The Journal of Animal & Plant Sciences, 21(2): 2011, Page: 182-185
ISSN: 1018-7081

USING OF FACTOR ANALYSIS SCORES IN MULTIPLE LINEAR REGRESSION MODEL FOR PREDICTION OF KERNEL WEIGHT IN ANKARA WALNUTS

E. Sakar, S. Keskin[*] and H. Unver[**]

Deptt. of Horticulture, Faculty of Agriculture Sanluirfa, Turkey, [*]Deptt. of Biostatistics, Faculty of Medicine Yuzuncu Yil University, Van, Turkey, [**] Kalecik Vocational School, Ankara, Turkey
Corresponding author e-mail: ebru.sakar09@gmail.com

ABSTRACT

Kernel weight is important for plant breeders to select high productive plants. The determination of relationships between kernel weight and some fruit-kernel characteristics may provide necessary information for plant breeders in selection programs. In the present study, the relationships between kernel weight (KW) and 7 fruit-kernel characteristics: Fruit Length, (FL,), Fruit Width (FW) Fruit Height (FH) Fruit Weight (FWe) Shell Thickness (ST), Kernel Ratio (KR) and Filled-firm Kernel Raito (FKR,), were examined by the combination of factor and multiple linear regression analyses. Firstly, factor analysis was used to reduce large number of explanatory variables, to remove multicolinearty problems and to simplify the complex relationships among fruit-kernel characteristics. Then, 3 factors having Eigen values greater than 1 were selected as independent or explanatory variables and 3 factor scores coefficients were used for multiple linear regression analysis. As a result, it was found that three factors formed by original variables had significant effects on kernel weight and these factors together have accounted for 85.9 % of variation in kernel weight.

Key words: Walnut, **c**ommunality, eigenvalues, varimax rotation, determination coefficient

INTRODUCTION

The determination of the relationships between kernel weight and fruit characteristics plays an important role in plant breeding research. Multiple regression and factor analyses have been used to interpret the multivariate relationships between kernel weight and fruit-kernel characteristics. This statistical tool is useful for predicting assumed dependent variable. Factor analysis is applied to a single set of variables to discover which variables are relatively independent of one another. It reduces many variables to a few factors. It also produces several linear combinations of observed variables which are called as factors. The factors summarize the pattern of correlations in the observed data. Because there are normally fewer factors than observed variables and because factor scores are nearly uncorrelated, use of factor scores in other analyses may be very helpful (Tabachnick and Fidell 2001).

The main objective of the present study was; using 3 a multivariate statistical approach, factor analysis, to classify predictor variables according to interrelationships and to predict kernel weight in Ankara walnuts. For this purpose, factor analysis scores of 7 fruit-kernel characteristics were used as independent variables in multiple linear regression model for prediction of kernel weight.

MATERIALS AND METHODS

Data on characteristics: Fruit Length, (FL), Fruit Width (FW) Fruit Height (FH) Fruit Weight (FWe) Shell Thickness (ST), Kernel Weight (KW), Kernel Ratio (KR) and Filled-firm Kernel Raito (FKR), were collected from 365 Ankara walnut samples.

Statistical analysis: Kolmogorov-Simirnov normality test was applied for all variables. After normality test, it was determined that all variables were normally distributed. Factor analysis was performed on 7 fruit-kernel characteristics to rank their relative significance and to describe their interrelation patterns to kernel weight.

Factor analysis: For the factor analysis, basic factor analysis equation can be represented in matrix form as:

$\mathbf{Z_{px1}} = \boldsymbol{\lambda}_{\mathbf{pxm}}\mathbf{F}_{mx1} + \boldsymbol{\varepsilon}_{\mathbf{px1}}$: where **Z** is a px1 vector of variables, $\boldsymbol{\lambda}$ is a pxm matrix of factor loadings **F** is a mx1 vector of factors and $\boldsymbol{\varepsilon}$ is a px1 vector of error (unique or specific) factors (Sharma, 1996). It will be assumed that factors were not correlated with the error components. Because of differences in were the units of each variables used in factor analysis, variables were standardized and correlation matrix of variables was used to obtain Eigen values. Loadings were correlation coefficients between variables and factors. Varimax rotation was used to facilitate interpretation of factor loadings (L_{ik}). Coefficients (C_{ik}), were used to obtain factor scores for selected factors. Factors with Eigen values greater than 1

out of 8 factors were employed in multiple regression analysis (Sharma, 1996; Tabachnick and Fidell, 2001; Johnson and Wichern, 2002).

Multiple linear regression analysis: Score values of selected factors were considered as independent variables for predicting kernel weight.

The regression equation is presented as;

$KW = a + b_1FS_1 + b_2FS_2 + b_3FS_3 + e$

where 'a' is regression constant (its value is zero), 'b_1', 'b_2' and b_3 are regression coefficients of Factor Scores (FS). FS is factor scores and e is the error term of the regression model. Regression coefficients were tested by using t test. Determination coefficient (R^2) was used as predictive success criteria for regression model (Draper and Smith 1998). All data were analyzed using MINITAB (ver:14) statistical package (Anonymous, 2000).

RESULTS AND DISCUSSION

Descriptive statistics and Pearson correlation coefficients for all characteristics are presented in Table 1 and Table 2, respectively. Since most of the correlation coefficients among variables were significant ($P \leq 0.01$ or $P \leq 0.05$). Correlation coefficients may be factorable.

Factor analysis revealed that 3 of the 7 factors have Eigen values greater than one and were selected as independent variables for multiple regression model (Tabachnick and Fidell, 2001; Johnson and Wichern, 2002).

Three selected factors explained 84.84 % (5.939/7) of the total variation of variables in factor analysis. Furthermore, communality values for variables were high. For example, communality for FWe was 94.0%, indicating that 94.0% of the variance in FWe is accounted for by Factor 1, 2 and 3. The proportion of variance in the set of variables accounted for by a factor is the sum of square loading for the factor (SSL) (variance of factor) divided by the number of variables (if rotation is orthogonal).

The proportion of variance is (3.187/7) = 0.4552 i.e for the first factor.45.52 % of the variance in the variables is accounted for by the first factor. The second and third factor accounts for 23.63 and 15.69 % of total variation, respectively. Because rotation is orthogonal, the three factors together accounted for 85.84 % of the variances in variables. The proportion of variance in the solution accounted for by a factor, in other words the proportion of covariance, is the Sum of Loadings for the factor divided by the sum of communalities.

For the selected three Factors, factor loading and factor score coefficients are presented in Table 3. After orthogonal rotation, the values of loading are correlations between variables and corresponding factors. The bold marked loads indicate the highest correlations between variables and corresponding factors. The greater loading, the more the variables is pure measure of factor. For instance, FL, FW, FH and FWe which showed the highest correlation with Factor 1 were considered as a group. Similarly, factor 3 showed highest correlation with only FKR.

The highest values of communalities indicated that the variances of variables were efficiently reflected in multiple regression analysis. All the 7 variables were included in the three selected factors. But only some of variables possessed high loads within each factor. FL, FW, FH and FWe possessed the highest loads in Factor 1, KR and ST in Factor 2, while FKR in Factor 3. Factors were interpreted from the variables that werehighly correlated with them. Thus, the first factor primarily has Fruit measurements; the second factor has primarily Kernel and Shell measurements while the third factor had only FKR. For factor 1, samples which scored high fruit measurements tended to assign high value to FL, FW, FH and FWe.

Factor score coefficients in Table 3 were used to obtain Factor score values. Factor score values for selected three factors were used as independent variables in multiple linear regression analysis to determine significant factors for kernel weight. All of the selected factors (Factor 1, 2 and 3) were found to have significant linear relationships with kernel weight ($p \leq 0.01$). The 85.9 % of variance in kernel weight was explained by Factor 1, 2 and 3 (Table-4).

As seen from Table 4, all three factors had positive and significant effect on kernel weight. Thus, kernel weight would to increase when the values of Factors scores increase. Increase in significant variables in Factor 1, namely, FL, FW, FH and FWe increase in kernel weight. Similarly, increasing in Factor results in 3 increases kernel weight. On the other hand, an increase in significant variables in factor 2 scores, indirectly increases of KR and decreases ST would bring an increase in kernel weight.

There is no study included the same variable and statistical model as in our study. Therefore the results were discussed with indirectly related studies. Sen (1983) pointed out that correlation coefficients between FW and KR as well as between FW and KR were very high in walnut. In addition, Sen (1985) stated that there is high significant correlation between KW and FL. Akça and Sen (1992) underlined that there were statistically significant correlations between KW and FL, FW, KR.

Firouz and Bayazid (2003) indicated that the correlation between average small diameter and average seed weight, kernel weight and percentage, shell weight was positive and significant. The correlation between average seed weight, and average kernel weight and percentage, shell weight and thickness was also positive and significant. There were significant and positive correlations between average kernel weight and average

shell weight and kernel percentage. The correlation between average shell weight and average shell thickness was significant and positive, whereas its correlation with average kernel percentage was significant and negative.

The results of this study are largely in agreement with those of previous studies. Yang *et al*. (2001) determined nut width, average nut weight, nut shell thickness, nut kernel percent age, per kernel weight, total fat content of kernel, total protein content and kernel yield per m^2 tree-crown projection area by principal component analysis. According to more than 86.29% of the cumulative variance proportion, the results proposed four principal components and its function equations which reflected the main economic characters of walnut. In the present study three factors together have accounted for 84.84 % of variation in kernel weight.

Table 1. Descriptive statistics for fruit-kernel characteristics

	Mean	SE	Mini.	Maxi.
Fruit Length (FL. mm)	35.679	0.226	26.52	49.85
Fruit width (FW. mm)	29.549	0.151	22.39	39.29
Fruit height (FH. mm)	30.975	0.173	23.17	44.40
Fruit weight (FWe. g)	10.032	0.136	4.3	20.2
Kernel weight (KW. g)	4.630	0.063	1.9	8.6
Kernel Ratio (KR. %)	46.450	0.341	2.86	70.16
Shell Thickness (ST. mm)	1.407	0.013	0.69	2.17
Filled-firm Kernel Raito (FKR. %)	89.950	0.916	0	100

Table 2. Pearson correlation coefficients among all characteristics

	FL	FW	FH	FWe	KW	KR	ST	FKR
FL	1							
FW	0.661**	1						
FH	0.575**	0.879**	1					
FWe	0.705**	0.805**	0.823**	1				
KW	0.654**	0.793**	0.796**	0.867**	1			
KR	-0.106*	-0.046	-0.043	-0.218**	0.246**	1		
ST	0.146**	0.062	0.120*	0.420**	0.080	-0.576**	1	
FKR	0.155**	0.140**	0.169**	0.382**	0.368**	0.030	0.132*	1

*: p<0.05. **. P<0.01

Table 3. Results of Factor analysis

	Factor Score Coefficients (c_{ik})			Rotated Factor Loadings (l_{ik}) and Communalities			
Variables	$Factor_1$	$Factor_2$	$Factor_3$	$Factor_1$	$Factor_2$	$Factor_3$	Communality
FL	0.264	-0.007	0.074	**0.800**	0.091	-0.049	0.651
FW	0.331	-0.092	0.129	**0.950**	-0.028	-0.008	0.903
FH	0.312	-0.072	0.079	**0.923**	0.005	-0.057	0.856
FWe	0.237	0.099	-0.154	**0.874**	0.291	-0.302	0.940
KR	0.037	-0.561	-0.170	-0.040	**0.886**	-0.120	0.801
ST	-0.057	0.538	-0.105	0.104	**-0.881**	-0.169	0.814
FKR	-0.102	-0.049	-0.944	0.122	0.017	**0.979**	0.974
Variance				3.187	1.654	1.098	5.939

Table 4. Results of multiple regression analysis

Predictor	Coef	SE	t	p
F1	0.856	0.020	43.29	0.001
F2	0.161	0.020	-8.17	0.001
F3	0.318	0.020	-16.10	0.001
	S = 0.377	R^2-Sq = 85.9%		
	R^2-Sq(adj) = 85.8%	(P<0.01)		

Conclusion: Various univariate models and approaches can be used to determine relationships between kernel weight and some fruit-kernel characteristics for selecting high productive species in plant breeding programs. But sometimes these models or approaches may fail to define complex relationships among variables related to kernel weight. Multivariate models or combining multivariate and univariate models may be capable to determine relationships among large number variables. In the present study, the relationships between kernel weight and some fruit-kernel characteristics have been examined using both multivariate and univariate approaches. Thus it was found that kernel weight can be predicted 85.6 % successfully by using of some fruit-kernel characteristics. The success of this approach has not been compared with other models due to lack of the modeling studies on kernel weight by using similar variables and approaches. Furthermore the present model may be useful to eliminate multicolinearity problems among large number of

variables. In addition, this approach gets easy doing multiple regression models by reducing large number of variables and interpretation of multiple regression model results by removing indirect effect of related explanatory variables. The present study is one of the pioneer and studies and hopefully, will be useful for future studies of similar nature.

REFERENCES

Akça, Y. and S. M., Sen (1992). Cevizlerde (Juglans regia L.) Onemli Seleksiyon Kriterleri arasındaki Iliskiler. YY Ü. ZF Derg., 2(2): 67- 75.

Anonymous (2000). MINITAB Statistical software, Minitab Inc. USA.

Draper, N. R. and H. Smith (1998). Applied Regression Analysis, John Wiley and Sons, Inc., New York, (USA). 706 pp

Firouz, M., and Y. Bayazid (2003) Evaluatıon of Walnut's Seed Propertıes in Kurdestan Province**.** Iranian J. Forest and Poplar Res. 11(4): 565-584.

Johnson, R. A. and D.W Wichern (2002). Applied Multivariate Statistical Analysis. Prentice Hall, upper Saddle River, New Jersey, (USA). 766 pp

Sen, S. M., (1983). Cevizlerde Önemli Meyve Kalite Faktörleri Arasındaki _liskiler. II. Meyve Agırlıgı ile Kabuk Kalınlıgı ve Kabuk Kırılması Arasındaki _liskiler. Atatürk Üniv. Ziraat Fak. Derg., 14(3): 17-28,

Sen, S. M., (1985). Cevizlerde Kabuk Kalınlıgı, Kabuk Kırılma Direnci, Kabukta Yapısma ve Kabuk Dikine Kırılma Direnci ile Diger Bazı Kalite Faktörleri Arasındaki _liskiler. Doga Bilim Derg., 9(1): 10-24.

Sharma, S. (1996). Applied Multivariate Techniques, John Wiley and Sons, Inc., New York, (USA). 493 pp

Tabachnick, B. G. and L. S Fidell (2001). Using Multivariate Statistics. Allyn and Bacon A Pearson Education Company Boston, (USA). 966 pp.

Yang, J., B. W. Guo and G, Q. Zhang (2001). The studies of principal component analysis on the main economic character and superior variety selection of walnut**.** J. Agri. University of Hebei. DOI: cnki: ISSN:1000-1573.0.2001-04-012.