

Chapter– 18

STATISTICS AND PROBABILITY

Ch.Id:-ASU/GRF/EB/RAPOET/2022/Ch-18

DOI: <https://doi.org/10.52458/9789391842765.2022.eb.grf.asu.ch-18>

¹**SHIVANI**

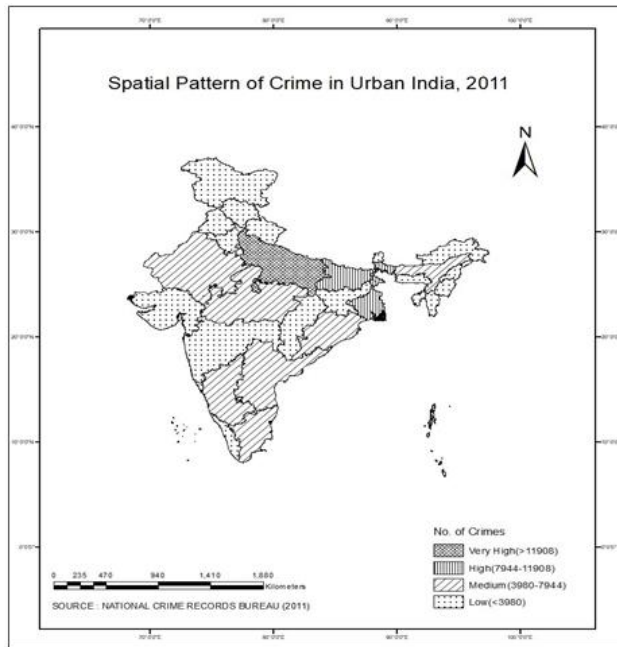
¹Apeejay Stya University, Sohna, Gurugram

²**Dr. GARIMA SHARMA**

²Apeejay Stya University, Sohna, Gurugram

INTRODUCTION TO THE TOPIC

This portfolio provides the statistical analysis on the topic urban crime in India in the year 2011. The main purpose of this study is to analyze the pattern of crime on the country level. To investigate the relation between population growth, the number of vehicles, and the growth of crime, appropriate statistical tools like correlation and regression methods were used.



Data

Below table shows the data used in this portfolio

States		Murder	Rape	Kidnapping & Abduction	Dacoity
1	Andhra Pradesh	2808	1442	2154	126
2	Arunachal Pradesh	65	42	93	13
3	Assam	1303	110	3764	305
4	Bihar	3198	934	4268	556
5	Chhattisgarh	1110	1053	472	68
6	Goa	48	29	28	2
7	Gujrat	1126	439	1214	221
8	Haryana	1062	733	959	167
9	Himachal Pradesh	130	168	212	1
10	Jammu and Kashmir	110	277	1077	14
11	Jharkhand	1747	784	941	309
12	Karnataka	1820	636	1395	214
13	Kerala	365	1132	299	71

14	Madhya Pradesh	2511	3406	1288	118
15	Maharashtra	2818	110	110	773
16	Manipur	78	53	110	1
17	Meghalaya	110	130	87	49
18	Mizoram	26	77	6	1
19	Nagaland	46	23	34	7
20	Odisha	1477	1112	1139	417
21	Punjab	842	479	681	28
22	Rajasthan	1461	1800	3204	28
23	Sikkim	14	16	10	0
24	Tamil Nadu	1877	677	1984	101
25	Tripura	163	205	154	11
26	Uttar Pradesh	4951	2042	8500	379
27	Uttarakhand	178	129	314	13
28	West Bengal	2109	2363	4285	236
29	A & N Islands	14	13	15	1
30	Chandigarh	24	27	58	6
31	D & N Haveli	14	4	9	7
32	Daman & Diu	6	1	3	4
33	Delhi	543	572	3767	33
34	Lakshadweep	NA	NA	NA	NA
35	Puducherry	32	7	12	5

STATISTICAL ANALYSIS OF DATA

In this section I would be analyzing the data completely using simple statistical tools and would be giving my suggestions after studying the analysis report. To begin with I have used the most common statistical tools- Mean, Mode, Median and Standard Deviation to analyses the data. The calculations are shown in the table below:

	States	Murder	Rape	Kidnapping & Abduction	Dacoity
1	Andhra Pradesh	2808	1442	2154	126
2	Arunachal Pradesh	65	42	93	13
3	Assam	1303	110	3764	305
4	Bihar	3198	934	4268	556
5	Chhattisgarh	1110	1053	472	68
6	Goa	48	29	28	2
7	Gujrat	1126	439	1214	221
8	Haryana	1062	733	959	167
9	Himachal Pradesh	130	168	212	1
10	Jammu and Kashmir	110	277	1077	14
11	Jharkhand	1747	784	941	309
12	Karnataka	1820	636	1395	214
13	Kerala	365	1132	299	71
14	Madhya Pradesh	2511	3406	1288	118
15	Maharashtra	2818	110	110	773
16	Manipur	78	53	110	1
17	Meghalaya	110	130	87	49
18	Mizoram	26	77	6	1
19	Nagaland	46	23	34	7
20	Odisha	1477	1112	1139	417
21	Punjab	842	479	681	28
22	Rajasthan	1461	1800	3204	28
23	Sikkim	14	16	10	0
24	Tamil Nadu	1877	677	1984	101
25	Tripura	163	205	154	11
26	Uttar Pradesh	4951	2042	8500	379
27	Uttarakhand	178	129	314	13
28	West Bengal	2109	2363	4285	236
29	A & N Islands	14	13	15	1
30	Chandigarh	24	27	58	6

31	D & N Haveli	14	4	9	7
32	Daman & Diu	6	1	3	4
33	Delhi	543	572	3767	33
34	Lakshadweep	0	0	0	0
35	Puducherry	32	7	12	5
	Total	34186	21025	55211	4285
	Mean	976.74	600.71	1577.45	122.42
	Median	543	277	472	33
	Mode	14	110	110	1
	Range	4945	2041	8497	375
	Standard deviation	1179.72	784.22	1807.87	179.035
	Skewness	1.92	1.87	2.21	2.01
	Kurtosis	4.68	3.74	6.38	4.203

From the above data we can conclude that:-

- **Mean** - Mean -Looking at the average number of cases that have been registered under different Acts, it is observed that the highest average no.of cases are registered under the act kidnapping and abduction which is equal to 1577.45, followed by murder which is 976.74 and then by rape cases which is nearly 600.71 and at the minimum by dacoity which is 122.42.This shows that Kidnapping and abduction are the most happening crimes in India and the govt. needs to focus more importance to the matter of prevention of this crime.
- **Median** - Talking about median we know that median is the central value which divides the data set into two parts: lower and upper part. Of these cities, 50% have rates that are lower than the calculated heinous crime while 50% have rates that are higher. From the above data we can say that 50% of cites have kidnapping and abduction rate more than 472,murder rate more than 543 ,rape cases more than 277 and dacoity more than 33.
- **Mode**- After calculating mode for every crime I can say that for the murder the mode is 14 which indicates that number of crimes in different cities is 14.The mode is the number that occurs most often in a data set and concerning the given data it shows the number of crimes happening in the states. It does not give us an idea of the overall crime situation. From the obtained mode I can simply say that no. of murder cases for most of the states are low as compare to the average. For crime- rape and kidnaping the mode is 110 which indicates that most of the rape happening in states is 110. For the dacoity the mode is 1.

- **Standard deviation**-The standard deviation of a dataset gives you a measure of how spread out it is. On average, it helps you ascertain how close each point is to the mean. A high standard deviation signifies a high deviation of data points from the mean. A moderate SD signifies a moderate deviation of data points from the mean and a low (can be even zero) signifies that the data points are close to the mean. After calculating the SD for each crime, I found that for Kidnapping which has SD of 1807.87, the leads are vastly spread and there is maximum difference between the crime cases for different states. The SD of Murder came out to be 1179.72 followed by the rape crime with SD 784.22 and then Dacoity with SD 179.03 having the minimum difference between the crime cases in different states. Now based on the mean, mode, median and standard deviation I analyzed the data using some more statistical tools such as range, skewness and kurtosis.
- **Range:** - The range of a set of data is the result of subtracting the smallest value from largest value. It gives a rough idea of how the outcome of the data set will be before you look at it actually. Since it only depends on two of the observations, it is most useful in representing the dispersion of small data sets. After calculating the range for each crime, I found that for kidnapping the range is 8497 which is far greater than the murder with the range 4945 and rape with range 2041 followed by dacoity with range 375 only.
- **Skewness and Kurtosis** - Skewness refers to a distortion or asymmetry that deviates from the symmetrical bell curve, or normal distribution, in a set of data. With the calculated mean, mode, median and standard deviation I can say that:
 - If $\text{mean} > \text{mode}$ or $\text{mean} > \text{median}$ then the sample is positively skewed.
 - If $\text{mean} < \text{mode}$ or $\text{mean} < \text{median}$ then the sample is negatively skewed.
 - If the skewness is greater than 0.5 then we can say that the sample is positively skewed, and it would have a long-left tail in the normal distribution.
 - If the skewness is less than -0.5 then we can say that the sample is negatively skewed, and it would have a long-right tail in the normal distribution.
 - Kurtosis is another measure to define the shape of the normal distribution. Kurtosis is a statistical measure that defines how heavily the tails of a distribution differ from the tails of a normal distribution.
- If kurtosis is greater than 3 then we can say that it is a leptokurtic distribution
- If kurtosis is less than -3 then we can say that it is a platykurtic distribution

- If kurtosis is between -3 and 3 or close to zero then we can say that it is a mesokurtic distribution

After calculating the skewness and the kurtosis for each crime I found out that the maximum measure of skewness and kurtosis can be seen for the kidnapping with 2.21 skewness and 6.38 kurtosis. I can say that data is positively skewed and it is leptokurtic distribution. For Dacoity the measure of skewness and kurtosis are 2.01 and 4.203 respectively. Again, the data is positively skewed, and it is leptokurtic distribution. For murder the measure of skewness and kurtosis are 1.92 and 4.69 respectively. Again, the data is positively skewed, and it is leptokurtic distribution.

The minimum measure of skewness and kurtosis can be seen for the rape cases. The measure is 1.87 and 3.74. Again it is positively skewed, and it is leptokurtic distribution.

The correlation between different heinous crimes:

Crime	Murder	Rape	Kidnapping and abduction	Dacoity
Murder	1.000	0.691	0.774	0.7508
Rape	0.691	1.000	0.5795	0.2766
Kidnapping and abduction	0.774	0.5795	1.000	0.4456
Dacoity	0.7508	0.2766	0.4456	1.0000

It is observed that there is a positive correlation among all the different crimes under this category. Further, looking at the correlation relationship, it is observed that dacoity crime has a positive and high degree of correlation with an attempt to murder, kidnapping for ransom, and rape. This indicates that the act of dacoity also relates the possibility of crimes such as an attempt to murder, riot, kidnapping for ransom, and rape crimes. Murder crime is reported to be fairly correlated with kidnapping for ransom. This indicates that there is a collateral possibility of happening of crimes (kidnapping for ransom, and rape cases) when an attempt to murder crime takes place.

CONCLUSION

From the above calculations and analysis, it can be concluded that:

- The top 5 crime rated states are UP(15872), West Bengal(8993), Bihar(8956), Madhya Pradesh(7323) and Andhra Pradesh(6830).
- Among the four crimes - kidnapping and abduction(55211) crime is the maximum followed by Murder(34186), rape(21025) and dacoity(4285).

- The crimes in each of the state vary with each other and the maximum crime cases to the overall crime rate for a crime is obtained from only four or five states data among the rest. The states for which there is many crimes need to follow a new crime control strategy and the least crime case states should maintain the strategy they have been following.
- All the four crimes had a positive skewness and a larger measure of kurtosis, we can conclude that there is a risk of high crime rate in near future. But based on the value of the these we cannot determine the actual crime rate.

BIBLIOGRAPHY

- <https://www.cuemath.com/data/harmonic-mean/>
- https://www.investopedia.com/terms/d/descriptive_statistics.asp
- <https://www.simplilearn.com/what-is-descriptive-statistics-article>
- <https://brownmath.com/stat/shape.htm>
- <https://brownmath.com/stat/shape.htm>
- <https://www.sarthaks.com/401341/explain-the-various-methods-of-measuring-skewness>
- <https://www.sarthaks.com/401341/explain-the-various-methods-of-measuring-skewness>