Egypt. J. Agric. Res., 76 (4) ,1998.

EVALUATION OF STATISTICAL METHODS FOR DETERMINING THE RELATIVE CONTRIBUTION OF YIELD FACTORS IN WHEAT

S.M. NASR

Central Laboratory for Design and Statistical Analysis, Agricultural Research Center Giza, Egypt .

(Manuscript received 21 October 1997)

Abstract

Three statistical procedures of relating yield components to yield namely; multiple linear regression (in the form of full model, stepwise and pathe analysis), factor and cluster analysis were applied to 10 yield factors to evaluate these statistical techniques. A set of six widely grown wheat varities, namely; (Giza 155, Giza 157, Giza 163, Sakha 8, Sakha 61 and Sakha 69) were used. They were planted at Sids during two successive seasons of 1995/96 and 1996/97 in a randomized complete block design with four replications.

The most important results obtained from this investigation can be summarized as follows:

1. Simple correlation coefficients revealed that No. of spikes/m2, spike weight, No. of plants/m2 and plant weight had the greatest influence on grain yield. Path analysis indicated that No. of spikes/m2, plant weight and No. of plants/m2 were the three factors that exerted the greatest influence both directly and indirectly upon grain yield. When studying correlations or path-analysis it is important to recognize the nature of population under consideration. The magnitude of correlation coefficient can often be influenced by the choice of individuals upon which the observations are made. For this reason path analysis may not be applicable to in some cases.

2. Multiple linear regression and stepwise regression had the same adjusted R^2 (93%) denoting similar accuracy. The test of multicollinearity and linearity in the two procedures between variables indicated that, these models showes highly multiple coefficient of determination ($R2 = 0.94$) for prediction only and were not valid to interpret the partial correlation coefficient.

3. Factor analysis and cluster analysis would seem to be more suitable and efficient than the other techniques. They provide more information about cluster of intercorrelated variables. The two procedures indicated greatest relationships between (No. of spikes/m2 & plant weight),(1000-kernels weight & grain yield/spike) and (spike length & No. of spikelets/spike). These relationships will enable wheat breeders to realize high grain yield.

INTRODUCTION

The yield of wheat is the integrated effect of many variables that affect plant growth throughout the season. Growth analysis and relative contribution studies may help in interpretatiing the results and may aid the breeder for getting better varieties and good evaluation of the agricultural practices. Walton (1972) criticized some statistical techniques (correlation, path coefficient, stepwise rgression analysis and multiple linear regression) and explained that false conclusion could be obtained. He postulated that biologists must seek right assistance from statistical methodology and suggested factor analysis as a new technique to identify growth and plant characters related to yield in spring wheat.

Correlation coefficient is not only an important statistical procedure to facilitate breeding programs for high yield, but it is also important to examine the direct and indirect contribution of the yield component variables. Path coefficient analysis could be used in this respect. It divides correlation coefficients into direct and indirect effects through alternate path ways (Dewey and Lu, 1959). Bhowmik, et al. (1989) reported that, number of spikes/plant and spike length showed highly significant positive correlation with grain yield both at genotypic and phenotypic levels. Path coefficient analysis revealed that number of spikes/plant, spike length and kernel weight were related with grain yield through direct effects. Amin, et al. (1990) reported that, yield was positively and significantly correlated with plant height and number of spikes/m2. Plant height, number of spikes/m2 and 1000 grain weight exhibited maximum positive and direct effect on yield. Petrovic, et al (1994) pointed out that, there was significant correlation between grain weight per spike and the number of grains per spikelet, number of spikelets per spike and number of grains per spike, and number of grains per spike and the number of grains per spikelet.

Shafshak et al. (1989) in comparison between the full model regression and the stepwise regression procedure concluded that the coefficient of determination for full model regression and partial correlation were higher than stepwise regression. The latter had higher errors of estimates than the former. Nasr and Geweifel (1991) reported that stepwise coefficient of determination was tended to equal full model, and error of estimates was lower for stepwise analysis indicating that stepwise analysis was more efficient that full model regression.

Nasr and leilah (1993) used four statistical techniques to determine the relative contribution for sugar beet. The results indicated that, factor analysis approach was more efficient than other procedures. It can help plant breeders to determine the nature and sequence of characters to be selected in breeding programs. Hammed (1993) using factor analysis based on grain yield and other traits showed three groups to be the most importanty yield components.

Yau and Ferrara. (1994) reported that, cluster analysis is a useful technique for describing differential yield responses of best breeding lines grown in diverse environments.

The present investigation was meant to evaluate three statistical methods for determining the relative contribution of yield factors in wheat grain yield.

MATERIALS AND METHODS

Six varieties of wheat namely; (Giza 155, Giza 157, Giza 163, Sakha 8, Sakha 61 and Sakha 69) were planted at Sids during two successive seasons (1995/96 and 1996/97) in a randomized complete block design with four replications. The plot area was 12 m2. Observations were taken from 10 random plants from the center of plots. Observations were taken from 10 random plants from the center of plots. Observations on 10 yield components were recorded namely; number of spikes/m2, spike length (cm), spike weight (gm.), number of grains/spike, number of spikelets/ spike, 1000-kernels weight (gm.), grain yield/spike (gm), number of plants/m2 and plant weight (gm.). In addition, four square meters were harvested from the middle of the experimental plots to estimate grain yield (ton / faddan) and straw yield (ton/faddan.).

The normal agricultural practices were done as usually performed by wheat growers in the district.

Statistical techniques :

The combined data for two seasons of yield and yield components were subjected to the following statistical procedures.:

1. Basic statistics and correlation matrix: Arithmatic mean, standard deviation, standard error, minimum, mixmum and simple correlation cofficient were computed for various characters according to Snedecor and Cochran (1980).

2. Path coefficient analysis, as applied by Dewey and Lu (1959) and Douglas, and Lu (1985) were used for partitioning the total correlation coefficient between yield (y) and its primary components into direct and indirect effects.

3. Multiple linear regression: between grain yield and the studied characters were done according to Draper and Smith (1966) was used Partial coefficient of determination (R2) was estimated for each component to evaluate the relative contribution and to construct a prediction model for grain yield (y) of wheat according to the following formula:

$$Y = a + b1\ X\ 1 + b2X2 + \dots\dots\dots\dots\dots\dots\ bn\ X\ n.$$

4. Stepwise multiple linear regression : was carried-out to determine the variables accounting for the majority of the total yield variability (Draper and Smith, 1966). It was used to compute a sequence of multiple regression equations in stepwise manner. At each step, one variable was added to the regression equations. It was the one that reduces the error sum of squares. Equivalently, It was the variable having the highest partial correlation with the dependent variable and which if added had the highest value in the regression analysis of variance. Moreover, variables were forced into the regression equation and automatically removed.

5. Factor analysis: method according to Cattel (1965). This method basically reduces a large numbers of correlated variables to a small number of uncorrelated factors. When the contribution of a factor to the total percentage of the trace was less than 10%, the process stopped. After extraction, the matrix of factor was submitted to a varimax orthogonal rotation, as applied by Harman (1976). The effect of rotation is to accentuate the larger loading in each factor and to suppress the minor loading coefficient and in this way to improve the opportunity of achieving a meaningful biological interpretation of each factor. A communality (h2) is the amouunt of the variance of a variable accounted for the common factors together. Since the purpose was to determine the way in which yield components are related to each other, yield (y) was not included in this structure.

6. Cluster analysis: Based on the coefficient of correlation of yield components was performed by the unpaired group method. Output of this analysis was used to derive a dendrogram. Clustering begins with the fusion of two most similar characters and proceeds until all characters are fused into clusters and/or all cluster are fused. The clustering process can be represented in the form of a dendrogram (or tree) in which the top branch indicates the highest fusion level, the next lower branch, the next highest fusion level, and so on. For reference purposes, the fusion levels will be designated 1, 2, ... from top to bottom, respectively (Joseph, et al., 1992).

RESULTS AND DISCUSSION

Data of simple correlation coefficients matrix, means, standard error, minimum and maximum for 11 characters of wheat presented in Table (1). The obtained data that, the relationship between all possible pairs of the 11 traits were highly significant (0.01) in all cases. In addition a number of interesting relationships can be observed from Table (1). The most important relationships to the wheat breeder are that between grain yield and No. of spikes/m2 (0.880), spike weight (0.701), No. of plants/m2 (0.762), plant weight (0.892) and straw yield (0.844) which showed highly significant positive correlation. This indicated that these characters had the greatest influence on grain yield. These results are in agreement with those reported by Bhowmik, et al. (1989) Another correlation worthy of some attention is that between grain yield / spike and spike length (0.864), No. of grains/spike (0.899), No. of spikelets/spike (0.804) and 1000-kernels weight (0.923). These results are in harmony with those reported by Petrovic, et al. (1994). Close associations were found also between number of spikes/m2 and plant weight.

It is apparent that many of the characters are correlated because of a mutual association with other characters. At this point, the Path-coefficient analysis provides an effective means of separating direct and indirect causes of association and permits a critical examination of the specific forces acting to produce a given correlation and measures the relative importance of each causal factor.

An examination of the correlation components indicated that No. of spikes/m2, plant weight and No. of plants/m2 were the three factors that exerted the greatest influence both directly and indirectly upon grain yield (Direct effect % = 39.548, 23.783 and 8.918 respectively), (Table 2). These three traits were important components in every correlation that involved grain yield. The second important components were No. of grains/spike, grain yield/spike and No. of spikelets/spike (Direct effect % = 6.396, 5.585 and 4.775 respectively). Results of path analysis indicated that, the major components of spike (spike length, spike weight and 1000-kernels weight) had small indirect effects and unimportant components effect upon grain yield/faddan which is not consitent with previous studies of bhowmik, et al., 1989 and Amin,et al. 1990. However, some of the assumed pathways may not be realistic. Therefore, when studying correlations or path-analysis is of prime importance to recognize both the nature of population under consideration and the magnitude of the correlation coefficient that is often be influenced by the choice of individuals upon which the observations are made. The effect of selection of parent clones upon the variation in the population under study is not known (Douglas and Lu 1985).

Table 1. A matrix of simple correlation coefficients, means standard deviation, standard error, minimum and maximum values for 11 characters of wheat. (combiend data of two seasons).

Characters	1	2	3	4	5	6	7	8	9	10	11
1- No. of spikes/m2	1.000										
2- Spike length cm.	.331	1.000									
3- Spike weight gm.	.437	.790	1.00								
4- No. of grains/spike	.308	.825	.619	1.000							
5- No. of spikelets/spike	.245	.843	.795	.749	1.000						
6- 1000-Kernels weight gm.	.349	.817	.660	.849	.744	1.000					
7- Grain yield/spike gm.	.298	.864	.720	.899	.804	.923	1.000				
8- No. of plants/m2	.495	.835	.741	.802	.771	.812	.864	1.000			
9- Plant weight gm.	.929	.390	.490	.261	.326	.406	.339	.558	1.000		
10- Straw yield ton/fadd.	.851	.459	.457	.419	.350	.489	.427	.582	.880	1.000	
11- Grain yield ton/fadd.	.880	.632	.701	.509	.555	.604	.592	.762	.892	.844	1.000
Mean	464.30	8.37	1.90	37.66	14.98	36.88	1.40	273.68	4.83	3.32	1.69
Standard error	7.19	0.05	0.01	0.29	0.09	0.13	0.01	1.63	0.06	0.05	0.03
Minimum	265.00	6.40	1.60	27.90	11.10	31.80	0.89	205.00	2.87	1.63	0.66
Max	777.00	10.90	2.41	46.20	18.70	41.30	1.86	328.00	7.11	5.35	2.80

Table 2. Direct and indirect effects for yield factors of grain yield of wheat according to path analysis and percentage of direct effect.

Characters	Direct effect PYi	Indirect effect	Total correlation	Direct effect %
1- No. of spikes/m2	0.439	0.441	0.880	39.549**
2- Spike length cm.	0.040	0.592	0.632	3.604
3- Spike weight gm.	0.043	0.655	0.698	3.873
4- No. of grains/spike	0.071	0.437	0.508	6.396*
5- No. of spikelets/spike	0.053	0.502	0.555	4.775*
6- 1000-Kernels weight gm.	0.027	0.631	0.604	2.432
7- Grain yield/spike gm.	0.062	0.530	0.592	5.585*
8- No. of plants/m2	0.099	0.663	0.762	8.918**
9- Plant weight gm.	0.264	0.628	0.892	23.783**
10- Straw yield ton/fadd.	0.022	0.866	0.844	1.982

Data in Table (3) shows the prediction model by using multiple linear regression for grain yield ton/faddan of wheat and its attributes. The prediction equation was formulated as follows:

$$Y = -2.o81 + 0.002 X1 + 0.039 X2 + 0.403 X3 + 0.005 X4 + 0.005 X5 + 0.009 X6 + 0.026 X7 + 0.004 X8 + 0.128 X 9 + 0.026 X 10$$

The relative contribution for all yield factor explained 94% of the total variation in grain yield, and 6% could be due to residual. No. of spikes/m2, spike weight gm, No. of plant/m2 and plant weight gm. had the highest partial coefficient of determination ($R2$ = 44.513%, 15.482%, 19.858% and 22.550% respectively). The other characters had small contribution of the total yield variance.

To test the validity of this procedure, we can calculate variance inflation factor (VIF). It is a potent aid for detecting the levels of multicollinearity. VIF's are the diagonal elements of VIF tell us the degree to which each independent variable is explained by the other independent variables. Thus large VIF values denote high collinearity. Values exceeding 5.00 are considered likely to cause difficulty in coefficient estimation due to multicollinearity. According to results in Table (3), it is clear that regression coefficients were not stable. The instabilities were likely due to linear relationships among yield components (correlations, Table 1.), a condition referred to as multicollinearity. However, in data sets containing measurement of uncontrolled variables, multicollinearity is often strong enough to make unclear results (Hoerl and Kennard, 1970). In our data (Table 3.) the VIF's for No.of spikes/m2, spike length cm., No. of grains/spike, 1000-kernels weight gm., grain yield/spike gm., No. of plants/m2, Plant weight gm. and straw yield ton/fad. exceeded the value (5.0). Thus, evidence from the VIF calculation confirmed a part of the instability of the regression coefficients due to interrelations among the explantory variables.

These results indicated that, using this model with the highly multiple R squared (R2 = 0.941) for prediction only (i.e., make no try to interpret the partial regression coefficients) and using a more sophisticated method of analysis such as ridge regression, principal components or factor analysis to obtain a model that more clearly reflects the simple effects of the predictors (Joseph, et al. 1992).

Table 3. The relative contribution of 11 characters in predicting grain yield for wheat by using multiple regression analysis.

Characters	Direct effect PYi	Indirect effect	Total correlation	Direct effect %
1- No. of spikes/m2	0.002	0.001	44.513**	8.435#
2- Spike length cm.	0.039	0.026	6.056	6.471#
3- Spike weight gm.	0.403	0.082	15.482**	3.892
4- No. of grains/spike	0.005	0.005	4.182	7.117#
5- No. of spikelets/spike	0.005	0.016	1.478	4.734
6- 1000-Kernels weight gm.	0.009	0.011	3.389	7.708 #
7- Grain yield/spike gm.	0.026	0.0133	1.199	14.500 #
8- No. of plants/m2	0.004	0.001	19.858**	6.680 #
9- Plant weight gm.	0.128	0.031	22.550**	11.392 #
10- Straw yield ton/fadd.	0.026	0.025	4.143	6.347 #
Y - intercept	=-2.081	0.319		
Multiple R	= .970			
Adjusted R squared	= .939			
R squared	= .941			
Standard error of est.	= .224			

** Significant at 1% probability level
# High collinearity

Stepwise multiple **regression analysis** was used to determine the best variables that mostly reduced the variance of yield. This is done by introducing the variables in order of importance. Table (4) shows the accepted and removed variables and reduction in yield variance caused by each variables. The accepted variables had the highest coefficient of multiple determination with the yield adjusted for variables already added. As it can be seen from Table (4), the accepted variables were No. of spikes/m2, spike length cm., spike weight gm., No. of plants/m2 and Plant weight gm. Those variables were responsible for 45.259%, 8.633%, 15.091%, 21.928% and 24.142% respectively of yield variance. Thhe prediction equation was formulated as follows:

$$Y = -2.315 + 0.002X_1 + +.056X2 + 0.393\ X_3 + 0.004X_4 + 0.137\ X_5$$

The other four characters are removed variables because their contribution was insignificant. According to these results 93.9% of the total variation in yield could be attributed to the five accepted variables (Table 4).

Therefore, these five traits have to be ranked the first in any breeding program for improving yield.

Table 4. Accepted and removed variables according to stepwise analysis and their relative contribution in grain yield variance in wheat.

Characters	Regression coefficient	Standard error	Partial R2%	Variance of inflation factor VIF
Accepted variables :				
1- No. of spikes/m2	0.002	0.001	45.259**	4.414
2- Spike length cm.	0.056	0.021	8.633*	4.533
3- Spike weight gm.	0.393	0.071	15.091*	3.006
4- No. of grains/spike	0.004	0.001	21.928**	4.237
5- Plant weight gm.	0.137	0.026	24.142**	3.165
Removed variables:				
1- No. of grains/spike	0.046			5.155#
2- No. of spikelets/spike	0.015			5.575#
3- 1000-kernels weight gm.	0.013			4.627
4- Grain yield/spike gm.	0.017			6.984#
5- Straw yield ton/fadd.	0.054			6.303#
Y - intercept	=-2.315	0.092		
Multiple R	=0.969			
R squared	=0.939			
Adjusted R squared	=0.928			
Standard error of est.	=0.123			

*,** Significant at 5% and 1% probability levels, respectively.

High collinearity

To evaluate the results of stepwise-analysis, assumptions must be tested. Meeting the assumptions of regression analysis is essential to ensure both that the results obtained were truly representative of the sample and that we have obtained the best possible results. The principal measure used to test the assumption is the residual - the difference between the actual dependent variable values and its predicted values (Joseph, et al. 1992). For comparison purposes, we use the **standardized partial regression plots.** When using more than one predictor variables, that each predictor variable's relationship is linear to ensure its best representation in the equation. To do this, we use the partial regression plot for each predictor variable in the equation. In Figure (1-a,b,c,d,e,f), we can see that the relationships for No. of spikes/m2 is quite well defined; thus it has strong and significant effects in the regression equation. Spike length cm., spike weight gm., No.of plants/m2 and plant weight gm are less well defined, both in slope and scatter of the points, thus explaining there lesser effect in the equation. For all five components, nonlinear pattern is shown. Therefore, this equation is not valid to interpret the partial correlation coefficients (Joseph, et al. 1992).

The results of *Factor analysis* are presented in Table 5. Factors were constructed using the principal factor analysis technique to establish the dependent relationship between morphological characteristics and yield components. Factor analysis divided the 10 variables into two main factors. For purposes of interpretation, only those factor loadings greater than 0.5 were considered important. A summary of the compoistion of variables of the two factors with loadings are given in Table (5). The two factors explained 86.60% of the total variability in the dependence structure.

Factor 1 included seven variables which accounted for 66.50% of the total variance. The seven variables were spike length cm., spike weight gm., No.of grains/spike, No.of spikelets/spike, 1000-kernels weight gm, grain yield/spike gm and No.of plants/m2. The seven variables were positively correlated with the factor. These variables were of almost equal importance and highly communality with factor 1. Thus, this factor may be regarded as spike factor.

Factor 2 was called a growth factor because it consisted of No.of spikes/m2, plant weight gm, and straw yield ton/fadd. It accounted only 20.10% of the total variance in the dependence structure (Table 5). All factor loading of factor 2 were positive and had a significant positive correlation between the three variables (R = 0.582). The sign of the loading indicates the direction of the relattionship between the factor and variable.

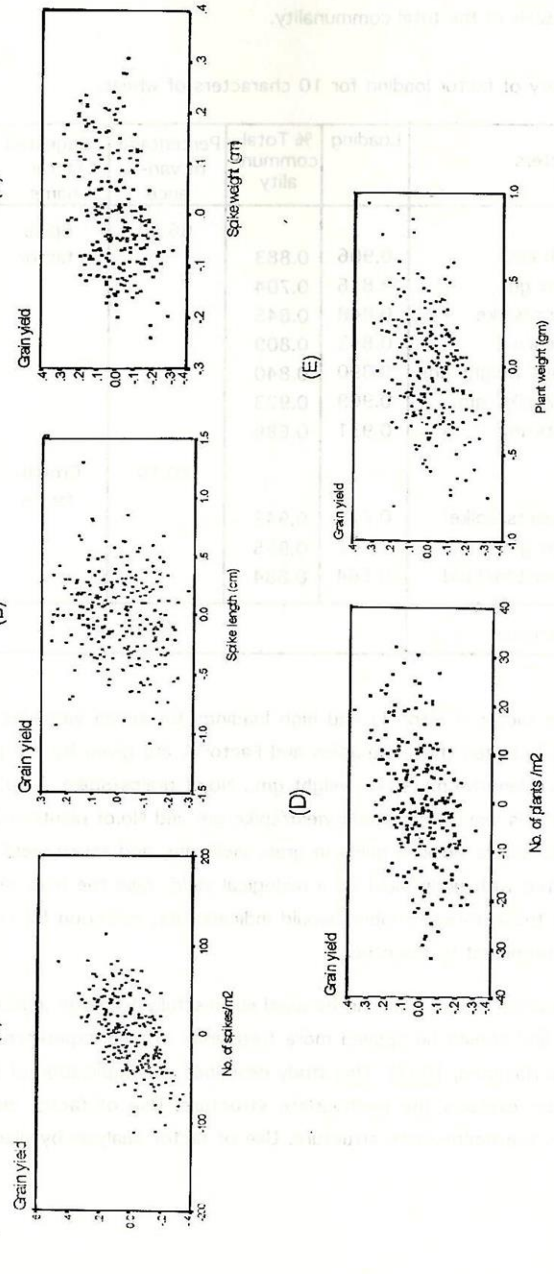

Fig. 1. Standaridized partial regression plot for No. of spikes/m2, spike length, spike weight, and plant weight (gm) with grain yield

The results of the present investigation indicates that the estimated commonalties (Table 5) were adequate for conclusion where the two factors together accounted for 86.60% of the total communality.

Table 5. Summary of factor loading for 10 characters of wheat.

Characters	Loading	% Total commun-ality	Percentage of vari-ance	Suggested factor name	Latent root
Factor 1:			66.50	Spike factor	6.646
1- Spike length cm.	0.906	0.883			
2- Spike weight gm.	0.836	0.704			
3- No. of grains/spike	0.848	0.845			
4- No. of spikes/m2	0.843	0.809			
5- 1000-kernels weight gm.	0.890	0.840			
6- Grain yield/spike gm.	0.909	0.923			
7- No. of plants/m2	0.931	0.686			
Factor 2:			20.10	Growth factor	2.071
8- No. of spikelets/spike	0.779	0.942			
9- Plant weight gm.	0.742	0.955			
10- Straw yield ton/fadd.	0.664	0.884			
Cumulative variance	86.60				

The spike factor (Factor 1) had high loadings for seven variables (Table 5). The correlation between these variables and Factor 1 are given by the plant types with more (spike length cm., spike weight gm., No.of grains/spike, No.of spikelets/ spike, 1000-kernels weight gm, grain yield/spike gm and No.of plants/m2) will enable breeders to realize desired gains in grain yield gm, and straw yield ton/fadd.) usually associated with grain yield on a biological yield. Also the high simple correlation between these three variables would indicate that selection for rather variable will be a determinent to the other.

Factor analysis is one that can be used successfully for large amounts of multivariate data, and should be applied more frequently in field experiments (Joseph, et al. 1992 and Hammed, 1993). This study describes one application of factor analysis that further explains the multivariate structure. Use of factor analysis that further explains the multivariate structure. Use of factor analysis by plant breeders has the potential of increasing the comprehension of the causal relationships of variables, and can help to determine the nature and sequence of traits to be selected in a breeding program.

Table (6) contains the results of cluster analysis, including the cases being combined at each stage of the process and the clustering coefficient. The coefficient (fourth column) is the correlation similarity coefficient between the two caefficient between the two cases or clusters being combined. Small value of cluster appear indicat that fairly homogeneous cluster are being merged. The seventh column shows average distance between clusters.

Table 6. Agglomeration Schedule of hierarchical cluster analysis using average Linkage (between groups).

Stage	Clusters Combined		Coefficient	Cluster appears		Distance
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	No. of spikes/m2	Plant weight	0.929	0	0	4
2	1000-kernels weight	Grain yield/spike	0.922	0	0	3
3	No.of grains/spike	1000-kernels weight	0.874	0	2	6
4	No. of spikes/m2	Straw yield	0.865	1	0	9
5	Spike length	No. of spikelets/spike	0.842	0	0	7
6	No. of grains/spike	No. of plants/m2	0.825	3	0	7
7	Spike length	No. of grains/spike	0.801	5	6	8
8	Spike length	Spike weight	0.721	7	0	9
9	No. of spikes/m2	Spike length	0.395	4	8	0

Figure (2) presents the tree of the clustered entries derived from the cluster analysis. There are two main clusters at average distance 25. Cluster 1 was comprised of variables (No. of spikes/m2, plant weight gm. and straw yield ton/fadd) developed by biological yield. Cluster 2 contained variables (spike length cm., No. of spikelets/spike, 1000-kernels weight gm., grain yield/spike gm., No. of grains/spike, No.of plants/m2 and spike weight gm) developed by spike factors. At the average distance 10, cluster 2 can be divided into two sublcluster (A & B). Subcluster

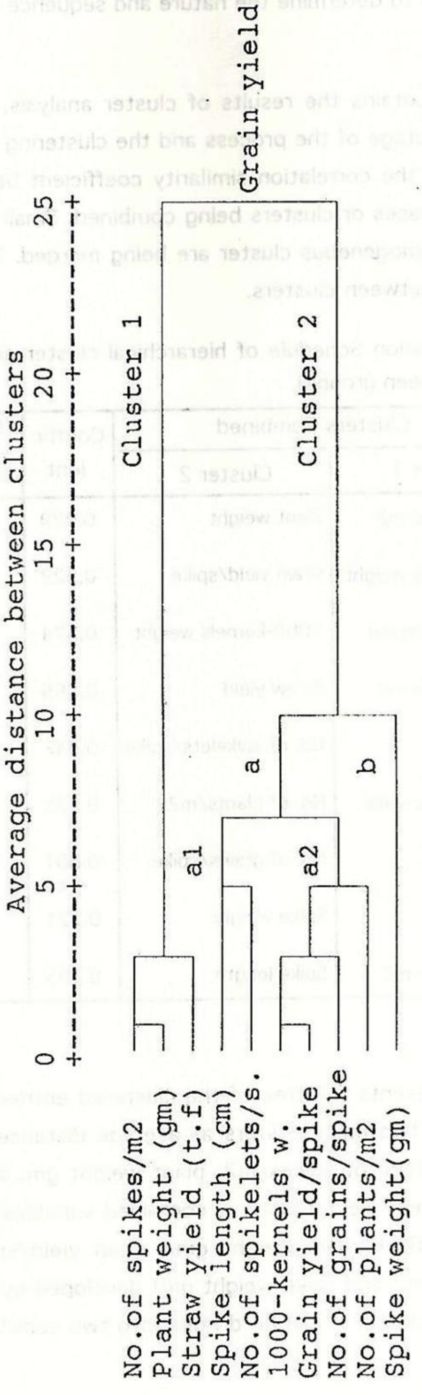

Fig. 2. Dendrogram based on cluster analysis of 10 characters of wheat.

(B) which contained spike weight gm. Subculture (A) at averae distance 7.5 presented two subsubclusters, (A1) and (A2). From these results it can be noted that, the cluster dendrogram (tree) contained the fourth fusion level. The first fusion including variables No.of spikes/m2, plant weight gm. and straw yield ton/fadd was one of the first clusters to be formed. All three variables in this cluster were closely related (average distance = 3). The second fusion consisted of spike length cm. and No. of spikelets/spike at average distance = 5.1000-kernels weight gm., grain yield/ spike gm., No.of grains/spike and No.of plants/m2 were found in the third fusion at average distance = 5. These three fusions had low average distance (less than 10), These rese led us to the importance of groups in wheat breeding programs for yield. Conversely, spike weight (gm) was not fused with other variables at the highest fusion level (average distance = 5, it had long average distance = 10, this means that, it was closely related to grain yield directly.

Cluster analysis had provided additional additional information, however, grouping the ten correlated variables into four clusters provides a quick visual overview of the clustering process, as well as showing which variable found in each cluster and the reltive importance of these variables for grain yield.

From the previous results, it can be concluded that, cluster analysis is the name of a group of multivariate techniques whose primary purpose is to identify the similarity from the characteristics they possess. They identify and classify variables so that same variables are very similar to other in the cluster. The resulting object is that clusters should then exhibit high internal (withincluster) homogeneity and high external (between-cluster) heterogeneity (Joseph et al. 1992).

So, It can be recommended from the previous results that:

1. It is essential to detect characters having the greatest influence on yield and their relative contribution to yield variation. This is useful in designing and evaluating breeding programs and agronomic systems.

2. The important characters over all statistical procedures used were No. of spikes/ m2, No. of plants/m2, spike weight and plant weight.
3. The results of statistical procedures indicated that, factor analysis and cluster analysis would seem to be more suitable and efficient than other procedures. They provide more information about cluster of intercorrelated variables. They can help plant breeders to determine the nature and sequence of characters to be selected in breeding programs.

4. No difference were detected between the accuracy and precision of the full model regression and stepwise procedures, where R squared and standard error of estimate values are equal.

REFERENCES

1. Amin, M.R.; Hoque, M.M.; Shaheed, M.A.; Sarker, A.K.D.; Kabir, Z. 1990. Genetic variability, character association and path analysis in wheat Bangladesh-Journal of Agricultural Research (Bangladesh). (1990). v. 15 (2) p. 1-5.

2. Bhowmik, A., Ali, M.S.; Sadeque, Z., Saifuddin, K. 1989, Correlation and path analysis in wheat (Triticum aestivum L.) Bangladesh - Journal of plant Breeding and Genetics. v. 2 (1,2) P. 23-26.

3. Cattel, R.B. 1965. Factor analysis on introduction to essentials. Purpose and underlying models. Biometrics, 21: 190-215.

4. Dewey, D.R. and K.H. Lu (1959). A correlation and path coefficient analysis of components of crested wheat grass seed production. Agron. J. 51: 515-518.

5. Douglas, R.D. and K.H. Lu. 1985. A correlation and path-coefficient analysis of components of crested wheat grass seed production. Agron. J. p. 515-518.

6. Draper, N.R. and H. Smith. 1966. Applied Regression Analysis. John Wilay & Sons Jny. Ny.

7. Hammed, M.Y. 1993. Factor Analysis in bread wheat. Mesopotamia Journal of Agriculture. (Iraq). v. 25 (no. 3) p. 53-57.

8. Harman, H.H. 1976. Modern factor analysis. 3rd Ed. Univ. of Chicago Press, Chicago.

9. Hoerl, A.E., and R.W. Kennard. 1970. Ridge regression: Applications to non-orthogonal problems. Technometrics 12: 69-82.

10. Joseph F.H.J., R.E. Anderson, R.L. Tatham. 1992. Multivariate data analysis. Macmillan Publishing Company. New York.

11. Nasr, S.M. and H.G.M. Geweifel. 1991. Evaluation of the relative contribution of peanut yield factors by using four statistical procedures. J. Agric. Scl. Mansoura Univ, 16 (7): 1483-1490.

12. Nasr. S.M. and A.A. Leilah. 1993. Integrated analysis of the relative contribution of some variables in sugar beet using some statistical techniques. Bull.

Fac. of Agric., Univ. of Cairo, vol (44): 253-266.

13. Petrovic, J., Cermin, L., Brezinova-Polakova, B. 1994. The heterosis effect and phenotype correlations of the quantitative characters of wheat. Polnohospodarstvo (Slovak Republic). v. 40 (9) p. 660-671.

14. Shafshak. S.E. Abd El-Halim. A.A., Saad. A.M.M. and Ahmed. F.A. 1989. Integrated regression analysis of maize yield factors using minimum sample size. Egypt. J. Apple. sci. 4 (1855-1866).

15. Snedecor, G.W. and G.W. Cochran. 1980. Statistical method. 7th Ed. The lowaState Press Amess, Lowa, U.S.A.

16. Walton, P.D. 1972. Factor analysis of yield in spring wheat (Triticum aestivum L.).

17. Yau, S.K.; Ortiz-Ferrara, G. 1994. Regression and cluster analysis of elite wheat lines grown in rain-fed environments. Journal-of-Genetics-and Breeding. v. 48 (2) p. 183-189.

تقييم طرق إحصائيه لتقدير المساهمة النسبية لمكونات محصول القمح

سعيد محمد نصر

المعمل المركزى لبحوث التصميم والتحليل الإحصائى - مركز البحوث الزراعية - الجيزة.

أجرى هذا البحث بهدف تقييم ثلاث طرق إحصائية تستخدم فى تقدير المساهمة النسبية لعوامل المحصول وهى تحليل الأنحدار الخطى المتعدد (فى صوره المختلفة وهى الإنحدار المتعدد الكامل والإنحدار المتعدد المحلى والأنحدار القياسى الجزئى) وتحليل العامل بالإضافة إلى التحليل العنقودى وذلك على عشرة صفات من مكونات محصول القمح وأستخدم لذلك سته أصناف من القمح هى جيزة ١٥٥ ، جيزة ١٥٧ ، جيزة ١٦٣ ، سخا ٦١ ، سخا ٧٩ . وقد إقيمت التجربة بمحطة بحوث سدس خلال موسمى ٩٥ / ١٩٩٦ ، ٩٦ / ١٩٩٧ فى تصميم قطاعات كاملة العشوائية فى أربع مكررات. وتتلخص أهم النتائج المتحصل عليها فيما يلى :

١ - أظهرت مصفوفة الإرتباط أن عدد السنابل فى المتر المربع ووزن السنبلة وعدد النباتات فى المتر المربع بالإضافة إلى وزن النبات هى أهم العوامل المؤثرة على محصول الحبوب. أما تحليل معامل المرور فقد حدد ثلاثه عوامل هامه ذات تأثير مباشر على محصول الحبوب وهى عدد السنابل فى المتر المربع ووزن النبات وعدد النباتات فى المتر المربع. إلا أنه يؤخذ على طريقتى الإرتباط ومعامل المرور أن نتائجهما تتأثر بطريقة إختيار النباتات الفردية وبالتالى فالتقديرات غالبا ما تكون متحيزه حيث تختلف النتائج من عشيرة إلى آخرى.

٢ - أشارت النتائج أيضا بعدم وجود فروق جوهرية بين طريقتى تحليل الإنحدار المتعدد والإنحدار المرحلى حيث كانت قيمه معامل التحديد المعدل لكل منهما متساوى تقريبا (٩٣٪) مما يشير إلى تساوى كفاءة الأسلوبين إلا أنه لوحظ عند إجراء إختبار التأثير الخطى المشترك البسيط والمتعدد بين مكونات المحصول لقياس كفاءة النماذج الناتجة من هذين الأسلوبين لوحظ وجود إرتباط ذاتى عالى المعنوية بين عوامل المحصول بالرغم من إرتفاع قيمه معامل التحديد التى وصلت إلى ٩٤٪ مما يؤكد عدم الإعتماد على هذين الأسلوبين فى تقدير المساهمة النسبية لعوامل المحصول الناتجة منهما ويكتفى الأستفادة من معادلات الإنحدار الناتجة منهما فى عمليات التنبؤ بالمحصول فقط.

٣ - أظهرت النتائج المتحصل عليها من طريقتى تحليل العامل والتحليل العنقودى إنهما أنسب الطرق حيث أعطت معلومات أكثر وضوحا عن طبيعة العلاقات المتداخلة بين عوامل المحصول مما يفيد مربى القمح فى برامج التربية والإنتخاب لتحسين المحصول مثل العلاقة بين (عدد السنابل فى المتر المربع، وطول النبات) و (وزن ١٠٠٠ حبة، محصول السنبلة)، و (طول السنبلة وعدد السنيبلات فى السنبلة) حيث لوحظ وجود إرتباط ذاتى عالى المعنوية بين هذه الصفات.