

1969

Exact Test for Simple Correlation In Analysis of Dispersion

James E. Dunn

University of Arkansas, Fayetteville

Follow this and additional works at: <http://scholarworks.uark.edu/jaas>

 Part of the [Mathematics Commons](#)

Recommended Citation

Dunn, James E. (1969) "Exact Test for Simple Correlation In Analysis of Dispersion," *Journal of the Arkansas Academy of Science*: Vol. 23 , Article 35.

Available at: <http://scholarworks.uark.edu/jaas/vol23/iss1/35>

This article is available for use under the Creative Commons license: Attribution-NoDerivatives 4.0 International (CC BY-ND 4.0). Users are able to read, download, copy, print, distribute, search, link to the full texts of these articles, or use them for any other lawful purpose, without asking prior permission from the publisher or the author.

This Article is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Journal of the Arkansas Academy of Science by an authorized editor of ScholarWorks@UARK. For more information, please contact scholar@uark.edu, ccmiddle@uark.edu.

AN EXACT TEST FOR SIMPLE CORRELATION IN ANALYSIS OF DISPERSION

James E. Dunn

Dept. of Mathematics, University of Arkansas

Suppose an experimental design results in p -variate experimental units Y_1, Y_2, \dots, Y_n so that under usual assumptions of independence and normality, i.e. $Y_i \sim N(\mu_i, \Sigma)$ for $i = 1, 2, \dots, n$, the techniques of analysis of dispersion (multivariate analysis of variance) are applicable for testing linear hypotheses with respect to $\mu_1, \mu_2, \dots, \mu_n$. But suppose the purpose of the experiment was not to test for differences in the mean but rather to examine the associations among the variables, i.e. a correlation analysis is called for. For example, an experimental design may have been used to provide data for path analysis (cf. Teng, 1969) in which case it is desirable to examine the correlations among the causal variables. Under these circumstances, the treatment means μ_1, \dots, μ_n simply serve as nuisance parameters to prevent one from proceeding to estimate the correlation matrix by standard methods.

On the other hand, the appropriate analysis of dispersion serves to adjust out these effects and will always result in an error matrix E of residual sums of squares and products and an associated error degrees of freedom v . As is well known, this matrix provides the starting point for analysis of covariance, but the purpose of this note is to point out that E also provides a means for making the most common correlation test of all; namely, the test of statistical independence between any pair of variables.

For the sake of notation, denote the elements of Σ by σ_{ij} and the elements of the associated correlation matrix P by ρ_{ij} where $P = D^{-1} \Sigma D^{-1}$ where $D = \text{diag.}(\sqrt{\sigma_{11}}, \sqrt{\sigma_{22}}, \dots, \sqrt{\sigma_{pp}})$. Denoting the elements of E obtained from analysis of dispersion by e_{ij} , then the usual estimator of P is

$$R = D^{-1} E D^{-1}$$

with elements $r_{ij} = \frac{e_{ij}}{\sqrt{e_{ii}} \sqrt{e_{jj}}}$ where $D = \text{diag.}(\sqrt{e_{11}}, \sqrt{e_{22}}, \dots, \sqrt{e_{pp}})$.

Formally, we want to test $H_0: \rho_{ij} = 0$. Fortunately, we can start with a relatively recent result (cf. Rao, 1965) which was not available to R. A. Fisher when he published his transformation of the estimated correlation coefficient into a t statistic (cf. Kendall, 1958), namely that E has the Wishart distribution with density function

$$f(E) = \frac{\exp\left(-\frac{1}{2} \text{tr}(E \Sigma^{-1})\right) |E|^{\frac{v-p-1}{2}}}{\pi^{\frac{p(p-1)}{4}} 2^{\frac{vp}{2}} \left|\Sigma\right|^{\frac{v}{2}} \prod_{i=1}^p \Gamma\left(\frac{v-i+1}{2}\right)}$$

for any analysis of dispersion so long as E is positive definite. Since the marginal distribution of any 2×2 partition, say E_2 , on the principal diagonal of E also has a Wishart distribution with $p = 2$, we may write

$$f(e_{ii}, e_{ij}, e_{jj}) = \frac{\exp\left(-\frac{1}{2} \text{tr}(E_2 \Sigma_2^{-1})\right) |E_2|^{\frac{v-3}{2}}}{\pi^{\frac{1}{2}} 2^v \left|\Sigma_2\right|^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right) \Gamma\left(\frac{v-1}{2}\right)}$$

$$\text{where } E_2 = \begin{bmatrix} e_{ii} & e_{ij} \\ e_{ij} & e_{jj} \end{bmatrix} \text{ and } \Sigma_2 = \begin{bmatrix} \sigma_{ii} & \sigma_{ij} \\ \sigma_{ij} & \sigma_{jj} \end{bmatrix}$$

But when H_0 is true $\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}} \sqrt{\sigma_{jj}}} = 0$ so that $\sigma_{ij} = 0$ and the joint density simplifies to

$$f(e_{ii}, e_{ij}, e_{jj}) = \frac{\exp\left(-\frac{e_{ii}}{2\sigma_{ii}} - \frac{e_{jj}}{2\sigma_{jj}}\right) (e_{ii} e_{jj})^{\frac{v-3}{2}} \left(1 - \frac{e_{ij}^2}{e_{ii} e_{jj}}\right)^{\frac{v-3}{2}}}{\pi^{\frac{1}{2}} 2^v (\sigma_{ii} \sigma_{jj})^{\frac{v}{2}} \Gamma\left(\frac{v}{2}\right) \Gamma\left(\frac{v-1}{2}\right)}$$

where $e_{ii} > 0$, $e_{jj} > 0$, and $-1 < e_{ij} < 1$. Making the transformation

Correlation In Analysis Of Dispersion

$$r_{ij} = \frac{e_{ij}}{\sqrt{e_{ii}} \sqrt{e_{jj}}}, \quad x = e_{ii}, \quad y = e_{jj}$$

with Jacobian $|J| = \sqrt{xy}$, the marginal density for r_{ij} is

$$\begin{aligned} m(r_{ij}) &= \frac{(1 - r_{ij}^2)^{\frac{\nu-3}{2}}}{\pi^{\frac{1}{2}} \Gamma(\frac{\nu}{2}) \Gamma(\frac{\nu-1}{2})} \int_0^\infty \int_0^\infty \left(\frac{x}{2\sigma_{ii}}\right)^{\frac{\nu}{2}-1} \left(\frac{y}{2\sigma_{jj}}\right)^{\frac{\nu}{2}-1} \frac{\exp\left(-\frac{x}{2\sigma_{ii}} - \frac{y}{2\sigma_{jj}}\right)}{4\sigma_{ii}\sigma_{jj}} dx dy \\ &= \frac{(1 - r_{ij}^2)^{\frac{\nu-3}{2}}}{B\left(\frac{1}{2}, \frac{\nu-1}{2}\right)} \quad -1 < r_{ij} < 1 \end{aligned}$$

Transforming once more, let

$$t = \frac{r_{ij} \sqrt{\nu-1}}{\sqrt{1-r_{ij}^2}} \quad (2)$$

with Jacobian $|J| = \frac{\nu-1}{(t^2 + \nu-1)^{\frac{3}{2}}}$ so that after simplification

$$g(t) = \frac{1}{\sqrt{\nu-1} B\left(\frac{1}{2}, \frac{\nu-1}{2}\right) \left(1 + \frac{t^2}{\nu-1}\right)^{\frac{\nu}{2}}} \quad -\infty < t < \infty$$

i.e. the t distribution with $\nu-1$ degrees of freedom.

Hence, to test $H_0: \rho_{ij} = 0$ in the presence of nuisance location parameters, perform the appropriate analysis of dispersion to obtain E with ν degrees of freedom. Compute the estimated correlation matrix R from E as shown in (1) and select the appropriate element r_{ij} . Transform to the t statistic as given by (2) and reject H_0 at the α -level of significance if $|t| > t_{(\nu-1)1-\frac{\alpha}{2}}$ where $t_{(\nu-1)1-\frac{\alpha}{2}}$ is the $1 - \frac{\alpha}{2}$ critical value of the t -distribution with $\nu-1$ degrees of freedom.

REFERENCES

- Kendall, M. G. and Stuart, A. 1958. *The Advanced Theory of Statistics, Vol. 1.* Hafner Publishing Company, pp. 383-390.
- Rao, C. R. 1965. *Linear Statistical Inference and Its Applications.* John Wiley & Sons, Inc., pp. 464-465.
- Teng, Yao-Chung. 1969. *Effects of row spacing, rates of seeding and nitrogen fertilization and varieties on yield and other agronomic characteristics of rice.* Unpublished M.S. thesis, University of Arkansas.