Theor Appl Genet (1991) 81: 559–561

Weighted regression analysis for comparing varietal adaptation

D. S. Virk [1], P. S. Virk [2], B. K. Mangat [1] and G. Harinarayana [3]

[1] Department of Plant Breeding, Punjab Agricultural University, Ludhiana, India
[2] School of Biological Sciences, University of Birmingham, B15 2TT, UK
[3] College of Agriculture, Pune, India

Received May 15, 1990; Accepted October 30, 1990
Communicated by B. R. Murty

Summary. The normally used joint linear regression analysis (OLS) is not appropriate for comparing estimates of stability parameters of varieties when the error variances of site means are heterogeneous. Weighted regression analysis (WLS), in these situations, yields more precise estimates of stability parameters. A comparison of the two analytical methods using the grain yield (kg ha^{-1}) data of 12 varieties and one hybrid of pearl millet [*Pennisetum typhoides* (Burm.) S. & H.], tested at 26 sites in India, revealed that the weighted regression analysis yields more efficient estimates of regression coefficients (b_i) than the ordinary regression analysis, and that the standard errors of b_i values were reduced by up to 43%. The estimated b_is differed with the two procedures. The number of varieties with b_is significantly deviating from unity was not only more (five varieties) with weighted regression analysis than the ordinary regression analysis (one variety), but the classification of varieties as possessing general or specific adaptation differed with the two procedures.

Key words: *Pennisetum typhoides* – Pearl millet – Weighted regression analysis – Stability analysis

Introduction

Joint regression analysis (Yates and Cochran 1938) and its several modifications (Finlay and Wilkinson 1963; Eberhart and Russell 1966; Perkins and Jinks 1968) are commonly used to assess the nature of genotype × environment interactions and to compare the relative stabilities of varieties, although the assumption of homogeneity of error variances across environments is rarely tested or satisfied. In the event of significant heterogeneity among error variances, comparisons will be appropriate only if the varietal means are weighted with their variances. In this paper, the effectiveness of weighted linear regression analysis (WLS) with ordinary regression analysis (OLS) for computing stability parameters by using grain yield data of pearl millet varieties is demonstrated.

Statistical methodology

When the error variances of site variety means are homogeneous (Eberhart and Russell 1966; Perkins and Jinks 1968), the significance of genotype × environment interaction is tested by partition of the interaction sum of squares (SS) into SS due to linear regression and deviations from regression by the maximum likelihood procedure of least squares, which assumes equal weights for all site means. In the event that error variances (V_{ij}) of the Y_{ij} mean values are heterogeneous, the basic assumption of analysis of variance, i.e., identical and independent distribution of errors, is violated. The assumption that deviations from the regression line are normally distributed with a common variance no longer holds. In such cases, a weighted least-squares regression analysis is appropriate (Steel and Torrie 1980).

Weighted regression analysis of individual varieties

The weighted regression analysis is performed by a two-way table of Y_{ij} mean grain yield values with $i = 1 \ldots v$ varieties and $j = 1 \ldots n$ environments. The variance of each variety mean (V_{ij}) is computed from the variety × replication mean squares (σ_e^2) as σ_e^2/r, where r is the number of replicate blocks. The information of variety mean yield is the reciprocal of V_{ij} and is used as weight (W_{ij}), i.e., $W_{ij} = r/\sigma_e^2$.

Thus, the weighted total sum of squares (SS) for the i^{th} variety is

$$SS\ Y_i = \sum_{j=1}^{n} W_{ij}\, Y_{ij}^2 - \left(\sum_{j=1}^{n} W_{ij}\, Y_{ij}\right)^2 \Big/ \sum_{j=1}^{n} W_{ij}$$

and $\bar{Y}_i = \left(\sum_{j=1}^{n} W_{ij}\, Y_{ij}\right) \Big/ \sum_{j=1}^{n} W_{ij}$ (is the weighted mean of i^{th} variety across j environments).

The regression coefficient is

$$b_i = \sum_{j=1}^{n} W_{ij}\, Y_{ij}\, e_j \Big/ \sum_{j=1}^{n} W_{ij}\, e_j^2,$$

where

$$e_j = \bar{Y}_{.j} - \bar{Y}_{..}, \quad \sum_{j=1}^{n} e_j = 0.$$

The reduction in SS due to regression is:

$$\text{Regression SS} = \left[\sum_{j=1}^{n} W_{ij}\, Y_{ij}\, e_j\right]^2 \Big/ \sum_{j=1}^{n} W_{ij}\, e_j^2 \quad \text{for } 1\ df$$

and the deviation SS is obtained by subtracting the regression SS from the total SS Y_i for $n-2$ df. The weighted sum of deviations from regression is zero. These weighted regression SS and deviation SS have a χ^2 distribution for 1 and $n-2$ degrees of freedom, respectively.

Example: analysis of grain yield data of pearl millet varieties

The grain yield (kg ha^{-1}) data of 12 open-pollinated composite varieties and one hybrid (MBH 130), along with one variable check variety of pearl millet, evaluated at 26 sites with three replications at each site during 1987–88 under the All-India Coordinated Pearl Millet Improvement Project, were analyzed by the above two methods.

The coefficient of variation for grain yield ranged from 4 to 33% over test sites, reflecting the wide range (208–108,900) of error variances (V_{ij}) of varietal means at different locations. Bartlett's χ^2 test, which had $\chi^2 = 456.81$ ($25\ df$, $P \leq 0.01$), confirmed the heterogeneity of the grain yield and the effects of nonorthogonality in the data. The correlation between error variances and e_j values was not significant, which ruled out the contribution of scaler effects towards the heterogeneity.

The unbiased estimates of the variety SS, environmental SS, and genotype × environment interaction SS can be obtained by sequential model fitting using the weighted least-squares procedure (Dunteman 1984). The general analysis of variance for the weighted least-squares analysis could not be done because the weights (W_{ij}) for grain yield were too small due to large error variances, and because even the mainframe computer could not handle large matrices generated by such data. Therefore, the weighted regression analysis of individual varieties has been reported. The ordinary joint regression analysis revealed significant differences among varieties and environments along with significant genotype × environment interaction. The SS due to heterogeneity among regressions and that due to deviations from regressions were also significant when tested against their corresponding error terms, indicating that both linear and non-linear components were important, as also reported by Virk et al. (1988).

The variety mean grain yield and estimates of regression coefficients from the ordinary (OLS) and weighted least-squares (WLS) analyses are given in Table 1. The

Table 1. Mean ($\bar{Y}_i$) grain yield (kg ha^{-1}), ordinary (OLS), and weighted regression (WLS) coefficients (b_i), standard errors of regressions (SE), deviation mean squares, and coefficients of determination (r_i^2) for pearl millet varieties

Variety	$b_i \pm$SE		Deviation MS		r_i^2		$\bar{Y}_i$
	OLS	WLS	OLS 10^4	WLS$^+$	OLS	WLS	
MP 122	0.91 ±0.05	0.89* ±0.04	114	3.40**	0.93	0.96	1745
MP 131	1.09 ±0.07	1.02 ±0.04	266**	4.16**	0.90	0.97	1826
MP 143	1.05 ±0.06	0.99 ±0.04	158**	3.63**	0.93	0.97	1847
MP 153	1.02 ±0.06	1.15** ±0.05	193**	7.67*	0.91	0.95	1834
MP 154	0.92 ±0.10	1.15 ±0.08	429**	17.32**	0.79	0.89	1851
MP 155	1.08 ±0.04	1.07** ±0.02	89	1.45	0.96	0.99	2007
MP 156	1.12* ±0.05	1.04 ±0.04	137*	3.29**	0.95	0.97	1865
MP 158	1.11 ±0.06	1.06 ±0.06	195**	8.69**	0.93	0.93	1908
MP 159	1.00 ±0.06	1.06 ±0.04	189**	4.00**	0.91	0.97	1817
MP 161	0.89 ±0.08	0.84** ±0.05	298**	5.71**	0.84	0.94	1614
MP 162	0.84 ±0.08	0.88* ±0.05	287**	7.79**	0.88	0.92	1640
WC-C-75	0.96 ±0.07	0.97 ±0.04	247**	4.20**	0.88	0.96	1764
MBH 130	1.00 ±0.07	0.90 ±0.06	229**	8.83**	0.90	0.91	1912
LSD (0.05)							102

*, ** Significant at the 5% and 1% probability level, respectively.
$^+$ Tested as χ^2 for 24 df

standard errors of b_i values are invariably smaller (up to 43%) following WLS than those from OLS, showing the superiority of WLS in efficiency of parameter estimates. The significance of the deviation mean squares (MS) was unaltered, irrespective of the method of analysis, except for MP 122, where the deviation mean squares was significant ($P \leq 0.01$) in WLS but not in OLS. There was no correlation between the deviation mean squares from the two procedures ($r = 0.02$; $P > 0.05$).

The coefficients of determination (r_i^2) as proposed by Pinthus (1973), measuring the dispersion of points around the best-fitting linear regression, were consistently higher in WLS than in OLS, WLS accounting for 89–99% of the variation compared to 79–96% in OLS.

The regression coefficients from the OLS and WLS differed in their ranks, except for MP 122 and WC-C-75. The rank correlation of b_i values was $r_s = -0.0007$; $P > 0.001$. While in OLS analysis MP 155 was classified as one possessing general adaptation, as it had the highest mean grain yield of 2,007 kg ha^{-1}, $b_i \approx 1.00$, and deviation MS ≈ 0.0, the weighted regression analysis indicated that it to had specific adaptation to favorable environments, since it had $b_i > 1.0$ with non-significant deviation MS. The examination of mean grain yields for MP 155 over different sites confirmed its specific adaptation. This demonstrates that its overall high yield was largely due to its better performance in the favorable environments. Thus, the weighted regression procedure proved useful in better identifying the nature of adaptation. A change in the estimates of b_is was also detected in some other varieties.

The estimates of b_is for varieties MP 153 and MP 155 revealed their above average response ($b_i > 1.0$), while varieties MP 122, MP 161, and MP 162 had below average response with $b_i < 1.0$ in the weighted regression analysis, unlike their average response uner OLS procedure.

It may be concluded that the weighted regression analysis of genotype × environment interactions in the event of heterogeneous error variances is superior to the ordinary regression procedure for stability analysis and that it better discriminates between varieties for their adaptation, in addition to yielding more efficient estimates of the regression coefficients as measures of adaptation.

Acknowledgements. Appreciation is expressed to all the cooperating scientists in the All-India Coordinated Pearl Millet Improvement Project.

References

Dunteman GH (1984) Introduction to linear models. Sage Publications, London

Eberhart SA, Russell WA (1966) Stability parameters for comparing varieties. Crop Sci 6:36–40

Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant breeding program. Aust J Agric Res 14:742–754

Perkins JM, Jinks JL (1968) Environmental and genotype-environmental components of variability. III. Multiple lines and crosses. Heredity 23:339–356

Pinthus MJ (1973) Estimate of gentoypic value: a proposed method. Euphytica 22:121–123

Steel RGD, Torrie JH (1980) Principles and procedures of statistics – a biometrical approach, 2nd edn. McGraw-Hill, London

Virk DS, Virk PS, Mangat BK, Harinarayana G (1988) Linear and non-linear regression analysis of genotype × environment interactions in pearl millet. Theor Appl Genet 75:736–740

Yates F, Cochran WG (1938) The analysis of groups of experiments. J Agric Sci 28:536–580