Justin Mann, Biostatistics with R, 2016

Paired t-test prototype

#Paired t-test prototype

#Justin Mann

#2/16/2016

#A paired t-test compares two sets of measurements,

#which are exactly matched for each individual of a population

**Hypotheses**:

H0: mean(x1) = mean(x2)

H1: mean(x1) ≠ mean(x2)

#Read in data: I'm using "iris," a data set built into r.

iris

#The following test will compare Sepal.Length between "versicolor" and "virginica"

#Assign subsets of two equal-length vectors

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

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

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

**Assumptions**:

#Observed values X1,1, X1,2, X1,3, ... X1,n are a random sample exactly matched with

#Observed values X2,1, X2,2, X2,3, ... X2,n across individuals 1,2,3, ... ,n.

#Let di= X2,i -X1,i for each individual i are a random sample from ~N(mu,Sigma).

#Variance of the population is unknown.

#Note: paired t-test is reasonably robust for deviations from normal distribution.

#verify that data vectors are the same length

length(versicolor)

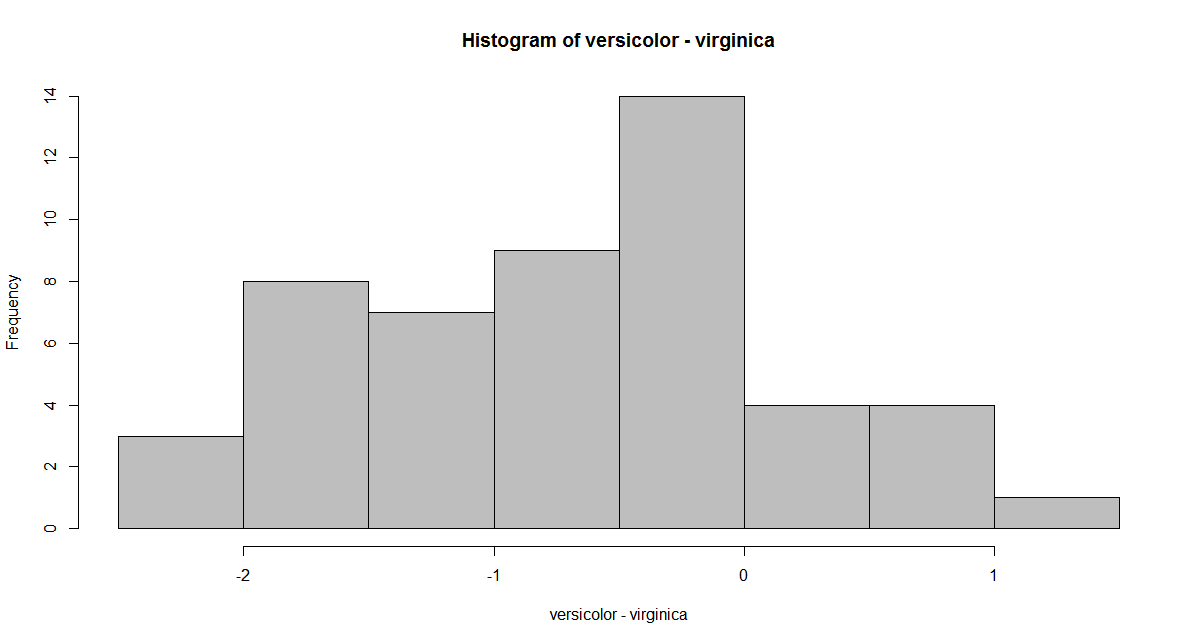
[1] 50

length(virginica)

[1] 50

#Visual confirmation that differences between pairs are normally distributed

hist(versicolor - virginica, col="grey")



#Below are the manual calculations for paired t-test

#Assign vectors versicolor and virginica to variables "x1" and "x2" respectively.

x1 <- versicolor

x2 <- virginica

#assign variable "n" to sample size

n <- length(versicolor)

#assign mean and standard deviation variables

x1bar <- mean(versicolor)

x2bar <- mean(virginica)

s1 <- sd(versicolor)

s2 <- sd(virginica)

#dbar is the mean difference between paired datum

d <- x1-x2

dbar <- mean(d)

s\_d <- sqrt(var(d))

#Manual calculatation of test statistic, "t"

t <- dbar/(s\_d/sqrt(n))

t

[1] -5.275345

#Probability (P) value (two sided case)

degf <- n-1

P <- 2 \* pt(t, degf)

P

[1] 2.989652e-06

#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] -0.900371 -0.403629

#Run the test using the built in function

t.test(versicolor,virginica,paired=TRUE, conf.level=0.95)

Output:

Paired t-test

data: versicolor and virginica

t = -5.2753, df = 49, p-value = 2.99e-06

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.900371 -0.403629

sample estimates:

mean of the differences

-0.652