DOCUMENT RESUME

ED 095 217 TM 003 894

AUTHOR	Pandey, Tej N.; Hubert, Lawrence J.
TITLE	A Comparison of Interval Estimation of Coefficient Alpha Using the Feldt and the Jackknife Procedures.
PUB DATE	[Apr 74]
NOTE	13p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974)
EDRS PRICE	MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS	*Comparative Analysis; Item Sampling; *Statistical Analysis; Statistics; *Test Reliability
IDENTIFIERS	Coefficient Alpha

ABSTRACT

This investigation had two major purposes. The first was to explore the use of an inferential technique called Tukey's Jackknife in establishing a confidence interval about cooefficient alpha reliability. The second purpose was to study the robustness of the Feldt and the jackknife procedures when the data fails to satisfy usual normality assumptions. Using a linear model of test score, computer simulated data representing a matched item-examinee sample from their respective populations was employed. The two jackknife procedures, is this situation, were less conservative. For log-normal distributions of examinee score components, Feldt's procedure gave reasonable interval estimates for low values (=.60) of population alpha, but for higher values of population alpha (=.90), the Feldt procedure was not found to be robust. The two Jackknife procedures, on the other hand, were found to be relatively robust over the entire range of investigation of values of alpha. None of the procedures was found useful in interval estimation of coefficient alpha when the distribution of examinee score components were of extreme double-exponential form. (RC)

A COMPARISON OF INTERVAL ESTIMATION OF COEFFICIENT ALPHA USING THE FELDT AND THE JACKKNIFE PROCEDURES

by

Tej N. Pandey
California State Department of Education

and

Lawrence J. Hubert
University of Wisconsin, Madison

Paper Read At
American Educational Research Association, 1974
Chicago

A COMPARISON OF INTERVAL ESTIMATION OF COEFFICIENT ALPHA USING THE FELDT AND THE JACKKNIFE PROCEDURES

by
Tej N. Pandey[1] and Lawrence J. Hubert
University of Wisconsin, Madison

INTRODUCTION

The calculated reliability of a test based on a small sample of subjects is an estimate of the population reliability, and hence, is subject to sampling fluctuation. Until quite recently little application has been made of statistical inference techniques to the reliability coefficient. Using analysis of variance framework, Ebel (1951) and Jackson and Ferguson (1941) made the first attempts to relate reliability coefficient estimates to the well known F-distribution. However, Kristoff (1963, 1970) and Feldt (1965) presented complete sampling theory of reliability estimates and also methods to apply it.

Kristoff derived the sampling distribution of a maximum likelihood estimate of the population value of alpha using the technique of transformation of variables. The sampling theory derived by Feldt, which is also of concern here, was based on normality assumptions regarding the true and error score distributions. To be more precise, consider the following components of variance model employed by Feldt,

$$x_{ij} = \mu + t_i + a_j + e_{ij}; \quad i = 1,\ldots,n; \; j = 1,\ldots,k,$$

and μ is some overall constant, $t_i \sim N(o, \sigma t^2)$, $[a_j \sim N(o, \sigma a^2)]$, $e \sim N(o, \sigma e^2)$, and the n+k+nk random variables $\{t_i\}$, $\{a_j\}$, $\{e_{ij}\}$ are mutually independent. Under these constraints the usual F-ratio provides a suitable test of $Ho: \sigma t^2 = \sigma o^2$ against $Hi: \sigma t^2 \neq \sigma o^2$; however, whenever t_i has a positive kurtosis, Scheffé (1959) has shown that the true significance level of the proposed F-test is higher than the assumed value of the significance level; and frequently, it is substantially higher.

---

[1]Presently at the Office of Program Evaluation and Research, California Department of Education, 721 Capitol Mall, Sacramento, California.

It is, quite conceivable that real data commonly found in educational and social sciences does not always meet the rigorous assumption of normality. Consequently, Feldt's statistic (W) may have limited applicability if it fails to be robust to deviations from normality.

The present investigation had two major purposes. The first was to explore the use of a relatively new inferential technique called Tukey's Jackknife (Miller, 1964), in establishing a confidence interval about α. The second purpose was to study the robustness of the Feldt and the jackknife procedures when the data fails to satisfy usual normality assumptions. Besides the claims of being a competitor of the usual normal theory F-test (see Arvesen, 1969), the jackknife procedure has been shown in many situations to behave robustly against non-normality.

<u>The jackknife Procedure</u>

The jackknife procedure used in obtaining an interval estimate of a parameter is based upon Quenouille's (1949, 1956) work on reduction of bias in an estimator. Tukey's contribution is in the extention of Quenouille's procedure as an infercutial technique that may be appropriate when either the distributional assumptions are in question or distribution theory is impossible to derive. The name "jackknife" procedure due to Tukey naturally suggests that it is a generally applicable tool like the boy scout's jackknife, though many of its jobs could be better accomplished using specialized tools if such tools were available. However, Miller (1964, 1968) has shown that Tukey's jackknife is a valid mathematical technique for constructing confidence intervals.

The jackknife procedure can be defined as dividing a set of observations into a set of mutually exclusive and exhaustive groups, obtaining estimates from combination of these groups, and finally averaging these estimates. To be more precise, let $X_1, X_2, \ldots, X_N$ be N independent observations, identically distributed random variables, having an unknown parameter Θ of the common density function F_{Θ}. Furthermore, we assume that a method for estimating Θ is available. In the jackknife procedure, N observations are first grouped into t groups of n observations each such that N=tn, i.e., the t groups are as follows:

$$X_1, \ldots, X_n;\ X_{n+1}, \ldots, X_{2n};\ \ldots ; X_{(t-1)n+1}, \ldots, X_{tn}.$$

As a notational conventions, let $\hat{\theta}_{-0}$ be some estimate of the parameter Θ based on all t groups, and let $\hat{\theta}_{-i}$ be the estimate of Θ based upon the deletion of the i-th group of observations, i.e., based upon (N-n) observations. New estimates of Θ, called pseudo-values, are formed by taking the following combination of $\hat{\theta}_{-0}$ and $\hat{\theta}_{-i}$:

$$\hat{\theta}_{*i} = t\,\hat{\theta}_{-0} - (t-1)\,\hat{\theta}_{-i} \quad \text{for } i = 1, \ldots, t.$$

The jackknife estimate of Θ is the mean of the pseudo-values

$$\hat{\theta}_{*.} = \sum_{i=1}^{t} \frac{\hat{\theta}_{*i}}{t}$$

An estimate of the standard error of the jackknife estimate is given by:

$$s_{\hat{\theta}_{*.}} = \left[\sum_{i=1}^{t} \frac{(\hat{\theta}_{*i} - \hat{\theta}_{*.})^2}{t(t-1)} \right]^{1/2}$$

The jackknife estimate of Θ possesses the interesting property that if Θ is biased of the order $1/N$, then $\hat{\theta}_{*.}$ reduces the bias to the order $1/N^2$. Moreover, Tukey (1958) suggested that in many situations the t pseudo-values $\hat{\theta}_{*1}, \hat{\theta}_{*2}, \ldots, \hat{\theta}_{*t}$, could be treated as t approximately independent, identically distributed observations from which an approximate confidence interval of Θ could be constructed from the student - t distribution. Tukey's proposal implies that the quantity

$$\frac{\hat{\theta}_{*.} - \theta}{s_{\hat{\theta}_{*.}}}$$

is approximately distributed as a Student - t with t - 1 degrees of freedom. Thus a 100r% confidence interval for Θ is

$$\left[\hat{\theta}_{*.} - s_{\hat{\theta}_{*.}}\, t_{(t-1),(1+\gamma/2)}\,;\ \hat{\theta}_{*.} + s_{\hat{\theta}_{*.}}\, t_{(t-1)(1+\gamma)/2}\right] = [L, U]$$

Jackknifing Transformation of Statistics

Miller (1964) in discussing Tukey's conjucture, actually proved that at least for the transformation of means, the pseudo-values are asymptolically distributed normally. However, in the jackknife estimate a slightly stronger assumption of bounded second derivative near the origin was required as compared to the requirement of first derivative near the origin for the unjackknifed estimate of the transformation of means.

In a subsequent paper, Miller (1968) extended his discussion of the mean to the case where θ is the sample variance, or a transformation of the sample variance. For a moderate sample size, his results have indicated that the jackknife t-test is a valid competitor to the F-test if the data are normal, and moreover, the jackknife gives almost correct significance levels if the data are not normal, unlike the F-test.

Miller's theorems on the jackknifing of means and variances were extended to the consideration of **U**-statistics, or a function of several **U**-statistics by Arvesen (1969). **U**-statistics embrace a large class of statistics including the sample mean, variance, and estimates of variance components in ANOVA models.

Application of jackknife for Interval Estimation

It has been pointed out by Mosteller and Tukey (1968) that the jackknife technique can be applied to the same data base in a variety of ways. This variety van occur in two ways. One of these is the particular function of the estimator to be jackknifed. For example, one can jackknife log $\hat{\theta}$ or $\hat{\theta}^{1/2}$ instead of $\hat{\theta}$. There may be some advantage in jackknifing one expression rather than other, or one function can lead towards desirable results whereas the other function may not be useful at all. Rogers (1971) while jackknifing disattenuated correlation coefficient $\{r(T_X, T_Y)\}$ found that performance of the jackknife procedure on the statistic $\{r(T_X, T_Y)\}^{7/5}$ was slightly superior to the performance of the jackknife procedure on $r(T_x, T_y)$ in certain situations. The transformations referred to here are known as variance stabilizing transformations. Arvesen (1969) has shown that log transformation is a variance. Stabilizing transformation of the ratio of variance component estimates in a two-way Model II ANOVA. Arvesen confirmed empirically that it was more useful to jackknife log of ratio of variance component estimates as compared to simple ratio of component estimates in ANOVA to obtain interval estimates. The problem, however, is that for many useful statistics variance stabilizing transformations are unknown.

The other way due to which variety in jackknifing occurs is the choice of forming sub-groups. When the data is in the form of a vector, Miller (1964) suggested to keep the number of observations (N) equal to the number of groups (t) , resulting in n=1, i.e., one observation per group. For data in the form of a matrix, Cronbach, Rajaratnam, Gleser, and Nanda (1972) as well as Collins (1970) used as many groups as the number of observations in the matrix. This investigation uses the foregoing as well as other procedure of group formation.

Method

To determine which particular statistic would "polish up" the behavior of the jackknife technique in giving interval estimates of coefficient alpha, the following five functions were jackknifed. It has been shown by Pandey (1973) that each of these statistics is a U-statistic:

1. $\hat{\alpha} = 1 - MS_P/MS_{PxU}$
2. $\log_e(\hat{\alpha}) = \log_e\{1 - MS_P/MS_{PxU}\}$
3. $z_{\hat{\alpha}} = \frac{1}{2}\log_e\{(1+\hat{\alpha})/(1-\hat{\alpha})\}$
4. $\hat{\theta}_\ell = MS_P/MS_{PxU}$
5. $\log_e(\hat{\theta}_\ell) = \log_e\{MS_P/MS_{PxU}\}$.

where $\hat{\alpha}$ = estimate of coefficient α, MSp and MSpxu are the mean squares for persons and persons by units in a two-way Model II analysis of variance.

Regarding choice of functions to be jackknifed, choice of $\hat{\alpha}$ is obvious because it is a natural function to be jackknifed. Loge $\hat{\alpha}$ was chosen because log transformation is known to be a variance stabilizing transformation for sample variance. $\hat{\theta}_l$ and $\log_e \hat{\theta}_l$ were used by Arvesen and Schmitz (1970) and $Z_{\hat{\alpha}}$ was intuitively considered to be a good transformation because it is a variance stabilizing transformation for sample correlation coefficients.[2]

For each of the statistic jackknifed, two methods of group formation were used. These are the "technique of eliminating both rows and columns" and the "technique of eliminating rows only" computation of pseudo-values and their variance are given in Pandey (1973).

DATA AND EXPERIMENTAL DESIGN

The continuously scored data representing a matched item-examinee sample from their respective populations was simulated on UNIVAC 1108 computer, the characteristics of which could be manipulated as desired by the experimenter. The following model representing a two-dimensional array of observations of size nxk was used to define the test scores:

$$x_{ij} = \mu + t_i + a_j + e_{ij}.$$

In this model μ is an arbitrary constant, t_i is the effect associated with examinee i, a_j is the effect associated with item j, and e_{ij} is a random error corresponding to a particular observation x_{ij}. The size of the array, value of μ, nature of the distributions of a_j and e_{ij}, and associated parameters were kept constant throughout the experiments, whereas the reliability of the fixed length test and the distributional form of the examinee effect t_i were manipulated systematically. Using three distributional forms of t_i (Normal, Log-normal, and Double-exponential) and three values of population coefficient alpha (.60, .75, and .90) resulted in nine simulation experiments. Each of

[2]Lord (1974) has recently shown that $Z_{\hat{\alpha}}$ is the correct variance stabilizing transformation of the stepped-up reliability coefficient.

the experiments was performed using 1000 replications; the number of times the computed confidence interval did not enclose the population value of coefficient alpha, and the mean interval length were recorded for the Feldt and the ten jackknife procedures at three values of confidence coefficients (.90, .95, .99). Various methods were compared with respect to their empirical significance values, as well as the tightness of the confidence bounds. The tables in the Appendix provide in part the results of the simulation experiments.

RESULTS AND DISCUSSION

The jackknife procedure using technique of elimination of row only, involving Fisher's z-transformation on alpha and $\log(MSp/MS_{PxU})$ were found to have potential for interval estimation of coefficient alpha for the normal and log-normal distributions of examinee score components, the Feldt procedure was found to give slightly conservative estimates of alpha, consistent with the earlier results of Feldt (1965). The two jackknife procedures, in this situation were less conservative. For log-normal distribution of examinee score components, Feldt's procedure gave reasonable interval estimates for low values (=.60) of population alpha, but for higher values of population alpha (=.90), the Feldt procedure was not found to be robust. The two jackknife procedures, on the other hand, were found to be relatively robust over the entire range of investigation of values of alpha. None of the procedures was found useful in interval estimation of coefficient alpha when the distribution of examinee score components were of extreme double-exponential form.

Collins (1970) used the jackknife technique to study generalizability coefficients. Based on his results, he recommended against using the jackknife as an inferential technique for generalizability coefficients. The results of the present investigation are similar to those of Collins for the technique of row and column elimination. However, the technique of row elimination only (not studied by Collins) proved useful in constructing confidence interval for coefficient alpha. Since the technique was found relatively robust, it is recommended that jackknife interval estimates of coefficient alpha be used in computerized item analysis packages.

REFERENCES

Arvesen, J. N. Jackknifing U-statistics. Annals of Mathematical Statistics, 1969, 40, 2076-2100.

Arvesen, J. N. and Schmitz, Thomas H. Robust procedures for variance component problems using the jackknife, Biometrics, 1970, 26, 677-686.

Collins, J. R. Jackknifing Generalizability, Doctoral Dissertation. University of Colorado, Boulder, Colorado, 1970.

Cronbach, L. J., Gleser, G. C., Nanda, H. and Rajaratnam, N. The Dependability of Behavioral Measurements, Wiley, New York, 1972.

Ebel, H. W. Estimation of reliability of ratings. Psychometrika, 1951, 16, 407-424.

Feldt, L. S. The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty, Psychometrika, 1965, 30, 357-370.

Jackson, R. W. and Gerguson, G. A. Studies on the Realiability of Tests. Bulletin No. 12 Department of Educational Research, Ontario College of Education, Toronto: University of Toronto Press, 1941.

Kristof, W. The statistical theory of stepped-up reliability coefficients when a test has been divided into several equivalent parts. Psychometrika, 1963, 28, 221-238.

Kristof, W. On the sampling theory of reliability estimation, Journal of Mathematical Psychology, 1970, 7, 371-377.

Miller, R. G. Jr. A trustworthy jackknife, Annals of Mathematical Statistics, 1964, 35, 1594-1605.

Miller, R. G. Jr. Jackknifing variances, Annals of Mathematical Statistics, 1968, 39, 567-582.

Mosteller, F. and Tukey, J. W. Data Analysis, including statistics. Handbook of Social Psychology, G. Lindzey and E. Aronsom, Eds., Addison-Wesley, Reading, Mass., 1968.

Pandey, T. N. The robustnest of interval estimation of coefficient alpha using the jackknife procedure, Unpublished Dissertation, University of Wisconsin, Madison, 1973.

Quenouille, M. Approximate tests of correlation in time series, Journal of Royal Statistical Society, Ser. B, 1949, 11, 68-84.

Quenouille. M. Notes on bias in estimation. Biometrika, 1956, 43, 353-360.

Rogers, W. T. Jackknifing disattenuated correlations. Unpublished Ph.D. Thesis, University of Colorado, 1971.

Scheffe, H. The analysis of Variance, Wiley, New York, 1959.

Tukey, J. W. Bias and confidence in not-quite large samples (abstract). Annals of Mathematical Statistics, 1958, 29, 614.

Table 1

Results of Feldt's Normal Theory Procedure

Simulation No.	Distribution	Population Alpha	.99 Lower Tail	.99 Upper Tail	.99 Error Freq.	.99 Ave. Interval	.95 Lower Tail	.95 Upper Tail	.95 Error Freq.	.95 Ave. Interval	.90 Lower Tail	.90 Upper Tail	.90 Error Freq.	.90 Ave. Interval
			Nominal Confidence Coefficient											
1	N	.60	6	0	6	.6893	26	27	53	.5176	54	52	106	.4321
2	N	.75	6	6	12	.4324	23	24	47	.3247	54	50	104	.2710
3	N	.90	2	3	5	.1724	23	22	45	.1295	52	63	115	.1081
4	LN	.60	8	0	8	.6910	30	26	56	.5190	56	48	104	.4332
5	LN	.75	8	8	16	.4358	27	27	54	.3273	56	58	114	.2732
6	LN	.90	15	11	26	.1779	37	38	75	.1336	64	77	141	.1115
7	E	.60	1	0	1	.9708	103	107	210	.7290	127	211	338	.6086
8	E	.75	40	110	150	.7121	121	237	358	[illegible]	114	362	476	.4464
9	E	.90	197	115	312	.3371	303	210	513	.2532	203	411	614	.2113

Table 2

Results on Row Only Elimination Procedure: Function Jackknifed $z_{\hat{\alpha}}$

Simulation No.	Distribution	Population Alpha	.99 Lower Tail	.99 Upper Tail	.99 Error Freq.	.99 Ave. Interval	.95 Lower Tail	.95 Upper Tail	.95 Error Freq.	.95 Ave. Interval	.90 Lower Tail	.90 Upper Tail	.90 Error Freq.	.90 Ave. Interval
			Nominal Confidence Coefficient											
1	N	.60	2	5	7	.7047	21	30	51	.5248	50	45	95	.4362
2	N	.75	4	3	7	.4759	17	21	38	.3458	48	48	96	.2845
3	N	.90	2	7	9	.1988	14	36	50	.1404	42	52	94	.1144
4	LN	.60	3	8	11	.7094	25	37	62	.5282	51	56	107	.4390
5	LN	.75	4	8	12	.4880	20	23	43	.3541	43	51	94	.2913
6	LN	.90	1	8	9	.2228	17	38	55	.1555	40	74	114	.1261
7	E	.60	81	114	195	.8207	106	143	249	.6224	132	180	312	.5210
8	E	.75	36	83	119	.7858	93	165	258	.5859	111	232	343	.4867
	E	.90	78	123	201	.4702	129	236	369	.3325	197	269	466	.2704

Table 3

Comparision of the Feldt, Row-1, and Row-2 Methods

Method	Confidence Coefficient	Average Empirical Error Frequencies				Average Interval Lengths	
		Normal		Log-Normal		Normal	Log-Normal
		Frequency	% Deviation	Frequency	% Deviation		
Feldt	.99	7.6	24.0	16.7	67.0	.4314	.4349
Row-1	.99	7.7	23.3	10.7	6.7	.4598	.4734
Row-2	.99	11.7	16.7	11.3	13.3	.5214	.5386
Feldt	.95	48.3	3.3	61.7	23.3	.3239	.3266
Row-1	.95	46.3	7.3	53.3	6.7	.3370	.3459
Row-2	.95	51.7	3.3	56.3	12.7	.3612	.3715
Feldt	.90	108.3	8.3	119.7	19.7	.2704	.2726
Row-1	.90	95.0	5.0	105.0	5.0	.2783	.2855
Row-2	.90	104.0	4.0	111.0	11.0	.2924	.3002