223hw2

```r
#223 hw 2
# zihao Huang
#2/12/2018


# 4.10
mu<-matrix(c(3,1,4))
sg<-matrix(c(6,1,-2,1,13,4,-2,4,4),nrow=3)
#a
z<-matrix(c(2,-1,3),nrow=1)
ex<-(z%*%mu)
varz<-z%*%sg%*%t(z)
## distribution is N(17,21)

#b
z1<-matrix(c(1,1,1),nrow=1)
z2<-matrix(c(1,-1,2),nrow=1)
Z<-rbind(z1,z2)
(Z%*%mu)
```

```
##      [,1]
## [1,]    8
## [2,]   10
```

```r
Z%*%sg%*%t(Z)
```

```
##      [,1] [,2]
## [1,]   29   -1
## [2,]   -1    9
```

```r
## it is following normal distribution with:
## mean is  8    variance is 29 -1
##         10                 -1  9

#c
## y2 is following N(1,13)

#d
## y1 and y3's joint distribution are following normal distribution with:
## mean is  3    variance is 6 -2
##          4               -2  4

#e
J<-matrix(c(1,0,0,0,0,1,0.5,0.5,0),byrow = T,nrow=3)
## The joint distribution follows with
## mu
(J%*%mu)
```

```
##      [,1]
## [1,]    3
## [2,]    4
## [3,]    2
```

```
## variance
(J%*%sg%*%t(J))
```

```
##      [,1] [,2] [,3]
## [1,]  6.0   -2 3.50
## [2,] -2.0    4 1.00
## [3,]  3.5    1 5.25
```

```
#4.11
mu<-matrix(c(3,1,4))
sg<-matrix(c(6,1,-2,1,13,4,-2,4,4),nrow=3)
t(mu-mean(mu))%*%mu
```

```
##          [,1]
## [1,] 4.666667
```

```
#b
c<-eigen(sg)$vectors
d0.5<-diag(diag(diag(sqrt(eigen(sg)$values))))
solve(c%*%d0.5%*%solve(c))
```

```
##             [,1]        [,2]       [,3]
## [1,]  0.46496849 -0.06966936  0.1701484
## [2,] -0.06966936  0.32645938 -0.1657086
## [3,]  0.17014841 -0.16570861  0.6916013
```

```
##[  0.46496849 -0.06966936  0.1701484][y-3]
##[ -0.06966936  0.32645938 -0.1657086][y-1]
##[  0.17014841 -0.16570861  0.6916013][y-4]
#c
##The distribution of part c is chi-square (3),since rank=3

#4.14
#The following variables are independent:
##(b),(c),(d), since their covariance are 0.

#4.16
muy<-matrix(c(2,-1),ncol=1)
mux<-matrix(c(3,1),ncol=1)
syy<-matrix(c(7,3,3,6),ncol=2)
syx<-matrix(c(-3,0,2,4),ncol=2)
sxy<-matrix(c(-3,2,0,4),ncol=2)
sxx<-matrix(c(5,-2,-2,4),ncol=2)
syx%*%solve(sxx)
```

```
##      [,1] [,2]
## [1,] -0.5 0.25
## [2,]  0.5 1.25
```

```
##E(y|x) is| 2| + | -0.5 0.25||x1-3|
##         |-1|   |  0.5 1.25||x2-1|
syy-syx%*%solve(sxx)%*%sxy
```

```
##      [,1] [,2]
## [1,]    5    2
## [2,]    2    1
```

```
##[1,]cov(y|x)=    5    2
##[2,]             2    1
eyx<-muy+syx%*%solve(sxx)%*%c(2-3,4-1)
##E(y|x=2,4) is |3.25|
##              |2.25|
(covyx<-syy-syx%*%solve(sxx)%*%sxy)
```

```
##      [,1] [,2]
## [1,]    5    2
## [2,]    2    1
```

```
#5.11
y<-matrix(c(3,6,5,10,10,12,14,9),nrow=4)
matvar<-var(y)
n<-nrow(y)
p<-ncol(y)
meany<- apply(y,2,mean)
meand<-meany-c(6,11)
(T2<-nrow(y)*t(meand)%*%solve(matvar)%*%meand)
```

```
##            [,1]
## [1,] 0.06103286
```

```
##T value is 0.061.
critval <- p*(n-1)/(n-p)*qf(.95,p,n-p)
(pvalT2 <- 1-pf((n-p)/((n-1)*p)*T2,p,n-p))
```

```
##           [,1]
## [1,] 0.9800613
```

```
##T2 is 0.061, pvalue is 0.9801>0.05, we cannot reject the null hypothesis
##and accept H0: mean=(6,11)

#5.12
mu0<-matrix(c(30,25,40,25,30),nrow=1)
y12<-read.table("C:/Users/simon/Desktop/STAT223/T3_6_PROBE.DAT",header=F)
y<-as.matrix(y12[,2:6])
matvar<-var(y)
n<-nrow(y)
p<-ncol(y)
meany<- apply(y,2,mean)
meand<-meany-c(30,25,40,25,30)
T2<-nrow(y)*t(meand)%*%solve(matvar)%*%meand
##T2 is 85.3327
critval <- p*(n-1)/(n-p)*qf(.95,p,n-p)
pval <- 1-pf((n-p/(p*(n-1))),p,n-p)
##p-value is 0.005701325<0.05, we reject the null hypothesis and accept
##alternative hypothesis that the means are not equal to mu0(30,25,40,25,30).
t<-c()
for (i in 1:p){
t[i]<-meand[i]/sqrt(matvar[i,i]/n)
}
## t is 2.5039076  0.2665219 -2.5156788  0.9510448  0.3161290
(p.value =2*pt(-abs(t), n-1))
```

```
## [1] 0.03123701 0.79525232 0.03061322 0.36400147 0.75840431
```

```
##The p-value from t test for each variables are 0.031 0.795 0.031 0.364 0.758.
##The p-value for only 1st and 3rd variables are significant.
##This shows that only the mean of 1st and 3rd row
##are different from the mean of H0,
##and other mean of rows are not significant to H0.

#5.16
y16<-read.table("C:/Users/simon/Desktop/STAT223/T5_5_FBEETLES.DAT",header=F)
y1<-y16[1:19,3:6]
y2<-y16[20:39,3:6]
s1<-var(y1)
s2<-var(y2)

meany1<- apply(y1,2,mean)
meany2<- apply(y2,2,mean)
meandiff<-(meany1-meany2)
n1<-19
n2<-20
p<-4
sp<-((n1-1)*s1+(n2-1)*s2)/(n1+n2-2)
T2<-(n1*n2/(n1+n2))*t(meandiff)%*%solve(sp)%*%(meandiff)
## T2 is 133.4873
a <- 1/(p*(n2+n1-2)/(n1+n2-p-1))
1-pf(a*T2, p, n1+n2-p-1)
```

```
##              [,1]
## [1,] 7.521805e-11
```

```
critval <- p*(n1+n2-2)/(n1+n2-p-1)*qf(.95,p,n1+n2-p-1)
##p-value is 7.521805e-11<0.05, we reject the null hypothesis and accept that
##their group means are significant to each other.
t<-c()
for (i in 1:4){
  t[i]<-meandiff[i]/sqrt(sp[i,i]*(1/n1+1/n2))
}
##t is 3.887946 -3.865239 -5.691131 -5.042625
(p.value =2*pt(-abs(t), n1+n2-2))
```

```
## [1] 4.049179e-04 4.326232e-04 1.645393e-06 1.236427e-05
```

```
##p-value are 4.049179e-04 4.326232e-04 1.645393e-06 1.236427e-05, these
##values means all the variables reject H0 and they are significant between two groups.
a<-t(solve(sp)%*%meandiff)
##a is 0.345249 -0.1303878 -0.1064338 -0.1433533

#5.20
y20<-read.table("C:/Users/simon/Desktop/STAT223/T5_8_GOODS.DAT",header=F)
y1<-y20[1:9,3:6]
y2<-y20[10:19,3:6]
n1<-9
n2<-10
s1<-var(y1)
s2<-var(y2)
meany1<- apply(y1,2,mean)
meany2<- apply(y2,2,mean)
```

```
meandiff<-(meany1-meany2)
sp<-((n1-1)*s1+(n2-1)*s2)/(n1+n2-2)
T2<-(n1*n2/(n1+n2))*t(meandiff)%*%solve(sp)%*%(meandiff)
##T2 is 18.46248
a<-t(solve(sp)%*%meandiff)
##a is -0.05689601 -0.009709542 -0.2421341 -0.07128274
(pval <- 1-pf((n1+n2-p/(p*(n1+n2-1))),p,n1+n2-p))
```

```
## [1] 9.926975e-06
```

```
##p-value is 8.73739e-06<0.05, we reject the null hypothesis and conclude that
##the two gourps have different means.

#5.21
y21<-read.table("C:/Users/simon/Desktop/STAT223/T5_9_ESSAY.DAT",header=F)
y21
```

```
##    V1  V2 V3  V4 V5
## 1   1 148 20 137 15
## 2   2 159 24 164 25
## 3   3 144 19 224 27
## 4   4 103 18 208 33
## 5   5 121 17 178 24
## 6   6  89 11 128 20
## 7   7 119 17 154 18
## 8   8 123 13 158 16
## 9   9  76 16 102 21
## 10 10 217 29 214 25
## 11 11 148 22 209 24
## 12 12 151 21 151 16
## 13 13  83  7 123 13
## 14 14 135 20 161 22
## 15 15 178 15 175 23
```

```
d1<-y21[,2]-y21[,4]
d2<-y21[,3]-y21[,5]
d<-cbind(d1,d2)
sd<-var(d)
n<-15
p<-2
meandiff <- apply(d,2,mean)
(T2 <- n*t(meandiff)%*%solve(sd)%*%meandiff)
```

```
##          [,1]
## [1,] 15.19123
```

```
##T2 is 15.19123
critval <- p*(n-1)/(n-p)*qf(.95,p,n-p)
(pval <- 1-pf((n-p/(p*(n-1))),p,n-p))
```

```
## [1] 0.0004290233
```

```
##P-value is 0.0004<0.05, we reject the null hypothesis and conclude that
##the mean of distance is significant to 0.
a<-t(solve(sd)%*%meandiff)
##a is -0.03601558 0.04770629
t<-c()
```

```r
for (i in 1:2){
  t[i]<-meandiff[i]/sqrt(sd[i,i]*(1/n))
}
##t is -3.837148 -2.436207
(p.value =2*pt(-abs(t), n1+n2-2))
```

```
## [1] 0.001319992 0.026135975
```

```r
##p-value are 0.0001020989 0.0044033449
##They are all smaller than 0.05, all the variables reject H0
##and they are sgnificant between two groups.
```