SOME THEOREMS ON QUADRATIC FORMS APPLIED IN THE STUDY OF ANALYSIS OF VARIANCE PROBLEMS, I. EFFECT OF INEQUALITY OF VARIANCE IN THE ONE-WAY CLASSIFICATION

By G. E. P. Box

Imperial Chemical Industries, Manchester, England and North Carolina State College

1. Summary and introduction. This is the first of two papers describing a study of the effect of departures from assumptions, other than normality, on the null-distribution of the F-statistic in the analysis of variance. In this paper, certain theorems required in the study and concerning the distribution of quadratic forms in multi-normally distributed variables are first enunciated and simple approximations tested numerically. The results are then applied to determine the effect of group-to-group inequality of variance in the one-way classification. It appears that if the groups are equal, moderate inequality of variance does not seriously affect the test. However, with unequal groups, much larger discrepancies appear. In a second paper, similar methods are used to determine the effect of inequality of variance and serial correlation between errors in the two-way classification.

2. Distribution of quadratic forms in multi-normal variates. In what follows we write $\chi^2(\nu)$ to denote a quantity distributed as χ^2 with ν degrees of freedom and $F(\nu_1, \nu_2)$ to denote a quantity distributed as the Fisher-Snedecor F with ν_1 and ν_2 degrees of freedom.

By obvious extension of a theorem due to Cochran [1] we have

THEOREM 2.1. *If z denotes a column vector of p random variables $z_1, z_2, \cdots, z_p$ having expectation zero and distributed in a multi-normal distribution with $p \times p$ variance-covariance matrix V, and if $Q = z'Mz$ in any real quadratic form of rank $r \leqq p$, then Q is distributed like a quantity*

$$X = \sum_{j=1}^{r} \lambda_j \chi^2(1) \tag{2.1}$$

where each χ^2 variate is distributed independently of every other and the λ's *are the r real nonzero latent roots of the matrix*

$$U = VM. \tag{2.2}$$

It readily follows that

THEOREM 2.2. *The* sth *cumulant* $K_s(Q)$ *is given by*

$$K_s(Q) = 2^{s-1}(s-1)! \sum_{j=1}^{r} \lambda_j^s. \tag{2.3}$$

Received 6/15/53.

Calculation of this quantity is often facilitated by using the relation

$$\sum_{j=1}^{r} \lambda_j^s = \mathrm{tr}(VM)^s = \sum_{a=1}^{p} \sum_{b=1}^{p} \cdots \sum_{s=1}^{p} u_{ab} u_{bc} \cdots u_{sa} \tag{2.4}$$

whence the first few cumulants of Q may readily be derived without actually determining the λ's. In particular

$$\mathrm{K}_1(Q) = \sum_{a=1}^{p} u_{aa} \tag{2.5}$$

$$\mathrm{K}_2(Q) = 2 \sum_{a=1}^{p} \sum_{b=1}^{p} u_{ab} u_{ba} . \tag{2.6}$$

When the λ_j are all positive, the following χ^2 series due to Robbins and Pitman [2] may be used to obtain the distribution of $X = \sum_{j=1}^{r} \lambda_j \chi^2(\nu_j)$.

THEOREM 2.3. *If X_0 is some constant greater than zero then*

$$P_1 \leqq \Pr\{X > X_0\} \leqq P_2 \tag{2.7}$$

where

$$P_1 = \sum_{l=0}^{n} c_l \Pr\left\{\chi^2(\nu + 2l) > \frac{X_0}{a_1}\right\} + \left(1 - \sum_{l=0}^{n} c_l\right) \Pr\left\{\chi^2(\nu + 2n + 2) > \frac{X_0}{a_1}\right\} \tag{2.8}$$

$$P_2 = \sum_{l=0}^{n} c_l \Pr\left\{\chi^2(\nu + 2l) > \frac{X_0}{a_1}\right\} + \left(1 - \sum_{l=0}^{n} c_l\right) \tag{2.9}$$

and the constants c_l are such that $\sum_{l=0}^{\infty} c_l = 1$ and are defined by the identity

$$\prod_{j=2}^{r} \{a_j^{-\nu_j/2}[1 - (1 - a_j^{-1})w]^{-\nu_j/2}\} = \sum_{l=0}^{\infty} c_l w^l \tag{2.10}$$

$a_1 = \lambda_1$ being the smallest of the λ_j, $a_j = \lambda_j/\lambda_1$ $(j \neq 1)$, and $\sum_{j=1}^{r} (\nu_j) = \nu$.

If the component degrees of freedom ν_j are even, a finite χ^2 series, derived below, may be used whether the λ_j are positive or not.

THEOREM 2.4. *The exact distribution of $X = \sum_{j=1}^{r} \lambda_j \chi^2(\nu_j)$, where the $\nu_j = 2g_j$ are even integers, is a weighted finite sum of χ^2 distributions,*

$$\Pr(X > X_0) = \sum_{j=1}^{r} \sum_{s=1}^{g_j} \alpha_{js} \Pr\{\chi^2(2s) > X_0/\lambda_j\} \tag{2.11}$$

and α_{js} is a constant involving only the λ's and is given by

$$\alpha_{js} = f_j^{(g_j - s)}(0)/(g_j - s)! \tag{2.12}$$

where $f_j^{(h)}(0)$ is obtained by differentiating $f_j(y)$ h times with respect to y and then putting $y = 0$ and

$$f_j(y) = \prod_{i \neq j}^{r} \left[\frac{\lambda_j - \lambda_i}{\lambda_j} + y \frac{\lambda_i}{\lambda_j}\right]^{-\nu_i/2} \tag{2.13}$$

In the special case $r = 2$, a series of this type has been used by Satterthwaite [3]. The general theorem is conveniently proved as follows.

PROOF. Since $g_j = \nu_j/2$ is an integer, the characteristic function

$$\varphi(t) = \prod_{j=1}^{r} (1 - 2it\lambda_j)^{-g_j} \tag{2.14}$$

can be resolved into partial fractions

$$\prod_{j=1}^{r} (1 - 2it\lambda_j)^{-g_j} = \sum_{j=1}^{r} \sum_{s=1}^{g_j} \alpha_{js}(1 - 2it\lambda_j)^{-s} \tag{2.15}$$

where the α_{js} are constants not containing t. We recognise this expression as the characteristic function of a quantity v whose probability density function is

$$p(v) = \sum_{j=1}^{r} \sum_{s=1}^{g_j} \alpha_{js}\, p\{\lambda_j \chi^2(2s)\}. \tag{2.16}$$

Hence X is distributed like v and equation (2.11) follows at once.

To find the values for the constants, (2.15) is written in the form

$$\begin{aligned}(1 - 2it\lambda_i)^{-g_i} \prod_{j \neq i}^{r} (1 - 2it\lambda_j)^{-g_j} &= \sum_{s=1}^{g_i} \alpha_{is}(1 - 2it\lambda_i)^{-s} \\ &\quad + \sum_{j \neq i}^{r} \sum_{w=1}^{g_j} \alpha_{jw}(1 - 2it\lambda_j)^{-w}.\end{aligned} \tag{2.17}$$

Putting $y = 1 - 2it\lambda_i$ and multiplying both sides of the identity by y^{g_i} we have

$$\begin{aligned}\prod_{j \neq i}^{r} \left\{\frac{\lambda_i - \lambda_j}{\lambda_i} + y\,\frac{\lambda_j}{\lambda_i}\right\}^{-g_j} &= \sum_{s=1}^{g_i} \alpha_{is}\, y^{g_i - s} \\ &\quad + \sum_{j \neq i}^{r} \sum_{w=1}^{g_i} \alpha_{jw} \left\{\frac{\lambda_i - \lambda_j}{\lambda_i} + y\,\frac{\lambda_j}{\lambda_i}\right\}^{-w} y^{g_i}.\end{aligned} \tag{2.18}$$

If y is put equal to 0 we have

$$\alpha_{ig_i} = \prod_{j \neq i}^{r} \left\{\frac{\lambda_i}{\lambda_i - \lambda_j}\right\}^{g_j}. \tag{2.19}$$

To obtain the remaining constants we differentiate both sides of identity h times and then put $y = 0$. There will be no contribution from the second member on the right-hand side of (2.18) and the term $\sum_{s=1}^{g_i} \alpha_{is} y^{g_i - s}$ will contribute $h!\alpha_{ig_i - h}$. Thus

$$\alpha_{ig_i - h} = \frac{f_i^{(h)}(0)}{h!} \quad \text{where} \quad f_i(y) = \prod_{j \neq i}^{r} \left\{\frac{\lambda_i - \lambda_j}{\lambda_i} + y\,\frac{\lambda_j}{\lambda_i}\right\}^{-g_j}. \tag{2.20}$$

In practice the constants can be most easily found as follows.

$$f_i(y) = \prod_{j \neq i}^{r} \left\{\frac{\lambda_i}{\lambda_i - \lambda_j}\right\}^{g_j} \prod_{j \neq i}^{r} \left\{1 + y\,\frac{\lambda_j}{\lambda_i - \lambda_j}\right\}^{-g_j} \tag{2.21}$$

$$= \prod_{j \neq i}^{r} \left\{\frac{\lambda_i}{\lambda_i - \lambda_j}\right\}^{g_j} \exp\left\{-\sum_{j \neq i}^{r} g_j \log\left[1 + y\,\frac{\lambda_j}{\lambda_i - \lambda_j}\right]\right\}. \tag{2.22}$$

Since t can always be chosen so that $|y| < 1$, we can expand each side of the equation and equate coefficients.

$$
\begin{aligned}
(2.23)\qquad \sum_{h=0}^{\infty} \alpha_{ig_i-h}\, y^h &= \sum_{h=0}^{\infty} \frac{f_i^{(h)}(0)}{h!}\, y^h \\
&= \prod_{j\neq i}^{r} \left\{\frac{\lambda_i}{\lambda_i - \lambda_j}\right\}^{g_j} \exp\left\{\sum_{h=1}^{\infty} \frac{y^h}{h} \sum_{j\neq i}^{r} g_j \left[\frac{-\lambda_j}{\lambda_i - \lambda_j}\right]^h\right\}.
\end{aligned}
$$

The relation between α_{ig_i-h} and the coefficient of y^h on the right-hand side of (2.23) is the same as that between the hth moment about the origin μ_h' and the hth cumulant K_h. The well known equalities expressing the moments in terms of the cumulants may be used, therefore, in calculating the coefficients required above. For if we write

$$
(2.24)\qquad K_{ih} = (h-1)! \sum_{j\neq i}^{r} \left\{ g_j \left[\frac{-\lambda_j}{\lambda_i - \lambda_j}\right]^h \right\}
$$

then

$$
(2.25)\qquad \alpha_{ig_i-h} = (\mu_{1h}'/h!)\alpha_{ig_i}
$$

$$
(2.26)\qquad \alpha_{ig_i} = \prod_{j\neq i}^{r} \left\{\frac{\lambda_i}{\lambda_i - \lambda_j}\right\}^{g_j}.
$$

3. Investigation of the accuracy of a simple approximation to the distribution of a nonnegative quadratic form. Welch [4], [5] and Fairfield Smith [6] have represented the distribution of a particular nonnegative quadratic form, when $r = 2$, by that of $Z = g\chi^2(h)$, the constants g and h being chosen so that the distribution has the same first two moments as Q. Satterthwaite [3] has suggested its use in the general case ($r \geqq 2$) when we have

THEOREM 3.1.

$$
Q = z'Mz = \sum_{j=1}^{r} \lambda_j \chi^2(\nu_j)
$$

is distributed approximately as $g\chi^2(h)$ *where*

$$
(3.1)\qquad g = \tfrac{1}{2}\frac{K_2(Q)}{K_1(Q)} = \frac{\sum \nu_j \lambda_j^2}{\sum \nu_j \lambda_j} \quad\text{and}\quad h = \frac{2\{K_1(Q)\}^2}{K_2(Q)} = \frac{(\sum \nu_j \lambda_j)^2}{\sum \nu_j \lambda_j^2}.
$$

It should be noted that the effective degrees of freedom, h, are necessarily *less* than the number appropriate if the λ_j were all equal. For if

$$
\nu_1, \cdots, \nu_j, \cdots, \nu_r;\quad \lambda_1, \cdots, \lambda_j, \cdots, \lambda_r;
$$

and μ are any positive real numbers, then

$$
(3.2)\qquad \sum_{j=i}^{r} \nu_i(\lambda_j - \mu)^2 \geqq 0,
$$

and if μ is the weighted mean of the λ's, that is if

TABLE 1

Comparison of Approximate and Exact Distributions of Some Quadratic Forms

Values of λ_j					Degrees of freedom					Exact probability (%) of exceeding approx. $100\alpha\%$ significance point					
λ_1	λ_2	λ_3	λ_4	λ_5	ν_1	ν_2	ν_3	ν_4	ν_5	$100\alpha\% = 5.00$	10.00	25.00	50.00	75.00	95.00
1	2	3			2	4	6			5.04	9.94	24.56	49.74	75.07	95.96
3	2	1			2	4	6			5.08	9.87	24.41	49.41	75.27	95.69
10	5	1			2	4	6			5.10	9.75	23.91	49.95	75.63	96.54
10	2	1			2	4	6			5.20	9.54	22.66	47.61	77.25	98.16
10	2	1			2	2	2			5.08	9.68	23.56	48.68	76.95	97.82
10	5	1			4	8	12			5.16	9.92	24.22	49.06	75.16	98.22
5	4	3	2	1	2	2	4	6	6	5.15	10.15	24.40	49.32	75.14	95.72

$$\mu = \left(\sum_{j=1}^{r} \nu_j \lambda_j\right) \Big/ \sum_{j=1}^{r} \nu_j, \tag{3.3}$$

then (3.2) is equivalent to

$$\sum_{j=1}^{r} \nu_j \lambda_j^2 - \left(\sum_{j}^{r} \nu_j \lambda_j\right)^2 \Big/ \sum_{j=1}^{r} \nu_j \geqq 0, \tag{3.4}$$

that is

$$\left(\sum_{j=1}^{r} \nu_j \lambda_j\right)^2 \Big/ \sum_{j=1}^{r} \nu_j \lambda_j^2 \leqq \sum_{j=1}^{r} \nu_j. \tag{3.5}$$

Although an approximation of this kind has often been used (see for example Patnaik [7]), investigations of its accuracy seem limited to the case $k = 2$ studied by Satterthwaite [3].

Table 1 shows the exact probabilities (calculated from the finite series of Theorem 2.4) of exceeding the significance points obtained from the approximation for a number of particular quadratic forms. This brief investigation suggests that the simple approximation is fairly good over a wide range of values of ν and λ. However, when small differences in probability were to be examined, it would be necessary to apply the method with caution and make checks by the exact methods.

4. Distribution of the ratio of independently distributed quadratic forms in multi-normal varieties. By canonical reduction of numerator and denominator, the ratio $Y = Q_1/Q_2$ is seen to be distributed like the quantity

$$X_1/X_2 = \left\{\sum_{j'=1}^{r'} \lambda'_{j'} \chi^2(\nu'_{j'})\right\} \Big/ \left\{\sum_{j=1}^{r} \lambda_j \chi^2(\nu_j)\right\}. \tag{4.1}$$

By representing numerator and denominator by infinite χ^2 series, Pitman and Robbins [2] have obtained the distribution of Y, when the λ's are all positive, as an infinite series in which each term contains a probability calculated from the F distribution (or more conveniently from the incomplete B-function). In our application it will always be possible to arrange that component χ^2's (at least of the denominator) have even degrees of freedom.

Employing the Robbins-Pitman [2] infinite series in the numerator and the finite series in the denominator, we readily obtain Theorem 4.1 (for example by using Cramer's theorem [8] concerning the characteristic function of a ratio).

THEOREM 4.1. *If the* λ'*'s of the numerator are all positive and if* $\lambda_1' = a'$ *is the smallest of the* λ'*'s and* $\sum_{j'=1}^{r'} \nu'_{j'} = \nu'$, *then*

$$P_1 \leqq P_r\,(Y > Y_0) \leqq P_2 \tag{4.2}$$

where

$$\begin{aligned} P_1 = \sum_{l'=0}^{n'} \sum_{j=1}^{r} \sum_{s=1}^{g_j} c_{l'}' \,\alpha_{js}\{I_{x_j}(s, \tfrac{1}{2}\nu' + l')\} \\ + \left\{1 - \sum_{l'=0}^{n'} c_{l'}'\right\} \sum_{j=1}^{r} \sum_{s=1}^{g_j} \alpha_{js}\{I_{x_j}(s, \tfrac{1}{2}\nu' + n' + 1)\}, \end{aligned} \tag{4.3}$$

$$P_2 = \sum_{l'=0}^{n'} \sum_{j=1}^{r} \sum_{s=1}^{g_j} c_{l'}' \,\alpha_{js}\{I_{x_j}(s, \tfrac{1}{2}\nu' + l')\} + \left\{1 - \sum_{l'=0}^{n'} c_{l'}'\right\}. \tag{4.4}$$

The c'*'s are obtained from* (2.10), *the* α_{js} *from* (2.24), (2.25) *and* (2.26), $I_x(pq)$ *is the incomplete Beta function integral, and* $x_j = \{1 + (\lambda_j/a_1')Y_0\}^{-1}$.

If both numerator and denominator of (4.1) have even degrees we may use the finite series in both numerator and denominator and obtain

THEOREM 4.2. *If* $\nu_j = 2g_j$ *and* $\nu_{j'}' = 2g_{j'}'$ *are even, then*

$$\Pr\,(Y \leqq Y_0) = \sum_{j'=1}^{r'} \sum_{s'=1}^{g'_{j'}} \sum_{j=1}^{r} \sum_{s=1}^{g_j} \alpha_{j's'}' \,\alpha_{js}\{I_{x_{jj'}}(s, s')\} \tag{4.5}$$

where the $\alpha_{j's'}'$ *and* α_{js} *are obtained from* (2.24), (2.25) *and* (2.26), *and* $x_{jj'} = \{1 + \lambda_j Y_0/\lambda_{j'}'\}^{-1}$.

Alternatively, if the forms are nonnegative it is usually simpler to use the following.

THEOREM 4.3. *If* $\lambda_1', \lambda_2', \cdots, \lambda_r'$ *and* $\lambda_1, \lambda_2, \cdots, \lambda_r$ *are all positive and the* ν_j *and* ν_j' *are all even then*

$$\Pr\left[\left\{\sum_{j'=1}^{r'} \lambda_{j'}' \,\chi^2(\nu_{j'}')\right\} \Big/ \left\{\sum_{j=1}^{r} \lambda_j \,\chi^2(\nu_j)\right\} > Y_0\right] = \sum_{i=1}^{r'} \sum_{s=1}^{g_i} \alpha_{is} \tag{4.6}$$

where the α_{is} $(i = 1, 2, \cdots r';\ s = 1, \cdots g_j)$ *are constants calculated from* (2.24), (2.25) *and* (2.26) *for the form* $\sum_{i=1}^{r'+r} \zeta_i\chi^2(\nu_i)$ *in which* $\zeta_1 = \lambda_1'$, $\zeta_2 = \lambda_2'$, $\cdots$, $\zeta_{r'} = \lambda_{r'}'$; $\zeta_{r'+1} = -Y_0\lambda_1$, $\zeta_{r'+2} = -Y_0\lambda_2$, $\cdots$, $\zeta_{r'+r} = -Y_0\lambda_r$.

PROOF. Since the quadratic forms are nonnegative, the left-hand side of (4.6) may be written

$$P = \Pr\left[\left\{\sum_{j'=1}^{r'} \lambda_{j'}' \,\chi^2(\nu_{j'}') - \sum_{j=1}^{r} Y_0\,\lambda_j \,\chi^2(\nu_j)\right\} > 0\right] \tag{4.7}$$

which is of the form

TABLE 2

Comparison of Approximate and Exact Distributions of Ratios of Quadratic Forms

Numerator								Denominator										Exact probability (%) of exceeding approx. 100α% significance point			
Degrees of freedom				Values of λ's				Degrees of freedom					Values of λ's								
ν'_1	ν'_2	ν'_3	ν'_4	λ'_1	λ'_2	λ'_3	λ'_4	ν_1	ν_2	ν_3	ν_4	ν_5	λ_1	λ_2	λ_3	λ_4	λ_5	100α% = 5.00	10.00	25.00	50.00
2	2			1	2			2	4	6			3	2	1			4.71	9.63	24.60	49.96
2	2			1	2			2	4	6			10	5	1			4.24	9.15	24.36	50.00
2	2			1	3			2	4	6			10	5	1			4.26	9.12	24.19	49.93
2	2			1	2			4	8	12			3	2	1			4.88	9.76	24.59	49.91
2	2			1	3			4	8	12			10	5	1			4.72	9.55	24.31	49.80
2	2	2	2	1	2	3	4	2	2	4	6	6	5	4	3	2	1	5.03	9.95	24.86	50.05

$$\Pr\left[\sum_{i=1}^{r'+r} \zeta_i \chi^2(\nu_i) > 0\right]. \tag{4.8}$$

Using (2.11)

$$P = \sum_{i=1}^{r'} \sum_{s=1}^{g_i} \alpha_{is} \Pr\{\chi^2(2s) > 0\} + \sum_{j=r+1}^{r'+r} \sum_{w=1}^{g_j} \alpha_{jw} \Pr\{\chi^2(2w) < 0\}. \tag{4.9}$$

But

$$\Pr\{\chi^2(2s) > 0\} = 1 \quad \text{and} \quad \Pr\{\chi^2(2w) < 0\} = 0. \tag{4.10}$$

Therefore,

$$P = \sum_{i=1}^{r'} \sum_{s=1}^{g_i} \alpha_{is}. \tag{4.11}$$

It will be noted that when this series is applicable, the required probability may be obtained directly without the rather tedious interpolation in the B-function tables required by the method of Theorem 4.2.

5. Use of Theorem 4.3 with quadratic forms that are not independent. This method may be used in suitable cases even if the quadratic forms Q_1 and Q_2 are not distributed independently. For if Q_1 and Q_2 are nonnegative, we may write

$$\begin{aligned} P &= \text{Ps}\{Q_1/Q_2 > \varphi\} = \Pr\{Q_1 - \varphi Q_2 > 0\} = \Pr(z'M_1 z - \varphi z'M_2 z) > 0 \\ &= \Pr\{z'Mz > 0\} = \Pr\left\{\sum_{i=1}^{r'} \zeta_i \chi^2(\nu_i) + \sum_{j=1}^{r} \zeta_j \chi^2(\nu_j) > 0\right\} \end{aligned} \tag{5.1}$$

where ζ_i $(i = 1, \cdots, r')$ is a positive latent root, repeated ν_i times, of the matrix $M = M_1 - M_2$ and ζ_j $(j = 1, \cdots, r)$ is a negative latent root of the same matrix repeated ν_j times. In certain investigations (for example in the study of the two-way classification of the analysis of variance table) it is possible to ensure that the ν_i and ν_j are all even, whence we may apply Theorem 4.3 and obtain

$$P = \sum_{i=1}^{r'} \sum_{s=1}^{\nu i/2} \alpha_{is}. \tag{5.2}$$

6. The accuracy of a simple approximation to the distribution of the ratio of independent nonnegative quadratic forms. Since approximation to the distribution of a positive quadratic form Q by $g\chi^2(h)$ is fairly satisfactory, we may attempt to approximate the ratio of two independent quadratic forms Q' and Q by fitting χ^2 distributions in both numerator and denominator, in the manner described.

THEOREM 6.1. *If Q' is distributed approximately as $g'\chi^2(h')$ and Q as $g\chi^2(h)$, a quantity whose distribution approximates to that of the ratio Q'/Q is $bF(h', h)$, where $b = (g'h')/(gh)$ and the g's and h's are found from* (3.1). *In fact*

$$b = \mathrm{K}_1(Q')/\mathrm{K}_1(Q), \qquad h' = 2\{\mathrm{K}_1(Q')\}^2/\mathrm{K}_2(Q'), \qquad h = 2\{\mathrm{K}_1(Q)\}^2/\mathrm{K}_2(Q). \tag{6.1}$$

In Table 2 are shown the exact probabilities (calculated from the finite series of Theorem 4.3) of exceeding the significance points obtained from the approximation. The approximation is not of great accuracy, but may be usefully employed to supplement the accurate (but less suggestive) exact methods.

7. The one-way analysis of variance classification. Data, which it is desired to test for group to group homogeneity of mean value, often are obtained in circumstances where group-to-group homogeneity of variance is not to be expected. To quote one of many examples; in biological work where each observation is the response observed with a particular animal and the subject of enquiry is the comparison of the effects of treatments applied to the animal, the application of the treatment itself would often be expected to cause extra variability, and the extent of this extra variability would vary with the type and manner of treatment applied.

The problem of the effect of unequal group variances was considered in the case of the t test by Welch [5]. He obtained approximate probabilities from which it appeared that the effect was small when the groups were of equal size, but larger when they were different in size. Later some exact probabilities for this case were found by Hsu [9] and another investigation by a different approximate method was made by Grunow [10]. Both these investigations confirmed Welch's results. Quensel [11] considered the one-way analysis of variance classification more generally and obtained an approximate expression for the variance of the F criterion when the group variances differed. He concluded that the test would not be greatly affected if the group sizes were equal.

David and Johnson [12], [13], [14] have discussed the general problem of the power function of analysis of variance criteria when the observations are distributed independently but do not necessarily follow the normal distribution or have constant variance. As a special case they consider the one-way classifi-

TABLE 3

Analysis of Variance. Group Variances Unequal

Source	Deg. fr.	Sum of squares Q	Expectation of Q	Null distribution of Q
Between groups	$k-1$	$Q_B = \sum_{t=1}^{k} n_t(\bar{y}_{t.} - \bar{y}_{..})^2$	$\sum_{t=1}^{k} n_t\gamma_t^2 + \sum_{t=1}^{k}\left(1 - \frac{n_t}{N}\right)\sigma_t^2$	$\sum_{t=1}^{k-1} \lambda_t\,\chi^2(1)$.
Within groups	$N-k$	$Q_W = \sum_{t=1}^{k}\sum_{i=1}^{n_t} (y_{ti} - \bar{y}_{t.})^2$	$\sum_{t=1}^{k} (n_t - 1)\,\sigma_t^2$	$\sum_{t=1}^{k} \sigma_t^2\,\chi^2\,(n_t - 1)$

cation in which the observations are normally distributed but the variances differ from group to group. Their method is different from that given here and is an approximate one. At the time of writing they have published few numerical results and these [14] are confined to the case in which the sizes of the groups are all equal. Confirming the results of Quensel, only slight changes in probability, from those expected if the assumptions were true, have been found.

Using the theorems on quadratic forms discussed above, the required probabilities may be found exactly, while the simple F approximation enables the nature of the effects found to be presented in a readily appreciated form.

Suppose we have $N = \sum_{t=1}^{k} n_t$ observations classified into k groups. The ith observation in the tth group is y_{ti}, the mean of the tth column $\bar{y}_{t.}$, and the grand mean $\bar{y}_{..}$, and there are n_t observations in the tth group. Then in the usual analysis of variance, the quantities Q_B and Q_W shown in the third column of Table 3 are calculated and are associated with degrees of freedom shown in the second column of the table.

It is usually assumed that

$$y_{ti} = \eta_t + z_{ti} \tag{7.1}$$

where $\eta_t = \alpha + \gamma_t$ is the population mean for the tth group, $\sum_{t=1}^{k} n_t\gamma_t = 0$ and the z_{ti} are errors distributed normally and independently about zero with the same variance σ^2. We retain all the assumptions except the last, and, instead of supposing the variance constant, we postulate that $\mathcal{E}\,(z_{ti}^2) = \sigma_t^2$ where $\sigma_1^2, \sigma_2^2, \cdots, \sigma_t^2, \cdots, \sigma_k^2$ are not necessarily all equal. Then

$$Q_B = \sum_{t=1}^{k} n_t(\gamma_t + \bar{z}_{t.} - \bar{z}_{t..})^2, \tag{7.2}$$

$$\mathcal{E}(Q_B) = \sum_{t=1}^{k} n_t\gamma_t^2 + \mathcal{E}\sum_{t=1}^{k} n_t(\bar{z}_{t.} - \bar{z}_{..})^2. \tag{7.3}$$

We notice that when the null hypothesis is true, Q_B is a quadratic form in the variables$_1$ $\bar{z}, \cdots, \bar{z}_{t.}, \cdots, \bar{z}_k$. The matrix $M = \{m_{ts}\}$ of this form is evidently $N^{-1}\{\delta_{ts}n_tN - n_tn_s\}$, where δ_{ts} is the Kronecker delta. Also the variables follow the multi-normal distribution with diagonal variance-covariance matrix V whose tth element is σ_t^2/n_t. It follows that the matrix U of Theorem 2.1 is

$$U = VM = N^{-1}\{\delta_{ts}\sigma_t^2N - \sigma_t^2n_s\}. \tag{7.4}$$

Using (7.3) and Theorems 2.1 and 2.2, the expectations and null-distribution of Q_B are those shown in the last two columns of Table 3, $\lambda_1, \cdots, \lambda_{k-1}$ being the latent roots of the matrix U. Again, from (7.1),

$$Q_E = \sum_{t=1}^{k} \sum_{i=1}^{n_t} (\bar{y}_{ti} - \bar{y}_{t.})^2 = \sum_{t=1}^{k} \sum_{i=1}^{n_t} (z_{ti} - \bar{z}_{t.})^2. \tag{7.5}$$

Also, since $\sum_{i=1}^{n_t} (z_{ti} - \bar{z}_{t.})^2$ is distributed independently of $\bar{z}_{t.}$ like $\sigma_t^2\chi^2(n_t - 1)$, it follows that Q_W is distributed like $\sum_{t=1}^{k} \sigma_t^2\chi^2\,(n_t - 1)$ independently of Q_B. We may now employ Theorem 4.1 to obtain the exact probability that the ratio of mean squares will exceed the significance points of the tabled F distribution, for any chosen set of group variances. In addition, an approximate value of this probability may be obtained using Theorem 6.1 with equations (2.5) and (2.6). We find that the ratio of mean squares is distributed approximately as $bF(h', h)$ where

$$b = \frac{N - k}{N(k - 1)} \sum_t (N - n_t)\sigma_t^2 / \sum_t (n_t - 1)\sigma_t^2, \tag{7.6}$$

$$h' = \{\sum_t (N - n_t)\sigma_t^2\}^2 / \{\sum_t n_t \sigma_t^2\}^2 + N \sum_t (N - 2n_t)\sigma_t^4\}, \tag{7.7}$$

$$h = \{\sum_t (n_t - 1)\sigma_t^2\}^2 / \{\sum_t (n_t - 1)\sigma_t^4\}. \tag{7.8}$$

A number of calculations of both exact and approximate probabilities are shown in Table 4. It is seen that in the cases studied the approximation, although not

TABLE 4

Probabilities of Exceeding 5% Point when Variances are Unequal in the One-Way Analysis of Variance Table

	Group variances					Number of Observations in Groups					Total	Probability (%) of exceeding 5% point		Values in approximating distribution, b F (h', h),		
	σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_5^2	n_1	n_2	n_3	n_4	n_5	N	Exact	Approx.	b	h'	h
(1)	1	1	1			Any					15	5.00		1	2	12
(2)	1	2	3			5	5	5			15	5.58	5.78	1	1.85	10.29
(3)	1	2	3			3	9	3			15	5.55	5.72	1	1.74	11.08
(4)	1	2	3			7	5	3			15	9.25	9.57	1.28	1.86	10.00
(5)	1	2	3			3	5	7			15	4.03	3.35	0.80	1.82	10.89
(6)	1	1	3			5	5	5			15	5.87	5.82	1	1.72	9.09
(7)	1	1	3			7	5	3			15	10.70	9.78	1.35	1.67	9.14
(8)	1	1	3			9	5	1			15	17.41	18.04	1.93	1.62	12.00
(9)	1	1	3			1	5	9			15	1.31	1.73	0.60	1.85	10.32
(10)	1	1	1	1	1	Any					25	5.00		1	4	20
(11)	1	1	1	1	3	5	5	5	5	5	25	7.42	6.86	1	3.21	15.08
(12)	1	1	1	1	3	9	5	5	5	1	25	14.64	15.56	1.48	3.04	20
(13)	1	1	1	1	3	1	5	5	5	9	25	2.49	2.60	0.73	3.40	15.43

of great accuracy, faithfully indicates the order and direction of the effects and enables a clear idea to be gained of the general effects to be expected.

8. Equal groups. For equal groups the comparison of mean squares is unbiased in the sense that the expectations of the mean squares are equal when the null hypothesis is true. In fact

$$\mathcal{E}(Q_B)/(k-1) = \mathcal{E}(Q_W)/(N-k) = \bar{\sigma}^2 \qquad \text{where } \bar{\sigma}^2 = (\sum \sigma_t^2)/k.$$

The mean squares are distributed independently but their ratio does not follow the distribution $F(k-1, N-k)$. Instead, applying the approximation, we find after a little reduction of (7.7) and (7.8) that the ratio of mean squares is distributed approximately as $F\{(k-1)\epsilon', (N-k)\epsilon\}$ where ϵ' and ϵ, the factors by which the degrees of freedom are reduced, are given by

$$\epsilon' = \left\{1 + \frac{k-2}{k-1}c^2\right\}^{-1}, \qquad \epsilon = (1 + c^2)^{-1}, \tag{8.1}$$

and c is the coefficient of variation of the variances. That is to say, c^2 is the variance of the variances divided by the square of the mean variance $\bar{\sigma}^2$:

$$c^2 = \frac{1}{k}\sum_{t=1}^{k}(\sigma_t^2 - \bar{\sigma}^2)^2/(\bar{\sigma}^2)^2. \tag{8.2}$$

Since, when the variances are unequal, ϵ' and ϵ are less than unity, one would expect that the significance of effects would be somewhat overestimated. Comparison in Table 4, of rows (2) and (6) with (1), and of row (11) with row (10), confirms this, and shows that for the differences in variance considered, only moderate discrepancies in probability occur.

Now the σ^2's are essentially positive and it is easily seen that

$$0 \leqq c^2 \leqq k - 1, \tag{8.3}$$

and if the variances range from a lower value σ^2 to an upper value $a\sigma^2$, then the largest possible value for c is attained when $k - 1$ of the variances are equal to σ^2 and the remaining one is equal to $a\sigma^2$, and that in this case

$$c^2 = (k-1)(a-1)^2/(a-1+k)^2. \tag{8.4}$$

TABLE 5

Values of c^2, ϵ' and ϵ

Largest Variance is a Times Larger than Each of the Remaining Variances

k	3 groups			6 groups			10 groups		
a	3	6	10	3	6	10	3	6	10
c^2	0.32	0.78	1.12	0.31	1.03	1.80	0.25	1.00	2.02
ϵ'	0.86	0.72	0.64	0.80	0.55	0.41	0.82	0.52	0.36
ϵ	0.76	0.56	0.47	0.76	0.50	0.36	0.80	0.50	0.33

Values for c^2 greater than one or at most two probably would be extremely rare in practice, although from the inequality (8.3) it is seen that values of c^2 as great as $k - 1$, and hence values of ϵ as small as $1/k$, *could* occur. Some idea of the discrepancy arising in a very unfavorable case is obtained by considering the example $k = 7$, $n_1 = n_2 = \cdots = n_7 = 3$, $a = 10$; the exact probability of exceeding the assumed 5 per cent point is then 12.0 per cent.

9. Unequal groups. It will be observed that the more serious discrepancies in Table 4 arise when the groups are unequal. This is because with unequal groups the comparison of mean squares is usually biased. If we write $\dot{\sigma}^2$ for the *weighted* mean variance $\sum \nu_t \sigma_t^2 / \sum \nu_t$, and $\bar{\sigma}^2$ for the unweighted mean variance $\sum \sigma_t^2/k$, where the weight $\nu_t = n_t - 1$ is the number of degrees of freedom in the tth group then the expression (7.6) for the bias coefficient reduces to the form

$$b = 1 + \frac{1 - 1/N}{1 - 1/k}\left\{\frac{\bar{\sigma}^2}{\dot{\sigma}^2} - 1\right\}. \tag{9.1}$$

The bias is seen to depend upon the ratio of the unweighted and weighted means of the variances.

In this connection it is of interest to consider the examples of rows (2), (3), (4) and (5) in Table 4. In each case the total number of observations is 15 and the unweighted mean variance is 2. In (2) the numbers in the groups are equal, the weighted and unweighted means agree, and there is no bias, the discrepancy in probability being small. In (3) the numbers are unequal but the distribution is symmetrical, the weighted and unweighted means again agree, and again the discrepancy is small and of the same order as that found before. In (4) the weighted mean variance of 1.67 is lower than the unweighted mean variance of 2, causing an upward bias and a marked discrepancy in the direction of over-estimation of significance. In (5) on the other hand, the weighted mean of 2.33 exceeds 2, causing a downward bias corresponding with a discrepancy in probability resulting in underestimation of significance.

We have seen that in the case of equal groups, the discrepancy, as measured by a reduction in the degrees of freedom of the approximating F distribution, is dependent on the *spread* of the distribution of variances as measured by the coefficient of variation. The feature of the distribution of variances which affects the bias, on the other hand, is related to the "skewness" of that distribution as measured by the ratio of weighted and unweighted means.

The factors which multiply the degrees of freedom in the approximation may be written in this case of unequal groups

$$\epsilon' = [1 + \{c(\lambda)\}^2]^{-1}, \qquad \epsilon = [1 + \dot{c}^2]^{-1}, \tag{9.2}$$

where $c(\lambda)$ is the coefficient of variation of the λ's and $\dot{c}$ is the weighted coefficient of variation of the variances. That is to say, $\dot{c}^2$ is the weighted variance of the variances divided by the weighted mean variance $\dot{\sigma}^2$:

$$\dot{c}^2 = (N - k)^{-1} \sum_{t=1}^{k} \nu_t(\sigma_t^2 - \dot{\sigma}^2)^2/(\dot{\sigma}^2)^2. \tag{9.3}$$

The variation among the λ's will be similar although somewhat less in extent than the variation among the σ_t^2's so that, as before, the coefficients ϵ' and ϵ will depend upon the *spread* of the σ_t^2's.

Study of Table 4, particularly rows (4), (7), (8), (9), (12), and (13), shows that quite large discrepancies can occur when the groups are unequal for even moderate variations of variance. Furthermore, it is clear that these discrepancies will persist in larger samples, for as the sample sizes are increased proportionately the bias coefficient will tend to the fixed limit

$$1 + (k/k - 1)\{\bar{\sigma}^2/\dot{\sigma}^2 - 1\}. \tag{9.4}$$

Acknowledgement. I am indebted to Mrs. Margaret Edmondson for valuable assistance with the computations.

REFERENCES

[1] W. G. Cochran, "The distribution of quadratic forms in a normal system, with applications to the analysis of covariance," *Proc. Cambridge Philos. Soc.*, Vol. 30 (1934), pp. 178–191.

[2] H. Robbins and E. J. G. Pitman, "Application of the method of mixtures to quadratic forms in normal variates," *Ann. Math. Stat.*, Vol. 20 (1949), pp. 552–560.

[3] F. E. Satterthwaite, "Synthesis of Variance," *Psychometrika*, Vol. 6 (1941), pp. 309–316.

[4] B. L. Welch, "The specification of rules for rejecting too variable a product, with particular reference to an electric lamp problem," *J. Roy. Stat. Soc., Suppl.*, Vol. 3 (1936), pp. 29–48.

[5] B. L. Welch, "The significance of the difference between two means when the population variances are unequal," *Biometrika*, Vol. 29 (1937), pp. 350–362.

[6] H. Fairfield Smith, "The problem of comparing the results of two experiments with unequal errors," *Journal of the Council for Scientific and Industrial Research* (Australia), Vol. 9, pp. 211–212.

[7] P. B. Patnaik, "The non-central χ^2 and F-distributions and their applications," *Biometrika*, Vol. 36 (1949), pp. 202–232.

[8] H. Cramér, *Random Variables and Probability Distributions*, Cambridge University Press, 1937, p. 46.

[9] P. L. Hsu, "Contributions to the theory of 'Students' t-test as applied to the problem of two samples," *Statistical Research Memoirs*, Vol. 2 (1938), pp. 1–24.

[10] D. G. C. Grunow, "Test for the significance of the difference between means in two normal populations having unequal variances," *Biometrika*, Vol. 38 (1951), pp. 252–256.

[11] C. E. Quensel, "The validity of the z-criterion when the variates are taken from different normal populations," *Skand. Aktuarietids.*, Vol. 30 (1947), pp. 44–55.

[12] F. N. David and N. L. Johnson, "The effect of non-normality on the power function of the F-test in the analysis of variance," *Biometrika*, Vol. 38 (1951), pp. 43–57.

[13] F. N. David and N. L. Johnson, "A method of investigating the effect of non-normality and heterogeneity of variance on tests of the general linear hypothesis," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 382–392.

[14] F. N. David and N. L. Johnson, "The sensitivity of analysis of variance tests with respect to random between groups variation," *Trabajos de Estadistica*, Vol. 2 (1951) pp. 179–188.