#Prototype to run a paired t-test
#Matthew Lundquist
#1/30/2016

#This a test is to compare related (paired) samples

Hypotheses:
H_0 – mean(X_1) = mean(X_2)
H_1 – mean(X_1) ≠ mean(X_2)

#Read in the data
iris <- read.table("iris.txt")

#I want to compare Sepal.Length between "setosa" and "versicolor"

#Subset into two equal-length vectors

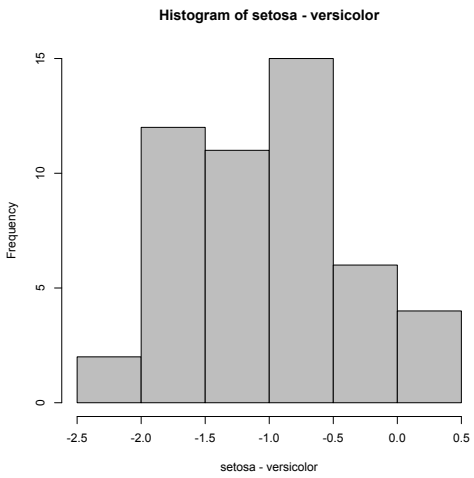
setosa <- iris$Sepal.Length[iris$Species == "setosa"]
versicolor <- iris$Sepal.Length[iris$Species == "versicolor"]

#A paired t-test is a parametric test. Must check the assumptions:

#Data vectors are the same length
length(setosa)
[1] 50

length(versicolor)
[1] 50

#Differences between pairs are normally distributed
hist(setosa - versicolor, col="grey")




```
#Run t-test manually

#Get the data together
x1 <- setosa
x2 <- versicolor

#sample size
n <- length(setosa)

#means and standard deviations
x1bar <- mean(setosa)
x2bar <- mean(versicolor)
s1 <- sd(setosa)
s2 <- sd(versicolor)


#Need dbar for paired test
d <- x1-x2
dbar <- mean(d)
s_d <- sqrt(var(d))

#Calculate test statistic
t <- dbar/(s_d/sqrt(n))
t
[1] -10.1459

#Probability value (two sided case)
degf <- n-1
P <- 2 * pt(t, degf)
```

P
[1] 1.241915e-13

#Confidence intervals
alpha <- 0.05

#T distribution
c1 <- qt(alpha/2,degf)
c1 <- -c1

#Confidence Intervals
ci_l <- dbar-c1*(s_d/sqrt(n))
ci_u <- dbar+c1*(s_d/sqrt(n))

CI <- c(ci_l, ci_u)
CI
[1] -1.114203 -0.745797

#Run the test using the built in function
t.test(setosa,versicolor,paired=TRUE, conf.level=0.95)

Paired t-test
data: setosa and versicolor
t = -10.146, df = 49, p-value = 1.242e-13
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.114203 -0.745797
sample estimates:
mean of the differences
                -0.93
P < 0.05, reject H_0