocedures, namelijk die van de geïtereerde kleinste kwadraten (I-MINQUE), de methode van de grootste aannemelijkheid (ML) en de methode van de grootste aannemelijkheid van de rest-contrasten (REML).

Hoofdstuk 9 beschrijft het gebruik van variantiecomponenten bij het voorspellen van stochastische effecten. De in de veefokkerij op grote schaal gebruikte beste lineaire zuivere voorspeller (BLUP) van stochastische effecten, wordt in de plantenveredeling nog lang niet in die mate toegepast. Hoe deze methode kan worden gebruikt bij het rassenonderzoek wordt beschreven in de laatste paragraaf 9.2.

## CURRICULUM VITAE

Léon Robert Verdooren werd op 10 januari 1936 geboren te Soerabaja in het toenmalige Nederlands-Indië. Hij behaalde in 1952 het eindexamen HBS-B aan het St. Canisius College te Nijmegen. In september 1952 begon hij zijn studie aan de Landbouwhogeschool te Wageningen. In september 1958 behaalde hij het examen van Landbouwkundig Ingenieur (met lof) in de richting Cultuurtechniek met als afstudeervakken de Cultuurtechiek, de Wiskundige Statistiek en de Weg- en Waterbouwkunde.

Tijdens zijn studie was hij vanaf mei 1957 aangesteld als student-assistent bij de afdeling Wiskunde LH. Na zijn afstuderen verbleef hij daar als medewerker tot februari 1962. Hij werd toen benoemd tot hoofd van de afdeling Wiskundige Statistiek van het Instituut voor Rassenonderzoek van Landbouwgewassen (IVRO) te Wageningen. In augustus 1969 keerde hij terug naar de afdeling Wiskunde LH als Wetenschappelijk Hoofdmedewerker. Sinds januari 1986 is hij Universitair Hoofddocent in de Statistische Proeftechniek bij de Vakgroep Wiskunde.