Theoretical and Applied Genetics 43, 226–231 (1973)

Multivariate Statistical Analysis of Grain Yield and Agronomic Characters in *Durum* Wheat[1]

J. LEE and **P. J. KALTSIKES**

Plant Science Department, University of Manitoba, Winnipeg, Manitoba (Canada)

Summary. Correlation, stepwise multiple regression and factor analyses were conducted on grain yield and a number of agronomic characters in the parental, F_1 and F_2 families originating from a 10×10 diallel cross in durum wheat. For the F_1 diallel, the correlation analysis indicated that the number of spikes and kernels per plant and 1,000 kernel weight had the highest correlations with grain yield; for the F_2 diallel, the number of spikes and kernels per meter, 1000 kernel weight and plant height showed most striking correlations with same.

Stepwise multiple regression analysis indicated that, for the F_1 diallel, number of kernels per plant, 1000 kernel weight and days to maturity were the most potent predictor variables for grain yield, accounting for 78% of its variability. For the F_2 diallel, the number of kernels and number of spikes per meter, 1000 kernel weight and number of kernels per spike were the most potent predictors for grain yield, accounting for 91% of its variability. Five common factors were extracted which explained 98.8% and 98.1% of the total variance in the F_1 and F_2 diallel, respectively. However, the importance of each of the five factors and the characters which loaded highly on each of them differed from generation to generation.

Introduction

In a plant breeding program, a breeder usually records data and makes his selections on the basis of a large number of agronomic characters among which significant positive and negative correlations may exist (Krishnamurthy, 1958; Johnson and Schmidt, 1968; Lebsock and Amaya, 1969; Hsu and Walton, 1970; Kaltsikes and Larter, 1970; Kaltsikes and Lee, 1971). Therefore, any analytical method which could result in a reduction of the number of agronomic characters to be recorded without sacrificing a significant amount of information would be a major asset to the breeder. Traditionally, correlation studies have been employed for this purpose. However, a correlation between two characters does not necessarily imply a cause-and-effect relationship. In fact, many unknown factors could have produced their numerical association. Alternatively, multiple regression analysis can be useful when the main interest is on the prediction of the response variable from a set of predictor variables. Grain yield is logically chosen as the response variable while other agronomic characters are the predictor variables. Furthermore, factor analysis, a multivariate statistical technique, is useful in explaining the inter-correlations among a set of selected variables (Lawley and Maxwell, 1963; Harman, 1967). It also helps in ascertaining the number and nature of common causative influences on which more intensive work can be concentrated. Practically, factor analysis can be used to select a set of fewer characters based on the structural interrelationships among the original set of characters.

[1] Contribution No. 323 from Plant Science Department, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada.

The present study was undertaken to provide information for two generations of a diallel cross of durum wheat (*Triticum turgidum* L. var. *durum*) with the following objectives: 1.) To identify agronomic characters which are important predictors for grain yield by means of a multiple stepwise regression analysis and 2.) To ascertain whether a smaller set of common causative influences (factors) could be isolated by means of factor analysis which would explain the inter-relationships among the original set of characters.

Materials and Methods

The ten cultivars of durum wheat (*Triticum turgidum* L. var. *durum*) used in this study and their respective country of origin were: Adur (France), Candeal Selection (Argentina), DT-310 (Canada), Iumillo (Italy), Kharkov Kaja (Russia), Leeds (USA), Madif (Italy), MY-54 (Mexico), Narodnaja (Russia) and Stewart 63 (Canada). These cultivars were crossed in a diallel fashion with reciprocal families bulked to yield 45 F_1 and subsequently 45 F_2 families. Hercules, a high performance, locally adapted cultivar, was also included in the study. Altogether, there were 56 entries (11 inbreds and 45 hybrids) in each of the two diallel generations. Seeds were sown in May 1971 at Winnipeg, Manitoba and Swift Current, Saskatchewan. At each location, the F_1 and F_2 diallel experiments were separately laid out in a randomized complete block design with two replications as follows: Each F_1 plot consisted of a single 3-meter row with 15 seeds space-planted. The number of plants surviving to harvest ranged from 8 to 15. Each F_2 plot consisted of three 3-meter rows with 160 seeds sown per row. A guard row was sown between plots to minimize inter-plot competition. Duplicate plots were sown for each of the eleven inbreds. Thus, each replicate for the F_2 diallel consisted of 67 plots. The inter-row distance for both diallels was 30 cm.

The following characters were measured from each plot:

1.) Grain yield. Yield observations consisted of the weight (gm) of seed from each plot. In the F_1 diallel,

Table 1. *Phenotypic correlation matrix for 12 characters in a 10 × 10 F_1 diallel of durum wheat*

	Yield/plant	Yield/plot	Spikes/plant	1,000 KW	Spikelets/spike	Florets/spike	Kernels/spike	Proportion of florets with seeds	Kernels/spikelet	Kernels/plant	Days to maturity
Yield/plot	0.88***	—									
Spikes/plant	0.80***	0.91***	—								
1,000 KW	0.74***	0.73***	0.62***	—							
Spikelets/spike	0.03	0.06	0.07	0.06	—						
Florets/spike	0.01	0.06	0.02	0.01	0.59***	—					
Kernels/spike	0.45***	0.45***	0.40***	0.35***	0.46***	0.65***	—				
Florets with seeds	0.57***	0.50***	0.48***	0.44***	−0.10	0.32***	0.50***	—			
Kernels/spikelet	0.49***	0.47***	0.41***	0.35***	−0.02	0.40***	0.87***	0.64***	—		
Kernels/plant	0.81***	0.88***	0.94***	0.62***	0.21***	0.24***	0.65***	0.54***	0.61***	—	
Days to maturity	0.67***	0.59***	0.51***	0.62***	−0.05	0.02	0.38***	0.47***	0.46***	0.53***	—
Plant height	0.58***	0.56***	0.47***	0.64***	0.27***	0.14*	0.40***	0.35***	0.31***	0.52***	0.45***

*** Significant $P \leq 0.001$. — * Significance $P \leq 0.05$. — n = 224.

Table 2. *Phenotypic correlation matrix for 11 characters in a 10 × 10 F_2 diallel of durum wheat*

	Yield/plot	Spikes/meter	1,000 KW	Spikelets/spike	Florets/spike	Kernels/spike	Proportion of florets with seeds	Kernels/spikelet	Kernels/meter	Days to maturity
Spikes/meter	0.90***	—								
1,000 KW	0.85***	0.73***	—							
Spikelets/spike	0.61***	0.54***	0.61***	—						
Florets/spike	0.76***	0.68***	0.69***	0.75***	—					
Kernels/spike	0.76***	0.66***	0.67***	0.71***	0.94***	—				
Florets with seeds	0.17**	0.11	0.11	0.08	0.07	0.40***	—			
Kernels/spikelet	0.63***	0.55***	0.51***	0.33***	0.79***	0.89***	0.50***	—		
Kernels/meter	0.91***	0.96***	0.75***	0.63***	0.82***	0.83***	0.22***	0.71***	—	
Days to Maturity	0.50***	0.47***	0.53***	0.55***	0.61***	0.59***	0.08	0.44***	0.54***	—
Plant Height	0.80***	0.73***	0.78***	0.72***	0.69***	0.67***	0.10	0.44***	0.76***	0.52***

*** Significant $P \leq 0.001$. — ** Significant $P \leq 0.01$. — n = 268.

grain yield per plant was derived by dividing plot yield by the number of plants surviving to harvest.

2.) Spike number. In the F_1 diallel, the number of fertile spikes per plant was determined by dividing the total number of fertile spikes by the number of plants surviving to harvest. In the F_2 diallel, the number of spikes per linear meter was determined by direct count.

3.) 1,000 Kernel weight (gm).

4.) Number of spikelets per spike.

5.) Number of florets per spike.

6.) Number of kernels per spike.

7.) Proportion of florets bearing seeds.

8.) Number of kernels per spikelet. Characters (4) through (8) were taken from the primary spike. Two and ten primary spikes were sampled from each F_1 and F_2 plot, respectively.

9.) Kernel number. In the F_1 diallel, the number of kernels per plant was obtained by dividing the total number of kernels by the number of plants surviving to harvest; in the F_2, the number of kernels per linear meter was estimated by multiplying the number of kernels per spike by number of spikes per linear meter.

10.) Maturity. The number of days to physiological maturity.

11.) Plant Height. Height (cm) was taken from the ground to the tip of the tallest spike, excluding own, on individual plants in the F_1 and on 10 randomly sampled spikes in the F_2.

Values for each of the above characters were averaged over each plot (224 and 268 F_1 and F_2 plots in all, respectively) prior to all statistical analyses. The F_1 and F_2 diallels were separately subjected to the product-moment correlation and stepwise multiple regression analyses as described by Draper and Smith (1967). According to this method, the multiple regression equation and multiple co-coefficient of determination were obtained by adding one independent variable at a time according to their relative importance in determining the response variable (grain yield). Termination of analysis occurred when the introduction of a new independent variable resulted in explaining less than 1% of variance of the response variable. The same two sets of data were also each subjected to factor analysis. The extraction of the original factor matrix was by principal factoring with iterations (Cattell, 1965). The initial estimate of the communality was the squared multiple correlation of each variable with every other in the original correlation matrix. Derivation of solution was through iterations until the communalities converged with a 0.001 criterion. The criterion used to determine the number of factors to be retained was an eigenvalue of ≥ 0.20. The factor matrix so obtained was rotated to simple structure by the varimax method (Harman, 1967).

Results

I. Correlation Analysis

F_1 diallel. All of the agronomic characters, excepting the number of spikelets and florets per spike, were positively correlated with grain yield per plant and per plot (Table 1). The most striking corrrelations were the number of spikes per plant and number of kernels per plant vs. grain yield per plot, followed by the correlation between 1000 kernel weight and grain yield per plant and per plot. Most of the agronomic characters included in the present investigation showed significant positive correlations amongst themselves. The only significant negative correlation was between the proportion of florets with seeds and number of florets per spike.

F_2 diallel. All of the agronomic characters were highly associated with grain yield per plot (Table 2). The most striking correlations with grain yield were the number of spikes per linear meter ($r = 0.90$), the number of kernels per meter ($r = 0.91$), 1000 kernel weight ($r = 0.85$) and plant height ($r = 80$). Most of the agronomic characters were highly correlated amongst themselves. Whereas the proportion of florets with seed showed significant negative correlation with the number of florets per spike in the F_1 diallel ($r = -0.32$), these two characters were independent of one another in the F_2 diallel ($r = 0.07$).

II. Stepwise Multiple Regression Analysis

F_1 diallel. The number of kernels per plant (X_{10}), 1000 kernel weight (X_4) and days to physiological maturity (X_{11}) were the most potent predictors for grain yield, Y (Table 3); together they accounted for 78% of the variability for grain yield per plot. When all of the remaining predictor variables were forced into the regression equation, only an additional 2% of the variability of grain yield was explained. Clearly the other characters were of limited value in predicting grain yield. The best multiple linear regression equation derived from the analysis was therefore $Y = 0.04\ X_{10} + 0.60\ X_4 + 1.12\ X_{11}$ ($R^2 = 0.78^{**}$). The multiple coefficient of determination was only moderately high since 22% of the variance for grain yield was not accounted for by the above regression equation.

F_2 diallel. The number of kernels per linear meter (X_9), 1000 kernel weight (X_3), number of spikes per linear meter (X_2) and number of kernels per spike (X_6) were the most potent predictors for grain yield, Y (Table 3); together, they accounted for 91% of the variability for grain yield. When the remaining characters were forced into the regression equation, only an additional 1% of the variability of grain yield was

Table 3. *Partial regression coefficients from the stepwise multiple regression analysis on grain yield per plant in a 10 × 10 diallel of durum wheat*[+]

F_1 diallel (n = 224)

Kernels/plant	1,000 KW	Days to maturity	R^2
0.06***	—	—	0.65***
0.04***	0.81***	—	0.75***
0.04***	0.60***	1.12***	0.78***

F_2 diallel (n = 268)

Kernels/meter	1,000 KW	Spikes/meter	Kernels/spike	R^2
0.28***	—	—	—	0.83***
0.19***	28.45***	—	—	0.89***
0.11***	28.09***	5.19***	—	0.90***
−0.23***	23.48***	19.28***	38.86***	0.91***

[+] Characters explained less than 0.01 of the variance for yield per plant were not included in table.
*** Significant $P \leq 0.001$

Table 4. *Varimax rotated factor matrices for yield and agronomic characters in a 10 × 10 diallel of durum wheat*

Character	Factor 1	2	3	4	5	Communality
			F_1 diallel			
Yield/plant	0.654	0.197	0.576	0.016	0.156	0.823
Yield/plot	0.794	0.167	0.495	0.014	0.077	0.909
Spikes/plant	0.930	0.116	0.290	0.023	0.103	0.973
1,000 Kernel weight	0.413	0.096	0.749	0.017	0.086	0.748
Spikelets/spike	0.039	0.025	0.056	0.877	−0.093	0.783
Florets/spike	0.023	0.506	−0.001	0.594	−0.650	1.000
Kernels/spike	01250	0.802	0.195	0.493	0.089	0.993
Florets with seeds	0.297	0.431	0.278	−0.083	0.762	0.938
Kernels/spikelet	0.247	0.966	0.198	−0.073	0.096	1.000
Kernels/plant	0.860	0.350	0.288	0.187	0.097	0.989
Days to maturity	0.325	0.254	0.581	−0.093	0.105	0.527
Plant height	0.257	0.115	0.693	0.249	0.082	0.625
			F_2 diallel			
Yield/plot	0.799	0.264	0.317	0.117	0.132	0.839
Spikes/meter	0.942	0.176	0.223	0.070	0.165	1.000
1,000 Kernel weight	0.610	0.299	0.232	0.058	0.265	0.588
Spikelets/spike	0.318	0.852	0.213	0.033	0.234	0.928
Florets/spike	0.404	0.450	0.825	−0.045	0.220	1.000
Kernels/spike	0.399	0.426	0.652	0.394	0.251	0.984
Florets with seeds	0.050	0.026	0.096	0.764	0.004	0.596
Kernels/spikelet	0.338	−0.018	0.761	0.533	0.195	1.000
Kernels/meter	0.830	0.265	0.397	0.179	0.202	0.987
Days to maturity	0.272	0.300	0.260	0.036	0.613	0.608
Plant height	0.600	0.472	0.180	0.059	0.224	0.668

explained. The best multiple linear regression equation derived from the analysis was $Y = -0.23\ X_9 + 23.48\ X_3 + 19.28\ X_2 + 33.86\ X_6$ ($R^2 = 0.91^{**}$). According to this regression equation 9% of the variability for grain yield could not be explained by the above four predictor variables.

III. Factor Analysis

F_1 diallel. The factor loadings and communalities in the varimax rotated matrix are presented in Table 4*. Five common causative influences (factors) which explained the structural inter-relationships among yield and agronomic characters were extracted. Together these five factors accounted for 98.8% of the variability for the twelve correlated characters. The communalities or amount of variance of a character accounted for by all factors taken together, ranged from 0.53 (days to maturity) to 1.00 (florets per spike and kernels per spikelet). For the purpose of interpretation, only those factor loadings greater than 0.6 were considered important. With this criterion, no characters loaded on more than one factor (Table 5).

The most important factor (factor 1) contained the characters grain yield per plant and per plot and number of spikes and kernels per plant (Table 5). This result implies that the expressions of the four characters were simultaneously influenced by some common underlying force. The magnitude of the influence of a factor on a character is the factor loading for that character; that is, the proportion of the variance of a character accounted for by a factor is the square value of the factor loading. Thus, factor 1 accounted for 82% of the variance for the number of spikes per plant. Similarly, factor 2 contained number of kernels per spikelet and per spike; factor 3 contained 1000 kernel weight and plant height; and factors 4 and 5 contained the number of spikelets per spike and the proportion of florets with seeds, respectively.

F_2 diallel. Five factors were extracted from the correlation matrix (Table 4) which explained 98.1% of the variation for the eleven correlated characters. The communalities ranged from 0.59 (1000 kernel weight) to 1.00 (number of spikes per meter, number of florets per spike, and number of kernels per spikelet). The most important factor (factor 1) contained grain yield per plot, number of spikes and kernels per meter, 1000 kernel weight and plant height. Factor 2 contained the number of spikelets per spike; factor 3 contained number of florets and kernels per spike and number of kernels per spikelet; factors 4 and 5 contained the proportion of florets with seeds and days to maturity, respectively.

Table 5. *Summary of factor loadings for yield and agronomic characters in a 10 × 10 diallel of durum wheat*

Factor	Characters of F_1 diallel
1	Grain yield per plant and per plot; spikes and kernels per plant
2	Kernels per spikelet and per spike
3	1,000 kernel weight and plant height
4	Spikelets per spike
5	Proportion of florets with seeds
	Characters of F_2 diallel
1	Grain yield per plot; spikes and kernels per meter; 1,000 kernel weight; plant height
2	Spikelets per spike
3	Florets and kernels per spike; kernels per spikelet
4	Proportion of florets with seeds
5	Days to maturity

* The importance of a particular factor as indicated by the proportion of variance accounted for by that factor in the initial or unrotated factor matrix is of no interest in the terminal or rotated factor matrix as the result of rotation. Nevertheless, the relative importance of the factors is still reflected by their order (i.e. factor 1 is relatively the most important factor, etc. . . .).

Discussion

In both generations most of the agronomic characters were highly associated with grain yield and amongst themselves. Although plant height was correlated with yield in both generations in the present study, this association does not occur with all genotypes of durum wheat. Kaltsikes and Lee (1971) found no association between height and yield from a 6-parent diallel in durum wheat while Kaltsikes and Larter (1970) found significant correlation between these two characters from a study of five inbreds of durum wheat grown in the Canadian western cooperative test. Lebsock and Amaya (1969) found correlation between height and yield only in certain crosses and generations in durum wheat.

Examination of the correlation matrix showed that correlation between number of spikes per plant and grain yield was among the highest in the F_1 data (Table 1) and yet this character accounted for less than 1% of the variance for grain yield (Table 3). However the correlation between number of kernels and spikes per plant was 0.94. The introduction of number of kernels into the regression equation was evidently sufficient in explaining the variation in yield which was due to these two variables and once one was introduced the other became numerically superfluous. From the biological point of view both characters are naturally important yield components. A similar result was found with the F_2 generation. Here although number of spikes per meter had a correlation coefficient of 0.90 with grain yield it only accounted for 1% of the variability for grain yield (Table 3).

The results obtained from the stepwise multiple regression analysis for both generations indicated that the two most potent predictors for grain yield were the number of kernels per plant (F_1) and per linear meter (F_2) and kernel weight. Days to physiological maturity was an important predictor variate for grain yield only in the F_1 generation while the number of spikes per linear meter and number of kernels per spike were important predictors for grain yield in the F_2 generation. Based on the present investigation, it is suggested that the number of kernels per plant or per linear meter, 1000 kernel weight, and to a lesser extent, days to physiological maturity and number of kernels per spike be given due importance as predictors for grain yield in selection programmes in durum wheat.

Factor analysis of the results of the yield and agronomic variates conducted on data from the two generations indicated that the importance of each of the factors extracted and the component variates belonging to individual factors differed from generation to generation. In the F_1, it was shown that grain yield and number of spikes and kernels per plant were simultaneously influenced by a common underlying factor whereas in the F_2, grain yield, number of spikes and kernels per meter, 1000 kernel weight and plant height were simultaneously influenced by a common underlying factoı. Other differences can be seen from Table 5. These results were not completely unexpected since the correlation matrices for the two generations were apparently heterogeneous with respect to a number of characters. The discrepancies of results between the two generations can be attributed to several possible causes: 1.) There is no genetic segregation in the F_1 whereas segregation occurred in the F_2; therefore, the latter generation was subjected to genetic sampling error; 2.) Plot size varied considerably between the two generations; therefore, each generation was subjected to varying degrees of sampling errors; 3.) F_1 material was space-planted whereas solid-planting was used in the F_2 material.

In a discussion of "logical correlated characters" in numerical taxonomy, Sokal and Sneath (1963) stated "we must exclude as redundant any property (character) which is a logical consequence of another". In a real sense, this applies to structurally correlated characters as well. From this point of view, perhaps the most important aspect of multiple regression and factor analyses as statistical tools is their ability to reduce redundance in character-taking. Stepwise regression analysis identified three characters in the F_1 and four characters in the F_2 as the most potent predictors for grain yield. Then the remaining variates included in the present investigation could be thought of as redundant information. Similarly, from the analysis of the F_1 population, there should be no reason to take all four characters (grain yield per plot and per plant, number of spikelets and kernels per plant) when measurement of any one will suffice. Thus, selecting one character from each factor with high loadings on more than one will in all likelihood preserve most of the information if the primary objective were to isolate common factors which could better explain the inter-relationship among a set of correlated characters. On the other hand, if the primary interest were to predict the performance of the response character (i.e. grain yield), stepwise multiple regression analysis would be a more appropriate statistical method. From the practical point of view a plant breeder initiating a breeding program with a new crop species about which a limited amount of information is available would be well advised to first conduct a multivariate statistical analysis of as many characters as he can handle. This accomplished and the important predictor variables and common causes (factors) identified he can then concentrate on identifying superior genotypes.

Acknowledgements

Financial assistance from the National Research Council of Canada and from the Canadian International Development Agency — International Development Research Centre — is gratefully acknowledged.

Literature

1. Cattell, R. B.: Factor analysis: An introduction to essentials. I. The purpose and underlying models. Biometrics **21**, 190—215 (1965). — 2. Draper, N. R., Smith, H.: Applied regression analysis. New York: Wiley 1967. — 3. Harman, H. H.: Modern factor analysis. Chicago: The University of Chicago Press 1967. — 4. Hsu, P., Walton, P. D.: Inheritance of morphological and agronomic characters in spring wheat. Euphytica **19**, 54—60 (1970). — 5. Johnson, V. A., Schmidt, J. W.: Hybrid wheat. Advan. Agron. **20**, 199—233 (1968). — 6. Kaltsikes, P. J., Larter, E. N.: The interaction of genotype and environment in durum wheat. Euphytica **19**, 236—242 (1970). — 7. Kaltsikes, P. J., Lee, J.: Quantitative inheritance in durum wheat. Can. J. Genet. Cytol. **13**, 210—218 (1971). — 8. Krishnamurthy, K.: Physiological sources of variation in wheat yield. Varietal analylysis of yield structures. Mysore J. Agric. Sci. **2**, 203—213 (1958). — 9. Lawley, D. N., Maxwell, A. E.: Factor analysis as a statistical method. London: Butterworths 1963. — 10. Sokal, R. R., Sneath, P. H. A.: Principles of numerical taxonomy. San Francisco: Freeman 1963. — 11. Lebsock, K. C., Amaya, A.: Variation and covariation of agronomic traits in durum wheat. Crop Sci. **9**, 312—315 (1969).

Received May 23, 1972

Communicated by W. Seyffert

J. Lee, Graduate Student
Dr. P. J. Kaltsikes, Assistant Professor
Plant Science Department
University of Manitoba
Winnipeg, Manitoba R3T 2N2
(Canada)