CSE 544 Project Spring 2022

Team number: 13

Team members:

Tirth Mehta (114343246)

Anshuman Funkwal (113141368)

Karan Sheth (114362694)

Pranave Sethuraj (113275364)

Submit a Github link in the [Google Sheet](). Please access the sheet from your cs account. Make your repo private, visible only to your group members. After the submission deadline, do not make any commits to the repo. Otherwise, you will lose points. Only AFTER the deadline, you will be asked to make the repo public so that TAs can grade your project. The github repo should include (i) all your py code with self-explanatory filenames and required comments; (ii) pdf of required answers, plots, and results of hypothesis (TAs are not required to run your code so all results should be in pdf); and (iii) your X dataset.

Q1)

<u>Cases dataset:</u>

The cumulative total death and cases column are processed (For each state, we subtracted the consecutive rows to get the value for a particular day) before outlier and other analysis

Outlier analysis:

MT_total_death:

Number of outliers: 76
Q1: 0.0
Q3: 5.0
Q1 − 1.5*IQR : -7.5
Q3 + 1.5*IQR : 12.5

MN_total_death

Number of outliers: 75
Q1: 1.0
Q3: 17.0
Q1 − 1.5*IQR : -23.0
Q3 + 1.5*IQR : 41.0

MT_total_cases

Number of outliers: 79
Q1: 2.0
Q3: 230.25
Q1 − 1.5*IQR : -340.375
Q3 + 1.5*IQR : 572.625

MN_total_cases
Number of outliers: 23
Q1: 0.0
Q3: 970.0
Q1 − 1.5*IQR : -1455.0

Q3 + 1.5*IQR : 2425.0
Here, MT and MN represent the states Montana and Minnesota respectively.
After preprocessing the data and performing outlier analysis, we observed few negative values in the dataset. As negative values for the number of cases and deaths does not make sense, we replaced them with zero values.

Q2(a)

<u>Wald's Test</u>

Montana Daily Death

Null hypothesis (H0): MT mean of february month death = MT mean of march month death
Alternate hypo (H1): MT mean of february month death != MT mean of march month death

Sample mean death for feb month: 2.1923076923076925
Sample mean death for mar month: 2.103448275862069
W: 0.3299411905963097
Accept the Null hypothesis

Montana Daily Cases

Null hypothesis (H0): MT mean of february month cases = MT mean of march month cases
Alternate hypo (H1): MT mean of february month cases != MT mean of march month cases

Sample mean cases for feb month: 202.5
Sample mean cases for mar month: 148.58620689655172
W: 19.31852871948626
Reject the Null hypothesis

Minnesota Daily Death

Null hypothesis (H0): MN mean of february month death = MN mean of march month death
Alternate hypo (H1): MN mean of february month death != MN mean of march month death

Sample mean death for feb month: 9.653846153846153
Sample mean death for mar month: 7.9655172413793105
W: 2.7707257763569073
Reject the Null hypothesis

Minnesota Daily Cases

Null hypothesis (H0): MN mean of february month cases = MN mean of march month cases
Alternate hypo (H1): MN mean of february month cases != MN mean of march month cases

Sample mean cases for feb month: 828.8076923076923
Sample mean cases for mar month: 1133.9655172413793
W: 54.048581213448465
Reject the Null hypothesis

<u>Z-Test</u>
Montana Daily Death

Null hypothesis (H0): MT mean of february month death = MT mean of march month death
Alternate hypo (H1): MT mean of february month death != MT mean of march month death

Sample mean death for feb month: 2.1923076923076925
Sample mean death for mar month: 2.103448275862069
Z: 0.9925760422917541
Accept the Null hypothesis

Montana Daily Cases

Null hypothesis (H0): MT mean of february month cases = MT mean of march month cases
Alternate hypo (H1): MT mean of february month cases != MT mean of march month cases

Sample mean cases for feb month: 202.5
Sample mean cases for mar month: 148.58620689655172
Z: 10.11618971404372
Reject the Null hypothesis

Minnesota Daily Death

Null hypothesis (H0): MN mean of february month death = MN mean of march month death
Alternate hypo (H1): MN mean of february month death != MN mean of march month death

Sample mean death for feb month: 9.653846153846153
Sample mean death for mar month: 7.9655172413793105
Z: 5.243596730030624
Reject the Null hypothesis

Minnesota Daily Cases

Null hypothesis (H0): MN mean of february month cases = MN mean of march month cases
Alternate hypo (H1): MN mean of february month cases != MN mean of march month cases

Sample mean cases for feb month: 828.8076923076923
Sample mean cases for mar month: 1133.9655172413793
Z: 12.810418919546303
Reject the Null hypothesis

T-Test

Montana Daily Death

Null hypothesis (H0): MT mean of february month death = MT mean of march month death
Alternate hypo (H1): MT mean of february month death != MT mean of march month death

Sample mean death for feb month: 2.1923076923076925
Sample mean death for mar month: 2.103448275862069
Alpha: 0.05 , n: 26
T: 0.27220405470995673
Accept the Null hypothesis

Montana Daily Cases

Null hypothesis (H0): MT mean of february month cases = MT mean of march month cases
Alternate hypo (H1): MT mean of february month cases != MT mean of march month cases

Sample mean cases for feb month: 202.5
Sample mean cases for mar month: 148.58620689655172
Alpha: 0.05 , n: 26
T: 3.30230309416593
Reject the Null hypothesis

Minnesota Daily Death

Null hypothesis (H0): MN mean of february month death = MN mean of march month death
Alternate hypo (H1): MN mean of february month death != MN mean of march month death

Sample mean death for feb month: 9.653846153846153
Sample mean death for mar month: 7.9655172413793105
Alpha: 0.05 , n: 26
T: 1.3952053035678755
Accept the Null hypothesis

Minnesota Daily Cases

Null hypothesis (H0): MN mean of february month cases = MN mean of march month cases
Alternate hypo (H1): MN mean of february month cases != MN mean of march month cases

Sample mean cases for feb month: 828.8076923076923
Sample mean cases for mar month: 1133.9655172413793
Alpha: 0.05 , n: 26
T: 7.188516389816137

Reject the Null hypothesis

<u>Wald's Two sample Test</u>
Montana Daily Death

Null hypothesis (H0): MT mean of february month death = MT mean of march month death
Alternate hypo (H1): MT mean of february month death != MT mean of march month death

Sample mean death for feb month: 2.1923076923076925
Sample mean death for mar month: 2.103448275862069
W: 0.2243665520023155
Accept the Null hypothesis

Montana Daily Cases

Null hypothesis (H0): MT mean of february month cases = MT mean of march month cases
Alternate hypo (H1): MT mean of february month cases != MT mean of march month cases

Sample mean cases for feb month: 202.5
Sample mean cases for mar month: 148.58620689655172
W: 15.003792394225187
Reject the Null hypothesis

Minnesota Daily Death

Null hypothesis (H0): MN mean of february month death = MN mean of march month death
Alternate hypo (H1): MN mean of february month death != MN mean of march month death

Sample mean death for feb month: 9.653846153846153
Sample mean death for mar month: 7.9655172413793105
W: 2.1006286056628
Reject the Null hypothesis

Minnesota Daily Cases

Null hypothesis (H0): MN mean of february month cases = MN mean of march month cases
Alternate hypo (H1): MN mean of february month cases != MN mean of march month cases

Sample mean cases for feb month: 828.8076923076923
Sample mean cases for mar month: 1133.9655172413793
W: 36.22080596026721
Reject the Null hypothesis

Two-sample unpaired t-test

Montana Daily Death

Null hypothesis (H0): MT mean of february month death = MT mean of march month death
Alternate hypo (H1): MT mean of february month death != MT mean of march month death

Sample mean death for feb month: 2.1923076923076925
Sample mean death for mar month: 2.103448275862069
Alpha: 0.05 , n+m-2: 53
T: 0.1401015997350352
Accept the Null hypothesis

Montana Daily Cases

Null hypothesis (H0): MT mean of february month cases = MT mean of march month cases
Alternate hypo (H1): MT mean of february month cases != MT mean of march month cases

Sample mean cases for feb month: 202.5
Sample mean cases for mar month: 148.58620689655172
Alpha: 0.05 , n+m-2: 53
T: 2.358955563502015
Reject the Null hypothesis

Minnesota Daily Death

Null hypothesis (H0): MN mean of february month death = MN mean of march month death
Alternate hypo (H1): MN mean of february month death != MN mean of march month death

Sample mean death for feb month: 9.653846153846153
Sample mean death for mar month: 7.9655172413793105
Alpha: 0.05 , n+m-2: 53
T: 1.0437534491381708
Accept the Null hypothesis

Minnesota Daily Cases

Null hypothesis (H0): MN mean of february month cases = MN mean of march month cases
Alternate hypo (H1): MN mean of february month cases != MN mean of march month cases

Sample mean cases for feb month: 828.8076923076923
Sample mean cases for mar month: 1133.9655172413793
Alpha: 0.05 , n+m-2: 53
T: 3.9462919788862987
Reject the Null hypothesis

Q2(b)

1 Sample KS Test

Computed parameters of Poisson, Geometric and Binomial using Oct-Dec 2021 data(both for cases and deaths) of 1st State MN using MME.

Poisson lambda Guess: 232.20689655172413
Geometric_p Guess: 0.004306504306504306
Binomial_n Guess: -0.26032489378300255 **Binomial_p Guess**:-891.988826643999

Used Oct-Dec 2021 data(both cases and deaths) of 2nd State MT as Test Data to check for equality with Poisson, Geometric and Binomial Distribution.

Results:

a. **1-sample KS Test for cases**

   Poisson Distribution
   H0: CDF of MT cases = CDF of Poisson Distribution with parameter (232.20689655172413)
   H1: CDF of MT cases != CDF of Poisson Distribution with parameter (232.20689655172413)
   Since 0.75(KS Statistic) > 0.05(Critical Value), **we reject H0**.

   Geometric Distribution
   H0: CDF of MT cases = CDF of Geometric Distribution with parameter (0.004306504306504306)
   H1: CDF of MT cases != CDF of Geometric Distribution with parameter (0.004306504306504306)
   Since 0.6293358198547704(KS Statistic) > 0.05(Critical Value), **we reject H0.**

   Binomial Distribution
   H0: CDF of MT cases = CDF of Binomial Distribution with parameter ((-0.26032489378300255, -891.988826643999))
   H1: CDF of MT cases != CDF of Binomial Distribution with parameter ((-0.26032489378300255, -891.988826643999))
   Since 1.0(KS Statistic) > 0.05(Critical Value), **we reject H0**.

b. **1-sample KS Test for deaths**

   Poisson Distribution

H0: CDF of MT deaths = CDF of Poisson Distribution with parameter (1.5862068965517242)
H1: CDF of MT deaths != CDF of Poisson Distribution with parameter (1.5862068965517242)
Since 0.25640942597879135(KS Statistic) > 0.05(Critical Value), **we reject H0.**

Geometric Distribution
H0: CDF of MT deaths = CDF of Geometric Distribution with parameter (0.6304347826086957)
H1: CDF of MT deaths != CDF of Geometric Distribution with parameter (0.6304347826086957)
Since 0.3333333333333333(KS Statistic) > 0.05(Critical Value), **we reject H0**.

Binomial Distribution
H0: CDF of MT deaths = CDF of Binomial Distribution with parameter ((-0.30022701475595914, -5.2833583208395805))
H1: CDF of MT deaths != CDF of Binomial Distribution with parameter ((-0.30022701475595914, -5.2833583208395805))
Since 1.0(KS Statistic) > 0.05(Critical Value), **we reject H0**.

2 Sample KS Test

D1 = Oct-Dec Data of 1st State MN
D2 = Oct-Dec Data of 2nd State MT

Results

a. **2 sample KS Test for field cases**

H0: CDF of MN cases = CDF of MT cases
H1: CDF of MN cases != CDF of MT cases
0.2068965517241379 > 0.05, therefore reject H0

b. **2 sample KS Test for field deaths**

H0: CDF of MN deaths = CDF of MT deaths
H1: CDF of MN deaths != CDF of MT deaths
0.2068965517241379 > 0.05, therefore reject H0

Permutation Test

D1 = Oct-Dec Data of 1st State MN

D2 = Oct-Dec Data of 2nd State MT

<u>Results</u>

a. H0: D1(MN cases) = D2(MT cases)
H1: D1(MN cases) != D2(MT cases)
p value : 0.168
0.168 > 0.05, therefore accept H0

b. H0: D1(MN deaths) = D2(MT deaths)
H1: D1(MN deaths) != D2(MT deaths)
p value : 0.075
0.075 > 0.05, therefore accept H0

Q2(c)

For the first 28 days of data, we estimated the lambda value using MME for the Poisson distribution. We estimated beta as 1/lambda and substituted it in the previous using this lambda. When we discover the posterior, the distribution is gamma, and we determine the powers of exponential and lambda in each loop and put them in a table. The gamma distribution was then plotted with MAP = alpha/beta.

posterior $\propto$ Likelihood $\times$ prior

$$\text{posterior} \propto \prod_{i=1}^{n} e^{-\lambda} \cdot \frac{\lambda^{x_i}}{x_i!} \times \frac{e^{-\lambda/\beta}}{\beta}$$

$$\text{posterior} \propto e^{-n\lambda} \cdot \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \times \frac{e^{-\lambda/\beta}}{\beta}$$

$$\text{posterior} \propto e^{(-n\lambda - \lambda/\beta)} \cdot \lambda^{\sum_{i=1}^{n} x_i} \times C$$

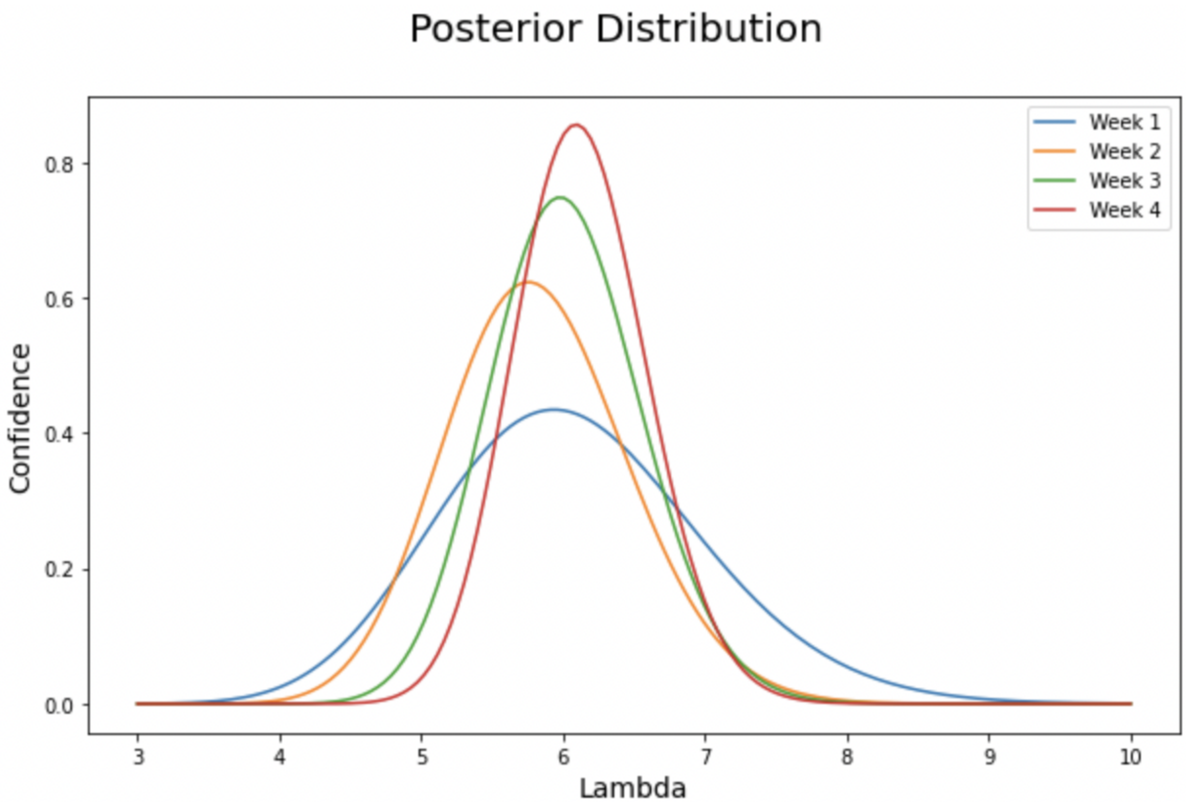
$$\approx \text{Gamma}(\alpha, v) = \frac{v^{\alpha} \cdot \lambda^{\alpha-1}}{\Gamma(\alpha)} \cdot e^{-v\lambda}$$

$$\therefore \alpha - 1 = \sum_{i=1}^{n} x_i, \quad V = n + \frac{1}{\beta}$$

Conjugate :

$$\text{Gamma}\left(\sum_{i=1}^{n} x_i + 1, \left(wn + \frac{1}{\beta}\right)\right)$$

where w = week no.



MAP values

Week-1 Lambda (X-Axis) 5.959731543624161	Confidence(Y-axis) 0.4345305575622842
Week-2 Lambda (X-Axis) 5.771812080536913	Confidence(Y-axis) 0.6230489705440925
Week-3 Lambda (X-Axis) 5.959731543624161	Confidence(Y-axis) 0.748060442642361
Week-4 Lambda (X-Axis) 6.100671140939598	Confidence(Y-axis) 0.8559050696700943

1. The above graph shows that after week 1 as the weeks progresses, the number of deaths are increasing and thus the MAP for the Lambda parameter is increasing.
2. Confidence in MAP value is increasing as the weeks are increasing.

Q2(d)

Montana State

MSE for Vaccination in Montana State with AutoRegression(3) is 4156566.273026713
MSE for Vaccination in Montana State with AutoRegression(5) is 8420263.939006817
MSE for Vaccination in Montana State with EWMA(0.5) is 3179076.577316552
MSE for Vaccination in Montana State with EWMA(0.8) is 3628224.748761411

MAPE for Vaccination in Montana State with AutoRegression(3) is 129.70332215458674
MAPE for Vaccination in Montana State with AutoRegression(5) is 187.47534904433886
MAPE for Vaccination in Montana State with EWMA(0.5) is 33.96585786448997
MAPE for Vaccination in Montana State with EWMA(0.8) is 40.98141845554573

Minnesota State

Assumption as suggested by TA on Piazza - Removed rows where distribution of vaccine is zero in the fourth week of may because it will give division by zero error while calculating MAPE

MSE for Vaccination in Minnesota State with AutoRegression(3) is 552701626.9075754
MSE for Vaccination in Minnesota State with AutoRegression(5) is 950049402.0830244
MSE for Vaccination in Minnesota State with EWMA(0.5) is 1301349510.7892516
MSE for Vaccination in Minnesota State with EWMA(0.8) is 1818024694.002908

MAPE for Vaccination in Minnesota State with AutoRegression(3) is 59.39017000731837
MAPE for Vaccination in Minnesota State with AutoRegression(5) is 77.44296257089414
MAPE for Vaccination in Minnesota State with EWMA(0.5) is 95.69075502666378
MAPE for Vaccination in Minnesota State with EWMA(0.8) is 101.27595925528125

Q2(e)

Paired T-test for September 2021 between two States

1) Paired t-test for mean of vaccine distribution between Minnesota and Montana is 5.448478346062849 which is greater than t_value(2.043) hence reject null hypothesis

Paired T-test for November 2021 between two States

2) Paired t-test for mean of vaccine distribution between Minnesota and Montana is 3.728625815262307 which is greater than t_value(2.043) hence reject null hypothesis

Q3 Exploratory Dataset

First Inference Part A

H0 : Passenger Count Pre Covid Lockdown(Jan-Feb 2020) = Passenger Count Post Covid Lockdown(2 month durations starting June 2020 to present)

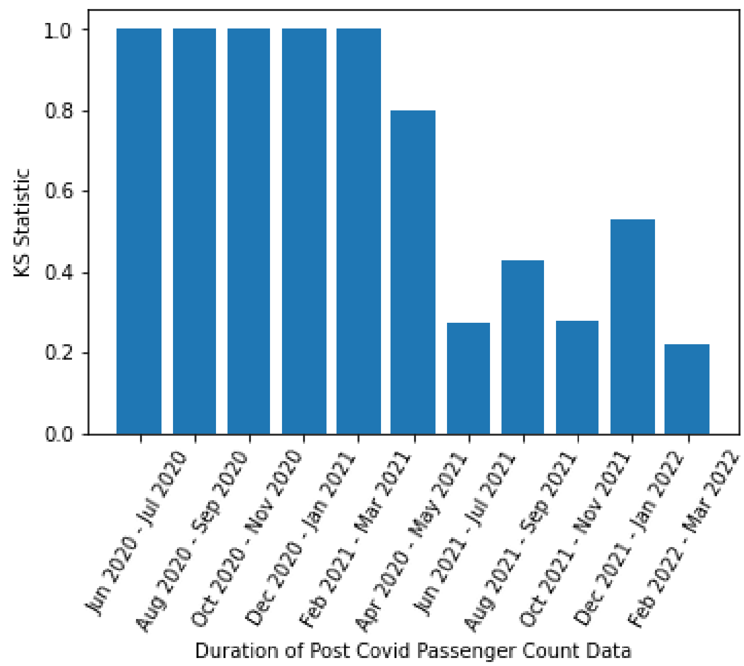
H1 : Passenger Count Pre Covid Lockdown(Jan-Feb 2020) != Passenger Count Post Covid Lockdown(2 month durations starting June 2020 to present)

Lockdown was imposed in most of the US states between March - May 2020.

D1 = Passenger Count Pre Covid Lockdown(Jan-Feb 2020)
D2 = Passenger Count Post Covid Lockdown(2 month durations starting June 2020 to present)

Results:

The passenger count starts becoming similar to the count pre covid lockdown as the time post lockdown lifted increases. This is inferred from the decreasing value of KS statistic in the graph below.



H0: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) = CDF of Post Covid Passenger Count Data(Jun 2020 - Jul 2020)
H1: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) != CDF of Post Covid Passenger Count Data(Jun 2020 - Jul 2020)

KS Statistic : 1.0
1.0 > 0.05, therefore reject H0

H0: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) = CDF of Post Covid Passenger Count Data(Aug 2020 - Sep 2020)
H1: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) != CDF of Post Covid Passenger Count Data(Aug 2020 - Sep 2020)

KS Statistic : 1.0
1.0 > 0.05, therefore reject H0

H0: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) = CDF of Post Covid Passenger Count Data(Oct 2020 - Nov 2020)

H1: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) != CDF of Post Covid Passenger Count Data(Oct 2020 - Nov 2020)

KS Statistic : 1.0
1.0 > 0.05, therefore reject H0

H0: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) = CDF of Post Covid Passenger Count Data(Dec 2020 - Jan 2021)
H1: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) != CDF of Post Covid Passenger Count Data(Dec 2020 - Jan 2021)

KS Statistic : 1.0
1.0 > 0.05, therefore reject H0

H0: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) = CDF of Post Covid Passenger Count Data(Feb 2021 - Mar 2021)
H1: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) != CDF of Post Covid Passenger Count Data(Feb 2021 - Mar 2021)

KS Statistic : 1.0
1.0 > 0.05, therefore reject H0

H0: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) = CDF of Post Covid Passenger Count Data(Apr 2020 - May 2021)
H1: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) != CDF of Post Covid Passenger Count Data(Apr 2020 - May 2021)

KS Statistic : 0.7983050847457639
0.7983050847457639 > 0.05, therefore reject H0

H0: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) = CDF of Post Covid Passenger Count Data(Jun 2021 - Jul 2021)
H1: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) != CDF of Post Covid Passenger Count Data(Jun 2021 - Jul 2021)

KS Statistic : 0.27507641011392125
0.27507641011392125 > 0.05, therefore reject H0

H0: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) = CDF of Post Covid Passenger Count Data(Aug 2021 - Sep 2021)
H1: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) != CDF of Post Covid Passenger Count Data(Aug 2021 - Sep 2021)

KS Statistic : 0.4292859127535431

0.4292859127535431 > 0.05, therefore reject H0

H1: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) = CDF of Post Covid Passenger Count Data(Oct 2021 - Nov 2021)
H0: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) != CDF of Post Covid Passenger Count Data(Oct 2021 - Nov 2021)

KS Statistic : 0.2786885245901639
0.2786885245901639 > 0.05, therefore reject H0

H0: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) = CDF of Post Covid Passenger Count Data(Dec 2021 - Jan 2022)
H1: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) != CDF of Post Covid Passenger Count Data(Dec 2021 - Jan 2022)

KS Statistic : 0.5273373428102784
0.5273373428102784 > 0.05, therefore reject H0

H0: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) = CDF of Post Covid Passenger Count Data(Feb 2022 - Mar 2022)
H1: CDF of Pre Covid Passenger Count Data(Jan-Feb 2020) != CDF of Post Covid Passenger Count Data(Feb 2022 - Mar 2022)

KS Statistic : 0.2203389830508476
0.2203389830508476 > 0.05, therefore reject H0

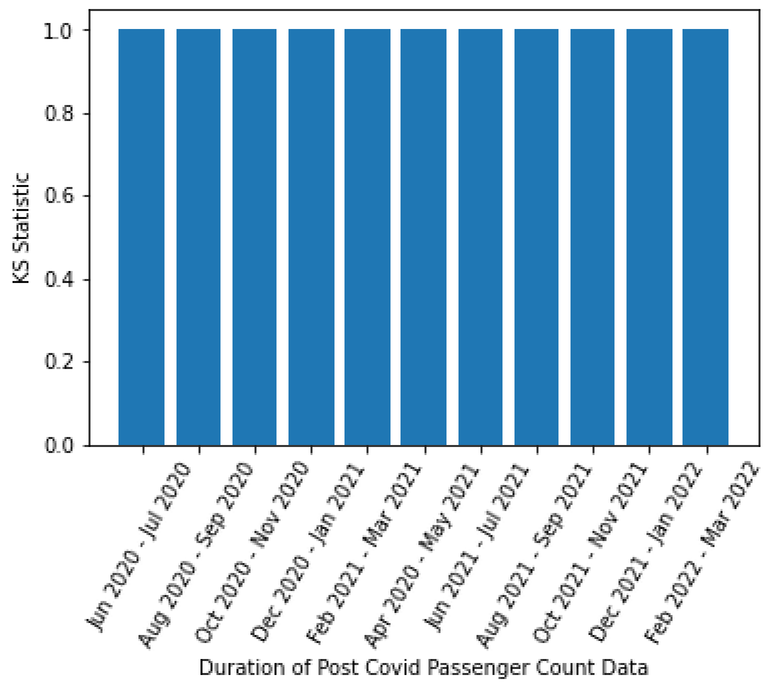
First Inference Part B

Data used from June 2020 to present

H0 : US Covid cases Count in 2 month duration = Passenger Count Post Covid Lockdown(2 month duration)

H1 : US Covid cases Count in 2 month duration != Passenger Count Post Covid Lockdown(2 month duration)

Result : As expected the two datasets should not be similarly distributed. This trend is maintained throughout.



H0: CDF of Total Cases in Jun 2020 - Jul 2020 = CDF of Post Covid Passenger Count Data(Jun 2020 - Jul 2020)
H0: CDF of Total Cases in Jun 2020 - Jul 2020 != CDF of Post Covid Passenger Count Data(Jun 2020 - Jul 2020)

KS Statistic : 1.0000000000000007
1.0000000000000007 > 0.05, therefore reject H0

H0: CDF of Total Cases in Aug 2020 - Sep 2020 = CDF of Post Covid Passenger Count Data(Aug 2020 - Sep 2020)
H0: CDF of Total Cases in Aug 2020 - Sep 2020 != CDF of Post Covid Passenger Count Data(Aug 2020 - Sep 2020)

KS Statistic : 1.0000000000000007
1.0000000000000007 > 0.05, therefore reject H0

H0: CDF of Total Cases in Oct 2020 - Nov 2020 = CDF of Post Covid Passenger Count Data(Oct 2020 - Nov 2020)
H0: CDF of Total Cases in Oct 2020 - Nov 2020 != CDF of Post Covid Passenger Count Data(Oct 2020 - Nov 2020)

KS Statistic : 1.0000000000000007
1.0000000000000007 > 0.05, therefore reject H0

H0: CDF of Total Cases in Dec 2020 - Jan 2021 = CDF of Post Covid Passenger Count Data(Dec 2020 - Jan 2021)
H0: CDF of Total Cases in Dec 2020 - Jan 2021 != CDF of Post Covid Passenger Count Data(Dec 2020 - Jan 2021)

KS Statistic : 0.9999999999999992
0.9999999999999992 > 0.05, therefore reject H0

H0: CDF of Total Cases in Feb 2021 - Mar 2021 = CDF of Post Covid Passenger Count Data(Feb 2021 - Mar 2021)
H0: CDF of Total Cases in Feb 2021 - Mar 2021 != CDF of Post Covid Passenger Count Data(Feb 2021 - Mar 2021)

KS Statistic : 0.9999999999999989
0.9999999999999989 > 0.05, therefore reject H0

H0: CDF of Total Cases in Apr 2020 - May 2021 = CDF of Post Covid Passenger Count Data(Apr 2020 - May 2021)
H0: CDF of Total Cases in Apr 2020 - May 2021 != CDF of Post Covid Passenger Count Data(Apr 2020 - May 2021)

KS Statistic : 1.0000000000000007
1.0000000000000007 > 0.05, therefore reject H0

H0: CDF of Total Cases in Jun 2021 - Jul 2021 = CDF of Post Covid Passenger Count Data(Jun 2021 - Jul 2021)
H0: CDF of Total Cases in Jun 2021 - Jul 2021 != CDF of Post Covid Passenger Count Data(Jun 2021 - Jul 2021)

KS Statistic : 1.0000000000000007
1.0000000000000007 > 0.05, therefore reject H0

H0: CDF of Total Cases in Aug 2021 - Sep 2021 = CDF of Post Covid Passenger Count Data(Aug 2021 - Sep 2021)
H0: CDF of Total Cases in Aug 2021 - Sep 2021 != CDF of Post Covid Passenger Count Data(Aug 2021 - Sep 2021)

KS Statistic : 1.0000000000000007
1.0000000000000007 > 0.05, therefore reject H0

H0: CDF of Total Cases in Oct 2021 - Nov 2021 = CDF of Post Covid Passenger Count Data(Oct 2021 - Nov 2021)
H0: CDF of Total Cases in Oct 2021 - Nov 2021 != CDF of Post Covid Passenger Count Data(Oct 2021 - Nov 2021)

KS Statistic : 1.0000000000000007

1.0000000000000007 > 0.05, therefore reject H0

H0: CDF of Total Cases in Dec 2021 - Jan 2022 = CDF of Post Covid Passenger Count Data(Dec 2021 - Jan 2022)
H0: CDF of Total Cases in Dec 2021 - Jan 2022 != CDF of Post Covid Passenger Count Data(Dec 2021 - Jan 2022)

KS Statistic : 0.9999999999999993
0.9999999999999993 > 0.05, therefore reject H0

H0: CDF of Total Cases in Feb 2022 - Mar 2022 = CDF of Post Covid Passenger Count Data(Feb 2022 - Mar 2022)
H0: CDF of Total Cases in Feb 2022 - Mar 2022 != CDF of Post Covid Passenger Count Data(Feb 2022 - Mar 2022)

KS Statistic : 0.9999999999999989
0.9999999999999989 > 0.05, therefore reject H0

Second Inference

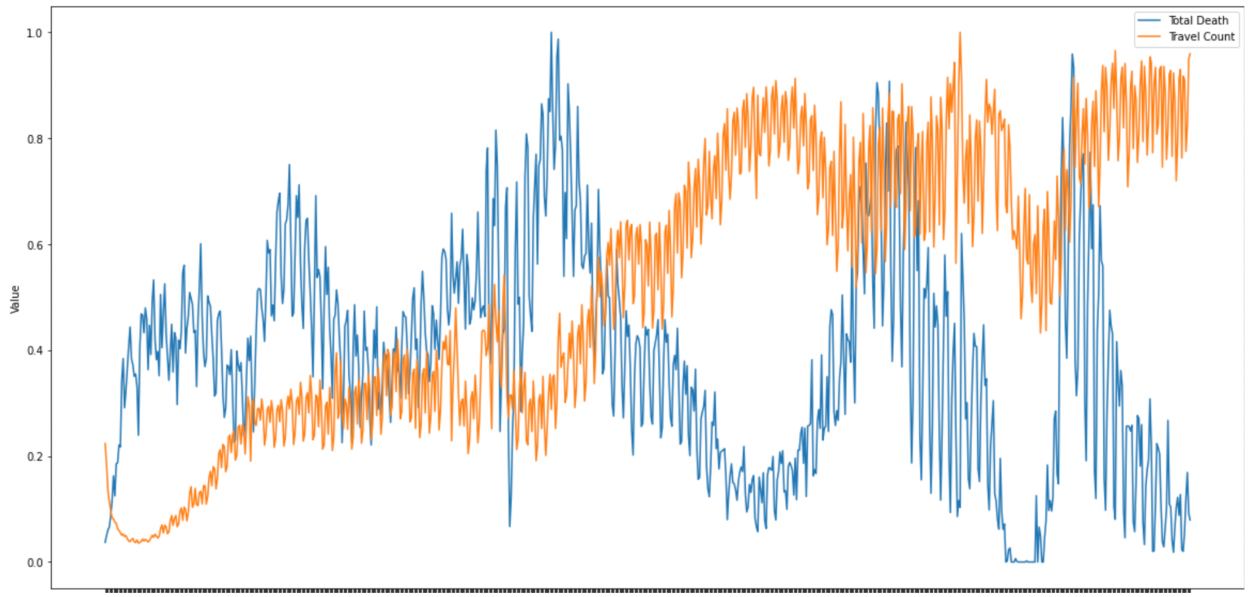
Using Pearson Correlation to find whether X&Y are linearly dependent or not.

H0: Total no. of passengers traveling via air is linearly dependent on no. of deaths dues to COVID.
H1: Total no. of passengers traveling via air is not linearly dependent on no. of deaths dues to COVID.

Steps Followed:
1. Selected Data from the start of the first lockdown imposed in the US (i.e. March 19, 2020).
2. Found Pearson Correlation between total deaths and travel count.
Pearson Correlation = -0.3357556690738906

This tells us that as the no. of deaths increases travel count decreases which totally makes sense.

As the graph suggests that some portion of the data is highly linearly correlated and some portion is not, maybe due to introduction of vaccines or some other factor. Hence, I choose to take arbitrary value to 0.30 as the threshold hence we can say that there is some linear relationship between Total no. of passengers and no. of deaths.



Third Inference

Wald's 2-Sample Test:

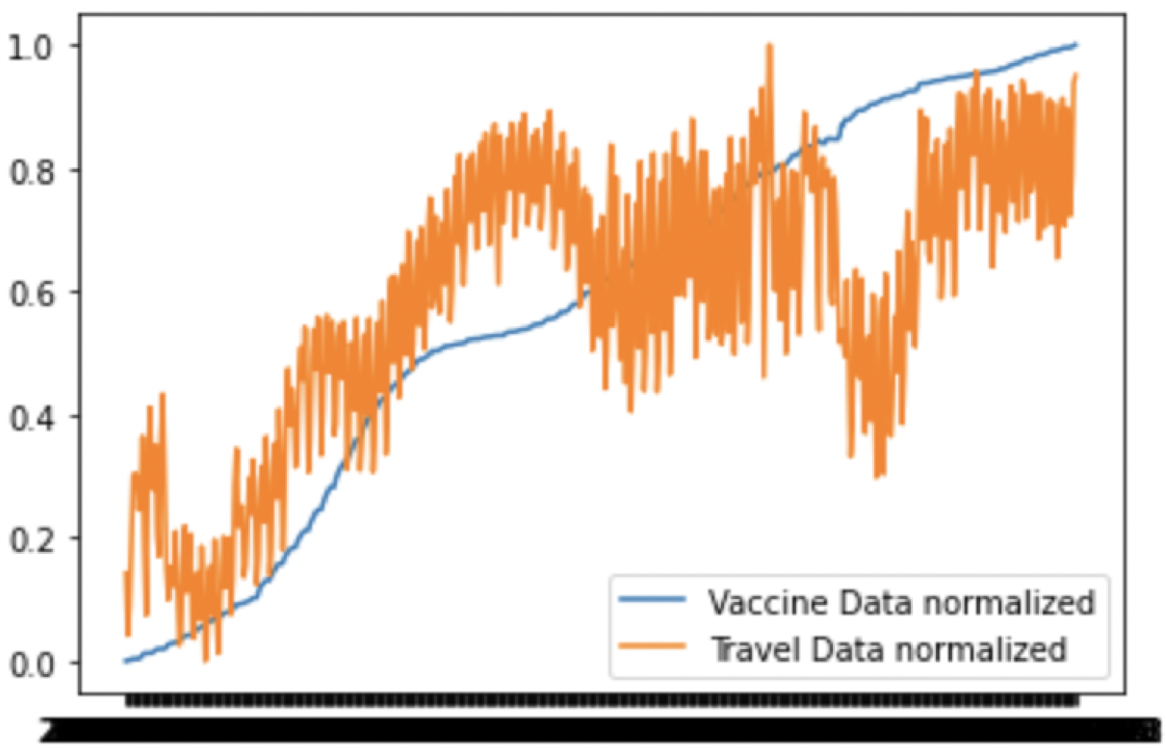
Wald's test is performed on the normalized data of number of passengers as well as number of vaccinations

H0: delta=0 (Traveling of passengers through flight is proportional to the number of vaccination)
H1: delta!=0 (Traveling of passengers through flight is not proportional to the number of vaccination)

Result:

Wald's statistic = 0.025732354918676974 which is less than 1.96 (taking alpha=0.05), hence we accept the null hypothesis.

Thus, the number of passengers traveling through flight is proportional to the number of vaccinations.


1.0
0.8
0.6
0.4
0.2
0.0
Vaccine Data normalized
Travel Data normalized