DOCUMENT RESUME

ED 310 116 TM 013 689

AUTHOR van der Burg, Eeke; de Leeuw, Jan
TITLE Use of the Multinomial Jackknife and Bootstrap in Generalized Nonlinear Canonical Correlation Analysis. Research Report 87-4.
INSTITUTION Twente Univ., Enschede (Netherlands). Dept. of Education.
PUB DATE 87
NOTE 50p.
AVAILABLE FROM Mediatheek, Faculteit Toegepaste Onderwijskunde, Universiteit Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Correlation; Elementary Secondary Education; *Error of Measurement; *Estimation (Mathematics); Mathematical Models; Monte Carlo Methods; Statistical Analysis
IDENTIFIERS *Bootstrap Methods; Eigenvalues; *Jackknifing Technique; Multinomial Models; Nonlinear Transformations; OVERALS Technique

ABSTRACT

The estimation of mean and standard errors of the eigenvalues and category quantifications in generalized non-linear canonical correlation analysis (OVERALS) is discussed. Starting points are the delta method equations. The jackknife and bootstrap methods are compared for providing finite difference approximations to the derivatives. Examining the basic properties of the jackknife method indicates that the vector of profile proportions is perturbed by leaving out single observations. The grid of perturbed values is used to estimate relevant derivatives. Bootstrapping means resampling with replacement from the original sample. Both procedures, bootstrapping and jackknifing, are used to compute pseudo-value means and standard errors for four different data sets: (1) the characteristics of 36 kinds of marine mammals; (2) data describing the attributes of 47 countries; (3) data from a study of school choice for 520 children leaving elementary school; and (4) a sample of 4,863 secondary students from the Second International Mathematics Study. For the small data sets the jackknife and bootstrap were used; for the larger sets, Monte Carlo versions of both were used. The jackknife method appeared less imprecise than the bootstrap method, and jackknife approximations were less stable for smaller samples. It is concluded that the bootstrap method performed better than did the jackknife method. For large samples, the bootstrap procedure works quite well for computing confidence intervals, and eigenvalues computed from OVERALS seem quite stable. Eight tables present the values from the four data sets, and four figures illustrate category quantifications for all data sets. (Author/SLD)

Use of the Multinomial Jackknife and Bootstrap in Generalized Nonlinear Canonical Correlation Analysis

Research Report 87-4

Eeke van der Burg
Jan de Leeuw

University of Twente
Division of Educational Measurement and Data Analysis

Use of the Multinomial Jackknife and Bootstrap in Generalized Nonlinear Canonical Correlation Analysis

Eeke van der Burg

Jan de Leeuw

ABSTRACT

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

Keywords: canonical correlation analysis, delta method, jackknife, bootstrap, confidence interval, nonlinear transformation.

INTRODUCTION

Nonlinear canonical correlation analysis with k sets of variables, OVERALS, is a multivariate technique in the sense of Gifi (1981, chap. 6). It is described by De Leeuw (1984a), Van der Burg, De Leeuw and Verdegaal (1984, 1986) and Verdegaal (1985). The k sets of variables are related in a linear way, as in ordinary canonical correlation analysis, but at the same time the variables are transformed nonlinearly. 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 (Van der Burg et al., 1984, 1986)

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 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. (De Leeuw,

1984a; De Leeuw, 1985; Tijssen, 1985; Van der Burg et al., 1986). 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.

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 p of proportions, based on a simple random sample of size n. Thus we suppose that p is a realization of the random variable $\underline{p}$, where $n\underline{p}$ has a multinomial distribution with parameters (n,π). We imbed the variable $\underline{p}$ in a sequence $\underline{p}_n$, where $n\underline{p}_n$ is also multinomial with parameters (n,π). For the expected value and the dispersion we have $E(\underline{p}_n)=\pi$ and $C(\underline{p}_n)=n^{-1}(\Pi-\pi\pi')$, where $\Pi=\text{diag}(\pi)$. We also have convergence in law to a normal distribution, in the

sense that $\underline{z}_n = n^{1/2}(\underline{p}_n - \pi) \rightarrow N(0,V)$, with V short for $\Pi - \pi\pi'$.

Now suppose ϕ is a real valued function defined for all p (or all p close to π), and twice continuously differentiable at π. Then the delta method (Rao, 1965, section 6.a.2, contains a nice discussion) states that $n^{1/2}(\phi(\underline{p}_n) - \phi(\pi)) \rightarrow N(0, g'Vg)$, where g is the vector of partials of ϕ at π. We can easily understand this result by writing $\underline{p}_n$ as $\underline{p}_n = \pi + n^{-1/2}\underline{z}_n$ and then developing a Taylor series for $\phi(\underline{p}_n)$ in $n^{-1/2}\underline{z}_n$ around π. This gives:

$$(1) \qquad \phi(\underline{p}_n) = \phi(\pi) + n^{-1/2} g'\underline{z}_n + (2n)^{-1}\underline{z}_n'H\underline{z}_n + o_p(n^{-1}).$$

A sequence of random variables $\underline{x}_n$ is $o_p(n^{-1})$ if $n\underline{x}_n$ converges in probability to zero. Matrix H contains the second order partials of ϕ at π. The variance of $\phi(\underline{p}_n)$ is given by

$$(2) \qquad \mathrm{Var}_\pi(\phi(\underline{p}_n)) = E_\pi((\phi(\underline{p}_n))^2) - (E_\pi(\phi(\underline{p}_n)))^2.$$

The second term of (2) is obtained by taking expectations of (1). As the random variable $\underline{z}_n$ converges in law to N(0,V) (the necessary conditions for $E_n(o_p(n^{-1})) = o(n^{-1})$ (cf. Serfling, 1980, section 1.4) are satisfied) we get

$$(3) \qquad E_\pi(\phi(\underline{p}_n)) = \phi(\pi) + (2n)^{-1}\mathrm{tr}HV + o(n^{-1}).$$

The squared expectation is

$$(4) \qquad (E_\pi(\phi(p_n)))^2 = (\phi(\pi))^2 + n^{-1}\phi(\pi)\mathrm{tr}HV + o(n^{-1}).$$

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

$$(5) \qquad E_\pi((\phi(p_n))^2) = (\phi(\pi))^2 + n^{-1}\phi(\pi)\mathrm{tr}HV + n^{-1}g'Vg + o(n^{-1}).$$

Subtraction of (5) and (4) gives

$$(6) \qquad \mathrm{Var}_\pi(\phi(p_n)) = n^{-1}g'Vg + o(n^{-1}),$$

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 $n^{-1}\hat{g}\hat{V}\hat{g}$ where $\hat{g}$ estimates g and $\hat{V}$ estimates V. Usually $\hat{V}=V_p=P-pp'$ with $P=\mathrm{diag}(p)$, and $\hat{g}=g_p$ the partials of ϕ at p. If we have an estimate $\hat{\sigma}$ of the standard error, then we also have an approximate confidence interval (95%) of the form $(\phi(p_n)-1.96\hat{\sigma},\ \phi(p_n)+1.96\hat{\sigma})$.

We can also evaluate the bias of $\phi(p_n)$ as an estimate of $\phi(\pi)$. To do so, expression (3) is used:

$$(7) \qquad E_\pi(\phi(p_n)-\phi(\pi)) = (2n)^{-1}\mathrm{tr}HV + o(n^{-1}),$$

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

In many cases, and for instance in generalized canonical corre-

lation 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 (Gifi, 1981, chap. 12; De Leeuw, 1984b; Schriever, 1985) the delta method can be applied, but in other cases it simply is not feasible. In some such cases we can use a resampling method such as bootstrap and jackknife (Efron (1982) gives an overview), which we view here as methods to approximate the relevant partials. Gifi (1981, chap. 13) concentrates on the bootstrap. In this paper we use both jackknife and bootstrap and compare the results.

THE MULTINOMIAL JACKKNIFE

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

$$p(k) = p + (n-1)^{-1}(p-e_k). \tag{8}$$

with e_k the k^{th} unit vector. The jackknife value for the observation is $\phi(p(k))$, and the pseudo value $\phi_k(p)$ is defined as

$$\phi_k(p) = n\phi(p) - (n-1)\phi(p(k)). \tag{9}$$

The jackknife method uses the average pseudo value $\sum p_k \phi_k(p)$ as an estimate of $\phi(\pi)$ (p_k are the elements of vector p), and uses the variance of the pseudo values as an estimate of g'Vg. We shall explain why this is a reasonable procedure. Observe for the moment that no derivatives need to be computed.

Write ϕ_S for the average of the pseudo values, and σ_S^2 for their variance. Then combination of (8) and (9) using a Taylor series for $\phi(p(k))$ in $(p-e_k)$ around p gives

$$(10) \qquad \phi_k(p) = \phi(p) - g_p'(p-e_k) - (2n-2)^{-1}(p-e_k)'H_p(p-e_k) + o((n)^{-1}).$$

Thus

$$(11) \qquad \phi_S(p) = \phi(p) - (2n-2)^{-1}trV_pH_p + o((n)^{-1}).$$

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

$$(12) \qquad (n-1)(E_\pi(\phi_S(p_n)) - \phi(\pi)) \to 0.$$

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

To compute σ_S^2 we observe that

$$(13) \qquad \phi_k(p) - \phi_S(p) = -g'(p-e_k) + o(1).$$

From (13) it follows that

$$\sigma_{\S}^2(p) \rightarrow g_p' V_p g_p. \tag{14}$$

Thus $\sigma_{\S}^2/n$ is asymptotically equal to the delta method variance estimate. More precise expansions can be found in De Leeuw (1985).

The only computations that are required is that the analysis technique is repeated for each observation that is successively dropped. For each jackknife sample a pseudo value 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 JACKKNIFE

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 pseudo values. Instead we can estimate the average pseudo value and the variance of the pseudo values 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

jackknife.

It must be remarked that the delta method provides us with an approximation to the standard error. The jackknife in our interpretation, gives an approximation to the delta method approximation of the standard error. And the Monte Carlo method approximates the jackknife 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 jackknife 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 nonstochastic context.

THE MULTINOMIAL BOOTSTRAP

If we look at the basic properties of the jackknife, 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 reguarly 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 gets the same weight, i.e. occurs once in the distribution of all possible jackknife sample f we use the corresponding perturbations of the profile proportions we need weights $p_1,..,p_m$ 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 (Efron, 1982). 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 centered around the sample value p, and that t. have weights according to their similarity with the sample value. Suppose

$$(15) \qquad q = (n_1/n, \ldots, n_m/n)$$

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

$$(16) \qquad w_q(p) = \frac{n!}{n_1! \ldots\ldots n_m!} p_1^{n_1} \ldots\ldots p_m^{n_m} .$$

The bootstrap pseudo values are defined as

$$(17) \qquad \phi_q(p) = 2\phi(p) - \phi(q).$$

The average bootstrap pseudo value is

(18) $\phi_{\S}(p) = \sum_q w_q(p)\phi_q(p),$

and the bootstrap variance is

(19) $\sigma^2_{\S}(n) = \sum_q w_q(p)(\phi_q(p) - \phi_s(p))^2.$

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

(20) $\phi_q(p) = \phi(p) - n^{-1/2}g_p'z - (2n)^{-1}z'H_pz + o(n^{-1}).$

Thus the average pseudo value is

(21) $\phi_{\S}(p) = \sum w_q(p)\phi_q(p) = \phi(p) - (2n)^{-1}trH_pV_p + o(n^{-1}),$

as $\sum w_q(p)z=0$ and $\sum w_q(p)zz'=V_p$. Combining (21) with (7) shows that the expected value of $\phi_{\S}(P_n)$ is

(22) $E_\pi\phi_{\S}(P_n) = E_\pi(\phi(P_n) - (2n)^{-1}trHV + o_p(n^{-1})).$

Combining (22) with (3) gives

(23) $n(E_\pi\phi_{\S}(P_n) - \phi(\pi)) \to 0.$

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

$$(24) \qquad \phi_q(p) - \phi_{\S}(p) = -n^{-1/2} g_p' z - (2n)^{-1} z' H p z + (2n)^{-1} \mathrm{tr} H_p V_p + o(n^{-1}).$$

Then the pseudo value variance is

$$(25) \qquad \sigma^2_{\S}(p) = n^{-1} g_p' V_p g_p + o(n^{-1}).$$

The expected value of $\sigma^2_{\S}(p_n)$ converges to the delta method variance

$$(26) \qquad n(E_\pi \sigma^2_{\S}(p_n)) \to g'Vg.$$

Thus the bootstrap variance is an asymptotically unbiased estimator of $n^{-1}g'Vg$. This result means that the variance of the bootstrap values (and the pseudo values) estimates the delta method variance. The jackknife pseudo value variance estimates g'Vg, so that the estimates of bootstrap and jackknife differ in a factor n.

For large n a random version of the jackknife 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 is done on what is large enough. Only Borsboom & Van Pelt (unpublished) did some research on this subject with regard to a computer program for nonlinear canonical correlation analysis (CANALS, cf. Van der Burg & De Leeuw, 1983). 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 .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 is implemented in a computer program by Borsboom & Visser (1987). We repeated the analysis of Borsboom & 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 jackknife and bootstrap procedures are used to compute pseudo value means and standard errors for four different data sets. For the smaller data sets (i.e. Whales and Russett) a complete jackknife and a random bootstrap procedure were applied, while in the case of the FYTY and SIMS data sets a Monte Carlo version of jackknife and bootstrap were used. For all data sets the generalized canonical correlations (i.e. eigenvalues) were employed. In addition for the Russett and 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 (respectivily category quantifications), computed from the original data matrix (or sample), in single precision (SP) and double precision (DP). For single precision the convergence of the

OVERALS program is computed with accuracy to 10^{-5}, while 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 jackknife (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, Vescia, 1985) consists of fifteen variables decribing 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 (1985). 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 considered was based on the Russett data set (Russett, 1964). These data contain three sets of va iables dealing with attributes of 47 countries. The first set contains two variables concerning ownership of land, the second set contains the variables gross national product and percentage of people working in agriculture, while the third set includes the four variables that are indicators of political instability. The multichotomized scores originally established by Gifi (1981, chap. 7) were used in

this study. The data were analyzed with numerical, single nominal, and multiple nominal options.

The third example (designated FYTY for From Year to Year) is based on a sample from a large school career survey (for references see De Leeuw & Stoop, 1979). In the example we consider, there are 520 school children 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 time period, the order considered was: subset 3, subset 2, subset 1. The FYTY data were analysed in the same three ways used with the Russett data.

The fourth example also comes from a school survey, in this case the SIMS (Second International Mathematics Study) Project (Pelgrum, Eggen & Plomp, 1984 and 1986). For this example the complete data base (4863 school children) as well as a sample from that base (1000 school children) were considered. The eigenvalues corresponding to the complete data base were treated as population parameters, while the sample data were used to estimate these population parameters. The variables considered in this example were divided into three subsets. One subset contains the variables type of school and father's education, another subset contains three attitudes towards mathematics, while 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 1 contains bootstrap estimates for the Whales example. As earlier specified, the bootstrap procedure was implemented with 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 are presented in Table 1. It may be noted that the bootstrap pseudo value means show only minor differences, even for the extreme cases (i.e. number of samples of 10 and 120). However, the estimated standard error tended to increase for number of samples 10 to 40, while remaining rather stable for number of samples greater than 40. In addition, it was found that differences were negligible between the 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.

INSERT TABLE 1 ABOUT HERE

The eigenvalues computed from the Whales data are presented in

Table 2. 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, Gifi, 1985) was used. The eigenvalues from ANACOR

INSERT TABLE 2 ABOUT HERE

---

are precisely the same as the OVERALS eigenvalues. However the variances do differ (see Table 3). Although the APL eigenvalues are slightly different from those obtained under the above mentioned approaches, those differences hardly seem worth mentioning. The jackknife and bootstrap results are presented in column 4 to 11 of Table 2. The bootstrap results are based on 120 samples, while the JSP and JDP results are based on 36 samples (a complete jackknife). The estimated means do not substantially differ from each other, but the estimated standard errors do manifest sizable differences. The three jackknife estimates are very similar, however the boot-

strap estimates are considerably lower in magnitude than the jackknife estimates. It is clear from Table 1 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 jackknife 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 (Table 3). The first two bootstrap estimates are sytematically lower than the CA values, while the corresponding jackknife estimates are higher. Only the third variance is similar for the bootstrap and jackknife (and lower than

---

INSERT TABLE 3 ABOUT HERE

---

for CA). Assuming that CA gives the more precise value (direct computation instead of approximation), we find that the jackknife overestimates and the bootstrap underestimates the standard error. We do not make conclusions on the third variance as the smallest eigenvalue normally is much less precise than the larger ones.

RESULTS FOR RUSSETT

A complete jackknife and a random bootstrap were applied to the Russett data. As the number of observations (47) is still rather small we may expect different results from the bootstrap and jackknife analyses. The results for the eigenvalues are presented in Table 4. One of the most striking aspects of these results is that the standard errors are considerably larger than those obtained with the Whales data. It should be realized however, that homogeneity analysis is known to be a very stable technique, and most likely an example with more variables in the sets will be less stable. JSP means appear to be rather strange, especially if we keep in mind that the data are not a sample but the population. JDP estimates seem to be a little less unexpected. Estimation by bootstrap is more like the eigenvalues based on the complete data set than the estimation by jackknife. But in any case the results appear discouraging. Bootstrap standard errors are smaller in general than jackknife standard errors but they are too large for reasonable confidence intervals.

INSERT TABLE 4 ABOUT HERE

A look at results for category quantifications of two variables (LABO and DEMO) is even more perplexing (Table 5). The bootstrap

estimates provide at least a rough approximation of the corresponding population results. However the jackknife estimates are quite different from the population values. In addition, the corresponding standard errors seem extrordinarily large. A direct look at the bootstrap and jackknife results (without computing pseudo values) shows their values to be rather stable (Figures 1 and 2). Every dash (-) corresponds to a category quantification resulting from one bootstrap/jackknife sample (called bootstrap/jackknife sample value), while the symbol 0 represents the original category quantifications of the Russett data. On the horizontal axis the original category scores are scaled, while on the vertical axis the category quantifications are scaled. As the jackknife and bootstrap

---

INSERT FIGURES 1 AND 2 ABOUT HERE

---

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 comparable for the two techniques. It may be noted that DEMO is a much more stable variable than is LABO. For the latter variable, cateories 1 & 2, and 4 & 5 respectivily show considerable overlap. However, overall the results appear rather stable.

Because the results for the Whales and the Russett examples were disappointing, additional examples involving larger data sets were

considered.

RESULTS FOR FYTY

The FYTY data were analyzed by the Monte Carlo version of jackknife and bootstrap. The resulting eigenvalue estimates are presented in Table 6. Considering the FYTY data as a sample from a large data base, we can expect eigenvalues obtained from the original sample to be larger than estimates obtained from bootstrap and jackknife. This is because sample estimates tend to overestimate the population parameters while the jackknife and bootstrap pseudo value 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 jackknife estimates were rather similar to one other but do not seem to show any relation with the sample eigenvalues. Here the JSP estimation seems to have outperformed the JDP estimation, but that is not really true. The JSP values have been produced with three digits accuracy, which was not sufficiently precise. For instance the second estimate of the population mean (multiple nominal), seems very good for JSP, and somewhat worse for JDP. This difference is due to the number of digits provided in the computer output. As all jackknife 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 jackknife estimates are the same. The JDP estimate, in

which 6 digits accuracy is provided, is much larger, but must be more precise. Thus in this case the JSP results may be somewhat suspect. For this reason we look only at the JDP results. For the following example we used output with 6 digits accuracy to avoid this problem.

INSERT TABLE 6 ABOUT HERE

The bootstrap estimates of the population mean appear better than the jackknife 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 smaller ones are also computed). This phenomenon may provide an explanation for the difference in accuracy obtained for first and second eigenvalues.

Due to the large number of observations standard errors are smaller than found for the Russett example. However especially in the case of jackknife these standard errors are still unacceptably large.

Untill now it would seem that the bootstrap method is preferable to the jackknife method, since it apprears 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 warrented. For this reason the next example is based on 1000 observations.

INSERT TABLE 7 ABOUT HERE

However, before this next example is presented some category quantifications of the FYTY data are considered. To do this, the variables PRE and TON are used (Table 7). 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 6). As category quantifications are proportional to the square root of eigenvalues, their standard errors will increase at a corresponding rate. In Table 7 we see that again the JSP and JDP results agree rather well, but the correspondence between sample eigenvalues and the estimated population means is almost nonexistent for the jackknife. Bootstrap results are very similar to the sample values for TON, however results for PRE show less similarity (but not as bad as jackknife estimates). Looking at the category quantifications of the jackknife and bootstrap samples directly (like we did for the Russet example), we find that those variables behave very stably (Figures 3 and 4). In the case of TON all fourty dashes (jackknife/bootstrap sample values) fall together both for

INSERT FIGURES 3 AND 4 ABOUT HERE

jackknife and bootstrap, while for PRE we find some variability amoung the category quantifications. Note that in 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.

RESULTS FOR SIMS

In Table 8 we find results for the eigenvalues of the SIMS data. The data base 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 jackknife approaches were used (on the sample) for reestimation of the population values (with the computer results specified to 6 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 jackknife is used. However, this did not occur in half the cases of Table 8. Since standard errors appear to be reasonably small, it is suggested that confidence intervals for bootstrap results be computed.

DISCUSSION

In comparing jackknife with bootstrap one is left with the impression that the jackknife method is far more imprecise than the bootstrap method. Apparently jackknife approximations are less stable for smaller samples. The use of double precision computations tends to improve the jackknife 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 jackknife 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 jackknife and bootstrap method converge to different values.

Jackknife 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. Estimation based on bootstrap is more robust than estimation based on jackknife. Results for category quantifications are discussed further on.

For smaller samples the estimated standard errors of eigenvalues (both for jackknife and bootstrap) are unacceptably large (varying from .025 to 1.05). 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 jackknife than for bootstrap, as bootstrap estimates are nearer to the population or sample value, than jackknife estimates are. Even for the largest sample considered, more deviations (i.e. population mean not in 95% confidence interval) were found with jackknife than with bootstrap. Deviations occur for second eigenvalues (ordinal and numerical) of JDP and for all second eigenvalues of JSP (Table 8). Many eigenvalue routines are less precise for the smallest computed eigenvalue. As jackknife is sensitive to precision this may be the cause.

Estimates for category quantifications seem rather strange. Category quantifications correspond to the square root of eigenvalues, and thus will be less precise. However the estimates are sometimes highly inaccurate, even for very stable jackknife or bootstrap sample values. Again in this case the bootstrap seems to perform much better than the jackknife (Table 7). In several plots stability of results is illustrated (Figures 1 to 4). Only in the case of LABO are the results unstable. This variable has a much smaller weight than the other variables used for illustration. It may be that this smaller weight results in a lower level of stability.

In the case of the example 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 pseudo value means which are presented in Table 8. A similar outcome was found (although the differences were less) for jackknife.

Using pseudo values tends to provide a better approximation than obtained by averaging bootstrap and jackknife results directly.

If we compare other studies using jackknife and bootstrap, we find that Boomsma (1986) 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 jackknife 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. jackknife and bootstrap sample values) leads to more positive conclusions which are more in agreement with other's earlier experience with the examples here considered (cf De Leeuw & Stoop, 1979; Gifi, 1981, chap.7).

ÆFERENCES

Borsboom, G.J.J.M. and Visser, R.A. (1987). Interpreting analysis of qualitative multivariate data using Bootstrap methods. Contribution to the ZUMA conference, march, 1987.

Boomsma, A. (1986). On the Use of Bootstrap and Jackknife in Covariance Structure Analysis. In COMPSTAT Proceedings in Computational Statistics. F. De Antoni, N. Lauro, and A. Rizzi (Eds.). Heidelberg: Physica-Verlag.

De Leeuw, J. (1984a). The Gifi-system of nonlinear multivariate analysis. In: Data Analysis and Informatics. E. Diday et al. (eds.). Amsterdam: North Holland, 231-243.

De Leeuw, J. (1984b). Statistical Properties of Multiple Correspondence Analysis. Department of Data Theory, University of Leiden, RR-85-06.

De Leeuw, J. (1985). Jackknife and Bootstrap in Multinomial Situations. Department of Data Theory, University of Leiden, RR-85-16

De Leeuw, J. & Stoop, I. (1979). 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: Van Achteren naar Voren. J.L. Peschar (ed.), s'Gravenhage: Staatsuitgeverij.

Efron, B. (1982). The Jackknife, the Bootstrap, and other Resampling Plans. CBMS-NSF Regional Conference Series in Applied Mathematics, no. 38, Philadelphia: SIAM.

Gifi, A. (1981). Nonlinear Multivariate Analysis. Department of Data Theory, University of Leiden. Reissued: Leiden: DSWO Press, 1987

Gifi, A. (1985). ANACOR. Department of Data Theory, University of Leiden, UG-85-04.

Pelgrum, W.J. and Eggen, Th.I.H.M. (1983). Tweede Wiskunde Projeckt: Opzet en Uitvoering (Second Mathematics Project: Plans and Performance). Enschede: Universiteit Twente, Toegepaste Onderwijskunde.

Pelgrum, W.J., Eggen, Th.I.H.M., and Plomp, Tj. (1985). Evaluation of mathematics education on the national level. In: B.P.M. Creemers (ed.). Evaluation Research: Reflections and Studies. The Hague: SVO.

Rao, C.R. (1965)). Linear Statistical Inference and its Applications. New York: Wiley.

Russett, B.M. (1964). Inequality and instability: the relation of land tenure to politics. World Politics, 21, 442-454

Serfling, R.J. (1980) Approximation Theorems of Mathematical Statistics. New York: Wiley.

Schriever, B.M. (1985). Order Dependence. Doctoral Dissertation, Free University of Amsterdam.

Tijssen, R. (1985). A new approach to non-linear canonical correlation analysis. Department of Psychology, University of Leiden, PRM-85-01.

Van der Burg. E. (1985). HOMALS classification of whales, porpoises and dolphins. In: Data Analysis in Real Life Environment: Ins and Outs of Solving Problems. J.F. Macrotorchino, J.M. Proth and J. Janssen (eds.). Amsterdam: North Holland, 25-36.

Van der Burg, E. and De Leeuw, J. (1983). Non-linear Canonical Correlation. British Journal of Mathematical and Statistical Psychology, 36, 54-80.

Van der Burg, E. De Leeuw , J. and Verdegaal, R. (1984). Non-linear canonical correlation with M sets of variables. Department of Data Theory University of Leiden, RR-84-12

Van der Burg , E. De Leeuw, J. and Verdegaal, R. (1986). Homogeneity analysis with k sets of variables: An alternating least squares method with optimal scaling features. Department of Education, University of Twente, RR-86-5. To appear in Psychometrika.

Verdegaal, R. (1985). Meer-sets analyse voor kwalitatieve gegevens (Multi-set analysis for qualitative data). Department of Data Theory, University of Leiden, RR-85-14.

Verdegaal, R. (1986). OVERALS. Department of Data Theory, University of Leiden, User's Guide UG-86-01.

Vescia, G. (1985). Automatic classification of cetaceans by similarity aggregation. In: Data Analysis of Real Life Environment: Ins and Outs of Solving problems. J.F. Macrotorchino, J.M. Proth & J. Janssen (eds.). Amsterdam: North Holland, 15-24.

AUTHORS' NOTE

The addresses of the authors are: Eeke van der Burg, University of Twente, Department of Education, P.O. 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.

TABLE 1

WHALES, multiple nominal. Estimated population means and estimated standard errors for bootstraps with different sample sizes. SP=single precision, DP=double precision.

Whales Bootstr	pop.mean SP	pop.mean DP	st.error SP	st.error DP	Bootstr	pop.mean SP	pop.mean DP	st.error SP	st.error DP
10	.601	.601	.031	.031	70	.608	.608	.042	.042
	.397	.398	.058	.058		.395	.396	.041	.041
	.306	.305	.022	.017		.298	.299	.026	.025
20	.602	.602	.029	.029	80	.608	.607	.043	.043
	.393	.393	.049	.049		.394	.394	.044	.044
	.308	.308	.020	.018		.297	.298	.025	.025
30	.605	.605	.032	.032	90	.610		.045	
	.395	.396	.046	.046		.394		.046	
	.302	.303	.024	.023		.297		.025	
40	.606	.606	.041	.041	100	.612		.047	
	.391	.392	.043	.043		.394		.044	
	.298	.299	.027	.026		.296		.025	
50	.609	.609	.040	.040	110	.612		.046	
	.391	.392	.042	.042		.393		.043	
	.299	.300	.027	.026		.295		.025	
60	.610	.610	.039	.039	120	.613		.045	
	.394	.395	.041	.041		.393		.042	
	.298	.299	.027	.026		.295		.026	

TABLE 2

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

Whales	eigenvalue		population mean				standard error			
	DP[1]	APL	JSP	JDP	JAPL	BSP	JSP	JDP	JAPL	BSP
MULT	.635	.637	.618	.611	.614	.613	.056	.056	.056	.045
NOM	.413	.415	.385	.398	.400	.393	.060	.060	.061	.042
	.317	.320	.279	.282	.285	.295	.028	.028	.026	.026

[1] Eigenvalues SP=DP=CA (Correspondence Analysis)

TABLE 3

WHALES, estimated variances for jackknife (36 samples), bootstraps (120 samples), and correspondence analysis. SP=single precision, DP=double precision.

Whales	JSP	JDP	BSP	CA
estimated variances	.0032	.0031	.0020	.0025
	.0037	.0036	.0018	.0031
	.0008	.0008	.0007	.0012

TABLE 4

RUSSETT, three measurement levels. Eigenvalues, estimated population means and estimated standard errors. SP=single precision, DP=double precision, J=jackknife (47 samples), B=bootstrap (40 samples).

Russett	eigenvalue		population mean			stand. error		
	SP	DP	JSP	JDP	BSP	JSP	JDP	BSP
single nominal	.765	.770	.420	.703	.673	.065	.058	.032
	.710	.706	.924	.616	.638	.105	.089	.034
numerical	.687	.687	.674	.556	.660	.057	.057	.048
	.461	.463	.479	.412	.400	.095	.076	.047
multiple nominal	.815	.815	.710	.690	.732	.049	.049	.029
	.736	.734	.450	.443	.625	.053	.054	.025

TABLE 5

RUSSETT, single nominal. Frequencies, category quantifications, estimated population means and estimated standard errors for two variables. SP=single precision, DP=double precision, J=jackknife (47 samples), B=bootstrap (40 samples).

Russett freq.	cat.quant. SP	DP	population mean JSP	JDP	BSP	standard error JSP	JDP	BSP
LABO								
10	-1.213	-1.157	-3.226	-.556	-1.240	.969	.835	.699
11	-.970	-1.019	.229	-1.969	-1.191	.931	.805	.476
14	.622	.606	1.120	.465	.662	.493	.445	.455
11	1.176	1.200	1.028	1.831	1.410	.706	.587	.392
DEMO								
1	1.147	1.106	1.434[1]	0.5551	1.0822	1.3591	.9941	.677[2]
19	-1.090	-1.085	-.674	-.480	-1.049	.322	.294	.305
12	-.400	-.410	-1.272	-1.610	-.553	.614	.529	.544
19	1.171	1.173	1.406	1.448	1.284	.215	.207	.261

[1] 519 observed values
[2] 511 observed values

TABLE 6

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

FYTY	eigenvalue		population mean			stand. error		
	SP	DP	JSP	JDP	BSP	JSP	JDP	BSP
single	.735	.734	1.202	1.060	.731	.061	.056	.021
nominal	.399	.400	.269	.167	.367	.034	.025	.017
numerical	.695	.695	.851	.679	.692	.016	.014	.017
	.341	.341	.639	.567	.326	.057	.054	.013
multiple	.742	.742	.690	.695	.729	.014	.012	.018
nominal	.555	.555	.555	.626	.535	.019	.026	.016

TABLE 7

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

FYTY freq.	cat.quant SP	DP	population mean JSP	JDP	BSP	standard error JSP	JDP	BSP
TON								
42	-.586	-.586	-2.052	-2.065	-.584	.175	.173	.128
208	-.853	-.853	-1.774	-1.748	-.851	.077	.075	.087
165	.056	.056	2.119	2.132	.080	.125	.122	.146
105	1.853	1.853	.317	.291	1.867	.125	.124	.125
PRE								
49	-.622	-.622	-43.102	-40.287	-.617	6.574	5.641	.581
147	-.592	-.592	-23.000	-20.275	-.694	4.157	3.696	.457
181	-.184	-.184	8.989	7.886	-.126	2.587	2.143	.295
111	.298	.298	33.994	32.606	.200	4.147	3.598	.676
32	3.680	3.680	5.795	-.212	4.275	3.022	2.350	.811

TABLE 8

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=jackknife, B=bootstrap (both 40 samples).

SIMS	eigenvalue		population mean			stand. error		
	POP[1]	SAM[1]	JSP	JDP	BSP[2]		JDP	BSP
multiple nominal	.615	.625	.615	.620	.619	.011	.011	.011
	.450	.463	.566	.463	.454	.013	.013	.012
single nominal	.613	.622	.621	.622	.616	.014	.013	.012
	.352	.377	.171	.377	.371	.015	.013	.010
single ordinal	.613	.622	.728	.628	.619	.014	.014	.012
	.352	.377	.602	.697	.375	.094	.092	.012
numerical	.506	.615	.609	.608	.613	.012	.012	.012
	.350	.363	.277	.321	.359	.010	.009	.011

1 No difference between eigenvalues SP and DP, except for the first eigenvalue of single nominal (SP,SAM): .621.

2 BDP with convergence to 10^{-5} gives the same results.

FIGURE CAPTIONS

Figure 1. RUSSETT. Category quantifications for LABO based on three methods: JSP, JDP and BSP (O orginal sample value, - jackknife/ bootstrap sample value).

Figure 2. RUSSETT. Category quantifications for DEMO based on three methods: JSP, JDP and BSP (O original sample value, - jackknife/ bootstrap sample value).

Figure 3. FYTY. Category quantifications for TON based on three methods: JSP, JDP and BSP (O original sample value, - jackknife/ bootstrap sample value).

Figure 4. FYTY. Category quantifications for TON based on three methods: JSP, JDP and BSP (O original sample value, - jackknife/ bootstrap sample value).

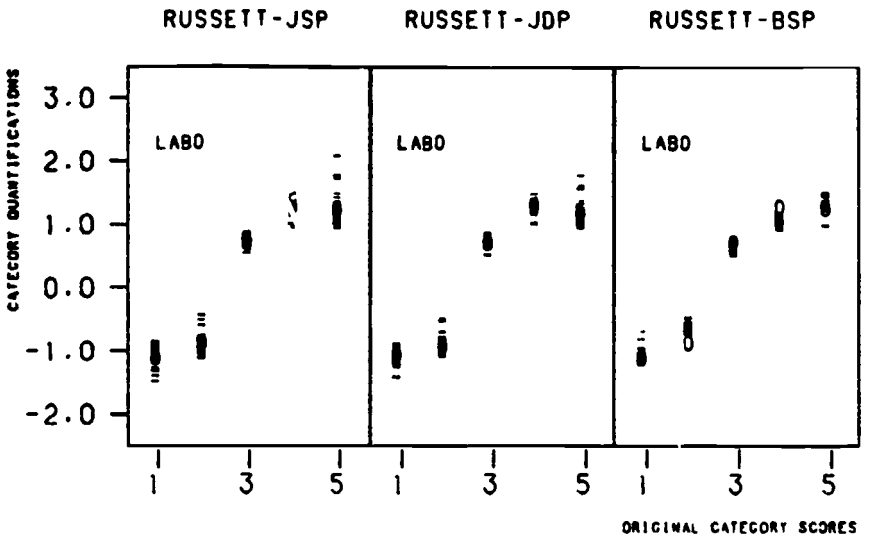

RUSSETT-JSP
RUSSETT-JDP
RUSSETT-BSP
CATEGORY QUANTIFICATIONS
3.0
2.0
1.0
0.0
-1.0
-2.0
LABO
LABO
LABO
1
3
5
1
3
5
1
3
5
ORIGINAL CATEGORY SCORES

FIGURE 1

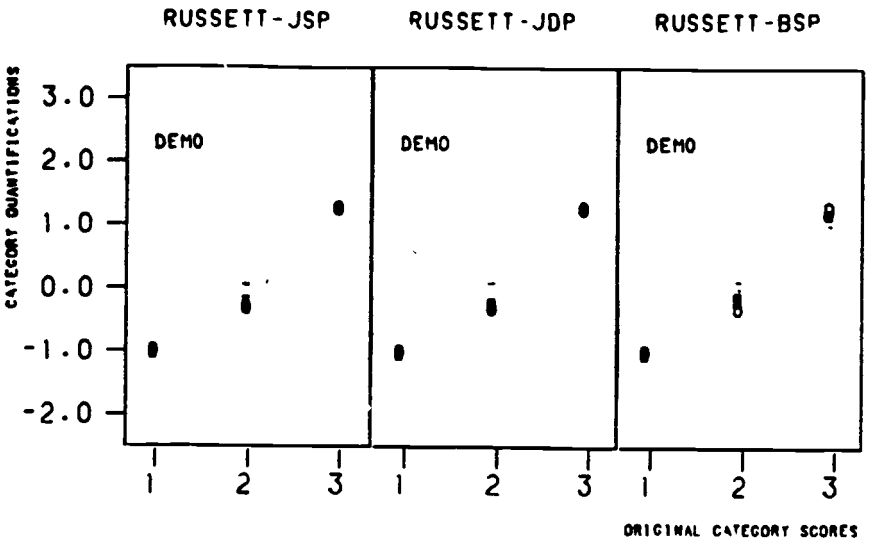

RUSSETT-JSP
RUSSETT-JDP
RUSSETT-BSP
CATEGORY QUANTIFICATIONS
3.0
2.0
1.0
0.0
-1.0
-2.0
DEMO
DEMO
DEMO
1
2
3
1
2
3
1
2
3
ORIGINAL CATEGORY SCORES

FIGURE 2

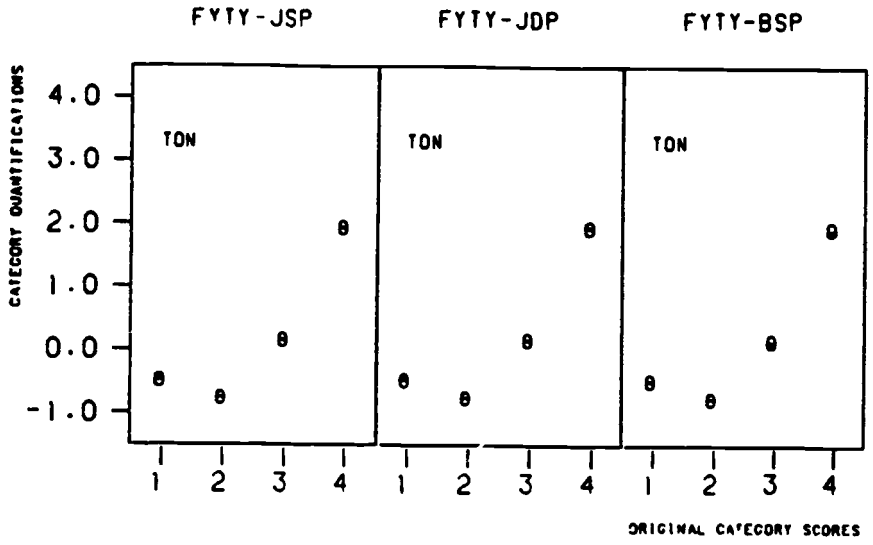

FYTY-JSP
FYTY-JDP
FYTY-BSP
CATEGORY QUANTIFICATIONS
4.0
3.0
2.0
1.0
0.0
-1.0
TON
TON
TON
1 2 3 4 1 2 3 4 1 2 3 4
ORIGINAL CATEGORY SCORES

FIGURE 3

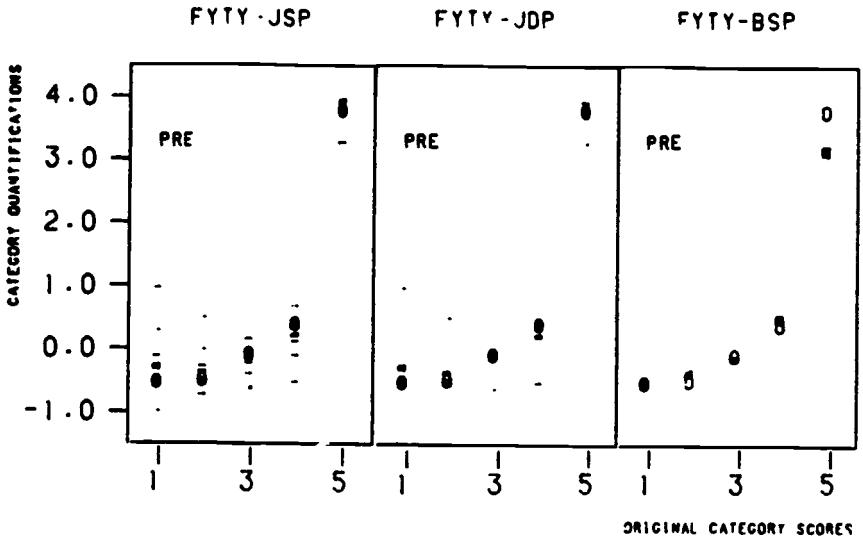

FYTY-JSP
FYTY-JDP
FYTY-BSP
CATEGORY QUANTIFICATIONS
4.0
3.0
2.0
1.0
0.0
-1.0
PRE
PRE
PRE
1
3
5
1
3
5
1
3
5
ORIGINAL CATEGORY SCORES

FIGURE 4

Titles of Recent Research Reports

RR-86-1 W.J. van der Linden, The use of test scores for classification decisions with threshold utility

RR-86-2 H. Kelderman, Item bias detection using the loglinear Rasch model: Observed and unobserved subgroups

RR-86-3 E. Boekkooi-Timminga, Simultaneous test construction by zero-one programming

RR-86-4 W.J. van der Linden, & E. Boekkooi-Timminga, A zero-one programming approach to Gulliksen's matched random subtests method

RR-86-5 E. van der Burg, J. de Leeuw, & R. Verdegaal, Homogeneity analysis with k sets of variables: An alternating least squares method with optimal scaling features

RR-86-6 W.J. van der Linden, & T.J.H.M. Eggen, An empirical Bayes approach to item banking

RR-86-7 E. Boekkooi-Timminga, Algorithms for the construction of parallel tests by zero-one programming

RR-86-8 T.J.H.M. Eggen, & W.J. van der Linden, The use of models for paired comparisons with ties

RR-86-9 H. Kelderman, Common item equating using the loglinear Rasch model

RR-86-10 W.J. van der Linden, & M.A. Zwarts, Some procedures for computerized ability testing

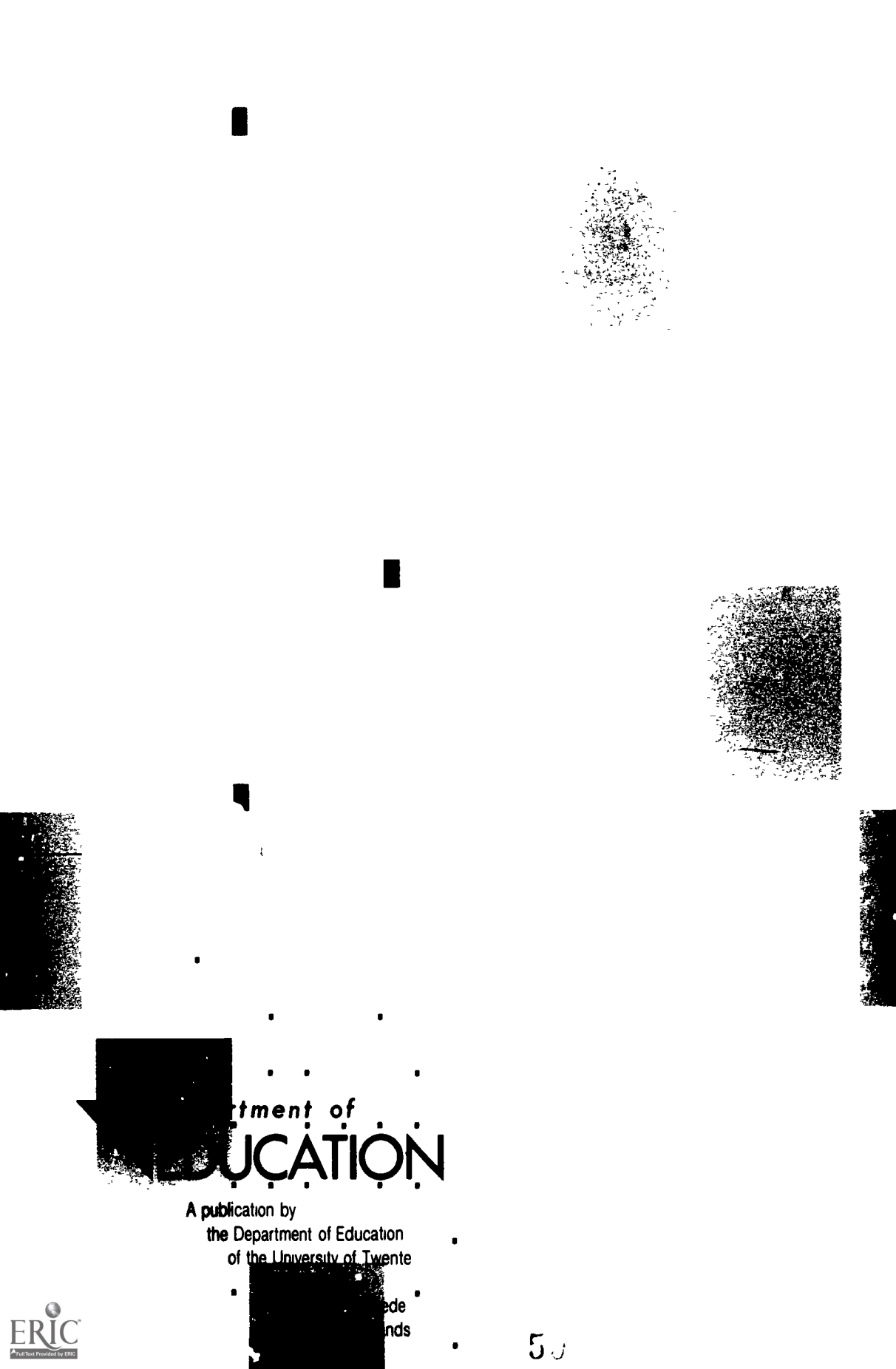
RR-87-1 R. Engelen, Semiparametric estimation in the Rasch model

RR-87-2 W.J. van der Linden (Ed.), IRT-based test construction

RR-87-3 R. Engelen, P. Thommassen & W. Vervaat, Ignatov's theorem: A new and short proof

RR-87-4 E. van der Burg & J. de Leeuw, Use of the multinomial jackknife and bootstrap in generalized nonlinear canonical correlation analysis

Research Reports can be obtained at costs from Mediatheek, Faculteit Toegepaste Onderwijskunde, Universiteit Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.



tment of

UCATION

A publication by
the Department of Education
of the University of Twente

ede
nds