homework5

Yi Chen

2/21/2020

Homework 5

problem 1

question a

```
## load the data
data <- read.table('T11-7.DAT',header = F)
zone <- as.numeric(as.factor(data$V6))
```

In this case, we need to compare the mean of five variables ($p = 5$) in three different groups ($g = 3$) using multivariate analysis of variance (MANOVA). Let τ_l represent the l-th treatment effect. In this case the treatment effect comes from the difference in zone. With constraint of $\sum_{l=1}^{g} n_l \tau_l = 0$ (n_l is the number of observation in the l-th treatment group), we have the hypohtesis of there are **no** treatment effects.

$$H_0 : \tau_1 = \tau_2 = \tau_3 = 0$$

This hyphothesis is tested by considering the relative sizes of the treatment and residual sums of squares and cross producets.

question b

```
library(dplyr)
n1 = as.numeric(table(zone)[1])
n2 = as.numeric(table(zone)[2])
n3 = as.numeric(table(zone)[3])
(n = c(n1,n2,n3))
```

```
## [1] 11 38  7
```

```
g = length(unique(data$V6))
p = ncol(data) -1
## calculate the mean of variable for each group
(group.bar <- data %>% group_by(V6) %>% summarise_all(funs(mean)))
```

```
## # A tibble: 3 x 6
##   V6          V1    V2    V3    V4    V5
##   <fct>    <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 SubMuli  4.45  33.1 0.171  6.56  5.48
## 2 Upper    7.23  22.3 0.432  4.66  5.77
## 3 Wilhelm  3.23  43.6 0.117  6.80 11.5
```

```
## calculate the overall mean
(x.bar <- colMeans(data[,1:5]))
```

```
##         V1         V2         V3         V4         V5
##  6.1803571 27.0464286  0.3414286  5.2991071  6.4335714
```

```
## calculate treatment sum of squre (B)
B <- 0
for (i in 1:g){
  B <- B + n[i] * t(as.matrix(group.bar[i,2:(p+1)] - x.bar)) %*% (as.matrix(gro
up.bar[i,2:(p+1)] - x.bar))
}
## calculate the residual sum of square (W)
W <- 0
for (i in 1:g){
  group.cov <- cov(data %>% filter(zone==i) %>% select(1:5))
  W <- W + (n[i]-1) * group.cov
}
## calculate the wilks' lambda
(Lambda = det(W)/det(B + W))
```

```
## [1] 0.115911
```

question c

```
# F statistics
(F.statistics <- ((sum(n)-p-2)/p)* ((1- sqrt(Lambda))/sqrt(Lambda)))
```

```
## [1] 18.98484
```

question d

```
# F-critical
alpha = 0.05
(F.critical <- qf(1 - alpha, 2*p, 2*(sum(n)-p-2) ))
```

```
## [1] 1.928687
```

question e

Since the F statistics is bigger than the F ciritical value, we can conclude that: we need to reject the H_0 hyphothesis. In other words, there exist statistically significant treatment effects.

I also use the package to double check the result. We get the same conclusion.

```
dependent.vars = as.matrix(data[,1:5])
group.indicator = as.factor(data$V6)
(ex.manova = summary(manova(dependent.vars ~ group.indicator), test = "Wilks"))
```

```
##                 Df   Wilks approx F num Df den Df    Pr(>F)
## group.indicator  2 0.11591   18.985     10     98 < 2.2e-16 ***
## Residuals       53
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

question f

We have calculate the corresponding B and W in the question b for calculate lambda. Here we report the result

```
B
```

```
##            V1         V2          V3         V4          V5
## V1  135.67315 -647.33656  11.4925598 -80.479227 -113.841359
## V2 -647.33656 3186.68117 -53.8000232 373.774403  648.788140
## V3   11.49256  -53.80002   0.9844204  -6.924981   -8.529018
## V4  -80.47923  373.77440  -6.9249811  48.803422   56.524562
## V5 -113.84136  648.78814  -8.5290178  56.524562  209.294200
```

```
W
```

```
##            V1         V2         V3         V4           V5
## V1 187.575243  -34.81237 -6.8479884 -21.133755   79.6722871
## V2 -34.812372 4221.15811 20.1123090  83.721918 -287.5114258
## V3  -6.847988   20.11231  4.4356653   8.637653   -0.3915679
## V4 -21.133755   83.72192  8.6376526  57.040433   33.2150163
## V5  79.672287 -287.51143 -0.3915679  33.215016  338.0228861
```

To double check, I compare these values with the results from package.

```
## check B
all(round(B,2)==round(ex.manova$SS[1]$group.indicator,2))
```

```
## [1] TRUE
```

```
## check W
all(round(W,2)==round(ex.manova$SS[2]$Residuals,2))
```

```
## [1] TRUE
```

question g

```
solve(W) %*% B
```

```
##            V1          V2          V3         V4         V5
## V1  1.0450289  -5.2450064  0.08579106 -0.5929236 -1.1557197
## V2 -0.1900605   0.9743514 -0.01538732  0.1057064  0.2322062
## V3  8.2501011 -38.1037641  0.71213845 -5.0251030 -5.5653725
## V4 -1.6610066   7.0589302 -0.14983832  1.0755359  0.4606293
## V5 -0.5719877   3.2465976 -0.04299261  0.2853781  1.0373734
```

```
(e.values <- eigen(solve(W) %*% B, only.values = TRUE)$values)
```

```
## [1]  4.178414e+00  6.660138e-01  7.468798e-16  1.227391e-16 -1.164161e-16
```

```
lambda.new <- 1
for (i in 1:min(p,g-1)){
  lambda.new <- lambda.new * (1/(1+e.values[i]))
}
lambda.new
```

```
## [1] 0.115911
```

```
round(lambda.new,5) == round(Lambda,5)
```

```
## [1] TRUE
```

The results match with each other.