KM 16(1984)
pag 41 - 52

AN EXAMPLE OF A ROBUSTNESS STUDY FOR LINEAR HYPOTHESES OF CORRELATIONS

Ab Mooijaart*

Rien van der Leeden*

Abstract

In this study an example of linear hypotheses about correlations will be discussed. The statistical distribution of the variables may be either normal or non-normal. It will be shown that by incorrectly assuming a normal distribution for the variables a linear hypothesis test may be highly non-robust.

Key words: non-normality, normality correlations, linear hypothesis, robustness, example.

*Department of Psychology, Leiden University, Hooigracht 15, 2312 KM LEIDEN, The Netherlands, tel. 071 - 148333, tst. 5126.

1. Introduction

Most studies for covariances and correlations assume normality for the manifest variables. Hypotheses testing for correlations rely, under the assumption of normality, heavily on the result given already in 1898 by Pearson & Filon.

Recently, however, several researchers have been interested in correlation studies not based upon this assumption. To mention a few very recent papers: Steiger & Hakstian (1982, 1983), De Leeuw (1983), Bentler (1983), Steiger & Browne (1984). A very detailed discussion of studies for covariances, which is related to correlational studies, can be found in Browne (1982, 1984).

The importance of the Steiger & Hakstian paper is that they ".... demonstrate that the non-robustness of the normal theory (NT) procedures occurs because .... the asymptotic variance-covariance structure of the correlation coefficients changes .... as a complicated function of the moments of the parent distribution" (Steiger & Hakstian, 1982, page 208).

In this paper an example is discussed in which it is shown what effect non-normality has upon different tests. It is shown that tests, in which normality is assumed, may be very sensitive to the skewness of the variables. Tests which do not assume normality are less sensitive for the skewness of the variables. Moreover, for our example it holds that in cases with highly skewed variables the sample size must be (very) large for using asymptotic properties.

Of course, this is just one particular example. However, more extensive studies are not known from the literature.

2. Testing linear hypotheses

Let there be m variables, then there are $t=m(m-1)/2$ intercorrelations. The vector of population correlations will be denoted by ρ and the corresponding vector of sample correlations by r (sample size is N). Then the linear hypothesis to be tested can be written as:

$$H_0 : \rho = X\beta = X_1 + X_2\beta_2 ,$$

in which ρ and X_1 are of the order (tx1) (where X_1 is a vector of constants, mostly set equal to zero), X_2 of the order (txq) (a matrix with elements mostly zero or one) and β_2 is the vector of unknown parameters of the order (qx1)

The loss-function we choose for estimating β is:

$$(1) \qquad X^2 = (r - X\beta)'\hat{\Sigma}^{-1}(r - X\beta) ,$$

in which $\hat{\Sigma}$ is a consistent estimator of the variance-covariance matrix of the sample correlations r. How this matrix looks like will be discussed in section 3. It is well-known from standard literature that under very mild conditions NX^2 is asymptotic chi-square distributed with degrees of freedom t-q (See for an overview of many minimum chi-square methods Ferguson (1958)).

An analogous form of (1) can be written as:

$$X^2 = (r - X_1 - X_2\beta_2)' \hat{\Sigma}^{-1}(r - X_1 - X_2\beta_2)$$

$$(2) \qquad = (s - X_2\beta_2) \hat{\Sigma}^{-1}(s - X_2\beta_2) ,$$

where obviously $s=r-X_1$.

The estimator of β_2 from minimizing (2) is given by:

$$(3) \qquad \hat{\beta}_2 = (X_2' \hat{\Sigma}^{-1} X_2)^{-1} X_2' \hat{\Sigma}^{-1} s ,$$

with variance covariance matrix of the estimators equal to:

$$(4) \qquad ACOV(\hat{\beta}_2) = N (X_2' \hat{\Sigma}^{-1} X_2)^{-1} .$$

An interesting alternative way for estimating β_2 is by solving the following linear set of equations:

$$(5) \qquad \begin{pmatrix} \hat{\Sigma} & X_2 \\ X_2' & \theta \end{pmatrix} \begin{pmatrix} \lambda \\ \beta_2 \end{pmatrix} = \begin{pmatrix} s \\ \theta \end{pmatrix} ,$$

in which θ is a vector of zeros of the order (qx1).

The proof for this is very simple. From (5) it follows:

$$(6a) \qquad \hat{\Sigma}\lambda + X_2\beta_2 = s$$

$$(6b) \qquad X_2'\lambda = \theta$$

and the substitution of λ from (6a) in (6b) yields the same estimator $\hat{\beta}_2$ as in (3).

Also it is easy to compute X^2 from (5). From (1) and (6a) we see that:

(7) $\quad X^2 = (s - X_2\beta_2)'\lambda = s'\lambda \ .$

This procedure is interesting because we do not have to invert two matrices as is necessary by using (3), but just solving a linear set of equations which mostly will save a lot of computation time.

3. Asymptotic covariances of correlations

In this section we will give the asymptotic covariances of the correlations. For a more detailed discussion of these matters we refer to Steiger & Hakstian (1982) and for a historical overview to Steiger & Hakstian (1983). From the literature we know:

$$\begin{aligned}(8)\qquad NCov(r_{ij},r_{k\ell}) &= \rho_{ijk\ell} - \tfrac{1}{2}\rho_{ij}(\rho_{iik\ell} + \rho_{jjk\ell}) \\ &\quad - \tfrac{1}{2}\rho_{k\ell}(\rho_{kkij} + \rho_{\ell\ell ij}) \\ &\quad + \tfrac{1}{4}\rho_{ij}\rho_{k\ell}(\rho_{iikk} + \rho_{ii\ell\ell} + \rho_{jjkk} + \rho_{jj\ell\ell}),\end{aligned}$$

with $\rho_{ijk\ell} = \dfrac{\sigma_{ijk\ell}}{\sqrt{\sigma_{ii}\sigma_{jj}\sigma_{kk}\sigma_{\ell\ell}}}$ where $\sigma_{ijk\ell} = E(X_i-\mu_i)(X_j-\mu_j)(X_k-\mu_k)(X_\ell-\mu_\ell)$ and $\mu_i = EX_i$. A way of computing the covariance matrix efficiently is given in Mooijaart (1984).

If we, in addition, assume the variables to be normally distributed then it holds:

$$\rho_{ijk\ell} = \rho_{ij}\rho_{k\ell} + \rho_{ik}\rho_{j\ell} + \rho_{i\ell}\rho_{jk}$$

and so we can write for the covariances of the correlations:

$$\begin{aligned}(9)\qquad NCov(r_{ij},r_{k\ell}) &= \rho_{ik}\rho_{j\ell}+\rho_{i\ell}\rho_{jk} - (\rho_{ij}\rho_{ik}\rho_{i\ell}+\rho_{ij}\rho_{jk}\rho_{j\ell}+\rho_{ik}\rho_{jk}\rho_{k\ell}+ \\ &\quad \rho_{i\ell}\rho_{j\ell}\rho_{k\ell}) + \tfrac{1}{2}\rho_{ij}\rho_{k\ell}(\rho^2_{ik}+\rho^2_{i\ell}+\rho^2_{jk}+\rho^2_{j\ell})\end{aligned}$$

which is the Pearson-Filon result for the normal distribution.

4. An example

In this section we discuss the results of a study in which a linear hypothesis is tested several times in samples from a well-known population. This population consists of variables with all intercorrelations equal to .5 . So in each sample the hypothesis is tested that all correlations are equal to .5 . Samples are drawn from the population with different sample sizes. We choose N=100, 200, 400 and 1000. For each sample 300 replications were taken. This makes it possible to see how the test statistic is distributed. According to the theory this statistic should be chi-square distributed with 6 degrees of freedom. The variables are chosen to be log-normal distributed. The probability density function of X is given by:

$$(10) \qquad P_X(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\tfrac{1}{2}\left(\frac{\log x - \zeta}{\sigma}\right)^2\right] .$$

This means that X is defined as

$$(11) \qquad U = (\log x - \zeta)/\sigma ,$$

in which U is a unit normal variable. (See for a detailed discussion of the log-normal distribution Aitchison & Brown (1957) or Johnson & Kotz (1970).) From (11) it is clear how data can be generated from a log-normal distribution. We randomly draw a value from an unit normal distribution and then set X equal to:

$$(12) \qquad X = \exp\left[\sigma U + \zeta\right]$$

Some characteristics (mean and standard deviation) of the log-

normal distributions are

$$(13a) \qquad \mu_1' = \exp\left[\zeta+\tfrac{1}{2}\sigma^2\right] ,$$

$$(13b) \qquad \sigma(X) = \exp\left[\zeta\right]\sqrt{\omega(\omega-1)} ,$$

$$\text{where} \qquad \omega = \exp\left[\sigma^2\right] .$$

The reason why we choose the log-normal distribution is that it is a skewed distribution. In our study we generate data from a log-normal distribution with variance equal to 1. This means from (13b) that:

$$(14) \qquad \zeta = -\tfrac{1}{2}\log\left[\omega(\omega-1)\right] .$$

So the skewness of the distribution is determined by the parameter σ only. The skewness of the distribution increases rapidly with σ. For instance, (see table 2, Johnson & Kotz, 1970, page 118) we find for the well-known measure of skewness α_3: $\alpha_3 = .3$ for $\sigma = .1$; $\alpha_3 = .61$ for $\sigma = .2$; $\alpha_3 = .95$ for $\sigma = .3$; $\alpha_3 = 1.32$ for $\sigma = .4$; $\alpha_3 = 1.75$ for $\sigma = .5$; $\alpha_3 = 2.26$ for $\sigma = .6$; $\alpha_3 = 6.18$ for $\sigma = 1$. In our study we vary the skewness of the variables by choosing $\sigma = .4$, $.5$ and $.6$ and we will see what the effect of the skewness is on the test statistics (the X^2-values).

So our procedure is as follows (for different σ and sample sizes): we generate 4 independent X-values by (12), and collect them in a vector x. Then we pre-multiply this vector with $K\Lambda^{\frac{1}{2}}$, in which K is a matrix which columns consist of the eigenvectors and Λ a diagonal matrix with the eigenvalues of the population correlation matrix. The result is that $K\Lambda^{\frac{1}{2}}x$ can be

conceived as a vector with elements coming from the population with the known correlation matrix.

In our study we are interested in two things: first: if we choose α = 5 % how often (from the 300 replications) is the test of the hypotheses rejected (the expected values are, of course, 5 % of 300, is 15 times). Second: according to the theory the distribution of the X^2-values must be chi-square distributed with 6 degrees of freedom. To test this the chi-square distribution with 6 degrees of freedom is divided in 20 intervals of 5 % and it is counted how many X^2-values fall in these intervals. Then it can be tested with a chi-square test with 19 degrees of freedom whether the X^2-values are chi-square distributed.

Results:

Table ! gives for different sample sizes and different σ values the percentages of rejecting the hypothesis that all intercorrelations are equal to .5 .

Table 1

Percentages of Rejecting the Hypothesis with α=5%

		Normal				Non-normal			
N		100	200	400	1000	100	200	400	1000
	.4	33.0	25.67	27.0	23.33	23.67	11.33	10.0	7.0
σ	.5	43.0	42.0	46.0	39.0	28.33	14.0	13.67	7.67
	.6	52.67	56.33	58.67	55.0	33.33	20.0	17.33	9.33

The left hand side in Table 1 is based upon the assumption that the variables are normal distributed, the right hand side

is not based upon the normality assumption. So the left and right hand side are based upon formula (9) and (8), respectively.

From Table 1 we see that the percentages of rejecting the hypothesis in the non-normal case are smaller than in the normal case. This is what we expected because the variables are non-normal. Further we see that the percentages increase with σ. This means that the hypothesis is rejected too often for increasing skewness of the variables. We also see that, in particular in the non-normal case, the percentages decrease with the sample size. This is what should happen because asymptotically the percentages should be 5 %. In particular in the non-normal case we see that for N=1000 the percentages are quite close to 5 %.

In Table 2 the results are given for testing the hypothesis if the X^2-values are chi-square distributed with 6 degrees of freedom. All values in the table are chi-square distributed with 19 degrees of freedom.

Table 2

Test Statistics for Chi-square Distribution
(degrees of freedom is 19)

		Normal				Non-normal			
N		100	200	400	1000	100	200	400	1000
	.4	524.8	345.6	361.5	277.3	256.0	54.3	29.3[b]	13.9[a]
σ	.5	938.7	924.9	1097.1	789.1	380.1	91.7	59.3	27.3[b]
	.6	1470.5	1702.5	1846.9	1604.8	541.3	184.5	110.5	34.3[c]

a) 10% critical X^2-value: 27.2

b) 5% critical X^2-value: 30.1

c) 1% critical X^2-value: 36.2

From Table 2 it is clear that under the assumption of normality the X^2-values are not chi-square distributed. In the non-normal case we clearly see that for increasing sample size the distribution of the X^2-values becomes closer to the expected chi-square distribution. In particular if the variables are moderate skewed ($\sigma = .4$) a sample size between 400 and 1000 is sufficient for getting the asymptotic chi-square distribution.

Conclusions

--- If variables are non-normal distributed, tests based upon the normality assumption lead too often to rejecting the hypothesis.

--- Obviously, for larger sample size the X^2-values are chi-square distributed.

--- For increasing skewness of the variables the sample size must increase also in order to find the X^2-values chi-square distributed.

References

Aitchison, J.R. & Brown, J.A.C., (1957), *The Lognormal Distribution*. London: Cambridge University Press.

Bentler, P.M., (1983), Simultaneous equation systems as moment structure models: With an introduction to latent variable models. *Journal of Econometrics*, *22*, 13-42.

Browne, M.W., (1982), Covariance structures. In: D.M. Hawkins (ed.), *Topics in Applied Multivariate Analysis*. London: Cambridge University Press.

Browne, M.W., (1984), Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62-83.

De Leeuw, J., (1983), Models and methods for the analysis of correlation coefficients. *Journal of Econometrics*, *22*, 113-137.

Ferguson, T.S., (1958), A method for generating best asymptotically normal estimates with application to the estimation of bacterial densities. *Annals of Mathematical Statistics*, *29*, 1046-1062.

Johnson, N.L. & Kotz, S., (1970), *Continuous Univariate Distribution-1*. Boston: Houghton Mifflin Company.

Mooijaart, A., (1984), Factor analysis for non-normal variables (accepted for publication).

Pearson, K. & Filon, L.N.G., (1898), Mathematical contributions to the theory of evolution: IV. On the probable error of frequency constants and on the influence of random selection of variation and correlation. *Philosophical Transactions of the Royal Society of London, Series A*, *191*, 229-311.

Steiger, J.H. & Browne, M.W., (1984), The comparison of interdependent correlations between optimal linear composites. *Psychometrika*, *49*, 11-24.

Steiger, J.H. & Hakstian, A.R., (1982), The asymptotic distribution of elements of a correlation matrix: Theory and application. *British Journal of Mathematical and Statistical Psychology*, *35*, 208-215.

Steiger, J.H. & Hakstian, A.R., (1983), A historical note on the asymptotic distribution of correlations. *British Journal of Mathematical and Statistical Psychology*, *36*, 157.