NOTES

This section is devoted to brief research and expository articles, notes on methodology and other short items.

NOTE ON AN APPROXIMATE FORMULA FOR THE SIGNIFICANCE LEVELS OF Z

By W. G. Cochran

1. **Introduction.** An important part has been played in modern statistical analysis by the distribution of $z = \frac{1}{2} \log \frac{s_1^2}{s_2^2}$, when s_1^2 and s_2^2 are two independent estimates of the same variance. In particular, all tests of significance in the analysis of variance and in multiple regression problems are based on this distribution. Complete tabulation of the frequency distribution of z is a heavy task, because the distribution is a two-parameter one, the parameters being the number of degrees of freedom, n_1 and n_2 in the estimates s_1^2 and s_2^2. Thus each significance level of z requires a separate two-way table. Fisher constructed a table of the 5 percent points in 1925 [1], and this has since been extended by several workers [2] to the 20, 1, and 0.1 percent level for a somewhat wider range of values of n_1 and n_2.

With his original table, Fisher gave an approximate formula for the 5 percent values of z, for high values of n_1 and n_2 outside the limits of his table. The formula reads:

$$z \text{ (5 percent)} = \frac{1.6449}{\sqrt{h-1}} - 0.7843\left(\frac{1}{n_1} - \frac{1}{n_2}\right), \tag{1}$$

$$\text{where } \frac{2}{h} = \frac{1}{n_1} + \frac{1}{n_2}.$$

The constant 1.6449 is the 5 percent significance level for *a single tail* of the normal distribution, and the constant 0.7843 will be found to be $\frac{1}{6}\{2 + (1.6449)^2\}$. Thus the general formula for the significance levels of z derivable from (1) is

$$z = \frac{x}{\sqrt{h-1}} - \left(\frac{x^2 + 2}{6}\right)\left(\frac{1}{n_1} - \frac{1}{n_2}\right),$$

where x is a normal deviate with unit standard error. By inserting the appropriate significance level of x, this formula has been extended [2] to the tables of the 20, 1, and 0.1 percent levels of z and commonly appears with all published tables of z. The objects of this note are to indicate the derivation of the formula and to suggest an improvement upon it in the latter cases.

2. The transformation of the z-distribution to normality. For high values of n_1 and n_2, the distribution of z approaches the normal distribution, the principal deviation being a slight skewness introduced by the inequality of n_1 and n_2. It is therefore natural to seek an approximate formula for the distribution of z by examining its relation to the normal distribution. For the z-distribution the ratio $\kappa_r/\kappa_2^{r/2}$, where κ_r is the r^{th} cumulant, is of the order $n^{-(\frac{1}{2}r-1)}$, where n is the smaller of n_1 and n_2. This property is common to a large number of distributions which tend to normality; for example, the distribution of the mean of a sample of size n from any distribution with finite cumulants. Fisher and Cornish [3] have recently given a method, applicable to all distributions with this property, for transforming the distribution to a normal distribution to any desired order of approximation. They also obtained explicit expressions for the significance levels of the original distribution in terms of the significance levels of the normal distribution, discussing the z-distribution as a particular example. The relation between z and the normal deviate x at the same level of probability was found to be

$$(2)\quad z = \frac{x}{\sqrt{h}} - \tfrac{1}{6}(x^2+2)\left(\frac{1}{n_1}-\frac{1}{n_2}\right) + \frac{1}{\sqrt{h}}\left\{\frac{x^3+3x}{12\,h} + \frac{x^3+11x}{144}h\left(\frac{1}{n_1}-\frac{1}{n_2}\right)^2\right\},$$

the three terms on the right hand side being respectively of order $n^{-\frac{1}{2}}$, n^{-1}, and $n^{-\frac{3}{2}}$, so that terms of order n^{-2} are neglected.[1]

If this equation is compared with equation (1), the latter appears at first sight to be the approximation of order n^{-1} to the z-distribution, except that the divisor of x is $\sqrt{h-1}$ in (1) and $\sqrt{h}$ in (2). Computation of a few values shows that at the 5 percent level, equation (1) is the better approximation. For example, for $n_1 = 40$, $n_2 = 60$, (1) gives z (5 percent) $= .2334$, (2) gives .2309, and the exact value is .2332.

Since

$$\frac{x}{\sqrt{h-1}} = \frac{x}{\sqrt{h}} + \frac{x}{2h\sqrt{h}} + \text{terms of order } n^{-2},$$

Fisher's approximation differs from (2) by including a correction term of order $n^{-\frac{3}{2}}$. Inspection of the true correction terms of this order in equation (2) shows that for finite values of n_1 and n_2 the term $\frac{x^3+11x}{144}\sqrt{h}\left(\frac{1}{n_1}-\frac{1}{n_2}\right)^2$ is considerably smaller than the term $\frac{x^3+3x}{12h\sqrt{h}}$, since the former has a smaller numerical coefficient and involves the difference between $\frac{1}{n_1}$ and $\frac{1}{n_2}$. Thus Fisher's formula gives a close approximation to the true formula of order $n^{-\frac{3}{2}}$, provided that $\frac{x}{2}$ is approximately equal to $\frac{x^3+3x}{12}$; i.e. if $\frac{x^2+3}{6}$ is approximately equal

[1] Fisher and Cornish also gave the two succeeding terms.

to 1. For the 5 percent level, $x = 1.6449$, and $\frac{x^2 + 3}{6} = 0.951$. Thus at the 5 percent level the use of $\sqrt{h - 1}$ in (1) instead of $\sqrt{h}$ extends the validity of Fisher's approximation from order n^{-1} to order $n^{-\frac{3}{2}}$.

This ingenious device, however, requires adjustment at other levels of significance. The values of $(x^2 + 3)/6$ at the principal significance levels are shown below.

Significance level—%	40	30	20	10	5	1	0.1
$\lambda = (x^2 + 3)/6$	0.51	0.55	0.62	0.77	0.95	1.40	2.09

If $\sqrt{h - 1}$ in formula (1) is replaced by $\sqrt{h - \lambda}$, with the above values of λ, Fisher's formula will be approximately valid to order $n^{-\frac{3}{2}}$ at all levels of significance. In particular, for the tables already published of the 20, 1 and 0.1 percent points, λ may be taken as 0.6, 1.4 and 2.1 respectively. The values of z given by the use of $\sqrt{h - 1}$ and $\sqrt{h - \lambda}$ are compared below for $n_1 = 24$, $n_2 = 60$.[2]

Significance Level	Approximate formula		Exact value
	$\sqrt{h - 1}$	$\sqrt{h - \lambda}$	
20%	.1346	.1337	.1338
1%	.3723	.3748	.3746
0.1%	.4875	.4966	.4955

The use of $\sqrt{h - \lambda}$ gives values practically correct to 4 decimal places, except for the 0.1 level of significance, at which the higher terms become more important.

With the aid of this formula, complete tabulation of the z-distribution for a given pair of high values of n_1 and n_2 is relatively simple. If very low probabilities at the tails are required, the further approximations given by Fisher and Cornish [3] may be used.

REFERENCES

[1] R. A. Fisher. *Statistical Methods for Research Workers.* Edinburgh, Oliver and Boyd. 1st Ed. 1925.

[2] R. A. Fisher and F. Yates. *Statistical Tables.* Edinburgh, Oliver and Boyd. 1938.

[3] E. A. Cornish and R. A. Fisher. "Moments and Cumulants in the Specification of Distributions," *Revue de l'Institut International de Statistique,* Vol. 4 (1937).

Iowa State College,
Ames, Iowa.

[2] The numerical terms in the approximate formula given for the 20 percent points on p. 28 of Fisher and Yates' *Statistical Tables* are in error. Their formula should read:

$$z = \frac{0.8416}{\sqrt{h - 1}} - 0.4514\left(\frac{1}{n_1} - \frac{1}{n_2}\right)$$