STA 104: Take Home Project
Looking into States that are being affected from Covid-19

Aman Singh & Peter Mueller
amasin@ucdavis.edu
petermueller@ucdavis.edu
UC Davis
Instructor: Amy T. Kim

2/28/2021

Ⓘ **Introduction**

Covid-19 has vastly changed the way of how the world works. Millions of people in America are all unable to do their work properly and have been forced to work at home. We are currently analyzing an data set called CovidA. which lists all the months of states as will as there monthly deaths. We chose to instead subset the dataset and work with 4 states. We chose California, New York City, Florida, and Hawaii. We chose California because that is where my partner and I are located and we would like to know more about California compared to other states. We chose New York City because NYC is a typically very popular spot for visitors and we were curious on how the analysis will work with an heavily populated city with Corona. We chose Hawaii because despite the fact that is an very famous spot for tourists, people have not been going there because of heavy fines indicted by the government of Hawaii, which leaves us curious on how they have been reacting to Corona deaths. We also chose Florida largely because it is one of the only states where nobody believed in the Coronavirus. All the bars, indoor dining and tourism is rampant over there with the state not caring if anyone wears a mask. We are curious on how Florida is doing in terms of the virus compared to states that are very progressive and strict on guidelines such as California and Hawaii.

Ⓘ **Summary of Data**

Table 1: Summary Statistics

|	California	Florida	Hawaii	New York City
Group Mean	4256.750	2465.00000	59.40000	3718.200
Group SD	5673.232	1863.53750	42.52999	6302.866
Rank Mean	14.250	13.71429	3.20000	12.400
Sample Size	4.000	7.00000	5.00000	5.000

%
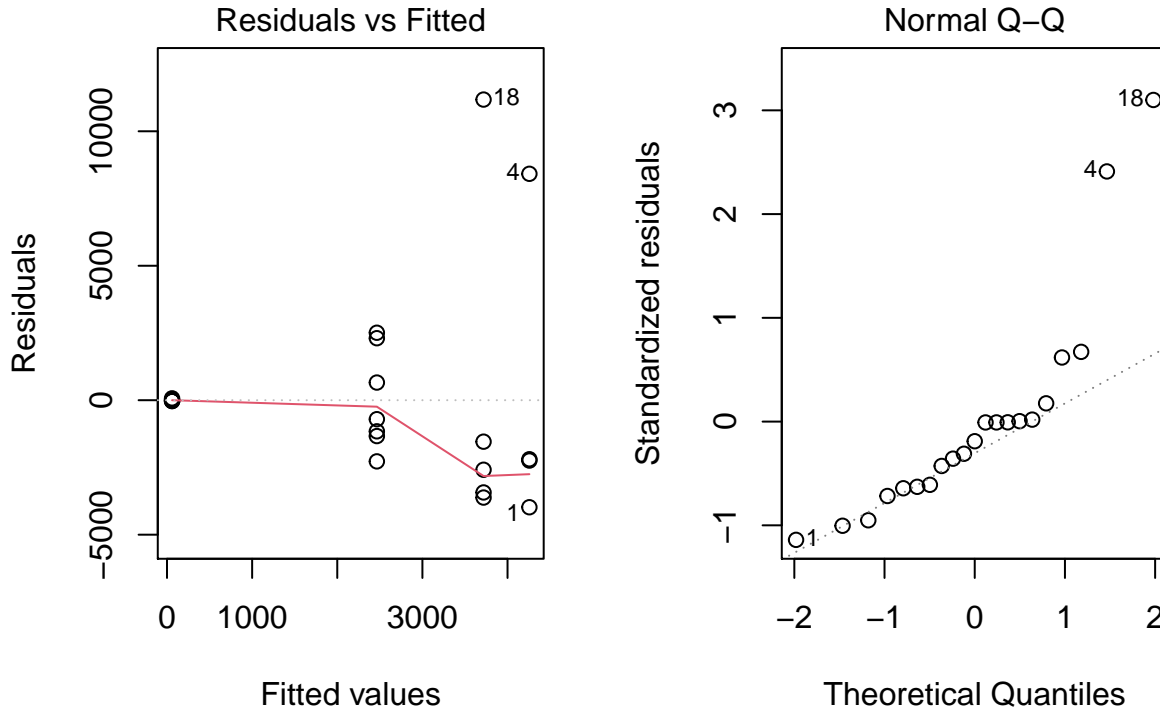
We see from the Shapiro- Wilks test based on the hypothesis that the variable is normally distributed in the dataset. We see that since the p-value is really low, we reject the null hypothesis and confirm that that the variable is not normally distributed. We also use the Levene test to test if K samples have equal variances and find that we fail to reject the null hypothesis thus concluding that no violations of any assumptions have

happened.

In this forthcoming graph, I purposely ended up changing the dataset where I made an replica of the subsetted graph that we had from the CovidA dataset. I than changed the months to where month 1 was January of 2020 and month 14 was February of 2021. This helped in viewing the graphs as there was not data for every month and when I first did the graph, specifically for NYC, there was no data of February of 202 but there was of February 2021. However, the reason why I changed it was because that data of 2021 was showing up earlier in the 2020 panel which was messing up the graph thus the reason as to why I changed the index.

Monthly Covid Deaths compared to States
Analysis of California, Florida, Hawaii & New York City Covid–19 Deaths
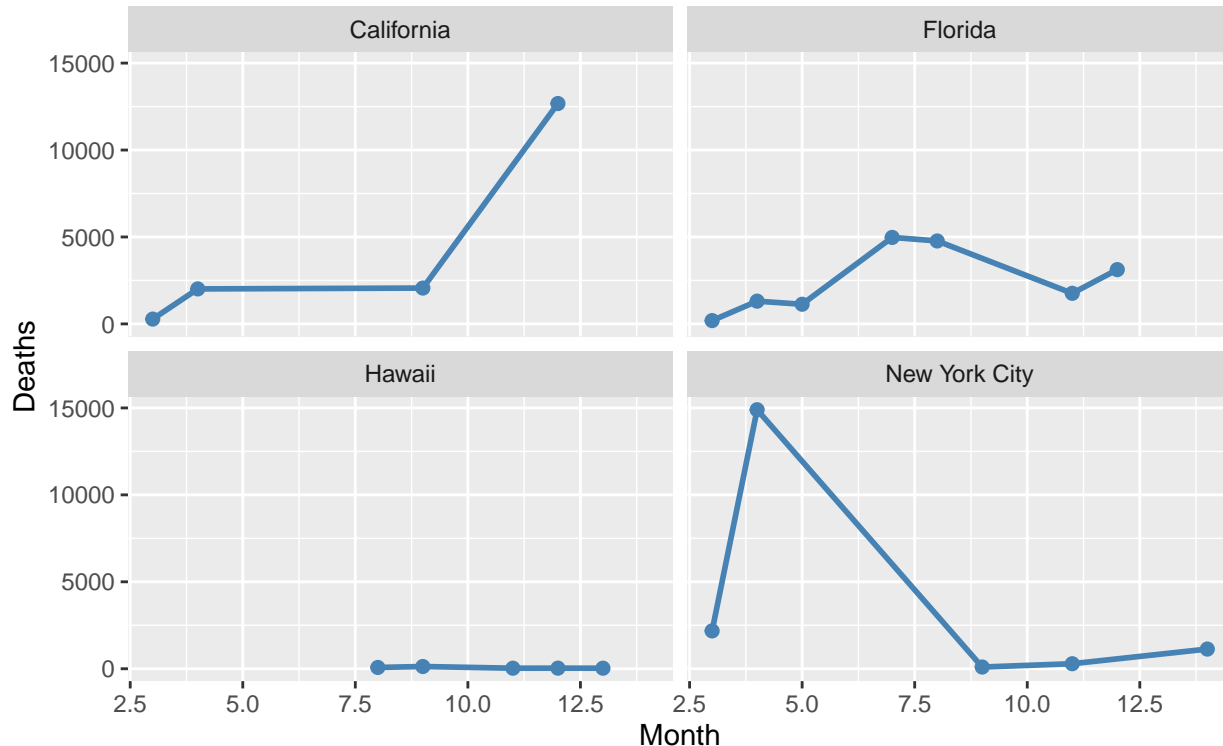
We see here that surprisingly, there were a lot of deaths in California, despite it being a much more Democratic and very progressive state. While they had it mostly low, the sudden increase in deaths was due to the fact that it was December, and people were probably visiting their friends and loved ones regardless which led to an increase in deaths.

(III) **Analysis**

(IV) **Interpretation**

(V) **Conclusion**

Code Appendix

```r
# cuttingoffcode
library(knitr)
opts_chunk$set(tidy.opts = list(width.cutoff = 70), tidy = TRUE)
# importing dataset
library(readr)
CovidA <- read_csv("CovidA.csv")
# subsetting data
CovidA_sub = subset(CovidA, CovidA$State == "California" | CovidA$State ==
    "New York City" | CovidA$State == "Florida" | CovidA$State == "Hawaii")
CovidA_sub$Rank = rank(CovidA_sub$Death, ties = "average")
Group.order = aggregate(Death ~ State, data = CovidA_sub, mean)$State
Xi = aggregate(Death ~ State, data = CovidA_sub, mean)$Death
si = aggregate(Death ~ State, data = CovidA_sub, sd)$Death
Ri = aggregate(Rank ~ State, data = CovidA_sub, mean)$Rank
ni = aggregate(Death ~ State, data = CovidA_sub, length)$Death
results = rbind(Xi, si, Ri, ni)
rownames(results) = c("Group Mean", "Group SD", "Rank Mean", "Sample Size")
colnames(results) = as.character(Group.order)
SR.2 = var(CovidA_sub$Rank)
Ano = aov(Death ~ State, data = CovidA_sub)
par(mfrow = c(1, 2))
plot(Ano, 1)
plot(Ano, 2)
shapiro.test(CovidA_sub$Death)
library(lawstat)
levene.test(CovidA_sub$Death, as.factor(CovidA_sub$State))
library(ggplot2)
library(dplyr)
# converting months as to not get confused by year so month 1 is
# January 2020 and month 14 is February 2021
CovidA_sub_graph = CovidA_sub
CovidA_sub_graph$Month[16] = 13
CovidA_sub_graph$Month[21] = 14
# ggplot graph with facetwrap
ggplot(data = CovidA_sub_graph, aes(Month, Death)) + geom_line(color = "steelblue",
    size = 1) + geom_point(color = "steelblue", size = 2) + labs(title = "Monthly Covid Deaths compared
    subtitle = "Analysis of California, Florida, Hawaii & New York City Covid-19 Deaths",
    y = "Deaths", x = "Month") + facet_wrap(~State)
```