Gurjus Singh

MSDS 401: Applied Stats With R

The Rise of COVID-19

1.)

Some summary statistics I found for the data set are that the sum for all the cases in the dataset was 22,977,399 for the entire year since COVID started.

A close up of a sign

Description automatically generated

The total deaths were 800,321.

A picture containing food

Description automatically generated

This makes a fatality rate of 800,321/22,977,399 which equals 3.48%.

The standard deviation of the number of cases 3,773.715

A close up of a sign

Description automatically generated

The max number of deaths is 4,928 while the minimum is -1918

A close up of a logo

Description automatically generated

A close up of a logo

Description automatically generated

The max of the dataset of the number of cases for 1 day is 78,427

A picture containing clock

Description automatically generated

The minimum of the cases per day is -2461

A picture containing object, clock

Description automatically generated

The fatality rate for the USA is at 3.11% that is deaths over total cases makes a 3% death rate.

A close up of a logo

Description automatically generated

In my first data visualization I show a line graph of the cases per USA starting in January and going till August. As you can see the number of cases rise and now starting to ease out as we enter August.

A screenshot of a cell phone

Description automatically generated

Here are histograms for deaths and cases for each country over all the days from January – August

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generated

2.)

Two countries I got the 95% confidence interval for using t.test are India and USA cases and deaths.

For India the confidence interval deaths and cases are below:

Cases (10748.44, 16059.68)

A screenshot of a cell phone

Description automatically generated

Deaths(207, 295)

A screenshot of a cell phone

Description automatically generated

USA:

Deaths (682, 890)

A screenshot of a cell phone

Description automatically generated

Cases (22313, 28123)

A screenshot of a cell phone

Description automatically generated

3.)

The relationship is between cases and death rate is mildly correlated it is not highly correlated in that the correlation is not close to 1 but in the middle at 0.738.

A picture containing drawing

Description automatically generated

A close up of a logo

Description automatically generated

4.)

Time series of number of cases

A screenshot of a cell phone

Description automatically generated

Time series of number of deaths

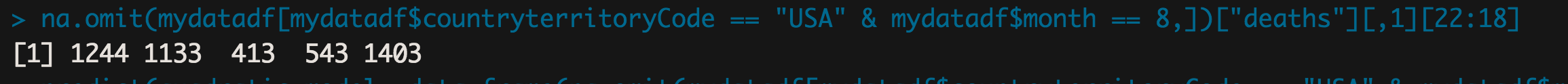
A screenshot of a cell phone

Description automatically generated

Here is the how the model fits the data using the lm function and also making it non-linear.

A close up of a logo

Description automatically generated

Training data based on deaths 5 days

Prediction of next five days versus actual data

A close up of a sign

Description automatically generated