

Data Analysis Exam – Fall 2022

There are two datasets used for the Fall 2022 Data Analysis Exam. One data set is the daily (from 04/01/2020 to 06/29/2021) COVID19 counts at the ZIP-level in the San Diego County. Another data set is a partial demographic variables including Zip, Total_Pop, gender, race (White, Black, Asian, American_Indian, Native_Hawaiian, Other_race), Hispanic related variables (Total_Hispanic, Non_Hisp_White, Non_Hisp_Black, Non_Hisp_Asian), unemployedCvillian, PopBelowpoverty, and popUninsured. Perform data analyses using any methods that you think are appropriate for this data set. You should specify what meaningful questions can be addressed from these data regarding the COVID 19 counts. You may not need to include all variables in your analysis or restrict your analysis to any one single method. However, important variable(s) should not be omitted. Your goal should be to use your broad knowledge of applied statistical methods to glean as much knowledge as you can from these data. Analyses that are statistically sound but provide only a superficial investigation of the data will not be sufficient for a passing grade. Note the two data sets are essential and suitable for the Data Analysis Exam. If you want to gather more related variables (like Census data, spatial locations, or other resources) that you think they will help you to conduct more meaningful data analyses, you are welcome to do so. However, the exam committee recommend you to work on the two provided data sets first.