DOCUMENT RESUME

ED 109 191 TM 004 641

AUTHOR Perrin, David W.; Whitney, Douglas R.
TITLE Methods for Smoothing Expectancy Tables Applied to the Prediction of Success in College. Research Report No. 79.
REPORT NO RR-79
PUB DATE [Apr 75]
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D. C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *College Freshmen; Comparative Analysis; *Expectancy Tables; *Grade Point Average; *Grade Prediction; Higher Education; Prediction; *Statistical Analysis
IDENTIFIERS Smoothing Methods

ABSTRACT

Six methods for smoothing expectancy tables were compared using data for entering students at 86 colleges and universities. Linear regression analyses were applied to ACT scores and high school grades to obtain predicted first term grade point averages (FGPA's) for students entering each institution in 1969-70. Expectancy tables were constructed for each institution using the relative frequency with which students in each predicted FGPA interval obtained FGPA's of "B or better" and "C or better". The methods used to smooth the 1969-70 expectancy tables were compared on how accurately the smoothed tables corresponded to the 1971-72 and 1972-73 observed data. The smoothed tables were substantially more accurate than the 1969-70 observed data on which they were based. Two maximum likelihood methods yielded the most accurate tables. Estimated relative frequencies were more accurate for the "B or better" level than the "C or better" level. (Author)

Session 7.04

ED109191

METHODS FOR SMOOTHING EXPECTANCY TABLES APPLIED TO THE PREDICTION OF SUCCESS IN COLLEGE

David W. Perrin
Douglas R. Whitney

Research Report No. 79

Presented at the Annual Meeting
American Educational Research Association
Washington, D.C.
April, 1975

Evaluation and Examination Service
University of Iowa
Iowa City, Iowa

April, 1975

TM 004 641

METHODS FOR SMOOTHING EXPECTANCY TABLES

APPLIED TO THE PREDICTION OF SUCCESS IN COLLEGE

ABSTRACT

Six methods for smoothing expectancy tables were compared using data for entering students at 86 colleges and universities. Linear regression analyses were applied to ACT scores and high school grades to obtain predicted first term grade point averages (FGPA's) for students entering each institution in 1969-70. Expectancy tables were constructed for each institution using the relative frequency with which students in each predicted FGPA interval obtained FGPA's of "B or better" and "C or better". The methods used to smooth the 1969-70 expectancy tables were compared on how accurately the smoothed tables corresponded to the 1971-72 and 1972-73 observed data. The smoothed tables were substantially more accurate than the 1969-70 observed data on which they were based. Two maximum likelihood methods yielded the most accurate tables. Estimated relative frequencies were more accurate for the "B or better" level than the "C or better" level.

METHODS FOR SMOOTHING EXPECTANCY TABLES

APPLIED TO THE PREDICTION OF SUCCESS IN COLLEGE

In many situations involving the selection of qualified applicants, the measure of primary interest is not a predicted score on some criterion, but the probability that an individual will obtain a criterion score equal to or greater than some pre-defined level. A college or university, for example, may be more interested in estimating the number of applicants with a given predicted grade-point average (PGPA) which may be expected to achieve an overall grade of C or better than in knowing each applicant's PGPA.

Although an expectancy table may be formed irrespective of the relationship between actual and predicted grade point averages, the concern in this paper is with the large class of bivariate relationships for which the first term grade point average (FGPA) and the predicted grade point average (PGPA) have a monotonic nondecreasing relationship. That is, the expected value of FGPA for any PGPA value is not less than the expected value of FGPA for any higher PGPA. This allows for a wide variety of bivariate distributions including all positive linear and certain nonlinear relationships. Although monotonic nondecreasing relationships may not occur in special circumstances involving students with extremely low or high PGPA's (Novick and Jackson, 1974), this assumption seems generally justifiable throughout the entire range of grade point averages for most colleges.

Note: The authors are grateful to Dr. E. James Maxey and the American College Testing Program for their support and assistance.

To construct an expectancy table, two measures must be available for some large number of observations (the experience pool) representative of the population for which the prediction is to be made. Typically, when predicting success in college, a previous freshman class is used as the experience pool. If the number of observations in the experience pool is sufficiently large, the expectancy table may be constructed directly from the observed data. One advantage of this method, as compared to others discussed below, is that it is relatively easy and requires only that individuals in the experience pool be assigned to the appropriate cell of the table representing a combination of PGPA and FGPA values. The cell frequencies are usually converted to relative frequencies or cumulative relative frequencies which are used to make predictions.

There are two potential limitations in this method of constructing expectancy tables. The first is that there are no absolute rules to determine what constitutes a "sufficiently large" number of observations in the experience pool. The second is that reversals will often occur. That is, the data reported in the table will be inconsistent with the assumptions about the relationship between FGPA and PGPA. The usual explanation for these reversals is that they are due to the sampling fluctuations of cell estimates based on few observations.

Methods for Smoothing Expectancy Tables

Expectancy tables can be derived using statistical theory rather than basing the table solely on the observed frequencies. The advantage of using a theoretical basis to smooth an expectancy table constructed from observed data is that it is possible to base estimated relative frequencies not only upon the observations in each cell but also upon the observations in adjacent cells. Such a procedure would be expected to have an effect similar to increasing the number of observations in a particular cell and would remove reversals resulting in cell entries which more accurately describe the underlying relationship.

One limitation in using a theoretical approach is that somewhat more statistical sophistication is required. A second limitation is that actual characteristics of the relationship between the variables would be lost if tables were derived using an <u>incorrect</u> theoretical basis (Schraeder, 1965). In case of college grade point averages, however, this appears unlikely.

The six smoothing methods investigated included a linear regression procedure similar to that used by the American College Testing (ACT) Program (ACT, 1971), an isotonic regression method attributed to Brunk (Ayer, Brunk, Ewing, Reid and Silverman, 1955), and four methods previously used primarily in bioassay problems. Two of the bioassay methods were iterative maximum likelihood procedures (Finney, 1971); one requiring a normal and the other a logistic transformation of the observed relative frequencies. The other two bioassay methods were noniterative minimum χ^2 procedures. One required a normal transformation (Berkson, 1955) and the other a logistic transformation (Berkson, 1944) of the observed relative frequencies. Perrin (1974, 1975) provides a complete description of the methods as applied in this study.

The linear regression method assumes a linear relationship between PGPA and FGPA and normal, homoscedastic conditional distributions of FGPA at each PGPA level. Least squares techniques provide the estimated means and standard deviations of the conditional FGPA distributions. These statistics are used in conjunction with the FGPA level of interest (e.g., 2.0 for C or better) to compute a standard deviate for each PGPA level. These deviates are referred to normal probability tables to estimate the relative frequencies of attaining each FGPA.

The isotonic regression method requires only the observed relative frequencies exceeding the chosen FGPA level at each PGPA level. When there are no reversals, these observations are the estimated relative frequencies. When the observed relative frequencies for two adjacent PGPA levels are reversed, a weighted average of the two observations is computed and substituted for each relative frequency. This weighted averaging procedure is continued until all reversals are removed. The numbers in this series, some of which are observed relative frequencies and some of which are weighted averages, are the estimated relative frequencies resulting from the isotonic regression method. Additionally, it is necessary to extrapolate or interpolate to obtain estimates for PGPA levels for which there were no observations in the base year.

The maximum likelihood methods used either an inverse normal or logistic transformation on the observed relative frequencies. The resulting values are iteratively regressed on PGPA until the difference between the sum of the differences between regression estimates for two successive iterations is sufficiently small (for this study 10^{-6}). Using the inverse of the original transformation, the regression estimates from the final iteration are transformed into the estimated relative frequencies.

The minimum χ^2 methods also depend on tranformations of the observed relative frequencies. However, the use of additional approximations suggested by Berkson (1944, 1955) eliminates the need for iteration. Using weights which are determined by the form of the transformation, the transformed values are regressed on PGPA. The inverse of the original transformation on the values resulting from the regression yields the estimated relative frequencies.

Procedures

The data for this study were the records of entering freshman students at a sample of institutions which had participated in one of the ACT Research Services for the years 1969-70, 1971-72 and 1972-73. (Henceforth, 1969-70 will be referred to as the base year and 1971-72 and 1972-73 as, respectively, validation year one and two.) The academic year 1970-71 was not included because, in the standard ACT procedures, predictions based on the base year data would not have been available for most freshman entering in 1970.

From a list of all institutions that had participated in either Research Service program for all three years, a stratified random sample of 86 institutions was drawn. The sample included 22 two-year institutions, 21 institutions granting only the bachelor's degree, 21 institutions granting at least one graduate degree other than the doctorate, and 22 institutions granting the doctorate (types 1, 2, 3 and 4 respectively). For each student completing the first term in one of these schools during the base year or either of the validation years, ACT provided the student's four ACT Assessment subtest scores, average of four self-reported high school grades, first term grade-point average (FGPA) at the institution, and sex.

Prediction of FGPA

Following the standard ACT procedure, regression analyses were conducted on the base year data separately by sex if both the numbers of men and women completing the first term at an institution exceeded 100. If either the number of men or women was less than 100, a single regression analysis was conducted for that institution. The separate regressions by sex were used whenever possible because the multiple correlation between FGPA and the five independent measures is typically higher for women than for men (ACT, 1971).

The regression analysis used five independent variables (four ACT Assessment subtest scores and the average of the self-reported high school grades) and was comparable to that used by ACT in its Basic Research Service. The regression weights calculated by this procedure were used to compute a PGPA for each student in the base and validation years. Similar analyses were conducted using a nonlinear regression procedure (Perrin, 1974) but are not reported since the improvement in predictive accuracy was slight and the conclusions concerning methods were similar to those reported here.

Constructing and Smoothing Expectancy Tables

Following the computation of PGPA, three expectancy tables (one for each year) were constructed for each institution. The base year table for each school was smoothed by application of the methods independently to each column of the table. Thus, for each base year table, six different "smoothed" tables were constructed. Each of these six tables contained relative frequencies, estimated by one of the methods, of obtaining a FPGA equal to or greater than C and B.

Criterion Measures

The criterion of interest was 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. The criterion measure for this study was

$$\left[\sum_{i=1}^{k} k^{-1}(p_i - \hat{\pi}_i)^2\right]^{1/2} \qquad (1)$$

where p_i is the validation year observed relative frequency, $\hat{\pi}_i$ is the estimated relative frequency resulting from the application of one of the smoothing methods, and k is the number of PGPA levels for which observations were available in the validation year. That is, the criterion measure was the positive square root of the mean squared error for k levels of PGPA. A small value for the criterion measure identified the method(s) producing the smallest estimation errors across all PGPA levels. Similar results, using two other criterion measures, are reported by Perrin (1974).

Analysis of Data

A four factor mixed analysis of variance (ANOVA) procedure was applied to the criterion measure (Lindquist, 1953). Factors in the analysis were institutional type (four levels), expectancy table construction method (seven levels), FGPA value (two levels), and validation year (two levels). Institutional type was considered a "between" effect and all other factors were treated as "within" effects. The unit of analysis was school within types. One construction method was the observed relative frequencies for the base year; the other six methods were those described above. Each of the methods yielded a criterion measure for two levels of FGPA ("C or better" and "B or better") for each validation year for each school.

In addition to the effects usually tested in such an analysis, three questions were of special interest. First, did the base year observed relative frequencies provide estimates which were as accurate as those provided by the six smoothing methods? Second, was there a difference in accuracy between the linear regression method presently used by ACT and the other five smoothing methods? Third, was there a difference in accuracy between the bioassay methods and the simpler isotonic regression method? Additional analyses were conducted to answer each of these questions using pre-planned contrasts.

Smoothing Methods

The results of the analyses conducted to evaluate the effects of the smoothing methods are summarized in Table 1. The mean square ratio for the main effect of methods exceeded the value of the 99th percentile in the appropriate F-distribution. Additionally, the value of the statistic comparing the criterion measure mean for base year relative frequencies with the average of the criterion measure means for the six smoothing methods exceeded the appropriate 99th percentile value. The criterion measure means for each method appear in Table 2. The results indicated that the smoothing methods provided more accurate estimates of the relative frequencies observed in the validation years than did unsmoothed base year relative frequencies. Thus, it is recommended that expectancy tables used to estimate relative frequencies of success in college should be smoothed rather than simply based on observed relative frequencies.

Insert Table 1 about here

Insert Table 2 about here

The comparison of the average criterion measure for the linear regression method with the average criterion measure for the other five smoothing methods also yielded a statistic greater than the appropriate 99th percentile value. This suggests that the method presently used by ACT is less accurate than the average accuracy of the other smoothing methods. Generally, conclusions and recommendations based on main effects would not be made when interactions involving that factor were observed. In this case, however, examination of simple effects at each level of the interacting factors did <u>not</u> contradict the conclusions drawn from the main effects for methods. (That is, the interaction was not disordinal.) Therefore, it is recommended that one of the other smoothing methods be used in preference to the linear regression method.

Which of the other smoothing methods should be adopted was not entirely clear. However, the maximum likelihood methods provided more accurate estimates of the validation years' relative frequencies than did the other smoothing methods. For institutions without the computational facilities or inclination to apply the maximum likelihood methods, the simpler isotonic regression method could be expected to provide estimates nearly as accurate as the more complex methods.

FGPA Level

Estimation of relative frequencies was more accurate for the "B or better" level than for the "C or better" level. The differences between the mean criterion measures for the "B or better" level and those for the "C or better" level exceeded the appropriate Scheffé 1% critical difference (Scheffé, 1959).

Validation Year

All criterion measure means for validation year one were lower than those for validation year two. (That is, accuracy declined over time.) However, the magnitude of these differences did not exceed the appropriate Scheffé critical difference. The criterion measure means for the FGPA level by validation year suggested that most of the difference in predictive accuracy between validation years was due to a loss of accuracy at the "B or better" level. Thus, although predictive accuracy was generally lower for the "C or better" level, accuracy at that level declined less between validation years one and two than for the "B or better" level.

Institutional Type

The results for the main effect of institutional type and for the simple effects of institutional type for each method indicated that the most accurate estimation of relative frequencies occurred at Type 4 and Type 1 institutions. The mean and median sizes of the entering freshman classes for each of the institutional types were inversely related to the accuracy indexes. These were considered likely to have contributed to the differences in estimation accuracy among the institutional types. That is, the smoothing methods and base year observed frequencies yielded the more accurate estimation for institutions with the largest entering enrollments.

Conclusion

Because of the widespread use of expectancy tables by colleges and universities, any improvement in accuracy would be beneficial. Of special importance are the uses of expectancy tables for guidance, admissions, and planning purposes. Application of the best smoothing method appears likely to result in a substantial improvement in the accuracy of the information used in individual decisions. Additionally, application of the best smoothing method would improve estimates of the number of students likely to be dropped for scholastic reasons. In any of these situations, more accurate information should result in better decisions.

References

American College Testing Program. Your college freshmen. Iowa City: ACT, 1971.

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.

Berkson, J. Application of the logistic function to bioassay. Journal of the American Statistical Association, 1944, 39, 357-365.

Berkson, J. Estimate of the integrated normal curve by minimum normit chi-square with particular reference to bioassay. Journal of the American Statistical Association, 1955, 50, 529-549.

Finney, D. J. Probit analysis. Cambridge: Cambridge University Press, 1971

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin, 1953.

Novick, M. R., & Jackson, P. H. Statistical methods for educational and psychological research. New York: McGraw Hill, 1974.

Perrin, D. W. A comparison of methods for smoothing expectancy tables as applied to the prediction of success in college. Unpublished doctoral dissertation, University of Iowa, 1974.

Perrin, D. W. Description and illustration of methods for smoothing expectancy tables. Research Report No. 80. Iowa City: Evaluation and Examination Service, 1975.

Scheffé, H. The analysis of variance. New York: John Wiley, 1959.

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

TABLE 1

ANALYSIS OF VARIANCE

SUMMARY TABLE

Source	df	Mean Squares	Mean Square Ratios
Between Schools	85		
D(Institutional Type)	3	0.3492	6.79*
Error	82	0.0514	
Within Schools	2322		
A(Method)	6	0.3439	154.02*
AD(Method x Type)	18	0.0085	3.79*
Error	492	0.0022	
B(FGPA Level)	1	1.3277	65.20*
BD(FGPA Level x Type)	3	0.0568	2.79*
Error	82	0.0204	
C(Validation Year)	1	0.0274	1.85
CD(Year x Type)	3	0.0092	0.62
Error	82	0.0148	
AB(Method x FGPA Level)	6	0.0209	25.78*
ABD(Method x FGPA Level x Type)	18	0.0009	1.13
Error	492	0.0008	
AC(Method x Year)	6	0.0005	0.64
ACD(Method x Year x Type)	18	0.0005	0.73
Error	492	0.0007	
BC(FGPA Level x Year)	1	0.0864	7.11*
BCD(FGPA Level x Year x Type)	3	0.0214	1.76
Error	82	0.0121	
ABC(Method x FGPA Level x Type)	6	0.0003	1.03
ABCD(Method x FPGA Level x Year x Type)	18	0.0004	1.48
Error	492	0.0003	
Total	2407		

*p<.01.

TABLE 2

CRITERION MEASURE MEANS

Smoothing Method	FGPA Level		Institutional Type				All
	B or Better	C or Better	1	2	3	4	
Linear Regression	.179	.234	.204	.211	.215	.188	.206
Isotonic Regression	.165	.221	.196	.213	.206	.158	.193
Maximum Likelihood (Normal)	.160	.214	.188	.203	.199	.158	.187
Maximum Likelihood (Logistic)	.157	.215	.189	.202	.199	.156	.186
Minimum χ^2 (Normal)	.180	.208	.191	.210	.214	.162	.194
Minimum χ^2 (Logistic)	.184	.205	.190	.211	.213	.164	.194
Base Year Relative Frequencies	.247	.304	.259	.319	.303	.223	.275
All Methods	.182	.229	.202	.225	.221	.173	