In: Bulletin of the International Statistical Institute, Proceedings of the 40th Session, Warsaw. ISI: Voorburg

TESTING THE DIFFERENCES BETWEEN OBSERVED AND IMPLIED CORRELATIONS IN SOME PATH ANALYSIS MODELS

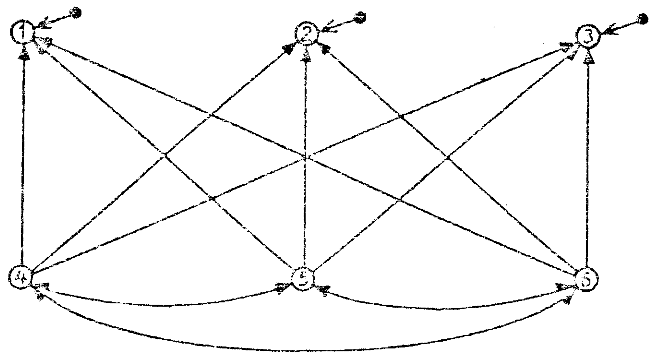
NANNY WERMUTH

UNIVERSITÄT MAINZ, FEDERAL REPUBLIC OF GERMANY

Whenever it is suspected that the observed correlations in a group of variables are confounded by the influence of background factors, it may be appropriate to investigate this possibility within the context of a path analysis model (Wright, 1934). For medical data, typical background or conditional variables are factors in the anamnesis or in the constitution of a patient, or repeated measurements of of the same type that have been recorded at a different location or at another time. If a particular type of path analysis model fits the data, then the observed correlations for a group of variables may be reproduced or explained by their correlations with the background factors.

Assume that the interrelations among variables 1, 2, and 3 are thought to be reproducible by the relations to variables 4, 5, and 6. Then, the appropriate path analysis model is the following (Figure 1): There are no direct paths between

Figure 1: Path diagram for a model with conditional independence of variables 1, 2, 3 given variables 4, 5, 6



variables 1, 2, and 3, but paths are leading to each of these three variables originating from variables 4, 5, and 6, and from residual factors. Furthermore, correlations among the background factors 4, 5, and 6 are permitted (there are paths with two headed arrows connecting them).

The correlations r^*_{ij} implied by this model for the variable pairs (1,2), (1,3), and (2,3) are

$$(1) \quad r^*_{ij} = \sum_{l=4,5,6} r_{il} p_{lj} \quad ,$$

where r_{il} are observed correlations, and p_{lj} are the standardized estimated linear regression coefficients from regressing variables 1, 2, and 3 on variables 4, 5, and 6. These path coefficients are readily computed using the sweep-operator (Dempster, 1972) on indices 4, 5, and 6 of the correlation matrix.

The implied correlations given in (1) are identical to the correlations implied by a particular covariance selection model (Dempster, 1972; Wermuth, 1975 a, b). For this covariance selection model, zero concentrations are fitted to the variable pairs (1,2), (1,3), (2,3). The model implies conditional independence of variables 1, 2, and 3 given variables 4, 5, and 6, and it is denoted as 1456/2456/3456 (in analogy to similar models for contingency tables).

Thus, a simple likelihood-ratio test is available from the theory of covariance selection to test the deviations between observed and implied correlations in the above described path analysis model, provided the correlations are a sample from a multivariate normal distribution. More precisely, let n be the sample size, R the observed correlation matrix, and $\hat{P}$ be the implied correlation matrix, then

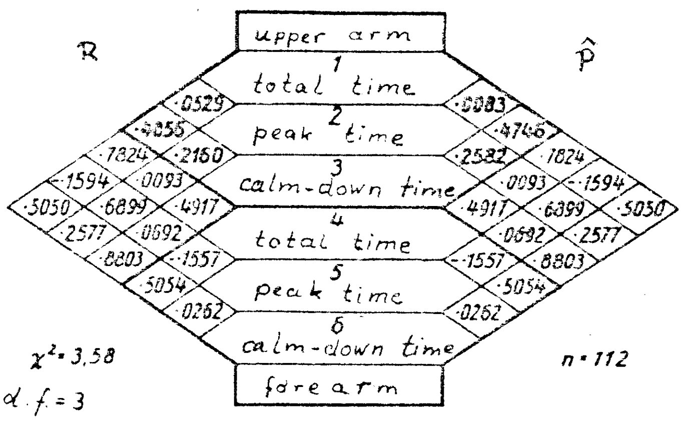
$$(2) \quad -2\ln\left(\frac{|\hat{P}|}{|R|}\right)^{-\frac{n}{2}} \sim \chi^2$$

with degrees of freedom equal to the number of zero concentrations for $\hat{P}$.

Numerical example

The following data (Figure 2) have been collected for the project "Angiology" of SFB 36, supported by the German Research Council (DFG). For 112 patients the correlations

Figure 2: Observed correlation matrix R and correlation matrix $\hat{P}$ implied by model 1456/2456/3456



between total time (1), peak time (2), and calm-down time (3) of pulse waves on the upper arm, and on the forearm (variables 4, 5, 6) are given. The path coefficients were calculated to be

$$(p_{14}, p_{15}, p_{16}) = (0.696, -0.055, 0.155)$$

$$= \begin{pmatrix} 1 & -0.1557 & 0.5054 \\ & 1 & 0.0264 \\ & & 1 \end{pmatrix}^{-1} (0.7824, -0.1594, 0.5050);$$

$$(p_{24}, p_{25}, p_{26}) = (-0.008, 0.682, 0.244);$$

$$(p_{34}, p_{35}, p_{36}) = (0.076, 0.059, 0.840) .$$

It can be seen that the resulting implied correlations $r^*_{12} = 0.008$, $r^*_{13} = 0.475$, $r^*_{23} = 0.258$ differ only little from the observed correlations. The likelihood-ratio test affirms this impression. Thus, the observed correlations on the upper arm can be explained by correlations with the corresponding measurements on the forearm.

Summary

Using a numerical example it is shown how to investigate the effects of background factors with a path analysis model, and how to use the theory of covariance selection to test the differences between observed and implied correlations.

References

Dempster, A.P. (1972), Covariance selection, Biometrics 28, 157

Wermuth, N. (1975a), Analogies between multiplicative models in contingency tables and covariance selection, Biometrics (to appear)

Wermuth, N. (1975b), Model search among multiplicative models, Biometrics (submitted)

Wright, S. (1934), The method of path coefficients, Ann. Math. Stat. 5, 161