STAT 488: Multivariate Statistical Analysis — Homework 3

Charles Hwang

3/23/2022

Problem 5.1

```
rm(list=ls())
X<-matrix(c(2,8,6,8,12,9,9,10),nrow=4,ncol=2)                              # Problem 5.1(a)
x<-matrix(c(mean(X[,1]),mean(X[,2])))
n<-dim(X)[1]
s11<-((X[1,1]-x[1])^2+(X[2,1]-x[1])^2+(X[3,1]-x[1])^2+(X[4,1]-x[1])^2)/(n-1)
s12<-s21<-((X[1,1]-x[1])*(X[1,2]-x[2])+(X[2,1]-x[1])*(X[2,2]-x[2])+(X[3,1]-x[1])*(X[3,2]-x[2])+(X[4,1]-
s22<-((X[1,2]-x[2])^2+(X[2,2]-x[2])^2+(X[3,2]-x[2])^2+(X[4,2]-x[2])^2)/(n-1)
n*t(x-matrix(c(7,11)))%*%solve(matrix(c(s11,s12,s21,s22),nrow=2))%*%(x-matrix(c(7,11)))
```

```
##          [,1]
## [1,] 13.63636
```

```
(n-1)*dim(X)[2]/abs(diff(dim(X)))                                          # Problem 5.1(b)
```

```
## [1] 3
```

```
dim(X)[2]
```

```
## [1] 2
```

```
abs(diff(dim(X)))
```

```
## [1] 2
```

```
# We can see the distribution of T^2 for Problem 5.1(a) is 3*F_(2,2).
(n-1)*dim(X)[2]/abs(diff(dim(X)))*qf(1-0.05,dim(X)[2],abs(diff(dim(X)))) # Problem 5.1(c)
```

```
## [1] 57
```

We fail to reject H_0 at the $\alpha = 0.05$ level. There is insufficient evidence $(13.6363636 < 57)$ that $\mu \neq \begin{bmatrix} 7 \\ 11 \end{bmatrix}$.

Problem 5.2

```
rm(list=ls())
X<-matrix(c(6-9,10-6,8-3,6+9,10+6,8+3),nrow=3)
x<-matrix(c(mean(X[,1]),mean(X[,2])))
n<-dim(X)[1]
s11<-((X[1,1]-x[1])^2+(X[2,1]-x[1])^2+(X[3,1]-x[1])^2)/(n-1)
s12<-s21<-((X[1,1]-x[1])*(X[1,2]-x[2])+(X[2,1]-x[1])*(X[2,2]-x[2])+(X[3,1]-x[1])*(X[3,2]-x[2]))/(n-1)
s22<-((X[1,2]-x[2])^2+(X[2,2]-x[2])^2+(X[3,2]-x[2])^2)/(n-1)
n*t(x-matrix(c(9,5)))%*%solve(matrix(c(s11,s12,s21,s22),nrow=2))%*%(x-matrix(c(9,5)))
```

```
##          [,1]
```

```
## [1,] 34.77778
```

Problem 5.18

```r
rm(list=ls())
ct<-read.table("/Users/newuser/Desktop/Notes/Graduate/STAT 488 - Multivariate Statistical Analysis/T5-2
x<-matrix(c(mean(ct$V1),mean(ct$V2),mean(ct$V3)))                         # Problem 5.18(a)
s<-matrix(c(var(ct$V1),cov(ct$V1,ct$V2),cov(ct$V1,ct$V3),cov(ct$V2,ct$V1),var(ct$V2),cov(ct$V2,ct$V3),c
n<-dim(ct)[1]
p<-dim(ct)[2]
n*t(x-matrix(c(500,50,30)))%*%solve(s)%*%(x-matrix(c(500,50,30)))
```

```
##          [,1]
## [1,] 223.3102
```

```r
(n-1)*p/(n-p)*qf(1-0.05,p,n-p)
```

```
## [1] 8.333483
```

```r
# We reject H0 at the alpha = 0.05 level. There is sufficient evidence (223.31 > 8.33)
# that mu' =/= [500, 50, 30]. There is strong reason to believe the scores in Table 5.2
# are different than the average scores for thousands of college students over the
# last ten years because of the large difference between T^2 and critical F-value.
# e_1 = 5878.791606449943, e_2 = 63.835101127801, e_3 = 14.598052422257  # Problem 5.18(b)
2*sqrt(5878.791606449942523*(n-1)*p/(n*(n-p))*qf(1-0.05,p,n-p))
```

```
## [1] 47.45999
```

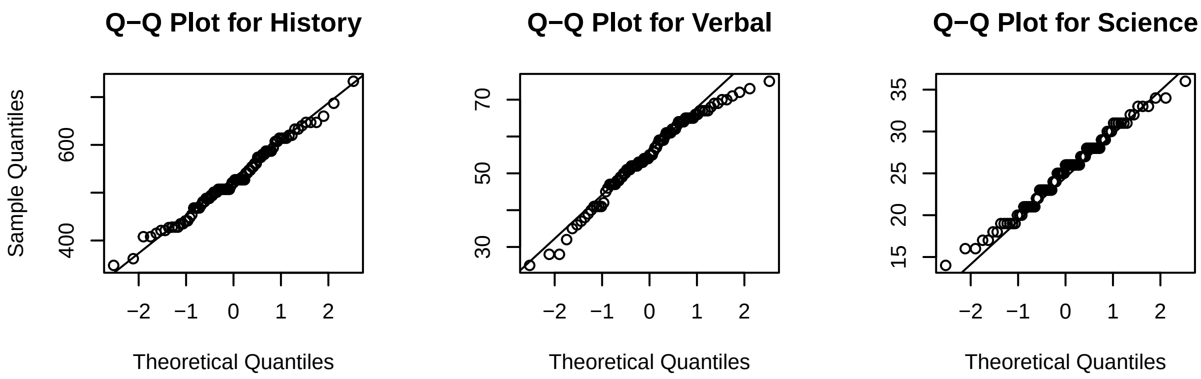
```r
2*sqrt(63.835101127800784*(n-1)*p/(n*(n-p))*qf(1-0.05,p,n-p))
```

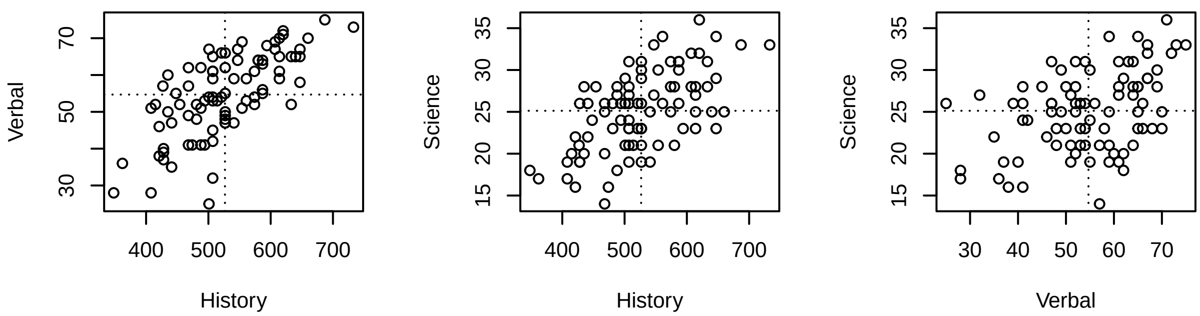
```
## [1] 4.945537
```

```r
2*sqrt(14.598052422256693*(n-1)*p/(n*(n-p))*qf(1-0.05,p,n-p))
```

```
## [1] 2.365
```

```r
par(mfrow=c(2,3))                                                          # Problem 5.18(c)
qqnorm(ct$V1,main="Q-Q Plot for History")
qqline(ct$V1)
qqnorm(ct$V2,main="Q-Q Plot for Verbal",ylab="")
qqline(ct$V2)
qqnorm(ct$V3,main="Q-Q Plot for Science",ylab="")
qqline(ct$V3)
plot(ct$V1,ct$V2,xlab="History",ylab="Verbal")
abline(v=mean(ct$V1),h=mean(ct$V2),lty=3)
plot(ct$V1,ct$V3,xlab="History",ylab="Science")
abline(v=mean(ct$V1),h=mean(ct$V3),lty=3)
plot(ct$V2,ct$V3,xlab="Verbal",ylab="Science")
abline(v=mean(ct$V2),h=mean(ct$V3),lty=3)
```





These data appear to be normal. We can see there is little variance in the Q-Q plots and the scatterplots do not appear to have any outliers or unduly influencing points.

Problem 6.1

```
rm(list=ls())
ef<-read.table("/Users/newuser/Desktop/Notes/Graduate/STAT 488 - Multivariate Statistical Analysis/T6-1
d<-matrix(c(mean(ef$V1-ef$V3),mean(ef$V2-ef$V4)))
S<-matrix(c(var(ef$V1-ef$V3),cov(ef$V1-ef$V3,ef$V2-ef$V4),cov(ef$V1-ef$V3,ef$V2-ef$V4),var(ef$V2-ef$V4)
n<-dim(ef)[1]
p<-dim(d)[1]
(n-1)*p/(n-p)*qf(1-0.05,p,n-p)
```

```
## [1] 9.458877
```

```
cat("(",d[1]-sqrt((n-1)*p/(n-p)*qf(1-0.05,p,n-p)*S[1,1]/n),", ",d[1]+sqrt((n-1)*p/(n-p)*qf(1-0.05,p,n-p
```

```
## (-22.45327, 3.726)
```

```
cat("(",d[2]-sqrt((n-1)*p/(n-p)*qf(1-0.05,p,n-p)*S[2,2]/n),", ",d[2]+sqrt((n-1)*p/(n-p)*qf(1-0.05,p,n-p
```

```
## (-5.700119, 32.24557)
```

```
(t(d)-matrix(c(0,0),nrow=1))%*%solve(S)%*%(d-matrix(c(0,0)))
```

```
##          [,1]
## [1,] 1.239937
```

```
(n-1)*p/(n*(n-p))*qf(1-0.05,p,n-p)
```

```
## [1] 0.8598979
```

We can see the point $\delta = 0$ falls outside the 95 percent confidence region (as calculated by (6-9)) as mentioned. This is consistent with the test for $H_0 : \delta = 0$ because the 95 percent simultaneous confidence intervals encompass more than one linear combination of δ_1 and δ_2.

Problem 6.2

```
cat("(",d[1]-qt(1-0.05/(2*p),n-1)*sqrt(S[1,1]/n),", ",d[1]+qt(1-0.05/(2*p),n-1)*sqrt(S[1,1]/n),")",sep=
```

```
## (-20.57311, 1.845835)
```

```
cat("(",d[2]-qt(1-0.05/(2*p),n-1)*sqrt(S[2,2]/n),", ",d[2]+qt(1-0.05/(2*p),n-1)*sqrt(S[2,2]/n),")",sep=
```

```
## (-2.974903, 29.52036)
```

We can see the 95 percent Bonferroni simultaneous confidence intervals for δ are δ_1 : (-20.5731073, 1.8458345) and δ_2 : (-2.9749034, 29.520358). The 95 percent simultaneous confidence intervals from Example 6.1 (page 277) are δ_1 : (-22.4532723, 3.7259996) and δ_2 : (-5.7001193, 32.2455738), so in comparing the two pairs of intervals, we can see the Bonferroni simultaneous confidence intervals are more narrow.

Problem 6.16

```
rm(list=ls())
stiff<-read.table("/Users/newuser/Desktop/Notes/Graduate/STAT 488 - Multivariate Statistical Analysis/T4
x<-matrix(c(mean(stiff$V1),mean(stiff$V2),mean(stiff$V3),mean(stiff$V4)))
s11<-var(stiff$V1)
s12<-s21<-cov(stiff$V1,stiff$V2)
s13<-s31<-cov(stiff$V1,stiff$V3)
s14<-s41<-cov(stiff$V1,stiff$V4)
s22<-var(stiff$V2)
s23<-s32<-cov(stiff$V2,stiff$V3)
s24<-s42<-cov(stiff$V2,stiff$V4)
s33<-var(stiff$V3)
s34<-s43<-cov(stiff$V3,stiff$V4)
s44<-var(stiff$V4)
S<-matrix(c(s11,s12,s13,s14,s21,s22,s23,s24,s31,s32,s33,s34,s41,s42,s43,s44),nrow=4)
C<-matrix(c(1,0,0,-1,1,0,0,-1,1,0,0,-1),nrow=3)
n<-dim(stiff)[1]
q<-dim(x)[1]
n*t(C%*%x)%*%solve(C%*%S%*%t(C))%*%(C%*%x)
```

```
##          [,1]
## [1,] 254.7212
```

```
(n-1)*(q-1)/(n-q+1)*qf(1-0.05,q-1,n-q+1)
```

```
## [1] 9.53891
```

```
# We reject H0 at the alpha = 0.05 level. There is sufficient evidence (254.7212 > 9.5389)
# that treatment effects exist between the variables.
c<-matrix(c(1,1,-1,-1)) # Testing for (mu_1 + mu_2) - (mu_3 + mu_4)
t(c)%*%x-sqrt((n-1)*(q-1)/(n-q+1)*qf(1-0.05,q-1,n-q+1)*t(c)%*%S%*%c/n)
```

```
##          [,1]
## [1,] 247.0397
```

```
t(c)%*%x+sqrt((n-1)*(q-1)/(n-q+1)*qf(1-0.05,q-1,n-q+1)*t(c)%*%S%*%c/n)
```

```
##         [,1]
## [1,] 596.027
```

We can see the 95 percent simultaneous confidence interval for comparing "dynamic" (x_1, x_2) and static (x_3, x_4) methods is (247.0396637, 596.027003).

Problem 6.19

```
rm(list=ls())
milk<-read.table("/Users/newuser/Desktop/Notes/Graduate/STAT 488 - Multivariate Statistical Analysis/T6-
gas<-milk[milk$V4=="gasoline",]                                            # Problem 6.19(a)
die<-milk[milk$V4=="diesel",]
x1<-matrix(c(mean(gas$V1),mean(gas$V2),mean(gas$V3)))
x2<-matrix(c(mean(die$V1),mean(die$V2),mean(die$V3)))
S1<-matrix(c(var(gas$V1),cov(gas$V1,gas$V2),cov(gas$V1,gas$V3),cov(gas$V2,gas$V1),var(gas$V2),cov(gas$V
S2<-matrix(c(var(die$V1),cov(die$V1,die$V2),cov(die$V1,die$V3),cov(die$V2,die$V1),var(die$V2),cov(die$V
n1<-dim(gas)[1]
n2<-dim(die)[1]
n<-dim(milk)[1]
p<-dim(milk[c("V1","V2","V3")])[2]
npn1<-abs(diff(dim(milk)))
Sp<-(n1-1)/(n-2)*S1+(n2-1)/(n-2)*S2
t(x1-x2-matrix(c(0,0,0)))%*%solve((1/n1+1/n2)*Sp)%*%(x1-x2-matrix(c(0,0,0)))
```

```
##          [,1]
## [1,] 50.91279
```

```
(n-2)*p/npn1*qf(1-0.01,p,npn1)
```

```
## [1] 12.93096
```

```
# We reject H0 at the alpha = 0.01 level. There is sufficient evidence (50.912 > 12.931)
# that the mean cost vectors of gasoline and diesel trucks are different.
14.04*solve(Sp)%*%(x1-x2)                                                  # Problem 6.19(b)
```

```
##           [,1]
## [1,]  3.576623
## [2,] -1.880006
## [3,] -4.476368
```

```
x1[1]-x2[1]-sqrt((n-2)*p/npn1*qf(1-0.01,p,npn1)*(1/n1+1/n2)*Sp[1,1]) # Problem 6.19(c)
```

```
## [1] -1.704346
```

```
x1[1]-x2[1]+sqrt((n-2)*p/npn1*qf(1-0.01,p,npn1)*(1/n1+1/n2)*Sp[1,1])
```

```
## [1] 5.930264
```

```
x1[2]-x2[2]-sqrt((n-2)*p/npn1*qf(1-0.01,p,npn1)*(1/n1+1/n2)*Sp[2,2])
```

```
## [1] -7.022268
```

```
x1[2]-x2[2]+sqrt((n-2)*p/npn1*qf(1-0.01,p,npn1)*(1/n1+1/n2)*Sp[2,2])
```

```
## [1] 1.72292
```

```
x1[3]-x2[3]-sqrt((n-2)*p/npn1*qf(1-0.01,p,npn1)*(1/n1+1/n2)*Sp[3,3])
```

```
## [1] -13.52648
```

```r
x1[3]-x2[3]+sqrt((n-2)*p/npn1*qf(1-0.01,p,npn1)*(1/n1+1/n2)*Sp[3,3])
```

```
## [1] -3.628618
```

```r
# We can see from the 99 percent simultaneous confidence intervals for each pair of
# variables that the cost of x3 = capital appears to be the most different.
# It appears that all assumptions listed in (6-19) have been met.    # Problem 6.19(d)
milks<-read.table("/Users/newuser/Desktop/Notes/Graduate/STAT 488 - Multivariate Statistical Analysis/T
gass<-milks[milks$V4=="gasoline",]
dies<-milks[milks$V4=="diesel",]
x1s<-matrix(c(mean(gass$V1),mean(gass$V2),mean(gass$V3)))
x2s<-matrix(c(mean(dies$V1),mean(dies$V2),mean(dies$V3)))
S1s<-matrix(c(var(gass$V1),cov(gass$V1,gass$V2),cov(gass$V1,gass$V3),cov(gass$V2,gass$V1),var(gass$V2),c
S2s<-matrix(c(var(dies$V1),cov(dies$V1,dies$V2),cov(dies$V1,dies$V3),cov(dies$V2,dies$V1),var(dies$V2),c
n1s<-dim(gass)[1]
n2s<-dim(dies)[1]
ns<-dim(milks)[1]
ps<-dim(milks[c("V1","V2","V3")])[2]
Sps<-(n1s-1)/(ns-2)*S1s+(n2s-1)/(ns-2)*S2s
t(x1s-x2s-matrix(c(0,0,0)))%*%solve((1/n1s+1/n2s)*Sps)%*%(x1s-x2s-matrix(c(0,0,0)))
```

```
##         [,1]
## [1,] 52.3836
```

```r
(ns-2)*ps/abs(diff(dim(milks)))*qf(1-0.01,ps,abs(diff(dim(milks))))
```

```
## [1] 12.99521
```

We reject H_0 at the $\alpha = 0.01$ level. There is sufficient evidence ($52.3835991 > 12.9952069$) that the mean cost vectors of gasoline and diesel trucks (excluding observations 9 and 21) are different.