DA Lab 4

Solving Problems from **Testing of Hypothesis Pdf** using code

Name: Dheeraj Chaudhary

Roll: 17BCS009

```
library(dplyr)
```

Z-Test for Testing of Hypothesis

```
####### Two sample Z-Test for Test of significant in difference between means #########
#### Here Given values are MEAN, SSAMPLE SIZE and STANDARD DEVIATION of Both Samples ######

##### Considering the Examples given in DR. ATHE's TESTING OF HYPOTHESIS Pg.No. 35  #####

# There is no significant difference between the means of two paddy fields

# Given DATA for Field No 1
field_1_Sample_Size <- 60
StandDeviation_1 <- 1.15
field_1_Sample_MEAN <- 18.5

# Given DATA for Field No 2
field_2_Sample_Size <- 60
StandDeviation_2 <- 1.15
field_2_Sample_MEAN <- 20.3

# calculating the Z-statistic at 5% LOS (i.e, Tabulated value is 1.96)
Z_Calculated <- (field_2_Sample_MEAN - field_1_Sample_MEAN) /
  sqrt((StandDeviation_1^2/ field_1_Sample_Size) + (StandDeviation_2^2 /
field_2_Sample_Size))

Z_Calculated
```

```
##### OUTPUT

          [1] 8.573049

## SINCE we got calculated value as 8.57 which is greater than tabulated value(1.96) So
reject Null Hypothesis that means both the fields have significant difference
```

T-Test for Testing of Hypothesis

```
####### Two sample T-Test for Test of significant in difference between means #########

#### Here Given values are Potato Plant yield Tubes fro two different varieties ######


##### Considering the Examples given in DR. ATHE's TESTING OF HYPOTHESIS Pg.No. 41  #####



#Consider Null Hypothesis as mean number of tubes of the variety_1 significantly differ from
the variety_2


# Given Data is

Variety_1 <- c(2.2, 2.5, 1.9, 2.6, 2.6, 2.3, 1.8, 2.0, 2.1, 2.4, 2.3)

Variety_2 <- c(2.8, 2.5, 2.7, 3.0, 3.1, 2.3, 2.4, 3.2, 2.5, 2.9)


# Create a data frame for the above both variety

my_data <- data.frame(

  Types=c(rep("Variety_1",11),rep("Variety_2",10)),

  num_tubes = c(Variety_1,  Variety_2)

)

my_data

#### OUTPUT

    Types num_tubes

1  Variety_1       2.2
2  Variety_1       2.5
3  Variety_1       1.9
4  Variety_1       2.6
5  Variety_1       2.6
6  Variety_1       2.3
7  Variety_1       1.8
8  Variety_1       2.0
9  Variety_1       2.1
10 Variety_1       2.4
11 Variety_1       2.3
12 Variety_2       2.8
13 Variety_2       2.5
14 Variety_2       2.7
15 Variety_2       3.0
16 Variety_2       3.1
17 Variety_2       2.3
```

```
18 Variety_2       2.4
19 Variety_2       3.2
20 Variety_2       2.5
21 Variety_2       2.9
```

```
# calculating the T- Test at 5% LOS (i.e, Tabulated value is 2.09)

group_by(my_data, Types) %>%

summarise(

  sample_size = n(),

  sample_mean = mean(num_tubes, na.rm = TRUE),

  sample_sd = sd(num_tubes, na.rm = TRUE)

)

#### OUTPUT

# A tibble: 2 x 4

  Types     sample_size sample_mean sample_sd
  <fct>           <int>       <dbl>     <dbl>
1 Variety_1          11        2.25     0.273
2 Variety_2          10        2.74     0.310


t.test(Variety_1, Variety_2)

#### OUTPUT

        Welch Two Sample t-test


data:  Variety_1 and Variety_2
t = -3.8625, df = 18.091, p-value = 0.001132
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.7634435 -0.2256474
sample estimates:
mean of x mean of y
 2.245455  2.740000


##So in the output we can see that calculated value (3.8) is greater than Tabulated value
(2.09) at 5%LOS, and hence. We conclude that the mean number of tubes of the variety_1
significantly not differ from the variety_2
```

F-Test for Testing of Hypothesis

```
####### Two sample F-Test for Test of significant in difference between means #########

#### Given values are two types of water to irrigate the gram plants ######


##### Considering the Examples given in DR. ATHE's TESTING OF HYPOTHESIS Pg.No. 45  #####



# Consider Null Hypothesis as The variances of the two systems of irrigation are homogeneous
```

```r
# Given Data is
Tap_water <- c(3.5, 4.2, 2.8, 5.2, 1.7, 2.6, 3.5, 4.2, 5.0, 5.2)
Saline_water <- c(1.9, 2.6, 2.3, 4.3, 4.0, 4.2, 3.8, 2.9, 3.7)


# Create a data frame for the above both variety
my_data <- data.frame(
  Types_of_irrigation=c(rep("Saline_water",9),rep("Tap_water",10)),
  Height_of_Plants = c(Saline_water, Tap_water)
)


my_data
#### OUTPUT
   Types_of_irrigation Height_of_Plants
1         Saline_water              1.9
2         Saline_water              2.6
3         Saline_water              2.3
4         Saline_water              4.3
5         Saline_water              4.0
6         Saline_water              4.2
7         Saline_water              3.8
8         Saline_water              2.9
9         Saline_water              3.7
10           Tap_water              3.5
11           Tap_water              4.2
12           Tap_water              2.8
13           Tap_water              5.2
14           Tap_water              1.7
15           Tap_water              2.6
16           Tap_water              3.5
17           Tap_water              4.2
18           Tap_water              5.0
19           Tap_water              5.2



# Calculating the F-statistic at 5% LOS (i.e, Tabulated value is 3.23)
group_by(my_data, Types_of_irrigation) %>%
  summarise(
    sample_size = n(),
    sample_mean = mean(Height_of_Plants, na.rm = TRUE),
    sample_sd = sd(Height_of_Plants, na.rm = TRUE)
  )
#### OUTPUT
# A tibble: 2 x 4
  Types_of_irrigation sample_size sample_mean sample_sd
  <fct>                     <int>       <dbl>     <dbl>
1 Saline_water                  9        3.3      0.889
2 Tap_water                    10        3.79     1.19


res.ftest <- var.test(Tap_water, Saline_water)
res.ftest
```

```
#### OUTPUT

     F test to compare two variances

data:  Tap_water and Saline_water
F = 1.7875, num df = 9, denom df = 8, p-value = 0.4254
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4102334 7.3321737
sample estimates:
ratio of variances
          1.787482
```

```
## So in the output we can see that calculated value(1.78) is less than tabulate value(which is 3.23) So accept the null Hypothesis, hence we conclude that the variances of the two systems of irrigation are homogeneous
```

Chi-Square Test for Hypothesis

```
####### Chi-Square Test to Test the significant difference between means #########

## Given values are number of two types plants with two charactor as their leaf colour ###

##### Considering the Examples given in DR. ATHE's TESTING OF HYPOTHESIS Pg.No. 49  #####

# Consider Null hypothesis attributes as flower_colour and Shape_of_leaf are independent of each other

# Calculating the Chi-square test at 5% LOS (i.e, Tabulated value is 3.84)

data <- as.table(rbind(c(99, 36), c(20, 5)))

dimnames(data) <- list(Flowe_color = c("White_Flower", "Red_Flower"),

                    Shape_of_leaf = c("Flat_leaf","Cirled_leaf"))

(Xsq <- chisq.test(data))

#### OUTPUT

       Pearson's Chi-squared test with Yates' continuity correction

data:  M
X-squared = 0.20429, df = 1, p-value = 0.6513

# Observed counts (same as data) will be

Xsq$observed

#### OUTPUT

            Shape_of_leaf
Flowe_color    Flat_leaf Cirled_leaf
  White_Flower        99          36
  Red_Flower          20           5
```

```
# expected counts under the null

Xsq$expected

#### OUTPUT

              Shape_of_leaf
Flowe_color     Flat_leaf Cirled_leaf
  White_Flower 100.40625    34.59375
  Red_Flower    18.59375     6.40625


## So we can see the output as Calulated value(0.20) which is less than tabulated value(3.84) at 5% LOS. So Null hypothesis is accepted. And hence we conclude that two characters, flower colour and shape of leaf are independent of each other
```