Nonparametric inference in nonlinear principal components analysis: Exploration and beyond

Linting, M.

References

Agresti, A., & Coull, B. A. (1998). Approximate is better than 'exact' for interval estimation of binomial proportions. *The American Statistician*, *52*, 119–126.

Ainsworth, M. D. S., & Bell, S. M. (1970). Attachment, exploration, and separation illustrated by the behavior of one-year-olds in a strange situation. *Child Development*, *41*, 49–67.

Anderson, M. J., & Ter Braak, C. J. F. (2003). Permutation tests for multifactorial analysis of variance. *Journal of Statistical Computation and Simulation*, *73*, 85–113.

Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics*, *34*, 122–148.

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis (2nd ed.).* New York: Wiley.

Arsenault, L., Tremblay, R. E., Boulerice, B., & Saucier, J. (2002). Obstetrical complications and violent delinquency: Testing two developmental pathways. *Child Development*, *73*, 496–508.

Beishuizen, M., Van Putten, C. M., & Van Mulken, F. (1997). Mental arithmetic and strategy use with indirect number problems up to hundred. *Learning and Instruction*, *7*, 87–106.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, *57*, 289–300.

Bowlby. (1969). *Attachment and loss, volume 1: Attachment.* New York: Basic Books.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *26*, 123–140.

Buja, A. (1990). Remarks on functional canonical variates, alternating least squares methods and Ace. *The Annals of Statistics*, *18*, 1032–1069.

Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research*, *27*, 509–540.

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika*, *31*, 33–42.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences, (2nd ed.).* Hillsdale, New Jersey: Lawrence Erlbaum.

Cohen, J. (1994). The earth is round ($p < 0.05$). *American Psychologist*, *49*, 997–1003.

De Haas, M., Algera, J. A., Van Tuijl, H. F. J. M., & Meulman, J. J. (2000). Macro and micro goal setting: In search of coherence. *Applied Psychology*, *49*, 579–595.

De Leeuw, J. (1988). Multivariate analysis with linearizable regressions. *Psychometrika*, *53*, 437–454.

De Leeuw, J., & Van der Burg, E. (1986). The permutational limit distribution of generalized canonical correlations. In E. Diday (Ed.), *Data analysis and informatics, IV* (pp. 509–521). Amsterdam, the Netherlands: Elsevier Science.

De Schipper, J. C., Tavecchio, L. W. C., Van IJzendoorn, M. H., & Linting, M. (2003). The relation of flexible child care to quality of center day care and children's socio-emotional functioning: A survey and observational study. *Infant Behavior & Development*, *26*, 300–325.

Diaconis, P. (1985). Theories of data analysis: From magical thinking through classical statistics. In D. C. Hoaglin, F. Mosteller, & J. W. Tukey (Eds.), *Exploring data tables, trends, and shapes* (pp. 1–36). New York: Wiley.

Dietz, E. J. (1983). Permutation tests for association between two distance matrices. *Systematic Zoology*, *32*, 21–26.

Dijksterhuis, G. B., & Heiser, W. J. (1995). The role of permutation tests in exploratory multivariate data analysis. *Food Quality and Preference*, *6*, 263–270.

Douglas, M. E., & Endler, J. A. (1982). Quantitative matrix comparisons in ecological and evolutionary investigations. *Journal of Theoretical Biology*, *99*, 777–795.

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans.* Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics.

Efron, B. (1988). Bootstrap confidence intervals: Good or bad? *Psychological Bulletin*, *104*, 293–296.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall.

Efron, B. E. (1983). Estimating the error rate of a prediction rule: improvements on cross-validation. *Journal of the American Statistical Association*, *78*, 316–313.

Eurelings-Bontekoe, E. H. M., Duijsens, I. J., & Verschuur, M. J. (1996). Prevalence of DSM-III-R and ICD-10 personality disorders among military conscripts suffering from homesickness. *Personality and Individual Differences*, *21*, 431–440.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272–299.

Fisher, R. A. (1935). *The design of experiments.* Edinburgh: Oliver and Boyd.

Fisher, R. A. (1938). *Statistical methods for research workers.* Edinburgh: Oliver and Boyd.

Fisher, R. A. (1940). The precision of discriminant functions. *Annals of Eugenics*, *10*, 422–429.

Gabriel, K. R. (1971). The biplot-graphic display of matrices with application to principal component analysis. *Biometrika*, *58*, 453–467.

Gabriel, K. R. (1981). Biplot display of multivariate matrices for inspection of data and diagnosis. In V. Barnett (Ed.), *Interpreting multivariate data* (pp. 147–174). Chichester, England: Wiley.

Gardner, S., & le Roux, N. J. (2003). Graphics and visualisation in practice: Biplots for exploring multi-dimensional reality. *Bulletin of the International Statistical Institute, 54th proceedings, Berlin: Germany*, *54*, 270–273.

Gifi, A. (1990). *Nonlinear multivariate analysis.* Chichester, England: Wiley.

Girshick, M. A. (1939). On the sampling theory of roots of determinantal equations. *Annals of Mathematical Statistics*, *10*, 203–224.

Glick, B. J. (1979). Tests for space-time clustering used in cancer research. *Geographical Analysis*, *11*, 202–208.

Gliner, J., Leech, N., & Morgan, G. (2002). Problems with null hypothesis significance testing (NHST): What do the textbooks say? *Journal of Experimental Education*, *71*, 83–92.

Good, P. I. (2000). *Permutation tests: A practical guide to resampling methods for testing hypotheses.* New York: Springer-Verlag.

Gower, J. C., & Blasius, J. (2005). Multivariate prediction with nonlinear principal components analysis: theory. *Quality & Quantity*, *39*, 359–372.

Gower, J. C., & Hand, D. J. (1996). *Biplots.* London: Chapman & Hall.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis.* New York: Academic Press.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*, 430–450.

Groenen, P. J. F., Commandeur, J. F., & Meulman, J. J. (1998). Distance analysis of large data sets of categorical variables using object weights. *British Journal of Mathematical and Statistical Psychology*, *51*, 217–232.

Guttman, L. (1941). The quantification of a class of attributes: A theory and a method of scale construction. In P. Horst (Ed.), *The prediction of personal adjustment* (pp. 319–348). New York: Social Science Research Council.

Heiser, W. J., & Meulman, J. J. (1983). Constrained multidimensional scaling including confirmation. *Applied Psychological Measurement*, *7*, 381–404.

Heiser, W. J., & Meulman, J. J. (1994). Homogeneity analysis: Exploring the distribution of variables and their nonlinear relationships. In

M. Greenacre & J. Blasius (Eds.), *Correspondence analysis in the social sciences: Recent developments and applications* (pp. 179–209). New York: Academic Press.

Hopman-Rock, M., Tak, E. C. P. M., & Staats, P. G. M. (2001). Development and validation of the observation list for early signs of dementia (OLD). *International Journal of Geriatric Psychiatry*, *16*, 406–414.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185.

Horney, K. (1945). *Our inner conflicts: a constructive theory of neurosis.* New York: Norton.

Horst, P. (1935). Measuring complex attitudes. *Journal of Social Psychology*, *6*, 369–374.

Hubert, L. J. (1984). Statistical applications of linear assignment. *Psychometrika*, *49*, 449–473.

Hubert, L. J. (1985). Combinatorial data analysis: Association and partial association. *Psychometrika*, *50*, 449–467.

Hubert, L. J. (1987). *Assignment methods in combinatorial data analysis.* New York: Marcel Dekker.

Hubert, L. J., & Schultz, J. (1976). Quadratic assignment as a general data analysis strategy. *British Journal of Mathematical and Statistical Psychology*, *29*, 190–241.

Huyse, F. J., Herzog, T., Lobo, A., Malt, U. F., Opmeer, B. C., & Stein, B. (2000). European consultation-liaison psychiatric services: the ECLN collaborative study. *Acta Psychiatrica Scandinavica*, *101*, 360–366.

Jolliffe, I. T. (2002). *Principal component analysis.* New York: Springer-Verlag.

Keselman, H., Cribbie, R., & Holland, B. (1999). The pairwise multiple comparison multiplicity problem: An alternative approach to familywise and comparisonwise type i error control. *Psychological Methods*, *4*, 58–69.

Killeen, P. R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science*, *16*, 345–353.

Killeen, P. R. (2006). Beyond statistical inference: a decision theory for science. *Psychonomic Bulletin and Review*, *13*, 549–562.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, *29*, 115–129.

Kruskal, J. B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society, Series B*, *27*, 251–263.

Kruskal, J. B., & Shepard, R. N. (1974). A nonmetric variety of linear factor analysis. *Psychometrika*, *39*, 123–157.

Lambert, Z. V., Wildt, A. R., & Durand, R. M. (1991). Approximating confidence intervals for factor loadings. *Multivariate Behavioral Research*, *26*, 421–434.

Landgrebe, J., Wurst, W., & Welzl, G. (2002). Permutation-validated principal components analysis of microarray data. *Genome Biology*, *3*, research/0019.

Lin, S. P., & Bendel, R. B. (1985). Algorithm AS 213: Generation of population correlation matrices with specified eigenvalues. *Applied Statistics*, *34*, 193–198.

Macdonald, R. R. (2005). Why replication probabilities depend on prior probability distributions: A rejoinder to Killeen (2005). *Psychological Science*, *16*, 1007–1008.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*, 209–220.

Markus, M. T. (1994). *Bootstrap confidence regions in nonlinear multivariate analysis.* Leiden, The Netherlands: DSWO Press.

Meulman, J. J. (1982). *Homogeneity analysis of incomplete data.* Leiden, The Netherlands: DSWO Press.

Meulman, J. J. (1984). *Correspondence analysis and stability.* (Unpublished report. Leiden: RR-84-01)

Meulman, J. J. (1992). The integration of multidimensional scaling and multivariate analysis with optimal transformations of the variables. *Psychometrika*, *57*, 539–565.

Meulman, J. J. (1993). Nonlinear principal coordinates analysis: minimizing the sum of squares of the smallest eigenvalues. *British Journal of Mathematical and Statistical Psychology*, *46*, 287–300.

Meulman, J. J. (1996). Fitting a distance model to homogeneous subsets of variables: Points of view analysis of categorical data. *Journal of Classification*, *13*, 249–266.

Meulman, J. J., & Heiser, W. J. (1983). *The display of bootstrap solutions in multidimensional scaling.* Murray Hill, NJ: Bell Laboratories. (Technical memorandum)

Meulman, J. J., Heiser, W. J., & SPSS. (2004). *SPSS Categories 13.0.* Chicago: SPSS Inc.

Meulman, J. J., Van der Kooij, A. J., & Babinec, A. (2002). New features of categorical principal components analysis for complicated data sets, including data mining. In W. Gaul & G. Ritter (Eds.), *Classification, automation, and new media* (pp. 207–217). Berlin: Springer.

Meulman, J. J., Van der Kooij, A. J., & Heiser, W. J. (2004). Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 49–70). London: Sage Publications.

Milan, L., & Whittaker, J. (1995). Application of the parametric bootstrap to models that incorporate a singular value decomposition. *Applied Statistics*, *44*, 31–49.

NICHD Early Child Care Research Network. (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly*, *11*, 269–306.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications.* Toronto: University of Toronto Press.

Nishisato, S. (1994). *Elements of dual scaling: An introduction to practical data analysis.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Noreen, E. W. (1989). *Computer intensive methods for testing hypotheses.* New York: John Wiley and Sons.

Ogasawara, H. (2004). Asymptotic biases of the unrotated/rotated solutions in principal component analysis. *British Journal of Mathematical and Statistical Psychology*, *57*, 353–376.

Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics and Data Analysis*, *49*, 974–997.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science*, *3*, 425–441.

Roskam, E. E. C. I. (1968). *Metric analysis of ordinal data in psychology.* Voorschoten, The Netherlands: VAM.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, *79*, 871–881.

Rousseeuw, P. J., Ruts, I., & Tukey, J. W. (1999). The bagplot: A bivariate boxplot. *The American Statistician*, *53*, 382–387.

Saporta, G., & Hatabian, G. (1986). Régions de confiance en analyse factorielle. [Confidence regions in factor analysis]. In E. Diday (Ed.), *Data analysis and informatics, IV* (pp. 499–508). Amsterdam: Elsevier.

SAS. (1992). *SAS/STAT Software: Changes and enhancements* (Tech. Rep.). SAS Institute Inc., Cary, North Carolina.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177.

Schönemann, P. H. (1966). A generalized solution of the orthogonal procrustes problem. *Psychometrika*, *31*, 1–10.

Shaffer, J. P. (1995). Multiple hypothesis testing. *Annual Review of Psychology*, *46*, 561–584.

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function, II. *Psychometrika*, *27*, 219–246.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function, I. *Psychometrika*, *27*, 125–140.

Shepard, R. N. (1966). Metric structures in ordinal data. *Journal of Mathematical Psychology*, *3*, 287–315.

Smouse, P. E., Long, J. C., & Sokal, R. R. (1985). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, *35*, 627–632.

Sokal, R. R. (1979). Testing statistical significance of geographical variation. *Systematic Zoology*, *28*, 227–232.

SPSS Inc. (2007). *CATPCA algorithm* (Tech. Rep.). Retrieved January 22, 2007, from http://support.spss.com/Tech/Products/SPSS/Documentation/... Statistics/algorithms/index.html; Use "guest" as user-id and password.

Ter Braak, C. J. F. (1992). Permutation versus bootstrap significance tests in multiple regression and ANOVA. In K. H. Jöckel, G. Rothe, & W. Sendler (Eds.), *Bootstrapping and related techniques* (pp. 79–86). Berlin: Springer Verlag.

Theunissen, N. C. M., Meulman, J. J., Den Ouden, A. L., Koopman, H. M., Verrips, G. H., Verloove-Vanhorick, S. P., & Wit, J. M. (2004). Changes can be studied when the measurement instrument is different at different time points. *Health Services and Outcomes Research Methodology*, *4*, 109–126.

Timmerman, M. E., Kiers, H. A. L., & Smilde, A. K. (in press). Estimating confidence intervals for principal component loadings: A comparison between the bootstrap and asymptotic results. *British Journal of Mathematical and Statistical Psychology.*

Van der Burg, E. (1988). *Nonlinear canonical correlation and some related techniques.* Leiden: DSWO Press, Leiden University.

Van der Burg, E., & De Leeuw, J. (1983). Nonlinear canonical correlation. *British Journal of Mathematical and Statistical Psychology*, *36*, 54–80.

Van der Burg, E., & De Leeuw, J. (1988). Use of the multinomial jackknife and bootstrap in generalized canonical correlation analysis. *Applied Stochastic Models Data Analysis*, *4*, 154–172.

Velicer, W. F. (1974). A comparison of the stability of factor analysis, principal component analysis, and rescaled image analysis. *Education and Psychological Measurement*, *34*, 563–572.

Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods*, *3*, 231–251.

Verhoeven, K., Simonsen, K., & McIntyre, L. (2005). Implementing false discovery rate control: Increasing your power. *OIKOS*, *108*, 643–647.

Vlek, C., & Stallen, P. J. (1981). Judging risks and benefits in the small and in the large. *Organizational Behavior and Human Performance*, *28*, 235–271.

Wagenmakers, E. J., & Grünwald, P. (2006). A Bayesian perspective on hypothesis testing: A comment on Killeen (2005). *Psychological Science*, *17*, 641–642.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association*, *22*, 209–212.

Winsberg, S., & Ramsay, J. O. (1983). Monotone spline transformations for dimension reduction. *Psychometrika*, *48*, 575–595.

Young, F., Takane, Y., & de Leeuw, J. (1978). The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling. *Psychometrika*, *43*, 279–281.

Young, F. W. (1981). Quantitative analysis of qualitative data. *Psychometrika*, *46*, 357–387.

Young, F. W., de Leeuw, J., & Takane, Y. (1976). Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. *Psychometrika*, *41*, 505–528.

Zeijl, E., Te Poel, Y., Du Bois-Reymond, M., Ravesloot, J., & Meulman, J. J. (2000). The role of parents and peers in the leisure activities of young adolescents. *Journal of Leisure Research*, *32*, 281–302.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*, 432–442.