

## Math 4650 - Spring 2021

### Project #1, due on March 30, 2021

The data set to be analyzed in this project is given in the csv file named "suicides1970.csv" on UNT canvas. It contains the distribution of suicides in the United States by month in 1970. You have been asked to provide a statistical analysis of the data and write a brief report on your findings and conclusions. Use the significance level  $\alpha = 0.05$  for all the tests to be conducted. **Your analysis and report should include at least the following parts:**

1. Provide the ecdf and box plots of the number of suicides.
2. Model the number of suicides in each month as a multinomial random variable with the appropriate probabilities and conduct a goodness-of-fit statistical test to see if the data is consistent with the hypothesis that the suicide rate is constant.
3. Group the data into four seasons: Winter (December, January, February); Spring (March, April, May); Summer (June, July, August); Autumn (September, October, November). Conduct an appropriate goodness-of-fit test to see if there is any evidence that the suicide rate varies seasonally.
4. Your written report should be typed not to exceed six pages.

You may work on this project by yourself or in a group of not exceeding 3 people. Please hand in one report for the group signed by all members. Each group member will receive the same grade and will be assumed to have participated fully in the project.