ON A SOURCE OF DOWNWARD BIAS IN THE ANALYSIS OF VARIANCE AND COVARIANCE

BY WILLIAM G. MADOW

Institute of Statistics, University of North Carolina

1. Summary. It is shown that if, in the analysis of variance, the experiments are not in a state of statistical control due to variations in the true means, then the test will have a downward bias. The power function of the analysis of variance test is obtained when this downward bias is present.

2. Introduction. To introduce the discussion of this bias let us consider the generalized Student's hypothesis.

Let $y_1, \cdots, y_{kN}$ be normally and independently distributed with variance σ^2, and let the expected value of $y_{i\nu}$, be $a_{i\nu}$.[1] Then the generalized Student's hypothesis is

(Null hypothesis) $$a_{i\nu} = a$$

and the class of alternative hypotheses against which the null hypothesis is tested is

(Class A) $$a_{i\nu} = a_i .$$

From the statement of the null hypothesis and the alternatives of Class A it follows that both the null hypothesis and the alternatives of Class A require that

$$a_{i1} = \cdots = a_{iN} . \tag{1.1}$$

Since our experiments are rarely in such perfect statistical control that (1.1) holds whether or not the null hypothesis is true, it becomes reasonable to investigate the existing F test when instead of the alternatives to the null hypothesis being of Class A, they are simply Class B:

(Class B) Equation (1.1) is false for at least one value of i.

Furthermore, for many practical purposes we would prefer to test the average null hypothesis:

(Average null hypothesis) $$\bar{a}_i = \bar{a},$$

where $N\bar{a}_i = a_{i1} + \cdots + a_{iN}$ and $k\bar{a} = \bar{a}_1 + \cdots + \bar{a}_k$, instead of the null hypothesis, the alternatives to the average null hypothesis being of Class C.

(Class C) The $a_{i\nu}$ can have any values such that not all the $\bar{a}_i$ equal $\bar{a}$.

[1] Throughout this paper the letter i will assume all integral values from 1 to k, the letters μ, ν will assume all integral values from 1 to N, the letters γ, η will assume all integral values from 1 to m, the letter α will assume all integral values from $n_1 + \cdots + n_{\gamma-1} + 1$ to $n_1 + \cdots + n_\gamma$, $(n_0 = 0)$, and α_1, α_2 will assume all integral values from 0 to ∞.

The F-test of the null hypothesis against the alternatives of Class A is, as is well known,

$$F = \frac{k(N-1)\sum_i (\bar{y}_i - \bar{y})^2}{(k-1\sum_{i,\nu}(y_{i\nu} - \bar{y}_i)^2)}$$

where $N\bar{y}_i = y_{i1} + \cdots + y_{iN}$ and $k\bar{y} = \bar{y}_1 + \cdots + \bar{y}_k$. To answer the questions formulated above concerning the F-test when the average null hypothesis or the alternatives of classes B or C are true, we must then calculate the distribution of F under these various conditions. This is done in Section 3.

A somewhat informal means of obtaining the conclusions is that of studying F itself. Taking the expected values of the numerator and denominator of F and defining

$$\phi_1^2 = \frac{N\sum_i(\bar{a}_i - \bar{a})^2}{(k-1)\sigma^2}$$

$$\phi_2^2 = \frac{1}{k(N-1)\sigma^2}\sum_{i,\nu}(a_{i\nu} - \bar{a}_i)^2$$

we obtain as the ratio of the two expected values

$$\bar{F} = \frac{1+\phi_1^2}{1+\phi_2^2}.$$

It is well known that, in general, the larger the value of N the more closely will F approximate $\bar{F}$. From this fact it is easy to see why if the null hypothesis is true, then $F \sim 1$, whereas if the null hypothesis is false but an alternative of Class A is true then

$$F \sim 1 + \phi_1^2 > 1$$

so that large values of F become more likely than if the null hypothesis were true. However, if an alternative of Class B is true then

$$F \sim \frac{1+\phi_1^2}{1+\phi_2^2}$$

so that if $\phi_1^2 < \phi_2^2$, smaller values of F occur more frequently than indicated by the null hypothesis. Thus we would tend to accept the null hypothesis more frequently than desired when it is false. Even when the null hypothesis is false so that $\phi_1^2 > 0$, the values of F will tend to be less if $\phi_2^2 > 0$ than if $\phi_2^2 = 0$ whether or not $\phi_1^2 < \phi_2^2$. Not only is the probability of an error of the first kind less than the value ϵ we may have previously selected, but also the power of the test is less than would be indicated by Tang's tables [1]. The lack of statistical control represented by variation of expected values within a class has the effect of making it less likely than the standard F-test indicates that the null hypothesis will be rejected whether it be true or false. Furthermore, even for relatively low values of ϕ_2^2, the reductions in the probabilities of rejection may be over 40 per cent as indicated by some examples given below.

If the average null hypothesis is true but (1.1) is false it follows that

$$F \sim \frac{1}{1 + \phi_2^2},$$

so that the full effect of the downward bias occurs in that case. Thus in cases where statistical control is lacking, to test the average null hypothesis by the F-test may well result in accepting the hypothesis when it is false. If the null hypothesis is rejected, however, then we can expect that the differences among the true means are even larger than indicated by Tang's tables.

To illustrate, it is shown in Section 4 that if $k = 5$ and $N = 7$, then the probability of rejecting the average null hypothesis when it is true, but (1.1) is false will not be the preassigned .05 but something less than .03 if $\phi_2^2 > .05$. Furthermore, if $\phi_2^2 > .07$, then the power of the F tests for this example will be reduced by at least 40 per cent whatever the value of ϕ_1^2.

The conclusions reached above remain valid for the analysis of variance and covariance in general. In the general case however, the value of the average null hypothesis in simplifying the analysis may be considerably reduced since the parameter ϕ_1^2 no longer vanishes when the average null hypothesis is true. For example, if $Ey_\nu = \beta_\nu x_\nu$, and if the average null hypothesis is $\bar{\beta} = 0$, where $N\bar{\beta} = \beta_1 + \cdots + \beta_N$, then upon calculating

$$\phi_1^2 = \frac{\left(\sum_\nu \beta_\nu x_\nu^2\right)^2}{\sigma^2 \sum_\nu x_\nu^2}$$

we see that ϕ_1^2 will not vanish in general if $\bar{\beta}$ vanishes.

Although as shown above the average null hypothesis may not have too great importance in the case of regression, yet if the "variance between treatments" is a function of arithmetic means of the random variables as in the "pure" analysis of variance the average null hypothesis may well be very useful. Simple examples of this are provided by the randomized block, Latin square, and similar designs.

The distributions that we shall need are given in Section 3. The inequalities on the basis of which the bias is demonstrated are obtained in Section 4.

It would be highly desirable to have Tang's tables extended so that they might provide the answers to the questions raised by this source of bias. In the absence of such extensions the inequalities of Section 4 may give some rough idea, but these inequalities are not sharp enough.

3. The calculation of the distributions. The following theorem was proved, although not explicitly stated, as part of an earlier note [2]. (Note the change from x_i to y_i as the notation for the random variable.)

THEOREM 1. *Let $y_1, \cdots, y_N$ be normally and independently distributed with variance σ^2 and means $a_1, \cdots, a_N$ and let $q_1, \cdots, q_m$ be quadratic forms*

$$q_\gamma = \sum_{\mu,\nu} a_{\mu\nu}^{(\gamma)} y_\mu y_\nu$$

in $y_1, \cdots, y_N$ of ranks $n_1, \cdots, n_m$. Then, if an orthogonal transformation

$$y_\nu = \sum_\mu c_{\nu\mu} z_\mu$$

exists such that

$$q_\gamma = \sum_\alpha z_\alpha^2, \tag{2.1}$$

it follows that the random variables q_γ/σ^2 are independently distributed in χ'^2 distributions with degrees of freedom $n_1, \cdots, n_m$ and parameters $\lambda_1, \cdots, \lambda_m$, where

$$\lambda_\gamma = \frac{1}{2\sigma^2} \sum_{\mu,\nu} a_{\mu\nu}^{(\gamma)} a_\mu a_\nu = \frac{Eq_\gamma}{2\sigma^2} - \frac{n_\gamma}{2}.$$

Various conditions for the existence of an orthogonal transformation satisfying (2.1) of Theorem 1 have been given. Among these are:

1. *Cochran's* [3] *condition.* If $\sum_\gamma q_\gamma = \sum_\nu y_\nu^2$ then a necessary and sufficient condition for the existence of an orthogonal transformation satisfying (2.1) is $\sum_\gamma n_\gamma = N$.

2. *Craig's* [4] *condition.* If A_γ denotes the matrix $(a_{\mu\nu}^{(\gamma)})$ then a necessary and sufficient condition for the existence of an orthogonal transformation satisfying (2.1) is $A_\gamma A_\eta = \delta_{\gamma\eta} A_\gamma$ where $\delta_{\gamma\eta}$ is the null matrix if $\gamma \neq \eta$ and the identity matrix if $\gamma = \eta$.

3. *Linear Hypothesis condition.* (Kolodziejczyk [5]) If λ be the likelihood ratio test of a linear hypothesis and if $E^2 = 1 - \lambda^{2/N}$, then $E^2 = q_1/(q_1 + q_2)$ and an orthogonal transformation exists satisfying (2.1) with $m = 2$.

To summarize some results obtained by Tang [1], let us state

THEOREM 2. *If $\chi_1'^2$ and $\chi_2'^2$ are independently distributed in distributions with n_1 and n_2 degrees of freedom and parameters λ_1 and λ_2, and if*

$$E^2 = \frac{\chi_1'^2}{\chi_1'^2 + \chi_2'^2},$$

then the probability density of E^2 is

$$p = p(E^2 \mid \lambda_1, \lambda_2, n_1, n_2) = e^{-\lambda_1-\lambda_2}(E^2)^{(n_1/2)-1}(1 - E^2)^{(n_2/2)-1}$$

$$\sum_{\alpha_1,\alpha_2} \frac{\lambda_1^{\alpha_1}\lambda_2^{\alpha_2}\Gamma\left(\frac{n_1+n_2}{2} + \alpha_1 + \alpha_2\right)}{\alpha_1 !\, \alpha_2 !\, \Gamma\left(\frac{n_1}{2} + \alpha_1\right)\Gamma\left(\frac{n_2}{2} + \alpha_2\right)} (E^2)^{\alpha_1}(1 - E^2)^{\alpha_2}. \tag{2.2}$$

By assigning certain values to λ_1 and λ_2 we obtain the following special cases of (2.2)

$$p_1 = p(E^2 \mid \lambda_1, 0, n_1, n_2) = e^{-\lambda_1}(E^2)^{(n_1/2)-1}(1 - E^2)^{(n_2/2)-1} \cdot \sum_{\alpha_1} \frac{\lambda_1^{\alpha_1}\Gamma\left(\frac{n_1 + n_2}{2} + \alpha_1\right)}{\alpha_1!\,\Gamma\left(\frac{n_1}{2} + \alpha_1\right)\Gamma\left(\frac{n_2}{2}\right)} (E^2)^{\alpha_1} \tag{2.3}$$

$$p_2 = p(E^2 \mid 0, \lambda_2, n_1, n_2) = e^{-\lambda_2}(E^2)^{(n_2/2)-1}(1 - E^2)^{(n_2/2)-1} \cdot \sum_{\alpha_2} \frac{\lambda_2^{\alpha_2}\,\Gamma\left(\frac{n_1 + n_2}{2} + \alpha_2\right)}{\alpha_2!\,\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2} + \alpha_2\right)} (1 - E^2)^{\alpha_2} \tag{2.4}$$

$$p_0 = p(E^2 \mid 0, 0, n_1, n_2) = \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} (E^2)^{(n_1/2)-1}(1 - E^2)^{(n_2/2)-1}. \tag{2.5}$$

It is noted that (2.3) is Tang's distribution (112) upon which the calculations of his tables were based. To see this we need only make the correspondence

This paper	*Tang*
λ_1	λ
n_1, n_2	f_1, f_2
α_1	i

We define ϵ to be the probability of an error of the first kind. Tang obtained the critical values E_ϵ^2 of E^2 by requiring that

$$P_I = \int_{E_\epsilon^2}^1 p_0\, dE^2$$
$$= \epsilon \qquad \text{say .01 or .05.}$$

Then he calculated

$$P_{II} = \int_0^{E_\epsilon^2} p_1(E^2 \mid \lambda_1, 0, n_1, n_2)\, dE^2$$

using the values of E_ϵ^2 obtained above. Hence $1 - P_{II}$ is the power of the test.

If, however, $\lambda_1 = 0$ but $\lambda_2 \neq 0$, then to find

$$P_{III} = \int_{E_\epsilon^2}^1 p_2(E^2 \mid 0, \lambda_2, n_1, n_2)\, dE^2$$

we could make the transformation $G^2 = 1 - E^2$ and find

$$P_{III} = \int_0^{1-E_\epsilon^2} p(G^2 \mid 0, \lambda_2, n_1, n_2)\, dG^2.$$

It is easy to verify that

$$p(G^2 \mid 0, \lambda_2, n_1, n_2) = p_1(E^2 \mid \lambda_2, 0, n_2, n_1)$$

if we put G in place of E^2 in the latter density. It follows that to calculate P_{III} it would be sufficient to have full tables of Tang's distribution since

$$P_{III} = \int_0^{1-E_\epsilon^2} p_1(E^2 \mid \lambda_2, 0, n_2, n_1)\, dE^2.$$

Tang's tables are not however sufficiently extensive. Furthermore, tables of (2.2) are also necessary. As yet these tables do not exist. However, some useful conclusions can be drawn from the inequalities obtained in the following section.

First, however, let us evaluate n_1, n_2, λ_1 and λ_2 for the generalized Student's hypothesis discussed in the introduction. It is easy to see that $n_1 = k - 1$ and $n_2 = k(N - 1)$. To evaluate λ_1 and λ_2 we note from Theorem 1 that we only need substitute Ey_{ij} for y_{ij} in q_1 and q_2 where

$$q_1 = N \sum_i (\bar{y}_i - \bar{y})^2$$

$$q_2 = \sum_{i,\nu} (y_{i\nu} - \bar{y}_i)^2.$$

Upon making these substitutions we obtain

$$\lambda_1 = \frac{N}{2\sigma^2} \sum_i (\bar{a}_i - \bar{a})^2$$

$$\lambda_2 = \frac{1}{2\sigma^2} \sum_{i,\nu} (a_{i\nu} - \bar{a}_i)^2.$$

Thus the various hypotheses concerning the a_{ij} influence the distribution of F or $E^2 = 1/(1 + Fn_1/n_2)$ by affecting the values of λ_1 and λ_2.

4. Limits of the values of p. It follows readily from (2.2) that,

$$p = \frac{\Gamma\left(\frac{n_1 + n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} (E^2)^{(n_1/2)-1}(1 - E^2)^{(n_2/2)-1} \tag{3.1}$$

$$\cdot e^{-\lambda_1-\lambda_2} \sum_{\alpha_1,\alpha_2} \frac{\lambda_1^{\alpha_1}\lambda_2^{\alpha_2}}{\alpha_1!\,\alpha_2!} (E^2)^{\alpha_1}(1 - E^2)^{\alpha_2}\, C_{\alpha_1\alpha_2}$$

where

$$C_{\alpha_1\alpha_2} = \frac{\Gamma\left(\frac{n_1 + n_2}{2} + \alpha_1 + \alpha_2\right)\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2} + \alpha_1\right)\Gamma\left(\frac{n_2}{2} + \alpha_2\right)\Gamma\left(\frac{n_1 + n_2}{2}\right)}.$$

Now if $a > 0, b > 0$, and j is an integer > 1, we have

$$\left(1+\frac{a}{b}\right)\left(1+\frac{a}{b+2}\right)\cdots\left(1+\frac{a}{b+2(j-1)}\right) < \left(1+\frac{a}{b}\right)^j.$$

Hence, it follows that

$$1 \le C_{\alpha_1\alpha_2} \le \left(\frac{n_1+n_2}{n_1}\right)^{\alpha_1}\left(\frac{n_1+n_2+2\alpha_1}{n_2}\right)^{\alpha_2}.$$

Substituting we see that

$$p_0 e^{-\lambda_1-\lambda_2}\cdot e^{\lambda_1 E^2+\lambda_2(1-E^2)} \le p \le p_0 e^{-\lambda_1-\lambda_2} \tag{3.2}$$
$$\cdot\exp\left\{\lambda_1 E^2\left(\frac{n_1+n_2}{n_1}\right)\exp\left[\frac{2\lambda_2(1-E^2)}{n_2}\right]+\lambda_2(1-E^2)\left(\frac{n_1+n_2}{n_2}\right)\right\}$$

and

$$p_1 e^{-\lambda_2+\lambda_2(1-E^2)} < p < p_1\exp\left[-\lambda_2+\lambda_2(1-E^2)\left(\frac{n_1+n_2}{n_2}\right)+2\frac{\lambda_2}{n_2}\right]. \tag{3.3}$$

Let $2n_i\phi_i^2 = \lambda_i,\ i = 1, 2.$

THEOREM 3. *Let* $\epsilon = \int_{E_\epsilon^2}^1 p_0\, dE^2$ *so that* ϵ *is the probability of an error of the first kind. Then, for all values of* ϕ_2^2

$$\epsilon > \int_{E_\epsilon^2}^1 p_2\, dE^2 \tag{3.4}$$

and if $E^2 > n_1/(n_1+n_2)$, *it follows that*

$$\epsilon > \epsilon\exp\{-2n_2\phi_2^2 + 2\phi_2^2(1-E_\epsilon^2)(n_1+n_2)\} > \int_{E_1^2}^1 p_2\, dE^2 > \epsilon e^{-n_2\phi_2^2}. \tag{3.5}$$

Furthermore, for all values of ϕ_2^2

$$\int_{E_\epsilon^2}^1 p_1\, dE^2 > \int_{E_\epsilon^2}^1 p\, dE^2, \tag{3.6}$$

and if $E^2 > (n_1+2)/(n_1+n_2)$, *it follows that*

$$\int_{E_\epsilon^2}^1 p_1\, dE_2 > \exp\{-2n_2\phi_2^2 + 2\phi_2^2(1-E_\epsilon^2)(n_1+n_2)\, 2\phi_2^2\}\int_{E_\epsilon^2}^1 p_1\, dE^2 \tag{3.7}$$
$$> \int_{E_\epsilon^2}^1 p\, dE^2 > e^{-2n_2\phi_2^2}\int_{E_\epsilon^2}^1 p_1\, dE^2.$$

Finally, if γ *can assume the two values* 0 *and* 2, *it follows that if*

$$\phi_2^2 > \frac{-\log\delta}{2(E_\epsilon^2(n_1+n_2)-(n_1+\gamma))} > 0, \tag{3.8}$$

then if $\gamma = 0$,

$$\int_{E_\epsilon^2}^1 p_2\, dE^2 < \epsilon\delta \tag{3.9}$$

and if $\gamma = 2$

$$\int_{E_\epsilon^2}^1 p \, dE^2 < \delta \int_{E_\epsilon^2}^1 p_1 \, dE^2. \tag{3.10}$$

Proof. To prove (3.4) and (3.6) it is only necessary to follow Daly's [6] procedure.[2] Since

$$\exp\{-2n_2\phi_2^2 + 2\phi_2^2(1 - E^2)(n_1 + n_2) + \gamma\phi_2^2\}$$

and

$$\exp\{-n_2\phi_2^2E^2\}$$

are decreasing functions of E^2, and

$$\exp\{-2n_2\phi_2^2 + 2\phi_2^2(1 - E^2)(n_1 + n_2) + \gamma\phi_2^2\} < 1$$

if

$$E^2 > \frac{n_1 + \gamma}{n_1 + n_2}$$

the inequalities (3.5) and (3.7) follow immediately from (3.2) and (3.3). Finally

$$\exp\{-2n_2\phi_2^2 + 2\phi_2^2(1 - E^2)(n_1 + n_2) + \gamma\phi_2^2\} < \delta < 1$$

if (3.8) is true, so that (3.9) and (3.10) follow.

From (3.8), (3.9) and (3.10) we can calculate either a lower limit for the bias, if we know ϕ_2, or the upper limit that ϕ_2 can have if we wish the bias to be not greater than some given amount. Thus these limits do not answer the important question of what is a value ϕ_2 such that if $\phi_2 < \phi$ then the bias is less than $(1 - \delta)\epsilon$. They only provide a value ϕ' of ϕ_2 such that if $\phi_2 > \phi'$ then the bias is at least $(1 - \delta)\epsilon$.

If, for example, $\delta = .5$ and $n_1 = 1$ as in the case of Students' ratio; we have if $\gamma = 0$

$$\phi_2^2 > \frac{.693}{2(n_2E^2 - 1)}$$

and if $\epsilon = .05$, then E^2 decreases steadily from .903 if $n_2 = 2$, to .063 if $n_2 = 60$ and the corresponding lower limits of ϕ_2^2 decrease from .43 to .12. Thus, if $\phi_2^2 > .43$ or .12 in these two cases, it follows that the probability of rejecting the average null hypothesis will be not .05 but something less than .025.

If $\delta = .6$ and $n_1 = 4$, $n_2 = 30$ then we can evaluate the lower limit of ϕ_2^2 for the example given in the introduction finding.

$$\phi_2^2 > \frac{.511}{2(.279)(34) - 8} = .05$$

implies a downward bias of at least 40 per cent of .05. Also, if $\phi_2^2 > .07$ then for

[2] The procedure followed is given in [6] on pp. 4, 5, equations (2.2) through Lemma 1.

any value of ϕ_1 the power of the analysis of variance test is reduced at least 40 per cent.

5. Conclusions. The rather sharp effects of a moderate lack of statistical control on the probabilities associated with the F-test indicates the importance of testing for statistical control outside of the industrial applications now made. Furthermore, it would seem advisable to investigate tests and designs that are less sensitive to the lack of control than is the F-test.

REFERENCES

[1] P. C. Tang, "The power function of the analysis of variance tests with tables and illustrations of their use," *Statistical Research Memoirs*, Vol. 2 (1938), pp. 126–157.

[2] William G. Madow, "The distribution of quadratic forms in non-central normal random variables," *Annals of Mathem. Stat.*, Vol. 9 (1940), pp. 100–104.

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

[4] A. T. Craig, "Note on the independence of certain quadratic forms," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 195–197.

[5] S. Kolodziejczyk, "On the important class of statistical hypotheses," *Biometrika*, Vol. 27 (1935), pp. 161–190.

[6] J. F. Daly, "On the unbiased character of likelihood-ratio tests for independence in normal systems," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 1–33. The procedure followed is given on pp. 4, 5, equations (2.2) through Lemma 1.