The problem in statistics of testing several variances. By M. S. BARTLETT, B.A., Queens' College. (Communicated by Mr J. WISHART.)

[*Received* 7 February, *read* 5 March 1934.]

1. The test of significance of the difference between two *means* was extended by means of Fisher's methods of analysis of variance to test jointly the differences between several means, the test being essentially the same as the test of significance of two variances. But though it is true in many cases that we are chiefly concerned with two independent estimates of variance, and the method of analysis of variance is available, the problem of jointly testing the *variances* obtained from several independent samples does not fit obviously into this scheme.

This problem has been tackled by Neyman and Pearson(1) by means of a criterion based on the principle of likelihood. The purpose of this note was primarily to suggest a possible alternative procedure which though having certain manifest disadvantages may prove to be of use when a relatively small number of samples, possibly unequal in size, is being considered. Some discussion has, however, also been added on the extension of the problem to the case of two variates.

2. If we have three independent sums of squares of a normal deviate x_1,

$$c_{11}(n_1),\ c_{11}(n_2),\ c_{11}(n_3) \equiv A,\ B,\ C$$

respectively, where n_1, n_2, n_3 denote the numbers of degrees of freedom, then from geometrical considerations it is apparent that the quantities A/B and $A+B$ are independent, since one depends on the angle between the radius vector representing $A+B$ and one of its components, and the other on its length. It follows that the ratios A/B, $C/(A+B)$ are independent; if D represents a fourth sum of squares, $D/(A+B+C)$ is independent of the first two ratios, and so on.

It is perhaps advisable to give a formal proof of this result. The joint distribution of A, B and C is known to be given by

$$p(A,B,C) \propto A^{\frac{1}{2}n_1-1} B^{\frac{1}{2}n_2-1} C^{\frac{1}{2}n_3-1} e^{-\frac{1}{2}(A+B+C)/\sigma^2} dA\, dB\, dC. \quad (1)$$

Writing $B/A = v_2$, $C/A = v_3$, and integrating with regard to A, we obtain

$$p(v_2, v_3) \propto \frac{v_2^{\frac{1}{2}n_2-1} v_3^{\frac{1}{2}n_3-1}\, dv_2\, dv_3}{(1+v_2+v_3)^{\frac{1}{2}(n_1+n_2+n_3)}}. \quad (2)$$

https://doi.org/10.1017/S0305004100016571 Published online by Cambridge University Press

Writing $B/A = u \equiv v_2$, $C/(A+B) = v = v_3/(1+v_2)$, we then have

$$p(u,v) \propto \frac{u^{\frac{1}{2}n_2-1} v^{\frac{1}{2}n_3-1} (1+u)^{\frac{1}{2}n_3}\, du\, dv}{(1+u+v[1+u])^{\frac{1}{2}(n_1+n_2+n_3)}}$$

$$\propto \frac{u^{\frac{1}{2}n_2-1}\, du}{(1+u)^{\frac{1}{2}(n_1+n_2)}} \cdot \frac{v^{\frac{1}{2}n_3-1}\, dv}{(1+v)^{\frac{1}{2}(n_1+n_2+n_3)}}$$

$$= p(u)\, p(v). \qquad (3)$$

3. These independent comparisons of variances are analogous to the independent comparisons possible with a number of means.

As a particular case, we have the t test for the difference of means of two samples independent of the test of the ratio of their variances, a fact pointed out by Neyman and Pearson (2).

Clearly, just as the likelihood criterion λ for the identity of the normal populations from which two samples were drawn splits up into the criteria corresponding to each of these two tests, that is,

$$\lambda = \lambda_1 \lambda_2 \qquad (4)$$

say, so above for the case of three sums of squares we shall have a similar relation (4), where now λ will represent the likelihood criterion found by Neyman and Pearson corresponding to their test of significance between these sums of squares, and λ_1, λ_2 the component criteria for A and B, C and $A+B$.

This process of splitting up the required comparison of A, B and C into independent comparisons, unlike the natural segregating off of the means, is evidently somewhat artificial, and any test based on this principle would not therefore appear to be completely efficient. Nevertheless, the test has the advantage of corresponding to known distributions, and, provided that not too many variances are being considered, may fairly rapidly be applied.

The point of making independent comparisons is that we are able after finding the corresponding probability levels to combine these levels into a composite test, by the suggestion originally made by Fisher for combining independent tests of significance (see (3), p. 97).

To do this, the exact probabilities are required, and Fisher's z table is not sufficient; we may, however, refer to the recently published tables of the incomplete beta function (4).

4. As an example, consider the sums of squares of deviations for three samples whose variation we wish to compare:

A, ·5004 ..... n_1, 9

B, ·1866 ..... n_2, 8

C, ·0728 ..... n_3, 6.

https://doi.org/10.1017/S0305004100016571 Published online by Cambridge University Press

It is evident that, by considering A, B, C in a different order, the final significance level will differ; the test would appear most efficient if we use as many degrees of freedom as possible by arranging the variances according to their degrees of freedom. If the degrees of freedom are anywhere the same, the convention would naturally be adopted of taking the corresponding variances as they first stand.

From the tables quoted, the significance levels are

$$A/B \ \ldots\ldots\ldots \ P_1 = \cdot 237,$$
$$C/(A+B) \ \ldots\ldots\ldots \ P_2 = \cdot 149.$$

Hence the combined significance level is obtained from

$$\chi^2 = -2(\log P_1 + \log P_2)$$
$$= 6 \cdot 71,$$

with 4 degrees of freedom, giving $P > \cdot 10$. The differences in variance thus cannot be said to be significant.

Since we are testing any departure from the expected ratios on our hypothesis of homogeneity, we consider both "tails" of the distribution; that is, we take P_1 or P_2 to be equal to $2P'$ or $2(1-P')$, whichever is less than unity, where P' is the corresponding probability level given in the table.

The value of this test in comparison with the test developed by Neyman and Pearson must obviously remain a matter for further consideration, in view of the disadvantages briefly indicated above; but it might also be noted that if the combined test does give a significant result, the independent probabilities we have found will be of use to us in examining where abnormal variation lies.

5. *Extension to two variates.*

Pearson and Wilks(5) have extended to the case of two variates the likelihood criteria developed by Neyman and Pearson for the case of k samples in one variate. Corresponding to the usual test of means in the case of one variate by an analysis of variance, they consider

$$L_2^2 = \frac{|c_{ij}(N-k)|}{|c_{ij}(N-1)|},$$

where $|c_{ij}| = c_{11}c_{22} - c_{12}^2$, c_{11} denoting a sum of squares for the first variate x_1, etc; N is the total number of members in the k samples; and $N-k$ and $N-1$ denote the number of degrees of freedom for the two (c_{ij}) considered. The notation for L_2^2 given above is sufficient to indicate that $c_{ij}(N-k)$ refer to the matrix elements when all the sample means have been eliminated, $c_{ij}(N-1)$ when the total mean only has been eliminated; $c_{ij}(n)$ is thus specified by its number of degrees of freedom n.

https://doi.org/10.1017/S0305004100016571 Published online by Cambridge University Press

The distribution of L_2 is shown to be

$$p(L_2) \propto L_2^{N-k-2}(1-L_2)^{k-2}dL_2, \tag{5}$$

the probability integral being thus an incomplete beta function. This can be seen by considering the moments of the factors of L_2^2, for

$$L_2^2 = \frac{c_{11}(N-k)}{c_{11}(N-1)} \frac{c_{22\cdot1}(N-k-1)}{c_{22\cdot1}(N-2)}, \tag{6}$$

where the numerator of the second factor has one less degree of freedom than that of the first. A sum of squares $c_{22\cdot1}$ is a function of the partial variate $x_{2\cdot1}$, the dependence of x_2 on x_1 having been eliminated (see (6), p. 261).

It will follow that any similar quantity

$$\frac{|c_{ij}(n_1)|}{|c_{ij}(n_1+n_2)|} = \frac{c_{11}(n_1)}{c_{11}(n_1+n_2)} \frac{c_{22\cdot1}(n_1-1)}{c_{22\cdot1}(n_1+n_2-1)} \tag{7}$$

is similarly distributed. Unfortunately we cannot use this fact to test the difference in variation for two samples, since the quantity in (7) is not in that case a suitable criterion to use. The relation (7) indicates, however, that the comparison of the variation in a second sample with that in the first, and that of a third sample with the total variation in the first two, will be independent in exactly the same way as for one variate; and the likelihood criterion for variation of the samples about their means will correspondingly factorize. That is,

$$\lambda = \lambda_1 \lambda_2,$$

as in equation (4).

It should be noticed that $c_{22\cdot1}(n_1-1)/c_{22\cdot1}(n_1+n_2-1)$ does not denote a straightforward comparison of the component partial variances for x_2, since this direct comparison will be of the form $c_{22\cdot1}(n_1-1)/c_{22\cdot1}(n_1+n_2-2)$. In fact the additional degree of freedom represents the difference in regression coefficients between the two samples, as can easily be verified algebraically. Hence we can write (7) finally as a product of three factors, the first a comparison of $c_{11}(n_1)$ with $c_{11}(n_2)$, the second that of $c_{22\cdot1}(n_1-1)$ with $c_{22\cdot1}(n_2-1)$, and the third that of $c_{22\cdot1}(1)$—denoting the difference in regression coefficients—with $c_{22\cdot1}(n_1+n_2-2)$. The third factor is of course independent of the second. The likelihood criterion λ for the variation for two samples will naturally split up into three corresponding factors.

Thus in problems where the association between the two variates can be regarded as one of *regression*, we are now in a position to consider possible alternatives to the tests developed by Pearson and Wilks. The set of variances for the first variate x_1, and the set of partial variances for the variate $x_{2\cdot1}$ could each be

https://doi.org/10.1017/S0305004100016571 Published online by Cambridge University Press

considered separately, either by means of the test developed by Neyman and Pearson, or by the method suggested in this note. The differences of the regression coefficients for the several samples will naturally be combined into a single test analogous to the test of sample means in one variate. That is, we may consider

$$c_{22\cdot1}(k-1) \equiv \sum^{k} c_{11}(b_{21}-B_{21})^2,$$

summed over the k samples, where B_{21} denotes the combined estimate from all the samples, and compare this quantity with the independent sum of squares $c_{22\cdot1}(N-2k)$, this test being independent of the mutual comparisons between the components of $c_{22\cdot1}(N-2k)$. We may notice that tests based on quantities $c_{22\cdot1}$ will be independent of the distribution of x_1, and can obviously be extended to meet the case of non-linear regression.*

6. As an example of the analysis of quantities $c_{22\cdot1}$, we may consider the data discussed by Pearson and Wilks ((5), p. 371), consisting of 5 samples of 12 joint observations on the tensile strength and hardness in aluminium die-castings. If we are to be able to predict "strength" x_2 from "hardness" x_1, the regression of x_2 on x_1 and the residual or partial variation $x_{2\cdot1}$ should both be stable. Assuming here the result found that the partial variation was stable within samples, we have, for the remaining analysis of $x_{2\cdot1}$, the following sums of squares $c_{22\cdot1}(n) = nv_{22\cdot1}$.

	$c_{22\cdot1}$	n	$v_{22\cdot1}$
(1) Partial variation within samples	245·24	50	4·905
(2) Differences between sample regression coefficients	14·416	4	3·604
(3) Difference between coefficient from means, and "mean" of coefficients in (2)	6·324	1	6·324
(4) Partial variation of sample means	236·42	3	78·81
Partial variation for whole	502·40	58	

Comparing (2) with (1), we see, without having to test the ratio, that the differences in the regression coefficients for the

* I should like to correct a misleading statement made in the paper of mine quoted (6); it was concluded there (p. 280) that we can only test the differences between regression coefficients provided that we could regard the values of the independent variate as fixed: I have since realized that this condition is unnecessary, or, equivalently, that we can *always* regard these values as fixed.

https://doi.org/10.1017/S0305004100016571 Published online by Cambridge University Press

5 samples are evidently consistent with our hypothesis of homogeneity; comparing (3) with (1) + (2), that the regression coefficient obtained from the means is also consistent; but on comparing (4) with (1) + (2) + (3), that the "partial variation" of mean strength from sample to sample is extremely heterogeneous. The same conclusion is therefore reached as in the paper (5) from which the example is taken, namely, that we could not use hardness as an index of strength while the component of strength not expressible in terms of hardness fluctuated so extremely between samples.

7. In problems where we wish to test the hypothesis of homogeneity without specifying in what way heterogeneity may arise, it is evident that the use of a single comprehensive test, such as the one developed on the principle of likelihood, has much to recommend it; it may in fact be regarded as a general test of goodness of fit.

If, however, we are definitely interested in specific factors, such, for example, as the means, it is evident that tests designed for those factors only will be more sensitive to possible discrepancies in them. In our choice of tests, this must therefore be borne in mind. If we do choose to test independent factors separately, we have seen that we can always if necessary combine our results into a single significance level.

REFERENCES.

(1) J. NEYMAN and E. S. PEARSON, *Bulletin de l'Académie Polonaise des Sciences et des Lettres*, A (1931), 460–481.

(2) —— —— *ibid.*, A (1930), 73–96.

(3) R. A. FISHER, *Statistical Methods for Research Workers* (4th ed. 1932).

(4) K. PEARSON, *Tables of the Incomplete Beta-Function* (1934).

(5) E. S. PEARSON and S. S. WILKS, *Biometrika*, 25 (1933), 353–378.

(6) M. S. BARTLETT, *Proc. Roy. Soc. Edin.*, 53 (1933), 260–283.

https://doi.org/10.1017/S0305004100016571 Published online by Cambridge University Press