# Data Analytics

## Testing of Hypothesis – Assignment

Name: - Ashish

Roll no.: - 17BCS006

```
install.packages("gapminder")
library(gapminder)
data("gapminder")
View(gapminder)
install. packages("dplyr")
library(dplyr)

# -------------------------------------- T-Test ------------------------------------------------------#

## One-Sample T-testing
mean(gapminder$lifeExp)
t.test (gapminder$lifeExp, mu = 5, alternative = "less", conf.level = 0.99)

## Tow-Sample T-testing
View(gapminder)
 df1 <- gapminder %>%
  select(country, lifeExp)%>%
  filter (country == "South Africa" | country == "Ireland")
 t.test (data = df1, lifeExp ~ country)

## Paired t-test
 t.test (gapminder$lifeExp, gapminder$pop, paired = TRUE, alternative = "two.sided")

# -------------------------------------- Z-Test ------------------------------------------------------#

## One Sample z-test
 library(gapminder)
 x <- rnorm(gapminder)
 z.test (x, sigma.x = 2)

## Two-Sample Z-Test
 x <- rnorm(gapminder)
 y <- rnorm(gapminder)
 z.test (x, sigma.x = 0.5, y, sigma. y = 0.5, mu=2)
 z.test (x, sigma.x=0.5, y, sigma. y=0.5, conf.level=0.90)
```

```
#----------------------------------------F-Test-------------------------------------------------------#
#  F test to compare two variances

 #* Method 1
 x <- gapminder$year
 y <- gapminder$lifeExp
 var.test (x, y, alternative = "two.sided")

 #* or Method 2
  var.test (lifeExp ~ country, data = df1, alternative = "two.sided")

 #* or Method 3
 res.ftest <- var.test (lifeExp ~ country, data = df1)
 res.ftest
  #*  Pearson's Chi-squared test

 chisq.test (data_frame$treatment, data_frame$improvement, correct=FALSE)

#------------------------------Chi-squared test--------------------------------------------------#

data_frame <- read.csv("https://goo.gl/j6lRXD")  #Reading CSV
 table (data_frame$treatment, data_frame$improvement)
 chisq.test(data_frame$treatment, data_frame$improvement, correct=FALSE)
 data("mtcars")
 table (mtcars$carb,  mtcars$cyl)
 chisq.test (mtcars$carb,  mtcars$cyl)
```

Output: ---

```
> #install.packages("gapminder")
> library(gapminder)
> data("gapminder")
> View(gapminder)
>
> #install.packages("dplyr")
> library(dplyr)
>
> ## ------------One-Sample T-testing------------##
> mean(gapminder$lifeExp)
[1] 59.47444
> t.test(gapminder$lifeExp, mu = 5, alternative = "less", conf.level = 0.9
9)

        One Sample t-test

data:  gapminder$lifeExp
t = 174.09, df = 1703, p-value = 1
alternative hypothesis: true mean is less than 5
99 percent confidence interval:
     -Inf 60.20308
sample estimates:
mean of x
 59.47444

>
> ## -------------Tow-Sample T-test--------------##
> View(gapminder)
>   df1 <- gapminder %>%
+     select(country,lifeExp)%>%
+     filter(country == "South Africa" | country == "Ireland")
>   t.test(data = df1, lifeExp ~ country)

        Welch Two Sample t-test

data:  lifeExp by country
t = 10.067, df = 19.109, p-value = 4.466e-09
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 15.07022 22.97794
sample estimates:
     mean in group Ireland mean in group South Africa
                  73.01725                   53.99317

>
> ##  -------------Paired t-test-----------------##
>   t.test(gapminder$lifeExp, gapminder$pop, paired = TRUE, alternative =
"two.sided")

        Paired t-test

data:  gapminder$lifeExp and gapminder$pop
t = -11.51, df = 1703, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -34645145 -24557161
sample estimates:
mean of the differences
              -29601153

>
> ## --------------One Sample z-test-------------##
>   library(gapminder)
>   x <- rnorm(gapminder)
>   z.test(x,sigma.x = 2)
```

```
        One-sample z-Test

data:  x
z = 0.47133, p-value = 0.6374
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -1.215464  1.985143
sample estimates:
mean of x
0.3848395

>
> ## ------------Two-Sample Z-Test--------------##
>   x <- rnorm(gapminder)
>   y <- rnorm(gapminder)
>   z.test(x,sigma.x = 0.5, y, sigma.y = 0.5, mu=2)

        Two-sample z-Test

data:  x and y
z = -5.1092, p-value = 3.236e-07
alternative hypothesis: true difference in means is not equal to 2
95 percent confidence interval:
 -0.04068658  1.09089916
sample estimates:
 mean of x  mean of y
 0.3479587 -0.1771476

>   z.test(x, sigma.x=0.5, y, sigma.y=0.5, conf.level=0.90)

        Two-sample z-Test

data:  x and y
z = 1.819, p-value = 0.06891
alternative hypothesis: true difference in means is not equal to 0
90 percent confidence interval:
 0.05027795 0.99993463
sample estimates:
 mean of x  mean of y
 0.3479587 -0.1771476

>
>
> ##############----------F-test---------------------#
>   #* Method 1
>   x <- gapminder$year
>   y <- gapminder$lifeExp
>   var.test(x, y, alternative = "two.sided")

        F test to compare two variances

data:  x and y
F = 1.7866, num df = 1703, denom df = 1703, p-value < 2.2e-16
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.624622 1.964654
sample estimates:
ratio of variances
          1.786567

>
>   #* or Method 2
>    var.test(lifeExp ~ country, data = df1, alternative = "two.sided")

        F test to compare two variances

data:  lifeExp by country
F = 0.43993, num df = 11, denom df = 11, p-value = 0.189
alternative hypothesis: true ratio of variances is not equal to 1
```

```
95 percent confidence interval:
 0.1266457 1.5281806
sample estimates:
ratio of variances
         0.4399289

>
>   #* or Method 3
>   res.ftest <- var.test(lifeExp ~ country, data = df1)
>   res.ftest

        F test to compare two variances

data:  lifeExp by country
F = 0.43993, num df = 11, denom df = 11, p-value = 0.189
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1266457 1.5281806
sample estimates:
ratio of variances
         0.4399289

>
>
>  chisq.test(data_frame$treatment, data_frame$improvement, correct=FALSE)

        Pearson's Chi-squared test

data:  data_frame$treatment and data_frame$improvement
X-squared = 5.5569, df = 1, p-value = 0.01841

>
> ###-------------Chi-squared test--------------------#
>   data_frame <- read.csv("https://goo.gl/j6lRXD")  #Reading CSV
>   table(data_frame$treatment, data_frame$improvement)

              improved not-improved
  not-treated       26           29
  treated           35           15
>  chisq.test(data_frame$treatment, data_frame$improvement, correct=FALSE)

        Pearson's Chi-squared test

data:  data_frame$treatment and data_frame$improvement
X-squared = 5.5569, df = 1, p-value = 0.01841

>   data("mtcars")
>   table(mtcars$carb, mtcars$cyl)

    4  6  8
  1 5  2  0
  2 6  0  4
  3 0  0  3
  4 0  4  6
  6 0  1  0
  8 0  0  1
>   chisq.test(mtcars$carb, mtcars$cyl)

        Pearson's Chi-squared test

data:  mtcars$carb and mtcars$cyl
X-squared = 24.389, df = 10, p-value = 0.006632

Warning message:
In chisq.test(mtcars$carb, mtcars$cyl) :
  Chi-squared approximation may be incorrect
```