SCALING THRESHOLD CHARACTERS

DANIEL GIANOLA AND H. W. NORTON

Department of Animal Science, University of Illinois, Urbana, Illinois 61801

Manuscript received May 5, 1981
Revised copy received July 24, 1981

ABSTRACT

A simple method of scaling ordered categorical responses having a joint distribution with an underlying normal variable is presented. Scores are developed that maximize heritability of the observed variate and that in the class of scores based on polychotomies: (1) maximize the correlation between score and the underlying genetic value to be predicted, and (2) minimize mean-square prediction error. Several examples suggest little is lost, in terms of heritability, by using equally spaced scores. The proposed scaling method discriminates among candidates for selection that would be tied if equally spaced scores are used and sometimes yields different rankings of candidates.

In many traits with a genetic basis such as litter size in sheep or swine, the expression of the character is a response in one of several mutually exclusive and exhaustive ordered categories. Several authors (e.g., Wright 1934; Dempster and Lerner 1950) have assumed an underlying normal distribution of phenotypes. The normal random variable has a joint distribution with the observed discrete variable, with the mapping being done via a set of fixed thresholds. Dempster and Lerner (1950) have shown that for binary variables, selection in the observed scale results in smaller genetic gains than would be realized from selection in the underlying scale. When the prevalence of the character is equal to the proportion of individuals required as parents, selection in the observed scale yields the same genetic gain as selection upon the underlying variate.

In many animal breeding applications where the threshold model is assumed, the categories of response are scored linearly (e.g., Berger and Freeman 1978) and the resulting phenotypic values are analyzed by linear model methodology. A discussion of problems associated with this approach was presented by Gianola (1980). While there is question as to the meaning of heritability estimates obtained from raw data, as opposed to those obtained from transformations such as probits (e.g., Lush, Lamoreux and Hazel 1948; Robertson and Lerner, 1949), it is possible to develop scores that are optimal in some sense. We describe a scoring procedure for categorical variates with an underlying normal distribution, and its implications on a selection program.

SCORING PROCEDURE

An animal or experimental unit is subjected to a set of conditions defining a population in a statistical sense and the expression of the character is a response

in one of \( m \) mutually exclusive and exhaustive ordered categories. The phenotype in the observed scale may be represented by an \( m \times 1 \) random vector, \( \mathbf{v} \), which has elements equal to 1 in the position corresponding to the category of response and equal to zero elsewhere. If \( \Pi_i \) is the probability of response in the \( i \)th category (with \( \sum_{i=1}^{m} \Pi_i = 1 \)), the covariance matrix of \( \mathbf{v} \) has elements \( \Pi_i (1 - \Pi_i) \) for \( i = 1, \ldots, m, \) and \(-\Pi_i \Pi_j \) for \( i \neq j \). It is assumed that there is an underlying normally distributed trait (\( y \)), with heritability \( h^2 \) in the narrow sense, having a joint distribution with the observed variate. In the underlying scale there is a set of \( m - 1 \) fixed thresholds \( t' = [t_1, t_2, \ldots, t_{m-1}] \) defining the \( m \) response categories. With \( t_0 = -\infty \) and \( t_m = \infty \), the experimental unit responds in the \( i \)th category if \( t_{i-1} < y < t_i \).

A common practice in animal breeding is to score response categories by a vector of weights \( \mathbf{a'} = [a_1, a_2, \ldots, a_m] \), so that the scored phenotype is \( \mathbf{a'} \mathbf{v} \). It can be shown (GIANOLA 1979) that the heritability of \( \mathbf{a'} \mathbf{v} \) is given by

\[
h_0^2 = h^2 \left[ \sum_{i=1}^{m} a_i (z_{i-1} - z_i) \right]^2 / \left[ \sum_{i=1}^{m} a_i^2 \Pi_i - \left( \sum_{i=1}^{m} a_i \Pi_i \right)^2 \right],
\]

where \( z_0 = z_m = 0 \) and \( z_i = \exp(-t_i^2/2)/\sqrt{2\Pi_i} \), with \( t_i = \Phi^{-1} \left( \frac{1}{h_i} \sum_{i=1}^{m} \Pi_i \right) \). The ratio \( h_0^2/h^2 \) is less than 1 and not invariant to \( \mathbf{a} \) except when \( m = 2 \). In this case, (1) reduces to \( h_0^2 = h^2 z^2 / \Pi (1 - \Pi) \), a well known result due to ROBERTSON (see DEMPSTER and LERNER 1950). It should be noted that \( h_0^2/h^2 \) is also invariant to location and scale changes in \( \mathbf{a} \) since

\[
\sum_{i=1}^{m} (a_i + k) (z_{i-1} - z_i) / \sigma = \sum_{i=1}^{m} a_i (z_{i-1} - z_i) / \sigma
\]

because \( \sum_{i=1}^{m} (z_{i-1} - z_i) = 0 \), and

\[
\text{Var} \left[ (\mathbf{a'} + 1'/\mathbf{k}) \mathbf{v} / \sigma \right] = \sigma^{-2} \text{Var} (\mathbf{a'} \mathbf{v}) = \sigma^{-2} \left[ \sum_{i=1}^{m} a_i^2 \Pi_i - \left( \sum_{i=1}^{m} a_i \Pi_i \right)^2 \right],
\]

where \( k \) and \( \mathbf{1} \) are scalars and \( \mathbf{1} \) is a vector of ones.

In general, one may wish to find a value of \( \mathbf{a} \) that maximizes

\[
F(a) = h_0^2 / h^2 = \left( \sum_{i=1}^{m} a_i x_i \right)^2 / \left[ \sum_{i=1}^{m} a_i^2 \Pi_i - \left( \sum_{i=1}^{m} a_i \Pi_i \right)^2 \right],
\]

where \( x_i = z_{i-1} - z_i \). These "optimal" scores applied to data would yield the linear transformation with maximum heritability. Differentiating (2) with respect to \( \mathbf{a} \), setting to zero and rearranging gives the following system of simultaneous equations

\[
\tilde{a}_i = \sum_{i=1}^{m} \tilde{a}_i \Pi_i + \frac{x_i \left[ \sum_{i=1}^{m} \tilde{a}_i^2 \Pi_i - \left( \sum_{i=1}^{m} \tilde{a}_i \Pi_i \right)^2 \right]}{\Pi_i \left( \sum_{i=1}^{m} \tilde{a}_i x_i \right)}
\]

for \( i = 1, \ldots, m \).
Differentiating again, the conditions for (3) to give a maximum are

\[ h_0^2 > h^2 x_i^2 / \Pi_i (1 - \Pi_i) \]
for \( i = 1, \ldots, m \).

In view of the invariance of \( h_0^2 / h^2 \) to location and scale changes in \( a \), only scales with mean value equal to zero, i.e., \( \sum_{i=1}^{m} a_i \Pi_i = 0 \), need to be considered. Hence (3) has solution

\[ \tilde{a}_i = K x_i / \Pi_i \],

where \( K \) is an arbitrary constant which can be conveniently set to \( K = 1 \). Further, \( x_i / \Pi_i \) can be interpreted as the mean value of the \( i \)th class of a standard normal variate with classes with proportions \( \Pi_j, j = 1, \ldots, m \). It also follows that the “optimal” scores are monotonic, i.e., \( \tilde{a}_1 < \tilde{a}_2 \ldots < \tilde{a}_m \). For any \( k \) such that \( 2 \leq i + k \leq m \), \( \tilde{a}_{i+k} - \tilde{a}_i > 0 \) if \( x_{i+k} / \Pi_{i+k} > x_i / \Pi_i \). Letting \( \phi (u) \) and \( \Phi (u) \) be the normal density and distribution functions, respectively, by Cauchy’s Mean Value theorem

\[ \frac{\phi' (c) - \phi' (t_i)}{\Phi (t_i) - \Phi (t_{i-1})} = -c \]

for \( t_{i-1} < c < t_i \). Since \( \phi' (c) / \phi' (c) = -c \), the inequality \( x_{i+k} / \Pi_{i+k} > x_i / \Pi_i \) can be written as \( c + \Delta > c \). As \( \Delta \) is positive, monotonicity of the scores is established.

Further, it can be verified that \( F(\tilde{a}) = \sum_{i=1}^{m} x_i^2 / \Pi_i \). Since \( x_i^2 / \Pi_i (1 - \Pi_i) < F(\tilde{a}) = h_0^2 / h^2 \), the conditions for a maximum are satisfied.

Maximizing (2) with respect to \( a \) is also analogous to the maximization of generalized squared distance in discriminant analysis (SNEDECOR and COCHRAN 1967). The problem requires solving

\[ \sum_{i} \tilde{a}_i = k^* x \],

where \( \sum = \text{Var}(\upsilon) \), \( k^* \) is a constant and \( x^{'} = [x_1, x_2, \ldots, x_m] \). In this case \( a \) would contain the coefficients of the canonical variate, and \( x \) would be the vector of mean differences between populations. From the form of \( \Sigma \tilde{a} \) and setting \( k^* = 1 \), equations (6) can be written as

\[ \Pi_i \left[ \tilde{a}_i - \sum_{i=1}^{m} \tilde{a}_i \Pi_i \right] = x_i \quad ; \quad i = 1, \ldots, m \],

and in view of the location invariance of the objective function, one can take \( \tilde{a}_i = x_i / \Pi_i \) as solution; this is, of course, identical to (5).

The scores are conditional in the probabilities and these are rarely known in advance. If maximum likelihood estimates of \( \Pi_i \) are used to solve (3), then by the invariance property (e.g., MOOD, GRAYBILL and BOES 1974) the resulting scores are maximum likelihood estimates of \( a \).
Genetic change of a merit function $T$ per generation after one round of truncation selection with a predictor $\hat{T}$ is given by

$$
\Delta T = i \rho(T,\hat{T}) \sigma_T ,
$$

where $i$ is the standardized selection differential or selection intensity, $\rho(T,\hat{T})$ is the correlation between merit function and predictor, and $\sigma_T$ is the standard deviation of the merit function. Selection intensity depends solely on the proportion selected ($\alpha$), e.g., in the multivariate normal distribution $i = z/\alpha$, where $z$ is the ordinate at the point of truncation that divides the normal curve into two sectors with areas $\alpha$ and $1 - \alpha$. Hence, differences in $\Delta T$ stemming from using alternative predictors must be examined in terms of differences in $\rho(T,\hat{T})$. Let $T = g$ be the additive genetic value in the underlying normal scale.

The problem is to predict the unobservable genetic value $(g)$ of a set of candidates for selection on the basis of a vector of observable records. If the candidates' $g$'s and their records are uncorrelated with those of other candidates, the best selection rule is the conditional expectation of $g$ given the data (COCHRAN 1951). This rule: (1) is unbiased and minimizes mean-square prediction error among all predictors, (2) maximizes the correlation between predictor and predictand, (3) maximizes the expected value of the predictand in the selected set upon truncation selection on the rule, and (4) maximizes genetic progress if the predictand and the records have a multivariate normal distribution.

In the underlying scale, $g$, and the phenotypic value of an individual, $y$, are assumed to be bivariate normal. By scaling and centering such that $E(y) = E(g) = 0$, $\sigma_y = 1$, $\sigma_g = h$, $\rho(y,g) = h$, the joint density of $g$ and $y$ can be written as

$$
\phi(g,y) = \exp\left\{-\frac{y^2}{2} - \frac{hyg + g^2}{1 - h^2}\right\}/\sqrt{2\pi(1-h^2)} .
$$

If $y$ were observable, the best selection rule would be $\hat{T}(y) = E(g|y) = h^2y$, so that $\Delta T(y) = ih^2$.

However, the thresholds polychotomize $y$ into a discrete variable $Y$ such that a response in the $i$th category, $Y = i$ ($i = 1, \ldots, m$), occurs if $t_{i-1} < y < t_i$. The joint distribution of $g$ and $Y$ is given by

$$
f_i(g) = \text{Prob} \{ t_{i-1} < y < t_i, g < \bar{G} < g + \Delta g \} ; i = 1, \ldots, m .
$$

The conditional distribution of $g$ given $Y = i$ is

$$
f(g|Y = i) = \int_{t_{i-1}}^{t_i} \phi(g,y) \, dy / \int_{t_{i-1}}^{t_i} \phi(y) \, dy , i = 1, \ldots, m ,
$$

where $\phi(y)$ is the standard normal density. The best selection rule for this case is then

$$
\hat{T}(Y = i) = E(g|Y = i) = h^2(\sqrt{2\pi})^{-1}(e^{-t_i^2/2} - e^{-t_{i-1}^2/2})/[\Phi(t_i) - \Phi(t_{i-1})] .
$$
Note that \( \hat{T}(Y = i) = h^2 x_i / \Pi_i = h^2 \tilde{a}_i \), where \( \tilde{a}_i \) are the “optimal” scores. Hence \( \hat{T}_c(Y) = h^2 \tilde{a}' \nu \), is the best selection rule when \( \gamma \) is not observable. The criterion \( \tilde{a}' \nu \) will yield the same ranking of candidates as \( \hat{T}(Y) \). Now, from results of Robertson (Dempster and Lerner 1950) and of Gianola (1979)

\[
\text{Cov}(\tilde{a}' \nu, g) = h^2 \tilde{a}' x
\]

and

\[
\Delta T(Y) = \Delta T(\tilde{a}' \nu) = ih^2 [F(\tilde{a})]^{1/2}.
\]

Therefore,

\[
\Delta T(y)/\Delta T(\tilde{a}' \nu) = [F(\tilde{a})]^{-1/2} = [h^2/h^2_0(\tilde{a})]^{1/2},
\]

which is always larger than one. From (13), it also follows that \( \Delta T(\tilde{a}' \nu) > \Delta T(\tilde{a}' \nu) \) for any \( \tilde{a} \) since \( F(\tilde{a}) > F(\tilde{a}) \).

We have shown that \( \tilde{a}' \nu \) is the best selection rule or best predictor when \( \gamma \) is not observable. The best predictor and best linear predictor are the same in this case. The best linear predictor can be obtained from the selection index equations

\[
\sum a = \text{Cov}(\nu, g) = h^2 \tilde{a},
\]

which are identical to (6). Hence, the set \( \tilde{a}_i = x_i / \Pi_i, i = 1, \ldots, m \) has the following properties in the class of predictors when only a polychotomy \( \{Y = i\} \) is observed: (1) maximizes heritability in the observed scale, (2) is unbiased and minimizes mean square prediction error, and (3) maximizes the correlation between \( \tilde{a}' \nu \) and \( g \). Since \( g \) and \( v \) do not follow a multivariate normal distribution, other properties of the best predictor in the multivariate normal case do not hold for \( \tilde{a}' \nu \): (1) maximization of genetic progress (this is achieved by \( \hat{T}(Y) = h^2 y \)), (2) maximization of the probability of overall correct ranking when the candidates for selection are independent and identically distributed, and (3) maximization of the probability of correct pairwise ranking when all genetic values to be predicted have the same expectation.

Several examples suggest that, in general, equally spaced scores will result in negligible losses in terms of heritability. Consider \( m = 3, \Pi_1 = \Pi_3 = 0.1587, \Pi_2 = 0.6826 \), so that the distribution in the observed scale is symmetric. In this case, any set of equally spaced scores will yield the maximum \( h^2_0 / h^2 = 0.7380 \), i.e., the “optimal” linear transformation of the scores would be about 74% as heritable as the variable in the underlying scale. With \( m = 4, \Pi_1 = \Pi_4 = 0.05 \) and \( \Pi_2 = \Pi_3 = 0.45 \), the maximum is \( h^2_0 / h^2 = 0.8142 \). The usual scores \( a_i = i, \) for \( i = 1, \ldots, 4 \) yield \( h^2_0 / h^2 = 0.8138 \). The increase in heritability resulting from the “op-
timal” scores is negligible because the mean values for the categories are almost equally spaced in the underlying scale. The proposed scaling procedure yields some increases in heritability in the outward scale when the outward distribution is highly skewed, as illustrated in Table 1. For example, if \( m = 3 \), \( \Pi_1 = \Pi_2 = 0.005 \) and \( \Pi_3 = 0.99 \), the “usual” scores \( a_1 = 1 \), \( a_2 = 2 \) and \( a_3 = 3 \) give \( h_o^2/h^2 = 0.06827 \), compared to 0.07237 for \( a_1 = 0 \), \( a_2 = 1 \) and \( a_3 = 6.53 \) resulting from the development proposed here.

The impact of this increase in heritability is less in sib selection or progeny testing programs. It is well known (e.g., Henderson 1963) that the correlation between the additive genetic value of a candidate for selection and the mean of single records of \( n \) full sibs is given by \( k(n, h_o^2) = \{nh_o^2/[4 + 2(n-1)h_o^2]\}^{1/2} \). If \( h^2 = 0.5 \), the usual scores give \( h_o^2 = 0.03414 \) and \( k(4, 0.03414) = 0.1802 \). With the “optimal” scores of \( h_o^2 = 0.03619 \) and \( k(4, 0.03619) = 0.1852 \). Thus, a 6% larger heritability in the outward scale results in only a 2.8% difference in the accuracy of prediction when \( n = 4 \). As \( n \) gets larger, this difference becomes smaller, as the family mean tends rapidly towards normality.

It is interesting to observe that “optimal” scores may alter the ranking of candidates for selection. Consider a situation where the progeny of three sires are scored for a trichotomous variate for which the assumption of underlying normality is plausible. In case (a) of Table 2, the three sires would be ranked the same with the “usual” equally spaced scores, but the “optimal” weights discriminate among these candidates for selection. Taking into account the varying amounts of information also would have changed sire rankings with the “usual” scores. In case (b), the usual scores yield \( D > E > C \), but with the “optimal” scores, the ranking is changed, \( E > D > C \). This change of ranking also occurs in real-life data from performance-tested cattle and may be important (Fernando et al. 1981).

Rae (1950) considered the problem of developing a scale that would maximize the regression of offspring on dam in the context of scoring fleece quality in sheep. Data were arranged as a \( p \times p \) contingency table with \( p \) grades for dams on one axis and those of the offspring on the other; the entries were the number of dam-

### Table 1

<table>
<thead>
<tr>
<th>Cases</th>
<th>( h_o^2/h^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h_o^2/h^2 )</td>
<td>&quot;Usual&quot; scores</td>
</tr>
<tr>
<td>No. of categories</td>
<td>( a_i = 1 )</td>
</tr>
<tr>
<td>3 (0.005, 0.005, 0.990)</td>
<td>0.06627</td>
</tr>
<tr>
<td>3 (0.450, 0.150, 0.400)</td>
<td>0.72160</td>
</tr>
<tr>
<td>4 (0.005, 0.005, 0.005, 0.985)</td>
<td>0.00033</td>
</tr>
<tr>
<td>4 (0.400, 0.100, 0.400, 0.100)</td>
<td>0.79423</td>
</tr>
<tr>
<td>5 (0.010, 0.005, 0.005, 0.010, 0.970)</td>
<td>0.14277</td>
</tr>
<tr>
<td>6 (0.0001, 0.0001, 0.0001, 0.0001, 0.0001, 0.9995)</td>
<td>0.00546</td>
</tr>
</tbody>
</table>
### TABLE 2

Effect of equally spaced versus optimal scores on sire rankings for hypothetical examples (m = 3)

<table>
<thead>
<tr>
<th>Sire</th>
<th>No. of daughters in category</th>
<th>Probability of configuration</th>
<th>Mean score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(b)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>0</td>
<td>12</td>
</tr>
</tbody>
</table>

* Multinomial sampling assumed.

Offspring pairs corresponding to a particular combination of scores. "Optimal" scores were found by solving a determinantal equation. The method was applied to 547 daughter-dam pairs and yielded a heritability estimate of 46%, as opposed to an estimate of 26% with equally spaced scores. Because there was no exact test of significance available to assess differences between scales, RAE (1950) suggested that the method should be entertained with larger data sets before inferences concerning the efficiency of the "optimal" scale could be drawn. A clear shortcoming of the method proposed by RAE is that the "optimal" scores are not monotonic. This is a result of the fact that the method does not take advantage of the ordering of the categories of response in any way. However, it merits further attention for the case of unordered response categories.

**LITERATURE CITED**


Corresponding editor: J. F. Kidwell