The Importance Of The Eigenvalues To Define A Configuration Of The Categories Of A Group Of Qualitative Variables

Colnago Efrem

Department Of Statistics, Probability, Applied Statistics
University Of Roma "La Sapienza", Piazzale Aldo Moro 5, Roma 00185 Italy
Efrem@Pow2.Sta.Uniroma1.It

Keywords: Categorical data, Burt's matrix, Quantification,Residual matrix, Linear and bilinear model

Abstract

There exist several methodologies for analizing multivariate qualitative data(Scaling,Analysis of Corrispondences,Dual scaling,Categorical Data,Multidimensional Scaling,and at last Data Mining).Our aim consist in finding a suitable constrains for discriminating and ordering the categories of the multivarisate set ofqualitative variables according to unidimensional scale(categorical multidimensional data).Hence the problem: exist an acceptable unidimensional scaling for this type of the data and what constrains are needed to arrive an optimal quantifications of the categories?Nishisato (1993) defined in bivariate analysis,the minimum reliable level of the first eigenvalue to discriminate the categories.We will demonstrate,in our application in the multivariate analysis if this approach is reliable.He defines also two type of data matrix in quantifications. One is an"incidence data" matrix,where that exist a second component that influences the data of the first and another "dominance data" matrix ,where this influences does'nt exist. Since we start from a Burt's matrix that usually is of the type "incidence data",then we want verify with an example the influences of the second eigenvalue on the configuration of the first one.We introduce the PCA,as a model with linear and bilinear component (Gabriel 1971)and decompone the initial matrix into a sum of matrices related to the first eigenvalue and of residual traditional matrices.If we eliminate the contribution of the second and other successive eigenvalues we build a special residual matrices,that holds always the first factor.If we can show the independence of the configurations of the first component from the others in special residual matrices,we can demonstre the not influence of the second factor. So two methodological problems exist :the first eigenvalue can optimally order the categories (at what level ?).Which is the acceptance intervall of λ?,based on the test χ^2(F).In this connection we introduce not negative matrix F formed by the configurations of the first eigenvalues in initial and successive special residual matrices.Besides we proceed, demonstrating the not influence of the second component in traditional matrix with a Greenacre test and afterwards also in special residual matrices With a $\chi(F)^2$ on previous special multiway tables of configurations.We will prove,as (Guttman 1944) in scale opinions,(Nishisato 1993) in Dual Scaling,(Kop and Tournois 1995) in multidimensional scaling,when the informations of the initial data matrix are sufficient to discriminate optimally the categories.(optimal quantification).

Reference

GABRIEL R.R.(1971) :-The least squares approximation of a matrix by additive and multiplicative models.{J.K.Pt:Soc n.40 p.186-199}.

GREENACRE MICHAEL J:(1988):-Corrispondence analysis of multivariate categorical data by least squares.-in {Biometrica 457 Great Britain}

Guttman Louis(1944)-"A basic for scaling qualitatrive data"-{American Sociological review-NY n.9 p.139-150}

J.L.KOP et J.TOURNOIS(1995):-Where is the first component?Another perspective on the convergence between multidimensional scaling and components analysis of correlations-{Psychometrika Society Meeting Leiden 1995}

NISHISATO SHIZUHITO(1993):-In quantifying different types of categorical data-{Psychometrika n.58 p.617-628}.