Factor Analysis

Martin Sewell

Department of Computer Science

University College London

May 2007 (updated August 2008)

Factor analysis is a statistical technique, the aim of which is to simplify a complex data set by representing the set of variables in terms of a smaller number of underlying (hypothetical or unobservable) variables, known as factors or latent variables. The origins of factor analysis can be traced back to Pearson (1901) and Spearman (1904), the term was first introduced by Thurstone (1931) and a good text on the subject is Gorsuch (1983). The technique is a branch of multivariate analysis and may also be described as unsupervised learning and an exercise in modelling. The observed variables are modelled as linear combinations of the factors, plus 'error' terms. The methodology relies on the (often unjustified) assumption that n factors exist.

Steps in performing factor analysis:

1. Collect data.

2. Generate a variance-covariance matrix of the observed variables.

3. Select the number of factors.

4. Extract the initial set of factors.

5. Perform factor rotation to a terminal solution.

6. Interpret the factor structure.

7. Construct factor scores to use in further analyses.

References

GORSUCH, Richard L., 1983. *Factor Analysis*. Second ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

PEARSON, Karl, 1901. On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine, Series 6*, **2**(11), 559–572.

SPEARMAN, C., 1904. "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, **15**(2), 201–292.

THURSTONE, L. L., 1931. Multiple Factor Analysis. *Psychological Review*, **38**(5), 406–427.