
Algebraic and Geometric Basis of Principal Components: An Overview

Pramit Pandit*, K. N. Krishnamurthy and K. B. Murthy

*Department of Agricultural Statistics, Applied Mathematics and Computer Science,
University of Agricultural Sciences, Bengaluru, Karnataka, India*

*E-mail: pramitpandit@gmail.com; kkmurthy13@gmail.com;
kkmurthy2005@gmail.com*

**Corresponding Author*

Received 21 May 2019; Accepted 10 June 2020;
Publication 13 October 2020

Abstract

Principal Component Analysis is considered as a dimension-reduction tool which may be used to reduce a large set of possibly correlated variables to hopefully a smaller set of uncorrelated variables that still accounts for most of the variation of the original large set. To understand the inner constructs of principal components, concepts of algebraic as well as geometric basis of principal components are prerequisites. Hence, in the current study, an attempt has been made to provide a step by step and vivid discussion of the basis of principle components and its various important properties.

Keywords: Algebraic basis, basis of principal components, geometric basis, principal components, properties of principal components.

1 Introduction

Principal component analysis (PCA) is a statistical method in which an orthogonal transformation is used to convert a set of observations of possibly correlated variables into a (hopefully, smaller) set of observations of linearly uncorrelated variables, called principal components (Jackson, 1991;

Journal of Reliability and Statistical Studies, Vol. 13, Issue 1 (2020), 73–86.

doi: 10.13052/jrss0974-8024.1314

© 2020 River Publishers

Harris, 2001). The method of principal components was given by Karl Pearson in the year of 1901 (Pearson, 1901), however the general procedure being used nowadays is due to Harold Hotelling, whose pioneering paper showed up later in 1933 (Hotelling, 1933). This transformation is defined such that the first principal component has the largest possible variance and accounts for as much of the variability present in the original data as possible (Anderson, 2003; Johnson and Wichern, 2007). Each succeeding principal component in turn has the next highest possible variance under the constraint that it is orthogonal to the preceding principal component(s) so that the resulting vectors become an uncorrelated orthogonal basis set (Hardle and Simar, 2014). In regression analysis, a test without principal component analysis may be ineffective or even impossible if the number of independent variables is large compared to the number of observations (Timm, 2002). Besides, substantially higher correlations among the independent variables may lead to unstable estimates of regression coefficients (Gujarati et al., 2011). In such cases, these variables can be reduced to a smaller number of principal components resulting in a better test or more stable estimates of regression coefficients (Rencher, 2012). In case of MANOVA, if p (number of dimensions) is close to ν_E (error degrees of freedom) so that a test has a low power, or that, $p > \nu_E$, making the determinants of Wilks' Λ negative, the dependent variables should be substituted with a smaller number of principal components in order to carry out the analysis (Rencher, 2012). In addition to these applications, depending on the fields, it is analogous to the discrete Karhunen–Loève transform in signal processing (Ahmed et al., 1974), proper orthogonal decomposition in mechanical engineering (Chatterjee, 2000), singular value decomposition in linear algebra (Bunch and Nielsen, 1978), Eckart–Young theorem in psychometrics (Johnson, 1963), empirical orthogonal functions in meteorological sciences (Hannachi et al., 2007) and so on. It should be noted that in the term principal components, the adjective ‘principal’ is used to describe the kind of components – main, primary, fundamental, major, and so on. The noun ‘principal’ as a modifier for components, is not used (Rencher, 2012). To understand the inner constructs of principal components, concepts of algebraic as well as geometric basis of principal components are prerequisites. Hence, in the current study, an attempt has been made to provide a step by step and vivid discussion of the basis of principle components and its various important properties.

2 Algebraic Basis of Principal Components

Let \vec{X} be an n-dimensional vector of variables and n is assumed to be very large. Now the interest is emphasized on reducing this n number of variables (may be correlated) into m ($m \leq n$) number of uncorrelated variables (Rencher, 2012).

$$\vec{X} = (X_1, X_2, X_3, \dots, X_m, X_{m+1}, \dots, X_n)$$

In other words, (n-m) number of variables have to be deleted in order to decrease redundancy. Now, if the first m number of variables are kept by eliminating the last (n-m) number of variables, it will contribute sufficient increase in error sum of squares, as here, the condition $V(X_1) \geq V(X_2) \geq \dots \geq V(X_m) \gg V(X_{m+1}) \geq \dots \geq V(X_n)$ is not ascertained. So, to ensure the condition $V(X_1) \geq V(X_2) \geq \dots \geq V(X_m) \gg V(X_{m+1}) \geq \dots \geq V(X_n)$, transformation of \vec{X} vector is obvious. Now, a way of transformation of the \vec{X} vector has to be thought of in such a way that the aforesaid variability condition is ascertained to ensure minimum increase in error sum of squares if the last (n-m) variables are eliminated. In other words, the objective of this transformation is to maximise the decrease of variance so that the last (n-m) number of variables can be easily chopped off.

First of all, let \vec{q} be an n-dimensional unit vector. So, by definition, Euclidian norm of \vec{q} vector, $\|\vec{q}\| = 1$.

Now, a projection of \vec{X} vector onto the n-dimensional unit vector \vec{q} has been made.

\therefore Projection of \vec{X} onto \vec{q} , $C = \vec{X}^T \vec{q} = \vec{q}^T \vec{X}$ (Stark and Yang, 1998), which is subjected to $\|\vec{q}\| = (\vec{q}^T \vec{q})^{1/2} = 1$.

Again, $V(C) = \vec{q}^T S \vec{q}$, where S is the sample variance-covariance matrix of variables of \vec{X} vector. This matrix S is beyond researchers' control. So what can be done is that this n-dimensional unit vector \vec{q} can be utilised as a search to get the desired form.

Let variance probe,

$$\Psi(\vec{q}) = \vec{q}^T S \vec{q} = \sigma^2 \tag{1}$$

Now, as a corollary to Rolle's Mean Value theorem (Riedel and Sahoo, 1998), for a very small change $\vec{\delta q}$, $\Psi(\vec{q})$ can be approximated to $\Psi(\vec{q} + \vec{\delta q})$, i.e.

$$\Psi(\vec{q}) = \Psi(\vec{q} + \vec{\delta q}) \tag{2}$$

$$\begin{aligned}
\text{From (1), } \Psi(\vec{q} + \vec{\delta}q) &= (\vec{q} + \vec{\delta}q)^T S(\vec{q} + \vec{\delta}q) \\
&= \vec{q}^T S\vec{q} + 2\vec{\delta}q^T S\vec{q} + \vec{\delta}q^T S\vec{\delta}q \\
&= \vec{q}^T S\vec{q} + 2\vec{\delta}q^T S\vec{q} \tag{3}
\end{aligned}$$

(As the quantity $\vec{\delta}q^T S\vec{\delta}q$ is infinitely small)
 Putting the value of (1) and (3) in (2),

$$\vec{q}^T S\vec{q} = \vec{q}^T S\vec{q} + 2\vec{\delta}q^T S\vec{q}$$

Or,

$$\vec{\delta}q^T S\vec{q} = 0 \tag{4}$$

Now, as \vec{q} is an unit vector, even after perturbation, $(\vec{q} + \vec{\delta}q)$ also remains as an unit vector.

$$\therefore \left\| (\vec{q} + \vec{\delta}q) \right\| = 1$$

Squaring both sides, it can be obtained that

$$\left\| (\vec{q} + \vec{\delta}q) \right\|^2 = 1$$

Or,

$$(\vec{q} + \vec{\delta}q)^T (\vec{q} + \vec{\delta}q) = 1$$

Or,

$$\vec{q}^T \vec{q} + 2\vec{\delta}q^T \vec{q} + \vec{\delta}q^T \vec{\delta}q = 1$$

Or,

$$\vec{\delta}q^T \vec{q} = 0 \tag{5}$$

(As the quantity $\vec{\delta}q^T \vec{\delta}q$ is infinitely small and $\vec{q}^T \vec{q} = 1$ for \vec{q} being an unit vector).

Now, using Lagrangian multiplier λ (Bertsekas, 2014), from (4) and (5),

$$\vec{\delta}q^T S\vec{q} - \lambda (\vec{\delta}q^T \vec{q}) = 0$$

Or,

$$\vec{\delta}q^T [S\vec{q} - \lambda\vec{q}] = 0$$

As this $\vec{\delta q}^T$ is certainly a non-zero quantity, so

$$[S\vec{q} - \lambda\vec{q}] = 0$$

Or,

$$[S - I\lambda]\vec{q} = 0 \quad (6)$$

where I is nxn identity matrix.

Now, this Equation (6) is a well-known form in linear algebra (Zhang, 2011), more specifically in matrix theory, where λ is the eigen value of S matrix and \vec{q} is the corresponding eigen vector (Searle and Khuri, 2017). As S is a square matrix of order n, there will be n number of eigen values and n corresponding eigen vectors. Arranging eigen values in the decreasing order (i.e. $\lambda_1 \geq \lambda_2 \geq \dots \lambda_m \geq \lambda_{m+1} \geq \dots \lambda_n$), it can be obtained (Johnstone, 2001),

$$S\vec{q}_j = \lambda_j\vec{q}_j \quad (\text{where, } j = 1, 2, \dots, n) \quad (7)$$

Now, two matrices Q and Δ are defined as $Q = [\vec{q}_1, \vec{q}_2, \vec{q}_3, \dots, \vec{q}_n]$ and $\Delta = \text{diag} (\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n)$, respectively. Compacting the n number of Equations (7) in a single equation, it can be obtained,

$$SQ = Q\Delta \quad (8)$$

Again, $\vec{q}_1, \vec{q}_2, \vec{q}_3, \dots, \vec{q}_n$ are eigen vectors of S matrix. As variance-covariance matrix S is a symmetric matrix, its eigen vectors are orthogonal to each other. Hence, matrix Q, consisting of n-orthogonal eigen vectors, is an orthogonal matrix, satisfying

$$\begin{aligned} \vec{q}_j^T \vec{q}_i &= 1, \quad \text{for } i = j \\ &= 0, \quad \text{for } i \neq j \end{aligned}$$

and

$$Q^T Q = I \quad \text{i.e.} \quad Q^T = Q^{-1}.$$

Pre-multiplying Q^T to both the sides of (8),

$$Q^T S Q = \Delta \quad (9)$$

and an expanded form of this Equation (9) will be

$$\begin{aligned} \vec{q}_j^T S \vec{q}_k &= \lambda_j, \quad \text{for } k = j \\ &= 0, \quad \text{for } k \neq j \end{aligned} \quad (10)$$

As the left hand side of the Equation (9) is similar to the Equation (1) as mentioned earlier, Variance-Covariance matrix of transformed vector C will be,

$$(i) \quad \Sigma = \Delta \quad (11)$$

From equality property of matrices (Aitken, 2016), the following properties can be obtained,

$$(ii) \quad \sigma_{ij} = \lambda_i, \quad \text{for } i = j \quad (12)$$

$$= 0, \quad \text{for } i \neq j,$$

$$(iii) \quad \sum_{i=1}^n \sigma_{ij} = \sum_{i=1}^n \lambda_i \quad (13)$$

and

$$(iv) \quad \text{proportion of variability explained by } i^{\text{th}} \text{ PC} = \frac{\lambda_i}{\sum_{i=1}^n \lambda_i} \quad (14)$$

3 Geometric Basis of Principal Component

Suppose for 2 variables X_1 and X_2 , the following scatterplot has been obtained (Figure 1).

Now, from the scatterplot (Figure 1), it can be clearly understood that X_1 and X_2 are highly correlated. It can also be observed that variability across X_1 is higher than variability across X_2 , however both are substantial values (Figure 2).

Keeping the origin rigid, if the axes are rotated θ^0 anti-clockwise (Figure 3), the rotation will yield Z_1, Z_2 as new axes (Eisenhart, 2005).

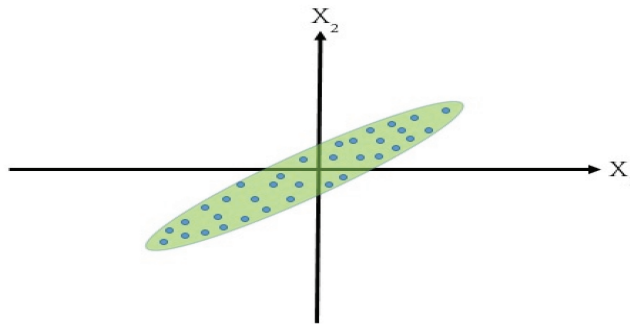


Figure 1 Scatterplot of X_1 and X_2 variable.

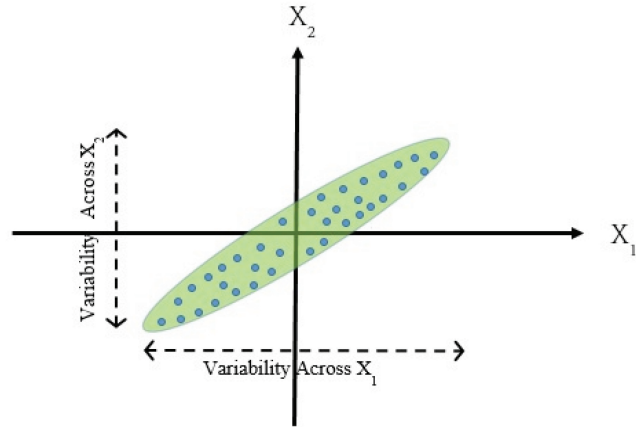


Figure 2 Variability across X_1 and X_2 axis in the scatterplot.

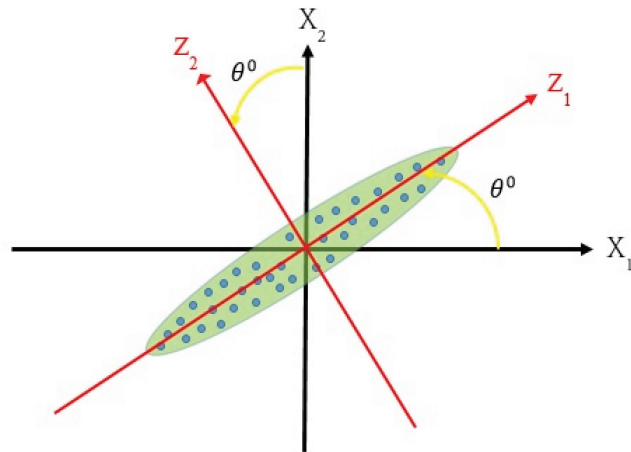


Figure 3 Rotated scatterplot of X_1 and X_2 variable.

After rotation, it can be seen that variability across Z_1 is much higher than Z_2 . In addition to that, Z_1 and Z_2 are observed to be almost independent. Now, if $V(Z_1) \gg V(Z_2)$, then Z_1 dimension alone will be able to provide sufficient information as much as available in the original data set.

Now, a particular point P is considered, whose coordinate is (x_1, x_2) in X_1X_2 plane and (z_1, z_2) in Z_1Z_2 plane (Figure 4). From P , perpendicular lines PM and PA have been drawn on Z_1 axis and Z_2 axis respectively. From P , another perpendicular line PL has been drawn on X_1 axis. From the point

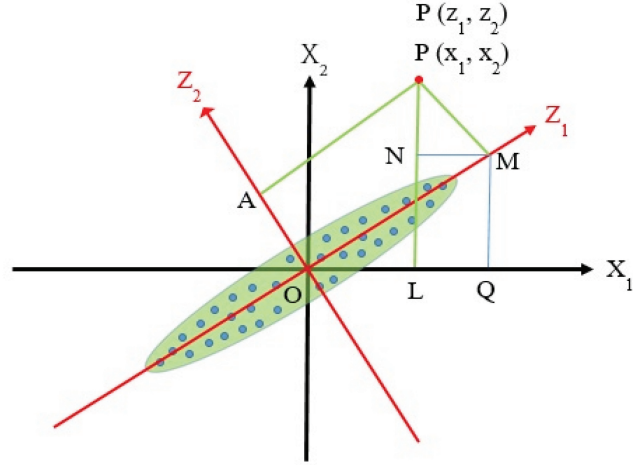


Figure 4 Coordinates of point P with respect to unrotated and rotated axes.

M, two perpendicular line MQ and MN have been drawn on X_1 axis and PL line, respectively.

Now,

$$OL = x_1 \quad PL = x_2$$

$$OM = z_1 \quad PM = z_2$$

Again,

$$\begin{aligned} OL &= OQ - LQ \\ &= OQ - MN \\ &= OM \cos\theta - PM \sin\theta \\ &= z_1 \cos\theta - z_2 \sin\theta \end{aligned} \tag{15}$$

(From $\triangle OMQ$ and $\triangle PMN$ respectively)

and

$$\begin{aligned} PL &= PN + NL \\ &= PN + MQ \\ &= PM \cos\theta + OM \sin\theta \\ &= z_2 \cos\theta + z_1 \sin\theta \end{aligned} \tag{16}$$

(From $\triangle PMN$ and $\triangle OMQ$ respectively)

Solving Equation no. (15) and (16), it can be obtained,

$$\begin{aligned} z_1 &= x_1 \cos\theta + x_2 \sin\theta \\ z_2 &= x_2 \cos\theta - x_1 \sin\theta \end{aligned}$$

In matrix form, which can be rewritten as (Aitken, 2016),

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Or,

$$\vec{z} = A^T \vec{x}$$

Or,

$$\vec{z} = [\vec{a}_1 \vec{a}_2] \vec{x}$$

Where,

$$\vec{a}_1 = \begin{bmatrix} \cos\theta \\ -\sin\theta \end{bmatrix}, \quad \vec{a}_2 = \begin{bmatrix} \sin\theta \\ \cos\theta \end{bmatrix}$$

$$\therefore \|\vec{a}_1\| = 1 \quad \text{and} \quad \|\vec{a}_2\| = 1$$

So, it can be observed that \vec{a}_i 's are unit vectors and $A^T A = I = A A^T$ i.e. transformation matrix A is an orthogonal matrix. Similarly, for n number of variables also, this can be generalised as,

$$Z_j = \sum_{i=1}^n a_{ji} X_i, \text{ subjected to the condition } \vec{a}_j^T \vec{a}_j = 1.$$

As the ultimate objective of this transformation is to maximise the decrease of variance in such a manner that the last few number of variables can be easily chopped off, maximisation of $V(Z_j)$ is required subjected to the condition, $\vec{a}_j^T \vec{a}_j = 1$.

Now,

$$V(Z_j) = \vec{a}_j^T S \vec{a}_j$$

here, S is the sample dispersion matrix of X_1, X_2, \dots, X_n variables.

In other words, maximisation of $\vec{a}_j^T S \vec{a}_j$ is needed to be subjected to the condition,

$$\vec{a}_j^T \vec{a}_j - 1 = 0.$$

Using Lagrangian multiplier, a function has been defined as,

$$L = \vec{a}_j^T S \vec{a}_j - \lambda(\vec{a}_j^T \vec{a}_j - 1)$$

Partially differentiating L with respect to a_j and equating it to zero, it can be obtained,

$$\frac{\delta L}{\delta a_j} = 0$$

Or,

$$2S\vec{a}_j - 2\lambda\vec{a}_j = 0$$

Or,

$$(S - \lambda I)\vec{a}_j = 0 \quad (17)$$

Equation (17) is similar to the Equation (6) mentioned earlier, from which in the same fashion, the same results can be obtained,

- (i) $\sum = \Delta$,
- (ii) $\sigma_{ij} = \lambda_i$, for $i = j$
 $= 0$, for $i \neq j$,

$$(iii) \sum_{i=1}^n \sigma_{ij} = \sum_{i=1}^n \lambda_i$$

and

- (iv) proportion of variability explained by i^{th} PC = $\frac{\lambda_i}{\sum_{i=1}^n \lambda_i}$.

4 Conclusion

In this study, both algebraic and geometric basis of principal components have been discussed thoroughly, which may considerably help in understanding the inner constructs of principal components. From both the approaches, it has been found that eigen values and elements of the corresponding eigen vectors of sample dispersion matrix are the variances and the coefficients of the original variables, respectively, of the corresponding newly formed principal components.

References

- Ahmed, N., Natarajan, T. and Rao, K.R. (1974). Discrete cosine transform, IEEE Transactions on Computers, 23(1), pp. 90–93.
- Aitken, A.C. (2016). Determinants and Matrices, Brousson, Read Books Ltd.

- Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis* (3rd Edition), New Jersey, John Wiley & Sons.
- Bertsekas, D.P. (2014). *Constrained Optimization and Lagrange Multiplier Methods*, New York, Academic Press Inc.
- Bunch, J.R. and Nielsen, C.P. (1978). Updating the singular value decomposition, *Numerische Mathematik*, 31(2), pp. 111–129.
- Chatterjee, A. (2000). An introduction to the proper orthogonal decomposition, *Current Science*, 78(7), pp. 808–817.
- Eisenhart, L.P. (2005). *Coordinate Geometry, USA*, Dover Publications Inc.
- Gujarati, D.N., Porter, D.C. and Gunasekar, S. (2011). *Basic Econometrics*, New York, McGraw-Hill.
- Hannachi, A., Jolliffe, I.T. and Stephenson, D.B. (2007). Empirical orthogonal functions and related techniques in atmospheric science: a review, *International Journal of Climatology*, 27, pp. 1119–1152.
- Hardle, W.K. and Simar L. (2014). *Applied Multivariate Statistical Analysis* (4th Edition), New York, Springer-Verlag New York, Inc.
- Harris, R.J. (2001). *A Primer of Multivariate Statistics* (3rd Edition). Mahwah, Lawrence Erlbaum Associates, Inc., Publishers.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology*, 24, pp. 417–441.
- Jackson, J.E. (1991). *A User's Guide to Principal Components*, New Jersey, John Wiley & Sons, Inc.
- Johnson, R.A. and Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis*. Upper Saddle River, New Jersey, Pearson Education, Inc.
- Johnson, R.M. (1963). On a theorem stated by Eckart and Young, *Psychometrika*, 28(3), pp. 259–263.
- Johnstone, I.M. (2001). On the distribution of the largest eigenvalue in principal components analysis, *The Annals of Statistics*, 29(2), pp. 295–327.
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space, *Philosophical Magazines*, 2, pp. 559–572.
- Rencher, A.C. (2012). *Methods of Multivariate Analysis* (3rd Edition), New Jersey, John Wiley & Sons.
- Riedel, T. and Sahoo, P.K. (1998). *Mean Value Theorems and Functional Equations*, Singapore, World Scientific Publishing Co. Pte. Ltd.
- Searle, S.R. and Khuri, A.I. (2017). *Matrix Algebra Useful for Statistics* (2nd Edition), New Jersey, John Wiley & Sons, Inc.

- Stark, H. and Yang, Y. (1998). *Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets, and optics*, New York, John Wiley & Sons, Inc.
- Timm, N.H. (2002). *Applied Multivariate Analysis*. New York, Springer-Verlag New York, Inc.
- Zhang, F. (2011). *Matrix Theory: Basic Results and Techniques* (2nd Edition), New York, Springer Science & Business Media.

Biographies



Pramit Pandit obtained his Bachelor's degree (Hons.) in Agriculture from Uttar Banga Krishi Viswavidyalaya and Master's (Ag.) majoring in Agricultural Statistics from University of Agricultural Sciences, Bengaluru. He was the recipient of ICAR-Junior Research Fellowship during his Master's degree programme for securing All India basis 2nd rank in AIEEA-UG-2016 examination in Statistical Sciences. He also secured All India basis 1st rank in AICE-JRF/SRF(PGS)-2018 in Agricultural Statistics and qualified ICAR-NET-2018. He was awarded with the prestigious UAS Gold Medal 2019 along with the Professor G. Gurusurthy Memorial Gold Medal, Sri Godabanahal Thuppamma Basappa Mallikarjuna Gold Medal and Sri Nijalingappa's 77th Birthday Commemoration Gold Medal for his exemplary academic excellence. He was also the recipient of best 'Best M.Sc. Thesis Award' for his research work on, 'Statistical Models for Insect Count Data On Rice', conducted under the supervision of Prof. K. N. Krishnamurthy.



K. N. Krishnamurthy received his B.Sc., M.Sc. and M.Phil. degrees in Statistics from Bangalore University and Ph.D. degree in Statistics from Himalayan University. Prof. Krishnamurthy is a recipient of the Best Teacher award from the National Institute for Education & Research, New Delhi during 2017. He is currently working as Head of the department as well as University Head of the Department of Agricultural Statistics, Applied Mathematics & Computer Science, University of Agricultural Sciences, GKVK, Bengaluru. He has 38 years of experience in teaching at the University.



K. B. Murthy received his B.Sc. and M.Sc. degrees in Mathematics from Mysore University and Ph.D. degree in Mathematics from Himalayan University. He has specialised in graph theory and applied sciences and has a teaching experience of over 25 years. He has published many papers in National and International Journals and also participated in many International conferences. He is a member of Board of Studies of University of Agricultural Sciences, GKVK, Bengaluru. Dr. K. B. Murthy is the recipient of the GAURAVACHARYA-2017 award for the significant contribution in the field of graph theory and education from National Institute for Education and Research, New Delhi.

