Statistical validation

USE OF THE MULTINOMIAL JACK-KNIFE AND BOOTSTRAP IN GENERALIZED NON-LINEAR CANONICAL CORRELATION ANALYSIS

EEKE VAN DER BURG

University of Twente, Department of Education, PO Box 217, 7500 AE Enschede, The Netherlands

AND

JAN DE LEEUW[*]

University of Leiden, Department of Data Theory, Middelstegracht 4, 2312 TW Leiden, The Netherlands

SUMMARY

In this paper we discuss the estimation of mean and standard errors of the eigenvalues and category quantifications in generalized non-linear canonical correlation analysis (OVERALS). Starting points are the delta method equations, but the jack-knife and bootstrap are used to provide finite difference approximations to the derivatives.

KEY WORDS   Canonical correlation analysis   Delta method   Jack-knife   Bootstrap
            Confidence interval   Non-linear transformation

INTRODUCTION

Non-linear canonical correlation analysis with k sets of variables, OVERALS, is a multivariate technique in the sense of Gifi (Reference 1, chap. 6). It is described by de Leeuw,[2] van der Burg, de Leeuw and Verdegaal,[3,4] Verdegaal[5] and van der Burg and de Leeuw.[6] In general the purpose of k sets canonical correlation is to find linear compounds of sets of variables that are related optimally in some way.[7-9] In the literature several optimization criteria are discussed.[10] In OVERALS the k sets of variables are also related in a linear and optimal way, but at the same time the variables are *transformed non-linearly*. This can be formulated as a least squares problem minimizing the sum of squared deviations between unknown object scores and linear combinations of transformed variables, organized in sets.[3,4,16]

   In current implementations of OVERALS the variables are categorical, i.e. they assume only a small number of possible values. The technique assigns a numerical score to each category, the so-called category quantification. Scores are assigned in such a way that the sum of the t largest *eigenvalues (generalized canonical correlations)* is maximized, while at the same time the measurement characteristics are respected. Thus for ordinal variables we impose ordinal restrictions on the category quantifications, for numerical variables the category quantifications must even be linear with the original scores. For nominal variables there are no additional restrictions.

---

[*]The author has recently moved to UCLA, Department of Psychology, 7619 Franz Hall, Los Angeles, CA 90024–1563, U.S.A.

The scoring system we just outlined gives a single quantification for each variable. It is also possible to obtain multiple quantifications for each variable by using copies.[2,4,6,11,12] This means that a variable occurs more than once in a set. When only nominal characteristics are employed for all copies, the measurement level is called multiple nominal. In case only one quantification is dealt with the measurement level is called single nominal, single ordinal or numerical.

Because variables are categorical the profile for each individual only assumes a finite number of values. The quantities computed in canonical correlation analysis (category quantifications, object or profile scores, canonical correlations, correlations between quantified variables and canonical variables, and so on) are functions of the profile frequencies.

De Leeuw and van der Burg[13] did a small study on the significance of eigenvalues in relation to OVERALS. This time we are interested in stability of eigenvalues and category quantifications. Therefore we want to compute confidence intervals. For this reason the delta method is studied to obtain estimations of the means and standard errors of the two statistics. With the help of the bootstrap and jack-knife resampling methods we estimate the delta method approximations. We compare the results for the two resampling methods.

THE DELTA METHOD

We shall develop our statistical methods in a general multinomial context, not necessarily in terms of profile frequencies or proportions, and not directly applied to OVERALS output. The data are a vector $\mathbf{p}$ of proportions, based on a simple random sample of size n. Thus we suppose that $\mathbf{p}$ is a realization of the random variable p, where np has a multinomial distribution with parameters $(n, \boldsymbol{\pi})$. We embed the variable p in a sequence p_n, where np_n is also multinomial with parameters $(n, \boldsymbol{\pi})$. For the expected value and the dispersion we have $E(p_n) = \boldsymbol{\pi}$ and $C(p_n) = n^{-1}(\mathbf{Pi} - \boldsymbol{\pi}\boldsymbol{\pi}^{\mathrm{T}})$, where $\mathbf{Pi} = \mathrm{diag}(\boldsymbol{\pi})$. We also have convergence in law to a normal distribution, in the sense that $z_n = n^{1/2}(p_n - \boldsymbol{\pi}) \to \mathrm{N}(0, \mathbf{V})$ with $\mathbf{V}$ short for $\mathbf{Pi} - \boldsymbol{\pi}\boldsymbol{\pi}^{\mathrm{T}}$.

Now suppose Φ is a real valued function defined for all $\mathbf{p}$ (or all $\mathbf{p}$ close to $\boldsymbol{\pi}$), and twice continuously differentiable at $\boldsymbol{\pi}$. Then the delta method (Reference 14, section 6.a.2, contains a nice discussion) states that $n^{1/2}(\Phi(p_n) - \Phi(\boldsymbol{\pi})) \to \mathrm{N}(0, \mathbf{g}^{\mathrm{T}}\mathbf{Vg})$, where $\mathbf{g}$ is the vector of partials of Φ at $\boldsymbol{\pi}$. We can easily understand this result by writing p_n as $p_n = \boldsymbol{\pi} + n^{-1/2}z_n$ and then developing a Taylor series for $\Phi(p_n)$ in $n^{-1/2}z_n$ around $\boldsymbol{\pi}$. This gives

$$\Phi(p_n) = \Phi(\boldsymbol{\pi}) + n^{-1/2}\mathbf{g}^{\mathrm{T}}z_n + (2n)^{-1}z_n^{\mathrm{T}}\mathbf{H}z_n + o_p(n^{-1}). \tag{1}$$

A sequence of random variables x_n is $o_p(n^{-1})$ if nx_n converges in probability to zero. The matrix $\mathbf{H}$ contains the second order partials of Φ at $\boldsymbol{\pi}$. The variance of $\Phi(p_n)$ is given by

$$\mathrm{var}_{\boldsymbol{\pi}}(\Phi(p_n)) = \mathbf{E}_{\boldsymbol{\pi}}((\Phi(p_n))^2) - (E_{\boldsymbol{\pi}}(\Phi(p_n)))^2 \tag{2}$$

The second term of (2) is obtained by taking expectations of (1). As the random variable z_n converges in law to $\mathrm{N}(0, \mathbf{V})$ we obtain

$$E_{\boldsymbol{\pi}}(\Phi(p_n)) = \Phi(\boldsymbol{\pi}) + (2n)^{-1}\mathrm{tr}\mathbf{HV} + o(n^{-1}) \tag{3}$$

The squared expectation is

$$(E_{\boldsymbol{\pi}}(\Phi(p_n)))^2 = (\Phi(\boldsymbol{\pi}))^2 + n^{-1}\Phi(\boldsymbol{\pi})\mathrm{tr}\mathbf{HV} + o(n^{-1}) \tag{4}$$

The first term of (2) is obtained by squaring (1) and taking expectations:

$$E_{\boldsymbol{\pi}}(\Phi(p_n))^2 = (\Phi(\boldsymbol{\pi}))^2 + n^{-1}\Phi(\boldsymbol{\pi})\mathrm{tr}\mathbf{HV} + n^{-1}\mathbf{g}^{\mathrm{T}}\mathbf{Vg} + o(n^{-1}) \tag{5}$$

Subtraction of (4) from (5) gives

$$\text{var}_\pi(\Phi(p_n)) = n^{-1}\mathbf{g}^T\mathbf{V}\mathbf{g} + o(n^{-1}) \tag{6}$$

which corresponds to the delta method variance. Result (6) makes it possible to estimate the standard error of statistics of the form $\Phi(p_n)$. The estimate is $n^{-1}\hat{\mathbf{g}}^T\hat{\mathbf{V}}\hat{\mathbf{g}}$ where $\hat{\mathbf{g}}$ estimates $\mathbf{g}$ and $\hat{\mathbf{V}}$ estimates $\mathbf{V}$. Usually $\hat{\mathbf{V}} = \mathbf{V}_p = \mathbf{P} - \mathbf{p}\mathbf{p}^T$ with $\mathbf{P} = \text{diag}(\mathbf{p})$, and $\hat{\mathbf{g}} = \mathbf{g}_p$ the partials of Φ at $\mathbf{p}$. If we have an estimate $\hat{\sigma}$ of the standard error, then we also have a confidence interval (95 per cent) of the form $(\Phi(p_n) - 1\cdot96\hat{\sigma}, \Phi(p_n) + 1\cdot96\hat{\sigma})$.

We can also evaluate the bias of $\Phi(p_n)$ as an estimate of $\Phi(\pi)$. For that reason we go back to (3). This expression shows that

$$E_\pi((\Phi(p_n)) - \Phi(\pi)) = (2n)^{-1}\text{tr}\mathbf{H}\mathbf{V} + o(n^{-1}) \tag{7}$$

so that estimates of $\mathbf{H}$ and $\mathbf{V}$ (e.g. $\mathbf{H}_p$ and $\mathbf{V}_p$) give an estimate of $\Phi(\pi)$ with bias of order n^{-1}.

In many cases, and for instance in generalized canonical correlation analysis, the quantities of interest are defined by very complicated implicit functions. This makes it extremely tedious to compute first-order derivatives, while second-order derivatives are usually well nigh impossible to obtain. In some special cases, such as correspondence analysis, the delta method can be applied (Reference 1, chap. 12; References 15–17 and Reference 18, appendix B), but in other cases it simply is not feasible. In such cases we can use a resampling method such as bootstrap (Reference 19 gives an overview), which we view here as methods to approximate the relevant partials. Gifi (Reference 1, chap. 13) concentrates on the bootstrap. In this paper we use both jack-knife and bootstrap and compare the results.

THE MULTINOMIAL JACK-KNIFE

Suppose we drop one observation from the sample. If it has profile number k, then the vector of profile proportions changes to

$$\mathbf{p}(k) = \mathbf{p} + (n - 1)^{-1}(\mathbf{p} - \mathbf{e}_k) \tag{8}$$

with $\mathbf{e}_k$ the kth unit vector. The jack-knife value for the observation is $\Phi(\mathbf{p}(k))$, and the pseudovalue $\Phi_k(\mathbf{p})$ is defined as

$$\Phi_k(\mathbf{p}) = n\Phi(\mathbf{p}) - (n - 1)\Phi(\mathbf{p}(k)) \tag{9}$$

The jack-knife method uses the average pseudovalue $\Sigma p_k\Phi_k(\mathbf{p})$ as an estimate of $\Phi(\pi)$ (p_k are the elements of the vector $\mathbf{p}$), and uses the variance of the pseudovalues as an estimate of $\mathbf{g}^T\mathbf{V}\mathbf{g}$. We shall explain why this is a reasonable procedure. Observe for the moment that no derivatives need to be computed.

Write Φ_* for the average of the pseudovalues, and σ_*^2 for their variance. Then combination of (8) and (9) using a Taylor series for $\Phi(\mathbf{p}(k))$ in $(\mathbf{p} - \mathbf{e}_k)$ around $\mathbf{p}$ gives

$$\Phi_k(\mathbf{p}) = \Phi(\mathbf{p}) - \mathbf{g}_p^T(\mathbf{p} - \mathbf{e}_k) - (2n - 2)^{-1}(\mathbf{p} - \mathbf{e}_k)^T\mathbf{H}_p(\mathbf{p} - \mathbf{e}_k) + o(n^{-1}) \tag{10}$$

Thus

$$\Phi_*(\mathbf{p}) = \Phi(\mathbf{p}) - (2n - 2)^{-1}\text{tr}\mathbf{V}_p\mathbf{H}_p + o(n^{-1}) \tag{11}$$

If we combine this result with the bias estimate provided by the delta method, given in (7), we see that

$$(n - 1)(E_\pi(\Phi_*(p_n)) - \Phi(\pi)) \to 0 \tag{12}$$

Thus the average pseudovalue corrects for bias, in the same way as the delta method adjustment requiring second-order derivatives.

To compute σ_*^2 we observe that

$$\Phi_k(\mathbf{p}) - \Phi_*(\mathbf{p}) = -\mathbf{g}^{\mathrm{T}}(\mathbf{p} - \mathbf{e}_k) + o(1) \tag{13}$$

From (13) it follows that

$$\sigma_*^2(\mathbf{p}) \to \mathbf{g}_p^{\mathrm{T}} \mathbf{V}_p \mathbf{g}_p \tag{14}$$

Thus σ_*^2/n is asymptotically equal to the delta method variance estimate. More precise expansions can be found in Reference 11.

The only computations that are required are that the analysis technique is repeated for each observation that is successively dropped. For each jack-knife sample a pseudovalue is computed. This gives new estimates of the relevant quantities. Their mean is the improved estimate, their variance is an estimate of the stability. If the number of profiles is much smaller than the number of observations, then it is more efficient to organize the computations in terms of profiles (as above), because repeating exactly the same analysis for observations with the same profile is avoided.

THE RANDOM JACK-KNIFE

Because generalized canonical correlation analysis with a large number of observations (and/or profiles) can be very expensive, it is often not feasible to compute all pseudovalues. Instead we can estimate the average pseudovalue and the variance of the pseudovalues by Monte Carlo methods. This amounts to leaving out one observation at random, and repeating this a number of times. If the sampling is repeated this obviously converges to the theoretical jack-knife.

It must be remarked that the delta method provides us with an approximation to the standard error. The jack-knife in our interpretation, gives an approximation to the delta method approximation of the standard error. And the Monte Carlo method approximates the jack-knife approximations. Thus there are three levels of approximation involved. It does not follow, of course, that approximation of the true standard error becomes progressively worse, because there can be complicated interactions between the three approximation processes.

We also emphasize that it is not necessary to present the jack-knife as an approximation of the delta method. It can also be interpreted in its own right as a method to study stability, indeed the idea of investigating the effect of 'leaving-one-out' also makes sense in a non-stochastic context.

THE MULTINOMIAL BOOTSTRAP

If we look at the basic properties of the jack-knife, as we have presented it, we see that the vector of profile proportions is perturbed by leaving out single observations. We apply our technique to all these perturbed vectors, which are located regularly around the observed vector, and we use this grid of perturbed values to estimate the relevant derivatives in a clever way. Because observations are assumed to be equally important, each perturbation of the sample is given the same weight, i.e. it occurs once in the distribution of all possible jack-knife samples. If we use the corresponding perturbations of the profile proportions we need weights $p_1, \ldots, p_m$ for the different profiles.

The bootstrap is based on a different set of perturbations. In fact bootstrapping means resampling with replacement from the original sample, taking n observations.[19] It means that we look at all vectors of profile frequencies adding up to n. It also means that the perturbations of the profile frequencies are centred around the sample value $\mathbf{p}$, and that they have weights according to their similarity with the sample value. Suppose

$$\mathbf{q} = (n_1/n, ..., n_m/n) \tag{15}$$

is a bootstrap perturbation of the profile proportions. Then the probability to occur in the distribution of all possible bootstrap samples is

$$w_p(\mathbf{p}) = \frac{n!}{n_1! ... n_m!} \ p_1^{n_1} ... p_m^{n_m}$$

The bootstrap pseudovalues are defined as

$$\Phi_q(\mathbf{p}) = 2\Phi(\mathbf{p}) - \Phi(\mathbf{q})$$

The average bootstrap pseudovalue is

$$\Phi_*(\mathbf{p}) = \Sigma_q w_q(\mathbf{p})\Phi_q(\mathbf{p}) \tag{18}$$

and the bootstrap variance is

$$\sigma_*^2(\mathbf{p}) = \Sigma_q w_q(\mathbf{p})[\Phi_q(\mathbf{p}) - \Phi_*(\mathbf{p})]^2 \tag{19}$$

Substitute $\mathbf{q} = n^{-1/2}\mathbf{z} + \mathbf{p}$ in the pseudovalues (17) and develop a Taylor series in $n^{-1/2}\mathbf{z}$ around $\mathbf{p}$. This gives

$$\Phi_q(\mathbf{p}) = \Phi(\mathbf{p}) - n^{-1/2}\mathbf{g}_p^T\mathbf{z} - (2n)^{-1}\mathbf{z}^T\mathbf{H}_p\mathbf{z} + o(n^{-1}) \tag{20}$$

The the average pseudovalue is

$$\Phi_*(\mathbf{p}) = \Sigma w_q(\mathbf{p})\Phi_q(\mathbf{p}) = \Phi(\mathbf{p}) - (2n)^{-1}\text{tr}\mathbf{H}_p\mathbf{V}_p + o(n^{-1}) \tag{21}$$

as $\Sigma w_q(\mathbf{p})\mathbf{z} = 0$ and $\Sigma w_q(\mathbf{p})\mathbf{z}\mathbf{z}^T = \mathbf{V}_p$. The expected value of $\Phi_*(p_n)$ is

$$E_\pi\Phi_*(p_n) = E_\pi(\Phi(p_n)) - (2n)^{-1}\text{tr}\mathbf{H}\mathbf{V} + o(n^{-1}) \tag{22}$$

Combining (22) with (3) gives

$$n(E_\pi\Phi_*(p_n) - \Phi(\pi)) \rightarrow 0 \tag{23}$$

Thus the average pseudovalue is an estimator of the population mean $\Phi(\pi)$ with bias of order n^{-1}. For the variance of the pseudovalues we subtract (21) from (20) which gives

$$\Phi_q(\mathbf{p}) - \Phi_*(\mathbf{p}) = -n^{-1/2}\mathbf{g}_p^T\mathbf{z} - (2n)^{-1}\mathbf{z}^T\mathbf{H}_p\mathbf{z} + (2n)^{-1}\text{tr}\mathbf{H}_p\mathbf{V}_p + o(n^{-1}) \tag{24}$$

The the pseudovalue variance is

$$\sigma_*^2(\mathbf{p}) = n^{-1}\mathbf{g}_p^T\mathbf{V}_p\mathbf{g} + o(n^{-1}) \tag{25}$$

The expected value of $\sigma_*^2(p_n)$ converges to the delta method variance

$$n(E_\pi\sigma_*^2(p_n)) \rightarrow \mathbf{g}^T\mathbf{V}\mathbf{g} \tag{26}$$

Thus the bootstrap variance is an asymptotically unbiased estimator of $n^{-1}\mathbf{g}^T\mathbf{V}\mathbf{g}$. This result means that the variance of the bootstrap values (and the pseudovalues) estimates the delta method variance. The jack-knife pseudovalue variance estimates $\mathbf{g}^T\mathbf{V}\mathbf{g}$, so that the estimates of bootstrap and jack-knife differ by a factor n.

For large n a random version of the jack-knife is necessary. For the bootstrap a random version is nearly always obligatory, as there are n^n bootstrap samples possible. If the number of bootstrap samples is r, we need of course to take r large enough. Not very much research has been done on what is large enough. Only Borsboom and van Pelt (unpublished) did some research on this subject with regard to a computer program for non-linear canonical correlation analysis (CANALS[20]). They took bootstrap samples (of 4241 observations) adding one at a time and recomputing standard errors of category quantifications. When the difference between previous and current standard errors was smaller than 0·01 they stopped taking bootstrap samples. Using the category with the largest standard error as a criterion, they judged that about 40 samples is enough for CANALS in this example. Their method has been implemented in a computer program by Borsboom and Visser.[21] We repeated the analysis of Borsboom and van Pelt with the Whales data (for a description see further on). We added 10 samples each time, starting at 10 and ending with 120.

EXAMPLES

In this section the jack-knife and bootstrap procedures are used to compute pseudovalue means and standard errors in order to compute confidence intervals and to estimate the bias for three different data sets. For the smaller data set (i.e. Whales) a complete jack-knife and a random bootstrap procedure was applied, whereas in the case of the FYTY and SIMS data sets Monte Carlo versions of jack-knife and bootstrap were used. For all data sets the generalized canonical correlations (i.e. eigenvalues) were employed. In addition, for the FYTY data some category quantifications were also considered. An overview of the results of these analyses is presented in tables found in the following sections of this paper. Many of these tables have a column for eigenvalues (respectively category quantifications), computed from the original data matrix (or sample), in single precision (SP) and double precision (DP). For single precison the convergence of the OVERALS program[30] is computed with accuracy to 10^{-5}, whereas for double precision it is computed with accuracy to 10^{-8}. The OVERALS convergence criterion is specified in terms of the difference between the sum of eigenvalues for two consecutive iterations in the alternating least squares procedure. Many tables which follow have three columns for estimated population means, of which there are two for jack-knife (single and double precision, JSP and JDP) and one for bootstrap (single precision, BSP). In addition many tables also contain estimated standard errors corresponding to the estimated means.

The first example considered (the Whales data set,[22]) consists of fifteen variables describing characteristics of 36 whales, porpoises and dolphins (e.g. form of the head, kind of feeding, place of blow hole, colour, etc.). Using the twelve variables without missing scores we repeated an analysis described by van der Burg.[23] This means that the program OVERALS was used for homogeneity analysis (multiple correspondence analysis): twelve sets each consisting of one variable treated as multiple nominal.

The second example (designated FYTY for From Year to Year) is based on a sample from a large school career survey (for references see Reference 24). In the example we consider, there are 520 schoolchildren and six variables divided into three subsets. The variables are choice of school after primary education (set 1), achievement test score and teacher's advancement recommendation (set 2), educational level of father and mother, and profession of father (set 3). The different sets were measured at different points in the time period; the order considered was subset 3, subset 2, subset 1. The FYTY data were analysed with numerical, single nominal and multiple nominal options.

The third example also comes from a school survey, in this case the SIMS (Second International Mathematics Study) Project.[25,26] For this example the complete database (4863 schoolchildren) as well as a sample from that base (1000 schoolchildren) were considered. The eigenvalues corresponding to the complete database were treated as population parameters, whereas the sample data were used to estimate these population parameters. The variables considered in this example were also divided into three subsets. One subset contains the variables type of school and father's education, another subset contains three attitudes towards mathematics, and the remaining subset includes a mathematics test score. Four measurement levels were considered for the SIMS data example. Single nominal, ordinal and numerical, and multiple nominal.

RESULTS FOR WHALES

Table I contains bootstrap estimates for the Whales example. As earlier specified, the bootstrap procedure was implemented for various numbers of samples with increments of size 10. The number of samples ranged from 10 to 120. Results for this example were computed in both single and double precision. The estimated population means and standard errors of the first three eigenvalues are presented in Table I. It may be noted that the bootstrap pseudovalue means (estimated population means) show only minor differences, even for the extreme cases (i.e. numbers of samples of 10 and 120). However, the estimated standard error tended to increase for numbers of samples from 10 to 40, while remaining rather stable for numbers of samples greater than 40. In addition, it was found that differences were negligible between the

Table I. Whales, multiple nominal. Estimated populations means and estimated standard errors of three eigenvalues for bootstraps with different sample sizes. SP = single precision, DP = double precision

No.	Population mean		Standard error		No.	Population mean		Standard error	
	SP	DP	SP	DP		SP	DP	SP	DP
10	0·601	0·601	0·031	0·031	70	0·608	0·608	0·042	0·042
	0·397	0·398	0·058	0·058		0·395	0·396	0·041	0·041
	0·306	0·305	0·022	0·017		0·298	0·299	0·026	0·025
20	0·602	0·602	0·029	0·029	80	0·608	0·607	0·043	0·043
	0·393	0·393	0·049	0·049		0·394	0·394	0·044	0·044
	0·308	0·308	0·020	0·018		0·297	0·298	0·025	0·025
30	0·605	0·605	0·032	0·032	90	0·610		0·045	
	0·395	0·396	0·046	0·046		0·394		0·046	
	0·302	0·303	0·024	0·023		0·297		0·025	
40	0·606	0·606	0·041	0·041	100	0·612		0·047	
	0·391	0·392	0·043	0·043		0·394		0·044	
	0·298	0·299	0·027	0·026		0·296		0·025	
50	0·609	0·609	0·040	0·040	110	0·612		0·046	
	0·391	0·392	0·042	0·042		0·393		0·043	
	0·299	0·300	0·027	0·026		0·295		0·025	
60	0·610	0·610	0·039	0·039	120	0·613		0·045	
	0·394	0·395	0·041	0·041		0·393		0·042	
	0·298	0·299	0·027	0·026		0·295		0·026	

E. VAN DER BURG AND J. DE LEEUW

two levels of precision. For the remaining examples 40 samples were considered when a random process was used in identifying samples. In addition, only single precision computation was used with the bootstrap.

The eigenvalues computed from the Whales data are presented in Table II. Four methods were considered: SP, DP, APL and CA. The first two use the Fortran OVERALS program, and result in similar eigenvalues. The third method is based on an APL version of the OVERALS algorithm. Since APL is quite precise (16 digits accuracy) the results obtained under this method may be considerably more accurate than results from the Fortran program. The fourth method considered was correspondence analysis (CA). In those cases in which all variables are treated as multiple nominal, the OVERALS problem may be reformulated in a format which is consistent with correspondence analysis. For this reason a correspondence analysis program (ANACOR[27] was used. The eigenvalues resulting from ANACOR are precisely the same as the OVERALS eigenvalues. However, the variances do differ. Although the APL eigenvalues are slightly different from those obtained under the above-mentioned approaches, those differences hardly seem worth mentioning. The jack-knife and bootstrap results are presented in columns 4 to 11 of Table II. The bootstrap results are based on 120 samples, whereas the JSP and JDP results are based on 36 samples (a complete jack-knife). The estimated means do not substantially differ from each other, but the estimated standard errors do manifest sizeable differences. The three jack-knife estimates are very similar; however, the bootstrap estimates are considerably lower in magnitude than the jack-knife estimates. It is clear from Table I that using more samples will not change the estimated standard error. Thus in this case the bootstrap results converge to values different from those of the jack-knife method. As the number of observations is very small in this example (i.e. 36) this may be due to the fact that asymptotic characteristics are not satisfied, and thus approximations may be imprecise. We can compare the results obtained under OVERALS with those obtained under CA. As it is possible to compute the first-order derivatives in case of correspondence analysis, the ANACOR program delivers variances (0·0025, 0·0031, 0·0012, respectively). The variances for the jack-knife and bootstrap are computed from Table II by squaring the standard error. This gives for JDP (0·0031, 0·0036, 0·0008) and for BSP (0·0020, 0·0018, 0·0007). The first two bootstrap estimates are systematically lower than the CA values, whereas the corresponding jack-knife estimates are higher. Only the third variance is similar for the bootstrap and jack-knife methods (and lower than for CA). Assuming that CA gives the more precise value (direct computation instead of approximation), we find that the jack-knife overestimates and the bootstrap underestimates the standard error. We do not make conclusions on the third variance, as the smallest eigenvalue is normally much less precise than the larger ones.

Table II. Whales, multiple nominal. Eigenvalues, estimated population means and estimated standard errors for jack-knife (J) (36 samples), bootstrap (B) (120 samples) and OVERALS-APL. SP = single precision, DP = double precision

	Eigenvalue		Population mean				Standard error			
	DP*	APL	JSP	JDP	JAPL	BSP	JSP	JDP	JAPL	BSP
Multiple	0·635	0·637	0·618	0·611	0·614	0·613	0·056	0·056	0·056	0·045
Nominal	0·413	0·415	0·385	0·398	0·400	0·393	0·060	0·060	0·061	0·042
	0·317	0·320	0·279	0·282	0·285	0·295	0·028	0·028	0·026	0·026

*Eigenvalues SP = DP = CA (correspondence analysis).

Another small example ($n = 47$) was tested. This time three sets of variables were available. The results were disappointing in the sense that standard errors were huge and estimations of means differed very much from the population values (we did not take a sample). As homogeneity analysis is known to be a stable technique it did not surprise us that the OVERALS results of this small example were worse than in case of the Whales data. However they looked really bad (worse for jack-knife than bootstrap). We supposed that this was mainly due to the small size of the data set, so that we turned over to larger sizes.

RESULTS FOR FYTY

The FYTY data were analysed by the Monte Carlo version of jack-knife and bootstrap. The resulting eigenvalue estimates are presented in Table III. Considering the FYTY data as a sample from a large database, we can expect eigenvalues obtained from the original sample to be larger than estimates obtained from bootstrap and jack-knife. This is because sample estimates tend to overestimate the population parameters, whereas the jack-knife and bootstrap pseudovalue means are supposed to reduce the bias found in sample estimates. The actual estimation outcomes that were obtained for the FYTY data are as follows. The jack-knife estimates were rather similar to one another but do not seem to show any relation with the sample eigenvalues. Here the JSP estimation (multiple nominal, second eigenvalue) seems to have outperformed the JDP estimation, but that is not really true. The JSP values have been produced with three digit accuracy, which was not sufficiently precise. As in this case all the jack-knife values differed from the sample value in the fourth and higher digits, we obtain the same values when we truncate at three digits. Consequently, sample and jack-knife estimates are the same. The JDP estimate, in which six digit accuracy is provided, is much larger, but must be more precise. Thus for FYTY eigenvalues the JSP results may be somewhat suspect. For this reason we concentrate on the JDP results. For the following example, we used output with six digits accuracy to avoid this problem.

The bootstrap estimates of the population mean appear better than the jack-knife estimates. The bootstrap estimates are indeed always smaller than the sample eigenvalues. For all measurement levels the first value provided a closer approximation than the second value. Many eigenvalue routines give more precise results for the larger eigenvalues (supposing that the

Table III. FYTY, three measurement levels. Eigenvalues, estimated population means and estimated standard errors. SP = single precision, DP = double precision, J = jack-knife, B = bootstrap (both 40 samples)

	Eigenvalue		Population mean			Standard error		
	SP	DP	JSP	JDP	BSP	JSP	JDP	BSP
Single	0·735	0·734	1·202	1·060	0·731	0·061	0·056	0·021
Nominal	0·399	0·400	0·269	0·167	0·367	0·034	0·025	0·017
Numerical	0·695	0·695	0·851	0·679	0·692	0·016	0·014	0·017
	0·341	0·341	0·639	0·567	0·326	0·057	0·054	0·013
Multiple	0·742	0·742	0·690	0·695	0·729	0·014	0·012	0·018
Nominal	0·555	0·555	0·555	0·626	0·535	0·019	0·026	0·016

smaller ones are also computed). This phenomenon may provide an explanation for the difference in accuracy obtained for first and second eigenvalues.

Owing to the large number of observations, standard errors are smaller than those found for the example with 47 cases (for which no tables are shown) and in the same range as for the Whales data. However, especially in the case of the jack-knife, these standard errors are still unacceptably large.

Until now it would seem that the bootstrap method is preferable to the jack-knife method, since it appears to provide more precise estimates with smaller variance. As the number of sample observations is still not large, an additional sample size which more closely approximates the asymptotic case seemed warranted. For this reason the next example is based on 1000 observations.

However, before this next example is presented some category quantifications of the FYTY data are considered. To do this, the variable PRE and TON are used (Table IV). Looking at the standard errors for these variables, it may be noted that TON is much more stable than PRE. We also see that the standard errors of the category quantifications are larger than those for the eigenvalues (Table III). As category quantifications are proportional to the square roots of eigenvalues, their standard errors will increase at a corresponding rate. In Table IV we see that the JSP and JDP results agree rather well, but the correspondence between sample eigenvalues and the estimated population means is almost non-existent for the jack-knife. Bootstrap results are very similar to the sample values for TON; however, results for PRE show less similarity (but not as bad as jack-knife estimates). A direct look at the bootstrap and jack-knife results (without computing pseudovalues) shows their values to be very stable (Figure 1). Every dash (—) corresponds to a category quantification resulting from one bootstrap/jack-knife sample (called a bootstrap/jack-knife sample value), and the symbol 0 represents the category quantifications of the original FYTY sample. On the horizontal axis the original category scores are scaled, and on the vertical axis the category quantifications are scaled. As the jack-knife and bootstrap sample value variances differ asymptotically by a factor n, we divided the bootstrap sample values (in deviation from their mean) by $n^{1/2}$ to make the results

Table IV. FYTY, single nominal. Frequencies, category quantifications, estimated population means and estimated standard errors for two variables. SP = single precision, DP = double precision, J = jack-knife, B = bootstrap (both 40 samples)

Frequency	Category quantifications		Population mean			Standard error		
	SP	DP	JSP	JDP	BSP	JSP	JDP	BSP
TON								
42	−0·586	−0·586	−2·052	−2·065	−0·584	0·175	0·173	0·128
208	−0·853	−0·853	−1·774	−1·748	−0·851	0·077	0·075	0·087
165	0·056	0·056	2·119	2·132	0·080	0·125	0·122	0·146
105	1·853	1·853	0·317	0·291	1·867	0·125	0·124	0·125
PRE								
49	−0·622	−0·622	−43·102	−40·287	−0·617	6·574	5·641	0·581
147	−0·592	−0·592	−23·000	−20·275	−0·694	4·157	3·696	0·457
181	−0·184	−0·184	8·989	7·886	−0·126	2·587	2·143	0·295
111	0·298	0·298	33·994	32·606	0·200	4·147	3·598	0·676
32	3·680	3·680	5·795	−0·212	4·275	3·022	2·350	0·811

comparable for the two techniques. It may be noted that both variables behave very stably. In the case of TON all forty dashes (jack-knife/bootstrap sample values) fall together both for jack-knife and bootstrap, whereas for PRE we find some variability among the category quantifications. Note that in the case of JDP, most of the variability that is present is due to a single outlying sample. This stability in the results is very striking.
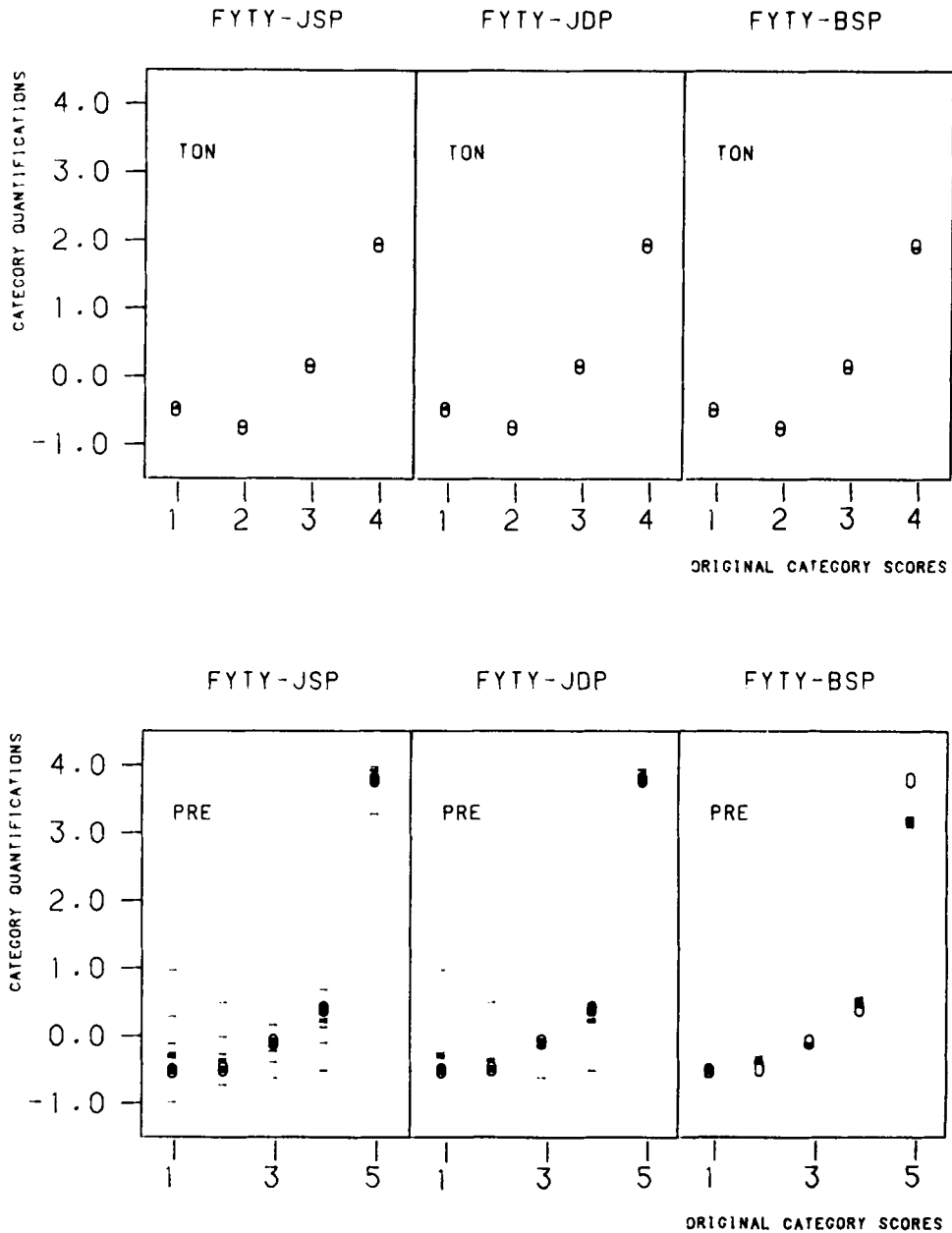

Figure 1.   FYTY. Category quantifications for TON and PRE based on three methods: JSP, JDP and BSP (0 original sample value, — jack-knife/bootstrap sample value)

Table V. SIMS, four measurement levels. Eigenvalues of population (POP) and sample (SAM), estimated population means and estimated standard errors. SP = single precision, DP = double precision, J = jack-knife, B = bootstrap (both 40 samples)

	Eigenvalue		Population mean			Standard error		
	POP*	SAM*	JSP	JDP	BSP†	JSP	JDP	BSP
Multiple	0·615	0·625	0·615	0·620	0·619	0·011	0·011	0·011
Nominal	0·450	0·463	0·566	0·463	0·454	0·013	0·013	0·012
Single	0·613	0·622	0·621	0·622	0·616	0·014	0·013	0·012
Nominal	0·352	0·377	0·171	0·377	0·371	0·015	0·013	0·010
Single	0·613	0·622	0·728	0·628	0·619	0·014	0·014	0·012
Ordinal	0·352	0·377	0·602	0·697	0·375	0·094	0·092	0·012
Numerical	0·606	0·615	0·609	0·608	0·613	0·012	0·012	0·012
	0·350	0·363	0·277	0·321	0·359	0·010	0·009	0·011

*No difference between eigenvalues SP and DP, except for the first eigenvalue of single nominal (SP, SAM): 0·621.
†BDP with convergence to 10^{-5} gives the same results.

RESULTS FOR SIMS

In Table V we find results for the eigenvalues of the SIMS data. The database itself consisted of 4863 observations and served as the population of interest. For the purpose of estimating the population eigenvalues a sample of 1000 observations was considered. The bootstrap and jack-knife approaches were used (on the sample) for re-estimation of the population values (with the computer results specified to six digits for both SP and DP). Note that bootstrap estimates are always between the population and the sample values. Theoretically this is to be expected. The same outcome is also expected when the jack-knife is used. However, this did not occur in half the cases of Table V. Since standard errors appear to be reasonably small, it is suggested that confidence intervals for bootstrap results be computed.

DISCUSSION

In comparing jack-knife with bootstrap one is left with the impression that the jack-knife method is far more imprecise than the bootstrap method. Apparently jack-knife approximations are less stable for smaller samples. The use of double precision computations tends to improve the jack-knife results, but does not eliminate the problem. For the bootstrap method it does not matter whether computations are SP or DP.

For smaller samples the jack-knife standard error is larger than the bootstrap standard error. It is possible that approximations are not precise enough because of the small sample size. It appears that the jack-knife and bootstrap method converge to different values.

Jack-knife results are sensitive to precision and thus to computational error. This can be concluded from the differences in SP and DP results. This finding does not vary with sample size. It concerns mainly the estimated population means of eigenvalues and not the corresponding standard errors. In relation to eigenvalues, the estimation based on bootstrap is more robust than estimation based on jack-knife.

For smaller samples the estimated standard errors of eigenvalues (both for jack-knife and

bootstrap) are unacceptably large (varying from 0·025 to 0·105 for the smaller sample for which no tables are shown). Such error variation would result in huge confidence intervals (with substantial overlap among intervals).

In all cases considered, the estimated population means for eigenvalues appear to be worse for jack-knife than for bootstrap, as bootstrap estimates are nearer to the population or sample value, than jack-knife estimates are. Even for the largest sample considered, more deviations (i.e. population mean not in the 95 per cent confidence interval) were found with jack-knife than with bootstrap. Deviations occur for second eigenvalues (ordinal and numerical) of JDP and for all second eigenvalues of JSP (Table V). Many eigenvalue routines are less precise for the smallest computed eigenvalue. As the jack-knife is sensitive to precision this may be the cause. The eigenvalue routines used are from Eispack.[28] It concerns the Fortran routines IMTQL2 and TRED2. The method followed is tridiagonalization of a symmetric matrix (TRED2) and singular value decomposition using the implicit QL method of this tridiagonal matrix (IMTQL2). We used single and double precision versions.

Estimates for category quantifications seem rather strange. Category quantifications correspond to the square roots of eigenvalues, and thus will be less precise. However the estimates are sometimes highly inaccurate, even for very stable jack-knife or bootstrap sample values. Again in this case the bootstrap seems to perform much better than the jack-knife (Table IV). In Figure 1 stability of sample results is illustrated.

In the case of the application based on the largest sample, the bootstrap means computed directly from bootstrap samples provided a less accurate estimate of the population value than was provided by pseudovalue means which are presented in Table V. A similar outcome was found (although the differences were less) for jack-knife. Using pseudovalues tends to provide a better approximation than obtained by averaging bootstrap and jack-knife results directly.

If we compare other studies using jack-knife and bootstrap, we find that Boomsma[29] concludes that both methods give very similar results for a sample size of 100. He estimates parameters for covariance structure analysis. We cannot confirm his findings in the case of OVERALS parameters.

In conclusion, we can say that the bootstrap method performed better than the jack-knife method. For larger samples the bootstrap procedure works quite well for computing confidence intervals. The use of 40 samples seems to be sufficient for estimation, but was not thoroughly investigated in our study. For larger samples eigenvalues computed from OVERALS seem quite stable. However, category quantifications seem to result in much wider confidence intervals. Studying results for category quantifications directly (i.e. jack-knife and bootstrap sample values) leads to more positive conclusions which are more in agreement with others' earlier experience with the example here considered.[24]

REFERENCES

1. A. Gifi, *Nonlinear Multivariate Analysis*, Department of Data Theory, University of Leiden, 1981; DSWO Press, Leiden, in press.
2. J. de Leeuw, 'The Gifi-system of nonlinear multivariate analysis', in E. Diday, M. Jambu, L. Lebart, J. Pages and R. Tomassone (eds), *Data Analysis and Informatics*, North Holland, Amsterdam, 1984, pp. 231–243.
3. E. van der Burg, J. de Leeuw and R. Verdegaal, 'Generalized nonlinear canonical correlation with *M* sets of variables', *RR-84-12*, Department of Data Theory, University of Leiden, 1984.
4. E. van der Burg, J. de Leeuw and R. Verdegaal, 'Homogeneity analysis with *k* sets of variables: an alternating least squares method with optimal scaling features', *RR-86-5*, Department of Education, University of Twente, 1986. To appear in *Psychometrika*.

5. R. Verdegaal, 'Meer sets analyse voor kwalitatieve gegevens (More sets analysis for qualitative data)', *RR-85-14*, Department of Data Theory, University of Leiden, 1985.
6. E. van der Burg and J. de Leeuw, 'Nonlinear canonical correlation with *k* sets of variables', *RR-87-8*, Department of Education, University of Twente, 1987, Submitted for publication.
7. P. Horst, 'Relations among *m* sets of measures', *Psychometrika*, **26**, 129–149 (1961).
8. J. D. Carroll, 'Generalization of canonical correlation analysis to three or more sets of variables', *Proceedings 76'th APA Convention*, 1968, pp. 227–228.
9. J. R. Kettenring, 'Canonical analysis of several sets of variables', *Biometrika*, **56**, 433–451 (1971).
10. J. P. van de Geer, 'Linear relations between *k* sets of variables', *Psychometrika*, **49**, 79–94 (1984).
11. J. de Leeuw, 'Jackknife and bootstrap in multinomial situations', *RR-85-16*, Department of Data Theory, University of Leiden, 1985.
12. R. Tijssen, 'A new approach to non-linear canonical correlation analysis', *PRM-85-01*, Department of Psychology, University of Leiden, 1985.
13. J. de Leeuw and E. van der Burg, 'The permutational limit distribution of generalized canonical correlations', in E. Diday, Y. Escoufier, L. Lebart, J. P. Pagès, Y. Schecktman and R. Tomassone (eds), *Data Analysis and Informatics, IV*, Elsevier Science Publishers (North-Holland), Amsterdam, 1986.
14. C. R. Rao, *Linear Statistical Inference and its Applications*, Wiley, New York, 1965.
15. M. E. O'Neill, 'Asymptotic distributions of the canonical correlations from contingency tables', *Australian Journal of Statistics*, **20**, 75–82 (1978).
16. J. de Leeuw, 'Statistical properties of multiple correspondence analysis', *RR-84-06*, Department of Data Theory, University of Leiden, 1984.
17. B. M. Schriever, *Order Dependence*, Series: CWI tracts no. 20, CWI, Amsterdam, 1986.
18. A. Z. Israëls, *Eigenvalue Techniques for Qualitative Data*, DSWO Press, Leiden, 1987.
19. B. Efron, *The Jackknife, the Bootstrap, and other Resampling Plans*, CBMS-NSF Regional Conference Series in Applied Mathematics, no. 38, SIAM, Philadelphia, 1982.
20. E. van der Burg and J. de Leeuw, 'Non-linear canonical correlation', *British Journal of Mathematical and Statistical Psychology*, **36**, 54–80 (1983).
21. G. J. J. M. Borsboom and R. A. Visser, 'Interpreting analysis of qualitative multivariate data using bootstrap methods', Contribution to the ZUMA conference, March 1987.
22. G. Vescia, 'Automatic classification of cetaceans by similarity aggregation', in J. F. Macrotorchino, J. M. Proth and J. Janssen (eds), *Data Analysis of Real Life Environment: Ins and Outs of Solving problems*, North Holland, Amsterdam, 1985, pp. 15–24.
23. E. van der Burg, 'HOMALS classification of whales, porpoises and dolphins', in J. F. Macrotorchino, J. M. Proth and J. Janssen (eds), *Data Analysis in Real Life Environment: Ins and Outs of Solving Problems*, North Holland, Amsterdam, 1985, pp. 25–36.
24. J. de Leeuw and I. Stoop, 'Sekundaire analyse "van jaar tot jaar" met behulp van niet lineaire multivariate technieken (Second analysis "From year to year" with the help of nonlinear multivariate techniques)', in J. L. Peschar (ed.), *Van Achteren naar Voren*, Staatsuitgeverij, s'Gravenhage, 1979.
25. W. J. Pelgrum and Th. I. H. M. Eggen, *Tweede Wiskunde Projeckt: Opzet en Uitvoering (Second Mathematics Project: Plans and Performance)*, Technische Hogeschool Twente, Toegepaste Onderwijskunde, Enschede, 1983.
26. W. J. Pelgrum, Th. I. H. M. Eggen and Tj. Plomp, 'Evaluation of mathematics education on the national level', in B. P. M. Creemers (ed.), *Evaluation Research: Reflections and Studies*, SVO, The Hague, 1985.
27. A. Gifi, 'ANACOR', *UG-85-04*, Department of Data Theory, University of Leiden, 1985.
28. B. T. Smith, J. M. Boyle, J. J. Dongarra, B. S. Garbow, Y. Ikebe, V. C. Klema and C. B. Moler, *Matrix Eigensystem Routines — EISPACK Guide*, Springer Verlag, Berlin, 1976.
29. A. Boomsma, On the use of bootstrap and jackknife in covariance structure analysis', in F. De Antoni, N. Lauro and A. Rizzi (eds), *COMPSTAT Proceedings in Computational Statistics*, Physica-Verlag, Heidelberg, 1986.
30. R. Verdegaal, 'OVERALS', *UG-86-01*, Department of Data Theory, University of Leiden, 1986.