Project Stage 2

This document was written and edited by:
Nikitha Narsing
Amulya Yadagani
Rashmi Hassan Udaya Kumar
Serena Wisnewski
Eric Cortes Aguilera

University of North Carolina at Greensboro

Task 1

Member:

Nikitha Narsing:

-

Amulya Yadagani:

-

Rashmi Hassan Udaya Kumar:

- Task 2 Member task 1
  The first step in recognizing what sort of distribution to fit our data to. A couple of observations:

  - The data are positive-valued, since they are measuring the number of covid-19 cases per day.

  - The data are discrete

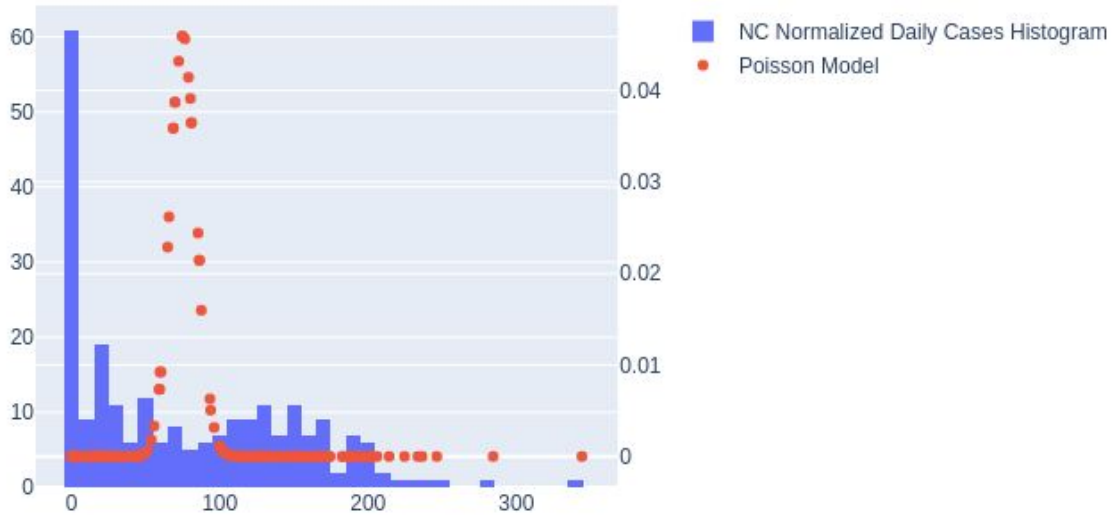
  - The data gives the number of covid-19 cases observed in period/time interval of a day

  - The probability of the occurrence of the cases each day is equiprobable.

  - The data is left skewed
  - It starts off high and has a long tail

  We know that the Poisson distribution models the probability of seeing a certain number of successes within a time interval, here we are measuring the number of covid-19 cases within a time interval of a day. Thus The Poisson distribution is a good fit

Poisson distribution the number of new cases in the state of NC



Measure of center = 75.52396557038008

Measure of spread = 5252.865980896412

Skewness = 0.6997251548048288

Kurtosis = -0.2861513970272158

Hypothesis

* Is being Male or Female a factor for higher covid cases

* Does belonging to a particular age group influence increase in covid cases

* Does belonging to a particular etnic group have a influence on the number of covid cases

Serena Wisnewski:

- Task 1
    - Chosen State: VA
        - New Case Per Day
            - Mean: 66.42
            - Median: 74.0
        - New Death Per Day:
            - Mean: 1.37
            - Median: 1.0
        - Compared to: NC, CO, AL, NY, VT
            - Besides NY's spike in March, VA COVID-19 trends are quite similar to the other states
        - Most Infected Counties:
            - Greensville County
            - Galax City
            - Emporia City
            - Northamption County
            - Franklin City

- Task 2
    - Distribution Fit for New Cases
        - Gamma: I chose gamma distribution because the new cases histogram had a positive skew, then gradually approached zero
        - Higher normalized educational acheivement results in lower normalized number of cases
        - Higher normalized broadband access results in lower normalized number of cases
        - Lower Normalized households with computers results in higher normalized cases
        -

Eric Cortes Aguilera:

- Part 1
    - After computing the weekly new cases as well as weekly new deaths for my chosen state Georgia, I obtained the following results: the case mean was 74.37, the case median was 46.0, the mode was 0.0, the death mean was 1.63, the death median was 2.0, and the death mode was 0.0. Compared to 5 states of similar population (Illinois, Pennnsylvania, Ohio, Michigan, and New Jersey) Georgia appeared to have a higher case mean then all states except Michigan. It had a higher case median than all the states except Illinois as well as Michigan, and the case mode was the same throughout all states. The death mean was less than all states except Ohio, the death median was higher than every state except Michigan, and the mode was the same throughout. This means that Georgia was one of the highest for weekly case mean but one of the lowest for weekly death mean. This may be due to less protective measures but better health care but as of now the cause is unknown.
- Part 2
    - The counties in the state of Georgia with the highest case rates per day are chattahoochee county, stewart county, echols county, appling county, atkinson county. This was accomplished by obtaining the mean daily cases and mean daily deaths for each county.

Task 2

Member tasks

Member:

Nikitha Narsing:

-

Amulya Yadagani:

-

Rashmi Hassan Udaya Kumar:

-

Serena Wisnewski:

- Task 2
    - Distribution Fit for New Cases
        - Gamma: I chose gamma distribution because the new cases histogram had a positive skew, then gradually approached zero
        - Higher normalized educational achievement results in lower normalized number of cases
        - Higher normalized broadband access results in lower normalized number of cases
        - Lower Normalized households with computers results in higher normalized cases
-

Eric Cortes Aguilera:

- Part 1
    - The distribution chosen for the cases of Georgia was the exponential distribution. This was done by first plotting the cases, then a histogram was plotted of the cases of Georgia. Once the histogram was plotted and displayed it was noticeable it followed an exponential for thus the reason why the exponential distribution was chosen. The distribution was then plotted on top of the histogram and it can be seen how it fits the graph quite nicely.
- Part 2
    - A poisson distribution model was made for both the cases and deaths for each state listed in part 2 of task 1. The state that best resembles the poisson distribution model for cases of Georgia is Illinois except that Illinois has two peaks and Georgia only has one. The death cases seemed to be all similar except for Illinois being the closest resemblance of Georgia's model.
- Part 4
    - The following table was obtained from part two of task 2:

	State	New Cases Per Day (Normalized)	New Deaths Per Day (Normalized)
0	GA	11.025809	0.247674
1	IL	8.783703	0.278009
2	MI	5.019988	0.290297
3	NJ	9.381419	0.764861
4	OH	5.019851	0.160152
5	PA	4.801392	0.260453

- The following table was obtained from part 4 of task 4:

	State	Owner-Occupied Unit	Renter-Occupied Unit	Average Household Size of Owner-Occupied Unit	Average Household Size of Renter-Occupied Unit
0	GA	2831192520	1649978358	2.666048	2.401018
1	IL	3114065055	1712294427	2.626158	2.322712
2	MI	981517725	448906677	2.583934	2.295574
3	NJ	767219481	425672145	2.656471	2.317941
4	OH	2589868692	1417129887	2.596776	2.291053
5	PA	2029265100	1075400061	2.578190	2.277586

- As can be noted georgia had the highest cases per day and as seen in the enhancement data georgia has the highest average household size for owner occupied houses and almost highest for rented occupied houses. Pennsylvania had the lowest cases per day and it had the lowest average for owner occupied houses. This points to the idea that average household size has an effect on the cases per day. Regarding the deaths, not much can be said at this point regarding these states the data seems to have no correlation.

- Part 5
  - Hypothesis
    - States that have a higher renter occupied housing will have higher cases and deaths
    - States that have a higher average household size will have higher cases and deaths

- States that have higher occupied houses will have higher cases and deaths