**Project Stage - III (Distributions and Hypothesis Testing)**

**Member Task – Viveka Reddy Erram**

**Fit a distribution to the number of COVID-19 new cases of a state of your choosing:**

For this task, I have selected the Florida state and Poisson distribution for plotting PMF values.

**Describe the type of distribution?**

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

**I have plotted a histogram for the normalized number of cases in Florida:**

Chart, histogram

Description automatically generated

**Calculating the PMF values using the Poisson distribution:**

**Chart, histogram

Description automatically generated**

**Poisson Distribution For number of cases across FL:**

**Line chart

Description automatically generated with low confidence**

**Statistics:**

**Table

Description automatically generated**

**Observations:**

* Mean is greater than median, so we can say it is positively skewed or right skewed distribution.
* Here we can see the distribution is not uniform.
* The Kurtosis value is positive, so the peak is high in the data.

From the above observations, we found the number of covid-19 cases per day is discrete and the distribution of the data is positively skewed with the tail to the right, thus we can say Poisson distribution is the best fit.

**Model a Poisson distribution of new COVID-19 cases and deaths of a state and compare to other 5 state?**

plotting the Poisson distribution for the normalized cases for the 6 states (FL, CA, NC,OH,TX,VA)

A picture containing table

Description automatically generated

plotting the Poisson distribution for the normalized Deaths for the 6 states (FL, CA, NC,OH,TX,VA)

**Chart, line chart, scatter chart

Description automatically generated**

**Perform correlation between Enrichment data variables and COVID-19 cases to observe any patterns.**

In this task, I have chosen large covid data and Employment enrichment data and merged them.

Later I have normalized the total number of cases and deaths per 100000 population in each county in the FL state which I have chosen.

I have selected three variables to correlate with Average Weekly Wage, Total Quarterly wages and ownership.

**Observations and Hypothesis:**

#### In Florida state, I have observed that the correlation between Total Quarterly wages, average weekly wage, establishment count.

There was a positive correlation with case rates and a negative correlation with death rates in locations with paid employees at any time during the year, and rates of cases and deaths differed in the sense that there was a positive correlation with case rates and a negative correlation with death rates. This illustrates that more employment data leads to more covid instances, and death rates have a faster increase in numbers when it comes to the number of establishments.

* If states have more Federal Government Ownership, then there is a lower chance of a greater number of Cases than Private ownership.
* Higher paid states with Total Quarterly Wages on the higher side in Private industry can indicate more numbers in Covid cases and/ death.
* Covid case numbers are greater if states had more employment in February in private ownership.