

STAT 435: Regression

Final Paper

Name: Chao Xue

Due: April 11, 2018

The premise of this final paper is to take what you have learned in STAT 435 and apply it to a real-world dataset that I have modified. The dataset: **communcrime.dat** relates to the total number of violent crimes per 100,000 population across various cities in the US. It is composed of a range of real-valued features relating to crime statistics, including state, proportion of ethnicity minorities, population age, SES metrics, immigration numbers, and various metrics relating to police enforcement.

ATTRIBUTE INFORMATION:

| | |
|---------------------|--|
| ViolentCrimesPerPop | total number of violent crimes per 100K population |
| state | US state (by number) |
| communityname | community name |
| racepctblack | % of population that is african american |
| racePctHisp | % of population that is of hispanic |
| agePct12t21 | % of population that is 12-21 in age |
| pctUrban | % of people living in areas classified as urban |
| medIncome | median household income |
| pctWPubAsst | % of households with public assistance income |
| PctPopUnderPov | % of people under the poverty level |
| PctUnemployed | % of people 16 and over unemployed |
| TotalPctDiv | % of population who are divorced |
| PctKids2Par | % of kids in family housing with two parents |
| PctImmigRec5 | % of immigrants who immigrated within last 5 years |
| PctImmigRec10 | % of immigrants who immigrated within last 10 years |
| PctRecImmig5 | % of population who have immigrated within the last 5 years |
| PctRecImmig10 | % of population who have immigrated within the last 10 years |
| PctNotSpeakEnglWell | % of people who do not speak English well |
| NumInShelters | number of people in homeless shelters |
| PctForeignBorn | % of people foreign born |
| PctUsePubTrans | % of people using public transit for commuting |
| LemasSwornFT | number of sworn full time police officers |
| LemasTotalReq | total requests for police |
| PolicReqPerOffic | total requests for police per police officer |
| OfficAssgnDrugUnits | number of officers assigned to special drug units |
| NumKindsDrugsSeiz | number of different kinds of drugs seized |

| | |
|---------------------|--------------------------------------|
| PolicAveOTWorked | police average overtime worked |
| LemasGangUnitDeploy | gang unit deployed |
| LemasPctOfficDrugUn | % of officers assigned to drug units |

Your goal is to perform the following tasks with the dataset presented as an analyst:

- Write up a short (2-page max) introduction describing the outcome and the relationship with the potential covariates in the dataset.
- Analyze the dataset and write up a full report describing the dataset, including Table 1, which should be a descriptive summary like that found any any publication or journal article
- Analyze and select a final regression model
- Interpret the model results
- Report issues you have found, recommendations, and other relevant pieces of information that may affect the results of the model, interpretation, etc.

The final paper, including the introduction, should be a maximum of 10-pages using font size 12, double spaced, excluding Tables and Figures. Therefore, you should be as concise as possible in your write up; however, descriptions should be in enough detail for me to reproduce your work. The paper is to be done **independent** of your classmates and any questions should be directed to me or Peter.