

A Study on Path Analysis of Groundnut (*Arachis hypogaea*) using Sub-Sampling Approach

Nandhini C., Patil S. G.

Received 18 March 2019 ; Accepted 22 April 2019 ; Published on 13 May 2019

Abstract In conventional data analysis, variables are represented by single valued vectors. Utilization of single valued variable may bring about a heavy loss of information. In most of agricultural experiments plant materials gets destroyed while taking observations and finally plant population of particular plot will be affected, to overcome this small number of samples can be used instead of large samples. In present study the genotypic correlation and direct and indirect effect of a variable is computed through path analysis using sub-sampling approach.

Keywords Groundnut, Path analysis, Genotypic correlation, Sub-sampling.

Introduction

The peanut also known as the groundnut (*Arachis hypogaea*) is a legume crop grown in the tropics and subtropics mainly for its edible seeds. It is important to both small and large commercial producers. It belongs to the family Fabaceae. Like most other legumes, peanuts harbor symbiotic nitrogen-fixing bacteria in root nodules. This capacity to fix nitrogen means peanuts require less nitrogen-containing fertilizer and also improve soil fertility, making

them valuable in crop rotations. An experiment was conducted in a RBD with three replications and compared 34 groundnut genotypes. For all the characters under study showed the highly significant variations among the 34 genotypes by the analysis of variance and dispersion (Zaman et al. 2010). Thirty groundnut genotypes was evaluated for both pod and haulm yield and also for haulm nutritive traits. Finally they found genotypes GAF 1723, ICGV 00064 and ICGV-IS 13998 have combined good pod and haulm yield with high haulm nutritive quality (Oteng-Frimpong et al. 2017).

The method of path analysis was originated and developed by Sewall Wright, a geneticist, in the 1920s (Wright 1921). Path analysis is a form of multiple regression analysis that is used to evaluate causal models by examining the relationships between a dependent variable and two or more independent variables which explains the variation with direct and indirect effects. Genotype \times environment (G \times E) interaction of groundnut crop was studied and its result shows significant G \times E interaction and shows significant differences between genotypes and locations for pod yield ($p < 0.01$) and observed high positive correlation between pod yield per plant and number of pods per plant (Asibuo et al. 2018).

Uprooting the plants is necessary while taking observations on root nodules / peg formation at different time intervals which leads to destruction of experimental material ultimately results in altering the plant population and yield. To avoid such loss of planting material only two samples can be utilized for path analysis using sub-sampling approach. The

Nandhini C., Dr Patil Santosh Ganapati*¹

¹Assistant Professor (Agricultural Statistics)

, Dept of Physical Sciences & Information Technology, AEC & RI,,
 Tamil Nadu Agricultural University, Coimbatore 641003, India
 e-mail: patilsgstat@tnau. ac.in

*Corresponding author

Table 1. ANOVA with sub-sampling in Randomized Block Design (RBD).

SV	df	SS	MS	E (MS)
Replication	r-1	$ts \sum .j. (\bar{Y}_{.j.} - \bar{Y}_{...})^2$	MS_{blk}	$\sigma \frac{2}{\epsilon} + s\sigma \frac{2}{\delta} ts \sum_j \beta_j^2 / (r-1)$
Treatments	t-1	$rs \sum_i (\bar{Y}_{i...} - \bar{Y}_{...})^2$	MS_{trt}	$\sigma \frac{2}{\epsilon} + s\sigma \frac{2}{\delta} + rs \sum_j \alpha_j^2 / (t-1)$
Plot error	(r-1) (t-1)	$s \sum_j \sum_i (\bar{Y}_{ij...} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{...})^2$	$MS_{plot\ error}$	$\sigma \frac{2}{\epsilon} + s\sigma \frac{2}{\delta}$
Sub-sampling error	tr (s-1)	$\sum_i \sum_j \sum_l (\bar{Y}_{ij.} - \bar{y}_{ij.})^2$	$MS_{ss\ error}$	$\sigma \frac{2}{\epsilon}$
Total	trs-1	SSTot		

present study was carried out to assess the genotypic correlation and path analysis for groundnut using sub-sampling approach.

Materials and Methods

ANOVA for sub-sampling

Data for a character is measured more than once is called multiobservation data which helps to increase the sample size per plot. Federer (1957) presented ANOVA procedures for sub-sample experimental data. Consider an experiment is laid out in Randomized Block Design with t treatments and r blocks. ANOVA model as :

$$Y_{ijl} = \mu + \alpha_i + \beta_j + \delta_{ij} + \epsilon_{ijl} \quad (1)$$

Where, Y_{ijl} denote i^{th} treatment in j^{th} block on l^{th} plant, α_i is i^{th} treatment ($i = 1, 2, \dots, t$), β_j is j^{th} block ($j = 1, 2, \dots, r$), $\delta_{ij} \sim N(0, \sigma \frac{2}{\delta})$ error due to plot and $\epsilon_{ijl} \sim N(0, \sigma \frac{2}{\epsilon})$ error due to sub-sample ($l = 1, 2, \dots, s$). ANOVA

VA for sub-sampling method given in Table 1.

Computational procedure

Components of variance and covariance: As mean

sum of squares, expectations of mean sum of products follow the same principles. Thus :

Geno-covariance - $\sigma_{gij} = \frac{MSP_{trt} - MSP_{plot\ error}}{rs} = \frac{\sigma_{eij} + rs \sigma_{gij} - \sigma_{eij}}{rs}$

where $i \neq j$

Genotypic variance = σ_{gij} if $i = j$.

Genotypic correlation : The genotypic correlation is the measure of association between the genotypes of individuals. It is used to describe how traits are associated at the genetic level and used to predict the effect of selection on one trait on changes in other traits. For a single trait, phenotypic variance is partitioned into genetic and environmental variance.

Genotypic covariance (G) = $\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$

$V = \text{diag}(\sqrt{G})$

Now genotypic correlation (ρ) = $V^{-1/2}GV^{-1/2}$

$$\rho = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{1Y} \\ \rho_{12} & 1 & \rho_{23} & \rho_{2Y} \\ \rho_{13} & \rho_{23} & 1 & \rho_{3Y} \\ \rho_{1Y} & \rho_{2Y} & \rho_{3Y} & 1 \end{bmatrix} \quad (2)$$

Table 2. ANCOVA table for RBD field data using sub-sampling approach. *Significant at 5% level of significance. Diagonal elements indicate the mean sum of squares and off diagonal elements indicates the mean sum of products, x_1 = Plant height, x_2 = Number of flowers per plant, x_3 = Number of pods per plant and Y = Pod yield per plant.

SV	df	MSP					
		x_1	x_2	x_3	Y		
Treatment	7	x_1	49.82*	4.36	-9.07	-19.94	
		x_2	4.36	3.26	-5.94	-4.30	
		x_3	-9.07	-5.94	122.28*	188.81	
		Y	-19.94	-4.30	188.81	392.00*	
Replication	2	x_1	40.52	-6.67	-9.02	-29.41	
		x_2	-6.67	6.52	20.64	49.30	
		x_3	-9.02	20.64	69.65	163.57	
		Y	-29.41	49.30	163.57	385.90	
Experimental error	14	x_1	7.74	1.30	-0.52	0.14	
		x_2	1.30	1.40	-0.64	-2.09	
		x_3	-0.52	-0.64	3.91	7.98	
		Y	0.14	-2.09	7.98	26.87	
Sub-sampling error	24	x_1	5.24	0.75	1.71	1.74	
		x_2	0.75	1.60	-0.08	-1.00	
		x_3	1.71	-0.08	11.48	16.96	
		Y	1.74	-1.00	16.96	27.56	

Path coefficients and residual effect : Consider linear relationship between dependent variable (Y) and independent variables x_1 , x_2 and x_3 .

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + R \tag{3}$$

Where, R is residual term.

We can rewrite the equation 3 as follows :

$$\frac{y-\bar{y}}{\sigma_y} = \beta_1 \frac{\sigma \times 1}{\sigma_y} \left(\frac{x_1 - \bar{x}_1}{\sigma_{x1}} \right) + \beta_2 \frac{\sigma \times 2}{\sigma_y} \left(\frac{x_2 - \bar{x}_2}{\sigma_{x2}} \right) + \beta_3 \frac{\sigma_{x3}}{\sigma_y} \left(\frac{x_3 - \bar{x}_3}{\sigma_{x2}} \right) + \frac{\sigma_R}{\sigma_y} \frac{R}{\sigma_R} \tag{4}$$

Which reduces to as :

$$Y = P_1 X_1 + P_2 X_2 + P_3 X_3 + P_R E_R \tag{5}$$

Table 3. Direct effect and its correlation with pod yield. Where r–correlation of corresponding variables with pod yield.

Variables	Direct effect	Correlation (r)
x_1	-0.153	-0.16
x_2	0.312	-0.08
x_3	0.963	0.87

Path coefficient can be defined as ratio of standard deviation of causal variable x_i to the standard deviation of Y.

$$p_i = \beta_{ii} \frac{\sigma_{x_i}}{\sigma_y}$$

Consider genotypic correlation matrix from the ANCOVA table as given bellow the direct and indirect path coefficients are calculated as :

$$\rho = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{1Y} \\ \rho_{12} & 1 & \rho_{23} & \rho_{2Y} \\ \rho_{13} & \rho_{23} & 1 & \rho_{3Y} \\ \rho_{1Y} & \rho_{2Y} & \rho_{3Y} & 1 \end{bmatrix} = \begin{bmatrix} B & A \\ A' & 1 \end{bmatrix} \tag{6}$$

Where, $B = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$ and $A = \begin{bmatrix} \rho_{1Y} \\ \rho_{2Y} \\ \rho_{3Y} \end{bmatrix}$

Direct effect = $P = B^{-1} A = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{23} & \rho_{23} & \rho_{3Y} \end{bmatrix}^{-1} \begin{bmatrix} \rho_{1Y} \\ \rho_{2Y} \\ \rho_{3Y} \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \\ P_3 \end{bmatrix}$

Where, P_i is the path coefficient from x_i to Y which.

For calculating indirect effect :

$$Q = (P_{3 \times 1} \ 1_{1 \times 3}) \odot I_{3 \times 3} = \begin{bmatrix} P_1 & 0 & 0 \\ 0 & P_2 & 0 \\ 0 & 0 & P_3 \end{bmatrix}$$

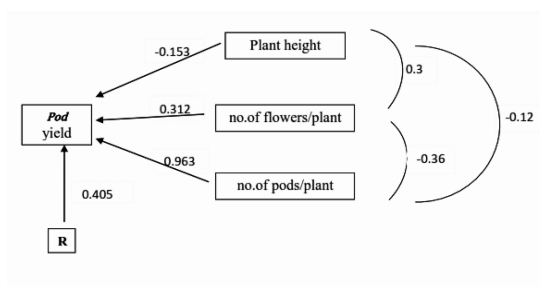


Fig. 1. Path diagram.

Where, \odot is Hadamard product

$$QB = \begin{bmatrix} P_1 & 0 & 0 \\ 0 & P_2 & 0 \\ 0 & 0 & P_3 \end{bmatrix} \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix} = \begin{bmatrix} P_1 & \rho_{12}P_1 & \rho_{13}P_1 \\ \rho_{12}P_2 & P_2 & \rho_{23}P_2 \\ \rho_{13}P_3 & \rho_{23}P_3 & P_3 \end{bmatrix} \quad (8)$$

The diagonal elements and off diagonal elements represent the direct and indirect causal effect on Y respectively.

Residual effect : The x_i and R are uncorrelated so variance of equation (5) yields.

$$1 = \sum_{i=1}^P P_i^2 + 2 \sum_{i,j=1}^P P_i P_j \rho_{ij} + P_R^2 \quad (\text{Where, } p \text{ causal factors, since } i \neq j)$$

$$P_R^2 = 1 - \sum_{i=1}^P P_i^2 + 2 \sum_{i,j=1}^P P_i P_j \rho_{ij} \quad (9)$$

The square root of P_R^2 shows the percentage of variability not accounted by the causal variables.

Results and Discussion

To study the effect of yield components on pod yield of eight groundnut varieties under drip irrigation a field experiment was conducted during Karthigaipattam (November–December) in Randomized Block Design with three replications. The experimental field is geographically located in the Western part (Kongu Nadu) of Tamil Nadu at 11°24' 34" N latitude and 77°39' 1" E longitude. The experimental site has a semi arid climate with moderate to high temperature throughout the year and relatively low rainfall. The data collected on plant height, number of flowers per

plant, number of pods per plant and pod yield per plant from the field.

From the Table 2, plant height, number of pods per plant and pod weight per plant showed significant at 5% level of significance. Calculation of genotypic correlation and path coefficients from Table 2 is given as :

Correlation matrix

	x_1	x_2	x_3	Y
x_1	1	0.35	-0.12	-0.16
x_2	0.35	1	-0.36	-0.08
x_3	-0.12	-0.36	1	0.87
Y	-0.16	-0.08	0.87	1

From the correlation matrix, the plant height and the number of flowers per plant were negatively correlated with both the number of pods per plant and pod yield. Number of pods per plant shows the highest positive correlation with pod yield (0.87).

Direct and indirect path coefficients

Path coefficients = x_i	x_1	x_2	x_3
	-0.153	0.108	-0.117
	-0.053	0.312	-0.344
	0.019	-0.112	0.963

From the above matrix, diagonal elements shows direct effect and off diagonal elements shows indirect effect. In this the number of pods per plant had greater positive direct effects on pod yield followed by number of flowers per plant and negative direct effects on pod yield by plant height. The variables having positive direct effects on grain yield were selected for further breeding program. The cause and effect relationship can be shown using path diagram as given in Fig. 1.

Residual effect

From the equation (9), the square root of P_R^2 gives path coefficient of R on Y. Here $P_R = 0.405$, the value of P_R multiplied by 100 gives the percentage of variability not explained by the variables. Therefore 40.5% of variability in Y was not explained

by the variables x_1 , x_2 and x_3 .

Conclusion

From the Table 3, correlation coefficient between plant height and pod yield ($r = -0.16$) is almost equal to its direct effect ($P_1 = -0.153$), then we can conclude that correlation explains the true relationship and a direct selection based on this character will be effective. Here we can see that correlation coefficient between number of plants per plant and pod yield is negative but it shows positive direct effect. In that case, the restricted simultaneous selection model to be followed to nullify the undesirable indirect effects in order to make use of the direct effect (Singh and Chaudhary 1979). Finally the variables (plant height, number of flowers per plant, number of pods per plant) explained only 59.5% of the variability in pod yield. To account fully for the variation in yield,

some other yield components need to be included in this analysis.

References

- Asibuo JY, Forpoh AS, Akromah R (2018) Genotype \times environment interactions of groundnut (*Arachis hypogaea* L.) for pod yield. *Ecol Genet and Genom* 7 : 27—32.
- Federer WT (1957) Variance and covariance analyses for unbalanced classifications. *Biometrics* 13 (3) : 333—362.
- Oteng-Frimpong R, Konlan SP, Denwar NN (2017) Evaluation of selected groundnut (*Arachis hypogaea* L.) lines for yield and haulm nutritive quality traits. *Int J Agron* 2017 (In press).
- Singh RK, Chaudhary BD (1979) Biometrical methods in quantitative genetic analysis. Kalyani Publishers, New Delhi, India.
- Wright S (1921) Correlation and causation. *J Agric Res* 20 : 557—585.
- Zaman M, Tuhina--Khatun M, Moniruzzaman M, Yousuf M (2010) Genetic divergence in groundnut (*Arachis hypogaea* L.). *Bangladesh J Pl Breed and Genet* 23 (1) : 45—49.