Sergiu Buciumas

Homework 2: Due Tuesday, June 21st by 10:00 PM

The first problem was written by Chris Williams of the University of Idaho. After that, I have chosen 5 book problems. **This first problem MUST be done in R.** The five book problems may be done using either SAS or R. Enjoy!

1. A small random sample was taken of single-family house prices in Moscow, ID and Seattle, WA.  The prices are listed below (in thousands of dollars):

Moscow: 25.0, 238.9, 20.5, 429.9, 279., 138., 345.

Seattle: 789., 250., 579.5, 489.5, 749.9, 2395., 295., 1975.

a. Perform a t-test and a Wilcoxon test to see if prices differ between the two cities.

#Ex1Hw2

library(perm)

moscow <- c(25.0, 238.9, 20.5, 429.9, 279., 138., 345)

seattle <- c(789., 250., 579.5, 489.5, 749.9, 2395., 295., 1975)

#a Perform a t-test and a Wilcoxon test to see if prices differ between the two cities

#t-test

t.test(moscow, seattle)



#Wilcoxon test

wilcox.test(moscow,seattle,exact=FALSE)



b. Perform an omnibus test of a difference in distributions between prices in the two cities.

#Perform an omnibus test of a difference in distributions between prices in the two cities.

ks.test(moscow,seattle)



c. Calculate the Hodges-Lehmann estimate of the shift parameter delta, and calculate a 95% confidence interval for delta. Verify the ka and kb values for the normal approximation by hand calculation.*[ Hint from BJB: see sections 2.10.2 and 2.10.5.]*

pwd <- outer(seattle,moscow,"-")

pwd

sortedc <- sort(c)

sorted

using A4

lower 13, upper 43

Ka = l 0.05 + 1 = 13 +1 = 14

Kb = u0.05= 43

Pwd(14) < delta <= 43

210.5 <delta <= 1545.1



d. Test for a difference in scale between the two house price distributions. Justify your choice of a scale test.

t.test(moscow, seattle, mu = 0.90)



we do not have evidence to reject null Hypothesis

ks.test(moscow, seattle)



e. Calculate the number of elements in the permutation distribution for the differences between city house prices.

RMD.test <- function(samp1,samp2,direction=c('two.sided','less','greater')[1],nsamp=10000){

devs1 <- samp1-median(samp1)

devs2 <- samp2-median(samp2)

devs <- c(devs1,devs2)

RMD <- mean(abs(devs1))/mean(abs(devs2))

if (direction[1]=='two.sided'){

RMD <- max(1/RMD, RMD)

}

RMDperms <- rep(NA,nsamp)

for (i in 1:nsamp){

tempdevs <- devs[sample(length(devs),length(devs),replace=FALSE)]

RMDperms[i] <- mean(abs(tempdevs[1:length(devs1)]))/mean(abs(tempdevs[-(1:length(devs1))]))

if (direction[1]=='two.sided') RMDperms[i] <- max(1/RMDperms[i], RMDperms[i])

}

if (direction[1]=='greater') pVal <- mean(RMDperms>=RMD)

if (direction[1]=='less') pVal <- mean(RMDperms<=RMD)

if (direction[1]=='two.sided') pVal <- mean(RMDperms>=RMD)

print(paste("Test statistic:",round(RMD,4)))

print(paste("Approximate p-value for ",direction[1],": ",pVal,sep=""))

}

RMD.test(moscow, seattle,nsamp=100000)

permTS(moscow, seattle,method="exact.mc",control=permControl(nmc=100000,seed=987971))

wilcox.test(moscow, seattle)

t.test(moscow, seattle)

ks.test(moscow, seattle)







f. Give a brief summary of your results for parts a-d to describe if any differences were observed in these tests.  If any differences were detected, what do your results tell you about the house prices in the two cities?

The additional exercises:

Chapter 2, Exercises 10, 15,

Chapter 2, Exercises 10 need help

**Chapter 2 Ex 15 Test for differences between the scale parameters using the Siegel-Tukey test.**

#prob Ch2, 15

install.packages("DescTools")

library(DescTools)

x <- c(21.9, 20.2, 19.4, 20.3, 19.6, 20.4, 18.4, 20.1, 22.0, 18.9)

length(x)

y <- c(20.2, 13.8, 21.8, 19.2, 19.6, 25.5, 17.0, 17.6, 19.5, 22.2)

length(y)

combinexy <- c(x, y)

combinexy

orderedxy <- order(combinexy)

orderedxy

sortedxy <- sort(combinexy)

sortedxy

SiegelTukeyTest(x, y)



**b. Test for differences between the scale parameters using the approximate RMD permutation test.**

#b RMD permutation test completed

RMD.test <- function(samp1,samp2,direction=c('two.sided','less','greater')[1],nsamp=10000){

devs1 <- samp1-median(samp1)

devs2 <- samp2-median(samp2)

devs <- c(devs1,devs2)

RMD <- mean(abs(devs1))/mean(abs(devs2))

if (direction[1]=='two.sided'){

RMD <- max(1/RMD, RMD)

}

RMDperms <- rep(NA,nsamp)

for (i in 1:nsamp){

tempdevs <- devs[sample(length(devs),length(devs),replace=FALSE)]

RMDperms[i] <- mean(abs(tempdevs[1:length(devs1)]))/mean(abs(tempdevs[-(1:length(devs1))]))

if (direction[1]=='two.sided') RMDperms[i] <- max(1/RMDperms[i], RMDperms[i])

}

if (direction[1]=='greater') pVal <- mean(RMDperms>=RMD)

if (direction[1]=='less') pVal <- mean(RMDperms<=RMD)

if (direction[1]=='two.sided') pVal <- mean(RMDperms>=RMD)

print(paste("Test statistic:",round(RMD,4)))

print(paste("Approximate p-value for ",direction[1],": ",pVal,sep=""))

}

#Test for differences between the scale parameters using the approximate RMD permutation test

RMD.test(x, y, direction="greater")



Chapter 3, Exercises 1, 6, 11. We will discuss exercise 6 some in class, as it is quite involved, and the data for exercise 6 are available through D2L.

Chapter3 Ex 1 The data are samples from three simulated distributions.

1. Apply the permutation F-test to the data.
2. Compare the results in part a with the results of the usual one-way analysis of variance.

#b.anova build

fit <- aov(c1 ~ factor(nrep))

summary(fit)

Chapter3 Ex 6

Chaper3 Ex 11

1. Population means of all groups 2-5 are the same, but different from group one. Because 2-10 have the same means not likely that a difference will be found among groups. Shortcomings of using LSD is that there are too many groups. Works best with 3 groups.
2. Bonferroni is more conservative than Tukey's HSD which is more conservative than the LSD. It would take more evidence to reject null for Bonferroni, so not likely differences would be found. Tukey does not take as much to reject null.