THE USE OF CUTTING SCORES IN SELECTION PROCEDURES

by

D. B. Owen and Loretta Li

Technical Report No. 129
ONR Contract

THE USE OF CUTTING SCORES IN SELECTION PROCEDURES

by

D. B. Owen and Loretta Li

Technical Report No. 129
Division of Mathematical Sciences
Department of Statistics
ONR Contract

June 15, 1979

Research sponsored by the Office of Naval Research
Contract N00014-76-C-0613
Project NR 042-341

DEPARTMENT OF STATISTICS
Southern Methodist University

THE USE OF CUTTING SCORES IN SELECTION PROCEDURES

by

D. B. Owen, Southern Methodist University, Dallas, Texas
and
Loretta Li, Bishop College, Dallas, Texas

1. INTRODUCTION

In 1939 Taylor and Russell produced a set of tables which gives the proportion of successful students (say) who are selected by the use of a single cutoff score, say, based on the overall SAT score. The tables were built around the assumption that the variables measuring success and aptitude were jointly bivariate normally distributed. There were several variations on this same general idea published by Jackson and Phillips (1945), Birnbaum (1950), Brown and Ghiselli (1953), and Stunkard and Hoyt (1952). A summary showing the differences and commonality of these procedures is given in the National Bureau of Standards Applied Mathematics Series No. 50 (1959), pp. xxxvii to xliv. In 1977, Thomas, Owen and Gunst extended the idea of the Taylor-Russell tables to two cutoff scores, say, SAT-V and SAT-M, where there may be a minimum requirement for both verbal and mathematical ability.

We have surveyed the educational and psychological literature of recent vintage and it appears that there is much concern over choosing cutoff scores, and the social and economic problems for people who are not selected. This is as it should be, but the concern has halted progress on the evaluation of cutoff scores and the mathematical techniques which underlie this area. In this paper we will first summarize by quotations what seems to be the prevalent thinking on these issues as published in the psychological and educational literature and then we will describe some new techniques which have been developed in the engineering literature which extend the Taylor-Russell cutoff score to the case where parameters are unknown. In all of the previous literature cited above, the parameters of the joint distributions were assumed known.

2. SUMMARY OF CURRENT LITERATURE SEARCH

We made a search of the literature to try to find what progress had been made and what problems have occurred with the use of the Taylor-Russell tables. The present concerns expressed in the psychological and educational literature with using cutoff scores to select individuals for college, for employment, etc., seems best represented by the following quotations.

- Lord and Novick (1968), p. 276, write: One must exercise care in using these tables (Taylor-Russell) because the situation they describe is highly artificial. The assumption of bivariate normality is certainly inaccurate in most testing applications. The basic situation we have described is highly artificial in another respect as well. We have supposed that we have no valid information on the applicants other than their test scores, and this is seldom the case. For example, in addition to knowing that a student has been graduated from high school, we generally know his class standing and perhaps have a record of prior work experience. It therefore seems that we should use the

Taylor-Russell tables to afford perspective, primarily, rather than to supply precise values."

Schrader in Payne (1967), p. 214, writes: "For general institutional planning, this method (Taylor-Russell) is probably the most realistic, since it takes the characteristics of the applicant group directly into account. However, in view of the well-known disadvantages of using a rigid cutting score in selection, tables prepared in this form should be used mainly for exploratory work."

Farr also in Payne (1967) p. 286, writes: "Hence, we are making a slight improvement in the percent of students who are successful at the loss of a considerable absolute number of such students." "In summary, I would observe that smaller quality programs based on strict selection among the applicants will not answer the problems of education, as they will produce only a slightly greater percentage of success among those admitted to the programs at a tremendous cost in absolute numbers of teachers provided." And also, p.287, "A second approach to evaluation is the comparison of the average performance of similar students under different programs of instruction."

Hoffman (1960), P. 116, and others approached this problem in the form of what he called "judgment; this may take the form of a recommendation concerning treatment or discharge, a decision that certain other data are necessary before final judgment is made, or a classification of the patient into a diagnostic category."
His "judgment" is a function (linear or otherwise) of all the available informations. No criterion is mentioned as to how to make the decisions based on the value of the judgment.
Cronbach and Gleser (1957), p. 46, approached this problem with the introduction of utility and payoff functions. "Our thinking is most consistent with the plan which assigns particular values to "hits" and "misses", and adjusts the cutting score to maximize expected utility."
Tiffin and McCormick (1965), p. 147, write: "The point of the above discussion is that the critical (hiring) score on a test must be varied with the tightness or looseness of the labor market. The tighter the market, the lower the critical score. The looser the market, the higher the critical score can be."
Cronbach (1970), p. 424, writes: "The choice of cutting score cannot be made scientifically, It is based on personal, social, and economic values, combined with practical considerations."

Cronbach and Gleser (1957), p. 76, write: "The Brogden linear relation or the Taylor-Russell function were connected specifically to coarse screening, and have been regarded as inapplicable to precise decisions or prediction for an individual." Further, p. 77, "When we may regard all successful men as making equal contribution to the institution, the Taylor-Russell tables are more appropriate for evaluating selective efficiency in fixed treatment than the linear function. Otherwise, the Taylor-Russell results are best regarded as a rough approximation to the linear relation."

Anastasi (1968), p. 133, writes: "In setting a cutoff score or a test, attention should be given to the percentage of false rejects (or false positives) as well as to the percentages of successes and failures within the selected group. In certain situations, the cutoff point should be set sufficiently high to exclude all but a few possible failures. This would be the case when the job is of such a nature that a poorly qualified worker could cause serious loss or damage. An example would be a commercial airline pilot. Under other circumstances, it may be more important to admit as many qualified persons as possible, at the risk of including more failures. In the latter case, the number of false rejects can be reduced by the choice of a lower cutoff score. Other factors that normally determine the position of the cutoff score include the available personnel supply, the number of job openings, and the urgency or speed with which the openings must be filled."

However, Farr [in Payne, (1967)] considers that a selection procedure is not very favorable in education or in employment selection because of the sacrifice of rejecting a large proportion of applicants, some of which will be successful. With the help of the Equal Employment Opportunity Commission, this problem cannot be ignored. An alternative method is to use the procedure to classify individuals rather than select them. The performance of applicants under different classifications is studied using regression analysis. In other words, the predicted scores of the applicant is used to determine which program is best for him.

There are some cautions concerning the use of the Taylor-Russell tables. In 1948, Max Smith listed the first one as the fact that the tables are obtained under the assumption that the joint distribution between the test scores and the criterion scores follows a normal bivariate distribution. Any departure from that distribution, for instance, a triangular distribution which occurs frequently in prediction of vocation, will not be appropriate to use the tables. Another caution concerns the source of the correlation coefficient used with the tables. The correlation coefficient with which one is supposed to enter the table is the correlation coefficient among the entire group of applicants. But very seldom will this be available. It will be estimated from the truncated group of those who were accepted and remained on the job for a certain period. This estimated correlation coefficient will be smaller than the correlation coefficient for the entire group and hence Smith (1948) concludes that we underestimate the increase in satisfactory employees by using the Taylor-Russell tables.

However, Chissom and Lanier (1975) found a significant multiple correlation for SAT-V scores, SAT-M scores, High School average, with college GPA. They found a multiple R of 0.57 for the three predictor variables with college GPA, with a P-value less than 0.01.

In addition Lord (1962) and (1963) considers the effect of errors of measurement on cutting scores.

3. NEW DEVELOPMENTS (PARAMETERS KNOWN)

We assume that we have a bivariate normal distribution with a performance variable Y (say, college GPA) and a lower specification limit on Y, which we will designate L. The quantity L may be the minimum GPA required for graduation. The proportion of the total population of GPA's (Y's) greater than L is designated γ. We propose to screen on the correlated variable X (say, total SAT score) so that we raise the proportion of Y's greater than L to δ, i.e., in mathematical terms

$$P\{Y \geq L\} = \gamma, \text{ and}$$

$$P\{Y \geq L \mid X \geq \mu_x - K_\beta \sigma_x\} = \delta,$$

where X and Y have a joint bivariate normal distribution with positive correlation, ρ.

The mean and standard deviation of X are μ_x and σ_x, respectively, and K_β is a standardized normal deviate corresponding to 100β% of a standardized normal distribution in the lower tail of the normal distribution.

Table I gives some representative values of a table constructed to meet this criterion. Further entries may be found in Owen, et al. (1975). We have tabulated values of β, i.e., the proportion of X's to be included in the screened population, in order to raise the proportions of Y's which meet specification to 0.95 in this case. You see, we have to know what proportion of Y is acceptable before screening and we have to know the correlation.

For example, if we wanted to raise the proportion of graduating students from 0.75 to 0.95 and the correlation ρ is 0.90 then we would

TABLE I

Tables of Values of β to Raise Proportion Successful from γ to $\delta = 0.95$

	Correlation = ρ				
γ	0.3	0.35	0.4	0.45	0.50
0.75	0.0035	0.0158	0.0429	0.0860	0.1429
0.80	0.0184	0.0531	0.1073	0.1759	0.2523
0.85	0.0880	0.1669	0.2561	0.3462	0.4318
0.90	0.3653	0.4746	0.5666	0.6425	0.7049

	Correlation = ρ				
γ	0.60	0.70	0.80	0.90	0.95
0.75	0.2812	0.4282	0.5661	0.6882	0.7432
0.80	0.4086	0.5511	0.6715	0.7696	0.8110
0.85	0.5806	0.6975	0.7863	0.8526	0.8784
0.90	0.7981	0.8612	0.9043	0.9331	0.9430

select the upper 68.82% of the X measurements, i.e., select all $X \geq \mu - 0.4908\sigma_x$.

In our example the original population can be divided into 4 parts:

(1) Those who are accepted by screening and would graduate, i.e., $P\{Y \geq L \text{ and } X \geq \mu_x - K_\beta\sigma_x\} = \delta\beta = 0.654$.

(2) Those who are rejected by screening but could have graduated, i.e., $P\{Y \geq L \text{ and } X < \mu_x - K_\beta\sigma_x\} = \gamma - \delta\beta = 0.096$. Note that these two add to $\gamma = P\{Y \geq L\}$.

(3) Those who are accepted by screening but would not graduate, i.e., $P\{Y < L \text{ and } X \geq \mu_x - K_\beta\sigma_x\} = \beta - \delta\beta = 0.034$.

(4) Those who are rejected by screening and would have failed, i.e., $P\{Y < L \text{ and } X < \mu_x - K_\beta\sigma_x\} = 1 - \gamma - \beta + \delta\beta = 0.216$. Note that these four proportions add to one.

This population is then divided into two populations, one of which is accepted by screening:

those who would graduate are

$$P\{Y \geq L \text{ given } X \geq \mu_x - K_\beta\sigma_x\} = \delta = 0.95,$$

and those who would fail to graduate are

$$P\{Y < L \text{ given } X \geq \mu_x - K_\beta\sigma_x\} = 1 - \delta = 0.05.$$

And the population which is rejected by screening:

those who could have graduated are

$$P\{Y \geq L \text{ given } X < \mu_x - K_\beta\sigma_x\} = \frac{\gamma-\delta\beta}{1-\beta} = 0.309,$$

and those who would fail to graduate are

$$P\{Y < L \text{ given } X < \mu_x - K_\beta\sigma_x\} = \frac{1-\gamma-\beta+\delta\beta}{1-\beta} = 0.691.$$

In the original population 25% fail to graduate, while in the population selected by screening only 5% fail to graduate.
On the other hand, in the rejected group 30.9% could have graduated.
Of course, the controversy in student selection procedures hinges on the fact that there are those individuals in the rejected group who could have graduated. There is no way to reduce this proportion to zero, and all of the double talk in the literature about this does not change that fact. The cutoff exists no matter how the issue is clouded and unless all applicants are accepted (and there is no selection), the proportion $\frac{\gamma-\delta\beta}{1-\beta}$ of rejected applicants could have graduated.

Now it might be well to digress to remark that we have assumed that we had a lower specification limit and a positive correlation. It is no trick to modify these rules to include cases of an upper specification limit and/or a negative correlation. The reader is referred to Owen, et al. (1975) for the rules for doing so.

All of this has been done under the assumption that the parameters of the bivariate normal distribution are known. This is comparable to the Taylor-Russell tables but with the change that the goal is first set and then the selection procedure is determined, rather than the other way around.

The extension which we now have developed is to the cases where the parameters of the bivariate distribution are unknown.

4. UNKNOWN PARAMETERS

When parameters are unknown, we proceed as follows:

(1) A preliminary sample of size n is obtained of paired values (x_1,y_1) .... (x_n,y_n) and the usual estimators of the parammeters are computed. We have to assume that these pairs come from a population which has not already been truncated by selection, otherwise we have the problem of underestimated correlation, among others.

(2) A lower $100\eta\%$ confidence limit on ρ is computed and called ρ^*. If this is positive, we proceed to step (3). If it is negative,we have additional steps to undertake. For a negative lower confidence limit we also compute an upper confidence limit. If this is positive,the procedure is stopped,since in that case there is a good chance the correlation is zero and nothing can be gained from the procedure. If the upper limit on the correlation is also negative, then a negative relationship between x and y is indicated and the procedures for negative correlation given by Owen et al. (1975) should be followed.

(3) A $100\eta\%$ lower confidence limit on $\gamma = P\{Y \geq L\}$ is computed, and we label it γ^*.

(4) Enter a table of the normal-conditioned on t-distribution (Table II) with parameters (and estimates) degrees of freedom = $n - 1$, γ^*, $\rho^* \sqrt{\frac{n}{n+1}}$, δ.

(5) All product (or applicants) are accepted if

$$X > \bar{x} - t_\beta \sqrt{\frac{n+1}{n}}\; s_x.$$

(6) We can then be $100(2\eta-1)\%$ sure that at least $100\delta\%$ of the Y's are above L in the selected population.

For example, if a preliminary sample of size 17 is taken and $r = 0.94$, then choosing $\eta = .95$ we obtain a 95% lower confidence limit on ρ to be $\rho^* = 0.8558$.

If $k = (\bar{y} - L)/s_y = 2.0$ then a 95% lower confidence limit on γ is $\gamma^* = 0.90$.

We enter the normal-conditioned on t-table with (17, 0.90, 0.8317, 0.95) for (n, γ, ρ, δ) and obtain $t_\beta = 1.388$. Our criterion is to select all students for which $X \geq \bar{x} - 1.428 s_x$. Then in the selected group we can be at least 90% sure that at least 95% of the students will be able to meet the graduation requirement.

If this screening is performed on a finite group of, say, M students, then the distribution of students in that group follow the binomial law with parameters M and δ. The situation is very similar to what is called prediction intervals in the literature, except that we say we are at least $100\ (2\eta-1)\%$ sure that the probability of z or less students failing

TABLE II

Values of t_β for the Normal-conditioned on t-distribution with $\delta = 0.95$ and $n = 17$

γ	0.7	0.75	0.80	0.85
0.80	0.130	0.292	0.445	0.591
0.85	0.519	0.660	0.793	1.002
0.90	1.093	1.208	1.318	1.422
0.92	1.469	1.570	1.666	1.756

is at least that given by the binomial distribution. Hence, if $M = 10$ for the example above with an $\eta = .95$, then we are at least 90% sure that the probability of zero failures in this group is 0.5999.

The proofs that the above procedures do in fact accomplish what is claimed for them are contained in papers by Owen and Boddie (1976) and Owen and Su (1977). An extension to two-sided entrance criteria is given by Li and Owen (1979). Extensive tables for consummating the various steps in the one-sided procedure appear in Odeh and Owen (1979).

5. THE NORMALITY ASSUMPTION

One of the problems which was mentioned several times in our survey of the literature above was the question on whether the measurements really had a joint bivariate normal distribution. Hensler, Mehrotra and Michalek (1977) addressed the problem of testing multivariate normality and, in particular, bivariate normality.

If the data should indicate that the variables are not normal, then a transformation should be applied to the data in an attempt to achieve normal variables. The Johnson System of curves may be used in this manner. See Johnson and Kotz (1970). Of course, the Johnson System is univariate, but it appears reasonable at this stage of development to use the Johnson procedure on the marginals which appear to be non-normal. It seems likely that this would then produce the joint bivariate normality which is required for using these procedures.

It is clear that the use of these screening procedures has many pitfalls when used to screen prospective students and all potential users should be aware of the concerns outlined in Section 2. The point is that screening does occur and hence there is a need to study its effects. Among problems not addressed in this paper are: (a) the estimate of the correlation coefficient may be available only from a selected group, and not the entire population. Lawley (1943-44) gives formulas for computing the effect on the parameters, but we do not know what the effect is on the lower confidence limit of the correlation coefficient; (b) the procedure for unknown parameters may not be viable in a situation where only limited numbers of students are available. There is a tradeoff between the pool of applicants and the size of student body selected. If the pool is much bigger than the groups to be selected then these procedures are viable. However, if the pool is only moderately bigger than the selected group then the procedures with unknown parameters are probably not viable.

All of this is to suggest that it is time the problems of selection were given more study.

REFERENCES

Anastasi, Anne. (1968). Psychological Testing (3rd.ed.). The Macmillan Co., New York, 133.

Birnbaum, Z.W. (1950). Effect of linear truncation of a multinormal population. Ann. Math. Statistics, 21, 272-279.

Brown, C.W. and Ghiselli, E.E. (1953). Per cent increase in proficiency resulting from use of selective devices. J. Appl. Psychology, 37, 341-344.

Cronbach, Lee J. (1970). Essentials of Psychological Testing (3rd.ed.) Harper & Row, New York, 424.

Cronbach, Lee J. and Gleser Goldine C. (1957). Psychological Tests and Personnel Decisions. University of Illinois Press, Urbana, 76,77.

Farr, David S. (1967). Article 41, Selection and a better use of tests. In Payne (Ed.) Educational and Psychological Measurement. Blaisdell Publishing Co., Waltham, Massachusetts, 284.

Hensler, G.L., Mehrotra, K.G., and Michalek, J.E. (1977). A goodness of fit test for multivariate normality. Communications In Statistics, A6, 1977, 33-41.

Hoffman, Paul I. (1960). The paramorphic representation of clinical judgment. Psychological Bulletin, Vol. 57, No. 2, pp. 116-131.

Johnson, Norman L. and Kotz, Samuel. (1970). Continuous Univariate Distributions -1. Houghton Mifflin, Massachusetts.

Jackson, Robert W.B. and Phillips, Alexander J. (1945). Prediction efficiencies by deciles for various degrees of relationship. Educational Research Series No. 11 (Ontario College of Education, University of Toronto.

Lawley, D.N. (1943-44). A note on Karl Pearson's selection formulae. Proc. Roy. Soc. Edinburgh, A62, 28-30, pt.1.

Li, Loretta and Owen, D.B. (1979). Two-sided screening procedures in the bivariate case. Technometrics, 20, 79-85.

Lord, Frederic M. and Novick, Melvin R. (1968). Statistical Theories of Mental Test Scores. Addison-Wesley Publishing Co., Reading, Massachusetts, 276.

Lord, Frederic M. (1962). Cutting scores and errors of measurement. Psychometrika, 27, 19-30.

Lord, Frederic M. (1963). Cutting scores and errors of measurement--a second case. Educational and Psychological Measurement, 23, 63-68.

National Bureau of Standards, Applied Mathematics Series 50. (1959). Tables of the Bivariate Normal Distribution Functions.

Odeh, R.E. and Owen, D.B. (1979). Tables for Normal Tolerance Limits, Sampling Plans, and Screening, Marcel Dekker, New York.

Owen, D.B. and Boddie, John. (1976). A screening method for increasing acceptable product with some parameters unknown. Technometrics, Vol. 18, No. 2, 195-199.

Owen, D.B., McIntire, D., and Seymour, E. (1975). Tables using one or two screening variables to increase acceptable product under one-sided specifications. Journal of Quality Technology, 7, 127-138.

Owen, D.B. and Haas, R. (1978). Tables of the normal conditioned on t-distribution. In (ED. H.A. David) Contributions to Survey Sampling and Applied Statistics. Academic Press, New York.

Owen, D.B. and Su, Yueh-ling Hsiao. (1977). Screening based on normal variables. Technometrics, 19, 65-68.

Schrader, William. (1967). Article 31, A taxonomy of expectancy tables. In Payne (Ed.) Educational and Psychological Measurement. Blaisdell Publishing Co., Waltham, Massachusetts, 214.

Smith, Max. (1958). Cautions concerning the use of the Taylor-Russell tables in employee selection, Journal of Applied Psychology, 32, 595-600.

Stunkard, D.L. and Hoyt, C.J. (1952). Tables for interpretation of correlation and reliability coefficients (Bureau of Educational Research, College of Education, University of Minnesota).

Taylor, H.C. and Russell, J.T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. J. Appl. Psychology, 23, 565-578.

Thomas, J. Gouras, Gunst, R.F. and Owen, D.B. (1977). Improving the use of educational tests as evaluation tools. Journal of Educational Statistics 2, 55-77.

Tiffin, Joseph and McCormick, Ernest T. (1965). Industrial Psychology (5th ed.) Prentice-Hall, Englewood Cliffs, N.J., 147.