DOCUMENT RESUME

ED 305 382 TM 012 911

AUTHOR Daniel, Larry G.

TITLE Use of the Jackknife Statistic To Establish the External Validity of Discriminant Analysis Results.

PUB DATE Jan 89

NOTE 22p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Houston, TX, January 27, 1989).

PUB TYPE Reports -- Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS *Discriminant Analysis; Educational Research; Estimation (Mathematics); Monte Carlo Methods; *Research Methodology

IDENTIFIERS Cross Validation; Discriminant Function Index; *External Validity; *Jackknifing Technique

ABSTRACT

That the jackknifing technique is superior to traditional techniques for assessing the external validity of statistical results of discriminant analysis is defended. Traditional approaches assessed include: (1) the empirical method, in which the discriminant function coefficients (DFCs) obtained in a given analysis are applied to predict group membership in the same sample used for deriving the DFCs; (2) the "holdout" method, in which statistical results are cross-validated by random splitting of the original sample into a group for deriving the discriminant function and a group for cross-validating it; (3) the Monte Carlo method; and (4) the random assignment method, whereby discriminant functions are computed based on repeated random assignment of actual cases from the original sample to groups. The jackknife statistic (JS) is similar to the "U-method" and focuses on the stability of the DFCs obtained in the original analysis. One case or subset of cases is eliminated from the original data set, and the discriminant function is computed using the remaining observations. The procedure is repeated, with each individual observation or unique subgroup, in turn, omitted. At each step, pseudo-values are computed, based on computation of the original and cases-minus-one DFCs. The values are averaged to provide a jackknifed estimate of the DFCs. A data set shows the JS's value and is used to assess the stability of the jackknifed DFCs for three predictors of teachers' (N=69) level of experience. The JS may be used to reduce bias in an estimator that is attributable to artifacts of the sample used. Since jackknife methods minimize sample splitting via sample omission and reuse, they are particularly useful with small samples. Four data tables are provided. (TJH)

Use of the Jackknife Statistic to Establish the External Validity of Discriminant Analysis Results

Larry G. Daniel

University of New Orleans 70148

Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX, January 27, 1989.

ABSTRACT

Several traditional statistical techniques for assessing the external validity of statistical results are discussed. The author presents reasons why the jackknife technique is superior to these traditional techniques. A small data set is used to illustrate the value of the jackknife statistic in determining the stability of discriminant function coefficients.

Use of the Jackknife Statistic to Establish the External Validity of Discriminant Analysis Results

Discriminant analysis is a powerful multivariate technique which may be used in educational research to classify individuals into groups or to identify specific dimensions or qualities which differentiate among individuals in various groups (Afifi & Clark; 1984). When employing discriminant analysis (or any other parametric statistical procedure), researchers are usually concerned with the validity of the obtained results with respect to the broader population of interest. As with any statistical technique, there is always the possibility that discriminant analysis results may simply capitalize on artifacts of the sample employed for the study, and as a result, may not be generalizable to the larger population of interest. Generalizability is particularly at risk in cases in which the sample size is extremely small or when the representativeness of the sample is questionable (Frank, Massey, & Morrison, 1965).

Researchers and statisticians have developed a number of approaches for assessing the external validity of statistical results, yet the value of many of the approaches is offset by certain weaknesses. In the present study, four traditional approaches to validation of discriminant analysis results are briefly discussed. Problems inherent to each of these approaches are presented. Two alternative methods, the U-method

and the jackknife statistic are discussed, with emphasis upon how these methods are superior to traditional methods. Selected variables from a small data set are used to perform a discriminant analysis to illustrate the value of the jackknife statistic in a concrete fashion.

Traditional Approaches to Assessing External Validity

Four traditional approaches to assessing the stability of discriminant function coefficients have been summarized in the literature (Afifi & Clark, 1984; Cooil, Winer, & Rados, 1987; Montgomery, 1975). These methods include the following:

(1) The "empirical" method. In this method, the discriminant function coefficients obtained in a given analysis are applied to predict group membership in the same sample used for deriving the coefficients. The degree of "goodness of fit" is assessed by determining the proportion of cases which have been correctly classified. Although this method is probably the most computationally straightforward of all validation techniques, it tends to produce very biased estimates of generalizability, particularly when the sample size is small. In general, use of the "empirical" method tends to overestimate classification probability since it employs the same sample for both deriving and validating the discriminant functions (Afifi & Clark, 1984).

(2) The "holdout" ("cross validation," "split half," or "invariance") method. Using this method, a researcher can cross-validate statistical results "by randomly splitting the original

sample. . .into two [approximately equivalent] subgroups: one for deriving the discriminant function and one for cross-validating it" (Afifi & Clark, 1984, p. 266). Ideally, discriminant function coefficients should be calculated for each of the subsamples, and then validated using the other sample. The invariance method is appealing for at least two reasons. First, it requires the use of a single sample, and consequently can be easily used within the domain of a single research study. Second, it minimizes the problem of bias inherent to the "empirical" method by using different samples to derive and validate results. For these reasons, the invariance method has been called "the most popular approach to cross-validation. . .in all of the social sciences" (Cooil et al., p. 271). The invariance method is problematic, however, when the sample size is small, since splitting an already small sample increases the risk that the function coefficients obtained in the even smaller groups are merely artifacts of the sample (Morrison, 1969).

(3) The "Monte Carlo" method. Using Monte Carlo methodology, researchers can randomly generate synthetic data from which discriminant functions are derived with the same degrees of freedom as the original data. These data can be used to validate the predictive discriminant function coefficients derived using the original data set. The Monte Carlo method is useful when all predictor variables are independent of one another, e.g., when uncorrelated factor scores are used as predictors (Crask & Perreault, 1977). In most cases in which

multiple predictors are used, however, the predictors will tend to be correlated. As a result, Monte Carlo methods are problematic in that it is difficult to reproduce the variance-covariance structure of the original data using randomly-generated data (Montgomery, 1975). However, a computer program available from Morris (1975) can be used for this purpose.

(4) The "random assignment" method. In this procedure, discriminant functions are computed based upon repeated random assignment of actual cases from the original sample to groups. Once several sets of discriminant functions are derived using the randomly assigned cases, the results of these classifications may be compared to the original sample's results (Montgomery, 1975). This method is appealing in that it uses actual rather than synthetic data, and therefore preserves the appropriate interrelationships among the variables. However, since this method relies upon random or chance classification, its use is questionable as an assessment in an "absolute" sense of the performance of discriminant function coefficients (Crask & Perreault, 1977).

Considering Alternatives to Traditional Validation Methods

As previously noted, traditional approaches to assessing the external validity or generalizability of discriminant analysis results are replete with a number of inherent weaknesses. As a result of these weaknesses, the several traditional validation techniques tend to produce biased estimates of the stability of the obtained results. Two less-frequently-used validation

methods, the "U-method" (Bartlett, 1952; Mantel, 1967) and the "jackknife statistic" (Gray & Schucany, 1972; Tukey, 1958) attempt to remedy the shortcomings associated with the traditional methods. The "U-method," which focuses upon classification errors, involves computation of a series of discriminant functions, each omitting one case or subset of cases from the original sample. At each step of the U-analysis, the obtained discriminant functions are used to classify the case(s) omitted at that step of the analysis.

The "jackknife statistic," although a similar technique, focuses upon the stability of the discriminant function coefficients obtained in the original analysis. In this technique, one case or subset of cases is eliminated from the original data set and the discriminant function is computed using the remaining observations. This procedure is repeated, with each individual observation or unique subgroup, in turn, omitted. In each step of the analysis, "pseudovalues" (Quenouille, 1956) are computed based upon the computation of the original and the cases-minus-one discriminant function coefficients. These pseudovalues are averaged to provide a "jackknifed" estimate of the discriminant function coefficients. Stability of the original values is assessed by determining whether they fall within confidence intervals for the jackknifed values.

U-method and jackknife approaches are superior to other traditional validation methods in that they make use of all of the data in a particular data set while eliminating bias in

estimates of stability by "averaging out" the effects of atypical or outlying cases within a given data set. Use of these techniques has been demonstrated to produce more conservative and less biased estimates of true population characteristics (Crask & Perreault, 1977). These techniques are particularly useful when sample size is small as they minimize sample splitting (Fenwick, 1987). The jackknife statistic offers a method for evaluating stability of discriminant function coefficients while the U-method estimates error rates in the classification of cases. The two methods may be used together or in isolation depending upon the researcher's purposes. In the present study, the use of the jackknife statistic will be illustrated.

Computing the Jackknife Statistic--An Overview of Procedure

According to Crask and Perreault (1977), "[t]he essence of the jackknife approach is to partition out the impact or effect of a particular subset of the data (e.g., a single case) on an estimate derived from the total sample" (p. 61). Generally, the jackknife statistic is derived by computing a statistical estimator (e.g., a discriminant function coefficient) using the entire population, and then computing the same estimator eliminating given subsets of the data. The averaged weighted value of the estimator when the analysis is run repeatedly with the various subsets of the data is used to compute the jackknifed value of the estimator. A brief explanation of the mathematical procedures involved in computing the jackknife statistic as explained by Crask and Perreault (1977) and based on the

pioneering work of Quenouille (1956) may be halpful.

In computing the jackknife statistic, a given sample of size N is partitioned into k subsets of size M ($kM = N$). All subsets must be of the same size (M) and may be as small as one case or as large as the largest multiplicative factor of N. A predictive estimator (e.g., a discriminant function coefficient), designated as theta-prime (θ') is then computed using all k of the subsamples from the original sample of size N. The same estimator is also computed with the i^{th} subset (i = 1 to k) omitted from the sample. This estimator is designated as θ_i' . This procedure is repeated k times with a different subset omitted each time.

Before computing the jackknifed estimator, weighted combinations of the θ' and θ_i' values are computed. These weighted values are called pseudovalues (Quenouille, 1956), and are designated by the letter J. The pseudovalues are computed using the equation:

$$(1)\quad J_i(\theta') = k\,\theta' - (k-1)\,\theta_i$$

where i = 1, 2, 3,. . . ., k.

The average of the pseudovalues is the jackknifed estimator:

$$(2)\quad J(\theta') = \Big[\sum_{i=1 \text{ to } k} J_i(\theta')\Big] / k$$

Tukey (1958) argued that a given set of pseudovalues could be regarded as an approximately normal distribution; hence the stability of a given jackknifed estimator may be evaluated by determining confidence intervals about the estimator, and then testing to determine whether the researcher can conclude that the

population estimator falls within those confidence interval bands. This test may be done by dividing the estimator by its associated standard error to obtain a Student *t*-value. The degrees of freedom for this *t*-value equal the numbers of partitions of the original sample (and, consequently, the number of pseudovalues) minus one. A jackknifed estimator is considered stable if its calculated *t*-value exceeds the *t*-critical value.

An Application of the Jackknife Technique

In the following example, selected variables from a small data set (Daniel & Okeafor, 1987) will be used to illustrate the jackknife technique as applied to the validation of discriminant function coefficients. The original study was designed to test the relationship between teachers' levels of teaching experience and the degree of confidence they placed in the professional performance of themselves and other teachers. Teachers at varying experience levels rated themselves, the typical beginning teacher, and the typical experienced teacher on three subscales of a "logic of confidence" measure (Okeafor, Licata, & I·ker, 1987). These subscales were "overlooking" (the degree to which the respondent felt undesirable behaviors of the teacher should be overlooked by superiors), "avoidance" (the degree to which the respondent felt administrators should avoid direct supervision of the teacher), and "professionalism" (the degree to which the respondent felt the teacher should be regarded as a professional). Data were analyzed using three one-way

multivariate analyses of variance, with teachers' level of experience (preservice, novice, or experienced) serving as the predictor variable for each set of ratings.

For the purposes of the present study, only the data pertaining to ratings of the typical beginning teacher will be used. In order to ease interpretation of results, the three levels of experience will be collapsed into two, with both preservice and novice teachers coded as "inexperienced." Although the theoretical soundness of collapsing these two categories into one may be debatable, the intent of the present study is to illustrate the usefulness of the jackknife statistic, and not necessarily to add substantively to the findings of the original study. The data used in the present study are presented in Table 1. The 69 cases were randomly assigned to 23 (<u>k</u>) groups of three persons each (<u>M</u>), for the purposes of performing the jackknife analysis.

INSERT TABLE 1 ABOUT HERE

Data from the entire sample (<u>N</u> = 69) were analyzed using the SPSSx DISCRIMINANT procedure. Standardized discriminant function coefficients derived from the analysis for the three predictor variables were -.57065 (professionalism subscale), -.61074 (avoidance subscale), and -.12889 (overlooking subscale). The DISCRIMINANT procedure was repeated 23 more times with one unique <u>k</u> group omitted from the sample in each repetition. Standardized discriminant function coefficients obtained for each of the

repetitions as well as for the original sample are presented in Table 2.

INSERT TABLE 2 ABOUT HERE

Using the data from Table 2, weighted pseudovalues (J_i, where *i* = 1 to 23) were computed for each of the 69 discriminant function coefficients obtained with the given subset *i* deleted at each of the 23 steps. These pseudovalues were computed using equation (1). Jackknifed discriminant function coefficients (average of the 23 pseudovalues for each discriminant function coefficient) were also computed. In addition, a calculated *t*-value was computed for each jackknifed coefficient using 22 degrees of freedom (number of pseudovalue repetitions minus one). Pseudovalues, jackknifed discriminant function coefficients, and associated *t*-values are presented in Table 3. Ninety-five percent confidence intervals for the jackknifed coefficients are presented in Table 4.

INSERT TABLES 3 AND 4 ABOUT HERE

Discussion

As previously noted, the jackknife statistic is useful in evaluating the stability of a given estimator by eliminating bias due to the inclusion of outlying or atypical cases in a given sample. In the present example, the stability of jackknifed discriminant function coefficients for three predictors of teachers' level of experience was assessed. The jackknifed

coefficients for all three of the predictors were quite close in value to the original coefficients obtained using the entire sample. Standard error confidence intervals and t-values were computed for each coefficient. Based upon these last computations, presented in Tables 3 and 4, the stability of the jackknifed coefficients for two of the three predictors (professionalism and avoidance subscales) was supported while the stability of the coefficient for the third predictor (overlooking subscale) was not. These findings indicate that the first two predictors may be considered as valid discriminators between the two groups of teachers, and that the results may be appropriately generalized to the larger population of interest. As indicated by the data presented in Table 3, the third variable (overlooking) tends to be unstable against changes in the composition of the sample, and therefore is a more biased indicator. Furthermore, the near-zero magnitude of the third predictor's discriminant function coefficient using the total sample, reported in Table 2, and using the jackknifed estimate, reported in Table 3, suggests that this variable has little predictive validity.

Summary

The jackknife statistic may be used to reduce the bias in an estimator which is attributable to artifacts of the sample employed for the study. Since jackknife methods minimize sample splitting through sample omission and reuse, they are particularly useful when sample size is small. The present study

has demonstrated an appropriate use of the jackknife statistic as a tool for assessing the stability of discriminant analysis results.

REFERENCES

Afifi, A. A., & Clark, V. (1984). Computer-aided multivariate analysis. Belmont, CA: Lifetime Learning Publications.

Bartlett, M. S. (1952). An inverse matrix adjustment arising in discriminant analysis. Annals of Mathematical Statistics, 22, 167.

Brillinger, D. R. (1966). The application of the jackknife to the analysis of sample surveys. Journal of the Market Research Society, 8, 74-80.

Cooil, B., Winer, R. S., & Rados, D. L. (1987). Cross-validation for prediction. Journal of Marketing Research, 24, 271-279.

Crask, M. R., & Perreault, W. D. (1977). Validation of discriminant analysis in marketing research. Journal of Marketing Research, 14, 60-68.

Daniel, L. G., & Okeafor, K. R. (1987, November). Teaching experience and confidence in teachers. Paper presented at the annual meeting of the Mid-South Educational Research Association, Mobile, AL. (ERIC Document Reproduction Service No. ED 292 763)

Fenwick, I. (1979). Techniques in market measurement: The jackknife. Journal of Marketing Research, 16, 410-414.

Frank, R. E., Massey, W. F., & Morrison, D. G. (1965). Bias in multiple discriminant analysis. Journal of Marketing Research, 2, 250-258.

Gray, H. L., & Schucany, W. R. (1972). The generalized jackknife statistic. New York: Marcel Dekker, Inc.

Green, P. E. (1964). Bayesian classification procedures in analyzing customer characteristics. _Journal of Marketing Research_, _1_, 44-50.

Mantel, N. (1967). Assumption-free estimators using U-statistics and a relationship to the jackknife method. _Biometrics_, _23_, 567-571.

Miller, R. G. (1964). A trustworthy jackknife. _Annals of Mathematical Statistics_, _35_, 1594-1605.

Montgomery, D. B. (1975). New product distribution: An analysis of supermarket buyer decisions. _Journal of Marketing Research_, _12_, 255-264.

Morris, J. D. (1975). A computer program to create a population with any desired centroid and covariance matrix. _Educational and Psychological Measurement_, _35_, 707-710.

Morrison, D. G. (1969). On the interpretation of discriminant analysis. _Journal of Marketing Research_, _6_, 156-163.

Okeafor, K. R., Licata, J. W., & Ecker, G. (1987). Toward an operational definition of the logic of confidence. _Journal of Experimental Education_, _56_, 47-54.

Quenouille, M. H. (1956). Notes on bias in estimation. _Biometrika_, _43_, 353-360.

Tukey, J. W. (1958). Bias and confidence in not-quite large samples. _Annals of mathematical statistics_, _29_, 614.

Table 1
DATA LISTING--SUBSCALE SCORES

CASE	K-GROUP	EXPER[1]	AVOID[2]	OVERLOOK[3]	PROFESS[4]
1	9	1	28	21	50
2	8	1	26	14	47
3	15	1	27	30	43
4	16	1	11	9	11
5	2	1	41	38	53
6	7	1	24	13	63
7	12	1	14	22	43
8	11	1	25	14	35
9	10	1	15	18	38
10	5	1	22	16	50
11	13	1	22	20	40
12	6	1	10	15	46
13	20	1	33	28	43
14	14	1	30	28	48
15	15	1	24	22	40
16	3	1	26	32	42
17	3	1	28	24	49
18	8	1	29	24	52
19	9	1	16	10	26
20	5	1	30	31	42
21	9	1	28	23	48
22	7	1	27	25	31
23	2	1	24	19	39
24	3	1	28	18	46
25	21	1	23	18	48
26	8	1	27	15	54
27	2	1	31	16	45
28	6	1	32	24	43
29	19	1	27	20	24
30	12	1	22	13	43
31	21	1	27	14	42
32	21	1	23	11	51
33	13	1	23	26	41
34	11	1	33	19	48
35	5	2	27	19	45
36	14	2	33	23	62
37	22	2	33	29	52
38	10	2	13	16	36
39	1	2	33	15	58
40	4	2	32	30	47
41	11	2	34	30	40
42	12	2	43	29	59
43	1	2	36	29	42
44	3	2	16	10	41
45	4	2	33	14	61
46	4	2	30	23	50

(continued next page)

Table 1 (continued)

47	19	2	40	32	43
48	17	2	42	18	64
49	19	2	40	24	50
50	17	2	40	24	60
51	18	2	37	24	53
52	18	2	26	23	39
53	16	2	28	13	53
54	13	2	28	25	49
55	17	2	29	24	48
56	15	2	32	26	54
57	23	2	32	21	45
58	6	2	22	12	61
59	20	2	41	18	53
60	10	2	19	18	48
61	22	2	21	23	38
62	1	2	42	36	56
63	20	2	42	24	59
64	23	2	32	30	33
65	23	2	32	29	55
66	14	2	23	25	49
67	22	2	24	27	42
68	18	2	40	27	53
69	16	1	21	20	45

[1] Level of experience (1 = experienced, 2 = novice)
[2] Avoidance subscale--rating of beginning teacher
[3] Overlooking subscale--rating of beginning teacher
[4] Professionalism subscale--rating of beginning teacher

Table 2

STANDARDIZED DISCRIMINANT FUNCTION VALUES WITH SUCCESIVE SUBSAMPLES DELETED FROM ORIGINAL SAMPLE

K-GROUP DELETED	PROFESS	AVOID	OVERLOOK
none	-.57	-.61	-.13
1	-.60	-.57	-.13
2	-.53	-.63	-.18
3	-.56	-.57	-.23
4	-.55	-.64	-.12
5	-.60	-.56	-.18
6	-.50	-.61	-.16
7	-.60	-.57	-.12
8	-.63	-.59	-.07
9	-.58	-.59	-.13
10	-.49	-.76	-.07
11*	-.56	-.65	-.04
12	-.62	-.52	-.16
13	-.56	-.63	-.12
14	-.49	-.69	-.09
15	-.55	-.58	-.19
16	-.68	-.50	-.30
17	-.56	-.62	-.13
18	-.60	-.60	-.13
19	-.54	-.61	-.10
20	-.56	-.57	-.20
21*	-.58	-.65	-.02
22*	-.54	-.72	-.03
23	-.63	-.57	-.10

*Actual function values for this repetition had positive signs. These "reflected" values were converted to negative values (multiplied by -1) so that they would be directly comparable to results from other repetitions.

Table 3

PSEUDOVALUES AND JACKKNIFED DISCRIMINANT FUNCTION COEFFICIENTS FOR PREDICTOR VARIABLES

REPETITION	PROFESS	AVOID	OVERLOOK
1	.15	-1.44	-.08
2	-1.52	-.20	1.08
3	-.86	-1.48	2.19
4	-.92	-.00	-.29
5	.17	-1.76	.99
6	-2.07	-.53	.54
7	.11	-1.55	-.31
8	.68	-1.03	-1.34
9	-.41	-1.08	-.02
10	-2.36	2.70	-1.39
11	-.76	.32	-2.12
12	.49	-2.68	.63
13	-.73	-.11	-.32
14	-2.43	1.12	-.95
15	-.95	-1.38	1.18
16	1.86	-2.98	3.65
17	-.72	-.46	-.19
18	.09	-.79	-.19
19	-1.16	-.54	-.73
20	-.80	-1.47	1.51
21	-.31	.21	-2.59
22	-1.16	1.78	-3.59
23	.73	-1.54	-.78
Jackknifed Coefficients	-.56	-.65	-.14
t-calc. values (df = 22)	2.66*	2.41*	.42
t-crit. values (*p* = .05)	2.07	2.07	2.07

*Indicates coefficient stability.

Table 4

UPPER AND LOWER BOUND (95% CONFIDENCE LEVEL) INTERVALS FOR JACKKNIFED COEFFICIENT VALUES

	PROFESS	AVOID	OVERLOOK
Original Coefficients	-.57	-.61	-.13
Jackknifed Coefficients	-.56	-.65	-.14
Lower Bound	-.98	-1.18	-.78
Upper Bound	-.14	-.11	.51