part 5

Miles Tweed

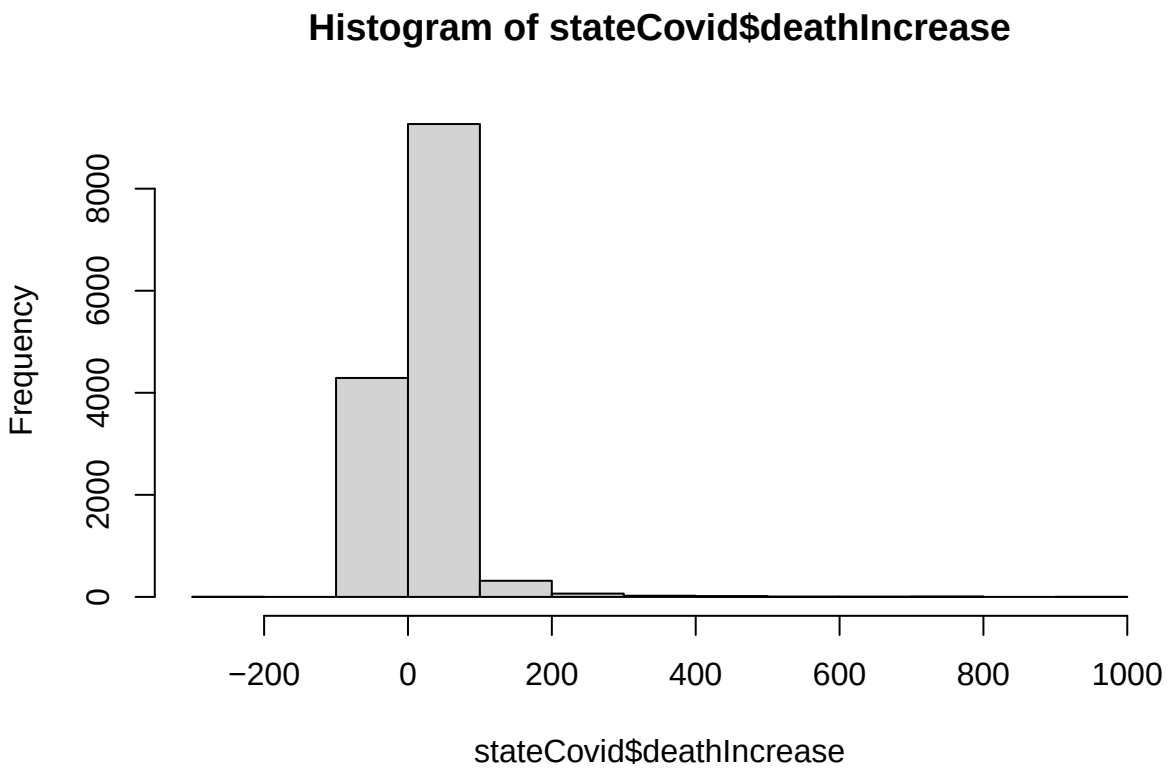
12/7/2020

Data Import

```
stateCovid <- read.csv(file = file.path(proj.dir, data.dir, "state_covid_data.csv"))
stateCovid <- stateCovid %>%
  mutate(date = lubridate::ymd(date))



stateCovid['month'] <- month(stateCovid$date)

stateCovid['week'] <- week(stateCovid$date)


# This shows that the data is skewed right with the daily increase having a median of 4 and a mean of 1
# Negative values? Perhaps corrections to cause of death?
summary(stateCovid$deathIncrease) # No NA values
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -213.00    0.00    4.00   16.38   15.00  951.00
```

```
hist(stateCovid$deathIncrease)
```



Weekly Dataset

The deathIncrease variable is the number of deaths added per day. We can use the sum to aggregate this data into weekly values.

```
stateCovidWeekly <- stateCovid %>%
  group_by(state,week) %>%
  select(state, week, deathIncrease) %>%
  summarise_at(vars(deathIncrease), funs(sum))
```

```
## Warning: `funs()` is deprecated as of dplyr 0.8.0.
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_warnings()` to see where this warning was generated.
```

```
print('Summary of deathIncrease variable for full sample population')
```

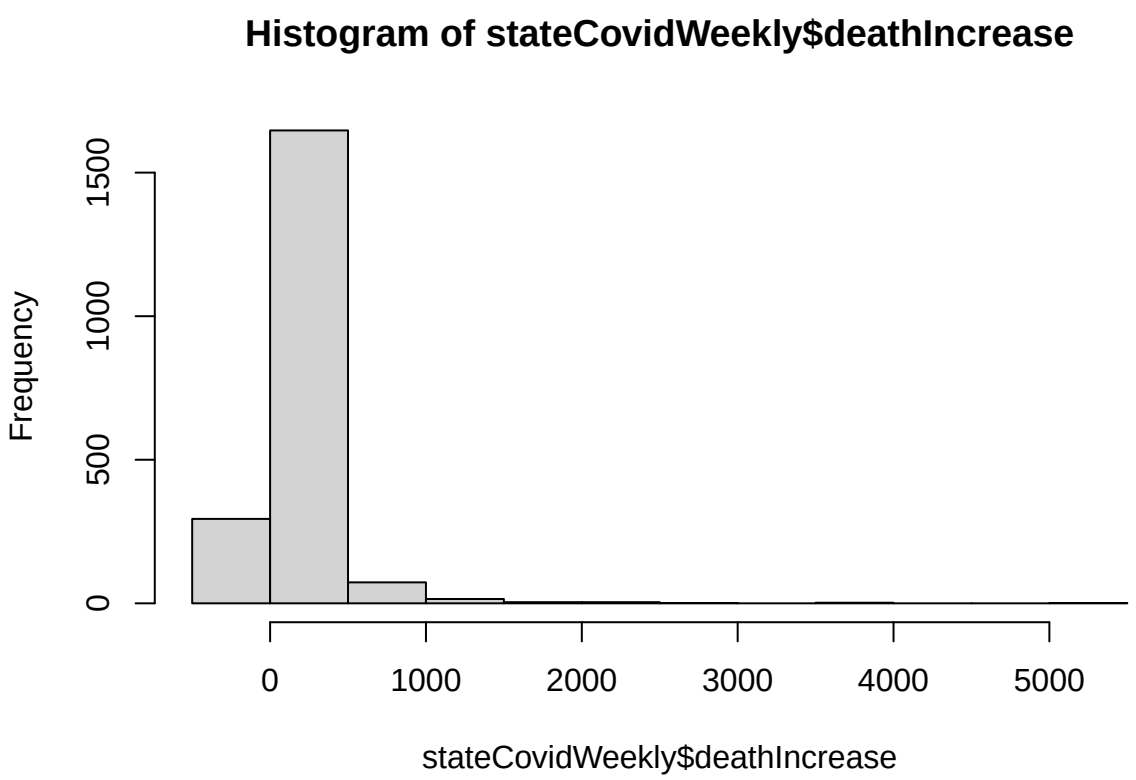
```
## [1] "Summary of deathIncrease variable for full sample population"
```

```
summary(stateCovidWeekly$deathIncrease)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
##   -66.0     5.0    38.0   112.3   106.0  5345.0
```

```
hist(stateCovidWeekly$deathIncrease)
```




```
#wk.samp.pop.med <- median(stateCovidWeekly$deathIncrease)


x.median <- function(x,i) { median(x$deathIncrease[i]) }

wk.samp.pop.med <- boot(data = stateCovidWeekly,
            statistic = x.median,
            R = 10000)



# Proportion of weeks over all states in which the increase is greater than the median increase
sam.pop.gt.med <-
  stateCovidWeekly %>%
    filter(deathIncrease > wk.samp.pop.med$t0)

state.week.num.gt.med <- length(sam.pop.gt.med$week)
state.week.num <- length(stateCovidWeekly$week)

samp.pop.prop.gt.med <- state.week.num.gt.med /state.week.num

# Unsurprisingly the proportion of values greater than the median is approximately 0.5
print(paste('Sample population median deaths per week calculated via bootstrap:',wk.samp.pop.med$t0))
```

```
## [1] "Sample population median deaths per week calculated via bootstrap: 38"
```

```
print(paste('Sample population proportion of weeks greater than median:',round(samp.pop.prop.gt.med,2)))
```

```
## [1] "Sample population proportion of weeks greater than median: 0.49"
```

Subset of states

We wish to analyse the proportion of weeks in which the death increase was above the median.

```
states <- stateCovidWeekly[,"state"] %>% unique()

set.seed(12)
st.sample.1 <- sample(states$state,8)

sam1.pop <-
  stateCovidWeekly %>%
  filter(state %in% st.sample.1)
```

Inference of population parameter

create an inference of the population parameter. This is proportion of weeks in which the death increase was above the mean or, in other words, the number of weeks in which the increase was greater than the mean divided by the total number of weeks in the sample.

```
# Proportion of weeks from sample 1 states in which the increase is greater than the median increase
sam1.pop.gt.med <-
  sam1.pop %>%
  ungroup() %>%
  filter(deathIncrease > wk.samp.pop.med$t0)

sam1.state.week.num.gt.med <- length(sam1.pop.gt.med$week)
sam1.state.week.num <- length(sam1.pop$week)

sam1.pop.prop.gt.med <- sam1.state.week.num.gt.med /sam1.state.week.num

print('Sample 1:')
```

```
## [1] "Sample 1:"
```

```
print(st.sample.1)
```

```
## [1] "AL" "MN" "ID" "MO" "SD" "AZ" "VI" "MP"
```

```
print(paste('Proportion of weeks with death increase above the median',round(sam1.pop.prop.gt.med,2)))
```

```
## [1] "Proportion of weeks with death increase above the median 0.47"
```

```
s1 <- sqrt((sam1.pop.prop.gt.med * (1 - sam1.pop.prop.gt.med))/sam1.state.week.num)

sam1.cI95.low <- sam1.pop.prop.gt.med - (1.96 * s1)
sam1.cI95.high <- sam1.pop.prop.gt.med + (1.96 * s1)

print(paste('The standard deviation for sample 1 is',round(s1,2)))
```

```
## [1] "The standard deviation for sample 1 is 0.03"
```

```r
print(paste('The 95% confidence interval for sample 1:',round(sam1.cI95.low, 2),'to',round(sam1.cI95.hig
```

```
## [1] "The 95% confidence interval for sample 1: 0.41 to 0.53"
```

Other subsets (Regional)

North East

```r
northeast.abb <- c("ME", "NH", "VT", "MA", "RI", "CT", "NY", "PA", "NJ")

set.seed(12)
st.sample.NE <- sample(northeast.abb,8)

samNE.pop <-
  stateCovidWeekly %>%
  filter(state %in% st.sample.NE)


samNE.pop.gt.med <-
  samNE.pop %>%
  ungroup() %>%
  filter(deathIncrease > wk.samp.pop.med$t0)

samNE.state.week.num.gt.med <- length(samNE.pop.gt.med$week)
samNE.state.week.num <- length(samNE.pop$week)

samNE.pop.prop.gt.med <- samNE.state.week.num.gt.med /samNE.state.week.num

print('Sample NE:')
```

```
## [1] "Sample NE:"
```

```r
print(st.sample.NE)
```

```
## [1] "NH" "NJ" "NY" "VT" "RI" "ME" "PA" "CT"
```

```r
print(paste('Sample size:', length(samNE.pop$week)))
```

```
## [1] "Sample size: 294"
```

```r
print(paste('Proportion of weeks with death increase above the median',round(samNE.pop.prop.gt.med,2)))
```

```
## [1] "Proportion of weeks with death increase above the median 0.41"
```

```r
sNE <- sqrt((samNE.pop.prop.gt.med * (1 - samNE.pop.prop.gt.med))/samNE.state.week.num)

samNE.cI95.low <- samNE.pop.prop.gt.med - (1.96 * sNE)
samNE.cI95.high <- samNE.pop.prop.gt.med + (1.96 * sNE)

print(paste('The standard deviation for sample NE is',round(sNE,2)))
```

```
## [1] "The standard deviation for sample NE is 0.03"
```

```r
print(paste('The 95% confidence interval for sample NE:',round(samNE.cI95.low,2),'to',round(samNE.cI95.h
```

```
## [1] "The 95% confidence interval for sample NE: 0.35 to 0.46"
```

Mid-West

```
midwest.abb <- c("WI", "MI", "IL", "IN", "OH", "ND", "SD", "NE", "KS", "MN", "IA", "MO")

set.seed(12)
st.sample.MW <- sample(midwest.abb,8)

samMW.pop <-
  stateCovidWeekly %>%
  filter(state %in% st.sample.MW)

samMW.pop.gt.med <-
  samMW.pop %>%
  ungroup() %>%
  filter(deathIncrease > wk.samp.pop.med$t0)

samMW.state.week.num.gt.med <- length(samMW.pop.gt.med$week)
samMW.state.week.num <- length(samMW.pop$week)

samMW.pop.prop.gt.med <- samMW.state.week.num.gt.med /samMW.state.week.num

print('Sample MW:')
```

```
## [1] "Sample MW:"
```

```
print(st.sample.MW)
```

```
## [1] "MI" "MN" "SD" "OH" "KS" "IN" "MO" "IL"
```

```
print(paste('Sample size:', length(samMW.pop$week)))
```

```
## [1] "Sample size: 290"
```

```
print(paste('Proportion of weeks with death increase above the median',round(samMW.pop.prop.gt.med,2)))
```

```
## [1] "Proportion of weeks with death increase above the median 0.71"
```

```
sMW <- sqrt((samMW.pop.prop.gt.med * (1 - samMW.pop.prop.gt.med))/samMW.state.week.num)

samMW.cI95.low <- samMW.pop.prop.gt.med - (1.96 * sMW)
samMW.cI95.high <- samMW.pop.prop.gt.med + (1.96 * sMW)

print(paste('The standard deviation for sample MW is',round(sMW,2)))
```

```
## [1] "The standard deviation for sample MW is 0.03"
```

```
print(paste('The 95% confidence interval for sample MW:',round(samMW.cI95.low,2),'to',round(samMW.cI95.
```

```
## [1] "The 95% confidence interval for sample MW: 0.65 to 0.76"
```

South

```
south.abb <- c("DE", "MD" ,"DC", "VA", "WV", "NC","SC","GA","FL","KY","TN","MS", "AL", "OK", "TX", "AR"

set.seed(12)
st.sample.S <- sample(south.abb,8)
```

```r
samS.pop <-
  stateCovidWeekly %>%
  filter(state %in% st.sample.S)

samS.pop.gt.med <-
  samS.pop %>%
  ungroup() %>%
  filter(deathIncrease > wk.samp.pop.med$t0)

samS.state.week.num.gt.med <- length(samS.pop.gt.med$week)
samS.state.week.num <- length(samS.pop$week)

samS.pop.prop.gt.med <- samS.state.week.num.gt.med /samS.state.week.num

print('Sample S:')
```

```
## [1] "Sample S:"
```

```r
print(st.sample.S)
```

```
## [1] "MD" "KY" "SC" "TN" "WV" "AL" "LA" "GA"
```

```r
print(paste('Sample size:', length(samS.pop$week)))
```

```
## [1] "Sample size: 288"
```

```r
print(paste('Proportion of weeks with death increase above the median',round(samS.pop.prop.gt.med,2)))
```

```
## [1] "Proportion of weeks with death increase above the median 0.73"
```

```r
sS <- sqrt((samS.pop.prop.gt.med * (1 - samS.pop.prop.gt.med))/samS.state.week.num)

samS.cI95.low <- samS.pop.prop.gt.med - (1.96 * sS)
samS.cI95.high <- samS.pop.prop.gt.med + (1.96 * sS)

print(paste('The standard deviation for sample S is',round(sS,2)))
```

```
## [1] "The standard deviation for sample S is 0.03"
```

```r
print(paste('The 95% confidence interval for sample S:',round(samS.cI95.low,2),'to',round(samS.cI95.high
```

```
## [1] "The 95% confidence interval for sample S: 0.67 to 0.78"
```

West

```r
west.abb <- c("ID","MT","WY","NV","UT","CO","AZ","NM","AK","WA","OR","CA","HI")

set.seed(12)
st.sample.W <- sample(west.abb,8)

samW.pop <-
  stateCovidWeekly %>%
  filter(state %in% st.sample.W)

samW.pop.gt.med <-
  samW.pop %>%
  ungroup() %>%
```

```
  filter(deathIncrease > wk.samp.pop.med$t0)

samW.state.week.num.gt.med <- length(samW.pop.gt.med$week)
samW.state.week.num <- length(samW.pop$week)

samW.pop.prop.gt.med <- samW.state.week.num.gt.med /samW.state.week.num

print('Sample W:')
```

```
## [1] "Sample W:"
```

```
print(st.sample.W)
```

```
## [1] "MT" "WA" "AZ" "UT" "CA" "NV" "HI" "WY"
```

```
print(paste('Sample size:', length(samW.pop$week)))
```

```
## [1] "Sample size: 294"
```

```
print(paste('Proportion of weeks with death increase above the median',round(samW.pop.prop.gt.med,2)))
```

```
## [1] "Proportion of weeks with death increase above the median 0.43"
```

```
sS <- sqrt((samS.pop.prop.gt.med * (1 - samS.pop.prop.gt.med))/samS.state.week.num)

samS.cI95.low <- samS.pop.prop.gt.med - (1.96 * sS)
samS.cI95.high <- samS.pop.prop.gt.med + (1.96 * sS)

print(paste('The standard deviation for sample S is',round(sS,2)))
```

```
## [1] "The standard deviation for sample S is 0.03"
```

```
print(paste('The 95% confidence interval for sample S:',round(samS.cI95.low,2),'to',round(samS.cI95.hig
```

```
## [1] "The 95% confidence interval for sample S: 0.67 to 0.78"
```