ON TESTING A SET OF CORRELATION COEFFICIENTS FOR EQUALITY

By D. N. Lawley

University of Edinburgh

0. Summary. The problem of testing a set of correlation coefficients for equality is discussed. We first generalize a result of Anderson and then provide a criterion of large-sample χ^2 type.

1. The generalization of a result of Anderson. In his preceding paper Anderson [1] has discussed (see Section 5 and Appendix A) the hypothesis that the $p - 1$ smallest latent roots of a correlation matrix of order p are all equal to some unknown value λ, which is equivalent to the hypothesis that the variates are equally correlated with correlation coefficient ρ, where $\lambda = 1 - \rho$. The set of p variates is assumed to follow a multivariate normal distribution.

Let r_{ij} $(i, j = 1, 2, \cdots, p)$ be the sample correlation coefficient between the ith and jth variates found from a random sample of size $n + 1$. Then, adopting Anderson's procedure, we put $y_{ij} = (r_{ij} - \rho)n^{\frac{1}{2}}$ $(i \neq j)$. Taking $i < j$, we have a set of $\frac{1}{2}p(p - 1)$ variates which are asymptotically normally distributed such that

$$E(y_{ij}^2) = \lambda^2(1 + \rho)^2,$$

$$E(y_{ij}y_{ik}) = \tfrac{1}{2}\lambda^2\rho(2 + 3\rho) \qquad (j \neq k),$$

$$E(y_{ij}y_{hk}) = 2\lambda^2\rho^2 \qquad \text{(no subscripts equal)}.$$

The results obtained by Anderson show that Bartlett's criterion [2] for testing the hypothesis is asymptotically equal to

$$(1.1) \quad (1/2\lambda^2)\{\textstyle\sum y_{ij}^2 - (2/p) \sum y_{ij}y_{ik} + [(p - 2)/p^2(p - 1)](\sum y_{ij})^2\},$$

where the summations are over all pairs of unequal suffices.

For the case where $p = 3$ Anderson has proved that the above expression is distributed asymptotically as $(1 - \lambda^2/3)\chi_2^2$, where χ_r^2 denotes a χ^2 variate with r degrees of freedom. A knowledge of this result gave me the idea of generalizing it for any value of p.

In the ensuing algebra it will be convenient to write

$$Y_k = \sum_i y_{ik}, \qquad \bar{y}_k = Y_k/(p - 1),$$

$$Y = \sum_{i<j} y_{ij}, \qquad \bar{y} = 2Y/\{p(p - 1)\},$$

$$y_{(ij)} = y_{ij} - [(p - 1)/(p - 2)](\bar{y}_i + \bar{y}_j) + [p/(p - 2)]\bar{y}.$$

Received June 26, 1961; revised April 26, 1962.

It is easy to show that

$$\sum_{i \neq j} y_{(ij)}^2 = \sum_{i \neq j} (y_{ij} - \bar{y})^2 - [2(p-1)^2/(p-2)] \sum_k (\bar{y}_k - \bar{y})^2.$$

Hence expression (1.1) may be put in the form

$$(1.2) \quad \begin{aligned} &(1/2\lambda^2)\{\sum_{i \neq j} (y_{ij} - \bar{y})^2 - [2(p-1)^2/p] \sum_k (\bar{y}_k - \bar{y})^2\} \\ &\qquad = (1/\lambda^2)\{\sum_{i<j} y_{(ij)}^2 + [2(p-1)^2/p(p-2)] \sum_k (\bar{y}_k - \bar{y})^2\}. \end{aligned}$$

We now construct new variates x_{ij} $(i \neq j)$ given by

$$\lambda x_{ij} = \alpha y_{ij} + \beta(Y_i + Y_j) + \gamma Y,$$

where α (positive), β and γ are chosen such that the x_{ij} (for $i < j$) are uncorrelated and such that each has unit variance. Define X_k, $\bar{x}_k$, X, $\bar{x}$ and $x_{(ij)}$ in the same way as Y_k, $\bar{y}_k$, etc. A simple method of determining the values of α, β and γ is obtained by noting that (for $p > 3$)

$$\lambda(x_{12} + x_{34} - x_{13} - x_{24}) = \alpha(y_{12} + y_{34} - y_{13} - y_{24}),$$

$$\lambda(X_1 - X_2) = \{\alpha + (p-2)\beta\}(Y_1 - Y_2),$$

$$\lambda X = \{\alpha + 2(p-1)\beta + \tfrac{1}{2}p(p-1)\gamma\}Y.$$

By equating in each of these relations the variances of the two sides we find that $\alpha = 1$, $\{1 + (p-2)\beta\}^2 = 2/\{p - (p-2)\lambda^2\}$,

$$1 + 2(p-1)\beta + \tfrac{1}{2}p(p-1)\gamma = 1/\{1 + (p-1)\rho\}.$$

We have also $y_{(ij)} = \lambda x_{(ij)}$, $(\bar{y}_i - \bar{y})^2 = \lambda^2(\bar{x}_i - \bar{x})^2/\{1 + (p-2)\beta\}^2 = \tfrac{1}{2}\lambda^2\{p - (p-2)\lambda^2\}(\bar{x}_i - \bar{x})^2$. Hence expression (1.2) may be transformed into

$$(1.3) \quad \sum_{i<j} x_{(ij)}^2 + \{1 - [(p-2)/p]\lambda^2\}[(p-1)^2/(p-2)] \sum_k (\bar{x}_k - \bar{x})^2.$$

Now consider the identity

$$\sum_{i<j} (x_{ij} - \bar{x})^2 = \sum_{i<j} x_{(ij)}^2 + [(p-1)^2/(p-2)] \sum_k (\bar{x}_k - \bar{x})^2.$$

The left hand side of this is distributed (asymptotically) as χ^2 with $\tfrac{1}{2}p(p-1) - 1 = \tfrac{1}{2}(p+1)(p-2)$ degrees of freedom. The second term on the right hand side is distributed as χ_b^2, where $b = p - 1$, since each of the p variates $\bar{x}_k$ has variance $1/(p-1)$ and the correlation coefficient between any two of them is $1/(p-1)$. It follows that the first term on the right hand side is distributed, independently of the second, as χ_a^2, with $a = \tfrac{1}{2}p(p-3)$. Inspection of expression (1.3) shows that Bartlett's criterion for $p > 3$ is asymptotically of the form $\chi_a^2 + \{1 - [(p-2)/p]\lambda^2\}\chi_b^2$.

2. A criterion of large-sample χ^2 type. For testing the hypothesis that all correlation coefficients are equal it would in practice be useful to have a criterion whose limiting distribution is of χ^2 type. Such a criterion, with $\frac{1}{2}(p+1)(p-2)$ degrees of freedom, is provided by the expression

$$\begin{aligned}\sum_{i<j} (x_{ij} - \bar{x})^2 &= (1/\lambda^2)\sum_{i<j} y_{(ij)}^2 + (2/\lambda^2)[(p-1)^2/(p-2)] \sum_k (\bar{y}_k - \bar{y})^2/\{p - (p-2)\lambda^2\} \\ &= (1/\lambda^2)\{\sum_{i<j} (y_{ij} - \bar{y})^2 - \mu\sum_k(\bar{y}_k - \bar{y})^2\},\end{aligned}$$

where $\mu = (p-1)^2(1-\lambda^2)/\{p-(p-2)\lambda^2\}$. If we write

$$\bar{r}_k = \sum_i r_{ik}/(p-1) \quad (i \neq k), \qquad \bar{r} = 2\sum_{i<j} r_{ij}/\{p(p-1)\},$$

this expression may be put into the form

$$(n/\lambda^2)\{\sum_{i<j} (r_{ij} - \bar{r})^2 - \mu\sum_k(\bar{r}_k - \bar{r})^2\}.$$

In practice the estimate $1 - \bar{r}$ would have to be substituted for the unknown parameter λ. This substitution does not affect the limiting distribution.

On investigation the above statistic is found to be asymptotically equal to -2 times the logarithm of the likelihood ratio criterion. That this is not true of the criterion previously discussed accounts for its limiting distribution not being of χ^2 type. The exact likelihood ratio criterion is difficult to evaluate and complicated in form.

REFERENCES

[1] Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist.* **34** 122–148.

[2] Bartlett, M. S. (1954). A note on the multiplying factors for various χ^2 approximations. *J. Roy. Statist. Soc. Ser. B.* **16** 296–298.