ANALYSIS OF VARIANCE OF DEPENDENT DATA

by

Bruno Baldessari
Department of Statistics
University of North Carolina
and
Universita di Roma

Institute of Statistics Mimeo Series No. 467

March 1966

CONTENTS:

This research was supported by the Mathematics Division of The Air Force Office of Scientific Research Contract No. AF-AFPSR-760-65

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
Chapel Hill, N. C.

1. Introduction and notation

The aims of this article are: a), show that analysis of variance (ANOVA) can be performed also when the elements of the "sample" are dependent and b), establish the more general type of dependence which allows us to do an ANOVA.

To make these statements precise we need some definition and notation.

Definition: By a "sample of n dependent elements" we mean a random sample of 1 element drawn from the r. vt. (random vector) $X^* = (X^*_1, \ldots, X^*_n)'$ in which the n r.v. (random variables) X^*_i, $i = 1, \ldots, n$, may be dependent [3].

We will denote this random sample by the r. vt. $X = (X_1, \ldots, X_n)'$ and it is clear that the d.f. (distribution function) of X is the same as that of X^* so that the dependence of the n r.v. $X_1, \ldots, X_n$ is defined by the dependence of the r. v. $X^*_1, \ldots, X^*_n$. If the r. vt. X^* is $N(\mu,V)$, (multivariate normal with mean vector μ and variance-covariance matrix V), where V is p.d., (positive definite), then X is $N(\mu,V)$ and the dependence of $X_1, \ldots, X_n$ is specified by V.

In this article we extend the validity of ANOVA to the case in which X is $N(\mu,V)$ so that the elements of the sample may be dependent. More precisely, we extend the validity of the methods of the classical ANOVA, (in which X is suppose to be $N(\mu,\alpha I)$, $\alpha > 0$, I is the n x n identity matrix) built up to test q null hypotheses $H_1, \ldots, H_q$ which specifies q distinct linear relations on the parameters of the general linear model of the data X, against the q ulternative hypothesis: "not H_1", ..., "not H_q". Of the mean vector μ we require, only, that the set of all a priori possible vectors μ contains at least one vector μ^* of the form : $\mu^* = \mu \mathbf{1}$, where $\mathbf{1}$ is the n-vector: $\mathbf{1} = (1, \ldots, 1)'$. We will supposethat $H_1, H_2, \ldots H_q$ are such that the total sum squares $X'(I - n^{-1}U)X$, $(U = 1\,1')$, can be decompose into $(q+1)$ positive semidefinite quadratic forms: $X' S_o X$, $X' S_1 X$, ..., $X' S_q X$, such that:

a). $\sum_{j=0}^{q} S_j = I - n^{-1} U$;

b). $S_j^2 = S_j$, $j=0, \ldots, q$;

c). $X' S_j X$ is distributed like $\alpha\, \chi'^2 (n_j, 2^{-1} \mu' S_j \mu)$, (noncentral chi-square with n_j degree of freedom and noncentrality parameter $2^{-1} \mu' S_j \mu$), where n_j = rank (S_j), $j = 0, \ldots, q$;

d). $X' S_o X, \ldots, X' S_q X$, are mutually independent;

e). $\dfrac{X' S_j X}{n_j} \; \dfrac{n_o}{X' S_o X}$ is distributed like

$F'(n_j, n_o, \mu' S_j \mu, \mu' S_o \mu)$, (Snedecor's double noncentral F distribution with n_j and n_o degrees of freedom and noncentrality parameters $\mu' S_j \mu$, $\mu' S_o \mu$), $j = 0, \ldots, q$ and for every μ.

Actually, if X is $N(\mu, \alpha I)$ then conditions a) and b) imply conditions c) and d) which imply condition e).

In the present extension we regard the matrices $S_o, \ldots, S_q$ as fixed; typically, if X is $N(\mu, \alpha I)$ the matrices S_j, $j = 0, \ldots, q$, will be such that $X' S_j X$, $j = 0, \ldots, q$, are, in some sense, the "best" statistics to use in order to test $H_1, \ldots, H_q$ against these alternatives.

We suppose that the fixed matrices $S_o, \ldots, S_q$, satisfy conditions

a) and b) and our generalization of ANOVA consists in the fact that <u>we establish the set of ALL p.d. variance-covariance matrices V such that, if X is $N(\mu,V)$ then conditions c), d), e) are also still satisfied</u>. In fact this set of matrices is the set of p.d. matrices V of structure:

$$V^* = 2^{-1}(A+A') + \alpha(I-U), \tag{1}$$

where the n x n matrix A is:

$$A = \begin{pmatrix} a_1, & \dots, & a_1 \\ \dots & \dots & \dots \\ a_n, & \dots, & a_n \end{pmatrix} ,$$

with : $a_i > 0$, $i = 1, \dots, n$.

The proof of this fact is based on two theorems which actually show that if X is $N(\mu,V)$ and if a) and b) are satisfied then, Theorem I): conditions c) and d) are equivalent to (1), and, (Theorem 2): conditions d) and e) are equivalent to (1). We also note that, if X is $N(\mu,V)$, formula (1) is equivalent to saying that $X'(I - n^{-1}U)X$ is distributed as $\alpha\,\chi'^2(n-1, 2^{-1}\mu'[I-n^{-1}U]\mu)$, as follows immediately from the proof of Proportion 1 of [3].

In the following we write $\langle A \rangle \Leftrightarrow \langle B \rangle$ to mean that the statement $\langle A \rangle$ is equivalent to the statement $\langle B \rangle$, and $\langle D\{x\} = D\{Y\} \rangle$ for the statement "the distribution of the r.v. X is the same as the distribution of the r.v. Y."

To avoid cumbersome formulas we will write $\langle I\{X'S_jX\} \rangle$ for the statement "the r.v. $X'S_jX$, $j = 0, \dots, q$ are mutually independent and, also: S for $I-n^{-1}U$; χ'^2_{μ,S_j} for $\chi'^2(n_j, 2^{-1}\mu' S_j \mu)$; F'_{μ,S_j,S_o} for $F'(n_j, n_o, \mu' S_j \mu, \mu' S_o \mu)$.

2. Conditions c), d) and matrix V.

Theorem 1. If X is $N(\mu,V)$ and matrices $S_o, \ldots, S_q$ satisfy a) and b), then:

$$< V=V^* > \iff < D\{X' S_j X\} = D\{\alpha \chi'^2_{\mu,S_j}\},\ j = 0, \ldots, q, \text{ and } I\{X' S_j X\} > \qquad (2)$$

Proof of the implication: $\Leftarrow$.

Conditions a) and b) imply that: $\sum_{j=0}^{q} \text{rank } (S_j) = \sum_{j=0}^{q} n_j =$ rank $(S) = n - 1$, so that we have:

$$< D\{X' S_j X\} = D\{\alpha X'^2_{\mu,S_j}\},\ j = 0, \ldots, q, \text{ and } I\{X' S_j X\} > \Rightarrow$$

$$< D\{X'(\sum_{j=0}^{q} S_j)X\} = D\{\alpha X'^2_{\mu,\Sigma S_j}\} > \iff < D\{X' SX\} = D\{\alpha X'^2_{\mu,S}\} > \iff$$

$$< V = V^* > .$$

The last equivalence follows from the proof of Proportion 1 of [3].

Proof of the implication: $\Rightarrow$.

First we prove that:

$$< V = V^* > \Rightarrow < D\{X' S_j X\} = D\{\alpha \chi'^2_{\mu,S_j}\},\ j = 0, \ldots, q> .$$

In fact, from the proof of Proportion 1 of [3], Corollary 5.1 of [5], and from condition a) we have:

$$< V = V^* > \iff < D\{X' SX\} = D\{\alpha \chi'^2_{\mu,S}\} > \iff$$

$$< \alpha^{-1} S V S = S > \iff < \alpha^{-1}(S_o + \ldots + S_q)V(S_o + \ldots + S_q) = S_o + \ldots + S_q >$$

so that, from condition b) we have:

$$< V = V^{*} > \;\Rightarrow\; < \alpha^{-1} S_j (S_o + \ldots + S_q)\, V\, (S_o + \ldots + S_q)\, S_j = S_j (S_o + \ldots + S_q) S_j > \;\Leftrightarrow$$

$$< \alpha^{-1} S_j\, V S_j = S_j,\; j=0,\ldots,q > \;\Leftrightarrow\; < D\{X' S_j X\} = D\{\alpha\, \chi'^{2}_{\mu, S_j}\},\; j=0,\ldots,q > .$$

Now we prove that:

$$< V = V^{*} > \;\Rightarrow\; < I\{X' S_j X\} > ,$$

and this will complete the proof of Theorem 1.

From what we have just proved we have:

$$< V = V^{*} > \;\Leftrightarrow\; < \alpha^{-1} (S_o + \ldots + S_q)\, V\, (S_o + \ldots + S_q) = S_o + \ldots + S_q > ,$$

so that, from condition b) we have:

$$< V = V^{*} > \;\Rightarrow\; < \alpha^{-1} S_j (S_o + \ldots + S_q)\, V\, (S_o + \ldots + S_q)\, S_k = S_j (S_o + \ldots + S_q)\, S_k,\; j \neq k >$$

$$< S_j\, V\, S_k = 0,\; j \neq k > \;\Leftrightarrow\; < I\{X' S_j X\} > .$$

The last equivalence follows from Craig's condition [4] in Aitken's generalization [1] from which we have that $S_j\, V\, S_k = 0$, $j \neq k$ implies that the r.v. $X' S_j X$, $j = 0, \ldots, q$ are pairwise independent. (In our particular case, pairwise independence implies mutual independence as can be seen from the relevent characteristic functions.)

3. Conditions d), e) and matrix v.

Theorem 2. If X is $N(\mu,v)$ and matrices $S_o, \ldots, S_q$, satisfy a) and b), then:

$$<V=V^{*}> \;<=>\; <D\{\frac{X^{'}S_jX}{n_j}\;\frac{n_o}{X^{'}S_oX}\} = D\{^{F^{'}}\mu,S_j,S_o\},\ j=1,\ldots,q \text{ for all } \mu, \text{and } I\{X'S_jX\}> \tag{3}$$

The proof of the implication => is obvious from (2).

Proof of the implication <=.

We may suppose $\mu = \mu^{*} = \mu 1$ because in the second statement of (3) we have "for all μ" and because of our hypothesis that in the set of all a priori possible vectors μ there is at least one μ of the form μ^{*}.

It is a simple calculation to verify that: $S\,\mu^{*} = \mu^{*}\,S = 0$. From this it follows that: $S_j\,\mu^{*} = \mu^{*}\,S_j = 0$, $j = 0, \ldots q$. In fact, by condition b) we have:

$$< S\,\mu^{*} = 0 > <=> <(S_o+ \ldots + S_q)\,\mu^{*} = 0 > => < S_j(S_o+ \ldots +S_q)\,\mu^{*} = 0,$$

$$j = 0, \ldots, q > <=> < S_j\,\mu^{*} = 0,\ j = 0, \ldots, q > ,$$

and similarly for : $< \mu^{*1}\,S_j = 0,\ j = 0, \ldots, q >$.

Now we show that, with $\mu = \mu^{*}$, and for every p.d. V, we have:

$$< I\{X^{'}S_jX\}> => < D\{\sum_{k=0}^{n_j} \lambda_{jk}\,\chi^2_{jk}\,(1)\},\ j=0, \ldots,q >, \tag{4}$$

where λ_{jk}, $j = 0, \ldots, q$, $k = 1, \ldots, n_j$, are positive constants and where χ^2_{jk} (1) is a (central) chi-square r.v. with 1 degree of freedom and, also, the chi-squares are mutually independent.

In fact, let $\varphi(t_o, \ldots, t_q)$ be the characteristic function of the r. vt. $(X^1 S_o X, \ldots, X^1 S_q X)$, so that we have:

$$\varphi(t_o, \ldots, t_q) \propto \int \exp\{i \sum_{j=0}^{q} t_j x' S_j x - \tfrac{1}{2} (x-\mu^*)' V^{-1} (x-\mu^*)\}, dx ,$$

where $\int$ denotes n-fold integration and $dx = dx_1 \ldots dx_n$.

From the relations: $\mu^* S_j = S_j \mu^* = 0$, $j = 0, \ldots, q$, it follows that with the transformation $y = x - \mu^*$ we have;

$$\varphi(t_o, \ldots, t_q) \propto \int \exp \{i \sum_{j=0}^{q} t_j y' S_j y - \tfrac{1}{2} y' V^{-1} y\} d y.$$

If V is positive definite then there exists a matrix T such that: $|T| \neq 0$, $V = T T'$, and the transformation $y = Tx$ gives:

$$\varphi (t_o, \ldots, t_q) \propto \int \exp \{i \sum_{j=0}^{q} t_j x' T' S_j T x - \tfrac{1}{2} x' x\} dx .$$

The matrices $T' S_j T$, $j = 0, \ldots, q$ are symmetric and

$< I\{X' S_j X\} > \; <=> \; < S_j V S_k = 0, j \neq k >$, so that:

$$(T' S_j T) (T' S_k T) = T' S_j V S_k T = 0 = T' S_k V S_j T =$$

$(T' S_k T) (T' S_j T)$, $j \neq k$, and this implies that a matrix P exists such that: $P' P = I$, $P' S_j P = \Lambda_j$, with Λ_j diagonal, $j = 0, \ldots, q$. Then the transformation $x = P y$ shows that:

$$\varphi(t_0, \ldots, t_q) \propto \int \exp\{i \sum_{j=0}^{q} t_j y' \Lambda_j y - \frac{1}{2} y' y\} \, d y.$$

Now Λ_j is positive semidefinite so its diagonal elements are non-negative and there are exactly n_j positive elements in the diagonal of Λ_j because: $\operatorname{rank}(\Lambda_j) = \operatorname{rank}(S_j) = n_j$, $j = 0, \ldots, q$. Also Λ_j and Λ_k, $j \neq k$, do not have any positive elements in the same row: in fact, $\Lambda_j \Lambda_k = 0$. From this it follws that:

$$\varphi(t_0, \ldots, t_q) \propto \prod_{j=0}^{q} \prod_{k=1}^{n_j} (1 - 2 \lambda_{jk} i t_j)^{-\frac{1}{2}}, \qquad (5)$$

where λ_{jk} is the k-th positive element in the diagonal of Λ_j.

Since $\varphi(0, \ldots, 0) = 1$, we see that $\varphi(t_0, \ldots, t_q)$ is exactly equal to the right hand side of (5), so that formula (3) is proved.

Now we prove the implication: $\Leftarrow$. In fact, from (3) and (5) we have:

$$< D\{\frac{X' S_j X}{n_j} \; \frac{n_0}{X' S_0 X}\} = D\{F'_{\mu, S_j, S_0}\}, \; j = 1, \ldots, q, \text{ for all } \mu, \text{ and } I\{X' S_j X\}> \; \Leftrightarrow$$

$$\left\langle D \left\{ \frac{\sum_{k=1}^{n_j} \lambda_{jk} \chi^2_{jk}(1)}{\sum_{k=1}^{n_0} \lambda_{0k} \chi^2_{0k}(1)} \right\} = D\left\{\frac{n_j}{n_0} F(n_j, n_0)\right\}, \; j = 1, \ldots, q, \text{ and } I\{X' S_j X\} \right\rangle,$$

where $F(n_j, n_o)$ is a (central) Snedecor's F r.v. with n_j and n_o degrees of freedom. From that we have:

$$< D\{\frac{X'S_jX}{n_j} \frac{n_o}{X'S_oX}\} = D\{F'_{\mu,S_j,S_o}\},\ j = 1, \dots, q, \text{ for all } \mu, \text{ and } I\{X'S_jX\}> \Rightarrow$$

$$\left\langle D\left\{\frac{\sum_{k=1}^{n_j} \lambda_{jk}\ \chi^2_{jk}(1)}{\sum_{k=1}^{o} \lambda_{ok}\ \chi^2_{ok}(1)}\right\} = D\left\{\frac{\chi^2(n_j)}{\chi^2(n_o)}\right\},\ j = 1, \dots, q, \text{ and } I\{X'S_jX\}\right\rangle \Leftrightarrow$$

$$< \lambda_{j1} = \dots = \lambda_{j\,n_j} = \lambda_{o1} = \dots = \lambda_{on_o} = \alpha > 0,\ j = 1, \dots, q, \text{ and } I\{X'S_jX\}> \Leftrightarrow .$$

$$< D\{X'S_jX\} = D\{\alpha\, \chi^2(n_j)\},\ j = 0, \dots, q, \text{ and } I\{X'S_jX\} > \Rightarrow$$

$$< D\{X' \sum_{j=0}^{q} S_j\ X\} = D\{\alpha\, \chi^2(n-1)\} > \Leftrightarrow < D\{X'S\ X\} = D\{\alpha\, \chi^2(n-1)\}> \Leftrightarrow .$$

$$< V = V^* >,$$

in which the first equivalence symbol is justified by the proportion of [2].

This completes the proof of theorem 2.

A C K N O W L E D G M E N T S

The author is grateful to Professor N. L. Johnson for helpful discussions and for revising this article.

R E F E R E N C E S

[1] Aitken, A. C. Statistical independence of quadratic forms in normal varieties. Biom., 37, (1950), 93.

[2] Baldessari, B. Observation sur le rapport de combinaisons lineaires de chi-deux. Pub. de l'Institut de Statis. de l'Univ. de Paris, (1965).

[3] Baldessari, B. Observations sur les l'echantillons gaussiens dépendents, Pub. de l'Institut de Statis. de l'Univ. de Paris,(1965).

[4] Craig, A. T. Not on the independence of certain quadratic forms, Ann. Math. Stat., 14, (1943), 195.

[5] Traybill, A. F. and Marsaglia, G. Independent matrices and quadratic forms in the general linear hypothesis, Ann. Math. Stat., 28, (1957), 678.