

## R Assignment 10.5

6th December 2020

Use the file “coronavirus-cases\_latest.csv”, which contains the daily Coronavirus confirmed cases in the UK to 12 November 2020 (see Lecture 10.2), to find out if there is any significant difference in the daily Coronavirus confirmed cases between: (1) London and Birmingham; (2) London and Manchester, and (3) Birmingham and Manchester. Formulate your tests, draw a conclusion from each test and make an overall comment on the results of your tests.

```
dataframe_corona <- read.csv(file = 'coronavirus-cases_latest.csv') #read data from csv file and store into dataframe
head(dataframe_corona ) #displaying first 6 elements of the data via head
```

```
##   Area.name Area.code Area.type Specimen.date Daily.lab.confirmed.cases
## 1      Adur E07000223      ltla   12/11/2020                0
## 2      Adur E07000223      ltla   11/11/2020                3
## 3      Adur E07000223      ltla   10/11/2020               11
## 4      Adur E07000223      ltla    9/11/2020                8
## 5      Adur E07000223      ltla    8/11/2020                6
## 6      Adur E07000223      ltla    7/11/2020                2
##   Cumulative.lab.confirmed.cases Cumulative.lab.confirmed.cases.rate
## 1                        473                735.6
## 2                        473                735.6
## 3                        470                730.9
## 4                        459                713.8
## 5                        451                701.4
## 6                        445                692.1
```

```
#subsetting the data from dataframe_corona according to city London and fetching the number of confirmed cases
london_confirmed_cases = subset(dataframe_corona, Area.name == "London")$Daily.lab.confirmed.cases
```

```
#finding the average of the daily confirmed cases of london city
mean(london_confirmed_cases)
```

```
## [1] 455.7319
```

```
#subsetting the data from dataframe_corona according to city Birmingham and fetching the number of confirmed cases
birmingham_confirmed_cases = subset(dataframe_corona, Area.name == "Birmingham")$Daily.lab.confirmed.cases
```

```
#finding the average of the daily confirmed cases of Birmingham city
mean(birmingham_confirmed_cases)
```

```
## [1] 117.8911
```

```
#subsetting the data from dataframe_corona according to city Manchester and fetching the number of confirmed cases
manchester_confirmed_cases = subset(dataframe_corona, Area.name == "Manchester")$Daily.lab.confirmed.cases
```

```
#finding the average of the daily confirmed cases of Manchester city
mean(manchester_confirmed_cases)
```

```
## [1] 97.10196
```

This means that the average of Coronavirus confirmed cases on daily basis is more in London as compared to Birmingham and Manchester.

To see if the difference between the two mean is statistically significant. We can use an independent samples t-test:

```
t.test(london_confirmed_cases,birmingham_confirmed_cases) #Solution 1
```

```
##
##  Welch Two Sample t-test
##
## data:  london_confirmed_cases and birmingham_confirmed_cases
## t = 8.8388, df = 294.13, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  262.6170 413.0647
## sample estimates:
## mean of x mean of y
##  455.7319  117.8911
```

The p-value between london and birmingham is  $< 0.000000000000000022$  which means it is way more smaller then the significance level of 0.05 so we reject the null hypothesis. Hence, there is significant difference between the daily Coronavirus confirmed cases between london and birmingham.

```
t.test(london_confirmed_cases,manchester_confirmed_cases) #Solution 2
```

```
##
##  Welch Two Sample t-test
##
## data:  london_confirmed_cases and manchester_confirmed_cases
## t = 9.4139, df = 290.32, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  283.6512 433.6086
## sample estimates:
## mean of x mean of y
##  455.73188  97.10196
```

The p-value between london and manchester is  $< 0.000000000000000022$  which means it is way more smaller then the significance level of 0.05 so we reject the null hypothesis. Hence, there is significant difference between the daily Coronavirus confirmed cases between london and manchester.

```
t.test(birmingham_confirmed_cases,manchester_confirmed_cases) #Solution 3
```

```
##
##  Welch Two Sample t-test
##
## data:  birmingham_confirmed_cases and manchester_confirmed_cases
## t = 2.2181, df = 1010.6, p-value = 0.02677
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.396977 39.181203
## sample estimates:
## mean of x mean of y
## 117.89105  97.10196
```

The p-value between Birmingham and Manchester is 0.02677 which means it is smaller then the significance level of 0.05. so we reject the null hypothesis. Hence, there is significant difference between the daily Coronavirus confirmed cases between Birmingham and Manchester.

From above analysis, it can be concluded that there is significant difference in the daily Coronavirus confirmed cases between: (1) London and Birmingham; (2) London and Manchester, and (3) Birmingham and Manchester.