Question 6:-

If a disease is likely to spread in a particular weather condition (data given in the disease index sheet), then access of that disease should be more in the months having suitable weather conditions. Help the analyst in coming up with a statistical test to support the claim for two districts for which the sample of weather and disease access data is provided in the datasheet. Identify the diseases for which you can support this claim. Test this claim both for temperature and relative humidity at 95% confidence.

Solution:-

We need to compare the diseases affected in any two districts with respect to weather and humidity.

#Lets Assume

Avg disease access when conditions are not favorable to disease = μ1

Avg disease access when conditions are favorable to disease= μ2

#Hypothesis

H0: null hypothesis says: μ2 <= μ1

Ha: alternative hypothesis says: μ2 > μ1

Approach

We will Two Sample T-test to compare sample mean of the given data sheets. This will verify which hypothesis we will support and which one is to reject.

##############

```
install.packages("janitor")
library(janitor)

install.packages("dplyr")
library(dplyr)


bdata <- read.csv("clipboard", sep = "\t", header = T)
names(bdata)

bdata1 <- subset(bdata, Temperature <= 24 & Relative.Humidity >=80)
```

```
data$D1_group <- ifesle(Temperature <= 24 & Relative.Humidity >=80, "favourable",
"unfavourable")

bdata2 <- subset(bdata, !(Temperature <= 24 & Relative.Humidity >=80))

bdata1$weather <- "favourable"
bdata2$weather <- "unfavourable"
newdata <- rbind (bdata1, bdata2)

test1 <- t.test(newdata$D1~newdata$weather, conf.level = 0.95, paired = FALSE, var.eq =
TRUE)
```

test1

```
Two Sample t-test

data:  newdata$D1 by newdata$weather
t = 2.7615, df = 22,
p-value = 0.01139 (99%)
alternative hypothesis: true difference in means between group favourable and group
unfavourable is not equal to 0
95 percent confidence interval:
  6.424349 45.175651
sample estimates:
  mean in group favourable
                  37.66667
mean in group unfavourable
                  11.86667
```

Result:- The p-value of 0.01139 suggests some evidence against the null hypothesis that there is no significant difference between the groups being compared.

With a significance level of 0.05 (as is commonly used), the p-value is greater than the threshold, indicating that we cannot reject the null hypothesis at a standard level of significance. However, if a more liberal significance level (e.g. 0.1) were used, we may be able to reject the null hypothesis and conclude that there is a significant difference between the groups.

```
test2 <- t.test(newdata$D2~newdata$weather, mu = 0, alternative = "two.sided", conf.level =
0.95, paired = FALSE, var.eq = TRUE)
```

test2

```
Two Sample t-test
```

data: newdata$D2 by newdata$weather
t = 3.465, df = 22, p-value = 0.002201
alternative hypothesis: true difference in means between group favourable and group unfavourable is not equal to 0
95 percent confidence interval:
 7.503318 & 29.874459
sample estimates:
 mean in group favourable mean in group unfavourable
                27.555556   8.866667

Results:- The p-value of 0.002201 suggests strong evidence against the null hypothesis that there is no significant difference between the groups being compared.

With a significance level of 0.05 (as is commonly used), the p-value is smaller than the threshold, indicating that we can reject the null hypothesis and conclude that there is a significant difference between the groups.----

test3 <- t.test(newdata$D3~newdata$weather, mu = 0, alternative = "two.sided", conf.level = 0.95, paired = FALSE, var.eq = TRUE)

test3

Two Sample t-test

data: newdata$D3 by newdata$weather
t = 3.5788, df = 22, p-value = 0.001675
alternative hypothesis: true difference in means between group favourable and group unfavourable is not equal to 0
95 percent confidence interval:
 11.85836 & 44.54164
sample estimates:
 mean in group favourable mean in group unfavourable
                39.66667   11.46667

Results:- The p-value of 0.001675 suggests strong evidence against the null hypothesis that there is no significant difference between the groups being compared.

With a significance level of 0.05 (as is commonly used), the p-value is smaller than the threshold, indicating that we can reject the null hypothesis and conclude that there is a significant difference between the groups.
---

test4 <- t.test(newdata$D4~newdata$weather, mu = 0, alternative = "two.sided", conf.level = 0.95, paired = FALSE, var.eq = TRUE)

test4

Two Sample t-test

data: newdata$D4 by newdata$weather
t = 3.9257, df = 22, p-value = 0.0007229
alternative hypothesis: true difference in means between group favourable and group unfavourable is not equal to 0
95 percent confidence interval:
 8.050571 26.082762
sample estimates:
 mean in group favourable mean in group unfavourable
                26.000000   8.933333

Results:- The p-value of 0.0007229 suggests strong evidence against the null hypothesis that the mean values of the two groups being compared are equal.

With a significance level of 0.05 (as is commonly used), the p-value is smaller than the threshold, indicating that we can reject the null hypothesis and conclude that there is a significant difference in the means of the two groups.

---

test5 <- t.test(newdata$D4~newdata$weather, mu = 0, alternative = "two.sided", conf.level = 0.95, paired = FALSE, var.eq = TRUE)

test5

Two Sample t-test

Two Sample t-test

data: newdata$D5 by newdata$weather
t = 5.7852, df = 22, p-value = 8.062e-06
alternative hypothesis: true difference in means between group favourable and group unfavourable is not equal to 0
95 percent confidence interval:
 16.13775 34.17336
sample estimates:
 mean in group favourable mean in group unfavourable
                30.55556   5.40000

Results:- The p-value of 8.062e-06 suggests strong evidence against the null hypothesis that the mean values of D1 are equal between the different levels of weather in the newdata dataset.

With a significance level of 0.05 (as is commonly used), the p-value is much smaller than the threshold, indicating that we can reject the null hypothesis and conclude that there is a significant difference in the mean values of D1 between the different levels of weather.

```
test6 <- t.test(newdata$D7~newdata$weather, mu = 0, alternative = "two.sided", conf.level = 0.95, paired = FALSE, var.eq = TRUE)
test6
```

```
Two Sample t-test

data:  newdata$D7 by newdata$weather
t = 1.0663, df = 22, p-value = 0.2979
alternative hypothesis: true difference in means between group favourable and group unfavourable is not equal to 0
95 percent confidence interval:
 -12.47405  38.87405
sample estimates:
  mean in group favourable mean in group unfavourable
                  35.66667                   22.46667
```

Results:- The p-value of 0.2979 suggests weak or no evidence against the null hypothesis that there is no significant difference between the groups being compared.

With a significance level of 0.05 (as is commonly used), the p-value is greater than the threshold, indicating that we cannot reject the null hypothesis at a standard level of significance. This means that we do not have enough evidence to conclude that there is a significant difference between the groups.