MULTIVARIATE ANALYSIS FOR EVALUATING SESAME YIELD AND ITS CONTRIBUTING FACTORS

Abd El-Mohsen , A. A.
Agronomy Dept., Faculty of Agriculture, Cairo University, Giza, Egypt

ABESTRACT

Two field experiments were carried out in a commercial field at Abo Rawash village, Giza governorate, Egypt during 2004 and 2005 seasons to compare five statistical procedures including: simple correlation, path analysis, multiple linear regression, stepwise regression and factor analysis in determining the relationship between sesame seed yield and its contributing traits. Thirty sesame genotypes were used for this purpose. The studied characters were: flowering date, plant height, number of fruiting branches, stem height to the first capsule, fruiting zone length, number of capsules on main stem, number of capsules per plant, capsule density on main stem, 1000-seed weight and seed yield per plant. The simple correlation coefficients and path analysis of yield components revealed that components with the highest positive correlation to yield also had the highest positive direct effect to yield i.e., number of capsules on main stem and number of capsules per plant. Path analysis showed that, the residual effect (0.433) was high in magnitude which shows that some other important yield contributing characters which contribute to yield have to be included. Stepwise multiple regression analysis showed that 77.25% of the total variation in seed yield could be explained by the variation in number of capsules per plant and flowering date in sesame. The linear regression equation was (Y) = 10.951 - 0.110 X_1 + 0.114 X_7, where Y, X_1 and X_7 represent seed yield per plant, flowering date and number of capsules per plant, respectively. Besides, coefficient of determination (R^2), adjusted R-squared statistic and standard error of estimate values, mean absolute error (MAE) and Durbin-Watson (DW) statistic test showed no significant differences between the full model regression and stepwise multiple regression analysis technique. However, the efficiency expressed is due to the reduction in number of variables in the fitted model from all variables (full model regression) to two variables only (stepwise multiple regression). Factor analysis indicated that three factors could explain approximately 81.9% of the total variation. The first factor which accounted for about 41% of the variation was strongly associated with fruiting zone length, number of capsules on main stem, number of capsules per plant, and capsule density. The second factor which accounts for about 25% of the variation, was strongly associated and positive effects on days to flowering, 1000-seed weight, plant height and stem height to the first capsule, whereas the third factor had positive effects on number of fruiting branches only, which accounts for about 16% of the variation. Factor analysis technique was more efficient than other used statistical techniques. It provides more information about cluster of inter-correlated variables. It could be concluded that the five of statistical analysis techniques, agreed upon that high yield of sesame plants could be obtained by selecting breeding materials with high number of capsules on main stem, number of capsules per plant, plant height and increasing capsule density on the main stem.

Keywords: Sesame trait interrelationships, multivariate analysis, simple correlation, path analysis, multiple linear regression, stepwise regression, factor analysis.

INTRODUCTION

Sesame (*Sesamum indicum* L.) is one of the major summer oil crops particularly grown in light soils of newly reclaimed land. It is an important annual oilseed crop grown especially in developing countries as a rich source of oil, protein, calcium and phosphorus. Sesame oil is characterized for its stability and quality so the production of this important oil seed crop has to be stepped up by evolving a high yielding variety. Therefore the plant breeder must know the relationship between yield-contributing characters and their association with yield.

In most of the crop improvement programs, augmentation in seed yield is one of the major objectives. The information on association between seed yield with its components is a prerequisite for breeding program aimed at yield improvement. Yield is a dependable complex inherited character as a result of interaction of several contributing factors that may be related or unrelated (Subramanian and Subramanian, 1994).

Correlation coefficient which measures the simple linear relationship between two traits does not predict the success of selection. However, path coefficient analysis, a standardized partial regression, determines the relative importance of direct and indirect effects on seed yield (Bhatt, 1973). Path coefficient analysis are more informative and useful than simple correlation coefficients and widely used in crop breeding to determine the nature of relationships between yield and some yield components (Dewey and Lu, 1959; Kang *et al.*, 1983; Shrief, 1983; Gupta and Chopra, 1984; Shabana and Abu-Hagaza, 1984; Board *et al.*, 1997; Samonte *et al.*, 1998 and El-Deeb and Mohamed, 1999).

Nasr and Geweifel (1991), El-Sergany Dawlat (1992) and Leilah and Al-Khateeb (2005) reported that stepwise multiple regression was more efficient than the full model regression. It is used to determine the best predictive equation for yield.

Factor analysis is a multivariate analysis method which aims to explain the correlation between a large set of variables in terms of a small number of underlying independent factors. It is assumed that each of the variables measured depends upon the underlying factors but is also subject to random errors. Walton (1972) suggested factor analysis as a new technique to identify growth and plant characters related to yield in spring wheat. Ibrahim *et al.* (1984), El-Deeb and Mohamed (1999) and Mansouri and Najafabadi (2004) used factor analysis in sesame.

This study was undertaken in order to determine the dependence relationship between seed yield and yield component characters of several genotypes of sesame by using five statistical procedures including; simple correlation, path analysis, multiple linear regression, stepwise multiple linear regression and factor analysis.

MATERIALS AND METHODS

Two field experiments were carried out at Abo Rawash village, Giza governorate, Egypt during 2004 and 2005 seasons to compare among five statistical procedures including; simple correlation, path analysis, multiple

linear regression, stepwise regression, factor analysis were used to study the relationship between sesame seed yield and its components. Thirty sesame genotypes were used in this study. These were: twenty seven sesame promising mutant lines and three local Egyptian sesame commercial cultivars namely Giza 25, Giza 32 and Shandaweel 3.

Seeds of commercial cultivars (G. 25, G. 32 and Shandaweel 3) were obtained from Oil Crop Res. Institute of Field Crops, Agric. Res. Cent., Giza, Egypt. While seeds of the twenty-seven mutant promising lines originated by selection through a previous breeding program of Agronomy Dept., Fac. of Agric., Cairo University.

The trial was arranged in a randomized complete blocks design with three replications. Each replication includes 30 experimental plots. Plot size was four rows each 5 m long and 0.5 m apart with total area of 10 m^2. Hill spacing was 10 cm within the row. The thirty sesame genotypes were planted at the beginning of May for two seasons. Cultural practices were carried out as usually schemed.

Data on yield per plant and yield components traits were obtained by calculating the mean of ten representative plants. Observations on 10 yield related components: flowering date (X_1), plant height (X_2), number of fruiting branches (X_3), stem height to the first capsule (X_4), fruiting zone length (X_5), number of capsules on main stem (X_6), Number of capsules per plant (X_7), capsule density (X_8), 1000-seed weight (X_9) and seed yield per plant (Y) were recorded.

Statistical technique:

Normality distributions in each trait were checked out by the Wilk Shapiro test (Neter *et al.,* 1996). The data were analysed according to the randomized complete blocks design over years. A combined analysis of variance was conducted for the two seasons according to Gomez and Gomez (1984). Homogeneity test of variances were performed according to procedures reported by Gomez and Gomez (1984). Thus, if hypothesis that the two error variances are homogeneous cannot be rejected, the combined analysis of variance was computed. The combined data of yield and yield components over both seasons were used the following statistical procedures. In order to determine the relationships between examined traits and seed yield per plant, correlation coefficients were calculated with the MSTAT-C software package (Freed *et al.,* 1989). The path coefficient analysis was performed by examining seed yield per plant as a dependent variable for major contributor's traits to seed yield per plant via OpenStat version 1.9, a computer program, as suggested by William (2007). Modeling was performed according to the multiple linear (full model) and stepwise multiple linear regression method, backward variable selections were applied using Statgraphics Plus for windows (Manugistics, 1998) and SPSS computer software (1999).

The following analyses were performed:

1. Basic statistics: The raw data was compiled by taking the means of all the plants taken for each treatment and replication for different traits in both the experimental years. The pooled means of both seasons were subjected to

further statistical and biometrical analysis. Simple statistical parameters, viz. average, standard error, variance and coefficient of variation were analyzed according to Steel and Torrie (1980).

2. The estimates of simple correlation coefficients were calculated between seed yield and its components according to the method described by Steel and Torrie (1980). Correlations were partitioned into path coefficient using the technique outlined by Dewey and Lu (1959). This technique involves partitioning of the correlation coefficient to determine direct (unidirectional pathways 'P') and indirect influence through alternate pathways (pathway (P) x correlation coefficient (r)) of various variables over seed yield per plant. Seed yield was considered as the resultant variable and the others as casual variables. The statistical significance of correlations was calculated according to Gomez and Gomez (1984).

3. Regression models:

To describe the seed yield per plant of sesame (y), multiple linear regression was fitted using different variables. The general regression model applied was: $y = \sum_{i=1}^{p} a_i x_i$

Where x_i is the input variable used for each particular model, a_i is the coefficient to be determined, and p is the number of input variables used after the stepwise procedure.

3.1. Multiple linear regression analysis was computed as applied by Draper and Smith (1981).

To select the model that best described the seed yield per plant of sesame, the following were calculated for each model:

a. Variable coefficients (a_i) and their standard error (SE).

b. Value of statistic test t (t value) and the F value for each variable.

c. Coefficient of determination (R^2), adjusted R-squared (R^2 adj.), standard error of estimation, and mean absolute error (MAE):

$$R^2 adj. = 1 - \left(\frac{n-1}{n-p} \right)\left(1 - R^2\right)$$

Where n is the number of observations (in this case 30) and p the number of variables selected. The expression *(n-p)* represents the degrees of freedom that affect the R^2 adjusted. The MAE is given by:

$$MAE = \left(\frac{\sum_{j=1}^{n}|e_j|}{n} \right)$$

Where e_j is the difference between the estimated and the real percentage value of seed yield per plant.

d. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in data.

2468

3.2. The stepwise multiple linear regression as applied by Draper and Smith (1981), was used to compute a sequence of multiple regression equations in a stepwise manner. At each step, one variable was added to the regression equations, it was the one that caused the maximum reduction in the residual sum of squares. Equivalently, it was the variables that had the highest partial correlation with the dependent variable adjusted for the variables already added. Similarly, it was the variable which if added, had the highest *F* value in the regression analysis of variance. Moreover, variables were forced into the regression equation and automatically removed when the values were below.

4. The factor analysis method was discussed by Cattell (1965). The method consists of the reduction of a large number of correlated variables to a much smaller number of clusters of variables called factors. After the loading of the first factor were found, they were taken into account when the second factor was calculated. The process was repeated on the residual matrix to find further 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 loadings was submitted to a varimax orthogonal rotation, as applied by Kaiser (1958). The effect of rotation is to accentuate the larger loadings 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. Thus, factor analysis indicates both groupings and contribution percentage to total variation in the dependence structure. The factor loadings of the rotated matrix, the percentage variability explained by each factor and the communalities for each variable were determined, to know the way in which yield components were related to each other.

RESULTS AND DISCUSSION

The average values, minimal and maximal values together with statistics associated with means are given in Table (1) for the ten characters evaluated in this study. The range in general shows that there was wide variability in each character evaluated.

Table 1. Average value, minimum, maximum, standard deviation and coefficient of variation for sesame yield characteristics.

Character	Statistic				
	Average	Mini.	Maxi.	S.D.	C.V.%
Days to flowering	50.83	40.10	59.53	6.48	12.76
Plant height (cm)	162.85	150.37	189.65	8.71	5.35
Number of fruiting branches	1.92	1.30	2.80	0.44	22.85
Stem height to the first capsule (cm)	55.03	41.90	81.3	8.65	15.72
Fruiting zone length (cm)	107.61	92.99	119.83	6.56	6.09
Number of capsules on main stem	88.60	40.90	111.20	18.90	21.34
Number of capsules per plant	124.49	72.90	156.70	23.19	18.63
Capsule density	0.82	0.44	0.96	0.16	19.89
1000-seed weight (g)	3.46	2.70	4.20	0.26	7.72
Seed yield, g / plant	19.58	13.15	25.90	3.23	16.48

1. Simple correlation:

The matrix in Table (2) shows Pearson product moment correlations between each pair of variables. These correlation coefficients range between -1 and +1 and measure the strength of the linear relationship between the variables.

Table 2. A matrix of simple correlation coefficients for 10 characters of thirty sesame genotypes grown in 2004 and 2005 seasons.

Character	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
X_1	1.00								
X_2	0.34	1.00							
X_3	-0.07	-0.26	1.00						
X_4	0.40*	0.69**	-0.22	1.00					
X_5	-0.04	0.41*	-0.08	-0.36*	1.00				
X_6	-0.26	-0.41*	0.31	-0.79**	0.48**	1.00			
X_7	-0.14	-0.31	0.38*	-0.69**	0.46**	0.92**	1.00		
X_8	-0.25	-0.55**	0.37*	-0.75**	0.24	0.96**	0.88**	1.00	
X_9	0.28	0.46*	-0.07	0.65**	-0.23	0.60**	-0.50**	-0.61**	1.00
Y	-0.36	-0.30	0.47**	-0.61**	0.38*	0.82**	0.85**	0.78**	0.49*

*, ** Significant at 0.05 and 0.01 probability levels, respectively.
X_1 = flowering date, X_2 = Plant height, X_3 = Number of fruiting branches, X_4 = stem height to the first capsule, X_5= Fruiting zone length, X_6 = Number of capsules on main stem, X_7 = Number of capsules per plant, X_8 = Capsule density, X_9= 1000-seed weight, Y= Seed yield per plant.

The seed yield per plant was positively and significantly correlated with number of fruiting branches (r = 0.47), fruiting zone length (r = 0.38), number of capsules on main stem (r = 0.82), number of capsules per plant (r = 0.85) and capsule density (r = 0.78). However, the correlation coefficients between seed yield per plant with stem height to the first capsule and 1000-seed weight were significant and negative suggesting the fact that increasing seed yield per plant in sesame along with these characters will be difficult. Such results are in concurrence with the results of Sanjeevaiah and Joshi (1974), Uzo (1985), Singh *et al.* (1997) and Nimbalkar *et al.* (2005) in sesame. Asthana and Raj (1970) found similar results for seed yield and capsule number on the main stem, Khidir and Osman (1970) for plant height, branch number and capsule number with seed yield, and Osman and Khidir (1974) for days to full maturity with seed yield.

2. Path coefficient analysis:

Path analysis can provide an effective means of partitioning the correlation coefficient into direct and indirect effects. Table (3) shows the direct and indirect coefficients of nine characters on seed yield. Among the effects of the investigated characters on seed yield per plant, number of capsules on main stem showed the highest direct effect (P_{6y} = 0.622) followed by number of capsules per plant (P_{7y} = 0.591) and plant height (P_{2y} = 0.551) suggesting the importance of these characters in sesame. This was in agreement with the earlier reports of Thangavelu and Rajasekaran (1983). The characters, flowering date, stem height to the first capsule, fruiting zone

length and capsule density recorded high and negative direct effects. However, Gupta and Chopra (1984), while confirming the present computations with respect to the direct contribution of number of capsules and number of branches per plant, they reported negative direct effects between plant height and seed yield, which is contradictory to the present investigation. The variation in the result may be due to genotypic differences in the material used and the environmental conditions. From the above results, it can be inferred that the characters: number of capsules on main stem, number of capsules per plant and plant height are to be given prime importance as they revealed a significant positive correlation coefficient and a high positive direct effect compared to other traits.

Table 3. Direct (diagonal) and indirect (above and below diagonal) effects of yield components on yield of sesame genotypes.

Character	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	R_{iy}
X_1	-0.210	0.187	-0.080	-0.135	0.018	-0.151	-0.072	0.085	-0.010	-0.36
X_2	-0.072	0.551	-0.067	-0.250	-0.191	-0.255	-0.183	0.187	-0.022	-0.30
X_3	0.046	-0.143	0.191	0.079	0.037	0.192	0.224	-0.126	0.003	0.47
X_4	-0.089	0.380	-0.059	-0.363	0.168	-0.510	-0.419	0.341	-0.059	-0.61
X_5	0.008	0.225	-0.015	0.130	-0.467	0.298	0.271	-0.081	0.011	0.38
X_6	0.079	-0.255	0.059	0.299	-0.244	0.622	0.558	-0.327	0.028	0.82
X_7	0.029	-0.170	0.072	0.250	-0.214	0.572	0.591	-0.300	0.024	0.85
X_8	0.053	-0.303	0.070	0.272	-0.122	0.597	0.520	-0.341	0.029	0.78
X_9	-0.059	0.253	-0.013	-0.235	0.107	-0.373	-0.295	0.170	-0.048	-0.49

Values underlined denote direct effect.
Residual effect = 0.433
X_1= flowering date, X_2 = Plant height, X_3 = Number of fruiting branches, X_4 = stem height to the first capsule, X_5 = Fruiting zone length, X_6 = Number of capsules on main stem, X_7 = Number of capsules per plant, X_8 = Capsule density, X_9 = 1000–seed weight, Y = Seed yield per plant.

Besides, the residual effect (0.433) was high in magnitude which shows that some other important yield contributing characters are contributing to yield and should be taken into consideration.

3. Multiple linear regression analysis:
In the present experiment, the linear regression of seed yield per plant and other agronomic traits was analyzed. The results showed that the multiple linear regression could be used, since there was significant difference between seed yield per plant and other agronomic traits (F = 9.55, P < 0.01) (Table 4).

Since the P-value in the ANOVA table is less than the 0.05 level of probability, there is a statistically significant relationship between the variables at the 95.0% confidence level. The R-Squared statistic indicates that the model explains 81.12% of the variability in seed yield. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 72.63%.

Table 4. Analysis of multiple linear regression of seed yield per plant and other agronomic traits of sesame, over both 2004 and 2005 seasons.

Source	D.f.	Sum of Squares	Mean Squares	*F*-Ratio	*P*-Value
Model	9	245.04	27.22	9.55	0.0000
Residual	20	56.99	2.84		
Total (Corr.)	29	302.03			

R-squared = 81.12 percent, R-squared (adjusted for d.f.) = 72.63 percent, Standard Error of Est. = 1.68, Mean absolute error = 1.04, Durbin-Watson statistic = 1.74 (P=0.11)

The standard error of the estimate shows that the standard deviation of the residuals is 1.68. The mean absolute error (MAE) of 1.04 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in data. Since the *P*-value is greater than 0.05, there is no indication of serial autocorrelation in the residuals at the 95.0% confidence level.

In determining whether the model can be simplified, it is noticed that the highest *P*-value on the independent variables was 0.736, belonging to seed index (Table 5). Since the *P*-value is greater or equal to 0.05, that term is not statistically significant at the 95.0% or higher confidence level. Consequently, we should consider removing seed index from the model.

Table 5 shows the results of fitting a multiple linear regression model to describe the relationship between seed yield and nine independent variables. The equation of the fitted model was:

Seed yield per plant (Y) = 9.159 - 0.106 X_1 + 0.203 X_2 + 1.40571 X_3 - 0.135 X_4 - 0.230 X_5 + 0.106 X_6 + 0.082 X_7 - 6.722 X_8 - 0.582 X_9

The relative contribution i.e., coefficient of determination (R^2%) for yield factors explained 81.12% of the total variation in seed yield which could be linearly related to variation in all variables, and 18.88% could be due to the residual. In this analysis, all variables were added in the prediction equation. It is well known that as more variables were added, the interpretation of association become more complex. On the other hand, some variables may contribute a little to the accuracy of the prediction equation. In addition, given that the number of observations was much greater than the number of potential independent variables (x) under consideration, the addition of a new variable will always increase R^2 but it will not necessary increase the precision of the estimate of the response. At this point, the stepwise multiple linear regression analysis was carried out to determine the best variables accounted for most of variance in yield.

Table 5. Variables selected by the model of multiple linear regression with their corresponding coefficient values (a_i), standard error (SE), *t* value, *F* value and significance level (*p*) according to the model used.

Parameter	Estimate	Standard Error	*T* Statistic	*P*-Value
Constant	9.159	15.667	0.584	0.565
Days to flowering (x_1)	-0.106	0.056	-1.889	0.043
Plant height (x_2)	0.203	0.381	0.534	0.598
Number of fruiting branches (x_3)	1.405	0.831	1.689	0.106
Stem height to the first capsule (x_4)	-0.135	.373	-0.362	0.721
Fruiting zone length (x_5)	-0.230	0.409	-0.562	0.580
Number of capsules on main stem (x_6)	0.106	0.170	0.622	0.540
Number of capsules per plant (x_7)	0.082	0.039	2.072	0.031
Capsule density (x_8)	-6.722	17.039	-0.394	0.697
1000-seed weight (x_9)	-0.582	1.708	-0.341	0.736

4. Stepwise multiple linear regression analysis:

The results in (Table 6) show that the stepwise method of linear regression could be used, since there was significant difference between seed yield per plant and other agronomic traits ($F = 45.84$, $P < 0.01$). It was observed that two characters, viz. (number of capsules per plant and flowering date, had significant effect on seed yield per plant, and both *P* values were < 0.05 (Table 6). Variables removed were plant height, number of fruiting branches, stem height to the first capsule, fruiting zone length, number of capsules on main stem, capsule density and 1000-seed weight.

Stepwise regression, with backward elimination, shows the results of fitting a linear regression model to describe the relationship between seed yield per plant and 9 independent variables. According to this analysis the predication equation runs as follows: (Y) = 10.951 - 0.110 X_1 + 0.114 X_7 (R-squared =77.25 percent), where Y, X_1 and X_7 represent seed yield per plant, flowering date and number of capsules per plant, respectively.

This model is apparently sufficient to cover most of the variation in yield in that the adjusted R^2 equals 75.56%. Since the *P*-value in the ANOVA table is less than 0.05, there is a statistically significant relationship between the variables at the 95.0% confidence level. The R-Squared statistic indicates that the model as fitted explains 77.25% of the variability in seed yield per plant. The adjusted R-squared statistic, which is more suitable for comparing models with different numbers of independent variables, is 75.56%. The standard error of the estimate shows the standard deviation of the residuals to be 1.59. The mean absolute error (MAE) of 1.16 is the average value of the residuals. The Durbin-Watson (DW) statistic tests the residuals to determine if there is any significant correlation based on the order in which they occur in the data file. Since the *P*-value is greater than 0.05, there is no indication of serial autocorrelation in the residuals at the 95.0% confidence level.

In determining whether the model can be simplified, it was noticed that the highest *P*-value on the independent variables is 0.024, belonging to

flowering date. Since the *P*-value is less than 0.05, that term is statistically significant at the 95.0% confidence level. Consequently, we probably don't want to remove any variables from the model.

Table 6. Analysis of linear regression of seed yield per plant and other agronomic traits of sesame.

(A) ANOVA (c)

Model	D.f.	Sum of Squares	Mean Squares	*F*-Ratio	*P*-Value
1. Regression	1	218.83	218.83	73.64	0.000(a)
Residual	28	83.20	2.97		
Total	29	302.03			
2. Regression	2	233.32	116.66	45.84	0.000(b)
Residual	27	68.71	2.54		
Total	29	302.03			

(a) Predictors: (constant), flowering date.
(b) Predictors: (constant), flowering date and number of capsules per plant.
(c) Dependent variable: seed yield per plant.

(B) Coefficients (a_i)

Model	Unstandardized coefficients		Standardized coefficients		*P*-Value
	a_i	Standard Error	a_i	*T* Statistic	
1. (Constant)	4.830	1.740		2.76	0.010
Number of capsules per plant (x_7)	0.118	0.014	0.851	8.58	0.000
2.(Constant)	10.950	3.030		3.61	0.001
Number of capsules per plant (x_7)	0.114	0.013	0.821	8.85	0.000
Flowering date (x_1)	-0.110	0.046	-0.221	-2.38	0.024

Dependent variable: seed yield per plant (y), $y = 10.951 + 0.114 X_7 - 0.110 X_1$
R-squared = 77.25 percent, R-squared (adjusted for d. f.) = 75.56 percent, Standard Error of Est. = 1.59, Mean absolute error = 1.16, Durbin-Watson statistic = 1.98 ($P = 0.43$)

5. Factor analysis:

The total variance explained by factors is indicated in Table 7, only the first 3 factors, which account for 81.9% of the total variance, are important. A principal factor matrix after orthogonal rotation for these 3 factors is given in (Table 8). The values in the table, or loadings, indicate the contribution of each variable to the factors. For the purposes of interpretation only, those factor loadings greater than 0.5 were considered important, these values are highly lighted in bold in (Table 8).

Factor 1, which accounted for about 41% of the variation, was strongly associated with fruiting zone length, number of capsules on main stem, number of capsules per plant, and capsule density. This factor was regarded as productivity per plant factor since it included several traits which are components of yield. These variables had positive loadings in factor 1. The sign of the loading indicates the direction of the relationship between the factor and the variable.

Factor 2 which, accounts for about 25% of the variation were named a reproductive (fertility) and productivity per plant factors since it consisted of days to flowering, seed index, plant height and stem height to the first capsules, which are associated with fertility and productivity. Again all these

variables had positive loadings. The third factor was strongly associated with number of fruiting branches only, which accounts for about 16% of the variation.

This study indicated that selection of variables in the productivity per plant factor (factor 1) could enable breeders to better realize the desired increment in seed yield of sesame.

Table 7. Total variance explained for each factor based on nine different characters of 30 sesame genotypes.

Factor	% of variance	Cumulative %
Days to flowering	52.138	52.138
Plant height (cm)	17.621	69.759
Number of fruiting branches	12.149	81.908
Stem height to the first capsule (cm)	8.506	90.414
Fruiting zone length (cm)	4.763	95.177
Number of capsules on main stem	3.776	98.953
Number of capsules per plant	0.947	99.900
Capsule density	0.061	99.961
1000-seed weight (g)	0.039	100

Table 8. Principal factor matrix after varimax rotation for nine characters of 30 genotypes of sesame.

Variable	Factors			Communality
	Factor 1	Factor 2	Factor 3	
Days to flowering	-0.093	0.701	-0.169	0.528
Plant height (cm)	-0.083	0.817	0.489	0.913
Number of fruiting branches	0.248	0.082	0.828	0.754
Stem height to the first capsule (cm)	-0.686	0.611	0.113	0.857
Fruiting zone length (cm)	0.764	0.289	0.494	0.911
Number of capsules on main stem	0.911	-0.295	-0.198	0.956
Number of capsules per plant	0.904	-0.122	-0.279	0.909
Capsule density	0.789	-0.393	-0.364	0.910
1000-seed weight (g)	-0.543	0.580	-0.050	0.633
Variance	3.696	2.214	1.461	7.371
% Var.	0.411	0.246	0.162	0.819

Numbers in bold are those with factor loadings greater than 0.50.

REFERENCES

Asthana, K. S. and U. K. Raj (1970). Correlation studies in till (S*esamum indicum* L.). Allahabad Tmr., 44:385-386.

Bhatt, G. M. (1973). Significance of path coefficient analysis in determining the nature of character association. Euphytica, 22:89-97.

Board, J. E., M. S. Kang and B. G. Harville (1997). Path analysis identify indirect selection criteria for yield of late planted soybean. Crop Sci., 37:879-884.

Cattell, R. B. (1965). Factor analysis: An introduction to essentials. I. The purpose and underlying models. Biometrics, 21:195-215.

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.

Draper, N. R. and H. Smith (1981). Applied regression analysis .2nd ed. John Willy and Sons, New York.

El-Deeb, A. A. and N. A. Mohamed (1999). Factor and cluster analysis for some quantitative characters in sesame (S*esamum indicum* L.).The 34thAnnual Conference of Statistics, Computer Science and Operations Research 4-6, Dec., 34, Part (II):14-27.

El-Sergany, Dawlat, Z. (1992). Comparative study between full model and stepwise regression analysis for maize and sugar beet yield components. J. Agric. Sci. Mansoura Univ. 17 (7):2291-2295.

Freed, R., S. P. Einensmith, S. Gutez, D. Reicosky, V. W. Smail and P. Wolberg (1989). User's Guide to MSTAT-C Analysis of agronomic research experiments. Michigan State University, East Lansing, USA.

Gomez, A. K. and A. A. Gomez (1984). Statistical procedures for agricultural research. John Wiley & Sons. NewYork, USA.

Gupta, B. S. and D. P. Chopra (1984). Genetic variability, correlation and path coefficient analysis in sesamum. Indian j. Agric. Sci., 54:1030-1033.

Ibrahim, A. F., F. M. El-Rayes, A. I. Ragab and H. N. El-Rassas (1984). Factor analysis of seed yield components in sesame (*Sesamum indicum* L.). Proc. 2nd Mediterranean Conf. Genetic, Cairo, March:277-288.

Kaiser, H. F. (1958). The varimax criterion for analytic notation in factor. Psychometrika, 23:187-200.

Kang, M. S., J. D. Miller and P. Y. P. Tai (1983). Genetic and phenotypic path analysis and heritability in sugarcane. Crop Sci., 23:643-647.

Khidir, M. O. and E. G. H. Osman (1970). Correlation studies in some agronomic characters in sesame. Exp. Agric., 8:27-31.

Leilah, A. A. and S. A. Al-Khateeb (2005). Statistical analysis of wheat yield under drought conditions. J. of Arid Enviroments, 61:483-496.

Mansouri, S. and M. S. Najafabadi (2004). Study and systemic analysis on yield and yield components association for sesame (*Sesamum indicum* L.) breeding. Seed and Plant, 20:149-165

Manugistics, (1998). Statgraphics Plus for Windows. Reference manual. Manugistics, Rock ville, MD.

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. Sci. Mansoura Univ. 16 (7): 1483-1490.

Neter, J., M. Khutner, C. Nachtsheim and W. Wasserman (1996). Applied Linear Statistical Models.4th. Ed. Irwin Series. Time Mirror. Education Group, pp.111-121.

Nimbalkar, C. A., P. A. Navale, A. P. Baviskar and V. H. Bajaj (2005). Character association and path analysis in sesame. Madars Agric. J., 92 (4-6):323-235.

Oaizar, A., M. A. Rana and S. U. H. Siddiqui (1991). Sunflower seed yield as influenced by some agronomic and seed characters. Euphytica, 56:137-142.

Osman, E. G. H. and M. O. Khidir (1974). Relations of yield components in sesame. Exp. Agric., 10:97-103.

Samonte, S. O., L. T. Wilson and M. McClung (1998). Path analysis of yield and yield-related traits of fifteen diverse rice genotypes. Crop Sci., 38:1130-1136.

Sanjeevaiah, R. S. and M. S. Joshi (1974). Correlation and genetic variability in sesame. Curr. Sci., 11:144-145.

Shabana, R. and N. M. Abu-Hagaza (1984). Correlation and path coefficient analysis of some characters contributing to seed yield. 9th Inter. Congress, Stat., Comp. Sci., Social and Demographic Res. 31 March 10 April:61-81, Ain Shams Univ.

Shrief, S. A. (1983). Comparative performance of characters in some mutant lines and local cultivars of sesame (*Sesamum indicum* L.). M.Sc. Thesis, Fac. Agric., Cairo Univ.

Singh, P. K., R. K. Dixit and R. K. Yadav (1997). Estimates of genetic parameters, character association and path analysis in sesame. Crop Res., 13(1):115-119.

SPSS (1999). SPSS for Windows, version 10.0. SPSS Inc., Chicago, USA.

Steel, R. G. D. and J. H. Torrie (1980). Principles and procedures of statistics. 2. ed. New York: McGraw-Hill, 633 p.

Subramanian, S. and M. Subramanian (1994). Correlation studies and path coefficient analysis in sesame (*Sesamum indicum* L.). J. Agron. Crop Sci., 173:241-248.

Thangavelu, M. S. and S. Rajasekaran (1983). Correlation and path coefficient analysis in sesame *(Sesamum indicum* L.). Madras Agric. J., 70 (2):109-113.

Uzo, J. O. (1985). Yield, yield components and nutritional attributes of cultivated sesame, *(Sesamum. indicum* L.) and its endemic wild relatives in Nigeria. FAO Pl. Prod. And Prot. Paper: Sesame and Safflower: status and potentials, pp. 166-176.

Walton, P. D. (1972). Factor analysis of yield in spring wheat *(Triticum aestivum* L.). Crop Sci., 12:731-733.

William, M. (2007). OpenStat For Windows, version 1.9, Tutorial Manual. I. S. U., USA.

تحليل العامل المتعدد لتقييم محصول السمسم ومكوناته

أشرف عبد الأعلى عبد المحسن
قسم المحاصيل – كلية الزراعة – جامعة القاهرة

أجريت هـذه الدراسـة بقصـد تقيـيم خمسـة مـن الطـرق الاحصـائية المسـتخدمة لتحليـل واختصار المتغيرات المرتبطة فى دراسة العلاقة بين المحصول ومكوناته وهذه الطرق هى : الارتباط البسيط , تحليل المسار , تحليل الانحدار المتعدد , تحليل الانحدار المتعدد المرحلى , تحليل العامل , وطبقت هذه الاساليب الاحصائية على مكونات المحصول لعدد ثلاثون من التراكيب الوراثية من السمسم وهى: ثلاثة أصناف تجارية بالإضافة إلى سبعة وعشرون من الطفرات المبشرة. وذلك خلال موسمى 2004 , 2005 فى منطقة ابو رواش – محافظة الجيزة , وقد تم دراسة عشرة متغيرات هى : ميعاد التزهير ، ارتفاع النبات ، عدد الأفرع الثمرية على النبات ، الارتفاع حتى أول كبسولة ، طول المنطقة الثمرية ، عدد الكبسولات على الساق الرئيسى ، عدد الكبسولات على النبات , كثافة الكبسولات على الساق الرئيسى , وزن 1000 بذرة ، محصول النبات.

أظهرت مصفوفة الارتبـاط البسـيط وتحليـل المسـار أن الصـفات ذات الارتبـاط العـالى الموجـب بمحصول النبات كان لها ايضـا أعلـى تـأثير موجب على محصول النبات وهى عدد الكابسولات على الساق الرئيسـى ، عدد الكبسـولات علـى النبات. أوضحت النتـائج وجود علاقـة عاليـة المعنويـة وموجبـة بين محصول البذرة للنبات والمكونات تحت الدراسة وهى عدد الافرع الثمرية على النبات ، عدد الكبسولات على الساق الرئيسى ، عدد الكبسولات على النبات ، وكثافة الكبسولات. أمـا نتـائج تحليـل المسـار فقد اوضحت أن عدد الكبسولات على الساق الرئيسى , عدد الكبسولات على النبات, ارتفاع النبات لها اكبر الأثر على التباين الكلى , كذلك أظهرت النتائج أن تحليـل المسـار يتسـع لدراسة مزيد من المتغيرات ، حيث ان المتغيرات تحت الدراسة لم تفسر سوى 57% من التبـاين الكلى .

أظهرت نتائج تحليل الانحدار المتعدد المرحلـى أن صفتى عدد الكابسولات على النبات وميعاد التزهير فسرت 77.25% من التباين الكلى فى محصول النبات ، حيث كانت أفضل معادلة للتنبؤ هي:

$$(Y) = 10.951 - 0.110 \ X_1 + 0.114 \ X_7$$

أوضحت الدراسة أن طريقة الانحدار المتعدد المرحلى افضل طريقة تعطى أفضل معادلة للتنبؤ بالمحصول مقارنة بتحليل الانحدار المتعدد (النموذج الكامل) ولا يوجد فرق بينها وبين طريقة تحليل الانحدار المتعدد وذلك يتضح من خلال مقارنة قيمة معاملى التحديد والتحديد المعدل ، الخطأ القياسى , الخطأ القياسى المطلق , وإحصاء Durbin-Watson لكلا التحليلين ، كذلك تبين أيضا باستخدام طريقة تحليل الانحدار المتعدد المرحلى أن المتغيرات المقبولة كانت على درجة كبيرة من الأهمية في تحقيق الهدف فى الحصول على أهم المتغيرات التـى تفسـر نسـبة كبيـرة مـن التبـاين الكلـى والتـى تخـدم بـرامج التربية , في حين أن الانحدار المتعدد يأخذ كل المتغيرات فى الاعتبار.

أظهرت نتائج تحليل العامل تحليـل واختصـار للمتغيرات المرتبطة إلـى ثلاثة عوامل يضم الأول طول المنطقة الثمرية , عدد الكابسولات على الساق الرئيسى ، عدد الكبسولات علـى النبات وكثافة الكبسولات ويساهم بحوالى 41% من التباين الكلى , بينما اشتمل العامل الثانى على متغيرات ميعاد التزهير والارتفاع حتى أول كابسولة ووزن الألف بذرة ويساهم بحوالى 25% من التباين الكلى ، بينما يمثل عدد الأفرع الثمرية العامل الثالث والذي يساهم بـ 16% من التباين الكلى وتساهم العوامل الثلاثة مجتمعة بنحو   81.9% من التباين الكلى. أوضحت النتائج المتحصل عليها ان طريقة تحليل العامل كانت اكثر الطرق كفاءة وانها تعطى معلومات نسبية أكبر عن المتغيرات المتداخلة.

تخلص هذه الدراسة إلى أن الأساليب الإحصائية المستخدمة اتفقت على أنه للحصول على محصول عالى من نباتات السمسم ، يجب انتخاب مواد تجريبية ذات صفات عالية من عدد الكبسولات على النبات ، عدد الكبسولات على الساق الرئيسى ، ارتفاع النبات ، وزيادة كثافة الكبسولات على الساق الرئيسى .