International Journal of Electronics Engineering (ISSN: 0973-7383
Volume 11 • Issue 1 pp. 380-385 Jan 2019-June 2019 www.csjournals.com

Principal component analysis to reduce youth suicidal crime rate with different causes

Siddiqui Bushra Zareen[1], Dr. Mukta Dhopeshwarkar[2]
Research Scholar[1], Research Guide[2]
Dr. Babasaheb Ambedkar Marathwada, University Aurangabad

Abstract: This paper mainly deals with the causes of youth suicide crime by using Principal Component Analysis (PCA). PCA is a statistical approach used for reducing the number of variables in causes of suicide. In PCA, every cause in the training set is represented as a linear combination of weighted eigenvectors called eigenvalues. These eigenvectors are obtained from covariance matrix of a training data set. The weights are found out after selecting a set of most relevant Eigenvalues. Various test is performed by projecting a different causes onto the subspace spanned by the eigenvalues and then classification is done by measuring minimum Euclidean distance. A number of experiments were done to evaluate the youth suicide for different causes. In this thesis, we used a training database of youth suicide for different states with their causes with gender and agewise.
Keywords:- PCA, Eigenvector, Co-relation Metrics.

Introduction:-

Over the last decades or so, youth suicide has become a popular area of research in computer vision and one of the most successful applications of suicide analysis and understanding. Because of the nature of the problem, not only computer science researchers are interested in it, but neuroscientists and psychologists also. It is the general opinion that advances in computer vision research will provide useful insights to neuroscientists and psychologists into how human brain works, and vice versa [1]. Thegoalisto implement the system (model) for a particular cause of suicide and distinguish it from a various causes with different states and to find out due to which cause in that particular state the youth are committing suicide. It gives us efficient way to find the lower dimensional space. Further this can be extended to recognize the causes of a youth suicide and that causes which relates with emotional intelligence. There are different 26 causes responsible for youth suicide and with that 26 causes we have only worked on sentimental causes. Principal component analysis has been done which finds that in which state the youth suicidal rate is increasing due to which cause.

I. PROCESS

One of the simplest and most effective PCA approachesused in youth suicidal crime systems is the so-called eigenvector approach. This approach transforms different causes into a small set of essential characteristics, eigenvectors, which are the main componentsoftheinitialset. PCA is a statistical approach to find the principal features of a distributed data set based on the total variance.

Principal components method

In principal components analysis.first finds the set of orthogonal eigenvectors of the correlation or covariance matrix of the variables. The matrix of principal components is the product of the eigenvector matrix with the matrix of independent variables. The first principal component accounts for the largest percent of the total data variation. The second principal component accounts the second largest percent of the total data variation, and so on. The goal of principal components is to explain the maximum amount of variance with the fewest number of components.

Non-uniqueness of coefficients

The coefficients for the principal components are unique (except for a change in sign) if the eigenvalues are distinct and not zero. If an eigenvalue is repeated, then the "space spanned" by all the

principal component vectors corresponding to the same eigenvalue is unique, but the individual vectors are not. Therefore, the coefficients that Minitab displays in output and those in a book or another program may not agree, although the eigenvalues (variances of the components) will always be the same.

If the covariance matrix has rank r < p, where p is the number of variables, then there will be p – r eigenvalues equal to zero. Eigenvectors corresponding to these eigenvalues may not be unique. This can happen if the number of observations is less than p or if there is multi-collinearity.

Eigenvectors

Eigenvectors, which are comprised of coefficients corresponding to each variable, are the weights for each variable used to calculate the principal components scores. The eigenvectors are obtained as the columns of the orthogonal matrix in the spectral decomposition of the covariance or correlation matrix, **S** or **R**. More specifically, because **R** is symmetric, an orthogonal matrix **V** exists such that **V'RV** = **D** or, equivalently, **R** = **VDV'**, where **D** is a diagonal matrix whose diagonal elements are the eigenvalues. The eigenvectors are the columns of **V**.The eigenvectors originate from **R** = **V L V'**.

Term	Description
R	correlation matrix
V	eigenvector matrix
D	diagonal matrix of eigenvalues

Scores

Formula

Scores are the linear combinations of the original variables that account for the variance in the data. The scores are calculated as follows: **Z** = **YV**

Term	Description
Z	matrix of principal components scores (n × m)
Y	standardized data matrix (n × p) used with the correlation matrix method
V	matrix of eigenvectors (p × m)

Eigenvalue

Formula

The eigenvalues are the diagonal elements of the diagonal matrix in the spectral decomposition of the covariance or correlation matrix (see the topic "Eigenvectors"). The eigenvalues also represent the sample variances of the principal components **Z** = **V Y**.

Term	Description
Z	matrix of principal components scores (n × m)
Y	standardized data matrix (n × p) used with the correlation matrix method
V	matrix of eigenvectors (p × m)

Proportion

Formula

The proportion of sample variance explained by the k^{th} principal component is calculated as follows:

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p}$$

Term	Description
λ_k	the k^{th} eigenvalue
p	the number of variables

Cumulative proportion

Formula

The cumulative proportion of sample variance explained by the first k principal components is calculated as follows:

$$\frac{\lambda_1 + \lambda_2 + \ldots + \lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p}$$

Term	Description
λ_k	the k^{th} eigenvalue
p	the number of variables

Mahalanobis distance

Formula

The Mahalanobis distance measures the distance from each point in multivariate space to the overall mean or centroid, utilizing the covariance structure of the data.

$$Y_i = \sqrt{((Y_i - \overline{Y})S^{-1}(Y_i - \overline{Y}))}$$

Minitab displays a reference line on the outlier plot to identify outliers with large Mahalanobis distance values. The reference line is defined by the following formula:

$$\sqrt{p \times F \text{ inverse CDF}(.95, p, n\text{-}p\text{-}1)}$$

When n – p – 1 is $\leq$ 0, Minitab displays the outlier plot without the reference line.

Term	Description
Y_i	data value vector at row i
$\overline{Y}$	mean vector
S^{-1}	inverse of the covariance matrix
p	the number of variables
n	the number of nonmissing row

Results and Discussions:-

The principal component method is done on Maharashtra for 12 years of data in which only the sentimental causes are selected for which we have normalised the data and had performed PCA for male and female with sentimental causes.

Example:-

PCA of Maharashtra 2001 (Male):-

Principal Component Analysis:

Eigenanalysis of the Correlation Matrix

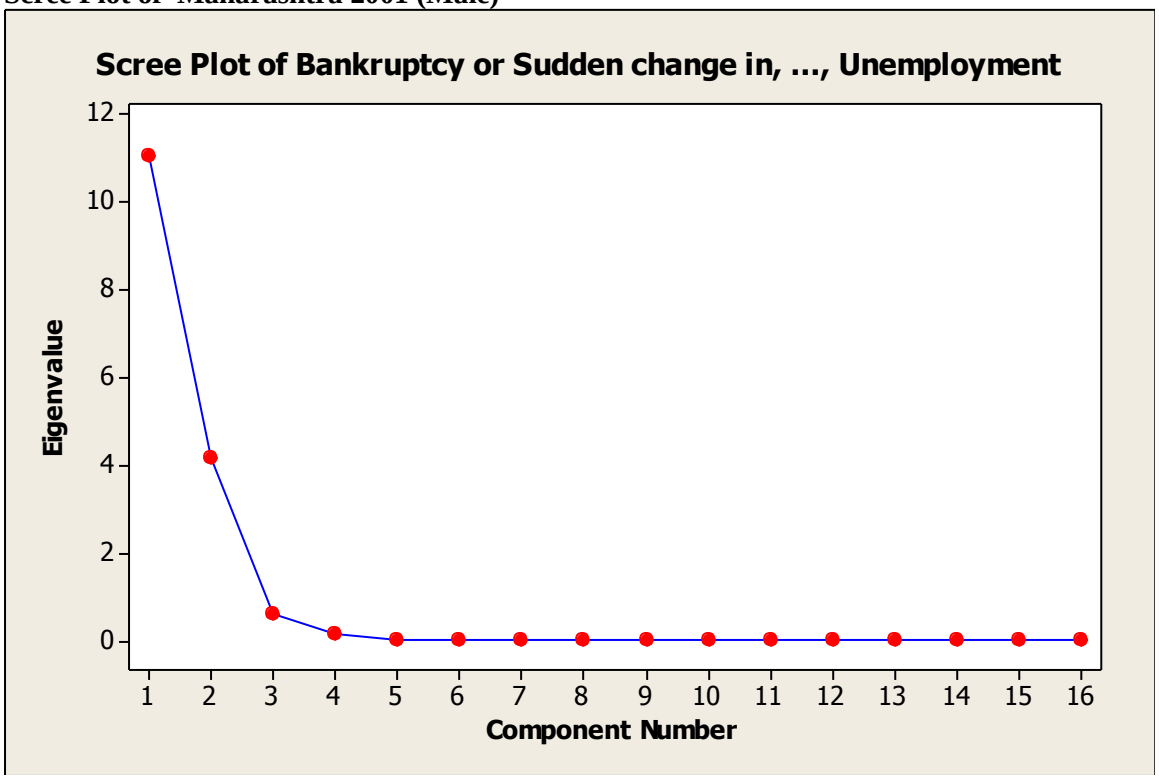
```
Eigenvalue   11.037   4.167   0.633   0.163   0.000   0.000   0.000   0.000
0.000
Proportion    0.690   0.260   0.040   0.010   0.000   0.000   0.000   0.000
0.000
Cumulative    0.690   0.950   0.990   1.000   1.000   1.000   1.000   1.000
1.000

Eigenvalue   0.000  -0.000  -0.000  -0.000  -0.000  -0.000  -0.000
```

Proportion	0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
Cumulative	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Variable	PC1	PC2
Bankruptcy or Sudden change in	0.268	0.210
Cancellation/Non-Settlement of	0.284	-0.145
Illness (Aids/STD)	0.295	-0.082
Cancer	0.201	0.364
Paralysis	0.041	0.377
Insanity/Mental Illness	0.296	0.079
Dowry Dispute	0.153	-0.401
Drug Abuse/Addiction	0.271	0.210
Failure in Examination	0.150	-0.414
Family Problems	0.300	0.028
Love Affairs	0.247	-0.279
Physical Abuse (Rape,IncestEtc	0.185	-0.344
Poverty	0.287	0.132
Professional/Career Problem	0.292	0.015
Property Dispute	0.266	0.228
Unemployment	0.295	-0.051

In these results, the first two principal components have eigenvalues greater than 1. The first three components explain 99.0% of the variation in the data.

Scree Plot of Maharashtra 2001 (Male)



PCA of Maharashtra 2001 (Female):-

Principal Component Analysis:

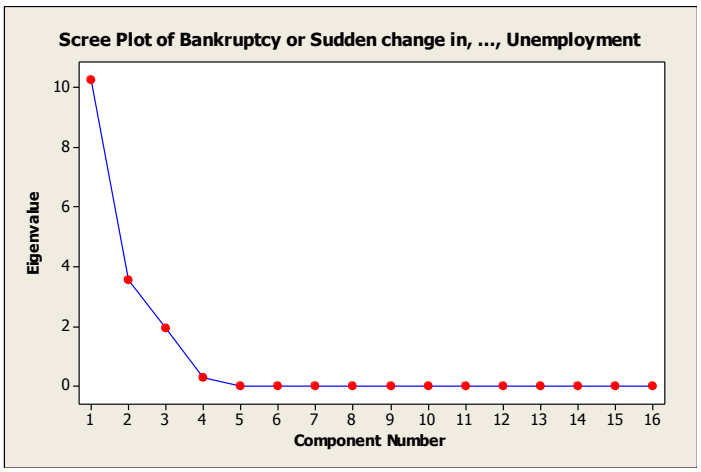
Eigenanalysis of the Correlation Matrix

```
Eigenvalue   10.239   3.556   1.939   0.266   0.000   0.000   0.000   0.000
-0.000
Proportion    0.640   0.222   0.121   0.017   0.000   0.000   0.000   0.000
-0.000
Cumulative    0.640   0.862   0.983   1.000   1.000   1.000   1.000   1.000
1.000

Eigenvalue  -0.000  -0.000  -0.000  -0.000  -0.000  -0.000  -0.000
Proportion  -0.000  -0.000  -0.000  -0.000  -0.000  -0.000  -0.000
Cumulative   1.000   1.000   1.000   1.000   1.000   1.000   1.000
```

In these results, the first three principal components have eigenvalues greater than 1. These first three components explain 98.3% of the variation in the data.

Variable	PC1	PC2
Bankruptcy or Sudden change in	0.224	0.336
Cancellation/Non-Settlement of	0.261	-0.258
Illness (Aids/STD)	0.291	-0.083
Cancer	0.287	0.041
Paralysis	0.096	0.379
Insanity/Mental Illness	0.307	0.037
Dowry Dispute	0.263	-0.269
Drug Abuse/Addiction	0.177	0.267
Failure in Examination	0.203	-0.363
Family Problems	0.305	-0.091
Love Affairs	0.248	-0.290
Physical Abuse (Rape,IncestEtc	0.287	-0.207
Poverty	0.299	0.129
Professional/Career Problem	0.208	0.274
Property Dispute	0.231	0.155
Unemployment	0.217	0.375

Scree Plot of Maharashtra 2001 (Female)



Conclusion:-

It is general belief in the minds of people that over the years suicide rate is increasing year by year. This has been tested in the study of Principal component analysis. It is observed that irrespective of different causes the suicide rate is increasing. For Maharashtra 2001for male the first two principal components have eigenvalues greater than 1. The first three components explain 99.0% of the variation in the data. The scree plot shows that the eigenvalues start to form a straight line after the two principal component. If 99.0% is an adequate amount of variation explained in the data, then you should use the first two principal components. For female the first three principal components have eigenvalues greater than 1. The first three components explain 99.0% of the variation in the data. The scree plot shows that the eigenvalues start to form a straight line after the three principal component. If 98.3% is an adequate amount of variation explained in the data, then you should use the first three principal components.

References:-

1. B.Moore,"Principalcomponentanalysisinlinearsystems:Controllability,observability,andmodelreduction,"*IEE ETransactionsonAutomaticControl*,vol.26,no.1,17-32,1981
2. J.Tzeng,"Split-and-combinesingularvaluedecompositionforlarge-scalematrix,"*JournalofAppliedMathematics*,2013
3. G.Tzimiropoulos,S.Zafeiriou,andM.Pantic,"Principalcomponentanalysisofimagegradientorientationsforfacer ecognition,"*2011IEEEInternationalConferenceonAutomaticFace&GestureRecognitionandWorkshops*,Santa Barbara,CA,2011,pp.553-558.
4. Li,H.,"Accurateandefficientclassificationbasedoncommonprincipalcomponentsanalysisformultivariatetimeser ies,"*Neurocomputing*,vol.171,pp.744-753,2016.
5. P.Kamencay,T.Trnovszky,M.Benco,R.Hudec,P.SykoraandA.Satnik,"AccuratewildanimalrecognitionusingPC A,LDAandLBPH,"*2016ELEKTRO*,StrbskePleso,2016,pp.62-67.
6. R.D.Santo,"Principalcomponentanalysisappliedtodigitalimagecompression,"*Einstein*,vol.10,no.2,pp.135-139,2012.
7. Q.DuandJamesE.Fowler,"Low-complexityprincipalcomponentanalysisforhyperspectralimagecompression,"*InternationalJournalofHighPerf ormanceComputingApplications*,vol.22,no.4,pp.438–448,2002
8. C.Lee,S.Youn,T.Jeong,E.LeeandJ.Serra-Sagristà,"HybridCompressionofHyperspectralImagesBasedonPCAWithPre-EncodingDiscriminantInformation,"in*IEEEGeoscienceandRemoteSensingLetters*,vol.12,no.7,pp.1491-1495,July2015.
9. C.W.WangandJ.H.Jeng,"ImagecompressionusingPCAwithclustering,"*2012InternationalSymposiumonIntelli gentSignalProcessingandCommunicationsSystems(ISPACS)*,NewTaipei,2012,pp.458-462.
10. A.VaishandM.Kumar,"WDRcodingbasedimagecompressiontechniqueusingPCA,"*2015InternationalConfere nceonSignalProcessingandCommunication(ICSC)*,Noida,2015,pp.360-365
11. S.T.Lim,D.F.W.YapandN.A.Manap,"Medicalimagecompressionusingblock-basedPCAalgorithm,"*Computer,Communications,andControlTechnology(I4CT),2014InternationalConferenc eon*,Langkawi,2014,pp.171-175.
12. M.-S.Wu,"Geneticalgorithmbasedondiscretewavelettransformationforfractalimagecompression,"*JournalofVisual CommunicationandImageRepresentation*,vol.25,no.8,pp.1835-1841,2014.
13. A.J.Hussain,D.Al-Jumeily,N.Radi,andP.Lisboa,"Hybridneuralnetworkpredictive-waveletimagecompressionsystem,"*Neurocomputing*,vol.151,no.3,pp.975-984,2015.
14. K.M.M.Prabhu,K.Sridhar,M.Mischi,andH.N.Bharath,"3-DwarpeddiscretecosinetransformforMRIimagecompression,"*BiomedicalSignalProcessingandControl*,vol.8,n o.1,pp.50-58,2013.

.