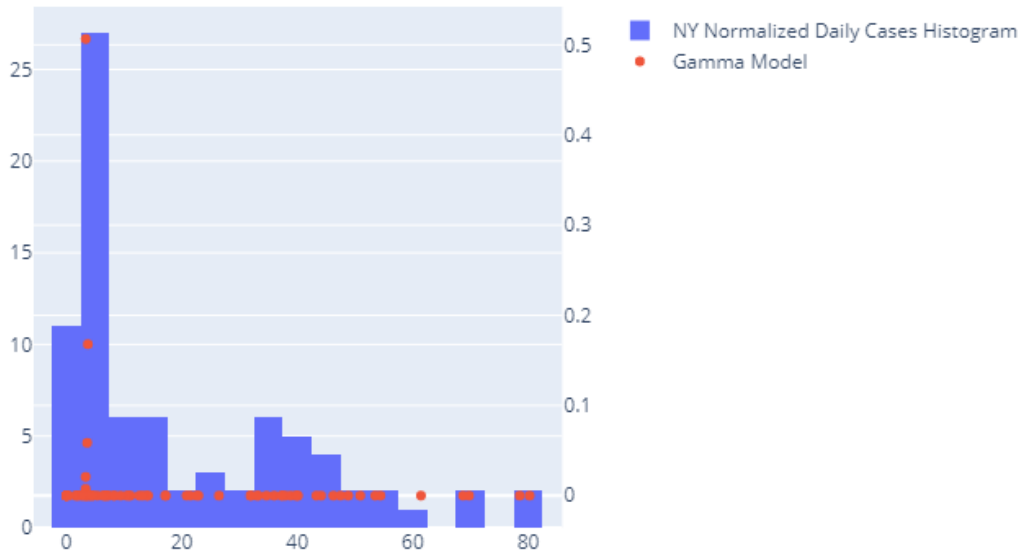


Stage 03 - Member Report

Aman Tej Vidapu, Graduate student

The first step is to plot a distribution to fit our data:

I tried plotting gamma distribution to the given data:



But, the distribution seems to be incorrect and by observing the data. we observe that:

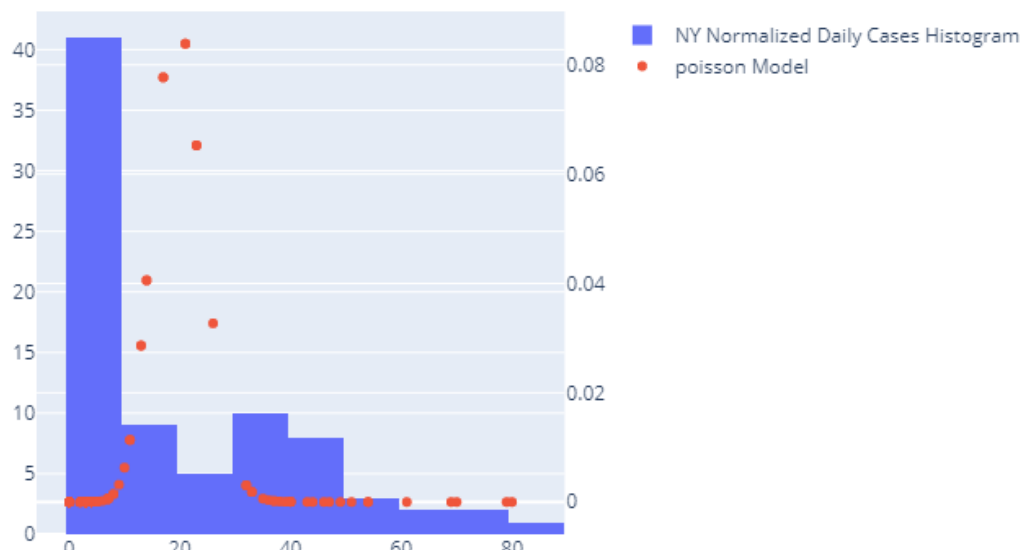
- The data is left skewed.
- It starts off high and has a long tail
- The data is positive-valued
- The data is discrete.
- The data gives the number of covid-19 cases/deaths observed in period/time interval of a day .
- Equiprobable probability of the occurrence for cases in a given timeline.

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

Statistics of the New York data:

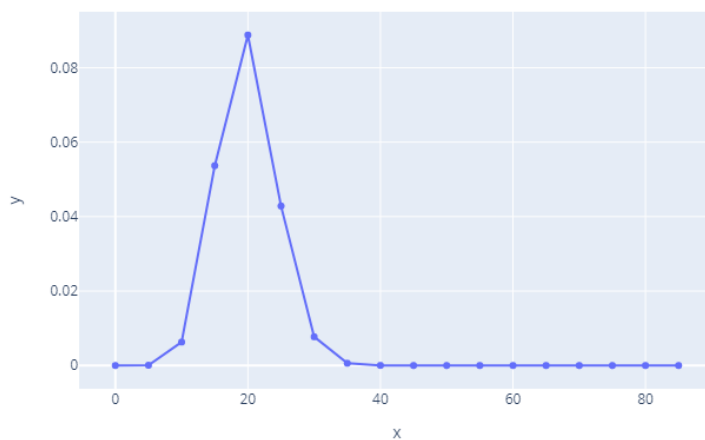
- Measure of centre: 19.843403874738804
- Measure of spread: 443.45191823532485
- Skewness: 1.1169348305588895
- Kurtosis: 0.3440613180342176

So, by observing the data we can say poisson distribution would be a good fit for the data. In order to get distribution, we have to take pmf values of the data. We can give normalised cases as data points and mu as mean of the normalised cases column.



In order to view poisson distribution clearly, Took a range(low-high) of values with step:

Poisson Distribution for Number of cases across NY



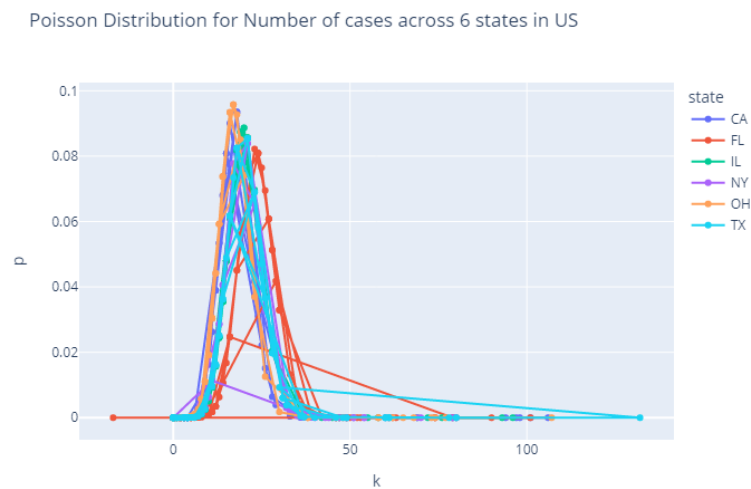
Now, comparing with 5 other states (California, Florida, Illinois, Ohio and Texas)

By Statistics:

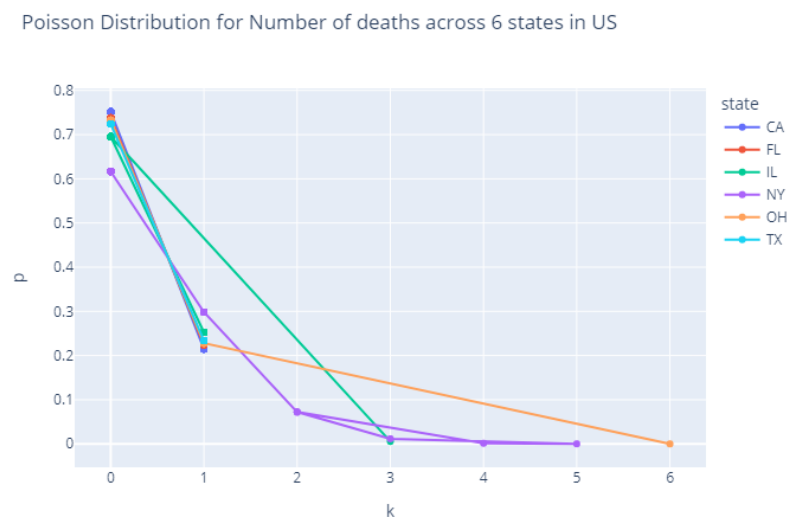
	State	mean	median	skew	var	kurtosis	Population
0	CA	18.0	8.0	2.0	639.0	4.996972	39512223
1	FL	24.0	18.0	1.0	530.0	1.355964	21477737

2	IL	20.0	14.0	2.0	467.0	3.351471	12671821
3	NY	20.0	9.0	1.0	443.0	0.344061	19453561
4	OH	17.0	9.0	2.0	500.0	3.631684	11689100
5	TX	20.0	14.0	2.0	496.0	7.448280	28995881

- By Cases (poisson distribution):



- By Deaths (poisson distribution):



Correlation:

Here we are taking three-four variables from enrichment data and grouping by state or county level as per our comparison then normalising per state/county population to understand correlation on normalised cases/deaths.

Hypothesis:

- Not having health insurance influence increase in covid cases.
- Work from home influence increase in cases
- Having private health insurance influence decreases in covid cases.