

part\_a

May 16, 2022

## 1 Part a Cases: Wald's test, Z-test, and t-test

Wald's test, Z-test, and t-test (assume all are applicable) to check whether the mean of COVID19 cases are different for Feb'21 and March'21 in the two states.

```
[32]: %cd /content/drive/Shareddrives/CSE544_Project/covid_dataset
      !ls
```

```
/content/drive/Shareddrives/CSE544_Project/covid_dataset
backup
COVID-19_Vaccinations_in_the_United_States_Jurisdiction.csv
covid_la_cleaned.csv
covid_la_cleaned_removed_outliers.csv
covid_md_cleaned.csv
covid_md_cleaned_removed_outliers.csv
United_States_COVID-19_Cases_and_Deaths_by_State_over_Time.csv
vacc_la_clean.csv
vacc_la_clean_removed_outliers.csv
vacc_md_clean.csv
vacc_md_clean_removed_outliers.csv
```

```
[33]: #importing all libraries
import pandas as pd
import numpy as np
from collections import Counter
import math
import csv

#global variable
Z_value = 1.96 # alpha = 0.05
t_value = 2.042 # alpha = 0.05 n =30
# Helper functions

# Function to calculate variance
def cal_variance(list):
    mean=sum(list)/len(list)
    sq_diff= [(x - mean) ** 2 for x in list]
    result= sum(sq_diff)/len(list)
```

```

    return result

# Function to calculate variance corrected
def cal_variance_corrected(list):
    mean=sum(list)/len(list)
    sq_diff= [(x - mean) ** 2 for x in list]
    result= sum(sq_diff)/(len(list)-1)
    return result

```

## 1.1 Dataset

```

[34]: # Taking LA and MD covid dataframes with only cases per day
df_covid_la = pd.read_csv('covid_la_cleaned_removed_outliers.csv', usecols =_
    ↳ ['submission_date', 'new_case'])
df_covid_md = pd.read_csv('covid_md_cleaned_removed_outliers.csv', usecols =_
    ↳ ['submission_date', 'new_case'])

[35]: # Taking only Feb21 and Mar21 Data for each state dataframe
df_covid_md["submission_date"] = pd.to_datetime(df_covid_md["submission_date"],_
    ↳ format="%Y-%m-%d")
df_covid_md_feb = df_covid_md[df_covid_md["submission_date"].dt.month.isin([2])_
    ↳ & df_covid_md["submission_date"].dt.year.isin([2021])]
df_covid_md_march = df_covid_md[df_covid_md["submission_date"].dt.month._
    ↳ isin([3]) & df_covid_md["submission_date"].dt.year.isin([2021])]
df_covid_la["submission_date"] = pd.to_datetime(df_covid_la["submission_date"],_
    ↳ format="%Y-%m-%d")
df_covid_la_feb = df_covid_la[df_covid_la["submission_date"].dt.month.isin([2])_
    ↳ & df_covid_la["submission_date"].dt.year.isin([2021])]
df_covid_la_march = df_covid_la[df_covid_la["submission_date"].dt.month._
    ↳ isin([3]) & df_covid_la["submission_date"].dt.year.isin([2021])]

```

## 1.2 One - Sample Hypothesis Testing

Test: \* Let mean of cases in Feb 2021 be used as true mean =  $\mu_0$ . \* Let mean of cases in March 2021 =  $\mu_1$ .

### 1.2.1 Wald's test for Maryland:

$H_0: \mu_1 = \mu_0$  vs  $H_1: \mu_1 \neq \mu_0$

if  $|W| > Z$ , we REJECT  $H_0$ .

where  $Z(\alpha=0.05/2) = 1.96$

```
[36]: # sample mean
sample_mean = df_covid_md_feb['new_case'].mean()

# mle estimator for a poisson distribution is the sample mean of the
↳ distribution itself
mle_estimator = df_covid_md_march['new_case'].mean()

# standard variation
std = np.sqrt(cal_variance(df_covid_md_march['new_case'].to_list()))

# Walds Test
W_md = np.abs( (mle_estimator - sample_mean) / std)

print("Wald's Test value for Maryland:", "{:.2f}".format(W_md))
if(W_md < Z_value):
    print("We do not reject the Hypothesis as |W| < Z")
else:
    print("We do reject the Hypothesis as |W| > Z")
```

Wald's Test value for Maryland: 0.09

We do not reject the Hypothesis as  $|W| < Z$

### 1.2.2 Wald's test for Louisiana:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|W| > Z$  , we REJECT  $H_0$ .

where  $Z(\alpha=0.05/2) = 1.96$

```
[37]: # sample mean
sample_mean = df_covid_la_feb['new_case'].mean()

# mle estimator for a poisson distribution is the sample mean of the
↳ distribution itself
mle_estimator = df_covid_la_march['new_case'].mean()

# standard variation
std = np.sqrt(cal_variance(df_covid_la_march['new_case'].to_list()))

# Walds Test
W_la = np.abs( (mle_estimator - sample_mean) / std)

print("Wald's Test value for Louisiana:", "{:.2f}".format(W_la))
if(W_la < Z_value):
    print("We do not reject the Hypothesis as |W| < Z")
else:
```

```
print("We do reject the Hypothesis as |W| > Z")
```

Wald's Test value for Louisiana: 1.84

We do not reject the Hypothesis as  $|W| < Z$

### 1.2.3 Z-test for Maryland:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|Z| > z\text{-val}$  , we REJECT  $H_0$ .

where  $z\text{-val}$  ( $\alpha=0.05/2$ ) = 1.96

```
[38]: # sample mean
true_mean = df_covid_md_feb['new_case'].mean()
sample_mean = df_covid_md_march['new_case'].mean()

# number of samples
n = len(df_covid_md_march['new_case'])

# standard variation
std = np.sqrt(cal_variance(df_covid_md_march['new_case'].to_list()))

# Z Test
Z_md = np.abs( (sample_mean - true_mean) / std / np.sqrt(n) )

print("Z Test value for Maryland:", "{:.2f}".format(Z_md))
if(Z_md < Z_value):
    print("We do not reject the Hypothesis as |Z'| < Z")
else:
    print("We do reject the Hypothesis as |Z'| > Z")
```

Z Test value for Maryland: 0.02

We do not reject the Hypothesis as  $|Z'| < Z$

### 1.2.4 Z-test for Louisiana:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|Z| > z\text{-val}$  , we REJECT  $H_0$ .

where  $z\text{-val}$  ( $\alpha=0.05/2$ ) = 1.96

```
[39]: # sample mean
true_mean = df_covid_la_feb['new_case'].mean()
sample_mean = df_covid_la_march['new_case'].mean()
```

```

# number of samples
n = len(df_covid_la_march['new_case'])

# standard variation
std = np.sqrt(cal_variance(df_covid_la_march['new_case'].to_list()))

# Z Test
Z_la = np.abs( (sample_mean - true_mean) / std / np.sqrt(n) )

print("Z Test value for Louisiana:", "{:.2f}".format(Z_la))
if(Z_la < Z_value):
    print("We do not reject the Hypothesis as |Z'| < Z")
else:
    print("We do reject the Hypothesis as |Z'| > Z")

```

Z Test value for Louisiana: 0.33  
We do not reject the Hypothesis as  $|Z'| < Z$

### 1.2.5 t-test for Maryland:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|T| > t\text{-val}$  , we REJECT  $H_0$ .

where  $t\text{-val}(\alpha=0.05/2, n = 30) = 2.042$

```

[40]: # sample mean
true_mean = df_covid_md_feb['new_case'].mean()
sample_mean = df_covid_md_march['new_case'].mean()

# number of samples
n = len(df_covid_md_march['new_case'])

# standard variation with corrected
std = np.sqrt(cal_variance_corrected(df_covid_md_march['new_case'].to_list()))

# t Test
t_md = np.abs( (sample_mean - true_mean) / std / np.sqrt(n) )

print("t Test value for Maryland:", "{:.2f}".format(t_md))
if(t_md < t_value):
    print("We do not reject the Hypothesis as |T'| < t-value")
else:
    print("We do reject the Hypothesis as |T'| > t-value")

```

t Test value for Maryland: 0.02

We do not reject the Hypothesis as  $|T'| < t\text{-value}$

### 1.2.6 t-test for Louisiana:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|T| > t\text{-val}$  , we REJECT  $H_0$ .

where  $t\text{-val}(\alpha=0.05/2, n = 30) = 2.042$

```
[41]: # sample mean
true_mean = df_covid_la_feb['new_case'].mean()
sample_mean = df_covid_la_march['new_case'].mean()

# number of samples
n = len(df_covid_la_march['new_case'])

# standard variation with corrected variance
std = np.sqrt(cal_variance_corrected(df_covid_la_march['new_case'].to_list()))

# t Test
t_la = np.abs( (sample_mean - true_mean) / std / np.sqrt(n) )

print("t Test value for Louisiana:", "{:.2f}".format(t_la))
if(t_la < t_value):
    print("We do not reject the Hypothesis as  $|T'| < t\text{-value}$ ")
else:
    print("We do reject the Hypothesis as  $|T'| > t\text{-value}$ ")
```

t Test value for Louisiana: 0.32

We do not reject the Hypothesis as  $|T'| < t\text{-value}$

## 1.3 Two - Sample Hypothesis Testing

Test: \* Let mean of cases in Feb 2021 =  $p_0$ . \* Let mean of cases in March 2021 =  $p_1$ .

### 1.3.1 2-Population Wald's test for Maryland:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|W| > Z$  , we REJECT  $H_0$ .

where  $Z(\alpha=0.05/2) = 1.96$

```
[42]: # sample mean
sample_mean_feb = df_covid_md_feb['new_case'].mean()
```

```

n = len(df_covid_md_feb['new_case'])

# mle estimator for a poisson distribution is the sample mean of the
↳distribution itself
sample_mean_mar = df_covid_md_march['new_case'].mean()
m = len(df_covid_md_march['new_case'])

# std denominator
variance_feb = cal_variance(df_covid_md_feb['new_case'].to_list())
variance_mar = cal_variance(df_covid_md_march['new_case'].to_list())
sqr_variance_feb = (variance_feb*variance_feb)/n
sqr_variance_mar = (variance_mar*variance_mar)/m
std = np.sqrt(sqr_variance_feb+sqr_variance_mar)

# Walds Test
W_md_two = np.abs( (sample_mean_mar - sample_mean_feb) / std)

print("2-Pop Wald's Test value for Maryland:", "{:.4f}".format(W_md_two))
if(W_md_two < Z_value):
    print("We do not reject the Hypothesis as |W| < Z")
else:
    print("We do reject the Hypothesis as |W| > Z")

```

2-Pop Wald's Test value for Maryland: 0.0012

We do not reject the Hypothesis as  $|W| < Z$

### 1.3.2 2-Population Wald's test for Louisiana:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|W| > Z$  , we REJECT  $H_0$ .

where  $Z(\alpha=0.05/2) = 1.96$

```

[43]: # sample mean
sample_mean_feb = df_covid_la_feb['new_case'].mean()
n = len(df_covid_la_feb['new_case'])

# mle estimator for a poisson distribution is the sample mean of the
↳distribution itself
sample_mean_mar = df_covid_la_march['new_case'].mean()
m = len(df_covid_la_march['new_case'])

# std denominator
variance_feb = cal_variance(df_covid_la_feb['new_case'].to_list())
variance_mar = cal_variance(df_covid_la_march['new_case'].to_list())

```

```

sqr_variance_feb = (variance_feb*variance_feb)/n
sqr_variance_mar = (variance_mar*variance_mar)/m
std = np.sqrt(sqr_variance_feb+sqr_variance_mar)

# Walds Test
W_la_two = np.abs( (sample_mean_mar - sample_mean_feb) / std)

print("2-Pop Wald's Test value for Louisiana:", "{:.4f}".format(W_la_two))
if(W_la_two < Z_value):
    print("We do not reject the Hypothesis as |W| < Z")
else:
    print("We do reject the Hypothesis as |W| > Z")

```

2-Pop Wald's Test value for Louisiana: 0.0044  
We do not reject the Hypothesis as |W| < Z

### 1.3.3 2-Population t-test for Maryland:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|T| > t\text{-val}$  , we REJECT  $H_0$ .

where  $t\text{-val}(\alpha=0.05/2, n = 30) = 2.042$

```

[44]: # sample mean
sample_mean_feb = df_covid_md_feb['new_case'].mean()
n = len(df_covid_md_feb['new_case'])

sample_mean_mar = df_covid_md_march['new_case'].mean()
m = len(df_covid_md_march['new_case'])

# std denominator
variance_feb = cal_variance_corrected(df_covid_md_feb['new_case'].to_list())
variance_mar = cal_variance_corrected(df_covid_md_march['new_case'].to_list())
sqr_variance_feb = (variance_feb*variance_feb)/n
sqr_variance_mar = (variance_mar*variance_mar)/m
std = np.sqrt(sqr_variance_feb+sqr_variance_mar)

t_md_two = np.abs( (sample_mean_mar - sample_mean_feb) / std)

print("2-Pop Wald's Test value for Maryland:", "{:.4f}".format(t_md_two))
if(t_md_two < t_value):
    print("We do not reject the Hypothesis as |T| < t-val")
else:
    print("We do reject the Hypothesis as |T| > t-val")

```

2-Pop Wald's Test value for Maryland: 0.0011



We do not reject the Hypothesis as  $|T| < t\text{-val}$

### 1.3.4 2-Population t-test for Louisiana:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|T| > t\text{-val}$  , we REJECT  $H_0$ .

where  $t\text{-val}(\alpha=0.05/2, n=30) = 2.042$

```
[45]: # sample mean
sample_mean_feb = df_covid_la_feb['new_case'].mean()
n = len(df_covid_la_feb['new_case'])

sample_mean_mar = df_covid_la_march['new_case'].mean()
m = len(df_covid_la_march['new_case'])

# std denominator
variance_feb = cal_variance_corrected(df_covid_la_feb['new_case'].to_list())
variance_mar = cal_variance_corrected(df_covid_la_march['new_case'].to_list())
sqr_variance_feb = (variance_feb*variance_feb)/n
sqr_variance_mar = (variance_mar*variance_mar)/m
std = np.sqrt(sqr_variance_feb+sqr_variance_mar)

t_la_two = np.abs( (sample_mean_mar - sample_mean_feb) / std)

print("2-Pop Wald's Test value for Louisiana:", "{:.4f}".format(t_la_two))
if(t_la_two < t_value):
    print("We do not reject the Hypothesis as  $|T| < t\text{-val}$ ")
else:
    print("We do reject the Hypothesis as  $|T| > t\text{-val}$ ")
```

2-Pop Wald's Test value for Louisiana: 0.0043

We do not reject the Hypothesis as  $|T| < t\text{-val}$

## 1.4 Conclusion:

We perform hypothesis testing on number of **CASES** per day for the Months of **Feb 2021** and **March 2021** for the states **MARYLAND** and **LOUISIANA**.

1. First we do a one-sample Hypothesis testing taking the true mean as sample mean of covid cases per day for the month of Feb21 and perform tests on March21 dataset. The following are the conclusions
  - Wald's Test:  
Maryland:  $W = 0.086 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Maryland  
Louisiana:  $W = 1.83 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Louisiana

- Z-Test:

Maryland:  $Z = 0.0155 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Maryland

Louisiana:  $Z = 0.329 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Louisiana

- t-Test:

Maryland:  $t = 0.0152 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Maryland

Louisiana:  $t = 0.324 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Louisiana

2. Now, we do a two-sample Hypothesis testing between the sample population of Feb21 and March21 dataset for cases per day.

- Wald's Test:

Maryland:  $W = 0.0011 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Maryland

Louisiana:  $W = 0.0044 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Louisiana

- t-Test:

Maryland:  $t = 0.0011 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Maryland

Louisiana:  $t = 0.0042 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Louisiana

All the tests are applicable as the number of samples is more than 30 for one-population tests. According to t-Test, if  $n > 30$  then it tends to follow a Normal Distribution under CLT rules.

We finally conclude by performing all the above tests that our hypothesis that the mean of covid cases per day in the month of Feb2021 is **NOT** different than the number of cases per day in the month of March2021. Hence both the parameters of the distributions are similar.

[46]: `W_md,W_la,Z_md,Z_la,t_md,t_la`

[46]: (0.08646682688793442,  
1.8366374272181543,  
0.015529900558504324,  
0.32986981982919544,  
0.015277365020134161,  
0.32450572543407613)

[47]: `W_md_two,W_la_two,t_md_two,t_la_two`

[47]: (0.001182310673129563,  
0.004447599844196726,  
0.001139643523995524,  
0.004283285879525227)

---



---

## 2 Part a DEATHS: Wald's test, Z-test, and t-test

Wald's test, Z-test, and t-test (assume all are applicable) to check whether the mean of death19 deaths are different for Feb'21 and March'21 in the two states.

```
[48]: %cd /content/drive/Shareddrives/CSE544_Project/covid_dataset
```

```
/content/drive/Shareddrives/CSE544_Project/covid_dataset
```

```
[49]: #importing all libraries
import pandas as pd
import numpy as np
from collections import Counter
import math
import csv

#global variable
Z_value = 1.96 # alpha = 0.05
t_value = 2.042 # alpha = 0.05 n =30

# Helper functions

# Function to calculate variance
def cal_variance(list):
    mean=sum(list)/len(list)
    sq_diff= [(x - mean) ** 2 for x in list]
    result= sum(sq_diff)/len(list)
    return result

# Function to calculate variance corrected
def cal_variance_corrected(list):
    mean=sum(list)/len(list)
    sq_diff= [(x - mean) ** 2 for x in list]
    result= sum(sq_diff)/(len(list)-1)
    return result
```

### 2.1 Dataset - Only Deaths

```
[50]: # Taking LA and MD death dataframes with only deaths per day columns
df_death_la = pd.read_csv('covid_la_cleaned_removed_outliers.csv', usecols =_
    ↳ ['submission_date', 'new_death'])
df_death_md = pd.read_csv('covid_md_cleaned_removed_outliers.csv', usecols =_
    ↳ ['submission_date', 'new_death'])
# Taking only Feb21 and Mar21 Data for each state dataframe
```

```

df_death_md["submission_date"] = pd.to_datetime(df_death_md["submission_date"],
↪format="%Y-%m-%d")
df_death_md_feb = df_death_md[df_death_md["submission_date"].dt.month.isin([2])
↪& df_death_md["submission_date"].dt.year.isin([2021])]
df_death_md_march = df_death_md[df_death_md["submission_date"].dt.month.
↪isin([3]) & df_death_md["submission_date"].dt.year.isin([2021])]
df_death_la["submission_date"] = pd.to_datetime(df_death_la["submission_date"],
↪format="%Y-%m-%d")
df_death_la_feb = df_death_la[df_death_la["submission_date"].dt.month.isin([2])
↪& df_death_la["submission_date"].dt.year.isin([2021])]
df_death_la_march = df_death_la[df_death_la["submission_date"].dt.month.
↪isin([3]) & df_death_la["submission_date"].dt.year.isin([2021])]

```

## 2.2 One - Sample Hypothesis Testing

Test: \* Let mean of Covid Deaths in Feb 2021 be used as true mean =  $\mu_0$ . \* Let mean of Covid Deaths in March 2021 =  $\mu_1$ .

### 2.2.1 Wald's test for Maryland Covid Deaths:

$H_0: \mu_1 = \mu_0$  vs  $H_1: \mu_1 \neq \mu_0$

if  $|W| > Z$ , we REJECT  $H_0$ .

where  $Z(\alpha=0.05/2) = 1.96$

```

[51]: # sample mean
sample_mean = df_death_md_feb['new_death'].mean()

# mle estimator for a poisson distribution is the sample mean of the
↪distribution itself
mle_estimator = df_death_md_march['new_death'].mean()

# standard variation
std = np.sqrt(cal_variance(df_death_md_march['new_death'].to_list()))

# Walds Test
W_md = np.abs( (mle_estimator - sample_mean) / std)

print("Wald's Test value for Maryland Covid Deaths:", "{:.2f}".format(W_md))
if(W_md < Z_value):
    print("We do not reject the Hypothesis as  $|W| < Z$ ")
else:
    print("We do reject the Hypothesis as  $|W| > Z$ ")

```

Wald's Test value for Maryland Covid Deaths: 3.21  
We do reject the Hypothesis as  $|W| > Z$

### 2.2.2 Wald's test for Louisiana: Covid Deaths

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|W| > Z$  , we REJECT  $H_0$ .

where  $Z(\alpha=0.05/2) = 1.96$

```
[52]: # sample mean
sample_mean = df_death_la_feb['new_death'].mean()

# mle estimator for a poisson distribution is the sample mean of the
# distribution itself
mle_estimator = df_death_la_march['new_death'].mean()

# standard variation
std = np.sqrt(cal_variance(df_death_la_march['new_death'].to_list()))

# Walds Test
W_la = np.abs( (mle_estimator - sample_mean) / std)

print("Wald's Test value for Louisiana:", "{:.2f}".format(W_la))
if(W_la < Z_value):
    print("We do not reject the Hypothesis as  $|W| < Z$ ")
else:
    print("We do reject the Hypothesis as  $|W| > Z$ ")
```

Wald's Test value for Louisiana: 0.67  
We do not reject the Hypothesis as  $|W| < Z$

### 2.2.3 Z-test for Maryland Covid Deaths:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|Z| > z\text{-val}$  , we REJECT  $H_0$ .

where  $z\text{-val} (\alpha=0.05/2) = 1.96$

```
[53]: # sample mean
true_mean = df_death_md_feb['new_death'].mean()
sample_mean = df_death_md_march['new_death'].mean()

# number of samples
```

```

n = len(df_death_md_march['new_death'])

# standard variation
std = np.sqrt(cal_variance(df_death_md_march['new_death'].to_list()))

# Z Test
Z_md = np.abs( (sample_mean - true_mean) / std / np.sqrt(n) )

print("Z Test value for Maryland Covid Deaths:", "{:.2f}".format(Z_md))
if(Z_md < Z_value):
    print("We do not reject the Hypothesis as |Z'| < Z")
else:
    print("We do reject the Hypothesis as |Z'| > Z")

```

Z Test value for Maryland Covid Deaths: 0.58  
We do not reject the Hypothesis as |Z'| < Z

## 2.2.4 Z-test for Louisiana Covid Deaths:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|Z| > z\text{-val}$  , we REJECT  $H_0$ .

where  $z\text{-val}$  ( $\alpha=0.05/2$ ) = 1.96

```

[54]: # sample mean
true_mean = df_death_la_feb['new_death'].mean()
sample_mean = df_death_la_march['new_death'].mean()

# number of samples
n = len(df_death_la_march['new_death'])

# standard variation
std = np.sqrt(cal_variance(df_death_la_march['new_death'].to_list()))

# Z Test
Z_la = np.abs( (sample_mean - true_mean) / std / np.sqrt(n) )

print("Z Test value for Louisiana Covid Deaths:", "{:.2f}".format(Z_la))
if(Z_la < Z_value):
    print("We do not reject the Hypothesis as |Z'| < Z")
else:
    print("We do reject the Hypothesis as |Z'| > Z")

```

Z Test value for Louisiana Covid Deaths: 0.12  
We do not reject the Hypothesis as |Z'| < Z

### 2.2.5 t-test for Maryland Covid Deaths:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|T| > t\text{-val}$  , we REJECT  $H_0$ .

where  $t\text{-val}(\alpha=0.05/2, n = 30) = 2.042$

```
[55]: # sample mean
true_mean = df_death_md_feb['new_death'].mean()
sample_mean = df_death_md_march['new_death'].mean()

# number of samples
n = len(df_death_md_march['new_death'])

# standard variation with corrected
std = np.sqrt(cal_variance_corrected(df_death_md_march['new_death'].to_list()))

# t Test
t_md = np.abs( (sample_mean - true_mean) / std / np.sqrt(n) )

print("t Test value for Maryland Covid Deaths:", "{:.2f}".format(t_md))
if(t_md < t_value):
    print("We do not reject the Hypothesis as |T'| < t-value")
else:
    print("We do reject the Hypothesis as |T'| > t-value")
```

t Test value for Maryland Covid Deaths: 0.57

We do not reject the Hypothesis as  $|T'| < t\text{-value}$

### 2.2.6 t-test for Louisiana Covid Deaths:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|T| > t\text{-val}$  , we REJECT  $H_0$ .

where  $t\text{-val}(\alpha=0.05/2, n = 30) = 2.042$

```
[56]: # sample mean
true_mean = df_death_la_feb['new_death'].mean()
sample_mean = df_death_la_march['new_death'].mean()

# number of samples
n = len(df_death_la_march['new_death'])

# standard variation with corrected variance
std = np.sqrt(cal_variance_corrected(df_death_la_march['new_death'].to_list()))
```

```

# t Test
t_la = np.abs( (sample_mean - true_mean) / std / np.sqrt(n) )

print("t Test value for Louisiana Covid Deaths:", "{:.2f}".format(t_la))
if(t_la < t_value):
    print("We do not reject the Hypothesis as |T'| < t-value")
else:
    print("We do reject the Hypothesis as |T'| > t-value")

```

t Test value for Louisiana Covid Deaths: 0.12  
 We do not reject the Hypothesis as |T'| < t-value

## 2.3 Two - Sample Hypothesis Testing

Test: \* Let mean of Covid Deaths in Feb 2021 =  $\mu_0$ . \* Let mean of Covid Deaths in March 2021 =  $\mu_1$ .

### 2.3.1 2-Population Wald's test for Maryland Covid Deaths:

$H_0: \mu_1 = \mu_0$  vs  $H_1: \mu_1 \neq \mu_0$

if  $|W| > Z$ , we REJECT  $H_0$ .

where  $Z(\alpha=0.05/2) = 1.96$

```

[57]: # sample mean
sample_mean_feb = df_death_md_feb['new_death'].mean()
n = len(df_death_md_feb['new_death'])

# mle estimator for a poisson distribution is the sample mean of the
  ↳ distribution itself
sample_mean_mar = df_death_md_march['new_death'].mean()
m = len(df_death_md_march['new_death'])

# std denominator
variance_feb = cal_variance(df_death_md_feb['new_death'].to_list())
variance_mar = cal_variance(df_death_md_march['new_death'].to_list())
sqr_variance_feb = (variance_feb*variance_feb)/n
sqr_variance_mar = (variance_mar*variance_mar)/m
std = np.sqrt(sqr_variance_feb+sqr_variance_mar)

# Walds Test
W_md_two = np.abs( (sample_mean_mar - sample_mean_feb) / std)

```



```

print("2-Pop Wald's Test value for Maryland Covid Deaths:", "{:.4f}".
    ↪format(W_md_two))
if(W_md_two < Z_value):
    print("We do not reject the Hypothesis as |W| < Z")
else:
    print("We do reject the Hypothesis as |W| > Z")

```

2-Pop Wald's Test value for Maryland Covid Deaths: 1.4701  
 We do not reject the Hypothesis as  $|W| < Z$

### 2.3.2 2-Population Wald's test for Louisiana Covid Deaths:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|W| > Z$ , we REJECT  $H_0$ .

where  $Z(\alpha=0.05/2) = 1.96$

```

[58]: # sample mean
sample_mean_feb = df_death_la_feb['new_death'].mean()
n = len(df_death_la_feb['new_death'])

# mle estimator for a poisson distribution is the sample mean of the
    ↪distribution itself
sample_mean_mar = df_death_la_march['new_death'].mean()
m = len(df_death_la_march['new_death'])

# std denominator
variance_feb = cal_variance(df_death_la_feb['new_death'].to_list())
variance_mar = cal_variance(df_death_la_march['new_death'].to_list())
sqr_variance_feb = (variance_feb*variance_feb)/n
sqr_variance_mar = (variance_mar*variance_mar)/m
std = np.sqrt(sqr_variance_feb+sqr_variance_mar)

# Walds Test
W_la_two = np.abs( (sample_mean_mar - sample_mean_feb) / std)

print("2-Pop Wald's Test value for Louisiana Covid Deaths:", "{:.4f}".
    ↪format(W_la_two))
if(W_la_two < Z_value):
    print("We do not reject the Hypothesis as |W| < Z")
else:
    print("We do reject the Hypothesis as |W| > Z")

```

2-Pop Wald's Test value for Louisiana Covid Deaths: 0.1449  
 We do not reject the Hypothesis as  $|W| < Z$

### 2.3.3 2-Population t-test for Maryland Covid Deaths:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|T| > t\text{-val}$  , we REJECT  $H_0$ .

where  $t\text{-val}(\alpha=0.05/2, n = 30) = 2.042$

```
[59]: # sample mean
sample_mean_feb = df_death_md_feb['new_death'].mean()
n = len(df_death_md_feb['new_death'])

sample_mean_mar = df_death_md_march['new_death'].mean()
m = len(df_death_md_march['new_death'])

# std denominator
variance_feb = cal_variance_corrected(df_death_md_feb['new_death'].to_list())
variance_mar = cal_variance_corrected(df_death_md_march['new_death'].to_list())
sqr_variance_feb = (variance_feb*variance_feb)/n
sqr_variance_mar = (variance_mar*variance_mar)/m
std = np.sqrt(sqr_variance_feb+sqr_variance_mar)

t_md_two = np.abs( (sample_mean_mar - sample_mean_feb) / std)

print("2-Pop Wald's Test value for Maryland Covid Deaths:", "{:.4f}".
      ↪format(t_md_two))
if(t_md_two < t_value):
    print("We do not reject the Hypothesis as  $|T| < t\text{-val}$ ")
else:
    print("We do reject the Hypothesis as  $|T| > t\text{-val}$ ")
```

2-Pop Wald's Test value for Maryland Covid Deaths: 1.4144

We do not reject the Hypothesis as  $|T| < t\text{-val}$

### 2.3.4 2-Population t-test for Louisiana Covid Deaths:

$H_0: P_1 = P_0$  vs  $H_1: P_1 \neq P_0$

if  $|T| > t\text{-val}$  , we REJECT  $H_0$ .

where  $t\text{-val}(\alpha=0.05/2, n =30) = 2.042$

```
[60]: # sample mean
sample_mean_feb = df_death_la_feb['new_death'].mean()
n = len(df_death_la_feb['new_death'])

sample_mean_mar = df_death_la_march['new_death'].mean()
```

```

m = len(df_death_la_march['new_death'])

# std denominator
variance_feb = cal_variance_corrected(df_death_la_feb['new_death'].to_list())
variance_mar = cal_variance_corrected(df_death_la_march['new_death'].to_list())
sqr_variance_feb = (variance_feb*variance_feb)/n
sqr_variance_mar = (variance_mar*variance_mar)/m
std = np.sqrt(sqr_variance_feb+sqr_variance_mar)

t_la_two = np.abs( (sample_mean_mar - sample_mean_feb) / std)

print("2-Pop Wald's Test value for Louisiana Covid Deaths:", "{:.4f}".
      ↪format(t_la_two))
if(t_la_two < t_value):
    print("We do not reject the Hypothesis as |T| < t-val")
else:
    print("We do reject the Hypothesis as |T| > t-val")

```

2-Pop Wald's Test value for Louisiana Covid Deaths: 0.1397  
 We do not reject the Hypothesis as |T| < t-val

## 2.4 Conclusion:

We perform hypothesis testing on number of **DEATHS** per day for the Months of **Feb 2021** and **March 2021** for the states **MARYLAND** and **LOUISIANA**.

1. First we do a one-sample Hypothesis testing taking the true mean as sample mean of death deaths per day for the month of Feb21 and perform tests on March21 dataset. The following are the conclusions
  - Wald's Test:
 

Maryland:  $W = 3.21 > 1.96$ , Therefore we **Reject** the hypothesis for Maryland

Louisiana:  $W = 0.66 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Louisiana
  - Z-Test:
 

Maryland:  $Z = 0.577 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Maryland

Louisiana:  $Z = 0.119 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Louisiana
  - t-Test:
 

Maryland:  $t = 0.567 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Maryland

Louisiana:  $t = 0.117 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Louisiana
2. Now, we do a two-sample Hypothesis testing between the sample population of Feb21 and March21 dataset for deaths per day.
  - Wald's Test:

Maryland:  $W = 1.47 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Maryland

Louisiana:  $W = 0.14 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Louisiana

- t-Test:

Maryland:  $t = 1.41 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Maryland

Louisiana:  $t = 0.13 < 1.96$ , Therefore we **Do Not Reject** the hypothesis for Louisiana

All the tests are applicable as the number of samples is more than 30 for one-population tests. According to t-Test, if  $n > 30$  then it tends to follow a Normal Distribution under CLT rules.

We finally conclude by performing all the above tests that our hypothesis that the mean of death deaths per day in the month of Feb2021 is **NOT** different than the number of deaths per day in the month of March2021. Hence both the parameters of the distribution are similar.

Though for Wald's test for Maryland was greater than 1.96, this means in ideal case Wald's test does not capture the idea of hypothesis testing well.

```
[61]: W_md,W_la,Z_md,Z_la,t_md,t_la
```

```
[61]: (3.21422385270194,  
      0.6654011413209963,  
      0.5772916458461961,  
      0.11950957295591827,  
      0.567904164192263,  
      0.11756619835199761)
```

```
[62]: W_md_two,W_la_two,t_md_two,t_la_two
```

```
[62]: (1.4700739216648995,  
      0.14494447930602344,  
      1.4143913482343333,  
      0.13973248500386345)
```

---

```
[ ]: !sudo apt-get install texlive-xetex texlive-fonts-recommended  
      ↪texlive-plain-generic  
      !jupyter nbconvert --to pdf /content/KNN.ipynb
```