J. Hortic. Sci.
Vol. 21(1), 2026

https://doi.org/10.24154/jhs.v21i1.4639

Original Research Paper

Application of MANOVA for simultaneous trait-based optimization of harvest and ripening stages in pumpkin (*Cucurbita moschata*)

Ranebennur S.[1*], Venugopalan R.[2], Padmini K.[2], Ranganath H.K.[1] and Sisira P.[2]

[1] Department of Agricultural Statistics, Applied Mathematics & Computer Science
College of Agriculture, University of Agricultural Sciences, Bengaluru - 560 065, India
[2] ICAR-Indian Institute of Horticultural Research, Bengaluru - 560 089, India
[*]Corresponding author Email: sukanyarsuccess@gmail.com

ABSTRACT

This study utilized a multivariate analysis of variance (MANOVA) approach to identify the optimal harvest stage and post-harvest ripening treatments for pumpkin (*Cucurbita moschata*) to maximize seed yield and quality. Twelve treatments as a combination of three harvesting stages (25, 35, and 45 days' post-pollination) and four post-harvest ripening durations (0, 10, 20, and 30 days) were analyzed in a randomized block design. Data were collected on seven seed yield traits *i.e.* total seed weight, number of filled and empty seeds, 100-seed weight, total seeds per fruit, seed yield per plant, and seed yield per plot. Principal component analysis (PCA) addressed multi-collinearity among traits by reducing dimensionality and identifying key contributors to variability. MANOVA revealed significant effects of treatments on the traits with treatment 12 (fruit harvested at 45 days after pollination and stored for 30 days) ranked as the most effective based on PCA-derived composite scores integrating all traits. This integrated statistical analysis highlights the potential to optimize agricultural practices for improved seed yield and quality in pumpkin cultivation.

Keywords: MANOVA, principal component analysis, pumpkin

INTRODUCTION

Pumpkin (*Cucurbita* spp.; 2n=2x=40), belonging to the family Cucurbitaceae, is commonly known as *Sitaphal*, *Kashiphal*, or *Kaddu*. Cucurbit seeds exhibit variation in size, shape, and structure, which are important taxonomic characteristics (Chakravarty & Hore, 1979). Economically important cucurbits such as cucumber, melon, and squash have increased the significance of research on optimal seed production (Welbaum, 1999). In several regions, cucurbit seeds, especially watermelon and squash, are also consumed as nutritious snack foods due to their high nutritional value (Madaan & Lal, 1984).

Seed maturity in cucurbits occurs over an extended period after pollination and is strongly influenced by developmental temperature. Unlike many crops, cucurbit seeds do not require desiccation to complete maturation. Seed maturity directly affects storage potential; seeds harvested prematurely are often immature and show poor storage quality. Since pumpkin fruits can remain viable for several days after harvest, seeds may continue to develop through nutrient transfer from the fruit pulp. Therefore, fruits harvested before full physiological maturity may still produce high-quality seeds if subjected to post-harvest ripening. Determining the correct harvest stage is essential for maximizing seed yield and quality while minimizing losses due to fruit rotting and adverse environmental conditions (Kumar et al., 2014; Singh et al., 2016).

Agricultural experiments often involve multiple correlated traits; therefore, multivariate analysis of variance (MANOVA) provides a more suitable approach than separate univariate ANOVAs (Oehlert, 2000). MANOVA simultaneously evaluates differences across multiple response variables, improving statistical power, accounting for interrelationships among traits, and controlling overall Type I error (Porter & O'Reilly, 2017). This approach is particularly useful for multi-season experiments conducted under consistent treatment structures.

In the present study, MANOVA was applied to simultaneously evaluate fruit harvest stage and post-harvest ripening treatments based on multiple seed yield and related traits. To address multicollinearity among variables such as total seed weight, number of

Table 1 : Multivariate analysis of variance (MANOVA) for p^{th} variable

Source of variation	Degrees offreedom	SSCPM (sum of squares and cross product matrix)
Treatment	v-1 = h	$H = b\ \sum_{i=1}^{v}(\bar{y}_{i.} = \bar{y}_{..})(\bar{y}_{i.} = \bar{y}_{..})'$
Replication	r-1 = t	$B = v\ \sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})(\bar{y}_{.j} - \bar{y}_{..})'$
Residual	(v-1)(r-1) =s	$R = \sum_{i=1}^{v}\sum_{j=1}^{b}(\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})(\bar{y}_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})'$
Total	vr-1	$T = \sum_{i=1}^{v}\sum_{j=1}^{b}(\bar{y}_{ij} - \bar{y}_{..})(\bar{y}_{ij} - \bar{y}_{..})'$

filled and empty seeds, 100-seed weight, total seeds per fruit, seed yield per plant, and seed yield per plot, principal component analysis (PCA) was employed. PCA reduces dimensionality by transforming correlated variables into independent principal components, thereby improving model robustness and interpretation (Ongala et al., 2016). This combined MANOVA–PCA approach enabled a comprehensive and reliable assessment of treatment effects and identification of the optimum harvest and ripening strategy.

With this background, the present study has been undertaken to identify the best effective stage of fruit harvest and pre storage duration treatments simultaneously based on all seed yield and associated traits and increase good quality seed in pumpkin.

MATERIALS AND METHODS

This study mainly utilizes secondary data on seven seed yield traits of pumpkin from experimental research field at ICAR-Indian Institute of Horticultural Research, Bangalore, India during 2021-22. Experiment was conducted in a randomized block design in a plot size of 3.5 meters by 3 meters (10.5 m), at spacing of 2 meters between rows and 1 meter between plants, and accommodated 8 plants. Each plant produces 2 fruits, giving a total yield of 16 fruits per plot. Each flower was date-tagged at anthesis to monitor fruit age and fruit harvested at 25, 35 and 45 days after pollination by insect (DAP) and then stored for 0, 10, 20 and 30 days at room temperature, before extracting the seeds. The experiment comprised twelve treatments formed by the combination of three fruit harvesting stages, namely 25, 35, and 45 days after flower anthesis, and four post-harvest ripening periods prior to seed extraction (0, 10, 20, and 30 days). Immediate seed extraction was carried out on the day of harvest for the 0-day ripening treatment, while fruits assigned to the remaining treatments were allowed to ripen for the specified duration before seed extraction. Observations recorded on seven seed yield traits *i.e.* total seed weight (g), number of filled seeds, number of empty seeds, 100 seed weight (g), total seeds per fruit (g), total seeds per plant (g), total seeds per plot (g).

Multivariate analysis of variance (MANOVA) is a statistical technique used to compare two or more dependent variables simultaneously across multiple groups. It extends analysis of variance (ANOVA), which compares the means of a single dependent variable among groups such as stages of fruit harvest and pre-storage duration treatments. By analyzing multiple dependent variables together, MANOVA determines whether significant differences exist among treatment means based on their combined effects.

Wilks' Lambda (Λ)

It measures the proportion of the variance in the dependent variables not accounted for by the independent variables. A smaller value of Λ indicates a stronger relationship between the independent and dependent variables.

$$\Lambda = \frac{|R|}{|H + R|}$$

Assuming the normal distribution, Wilk's Λ is distributed as $\chi^2_{p(v-1)}$ (α) when sample size is large.

To test the significance of treatment we have

if $-(n-1-\underline{(p+v)})\ \ln \Lambda > \chi^2_{p(v-1)}$ (α)

Reject the H_0 at significance level of α, where $\chi^2_{p(v-1)}$ (α) is the upper $(100\alpha)^{th}$ percentile of a chi-square distribution with p(v-1) df

p is the number of variables,

v is the number of treatments, and n is number of degrees of freedom error(s) and treatment (t) *i.e.* t+s.

Principal component analysis

Method of principal component analysis is a data reduction technique that transform set of highly correlated 'p' variables into linear combination of few uncorrelated principal components which account for most of the variation in the data set. Principal components (PC) are uncorrelated linear combination of p random variables $X_1, X_2, \ldots, X_p$ and it depend solely on the covariance matrix ($\sum$) or correlation matrix (ρ), whose variances are as large as possible.

If $X^1 = [X_1, X_2, \ldots, X_p]$ be any random vector with covariance matrix and eigen values $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_p \geq 0$

Then,

$$Z_1 = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p$$

$$Z_2 = a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p$$

.

.

$$Z_p = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p$$

The principal component $Z_1, Z_2, \ldots, Z_p$ are the linear combination of the variable $X_1, X_2, \ldots, X_p$, whose variance Var (Z_i) is as large as possible.

Var (Z_1) = λ_1 and the constants are the elements of corresponding eigen vector, scaled so that

$$a_{i1}^2 + a_{i2}^2 + \ldots + a_{ip}^2 = 1$$

The principal component extracted so that $PC_1(Z_1)$ accounts for large amount of total variation in the data.

i.e., Var (Z_1) = λ_1 subject to condition $a_1'a_1 = 1$

is the weighed linear combination of observed variable which is uncorrelated with and account for maximum amount of total variation not already account for

i.e., Var (Z_1) = λ_1 subject to condition $a_2'a_2 = 1$
Similarly, for i_{th} PC,

Var (Z_1) = λ_1 subject to condition $a_1'a_1 = 1$ and Cov (Z_i, Z_j) = 0

The advantage of PCA technique is that it can tackle the situation when the number of variables is larger than number of observations.

Composite index formation based on the principal component analysis

After standardization of the variable data, next step is to assigning of factor loading and weights. The corresponding factor loadings and weights of the treatment is computed by using principal component analysis. The weights of the variables will be estimated with the help of eigen values and principal component matrix.

Equation $|\sum - \lambda I| = 0$, is solved for obtaining eigen values

Where $\sum$ denotes pooled variance covariance matrix of transformed indictors. Solving the matrix equation $|\sum - \lambda I| = 0$, standardized eigen vectors are obtained from each eigen value.

The weight of the variable is estimated as,

$$W_i = \sum_j |L_{ij}| E_j$$

Where,

i = 1, 2, ..., variable and

j = 1, 2, ..., principal components

L_{ij} is the factor loading of i^{th} variable on j^{th} principal component.

E_j is the eigen value of j^{th} variable.

The index is determined using these weights in the following formula,

$$I_j = \frac{\sum_{i=1}^{n} X_i W_i}{\sum_{i=1}^{n} W_i}$$

Ranking of the treatments on the basis of formed composite scores

The composite development index was calculated using the formula for composite scores, which consolidates multiple individual measurements, items, or variables into a single score. This composite score summarizes overall performance, status, or characteristics represented by these components. Using the obtained value, treatments are ranked in descending order of the index. The treatment with maximum index is ranked as first and next highest and so on.

Table 2 : Correlation matrix of seed yield and associated traits

Seed yield trait	Total seed weight	No. of filled seeds	No. of empty seeds	100 seed weight	Total seeds/ fruit	Seed yield/ plant	Seed yield/ plot
Total seed weight	1						
No of filled seeds	0.79*	1					
No of empty seeds	-0.62*	-0.93***	1				
100seed weight	0.87***	0.74**	-0.75**	1			
Total seeds per fruit	0.77**	0.64*	-0.33	0.34	1		
Seed yield per plant	1***	0.79**	-0.62*	0.87***	0.77**	1	
Seed yield per plot	1***	0.79**	-0.62*	0.87***	0.77**	1***	1

*** Correlation is significant at 0.001 level (two tailed); ** Correlation is significant at 0.01 level (two tailed) * Correlation is significant at 0.05 level (two tailed)

RESULTS AND DISCUSSION

To perform MANOVA, it was essential first to analyse the relationship between traits such as total seed weight, no of filled seeds, no of empty seeds, 100 seed weight, total seeds per fruit, total seeds per plant, and total seeds per plot, considering the potential for multicollinearity. These traits were likely to be correlated, as shown in Table 2. For example, higher no of filled seeds might lead to a higher total seed weight, or a greater total number of seeds per plant might influence the total seeds per plot. However, such correlations could introduce collinearity, complicating statistical analyses.

To address this, principal component analysis (PCA) was often used. PCA reduced the dimensionality of the data by transforming the correlated variables into a set of uncorrelated components, known as principal components. These components captured the majority of the variance in the data while minimizing collinearity. By focusing on the principal components, researchers could simplify the analysis, reducing redundancy in the data and gaining clearer insights into the underlying factors driving variation in seed traits. This approach allowed for a more robust and interpretable model, ensuring that the relationships between traits were better understood without the confounding effects of collinearity.

Data presented in Table 2, provided correlations between various seed yield traits, revealing how these traits interacted and influenced overall yield. Correlation coefficients measured the strength and direction of these relationships, with values ranging from -1 to 1. A coefficient of 1 indicated a perfect positive correlation, whereas increase in one trait directly led to an increase in another. Conversely, a coefficient of -1 indicated a perfect negative correlation, whereas increase in one trait resulted in a decrease in another. The significance of these correlations was indicated by asterisks, highlighting the reliability of these relationships.

The analysis was executed using a dataset comprised of seed yield and associated traits of pumpkin. Since, these traits were measured using different units of measurement, the data pertaining to these traits were transformed to obtain unit-free standard normal values of the original dataset. Further, the method of principal component analysis (PCA) was employed on the standardized data to reduce the large volume of data. It also transforming a large set of correlated variables into a smaller set of uncorrelated variables, known as principal components. PCA yielded a few mutually orthogonal variables that accounted for most of the variation in the dataset without losing any information.

Each principal component constructed using PCA extracted the maximum variability present in the dataset. The variability extracted by the first seven principal components is presented in Table 3.

The variance associated with each principal component was represented by the corresponding eigen values (Table 3). The maximum variability was captured by the first principal component, which accounted for 75.41% of the variability. The second principal component captured 14.79% of the variability, and so on. From the cumulative percentage column in the Table, it was observed that the first two principal components together accounted for approximately 90.20% of the total variability in the data. This meant that these two components captured the vast majority of the dataset's variability, making them highly effective for reducing the dataset's dimensionality while retaining most of the information. Using just these two components significantly simplified the data, allowing for easier analysis and visualization with minimal loss of detail.

Table 3 : Eigen values and extraction of variability

PC No.	Eigen values	Variance (%)	Cumulative (%)
1	5.29	75.41	75.41
2	1.04	14.79	90.20
3	0.62	8.72	98.92
4	0.08	1.08	100.00
5	0.00	0.00	100.00
6	0.00	0.00	100.00
7	0.00	0.00	100.00

The percentage of variability explained by each principal component (PC) was indicated by the scree plot (Fig. 1). Maximum variability in the dataset was captured by the first two eigen values. The plot exhibited an "elbow" after the second component, where the slope of the line flattened. This is a common indication in PCA that the first few components are sufficient to describe most of the variability in the data, and additional components contribute very little. In this case, focusing on the first two components (PC1 and PC2) proved effective for dimensionality reduction, while the remaining components could be disregarded without losing significant information.

Principal component coefficients, or loadings, were the eigen vectors associated with each eigen value, providing information about the importance of each variable in forming the principal components (PCs). The factor loadings corresponding to the principal components are presented in Table 4.

The MANOVA revealed a statistically significant effect of the treatment groups on the multivariate

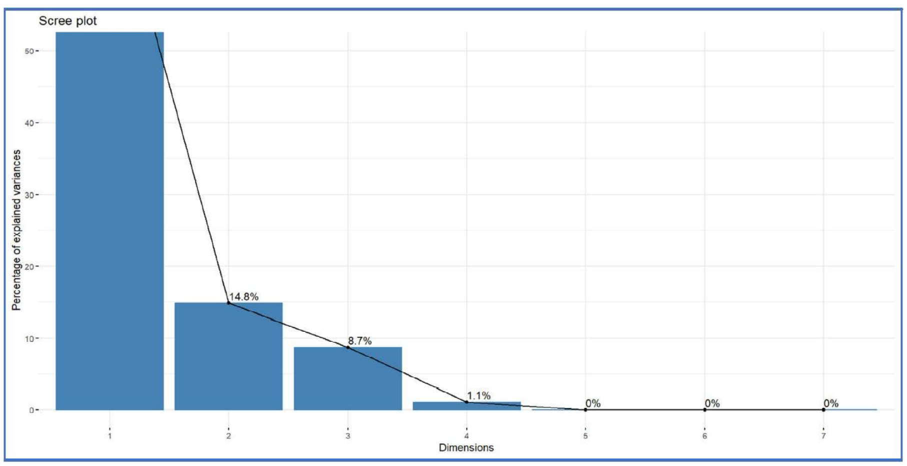

Fig. 1 : Scree plot for Eigen value of seed yield and associated traits data

Table 4 : Factor loadings corresponding in seven principal components

Trait	Principal components						
	1	2	3	4	5	6	7
Total seed weight	-0.42	0.16	0.19	-0.31	-0.24	0.78	0.00
No. of filled seeds	-0.39	-0.24	-0.50	0.10	0.69	0.21	0.00
No. of empty seeds	0.30	0.65	0.34	0.11	0.57	0.18	0.00
100 seed weight	-0.37	-0.23	0.54	0.72	0.00	0.00	0.00
Total seeds per fruit	-0.29	0.62	-0.48	0.42	-0.32	-0.10	0.00
Seed yield per plant	-0.42	0.16	0.19	-0.31	0.12	-0.39	0.71
Seed yield per plot	-0.42	0.16	0.19	-0.31	0.12	-0.39	-0.71

Table 5 : Summary of multivariate analysis of variance (MANOVA)

SV	df	Wilk's lambda	F	num df	den df	Pr(>F)
Treatment	11	1.38	6.98*	22	70	<0.05

* indicates significant five per cent level

response variables. The analysis includes the test statistic, F value, degrees of freedom, and p-value for twelve treatment groups (Table 5). The Pillai's trace statistic is 1.38, which is highly significant at the 0.05 level, indicating that the mean vectors for all twelve treatment groups differ significantly. With 11 degrees of freedom for the treatment factor and 22 and 70 degrees of freedom for the numerator and denominator, respectively, the F-statistic is 6.98. The p-value is less than 0.05, confirming that the treatment groups have a significant impact on the multivariate response and leading to the rejection of the null hypothesis.

The treatment factor, with 11 degrees of freedom, accounts for a substantial portion of the variance, having a sum of squares value of 182.58 (Table 6). The mean sum of squares for the treatment is 16.60, resulting in an F value of 9.02. The extremely low p-value (<0.05) indicates that the treatment effect is highly significant, suggesting that the observed differences across treatment groups are unlikely to be due to random chance. The error term, with 35 degrees of freedom, represents the residual variance not explained by the model, accounting for the remaining variability in PC2 not attributed to treatment or replication.

The treatment effect is statistically significant, with a p-value of less than 0.05, indicating that differences among the treatment groups have a significant impact on PC2 (Table 7). The F value of 5.56 demonstrates that the variance explained by the treatment is substantial compared to the residual (error) variance. The error term, with 35 degrees of freedom, represents the residual variance not explained by the model, accounting for the remaining variability in PC2 not attributed to treatment or replication.

Formation of composite scores

As discussed in methodology, the composite development index was calculated using the formula.

$$I_j = \frac{\sum_{i=1}^{n} X_i W_i}{\sum_{i=1}^{n} W_i}$$

For composite scores, which consolidates multiple individual measurements, items, or variables into a single score. This composite score summarizes overall performance, status, or characteristics represented by these components. The significance of the treatments on the composite scores was evaluated using multivariate analysis of variance (MANOVA). The MANOVA results revealed that the treatment factor had a statistically significant impact on both PC1 and

Table 6 : Summary of analysis of variance for PC1

SV	df	Sum of square	Mean sum of square	F value	Pr(>F)
Replication	1	1.05	1.05	0.58	> 0.05
Treatment	11	182.58	16.60	9.02*	< 0.05
Error	35	64.48	1.84		

* indicates significant at 5 per cent level

Table 7 : Summary of analysis of variance (ANOVA) for PC2

SV	df	Sum of square	Mean sum of square	F value	Pr(>F)
Replication	1	0.30	0.30	0.60	>0.05
Treatment	11	30.73	2.80	5.56*	<0.05
Error	35	17.60	0.50		

* indicates significant at 5 per cent level

PC2, with highly significant p-values. This indicates that the treatments influenced not only the individual response variables but also the broader patterns of variation captured by the principal components.

The significant MANOVA results suggest that the stage of fruit harvest and pre-storage duration treatments induce distinct and measurable changes in the response variables, as reflected in the composite scores. These findings are crucial for identifying which treatments lead to the most favourable outcomes, as defined by the principal components.

The composite scores reflect the performance or effectiveness of each treatment. Treatments are ranked based on the mean value of these composite scores. Treatment 12 (fruit harvested at 45 days after pollination and stored for 30 days) achieved the highest composite score of 106.53, earning the top rank (Table 8).

Table 8 : Composite scores and assign rank

Treatment	Mean of composite scores	Rank
12	106.53	1
8	106.12	2
7	103.16	3
5	100.85	4
9	100.84	5
10	95.35	6
6	94.57	7
1	87.53	8
4	87.42	9
3	81.06	10
2	73.63	11
11	70.97	12

By considering all above results, fruit harvested at 45 days after pollination and stored for 30 days is identified as the best stage of harvest and post-harvest ripening of fruits treatments, simultaneously based on all seed yield and associated traits performance using MANOVA approach in Pumpkin.

CONCLUSION

This study demonstrated the application of a MANOVA-based statistical framework, supported by PCA, to evaluate and identify the most effective stage of fruit harvest timing and post-harvest ripening duration for improving seed yield and associated traits in pumpkin. The results confirmed significant multivariate differences among treatment groups, with composite scores clearly distinguishing superior treatment combinations. Notably, fruit harvested at 45 days after pollination and stored for 30 days, consistently outperformed others, achieving the highest composite index score and thus representing the optimal strategy for seed production. These findings provide robust evidence to guide agronomic practices, contributing to higher seed quality and better yield outcomes. This methodological approach can be extended to similar horticultural crops where multiple correlated traits are influenced by harvest and post-harvest variables.

REFERENCES

Chakravarty, H. L., & Kore, D. K. (1979). Role of seed structure in cucurbit classification. *Journal of Cytology and Genetics*, *14*, 323-329.

Kumar, V., Tomar, B. S., Kaddi, G., & Kumar, S. (2014). Effect of stage of harvest and post-harvest ripening of fruits on hybrid seed yield and quality in pumpkin (*Cucurbita moschata*). *Indian Journal of Agricultural Sciences*, *84*(6), 737-741. https://doi.org/10.56093/ijas.v84i6.41473

Madaan, T. R., & Lal, B. M. (1984). Some studies on the chemical composition of cucurbit kernels and their seedcoats. *Plant Foods for Human Nutrition*, *34*(1), 81-86.

Oehlert, G. W. (2000). *A first course in design and analysis of experiments*. W. H. Freeman.

Ongala, J., Mwanga, D., & Nuani, F. (2016). On the use of principal component analysis in sugarcane clone selection. *Journal of the Indian Society of Agricultural Statistics*, *70*(1), 33-39.

Porter, H. F., & O'Reilly, P. F. (2017). Multivariate simulation framework reveals performance of multi-trait GWAS methods. *Scientific Reports*, *7*, Article 38837. https://doi.org/10.1038/srep38837

Singh, R., Singh, A. P., Mishra, R. K., & Singh, S. N. (2016). Effect of sowing dates and harvesting on fruit setting, fruit traits, seed yield and profitability of cucumber. *Economics Environment & Conservation*, *22*(2), 905-909.

Wellbaum, G. E. (1999). Cucurbit seed development and production. *Horticultural Technology*, *9*(3), 341-348.

(Received : 10.8.2025; Revised : 13.2.2026; Accepted : 20.2.2026)