

## Constructing categories for polychotomous variates

H. IWASAKI

*Department of Animal Science, College of Agriculture  
Kyoto University, Kyoto-shi 606, Japan*

### Summary

The objective of this study was to find, using numerical methods, a configuration of response categories which maximizes the ratio of heritability in a discrete scale to that in the underlying normal scale for polychotomous characters. It is shown that to maximize this ratio it is necessary to construct a symmetric discrete distribution with symmetrically spaced weights for the categories of response. The thresholds that maximize the ratio are located at the arithmetic mean of the expected values, on a normal scale, of contiguous categories. The value of the ratio for the « Optimal discrete distribution » found in this study is compared with those for some other discrete distributions.

*Key words* : Categories, polychotomous variates, heritability.

### Résumé

#### *Construction de catégories pour des variables polychotomiques*

Le but de cette étude est d'établir, par des méthodes numériques, une configuration des catégories de réponse qui maximise le rapport de l'héritabilité d'un caractère polychotomique définie sur une échelle discrète à celle définie sur l'échelle normale sous-jacente. On montre que, pour maximiser ce rapport, il est nécessaire de construire une distribution discrète symétrique, avec des scores répartis symétriquement dans les diverses catégories de réponse. Les seuils qui maximisent le rapport considéré sont égaux aux moyennes arithmétiques des espérances mathématiques, sur une échelle normale, des catégories adjacentes. On compare la valeur du rapport des héritabilités correspondant à la distribution discrète optimale trouvée dans cette étude à celles correspondant à quelques autres distributions discrètes.

*Mots clés* : Catégories, variables polychotomiques, héritabilité.

### I. Introduction

Threshold characters present a discrete phenotypic distribution and heritability on such a scale depends on the population frequencies in the different categories of response (DEMPSTER & LERNER, 1950 ; VINSON *et al.*, 1976 ; GIANOLA, 1979). As far as

the case of dichotomous phenotypes is concerned, it is already known that the ratio of heritability in a discrete scale to that in the underlying normal scale is maximum with an incidence of 0.5, and that the maximum expected heritability in a discrete scale is about 65 p. 100 of that in the underlying normal scale (e.g., FALCONER, 1981 ; JAMES & MCGUIRK, 1982). This finding, in other words, implies that the dichotomous distribution with a 50 p. 100 incidence is a discrete distribution which minimizes the loss in heritability to that of the underlying normal variate. In this context, however, no such theoretical evidence is available for cases of polychotomous characters with 2 or more thresholds.

Thus, the objective of this study was to evaluate the maximum of the ratio of heritability in the discrete scale to that in the underlying normal scale and to find the configuration of a discrete distribution which maximizes the ratio.

## II. Methodology

GIANOLA (1979) has developed the following general formula relating heritability in the narrow sense in the underlying normal scale to that in the outward scale :

$$h_o^2 = h_u^2 \left[ \sum_{i=1}^m a_i \{\phi(t_{i-1}) - \phi(t_i)\}^2 / \left\{ \sum_{i=1}^m a_i^2 \pi_i - \left( \sum_{i=1}^m a_i \pi_i \right)^2 \right\} \right] \quad (1)$$

where  $h_o^2$  is the heritability in an outward scale ;  $h_u^2$  is the heritability in the underlying normal scale ;  $m$  is the number of mutually exclusive and exhaustive categories of response ;  $\pi_i$  is the probability of response in the  $i^{\text{th}}$  category with  $\sum_{i=1}^m \pi_i = 1$  ;  $t_i$  is the  $i^{\text{th}}$  fixed threshold with  $\phi(t_i) = (2\pi)^{-1/2} \exp(-t_i^2/2)$  ;  $a_i$  is the weight for the  $i^{\text{th}}$  category.

Eq. (1) can be used for this study in the form of  $F(t, \mathbf{a}) = h_o^2/h_u^2$ . The problem is to find the maximum of  $F(t, \mathbf{a})$  with respect to  $\mathbf{t}' = [t_1, t_2, \dots, t_{m-1}]$  and  $\mathbf{a}' = [a_1, a_2, \dots, a_m]$ , and maximizing  $F(t, \mathbf{a})$  with respect to the unknown parameters requires solving :

$$\frac{\partial F(t, \mathbf{a})}{\partial \mathbf{t}} = \mathbf{0} \quad (2)$$

$$\frac{\partial F(t, \mathbf{a})}{\partial \mathbf{a}} = \mathbf{0} \quad (3)$$

From (2) the following system of equations is derived :

$$\begin{aligned} & -\phi(t_k) (a_{k+1} - a_k) \sum_{i=1}^m a_i \{\phi(t_{i-1}) - \phi(t_i)\} [-2t_k \left\{ \sum_{i=1}^m a_i^2 \pi_i - \left( \sum_{i=1}^m a_i \pi_i \right)^2 \right\} \\ & - \sum_{i=1}^m a_i \{\phi(t_{i-1}) - \phi(t_i)\} \{(a_{k+1} + a_k) - 2 \sum_{i=1}^m a_i \pi_i\}] = 0 ; k = 1, 2, \dots, m-1 \end{aligned} \quad (4)$$

On the other hand, without loss of generality (3) with  $\sum_{i=1}^m a_i \pi_i = 0$  leads to :

$$a_k = \{\phi(t_{k-1}) - \phi(t_k)\} / \pi_k ; k = 1, 2, \dots, m \tag{5}$$

as has been stated already by GIANOLA & NORTON (1981). Hence, using  $\sum_{i=1}^m a_i \pi_i = 0$  and (5) and rearranging (4) gives :

$$\begin{aligned} G(t_k) &= t_k - \left\{ \frac{\phi(t_{k-1}) - \phi(t_k)}{\pi_k} + \frac{\phi(t_k) - \phi(t_{k+1})}{\pi_{k+1}} \right\} / 2 \\ &= t_k - (a_k + a_{k+1}) / 2 \\ &= 0 ; k = 1, 2, \dots, m - 1 \end{aligned} \tag{6}$$

Now we take 3 successive standard normal variates  $t_{k-1}, t_k, t_{k+1}$  ( $t_{k-1} < t_k < t_{k+1}$ ) and their symmetrical points about zero  $-t_{k-1}, -t_k, -t_{k+1}$  ( $-t_{k-1} > -t_k > -t_{k+1}$ ). So, from

$$\phi(t_{k-1}) - \phi(t_k) = -\{\phi(-t_k) - \phi(-t_{k-1})\}$$

and

$$\int_{t_{k-1}}^{t_k} \phi(x) dx = \int_{-t_k}^{-t_{k-1}} \phi(x) dx ; k = 1, 2, \dots, m - 1$$

one can find sequentially if  $G(t_k) = 0$ , also  $G(-t_k) = 0$  for  $k = 1, 2, \dots, m - 1$  and consequently from (5) one can also find  $a_k = a_{m-k+1}$  for  $k = 1, 2, \dots, m$ . These facts show that a necessary condition to maximize  $h_v^2/h_u^2$  is to take thresholds to form a symmetric distribution with symmetrically spaced weights around the mean. The vector  $t$  maximizing  $h_v^2/h_u^2$  would have elements consisting of the arithmetic means between  $a_k$  and  $a_{k+1}$  ;  $k = 1, 2, \dots, m - 1$ . The procedure developed by BRENT (1973) was used to find an iterative solution to the system (6). The procedure is a modification of Newton's method which utilizes orthogonal transformations. Equal frequency values for response categories were taken to obtain initial  $t$ 's for iteration. The calculation of  $\pi_i$  was made on the basis of the method of CLENSHAW & CURTIS (1960).

Values of  $h_v^2/h_u^2$  are given in table 1 for the following cases : 1) « optimal distribution » (as given in this paper) with either normal or equally spaced scores, 2) equal percentile distribution, with normal or equally spaced scores, and 3) equally spaced thresholds with normal or equally spaced scores. Computation was performed by FACOM M-382 at the Data Processing Center of Kyoto University.

### III. Results and discussion

Table 1 presents the value of  $h_v^2/h_u^2$  in the situation of the « optimal distribution » with normal scoring of categories for a given number of categories, in comparison with those in other natural situations. For dichotomous responses, the maximum expected ratio of  $h_v^2/h_u^2$  is well known to be only 0.637. This work reveals that the maximum of the ratio in the trichotomous case is about 0.80 with an increment of 0.17 from that in the dichotomous case, and the maximum value in the pentachotomous case is above

0.90. The maximum, always less than 1, increases as the number of response categories increases. When the number of categories is as many as 10, the outward variable following the optimal distribution found may be as heritable as the underlying variable.

TABLE 1  
Value of  $h_o^2/h_u^2$  for several discrete distributions.

Type of distribution	Scoring method <sup>(a)</sup>	No. of categories							
		3	4	5	6	7	8	9	10
Optimal distribution	NS <sup>(b)</sup>	.809	.882	.920	.942	.955	.965	.972	.977
	ESS <sup>(c)</sup>	.809	.881	.917	.939	.953	.962	.969	.974
Equal percentile distribution	NS	.793	.860	.895	.914	.924	.930	.934	.936
	ESS	.793	.856	.886	.902	.912	.917	.920	.923
Distribution with equally spaced thresholds	NS	.738	.837	.890	.921	.940	.954	.963	.970
	ESS	.738	.837	.889	.920	.940	.954	.963	.970

(a) NS : Normal scoring of categories.

ESS : Equally spaced scoring of categories.

(b) Values were found from Eq. (6).

(c) Values were found from Eq. (4).

Differences in  $h_o^2/h_u^2$  between the optimal distribution and some distributions investigated in this study were found to exist although the magnitudes were small on the whole. In cases with less than 5 categories, the value of  $h_o^2/h_u^2$  for the equal percentile distribution is closer to that for the optimal distribution and higher than that for the distribution derived from equally spaced thresholds, and *vice versa* in the cases of more than 5. The 2 different scoring systems for categories, normal scoring and equally spaced scoring, had very little effect on the ratio value in all the situations as has been found by GIANOLA & NORTON (1981). This result is due to the fact that the differences between scores of adjacent categories are approximately the same with the 2 scoring procedures.

TABLE 2  
Probabilities of response to yield the maximum of  $h_o^2/h_u^2$  with normal scoring of categories.

No. of categories	Probability of response									
	$\pi_1$	$\pi_2$	$\pi_3$	$\pi_4$	$\pi_5$	$\pi_6$	$\pi_7$	$\pi_8$	$\pi_9$	$\pi_{10}$
3	.270	.459	.270							
4	.163	.336	.336	.163						
5	.106	.244	.297	.244	.106					
6	.073	.181	.245	.245	.181	.073				
7	.053	.137	.198	.220	.198	.137	.053			
8	.040	.106	.161	.191	.191	.161	.106	.040		
9	.031	.084	.132	.164	.175	.164	.132	.084	.031	
10	.024	.068	.109	.140	.157	.157	.140	.109	.068	.024

Table 2 shows probabilities of response for categories to yield the maximum of  $h_o^2/h_c^2$  with normal scoring of categories. The distributions show a bell-shaped symmetrical pattern as would be expected. These are the discrete distributions which fit « best » a  $N(0, 1)$  for a given number of categories, and in this situation of frequency distribution with normal scoring of categories, response to selection would be maximum.

In conclusion, the results obtained in this work suggest how to construct a discrete distribution from data on a normal one with the restriction of minimizing the loss in heritability for a given number of categories. The answer is that the discrete distribution should be symmetric (table 2) and that categories should be scored using normal scores. This finding may be of interest when a categorical presentation might be better adapted than a continuous one for practical reasons and may also be useful in studies measuring the loss of information occurring from rendering a normal distribution into a discrete one.

*Received July 16, 1984.*

*Accepted June 26, 1985.*

### Acknowledgements

The author wishes to thank the 2 referees and a scientific editor, and Prof. Y. YAMADA for their valuable suggestions on the manuscript.

### References

- BRENT R.P., 1973. Some efficient algorithms for solving systems of nonlinear equations. *SIAM J. Numer. Anal.*, **10**, 327-344.
- CLENSHAW C.W., CURTIS A.R., 1960. A method for numerical integration on an automatic computer. *Numer. Math.*, **2**, 197-205.
- DEMPSTER E.R., LERNER I.M., 1950. Heritability of threshold characters. *Genetics*, **35**, 212-236.
- FALCONER D.S., 1981. *Introduction to Quantitative Genetics*, 2nd ed., 275 pp., Longman, London.
- GIANOLA D., 1979. Heritability of polychotomous characters. *Genetics*, **93**, 1051-1055.
- GIANOLA D., NORTON H.W., 1981. Scaling threshold characters. *Genetics*, **99**, 357-384.
- JAMES J.W., MC GUIRK B.J., 1982. Regression of offspring on parent for an all-or-none trait. *Z. Tierzuchtg. Zuchtgsbiol.*, **99**, 308-314.
- VINSON W.E., WHITE J.M., KLEWER R.H., 1976. Overall classification as a selection criterion for improving categorically scored components of type in *Holsteins*. *J. Dairy Sci.*, **59**, 2104-2121.