DOCUMENT RESUME

ED 137 345 TM 006 146

AUTHOR Kolen, Michael J.; And Others
TITLE Methods of Smoothing Double-Entry Expectancy Tables Applied to the Prediction of Success in College. Research Report No. 91.
INSTITUTION Iowa Univ., Iowa City. Evaluation and Examination Service.
PUB DATE Mar 77
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *College Entrance Examinations; College Freshmen; *Expectancy Tables; *Grade Point Average; *Higher Education; Multiple Regression Analysis; Prediction; Predictor Variables; Probability; Secondary Education; *Statistical Analysis; *Success Factors
IDENTIFIERS ACT Assessment Program; *Smoothing Methods

ABSTRACT

Six methods for smoothing double-entry expectancy tables (tables that relate two predictor variables to probability of attaining a selected level of success on a criterion) were compared using data for entering students at 85 colleges and universities. ACT composite scores and self-reported high school grade averages were used to construct expectancy tables based on data for students entering each institution in 1969-1970. Tables were constructed using two levels of success--"C or better" and "B or better" first semester grade point averages. The tables were smoothed using each method and evaluated according to how closely the smoothed tables corresponded to observed data at the same institutions in 1971-72 and in 1972-73. The smoothed tables were more accurate than those based on 1969-70 observed relative frequencies. A linear regression of observed relative frequency on predictor value was most accurate; two extensions of an isotonic regression method were nearly as accurate. A commonly used regression method was found to be less accurate than most other methods. (Author)

20.08

METHODS OF SMOOTHING DOUBLE-ENTRY EXPECTANCY TABLES APPLIED TO THE PREDICTION OF SUCCESS IN COLLEGE

Michael J. Kolen
William M. Smith
Douglas R. Whitney

Research Report No. 91

Presented at the Annual Meeting
American Educational Research Association
New York, New York
April, 1977

Evaluation and Examination Service
The University of Iowa
Iowa City, Iowa

March, 1977

METHODS OF SMOOTHING DOUBLE-ENTRY EXPECTANCY TABLES

APPLIED TO THE PREDICTION OF SUCCESS IN COLLEGE

Michael J. Kolen, William M. Smith and Douglas R. Whitney

University of Iowa

Abstract

Six methods for smoothing double-entry expectancy tables (tables that relate two predictor variables to probability of attaining a selected level of success on a criterion) were compared using data for entering students at 85 colleges and universities. ACT composite scores and self-reported high school grade averages were used to construct expectancy tables based on data for students entering each institution in 1969-70. Tables were constructed using two levels of success--"C or better" and "B or better" first semester grade point averages. The tables were smoothed using each method and evaluated according to how closely the smoothed tables corresponded to observed data at the same institutions in 1971-72 and in 1972-73. The smoothed tables were more accurate than those based on 1969-70 observed relative frequencies. A linear regression of observed relative frequency on predictor value was most accurate; two extensions of an isotonic regression method were nearly as accurate. A commonly used regression method was found to be less accurate than most other methods.

METHODS OF SMOOTHING DOUBLE-ENTRY EXPECTANCY TABLES APPLIED TO THE PREDICTION OF SUCCESS IN COLLEGE

The probability that an individual will attain a certain level of success is often the measure of primary interest in guidance and placement decisions and in the selection of qualified applicants. Expectancy tables display the likelihood of success on some criterion measure given various levels of performance on one or more predictors related to success.

One factor that detracts from the usefulness of expectancy tables is that chance irregularities can occur when the tables are constructed directly from observed data, especially when relative frequencies are computed for categories containing few observations (Anastasi, 1976). These can make it appear that the probability of success is smaller for high ability students than for lower ability students when theory and experience indicate otherwise. When such apparent contradictions occur, it is desirable to "smooth" the expectancy table to remove the logical inconsistencies. Perrin & Whitney (1976) studied a number of methods for smoothing single-entry expectancy tables and concluded that the gains in accuracy which resulted warranted their use for tables used in college admissions, guidance, or planning purposes.

When two predictors are available, the probability of attaining a chosen level of success on the criterion can be displayed in a double-entry expectancy table (Wesman, 1966). Score intervals on the two predictors are listed at the margins of such double-entry expectancy tables. The values in the body of the table estimated the probability of achieving success given the row and column values on the predictors. These probabilities can be estimated directly from the observed relative frequencies or by the use of various smoothing techniques.

The concern of this paper is with situations in which it can be presumed that the two predictor variables have a monotonic nondecreasing relationship with probability of success and with each other. The problem examined here arose in the process of constructing expectancy tables using achievement test scores and high school percentile rank to estimate probabilities of attaining certain grade point averages. A number of the estimated probabilities in the tables constructed using the observed relative frequencies contradicted our beliefs about the nature of the underlying relationship among the variables. That is, the assumption of monotonicity seems reasonable for tables reflecting test scores and high school grades used to estimate the probability of attaining certain grade average levels in college.

American College Test (ACT) composite scores and self reported high school grade average (HSA) were used to estimate the proportion of students evidencing two levels of success in college. The two levels were earning a first term grade point average (GPA) of "C or better" and "B or better". Double-entry expectancy tables were constructed directly from the observed relative frequencies and also by applying various smoothing techniques to these relative frequencies. Two indexes reflecting the similarity between cross-validation year relative frequencies and estimated relative frequencies were used to evaluate the smoothing methods.[1]

Construction Methods

Method 1: Observed Relative Frequencies

The most common method for constructing double-entry expectancy tables

[1] The use of base year data to prepare tables for use with students in subsequent years suggest that, for each institution, joint probabilities exist and are identical from year-to-year. The degree to which each method's estimate resembles these parameters will be studied in another paper.

is to report the observed relative frequency of success at each combination of the two predictor variables. For example, if 30% of the individuals with ACT scores between 20 and 25 and HSA between 2.6 and 3.3 were successful in a particular year, a value of .30 would be placed in the appropriate cell of the table. Thus, the observed relative frequencies from one year are used to predict the relative frequencies in subsequent years.

Method 2a: Linear Regression on GPA

The original data was used to estimate the intercept, regression weights, and standard error of estimate for a mutliple linear regression model using GPA as the criterion variable. The entry for each cell in the expectancy table was estimated by substituting the center[2] of each ACT and HSA interval into the regression equation and calculating the standard normal deviate corresponding to the difference between the predicted GPA and the selected level of success. The cell entry is then the cumulative distribution function at the normal deviate value. This method assumes a planar relationship among GPA, HSA and ACT and normal, homoscedastic conditional distributions.

Method 2b: Curvilinear Regression on GPA

Method 2a was amended to include the square of each predictor variable and the cross-product term in the equation for predicting GPA. This method allowed for a wider range of possible relationships among GPA, ACT, and HSA than did Method 2a. Otherwise, it was identical to Method 2a.

[2] The center of each interval was defined as the mean of the interval under the assumption that the marginal distributions of ACT and HSA were normal. The mean of the interval in unit normal deviates from the overall mean is:

$$\mu_{interval} = \frac{y_1 - y_2}{p} \text{ where,}$$

y_1 and y_2 are the ordinates of the unit normal curve corresponding respectively, to the low and high endpoints of the interval, and p is the area under the unit normal curve that lies between the endpoints of the interval.

Method 3a: Linear Regression on Observed Relative Frequencies:

ACT and HSA were regressed on the relative frequencies in the unsmoothed table (each relative frequency was weighted by the number of cases in that cell). The center of the intervals were again used as the values for ACT and HSA in the computations. The estimated intercept and regression weights were used to generate the smoothed expectancy table. Any predicted relative frequencies less than zero or greater than one were set equal to the appropriate limit. This method assumes that a planar relationship existed among probability of success, HSA and ACT scores.

Method 3b: Curvilinear Regression on Observed Relative Frequencies.

Method 3a was amended to include the square of each predictor variable and the cross-product term in the equation for predicting the relative frequencies of success. This method (like 2a) allowed for a wider range of relationships among probability of success, ACT, and HSA.

Method 4: Extensions of the Isontonic Regression

The isotonic regression method (Ayer, Brunk, Ewing, Reed, and Silverman, 1955), which assumes only that the relationship between probability of success and the predictor variable is non-decreasing monotonic, is very straightforward in single-entry tables. After forming the expectancy table, the table is examined for reversals. Each reversal, where relatively fewer students with higher predictor values achieved success, is considered as a chance reversal because it violates the assumption that the relationship between predictor and criterion is non-decreasing monotonic. When such a reversal is encountered, the two (or more) relative frequencies involved in the reversal are weighted by the number of observations in each of the cells and averaged. This average replaces the observed relative frequencies in each cell involved in the reversal. The process continues until there are no reversals remaining.

The computations involved in extending this method to double-entry tables, while still preserving its mathematical qualities discussed by Ayer, et al (1955), appear to be intractable. The extensions reported here are attempts to extend the logic (though they do not necessarily retain the mathematical properties), of the single-entry method to smooth double-entry tables.

In the two extensions of this method, the only assumptions about the form of the relationship among ACT, HSA, and probability of success are that for any given ACT level, the relationship between HSA and the probability of success is non-decreasing monotonic and that for any given HSA level, the relationship between ACT and probability of success is non-decreasing monotonic (i.e., the conditional distributions of each predictor and criterion given the other predictor are non-decreasing monotonic). These assumptions do not, for example, indicate whether or not individuals with an ACT score of 25 and a HSA of 2.7 should have as high an estimated probability of success than individuals with an ACT score of 24 and a HSA of 3.0. The extensions to double-entry tables involved two methods of resolving this problem; other defensible (and possibly better) extensions exist.

Method 4a: Alternately Treating Rows and Columns as Single-Entry Tables

In the first phase, each row in the table was smoothed (if necessary) by the method described above for single-entry tables. After all rows had been adjusted, the table was examined column by column to determine if any reversals remained. Any reversals in the columns were adjusted by the same procedure that was applied to the rows. The second phase involved returning to the original table and smoothing first by columns and then by rows. Each of the two phases produced usable, although often different, solutions. Since there was no reason to prefer one solution over the other, the cell entries in the two resulting tables were averaged and examined to determine if any reversals existed. If reversals still remained, the entire procedure was repeated on the averaged table as often as necessary.[3] Cells that contained zero observations in the original data were treated as zero frequency and zero relative frequency. The cell frequencies involved in reversals were also averaged because, without this procedure, the solution would often not converge or would converge very slowly.

Figure 1 illustrates this smoothing method applied to a 3 x 4 expectancy table. In the second row in Figure 1b, the values .300, .200, and .200 in the unsmoothed table represent a reversal. These values were averaged (weighted by cell frequencies) to obtain the value .211. Also note that the cell frequencies were averaged to produce the new cell frequency values of 31.7. The resulting table (Figure 1f) was obtained by averaging the cell entries and frequencies in the tables in Figure 1c and Figure 1e. Since no reversals were present in the resulting expectancy table, no further smoothing was required.

Insert Figure 1 about here

Method 4b: Linear Regression Weights Used to Provide a Single Dimensional Ordering

If the cells of the table could be ordered in such a way that one would know, for example, that individuals with an ACT score of 25 and a HSA of 2.7 have at least as great an estimated probability of success as individuals

[3]Of the 170 tables smoothed in this study, 152 (89%) required no repetitions and the remaining 18 tables required only one repetition of the procedure. Thus, this procedure appears to converge rapidly.

with an ACT score of 24 and a HSA of 3.0, then the single-entry method could be employed. To approximate this information, the regression weights from Method 3a were used to provide an unequivocal ordering of the cells. If the predicted value for one cell was greater than the predicted value for another cell, then the probability of success for individuals with scores corresponding to the first cell was assumed to be at least as great as the probability of success for individuals with scores corresponding to the second. In this way a single-entry table was constructed. The single-entry smoothing method was then applied and the table reconstructed. This solution avoided some of the problems encountered in Method 4a.

An example of this method is shown in Figure 2. Based on the regression weights computed from the data shown in Figure 1a, the cells were ordered as in Figure 2a, smoothed using the single-entry method in Figure 2b, and reconstructed in Figure 2c. As would be expected, Methods 4a and 4b produced different results, although the results were usually similar for cells that initially contained a relatively large number of observations.

Insert Figure 2 about here

Procedure

The data for this study consisted of the records of entering freshman students at a sample (stratified by college type) of 85 institutions that participated in one of the ACT Research Services for the years 1969-70, 1971-72 and 1972-73. Perrin & Whitney (1976) provide a more complete description of the institutions. (Henceforth, 1969-70 will be referred to as the base year and 1971-72 and 1972-73 as, respectively, validation year one and two.) For each student completing the first term in one of these schools during the base year or either of the validation years, ACT Assessment composite score, average of four self-reported high school grades and first term grade-point average (GPA) at the institution was obtained.

Construction and Smoothing Expectancy Tables

Six 5 x 5 expectancy tables were constructed for each institution using the observed relative frequencies (one for each of the three years and at each of the two levels of success). The two base year tables for each school were smoothed using each of the six methods. In order to standardize the tables, the ACT and HSA variables were divided into five intervals, each approximately one standard deviation in width. The five ACT composite intervals (based on a mean of approximately 20 and standard deviation of 5) were 12.5 or below, 12.5 to 17.5, 17.5 to 22.5, 22.5 to 27.5 and 27.5 or above. The five HSA intervals (based on the mean of 2.6 and standard deviation of 0.7 reported in the ACT Basic Research Report of 1970-71) were 1.55 or below, 1.55 to 2.25, 2.25 to 2.95, 2.95 to 3.65, and 3.65 or above.

Criteria

The two criteria reflected the degree to which each set of relative frequencies estimated from the base year data corresponded to the relative frequencies observed in each of the validation years. In each case, smaller values of the index reflect more accurate estimation.

The first index weighted each of the errors in predicting the relative frequencies in the validation year equally. This index would be expected to identify the method(s) producing the smallest average predictive error across all cells of the table. The first criterion measure was:

$$D_1 = \left[\frac{\sum_{i=1}^{m} \sum_{j=1}^{n} (P_{ij} - \hat{\pi}_{ij})^2}{mn} \right]^{1/2}$$

where D_1 is the root mean-squared error in estimating relative frequencies, P_{ij} is the observed relative frequency of success in the cross-validation year at interval i of predictor 1 and, interval j of predictor 2, $\hat{\pi}_{ij}$ is the relative frequency predicted by the model, m is the number of intervals on variable 1, and n is the number of intervals on variable 2. If no observations resulted for a cell in the table in the validation year, then that term was not included and the denominator was reduced by one.

The second index associated greater seriousness with prediction error for cells of the table containing a larger number of observation. This index would be expected to identify the method(s) producing the smallest average prediction error for the individual observations in the validation year. The second criterion measure was:

$$D_2 = \left[\frac{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} (P_{ij} - \hat{\pi}_{ij})^2}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij}} \right]^{1/2}$$

Where D_2 is a weighted measure of the error in estimating frequencies, f_{ij} is the number of observations in the ijth cell of the validation year table, and the other symbols are as defined above. Interpretations of the relative validities of the methods of constructing double-entry expectancy tables were based on the values of D_1 and D_2.

Analysis of Data

A four factor mixed analysis of variance procedure was applied to each of the criterion measures (Myers, 1972). Factors in the analysis were size of class (actually, number of first term students with ACT data-five levels), expectancy table construction method (seven levels), GPA value (two levels), and validation year (two levels). Size of class was considered a "between" effect and all other factors were treated as "within" effects. The unit of analysis, school within types, was considered to be a "random" effect and all other factors were considered "fixed" effects.

Results

The results from the analysis of variance for the criterion measures are presented in Tables 1 and 2. All tests (including post-hoc tests) were conducted at the .01 level of significance. Unbiased estimates of the variance components (Myers, 1972) are also provided.[4]

Insert Tables 1 and 2 about here

The appropriate means for the main effects and interactions which surpassed the .01 level of significance are provided in Tables 3 and 4. The Tukey (1953) critical differences for comparisons between any pair of means are provided in Table 5. The following discussion is based on these pairwise comparisons. Since the interactions involving methods were essentially ordinal, the main effects are discussed prior to the interactions.

[4] Since there was only one observation per cell, no direct estimate of σ_e^2 was possible. It was assumed that the variance component for the highest order pooled interaction ($\sigma^2_{SMYG/Z}$) was equal to zero. In this way σ_e^2 was estimated by $MS_{SMYG/Z}$. Also, after the variance components were computed, the estimates that were negative were replaced by zero. Thus, these estimates are no longer unbiased.

Insert Tables 3, 4, and 5 about here

Main Effects

The relative frequencies estimated by any of the six smoothing methods were more accurate than the estimates provided using the observed relative frequencies (Method 1). In addition, Method 3a was more accurate than Method 2a for the D_1 criterion. Methods 3a, 3b, 4a and 4b were more accurate than Methods 2a and 2b for the D_2 criterion.

As would be expected, the accuracy of all methods increased as class size increased and the methods were generally more accurate for validation year 1 than for validation year 2. In addition, all methods were more accurate for the "B or better" level than for the "C or better" level.

Interaction of Methods and Class Size

The general tendency for accuracy to increase for all methods as class size increased held for both criterion measures. For the D_1 criterion, all of the smoothing methods were more accurate than was Method 1 across all class size levels. No substantial differences among the remaining methods were noted at any of the class size levels.

For the two smallest class size intervals, the relative frequencies estimated by any of the smoothing methods were more accurate than the estimates produced by Method 1 using the D_2 criterion measure. However, as class size increased, the relative accuracy of Method 2a (and, to some extent, Method 2b) decreased with respect to the remaining methods, including Method 1. No substantial differences among the means of Methods 3a, 3b, 4a, and 4b were observed at any of the class size levels. Based on these results, Methods 3a, 3b, 4a, and 4b provided more accurate estimated relative frequencies at a wider range of class size levels than did the remaining methods.

Interactions of Method with GPA Level, and/or Validation Year.

Based on the D_1 index, the smoothing methods resulted in more accurate estimation of validation year relative frequencies than did Method 1 at all GPA level and validation year combinations. The accuracy of the smoothing methods did not differ substantially for the tables computed using the "B or better" level of success. Method 3a was, however, superior to the other methods for the validation year 1 and "C or better" GPA combination. For the validation year 2 and "C or better" GPA combination, Method 3a was the most accurate while Methods 2a and 2b tended to be the least accurate of all the methods except Method 1.

The smoothing methods were also more accurate than Method 1 at both GPA levels according to the D_2 criterion measure. The accuracy of the smoothing methods did not differ substantially at the "B or better" GPA level. Methods 3a, 3b, 4a, and 4b were more accurate than Methods 2a and 2b at the "C or better" level. Thus, Method 3a was at least as accurate as the other methods for all combinations of class size, GPA, and validation year levels.

Accuracy Gained by Using Smoothing Methods.

In order to reflect the degree to which each smoothing method improved on the predictions from observed relative frequencies (Method 1), the average percentage gain in accuracy by using each of the smoothing methods is presented in Table 6. These values indicate that, overall, the use of each smoothing method resulted in a gain in predictive accuracy of about 25% for the D_1 index. The gains in predictive accuracy based on the D_2 index were about 20% for Methods 3a, 3b, 4a, 4b but only 10% for Methods 2a and 2b. Similar values were computed for the index (comparable to our D_1 index) used by Perrin and Whitney (1976); their smoothing methods for single-entry tables resulted in a gain in predictive accuracy of from 25% to 32%.

Insert Table 6 about here

Discussion

The use of smoothing methods resulted in a practically significant increase in predictive accuracy in both this study and that of Perrin and Whitney (1976). In the present study,the relative size of the estimated variance components for methods and for the interaction of methods with the other variables suggest that methods contributed substantially to the total variance of the model. The present study and that of Perrin (1974) suggest that Methods 2a and 2b result in a substantial increase in average predictive accuracy across cells (as reflected by the D_1 index). However, this increase in accuracy was surpassed by some of the other methods. These studies also suggest that Methods 2a and 2b result in only a minimal increase in predictive accuracy for individual observations (as reflected by the D_2 index). Both of these types of accuracy are desirable in most educational situations. Since Methods 2a and 2b did not provide a substantial gain in the latter type of accuracy, these commonly used regression methods (Schraeder, 1965) are less appropriate than are other methods.

Of the construction methods studied, Method 3a (linear regression on observed relative frequencies) was at least as accurate as any of the other methods for both criterion measures at all combinations of GPA, class size, and validation year levels. Method 3a would be generally preferred for double-entry tables like those studied.

The gain in predictive accuracy from using either extension of the isotonic regression method was nearly as great (overall and for most GPA, class size, and validation year levels) as was that from using Method 3a. Because Method 4b required the computation of the regression estimates used in Method 3a and subsequently employed the isotonic regression method for single-entry tables, Method 4b is relatively complex. Sc even though the use of Method 4b produced estimates nearly as accurate as those produced by Method 3a, the complexity of this method does not suggest its use. Because the computations involved in using Method 4a are relatively simple (in fact, clerical personnel could easily use this method aided only by a pocket calculator) this method would be preferred when access to a computer is not available. Method 4a would also be preferred when the more stringent statistical assumptions of Method 3a are not likely to be met.

Any of the smoothing methods studied remove logical contradictions between observed data and beliefs regarding the actual relationship among predictor and success measures. Smoothing would also be expected to increase the accuracy of prediction when constructing tables to be used for admission, guidance, or planning purposes. If such methods are used, however, the traditional regression methods (Methods 2a and 2b) can not be expected to be as accurate as the other methods studied. Methods 3a and 4a are especially recommended for this purpose.

References

Anastasi, A. A. Psychological Testing. 4th Ed. New York: MacMillan Publishing Company, Inc., 1976.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., & Silverman, E. An Empirical distribution function for sampling with incomplete information. Annals of Mathematical Statistics, 1955, 26, 641-647.

Myers, J. L. Fundamentals of experimental design. Boston: Allyn and Bacon, Inc., 1972.

Perrin, D. W. A comparison of methods for smoothing expectancy tables as applied to the prediction of success in College (Doctoral dissertation University of Iowa, 1974). Ann Arbor, Michigan: University Microfilms, 1974, No. 75-13805.

Perrin, D. W., & Whitney, D. R. Methods for smoothing expectancy tables applied to the prediction of success in college. Journal of Educational Measurement, 1976, 13, 223-231.

Schraeder, W. B. A taxonomy of expectancy tables. Journal of Educational Measurement, 1965, 2, 29-35.

Tukey, J. W. The problems of multiple comparisons. Ditto, Princeton University, 1953, pp. 396.

Wesman, A. G. Double-entry expectancy tables. Test Service Bulletin of the Psychological Corporation, 1966.

Predictor 1

Predictor 2	Low			High
High	10 .100	20 .600	80 .500	40 .800
	10 .300	80 .200	5 .200	20 .700
Low	80 .100	20 .400	5 .600	0 .00

Figure 1a. Unsmoothed data constructed by Method 1.

Predictor 1

Predictor 2	Low			High
High	10 .100	50 .520	50 .520	40 .800
	31.7 .211	31.7 .211	31.7 .211	20 .700
Low	80 .100	20 .400	2.5 .600	2.5 .600

Figure 1b. Rows adjusted.

Predictor 1

Predictor 2	Low			High
High	20.8 .184	50 .520	50 .520	40 .800
	20.8 .184	25.8 .284	17.1 .239	20 .700
Low	80 .100	25.8 .284	17.1 .239	2.3 .600

Figure 1c. Rows adjusted first and then columns.

Predictor 1

Predictor 2	Low			High
High	10 .200	20 .600	80 .500	40 .800
	10 .200	50 .240	5 .400	20 .700
Low	80 .100	50 .240	5 .400	0 .00

Figure 1d. Columns adjusted.

Predictor 1

Predictor 2	Low			High
High	10 .200	50 .520	50 .520	40 .800
	10 .200	50 .240	5 .400	20 .700
Low	80 .100	50 .240	2.5 .400	2.5 .400

Figure 1e. Columns adjusted first and then rows.

Predictor 1

Predictor 2				
High	15.4 .192	50 .520	50 .520	40 .800
	15.4 .192	37.9 .262	11.0 .320	20 .700
Low	80 .100	37.9 .262	9.8 .320	2.5 .500

Figure 1f. Averaging of two double-adjusted tables.

Figure 1. Illustration of smoothing method 4a.

TABLE 1

ANOVA SUMMARY TABLE FOR D_1 INDEX

Source	df	Mean Square	Mean Square Ratio	Estimated Variance Components
Between Schools	84			
Z(Class Size)	4	0.57151	17.13*	.000904
S(School)/Z(Size)	80	0.03336		.001010
Within Schools	2295			
M(Method)	6	0.28053	111.75*	.000701
ZM(Size X Method)	24	0.01327	5.28*	.000108
SM/Z(School X Method/Size)	480	0.00251		.000376
Y(Validation Year)	1	0.16749	9.88*	.000063
ZY(Size X Year)	4	0.00678	0.40	.000000***
SY/Z(School X Year/Size)	80	0.01695		.000548
G(GPA Level)	1	0.64152	24.77*	.000259
ZG(Size X GPA Level)	4	0.02227	0.86	.000000***
SG/Z(School X GPA Level/Size)	80	0.02590		.000849
MY(Method X Year)	6	0.00040	0.46	.000000***
ZMY(Size X Method X Year)	24	0.00064	0.74	.000000***
SMY/Z(School X Method X Year/Size)	480	0.00087		.000046
MG(Method X GPA Level)	6	0.04146	23.99*	.000100
SMG(Size X Method X GPA Level)	24	0.00187	1.08	.000001
SMG/Z(School X Method X GPA Level/Size)	480	0.00173		.000218
YG(Year X GPA Level)	1	0.00512	0.32	.000000***
ZYG(Size X Year X GPA Level)	4	0.03864	2.39	.000038
SYG/Z(School X Year X GPA Level/Size)	80	0.01614		.000521
MYG(Method X Year X GPA Level)	6	0.00242	3.75*	.000005
ZMYG(Size X Method X Year X GPA Level)	24	0.00030	0.46	.000000***
SMYG/Z(School X Method X Year X GPA Level/Size)	480	0.00065		.000000**
Error				.000645
Total	2379			.006371

*p<.01

**Is zero by assumption

***Negative estimates replaced by zero

TABLE 2

ANOVA SUMMARY TABLE FOR D_2 INDEX

	df	Mean Square	Mean Square Ratio	Estimated Variance Components
Between Schools	84			
Z(Class Size)	4	0.44860	15.85*	.000706
S(School)/Z(Size)	80	0.02830		.000946
Within Schools	2295			
M(Method)	6	0.05326	43.75*	.000131
ZM(Size X Method)	24	0.00886	7.28*	.000077
SM/Z(School X Method/Size)	480	0.00122		.000213
Y(Validation Year)	1	0.07604	9.39*	.000029***
ZY(Size X Year)	4	0.00292	0.36	.000000
SY/Z(School X Year/Size)	80	0.00810		.000267
G(GPA Level)	1	0.12149	11.15*	.000046***
ZG(Size X GPA Level)	4	0.00182	0.17	.000000
SG/Z(School X GPA Level/Size)	80	0.01089		.000360
MY(Method X Year)	6	0.00018	0.89	.000000***
ZMY(Size X Method X Year)	24	0.00011	0.54	.000000***
SMY/Z(School X Method X Year/Size)	480	0.00020		.000007
MG(Method X GPA Level)	6	0.01052	11.35*	.000024
SMG(Size X Method X GPA Level)	24	0.00122	1.32	.000003
SMG/Z(School X Method X GPA Level/Size)	480	0.00093		.000155
YG(Year X GPA Level)	1	0.01550	5.18*	.000005
ZYG(Size X Year X GPA Level)	4	0.01142	3.81*	.000014
SYG/Z(School X Year X GPA Level/Size)	80	0.00230		.000072
MYG(Method X Year X GPA Level)	6	0.00026	1.58	.000000
ZMYG(Size X Method X Year X GPA Level)	24	0.00017	1.06	.000000**
SMYG/Z(School X Method X Year X GPA Level/Size)	480	0.00016		.000000
Error				.000162
Total	2379			.003241

* $p < .01$

** Is zero by assumption

*** Negative estimates replaced by zero

80	10	10	20	80	20	5	5	80	0	20	40
.100	.300	.100	.400	.200	.600	.600	.200	.500	.000	.700	.800

Figure 2a. Ordering of cells, with regression weight: $b_{row} = .0367$, $b_{col} = .1893$, and $a = .1398$.

80	10	10	50	50	22	22	22	22	22	20	40
.100	.200	.200	.240	.240	.509	.509	.509	.509	.509	.700	.800

Figure 2b. Smoothed values.

Predictor 1

		Low			High
	High	10 .200	22 .509	22 .509	40 .800
Predictor 2		10 .200	50 .240	22 .509	20 .700
	Low	80 .100	50 .240	22 .509	22 .509

Figure 2c. Reconstructed, smoothed double-entry table.

Figure 2. Illustration of smoothing method 4b.

TABLE 3

MEANS ON CRITERION D_1

Smoothing Method	GPA Level: B or Better, Validation Year: Year 1	GPA Level: B or Better, Validation Year: Year 2	GPA Level: C or Better, Validation Year: Year 1	GPA Level: C or Better, Validation Year: Year 2	Class Size: 98-159	Class Size: 160-219	Class Size: 220-339	Class Size: 340-1149	Class Size: 1150-4193	All
1	.2452	.2712	.3358	.3360	.3510	.3558	.2995	.2764	.2027	.2971
2a	.2000	.2187	.2404	.2530	.2323	.2557	.2397	.2242	.1883	.2280
2b	.1967	.2204	.2340	.2493	.2461	.2559	.2345	.2130	.1759	.2251
3a	.2020	.2168	.2084	.2281	.2279	.2375	.2306	.2053	.1678	.2138
3b	.2062	.2218	.2200	.2408	.2424	.2487	.2327	.2071	.1801	.2222
4a	.1971	.2202	.2294	.2426	.2457	.2503	.2366	.2083	.1707	.2223
4b	.1978	.2137	.2276	.2424	.2458	.2480	.2386	.2061	.1634	.2204
All Methods	.2064	.2261	.2422	.2560	.2559	.2646	.2446	.2200	.1784	

TABLE 4

MEANS* ON CRITERION D_2

Smoothing Method	GPA Level: B or Better	GPA Level: C or Better	Class Size: 98-159	Class Size: 160-219	Class Size: 220-339	Class Size: 340-1149	Class Size: 1150-4193	All
1	.1655	.1937	.2284	.2240	.1897	.1512	.1041	.1796
2a	.1495	.1725	.1689	.1824	.1721	.1506	.1311	.1610
2b	.1496	.1739	.1780	.1842	.1725	.1466	.1274	.1617
3a	.1442	.1430	.1604	.1657	.1570	.1317	.1032	.1436
3b	.1462	.1500	.1685	.1729	.1567	.1284	.1139	.1481
4a	.1442	.1548	.1761	.1785	.1611	.1332	.0986	.1495
4b	.1422	.1535	.1767	.1749	.1609	.1307	.0961	.1479
All Methods	.1488	.1631	.1796	.1832	.1671	.1390	.1106	

*In addition, the overall means for validation years 1 and 2 were .1503 and .1616 respectively.

TABLE 5

TUKEY CRITICAL DIFFERENCES FOR COMPARISONS

OF PAIRS OF MEANS CONTAINED IN

SIGNIFICANT MAIN EFFECTS OR INTERACTIONS

Source	Criterion Measure	
	D_1	D_2
Z	.0398	.0367
M	.0133	.0093
ZM	.0370	.0258
MG	.0174	.0127
ZYG	---	.0251
MYG	.0163	---

TABLE 6

MEAN PERCENTAGE GAIN IN PREDICTIVE ACCURACY USING SMOOTHING METHODS COMPARED TO THAT OF METHOD 1

Method	Criterion Index	
	D_1	D_2
1	--	--
2a	23	10
2b	24	10
3a	28	20
3b	25	18
4a	25	17
4b	26	18