Theor. Appl. Genet. 53, 89-91

Limitations of Conventional Regression Analysis A Proposed Modification

M.M. Verma and G.S. Chahal
Department of Plant Breeding, Punjab Agricultural University, Ludhiana (India)
B.R. Murty
Nuclear Research Laboratory, Indian Agricultural Research Institute, New Delhi (India)

Summary. The conventional genotype-environment interaction analysis cannot detect the theoretically ideal genotype which has been defined as the one with relatively low sensitivity in the poor environments and high sensitivity in the favourable environments.

The computation of separate regression coefficients on the two regions of the response curve has been suggested to detect such genotypes. This procedure is simple and more convenient than the complicated curvilinear regression analysis.

Key words: Regression analysis – Genotype-environment interaction – Proposed modification

Introduction

Limitations of Regression Analysis

The regression technique of measuring genotype-environment interaction developed by Finlay and Wilkinson (1963) and later improved by Eberhart and Russell (1966) and Perkins and Jinks (1968a, b) has been extensively used in many crop plants to quantify the response of a set of genotypes to the varying environments. There is now sufficient evidence that mean performance and sensitivity to environment are two independent characters which can be manipulated (Bucio-Alanis, Perkins and Jinks 1969; Bains 1976). This technique enables the breeder to identify three kinds of genotypes for direct use in appropriate environments or for use in breeding programmes as parents:

(a) Parents with low sensitivity ($1 + \beta < 1$)
(b) Parents with average sensitivity ($1 + \beta = 1$)
(c) Parents with above average sensitivity ($1 + \beta > 1$)

However, a theoretically ideal genotype would be the one which possesses a relatively high yield and stable performance in the low-yielding environments as well as the capacity to respond to favourable environments (Fig. 1). In almost all crop plants varieties are known that either are ideally suited to poor environments or to rich environments. The availability of genotypes with different levels of mean performance and environmental sensitivity suggests that sensitivity has not only a separate genetic control but that in fact there may be two distinct sets of gene-systems controlling sensitivity in the contrasting environments in addition to some common genes. It can be argued that it is possible to envisage a genotype combining high mean performance with high or low sensitivity in each of the contrasting set of environments. In other words, genotypes of the following kinds, among others, should occur if a very large sample is tested:

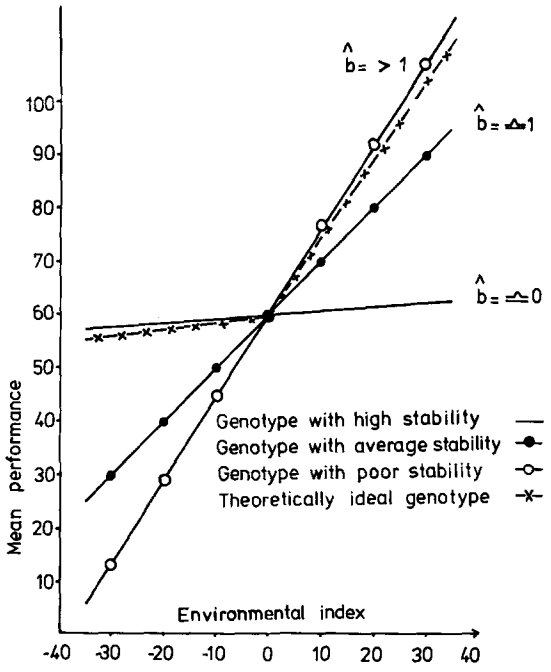

Fig. 1. Theoretically ideal genotype with low sensitivity in the poor environments and high sensitivity in the favourable environments and three kinds of regression lines in the conventional analysis

0040-5752/78/0053/0089/$ 1.00

(a) Genotypes with poor sensitivity under below average environmental conditions but high sensitivity under favourable conditions (theoretically ideal genotype).

(b) Genotypes with high sensitivity under below average conditions but with low sensitivity under favourable conditions.

It can be argued that the genotypes of both kinds do exist but cannot be identified by the regression technique in its present form considering all the environments together. The deviations from regression being rather purely statistical have considerable biological significance in that there may be a consistent trend not only in their magnitude but also in their direction under the below average and the above average environmental conditions. Accordingly, a near-ideal genotype, as defined above, may have a large magnitude of deviations around a single best-fitting regression line and may in fact be rejected as the direction and magnitude of deviations at different environments are not examined. Deviations in the desirable direction in the poor environments for a regression line based on favourable environments are, in fact, advantageous.

Proposed Change in Regression Analysis

The series of environments are first truncated at zero environmental index or around it so as to have two sub-sets of environments: one sub-set consisting of those with minus environmental indices and the other comprising environments with plus environmental indices including the one with minimum negative deviation for continuity of the two regression lines. Then the response curve of the genotypes with graded environments should be partitioned and the slopes determined separately for the two regions. It should be possible to detect genotypes of the categories listed in Table 1. Further, for each genotype the regression value in one analysis can be compared with the one in the second analysis in order to determine if the two are statistically different or whether there is no difference between the responses in the two sets of environments, provided sufficient residual degrees of freedom are available to detect differences between the two regression coefficients. Out of nine possible combinations, combinations (ii) and (iii) approach the ideal genotype depending upon the mean performance; the combinations (i) and (vii) are ideal for below average environmental conditions whereas combination (ix) is the best for favourable environments. Tall Indian varieties of wheat and rice can be said to match combination (vii) since under high fertility and favourable conditions these would tend to lodge and not respond to favourable environments in terms of yield per unit of area.

Table 1. Classification of genotypes on the basis of regression $(1 + \beta)$ in the two sets of environments

Category	Regression $(1 + \beta)$	
	Poor Environments	Favourable Environments
(i)	< 1.0	< 1.0
(ii)		$= 1.0$
(iii)		> 1.0
(iv)	$= 1.0$	< 1.0
(v)		$= 1.0$
(vi)		> 1.0
(vii)	> 1.0	< 1.0
(viii)		$= 1.0$
(ix)		> 1.0

Example

Data on mean final plant height (over 8 replicate individuals) of 10 inbred lines of *Nicotiana rustica* grown in eight environments was available from the Department of Genetics, University of Birmingham, England. Out of the eight environments studied, four had negative environmental indices, and the remaining four, positive indices. The environmental indices were −17.52, −9.45, −5.60, −2.84, 2.21, 9.89, 11.55 and 11.75. The first four environments were considered as poor environments. Set comprising the favourable environments included the environments with +ve index as well as the one with environmental index of −2.84 so as to provide continuity of the regression line. The regression slopes with respect to all the genotypes are presented in Table 2. Due to the small number of environments in the proposed analysis of two sets, many regression slopes did not reach the level of significance. However, it may be noted that the magnitude of sensitivity as measured by the regression technique is not constant over the two sets of environments and the rank correlation between the sensitivities in the two sets of environments was not significant (−0.515 with 8d.f.). Differential behaviour of at least some of the genotypes clearly indicates that the computation of the linear regression over all the environments has the masking effect on the detection of near-ideal genotypes, if they exist in the population. Genotype 2 approaches this ideal in the given example (Fig. 2). It may be noted that the difference between the genotypes 1 and 2 based on common regression on all environments is minimal. Genotype 2 is as good as genotype 1 in the favourable environments but is far superior to it in the unfavourable environments. Responses in these two sub-sets of environments are not dependent on each other as is evident from the rank correlation. It is, therefore, likely that the two genetic systems for these two types of responses are not mutually antagonistic.

Curvilinear response is not necessarily favoured in the

Table 2. Regression Coefficients ($1 + \beta_i$) in the two sets of environments

Genotype	All Environments (Eight)	Environments with -ve index (Four)		Environments with +ve index and the one with the least -ve index (Five)	
	$1 + \beta_i$	$1 + \beta_i$	Rank	$1 + \beta_i$	Rank
1.	0.9713[a]	0.7179[b]	7	1.1225	3
2.	0.8431[a]	0.4123[b]	9	1.1204[b]	4
3.	0.7974[b]	−0.0756	10	1.5630	2
4.	1.1120[a]	1.0202	6	1.5905[b]	1
5.	1.0692[a]	1.1355[b]	5	1.1026[b]	5
6.	0.5297[a]	0.5210	8	0.2223	10
7.	1.4230[a]	2.0265[b]	1	0.8283	8
8.	1.0415[a]	1.1425	4	0.9947	7
9.	1.1238[a]	1.2737[b]	3	1.0226	6
10.	1.1160[a]	1.8250[b]	2	0.4330	9

[a]Significant at 1% level; [b]Significant at 5% level
Number of environments given in parenthesis

favourable environments because we would like to minimize the deviations from linearity under these environments. The approach suggested in this paper is a simple and convenient method to detect the differences between genotypes in place of complicated curvilinear regression analysis.

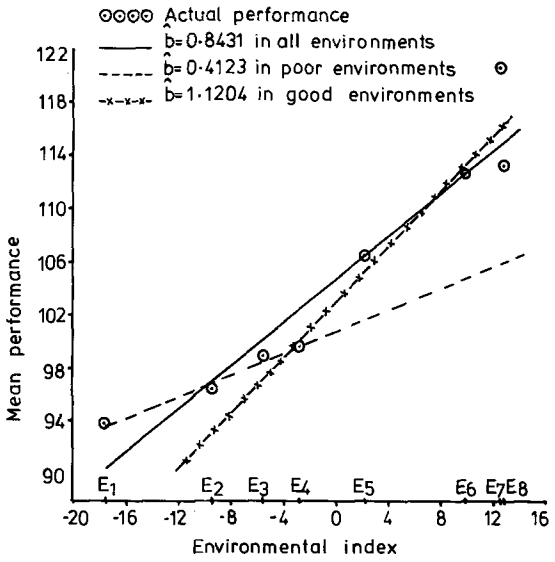

Fig. 2. Regression lines of the genotype 2 in the two separate analyses ($1 + \beta = 0.4123$ and 1.1204) compared with the regression line over all environments ($1 + \beta = 0.8431$)

Acknowledgement

Thanks are due to Dr. J.L. Jinks, Professor, Department of Genetics, University of Birmingham, Birmingham B15 2TT, England, for allowing the use of the data collected in the department.

Literature

Bains, K.S.: Parent dependent genotype X environment interaction in crosses of spring wheat. Heredity **36**, 163-171 (1976)

Bucio-Alanis, L.; Perkins, J.M.; Jinks, J.L.: Environmental and genotype-environmental components of variability. V. Segregating generations. Heredity **24**, 155-157 (1969)

Eberhart, S.A.; Russell, W.A.: Stability parameters for comparing varieties. Crop Sci. **6**, 36-40 (1966)

Finlay, K.W.; Wilkinson, G.N.: The analysis of adaptation in a plant breeding programme. Aust. J. Agri. Res. **14**, 742-754 (1963)

Perkins, J.M.; Jinks, J.L.: Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. Heredity **23**, 339-356 (1968a)

Perkins, J.M.; Jinks, J.L.: Environmental and genotype-environmental components of variability. IV. Non-linear interactions for multiple inbred lines. Heredity **23**, 525-535 (1968b)

Received June 6, 1978
Communicated by H.F. Linskens

M.M. Verma and G.S. Chahal
Department of Plant Breeding
Punjab Agricultural University
Ludhiana (India)

B.R. Murty
Nuclear Research Laboratory, IARI
New Delhi (India)