Research Proposal

Data

Crime has been prevalent in our society for a very long time. It continues to be a threat to us and our society and demands serious consideration if we hope to reduce the onset of the repercussions caused by it. Hundreds of crimes are recorded daily by the data officers working alongside the law enforcement authorities throughout the United States.

The motivation behind this topic for the research is that every aware citizen in today's modern world wants to live in a safe environment and neighborhood. However, it is a known fact that crime exists in our society in some form.

This dataset contains data on the rate of violent crime (crimes per 1,000 population) for California, its regions, counties, cities, and towns. Crime and population data are from the Federal Bureau of Investigations, Uniform Crime Reports.

Data Dictionary

There are 27 columns and 49277 rows. Following are the descriptions for each column:

1. Ind_id= Indicator ID
2. Ind_definition= Definition of indicator
3. Reportyear= years that the indicator was reported, 2000 to 2013
4. Race_eth_code= numeric code for race/ethnicity
5. Race_eth_name= name of race/ethnic group
6. Geotype= type of geographic unit, PL-place, CO-county, RE- region, CA-state
7. Geotypevalue= 5-digit FIPS place code, 5-digit FIPS county code, 2-digit region ID, 2-digit FIPS state code
8. Geoname= place name, county name, region name, state name
9. County_fips= Name of county that geotype is in
10. County_name= FIPS code of county that geotype is in,2-digit census state code (06) plus 3-digit census county code
11. Region_code= region code, 1-Bay Area, 2-Butte, 3-Central/Southeast Sierra, 4-Monterey Bay, 5-North Coast, 6-Northeast Sierra, 7-Northern Sacramento Valley, 8-Sacramento Area, 9-San Diego,10-San Joaquin Valley, 11-San Luis Obispo, 12-Santa Barbara, 13-Shasta, 14-Southern California.
12. Region_name= region name
13. Strata_name_code= 1= Type of violent crime
14. Strata_name= name of the strata
15. Strata_level_name_code= Code for the strata levels, 1= Aggravated assault, 2= Forcible rape, 3= Murder and non-negligent manslaughter, 4= Robbery, 5= Violent crime total
16. Strata_level_name= name of the strata levels
17. Numerator= Number of crimes by type of crime
18. Denominator= Total population as reported by the Federal Bureau of Investigation

19. Rate= Number of violent crimes per 1,000 population, only for violent crime total
20. Ll_95ci= Lower limit of 95% confidence interval
21. Ul_95ci= Upper limit of 95% confidence interval
22. Se= Standard error of the rate
23. Rse= Relative standard error (se/percent * 100) expressed as a percent
24. Ca_decile= California decile
25. Ca_rr= Rate ratio to California rate
26. Dof_population= Total population from the California Department of Finance
27. Version= Date/time stamp of version of data

Link- https://data.ca.gov/dataset/violent-crime-rate
I have downloaded the excel file "violent-crime-rate-california-2000-2013" from this link and converted it into "crimes.csv", which is further uploaded to my google colab notebook.

Research Design

First hypothesis: Does the total violent crime rate differ in every region of California?

Null Hypothesis

Ho=, the rate of violent crimes in each region of California is the same.

Alternate Hypothesis

Ha= the rate of violent crimes in at least one region is different from others.

Second hypothesis: Does the crime rate change between the year 2009 to the year 2013?

Null Hypothesis

Ho= crime rate is different between the year 2009 to 2013

Alternate Hypothesis

Ha= crime rate is same between the year 2009 to 2013

Following tests will be used to test the hypothesis:

- Normality test will specify if the data is normally distributed or not.
- Calculate a statistically significant p-value using the T-Test or Shapiro test.
- ANOVA Test will identify whether any of the groups differs from the others.
- Compare multiple normally distributed groups using Tukey's HSD test.

Audience

Although we cannot control what goes on around us, we can try to take a few steps to aid the government, and police authorities control it. Therefore this research will be valuable for the government, the police authorities, citizens, job seekers, and students. This research will also specify which region requires more thorough investigations by the police authorities on the crimes where the crime rate is much higher. This research will help government and police authorities implement more security measures for the citizens to have a safe environment in those regions.