Homework 1 – Roozbeh Dadashzadeh

# Q1 ----------------------------------------------------------------------

#Preprocessing The Data

library(faraway)

data(pima)

d<-pima

d$diastolic[d$diastolic==0]=NA

d$glucose[d$glucose==0]=NA

d$triceps[d$triceps==0]=NA

d$bmi[d$bmi==0]=NA

d$insulin[d$insulin==0]=NA

#Plotting

par(mfrow=c(1,2))

hist(d$age[which(!is.na(d$triceps))],

xlab = 'Observed triceps',

main = 'Age of Observed Triceps',

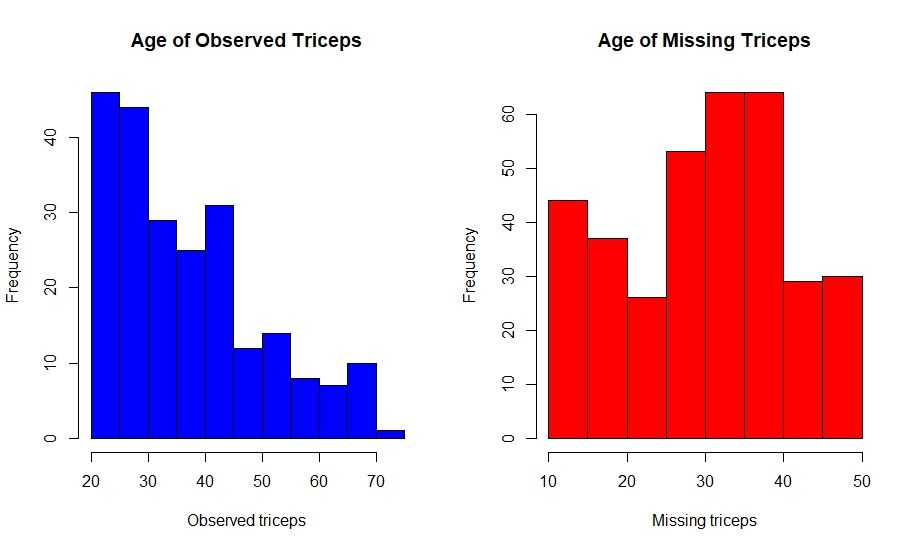
col = "blue",border="black")

hist(d$age[which(is.na(d$triceps))],

xlab = 'Missing triceps',

main = 'Age of Missing Triceps',

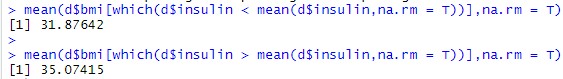
col = "red",border="black")



# Q2 ----------------------------------------------------------------------

#Calculating The Mean of BMI for Certain Values

mean(d$bmi[which(d$insulin < mean(d$insulin,na.rm = T))],na.rm = T)

mean(d$bmi[which(d$insulin > mean(d$insulin,na.rm = T))],na.rm = T)

# Q3 ----------------------------------------------------------------------

#Computing confidence intervals for insulin

Insulin = d$insulin[is.finite(d$insulin)]

Insulin\_mean = tapply(Insulin,d$test[which(is.finite(d$insulin))],mean)

Insulin\_sd = tapply(Insulin,d$test[which(is.finite(d$insulin))],sd)

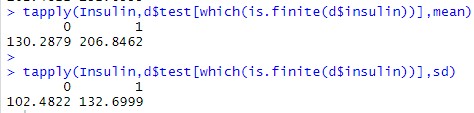
CI = function(x,y,z){a = x[seq(length(x))]-1.96\*y[seq(length(y))]/sqrt(z[seq(length(z))])

b = x[seq(length(x))]+1.96\*y[seq(length(y))]/sqrt(z[seq(length(z))])

c = c(a,b)

return(c)}

CI(Insulin\_mean,Insulin\_sd,

 c(length(which(d$test[is.finite(d$insulin)]==0)),length(which(d$test[is.finite(d$insulin)]==1))))

They don’t have overlapping confidence interval so there is a statistically significant difference between groups.