Optimal Cutting Scores Using A Linear Loss Function

Wim J. van der Linden
Twente University of Technology, Enschede, The Netherlands

Gideon J. Mellenbergh
University of Amsterdam, Amsterdam, The Netherlands

The situation is considered in which a total score on a test is used for classifying examinees into two categories: "accepted (with scores above a cutting score on the test) and "not accepted" (with scores below the cutting score). A value on the latent variable is fixed in advance; examinees above this value are "suitable" and those below are "not suit-able." Using a linear loss function, a procedure is described for computing a cutting score that minimizes the risk for the decision rule. The procedure is demonstrated with a criterion-referenced achievement test of elementary statistics administered to 167 students.

In this article the situation is considered in which (1) a latent variable can be dichotomized into the categories "suitable" and "not suitable"; (2) the latent variable is measured by an instrument composed of items that are scored either 0 or 1; (3) the total score on the measurement instrument is the unweighted sum of the item scores; (4) examinees are "accepted" if their total score is higher than a cutting score; otherwise, they are "not accepted."

Examples of such situations include mastery decisions in criterion-referenced measurement. A decision-theoretic approach of criterion-referenced measurement has been proposed by Hambleton and Novick (1973). More recently, Meskauskas (1976) has given an overview of the available models for setting cutting scores. The general decision framework for this situation has been described by Huyhn (1976). Beside mastery decisions in educational testing, other examples are acceptance-rejection decisions for job applicants or for special treatments such as psychotherapy.

An optimal decision is formulated using a *linear* loss function. The following notation will be used:

n = number of items in the measurement instrument;

x_i = observed total score of examinee i ($x_i = 0, 1, \ldots n$);

τ_i = true proportion of items answered correctly by examinee i ($0 \leqslant \tau_i \leqslant 1$);

$h(x)$ = probability density of x;

$k(\tau, x)$ = joint probability density of x and τ;

$p(\tau|x)$ = probability density of τ, given x;
c = cutting score on the measurement instrument ($c = 0, 1, \ldots n$);
d = cutting score on the latent variable ($0 \leqslant d \leqslant 1$);
$E(\tau|x)$ = regression function of τ on x;
σ_x^2 = variance of x;
μ_x = mean of x;
$\varrho_{xx'}$ = reliability coefficient of x.

Optimal Cutting Scores

In order to compute an optimal cutting score on the measurement instrument, a loss function should be specified. The decision situation can be represented as in Figure 1.
The general form of the loss function is:

$$L = \begin{bmatrix} \ell_{00}(\tau) & \text{for } \tau < d, & x < c \\ \ell_{10}(\tau) & \text{for } \tau \geq d, & x < c \\ \ell_{01}(\tau) & \text{for } \tau < d, & x \geq c \\ \ell_{11}(\tau) & \text{for } \tau \geq d, & x \geq c \end{bmatrix} \qquad [1]$$

In this formula, $\ell_{ij}(i, j = 0, 1)$ is a function of τ. The risk is the expected loss (Ferguson, 1967, p. 7):

Figure 1

Twofold Table for Dichotomous Decisions

	Latent Variable	
	Not suitable ($\tau < d$)	Suitable ($\tau \geq d$)
Accepted ($x \geq c$)	$\ell_{01}(\tau)$	$\ell_{11}(\tau)$
Decision **Not accepted** ($x \leq c$)	$\ell_{00}(\tau)$	$\ell_{10}(\tau)$

$$R = E \ L = \sum_{x=0}^{c-1} \int_0^d \ell_{00}(\tau) \ k(x,\tau) \ d\tau + \sum_{x=0}^{c-1} \int_d^1 \ell_{10}(\tau) \ k(x,\tau) \ d\tau +$$

$$\sum_{x=c}^{n} \int_0^d \ell_{01}(\tau) \ k(x,\tau) \ d\tau + \sum_{x=c}^{n} \int_d^1 \ell_{11}(\tau) \ k(x,\tau) \ d\tau \qquad [2]$$

An optimal cutting score on the measurement instrument is that value of c that minimizes the risk.

Huyhn (1976) and Mellenbergh, Koppelaar, and van der Linden (1976) have considered the following loss function:

$$L = \begin{cases} \ell_{00}(\tau) = \ell_{11}(\tau) = 0 \\ \ell_{01}(\tau) = a \\ \ell_{10}(\tau) = b \end{cases} \qquad [3]$$

Assuming that the probability density of x, given τ, is binomial and that the regression of τ on x is linear, these authors have described methods for obtaining the cutting score that minimizes the risk.

An obvious disadvantage of the loss function given in Equation 3 is that the loss is constant. For instance, a not-accepted examinee with a latent score just above the cutting score gives the same loss as a not-accepted examinee with a latent score far above the cutting score. This constant loss can be eliminated by using a linear loss function:

$$L = \begin{cases} \ell_{00}(\tau) = \ell_{10}(\tau) = b_0(\tau - d) + a_0 \\ \ell_{11}(\tau) = \ell_{01}(\tau) = b_1(d - \tau) + a_1 \end{cases} \quad (b_0 + b_1) > 0 \qquad [4]$$

The condition $(b_0 + b_1) > 0$ is needed for the mathematical derivations given below; this condition is not really restrictive in its applications.

Specifying a loss function of the type given in Equation 4, it should be noted that both parts of Equation 4 contain two different components and, hence, two different kinds of parameters:

1. The parameters a_0 and a_1 represent amounts of loss, independent of the scores on the latent variable. These are constants and are used when the classification of examinees as either accepted or not accepted is to be made;
2. both $b_0(\tau - d)$ and $b_1(d - \tau)$ represent amounts of loss dependent upon the latent variable τ. These are proportional to the difference between the latent score τ and the cutting score d. The values of the parameters b_0 and b_1 are constants of proportionality.

The values for a_0 and a_1 should be chosen relative to each other and to $b_0(\tau - d)$ and $b_1(d - \tau)$, in such a manner that the resulting loss function represents the psychological, social, and economic consequences of both possible decisions.

Substituting Equation 4 into Equation 2 gives:

$$R = \sum_{x=0}^{c-1} \int_0^1 \{b_0(\tau - d) + a_0\} \ k(x,\tau) \ d\tau +$$

$$\sum_{x=c}^{n} \int_0^1 b_1(d-\tau) + a_1\} \ k(x,\tau) \ d\tau \qquad [5]$$

Using

$$k(x, \tau) = p(\tau | x) \, h(x) \qquad [6]$$

$$\int_0^1 p(\tau | x) \, d\tau = 1 \qquad [7]$$

$$\int_0^1 \tau \, p(\tau | x) \, d\tau = E(\tau | x) \qquad [8]$$

it follows that

$$R = \sum_{x=0}^{c-1} [b_0 \{E(\tau | x) - d\} + a_0] \, h(x) -$$
$$\sum_{x=c}^{n} [b_1 \{E(\tau | x) - d\} - a_1] \, h(x) \qquad [9]$$

This is equivalent to

$$R = \sum_{x=0}^{n} [b_0 \{E(\tau | x) - d\} + a_0] \, h(x) -$$
$$\sum_{x=c}^{n} [(b_0 + b_1) \{E(\tau | x) - d\} + (a_0 - a_1)] \, h(x) \qquad [10]$$

Remembering $(b_0 + b_1) > 0$ and eliminating the constant term of Equation 10, R is minimal for the cutting score, c', that maximizes

$$R' = \sum_{x=c}^{n} [(b_0 + b_1) \{E(\tau | x) - d\} + (a_0 - a_1)] \, h(x) \qquad [11]$$

The density $h(x)$ is equal to or greater than zero for all values of x. Therefore, if $E(\tau | x)$ is a monotonically increasing function, R is maximal for $c' = x$ for which

$$(b_0 + b_1) \{E(\tau | x) - d\} + (a_0 - a_1) \qquad [12]$$

is positive for the first time. Using this fact, the optimal cutting score can be found if the regression function is specified.

A possible regression function is the linear regression function of classical test theory (Lord & Novick, 1968, p. 65):

$$E(\tau|x) = \rho_{xx'} \frac{x}{n} + (1 - \rho_{xx'}) \frac{\mu_x}{n} \qquad [13]$$

Substituting Equation 13 into Equation 12, setting the result equal to 0, and solving for x gives

$$x' = \mu_x + \frac{n\{d - (a_0 - a_1)/(b_0 + b_1)\} - \mu_x}{\rho_{xx'}} \qquad [14]$$

The cutting score is an integer; for the first integer smaller than x' Equation 12 is negative and for the first integer greater than x' this expression is positive. Therefore, the optimal cutting score is

$$c' = \text{entier } [x'] + 1 \qquad [15]$$

where the entier function replaces x' by the first integer smaller than x'.

Obviously, choosing another test model will give rise to another regression function and, hence, to another expression for the optimal cutting score.

A Numerical Example

The procedure was applied to a 20-item, three-choice criterion-referenced test in elementary statistics and administered to 167 sophomores majoring in psychology. The teachers of the course considered students as having mastered the subject matter if they could answer correctly at least 80% of the total domain of items. Therefore, d is fixed at 0.80.

De Bruyne (1976) split the test into two parallel subtests using the method of "matched random subtests" (Gulliksen, 1950, p. 207). The means and variances were computed for the total test score (X) and the subtests (X_1 and X_2). Furthermore, the reliability was computed as the correlation coefficient between scores on the subtests corrected for double test length by the Spearman-Brown formula. The results of these computations are shown in Table 1. The table shows that the subtests are parallel in the sense that their means and variances are approximately equal.

Table 1

Statistics Subtests (X_1 and X_2) and Test (X)
"Elementary Statistics" ($N = 167$; $n = 20$)

Statistic	Subtest		Test (X)
	X_1	X_2	
Mean	7.401	7.479	14.880
Variance	3.145	2.782	9.612
Correlation		0.623	
Reliability ($r_{xx'}$)			0.768

Table 2

Optimal Cutting Scores (c) for Different
Values of $(a_0 - a_1)/(b_0 + b_1)$

$(a_0 - a_1)/(b_0 + b_1)$	c	$(a_0 - a_1)/(b_0 + b_1)$	c
$\geq$ + 0.628	0	+ 0.205 − + 0.243	11
+ 0.590 − + 0.627	1	+ 0.167 − + 0.204	12
+ 0.551 − + 0.589	2	+ 0.129 − + 0.166	13
+ 0.513 − + 0.550	3	+ 0.090 − + 0.128	14
+ 0.474 − + 0.512	4	+ 0.052 − + 0.089	15
+ 0.436 − + 0.473	5	+ 0.013 − + 0.051	16
+ 0.398 − + 0.435	6	− 0.025 − + 0.012	17
+ 0.359 − + 0.397	7	− 0.063 − − 0.026	18
+ 0.321 − + 0.358	8	− 0.102 − − 0.064	19
+ 0.282 − + 0.320	9	− 0.140 − − 0.103	20
+ 0.244 − + 0.281	10	$\leq$ − 0.141	

The parameters μ_x and $\varrho_{xx'}$ of Equation 13 were estimated by the mean and $r_{xx'}$ from Table 1. Using Equation 15, the optimal cutting scores were computed for the whole range of possible values of the ratio $(a_0 - a_1)/(b_0 + b_1)$; the results are shown in Table 2.

Discussion

The values of the ratio $(a_0 - a_1)/(b_0 + b_1)$ in Table 2 are rather small. However, the values of a_0 and a_1, which are in the numerator of this ratio, should be considered relative to the range of possible values of $b_0(\tau - d)$ and $b_1(d - \tau)$. Obviously, choosing another interval of possible values for τ, for example $[0,n]$, will give rise to larger values for the ratio $(a_0 - a_1)/(b_0 + b_1)$ in Table 2.

Inspection of Equations 12 and 14 shows that for fixed $(b_0 + b_1)$ the optimal cutting score is dependent upon $(a_0 - a_1)$. The larger the difference, the lower the optimal cutting score. This makes sense: a larger value of a_0 relative to a_1 means a larger amount of constant loss for the decision "nonacceptance" relative to "acceptance" and, in that case, the optimal cutting score is lower; a smaller value of a_0 relative to a_1 gives rise to the opposite.

An interesting case is $a_0 = a_1$. Whenever this occurs, the difference $(a_0 - a_1)$ of Equation 12 can be dropped and the ratio $(a_0 - a_1)/(b_0 + b_1)$ of Equation 14 disappears; this is true for every pair of values of b_0 and b_1. If the amounts of constant loss for both decisions are equal or if there are no constant losses at all, there is no need to choose values for b_0 and b_1. In such a case, the optimal cutting score is that value of x for which $[E(\tau|x) - d]$ is positive for the first time. Or, using the linear regression function of classical test theory, the optimal cutting score is the first integer value above

$$x' = \mu_x + \frac{nd - \mu_x}{\rho_{xx'}}$$

[16]

Using Equation 16, the optimal cutting score for the example is: $c' = 17$.

From equations 14 and 15, it follows that the optimal cutting score, c', is dependent upon the reliability coefficient $\varrho_{xx'}$. For a positive numerator in the second term of the right-hand part of Equation 14, the effect of increasing the reliability coefficient is to lower the optimal cutting score; for a negative numerator, the opposite occurs. In the case of Equation 16, this can be interpreted as a reverse regression-toward-the-mean: increasing the reliability coefficient means a shift of x' in the direction of μ_x.

Since the optimal cutting score is a function of the reliability coefficient in Equation 14, the accuracy of the estimate of the optimal cutting score is dependent on the accuracy of the estimate of the reliability coefficient. From the theory of estimating the reliability coefficient for normally distributed test scores, it follows that the confidence interval of the coefficient will increase as the coefficient in the population and the sample size decrease (Lord & Novick, 1968, p. 207). Criterion-referenced tests are usually short, and it is possible that population reliability coefficients will be small. Therefore, for criterion-referenced tests with low reliability coefficients the estimate of the optimal cutting score from Equation 14 may be poor. Hence, it is suggested that Equation 14 be used with reliable tests and large samples.

References

De Bruyne, H. C. D. *Blokken in het onderwijs.* Groningen: Tjeenk Willink, 1976.

Ferguson, T. S. *Mathematical statistics: A decision theoretic approach.* New York: Academic Press, 1967.

Gulliksen, H. *Theory of mental tests.* New York: Wiley & Sons, 1950.

Hambleton, R. K. & Novick, M. R. Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement,* 1973, *10,* 159–170.

Huyhn, H. Statistical considerations of mastery scores. *Psychometrika,* 1976, *41,* 65–79.

Lord, F. M. & Novick, M. R. *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley, 1968.

Mellenbergh, G. J., Koppelaar, H., & van der Linden, W. J. *Dichotomous decisions based on dichotomously scored items.* Amsterdam: Psychologisch Laboratorium, 1976.

Meskauskas, J. A. Evaluation models for criterion-referenced testing. *Review of Educational Research,* 1976, *46,* 133–158.

Author's Address

Gideon J. Mellenbergh, Psychologisch Laboratorium, Universiteit van Amsterdam, Weesperplein 8, Amsterdam, The Netherlands.