Analysis of Crimes against women in India

Udit Narayan Sahoo, Biswajit Mandal

November 2023

Abstract

This project synthesizes its findings to propose evidence-based recommendations for policymakers, law enforcement agencies, and civil society to collectively address and combat crimes against women in India. By combining quantitative and qualitative approaches, this analysis seeks to contribute to the ongoing discourse on women's safety, fostering a more informed and effective approach towards creating a safer and more equitable society.

Contents

0.1 Introduction

Violence against women, defined by the United Nations as any gender-based act resulting in, or likely to result in, physical, sexual, or mental harm, constitutes a severe violation of human rights. It affects women of all ages, castes, and creeds, presenting a widespread public health concern. The repercussions of such violence extend across short-, medium-, and long-term horizons, impacting both physical and mental well-being. Beyond individual suffering, the societal and economic ramifications are profound.

This issue is entrenched in social, economic, developmental, legal, educational, human rights, and health contexts, contributing to morbidity and mortality among women. The International Day for the Elimination of Violence Against Women, observed on November 25^{th} annually, highlights the persistent challenge. Despite efforts, violence against women impedes progress toward equality, development, peace, and the realization of women and girls' human rights.

According to a WHO report, one in three women and girls experiences physical or sexual violence, often perpetrated by an intimate partner. In India, National Crime Records Bureau (NCRB) data emphasizes cruelty by husbands or relatives as the most prevalent crime against women. The NCRB classifies such crimes into two categories: those under the Indian Penal Code (IPC) and those under Special and Local Laws (SLL). This study focuses specifically on crimes under the IPC, highlighting seven severe offenses, including rape, kidnapping, dowry deaths, acid attacks, domestic violence, assaults, and human trafficking.

0.2 Objective

The analysis of the topic can be approached in the following ways:

- Comparison of crimes against women mentioned previously with the overall crime rate in the country over the past five years, from 2017 to 2021. This involves assessing whether crimes against women have increased or decreased during this period.
- Graphical observation of the proportion of crimes accompanied by a state-wise comparison of crimes against women in India for the year 2021.
- Examination of the effects of different social variables on the total number of crimes per lakh population through regression analysis.
- Exploration of the possibility of forecasting future crime rates at different time points, if feasible.

0.3 Source of Data

- The data on the number of Crimes against Women in various States of India from 2017 to 2021 is acquired from the official website of the National Crime Records Bureau, Ministry of Home Affairs, Government of India.
- Additionally, the data on the literacy rate of females and the workforce participation rate is sourced from the official website of the Ministry of Statistics and Programme Implementation (MoSPI).
- The information on police stations per lakh population for each state is sourced from the Bureau of Police Research and Development, Government of India.

- Additionally, data on female population and urbanization are collected from Google.

The search links for all the data are provided in the References section.

0.4 Methodology

1. The analysis of crime incidence employs various graphical methods. Firstly, line diagrams are utilized to illustrate the total number of crimes over the five-year period from 2017 to 2021. These line diagrams depict each of the seven previously mentioned crimes individually, as well as the overall number of crimes against women in India.

2. For the graphical analysis of the data on crimes against women in 2021, the following plots are employed:

   - A pie chart is used to display the proportion of each of the five crimes relative to the total number of crimes.

   - Principal component analysis is applied to the cross-sectional data of 2021, and the resulting bi-plot reveals which states were more susceptible to specific crimes in that year.

   - A vertical bar diagram is employed to plot the total number of crimes against women against the states, facilitating a comparison between states.

3. To investigate the impact of different social variables on the total number of crimes against women per lakh population, ordinary least square (OLS) is applied to the cross-sectional data of 2019. Here, crime is treated as an error-contaminated variable, considering the potential for under-reporting, and OLS is applied to an econometric model.

4. Finally, various measures and tests are employed to assess the randomness of the data, with attempts made to forecast it if feasible.

0.5 Graphical Analysis of Each Crime Against Women According to Year

In this section, our objective is to present several graphs that provide a comprehensive overview of various types of crimes occurring over a span of 5 years. Specifically, we focus on the total number of crimes associated with seven distinct types of crimes against women, and these figures are graphically depicted against the respective years.

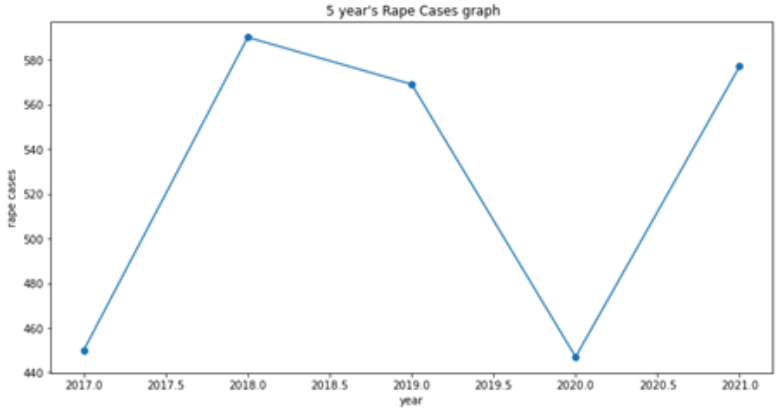

Figure 1: The total number of registered incidents of rape cases has exhibited a continuous increase since the year 2017 to 2018, reaching its peak in 2018 then slightly decreased in between 2018 to 2019, and decreased in 2020 due to the pandemic effect and then started continuously increasing.

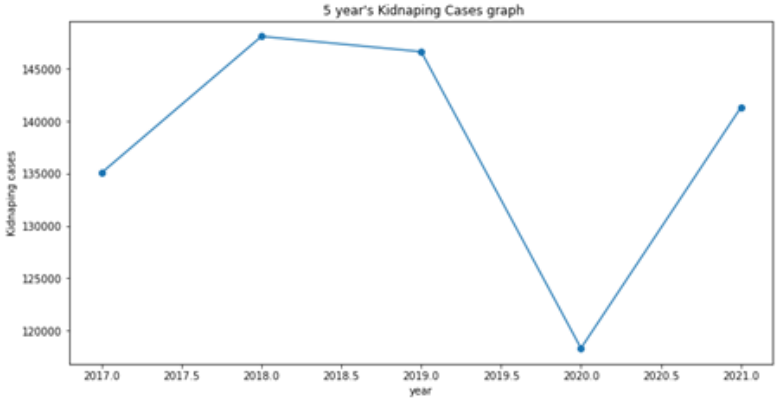

Figure 2: The total number of registered incidents of Kidnapping has exhibited a continuous increase since the year 2017 to 2018, reaching its peak in 2018 then slightly decreased in between 2018 to 2019, and completely decreased in 2020 due to the pandemic effect and then started continuously increasing.

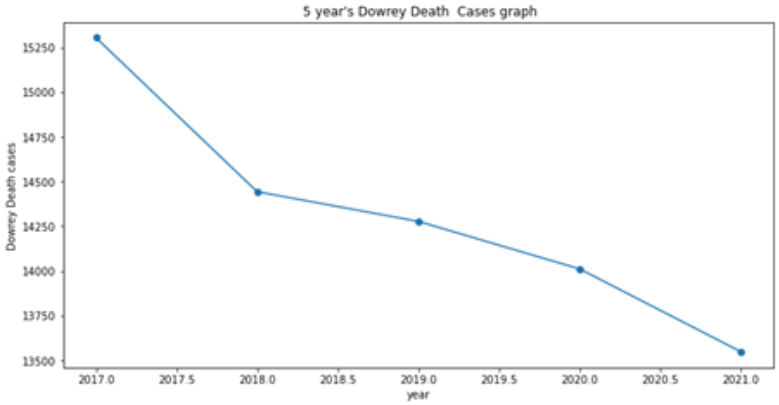

Figure 3: The total number of registered incident of Dowrey Death has exhibited a sudden downfall since the year 2017 to 2018, since then it started continuous decrease till 2021

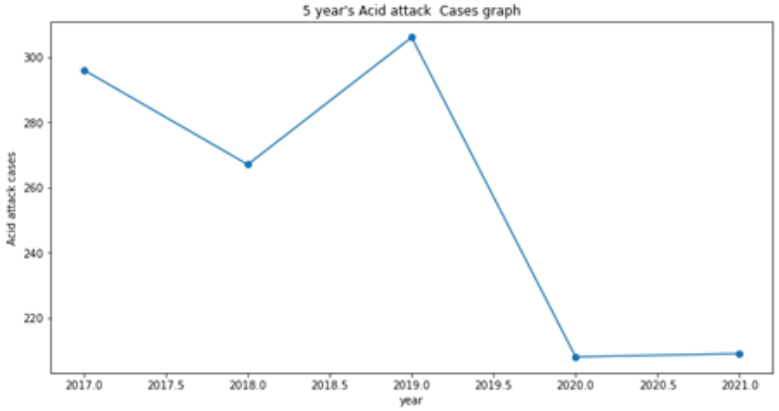

Figure 4: The total number of registered incidents of Acid Attack has exhibited a continuous decrease since the year 2017 to 2018, then reaching its peak in 2019 then a sudden downfall in between 2019 to 2020 due to the pandemic situation, since then it remains same as 2020.

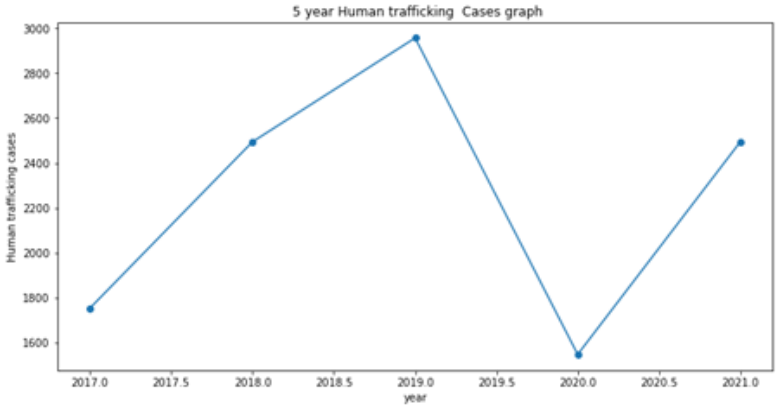

Figure 5: The total number of registered incidents of Human Trafficking has exhibited a continuous increase since the year 2017 to 2019 reaching its peak in 2019 then a sudden downfall in between 2019 to 2020 due to the pandemic situation, since then it started continuous increase till 2021.

0.6 Graphical Analysis of Crime Against Women in 2021

0.6.1 Proportion of each crime to the total crime for the year 2021 is shown by a pie chart

Considering the cross-sectional data of the year 2021 we can show which crime is contributing in which amount to the total number of crimes among the five crimes that are rape, kidnapping and abduction, dowry deaths, acid attack, trafficking by a pie chart.

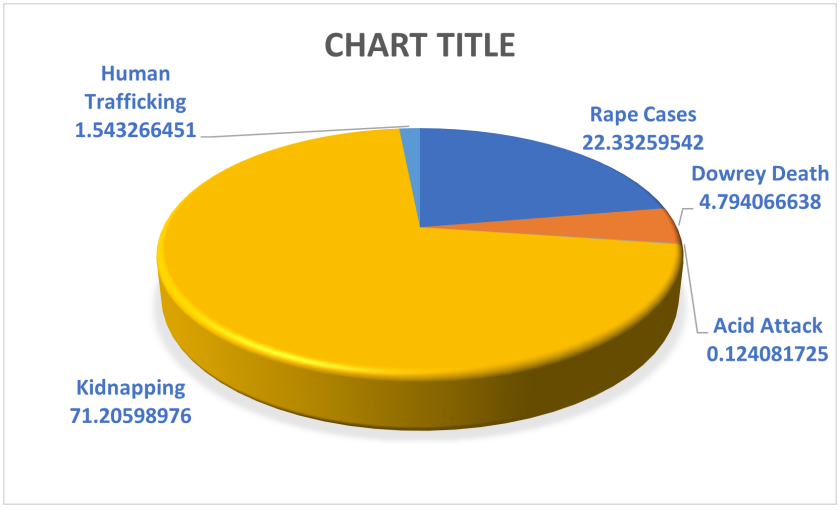

Figure 6: In 2021, it can be concluded that among the five crimes against women, Kidnapping holds the largest proportion, followed by Rape Cases and then Dowrey Death, Human Trafficking, Acid Attack respectively.

0.6.2 Bi-plot of Cross-Sectional data of the year 2021

In this section we apply Principal component analysis (PCA) on the cross sectional data of 2021 considering 29 states and 7 union territories of India. After applying PCA we construct the bi-plot.

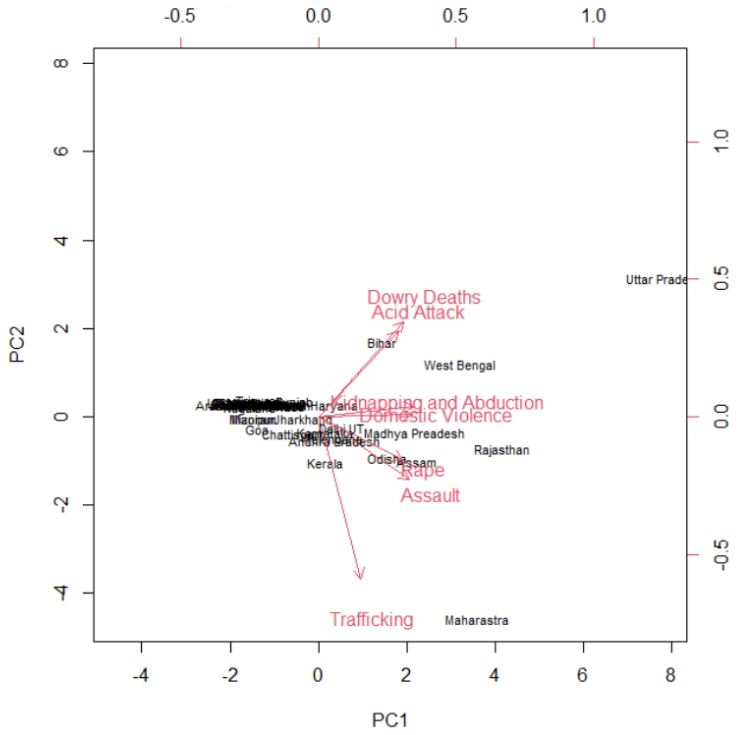

Figure 7: The state names represent the scores for the first two principal components. The pink arrows indicate the first two principal component loading vectors.

- Bihar exhibited a higher susceptibility to crimes such as Acid Attack and Dowry Deaths.
- West Bengal showed a higher susceptibility to crimes such as Domestic Violence and Kidnapping and Abduction.
- Uttar Pradesh demonstrated a higher susceptibility to crimes such as Acid Attack, Dowry Deaths, Kidnapping.
- On the other hand, Rajasthan, Madhya Pradesh, Assam, Odisha, and Delhi were more prone to crimes such as Rape cases.
- Maharashtra was prone to Human Trafficking.
- Similarly, states such as Goa, Manipur, Mizoram, Sikkim, Uttarakhand, and Union Territories like Daman and Diu, Lakshadweep, Puducherry exhibited a lower susceptibility to any crime and Some states were significantly less prone to any crime, and there may be an overlap in those regions.

0.6.3 State-wise comparison of total crime against women in 2021

In this section we are going to observe a vertical bar diagram to compare the total number of crimes against women among the states, happened in the year 2021.

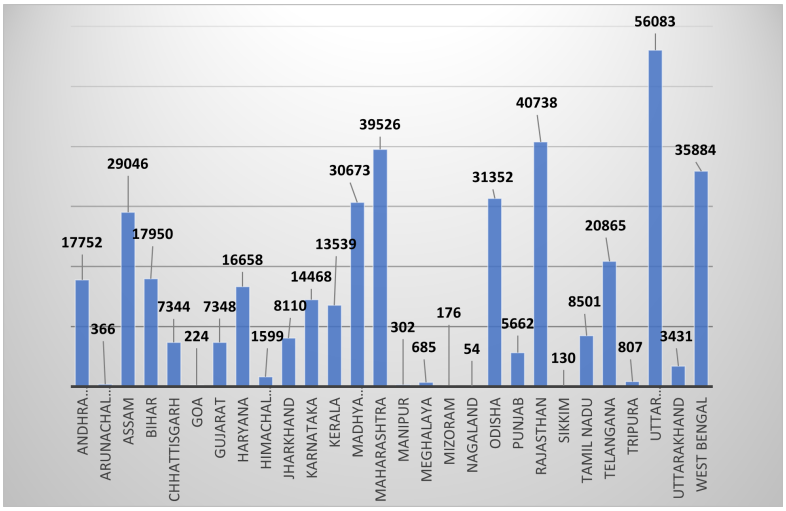

Figure 8: In 2021, the highest number of crimes against women occurred in Uttar Pradesh. Rajasthan got the 2^{nd} position, followed by Maharashtra in the 3^{rd} position. West Bengal got the 4^{th} position, while Odisha ranked the 5^{th} position in the total number of crimes against women in 2021.

0.7 Regression Analysis to find the effect of Different Social Variables on Crime Against Women

The impacts of certain social variables on the number of crimes against women are explored in this section, utilizing the ordinary least square (OLS) method. In this analysis, crime is treated as an error contaminated variable, taking into account the potential for underreporting. The OLS method is then employed to model and analyze the relationships within the data.

In order to explore the impact of social factors on crimes against women, we have chosen the total number of registered crimes per lakh female population as our response variable. We consider various social factors as covariates in our analysis. To assess the influence of these covariates on the dependent variable (crime), we will conduct regression analysis using cross-sectional data from the year 2021 across the 28 states and 8 Union territories of India.

However, a potential challenge arises from the underreporting of crimes against women in our country due to various social reasons. The officially registered number of crimes may be lower than the actual incidents that occurred. In India, people may be hesitant to report crimes to avoid involvement in police proceedings. Additionally, social stigma can lead to the non-registration of crimes such as rapes and assaults in Indian society. Therefore, it is crucial to acknowledge that the number of registered crimes may not accurately reflect the true extent of the issue. This consideration is essential to bear in mind throughout our analysis.

0.7.1 Description of Covariates

The description of the covariates for i^{th} state (i=1(1)36) are shown in this table —

Sl No.	Covariates	Descriptions of Covariates
1.	Total number of registered crimes against women per lakh female population (Y_i)	$Y_i = \frac{\text{Total number of crimes against women in the state}}{\text{Total female population in the state}} \times 100,000$
2.	Percentage of female population (X_{1i})	Percentage of female population among the total population of the state as per the 2011 census
3.	Female literacy rate (X_{2i})	Percentage of females who are literate among the total population in the state as per the 2011 census
4.	Workforce participation rate of females (X_{3i})	Rate of female population among the total population who work and earn in the state as per the 2011 census
5.	Number of police stations per lakh female population (X_{4i})	$X_{4i} = \frac{\text{Number of police stations in the state}}{\text{Total female population in the state}} \times 100,000$
6.	Urbanization rate (X_{5i})	Percentage of the total population residing in urban areas of the state as per the 2011 census

Table 1: In the census of the year 2011, Andhra Pradesh and Telangana were amalgamated as the erstwhile state of Andhra Pradesh. Consequently, the total number of registered crimes for Andhra Pradesh and Telangana are considered collectively as the total number of crimes for Andhra Pradesh. This arrangement results in a total of 28 states and 7 Union territories.

0.7.2 Model

We consider the regression equation as :

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \epsilon_i, i = 1(1)35 \tag{1}$$

Where,

- Y= total number of registered crimes against women per lakh female population
- X_1= Percentage of female population
- X_2= Female literacy rate
- X_3= Work force participation rate of female
- X_4= Number of police stations per lakh female population
- X_5= Urbanization rate
- β_j= regression coefficients, j=0(1)5
- The error term ϵ_i is independently and identically distributed ($\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$)

In light of underreporting, the response variable can be regarded as a variable contaminated by errors. Therefore, the actual model is now an econometric model,

$$Y_i^* = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i, \quad i = 1(1)35, \quad u_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_u^2)$$

Where, Y_i^*= True but not measurable response for the i^{th} State
And $Y_i = Y_i^* + v_i$, where v_i is the error term, $i = 1(1)35$, $v_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_v^2)$.
Therefore, $\epsilon_i = u_i + v_i$ and $\sigma_\epsilon^2 = \sigma_u^2 + \sigma_v^2$.
For the sake of simplicity, let us take $\sigma_v^2 = 1$. Thus, $v_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$.
To enhance the clarity and flow of the statement, the rearranged version could be:

To estimate the regression coefficients, ordinary least squares (OLS) is employed. To ensure the validity of the estimation, it is imperative to assess the compatibility of the data. This involves checking for normality, investigating the presence of heteroscedasticity in the error terms, and examining the potential multicollinearity among the independent variables. Notably, since we are dealing with cross-sectional data from the year 2021, there is no need to scrutinize autocorrelation between the error terms.

0.7.3 Checking for Normality

In order to check whether the error terms follow normal distribution or not we perform Shapiro Wilk normality test and get the results,
Value of test statistic= W = 0.93011,
p-value = $0.02818 (< 0.93011)$
Therefore, we fail to reject the null hypothesis at 1% level of significance and conclude that the error term follows a normal distribution.

0.7.4 Checking for Heteroscedasticity

In order to check that whether heteroscedasticity is present in the error term or not we perform Glejser test and get the results,
Value of test statistic=8.68
p-value= $0.122 (< 8.68)$
Therefore, we fail to reject the null hypothesis at 1% level of significance and conclude that heteroscedasticity is not present in the data that is the error terms are homoscedastic.

0.7.5 Checking for Multicollinearity

In order to check the presence of multicollinearity between the independent variables that is between the covariates we use the Variance Inflation Factor(VIF). Which is given by,

$$VIF_j = \frac{1}{1 - R_j^2}$$

where, R_j= multiple correlation obtaining regressing X_j on other independent variables.
If $VIF_j > 10$ then we can conclude that multicollinearity is present.
After calculating the VIF_j for i=1(1)5 we get the following results shown in the table :

Variable	VIF
$X1$	2.879536 (< 10)
$X2$	4.182366 (< 10)
$X3$	2.31875 (< 10)
$X4$	3.348450 (< 10)
$X5$	1.181621 (< 10)

Table 2: it can be observed that VIF for each of the 5 variables are less than 10. Therefore, we can conclude that multicollinearity is not present between the independent variables.

0.7.6 Results Obtained from OLS

The results obtained from OLS without underreporting is shown in the table :

Variables	**Estimated values of coefficient**	**Standard error**	**t-value**	**p-value**
Intercept	-36.2317	157.7137	-0.230	0.820
Percentage of female population ($X1$)	3.9176	3.3664	1.164	0.254
Female literacy rate ($X2$)	-1.0850	0.8439	-1.286	0.209
Workforce participation rate of female ($X3$)	-1.3317	0.8556	-1.556	0.130
Number of police stations per lakh female population ($X4$)	-0.9838	0.8786	-1.120	0.272
Urbanization rate ($X5$)	0.4500	0.4820	0.934	0.358

Table 3: Estimated Values of Coefficients without underreporting

Parameter	Value
Residual Standard error	40 (on 29 degrees of freedom)
Multiple R-squared	0.1842
Adjusted R-squared	0.04353
F-statistic	1.309 (on 5 and 29 DF)
p-value	0.2874
AIC	364.9675

Table 4: ANOVA Results without underreporting

The results obtained from OLS with underreporting is shown in Table :

Variables	Estimated values of coefficient	Standard error	t-value	p-value
Intercept	-29.7869	157.7966	-0.189	0.852
Percentage of female population ($X1$)	3.7886	3.3682	1.125	0.270
Female literacy rate ($X2$)	-1.0886	0.8444	-1.289	0.208
Workforce participation rate of female ($X3$)	-1.3197	0.8560	-1.542	0.134
Number of police stations per lakh female population ($X4$)	-0.9810	0.8791	-1.116	0.274
Urbanization rate ($X5$)	0.4439	0.4822	0.921	0.365

Table 5: Estimated Values of Coefficients after underreporting

Parameter	Value
Residual Standard error	40 (on 29 degrees of freedom)
Multiple R-squared	0.1842
Adjusted R-squared	0.04114
F-statistic	1.292 (on 5 and 29 DF)
p-value	0.2945
AIC	365.0042

Table 6: ANOVA Results after underreporting

The above results reveal that the AIC (Akaike Information Criterion) is lower for the model without underreporting. Consequently, the model without underreporting provides the best fit for the data.
The fitted regression equation obtained from the model without underreporting is given by,

$$Y_i = -36.2317 + 3.9176X_{1i} - 1.0850X_{2i} - 1.3317X_{3i} - 0.9838X_{4i} + 0.4500X_{5i}, \quad i = 1(1)35$$

The relationship between the social factors taken as covariates and the crimes against women per lakh population are discussed below.

1. Percentage of Female population and crimes against women :

   The positive slope observed between the total number of crimes against women per lakh female population and the percentage of female population in a state indicates that as the female population increases, the number of women exposed to potential crimes also increases. This relationship suggests that higher female populations are associated with a higher incidence of crimes against women. Which illustrates that states with relatively high populations also contribute significantly to the total number of crimes against women in the country.

2. Female literacy rate and crime against women :

   "Education is the vaccine for violence." This statement underscores the transformative power of education as a potent tool for women to mitigate the occurrence of violence against them. Our study reveals that female literacy rate has a negative impact on the total number of crimes against women per lakh female population. This suggests that an educated woman is often more aware and empowered to prevent potential violence.

   As highlighted in the introduction, NCRB data in India indicates that cruelty by husbands or their relatives accounts for the highest number of recorded cases in crimes against women. Education equips a woman with the means to preempt such situations. It enables her to secure employment, achieve financial independence, and make choices that may allow her to escape environments where violence is prevalent.

3. Work force participation rate of female and crime against women :

   The workforce participation rate of females exhibits a negative impact on crimes against women. This suggests that an increase in work opportunities for women contributes to their financial independence. Consequently, financially independent women are better positioned to access crime-controlling measures, leading to a decrease in crimes against them.

   Additionally, financial independence empowers women to resist prevalent societal issues such as the dowry system. As women gain economic autonomy, the incidence of crimes associated with practices like dowry deaths is likely to decrease naturally.

4. Number of police stations and crime against women :

   The number of police stations per lakh female population demonstrates a negative relationship with crimes against women. This suggests that an increase in the number of police stations within a state leads to enhanced police surveillance, even in smaller areas. As a consequence, with a higher number of police officers and personnel per lakh population, law enforcement agencies can conduct more frequent patrolling. This increased police presence has the potential to effectively control and reduce the incidence of crimes against women.

5. Urbanization rate and crime against women :

   The urbanization rate is positively correlated with crimes against women, indicating that the incidence of such crimes is higher in urban areas. It's important to note that crimes against women also occur in rural areas; however, the analysis is based on officially registered crimes. In urban areas, crimes tend to be reported and registered more frequently than in rural areas, contributing to the observed positive relationship between urbanization rates and crimes against women in the dataset.

0.8 Problem in Forecasting of the Number of Crimes for Future Time Point

In this section, we will attempt to forecast the data for future time points. To achieve this, we have analyzed the data on the number of rapes in India for 12 years, spanning from 2011 to 2021. Our goal is to predict the number of rapes at future time points.

For this purpose, to determine the order of the Moving Average (MA) process, we need to plot the autocorrelation function (ACF). Similarly, to find the order of the Autoregressive (AR) process, a plot of the partial autocorrelation function (PACF) is required. The ACF and PACF plots of the data are shown in figures below -

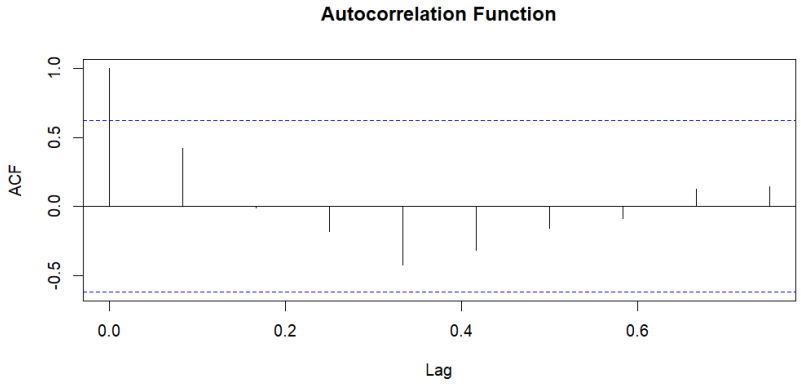

Figure 9: We can conclude that the autocorrelation is 1 only at lag zero. For all other lags, the autocorrelations either fall within the confidence band or are nearly zero. This observation leads us to the conclusion that our data is entirely random in nature.

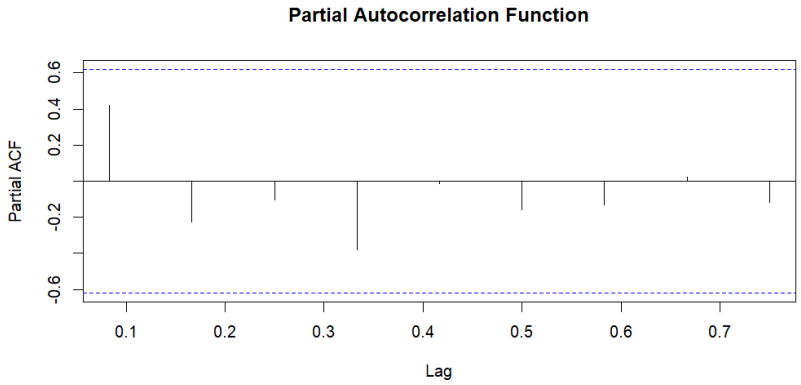

Figure 10: Observing that the partial autocorrelation for each lag is within the confidence band, we can conclude that our data is entirely random in nature based on this plot.

Therefore, we can not forecast the number of rapes in India for future time point.
Similarly, each crime category in India exhibits the same characteristic. Consequently, we cannot forecast any specific crime for any state using our dataset.

0.9 Conclusion

The empirical results from our regression analysis reveal several key relationships. There is a positive association between the female population of a state and crimes against women. Conversely, crimes exhibit a negative relationship with the female literacy rate, the workforce participation of females, and the number of police stations per lakh population. Additionally, crimes against women are positively associated with the urbanization rate.

In conclusion, it can be inferred that increasing the female literacy rate, enhancing the number of police stations per lakh population, and providing women with greater opportunities for work and financial independence could contribute to a reduction in crimes against women.

0.10 References

- "Time Series Analysis and Its Applications: With R Examples" by Robert H. Shumway and David S. Stoffer
- "Introduction to Time Series and Forecasting" by Peter J. Brockwell and Richard A. Davis
- "Applied Time Series Analysis for the Social Sciences" by John L. T. Cook and Linda C. Skinner
- "Forecasting: Principles and Practice" by Rob J Hyndman and George Athanasopoulos
- "Time Series Analysis: Forecasting and Control" by George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel
- "Practical Regression and Anova using R" by Julian J. Faraway
- "A Modern Approach to Regression with R" by Simon Sheather

- "Regression Analysis: Concepts and Applications" by Philip H. Kott
- "Regression Analysis by Example" by Samprit Chatterjee, Ali S. Hadi, and Bertram Price
- "Pattern Recognition and Machine Learning" by Christopher M. Bishop
- "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville
- "The Hundred-Page Machine Learning Book" by Andriy Burkov
- "Understanding Machine Learning: From Theory to Algorithms" by Shai Shalev-Shwartz and Shai Ben-David
- "Machine Learning for Dummies" by John Paul Mueller and Luca Massaron
- "Modeling and Forecasting Crime: A Review" by Jeffery T. Walker
- "Time Series Analysis for Crime Prevention: Assessing the Impact of Foot Patrol in Violent Crime Reduction" by Jerry H. Ratcliffe and George E. Tita
- "Time Series Analysis of Crime Rates" by John R. Hipp and Adam Boessen
- "Forecasting the Future of Predictive Crime Mapping" by Rachel L. Boba
- "Short-Term Forecasting of Crime" by Leslie Kennedy, James D. Green, and Lawrence F. Travis III
- "Spatio-Temporal Crime Hotspots and the Ambient Population" by Rachel L. Boba and Joel M. Caplan

0.11 Links used for this Projectwork

- NCRB crime data search links for years 2017 to 2021 in India :

  For 2017 :Click Here

  For 2018 :Click Here

  For 2019 :Click Here

  For 2020 :Click Here

  For 2021 :Click Here
- Data search link for literacy rate of female of each state in India : Click Here
- Data search link for workforce participation rate of women of each state in India : Click Here
- Data search link for police stations of each state in India : Click Here
- Data search link for female population of each state in India : Click Here
- Data search link for urbanization rate of each state in India : Click Here
- Crime in India: specification and estimation of violent crime index published in Journal of Productivity Analysis : Click Here
- Analysis of Growth and Identifications of the Determinants of Crime against Women: Insight from India published in Journal of International Women's Studies : Click Here

0.12 Acknowledgement

I extend my sincere and heartfelt gratitude to my respected sir: Mr. Prasenjit Banerjee (Department of Applied Statistics and and Analytics, Maulana Abul Kalam Azad University of Technology). His help, guidance, and encouragement was invaluable in shaping this endeavor. Without his support, the project couldn't have evolved to its current form within the limited time frame.

I would also like to express my heartfelt thanks to the other professors in our department. Their ongoing guidance and unwavering support, along with the valuable information they provided about the project, were instrumental in its success.

I am grateful to my parents for their cooperation and encouragement throughout this journey. Last but certainly not least, I extend my gratitude to all my friends who played a role, whether directly or indirectly, in helping me complete this project.

0.13 R Programming Code

The entire study has utilized the programming language R. The corresponding R code is provided below -

- R code for graphical analysis of each crime according to year:

```
## Reading of 2021 data
data21 = read.csv("C:/Users/bishu/Desktop/crime 2021.csv")
data21
head(data21)
crimes = as.matrix(data21$Total21)
states = as.matrix(data$State)  % Note: Corrected from 'data' to 'data21'
barplot(crimes)
rownames(data21) = data21[, 1]  % Make 1st colnames as rownames
data21[, 1] = NULL  % Remove the row names
head(data21)

    ##calculating total of each crime
totRape21=sum(data21$Rape)
totKandA21=sum(data21$KandA)
totDowryDeaths21=sum(data21$DowryDeaths)
totAcidAttack21=sum(data21$AcidAttack)
totDomesticViolence21=sum(data21$DomesticViolence)
totAssault21=sum(data21$Assault)
totTrafficking21=sum(data21$Trafficking)
totRape21
totKandA21
totDowryDeaths21
totAcidAttack21
totDomesticViolence21
totAssault21
totTrafficking21
sum21=totRape21+totKandA21+totDowryDeaths21+totAcidAttack21
+totDomesticViolence21+totAssault21+totTrafficking21
```

```
##changing column names
colnames(data21)=c("Rape","Kidnapping and Abduction",
"DowryDeaths","Acid Attack","DomesticViolence",
"Assault","Trafficking")
head(data21)

######################2020
##Reading of 2020 data
data20=read.csv("C:/Users/bishu/Desktop/crime 2020.csv")
data20
head(data20)
rownames(data20)=data20[,1] ##make 1st colnames as rownames
data20[,1]=NULL ##remove the row names
head(data20)


##calculating total of each crime
totRape20=sum(data20$Rape)
totKandA20=sum(data20$KandA)
totDowryDeaths20=sum(data20$DowryDeaths)
totAcidAttack20=sum(data20$AcidAttack)
totDomesticViolence20=sum(data20$DomesticViolence)
totAssault20=sum(data20$Assault)
totTrafficking20=sum(data20$Trafficking)
totRape20
totKandA20
totDowryDeaths20
totAcidAttack20
totDomesticViolence20
totAssault20
totTrafficking20
sum18=totRape20+totKandA20+totDowryDeaths20
+totAcidAttack20+totDomesticViolence20
+totAssault20+totTrafficking20

##Reading of 2019 data
data19=read.csv("C:/Users/bishu/Desktop/crime 2019.csv")
data19
head(data19)
crimes=as.matrix(data19$Total19)
states=as.matrix(data$State)
barplot(crimes)
rownames(data19)=data19[,1] ##make 1st colnames as rownames
data19[,1]=NULL ##remove the row names
head(data19)

##calculating total of each crime
totRape19=sum(data19$Rape)
totKandA19=sum(data19$KandA)
```

```
totDowryDeaths19=sum(data19$DowryDeaths)
totAcidAttack19=sum(data19$AcidAttack)
totDomesticViolence19=sum(data19$DomesticViolence)
totAssault19=sum(data19$Assault)
totTrafficking19=sum(data19$Trafficking)
totRape19
totKandA19
totDowryDeaths19
totAcidAttack19
totDomesticViolence19
totAssault19
totTrafficking19
sum19=totRape19+totKandA19+totDowryDeaths19
+totAcidAttack19+totDomesticViolence19
+totAssault19+totTrafficking19

##changing column names
colnames(data19)=c("Rape","Kidnapping and Abduction",
"Dowry Deaths","Acid Attack",
"DomesticViolence","Assault","Trafficking")
head(data19)

######################2018
##Reading of 2018 data
data18=read.csv("C:/Users/bishu/Desktop/crime 2018.csv")
data18
head(data18)
rownames(data18)=data18[,1] ##make 1st colnames as rownames
data18[,1]=NULL ##remove the row names
head(data18)

##calculating total of each crime
totRape18=sum(data18$Rape)
totKandA18=sum(data18$KandA)
totDowryDeaths18=sum(data18$DowryDeaths)
totAcidAttack18=sum(data18$AcidAttack)
totDomesticViolence18=sum(data18$DomesticViolence)
totAssault18=sum(data18$Assault)
totTrafficking18=sum(data18$Trafficking)
totRape18
totKandA18
totDowryDeaths18
totAcidAttack18
totDomesticViolence18
totAssault18
totTrafficking18
sum18=totRape18+totKandA18+totDowryDeaths18
+totAcidAttack18+totDomesticViolence18
+totAssault18+totTrafficking18
```

```
#####################2017
##Reading of 2017 data
data17=read.csv("C:/Users/bishu/Desktop/crime 2017.csv")
data17
head(data17)
rownames(data17)=data17[,1] ##make 1st colnames as rownames
data17[,1]=NULL ##remove the row names
head(data17)

##calculating total of each crime
totRape17=sum(data17$Rape)
totKandA17=sum(data17$KandA)
totDowryDeaths17=sum(data17$DowryDeaths)
totAcidAttack17=sum(data17$AcidAttack)
totDomesticViolence17=sum(data17$DomesticViolence)
totAssault17=sum(data17$Assault)
totTrafficking17=sum(data17$Trafficking)
totRape17
totKandA17
totDowryDeaths17
totAcidAttack17
totDomesticViolence17
totAssault17
totTrafficking17
sum17=totRape17+totKandA17+totDowryDeaths17
+totAcidAttack17+totDomesticViolence17
+totAssault17+totTrafficking17

##LINE DIAGRAM

##Taking the crime values in different vectors
year=c(2017,2018,2019,2020,2021)
Rape=c(totRape17,totRape18,totRape19,totRape20,totRape21)
KandA=c(totKandA17,totKandA18,totKandA19,totKandA20,totKandA21)
DowryDeaths=c(totDowryDeaths17,totDowryDeaths18,
totDowryDeaths19,totDowryDeaths20,totDowryDeaths21)
AcidAttack=c(totAcidAttack17,totAcidAttack18,
totAcidAttack19,totAcidAttack20,totAcidAttack21)
DomesticViolence=c(totDomesticViolence17,
totDomesticViolence18,totDomesticViolence19,
totDomesticViolence20,totDomesticViolence21)
Assault=c(totAssault17,totAssault18,totAssault19,
totAssault20,totAssault21)
Trafficking=c(totTrafficking17,totTrafficking18,
totTrafficking19,totTrafficking20,totTrafficking21)
sum=c(sum17,sum18,sum19,sum20,sum21)

##Line diagram for rape
plot(year,Rape,type="b",main="Total number of
```

```
rapes in India in last 5 years",pch=19,col="red",
xlab="Years",ylab="No of Rapes")

##line diagram for Kidnapping and abduction
plot(year,KandA,type="b",pch=18,col="blue",
main="Total no of Kidnapping and abductions
in India in last 5 years",
ylab="Kidnapping and abduction")

##Line diagram for Dowry deaths
plot(year,DowryDeaths,type="b",pch=18,
col="deeppink",main="Total no of Dowry
deaths in India in 5 years",ylab="Dowry Deaths")

##Line diagram for Acid attack
plot(year,AcidAttack,type="b",pch=18,col="green",
main="Total no of Acid Attacks in India in 5
years",ylab="Acid Attacks")

##Line diagram for Domestic Violence
plot(year,DomesticViolence,type="b",pch=18,
col="purple",main="Total no of Domestic
Violence in India in 5years",ylab="Domestic Violence")

##Line diagram for Assaults
plot(year,Assault,type="b",pch=18,col="brown",
main="Total no of Assaults in India in 5 years",ylab="Assaults")

##Line diagram for trafficking
plot(year,Trafficking,type="b",pch=18,col="black",
main="Total no of Human Trafficking in India in 5
years",ylab="Trafficking")

##Line diagram For total crime
sum1=c(sum(data17$Total17),sum(data18$Total18),
sum(data19$Total19),sum(data15$Total20),sum(data16$Total21))
plot(year,sum1,type="b",pch=18,col="red",
main="Total crime against Women
in India in 5 years",ylab="Total Crime")
```

- R code for graphical analysis of crime against women in 2021 -

```
###PIE CHART
x=c(totRape21,totKandA21,totDowryDeaths21,totAcidAttack21,
totDomesticViolence21,totAssault21,totTrafficking21)
X=c("Rape","Kidnapping & Abduction","Dowry Deaths",
"Acid Attack","DomesticViolence","Assault","Trafficking")
df=data.frame(X,x)
df
```

```
library(RColorBrewer)
color=brewer.pal(df$X, "Set1")
pie(df$x,labels=df$X,hole=0,radius=1,main="Proportion of
crimes of the year 2021",col=color,values="%")

##Principal component analysis of 2021 data
library(stats)
pca21=prcomp(data21,scale=TRUE)
pca21
names(pca21)
summary(pca21)
pca21$rotation=-pca21$rotation
pca21$rotation
pca21$x=-pca21$x
pca21$x

##Biplot
biplot(pca21,scale=0,cex=c(0.7,1),main="Biplot for number
of crimes in 2019")
```

- R Code for regression analysis:

```
        ##Reading of 2021 data
data21=read.csv("C:/Users/bishu/Desktop/crime 2021 new.csv")
data21
head(data21)
rownames(data21)=data21[,1] ##make 1st colnames as rownames
data21[,1]=NULL ##remove the row names
head(data21)

##reading of the data regarding covariates
covariates=read.csv("C:/Users/bishu/Desktop/covariates.csv")
head(covariates)
rownames(covariates)=covariates[,1] ##make 1st colnames as rownames
covariates[,1]=NULL ##remove the row names
colnames(covariates)=c("X1","X2","X3","X4","X5","pop")
head(covariates)

##Extracting the covariates as vectors from the data frame
y=matrix(data21$Total21/covariates$pop)*100000
X1=matrix(covariates$X1)
X2=matrix(covariates$X2)
X3=matrix(covariates$X3)
X4=matrix(covariates$X4/covariates$pop)*100000
X5=matrix(covariates$X5)
data1=data.frame(y,X1,X2,X3,X4,X5)
head(data1)

##Fitting model with underreporting
```

```
library(mvtnorm)
set.seed(1)

##estimate the model (with measurement error in y)
error_model=lm(y~., data = data1)
summary(error_model)
aic1=AIC(error_model)

##Fitting model withour underreporting
##estimate the model (without measurement error in y)
Y=data1$y-matrix(rnorm(n = 35, sd =1))
data2=data.frame(Y,X1,X2,X3,X4,X5)
head(data2)
nonerror_model=lm(Y~., data = data2)
summary(nonerror_model)
aic2=AIC(nonerror_model)

##printing AIC of both models
aic1
aic2

##print estimated coefficients
nonerror_model$coefficients
error_model$coefficients

##Calculating MSE for both the models
fit1=fitted(error_model)
fit2=fitted(nonerror_model)
MSE1=sum((y-fit1)^2)/35
MSE2=sum((Y-fit2)^2)/35
MSE1
MSE2

##Normality Check
shapiro.test(resid(nonerror_model))

##Heteroscedasticity check
library(skedastic)
glejser(nonerror_model)

##Multicolinearity
library(car)
vif(nonerror_model)
```

- • R code for problem in Forecasting:

```
    ##Reading the data of 2021
data21=read.csv("C:/Users/bishu/Desktop/crime 2021.csv")
data21
```

```
head(data21)
rownames(data21)=data21[,1] ##make 1st colnames as rownames
data21[,1]=NULL ##remove the row names
head(data21)

##Reading the data of 2020
data20=read.csv("C:/Users/bishu/Desktop/crime 2020.csv")
data20
head(data20)
rownames(data20)=data20[,1] ##make 1st colnames as rownames
data20[,1]=NULL ##remove the row names
head(data20)

##Reading the data of 2019
data19=read.csv("C:/Users/bishu/Desktop/crime 2019.csv")
data19
head(data19)
rownames(data19)=data19[,1] ##make 1st colnames as rownames
data19[,1]=NULL ##remove the row names
head(data19)

##Reading the data of 2018
data18=read.csv("C:/Users/bishu/Desktop/crime 2018.csv")
data18
head(data18)
rownames(data18)=data18[,1] ##make 1st colnames as rownames
data18[,1]=NULL ##remove the row names
head(data18)

##Reading the data of 2017
data17=read.csv("C:/Users/Dell/Desktop/crime 2017.csv")
data17
head(data17)
rownames(data17)=data17[,1] ##make 1st colnames as rownames
data17[,1]=NULL ##remove the row names
head(data17)

##taking the years in a vector
year=c(2011,2012,2013,2014,2015,2016,2017,2018,2019,2020,2021)
year

##Taking 11 years of data on rape of India
Inrape=c(1759,2042,1563,2500,1963,3050,3468,data17[27,1],
data18[27,1],data19[27,1],data20[27,1],data21[27,1],)
acf(\textit{In}rape,10,main="ACF of number of rapes in In")

##autocrrelation function
windows()
pacf(Inrape,10,main="PACF of number of rapes in India")
```

```
##partial autocorrelation function
##checking for randomness
Box.test(Inrape,lag=10,"Box-Pierce",fitdf=0) ##Portmanteau Test
Box.test(Inrape,lag=10,"Ljung-Box",fitdf=0) ##Ljung-Box Test
```