Scientific Research and Review Journal/ 1 (3): 88-100, 2013
© 2013 Available online at www.srrjournal.com
ISSN 0000 ©2013 SRRJ Journal
Science Explorer Publications

Correlation and regression analysis in barley

Ashraf A. Abd El-Mohsen

Agronomy Department, Faculty of Agriculture, Cairo University, El-Gamaa Street, P.O. Box12613 Giza, Egypt.

Corresponding author E-mail: dr_ashraf200625@yahoo.com, ashraf_stat91@yahoo.com.

ABSTRACT: Two field experiments were carried out at the Experimental Station, Faculty of Agriculture, Cairo University, during the two successive winter seasons of 2008/09 and 2009/10. Six cultivars were grown in a randomized complete blocks design with three replications and evaluated for eight characteristics. Combined analysis of variance was done from the mean data obtained for each characteristic over two seasons and correlation and regression analysis were carried out to better understand the relationship between yield and some yield components. Results indicated that seasons significantly affected all traits and interaction between seasons and cultivars was also significant. Highly significant differences and adequate genetic variability were observed among cultivars for all the eight characters. The results of the correlation coefficients of traits with grain yield revealed that the grain number per spike (r=0.84**), grain weight/spike (0.87**), 1000-grain weight (r=0.88**), number of spikes per square meter (r=0.68*) and spike length (r=0.67*) had the highest significant positive correlation with grain yield, indicating dependency of these characters on each other. Heading date was negatively and highly correlated with number of spikes per square meter (r= - 0.58*), number of grains per spike (r= - 0.87**), grain weight per spike (r= - 0.89**), thousand grain weight (r= - 0.87**) and spike length (r= - 0.75**). The criteria used in identifying the best subsets are based on monotone functions of the residual sum of squares (RSS) such as R^2, adjusted R^2 and Mallow's Cp. Results revealed that the best subset regression model, based on the three different criteria, the predicted equation for barley grain yield per fed (Y) was $Y = 3.12 – 0.006\ x_1 – 0.019\ x_2 + 0.0007\ x_3 + 0.020\ x_6 – 0.149\ x_7$. The simplified results from best subset regression analysis show that the highest coefficient of determination (R^2=96.5%), adjusted R^2 (93.6%) and lowest Mallows' conceptual predictive (Cp) value (5.8), and has five-independent variable model with all variables except number of grain per spike (X_4) and grain weight per spike (X_5). The Best Subset Multiple Regression analysis indicates that adding the variable number of grain per spike (X_4) and grain weight per spike (X_5) does not improve the fit of the model.
Key words: Barley, Best subset regression, Correlation, Grain yield and components, Mallow's Cp, Multiple regression analysis, Regression analysis.

INTRODUCTION

Barley (*Hordeum vulgare* L.) is one of the most important cereal crops in Egypt and is ancient as the origin of agriculture itself. It is considered as one of the most suitable cereal crops, which can survive and grow over a wide range of soils and under many adverse climatic conditions compared with many other cereal crops. It ranks fourth after wheat, rice and maize in the world's cereal production.

In most of the crop improvement programmes, raising grain yield is one of the major objectives. The information on association between grain yield with its components is prerequisite for breeding programmes aiming at yield improvement. Therefore, association and regression studies were undertaken in barley. In recent years, breeding of new barley varieties with high yield and good quality has been regarded as a very important research approach by agricultural science researchers in our country, and several new varieties of barley with high yield and good quality have been developed.

Yield in barley is a very complex trait and is a result of the interaction between various yield components. Knowledge of the association between yield-related traits is of immense importance to the selection of desired combinations of characters. Further, correlation analysis provides information about the correlated responses to selection of important plant characters. Correlation and regression analyses are multivariate tools that help to study the interrelationships and inter-dependence among traits.

In many crops, especially cereal crops, yield depends on some plant attributes such as plant height, number of leaves, stalk thickness and tillering capacity etc. These plant attributes are referred to as the

independent variables, covariates, predictors, or regressors, while yield is the corresponding dependent variable. Each of these regressors contributes to the variation in the yield of a variety, although the contribution varies from one variety to another.

Correlation analysis among yield and yield components is one of the prerequisite techniques to determine the influence of environment on productivity and yield potential. The information on the nature and magnitude of correlation coefficients help breeders to determine the selection criteria for simultaneous improvement of various characters along with yield. Determination of correlation coefficients between various barley characters helps to obtain best combinations of attributes for obtaining higher return per unit area.

The statistical technique that is used to establish the existence of linear relationship between the dependent variable and the independent variables is the Regression Analysis. If there is a single independent or predicator variable is referred to as simple linear regression, while if it involved more than one independent or predictor variables we have the case of Multivariate regression or multiple regression analysis.

The aim of Multiple Regression Analysis (MRA) is to find the best set of the independent variables which can explain dependent variable on condition that the assumptions are provided. The term regression is used to establishing the actual relationship between two or more variables. But scientific, social, economic and agricultural phenomena do not confine to two variables. In these studies we often need to give actual relationship between two or more than two variables (Agrawal 1991). For this purpose we choose the method of all possible regressions. This technique requires that investigator fit all the subset regression models involving one predictor variable, two predictor variables and so on. Each subset regression model was then evaluated according some suitable criterion like R^2, R^2-adjusted and Mallow's Conceptual predictive Cp statistic and the best subset regression model was selected (Draper and Smith 1998).

The objectives of this study were to (i) evaluate six barely cultivars for grain yield and their components. (ii) estimate the relationship between yield and some yield components and to determine the most important characters that can be used as selection criteria in a breeding program for yield improvement in barley crop. iii) compare different prediction models by using best subset regression, based on three different criteria.

MATERIALS AND METHODS

Location of Study and Plant Materials

The present study was carried out at the Agricultural and Research Station, Faculty of Agriculture, Cairo University, Giza, Egypt (30° 02'N Latitude and 31° 13' E Longitudes, Altitude 22.50 m), during 2008/2009 and 2009/2010 growing seasons.

Six barley cultivars viz., Giza 123, Giza 126, Giza 2000 (covered barley) and Giza 129, Giza 130, Giza 131 (naked barley) were obtained from the Agric. Res. Center, Ministry of Agriculture, Egypt.

Experimental design and plot arrangement

The experiments were designed in a Randomized Complete Blocks Design (RCBD) with three replications. Each replicate consisted of six plots, each devoted to one cultivar. Plots were 3 m^2 including five rows, 3 m long spaced 20 cm apart. All experimental plots were subjected to uniform agronomic practices.

Cultivation Practices

Sowing date was Dec. 3 and 5 in the first and second season, respectively. Sowing was done by hand in plots of 5 rows, 3 m long and 20 cm wide with plants spaced 5 cm apart within rows. Sowing rate was 60 kg seed/feddan for all genotypes. Grains were drilled in rows using dry method of planting. The preceding crop was maize in both seasons. Fertilizers were applied at the rate of 100 kg /fed ammonium nitrate (33.5% N) in two equal doses, the first dose was added at tillering stage and the second dose was added at shooting stage, while phosphorus and potassium were added at a rate of 150 kg/fed, calcium super phosphate (15.5% P_2 O_5) and 50 kg/fed potassium sulfate (48.5% K_2 O), respectively. Three irrigations were added during growth by flooding system. In all experiments, weeds were controlled by hand as needed. Other cultural practices were kept constant for all the treatments.

Data Collection

Data on different agronomic traits were collected on both of plant and plot basis. Measurements and observation of examined characters were done on ten plants randomly chosen in the middle-row of each plot. Eight different traits were measured including grain yield and morphological traits and yield components, plant height, days to 50% heading, number of spikes per square meter, number of grains per spike, grain weight per spike, thousand grain weight and spike length. At maturity, grain yield of the three middle rows of each plot was determined and converted into tons per feddan.

Statistical Analysis

Trait means and the coefficients of variation (CV) were determined for the studied traits. The analysis of variance (ANOVA) (Steel *et al* 1997) for the randomized complete blocks design was performed for each variable in each trial and combined analysis of variance was done for the data of the two studied seasons when error variances of both seasons were homogeneous. The least significant difference (LSD) test was used to compare treatment means using the computer program MSTAT-C (MSTAT-C 1991).

To analyze the relationships between grain yield and yield components accurately, correlation and regression analysis was performed for all genotypes using MINITAB (2005) 14 software statistical package. The data over two years subjected to estimate correlation and regression coefficients among measured characteristics.

Criteria for evaluating subset regression models

The regression model selection procedure is designed to help select the independent variables used in building a multiple regression model to predict a single quantitative dependent variable Y. The procedure considers all possible regressions involving different combinations of the independent variables. It compares models based on the coefficient of multiple determination R^2, adjusted R-Square, Mallows' *Cp*-statistic (Draper and Smith 1998 and Montgomery *et al* 2001), and the mean squared error. Of particular value for predictive model selection is the *Cp*-statistic proposed in Mallows (1973, 1995 and 1997).

All possible subsets, this method builds all one-variable models, all two-variable models, and so on, until the last all-variable model is generated. The method requires a powerful computer (because a lot of models are produced), and selection of any one of the criteria: R-squared, adjusted R-squared, Mallows' Conceptual predictive *Cp*.

To determine an appropriate subset of the predictor variables, there are several different criteria available. They include R^2, adjusted R^2, Mallow's Conceptual predictive *Cp*. In order to establish adaptation of the estimated regression model by empirical data we use standard error of the sample regression which represents the estimation of standard deviation of the random error σ_ε. It is marked by S_ε, and it is presented as square root of repetition, or:

$$S_\varepsilon = \sqrt{\sigma^2} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-k-1}} = \sqrt{\frac{SS_e}{n-k-1}}$$

where SSE is a sum of square root aberration of the empirical points of regression model (Error Sum of Squares).

The standard error of regression as absolute measure of the unexplained variability is not convenient for comparison. That is the reason why we use relative indicator coefficient of multiple determination R^2. It is presented as a measure of explained variability and is calculated by this equation:

$$R^2 = \frac{\sum (\hat{y}_i - \overline{y})^2}{\sum (y_i - \overline{y})^2} = \frac{SS_R}{SS_y}$$

The coefficient of multiple determination shows the percentage of variations of dependent variable Y which is described by common influence of independent variables which are involved in this model. During its calculation we should take care of the number of independent variables and of sample size. It is achieved by calculation of the adjusted coefficient of multiple determination:

$$R^2_{adj.} = 1 - \frac{n-1}{n-k-1}.(1-R^2)$$

where: n is the sample size and k number of independent variables.

Mallow's Conceptual predictive (*Cp*) Criterion:

Mallow's *Cp* is a technique for model selection in regression (Mallows 1973). The *Cp*-statistic is defined as a criterion to assess fits when models with different numbers of parameters are being compared. If model (p) is correct then *Cp* will tend to be close to or smaller than p. Therefore a simple plot of *Cp* versus p can be used to decide amongst models. This criterion is related to the mean-square error of a fitted value as follows:

$$Mallows\ C_p = \frac{RSS(P)}{S^2} - (n - 2p),$$

where n is the sample size, p is the number of covariates including β_0 , *RSS (p)* is the residual sum of squares from a model containing *p* parameters, and S^2 is the mean residual sum of squares from the model containing all possible covariates (Full model).

A model is good according to this criterion if *Cp* ≤ p. We may choose the smallest model for which *Cp* ≤ p, so a benefit of this criterion is that it can achieve for us a `good' model containing as few variables as possible.

RESULTS AND DISCUSSION

Means of Barley Yield and its Components

Basic statistical parameters: mean values, standard error, standard deviation, minimum and maximum values and coefficient of variation, for the six cultivars under investigation of all studied traits are presented in (Table 1). In the present investigation, there was a considerable variation with regard to all characteristics under study (Table 1).

The results shown in Table 1 show that the coefficient of variation was the highest for spike length, followed by grain weight per spike. Heading date had the lowest value, followed by number of spikes per square meter, number of grains per spike and plant height. Thousand grain weight and grain yield per feddan showed moderate values for the coefficient of variation (Table 1). Coefficient of variation also known as 'relative variability' calculated as percentage is a measure of how much variability exists for selection. Similar results have been reported by Zakova and Benkova (2006) and Sarkar *et al* (2010).

Table 1. Basic statistical parameters for yield and yield components in barley: mean values, standard deviation, standard error, minimum values (Min) maximum values (Max) and coefficient of variation (CV).

Character	Mean	SD	SE	Min.	Max.	CV (%)
Heading date (days)	76.64	5.03	1.45	70.00	84.33	6.56
Plant height (cm)	92.28	8.34	2.41	79.67	101.33	9.04
Number of Spikes /m^2	507.66	39.35	11.36	455.55	579.69	7.75
Number of grains/spike	55.45	5.01	1.45	46.36	60.44	9.03
1000-grain weight (g)	47.76	4.85	1.40	39.75	53.03	10.17
Grain weight /spike	2.39	0.29	0.09	1.97	2.08	12.52
Spike length (cm)	5.46	0.79	0.23	4.40	6.87	14.60
Grain yield (ton/fed)	1.71	0.19	0.05	1.44	1.98	11.01

Means of grain yield varied between 1.44 and 1.98 ton per feddan. Plant height ranged from 79.67 to 101.33 cm. Heading date was between 70 and 84.33 day, whereas the number of spikes per square meter was between 455.55 and 579.69. The number of grains per spike, thousand grain weight, grain weight per spike and spike length were between 46.36 and 60.44, 39.75 and 53.03 g, 1.97 and 2.08 g, 4.40 and 6.87 cm, respectively (Table 1).

Variance analysis

Results from the combined analysis of variance over two years revealed that grain yield and yield components had significant differences between years, cultivars and interaction of years x cultivars (Table 2).

Table 2. Combined analysis of variance for grain yield and yield components over two years.

S.O.V.	df	Mean squares							
		PH	DH	NS/m^2	Ng/S	GW/S	Sl	TSW	GY
Years (Y)	1	25.11*	13.36*	392.50*	60.03*	0.042	0.026	0.095	0.152*
Cultivars (C)	5	452.11**	164.56**	10127.44**	148.09**	0.578**	4.17**	152.50**	0.209**
C x Y	5	7.44*	9.03*	36.112*	15.62*	0.015	0.013	3.11	0.180*

PH: Plant height, DH: Heading date, NS/m^2: Number of spikes per square meter, NG/S: Number of grains per spike, GW/S: Grain weight per spike, TSW: Thousand grain weight, SL: Spike length, GY: Grain yield (ton/fed).
** = Significant at 1% level. * = Significant at 5% level. ns = Non-Significant.

The analysis of variance presented in Table (2) shows that, cultivars differed significantly at $p < 0.01$ for all characteristics, indicating significant genetic variability present for these traits among cultivars. The presence of variability in crop is important for genetic studies and consequently improvement and selection. Seasonal variations significantly affected all traits except grain weight per spike, spike length and 1000-grain weight, indicating that genotypes responded for this trait similarly to weather condition in both years (Table 2). Significant year (Y) effects ($p < 0.05$) indicated the presence of variability in the environmental variables (temperature, rainfall, humidity, sunshine) for both years of evaluation. The interaction between seasons and cultivars was also significant for all traits under study except grain weight per spike, spike length and 1000-grain weight (Table 2), indicating that differences between cultivars were affected by the growing season. Hamid *et al*

(2005), Zakova and Benkova (2006), Hasan *et al* (2010), Tamm and Kuuts (2010), Ibrahim *et al* (2011) and Zaefizadeh *et al* (2011a) recorded similar results in their studies with different barley genotypes.

Varietal performance

Grain yield is a complex character affected by several environmental, morphological and physiological characters. In the present study, yield and yield components were significantly affected by cultivars. Grain yields also depend upon other yield components. Results of the combined analysis of variance showed significant differences among the studied cultivars for yield and its components, (Table 3).

Comparison of the means of different traits indicated that the cultivars were superior in each trait. Data presented in Table 3 indicated that the barley cultivars: Giza 123, Giza 126, Giza 129, Giza 130, Giza 131 and Giza 2000 differed significantly in yield and yield components, via., plant height, days to heading, number of spikes per square meter, number of grains per spike, thousand grain weight and grain yield per spike.

Table 3. Variation among barley cultivars in grain yield and other agronomic characteristics.

Cultivar	Agronomic characteristics (mean values)							
	PH	HD	NS /m^2	NG/S	GW/S	TWS	SL	GY
Giza 123	86.50b	70.50c	574.84a	60.42a	2.79a	52.79a	5.73b	1.97a
Giza 126	100.50a	82.67a	459.30d	48.15c	1.99d	40.11c	4.43e	1.50d
Giza 129	88.00b	73.33bc	516.92bc	59.62a	2.71a	52.51a	6.82a	1.79a-c
Giza 130	100.00a	76.00b	498.02c	53.62b	2.26c	45.32b	5.20cd	1.63b-d
Giza 131	98.83a	83.33a	473.71d	51.93b	2.18c	45.21b	4.86de	1.53cd
Giza 2000	79.83c	74.00b	523.66b	58.97a	2.44b	50.63a	5.70bc	1.84ab

PH: Plant height, DH: Heading date, NS/m^2: Number of spikes per square meter, NG/S: Number of grains per spike, GW/S: Grain weight per spike, TSW: Thousand grain weight, SL: Spike length, GY: Grain yield (ton/fed). Means followed by the same letter within a column are not significantly different at the 5% level of probability based on the combined analysis over two years as indicated by the protected LSD test.

The mean grain yield of cultivars varied between 1.50 for Giza 126 to 1.97 for Giza 123. The cultivar Giza 123 produced the highest grain yield (1.97 ton/fed) which was at par with cultivars Giza 129 and Giza 2000. On the other hand the differences between Giza 126, Giza 130 and Giza 131 were not significant (Fig. 1). The higher grain yield of cultivar Giza 123 was due to the higher number of spikes per square meter, number of grains per spike, grain weight per spike and thousand grain weight. The differences between cultivars might be attributed to variation in translocation rate of photosynthates from leaves and storage organs to the grain. These results agree with those of Shaaban *et al* (1984), El-Sayed *et al* (1992), Ibrahim *et al* (2011) and Zaefizadeh *et al* (2011 b).

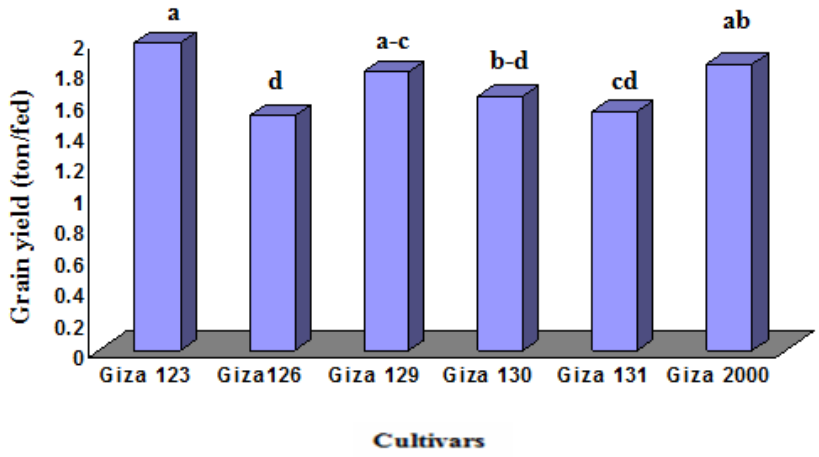

Fig 1. Comparison of mean grain yields (ton/fed) of the six barley cultivars.
Columns with the same letter are not statistically different at the p=0.05 according to LSD test.

As can be seen from Table 3, all cultivars produced statically different numbers of spikes per square meter. The cultivar Giza 123 produced significantly more spikes per square meter (574.84) than all other cultivars and was followed by Giza 2000 (523.66). Cultivars varied significantly regarding thousand grain weight

(Table 3). The cultivar Giza 123 was at par with Giza 129 and Giza 2000 which produced significantly the highest thousand grain weight (52.79 g).

When we look at number of grains per spike, Giza 123 (60.42) was highest, while Giza 126 (48.15) had the lowest values. The mean grain weight per spike varied among cultivars in the range of 2.79 for Giza 123 to 1.99 for Giza 126. The cultivar Giza 129 produced significantly highest spike length (6.82) and was followed by Giza 123, Giza 2000, Giza 130, Giza 131 and Giza 126 having average spike length of 5.73, 5.70, 5.20, 4.86 and 4.43, respectively.

It may be concluded that differences between barley cultivars may be due to genetical differences, as well as, cultivars response. It is noteworthy to mention that differences in yield potential of barley depend on the part of photosynthetic material partitioned to grains. It is worthy to mention that the results of varietal differences in yield and its components were in harmony with the results obtained by Hamid *et al* (2005), Zakova and Benkova (2006), Hasan *et al* (2010) and Tamm and Kuuts (2010).

Correlation analysis

Knowledge of the interrelationships between seed yield and other characters is important to effective selection (Ariyo 1995). Consistent with this, efforts were made to evaluate the nature of inter-relationships between different yield components.

The simple correlation coefficients were determined for eight character combinations with the objective to obtain information about the relationships among different character combinations. Knowledge of correlation is also required to obtain the expected response of other characters when selection is applied to a character of interest in a breeding program (Falconer 1989).

Coefficients of correlation grain yield and yield components from data obtained over years are presented in Table 4. The combined data over two years in Table 4 show that number of spikes per square meter, number of grains per spike, grain weight per spike, 1000-grain weight and spike length showed significant positive correlation with grain yield.

When we look at the relationship among traits, the results of the correlation coefficients revealed that the grain number per spike, grain weight/spike and 1000-grain weight had the highest significant positive correlation with grain yield, r= 0.84**, r= 0.87** and r=0.88** (Table 4), indicating dependency of yield on these characters. Furthermore, results also indicated that number of spikes per square meter and spike length was

Table 4. The correlation coefficients among the grain yield and yield components in barley.

Character	GY	PH	HD	NS/m^2	NG/S	GW/S	TSW
Plant height (cm)	-0.78**						
Heading date (days)	-0.90**	-0.73**					
Number of spikes /m^2	0.68*	-0.41^{ns}	-0.58*				
Number of grains/spike	0.84**	-0.82**	-0.87**	0.58*			
Grain weight /spike	0.87**	-0.72**	-0.89**	0.66*	0.94**		
1000-grain weight (g)	0.88**	-0.80**	-0.87**	0.64*	- 0.67*	0.93**	
Spike length (cm)	0.67 *	-0.69*	-0.75**	0.70**	0.81**	0.85**	-0.65*

PH: Plant height, DH: Heading date, NS/m^2: Number of spikes per square meter, NG/S: Number of grains per spike, GW/S: Grain weight per spike, TSW: Thousand grain weight, SL: Spike length, GY: Grain yield (ton/fed).
** = Significant at 1% level. * = Significant at 5% level. ns = Non-Significant

Significantly correlated with barley grain yield (r= 0.68*, r= -0.67*, respectively).These correlations show that selection for these traits plays an important role in improving grain yield in barley. These results are in agreement with those reported by Rasmusson and Chanel (1970), Singh (1999), Kole (2006) and Zaefizadeh *et al* (2011 b).

The correlation of spike length with number of grains per spike was highly significant and positive, suggesting that the longer the spike length the higher would be the number of grains per spike. Results of table 4 show that correlation of number of grains per spike with 1000-grain weight was significant and negative (r= -0.67*). We can conclude that by increasing number of grains per spike, 1000-grain weight decreases and vice versa. Field data showed that increasing the spike length increased the number of grains per spike and decreased the 1000 grain weight. 1000-grain weight is an important yield component (Petr *et al* 1979).

Simple correlation coefficients between grain yield and other variables are given in Table 4. There was significant negative relation between plant height and grain yield (r= - 0.78 **) and days to heading with grain yield (r= - 0.90**). This means that decreasing days to heading decreases grain yield, and increasing plant height decreases grain yield. This may be attributed to that long stems could result in plants lodging, which reduces grain yield (Gardner *et al* 1985). Increases in grain yield of spring barley are generally associated with reduced plant height and improved lodging resistance (Grausgruber *et al* 2002). Plant height is an important

morphological character directly linked with the productive potential of plant in terms of grain yield (Alam *et al* 2007). Zaefizadeh *et al* (2011 b) also detected negative correlation between plant height and grain yield.

In our study, heading date was negatively and highly correlated with plant height, number of spikes per square meter, number of grains per spike, grain weight per spike, thousand grain weight and spike length with values of (r= -0.73**, r= -0.58*, r= -0.87**, r= -0.89**, r= -0.87**, r= -0.75**, respectively). Also a negative correlation between plant height and 1000-gran weight is found. Zaefizadeh *et al* (2011 b) reported negative correlation between days to heading and plant height and between plant height and 1000-grain weight.

Regression analysis

Here we shall among other things distinguish between simple linear regression and multivariate regression model; furthermore discussed the assessing of model.

A Simple Linear Regression Model

Simple regression coefficients (b) of yield on the different characters were computed, together with their S_b values (the sample standard deviation of the regression coefficients). The significance of the coefficients obtained was tested by calculating t values (i.e. by the t test) as shown in Table 5. All the b values were positive, except that for plant height and days to heading, and all were highly significant except the b values for number of spikes per square meter and spike length were significant over the two seasons. Grain weight per spike gave the highest regression coefficient. Similar results had been reported by other authors (Zajac *et al* 1999, Yusaf *et al* 2003 and Zaefizadeh *et al* 2011 a)

Table 5. Regression coefficients (b values) of different component traits on grain yield in barley along with their standard errors, t values and linear regression equations.

Character (x_i)	Regression values (b values)	R^2-adjusted	Standard error of coefficients	T value	P -value	Linear regression equation $\hat{Y}$ = Grain yield ton/fed.
Plant height (cm)	-0.02**	0.5786	0.004	-4.01	0.002	$\hat{Y} = 3.34 - 0.02\ x_1$
Heading date (days)	-0.03**	0.8000	0.005	-6.71	0.000	$\hat{Y} = 4.30 - 0.03\ x_2$
Number of Spikes /m^2	0.001*	0.4122	0.0004	2.95	0.014	$\hat{Y} = 1.01 + 0.001\ x_3$
Number of grains/spike	0.03**	0.6881	0.006	5.03	0.001	$\hat{Y} = -0.05 + 0.02\ x_4$
Grain weight /spike	0.55**	0.7424	0.096	5.72	0.000	$\hat{Y} = 0.39 + 0.55\ x_5$
1000-grain weight (g)	0.03**	0.7576	0.005	5.95	0.000	$\hat{Y} = 0.07 + 0.03\ x_6$
Spike length (cm)	0.15*	0.3979	0.055	2.88	0.016	$\hat{Y} = 0.84 + 0.15\ x_7$

** = Significant at 1% level. * = Significant at 5% level.

From the simple regression (Table 5) it was found that the regression mean squares were highly significant, indicating presence of variation in the studied material and the importance of these characters to yield. Regression coefficient (b values) for number of spikes per square meter, number of grains per spike, grain yield per spike, thousand grain yield and spike length were positively significantly correlated with grain yield indicating that increase in these characters would increase the grain yield. Other traits including plant height and heading date showed significant and negative 'b' values suggesting that grain yield would be decreased with the increase of both characters.

Linear regression of number of spikes per square meter, number of grains per spike, grain yield per spike, thousand grain yield and spike length it leads to increase the yield by 0.001, 0.03, 0.55, 0.03 and 0.15 units, respectively. Presence of highly significant and positive correlation between number of spikes per square meter, number of grains per spike, grain yield per spike, thousand grain yield and spike length with grain yield shows that the results of regression analysis are in harmony with correlation results, while, plant height and heading date reduce the yield by 0.02 and 0.03 units, respectively. Coefficients of regression suggest that an increase of one centimeter in plant height may reduce grain yield by 0.02 ton/fed and decrease of one day in days to heading may reduce the grain yield by 0.03 ton/fed.

Linear relationships were observed in both thousand grain weight and days to heading and grain yield (Figs. 2 and 3). These two factors had the highest R^2 in the study. Both thousand grain weight and days to heading affected grain yield and played an important role in determining grain yield in barley.

Grain yield showed a positive and linear relationship with thousand grain weight (Fig. 2). The higher value of adjusted R^2 (adj. R^2= 0.75) suggests that 75% of the variation in grain yield could be explained by the variation in thousand grain weight. So this trait is the most important component of grain yield in barley.

The regression equation shows how one trait, dependent variable, changes if an independent variable changes for one unit. In that way, regression equation between grain yield (ton/fed) and thousand grain weight was established and was significant at the 1% level with a linear equation form:

$\hat{Y} = 0.07 + 0.03\ x_6$, and it shows that by increasing one unit in thousand grain weight, grain yield will increase on the average by 0.03 ton/fed (Table 5 and Figure 2).

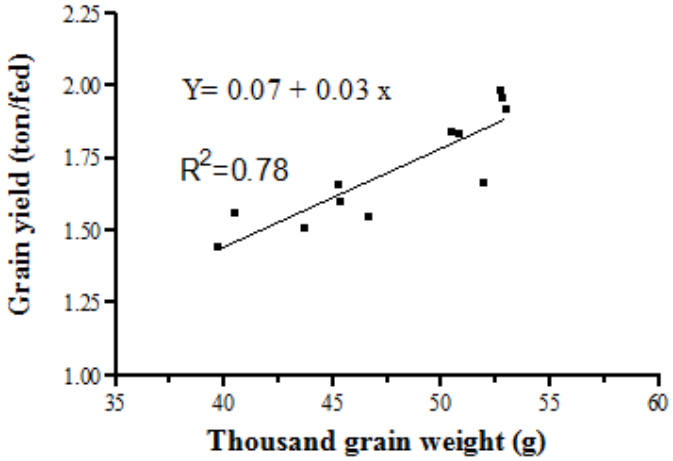

Fig 2. Functional relationship between thousand and grain yield (ton/fed).

Grain yield (ton/fed) when plotted against heading date yielded a straight line. This suggests that grain yield (ton/fed) is dependent on heading date and 80% (adj-R^2= 0.80) of the variation in grain yield could be explained by variation in heading date (Fig. 3)

Linear regression equation was established between days to heading and grain yield and was of significant regression coefficient on the level of 5% and 1 %, having the following form:

$\hat{Y} = 4.30 - 0.03\ x_2$,

which shows that a one-day increase in the number of days to heading, grain yield decreases by an average of 0.03 ton/fed (Table 5 and Figure 3).

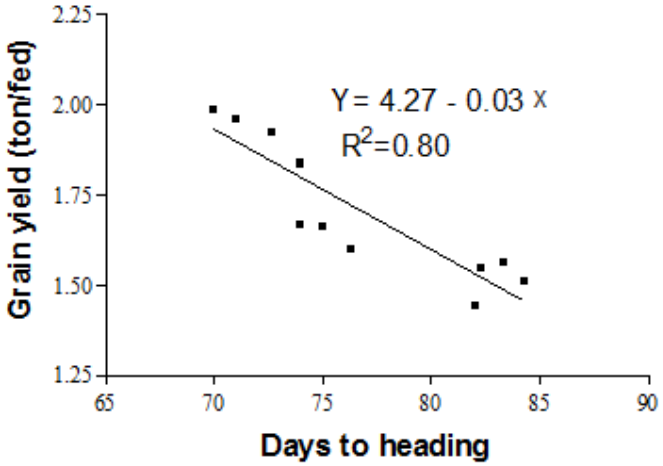

Fig 3. Functional relationship between days to heading and grain yield (ton/fed).

Multiple regression results

The regression of grain yield on other characters (plant height, heading date, number of spikes/m^2, number of grains/spike, grain weight /spike, 1000-grain weight and spike length) is presented in Table 6. The variation in response explained by all the predictors was 93.32% and the F-test was highly significant. These results, are in agreement with those reported by Yusaf *et al* (2003) and Zaefizadeh *et al* (2011 a).

The final equation of grain yield based on seven variables shown in Table 6 is:

$Y = 3.537 – 0.008\ x_1 – 0.017\ x_2 + 0.0006\ x_3 – 0.012\ x_4 + 0.275\ x_5 + 0.015\ x_6 – 0.149\ x_7$. In this equation Y is the grain yield; x_1, x_2, x_3 , x_4, x_5, x_6 and x_7 are plant height, heading date, number of Spikes/m^2, number of grains/spike, grain weight/spike, 1000-grain weight and spike length, respectively.

Table 6. Analysis of multiple linear regression of grain yield on other agronomic traits of barley, over both 2008/09 and 2009/10 seasons.

S.O.V.	DF	Sum of squares	Mean squares	F-ratio	P-value
Regression	7	0.38049	0.05436	22.95**	0.0045
Residual	4	0.00947	0.00237		
Total	11	0.38996			

**= Highly significant at 1% level of significance.

Multiple correlation coefficient R = 0.987, Coefficient of determination R^2=0.975, Adjusted R^2= 0.933

All possible regressions and "best subset" regression models

The Regression Model Selection procedure is designed to help select the independent variables to use in building a multiple regression model to predict a single quantitative dependent variable (Y). All possible subsets, this method builds all one variable models, all two-variable models, and so on, until the all variable model is generated. The method requires a powerful computer (because a lot of models are produced), and selection of any one of the criteria: R-squared, adjusted R-squared, Mallow's C*p*. The most commonly used criterion to help in choosing between alternative equations in multiple regression is the R^2 (adjusted or unadjusted), the F-ratio based on R^2, along with the statistical significance of the F-ratio (Schumacker 1994).

Best subset regression identifies the best fitting regression models that can be constructed with the independent variables screened. Best subset regression is an efficient way to identify models that achieve in predicting grain yield with as few independent variables as possible. Subset models estimate the actual regression coefficients and predict the future responses with smaller variance when comparing the full model using all independent variables. MINITAB examines all possible subsets of the independent variables, beginning with all models containing one independent variable, followed by all models containing two independent variables, and so on. By default, MINITAB displays the two best models for each number of independent variables as shown in Table 7. The table explains that each line of the output represents a different regression model. Independent variables that are present in the model are indicated by X. From the table, we explained several methods for determining the so-called "best" model according to various criteria. Table (7) shows these statistics for the 31 models fit (n = 12 for all models).

When comparing models with the same number of parameters, the model with the highest R^2 value should typically be selected. A larger R^2 value indicates that more of the variation in the response variable is explained by the model. However, R^2 never decreases when another predictor is added (adding variables to the model can never decrease the amount of variation explained). Thus other techniques are suggested when comparing models with different numbers of explanatory variables (Broersen 1986).

A researcher must balance the idea of increasing R^2 versus keeping the model simple. There are many additional statistics that have been developed in addition to R^2 to determine the "best" model, such as adjusted R^2 and Mallow's *Cp* criteria. Each of these statistics includes a penalty for including too many terms. While any one of these statistics is appropriate for some models, adjusted R^2 still tends to select models with too many terms. Best subsets techniques use several statistics to simultaneously compare several regression models with the same number of predictors (Miller 1984).

Keeping in view the correlation of different variables with grain yield, best subsets regression was done using the three main criteria for model fitting, viz., coefficient of determination (R^2) achieved by least square fit, adjusted-R^2, and Mallow's *Cp*-statistics (Table 7) (Draper and Smith, 1998). Based upon the selected criterion, the objective is to find the subset of independent variables that yields the lowest significance level among all possible subsets.

Mallows proposed the statistic as a criterion for selecting among many alternative subset regressions (Mallows 1973). Mallow's *Cp* compares the precision and bias of the full model to models with the best subsets of independent variables. It helps to strike an important balance with the number of

independent variables in the model. A model with too many independent variables can be relatively imprecise while one with too few can produce biased estimates (Wikipedia, the Free Encyclopedia, Mallow's *Cp* Available). A Mallow's *Cp* value that is close to the number of independent variables plus the constant indicates that the model is relatively precise and unbiased in estimating the true regression coefficients and predicting future responses.

Table 7. Models consisting of different combinations of variables selected using all-subsets regressions from 7 potentially important predictors of barley.

Vars	p	Selection Procedure			Variables						
		R^2	R^2-adjusted	Mallows Cp	PH	HD	NS/m^2	NG/S	GW/S	TSW	SL
1	2	81.8	80.0	21.9		X_2					
1	2	78.0	75.8	28.3						X_6	
1	2	76.6	74.2	30.6					X_5		
1	2	71.6	68.8	38.7				X_4			
1	2	61.7	57.9	55.1	X_1						
2	3	85.5	82.3	17.9		X_2				X_6	
2	3	85.3	82.1	18.1		X_2	X_3				
2	3	85.1	81.8	18.5	X_1	X_2					
2	3	84.1	80.6	20.1		X_2			X_5		
2	3	83.1	79.3	21.9		X_2		X_4			
3	4	89.3	85.3	13.6		X_2				X_6	X_7
3	4	88.8	84.6	14.5	X_1	X_2	X_3				
3	4	87.5	82.8	16.6			X_3			X_6	X_7
3	4	87.2	82.3	17.1		X_2	X_3			X_6	
3	4	87.2	82.3	17.2		X_2	X_3				X_7
4	5	94.3	91.0	7.5		X_2	X_3			X_6	X_7
4	5	93.5	89.8	8.7	X_1	X_2	X_3				X_7
4	5	91.5	86.6	12.0	X_1		X_3		X_5		X_7
4	5	90.7	85.3	13.4	X_1		X_3			X_6	X_7
4	5	90.2	84.5	14.2		X_2	X_3		X_5		X_7
5	6	96.5	93.6	5.8	X_1	X_2	X_3			X_6	X_7
5	6	95.7	92.1	7.1	X_1	X_2	X_3		X_5		X_7
5	6	94.4	89.7	9.3		X_2	X_3		X_5	X_6	X_7
5	6	94.3	89.6	9.3		X_2	X_3	X_4		X_6	X_7
5	6	93.7	88.4	10.4	X_1	X_2	X_3	X_4			X_7
6	7	96.6	93.1	7.1	X_1	X_2	X_3		X_5	X_6	X_7
6	7	96.6	92.4	7.7	X_1	X_2	X_3	X_4		X_6	X_7
6	7	96.5	92.2	7.8	X_1	X_2	X_3	X_4	X_5		X_7
6	7	94.4	87.6	11.3		X_2	X_3	X_4	X_5	X_6	X_7
6	7	93.7	86.2	12.3	X_1		X_3	X_4	X_5	X_6	X_7
7	8	97.6	93.3	8.0	X_1	X_2	X_3	X_4	X_5	X_6	X_7

PH: Plant height, DH: Heading date, NS/m^2: Number of spikes per square meter, NG/S: Number of grains per spike, GW/S: Grain weight per spike, TSW: Thousand grain weight, SL: Spike length.

The highlighted models have acceptable Mallow's *Cp*. So we can choose the model with good *Cp* and the smallest number of variables to get best adjusted R^2, which have the largest value of the model (93.6%). The best model includes plant height, heading date, number of spikes/m^2, thousand grain weight and spike length. Both numbers of grains per spike and grain weight per spike are not included.

The simplified results from best subset regression analysis show that the highest adjusted R^2 (93.6%) and lowest Mallow's *Cp* value (5.8), and the lowest S value (0.047) has five-independent variable model with all variables except number of grains per spike (X_4) and grain weight per spike (X_5). The multiple regression

indicates that adding the number of grain per spike (X_4) and grain weight per spike (X_5) does not improve the fit of the model.

Taking a closer look at the results of the different criteria, we observe that the model preferred by the adjusted R-squared criteria (that is, the model containing X_1, X_2, X_3, X_6 and X_7) has the first smallest value of Mallow's *Cp* (*Cp*=5.8). That is, although this model is the `best', according to the *Cp* criterion, it is still a `good' model. Also, the adjusted R-squared criterion is fairly large for this model: adjusted R^2= 93.6%, which is almost the same as the adjusted R^2 for the maximum model (93.3%). These findings are consistent with the results of Nikolopoulos *et al* (2007) and Abdullahi *et al* (2010).

A useful diagnostic plot for determining the best model according to the *Cp*-statistic is to plot the *Cp* values versus *p* for all models fit, and look to see where the *Cp* value crosses the *Cp* = *p* line. Mallow's *Cp* values of subset regressions for each P are given in the column 5 of the Table 7. A *Cp* plot is presented in Figure 4. It is clear from the figure that the regression model with *Cp* = 5.8 corresponding to p = 6 is preferred because of having fewer number of regressors and lies just below the *Cp* = P line showing small amount of bias as compared to the models of which the *Cp* values lie above the line.

By default, the table 7 shows the best model for each number of independent variables. For example, the best model involving only 5 independent variables includes variables X_1, X_2, X_3 X_6 and X_7, and gives an adjusted R-squared of 93.6%. The model with the best adjusted R-squared includes 5 variables, X_1, X_2, X_3 X_6 and X_7 . The model with the smallest *Cp* is X_1, X_2, X_3, X_6 and X_7. Since its *Cp* value is less than 6, that model appears to be the best.

Considering multiple correlation coefficient (R =0.982) of the five components under study, the determination of adjusted-R^2 shows that 93.6 percent of variation in grain yield was due to these five factors (Table 8). The coefficient of adjusted R^2 (0.936) represents the influence of the traits involved in the study on total variability of grain yield. The remaining 0.064%

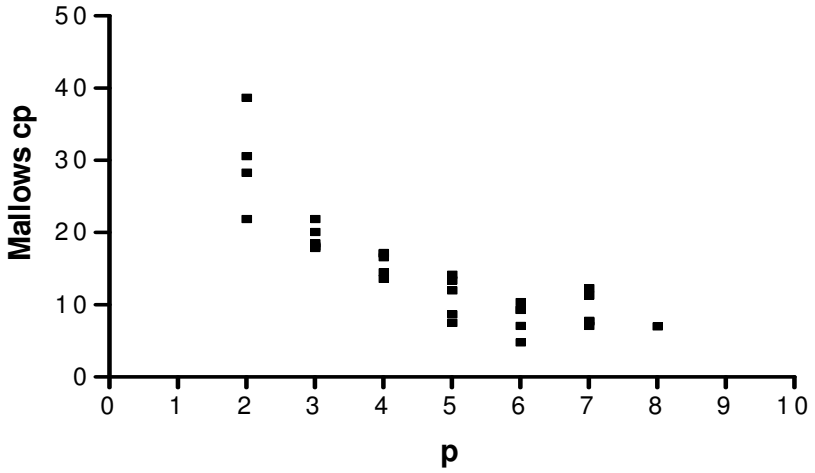

Fig 4. The plot of *Cp* against *p* obtained for all the subset regressions

Table 8. Analysis of the best subset multiple linear regression analysis of grain yield and other agronomic traits of barley, over both 2008/09 and 2009/10 seasons.

S.O.V.	df	Sum of squares	Mean squares	F-ratio	P-value
Regression	5	0.376337	0.075267	33.14 **	0.000
Residual	7	0.013627	0.002271		
Total	11	0.389964			

**= Highly significant at 1% level of significance.

Multiple correlation coefficient R=0.982, Coefficient of determination R^2= 0.965, Adjusted R^2= 0.936

Could be attributed to factors that were not included in this study. It would therefore, be advisable to lay emphasis on these traits while, selecting high yielding cultivars in barley. The findings were consistent with the results of Zaefizadeh *et al* (2011a). Mohammadi (2002) also reported the number of fertile tillers and the number of grains per spike as the important traits in showing the yield in the barley genotypes, although these two traits

and traits of grain filling time, days to spike appearance and plant height totally were showing the yield rate of 49% in the regression model.

After confirmation of the results based on the three criteria, grain yield (Y) model was developed as follows:

$Y = 3.12 – 0.006\ x_1 – 0.019\ x_2 + 0.0007\ x_3 + 0.020\ x_6 – 0.149\ x_7$. In this model Y is the grain yield; x_1, x_2, x_3, x_4, x_5, x_6 and x_7 are plant height, heading date, number of spikes/m^2, 1000-grain weight and spike length, respectively.

From the previous model, it is deduced that for every unit increase in plant height there is a decrease of 0.006 ton/fed decreases in yield, and a decrease of about 0.019 ton/fed was observed in the yield when the days to heading is increased by one unit. Similarly a decrease of about 0.149 ton/fed was noted for every unit increase in spike length. There is an increase of about 0.0007 in the yield when the number of spikes per square meter is increase by one unit and an increased of about 0.020 ton/fed was observed in yield when the one-thousand grain weight is increased by one unit.

CONCLUSION

From this investigation it is concluded that significant genotypic variation is present among barley cultivars to be subjected to selection for grain yield and its attributes.

The data obtained from this study could be useful for barley breeders and grain producers in order to increase grain yield. Therefore, the characteristics of plant height, time to heading, number of spikes per square meter, grain weight per spike and spike length can be used as selection criteria to increase grain yield in barley.

Correlation analysis mirrored the strength of the relationship between barley yield and its attributes, where they were positively and highly correlated with the grain yield except for plant height and days to heading, where it were negatively and highly correlated with grain yield barley. Moreover, the developed prediction equations are accurate because they have high R^2 and low SE%.

Regression analysis indicated the importance of plant height, days to heading, number of spikes per square meter, thousand grain weight and spike length in influencing grain yield in barley. High value of the adjusted coefficient of determination ($R^2 = 0.936$) indicates that the traits chosen for this study explained almost all grain yield variation.

So for increasing grain yield, a barley cultivar should have more number of spikes per square meter, more number of grains per spike, high grain weight per spike, high thousand grain weight and large spike length, because these characters are positively associated with grain yield. However, in case of selection for yield improvement the cultivar should have short plant height with less time required for days to heading.

REFERENCES

Abdullahi FB, Usman A, Cole AT. 2010. Constructing the best regression model for maiwa variety. Pakistan J. of Nutrition 9 (4): 380-386.

Agarwal BL.1991. Basic Statistics. Second edition, Wiley Eastern Limited, New Delhi.

Alam MZ, Haider SA, Paul NK. 2007. Yield and yield components of barley (*Hordeum vulgare* L.) in relation to sowing times. J. Bio. Sci.15: 139-145.

Ariyo, O. J. (1995). Correlations and path-coefficient analysis of components of seed yield in soybean. African Crop Sci. J., 3 (1): 29-33.

Broersen PMT. 1986. Subset regression with stepwise directed search. Appl. Statist. 35(2), 168-177.

Draper NR, Smith H.1998. Applied Regression Analysis (Third Edition). New York: Wiley.

El-Sayed AA, Noman MM, El–Rayes AM.1992. Evaluation of newly released barley cultivars to nitrogen fertilizer in sandy soils. Nile Valley Regional Program Barley Annual Report (1991/1992).

Falconer DS. 1989. Introduction to Quantitative Genetics, 2nd edition. Longman New York, USA.

Gardner FP, Pearce RB, Mitchell RL.1985. Physiology of Crop Plants. Iowa State University Press. Ames. USA.

Grausgruber H, Bointner H, Tumpold R, Ruckenbauer P.2002. Genetic improvement of agronomic and qualitative traits of spring barley. Plant Breeding 121: 411–416.

Hamid RB, Zeinalabedin TS, Seyed AM. 2005. Agronomic factors on selected hulless barley genotypes. J. of Agronomy 4 (4): 333-339.

Hasan K, Akar T, Kendal E, Sayım S.2010. Evaluation of grain yield and quality of barley varieties under rainfed conditions. African J. of Biotech. 9(6):7825-7830.

Ibrahim OM, Magda HM, Tawfik MM, Badr EA. 2011. Genetic diversity assessment of barley (*Hordeum vulgare* L.) genotypes using cluster analysis. Inter. J. of Acad. Res. 3(2): 81-85.

Kole PC. 2006. Variability, correlation and regression analysis in third somaclonal generation of barley. Barley Genetics Newsletter 36:44-47

Mallows CL. 1973. Some comments on *Cp*. Technometrics 15(4): 661–675.

Mallows CL. 1995. More comments on *Cp*. Technometrics 37(4): 362–372.

Mallows CL. 1997. *Cp* and prediction with many regressors: comments on Mallows. Technometrics 39(1):115–116.

Miller AJ. 1984. Selection of subsets of regression variables. J. Roy. Statist. Soc. Ser. A. 147: 389-425.

MINITAB .2005. MINITAB Reference Manual, Release 14 for Windows. PA: Minitab Inc. State College, Harrisburg, Pennsylvania, USA.

Mohammadi M. 2002. Physiological traits associated with the performance of two barley genotypes in normal conditions and drought stress. Seed and Plant Research Magazine, 17: 61-72.

Montgomery DC, Peck EA, Vining GG.2001. Introduction to Linear Regression Analysis, 3rd Ed., Wiley, New York, NY.
MSTAT-C program.1991. A software program for the design management and analysis of agronomic research experiments. Michigan State University.
Nikolopoulos K, Goodwin P, Patelis A, Assimatopoulos V. 2007. Forecasting with cue information: A comparison of multiple regressions with alternative forecasting approaches. European J. Operational Res. 180: 354-368.
Petr J, Hnilica P, Schmidt J.1979. Yield formation in spring barley: tillering, ear formation and grains per ear. Rostlinna Vyroba. 25(4): 433-444.
Rasmusson DC, Chanel RQ.1970. Selection for grain yield and components of yield in barley. Crop Sci. 10: 51-54.
Sarkar B, Verma RPS, Parsad R, Shoran J.2010. Diversity among barley germplasm collection in India. Indian J. Genet. 70(3): 234-239.
Schumacker RE.1994. A comparison of the Mallows Cp and principal component regression criteria for best model selection in multiple regression. Multiple Regression Viewpoints 21(1), 12-22.
Shaaban SA, El- Haroun MS, El-Tawail AYM.1984. Growth and yield response of two barley cultivars to irrigation frequency and nitrogen fertilizer. Annals . Agric. Sci., Fac. Of Agric. Ain Shams Univ., Cairo, Egypt 28(3): 1387-1413.
Singh BP.1999. Correlation study in barley (*Hordeum vulgare* L.). J. Appl. Biol. 9:143-145.
Steel R, Torrie J, Dicky D. 1997. Principles and Procedures of Statistics; A Biometrical Approach. 3rd Ed. W.C.B/McGraw-Hill, New York.
Tamm U, Kuuts H. 2010. About new varieties meeting the malting barley requirements. From: http://www.eau.ee/~aps/pdf/20044/tamm.pdf.
Wikipedia, the Free Encyclopedia, Mallow's CP Available: http://en.wikipedia.org/wiki/Mallows'_Cp.
Yusaf H, Asim SM, Zaman Q, Khan N.2003. All possible regression study of wheat crop. Pakistan j. of App. Sci. 3 (4): 236-239.
Zaefizadeh M, Khayatnezhad M, Gholamin R.2011a. Comparison of multiple linear regressions (MLR) and artificial neural network (ANN) in predicting the yield using its components in the hulless barley. American-Eurasian J. Agric. & Environ. Sci. 10 (1): 60-64.
Zaefizadeh M, Ghasemi M, Azimi J, Khayatnezhad M, Ahadzadeh B.2011b. Correlation analysis and path analysis for yield and its components in hulless barley. Advance in Environmental Biology 5(1): 123-126.
Zajac T, Gierdziewicz M, Oleksy A, Bieniek J. 1999. Estimation of yield component share in yield of spring barley and field bean depending on years of cultivation. Acta Agraria et Silvestria / Agraria 37:27-37.
Zakova M, Benkova M.2006. Characterization of spring barley accessions based on multivariate analysis. Communications in Biometry and Crop Sci. 1(2): 124–134.