Crime Data Analysis of Pakistan Using Principal Factor Approach

Urooj Shehzadi (um8735091@gmail.com)
College of Statistical and Actuarial Sciences, University of the Punjab, Lahore, Pakistan

Maryam Ilyas (maryam.stat@pu.edu.pk)
College of Statistical and Actuarial Sciences, University of the Punjab, Lahore, Pakistan

Aiman Tahir[1] (aimantahir78@gmail.com)
College of Statistical and Actuarial Sciences, University of the Punjab, Lahore, Pakistan

ABSTRACT

This study is designed to interpret the yearly crime data of Pakistan at the country level for the period January 1997 to December 2018 and at the provincial level for the period January 1998 to December 2018. In crime analysis, the crime variables are typically large in number and are linearly related to each other. It is often difficult to interpret this type of dataset. To cope with these two problems, the Principal Factor Analysis (PFA) and correlation analysis are usually performed. Correlation analysis observe the pattern of relationship between the set of variables and PFA is a technique for reducing the dimensionality of such datasets, increasing interpretability but at the same time minimizing information loss. This study constructs the new transformed crime factors of original crime incidents that are low in dimensions and provide most of the total variability of the original variables. Crime against person and crime against property are two principal factors which are identified in this case study. These two factors explain 84.066% of the total variability in the data analysis of Pakistan. The 79.819% of the total variance of original variables is captured by these two factors in the crime analysis of Punjab.

Keywords: *crime analysis, principal component analysis, factor analysis, dimension reduction*

1. INTRODUCTION

Crime is an illegal action which is specifically against the criminal law. It is forbidden and punishable by law. It is destructive not only to someone but also leaves the negative influences on society, community or state. As it

1. Corresponding author

is harmful to society, so it is important to identify crime incidents and its prevention measures (Vold, 1958). According to Lopez-Rey (1986), Crime is the violation of previously stated law. Crime can be categorized as civil or social crime. They are conducted everywhere in the world either at a micro level or at a larger scale. Murder, attempted murder, rape, kidnapping, dacoity, robbery, burglary, motor vehicles theft and cattle theft are most common crimes which are attempted in our societies.

To find out the prevention measures of crime, it is important to collect and compile the crime data to investigate how and why crime occurs. There is exist excessive literature on crime data. America reported a significant increase in the crime rate of murders conducted by the organized crime at a large scale since 2007. It has the highest homicide rate of 26.5 per 100000 population. In 2011, United Nations Office on Drug and Crime (UNODC) estimates America has approximately (12408899/100000) of crime followed by Germany 2112843 and France 1172547 (Harrendorf et.al, 2010). According to the British crime survey, an increase of 190% mobile phone theft between 1995 and 2000 is reported, representing 28% of all robberies in 2000-2001 compared to 8% in 1998-1999 (Harrington and Mayhew, 2002). Central and eastern Europe faced increase in drug and property offenses between 1990 and 2000. In addition to the figures mentioned above crime rate was highest during 1990s (Aromaa and Nevala, 2004). South Africa also has a high prevalence of shoplifting, commercial crime, residential and business burglaries, and theft from motor vehicles during 2009-2010). Several studies have been conducted which have analyzed the determinants of crimes (Khan et al., 2015; Cerulli et al., 2018). Incidence of crime by property and by self are positively correlated to the increase in poverty among population.

Pakistan is one of the countries having high crime rate. It is becoming an alarming situation day by day. The total number of reported crimes including dacoity, robbery, burglary, cattle theft, murder, attempted murder, kidnapping has increased by about 63 percent during the period 1996-2007 (Gillani et al., 2009). It is reported that eight of the ten districts of Punjab and two of Khyber Pakhtunkhwa (KPK) have highest number of crimes prevalence. Districts in which occurrence of crimes is highly reporting includes Lahore (5102), Faisalabad (2294) and Peshawar (1665). Percentages of crimes regarding property, robbery, dacoity, and criminal trespass went up increased by 10% and 11%, respectively.

To control the crime activities, there is need to explore the crime data. Crime analysis involves the large number of variables and these variables are interrelated. Several studies have been conducted to observe the pattern of relationship between the set of crime variables using correlation analysis, multivariate techniques and other statistical techniques. (Ahamad, 1967; Muchwanju, 2015; Norman, 2004; Syed and Ahmed, 2013; Bello et al., 2014; Abbas et al., 2018). The crime data have been analyzed most of the time by using the principal component analysis. There is abundant literature on it (Ali and Razzak, 2015; Syed and Ahmed, 2013; Haider and Ali, 2015; Olufolabo et al., 2015; Osuji et al., 2015; Olakorede et al., 2017). Correlation analysis examines which variables are highly or moderately correlated to each other. Highly correlated crimes can be used to predict the crime rate in place of each other. The moderate correlation between crimes against property or a person implies that crime rate would be predicted except of these variables. On the other hand, Multivariate strategies (Johnson and Wichern, 2002) provide useful features to recognize the pattern of relationship between the set of variables by utilizing their covariance structure. As the variables are not independent to each other in crime data matrix, one variable is linearly related to other. To extract the unique information from crime dataset, some transformed components or new factors of these set of original crime variables are formed by using Principal Component Analysis (PCA) (Jolliffe, 2003) and principal factor analysis (Kim and Mueller, 1978b; Yong, 2013). These new low dimensional factors are uncorrelated to each other and provide most of the variation of the dataset. In this aspect, computational cost is reduced and crime analysis becomes easy.

This paper explores the yearly crime dataset based on police report which is collected from Pakistan and Punjab Bureau of Statistics. This study is conducted to interpret the crime data of Pakistan at two levels as country level and province level. At first stage, 9 major crimes incidents of Pakistan for the period January 1997 to December 2018 are evaluated. At second level, 14 major crimes incidents of Punjab for the period January 1998 to 2018 December are interpreted. This study evaluates the most common crime variables, their pattern of relationship and new transformed factors of original crime incidents by utilizing the correlation analysis and principal factor analysis. These low dimensional transformed components instead of the original variables can be used for predicting the crime rate in future analysis.

This paper is distributed in different sections. Section 2 explains the sources of data collection and gives brief description of the crime data. Section 3 describes the methods which have been used in this paper. Section 4 analyzes the results. In section 5, the paper is concluded.

2. DATA COLLECTION AND DESCRIPTION

The yearly crime data of Pakistan and Punjab which relies on police report is collected from Pakistan Bureau of Statistics and Punjab Bureau of Statistics respectively. Pakistan Bureau of Statistics is official government department of Pakistan which is developed for collecting and compiling the statistical facts and figures for the researchers and planners to design the several social and economic policies. Punjab Bureau of statistics is data handling agency of government of Punjab which is working to assemble and organize the statistical information at provincial level. It provides user friendly data and ensure the transparency and integrity of the data.

In this paper, crime data is taken at two levels. At first stage, nine major crimes of Pakistan for the period January 1997 to December 2018 are collected. These nine crime indicators are murder, attempted murder, kidnapping/abduction, dacoity, robbery, burglary, cattle theft, other theft, others. At second level, fourteen major crime incidents of Punjab between the period January 1998 to December 2018 are utilized in this case study. Murder, attempted murder, hurt, rioting, assault on public servant, rape, kidnapping, dacoity, robbery, burglary, motor vehicle theft, cattle theft, ordinary theft, others are fourteen crimes which are observed at Punjab level.

3. METHODOLOGY

This section reviews the several approaches as principal component analysis, its model, principal factor analysis, KMO and Bartlett's test which are utilized in this paper to explore the crime data.

3.1 Principal component analysis

Principal Component Analysis (PCA) is one of the most frequently used multivariate technique. It concerns to explain the variance-covariance structure of a set of variables through a few linear combinations of these variables. Its main objectives are data reduction and interpretation (Johnson and Wichern, 2002). The classical PCA as an exploratory data analysis tool involves a data matrix $\boldsymbol{X}$ of order $n \times p$, whose j^{th} column is the vector $\boldsymbol{x}_j$ of observations on the j^{th} variable. The linear combination of the columns of matrix $\boldsymbol{X}$ with maximum variance are calculated using the following relation

$$\sum_{j=1}^{p} a_j \boldsymbol{x}_j = \boldsymbol{X}\boldsymbol{a} = \boldsymbol{Z} \qquad [\text{ i }]$$

where $\boldsymbol{a}$ is a vector of constants $a_1, a_2, \ldots, a_p$. These linear combinations are called principal components. This transformed set of variables is uncorrelated to each other (Jolliffe, 2003). The covariance matrix of the principal components is calculated by *Cov* ($\boldsymbol{Z}$) = $\boldsymbol{a}' \mathbf{S}\, \boldsymbol{a}$, where $\mathbf{S}$ is the sample covariance matrix associated with the dataset and ($'$) denotes transpose. Hence, identifying the linear combination with maximum variance is equivalent to obtaining a p -dimensional vector $\boldsymbol{a}$ which maximizes the quadratic form $\boldsymbol{a}' \mathbf{S}\, \boldsymbol{a}$ with the constraint $\boldsymbol{a}'\boldsymbol{a}$ =1.

Another attractive property of principal components is sequential variance maximization. It means that first principal component explains maximum variation of the data. The second principal component will have the maximum variation left over after the first principal component and so on. First few principal components can capture unique and most of the total variability of data matrix. Then these transformed variables are used for further analysis instead of the original variables.

There is no hard and fast rule about how many components to be retained. Several rules of thumb are used for this purpose. Scree plot (Catell, 1996) is one of these rules provide useful visual aid to determining an appropriate number of principal components. It is a plot of magnitude of eigenvalue versus its number i.e., $\hat{\lambda}_i$ versus *i.* The number of components is chosen to be the point at which the remaining eigenvalues are relatively small and all about the same size. A cumulative proportion of variation explained by principal components is also a useful criterion for determining the number of components to be retained in the analysis.

3.2 Principal factor analysis

Principal Factor Analysis (PFA) is an estimation method of common factor analysis. The essential purpose of factor analysis is to describe if possible, the covariance structure among many variables in terms of a few underlying but unobservable random quantities called factors or latent variables (Yong, 2013; Kim and Mueller, 1978b). Suppose all variables within a particular group are highly correlated among themselves but have relatively small correlation with variables in a different group.

The covariance structure of orthogonal factor model is composed of the portion of the variation contributed by the common factors which is called communality and portion of the variation due to the specified factor that is known as specific variance. It is implied by using the relation

$$\mathbf{S} = \tilde{\boldsymbol{L}}\tilde{\boldsymbol{L}}' + \boldsymbol{\psi} \qquad [\text{ ii }]$$

To estimate model [ii], PFA is one of the methods that is used for estimation purpose. It is performed to extract the common factors by utilizing the principal component strategy. In which sample covariance matrix $\boldsymbol{S}$ is specified in terms of its eigenvalue and eigenvector pairs $(\hat{\lambda}_1, \hat{\boldsymbol{a}}_1)$, $(\hat{\lambda}_{1,}, \hat{\boldsymbol{a}}_2)$,…, $(\hat{\lambda}_{1,}, \hat{\boldsymbol{a}}_p)$, where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq, \dots, \geq \hat{\lambda}_p$. Let $m < p$ be the number of common factor. Then the matrix of estimated factor loadings $\{\tilde{l}_{ij}\}$ is given by

$$\tilde{\boldsymbol{L}} = [\sqrt{\hat{\lambda}_1}\hat{\boldsymbol{a}}_1 \vdots \sqrt{\hat{\lambda}_2}\hat{\boldsymbol{a}}_2 \vdots \dots \vdots \sqrt{\hat{\lambda}_m}\hat{\boldsymbol{a}}_m] \qquad [\text{ iii }]$$

The communalities are estimated as

$$\tilde{h}_i^2 = \tilde{l}_{i1}^2 + \tilde{l}_{i2}^2 + \dots + \tilde{l}_{im}^2 \qquad [\text{ iv }]$$

Since, original factor loadings may not be readily interpretable so it is recommended to rotate them for achieving the simpler structure. After rotating the factor loadings, each variable loads highly on factor 1 and has small to moderate loadings on other factors.

To evaluate whether the dataset is appropriate for factor analysis, the two steps that are take into account. The number of samples and strength of relationship between variables are investigated before performing the factor analysis Pallant (2013). To determine the adequacy of sampling, Kaiser-Meyer-Olkin (KMO) (Kaiser 1970, 1974) is used and the strength of the relationship is accessed by using the Bartlett's test of sphericity (Bartlett, 1954).

3.3 Kaiser-Meyer-Olkin (KMO) Test:

It is a measure of how suited your data is for Factor Analysis. The test measures sampling adequacy for each variable in the model and for the complete model. The statistic is a measure of the proportion of variance among variables that might be common variance. The lower the proportion, the more suited your data is to Factor Analysis. The sampling is adequate or sufficient if the value of Kaiser Meyer Olkin (KMO) is larger than 0.5 (Field, 2000), according to Pallant (2013) the value of KMO is 0.6 and above. Kaiser (1974) recommends a bare minimum of 0.5 and the value between 0.5 and 0.7 are mediocre, value between 0.7 and 0.8 are good, value between 0.8 and 0.9 are great and value between 0.9 and above are superb (Hutcheson and Sofroniou, 1999).

3.4 Bartlett's Test of Sphericity:

It compares an observed correlation matrix to the identity matrix. Essentially, it checks to see if there is a certain redundancy between the variables that could be summarized with a few numbers of factors. The null hypothesis of the test is that the variables are orthogonal, i.e. not correlated.

The significant value less than 0.05 indicates that these data do not produce an identity matrix and are thus approximately multivariate normal and acceptable for further analysis (Pallant, 2013; Field, 2000).

3.5 Computational steps

Following are some computational steps which were involved in this analysis.

1. Collect the dataset from the Bureau of statistics explained in section 2.
2. Compute the correlation matrix of the set of variables.
3. Apply the KMO test and Bartlett's test of sphericity on correlation matrix computed in step 2 that shows dataset is appropriate for factor analysis.
4. Compute the eigen values and eigenvectors of the correlation matrix obtained in step 2.
5. Retain the number of factors by following the scree plot criterion and accessing the cumulative percentage of explained variation.
6. Compute factor loadings on the factors retained in step 5 by using the eq. [iii]
7. Rotate the factor loadings and construct the factors.
8. Obtain the communalities utilizing the eq. [iv]

4. ANALYSIS AND RESULTS

In this section, the results of the analysis of the data on the total number of crimes committed yearly from 1997 to 2018 in Pakistan and 1998 to 2018 in Punjab are presented.

4.1 Analysis at country level

Correlation matrix of the crime variables

Table 1

Correlation	Murder	Att. murder	Kidnapping	Dacoity	Robbery	Burglary	Cattle. Theft	Other. Theft	Others
Murder	1.000	.741	.543	.735	.683	.653	.430	.599	.409
Att. murder	.741	1.000	.294	.845	.684	.528	.632	.404	.134
Kidnapping. Abduction	.543	.294	1.000	.595	.802	.777	.035	.934	.949
Dacoity	.735	.845	.595	1.000	.856	.712	.367	.661	.473
Robbery	.683	.684	.802	.856	1.000	.646	.444	.847	.747
Burglary	.653	.528	.777	.712	.646	1.000	-.032	.733	.584
Cattle theft	.430	.632	.035	.367	.444	-.032	1.000	.227	.039
Other theft	.599	.404	.934	.661	.847	.733	.227	1.000	.908
Others	.409	.134	.949	.473	.747	.584	.039	.908	1.000

This correlation matrix displays different levels of correlation between the crimes. The table shows kidnapping/abduction and others crime have the highest correlation coefficient (.949) which implies that they are the most common crimes in Pakistan.

The null hypothesis that the correlation matrix is an identity matrix which is rejected at 5% level of significance (Bartlett's test of Sphericity; $\chi 2$ = 251.408 with 36 degrees of freedom, p-value = .000), this implies that the correlation in the dataset is appropriate for factor analysis. Also, KMO statistic = .763 reveals that adequate sampling is being used for this analysis.

Percentage of the explained variation

Table 2

Factors	Initial eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.781	64.237	64.237	5.781	64.237	64.237	4.498	49.980	49.980
2	1.785	19.828	84.066	1.785	19.828	84.066	3.068	34.086	84.066
3	.774	8.596	92.661						
4	.326	3.624	96.286						
5	.185	2.060	98.346						
6	.063	.703	99.049						
7	.040	.443	99.492						
8	.036	.397	99.889						
9	.010	.111	100.000						

Table 2 displays eigenvalues, percentage and cumulative percentage of explained variance which leads to decide how many factors (or components) would be retained. As rule of thumb, factors having eigenvalues greater than one are sufficient to be retained. It can be observed from the scree plot of eigenvalues on Figure 1 (see, Appendix A.1), and extraction sums of squares loadings on table 2, two components are extracted. These two factors explain 84.066% of the total variability of the data set as the first component accounts 49.980% maximum variation of the total variation and 2nd component captures 34.086% of the total variance.

Rotated component matrix

Table 3

	Components	
	1	2
Murder	.486	.710
Att. Murder	.210	.941
Kidnapping	.982	.105
Dacoity	.569	.721
Robbery	.748	.579
Burglary	.786	.300
Cattle theft	-.106	.809
Other theft	.922	.262
Others	.945	-.014

This rotation matrix is converged in 3 iteration. It is clear that variables i.e., kidnapping, robbery, burglary other theft and others define the factor 1 (high loadings on factor 1, small or negligible loadings on factor 2), while variables as murder, attempted. murder, dacoity and cattle theft loads the factor 2 (high loadings on factor 2, small or negligible loadings on factor 1).

Communalities

Table 4

	Initial	Extraction
Murder	1.000	.740
Att. murder	1.000	.929
Kidnapping	1.000	.974
Dacoity	1.000	.843
Robbery	1.000	.895
Burglary	1.000	.708
Cattle theft	1.000	.666
Other theft	1.000	.918
Others	1.000	.893

From Table 4, It can be seen that the most of the variation of all the original variables are captured by the retained two factors. As kidnapping/ abduction, attempted murder and other theft are best represented in the common factor space this is because a high proportion of their variances explained by the principal components.

4.2 Analysis at province level

Correlation matrix part # 1

Table 5

Correlation	Murder	Att. murder	Hurt	Rioting	Assault on public servant	Rape	Kidnapping
Murder	1.000	.844	.574	.276	.330	.033	.340
Att. murder	.844	1.000	.760	.604	.044	-.343	-.053
Hurt	.574	.760	1.000	.625	.103	-.364	-.118
Rioting	.276	.604	.625	1.000	-.312	-.703	-.611
Assault on public servant	.330	.044	.103	-.312	1.000	.646	.696
Rape	.033	-.343	-.364	-.703	.646	1.000	.923
Kidnapping	.340	-.053	-.118	-.611	.696	.923	1.000
Dacoity	.911	.679	.452	.025	.510	.308	.609
Robbery	.554	.170	.043	-.486	.717	.764	.938
Burglary	.595	.246	.038	-.300	.692	.717	.810
Motor vehicle theft	.431	.031	-.091	-.584	.733	.880	.980
Cattle theft	.461	.656	.537	.310	-.005	-.252	.033
Ordinary theft	.347	.034	.039	-.525	.596	.753	.907
Other	.287	.006	.144	-.252	.376	.502	.632

Correlation (part # 2)

Table 5 (cont'd)

Correlation	Dacoity	Robbery	Burglary	Motor veh -icles theft	Cattle theft	Ordinary theft	Others
Murder	.911	.554	.595	.431	.461	.347	.287
Att. murder	.679	.170	.246	.031	.656	.034	.006
Hurt	.452	.043	.038	-.091	.537	.039	.144
Rioting	.025	-.486	-.300	-.584	.310	-.525	-.252
Assault on public servant	.510	.717	.692	.733	-.005	.596	.376
Rape	.308	.764	.717	.880	-.252	.753	.502
Kidnapping	.609	.938	.810	.980	.033	.907	.632
Dacoity	1.000	.783	.727	.675	.492	.639	.462
Robbery	.783	1.000	.822	.966	.205	.903	.645
Burglary	.727	.822	1.000	.852	-.040	.577	.509
Motor vehicle theft	.675	.966	.852	1.000	.063	.878	.594
Cattle theft	.492	.205	-.040	.063	1.000	.312	.151
Ordinary theft	.639	.903	.577	.878	.312	1.000	.614
Other	.462	.645	.509	.594	.151	.614	1.000

This correlation matrix displays different levels of correlation between the crimes. The table shows robbery and motor vehicle theft having the highest correlation coefficient (.966) implying that they are the most common crimes in Punjab.

The null hypothesis that the correlation matrix is an identity matrix is rejected at 5% level of significance (Bartlett's test of Sphericity; $\chi 2 = 415.357$ with 91 degrees of freedom, p-value = .000), this implies that the correlation in the dataset is appropriate for factor analysis. Also, KMO statistic = .721 reveals that adequate sampling is being used for this analysis.

Percentage of the explained variation

Table 6

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.266	51.903	51.903	7.266	51.903	51.903	7.182	51.297	51.297
2	3.908	27.916	79.819	3.908	27.916	79.819	3.993	28.522	79.819
3	.939	6.704	86.524						
4	.651	4.647	91.171						
5	.532	3.803	94.973						
6	.234	1.670	96.643						
7	.216	1.544	98.187						
8	.138	.984	99.172						
9	.045	.321	99.493						
10	.031	.218	99.710						
11	.021	.147	99.857						
12	.011	.077	99.934						
13	.006	.044	99.979						
14	.003	.021	100.000						

Table 6 displays eigenvalues, percentage and cumulative percent of explained variance which directs to how many factors (or components) are being retained. As rule of thumb factors having eigenvalues greater than one are sufficient to be retained. From the scree plot of eigenvalues on Figure (see, Appendix A.2), and extraction sums of squares loadings on table 8, it can be seen that two components would be extracted. These two factors explain 79.819% of the total variance in the data set as the first component accounts 51.297% maximum variation of the total variance and 2nd component captures 28.522% of the total variance.

Rotated component matrix

Table 7

	Component	
	1	2
Murder	.418	.827
Att. murder	.007	.966
Hurt	-.078	.849
Rioting	-.589	.657
Assault on public servant	.755	.067
Rape	.885	-.381
Kidnapping	.986	-.065
Dacoity	.676	.688
Robbery	.969	.172
Burglary	.851	.186
Motor vehicles theft	.990	.007
Cattle theft	.054	.717
Ordinary theft	.891	.078
Others	.657	.111

Rotated component matrix is converged in 3 iterations. It is clear that variables assault on public servant, rape, kidnapping /abduction, robbery, burglary, motor vehicle theft, ordinary theft and others define factor 1 (high loadings on factor 1, small or negligible loadings on factor 2), while variables murder, attempted murder, hurt, dacoity, rioting and cattle theft determine factor 2 (high loadings on factor 2, small or negligible loadings on factor 1).

Communalities

Table 8

	Initial	Extraction
Murder	1.000	.859
Att. Murder	1.000	.932
Hurt	1.000	.727
Rioting	1.000	.778
Assault on public servant	1.000	.575
Rape	1.000	.928
Kidnapping	1.000	.976
Dacoity	1.000	.930
Robbery	1.000	.969
Burglary	1.000	.759
Motor vehicles theft	1.000	.980
Cattle theft	1.000	.517
Ordinary theft	1.000	.800
Others	1.000	.444

From Table 8, it can be seen the most of the variation of all the original variables are captured by the retained two factors. that motor vehicles theft, kidnapping, robbery, attempted murder, dacoity, and rape (.980, .976, .969, .932, .930, .928.) are respectively best represented in the common factor space and this is because a high proportion of their variances explained by the principal components.

CONCLUSION

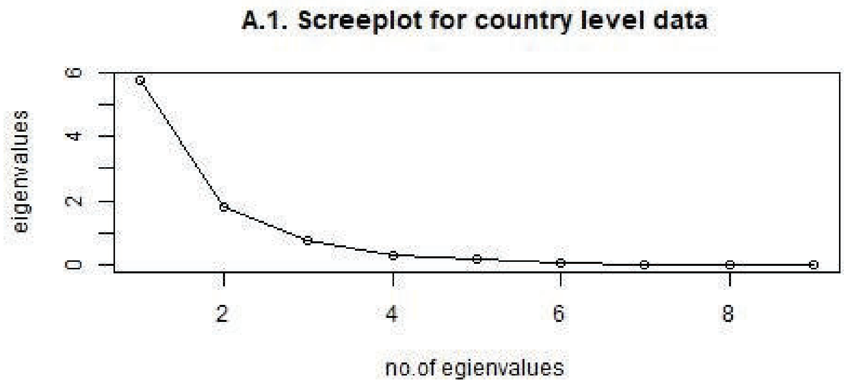
The purpose of this paper is to interpret the crime data and observe the interrelation of crimes to each other. It also reduces the dimensionality of dataset and ignore the redundancy of the variables so that data analysis become easy. The year-based crime data of Pakistan and Punjab is explored here. To examine the correlation structure between crimes how the crimes are related to each other and to overcome the high dimensionality of data, the correlation analysis and the principal factor analysis are performed. After carrying out the principal factor analysis at country and provincial level, two principal factors are constructed i.e., crime against person and crime against property. These two factors can be used for predicting the crime rate in future.

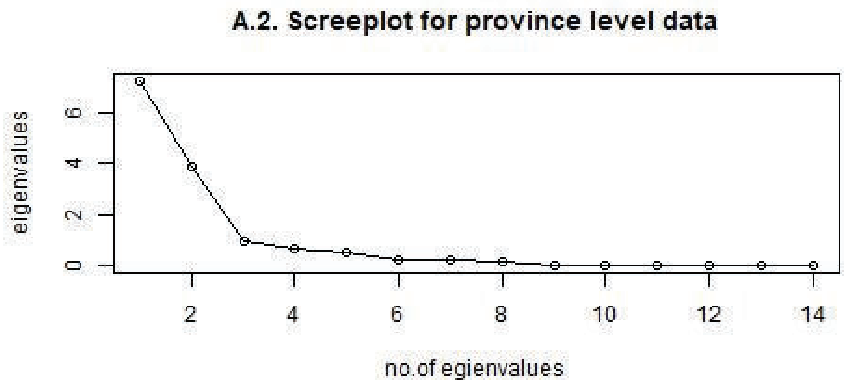
References

1. **Ahamad, B.**, 1967, *An analysis of crimes by the method of principal components*. Journal of the Royal Statistical Society: Series C (Applied Statistics), *16*(1):17-35.
2. **Abbas, S., Shouping, L., Sidra, F., & Sharif, A**., 2018, *Impact of Crime on Socio-Economic Development: A Study of Karachi*. Malaysian Journal of Social Sciences and Humanities (MJSSH), 3(3): 148-159.
3. **Ali, M., & Razzak, H.**, 2015, *Principal component analysis of Pakistan crime data.* Sci. Int.(Lahore), *27*(5): 4839-4844.
4. **Aromaa, K., & Nevala, S.**, 2004, *Crime and Crime Control in an integrating Europe*. Heuni.
5. **Bartlett, M. S.**, 1954, *A note on the multiplying factors for various χ2 approximations*. Journal of the Royal Statistical Society: Series B (Methodological), 16(2):296-298.
6. **Bello, Y., Batsari, Y. U., & Charanchi, A. S**., 2014, *Principal Component Analysis of Crime Victimizations in Katsina Senatorial Zone*. International journal of science and Technology, 3(4):192-202.
7. **Burger, J., Gould, C., & Newham, G.**, 2010, *The state of crime in South Africa: An analysis of the SAPS crime statistics for 2009/10*. South African Crime Quarterly, 34, 3-12.
8. **Cattell, R. B.**, 1966, *The scree test for the number of factors*. Multivariate behavioral research, 1(2):245-276.
9. **Cerulli, G., Ventura, M., & Baum, C. F.**, 2018, *The economic determinants of crime: an approach through responsiveness scores*. Report, Department of Economics, Boston College.
10. **Field, A.**, 2013, *Discovering statistics using IBM SPSS statistics*. sage.
11. **Gillani, S. Y. M., Rehman, H. U., & Gill, A. R.**, 2009, *Unemployment, poverty, inflation and crime nexus: Cointegration and causality analysis of Pakistan*. Pakistan Economic and Social Review, 79-98.
12. **Haider, A., & Ali, A.**, 2015, *Socio-economic determinants of crimes: a cross-sectional study of Punjab districts*. International Journal of Economics and Empirical Research, 3(11):550-560.
13. **Harrendorf, S., & Heiskanen, M.**, 2010, *International statistics on crime and justice*. S. Malby (Ed.). Helsinki: European Institute for Crime Prevention and Control, affiliated with the United Nations (HEUNI).
14. **Harrington, V., Mayhew, P., & Home Office, London (United Kingdom)**. *Research, Development and Statistics Directorate*;. (2001). Mobile phone theft.
15. **Hutcheson, G. D., & Sofroniou, N.**, 1999, *The multivariate social scientist: Introductory statistics using generalized linear models*. Sage.
16. **Jolliffe, I. T.**, 2003, *Principal component analysis*. Technometrics, 45(3): 276.
17. **Kaiser, H. F.**, 1970, *A second-generation little jiffy*. Psychometrika, 35(4): 401-415.
18. **Kaiser, H. F**., 1974, *An index of factorial simplicity*. Psychometrika, 39(1):31-36
19. **Khan, N., Ahmed, J., Nawaz, M., & Zaman, K.**, 2015, *The socio-economic determinants of crime in Pakistan: New evidence on an old debate.* Arab Economic and Business Journal, 10(2):73-81.
20. **Kim, J. O., Ahtola, O., Spector, P. E., & Mueller, C. W.**, 1978, *Introduction to factor analysis: What it is and how to do it* (No. 13). Sage.
21. **Lopez-Rey, M**., 1986, *The dimensions of crime*. Federal Probation, 50, 32.

22. **Muchwanju, C., Chelule, J. C., & Mung'atu, J.**, 2015, *Modelling crime rate using a mixed effect regression model*. American journal of Theoretical and Applied statistics, 4(6): 496-503.
23. **Norman, P., & Button, M.**, 2004, *International crime in the Mediterranean*. In Mediterranean security and beyond. Bruce George MP.
24. **Olakorede, N. M., Adams, S. O., & Olanrewaju, S. O.**, 2017, *Principal component analysis of crime data in Gwagwalada Area Command, Abuja from 1995-2015*. American Journal of Theoretical and Applied Statistics, 6(1):38-43.
25. **Olufolabo, O. O., Akintande, O. J., & Ekum, M. I.**, 2015, *Analyzing the Distribution of Crimes in Oyo State Using Principal Component Analysis*. IOSR J. Math. Ver. I, 11(3): 2278-5728.
26. **Osuji, G. A., Obubu, M., & Obiora-Ilouno, H. O.**, 2015, *An Investigation On Crime Rate In Southeastern Nigeria*. European Journal of Statistics and Probability, 3(4): 1-9.
27. **Pallant, J.**, 2020, *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge.
28. **Johnson, R. A., & Wichern, D. W.**, 2002, *Applied multivariate statistical analysis* (Vol. 5, No. 8). Upper Saddle River, NJ: Prentice hall.
29. **Syed, S. H., & Ahmed, E.**, 2013, *Poverty, Inequality, Political Instability and Property Crimes in Pakistan: A Time Series Analysis*. Asian Journal of Law and Economics, 4(1-2): 1-28.
30. **Vold, G. B.**, 1958, *Theoretical criminology*. Oxford University Press.
31. **Yong, A. G., & Pearce, S.**, 2013, *A beginner's guide to factor analysis: Focusing on exploratory factor analysis*. Tutorials in quantitative methods for psychology, 9(2):79-94.

Appendix


A.1. Screeplot for country level data
eigenvalues
0
2
4
6
2
4
6
8
no.of egienvalues


A.2. Screeplot for province level data
eigenvalues
0
2
4
6
2
4
6
8
10
12
14
no.of egienvalues