Sequential Path Model for Grain Yield in Soybean

Mohammad SEDGHI, Bahman AMANPOUR-BALANEJI

University of Mohaghegh Ardabili, Department of Agronomy and Plant Breeding, Faculty of Agriculture,
Ardabil, Iran; mosedghi2003@yahoo.com, amanpour.b@gmail.com

Abstract

This study was performed to determine some physiological traits that affect soybean's grain yield via sequential path analysis. In a factorial experiment, two cultivars (Harcor and Williams) were sown under four levels of nitrogen and two levels of weed management at the research station of Tabriz University, Iran, during 2004 and 2005. Grain yield, some yield components and physiological traits were measured. Correlation coefficient analysis showed that grain yield had significant positive and negative association with measured traits. A sequential path analysis was done in order to evaluate associations among grain yield and related traits by ordering the various variables in first, second and third order paths on the basis of their maximum direct effects and minimal collinearity. Two first-order variables, namely number of pods per plant and pre-flowering net photosynthesis revealed highest direct effect on total grain yield and explained 49, 44 and 47 % of the variation in grain yield based on 2004, 2005, and combined datasets, respectively. Four traits i.e. post-flowering net photosynthesis, plant height, leaf area index and intercepted radiation at the bottom layer of canopy were found to fit as second-order variables. Pre- and post-flowering chlorophyll content, main root length and intercepted radiation at the middle layer of canopy were placed at the third-order path. From the results concluded that, number of pods per plant and pre-flowering net photosynthesis are the best selection criteria in soybean for grain yield.

Keywords: correlation, grain yield, physiological traits, sequential path analysis, soybean

Introduction

The grain yield is a polygenically controlled character. Breeders try to select varieties with high yield potential. The selection on the basis of grain yield is usually not very effective and efficient, but selection based on its related characters could be more efficient. Recently, strategies to optimize yield in soybean have focused on specific production systems and the physiology and mechanisms involved in yield formation (Ball *et al.,* 2005; Carter and Boerma, 1979). Wright (1921), proposed a method called path analysis that partitions the estimated correlations in direct and indirect effects of traits on a basic variable. A path coefficient is a standardized partial regression coefficient, and measures the direct influence of a predictor variable on the response variable (Mohammadi *et al.*, 2003). This method has been studied in soybean (Akhter and Sneller, 1996; Ball *et al.*, 2001; Barbaro *et al.*, 2006; Kau and Modhova, 1972; Santos *et al.*, 1995), rice (Kumar *et al.*, 1999), green gram (Singh and Singh, 1973), corn (Mohammadi *et al.*, 2003) and potato (Asghari-zakaria *et al.*, 2007). Scientists in path analysis, consider the predictor characters as first-order variables to analyze their effects over a dependent variable such as yield (Kumar *et al.*, 1999; Mohammadi *et al.*, 2003). The estimation of the path coefficients can be adversely affected by the effects of multicolinearity between the traits, which appear when the random observations of the explanatory variables or linear combinations are correlated (Ferrari, 1989; Hair *et al.*, 1995; Somante *et*

al.., 1998). Ignorance of multicollinearity effects can bring forth undesirable results (Crus and Careiro, 2003). In this condition, the variances associated to the estimators of the path coefficients can therefore attain very high values, making the estimates little reliable (Carvalho, 1995; Crus and Careiro, 2003). Besides, the parameter estimates can assume values without any coherence with the biological phenomenon under study (Crus and Careiro, 2003). In order to lessen the adverse effects of multicollinearity, one can identify the variables that are causing the problems and eliminate them, to carry out the analysis with a smaller group (Barbaro *et al.*, 2006). Other solution is organizing and analyzing various predictor variables in first, second and third-order paths (Samonte *et al.*, 2004). Akhter and Sneller (1996), correlated yield with vegetative mass, height of plant, and number of main-stem nodes. Board *et al.*, (1997 and 1999a), indicated that seed m^{-2}, reproductive nodes m^{-2} and pods reproductive-node^{-1} served as the best selection criteria in soybean. Board *et al.* (1999b) proposed the pod number per reproductive-node as a selection criterion for high yield. These traits are highly correlated to plant density and growth season, which breeders should attend to climate changes and planting methods.

The objectives of the present study were to determine the effects of some physiological traits of soybean on grain yield via sequential path analysis and with a view to help breeders in the selection process of plants by the best controlled trait for high grain yield.

Materials and methods

Two cultivars of maturity group II and III respectively Harcor and Williams were sown on 19 May 2004 and 12 May 2005 at the research station of Tabriz university, Iran. There were four levels of nitrogen treatments (two levels with *bradyrhizobium japonicum* and two levels of urea application that each group contains a complementary urea application at R1-R2 growth phase about 50 kg N ha^{-1}) and two levels of weed management (weedy and non-weedy). The experiment was factorial that arranged in a randomized complete block design with three replications. In each replication, the size of the plot consisted of five rows with a length of three meters. The spacing between and within the rows were maintained at 60 and 8 cm, respectively and final density was 210,000 plant per hectare.

Grain yield (GY), pod number per plant (POD), photosynthesis rate at V6 stage (PreNP), photosynthesis rate at R2 stage (PostNP), plant height (Height), leaf area index (LAI), intercepted radiation at bottom layer of canopy (PAR 0), intercepted radiation at middle layer of canopy (PAR 1/2), chlorophyll content at V6 stage (PreCHL), chlorophyll content at R2 stage (PostCHL) and main root length (Root) were measured.

Pre (V6) and Post (R2) flowering chlorophyll content and leaflet photosynthetic rate were recorded with SPAD-502 (Minolta, Japan) and HCM-100 portable photosynthesis meter (WALZ, Germany), respectively. At R2 stage intercepted radiation was measured in three different layers of soybean's canopy (top, middle and bottom) with Sun scan instrument. At this stage leaf area of canopy was measured too. Observations were recorded on ten randomly selected plants from each plot for plant height, pods per plant, and main root length. Final harvest area for grain yield was 2 m^2.

Statistical Analysis

The datasets were first tested for Skewness and Kurtosis by Mstatc statistical software. Data from each trait were subjected to analysis of variance using SAS software. Test for homogeneity of error variance between various datasets obtained using Hartley's F$_{max}$ test (Ott, 1988). Correlation coefficients between various pairs of characters were computed. A preliminary analysis was performed by means of the conventional path model in which all traits were considered as first-order predictor variables with grain yield (GY) as the response variable. Sequential stepwise multiple regression was performed to organize the predictor variable into first, second and third-order paths on the basis of their respective contributions to the total variation of grain yield and minimal collinearity. With this procedure two variables i.e. POD and PreNP were selected as first-order variables. This procedure was again performed separately taking POD and PreNP as dependent variables to find out first-order variables for these two

response variables, which shall be, consequently, second-order variables for GY. Similar procedure was followed to determine the third-order variables for GY. Direct effects of yield characters in different order paths were estimated by the procedure described by Williams *et al.* (1990).

Results and discussion

At 2004 dataset, all characters except PAR 1/2, PAR 0, Height and Root showed significant correlation with GY. Among these four variables except Root, the rest showed significant correlation with GY in 2005 and combined datasets. Computed correlation coefficients between different pairs of characters from three datasets are presented in Tab.1 and 2. In these datasets, almost the highest correlations were between PreNP and POD with GY. Conventional path analysis (where the all traits were considered as first-order variables with GY as the response variable), and analysis of collinearity indicated collinearity in the model and inconsistent patterns of relationships among the variables. In this condition, interpretation of results and determination of actual contribution of each criterion on yield will be complicated. In this study, conventional path analysis showed that in some traits such as LAI and PAR 0, multicollinearity (VIF= 93 and 25, respectively) exists and using sequential path analysis can eliminate these effects. Also, PostNP had positive (0.52) and negative (- 0.48) direct effects at 2004 and 2005 datasets based on conventional path model, but the direct effects of this trait in sequential path analysis were high and positive in three datasets (Fig. 1). Some researchers investigated interrelationships among yield and its related traits without consideration or computing collinearity (Ahmad and Saleem, 2003; Akbar *et al.*, 2003; Ball *et al.*, 2001; Rauf *et al.*, 2004). In totally, two different ways suggested to eliminating or decreasing of multicollinearity. First, considering one trait from pair traits with sever correlations in the model (Barbaro *et al.*, 2006; Carvalho, 1995), and second, evaluation of degree of multicollinearity and perform a sequential path analysis (Aghari-zakaria *et al.*, 2007; Mohammadi *et al.*, 2003; Samonte *et al.*, 1998).

In present study, the two mentioned solutions were used and proper models were fitted. In sequential path analysis (Fig. 1), POD and PreNP were considered first-order variables, which explained 49, 44, and 47% of the variation in GY based on 2004, 2005, and combined datasets, respectively (Tab. 3). These two traits had high positive direct effects on GY. The path analysis at second-order variables over the first-order variables showed that 59% (2004), 33% (2005), and 42% (combined) of the total variation for POD were explained by two characters, namely PostNP and Height (tab. 3). Among these characters, PostNP had significant high and positive direct effect, while Height had significant negative direct effect on GY based on three datasets (Tab. 3). In the same order path, LAI and PAR 0 had significant positive and nega-

Tab.1. Correlation coefficients between traits measured in the 2004 (above diagonal) and 2005 (below diagonal) datasets

	LAI	PreCHL	PostCHL	PreNP	PostNP	PAR 1/2	PAR 0	Height	POD	GY	Root
LAI	-	0.54**	0.56*	0.5**	0.49**	0.38**	0.41**	-0.29*	0.66**	0.52**	-0.1
PreCHL	0.65**	-	0.83**	0.17	0.26	0.16	0.18	0.03	0.35*	0.32*	0.26
PostCHL	0.72**	0.76**	-	0.21	0.43**	0.14	0.19	-0.001	0.45**	0.44**	0.24
PreNP	0.54**	0.43**	0.72**	-	0.65**	-0.2	-0.24	0.2	0.36*	0.55**	0.04
PostNP	0.63**	0.59**	0.62**	0.61**	-	-0.14	0.02	0.05	0.6*	0.51**	0.28
PAR 1/2	0.55**	0.44**	0.39**	0.18	0.12	-	0.85**	-0.72**	0.36*	0.22	-0.36*
PAR 0	0.59**	0.28	0.25	0.11	0.03	0.6**	-	-0.77**	0.5**	0.19	-0.15
Height	-0.32*	-0.04	-0.15	0.03	0.23	-0.52**	-0.41**	-	-0.47**	-0.26	0.37**
POD	0.66**	0.6**	0.47**	0.21	0.51**	0.41**	0.48**	-0.18	-	0.62**	0.09
GY	0.60**	0.46**	0.59**	0.63**	0.56**	0.55**	0.33*	-0.29	0.38**	-	-0.11
Root	0.08	0.16	-0.08	-0.02	0.33*	-0.09	-0.07	0.53**	0.12	-0.006	-

* and ** significant at 0.05 and 0.01 probability levels

tive effects respectively on PreNP. These two characters explained 46 (2004), 33 (2005) and 42% (combined) of variation in PreNP (Tab. 3). Results of sequential path analysis when the third-order variables were used as pre-

acters often highlighted in these studies, were the yield components and the effects of physiological characters on yield investigated using path analysis model. In this study specific attention were made toward physiological traits

Tab. 2. Correlation coefficients between traits in combined dataset

	LAI	PreCHL	PostCHL	PreNP	PostNP	PAR 1/2	PAR 0	Height	POD	GY	Root
LAI	-										
PreCHL	0.51**	-									
PostCHL	0.49**	0.85**	-								
PreNP	0.44**	0.43**	0.51**	-							
PostNP	0.54**	0.42**	0.5**	0.63**	-						
PAR 1/2	0.47**	0.27**	0.21*	0.008	-0.005	-					
PAR 0	0.46**	0.05	0.02	-0.23*	-0.03	0.68**	-				
Height	-0.30**	-0.07	-0.12	0.04	0.11	-0.61**	-0.53**	-			
POD	0.55**	0.57**	0.54**	0.42**	0.54**	0.37**	0.25*	-0.31**	-		
GY	0.60**	0.46**	0.53**	0.63**	0.55**	0.25*	0.13	-0.31**	0.53**	-	
Root	-0.009	0.26*	0.18	0.18	0.33**	-0.19	-0.15	0.42**	0.17	0.004	-

* and ** significant at 0.05 and 0.01 probability levels.

dictors and second-order variables as response variables indicated that PreCHL and PAR 1/2 positively influenced LAI and accounted for more than 40% of observed variation in LAI (Tab. 3). These two characters had significant positive correlations with each other (Tab. 1 and 2). For this, these two characters exerted considerable indirect effects on LAI through each other. In the same order path, only PostCHL influenced PostNP with positive direct effects above than 0.50 and adjusted R^2 above than 24 % in combined dataset. Also, the Height of plant was affected by Root. The direct effect of Root on Height was above than 0.42 and more than 17 % of the Height variation was explained by Root (Tab. 3).

Path analyses performed in earlier studies on soybean considered the effects of population density and yield components as first-order variables and yield as the response variable and did not take into account the multicollinearity factor (Ball *et al.*, 2001; Barbaro *et al.*, 2006). The char-

and the PreNP as a physiological characters and POD as a yield component character due to higher direct effects on yield were considered as first-order variables. These characters can be used for selection of soybean genotypes for grain yield. Barbaro *et al.* (2006), found that POD number, plant height and number of reproductive nodes are important criteria for evaluating of different soybean varieties and suggested that pod number per plant as the best selection criterion. Ball *et al.* (2001), also demonstrated that pod number per plant is the most important variable for grain yield in soybean at constant densities.

In conclusion, character associations revealed by path analysis could be influenced by different factors such as variety, traits, climate and statistical methods. Therefore, the general applicability of the sequential path model for determining the effects of the most important traits on grain yield in soybean needs to more studies in different conditions.
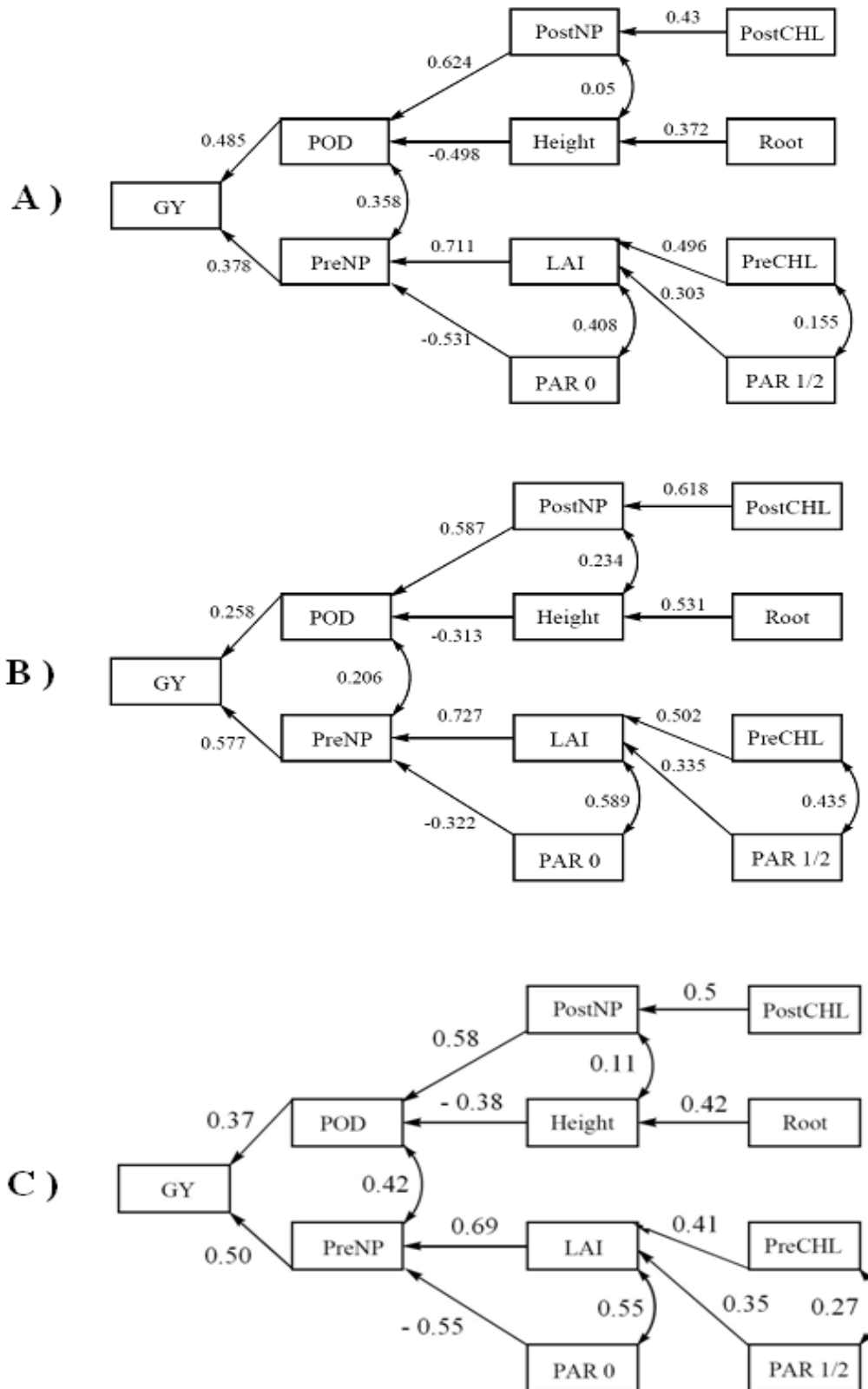
Fig. 1. Sequential path model indicating interrelationships among various characters contributing to grain yield (GY) for 2004 (A), 2005 (B) and combined dataset (C). *Abbreviations:* pod number per plant (POD), photosynthesis rate at V6 stage (PreNP), photosynthesis rate at R2 stage (PostNP), plant height (Height), leaf area index (LAI), intercepted radiation at bottom layer of canopy (PAR 0), chlorophyll content at V6 stage (PreCHL), chlorophyll content at R2 stage (PostCHL), main root length (Root) and intercepted radiation at middle layer of canopy (PAR 1/2).

Tab. 3. Direct effects of variables on soybean grain yield in sequential path model, variation inflation factor (VIF) and tolerance values (1/VIF) for the predictor variables in Model 1 (all predictor variables as first-order variables) and Model 2 (predictors grouped into first, second and third-order variables)

Response variables	Predictor variables	Direct effect			VIF						Tolerance						R^2 Adj.		
					M1			M2			M1			M2					
		2004	2005	total	2004	2005	total	2004	2005	total	2004	2005	total	2004	2005	total	2004	2005	total
GY	POD	0.485	0.258	0.37	3.414	2.433	2.29	1.147	1.044	1.211	0.293	0.411	0.437	0.872	0.958	0.826	0.49	0.44	0.47
	PreNP	0.378	0.577	0.50	2.98	2.868	2.367	1.147	1.044	1.211	0.335	0.349	0.423	0.872	0.958	0.826			
POD	PostNP	0.624	0.587	0.58	1.98	3.641	2.073	1.002	1.058	1.01	0.505	0.275	0.482	0.998	0.945	0.989	0.59	0.33	0.42
	Height	-0.498	-0.313	-0.38	3.01	2.726	2.029	1.002	1.058	1.01	0.332	0.367	0.493	0.998	0.945	0.989			
PreNP	LAI	0.711	0.727	0.69	2.566	6.329	2.693	1.199	1.531	1.268	0.39	0.158	0.37	0.834	0.653	0.789	0.46	0.33	0.42
	PAR 0	-0.531	-0.322	-0.55	5.586	2.55	2.37	1.199	1.531	1.268	0.179	0.392	0.422	0.834	0.653	0.789			
PostNP	PostCHL	0.43	0.618	0.50	3.459	3.728	3.546	1	1	1	0.289	0.268	0.282	1	1	1	0.17	0.37	0.24
Height	Root	0.372	0.531	0.42	1.518	1.227	1.176	1	1	1	0.659	0.815	0.851	1	1	1	0.12	0.27	0.17
LAI	PreCHL	0.496	0.502	0.41	3.353	2.825	3.878	1.025	1.234	1.078	0.298	0.354	0.258	0.976	0.81	0.927	0.36	0.49	0.36
	PAR 1/2	0.303	0.335	0.35	4.544	1.825	1.176	1.025	1.234	1.078	0.22	0.548	0.85	0.976	0.81	0.927			

References

Ahmad, A. and M. Saleem (2003). Path coefficient analysis in *Zea mays* L. International J. of Agriculture & Biology 5(3): 245-248.

Akbar, M., T. Mahmood, M. Anwar, M. Ali, M. Shafiq and J. Salim (2003). Linseed improvement through genetic variability, correlation and path coefficient analysis. International J. of Agriculture & Biology 5(3): 303-305.

Akhter, M. and C. H. Sneller (1996). Yield and yield components of early maturing soybean genotypes in the Mid-south. Crop Sci. 36: 877-882.

Asghari-Zakaria, R., M. Fathi and D. Hasan-Panah (2007). Sequential path analysis of yield components in potato. Potato Research 49:273-279.

Ball, A. R., W. R. McNew, D. E. Vories, C. T. Keisling and C. L. Purcell (2001). Path analyses of population density effects on short-season soybean yield. Agron J. 93:187-195.

Barbaro, I. M., M. A. Pessoa da Cruz Centurion, A. O. Di Mauro, S. H. Uneda-Trevisoli, N. H. Castro Arriel and M. M. Costa (2006). Path analysis and expected response in indirect selection for grain yield in soybean. Crop Breeding and Applied Biotechnology 7:151-159.

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

Board, J. E., M. S. Kang and B. G. Harville (1999a). Path analyses of the yield formation process for late planted soybean. Agron J. 91:128-135.

Board, J. E., M. S. Kang and B. G. Harville (1999b). Identification of soybean cultivars for late planted culture, p. 543. In: Kauffman H.E. Eds. Proc. Of 6th World soybean Conf., Chicago, IL. 4-7 Aug. Superior Printing, Chicago, IL.

Carter, T. E. and H. R. Boerma (1979). Implications on genotype x planting date and row spacing interactions in double-cropped soybean cultivar development. Crop Sci. 19: 607-610.

Carvalho, S. P. (1995). Metodos alternativos de estimacao dos coefecientes de trilha e indices de selecao sob multicolinearidadae. Editora UFV. Vicosa.

Cruz, C. D. and P. C. S. Carneiro (2003). Modelos biometricos aplicados ao melhoramento genetico. Editora UFV. Vicosa.

Ferrari, F. (1989). Estimadores viesados para modelos de regressao em presence de multicolineardade. Editora USP-ESALQ. Piracicaba.

Hair, J. R., R. E. Anderson, W. C. Tatham Black (1995). Multivariate data analysis with readings. Prentice Hall. Englewood. NJ.

Kau, R. H. and M. P. Madhava (1972). Association between yield and its components in soybean. Indian J. Genet Plant Breed. 32:276-280.

Kumar, A., M. G. Gangashetti and A. Dahiya (1999). Analysis of direct and indirect effects for quantitative traits in diallel crosses of maize. Ann. Biol. 15:173-176.

Kumar, I. and S. S. Saini (1973). Path analusis in short stature rice varieties. Indian J. Genet Plant Breed. 33:13-15.

Mohammadi, S. A., B. M. Prasanna and N. N. Singh (2003). Sequential path model for determining interrelationships among grain yield and related characters in maize. Crop Sci. 43:1690-1697.

Ott, L. (1988). An introduction to statistical methods and data analysis. 3rd ed. PWS-Kent, Boston.

Rauf, S., T. M. Khan, H. A. Sadaqat and A.I. Khan (2004). Correlation and path coefficient analysis of yield componenets in cotton. International J. of Agriculture & Biology 6(4):686-688.

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

Santos, C. A. F., M. S. Reis, C. D. Cruz, C. S. Sediyama and T. Sediyama (1995). Adequacao de modelos no estudo do coeficiente de trilha nos componentes primarios e secundarios de progenies F_6 de soja (*Glycine max* L.). Revista Ceres 42:111-121.

Singh, T. P. and K. B. Singh (1973). Association of grain yield and its components in segregating populations of green gram. Indian J. Genet Plant Breed. 33:112-117.

Williams, W. A., M. B. Jones and M. W. Demment (1990). A concise table for path analysis statistics. Agron. J. 82:1022-1024.

Wright, S. (1921). Correlation and causation. Journal of Agricultural Research 20:557-585.