ORTHOGONALITY IN ANALYSIS OF VARIANCE

By G. A. F. Seber[1]

University of Manchester

1. Introduction. Although the literature on the subject of analysis of variance is extensive (c.f. Plackett [7]) and goes back a long way, a recent paper by Darroch and Silvey [1] throws a fresh light on the idea of orthogonality which is usually associated with analysis of variance models.

Suppose we have a general linear model $\mathcal{G}:\mathbf{y} = \boldsymbol{\theta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is a random vector distributed as $N[\mathbf{0}, \sigma^2 I_n]$ and $\boldsymbol{\theta}$, the vector of means, belongs to Ω, a subspace of n-dimensional Euclidean space R^n. Consider a sequence of linear hypotheses $\mathcal{H}_i:\boldsymbol{\theta}$ belongs to ω_i a subspace of Ω $(i = 1, 2, \cdots, K)$. Then from Darroch and Silvey [1] we have the following definition: an experimental design is orthogonal relative to a general linear model $\mathcal{G}$ and linear hypotheses $\mathcal{H}_1, \mathcal{H}_2, \cdots, \mathcal{H}_K$ if the subspaces $\Omega, \omega_1, \omega_2, \cdots, \omega_K$ satisfy the conditions $\omega_i^\perp \cap \Omega \perp \omega_j^\perp \cap \Omega$ for all $i, j, i \neq j$, i.e. if the orthogonal complements of the ω_i with respect to Ω are mutually perpendicular. This definition expresses in general terms the well known orthogonality property of analysis of variance models; namely, that the sums of squares obtained by nesting the hypotheses are stochastically independent and are the same irrespective of the order of the nesting (c.f. Scheffé [11] and Kempthorne [2] p. 49).

In this paper we shall derive necessary and sufficient conditions for a general p-factor analysis of variance model, with unequal observations per cell, to be orthogonal.

2. Matrix conditions for orthogonality. A vector space can be represented in two ways: either as the null space $\mathfrak{N}[A]$ of a matrix A, i.e. $\Omega = \{\boldsymbol{\theta} \mid A\boldsymbol{\theta} = \mathbf{0}\}$ or as the range space $\mathfrak{R}[X]$ of a matrix X, i.e. $\boldsymbol{\theta}$ belongs to Ω if and only if there exists $\boldsymbol{\alpha}$ such that $\boldsymbol{\theta} = X\boldsymbol{\alpha}$. When X is of full rank we have the familiar regression model, while if X is not of full rank and the elements of X are either zero or one, we have the analysis of variance model in which $\boldsymbol{\alpha}$ is not unique and identifiability conditions $H\boldsymbol{\alpha} = \mathbf{0}$ say, are introduced (c.f. Scheffé [11] p. 17). Although the range space representation is the more familiar one, the identifiability conditions can cause theoretical difficulties and so it is often easier to use the null space representation, as shown in Rao [9], Roy and Roy [10] and in the theory below. We shall require the following lemma.

Lemma. *If*

$$\Omega = \{\boldsymbol{\theta} \mid A\boldsymbol{\theta} = \mathbf{0}\}, \qquad \omega_i = \{\boldsymbol{\theta} \mid A\boldsymbol{\theta} = \mathbf{0}, A_i\boldsymbol{\theta} = \mathbf{0}\}$$

where the rows of the matrix $[A' \vdots A_i']'$ *are linearly independent*, $(i = 1, 2, \cdots, K)$, *and* $AA_i' = 0$ *for* $i = 1, 2, \cdots, K$, *then* $\omega_i^\perp \cap \Omega \perp \omega_j^\perp \cap \Omega$ *if and only if* $A_iA_j' = 0$.

Received 29 July 1963.

[1] Now at London School of Economics.

PROOF. Since the null space of any matrix is the orthogonal complement of the range space of its transpose

$$\omega_i^\perp = \mathcal{R}[A' \vdots A_i'].$$

Thus

$$\begin{aligned} \omega_i^\perp \cap \Omega &= \mathcal{R}[A' \vdots A_i'] \cap \mathfrak{N}[A], \\ &= (\mathcal{R}[A'] \oplus \mathcal{R}[A_i']) \cap \mathfrak{N}[A], \\ &= (\mathcal{R}[A'] \cap \mathfrak{N}[A]) \oplus (\mathcal{R}[A_i'] \cap \mathfrak{N}[A]), \end{aligned}$$

where "$\oplus$" denotes "direct sum" as used in vector analysis. Now $\mathcal{R}[A_i'] \perp \mathcal{R}[A']$ and hence $\mathcal{R}[A_i'] \subset \mathfrak{N}[A]$. Also $\mathcal{R}[A'] \cap \mathfrak{N}[A] = (\mathfrak{N}[A])^\perp \cap \mathfrak{N}[A] = \mathbf{0}$. Thus $\omega_i^\perp \cap \Omega = \mathcal{R}[A_i']$ and $\omega_i^\perp \cap \Omega \perp \omega_j^\perp \cap \Omega$ if and only if $A_i A_j' = 0$. This proves the lemma.

3. The two factor analysis of variance. Consider a two factor analysis of variance with unequal observations per cell and the two factors at I and J levels respectively. Suppose we have K_{ij} observations per cell. This gives us the model

$$y_{ijk} = \theta_{ijk} + \epsilon_{ijk} \qquad \text{for } i = 1, 2, \cdots, I;\ j = 1, 2, \cdots, J;$$

$k = 1, 2, \cdots, K_{ij}$, and the ϵ_{ijk} are all distributed independently as $N(0, \sigma^2)$. Let

$$K_{i\cdot} = \sum_{j=1}^{J} K_{ij}, \qquad K_{\cdot j} = \sum_{i=1}^{I} K_{ij} \quad \text{and} \quad K_{\cdot\cdot} = \sum_{i=1}^{I} \sum_{j=1}^{J} K_{ij}.$$

Let

$$\begin{aligned} \mu_{ijk} &= \theta_{ijk} - \bar{\theta}_{ij\cdot}, \\ \gamma_{ij} &= \bar{\theta}_{ij\cdot} - \bar{\theta}_{i\cdot\cdot} - \bar{\theta}_{\cdot j\cdot} + \bar{\theta}_{\cdots}, \\ \alpha_i &= \bar{\theta}_{i\cdot\cdot} - \bar{\theta}_{\cdots} \\ \beta_j &= \bar{\theta}_{\cdot j\cdot} - \bar{\theta}_{\cdots}, \end{aligned}$$

where

$$\bar{\theta}_{ij\cdot} = \sum_{k=1}^{K_{ij}} \theta_{ijk}/K_{ij}, \qquad \bar{\theta}_{i\cdot\cdot} = \sum_{j=1}^{J} \sum_{k=1}^{K_{ij}} \theta_{ijk}/K_{i\cdot},$$

$$\bar{\theta}_{\cdots} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K_{ij}} \theta_{ijk}/K_{\cdot\cdot} \text{ etc.}$$

Thus we have imposed identifiability constraints $\sum_{i=1}^{I} \alpha_i K_{i\cdot}/K_{\cdot\cdot} = 0$, $\sum_{j=1}^{J} \beta_j k_{\cdot j}/K_{\cdot\cdot} = 0$, etc. In general we are interested in testing the following hypotheses $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$, where

$$\mathcal{G}: \mu_{ijk} = 0$$
$$\mathcal{H}_1: \mu_{ijk} = 0, \quad \gamma_{ij} = 0 \quad \text{(i.e. interactions zero)},$$
$$\mathcal{H}_2: \mu_{ijk} = 0, \quad \alpha_i = 0 \quad \text{(i.e. row effects zero)},$$
$$\mathcal{H}_3: \mu_{ijk} = 0, \quad \beta_j = 0 \quad \text{(i.e. column effects zero)},$$

and we wish to know under what conditions the above design is orthogonal with respect to $\mathcal{G}$, $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$. In the past it has usually been the practice to work in terms of the parameters $\bar{\theta}_{\cdots}$, α_i, β_j and γ_{ij} with the range space representation

$$\mathcal{G}: \bar{\theta}_{ij\cdot} = \bar{\theta}_{\cdots} + \alpha_i + \beta_j + \gamma_{ij}.$$

However when we endeavour to find conditions for orthogonality with this representation we run into difficulties with the identifiability constraints on the α_i, β_j and γ_{ij}. Although the identifiability conditions can be used to eliminate some of the parameters, and thus simplify the problem, this is not practicable for more general p-factor analysis of variance models. Thus we find it easier to work with the θ's i.e. the null space representation, as demonstrated by the following theorem.

THEOREM 1. *The design* $\mathcal{G}$, $\mathcal{H}_1$, $\mathcal{H}_2$, $\mathcal{H}_3$ *is orthogonal if and only if* $K_{ij} = K_{i\cdot}K_{\cdot j}/K_{\cdot\cdot}$ *for all* i, j.

PROOF. Let $\boldsymbol{\theta}$ be the column vector with elements θ_{ijk}; we can express the set of hypotheses as follows:

$$\mathcal{G}: A\boldsymbol{\theta} = \mathbf{0} \quad \text{and} \quad \mathcal{H}_u: A\boldsymbol{\theta} = \mathbf{0}, \quad A_u\boldsymbol{\theta} = \mathbf{0} \quad \text{for } u = 1, 2, 3.$$

For example we wish to express the conditions

$$\mu_{ijk} = \theta_{ijk} - \sum_{k=1}^{K_{ij}} \theta_{ijk}/K_{ij} = 0, \tag{3.1}$$

where $i = 1, 2, \cdots, I; j = 1, 2, \cdots, J; k = 1, 2, \cdots, K_{ij}$ in the form $A\boldsymbol{\theta} = \mathbf{0}$. The matrix A would be $K_{\cdot\cdot} \times K_{\cdot\cdot}$ and the row corresponding to Equation (3.1) would have the (r_0, s_0, t_0) element of the form

$$\delta_{ir_0}\delta_{js_0}\delta_{kt_0} - \delta_{ir_0}\delta_{js_0}/K_{r_0s_0}, \tag{3.2}$$

where δ_{ab} is the Kronecker delta. Similarly the row of A_1 corresponding to $\gamma_{i_1j_1} = 0$ has its (r_1, s_1, t_1) element as

$$\delta_{i_1r_1}\delta_{j_1s_1}/K_{r_1s_1} - \delta_{i_1r_1}/K_{r_1\cdot} - \delta_{j_1s_1}/K_{\cdot s_1} + 1/K_{\cdot\cdot}. \tag{3.3}$$

The (r_2, s_2, t_2) element of row $\alpha_{i_2} = 0$ for matrix A_2 is given by

$$\delta_{i_2r_2}/K_{r_2\cdot} - 1/K_{\cdot\cdot} \tag{3.4}$$

and the (r_3, s_3, t_3) element of row $\beta_{j_3} = 0$ for matrix A_3 is

$$\delta_{j_3s_3}/K_{\cdot s_3} - 1/K_{\cdot\cdot}. \tag{3.5}$$

Now by multiplying Equations (3.2) and (3.3), putting $r_0 = r_1$, $s_0 = s_1$, $t_0 = t_1$ and summing on r_0, s_0, $t_0 (t_0 = 1, 2, \cdots, K_{r_0 s_0}; r_0 = 1, 2, \cdots, I; s_0 = 1, 2, \cdots, J)$ we have $A_1 A' = 0$. Similarly $A_2 A' = A_3 A' = 0$ since (3.2) is the only term containing t_0 and this summed on t_0 is zero. Thus, by the lemma, the $\mathcal{H}_u$ are orthogonal if and only if $A_u A_v' = 0$ for all u, v $u \neq v$, and we now show that these matrix conditions hold if and only if $K_{ij} = K_{i\cdot} K_{\cdot j}/K_{\cdot\cdot}$ for all i, j.

A. *Sufficiency*. Let $K_{ij} = K_{i\cdot} K_{\cdot j}/K_{\cdot\cdot}$ for all i, j; then expression (3.3) becomes

$$K_{\cdot\cdot}(\delta_{i_1 r_1}/K_{r_1\cdot} - 1/K_{\cdot\cdot})(\delta_{j_1 s_1}/K_{\cdot s_1} - 1/K_{\cdot\cdot}). \tag{3.6}$$

Thus by multiplying (3.4) and (3.6) together, putting $r_2 = r_1$ and summing on r_1, s_1, t_1 we have that $A_1 A_2' = 0$. In a similar manner it can be shown that $A_1 A_3' = A_2 A_3' = 0$. Hence the design is orthogonal relative to $\mathcal{G}$, $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$.

B. *Necessity*. Given $A_2 A_3' = 0$, we multiply (3.4) and (3.5) together, set $s_3 = s_2$ and sum on r_2, s_2, t_2. This gives us an element of $A_2 A_3'$ and therefore

$$\begin{aligned} 0 &= \sum_{r_2, s_2, t_2} (\delta_{i_2 r_2}/K_{r_2\cdot} - 1/K_{\cdot\cdot})(\delta_{j_3 s_2}/K_{\cdot s_2} - 1/K_{\cdot\cdot}), \\ &= K_{i_2 j_3}/K_{\cdot j_3} - K_{i_2\cdot}/K_{\cdot\cdot}. \end{aligned}$$

Thus the conditions $K_{ij} = K_{i\cdot} K_{\cdot j}/K_{\cdot\cdot}$ are both necessary and sufficient for orthogonality and this completes the theorem. We would mention that with the identifiability constraints given above, it is well known that the conditions $K_{ij} = K_{i\cdot} K_{\cdot j}/K_{\cdot\cdot}$ are *sufficient* for orthogonality (Scheffé [11]). However the above method of proof not only proves that the conditions are necessary but also lends itself readily to generalisation.

4. The general p-factor model. Suppose we have a p-factor analysis of variance model with the rth factor at level I_r $(r = 1, 2, \cdots, p)$ and $n_{i_1 i_2 ** i_p}$ observations per cell. Let $\mathcal{G}$, $\mathcal{H}_r$, $\mathcal{H}_{rs}$, $\cdots$, $\mathcal{H}_{12**p}$ $(r, s = 1, 2, \cdots, p, r \neq s$ etc.) be the general model, the hypotheses of no main effects, the hypotheses of no first order interactions, $\cdots$, the hypothesis of no $(p - 1)$ order interaction respectively. Then we can extend Theorem 1 as follows.

THEOREM 2. *The above p-factor model is orthogonal with respect to* $\mathcal{G}$ $\mathcal{H}_r$, $\mathcal{H}_{rs}$ *etc. if and only if*

$$n_{i_1 i_2 ** i_p} = (n_{i_1 \cdots\cdot})(n_{\cdot i_2 \cdots}) \cdots (n_{\cdots\cdot i_p})/(n_{\cdots\cdot\cdot})^{p-1} \tag{4.1}$$

for all $i_1, i_2, \cdots, i_p$ *where*

$$n_{i_1 \cdots\cdot} = \sum_{i_2=1}^{I_2} \sum_{i_3=1}^{I_3} \cdots \sum_{i_p=1}^{I_p} n_{i_1 i_2 ** i_p} \text{ etc.}$$

PROOF. As the notation is complicated we shall only outline the method of proof.

Let A, A_r, A_{rs} etc. be the corresponding matrices in the null space representation of the above hypotheses. Then, as in Theorem 1 we have

$$AA_r' = AA_{rs}' = \cdots = AA_{12**p}' = 0,$$

and by the lemma we have orthogonality if and only if the matrix products A_rA_s', A_rA_{st}', $A_{st}A_{uv}'$ etc. are zero. It is seen that Conditions (4.1) are sufficient for this to happen by factorising the product of corresponding elements in the above matrix products into terms of the form

$$(\delta_{i_1r_1}/n_{r_1}.... - 1/n.....),\ (\delta_{i_2r_2}/n._{r_2}... - 1/n.....)\ \text{etc.}$$

That the conditions are also necessary follows by an inductive argument. For example in a 4-factor model from the relations $A_1A_2' = A_{12}A_3' = A_{123}A_4' = 0$ we can prove the following equations

$$n_{i_1i_2}.. = (n_{i_1}...)(n._{i_2}..)/(n....),$$
$$n_{i_1i_2i_3}. = (n_{i_1i_2}..)(n..._{i_3})/(n....),$$
$$n_{i_1i_2i_3i_4} = (n_{i_1i_2i_3}.)(n..._{i_4})/(n....),$$

which combined give

$$n_{i_1i_2i_3i_4} = (n_{i_1}...)(n._{i_2}..)(n.._{i_3})(n..._{i_4})/(n....)^3 .$$

Thus in conclusion we see that the Conditions (4.1) are necessary and sufficient for orthogonality.

5. General remarks. We observe that the above theorem would still apply if we included the hypothesis $\mathcal{H}_0 : \theta..... = 0$ among our hypotheses to be tested. This follows from the fact that if A_0 is the corresponding matrix for the null space representation of $\mathcal{H}_0$, then A_0A_r', A_0A_{rs}' etc. are all zero because of the identifiability conditions which underly our method of defining the main effects and interactions.

It can also be shown that the variance-covariance matrix for the estimates of the unknown parameters is of a diagonal block form, i.e. for the groups of parameters $\theta.....$, α_i's, β_j's, γ_{ij}'s etc. the estimate of any one parameter of one group is uncorrelated with that of any parameter of any other group. This is the more usual way of defining orthogonality of design (Yates [13], Kempthorne [2]) and although equivalent to the definition we have adopted, it is not as fundamental as it depends on the parameters used, and therefore to some extent on the identifiability conditions rather than on the basic structure underlying the design.

One also notes that conditions similar to $K_{ij} = K_i.K._j/K..$ of Theorem 1 are derived in Plackett [6] where the design of optimum multifactorial experiments is considered. This follows from the fact that in multifactorial designs in which the design matrix has been reduced to a matrix of full rank by a suitable transformation of parameters (Plackett [8]), the condition for the design to be optimum (optimum in the sense that certain contrasts can be estimated with equal maximum precision) is that the variance-covariance matrix is proportional to the unit matrix and not just of diagonal block form as required by orthogonality. Thus one would expect to obtain the same kind of necessary conditions as given above.

The literature on optimum designs is considerable (e.g. Kiefer [4]) and various criteria for optimality have been suggested (Kiefer [3]). However one feels intuitively that the role of orthogonality plays an important part in the structure of optimum designs as is illustrated for example by the Latin square (Wald [12]).

REFERENCES

[1] DARROCH, J. N. and SILVEY, S. D. (1963). On testing one or more hypothesis. *Ann. Math. Statist.* **34** 555–567.

[2] KEMPTHORNE, O. (1952). *Design and Analysis of Experiments.* Wiley, New York.

[3] KIEFER, J. (1958). On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. *Ann. Math. Statist.* **29** 675–699.

[4] KIEFER, J. (1961). Optimum designs in regression problems, II. *Ann. Math. Statist.* **32** 298–325.

[5] PLACKETT, R. I. and BURMAN, J. P. (1946). Design of optimum multifactorial experiments. *Biometrika* **33** 305–325.

[6] PLACKETT, R. L. (1946). Some generalisations in the multifactorial design. *Biometrika* **33** 328–332.

[7] PLACKETT, R. L. (1960). Models in the analysis of variance. *J. Roy. Statist. Soc. Ser. B.* **22** 195–217.

[8] PLACKETT, R. L. (1960). *Regression Analysis.* Oxford University Press.

[9] RAO, C. R. (1959). Some generalisations in the multifactorial design. *Biometrika* **46** 49–58.

[10] ROY, S. N. and ROY, J. (1959). A note on a class of problems in normal multivariate analysis of variance. *Ann. Math. Statist.* **30** 577–580.

[11] SCHEFFÉ, H. (1959). *The Analysis of Variance.* Wiley, New York.

[12] WALD, A. (1943). On the efficient design of statistical investigations. *Ann. Math. Statist.* **14** 134–140.

[13] YATES, F. (1933). The principles of orthogonality and confounding in replicated experiments. *J. Agric. Science.* **23** Part I 108–145.