Two Sample T-test

By: Samia Porto, Andrew Slavetskas

This **two-sample t-test** compares two independent sets of data that have equal or similar variances. The variances can be compared using another test, such as the **F-test** or the **Levene test**. After the variances are determined to be equal, this t-test can compare the means of those sets of data to determine whether the two means are statistically similar.

Two Sample t-Test with Equal Variances

Assumptions

- The values in both datasets are random and normally distributed
- Variances are approx equal but unknown
- X1 and X2 are independent

Null and Alternative Hypotheses

- H0 -> $\mu_1 = \mu_2$ — The null states that the two means are equal.
- H1 -> $\mu_1 \neq \mu_2$ — The alternative states that they are not equal

Prototype in R:- ***(Based on the example used for the F-test)***

```
# X=read.table("Pizza.txt",header=T)

#  attach(X)
#  X1=Time[Company=="A"]
   X1
[1] 20.4 24.2 15.4 21.4 20.2 18.5 21.5


#  X2=Time[Company=="B"]
   X2
[1] 20.2 16.9 18.5 17.3 20.5
```

Time	*Company*
20.4	A
24.2	A
15.4	A
21.4	A
20.2	A
18.5	A
21.5	A
20.2	B
16.9	B
18.5	B
17.3	B
20.5	B

Test Statistics:

Two Sided Case

```
# t.test(X1,X2,alternative="two.sided",var.equal=T,conf.level=0.95)

        Two Sample t-test

data:  X1 and X2
t = 1.1194, df = 10, p-value = 0.2892
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.533926  4.631068
sample estimates:
mean of x mean of y
 20.22857  18.68000
```

One Sided Case (Lower Tail)

```
# t.test(X1,X2,alternative="less",var.equal=T,conf.level=0.95)

        Two Sample t-test

data:  X1 and X2
t = 1.1194, df = 10, p-value = 0.8554
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
     -Inf 4.056003
sample estimates:
mean of x mean of y
 20.22857  18.68000
```

One Sided Case (Upper Tail)

```
> t.test(X1,X2,alternative="greater",var.equal=T,conf.level=0.95)

        Two Sample t-test

data:  X1 and X2
t = 1.1194, df = 10, p-value = 0.1446
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.9588602        Inf
sample estimates:
mean of x mean of y
 20.22857  18.68000
```

R gives results with a t statistic value, degrees of freedom (df), and the P-value. It also gives the confidence interval for the test and the actual means of the two data sets.

Explanation of the variables in the T-test command:

X1 and X2 are the two data sets.
Alternative is H_1 or the opposite of the null hypotheses; possible choices for this field are two.sided, upper, or lower.
Var.equal tells R that the variances are equal (T) or not (F). If this flag is set to F then this test is used for unequal variances.
Conf.level is the confidence level for the test; it equates to 1-α and sets the confidence intervals for the test.

Decision Rule:

If P-value > α, accept H_0, and reject the H_1
If P-value > α, accept H_1, and reject H_0

Conclusion :

If we assume that the error rate (α) is 0.05→ in this case P-value > α, we accept Null hypotheses. As in both samples have equal means.

Two Sample t-Test with Unequal Variances

** In the case of Two Sample t-Test with unequal variances, we follow the same steps as the previous example. The only difference is the prototype in R will follow the following format:

t.test(X1,X2, alternative=c("two.sided", "less", "greater"), var.equal=F,conf.level= (α-1))