```r
#Two sample t-test (for populations with EQUAL variance) prototype
#Justin Mann
#2/22/2016

#A test for two sets of measurments in which the variances are appoximately equal.

#Assumptions:
# Observed values x1.1...x1.n are a random sample from a normal distribution.
# Observed values x2.1...x2.n are a random sample from a normal distribution.
# Variances are approximately equal but unknown
# Both sample are independent.
## Note this test is reasonably robust for deviation from normal distribution in the two-tailed case when sample sizes are similar.

# Hypotheses:
#1) Null: Mu1 is equal to Mu2.
#2) Alternative: Mu1 is not equal to Mu2

#Paperwork

#read in data
iris

#assign variables
x1 <- iris$Sepal.Length[iris$Species=="setosa"]
x1

x2 <- iris$Sepal.Length[iris$Species=="versicolor"]
x2

#assign number of observations
n1 <- length(x1)
n1

n2 <- length(x2)
n2

#assign means
x1bar <- mean(x1)
x1bar
x2bar <- mean(x2)
x2bar

#assign standard deviations
s1 <- sqrt(var(x1))
s1
s2 <- sqrt(var(x2))
s2
```

```
#Pooled Sample Variance
#Both sample variances are pooled and adjusted for differences in sample size).
sp <- ((n1-1)*s1^2+(n2-1)*s2^2)/(n1+n2-2)
sp
[1] 0.1953408

#####Test Statistic#####
t <- (x1bar-x2bar)/sqrt(sp*(1/n1+1/n2))
t
[1] -10.52099

#Critical Values of the Test (probability of type 1 error) Note: two-tailed CVs are mirror values because
of normal distribution
alpha <- 0.05

c1 <- qt(alpha/2, n1+n2-2) #this is the two-sided lower critical value
c1
[1] -1.984467
c2 <- qt(1-alpha/2, n1+n2-2)#this is the two-sided higher critical value
c2
[1] 1.984467
abs_c <- abs(c1) #If using two-sided test, use absolute value of c1.

c3 <- qt(alpha, n1+n2-2) #one sided case lower CV
c3

c4 <- qt(1-alpha, n1+n2-2) #one sided case upper CV
c4

#Decision Rules:
#If abs(t) is > abs_c, then reject Null, otherwise accept Null
#If t < c3, then reject Null ** one sided case lower tail
#If t > c4, then reject Null ** one sided case upper tail

#Probability (P) Value

p1 <- 2*pt(t, n1+n2-2) #two sided case, if t < or equal to 0
p1
[1] 8.985235e-18

p2 <- 2*(1-pt(t, n1+n2-2)) #two sided case, if t > 0
p2

p3 <- pt(t, n1+n2-2) #one sided case lower tail
p3

p4 <- 1-pt(t, n1+n2-2) #one sided case upper tail
```

p4

#Confidence Intervals for the Difference in Mean:

```
#two sided case = (ci_a, ci_b)
ci_a <- x1bar-x2bar-abs_c*sqrt(sp*(1/n1+1/n2))
ci_a
```

[1] -1.105417

```
ci_b <- x1bar-x2bar+abs_c*sqrt(sp*(1/n1+1/n2))
ci_b
```

[1] -0.7545835

```
ci_l <- x1bar-x2bar-c3*sqrt(sp*(1/n1+1/n2)) #one sided lower tail
ci_l

ci_u <- x1bar-x2bar-c4*sqrt(sp*(1/n1+1/n2)) #one sided upper tail
ci_u

# Now, the built in R function
t.test(x1,x2,alternative = "two.sided",var.equal = TRUE, conf.level = 0.95)
```

Results of Hypothesis Test

Null Hypothesis: **difference in means = 0**

Alternative Hypothesis: **True difference in means is not equal to 0**

Test Name: **Two Sample t-test**

Estimated Parameter(s): **mean of x = 5.006**

mean of y = 5.936

Data: **x1 and x2**

Test Statistic: **t = -10.52099**

Test Statistic Parameter: **df = 98**

P-value: **8.985235e-18**

95% Confidence Interval: **LCL = -1.1054165**

UCL = -0.7545835