IT350 : Data Analytics
Lab Assignment 2
Hypothesis testing

Name: Chinmayi C. Ramakrishna **Date of Submission:** 8th Feb, 2021
Roll No.: 181IT113

Z-test: One sample

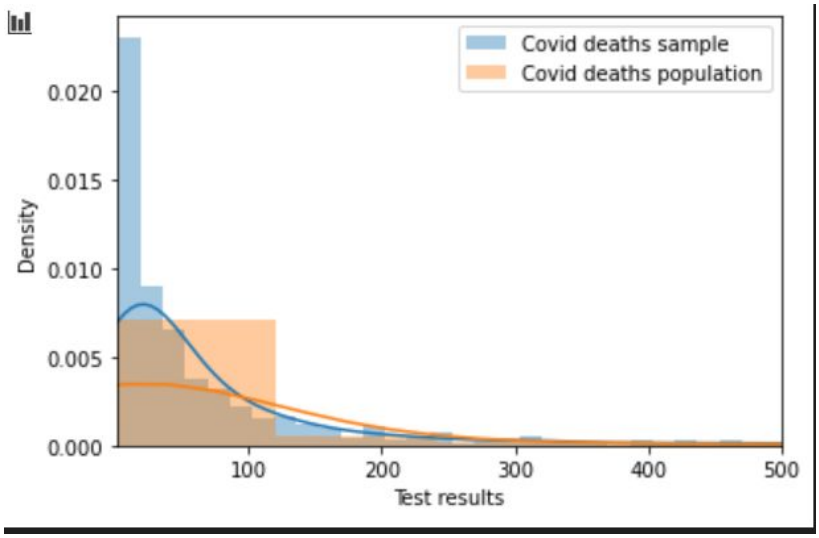

Fig 1. Comparison between covid_deaths sample and population

Step 1: State null hypothesis

Null hypothesis: covid deaths mean = 97.7

Step 2: State alternate hypothesis

Alternate hypothesis: Mean is lesser than 97.7

Step 3: Apply tail test for distribution

We need to apply the left-tailed test.

Step 4: Choose significance level

Significance level : 0.05

Step 5: Compute standard deviation of sample

```
Standard deviation of sample:  1.032991062450859
```

Step 6: Compute critical value at chosen significance level using z score

From the left-tailed z table
Critical value at 0.05 = 1.03- 0.8531 = 0.1769

Step 7: Find z-score

```
Sample mean:  71.53115264797508
Population mean:  97.70478983382209
Population standard deviation:  388.9082048843275
Sample size:  642
```

```
Z-score:  1.7052359753909032
```

Decision: Since z-score is more than the critical value, the null hypothesis can be rejected.

Step 1: State null hypothesis

Null hypothesis: covid deaths mean = 97.7

Step 2: State alternate hypothesis

Alternate hypothesis: covid deaths mean is greater than 97.7

Step 3: Apple tail test for distribution

We need to apply right-tailed test

Step 4: Choose significance level

Significance level : 0.05

Step 5: Compute standard deviation of sample

```
Standard deviation of sample:  1.032991062450859
```

Step 6: Compute critical value at chosen significance level using z score

From the right-tailed z table
Critical value at 0.05 = 1.03+ 0.3531= 1.3831

Step 7: Find z-score

```
Sample mean:  71.53115264797508
Population mean:  97.70478983382209
Population standard deviation:  388.9082048843275
Sample size:  642

Z-score:  1.7052359753909032
```

Decision: Since z-score is greater than the critical value, the null hypothesis can be rejected.

Therefore, covid deaths mean is greater than 97.7

Step 1: State null hypothesis

Null hypothesis: covid deaths mean = 97.7

Step 2: State alternate hypothesis

Alternate hypothesis: covid deaths mean ≠ 97.7

Step 3: Apple tail test for distribution

We need to apply the two-tailed test.

Step 4: Choose significance level

Significance level : 0.05

Step 5: Compute standard deviation of sample

```
Standard deviation of sample:  1.032991062450859
```

Step 6: Compute critical value at chosen significance level using z score

Critical value at 0.05 = 1.03 + 0.3461 = 1.3761

Step 7: Find z-score

```
Z-score:  1.7052359753909032
```

Decision: Since z-score is greater than critical value, the null hypothesis is rejected. Therefore, mean is not equal to 97.7.

z-test : Two sample

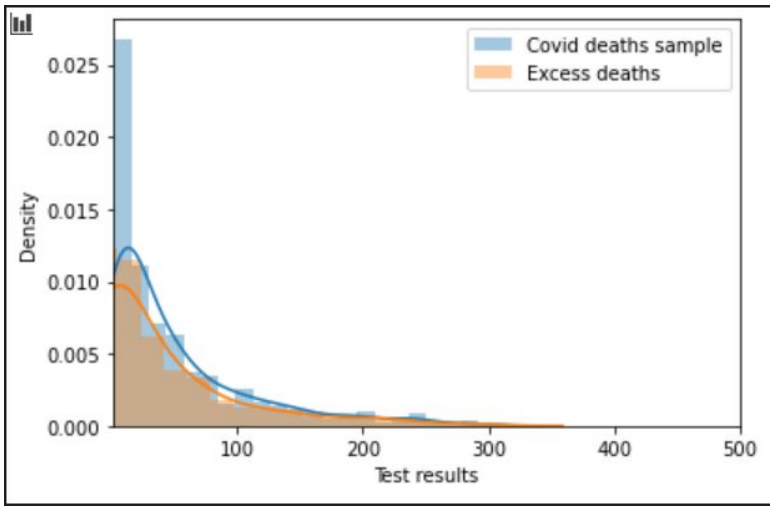

Fig 2. Covid deaths sample and excess deaths sample

Step 1: State null hypothesis

Null hypothesis: covid deaths mean = excess deaths mean

Step 2: State alternate hypothesis

Alternate hypothesis: covid deaths mean ≠ excess deaths mean

Step 3: Apple tail test for distribution

We need to apply the two-tailed test.

Step 4: Choose significance level

Significance level : 0.05

Step 5: Compute standard deviation of sample

```
Standard deviation of covid deaths sample:  1.032991062450859
Standard deviation of excess deaths sample:  2.3385788593429044
```

Step 6: Compute critical value at chosen significance level using z score

Critical value at 0.05 for covid deaths sample = 1.03 + 0.3461 = 1.3761
Critical value at 0.05 for excess deaths sample = 2.33+0.4898 = 2.8198

Step 7: Find z-score

```
Z-score:  227.3830017777252
```

Decision: Since z-score is greater than critical value, the null hypothesis is rejected. Therefore, both the samples do not have equal mean.

T-test

One sample t- test

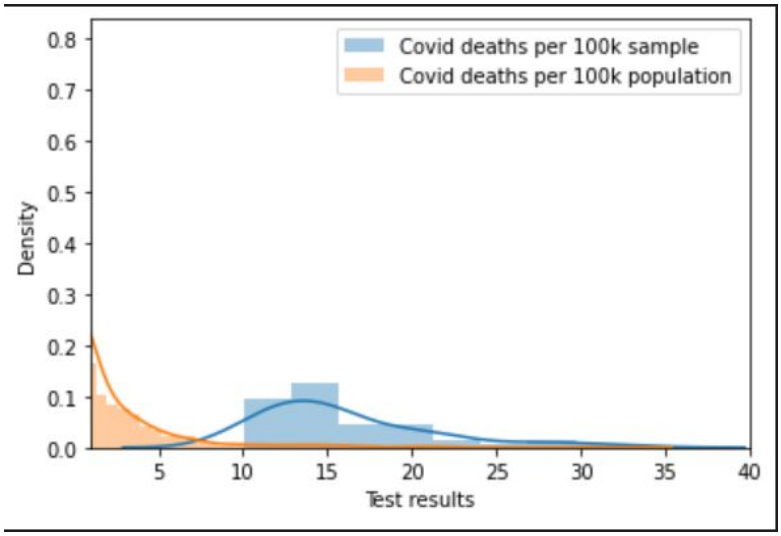

Fig 3. Covid deaths per 100k sample vs population

Step 1: State null hypothesis

Null hypothesis: Covid_deaths_per_100k has a mean of 2.90

Step 2: State alternate hypothesis

Alternate hypothesis: Covid_deaths_per_100k has a mean not equal to 2.90.

Step 3: Apple tail test for distribution

We need to apply two-tailed test

Step 4: Choose significance level

Significance level : 0.05

Step 5: Find degrees of freedom

```
Degrees of freedom: 48 -1 = 47
```

Step 6: Compute critical value at chosen significance level using degrees of freedom

From the two-tailed t table
Critical value at 0.05 = 2.39

Step 7: Find t-score

```
Mean of covid deaths per 100k sample:  16.072397716458443
Mean of population of covid deaths per 100k:  2.115330205271514
Standard deviation of sample of covid deaths per 100k:
5.197312906798237
Sample size of covid deaths per 100k:  48
t-score:  18.605268135712134
```

Decision: Since t-score is greater than the critical value, the null hypothesis can be rejected.

Two Sample t-test: Dependent sample

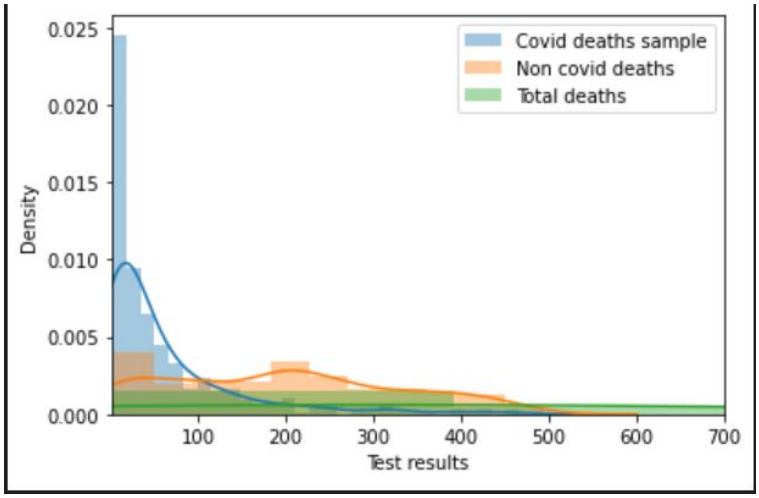

Fig 4. Deaths with covid, without covid and total deaths

Step 1: State null hypothesis

Null hypothesis: There is no difference in deaths with and without covid

Step 2: State alternate hypothesis

Alternate hypothesis: Covid deaths is greater than non covid deaths

Step 3: Apple tail test for distribution

We need to apply right-tailed test

Step 4: Choose significance level

Significance level : 0.05

Step 5: Find degrees of freedom

```
Degrees of freedom:  1325
```

Step 6: Compute critical value at chosen significance level using degrees of freedom

From the right-tailed t table
Critical value at 0.05 = 1.646

Step 7: Find t-score

```
Mean of covid deaths sample:  49.932565789473685
Mean of non covid deaths sample:  91.66295264623956
Mean of population of covid deaths:  97.70478983382209
Mean of population of non covid deaths:  806.0371456500488
Standard deviation of covid deaths:  58.73732097328938
Standard deviation of non covid deaths:  65.24505198110661
Sample size of covid deaths:  630
Sample size of non covid deaths:  697
t-score:  195.85796557225905
```

Decision: Since t-score is greater than the critical value, the null hypothesis can be rejected.

Two Sample t-test : Independent sample

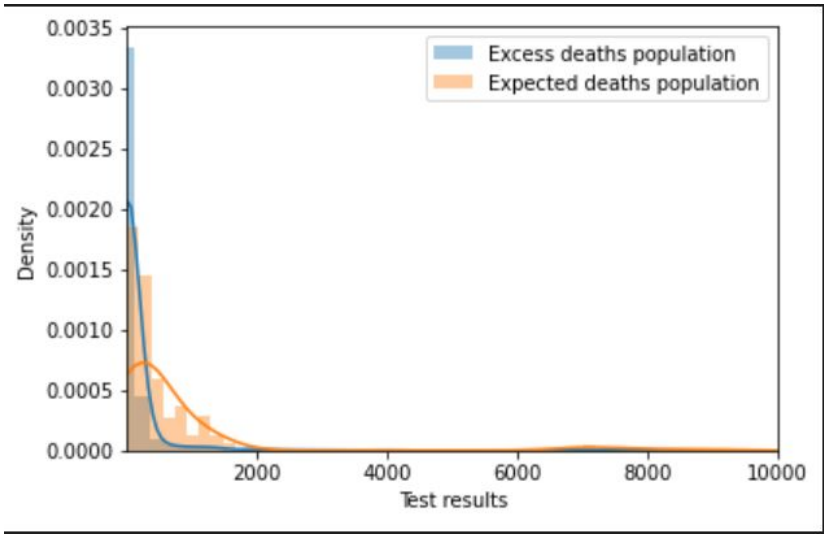

Fig 5. Excess deaths population vs Expected deaths population

Step 1: State null hypothesis

Null hypothesis: There is no difference in the mean of expected deaths and excess deaths

Step 2: State alternate hypothesis

Alternate hypothesis: mean of expected deaths is greater than the mean of excess deaths

Step 3: Apple tail test for distribution

We need to apply right-tailed test

Step 4: Choose significance level

Significance level : 0.05

Step 5: Find degrees of freedom

```
Degrees of freedom:  94+34 -2 = 126
```

Step 6: Compute critical value at chosen significance level using degrees of freedom

From the right-tailed t table
Critical value at 0.05 = 1.660

Step 7: Find t-score

```
Mean of excess deaths sample:  2662.1176470588234
Mean of expected deaths sample:  7.074468085106383
Mean of population of excess deaths:  132.95552884615384
Mean of population of expected deaths:  756.3810096153846
Standard deviation of excess deaths:  2549.2152753015307
Standard deviation of expected deaths:  0.29290395145326925
Sample size of expected deaths:  94
Sample size of excess deaths:  34
t-score:  32.28009442036821
```

Decision: Since t-score is greater than the critical value, the null hypothesis can be rejected.

F-test

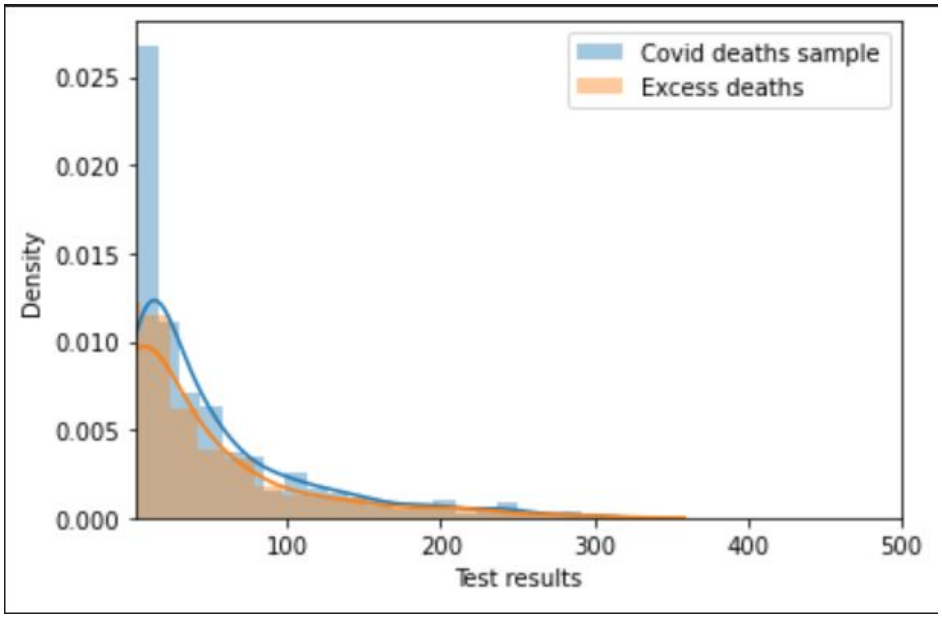

Fig 6. Covid deaths and Excess deaths

Chi square test

Step 1: State null hypothesis

Null hypothesis: Covid deaths and excess deaths are independent

Step 2: State alternate hypothesis

Alternate hypothesis: Covid deaths and excess deaths are not independent

Step 3: Choose significance level

Significance level : 0.05

Step 4: Find degrees of freedom

```
Degrees of freedom:  1
```

Step 5: Compute critical value at chosen significance level using degrees of freedom

From the right-tailed f table
Critical value at 0.05 = 3.841

Step 6: Find chi-square value

	excess_deaths	**expected_deaths**	**Total**
covid_deaths	228	507	**735**
non_covid_deaths	228	507	**735**
Total	**456**	**1014**	**1470**

Table 1. Excess deaths and covid deaths

```
Chi-square-score:  0.0
```

Decision: Since chi-square value is less than the critical value, the null hypothesis cannot be rejected.

Hence excess deaths and covid deaths are independent.

ANOVA test

The different groups considered are:

Covid deaths in the week 1 -10
Covid deaths in the week 11-30
Covid deaths in the week 31-52

Step 1: State null hypothesis

Null hypothesis: There is no variance between the means of the different groups

Step 2: State alternate hypothesis

Alternate hypothesis: Means are not equal

Step 3: Choose significance level

Significance level : 0.05

Step 4: Find degrees of freedom

K = 3
N = 258

```
Df1 = 2
Df2 = 255
```

Step 5: Compute critical value at chosen significance level using degrees of freedom df1 and df2

From the ANOVA f table
Critical value at 0.05 = 2.3473

Step 6: Find F value

```
F-score:  18.50506413762189
```

Decision: Since F-score value is greater than the critical value, the null hypothesis can be rejected.

Therefore, the covid deaths in the three different times have sufficient variance.