HYPOTHESIS TESTING IN R

- LARGE SAMPLE TESTS -

1. **ONE SAMPLE Z-TEST**
   Code :

```
######### 4. ONE SAMPLE Z-TEST  ########
#Given data : suppose that a student is interesting in estimating
#how many memes their professors know and love.
#So they go to class, and every time a professor uses a new meme,
#they write it down. After a year of classes,
#the student has recorded the following meme counts,
#where each count corresponds to a single class they took:
x<-c(3,7,11,0,7,0,4,5,6,2)
n<-length(x)
z_stat<-(mean(x)-3/(2/sqrt(n)))    #standard deviation is given = 2
z_stat
```

   Output :

```
> x<-c(3,7,11,0,7,0,4,5,6,2)
> n<-length(x)
> z_stat<-(mean(x)-3/(2/sqrt(n)))    #standard deviation is given = 2
> z_stat
[1] -0.2434165
```

2. **TWO SAMPLE Z-TEST**
   Code :

```r
######### 5. TWO SAMPLE Z-TEST ########
#Given data :  suppose that a student wants to figure out if biology profe
#English professors know more memes. The student writes a meme quiz
#and springs it on 14 unsuspecting biology professors
#and 18 unsuspecting English professors during office hours.
biology <- c(3, 7, 11, 0, 7, 0, 4, 5, 6, 2, 4, 7, 2, 9)
english <- c(8, 5, 4, 10, 4, 5, 7, 2, 6, 1, 2, 7, 0, 6, 4, 12, 5, 2)
# Creating data frame
test_results <- data.frame(
  score = c(biology, english),
  department = c(rep("biology", 14),
    rep("english",18)
  )
)
delta_0<-0
#an assumption
sigma_sq_1 <- 3
sigma_sq_2 <- 2
n1<-14
n2<-18
#calculating the z-statistic
z_stat <- (mean(biology) - mean(english) - delta_0) /
  sqrt(sigma_sq_1 / n1 + sigma_sq_2 / n2)
z_stat
```

Output :

```
> #calculating the z-statistic
> z_stat <- (mean(biology) - mean(english) - delta_0) /
+   sqrt(sigma_sq_1 / n1 + sigma_sq_2 / n2)
> z_stat
[1] -0.3756527
```

- <u>SMALL SAMPLE TESTS</u> -

1. **ONE SAMPLE T-TEST**

   Code :

```r
library("dplyr")

######### 1. ONE SAMPLE T-TEST (INDEPENDENT T-TEST) ########
#Given data : suppose that a student is interesting in estimating
#how many memes their professors know and love.
#So they go to class, and every time a professor uses a new meme,
#they write it down. After a year of classes,
#the student has recorded the following meme counts,
#where each count corresponds to a single class they took:
x<-c(3,7,11,0,7,0,4,5,6,2)
t.test(x,mu=3)
```

Output :

```
> t.test(x,mu=3)

        One Sample t-test

data:  x
t = 1.3789, df = 9, p-value = 0.2012
alternative hypothesis: true mean is not equal to 3
95 percent confidence interval:
 2.0392 6.9608
sample estimates:
mean of x
      4.5
```

2. **TWO SAMPLE T-TEST**

   Code :

```r
######### 2. TWO SAMPLE T-TEST (INDEPENDENT T-TEST) ########
#Given data is of the two populations being compared are
#“men who have not taken caffeine”
#and “men who have taken caffeine”.
placebo<-c(105,119,100,97,96,101,94,95,98)
caffeine<-c(96,99,94,89,96,93,88,105,88)
#Creating dataframe for both the population
pop_data<-data.frame(Types=c(rep("placebo",9),rep("caffeine",9)),
                     num_tubes = c(placebo,  caffeine))
pop_data
group_by(pop_data, Types) %>%
  summarise(sample_size = n(),
  sample_mean = mean(num_tubes, na.rm = TRUE),
  sample_sd = sd(num_tubes, na.rm = TRUE))
t.test(placebo,caffeine)
```

Output :

```
        Welch Two Sample t-test

data:  placebo and caffeine
t = 1.9948, df = 14.624, p-value = 0.06505
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4490961 13.1157627
sample estimates:
mean of x mean of y
100.55556  94.22222
```

```
> pop_data
      Types num_tubes
1   placebo       105
2   placebo       119
3   placebo       100
4   placebo        97
5   placebo        96
6   placebo       101
7   placebo        94
8   placebo        95
9   placebo        98
10 caffeine        96
11 caffeine        99
12 caffeine        94
13 caffeine        89
14 caffeine        96
15 caffeine        93
16 caffeine        88
17 caffeine       105
18 caffeine        88
> group_by(pop_data, Types) %>%
+   summarise(sample_size = n(),
+   sample_mean = mean(num_tubes, na.rm = TRUE),
+   sample_sd = sd(num_tubes, na.rm = TRUE))
# A tibble: 2 x 4
  Types    sample_size sample_mean sample_sd
  <fct>          <int>       <dbl>     <dbl>
1 caffeine           9        94.2      5.61
2 placebo            9       101.       7.70
```

3. **PAIRED T-TEST**
   Code :

```
######### 3. PAIRED T-TEST  ########
#Given data
x1=c(7.3, 6.8, 7.3, 4.8, 5.6, 6.2, 5.6, 4.5,
     6.3, 6.7, 5.4, 6.8, 6.5, 7.7, 8.3)
x2=c(9.1, 8.6, 7.1, 9.6, 9.7, 9.8, 9.1, 5.2,
     9.4, 9.3, 4.9, 10.1, 6.3, 10.2, 10.9)
t.test(x1,x1,paired = TRUE)
```

Output :

```
        Paired t-test

data:  x1 and x1
t = NaN, df = 14, p-value = NA
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 NaN NaN
sample estimates:
mean of the differences
                      0
```

4. **F-TEST**

Code :

```
################ 6. F-Test FOR COMPARING TWO VARIENCES ########
# Given data
y1=c(45, 87, 123, 120, 70)
y2=c(51, 71, 42, 37, 51, 78, 51, 49, 56, 47, 58)
var.test(y1,y2)
```

Output :

```
        F test to compare two variances

data:  y1 and y2
F = 7.7881, num df = 4, denom df = 10, p-value = 0.008108
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  1.74296 68.87739
sample estimates:
ratio of variances
          7.788141
```

5. **CHI-SQUARED TEST**

Code :

```
######### 7. CHI-SQUARED TEST FOR NOMINAL (CATEGIORICAL) DATA ########
A<-c(38,33,42,26,11)
B<-c(72,57,38,44,29)
M <- as.table(rbind(A,B))
M

(Xsq <- chisq.test(M))
Xsq$observed
Xsq$expected
```

Output :

```
   A  B  C  D  E
A 38 33 42 26 11
B 72 57 38 44 29
> (Xsq <- chisq.test(M))

        Pearson's Chi-squared test

data:  M
X-squared = 9.5785, df = 4, p-value = 0.04816

> Xsq$observed
   A  B  C  D  E
A 38 33 42 26 11
B 72 57 38 44 29
> Xsq$expected
         A        B        C        D        E
A 42.30769 34.61538 30.76923 26.92308 15.38462
B 67.69231 55.38462 49.23077 43.07692 24.61538
```