Pertanika J. Trop. Agric. Sci. 18(3): 201-207(1995) ISSN: 0126-6128

Component Analyses and their Implication on the Breeding of Soya Bean (*Glycine max* (L.) Merr)

O.J. ARIYO
Department of Plant Breeding & Seed Technology
University of Agriculture
Abeokuta, Nigeria

Keywords: soya bean breeding, correlation, factor analysis, selection index

ABSTRAK

Sepuluh ciri yang digabungkan dengan hasil biji benih dalam 20 genotip kacang soya, Glycine max *(L) merr, telah dianalisis menggunakan teknik-teknik korelasi, regresi dan analisis-analisis faktor. Korelasi linear berganda (R) 0.99 dengan koefisien penentuan 0.97 adalah direkodkan. Analisis regresi menunjukkan bahawa hari-hari mengeluarkan bunga, hari-hari kematangan, ketinggian untuk menuai, bilangan dahan setiap pokok, bilangan biji setiap lenggai, pembintilan dan panjang lenggai yang banyak dihasilkan kepada variasi adalah disebabkan oleh regresi. Hanya ketinggian untuk menuai dan hari-hari kematangan, secara positif dan signifikan berkaitan dengan hasil biji benih. Analisis faktor menghasilkan keputusan yang sama dengan analisis regresi dan kolerasi ciri-ciri tanaman. Empat faktor pertama diambil kira untuk 82.53% variasi dalam struktur bergantungan.*

ABSTRACT

The characters associated with seed yield in 20 genotypes of soya bean, Glycine max *(L) Merr. were analysed using techniques of correlation, regression and factor analyses. The multiple linear correlation (R) of 0.99 with a coefficient of determination of 0.97 was recorded. The regression analysis indicated that days to flowering, days to maturity, height at harvest, number of branches per plant, number of seeds per pod, nodulation and pod length contributed substantially to the variation due to regression. Only height at harvest and days to maturity were positively and significantly correlated with seed yield. Factor analysis produced a similar result to those of plant character correlation and regression analysis. The first four factors accounted for 82.53% of the variation in the dependence structure.*

INTRODUCTION

Seed yield in soya bean is a complex character influenced by the interplay of many other characters. Knowledge of the relationship of yield with its main components is important in plant breeding, particularly for indirect selection for quantitative traits, such as seed yield, that exhibit low heritability. Beside the yield components, physiological and morphological characters of soya bean plants are known to play a major and interdependent role in determining seed yield (Denis and Adams 1978; Bartual *et al.* 1985).

Correlation studies between characters and the use of multivariate analysis to determine the relative contribution of different characters to the total variation are of great value in determining the most effective breeding procedures (Bhatt (1976) in wheat; Ghaderi *et al.* (1979) in mung bean; Broich and Palmer (1980) and Ariyo (1995) in soya bean; Ariyo (1991a, 1991b, 1993 in okra).

The breeder has a number of desirable characters in mind when carrying out selection. To maximize improvement in the character of choice, selection is generally applied simultaneously to several other characters that influence the character of choice. Falconer (1960) reported that the most rapid improvement of economic value was expected from selection applied simultaneously to all the components. Use of selection index gives adequate weight to each of the desirable characters identified. Hazel and Lush (1942) and Falconer (1981) reported that selection based on such an index was more efficient than selecting individuals on various characters. In addition, Ariyo (1991a) reported that the selection index should be used in conjunction with yield data preferably obtained across contrasting environments to produce valuable results.

The objectives of this study were:

(a) to determine the relative importance of various characters of soya bean and the relationship among them, and

(b) to construct selection indices for seed yield.

MATERIALS AND METHODS

Twenty genotypes, consisting of early, medium and late maturing varieties of soya bean from the International Institute of Tropical Agriculture (IITA), Ibadan were grown in a randomized complete block design with three replications. The planting was done at the University of Agriculture, Abeokuta, in July, 1991. Each entry was grown in four-row plots of 6 x 3m but only the competitive plants in the two inner rows were observed. Following planting, a mixture of 4 l Galex and 1 l Gramaxone in water was sprayed per hectare to control weeds. Subsequent weeding was done manually.

The number of days to flowering was recorded as the date the plants of a genotype attained 50% flowering. Maturity was when the pods had turned brown just before shattering. Shattering was taken as the proportion of guard rows that shattered two weeks after the two inner competitive plants had been harvested to determine seed yield; nodulation was assessed on the size and number of nodules at full bloom. Lodging was scored on the proportion of the plants that fell down at harvest. Genotype means, averaged across replications, were used for statistical analysis.

Phenotypic a coefficients of correlation were calculated among all the characters evaluated following the procedure of Steel and Torrie (1968). Step-wise multiple regression analysis (forward selection) was performed as outlined by Draper and Smith (1966) by which the multiple-regression equation and multiple coefficient of determination (R2) were obtained by adding independnet variables, one at a time depending on their relative importance, in determining dependent variables. Analysis was terminated when the proportion of dependent variance explained by adding each of the remaining variables was not significant at 0.05 level of probability. In this study, seed yield was fitted as a linear function of the other ten characters. The sequential contribution of each character to the total variation in seed yield was determined by the forward selection procedure. On the basis of the relative importance of each character to seed yield, selection indices were contructed.

Data were also subjected to factor analysis according to the procedure of Cattell (1965). The analysis produced factor loading as well as communality for each character from the variance-covariance matrix of the 20 genotypes. The factors were ranked on the basis of the magnitude of variability explained in the dependent structure. When the contribution of a factor to the variability was less than 10%, the process was terminated. The particular combination of variables that form a factor accounted for more of the variance of the data as a whole than any other linear combination variable. Therefore, Factor 1 was the best combination of the linear relationships in the data. Factor 2 was the best linear combination of variables that accounted for most of the residual variance after the effect of Factor 1 had been removed. Subsequent factors contributed progressively less to the total variance.

Communality is the amount of variance of a variable accounted for by all factors collectively and it is the R^2 value obtained by regressing a variable on all other variables in the model (Lee and Kaltsikes 1973; Eckert and Westfall 1975).

RESULTS

Table 1 presents the correlation coefficients between the various characters evaluated. Only height at harvest was strongly correlated with seed yield (r=0.65), while number of days to maturity was moderately correlated with seed yield (r=0.44). Height at harvest was also positively correlated with days to flowering (r=0.78) and maturity (r=0.57) while pod shattering was negatively correlated with days to maturity (r=0.75) and height at harvest (r=0.58). Number of branches per plant exhibited a positive relationship with days to flowering (r=0.59), days to maturity (r=0.55), height at harvest (r=0.58), and lodging at harvest (r=0.44) but was negatively correlated with pod shattering (r=-0.62). On the

TABLE 1
Evaluation of correlations among the eleven characters of soya bean

Characters	Seed Yield/ plot (g)	Days to Flowering	Days to Maturity	Height at Harvest (cm)	Lodging at Harvest	Pod Shattering	Number of Branches/ Plant	Number of Pods/Plant	Number of Seeds/Pod	Nodulation
Days to Flowering	0.30									
Days to Maturity	0.44*	0.35								
Height at Harvest (cm)	0.65**	0.78**	0.57**							
Lodging at Harvest	0.09	0.22	0.34	0.06						
Shattering	-0.35	-0.30	-0.75**	-0.58**	-0.01					
Number of Branches/Plant	0.22	0.59**	0.55*	0.58**	0.44*	-0.62**				
Number of Pods/Plant	-0.10	0.52*	0.38	0.29	0.61**	-0.28	0.80**			
Number of Seeds/Pods	0.34	0.08	0.07	0.19	-0.49*	0.00	-0.13	-0.36		
Nodulation	0.28	0.12	0.46*	0.28	0.01	-0.53*	0.44*	0.39	-0.19	
Pod Length (cm)	0.29	-0.14	-0.45*	0.10	-0.10	0.14	-0.27	-0.31	-0.27	-0.24

* Significant at P = 0.05; ** Significant at P = 0.01

other hand, number of pods per plant showed a correlation of r=0.52, 0.61 and 0.80 with days to flowering, lodging at harvest and number of branches per plant respectively. Nodulation was positively correlated with days to maturity and number of branches per plant, but negatively correlated with shattering. Pod length was negatively correlated with days to maturity (r=-0.45).

Four factors were obtained by factor analysis (Table 2); together these accounted for 82.53% of the variance for all the 11 characters. This implies that the factor-analysis model used in this study was effective in illuminating the unique variance of each variable. Communalities ranged from 0.6728 to 0.9360. Generally, factor loadings of 0.7 - 0.9 would be considered high loadings and those from 0 to 0.2 as low loadings (Denis and Adams 1978). In this study, however, for the purpose of interpretation only characters exhibiting factor loadings of 0.5 or larger were considered important and no character was loaded on more than one factor. It should be noted that the choice of loading of 0.5 or greater is arbitrary and does not imply biological significance. Biological interpretation of factors depends largely on the genotypes evaluated, the particular sets of characters measured, and how well the researcher understands the biology of the organism (Fakorede 1979). These are limitations of factor analysis.

Factor 1, which accounted for 39.44% of the total variance, contained days to flowering, days to maturity, height at harvest, shattering, number of branches per plant and number of pods per plant. The influence of a factor on a trait is determined by the square of the factor loading for that trait (Lee and Kaltsikes 1973). Therefore, Factor 1 accounted for 33% of the variance due to seed yield in this study. In addition, Factor 1 related essentially to the physiological aspect of the crop. Apart from number of pods per plant, other characters loaded in Factor 1 had direct bearing on a crop's growth and development.

Lodging at harvest and number of seeds per pod were grouped under Factor 2, and this factor accounted for 19.34% of the variance in the data as a whole. Factor 3 comprised only pod length, while Factor 4 was nodualtion; most characters in Factor 1 were correlated with each

TABLE 2
Evaluation of communalities and the factor loadings of eleven characters

Traits	Communalities	Factor Loadings 1	2	3	4
Factor 1					
Days to Flowering	0.8380	0.6940	0.1088	0.2378	0.5371
Days to Maturity	0.7694	0.8130	0.1252	-0.2680	-0.1441
Plant Height	0.9331	0.7724	0.4738	0.2929	0.1621
Shattering	0.7938	-0.7530	-0.2327	0.1695	0.3792
Number of Branches/Plant	0.8258	0.8678	-0.2480	-0.0034	0.1061
Number of Pods/Plant	0.8759	0.7017	-0.5956	0.0384	0.1654
Factor 2					
Lodging at Harvest	-0.6102	0.4136	-0.2327	0.3402	0.1172
Seed Yield/Plant	0.7885	0.4371	0.6715	0.3542	-0.1450
Seeds/Pod	0.8923	-0.0681	0.7234	-0.4168	0.4333
Factor 3					
Pod Length	0.9359	-0.2913	0.2312	0.8469	-0.2834
Factor 4					
Nodulation	0.7518	0.5794	-0.0260	-0.2484	-0.5948
Percentage of Total Variation		39.44	19.34	13.07	10.68
Cumulative Percentage		39.44	58.78	71.85	82.53
Eigen Value		4.33	2.13	1.43	1.17

other while correlations between characters were observed across factors.

Table 3 presents linear regression analysis of yield components. The analysis showed that all except lodging at harvest, shattering and number of pods per plant contributed significantly to the total variation due to regression. The multiple linear correlation coefficient between seed yield and the other ten characters was R=0.99 given a coefficient of determination of 0.97.

Table 4 gives the selection indices, using various character combinations and their relative effectiveness. The high value of the coefficient of determination indicated that the ten characters accounted for most of the variation

TABLE 3
Multiple linear regression analysis of components of seed yield in soya bean

Source	DF	MS
Regression	10	205,820**
Days to flowering	1	194,393**
Days to maturity	1	271,429**
Height at harvest (cm)	1	673,161**
Lodging at harvest	1	46,325
Shattering	1	55,814
Number of branches/pods	1	179,221**
Number of pods/plant	1	4,476
Number of seeds/pod	1	188,524**
Nodulation	1	252,840**
Pod length (cm)	1	192,021**

in seed yield. The step-wise regression indicated that nine characters accounted for 97% of the total variation. The remaining character, shattering, did not meet the a = 0.5 significance level for entry into the model. Plant height alone accounted for 42.46% of the total variation due to regression. By inclusion of number of days to flowering, 53.62% of the total variation was explained. Adding number of branches per plant, lodging, number of seeds per pod, nodulation and pod length, 92.47% of the total variation was accounted for. Inclusion of days to maturity and number of pods per plant separately occasioned a marginal increase of about 2% in each case. It is generally accepted that correlation may not always adequately indicate the realtionship among variables. This is why multivariate rather than bivariate statistical methods are preferred.

DISCUSSION

The high correlation between days to flowering, days to maturity and height at harvest suggests that short varieties flowered earlier and matured earlier. However, high correlation between height at harvest and seed yield indicates that tall varieties yielded more than short varieties. The correlation between branches per plant, pods per plant and days to flowering indicates that late flowering plants produced more seed-bearing branches, which were possibly responsible for correlation between height at harvest and seed yield as only tall varieties could accommodate many branches. Late flowering varieties matured late and grew taller,

TABLE 4
Selection indices in soya bean for yield. Multiple regression equation

	R^2(%)
$Y1 = -11.62 + 0.45X_1$	42.46
$Y2 = 12.02 - 0.74X_1 - 0.86X_2$	53.62
$Y3 = 13.07 + 0.85X_1 - 0.69X_2 - 1.64X_3$	59.08
$Y4 = 6.86 + 0.9X_1 - 0.82X_2 - 2.61X_3 + 12.99X_4$	66.75
$Y5 = -8.29 + 0.85X_1 - 0.84X_2 - 2.44X_3 + 18.54X_4 + 4.95X_5$	73.10
$Y6 = -24.47 + 0.76X_1 - 0.62X_2 - 3.62X_3 + 23.39X_4 + 6.85X_5 + 3.13X_6$	82.04
$Y7 = -96.54 + 0.43X_1 - 0.15X_2 - 2.83X_3 + 17.35X_4 + 11.92X_5 + 4.92X_6 + 9.70X_7$	92.47
$Y8 = -222.87 + 0.06X_1 + 0.35X_2 - 2.54X_3 + 22.06X_4 + 13.17X_5 + 4.83X_6 + 15.73X_7 + 1.03X_8$	95.13
$Y9 = -243.86 - 0.08X_1 + 0.64X_2 - 1.41X_3 + 24.80X_4 + 13.01X_5 + 5.39X_6 + 16.38X_7 + 1.01X_8 + 13.01X_9$	97.10

X_1 = Height at harvest; X_2 = Days to flowering; X_3 = Number of branches/plant; X_4 = lodging;
X_5 = Number of seeds/pod; X_6 = Nodulation; X_7 = Pod length; X_8 = Days to maturity; X_9 = Number of pods/plant

besides producing more branches. That such varieties are susceptible to lodging suggests that a compromise must be struck between high yield and loss due to lodging at harvest. Correlation between pods per plant and days to flowering indicated that late flowering varieties which were also late maturing produced more pod-bearing branches. Of interest is the correlation between pods per plant and lodging at harvest. This implied that top-heavy varieties were likely to lodge under the weight of the pods.

The study identified days to flowering, days to maturity, number of branches per plant, height at harvest and pods per plant, some of these characters correlated among themselves, as important yield components in soya bean.

The results of regression analysis were complementary to the correlation studies by highlighting the relative weight of the various characters to the total variation. Height at harvest contributed the highest variance to yield followed by days to maturity, days to flowering, pod length and number of branches per plant in diminishing order of contribution. Similarly, the factor model identified the contribution of all these characters to the dependence structure. Days to flowering, days to maturity, height at harvest, number of branches per plant, number of pods per plant and pod length had high communalities, suggesting their importance in the dependence structure.

The various selection indices showed that height at harvest alone accounted for 42.46% of the total variation due to regression, demonstrating a multiple correlation coefficient (R) of 0.66. A selection index based on height at harvest and days to flowering gave the coefficient of determination of 53.62% of the total variable. The coefficient of determination continued to increase as more characters were entered, and terminated at R^2 of 97.10%. However, there was a marginal decrease in R^2 value as more characters were progressively added, indicating that earlier characters were more important than later entries. Although number of pods per plant, days to maturity and pod length were ranked higher in importance than lodging at harvest, nodulation and number of seeds per pod, these were nonetheless also important components of pod yield.

In breeding soya bean for high yield, therefore days to flowering, which to a large extent determines days to maturity and height at harvest, is a premium character. Other characters such as number of branches per plant, pod length and nodulation should also be considered. Similar observations were also reported by Denis and Adams (1978). Since it is better to apply selection generally to several characters that influence yield, a knowledge of the inheritance of the various characters in the selection index is required.

This study gives an insight into the association among traits observed in a set of genotypes. A statistical demonstration of association, whether by correlation, regression or factor analysis, however, does not provide information on the causative agents (genetic, physiological, morphological or environmental) (Fakorede 1979). Genetic analysis is necessary whenever the deterministic relationship is of interest.

Factor analysis has the limitations of the arbitrary and subjective nature of interpretation of factors and the dependence on the number and loadings of factors on the particular set of genotypes and variables. Despite these limitations, however, its data-reducing capacity gives it an advantage over correlation and regression analyses.

ACKNOWLEDGEMENT

The author acknowledges the contribution of the Grain Legume Improvement Programme, International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria towards the successful completion of this study.

REFERENCES

ARIYO, O.J. 1991a. Effectiveness of classical selection index in discriminating desirable varieties of okra (*Hibiscus esculentus*). *Indian J. Agric. Sci.* **60:** 793-795.

ARIYO, O.J. 1991b. Regression analysis of pod yield and yield components in okra *(Abelmoschus esculentus* (L). Moench). *J. Agric. Sci. Technol.* **1:** 72-74.

ARIYO, O.J. 1993. Genetic diversity in West African okra *(Abelmoschus caillei* (A. Chev.) stevels - Multivariate analysis of morphological and agronomic characteristics. *Genet. Res. Crop. Evol.* **40:** 25-32.

ARIYO, O.J. 1995. Correlation and path-coefficient analysis of components of seed yield in soybeans. *African Crop. Sci. J.* **3:** 29-33.

BARTUAL, R., E.A CARBONELL and D.E. GREEN. 1985. Multivariate analysis of a collection of Soybean cultivars for southwestern Spain. *Euphytica* **34:** 113-123.

BHATT, G.M. 1976. An application of multivariate analysis to selection for quality characters in wheat. *Austral. J. Agric.* **27:** 11-18.

BROICH, S.L. and R.G. PALMER. 1980. A cluster analysis of wild and domesticated soyabean phenotypes. *Euphytica* **29:** 23-32.

CATTELL, R.B. 1965. Factor analysis. An introduction to essentials 1. The purpose of underlying models. *Biometrics* **21:** 23-32.

DENIS, J.C. and M.W. ADAMS. 1978. A factor analysis of the plant variables related to yield in dry beans. 1. Morphological traits. *Crop. Sci.* **18:** 74-78.

DRAPER, N.R. and H. SMITH. 1966. *Applied Regression Analysis.* 2nd edn. New York: Wiley.

ECKERT, R.T. and R.D. WESTFALL. 1975. The factor analysis of multivariate data systems *Northwestern Forest. Tree Improvement Conf. Proc.* **22:** 41-52.

FAKOREDE, M.A.B. 1979. Inter-relationship among grain yield and agronomic traits in a synthetic population of maize. *Maydica* **24:** 181-192.

FALCONER, D.S. 1981. *Introduction to Quantitative Genetics.* 2nd edn. London: Longman.

GHADERI, A., M. SHSHEGAR and B. EHDAIE. 1979. Multivariate analysis of genetic diversity for yield and its components in mung bean. *J. Amer. Soc. Hort. Sci.* **104:** 728-731.

HAZEL, L.N. and J.L. LUSH. 1942. The efficiency of three methods of selection. *J. Hered.* **33:** 393-399.

LEE, J. and P.J. KALTSIKES. 1973. Multivariate statistical analysis of grain and agronomic characters in durum wheat. *Theor. Appl. Genet.* **43:** 226-231.

STEEL, R.G.B. and J.H. TORRIE. 1968. *Principles and Procedures of Statistics.* 2nd edn. New York: McGraw Hill.

(Received 26 June 1995)
(Accepted 3 January 1996)