Biometrika (1980), 67, 1, *pp.* 85–92
Printed in Great Britain

The Welch–James approximation to the distribution of the residual sum of squares in a weighted linear regression

BY SØREN JOHANSEN

Institute of Mathematical Statistics, University of Copenhagen

SUMMARY

A general discussion is given of the approximate distribution of the residual sum of squares in a linear model in which a weighted analysis is made with weights estimated empirically as the reciprocals of variance estimates. Applications to testing hypotheses are made and various generalizations indicated. The initial results simplify those of James (1951, 1954).

Some key words: Approximate distribution theory; Behrens–Fisher problem; Component of variance; Empirical weight; Linear model; Quadratic form; Residual sum of squares; Weighted analysis.

1. INTRODUCTION

If in a regression problem the variances are not equal it is common to use the reciprocal estimated variances as weights. The residual sum of squares Q has asymptotically a χ^2 distribution when the degrees of freedom tend to infinity.

Welch (1947, 1951) gave an approximation to the distribution of Q in the special case of the comparison of n means, by using a suitably chosen F distribution. James (1951, 1954) gave an improved approximation using the fractiles of a χ^2 distribution and extended the results to the general linear model.

We shall show here how the results for the general linear model can be considerably simplified by using the technique due to Welch (1951), and extend the results to multivariate models and variance component models.

Results will also be given on the variance of the fitted value, thereby extending the results of Jacquez, Mather & Crawford (1968) to the general linear model.

2. NOTATION AND MAIN RESULT

Let Y have an n-dimensional normal distribution with mean $\xi \in L_0$, a subspace of dimension $p < n$, and variance–covariance matrix $D = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_n^2)$. We assume that we have independent estimates $S = \operatorname{diag}(s_1^2, \ldots, s_n^2)$ which have χ^2 distributions with degrees of freedom $(f_1, \ldots, f_n)$ and scale parameters $(\sigma_1^2/f_1, \ldots, \sigma_n^2/f_n)$.

We want to test the hypothesis H: $\xi \in L_1$, where $L_1 \subset L_0$ is a subspace of dimension $m < p$.

The weighted least squares estimate is found for $i = 0, 1$ by minimizing $(Y-\xi)'S^{-1}(Y-\xi)$ for $\xi \in L_i$, and is given by the projection $P_i(S)$ onto L_i with respect to S^{-1}.

As a test statistic for the hypothesis H we use the residual sum of squares

$$\begin{aligned} Q &= Y'\{P_0(S)-P_1(S)\}'S^{-1}\{P_0(S)-P_1(S)\}\,Y \\ &= Y'S^{-1}\{P_0(S)-P_1(S)\}\,Y \\ &= Y'TY, \end{aligned}$$

say. We can then formulate the main result.

THEOREM 1. *Up to terms of order $1/f_i$, we have that*

$$E(Q) = p-m+2A+2B, \quad \operatorname{var}(Q) = 2(p-m)+14A+2B, \tag{1}$$

where

$$A = \Sigma_i(P_{0ii}-P_{1ii})(1-P_{1ii})/f_i, \quad B = \Sigma_i(P_{0ii}-P_{1ii})(1-P_{0ii})/f_i. \tag{2}$$

Here $P_\nu = P_\nu(D)$ $(\nu = 0, 1)$ and $P_{\nu ii}$ is the ith diagonal element of the matrix P_ν.

We first comment on this result and then give some applications and defer the generalizations to § 4.

Notice first that for testing a linear model in the full model we have that $P_0 = I$, the identity, and hence $B = 0$ and $A = \Sigma_i(1-P_{1ii})^2/f_i$. As a special case consider the comparison of n means. Welch (1951) derived (1) with $B = 0$ and $A = \Sigma_i(1-\sigma_i^{-2}/\Sigma_j\sigma_j^{-2})^2 f_i^{-1}$. This result we can easily find since the projection P_1 in this case is given by

$$\{P_1(Y)\}_i = \Sigma_j \sigma_j^{-2} Y_j / \Sigma_j \sigma_j^{-2}$$

from which we read off directly that $P_{1ii} = \sigma_i^{-2}/\Sigma_j\sigma_j^{-2}$. Thus in a sense Welch's result is true for the general linear model tested within the full model.

James (1951) considers the problem of testing n means equal to zero. In this case $P_0 = I$, $P_1 = 0$ and this gives $B = 0$ and $A = \Sigma_i 1/f_i$ which is consistent with James's result.

Notice how, in general, we can interpret the result in terms of the weighted regression with D^{-1} as weights. This gives rise to the projections P_0 and P_1 and P_{1ii}, say, is just the coefficient of Y_i in the expression for the fitted value in this regression $\{P_1(Y)\}_{ii}$.

A different interpretation can be given as follows. Define the residuals in the regression with D^{-1} as weights by $R_1 = (I-P_1)(Y)$, $R_0 = (I-P_0)(Y)$ and $R_{01} = (P_0-P_1)(Y) = R_1 - R_0$. Then, as is well known,

$$\operatorname{var}(R_1) = (I-P_1)(D), \quad \operatorname{var}(R_0) = (I-P_0)(D)$$

and

$$\operatorname{var}(R_{01}) = (P_0-P_1)(D),$$

which means that

$$A = \Sigma_i \operatorname{var}(R_{01i})\operatorname{var}(R_{1i})/[\{\operatorname{var}(Y_i)\}^2 f_i], \quad B = \Sigma_i \operatorname{var}(R_{01i})\operatorname{var}(R_{0i})/[\{\operatorname{var}(Y_i)\}^2 f_i]. \tag{3}$$

The results of Theorem 1 can be applied as Welch (1951) does to fit a distribution of the form $cF(f_1, f_2)$ such that it has the same mean and variance as Q. This is accomplished if we take

$$f_1 = p-m, \quad c = p-m+2(A+B)-6(A-B)/(p-m+2), \quad f_2 = \tfrac{1}{3}(p-m)(p-m+2)(A-B).$$

We have here used the approximations

$$E\{cF(f_1,f_2)\} = c(1-2/f_2)^{-1} = c(1+2/f_2),$$

$$\operatorname{var}\{cF(f_1,f_2)\} = c^2(1-2/f_2)^{-1}(1-4/f_2)^{-1}(1+2/f_1) - c^2(1-2/f_2)^{-2}$$

$$= 2c^2\{1+(f_1+6)/f_2\}f_1^{-1}.$$

It is difficult to compare the results of (2) with those of James (1954) because of the rather complicated form in which they are given in his paper.

The result can be formulated as follows:

$$\operatorname{pr}\{Q \leqslant \xi + h_1(\xi)\} = 1-\alpha+O(1/f_i^2),$$

where $\xi = \chi^2_{1-\alpha}(p-m)$ and $h_1(\xi)$ is of the order of $1/f_i$. From the relation

$$F_{1-\alpha}(f_1, f_2) = \chi^2_{1-\alpha}(f_1)\,[1 + \{\chi^2_{1-\alpha}(f_1) - f_1 + 2\}/(2f_2)]f_1^{-1}$$

it follows that in our notation

$$h_1(\xi) = \xi\{A + 7B + 3(A-B)\,\xi/(p-m+2)\}\{2(p-m)\}^{-1}.$$

We shall go through a few special cases to show how easily $h_1(\xi)$ is found with the present notation.

Consider, for instance, a two-way analysis of variance, where Y_{ij} has mean ξ_{ij} and variance σ^2_{ij} $(i = 1, \ldots, r;\ j = 1, \ldots, s)$. We want to test the hypothesis of an additive model, that is $\xi_{ij} = \alpha_i + \beta_j$. In order to find A and B we assume that $w_{ij} = \sigma_{ij}^{-2}$ is known and let $w_{i.} = \Sigma_j w_{ij}$, $w_{.j} = \Sigma_i w_{ij}$ and $w_{..} = \Sigma_{ij} w_{ij}$; then the estimate for ξ_{ij} is given by

$$\hat{\xi}_{ij} = \Sigma_i Y_{ij} w_{ij}/w_{.j} + \Sigma_j Y_{ij} w_{ij}/w_{i.} - \Sigma_{ij} Y_{ij} w_{ij}/w_{..}$$

which gives

$$P_{1ij,ij} = w_{ij}(1/w_{i.} + 1/w_{j.} - 1/w_{..}).$$

Since $P_0 = I$, we are testing inside the full model, we find $B = 0$ and

$$A = \Sigma_{ij}\{1 - w_{ij}(1/w_{i.} + 1/w_{.j} - 1/w_{..})\}^2/f_{ij},$$

$$h_1(\xi) = \frac{\xi}{2(r-1)(s-1)}\left\{1 + \frac{3\xi}{(r-1)(s-1)+2}\right\}\Sigma_{ij}\{1 - w_{ij}(1/w_{i.} + 1/w_{.j} - 1/w_{..})\}^2 f_{ij}^{-1}.$$

This should be compared with (4·19) of James (1954).

Another case of interest is a simple linear regression where the mean of Y_i is $\alpha + \beta t_i$ and we let σ_i^2 denote the variance. We want to test $\beta = 0$ and in order to find A and B we assume that $w_i = \sigma_i^{-2}$ is known. Then

$$\{P_0(Y)\}_i = \Sigma_j w_j Y_j/\Sigma_j w_j + (t_i - \bar{t})\,\Sigma_j w_j Y_j(t_j - \bar{t})/\text{SSD},$$

where $\bar{t} = \Sigma_j w_j t_j/\Sigma_j w_j$ and $\text{SSD} = \Sigma_j w_j(t_j - \bar{t})^2$. We also find that $\{P_1(Y)\}_i = \Sigma_j w_j Y_j/\Sigma_j w_j$. Hence

$$A = \Sigma_i w_i(t_i - \bar{t})^2(1 - w_i/\Sigma_j w_j)\,(\text{SSD} f_i)^{-1},$$

$$B = \Sigma_i w_i(t_i - \bar{t})^2\{1 - w_i/\Sigma_j w_j - (t_i - \bar{t})^2 w_i/\text{SSD}\}\,(\text{SSD} f_i)^{-1},$$

$$h_1(\xi) = \tfrac{1}{2}\xi[\Sigma_i w_i(t_i - \bar{t})^2\{8(1 - w_i/\Sigma_j w_j) - 7(t_i - \bar{t})^2 w_i/\text{SSD}\}\,(\text{SSD} f_i)^{-1} + \Sigma w_i^2(t_i - \bar{t})^4(\text{SSD}^2 f_i)^{-1}].$$

As a byproduct of the above analysis we can also obtain various other approximations. Let $P(S)$ denote the projection onto a subspace L with respect to S^{-1} and let $P = P(D)$, then we have Theorem 2.

THEOREM 2. *Let $\hat{Y} = P(S)\,Y$ denote the fitted values; then $E(\hat{Y}) = \xi$ and up to terms of order $1/f_i$, we have that*

$$\text{var}\,(\hat{Y})_{ij} = P_{ij}\sigma_j^2 + 2\Sigma_\nu P_{i\nu}(1 - P_{\nu\nu})\,P_{\nu j}\sigma_j^2/f_\nu, \tag{4}$$

$$E\{P(S)\,S\}_{ij} = P_{ij}\sigma_j^2 - 2\Sigma_\nu P_{i\nu}(1 - P_{\nu\nu})\,P_{\nu j}\sigma_j^2/f_\nu. \tag{5}$$

COROLLARY. *An estimate of* $\text{var}\,(\hat{Y})_{ij}$ *which is unbiased up to terms of order $1/f_i$ is given by*

$$P_{ij}(S)\,s_j^2 + 4\Sigma_\nu P_{i\nu}(S)\,\{1 - P_{\nu\nu}(S)\}\,P_{\nu j}(S)\,s_j^2/f_\nu. \tag{6}$$

These results generalize the formula for the variance in a simple regression model, as derived by Jacquez *et al.* (1968).

Finally we note that the expressions in Theorems 1 and 2 all depend on the true but unknown D. If we insert S in place of D, that is replace P_ν by $P_\nu(S)$, then a bias of order $1/f_i^2$ will result in A and B. To the accuracy we are working with this does not matter.

3. Proof of main result

We shall first prove the results for $D = I$ and then transform to the general case. Welch (1951) derived the first terms of the moment generating function but we shall only need the first two moments to get the desired accuracy. Thus we want to find first the mean and variance of Q given S:

$$E(Q \mid S) = E(Y'TY \mid S) = \operatorname{tr}(T), \quad \operatorname{var}(Q \mid S) = \operatorname{var}(Y'TY \mid S) = 2\operatorname{tr}(T^2). \tag{7}$$

We shall need the following approximation.

Lemma. *Let $P(S)$ be a projection on to the subspace L with respect to S^{-1}; then if $S^{-1} = I + U$, we have that*

$$P(S) = P + PU(I-P) - PUPU(I-P) + \varepsilon,$$

where $\varepsilon \in o(\|U\|^2)$.

Proof. Let $\xi \in L$ be equivalent to $\xi = X\beta$, $\beta \in R^p$, with rank $(X) = p$; then

$$\begin{aligned} P(S) &= X(X'S^{-1}X)^{-1}X'S^{-1} = X(X'X + X'UX)^{-1}X'(I+U) \\ &= X[(X'X)^{-1} - (X'X)^{-1}X'UX(X'X)^{-1} + (X'X)^{-1}\{X'UX(X'X)^{-1}\}^2 + \varepsilon]\,X'(I+U), \end{aligned}$$

from which the result follows for $P = X(X'X)^{-1}X'$.

Now we apply this result to $T = S^{-1}\{P_0(S) - P_1(S)\}$. This gives that

$$\begin{aligned} T = P_0 - P_1 &+ (P_0 - P_1)\,U(I-P_0) + (I-P_1)\,U(P_0 - P_1) \\ &- (I-P_1)\,UP_1\,U(I-P_1) + (I-P_0)\,UP_0\,U(I-P_0) + \varepsilon, \end{aligned}$$

and hence

$$\begin{aligned} \operatorname{tr}(T) = p - m + 0 &+ \Sigma_i\, U_{ii}(P_{0ii} - P_{1ii}) - \Sigma_{ij}\, U_i U_j P_{1ij}(\delta_{ji} - P_{1ji}) \\ &+ \Sigma_{ij}\, U_i U_j P_{0ij}(\delta_{ji} - P_{0ji}) + \varepsilon. \end{aligned} \tag{8}$$

Now we can find the expectation of Q as $E\{E(Q \mid S)\} = E\{\operatorname{tr}(T)\}$. Note therefore that

$$E(U_i) = E(S_i^{-1} - 1) = (1 - 2/f_i)^{-1} - 1 = 2/f_i,$$

$$E(U_i^2) = E(S_i^{-1} - 1)^2 = \{(1 - 2/f_i)(1 - 4/f_i)\}^{-1} - 2(1 - 2/f_i)^{-1} + 1 = 2/f_i$$

and that $E(U_i U_j)$ is of order $1/(f_i f_j)$ and therefore to be discarded if $i \neq j$.

Then

$$E(Q) = p - m + \Sigma_i\, 2\{P_{0ii} - P_{1ii} - P_{1ii}(1 - P_{1ii}) + P_{0ii}(1 - P_{0ii})\}/f_i = p - m + 2(A+B).$$

Similarly we find that

$$\begin{aligned} \operatorname{tr}(T^2) = \operatorname{tr}(P_0 - P_1) &+ 2\operatorname{tr}\{U(P_0 - P_1)\} + \operatorname{tr}\{U(P_0 - P_1)\,U(P_0 - P_1)\} \\ &- 2\operatorname{tr}\{UP_1\,U(P_0 - P_1)\} + 2\operatorname{tr}\{U(I - P_0)\,U(P_0 - P_1)\}, \end{aligned} \tag{9}$$

and hence

$$E\{\operatorname{var}(Q \mid S)\} = 2[p - m + \Sigma_i\, 2(P_{0ii} - P_{1ii})\{2 + P_{0ii} - P_{1ii} - 2P_{1ii} + 2(1 - P_{0ii})\}/f_i].$$

From (8) we also find that

$$\operatorname{var}\{E(Q\,|\,S)\} = \operatorname{var}\{\operatorname{tr}(T)\} = \Sigma_i\, 2(P_{0ii}-P_{1ii})^2/f_i$$

and adding these we end with

$$\operatorname{var}(Q) = E\{\operatorname{var}(Q\,|\,S)\}+\operatorname{var}\{E(Q\,|\,S)\} = 2(p-m)+14A+2B.$$

If $D \neq I$, consider the variable $\tilde{Y} = D^{-\frac{1}{2}}\, Y$ with mean $D^{-\frac{1}{2}}\,\xi$ and variance matrix I. Then let $\tilde{S} = D^{-\frac{1}{2}}\,SD^{-\frac{1}{2}}$. The projection $\tilde{P}(\tilde{S}) = D^{-\frac{1}{2}}\,P(S)\,D^{\frac{1}{2}}$ and $\tilde{P} = D^{-\frac{1}{2}}\,PD^{\frac{1}{2}}$ or $\tilde{P}_{ij} = \sigma_i^{-1}\,P_{ij}\,\sigma_j$ and in particular the diagonal elements are not changed. It is seen that Q as well as A and B are invariant under this transformation.

To prove Theorem 2 on $\operatorname{var}\{P(S)\,Y\}$ note that

$$\operatorname{var}\{P(S)\,Y\} = E[\operatorname{var}\{P(S)\,Y\,|\,S\}]+\operatorname{var}[E\{P(S)\,Y\,|\,S\}] = E\{P(S)\,P(S)'\}+0,$$

which by the Lemma can be written $E\{P+PU(I-P)\,UP+\ldots\}$, and hence

$$\operatorname{var}\{P(S)\,Y\}_{ij} = P_{ij}+2\Sigma_\nu\, P_{i\nu}(1-P_{\nu\nu})\,P_{\nu j}/f_\nu,$$

which shows (4) for the case $D = I$.

Finally, the Lemma applied to $P(S)\,S$ gives that $E\{P(S)\,S\} = E\{P-PUP+PUPUP+\ldots\}$, and hence

$$E\{P_{ij}(S)\,s_j^2\} = P_{ij}-2\Sigma_\nu\, P_{i\nu}(1-P_{\nu\nu})\,P_{\nu j}/f_\nu,$$

which proves (5). The general case $D \neq I$ is solved as before by applying (4) and (5) to the variables $\tilde{Y}$ and $\tilde{S}$. Finally, the corollary follows easily from Theorem 2.

4. Generalizations

The methods above give a much more general result. Let us first define $Z = S-I$; then $U = S^{-1}-I = (I+Z)^{-1}-I = -Z+Z^2+\ldots$. We can now collect the results of (7)–(9), and we find that to the order of approximation we need

$$E(Q) = p-m+E[\operatorname{tr}\{Z(P_0-P_1)\,Z(P_0-P_1)\}]+2E[\operatorname{tr}\{Z(P_0-P_1)\,Z(I-P_0)\}], \tag{10}$$

$$\operatorname{var}(Q) = 2(p-m)+E[\operatorname{tr}\{Z(P_0-P_1)\}]^2+6E[\operatorname{tr}\{Z(P_0-P_1)\,Z(P_0-P_1)\}] + 8E[\operatorname{tr}\{Z(P_0-P_1)\,Z(I-P_0)\}]. \tag{11}$$

This result does not depend on any particular form of the estimate S, only on the fact that certain moments exist.

We consider two generalizations. First let $V_1, \ldots, V_k$ denote an orthogonal decomposition of R^n with dimensions $m_1, \ldots, m_k$ and projections $Q_1, \ldots, Q_k$. Consider the family of distributions of Y on R^n given by the normal distribution with mean $\xi \in L_0$ a subspace of R^n and variance $\Gamma = \Gamma_1+\ldots+\Gamma_k$, where Γ_i is an arbitrary covariance matrix giving rise to a normal distribution with support on V_i. Note that this implies that $Q_i\,\Gamma_i\,Q_j = \delta_{ij}\,\Gamma_i$. We want to test the hypothesis that $\xi \in L_1 \subset L_0$.

Further we let S_i denote an estimate of Γ_i which has a Wishart distribution of the form $W_{m_i}(f_i, \Gamma_i)/f_i$. In this situation we can complete the reduction of the expressions (10) and (11) by means of the following Lemma.

Lemma. *Let S have the distribution $W_p(f, \Sigma)/f$ and let M and N denote symmetric $p\times p$ matrices; then*

$$E[\operatorname{tr}\{(S-\Sigma)\,M(S-\Sigma)\,N\}] = E[\operatorname{tr}\{(S-\Sigma)\,M\}\operatorname{tr}\{(S-\Sigma)\,N\}] = \{\operatorname{tr}(M\Sigma N\Sigma)+\operatorname{tr}(M\Sigma)\operatorname{tr}(N\Sigma)\}/f.$$

Proof. Let $U_1, \ldots, U_f$ be independently normally distributed with mean 0 and covariance matrix Σ, then

$$S = \sum_{i=1}^{f} U_i U_i'/f$$

and

$$\begin{aligned}
E[\operatorname{tr}\{(S-\Sigma) M (S-\Sigma) N\}] &= \Sigma_i \Sigma_j E[\operatorname{tr}\{(U_i U_i' - \Sigma) M (U_j U_j' - \Sigma)\} N/f^2] \\
&= E[\operatorname{tr}\{(U_1 U_1' - \Sigma) M (U_1 U_1' - \Sigma)\} N/f] \\
&= \{E(U_1' M U_1 U_1' N U_1) - \operatorname{tr}(\Sigma M \Sigma N)\}/f \\
&= \{\operatorname{tr}(\Sigma M \Sigma N) + \operatorname{tr}(\Sigma M)\operatorname{tr}(\Sigma N)\}/f.
\end{aligned}$$

Similarly

$$\begin{aligned}
E[\operatorname{tr}\{(S-\Sigma) M\}\operatorname{tr}\{(S-\Sigma) N\}] &= \Sigma_i \Sigma_j E[\operatorname{tr}\{(U_i U_i' - \Sigma) M\}\operatorname{tr}\{(U_j U_j' - \Sigma) N\}]/f^2 \\
&= E[\operatorname{tr}\{(U_1 U_1' - \Sigma) M\}\operatorname{tr}\{(U_1 U_1' - \Sigma) N\}]/f \\
&= \{E(U_1' M U_1 U_1' N U_1) - \operatorname{tr}(\Sigma M \Sigma N)\}/f
\end{aligned}$$

as was to be proved.

From this Lemma it follows that if $S = \Sigma_i S_i$ then $Z = S - I = \Sigma_i(S_i - Q_i)$ and it follows that the result of Theorem 1 holds with

$$A = \tfrac{1}{2}\Sigma_i[\operatorname{tr}\{(P_0 - P_1) Q_i (I - P_1) Q_i\} + \operatorname{tr}\{(P_0 - P_1) Q_i\}\operatorname{tr}\{(I - P_1) Q_i\}]/f_i, \tag{12}$$

$$B = \tfrac{1}{2}\Sigma_i[\operatorname{tr}\{(P_0 - P_1) Q_i (I - P_0) Q_i\} + \operatorname{tr}\{(P_0 - P_1) Q_i\}\operatorname{tr}\{(I - P_0) Q_i\}]/f_i. \tag{13}$$

As a simple example suppose that $Y_1, \ldots, Y_k$ have independent p-dimensional normal distributions with means $\xi_1, \ldots, \xi_k$ and covariance matrices $\Gamma_1, \ldots, \Gamma_k$. We want to test the hypothesis that $\xi_1 = \ldots = \xi_k$.

The estimate of the common mean is $\hat{\xi} = (\Sigma_i \Gamma_i^{-1})^{-1} \Sigma_i \Gamma_i^{-1} \xi_i$. To put it into the above framework let $R^{kp} = R^p \times \ldots \times R^p$ and let Q_i denote the projection into the ith copy of R^p considered as a subspace of R^{kp}. The projection $P(\Gamma)$ can be represented as a block matrix with (s,t)th block equal to $W^{-1} W_t$, where $W_t = \Gamma_t^{-1}$ and $W = \Sigma_t W_t$ $(s, t = 1, \ldots, k)$. Similarly the (s,t)th block of Q_i is $I_{p\times p}$ if $s = t = i$ and is zero otherwise. Hence we find for $P_0 = I$, that $B = 0$ and

$$\begin{aligned}
A &= \tfrac{1}{2}\Sigma_i(\operatorname{tr}[\{I - P(\Gamma)\} Q_i]^2 + [\operatorname{tr}\{I - P(\Gamma)\} Q_i]^2)/f_i \\
&= \tfrac{1}{2}\Sigma_i[\operatorname{tr}(I - W^{-1} W_i)^2 + \{\operatorname{tr}(I - W^{-1} W_i)\}^2]/f_i,
\end{aligned}$$

which is consistent with the result of James (1954).

As the final generalization we consider the case where $\Gamma_i = \sigma_i^2 Q_i$, and where σ_i^2 is estimated by s_i^2 which has a χ^2 distribution with f_i degrees of freedom and a scale parameter σ_i^2/f_i. Also let $s_1^2, \ldots, s_k^2$ be independent. Then we find that $Z = S - I = \Sigma_i(s_i^2 - 1) Q_i$ and that

$$\begin{aligned}
E[\operatorname{tr}\{(S-I) M (S-I) N\}] &= \Sigma_{i,j}\operatorname{tr}(Q_i M Q_j N) E\{(s_i^2 - 1)(s_j^2 - 1)\} \\
&= \Sigma_i 2\operatorname{tr}(Q_i M Q_i N)/f_i, \\
E[\operatorname{tr}\{(S-I) M\}\operatorname{tr}\{(S-I) N\}] &= \Sigma_{i,j}\operatorname{tr}(Q_i M)\operatorname{tr}(Q_j N) E\{(s_i^2 - 1)(s_j^2 - 1)\} \\
&= \Sigma_i 2\operatorname{tr}(Q_i M)\operatorname{tr}(Q_i N)/f_i.
\end{aligned}$$

Hence we find that

$$E(Q) = p - m + 2A + 2B, \quad \operatorname{var}(Q) = 2(p-m) + 14A + 2B + 2C,$$

where

$$A = \Sigma_i[\mathrm{tr}\{Q_i(P_0-P_1)Q_i(I-P_1)\}]/f_i, \quad B = \Sigma_i[\mathrm{tr}\{Q_i(P_0-P_1)Q_i(I-P_0)\}]/f_i,$$

$$C = \Sigma_i([\mathrm{tr}\{Q_i(P_0-P_1)\}]^2 - \mathrm{tr}\{Q_i(P_0-P_1)\}^2)/f_i.$$

As an application of this result consider the balanced incomplete block design given in Table 1(a). There are three subspaces of the 6-dimensional observations space which are of interest. These spaces are spanned by vectors of the form given in Table 1(b).

Table 1. *Analysis of a balanced incomplete block design*

(a) Design and observations

		Treatment		
		A	B	C
	1	Y_1	Y_2	
Block	2	Y_3		Y_4
	3		Y_5	Y_6

(b) The error strata and treatment space

Block space			Plot space			Treatment space		
α	α		α	$-\alpha$		α	β	
β		β	β		$-\beta$	α		γ
	γ	γ		γ	$-\gamma$		β	γ

(c) The projection onto the treatment space

$$\begin{array}{cccccc}
\frac{1}{2} & \phi & \frac{1}{2} & \phi & -\phi & -\phi \\
\phi & \frac{1}{2} & -\phi & -\phi & \frac{1}{2} & \phi \\
\frac{1}{2} & \phi & \frac{1}{2} & \phi & -\phi & -\phi \\
-\phi & -\phi & \phi & \frac{1}{2} & \phi & \frac{1}{2} \\
\phi & \frac{1}{2} & -\phi & -\phi & \frac{1}{2} & \phi \\
-\phi & -\phi & \phi & \frac{1}{2} & \phi & \frac{1}{2}
\end{array}$$

$$\phi = \tfrac{1}{2}(\sigma^2-\tau^2)/(3\tau^2+\sigma^2).$$

The projections onto the block space and plot space Q_b and Q_p are best described as block diagonal matrices with 2×2 matrices along the diagonal. These are

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

respectively. Thus the covariance matrix is

$$\Gamma = \tau^2 Q_b + \sigma^2 Q_p, \quad \sigma^2 > 0, \quad \tau^2 > 0,$$

and the projection with respect to $\Gamma^{-1} = \tau^{-2}Q_b + \sigma^{-2}Q_p$ onto the treatment space L_0 is given in Table 1(c).

If we want to test the hypothesis of no treatment effects, we also have to project onto the space L_1 spanned by the vector of 1's. Since L_1 is contained in the blockspace the projection can be written $\{P_1(Y)\}_i = \bar{Y}$. Now let s_1^2 and s_2^2 be estimates of τ^2 and σ^2 with f_1 and f_2 degrees of freedom. Then we can test the hypothesis of no treatment effect using the information from both strata by means of the statistic Q. Some calculations show that in this case the corrections to the mean and variance are given by

$$A = (2\sigma^2/f_1 + 6\tau^2/f_2)(3\tau^2+\sigma^2)^{-1}, \quad B = (1/f_1 + 1/f_2)\, 6\tau^2\sigma^2(3\tau^2+\sigma^2)^{-2},$$
$$C = 2(\sigma^4/f_1 + 9\tau^4/f_2)(3\tau^2+\sigma^2)^{-2}.$$

It is a pleasure to thank Anders Hald for useful comments on a first draft of the paper.

REFERENCES

JACQUEZ, J. A., MATHER, F. J. & CRAWFORD, C. R. (1968). Linear regression with non-constant, unknown error variances: Sampling experiments with least squares, weighted least squares and maximum likelihood estimators. *Biometrics* **24**, 607–26.

JAMES, G. S. (1951). The comparison of several groups of observations when the ratios of the population variances are unknown. *Biometrika* **38**, 324–9.

JAMES, G. S. (1954). Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika* **41**, 19–43.

WELCH, B. L. (1947). The generalisation of 'Student's' problem when several different population variances are involved. *Biometrika* **34**, 28–35.

WELCH, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika* **38**, 330–6.

[*Received April* 1979. *Revised June* 1979]