midterm2.R

benjamin

2020-12-01

# Midterm 2  
# Ben Reichert  
  
### 1  
  
#a  
pin <- 0.894  
q <- qnorm(0.975)  
nin <- 110  
cat("Mouthgaurd in: (",pin - q \* sqrt(pin\*(1-pin)/nin),",",pin + q \* sqrt(pin\*(1-pin)/nin),")\n")

## Mouthgaurd in: ( 0.8364728 , 0.9515272 )

#b  
pout <- 0.925  
q <- qnorm(0.975)  
nout <- 198  
cat("Mouthguard out: (",pout - q \* sqrt(pout\*(1-pout)/nout),",",pout + q \* sqrt(pout\*(1-pout)/nout),")")

## Mouthguard out: ( 0.8883126 , 0.9616874 )

#c  
# A p-value of 0.33 tells me that we fail to reject the null hypothesis that Curry's free throw ability  
# is the same mouthguard in as it is mouthguard out.  
  
  
### 2  
ren <- c(2.2,1.52,1.54,.77,.34,.45,.39,.29,.18,.16,.23,.24,.17,.08,.02,.02)  
hd <- c(1.84,.44,.3,.06,.2,.14,.1,.09,.06,.04)  
lren <- log(ren)  
lhd <- log(hd)  
  
#a  
# Null: the mean log percentage change in renal patients is not greater than that of heart patients.  
# Alternative: the mean log percentage change in renal patients is higher than in heart disease patients.  
  
#b  
print(t.test(lren,lhd,alt="greater"))

##   
## Welch Two Sample t-test  
##   
## data: lren and lhd  
## t = 1.0602, df = 21.693, p-value = 0.1504  
## alternative hypothesis: true difference in means is greater than 0  
## 95 percent confidence interval:  
## -0.326982 Inf  
## sample estimates:  
## mean of x mean of y   
## -1.327820 -1.854697

# T = 1.0602  
# Df = 21.693  
Delta.hat = mean(lren) - mean(lhd)  
std.error = sqrt(var(lren)/length(lren) + var(lhd)/length(lhd))  
Tw = Delta.hat / std.error  
df = (var(lren)/length(lren)+var(lhd)/length(lhd))^2 / ((var(lren)/length(lren))^2/(length(lren)-1) + (var(lhd)/length(lhd))^2/(length(lhd)-1))  
cat("T =",Tw,"\nDf =",df,"\n")

## T = 1.06023   
## Df = 21.69253

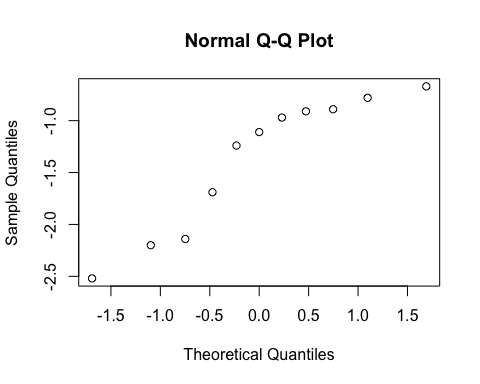
#c  
p <- 1 - pt(Tw, df=df)  
cat("P-value =",p,"\n")

## P-value = 0.1503513

# Since the P-value > the significance level we fail to reject the null hypothesis that the mean  
# log percentage change in renal patients is not greater than that of heart patients.  
  
#d  
lower <- Delta.hat - qt(0.975, df=df) \* std.error  
upper <- Delta.hat + qt(0.975, df=df) \* std.error  
cat("95% conf interval: (",lower,",",upper,")\n")

## 95% conf interval: ( -0.5045728 , 1.558326 )

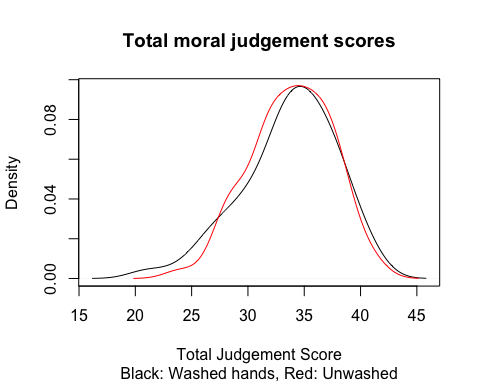
# This confidence interval tells me that we are 95% confident that distributions of red blood  
# cell count changes in renal anemia patients will be between -0.5 and 1.6 higher.  
  
  
### 3  
old <- c(.89,.49,.91,.8,.56,.79,.47,.5,1.08,1.65,1.94)  
yung <- c(2.13,1.16,2.6,1.58,1.53,1.7,2.67,2.64,2.19,2.54,4.46)  
dhat <- mean(yung)-mean(old)  
  
#a  
# We should test this data using a one-sample t-test because there is only one independent sample here,  
# which is the difference between the old and young roots' water absorption.  
  
#b  
qqnorm(old-yung)



# a) We can see from the qqnorm plot that old-young does not follow a normal distribution,  
# thus we should be hesitant to run a t-test on it.  
# b) You cannot take the log on negative data, and old-young would give negative.  
  
#c  
ratio <- old/yung  
low <- mean(ratio) - qt(.975, df=length(ratio)-1) \* sd(ratio) / sqrt(length(ratio))  
hi <- mean(ratio) + qt(.975, df=length(ratio)-1) \* sd(ratio) / sqrt(length(ratio))  
cat("95% conf interval: (",low,",",hi,")\n\n")

## 95% conf interval: ( 0.3147243 , 0.4980868 )

# This means that we are 95% confident that that ratio of young/old is between 0.31 and 0.50.  
  
  
### 4  
wash <- read.table("handwashing.txt",header=TRUE)  
  
#a  
# All 132 participants are the experimental units.  
# On each unit the measurements taken are the moral judgement scores from 1-7.  
# There are two independent samples, the group of 63 students who had their hands washed  
# and the group of 69 students who did not.   
  
#b  
Total <- rep(0,132)  
Twash <- rep(0,63)  
Tdirt <- rep(0,69)  
tw <- 1  
td <- 1  
for (i in 1:132) {  
 if (wash[i,1]) {  
 Total[i] <- sum(wash[i,])-1  
 Twash[tw] <- sum(wash[i,])-1  
 tw = tw+1  
 } else {  
 Total[i] <- sum(wash[i,])  
 Tdirt[td] <- sum(wash[i,])  
 td = td+1  
 }  
}  
plot(density(Twash),main="Total moral judgement scores",sub="Black: Washed hands, Red: Unwashed",  
 xlab="Total Judgement Score")  
lines(density(Tdirt),col="Red")



aveTW <- ave(Twash)  
aveTD <- ave(Tdirt)  
cat("Average total moral judgement score of washed hands:",aveTW[1],"\n")

## Average total moral judgement score of washed hands: 33.66667

cat("Average total moral judgement score of unwashed hands:",aveTD[1],"\n")

## Average total moral judgement score of unwashed hands: 33.84058

#c  
# Null: Handwashing does not decrease the total morality judgement score.  
# Alternative: Handwashing decreases the total morality judgement score.  
# Welch's T-test (one-tailed)  
D.hat = mean(Tdirt) - mean(Twash)  
s.error = sqrt(var(Twash)/length(Twash) + var(Tdirt)/length(Tdirt))  
TW = D.hat / s.error  
DF = (var(Twash)/length(Twash)+var(Tdirt)/length(Tdirt))^2 / ((var(Twash)/length(Twash))^2/(length(Twash)-1) + (var(Tdirt)/length(Tdirt))^2/(length(Tdirt)-1))  
p4 <- (1 - pt(Tw, df=df))/2  
cat("Test Statistic:",TW,  
 "\nDegrees of freedom:",DF,  
 "\nP-value:",p4,"\n")

## Test Statistic: 0.2539921   
## Degrees of freedom: 122.7299   
## P-value: 0.07517565

# Given that the p-value > our significance value of 0.05, we fail to reject the null hypothesis that  
# washing hands results in a lower total morality score. To put it into plain english, we failed to be  
# statistically confident that washing hands lowers morality score.   
  
#d  
low <- mean(Twash) - qt(.975, df=length(Twash)-1) \* sd(Twash) / sqrt(length(Twash))  
hi <- mean(Twash) + qt(.975, df=length(Twash)-1) \* sd(Twash) / sqrt(length(Twash))  
cat("95% conf interval of Total for handwashing: (",low,",",hi,")\n")

## 95% conf interval of Total for handwashing: ( 32.60779 , 34.72554 )

low <- mean(Tdirt) - qt(.975, df=length(Tdirt)-1) \* sd(Tdirt) / sqrt(length(Tdirt))  
hi <- mean(Tdirt) + qt(.975, df=length(Tdirt)-1) \* sd(Tdirt) / sqrt(length(Tdirt))  
cat("95% conf interval of Total for control: (",low,",",hi,")\n")

## 95% conf interval of Total for control: ( 32.9748 , 34.70636 )

Delta.hat = mean(Tdirt) - mean(Twash)  
std.error = sqrt(var(Tdirt)/length(Tdirt) + var(Twash)/length(Twash))  
Tw = Delta.hat / std.error  
df = (var(Tdirt)/length(Tdirt)+var(Twash)/length(Twash))^2 / ((var(Tdirt)/length(Tdirt))^2/(length(Tdirt)-1) + (var(Twash)/length(Twash))^2/(length(Twash)-1))  
lower <- Delta.hat - qt(0.975, df=df) \* std.error  
upper <- Delta.hat + qt(0.975, df=df) \* std.error  
cat("95% conf interval of diff in mean between two groups: (",lower,",",upper,")\n\n")

## 95% conf interval of diff in mean between two groups: ( -1.181474 , 1.529301 )

### 5  
don <- read.table("trump.txt",header=TRUE)  
trump.white <- subset(don, name.type == "white")  
trump.nonwhite <- subset(don, name.type == "nonwhite")  
  
#a  
nwvd <- trump.nonwhite[,"vote.diff"]  
nwvdm <- mean(nwvd)  
nwvdsd <- sd(nwvd)  
nwvdn <- length(nwvd)  
t <- nwvdm / (nwvdsd/sqrt(nwvdn))  
p <- 2 \* pt(t, df = nwvdn-1)  
cat("One-sample t-test of the hypothesis that μNW is 0 p-value:",p,"which is less than the significance value of 0.05",  
 "\nthus we reject the null hypothesis that μNW is zero and accept the alternative hypothesis that the mean vote",  
 "\ndifferential for a hypothetical population of Trump delegates with nonwhite names is not 0, meaning the delegates",  
 "\nwith non-white names did not succeed in getting 1/3 of all votes for Trump. Ultimately, we conclude that the",  
 "\n'color' of names negatively impacted Trumps votes.\n")

## One-sample t-test of the hypothesis that μNW is 0 p-value: 0.004213093 which is less than the significance value of 0.05   
## thus we reject the null hypothesis that μNW is zero and accept the alternative hypothesis that the mean vote   
## differential for a hypothetical population of Trump delegates with nonwhite names is not 0, meaning the delegates   
## with non-white names did not succeed in getting 1/3 of all votes for Trump. Ultimately, we conclude that the   
## 'color' of names negatively impacted Trumps votes.

#b  
wvd <- trump.white[,"vote.diff"]  
D.hat = mean(wvd) - mean(nwvd)  
s.error = sqrt(var(wvd)/length(wvd) + var(nwvd)/length(nwvd))  
TW = D.hat / s.error  
DF = (var(wvd)/length(wvd)+var(nwvd)/length(nwvd))^2 / ((var(wvd)/length(wvd))^2/(length(wvd)-1) + (var(nwvd)/length(nwvd))^2/(length(nwvd)-1))  
p <- (1 - pt(Tw, df=df))  
cat("Two-sample t-test of the hypothesis μW is the same as μNW p-value:",p,"which is greater than the significance",  
 "\nvalue of 0.05, thus we fail to reject the null hypothesis that the vote differential of Trump delegates",  
 "\nthat are white is the same as the vote differential of Trump delegates that are nonwhite. We therefore",  
 "\nconclude that μW is different from μNW, and ultimately that Trump's vote differential is contingent on",  
 "\nthe 'color' of names, and that the white named delegates were more successful at getting Trump votes.")

## Two-sample t-test of the hypothesis μW is the same as μNW p-value: 0.3999633 which is greater than the significance   
## value of 0.05, thus we fail to reject the null hypothesis that the vote differential of Trump delegates   
## that are white is the same as the vote differential of Trump delegates that are nonwhite. We therefore   
## conclude that μW is different from μNW, and ultimately that Trump's vote differential is contingent on   
## the 'color' of names, and that the white named delegates were more successful at getting Trump votes.

#c  
# I think the two-sample t-test is closer to having its assumptions met because we have two independent sample population  
# sets so doing a one-sample t-test would be leaving out data that could potentially influence the outcome. For this reason,  
# because the two-sample t-test includes all available data, I think it is closer to being accurate across this data.