JOURNAL OF THE TENNESSEE ACADEMY OF SCIENCE
VOLUME 38, NUMBER 1, JANUARY, 1963

A MULTIVARIATE ANALYSIS OF ADMISSIONS CRITERIA AT VANDERBILT ENGINEERING SCHOOL

RAEBURN STILES
Associate Professor of Applied Mathematics
H. E. WILLIAMS
Assistant Professor of Applied Mathematics

It has often been noted in recent years that the number of young people applying for admission to our colleges and universities is increasing rapidly. Present predictions indicate a continuing increase in applications for the next ten years at least, and perhaps longer. This presents a problem of expanding dimensions to college admissions officers – particularly in many private institutions where enrollment is limited. The problem stated briefly is that of how best to select from among the applicants students that have the best chances of doing successful college work.

Selection of students for engineering schools appears to be an especially complex task since such well-known predictions of college success as IQ scores and College Entrance Examination Board Scholastic Aptitude Tests, although quite useful, do not seem to have as high a degree of demonstrated validity for engineering as for a liberal arts program of studies. It was in this context that the authors decided to analyze statistically those variables obtainable from applications for admission to the Vanderbilt School of Engineering and determine the most efficient, weighted combinations of variables for predicting scholastic success in the school. Although the results of this study cannot be assumed to be strictly applicable to other engineering schools, it was felt that the information obtained would be of general interest to engineering school admission programs, and that presumably some of the ideas presented will be valid in other situations.

The entering freshman class of 1959 was selected for the analysis. Eight predictor variables, readily obtainable from applications for admission, were utilized in the study. The criterion of success used was the individual student's grade point average in the engineering school. At the time of the study, grades through the freshman and sophomore years were available for those students remaining in school. Since experience has indicated that the first two years are the most critical for evaluating successful students progress toward the B.E. degree, the 1959 entering class was deemed appropriate for the study. Both those students remaining in school and those no longer in school were included in the analysis in order to provide the fullest possible range of variability. This gave a total sample size of 163 cases. The variables utilized are as follows:

I. $CEEB_V$ – College Entrance Examination Board Scholastic Aptitude Tests, Verbal Battery Score.

II. $CEEB_M$ – College Entrance Examination Board Scholastic Aptitude Tests, Mathematics Battery Score.

III. $CEEB_T$ – College Entrance Examination Board Scholastic Aptitude Tests – total score – $CEEB_V + CEEB_M$.

IV. IQ – Intelligence Quotient, high school age levels, based on a mean of 100 and standard deviation of 15.

V. HS_{GP} – High school gradepoint average. The basis for quantification of this variable was A = 3 points; B = 2 points; C = 1 point; D = 0 points. This average had been computed by the admissions office and was recorded on the application.

VI. HS_{RA} – High school percentile rank in class.

VII. Age – The age of the applicant as of September 1959.

VIII. HS_{RE} – High school recommendation. The high school principal or senior counselor is required to write a recommendation for each applicant. This variable was difficult to quantify, but the procedure used was to set up a –1, 0, +1 trichotomous rating scale as follows:

+1=student is given a positive recommendation by the high school.

0=student is given a neutral recommendation or no recommendation.

–1=student is given a negative recommendation.

(It might be well to note that a suggestion for possible improvement of this variable is included later in the article.)

IX. GPA – Grade point average in the engineering school. This is the criterion variable. The quantification scheme was the same as for HS_{PG}. Presumably, a slightly more efficient prediction would have been obtained by scoring a grade of F as —1 to distinguish an F grade from a D. However, a GPA computed in this manner was not available from the machine tabulated grade record sheets used in the school.

The collected data was fed into an electronic computer programmed to compute the intercorrelations of the 9 variables, the means ($\bar{x}_i$), and the standard deviations (s_i) of the variables. These are given in Table 1. The intercorrelation matrix was then solved by the Fisher-Doolittle method to determine the Beta weights for the 8 predictor variables that give the best least squares linear fit to the data. The individual Beta weights were then tested at a 90% confidence level to determine if they were statistically different from zero.* Those variables whose Beta weights were not significantly different from zero were discarded since they did not add anything of significance to the prediction. New Beta weights were then computed for the predictor variables remaining. These beta weights (symbolized by B_i) are based on standard z scores $z_i = \frac{x_i - \bar{x}}{s_x}$, so it was then necessary to obtain corrected Beta weights for usage with the raw scores of the variables (symbolized by b_i). The prediction equation thus computed is, GPA (Predicted) $=$

$$- 2.4816 + .0011\ CEEB_V + .0027\ CEEB_M + .4774\ HS_{GP} + 1.1207\ HS_{RE} \qquad (1)$$

Both the Beta weights for standard scores and for raw scores are given in Table 2. This is done because the standard score Beta weights are a better indicator of the relative importance of each of the predictor variables to the overall prediction th are the raw score Beta weights. The relation tween the two is

$$b_i = \beta_i \times \frac{S_{GPA}}{s_i}$$

A measure of the validity of the prediction is gi by the multiple correlation coefficient,

$$R = \sqrt{\sum_{1}^{n} \beta_i\, r_{GPA,i}} = .6742,$$

where $r_{GPA,i}$ refers to the correlation betw the criterion, GPA, and the ith predictor varia The standard error of the prediction is given

$$s_P = s_{GPA} \sqrt{1 - R^2} = .5640$$

This enables us to compute a 50% confidence terval for the prediction in the following mann

GPA (Predicted) $\pm$ .6745 s_P or

GPA (Predicted) $\pm$ .3804

Listed below are four examples of the use equation (1) to predict GPA. The column to right of the predicated GPA contains a 50% co dence band for the prediction, computed by eq tion (5). The last column on the right conta the actual GPA of the student used in the examp

The reader has probably noticed that the v ables discarded from the prediction were CEE IQ, HS_{RA}, and age. Perhaps a word of explanat is in order concerning these variables. It sho not be inferred that the discarding of these v ables implies that they have no value at all as p dictors of engineering success. One could not bla the reader who reacted somewhat to an infere of this kind. Rather, the implication is that th four variables add nothing new to the predict

$CEEB_V$	$CEEB_M$	HS_{GP}	HS_{RE}	Predicted GPA	Confidence Band	Actual GPA
554	681	1.33	0	0.60	.2196– .9804	0.94
502	421	1.33	+1	0.96	.5796–1.3404	0.77
464	527	1.50	+1	1.29	.9096–1.6704	1.38
671	740	2.50	+1	2.56	2.1796–2.9404	2.70

that is not already being supplied and more ficiently supplied by the four variables remain in the equation. $CEEB_T$, for instance, is a be *single* predictor than either $CEEB_V$ or $CEEB_M$ Table 1). However, a weighted combination of latter two is a better predictor of the criterion t $CEEB_T$. All three of these could not be inclu in the prediction equation because $CEEB_T$ i linear combination of $CEEB_V$ and $CEEB_M$ and resulting system of equations would be inconsist On the other hand, since the ages of the vari

*The test of significance used was

$$t = \frac{\beta_i \text{ (standard scores)}}{\sqrt{\frac{(1 - R^2)\, C_{ii}}{(N - m - 1)}}}, \text{ degrees of freedom} = (N-m-1),$$

where R is the multiple correlation coefficient, C_{ii} is the ith main diagonal entry in the inverse of the correlation matrix, N is the number of cases, and m is the number of predictor variables.

Table 1

Matrix of Intercorrelations of the Predictor Variables

	$CEEB_V$	$CEEB_M$	$CEEB_T$	IQ	HS_{GP}	HS_{RA}	Age	HS_{RE}	GPA
$CEEB_V$	1.000	.4756	.8715	.4587	.3532	.3213	.0767	.1480	.4236
$CEEB_M$	.4756	1.0000	.8436	.4747	.3359	.3839	.0063	.2010	.5099
$CEEB_T$	.8715	.8436	1.0000	.5329	.4068	.4147	.0436	.2064	.5427
IQ	.4587	.4747	.5329	1.0000	.2875	.3664	.1536	.2522	.2788
HS_{GP}	.3532	.3359	.4068	.2875	1.0000	.7833	.0786	.3352	.5694
HS_{RA}	.3213	.3839	.4147	.3665	.7833	1.0000	.0387	.2956	.4976
Age	.0767	.0063	.0436	.1536	.0786	.0387	1.0000	.0930	.0356
HA_{RE}	.1480	.2010	.2064	.2522	.3352	.2956	.0930	1.0000	.2834
x_i	507.47	587.04	1093.91	117.06	1.7325	74.098	17.810	0.6503	1.2175
s_i	91.271	85.031	152.36	10.323	0.6391	21.789	0.6697	0.0477	0.7637

applicants were so nearly the same, the variability in ages was drastically restricted and the variable proved to have no predictive value in this particular study. (See Table 1.)

It is interesting to observe that the HS_{RE} variable produced a statistically significant Beta weight and remained in the prediction equation despite the inefficient quantifying scheme we were forced to use. It is felt that perhaps if a more efficient method of quantifying this variable were devised, it would prove to have markedly better validity as a predictor. A possible means for achieving this would be to include a five point rating scale, such as the one described in Table 3, with the blank grade transcript sent to the high schools by most colleges. This scale should be marked by the principal or senior counselor and returned with the transcript of grades. Upon rehashing the data, it was observed that the +1 ratings had little predictive validity while the 0 and −1 ratings, particularly the −1 ratings, were usually quite valid.

Let us observe at this point that the row of B_i constants indicates the relative importance of the variables in the prediction. Thus HS_{GP} is the strongest predictor, followed by $CEEB_M$, $CEEB_V$, and HS_{RE}, respectively.

Table 2.

Beta Weights

	$CEEB_V$	$CEEB_M$	HS_{GP}	HS_{RE}
B_i (standard scores)	.1303	.2983	.3995	.0700
b_i (raw scores)	.0011	.0027	.4774	1.1207

Table 3.

Sample 5 Point Rating Scale for High School Recommendation

The principal or senior counselor will mark the one phrase which best describes the applicant.

5 — It is felt that this student is superior college material in every respect.

4 — It is felt that this student is good college material in every respect.

3 — It is felt that this student has the minimum qualifications for college work.

2 — There is doubt as to whether this student is qualified for college work.

1 — It is felt that this student is *not* qualified for college work.

LITERATURE CITED

Dubois, Philip H. 1957. *Multivariate Correlation Analysis.* Harper and Brothers, New York, 425 pp.

Peters, Charles C., and Van Voorhis, Walter R. 1940. *Statistical Procedures and Their Mathematical Bases.* McGraw-Hill Book Company, Inc., New York, 478 pp.

Walker, Helen M., and Lev, Joseph. 1953. *Statistical Inferences.* Henry Holt and Company, New York, 573 pp.

DATES OF MAILING VOLUME 37 (1962)

Number 1, January	February 27
Number 2, April	April 25
Number 3, July	July 19
Number 4, October	September 27