Robust estimation of Cronbach's alpha

A. Christmann[1] and S. Van Aelst[2]

[1] University of Dortmund, HRZ, D-44221 Dortmund, Germany
[2] Ghent University, Krijgslaan 281 S9, B-9000 Gent, Belgium

Keywords: Cronbach's alpha, MCD, M-estimator, Robustness, S-estimator.

1 Abstract

Cronbach's alpha is a popular method to measure reliability, e.g. in quantifying the reliability of a score to summarize the information of several items in questionnaires. The alpha coefficient is known to be non-robust. We study the behaviour of this coefficient under various statistical models to identify situations, which can easily occur in practice, but under which the Cronbach's alpha coefficient is extremely sensitive to violations of the classicial model assumptions. Furthermore, we construct a robust version of Cronbach's alpha which is insensitive to a small proportion of data that belong to a different source. The idea is that the robust Cronbach's alpha reflects the reliability of the bulk of the data. For example, it should not be possible that some small amount of outliers makes a score look reliable if it is not.

2 Method

We consider the problem of constructing a measure of reliability for a set of items such as in a questionnaire. Cronbach (1951) proposed the coefficient alpha as a lower bound to the reliability coefficient in classical test theory (see also Kuder and Richardson, 1937). This popular measure has been investigated further by e.g. Feldt (1965) and Bravo and Potvin (1991).

Consider a series of items $Y_j = T_j + \varepsilon_j$ for $j = 1, \ldots, p$ where T_j are the true unobservable test scores and ε_j are the associated errors which are independent from the true test scores and distributed with zero mean. Then Cronbach's alpha is given by

$$\begin{aligned} \alpha_n^C &= \frac{p}{p-1}\left[\frac{\sum\sum_{i\neq j}\operatorname{Cov}(Y_i, Y_j)}{\operatorname{Var}\left(\sum_{j=1}^p Y_j\right)}\right] \qquad (1) \\ &= \frac{p}{p-1}\left[1 - \frac{\sum_{j=1}^p \sigma_j^2}{\sum_{j=1}^p\sum_{k=1}^p \sigma_{j,k}}\right] \quad , \end{aligned}$$

where σ_j^2 is the variance of item Y_j and $\sigma_{j,k}$ is the covariance of the pair (Y_j, Y_k). It has been shown that Cronbach's alpha is always a lower bound of reliability, see Gutman (1953).

Cronbach's alpha can be estimated by substituting empirical variances and covariances in expression (1) above. However it is well known that classical estimators such as empirical variances and covariances can be heavily influenced by a few erroneous observations. Therefore the resulting estimate of Cronbach's alpha can be completely misleading as soon as some mistaken observations are present. We want to avoid this problem and aim to construct a robust version of Cronbach's alpha in the sense that we develop a reliability measure that is able to resist outlying observations. The robust Cronbach's alpha will thus measure the reliability of the most central part of the observations while not being affected by outlying observations. A robust measure of reliability was proposed by Wilcox (1992) who used the midvariance and midcovariance as robust estimates for the variances and covariances in (1) .

In this paper we propose to estimate the covariance matrix of $Y = (Y_1, \ldots, Y_p)^t$ using a robust estimator and then we substitute the elements of this robust covariance estimate into (1) .

Many robust estimators of multivariate location and scatter have been investigated in the literature, such as M-estimators (Maronna 1976, Kent and Tyler 1991), the minimum volume ellipsoid and minimum covariance determinant estimator (Rousseeuw 1984), and S-estimators (Davies 1987, Lopuhaä 1989). Recently, robust multivariate statistical methods based on robust estimation of location and scatter have been developed and investigated such as factor analysis (Pison et al. 2002), discriminant analysis (Croux and Haesbroeck 2000), and multivariate regression (Rousseeuw et al. 2001).

An advantage of constructing a robust Cronbach's alpha as proposed in this paper is that it can be obtained immediately from the robust scatter matrix estimate computed for the robust multivariate analysis without any additional computational load. Hence, such robust versions of Cronbach's alpha based on the minimum covariance determinant estimator can easily be computed using statistical software packages, e.g. S-Plus, R, or SAS.

References

G. Bravo and L. Potvin (1991). Estimating the Reliability of Continuous Measures with Cronbach's Alpha or the Intraclass Correlation Coefficient: Toward the Integration of Two Traditions. *J. Clin. Epidemiol.*, 44, 381–390.

L.J. Cronbach (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika*, 16, 297–334.

C. Croux and G. Haesbroeck (2000). Principal Component Analysis based on Robust Estimators of the Covariance or Correlation Matrix: Influence Function and Efficiencies. *Biometrika*, 87, 603–618.

L. Davies (1987). Asymptotic Behavior of S-Estimators of Multivariate Location Parameters and Dispersion Matrices. *The Annals of Statistics*, 15, 1269–1292.

L.S. Feldt (1965). The Approximate Sampling Distribution of Kuder-Richardson Reliability Coefficient Twenty. *Psychometrika*, 30, 357–370.

L. Guttman (1953). Reliability Formulas That Do Not Assume Experimental Independence. *Psychometrika*, 18, 225–239.

K.T. Kent and D.E. Tyler (1991). Redescending M-estimates of Multivariate Location and Scatter. *The Annals of Statistics*, 19, 2102–2119.

G.F. Kuder and M.W. Richardson (1937). The Theory of the Estimation of Test Reliability. *Psychometrika*, 2, 151–160.

H.P. Lopuhaä (1989). On the Relation Between S-Estimators and M-Estimators of Multivariate Location and Covariance. *The Annals of Statistics*, 17, 1662–1683.

R.A. Maronna (1976). Robust M-Estimates of Multivariate Location and Scatter. *The Annals of Statistics*, 4, 51–67.

G. Pison, P.J. Rousseeuw, P. Filzmoser, and C. Croux (2002). Robust Factor Analysis. *Journal of Multivariate Analysis*, to appear.

P.J. Rousseeuw (1984). Least Median of Squares Regression. *Journal of the American Statistical Association*, 79, 871–880.

P.J. Rousseeuw, S. Van Aelst, K. Van Driessen, and J. Agullò (2001). Robust Multivariate Regression. Submitted.

R.R. Wilcox (1992). Robust Generalizations of Classical Test Reliability and Cronbach's Alpha. *British Journal of Mathematical and Statistical Psychology*, 45, 239–254.