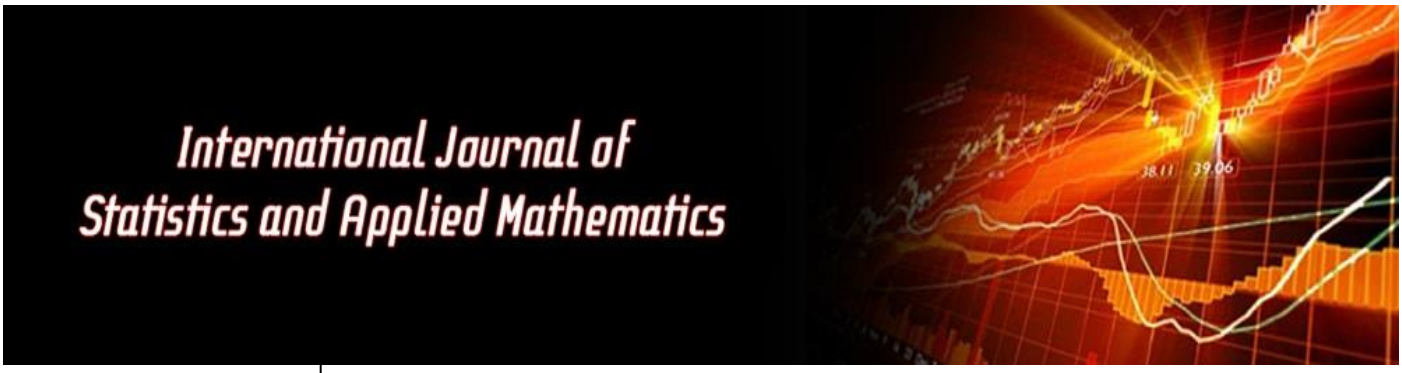

ISSN: 2456-1452
Maths 2022; 7(4): 15-20

www.mathsjournal.com
Received: 17-06-2022
Accepted: 22-07-2022

Dr. G Madhu Sudan
Assistant Professor, Department of Statistics, University of Allahabad, Prayagraj, Uttar Pradesh, India

Dr. B Hari Mallikarjuna Reddy
Assistant Professor, Department of Maths and Statistics, YSR Horticulture University, Andhra Pradesh, India

Neeraja Sesha Sai Bandreddi
Data Scientist, Liftlab, Hyderabad, Telangana, India

Corresponding Author:
Dr. G Madhu Sudan
Assistant Professor, Department of Statistics, University of Allahabad, Prayagraj, Uttar Pradesh, India

Estimate the principle components of dispersion matrix of the vector and their variances by R-programing

Dr. G Madhu Sudan, Dr. B Hari Mallikarjuna Reddy and Neeraja Sesha Sai Bandreddi

DOI: https://doi.org/10.22271/maths.2022.v7.i5a.875

Abstract
In the present research article, an attempt has been made by developing a simple R-Programming for estimating Principal Components of observation vector with their Variances. It is very hazardous task to calculate and interpret the dispersion matrix manually, by reducing the dimensions to the few principal components that explains main patterns. Hence the powerful software program "R" is applied here.

Keywords: Principal component analysis, vectors, variances, R-programming

Introduction
Principal Component Analysis (PCA) was invented in 1901 by Karl Pearson. Principal Component Analysis is a Statistical method for making sense of datasets with large number of measurements (which can be thought of as dimensions) by reducing the dimensions to the few principal components that explains main patterns. PCA involves a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called Principal Components.
In Population Genetics, PCA can be used to identify differences in ancestry among Population and Sample. This paper provides the usefulness of the principal components analysis method in agricultural, mainly estimating Principal Components of observation vector with their Variances it takes into account in our analysis.

Methodology
To obtain the estimate of Principal Component Analysis of 'X' and their variances for given data.
We have to apply the following steps

1. Find the column total for the given matrix A1
2. Divide each total by the maximum total
3. Write the values obtained in step-2 in a vector form adjacent to the matrix A let this vector be a1
4. Find the new matrix by multiplying the elements in each row A1 by the corresponding element in the vector a1
5. Repeat steps 1,2,3,4 on the new matrix and obtain another new matrix.

Continue this process till two consecutive vectors of type a1 are identical. Let us denote this vector by c1

6. Find $c_1^* = \frac{1}{length\ of\ c_1} \times C_1$
7. The first principal component is given by $C_1^{*1} X$
8. Find the largest root f1 given by $f_1 = \frac{first\ row\ of\ A_1 * C_1^*}{first\ element\ of\ C_1^*}$
9. Find the matrix f_1, C_1^*, C_1^{*1}

10. Find the matrix $A_2 = A_1 - f_1 C_1^* C_1^{*1}$
11. Repeat this process from 1 to 10 and find second principal component and then find $A_3 = A_2 - f_1 C_1^* C_1^{*1}$
12. Again repeat this process from steps 1 to 10 and find the third principal component C_3^{*1} And largest root f3

Simplifications

we have to simplify $A_1 = \begin{bmatrix} 34.01 & 10.50 & 1.77 \\ 10.50 & 23.01 & 3.43 \\ 1.77 & 3.43 & 4.59 \end{bmatrix}$ is the unbiased estimator of the dispersion matrix of the vector $X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$ based on 95 dof , we have to obtain the estimate of Principal Component Analysis of ' X' and their variances for given data.

First iteration:

					a_1
1*st iteration*		34.01	10.50	1.77	1.0000
		10.50	23.01	3.43	0.7982
		1.77	3.43	4.59	0.2112
	Totals	46.28	36.94	9.79	

Second iteration:

					a_2
2*nd iteration*		34.01	10.50	1.77	1.0000
		8.38	18.36	2.73	0.6919
		0.37	0.72	0.96	0.1280
	Totals	42.76	29.58	5.47	

Third iteration:

					a_3
3*rd iteration*		34.01	10.50	1.77	1.0000
		5.79	12.70	1.89	0.5846
		0.04	0.09	0.12	0.0950
	Totals	39.85	23.30	3.78	

Fourth iteration:

					a_4
4*th iteration*		34.01	10.50	1.77	1.0000
		3.39	7.42	1.10	0.4796
		0.004	0.008	0.11	0.0772
	Totals	37.40	17.93	2.88	

Fifth iteration:

					a_5
5*th iteration*		34.01	10.50	1.77	1.0000
		1.62	3.50	0.53	0.3946
		0.0043	0.0006	0.0009	0.0646
	Totals	35.63	14.06	2.30	

Sixth iteration:

					a_6
6*th iteration*		34.01	10.50	1.77	1.0000
		0.64	1.40	0.20	0.3436
		0.00002	0.00004	0.00006	0.0571
	Totals	34.65	11.90	1.97	

Seventh iteration:

					a_7
7*th iteration*		34.01	10.50	1.77	1.0000
		0.22	0.48	0.07	0.3209
		0.00001	0.00002	0.00003	0.0538
	Totals	34.23	10.98	1.84	

Eighth iteration:

					a_8
8*th iteration*		34.01	10.50	1.77	1.0000
		0.07	0.15	0.02	0.3126
		–	–	–	0.0526
	Totals	34.08	10.65	1.74	

Ninth iteration:

$$\begin{bmatrix} 9th\ iteration & 34.01 & 10.50 & 1.77 \\ & 0.02 & 0.04 & 0.007 \\ & – & – & – \\ Totals & 34.03 & 10.54 & 1.77 \end{bmatrix} \overset{a_9}{\begin{bmatrix} 1.0000 \\ 0.3099 \\ 0.0522 \end{bmatrix}}$$

Tenth iteration:

$$\begin{bmatrix} 10th\ iteration & 34.01 & 10.50 & 1.77 \\ & 0.005 & 0.015 & 0.002 \\ & – & – & – \\ Totals & 34.00 & 10.51 & 1.77 \end{bmatrix} \overset{a_{10}}{\begin{bmatrix} 1.0000 \\ 0.3091 \\ 0.0521 \end{bmatrix}}$$

Eleventh iteration:

$$\begin{bmatrix} 11th\ iteration & 34.01 & 10.50 & 1.77 \\ & 0.0021 & 0.0046 & 0.0006 \\ & – & – & – \\ Totals & 34.01 & 10.50 & 1.77 \end{bmatrix} \overset{a_{11}}{\begin{bmatrix} 1.0000 \\ 0.3088 \\ 0.0520 \end{bmatrix}}$$

Twelfth iteration:

$$\begin{bmatrix} 12th\ iteration & 34.01 & 10.50 & 1.77 \\ & 0.0065 & 0.0004 & 0.0006 \\ & – & – & – \\ Totals & 34.01 & 10.50 & 1.77 \end{bmatrix} \overset{a_{12}}{\begin{bmatrix} 1.0000 \\ 0.3088 \\ 0.0520 \end{bmatrix}}$$

$$C_1 = \begin{bmatrix} 1.0000 \\ 0.3088 \\ 0.0520 \end{bmatrix} \quad l_1 = \sqrt{1.000 + 0.0953 + 0.0027} = 1.0478$$

$$c_1^* = \frac{1}{1.0478} \begin{bmatrix} 1.0000 \\ 0.3088 \\ 0.0520 \end{bmatrix} = \begin{bmatrix} 0.9543 \\ 0.2946 \\ 0.4966 \end{bmatrix}$$

$$f_1 = \frac{\begin{bmatrix} 34.01 & 10.50 & 1.77 \end{bmatrix} \begin{bmatrix} 0.9543 \\ 0.2946 \\ 0.0496 \end{bmatrix}}{0.9543}$$

$$f_1 = \frac{35.6381}{0.9543} = 37.3445$$

$$c_1^{*1} = \begin{bmatrix} 0.9543 & 0.2946 & 0.0496 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$c_1^{*1} = 0.9543x_1 + 0.2946x_2 + 0.0496x_3$$

$$\therefore f c_1^* c_1^{*1} = \begin{bmatrix} 37.3445 \end{bmatrix} \begin{bmatrix} 0.9543 \\ 0.2946 \\ 0.0496 \end{bmatrix} \begin{bmatrix} 0.9543 & 0.2946 & 0.0496 \end{bmatrix} = \begin{bmatrix} 34.0089 & 10.5035 & 1.7698 \\ 10.5035 & 3.2432 & 0.5471 \\ 1.7698 & 0.5471 & 0.0972 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 34.01 & 10.50 & 1.77 \\ 10.50 & 23.01 & 3.43 \\ 1.77 & 3.43 & 4.59 \end{bmatrix} - \begin{bmatrix} 34.00 & 10.50 & 1.76 \\ 10.50 & 3.23 & 0.54 \\ 1.76 & 0.54 & 0.09 \end{bmatrix} = \begin{bmatrix} 0.011 & -0.0035 & 0.002 \\ -0.0035 & 19.7666 & 2.8829 \\ 0.0002 & 2.8829 & 4.4978 \end{bmatrix}$$

By the way of same Simplifications through iterations we will get

$$C_2 = \begin{bmatrix} -0.0001 \\ 1.0000 \\ 0.1459 \end{bmatrix} l_2 = 1.0158$$

$$c_2^* = \frac{1}{1.0105}\begin{bmatrix} -0.0001 \\ 1.0000 \\ 0.1459 \end{bmatrix} = \begin{bmatrix} -0.0001 \\ 0.9895 \\ 0.1443 \end{bmatrix}$$

$$f_2 = \begin{bmatrix} 34.01 & 10.50 & 1.77 \end{bmatrix} \frac{1}{1.0105}\begin{bmatrix} -0.0001 \\ 1.0000 \\ 0.1459 \end{bmatrix} = 19.7216$$

$$c_2^{*1} = -0.0001x_1 + 0.9845x_2 + 0.1443x_3$$

$$A_3 = \begin{bmatrix} 0.011 & -0.0035 & -0.002 \\ -0.0035 & 19.7666 & 2.8829 \\ 0.0002 & 2.8829 & 4.4978 \end{bmatrix} - f_2 c_2^* c_2^{*1}$$

$$A_3 = \begin{bmatrix} 0.0010 & -0.0001 & -0.0002 \\ -0.0010 & 0.4569 & 0.0655 \\ 0.0002 & 0.0655 & 4.0867 \end{bmatrix}$$

$$C_3 = \begin{bmatrix} 0.00007 \\ 1.00165 \\ 1.00000 \end{bmatrix} l_3 = 1.0000$$

$$c_3^* = \frac{1}{1.0000}\begin{bmatrix} 0.000072 \\ 0.001650 \\ 1.000000 \end{bmatrix} = \begin{bmatrix} -0.0001 \\ 0.9895 \\ 0.1443 \end{bmatrix}$$

$$f_3 = \begin{bmatrix} 0.001099 & -0.000102 & 0.000296 \end{bmatrix} \frac{1}{0.000072}\begin{bmatrix} 0.000072 \\ 0.001650 \\ 1.000000 \end{bmatrix} = 4.1096$$

$$c_3^{*1} = 0.000072x_1 + 0.001650x_2 + 1.0000x_3$$

$$\therefore c_1^{*1} = 0.9543x_1 + 0.2946x_2 + 0.0496x_3$$
$$c_2^{*1} = -0.0001x_1 + 0.9845x_2 + 0.1443x_3$$
$$c_3^{*1} = 0.000072x_1 + 0.001650x_2 + 1.0000x_3$$

The dispersion matrix is

$$\begin{bmatrix} 37.3445 & 0 & 0 \\ 0 & 19.7216 & 0 \\ 0 & 0 & 4.1098 \end{bmatrix}$$

R-Programing

```
Principal Component Analysis <- function(input Matrix)
{
# This function takes matrix as input and gives the pca variances as output.
mainMatrix <- inputMatrix; j <- nrow(mainMatrix) - (nrow(mainMatrix)-1)
adjMatrix <- as.matrix(colSums(mainMatrix) / max(colSums(mainMatrix)))
adjMatrix1 <- matrix (nrow = nrow(adjMatrix), ncol = ncol(adjMatrix))
adjMatrix1[is.na(adjMatrix1)] <- 0; output <- NULL
while(nrow(mainMatrix) >= j)
{
mainMatrix1 <- mainMatrix
while(all(round(adjMatrix1,5) == round(adjMatrix,5)) == FALSE)
{
adjMatrix1 <- round(adjMatrix,5)
mainMatrix1 <- round(mainMatrix1,5)
for (i in 1:nrow(mainMatrix1))
{
mainMatrix1[i,] <- mainMatrix1[i,] * adjMatrix1[i]
}
a11 <- round (as.matrix(colSums(round(ifelse(mainMatrix1 < 0, 0,mainMatrix1),5)) /
max (colSums(round(ifelse(mainMatrix1 < 0, 0,mainMatrix1),5)))),5)
adjMatrix <- round(as.matrix(colSums(round(mainMatrix1,5)) /
max(colSums(round(mainMatrix1,5)))),5)
}
p1 <- (nrow(mainMatrix) - (nrow(mainMatrix)-1))/sqrt(sum(adjMatrix^2)) * adj Matrix
p2 <- as.numeric (main Matrix [(nrow (main Matrix) - (nrow (main Matrix)-1)),] %*% p1)/p1[1]
output <- rbind(output,data.frame(t(p1),variance = p2, row.names = paste0("equn",j)))
p3 <- p2 * (p1 %*% t(p1)); mainMatrix <- mainMatrix - round(p3,5)
adjMatrix <- round(as.matrix (col Sums (main Matrix) / max(colSums(mainMatrix))),5)
j <- j + nrow (main Matrix) - (nrow(mainMatrix)-1)
}
return(round(output,4))
}
```

$$\begin{bmatrix} 37.3445 & 0 & 0 \\ 0 & 19.7216 & 0 \\ 0 & 0 & 4.1098 \end{bmatrix}$$

Conclusions

The variance-covariance matrix is a convenient expression of statistics in data describing patterns of variability and covariation. The variance-covariance matrix is widely used both as a summary statistic of data and as the basis for key concepts in many multivariate statistical models. Principal Component Analysis (PCA) is used to explain the variance-covariance structure of a set of variables through linear combinations. It is often used as a dimensionality-reduction technique. Finally we conclude that manually calculations are difficult that is the reason we approached statistical computing by R-Programming is the best and simple technique. The output as follows

The Principle Components that we have obtained are

$$\therefore c_1^{*1} = 0.9543x_1 + 0.2946x_2 + 0.0496x_3$$
$$c_2^{*1} = -0.0001x_1 + 0.9845x_2 + 0.1443x_3$$
$$c_3^{*1} = 0.000072x_1 + 0.001650x_2 + 1.0000x_3$$

The dispersion matrix is $\begin{bmatrix} 37.3445 & 0 & 0 \\ 0 & 19.7216 & 0 \\ 0 & 0 & 4.1098 \end{bmatrix}$

References

1. Cromwell JB, Hannan MJ, Lbys WC, Terraza M. Multivariate Tests for time series models. Danial Straumann, Estimation in conditionally Heteroscedastic Time Series Models; c1994.
2. Engle Robert F, Tin Bollersler. Modelling the Persistence of Conditional Variances. Econometrica Reviews. 1986;5:1-50.
3. Engle Robert F, Kenneth F Kronel. Multivariate Simultaneous Generalized ARCH USCD Mimeo; c1983.
4. Helmy AK, Taweel GHS. Authentication Scheme Based on Principal Component Analysis for Satellite Images. Int. J. Signal Processing, Image Processing and Pattern Recognition. 2009;2(3):1-14.
5. Principal Component Analysis (PCA), http://www.uga.edu/strata/software/pdf/ pcaTutorial.pdf, accessed; c2012.
6. SPSS (Statistical Package for the Social Sciences), 2012. p. 7. http://www.spss.com/accessed. http://www.profs.info.uaic.ro.
7. Madhusudan G, *et al*. Statistical Method of minimum Chi-Squareestimate with gene frequencies in Blood Group Systems Andhra Agriculture Journal (AAJ), The Andhra Agric. J. 2019;(Spl):42-45. ISSN NO:003-2950