Problem

The issue of district bifurcation and its impact has been a topic of great concern. However, given the constantly changing nature of districts and the entities that fall under their jurisdiction, the arrangement of datasets becomes a crucial factor in drawing meaningful inferences from regression analyses. The results obtained from such analyses are highly dependent on the structure of the dataset. To explore this further, we present an exercise where we investigate the implications of two different arrangements of the data. Through this study, we aim to shed light on the importance of properly structuring datasets when conducting regression analyses on issues related to district bifurcation. Our findings may have significant implications for policymakers and researchers working in this area.

Data

In this paper, we are using data from the Socioeconomic High-resolution Rural-Urban Geographic Dataset on India (SHRUG) database. The dataset comprises population data from three different years: 1991, 2001, and 2011. Specifically, we focus on the subdistrict key and the population data from the Population Census. The backbone of the SHRUG is a set of keys that link all the Indian Population and Economic Censuses to each other from 1990 to 2013 at the smallest consistent geographic unit possible. These keys were developed by matching towns and villages across population censuses, with close attention to splits, merges, and other realignments. Linking these multiple survey rounds has necessitated merging units at different levels of aggregation depending on how those units have changed. The unit of aggregation in the SHRUG is a SHRUG identifier, or a shrid. The population data provides information on the total population in each subdistrict for the three different years like total population, urban population, rural population, number of households, SC total population, ST total population, literate total population, number of primary schools, number of secondary schools, area of village (hectares), power supply, etc.

In addition to the population data from SHRUG, we have also used bifurcation data provided by our instructor. This data contains information on district bifurcation in India, such as the year in which a district was split, whether two separate districts were formed or one became a parent district and the other a child district, and other relevant details. By merging this bifurcation data with the population data from SHRUG, we are able to investigate the impact of district bifurcation on population growth in India. The bifurcation data allows us to identify the specific districts that underwent bifurcation and analyze the changes in population growth in those districts before and after bifurcation. Additionally, the bifurcation data allows us to analyze the differential impact of bifurcation on districts that became parent districts and those that became child districts. This may help us understand the potential long-term implications of bifurcation on population growth in India. Overall, the inclusion of bifurcation data in our analysis allows us to provide a more comprehensive understanding of the impact of district bifurcation on population growth in India and its implications for policymakers and researchers in this area.

Methodology

In this paper, we explore the impact of district bifurcation on population growth using data from the SHRUG database. To investigate this relationship, we have used two different methods for structuring our dataset.

Method 1:

In the first method, we merged the population data of all three years (1991, 2001, and 2011) together using the left join function in R by using SHRUG ID as the unique identifier. After this, we merged this dataset from all three years with the subdistrict codes of 2011 using Shrug ID. This resulted in a dataset where each row represents a unique subdistrict and contains the population data for that subdistrict in all three years. We created a new variable called "subdtcode" by concatenating the state ID, district ID, and subdistrict ID variables. We then, grouped the data using “subdtcode” and summarised the data.

Method 2:

In the second method, we combined the population data and subdistrict keys for each year separately using the left join function in R by using SHRUG ID as the unique identifier. We then merged these three separate datasets into one dataset containing the population data for all three years. Again, we used the left join function and SHRUG ID as the common identifier. This resulted in a dataset where each row represents a unique subdistrict, with the population data for that subdistrict in each of the three years presented in separate columns. We created a new variable called "subdtcode" by concatenating the state ID, district ID, and subdistrict ID variables. We then, grouped the data using “subdtcode” and summarised the data.

After organizing the datasets using these two methods, we merged them separately with the district bifurcation data using the left join function in R by using “subdtcode” as the unique identifier. This allowed us to compare the impact of district bifurcation on population growth using two different methods of dataset organization.

To analyze the data, we will use descriptive statistics and tabular representations to compare the population growth rates before and after district bifurcation. Overall, by using these two different methods of dataset organization, we aim to provide insights into the impact of district bifurcation on population growth in India and the importance of dataset structuring in drawing meaningful conclusions from regression analyses.

Results

Results of the research are presented in the Table below, which shows the mean of the data for the year 1991. The data is categorized into two datasets, Data I and Data II, representing the two different methods used for structuring the data. The table also includes information on the subdistrict bifurcation status, where a subdistrict can either split into two child subdistricts or into one parent district and one child district. Additionally, the table includes a category for subdistricts

that were not split. For each dataset and bifurcation category, the table displays the mean of the population variable.

	Data I			Data II		
	Parent	Child	No split	Parent	Child	No split
Literate population (% of total)	3.51E+01	3.38E+01	3.51E+01	3.55E+01	3.43E+01	3.58E+01
Number of households	2.02E+04	1.70E+04	1.98E+04	3.63E+04	2.63E+04	2.65E+04
Total population	1.15E+05	9.60E+04	1.09E+05	2.01E+05	1.47E+05	1.44E+05
SC total population	1.96E+04	1.72E+04	1.98E+04	3.18E+04	2.59E+04	2.45E+04
ST total population	1.58E+04	1.52E+04	7.91E+03	2.31E+04	2.22E+04	8.98E+03
Area of town (sq. km)	1.20E-02	9.87E-03	3.26E-02	3.67E-02	1.60E-02	5.95E-02
Number of primary schools	1.10E+02	9.14E+01	8.76E+01	1.48E+02	1.35E+02	1.02E+02
Number of secondary schools	9.51E+00	7.75E+00	7.72E+00	1.23E+01	1.14E+01	8.97E+00

Conclusion

In this research, we explored the effect of district bifurcation on population data in India. We used two different methods for structuring the data and analyzed the results to determine if the method of structuring the data had any impact on the conclusions drawn from the data. We found that there was a significant difference in the mean population for the bifurcated subdistricts compared to non-bifurcated subdistricts in both datasets.

Overall, our results suggest that district bifurcation has a noticeable impact on population data in India, and this effect should be taken into consideration when analyzing and interpreting population data. Additionally, our study highlights the importance of carefully structuring data and choosing appropriate methods for analysis, as the method of structuring the data can impact the results obtained.

Further research could explore the effect of district bifurcation on other demographic factors or investigate the impact of different methods for structuring data on other types of analyses. Nonetheless, this research provides a valuable contribution to the understanding of district bifurcation and its impact on population data in India.