A DECISION THEORETIC APPROACH TO STUDENT PLACEMENT

James J. Higgins
Kansas State University

James R. Schwenke
Bristol-Myers Company

ABSTRACT

A decision theoretic approach is proposed for scheduling a student into a future course when there is more than one possible choice. The statistical aspects of the decision theoretic placement procedure are discussed in detail. A linear model approach is given for utilizing items in a student's records to estimate the probabilities of attaining various levels of performance or achievement as measured by an appropriate criterion. The probabilities so obtained are combined with subjective judgements via a utility function to arrive at a placement decision. An example is included illustrating the procedure for placing junior high mathematics students into one of two possible math courses.

INTRODUCTION

In scheduling a student for a future class, the teacher or counselor may have more than one course in a subject area from which to choose. Faced with several options, how does the counselor choose the course that will be of most value to the student? Clearly, the student's abilities must be taken into consideration, but ability alone is not the only factor. Motivation, career goals, and feelings about the worth of the various courses should also be considered. The decision theoretic approach takes into account both the abilities and the aspirations of the student in arriving at a placement decision. An elementary account of decision theory can be found in the textbook by Winkler (1972). A more advanced treatment can be found in Ferguson (1967). An overview of the aspects of decision theory that are applicable to this paper is presented below.

Consider that a student is to be scheduled into one of two mathematics courses (e.g., algebra or pre-algebra) for a new school term. At the outset, the proposed placement procedure requires that the possible outcomes of having taken each course be defined in terms of an appropriate measure of performance or achievement. Grades (A,B,C,D,F) which are readily available to counselors would be one possible measure, but others could be used instead. The matter of which measure might be most appropriate in a given situation will not be considered. Rather, we will assume that a measure has been agreed upon in the context of the placement decision. From that point, the decision theoretic placement procedure proceeds in three steps.

1. Probabilities are obtained for each outcome and for each student to be placed. For instance, a given student may have probabilities .4, .3, .2, .1, and 0 of obtaining an A,B,C, D, or F, respectively, in algebra and probabilities .6,.3,.1,0, 0 in pre-algebra. Ideally, these probabilities will be estimated from the student's previous records using appropriate statistical techniques. Thus, the probabilities will reflect the student's ability to attain the various levels of performance or achievement. In cases where records are incomplete the counselor will have to use prior experience in arriving at reasonable estimates of these probabilities.

2. Each student, in consultation with a counselor, assigns a numerical value to each outcome to express the personal worth placed on that outcome. For instance, the grade of A in algebra might have a high value for a student whose interests

are in science but be of relatively less value to one whose primary interests are in art. These values, called <u>utilities</u>, are meant to be subjective. Unlike the probabilities, which represent ability, the utilities are numerical expressions of the student's academic goals in the subject area. If decision theory were applied in a business context, utilities might represent the anticipated profits of various outcomes. In the educational context, the numerical scale for the utilities is somewhat arbitrary. In Section 3, we propose a 100 point utility scale with the value of 100 going to the most valuable outcome to the student.

3. An expected utility is computed for each course, as shown in Table 1.1 (in which grades have been used as the measure of performance). The expected utilities allow the counselor to use both the student's ability and goals in arriving at a placement decision. The student is placed in the course having the greatest expected utility for the student. Since the expected utilities represent the abilities and personal goals of each student, no meaningful comparison of expected utilities could be made among students.

In developing the decision procedure, two tasks are evident. One must estimate the probabilities of obtaining the various grades (the P_{ij}'s in Table 1.1), and assign the utilities to the various grades (the V_{ij}'s in Table 1.1).

ESTIMATING OUTCOME PROBABILITIES

When a placement decision is to be made, data that can be used in the decision making process are usually available to a counselor. Often it is possible to use data in a multiple regression model to predict a student's performance or achievement scores in the courses in question. If this can be done, then it is also possible to predict outcome probabilities,

Table 1.1

CALCULATION OF EXPECTED VALUE

	Course 1		Course 2	
Grade	Probabilities	Utilities	Probabilities	Utilities
A	P_{11}	V_{11}	P_{21}	V_{21}
B	P_{12}	V_{12}	P_{22}	V_{22}
C	P_{13}	V_{13}	P_{23}	V_{23}
D	P_{14}	V_{14}	P_{24}	V_{24}
F	P_{15}	V_{15}	P_{25}	P_{25}
Expected Value	$E(V_1) = P_{11}V_{11} + \ldots$ $\ldots + P_{15}V_{15}$		$E(V_2) = P_{21}V_{21} + \ldots$ $\ldots + P_{25}V_{25}$	

provided that the outcomes are defined in terms of the performance or achievement criterion. Predicting outcome probabilities has not received much attention in the regression context, but it is a problem that may be of some interest, independent of the decision theoretic placement procedure. For instance, it may be important for a student to know the probability of receiving a passing grade in a certain course. Thus, the methodology discussed in this section has wider applicability than the placement decision context.

Consider a situation in which a decision is to be made between two courses within a specific curriculum. For purposes of illustration, let us assume that grades (A=4, B=3, C=2, D=1, F=0) are chosen as the measure of course performance. (The methodology described below can easily be extended to more courses or other measures of performance or achievement.) Let Y_1

and Y_2 denote the grade (on the 4 point numerical scale) that a student will earn in course 1 and 2, respectively. Let $X_1, X_2, \ldots, X_p$ denote variables from the student's records that will be used to predict grades. A common model for predicting grades, and the one that we will use here, is the multiple linear regression model

$$\begin{aligned} Y_1 &= a_0 + a_1 x_1 + \ldots + a_p x_p + \varepsilon_1 \\ Y_2 &= b_0 + b_1 x_1 + \ldots + b_p x_p + \varepsilon_2 \end{aligned} \tag{2.1}$$

where ε_1 and ε_2 are (approximately) normally distributed "error" variables with mean 0 and variances σ_1^2 and σ_1^2, respectively. Assuming that the error variables are independent from student to student, one may estimate the coefficients in (2.1) using standard regression techniques (Draper and Smith, 1981). The predicted values in the model (2.1) are rounded off to the nearest integer in order to predict grades, i.e., if:

Y_i	Predicted Grade
less than 0.50,	F = 0
0.50 to 1.49,	D = 1
1.50 to 2.49,	C = 2
2.50 to 3.49,	B = 3
3.50 or above,	A = 4.

It is not essential that rounding off be done this way. Rather, all one needs is some reasonable rule for translating predicted values in the regression model to predicted grades.

Once the coefficients of the models in (2.1) have been estimated, estimates of the outcome probabilities can be obtained. We note that for a given value of predictor variables $X_1, X_2, \ldots, X_p$, the variables Y_1 and Y_2 are approximately normally distributed with means

$$\mu_1 = a_0 + a_1x_1 + \ldots + a_px_p$$

$$\mu_2 = b_0 + b_1x_1 + \ldots + b_px_p$$

and variances σ_1^2 and σ_2^2, respectively. If these means and variances were known exactly, then outcome probabilities could be readily computed. For example, the probability of a B in Course 1 would be given by

$$P[2.50 < Y_1 \leq 3.49] = P[\frac{2.50 - \mu_1}{\sigma_1} < Z < \frac{3.49 - \mu_1}{\sigma_1}]$$

where Z is a standard normal variate. In other words, one simply computes standardized limits and refers to a standard normal probability table for the desired probability. We propose simply substituting estimated means and variances for the true quantities and then carrying out the above computations as indicated. The estimated means $\hat{\mu}_1$ and $\hat{\mu}_2$ are found by using the estimated coefficient $\hat{a}_0$, $\hat{a}_1$, ..., $\hat{a}_p$, $\hat{b}_0$, $\hat{b}_1$, ..., $\hat{b}_p$ in the regression model. The estimated variance $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ are simply the mean squared errors (residual mean squares) of the respective regression models. The proposed estimates of the outcome probabilities are not the minimum variance unbiased estimates, which appear to be cumbersome to compute (see Johnson and Kotz, 1970, page 73), but the proposed estimates are easy to compute and should be adequate for practical purposes.

The estimated probabilities are very useful measures of expected performance. They are easily understood by both students and parents, whereas means and variances might not be.

A multiple linear regression approach to estimating these probabilities is suggested. Novick and Jackson (1974) have a detailed discussion of this problem from another point of view.

They consider the analysis when the data are grouped into categories according to values of the prediction variables $X_1, X_2, \ldots, X_p$ and the multinomial distribution is the underlying probability model. If the assumptions are correct or nearly correct, the multiple linear regression approach is a useful way of "smoothing" the data, thus eliminating what Novick and Jackson call "inversions" in the estimated probabilities. This approach also avoids artificial grouping of the predictor variables $X_1, X_2, \ldots, X_p$ into categories and eliminates the problem of having categories with so few values that the probability estimates are unreliable. If the assumptions of the multiple linear regression model are not valid, the techniques of Novick and Jackson perhaps may be profitably employed.

THE UTILITY FUNCTION

Once the probabilities have been determined, utilities must be assigned to the various outcomes. Outcomes are first ranked in order of preference to the student. Typically, the highest level of performance or achievement in the most advanced course will have the highest ranking. Arbitrarily, a value of 100 is assigned to the most preferred outcome and a value of 0 to the least preferred. Values are assigned to the other outcomes according to their worth with respect to those that are the most and least preferred. Outcomes may have the same value, but if one is more preferred than another, it must have a greater value. Also, two students may have the outcomes listed in the same order of preference but not have the same values assigned to them. The values are assigned to the outcomes subjectively and reflect the worth of each to the student.

As mentioned before, the values given to the outcomes are called utilities, and the rule which the student and counselor use to determine the utilities is called the <u>utility function</u>. Winkler (1972) suggests a way of determining the utility function through a sequence of hypothetical lotteries. A more practical way to determine the utility function may be to express the value of the outcome in question as a percentage of the value of the most desired outcome. For example, if a student is not too concerned with grade point average but is more concerned with just passing the course, a B in a given course might be worth 95% of the value of an A. If the student wanted to earn a high grade point average, a B might be worth only 80% of the value of an A.

The method of eliciting the utility functions from students will depend on the extent to which decision theory is used in counseling. Utilities could be obtained through a one-to-one exchange with the counselors if decision theory were applied only in selected cases. If it were implemented on a larger scale some instrument for obtaining utilities would have to be developed. It seems likely that it will take some experimentation to arrive at the procedure that would be most useful to a school system. However, we believe that even crudely determined utility functions would be better than ignoring differences in values among the students when making placement decisions.

EXAMPLE

Let us consider students going from eighth grade to ninth grade and having two mathematics options to consider--algebra I or pre-algebra. Grades (A,B,C,D,F) are the measure of performance in these classes. Let Y_1 be the ninth grade algebra I grade point for a particular student and Y_2 be the ninth grade pre-algebra grade point. The data used in estimating the model coefficients were obtained from 104 algebra I students and 94 pre-algebra students in a Florida junior high school. In this particular school, the students were scheduled for the ninth grade in the middle of the eighth grade. For this example, seven scores relating to mathematics were used in the multiple regression model for estimating the required probabilities. They are, with the notation to be used:

- X_1 : 7th grade math computation stanine
- X_2 : 7th grade math concepts stanine
- X_3 : 7th grade math problem solving stanine
- X_4 : 1st semester 7th grade math grade point
- X_5 : 2nd semester 7th grade math grade point
- X_6 : 1st quarter 8th grade math grade point
- X_7 : 2nd quarter 8th grade math grade point.

Classroom grades were recorded as letter grades, as were the seventh and eighth grade math grades, X_4 through X_7. These letter grades were converted to numerical grades following the usual pattern A = 4.0, B = 3.0, C = 2.0, D = 1.0, and F = 0.0. Application of standard regression techniques resulted in the models for algebra I and pre-algebra as shown in Table 4.1.

Table 4.1
MODELS

Variable	Estimated Coefficient
Intercept Algebra I	-.20
Intercept Pre-Algebra	.41
X_1	.02
X_2	.02
X_3	.11
X_4	.22
X_5	.08
X_6	.09
X_7	.20

MSE = .66
R^2 = .22

The models used have equal coefficients for the variables X_1 - X_7 and equal error variances. Graphic analysis and various statistical tests seemed to show that the models were adequate in this case, but models with different coefficients and error variances might be appropriate in other cases.

Estimated probabilities of D or better, C or better, B or better, and A were plotted as a function of the estimated mean (Figure 4.1). Since there is a common variance for algebra I and pre-algebra models, only one plot is needed for both. To

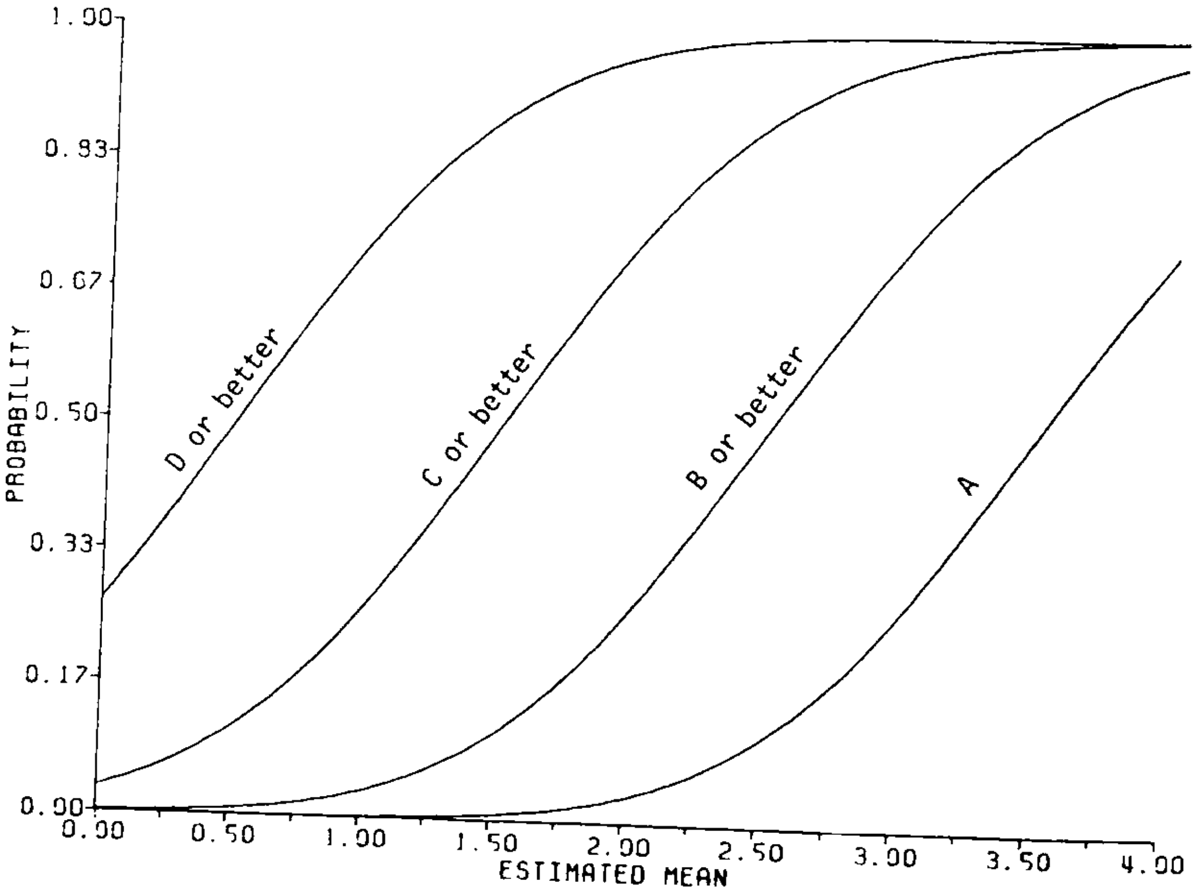

Figure 4.1 - Estimate Grade Probabilities, $\sigma^2 = 0.6586$

illustrate, if a student had predictor scores of $X_1 = 6$, $X_2 = 6$, $X_3 = 4$, $X_4 = 3$, $X_5 = 3$, $X_6 = 2$, and $X_7 = 4$, then the estimated mean for algebra I is

$$\hat{\mu} = -.20 + .02(6) + \ldots + .20(4) = 2.36.$$

The probability of a B or better is approximately .44 for this student.

The results of using the decision theoretic procedure for this student are given in Table 4.2. Notice that the student placed relatively high value on good grades in algebra I. The conclusion, based on the largest expected value to the student, would be to schedule the student into algebra I.

Table 4.2

RESULTS OF THE DECISION THEORETIC PROCEDURE

	Algebra I		Pre-Algebra	
Grade	Probabilities	Utilities	Probabilities	Utilities
A	0.08	100	0.26	80
B	0.36	90	0.46	70
C	0.42	80	0.24	50
D	0.13	0	0.03	0
F	0.01	0	0.01	0
Expected Values	74		65	

SUMMARY

Decision theoretic counseling is a structured, quantitative procedure for making placement decisions. Its potential value comes from the way that it brings together both data and beliefs in helping students choose the courses that appear to be of more benefit to them. Data provide estimated grade (or outcome) probabilities, and beliefs result in utility functions. The recommended courses are the ones with the greatest expected utility for the student. Unlike other placement procedures, subjective judgments, desires, and goals are brought to the forefront through the utility function rather than remaining in the background in some undefined way.

Regression models for estimating the probabilities of the various outcomes may be developed from grade records, standardized tests, and the like. The feasibility of carrying out this aspect of the decision theoretic process obviously will depend on the statistical expertise and computing facilities available to the school system. Fortunately, many of the smaller computers which are widely available to school systems have the capability of handling the computations necessary to do multiple regression. Although most regression analyses report means, variance, 95% prediction limits, and the like, a very useful way to summarize the data is to report outcome probabilities. Probabilities are generally easy to explain to students and parents, and are what is needed in the decision process.

Another approach to estimating outcome probabilities would be to obtain the desired estimates from the observed frequencies in a multifactor contingency table. This procedure may be satisfactory when the number of categories in the table is small (e.g., when this year's math grade is the only variable to be used in predicting next year's math grade). Otherwise, it would require rather large sample sizes to provide satisfactory estimates. In the above example, and others, satisfactory estimates of outcome probabilities can be obtained even with relatively small sample sizes by using the regression approach, provided the normality assumptions of the model are met.

There appear to be several ways to elicit utility functions from students. Under ideal circumstances, counselors would have time to interact with all students to determine utilities, but time constraints will probably make such interactions impractical. It may be possible for this aspect of the decision process to be carried out by teachers in the classrooms, provided an instrument was devised by the counselor to obtain the information necessary to construct utility functions. Of course, questions of validity and reliability of any such instrument would be an issue, but the difficulties do not seem to be insurmountable. However, more research would be needed on this point.

Although utility functions are expressions of individual beliefs, one might find that utilities are rather similar within certain groups due to such factors as educational, cultural, or income backgrounds. In these situations the possibility of grouping individuals in order to obtain a common utility function for each group could be explored.

The decision theory process would not have to be applied to all students in a system to be useful to counselors. Rather, it might be applied only in special circumstances or only with students who are in need of special advice. For example, if a student was having difficulty deciding between a foreign language elective or a science elective, a utility function and the expected utility might help clarify the choice. Moreover, the concept of expected utility might be useful to counselors even though it would not be applied directly to the student population. For instance, counselors could construct hypothetical utility functions to represent plausible attitudes and values of their students and compute expected utilities for various outcome probabilities as an aid in formulating general placement policies or recommendations.

It appears that implementation of the decision theoretic placement procedure could take many useful forms. It is beyond the scope of this paper to say which form would have the most potential benefit. Rather, the intent has been to introduce the decision theoretic approach to student placement through examples and illustrations and to suggest that it would be feasible for counselors to use this approach to meet specific objectives.

REFERENCES

Draper, Norman and Smith, Harry. Applied Regression Analysis. New York: John Wiley and Sons, Inc., 1981.

Ferguson, Thomas S. Mathematical Statistics, A Decision Theoretic Approach. New York: Academic Press, 1967.

Johnson, N.L. and Kotz, S. Continuous Univariate Distributions- 1. New York: John Wiley and Sons, Inc., 1970.

Novick, Melvin R., and Jackson, Paul II. Statistical Methods for Educational and Psychological Research. New York: McGraw-Hill, 1974.

Winkle, Robert L. Introduction to Bayesian Inference and Decision. New York: Holt Rinehart and Winston, Inc., 1972.