Theor. Appl. Genet. 54, 7-9 (1979)

Using Regression Coefficient as a Stability Parameter in Plant Breeding Programs*

K. Miezan, G.A. Milliken and G.H. Liang
Kansas State University, Manhattan, Kansas (U.S.A.)

Summary. Expanding the regression coefficient as stability parameter (Finlay and Wilkinson 1963) requires an unbiased interpretation of the parameter. Information on the covariances among the genotypes of the population must be specific, particularly when the assumed relatedness of the genotypes appears questionable. Such a problem, however, is not expected when the covariances between the genotypes are either zero or equal and possibly non zero.

Key words: Regression coefficient Stability parameter Covariance

Introduction

In plant breeding programs, genotypes are often evaluated in different environments (locations and years) before selecting or recommending certain genotypes. Genotypes tested in different environments almost invariably show genotypes-by-environment interactions; that is, the relative performances of the genotypes vary from one environment to another. Such differential response of genotypes in different environments makes it difficult for breeders to decide which genotypes should be selected. Different methods were proposed to solve the problems created by genotype-by environment interactions (çomstock and Moll 1963; Liang, Heyne, and Walter 1966; Sprague and Federer 1951). On the other hand, the regression approach originally proposed by Yates and Cochran (1938) and later used by Finlay and Wilkinson (1963) has also been widely used in comparing and measuring genotypic performance in various environments. In the regression approach, an environment index is measured by the mean performance of all genotypes grown in that environment, and performance of individual genotypes is regressed on the environmental index. Regression coefficients that have been so derived measure phenotypic stability; that is, phenotypes with regression coefficients of 1.0 have an average stability, whereas coefficients less than or greater than 1.0 indicate above average and below average stability. Many researchers have used this method to investigate genotype-by-environment interactions in different species (Breese 1969; Johnson, Shafer and Schmidt 1968; Jowett 1972; Perkins and Jinks 1968 and 1971), and others have discussed the limitations for using a regression on an environment index (Knight 1970; Witcombe and Whittington 1971).

This paper examines some theoretical problems associated with using and interpreting the regression coefficient as a stability parameter.

Experimental Theory

Assume the following model for any character, such as grain yield,

$$X_{ijk} = \mu + g_i + e_j (1 + b_i) + \epsilon_{ijk}$$

where X_{ijk} = phenotypic performance of the i^{th} genotype in the j^{th} environment in the k^{th} replicate,

μ = overall mean of the population.
g_i = effect of the i^{th} genotype, $i = 1,2,...,n$,
e_j = effect of the j^{th} environment, $j = 1,2,...,m$,
$1 + b_i$ = linear regression of the i^{th} genotype on the environment means and
ϵ_{ijk} = random error associated with the i^{th} genotype in the j^{th} environment in the k^{th} replicate.

When the mean of the j^{th} environment, $\overline{X}.j.$, is used as the environment index, (the subscripts will be dropped to

---

* Contribution No. 78-330-J, Kansas State Agric. Expt. Sta., Manhattan, Ks.

simplify the notation, i.e., use $\overline{X}$) then the stability parameter of the i^{th} genotype is defined as

$$1 + b_i = \frac{\mathrm{Cov}(X_i, \overline{X})}{\mathrm{Var}(\overline{X})}$$

where X_i is the performance of the i^{th} genotype in a given environment and

$$\overline{X} = \frac{\overset{n}{\underset{1}{\Sigma}} X_i}{n}$$

which is the environment index where n = number of genotypes. When i = 1, the quantity $1 + b_1$ can be expressed by

$$1 + b_1 = \frac{\mathrm{Cov}(X_1, \overline{X})}{\mathrm{Var}(\overline{X})}$$

where the covariance and variance are respectively:

$$\mathrm{Cov}(X_1, \overline{X}) = \mathrm{Cov}(X_1, \frac{1}{n} \sum_1^n X_i) = \frac{1}{n} \sum_1^n \mathrm{Cov}(X_1, X_i)$$

$$= \frac{1}{n}[\mathrm{Var}(X_1) + \sum_2^n \mathrm{Cov}(X_1, X_i)]$$

$$\text{where } \mathrm{Cov}(X_1, X_1) = \mathrm{Var}(X_1)$$

and

$$\mathrm{Var}(\overline{X}) = \mathrm{Var}(\frac{1}{n} \sum_1^n X_i) = \frac{1}{n^2} \mathrm{Var}(\sum_1^n X_i)$$

$$= \frac{1}{2^2} [\sum_1^n \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)]$$

and thus,

$$1 + b_1 = \frac{\mathrm{Var}(X_1) + \sum_2^n \mathrm{Cov}(X_1, X_j)}{\frac{1}{n} \sum_{i=1}^n \mathrm{Var}(X_i) + \frac{1}{n} \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)} \quad \dots\dots (1)$$

To evaluate the effectiveness of $1 + b_1$ as a measure of stability, we consider several special cases by applying various restrictions on the covariances.

1. Assume that $\mathrm{Cov}(X_i, X_j) = 0$ for any $i \neq j$, i.e.,

$$\sum_2^n \mathrm{Cov}(X_1, X_j) = \sum_{i \neq j} \mathrm{Cov}(X_i, X_j) = 0.$$

The assumption of zero covariances corresponds to assuming that the genotypes in the experiment are genetically unrelated.

Then equation (1) becomes $1 + b_1^* = \dfrac{\mathrm{Var}(X_1)}{\dfrac{\sum_{i=1}^n \mathrm{Var}(X_i)}{n}}$

Let $\overline{\mathrm{Var}(X)}$ denote $\dfrac{\sum_{i=1}^n \mathrm{Var}(X_i)}{n}$, then

$$1 + b_i^* = \frac{\mathrm{Var}(X_i)}{\overline{\mathrm{Var}(X)}} \quad \dots\dots\dots\dots\dots\dots\dots (2)$$

Therefore, $1 + b_i^*$ is the ratio of the variance of the genotype under study to the average variance of all the genotypes in the population. The interpretations of the regression coefficients are:

$1 + b_i^* = 1$ indicates average stability, i.e., the i^{th} genotype has the same response to environmental changes defined by $\mathrm{Var}(X_i)$ as the average genotype expressed by $\overline{\mathrm{Var}(X)}$.

$1 + b_i^* > 1$ indicates a below average stability as the i^{th} genotype responds more to environmental changes than the average genotype.

$1 + b_i^* < 1$ indicates an above average stability as all the genotypes on the average respond more to the environment than the i^{th} genotype.

If we define r_i as

$$r_i = \frac{\mathrm{Var}(X_i)}{\sum_{j=1}^n \mathrm{Var}(X_j) - \mathrm{Var}(X_i)}, \text{ then}$$

$$1 + b_i^* = \frac{n\,\mathrm{Var}(X_j)}{[\Sigma \mathrm{Var}(X_j) - \mathrm{Var}(X_i)] + \mathrm{Var}(X_i)} = \frac{r_i n}{r_i + 1}$$

$$(i = 1,2,\dots,n)\dots(3)$$

Note that if all of the variances are equal, than $1 + b_i^* = 1$.

From (3), it can be noted that:

(a) when $r_i \to 0$, then $b_i \to -1$, and that happens when $\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \dots + \mathrm{Var}(X_{i-1}) + \mathrm{Var}(X_{i+1}) + \dots + \mathrm{Var}(X_n) \to \infty$

(b) When $r_i \to \infty$, then $b_i^* \to n-1$ and that happens when $\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \dots + \mathrm{Var}(X_{i-1}) + \mathrm{Var}(X_{i+1}) + \dots + \mathrm{Var}(X_n) \to 0$.

Thus, by assuming the covariances among genotypes to be zero, we see that if we conduct an experiment including a genotype with a very large variance and another experiment without this genotype, we will get a very large to possibly small value of the regression coefficient. The upper limit of the b* value will depend on number of genotypes and the magnitude of their variances.

2. Assume that the covariances are all equal and possibly non zero, i.e.,

$$\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(X_1, X_3) = \dots = \mathrm{Cov}(X_1, X_j) = \dots = \mathrm{Cov}(X_1, X_n) = \dots = \mathrm{Cov}(X_{n-1}, X_n),$$

then $\sum_2^n \mathrm{Cov}(X_1, X_j) = (n-1)\,\mathrm{Cov}(X_1, X_j)$,

$\sum_{i \neq j} \mathrm{Cov}(X_i, X_j) = (n-1)\,\mathrm{Cov}(X_i, X_j)$

and

$$1 + b_1 = \frac{\mathrm{Var}(X_1) + (n-1)\,\mathrm{Cov}(X_1,X_j)}{\overline{\mathrm{Var}(X)} + (n-1)\,\mathrm{Cov}(X_1,X_j)}\,.$$

Hence, we obtain an interpretation similar to that with the zero covariances assumption,

$1 + b_i \leq 1$ implies $\mathrm{Var}(X_i) \leq \overline{\mathrm{Var}(X)}$, and

$1 + b_i > 1$ implies $\mathrm{Var}(X_i) > \overline{\mathrm{Var}(X)}$.

This is a more reasonable assumption than the zero covariances, because we can have such a relationship between genotypes just by random sampling from a finite number of genotypes.

3. If the covariances between X_1 and $X_2, X_3 \ldots, X_n$ are equal and positive but smaller than the covariances among $X_2, X_3 \ldots, X_n$ (also assumed to be equal and positive), i.e., $\mathrm{Cov}(X_1,X_2) = \mathrm{Cov}(X_1,X_3) = \ldots = \mathrm{Cov}(X_1,X_j) = \ldots = \mathrm{Cov}(X_1,X_n) < \mathrm{Cov}(X_2,X_3) = \ldots = \mathrm{Cov}(X_2,X_n) = \ldots = \mathrm{Cov}(X_h,X_p) = \ldots = \mathrm{Cov}(X_{n-1},X_n)$,

then

$$1 + b_1 \simeq \frac{\mathrm{Var}(X_1) + (n-1)\,\mathrm{Cov}(X_1,X_j)}{\overline{\mathrm{Var}(X)} + (n-2)\,\mathrm{Cov}(X_h,X_p) + 2\,\mathrm{Cov}(X_1,X_j)} \qquad (4)$$

so that,

$1 + b_i < 1$ does not necessarily mean that X_1 is a less stable genotype because for $\mathrm{Var}(X_i) \leq \overline{\mathrm{Var}(X)}$, $1 + b_i$ will always be less than 1. The only case in which b_i should be used is when $1 + b_i$ is greater than 1.0 which only indicates that $\mathrm{Var}(X_i)$ is greater than $\overline{\mathrm{Var}(X)}$.

4. If the covariances between X_1, and $X_2, X_3, \ldots, X_n$ are equal and positive but larger than the covariances between $X_2, X_3, \ldots, X_n$ (also assumed equal and positive), i.e., $\mathrm{Cov}(X_1,X_2) = \mathrm{Cov}(X_1,X_3) = \ldots = \mathrm{Cov}(X_1,X_h) = \ldots = \mathrm{Cov}(X_1,X_n) > \mathrm{Cov}(X_2,X_3) = \mathrm{Cov}(X_2,X_4) = \ldots = \mathrm{Cov}(X_2,X_h) = \ldots = \mathrm{Cov}(X_{n-1},X_n)$, then the form of $1 + b_1$ is the same as in (4) if n is large. The interpretations differ, however, for it could easily be demonstrated that unless $1 + b_i$ is less than 1 (indicating $\mathrm{Var}(X_i) < \overline{\mathrm{Var}(X)}$), the b value may not be very useful.

5. If no assumptions can be made about the covariances among the genotypes, then the interpretation of the b values will be very complicated and possibly not interpretable. This suggests that more interpretable results would be obtained by some other kind of environmental index.

Conclusion

Using the regression coefficient as a stability parameter may help plant breeders and geneticists select genotypes for wide adaptation. However, it appears that the assumptions underlying the use of the method are not always satisfied, particularly that of the random sample of the genotypes, as the breeder and the geneticist usually work with highly selected material. Under these conditions, it seems that one must be cautious in using the regression coefficient as stability parameter without some prior information about the covariances among the genotypes. This problem could, however, be removed by working with truly random samples from a large population of genotypes to satisfy the zero covariances assumptions, or random samples of genotypes from a finite population to satisfy the assumption of equal covariances. In any case, the random sample of genotypes is the most important assumption to satisfy before the regression coefficient can be used efficiently as a stability parameter.

Literature

Breese, E.L.: The measurement and significance of genotype-environment interactions in grasses. Heredity **24**, 27-44 (1969)

Comstock, R.E.; Moll, R.H.: Genotype-environment interaction. In: Statistical Genetics and Plant Breeding (eds. Hanson, W.D.; Robinson, H.F.) pp. 164-196, Washington, D.C.: Nat. Acad. Sci., Nat. Res. Council 1963

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

Johnson, V.A.; Shafer, L.; Schmidt, J.W.: Regression analysis of general adaptation in hard red winter wheat (*Triticum aestivum* L.). Crop Sci. **8**, 187-191 (1968)

Jowett, D.: Yield stability parameters for sorghum in East Africa. Crop Sci. **6**, 314-317 (1972)

Knight, R.: The measurement and interpretation of genotype-environment interactions. Euphytica **19**, 225-235 (1970)

Liang, G.H.; Heyne, E.G.; Walter, T.L.: Estimate of variety X environmental interactions in yield tests of three small grains and their significance on the breeding programs. Crop Sci. **6**, 135-139 (1966)

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

Perkins, J.M.; Jinks, J.L.: Specificity of the interaction of genotypes with contrasting environments. Heredity **26**, 463-474 (1971)

Sprague, G.F.; Federer, W.T.: A comparison of variance components in corn yield trials. II. Error, year X variety, location X variety and variety components. Agron. J. **48**, 535-541 (1951)

Witcombe, J.R.; Whittington, W.J.: A study of the genotype by environment interaction shown by germinating seeds of *Brassica napus*. Heredity **26**, 397-411 (1971)

Yates, F.; Cochran, W.G.: The analysis of groups of experiments. J. Agric. Sci., Camb. **28**, 556-580 (1938)

Received May 10, 1978
Accepted August 22, 1978
Communicated by R.W. Allard

Dr. G.A. Milliken
Department of Statistics
Kansas State University
Manhattan Kansas 66506 (U.S.A.)