DiFrank, Annie

Course#: ST537

HW#:2

Submission Date: 01/24/25

DiFrank\_HW2

2025-01-22

## Problem 1

Assumptions: The data is drawn from a multivariate normal distribution. Observations are independent and drawn randomly. The s is a good estimate of population covariance matrix.

# Given data  
x\_bar <- c(5.936, 2.770, 4.260, 1.326)  
s <- matrix(c(0.40, 0.09, 0.30, 0.05,  
 0.09, 0.10, 0.07, 0.05,  
 0.30, 0.07, 0.30, 0.05,  
 0.05, 0.05, 0.05, 0.08), nrow = 4, byrow = TRUE)  
n <- 50  
  
#Compute variance of (x\_bar1 - x\_bar2)  
var\_diff <- s[1, 1] + s[2, 2] - 2 \* s[1, 2]  
  
#Compute T^2  
T\_squared <- n \* (x\_bar[1] - x\_bar[2])^2 / var\_diff  
  
#Compute critical value  
alpha <- 0.05  
critical\_value <- qf(1 - alpha, df1 = 1, df2 = n - 1)  
  
#Output results  
list(T\_squared = T\_squared, Critical\_Value = critical\_value, Reject\_H0 = T\_squared > critical\_value)

## $T\_squared  
## [1] 1566.181  
##   
## $Critical\_Value  
## [1] 4.038393  
##   
## $Reject\_H0  
## [1] TRUE

At alpha = 0.05, we have statistically significant evidence to suggest we reject the null hypothesis that u1 =u2.

## Problem 2

**Part (a) We test if there is a difference among the mean reaction times across the four treatments.**

#### Hypotheses:

* Null hypothesis (Ho​): u1=u2=u3=u4, i.e. no overall treatment effect.
* Alternative hypothesis (Ha): At least one u1​ differs; there is an overall treatment effect.

We assume the data is drawn from a p=variable normal distribution and observations are random and independent; there is no carry-over effect between treatments.

dat <- read.table("data/T6-8.DAT", header = FALSE)  
colnames(dat) <- c("trt1", "trt2", "trt3", "trt4")   
head(dat)

## trt1 trt2 trt3 trt4  
## 1 869 860.5 691.0 601  
## 2 995 875.0 678.0 659  
## 3 1056 930.5 833.0 826  
## 4 1126 954.0 888.0 728  
## 5 1044 909.0 865.0 839  
## 6 925 856.5 1059.5 797

Perfoming Hotelling’s T^2 hypothesis test:

library(ICSNP)

## Loading required package: mvtnorm

## Warning: package 'mvtnorm' was built under R version 4.3.3

## Loading required package: ICS

mu0=rep(0,ncol(dat))  
HotellingsT2(X=dat,mu=mu0)

##   
## Hotelling's one sample T2-test  
##   
## data: dat  
## T.2 = 341.2, df1 = 4, df2 = 28, p-value < 2.2e-16  
## alternative hypothesis: true location is not equal to c(0,0,0,0)

According to our test, we have significant evidence (p value <2.2e-16) to reject the null hypothesis that there is no overall treatment effect (the mean vector is not 0). At least one of the four measures is a non-zero mean difference before and after treatment.

**Part (b) Construct 95% simultaneous CIs for contrasts representing the treatments mean difference u1-u2, u1-u3, and u1-u3. Interpret the results.**

Function for simultaneous CIs:

simul.CI <- function(data, level = 0.95){  
 xbar <- colMeans(data)  
 S<-cov(data)  
 n<-nrow(data)  
 p<-ncol(data)  
 crit<- (n-1)\*p/((n-p)\*n) \* qf(1-level,p,n-p,lower.tail = FALSE)  
 H<-sqrt(crit\*diag(S))  
 out<-data.frame(Estimate = xbar, Lower = xbar-H, Upper =xbar+H)  
 return(out)  
}

Doing simultaneous CIs:

sint<- simul.CI(dat)  
round(sint,3)

## Estimate Lower Upper  
## trt1 967.562 850.991 1084.134  
## trt2 876.891 785.365 968.416  
## trt3 828.625 746.707 910.543  
## trt4 716.625 648.981 784.269

Because 0 is not included in the confidence intervals, there is strong evidence that the true mean treatment difference is not 0 at 95% confidence.

**Part (c)  Repeat (a) by using a contrast matrix that describes interaction effect, and each of the two main effects.**

We assume there is no carry over affect between the treatments. We assume the data follows a normal distribution and was randomly sampled.

Function for testing multiple contrasts:

T2.contrast<- function(data.matrix, contrast.matrix,alpha = 0.05){  
 Xmat<- data.matrix  
 C<- contrast.matrix  
 n<- nrow(Xmat)  
 q<- nrow(C)  
 xbar<- colMeans(Xmat)  
 S<- cov(Xmat)  
 #intermediate quantities  
 invCSC<- solve(C%\*%S %\*% (t(C)))  
 Cxbar<- C %\*% xbar  
 #test statistic  
 T2<- n\*(n-q)/((n-1)\*q) \* (t(Cxbar)) %\*% invCSC %\*% (Cxbar)  
 #critical value  
 critical\_F = qf(alpha, df1=q,df2=n-q,lower.tail = FALSE)  
 #p value  
 pv<- pf(T2, df1=q, df2=n-q,lower.tail = FALSE)  
 #display results  
 results <- data.frame(T2=T2, Fcritical=critical\_F,df1=q,df2=n-q,pvalue=pv)  
 return(results)  
}

Contrast matrix and performing the test:

C<- cbind(c(1,1,1),-diag(1,3))  
C

## [,1] [,2] [,3] [,4]  
## [1,] 1 -1 0 0  
## [2,] 1 0 -1 0  
## [3,] 1 0 0 -1

#test   
T2.contrast(dat, C)

## T2 Fcritical df1 df2 pvalue  
## 1 42.36316 2.93403 3 29 1.008102e-10

There is strong evidence at alpha = 0.05 that the four treatment means are different (i.e. strong evidence to reject the null hypothesis that the 4 means are equal).

## Problem 3

#reading in data   
dat3<- read.table("data/T1-8.DAT", header = FALSE)  
colnames(dat3) <- c("DominantRadius", "Radius", "DominantHumerus" ,"Humerous","DominantUlna","Ulna")  
head(dat3)

## DominantRadius Radius DominantHumerus Humerous DominantUlna Ulna  
## 1 1.103 1.052 2.139 2.238 0.873 0.872  
## 2 0.842 0.859 1.873 1.741 0.590 0.744  
## 3 0.925 0.873 1.887 1.809 0.767 0.713  
## 4 0.857 0.744 1.739 1.547 0.706 0.674  
## 5 0.795 0.809 1.734 1.715 0.549 0.654  
## 6 0.787 0.779 1.509 1.474 0.782 0.571

**Part (a) For each bone, define the difference response between the dominant and non-dominant side. If the true mean difference is zero, it means that the mean mineral content is the same for dominant and non-dominant bones. We are interesting in formally testing whether the mean difference is different than zero.**

μD​=[μD1​​,μD2​​,μD3​​] is the vector of mean differences (Di for the three bones)

Null hypothesis: Ho: μD = 0 (the mean differences are all zero across bones)

Alternative hypothesis: Ha: μD /= 0 (at least one mean difference is significantly different from zero)

Test name: Hotelling’s T2 Test, where the sample mean vector is Dˉ, the sample covariance matrix S of the differences.

**Part (b) Performance of the test**

mu0=rep(0,ncol(dat3))  
HotellingsT2(X=dat3,mu=mu0)

##   
## Hotelling's one sample T2-test  
##   
## data: dat3  
## T.2 = 217.05, df1 = 6, df2 = 19, p-value = 2.22e-16  
## alternative hypothesis: true location is not equal to c(0,0,0,0,0,0)

At alpha=0.05, there is statistically significant evidence to reject the null (i.e. the true mean difference is not equal to 0).

**Part (c) Bonferroni 95% intervals for each component of the difference of the two mean vectors**

#differences  
diff\_rad <- dat3$DominantRadius - dat3$Radius  
diff\_hum <- dat3$DominantHumerus - dat3$Humerous  
diff\_ulna <- dat3$DominantUlna - dat3$Ulna  
  
#mean differences  
mean\_diff\_rad <- mean(diff\_rad)  
mean\_diff\_hum <- mean(diff\_hum)  
mean\_diff\_ulna <- mean(diff\_ulna)  
  
#standard errors  
se\_rad <- sd(diff\_rad) / sqrt(length(diff\_rad))  
se\_hum <- sd(diff\_hum) / sqrt(length(diff\_hum))  
se\_ulna <- sd(diff\_ulna) / sqrt(length(diff\_ulna))  
  
#bonferroni adjustment  
alpha = 0.05  
k <- 3 #number of tests  
adjusted\_alpha <- alpha / k  
t\_critical <- qt(1 - adjusted\_alpha / 2, df = length(diff\_rad) - 1)  
  
#CIs   
ci\_rad <- c(  
 mean\_diff\_rad - t\_critical \* se\_rad,  
 mean\_diff\_rad + t\_critical \* se\_rad  
)  
  
ci\_hum <- c(  
 mean\_diff\_hum - t\_critical \* se\_hum,  
 mean\_diff\_hum + t\_critical \* se\_hum  
)  
  
ci\_ulna <- c(  
 mean\_diff\_ulna - t\_critical \* se\_ulna,  
 mean\_diff\_ulna + t\_critical \* se\_ulna  
)  
  
#results   
cat("Bonferroni 95% Confidence Intervals:\n")

## Bonferroni 95% Confidence Intervals:

cat("Radius:", ci\_rad, "\n")

## Radius: -0.005671023 0.05663102

cat("Humerus:", ci\_hum, "\n")

## Humerus: -0.007854538 0.1235345

cat("Ulna:", ci\_ulna, "\n")

## Ulna: -0.0293945 0.0505145

For the radius, ulna, and humerus, the intervals include zero, indicating there is not statistically significant evidence to reject a null of zero difference between dominant and non-dominant mineral content.

## Problem 4

#reading in data   
dat4<- read.table(file="data/middleschool.txt", sep=",",header=TRUE)  
colnames(dat4)

## [1] "ID" "Math" "ELA" "Science"   
## [5] "SocialStudies" "school"

head(dat4)

## ID Math ELA Science SocialStudies school  
## 1 1 78.80 81.26 77.81 91.92 A  
## 2 2 96.64 90.50 87.93 100.00 A  
## 3 3 79.43 79.09 84.19 87.05 A  
## 4 4 85.19 80.48 85.48 89.87 A  
## 5 5 73.96 81.50 76.13 91.08 A  
## 6 6 79.42 77.55 78.05 88.53 A

**Part (a) Construct Bonferroni intervals for the mean performance in each of the four subjects, MATH, ELA, Science and Social studies, for school A.**

#Filter data for School A and School B  
school\_A <- dat4[dat4$school == "A", ]  
school\_B <- dat4[dat4$school == "B", ]  
  
#Function to calculate Bonferroni intervals  
bonferroni\_intervals <- function(data, subjects, alpha = 0.05) {  
 k <- length(subjects) # Number of comparisons  
 adjusted\_alpha <- alpha / k  
 t\_critical <- qt(1 - adjusted\_alpha / 2, df = nrow(data) - 1)  
   
 results <- data.frame(  
 Subject = subjects,  
 Mean = sapply(subjects, function(subj) mean(data[[subj]], na.rm = TRUE)),  
 Lower = NA,  
 Upper = NA  
 )  
   
 results$SE <- sapply(subjects, function(subj) {  
 sd(data[[subj]], na.rm = TRUE) / sqrt(nrow(data))  
 })  
   
 results$Lower <- results$Mean - t\_critical \* results$SE  
 results$Upper <- results$Mean + t\_critical \* results$SE  
 return(results)  
}  
  
# Subjects to analyze  
subjects <- c("Math", "ELA", "Science", "SocialStudies")  
  
# Bonferroni intervals for School A  
bonferroni\_A <- bonferroni\_intervals(school\_A, subjects)  
bonferroni\_A

## Subject Mean Lower Upper SE  
## Math Math 79.2064 77.20027 81.21253 0.7885224  
## ELA ELA 79.7858 78.59327 80.97833 0.4687337  
## Science Science 79.5381 78.02665 81.04955 0.5940867  
## SocialStudies SocialStudies 89.7189 88.49729 90.94051 0.4801630

**Part (b) Repeat (a) for school B.**

bonferroni\_B <- bonferroni\_intervals(school\_B, subjects)  
bonferroni\_B

## Subject Mean Lower Upper SE  
## Math Math 80.44173 78.99543 81.88803 0.5720234  
## ELA ELA 80.15713 79.36491 80.94935 0.3133300  
## Science Science 80.17173 79.10481 81.23866 0.4219785  
## SocialStudies SocialStudies 90.21747 89.40442 91.03051 0.3215648

**Part (c) Use (a) and (b) and comment on whether any school has better performance in any subject.**

For Math, ELA, Science, and Social Studies, the confidence intervals between each school overlap, so I wouldn’t make any strong conclusions on the differences between schools. Even though the mean is lower in each subject for school A, we are using 95% confidence, and our 95% confidence interval for each of these subjects in A contains the mean for school B.