AN ALTERNATIVE TO MARASCUILO’S "LARGE-SAMPLE MULTIPLE COMPARISONS" FOR PROPORTIONS

JACOB COHEN
New York University

An alternative to Marascuilo's $\chi^2$ analogue of Scheffé's theorem for performing tests on multiple linear contrasts among proportions coming from K independent populations is described. Based on the arcsin transformation of $p$, the proposed alternative is suitable for small samples, may be used when a sample $p$ is 0 or 1, and is computationally simpler. Tests of complex contrasts are also exemplified.

Marascuilo's (1966) recent article constitutes a most useful contribution to applied statistical methodology in providing a solution to a variety of frequently recurring problems which arise following omnibus null hypothesis tests, that is, where more than two independent samples are subjected to a test of the equality of their respective population parameters. By means of a $\chi^2$ analogue to Scheffé's theorem (1959), he showed how it is possible to perform the omnibus test of the equality of $K > 2$ parameter values by means of a statistic, $U'$, which is $\chi^2$ distributed; and, more important, how then to perform linear contrasts of the parameters subject to an experimentwise control of the $\alpha$ error risk. For general background, see Marascuilo (1966).

The present note suggests an alternative procedure for one of his applications, that of multiple comparisons among proportions, his Example 2. The alternative stays within the spirit of his procedure, but uses the arcsin transformation of the sample proportions instead of the proportions themselves. The advantages of the alternative are: (a) It is suitable for use with small samples. (b) It can be used when sample proportions are 0 or 1. (c) It is computationally simpler.

The difficulty which arises in statistical inference with the sample proportion $p$ is the same as that with the sample correlation coefficient—their standard errors are not only a function of the sample $N$, but also of their unknown population parameters. In classical inference, with large enough samples, the sample statistics are substituted for the unknown parameters with the expectation that the sampling error discrepancy is tolerably small. A superior solution, not dependent upon large samples, results when one can find a transformation of the function whose sampling distribution is normal (or, at least, known), and whose sampling error is free of any dependence on the unknown parameter value. This is exemplified by the widely known transformation of $r$ to $Z$ of Fisher, and was utilized by Marascuilo (1966, pp. 281–283) in his multiple comparison among correlation coefficients. In other words, the standard error of a sample $Z$ is a function solely of $N$, that is, $(N - 3)^{-1}$.

There is available a transformation of $\rho$, also due to Fisher but not as widely known among behavioral scientists, which has exactly the same function. It is approximately normally distributed, even for small samples, with standard error solely a function of $N$. It is, in general,

$$\Phi = 2 \arcsin \sqrt{\rho},$$

and when $\rho$ equals 0 or 1,

$$\Phi = 2 \arcsin \sqrt{1/(4N)} \text{ for } \rho = 0,$$

$$\Phi = \pi - \Phi \text{ for } \rho = 1.$$

The variance of any value of $\Phi$ is simply $1/N$. Values for $\Phi$, $\Phi_0$, and $\Phi_1$ are extensively tabled in Owen (1962).

This note proposes that Marascuilo's treatment of proportions be made consistent with his treatment of correlation coefficients, thus have the same desirable properties. Using his notation, and the data used in his Example 2, one is testing the statistical hypothesis:

$$H_0: p_1 = p_2 = \cdots = p_k = p_0$$

Supported by Grant No. MH 06137 from the National Institute of Mental Health of the United States Public Health Service.
by means of the arcsin transformation. For the proposed alternative, 
\[ U'_0 = \sum_{k=1}^{K} (\Phi_k - \Phi_0)^2 \]

Since \( W_k = 1/\text{var}(\Phi_k) \), and \( \text{var}(\Phi_k) = 1/N_k \) for this test \( W_k = N_k \), the sample size. Thus,
\[ U'_0 = \sum_{k=1}^{K} N_k (\Phi_k - \Phi_0)^2 \]

with
\[ \Phi_0 = \frac{\sum_{k=1}^{K} N_k \Phi_k}{\sum_{k=1}^{K} N_k} \]
a a sample-size-weighted mean of the \( \Phi_k \)'s. 

\( U'_0 \) is \( \chi^2 \) distributed with \( (K - 1) \) df.

For Marascuilo's (1966, Table 2) example, computations are summarized in Table 1. For this example, 
\[ \Phi_0 = \frac{76(1.3317) + 228(1.1167) + 168(0.9183)}{76 + 228 + 168} = 1.0807 \]

and
\[ U'_0 = 76(1.3317 - 1.0807)^2 + 228(1.1167 - 1.0807)^2 + 168(0.9183 - 1.0807)^2 = 9.51. \]

This value agrees well with that of Marascuilo's method (9.59), the usual Pearson \( \chi^2 \) formula (9.59), and Goodman's \( Y^2 \) statistic (9.58), all of which lead to the same decision, since the \( \alpha = .05 \) \( \chi^2 \) value for 2 df is 5.99. However, the interest lies in the follow-up multiple contrasts.

Since it is known that at least one linear contrast among the parameters is significantly different from zero, following Marascuilo, the general form of the \( (1 - \alpha)\% \) set of simultaneous confidence intervals about the simple contrasts is
\[ (\Phi_{k_1} - \Phi_{k_2}) \pm \sqrt{X_{K-1}^2(1 - \alpha)} \sqrt{\frac{1}{N_{k_1}} + \frac{1}{N_{k_2}}}. \]

For this example
\[ \sqrt{X_{K-1}^2(1 - \alpha)} = \sqrt{X_{2}^2(95)} = \sqrt{5.99} = 2.45. \]

The simple contrasts thus are
\[ (1.3317 - 1.1167) \pm 2.45 \sqrt{0.013158 + 0.004386} = - .1096 < \Phi_1 - \Phi_2 < .5396 \text{ Not Significant} \]
\[ (1.3317 - 0.9183) \pm 2.45 \sqrt{0.013158 + 0.005952} = .0748 < \Phi_1 - \Phi_3 < .7520 \text{ Significant} \]
\[ (1.1167 - 0.9183) \pm 2.45 \sqrt{0.004386 + 0.005952} = -.0508 < \Phi_2 - \Phi_3 < .4476 \text{ Not Significant} \]

The results are the same as Marascuilo's. Note that in Scheffe tests, such as this, the \( \alpha \) criterion is experimentwise, that is, the probability of any (one or more) contrasts being significant when the omnibus \( H_0 \) is true is no greater than \( \alpha \).

Marascuilo illustrated only simple contrasts for proportions in his article. If a contrast such as
\[ \Psi = a_1 \Phi_1 + a_2 \Phi_2 + \cdots + a_K \Phi_K \quad (\sum_{k=1}^{K} a_k = 0) \]
has more than two \( a_k \) values which are not zero, then \( \text{var}(\Psi) \) for independent parameters is estimated by
\[ \text{var}(\Psi) = \sum_{k=1}^{K} \frac{a_k^2}{N_k}, \]
and all linear contrasts \( \Psi \) are estimated by
\[ \Psi = \frac{\Phi_1 + \Phi_2}{2} - \Phi_2 = \frac{1}{2} \Phi_1 + \frac{1}{2} \Phi_2 - (1) \Phi_3. \]

For the above example, assume a post hoc interest in the complex contrast
\[ \Psi = \frac{\Phi_1 + \Phi_2}{2} - \Phi_3. \]

For the example,
\[ \Psi = 0.5(1.3317) + 0.5(1.1167) - .9183 = .3059 \]
is an unbiased estimate (expressed in units of \( \Phi \)).
The $a_k$ coefficients for $\Psi$ are $1/2, 1/2,$ and $-1$, respectively, and sum to zero as required. For the example,

$$\text{var}(\Psi) = \frac{(\frac{1}{2})^2}{76} + \frac{(\frac{1}{2})^2}{228} + \frac{(-1)^2}{168} = .010338.$$  
$$\sqrt{\text{var}(\Psi)} = .1017 \quad \text{and} \quad \sqrt{X^2_{.95}} = 2.45.$$  

The 95% interval for $\Psi$ is therefore

$.3059 \pm 2.45(.1017)$, or  
$.0567 < \Psi < .5550$ Significant.

That is, since the limits of the 95% interval for $\Psi$ are both positive, the null hypothesis that the $\Phi$ agreeing in Population 3 (high education level) is equal to that of the mean of the $\Phi$s of the other two populations can be rejected with risk no more than 5% in favor of the alternative that the former is greater.

One would ordinarily prefer to make the above statement with regard to the $p$'s rather than the $\Phi$s, as one can in simple contrasts. For complex contrasts, which always involve means of subsets of parameters, such a formulation would not be strictly valid, since the arcsin transformation is not a linear one. Thus, in general, a mean of two or more $\Phi$s of $p$'s is not equal to the $\Phi$ of the mean of the $p$'s. Fortunately, however, over the middle range of $p$'s (say, .25-.75), and outside these limits over short distances (say, .10-.15), the arcsin transformation is virtually linear in $p$. Therefore, for complex contrasts involving means over the middle range of $p$ or, generally, over short distances along the $p$ scale, conclusions drawn in terms of $p$'s do not risk serious inaccuracy. Such would be the case in the example.

In a similar way, other complex contrasts can be made. Note that the significance of the omnibus $U_0$ test theoretically guarantees that at least one of the possible contrasts is significant, but this need not be a simple contrast involving only the difference between two $p$'s; it may be a complex contrast involving any number of $p$'s from 3 to $K$.

As Marascuilo noted in connection with the test on $r$'s (1966, pp. 282 ff) the $Z$ transformation assures the validity of the test for small samples. The proposed alternative for tests on proportions similarly assures small-sample validity. The latter is a particularly important advantage of this alternative procedure, since the problem of multiple comparisons among $p$'s coming from samples at least one of which is small arises frequently in data analysis. Further, the inadequacy of the classical direct comparison of $p$'s when a sample $\bar{p}$ is 0 or 1 is apparent in that, by direct substitution, its sampling variance is estimated as zero, a very dubious procedure, and particularly so when $N$ is small. Finally, since the $W_k$'s in the present approach are simply the sample sizes, computation is simplified.

**REFERENCES**


(Received June 8, 1966)