

Biostat 653 Homework 2

David (Daiwei) Zhang

October 11, 2017

Problem 2

(1)

```
N <- 1000
mu <- c(0,0)
sig <- matrix(c(2, 0.5, 0.5, 1), 2, 2)
y <- mvrnorm(N, mu, sig)
head(y)
```

```
##           [,1]      [,2]
## [1,]  0.3320714 -0.8744666
## [2,]  0.3503837 -0.1306056
## [3,]  1.9641893  0.5817348
## [4,]  1.1061370 -0.3443937
## [5,] -0.4802603  0.0159834
## [6,] -0.4367549 -0.4588192
```

```
y <- t(y)
```

(3)

```
x <- matrix(1,2,1)
sig <- matrix(c(1,0,0,1), 2, 2)
num.iter <- 1e4
for(i in 1:num.iter){
  sig.inv <- solve(sig)
  mu <- 1/N / sum(sig.inv) * sum(colSums(sig.inv) %*% y)
  sig <- (y - mu) %*% t(y - mu) / N
}
mu.var <- 1/sum(sig.inv) / N
print(mu)
```

```
## [1] -0.01667662
```

```
print(sig)
```

```
##           [,1]      [,2]
## [1,]  1.9350772  0.5044371
## [2,]  0.5044371  1.0974171
```

```
print(mu.var)
```

```
## [1] 0.0009236566
```

(4)

Test statistic: $N(\mathbf{1}_2 \hat{\mu})^T \hat{\Sigma}^{-1} (\mathbf{1}_2 \hat{\mu}) \sim \chi_1^2$.

```
test.stat <- N * sum(sig.inv) * mu^2
pval <- 1 - pchisq(test.stat, 1)
print(pval)
```

```
## [1] 0.583196
```

Since the p-value is 0.583196, we do not reject the null hypothesis at $\alpha = 0.05$.

(6)

```
x <- matrix(1,2,1)
w <- matrix(c(1,0,0,1), 2, 2)
num.iter <- 1e4
for(i in 1:num.iter){
  mu <- 1/N / sum(w) * sum(colSums(w) %*% y)
  sig <- (y - mu) %*% t(y - mu) / N
  a <- sum(diag(sig)) / 2
  b <- sig[1,2]
  sig <- matrix(c(a,b,b,a), 2,2)
  w <- solve(sig)
}
mu.var <- 1/sum(sig.inv) / N
print(mu)
```

```
## [1] -0.01480577
```

```
print(sig)
```

```
##           [,1]      [,2]
## [1,] 1.5162437 0.5044336
## [2,] 0.5044336 1.5162437
```

```
print(mu.var)
```

```
## [1] 0.0009236566
```

(7)

We see that $\hat{\mu}$ is close to μ , and the off diagonal elements of $\hat{\Sigma}$ is close to those of Σ . However, the diagonal elements are the average of the true diagonal elements, which makes sense given the restriction of our model.

(8)

```
mu.var.model <- 1 / sum(w)/N
mu.var.robust <- (1/N)*(1/sum(w)) * sum((colSums(w)%*%(y-mu))^2) * (1/sum(w))*(1/N)
print(mu.var.model)
```

```
## [1] 0.001010339
```

```
print(mu.var.robust)
```

```
## [1] 0.001010339
```

The two variances are close here. However, if they are different, then the robust variance is better to use, since our model is not correct.

Problem 3

5.1.1

```
col.names = c("group", "id", "cho0", "cho6", "cho12", "cho20", "cho24")
cho <- read.table("../Downloads/cholesterol-data.txt", na.strings = ".", col.names = col.names)
head(cho)
```

```
##   group id cho0 cho6 cho12 cho20 cho24
## 1     1  1  178  246   295   228   274
## 2     1  2  254  260   278   245   340
## 3     1  3  185  232   215   220   292
## 4     1  4  219  268   241   260   320
## 5     1  5  205  232   265   242   230
## 6     1  6  182  213   173   200   193
```

5.1.2

```
means <- c()
vars <- c()
sds <- c()
for(i in 1:2){
  print(i)
  mean.this <- colMeans(cho[cho$group==i,-(1:2)], na.rm = TRUE)
  var.this <- colVars(as.matrix(cho[cho$group==i,-(1:2)]), na.rm = TRUE)
  sd.this <- colSds(as.matrix(cho[cho$group==i,-(1:2)]), na.rm = TRUE)
  means <- rbind(means, mean.this)
  vars <- rbind(vars, var.this)
  sds <- rbind(sds, sd.this)
}
```

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

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

```
rownames(means) <- c(1,2)
rownames(vars) <- c(1,2)
rownames(sds) <- c(1,2)
colnames(vars) <- colnames(means)
colnames(sds) <- colnames(means)
print(means)
```

```
##      cho0      cho6      cho12      cho20      cho24
## 1 226.0161 245.5323 252.0182 256.7955 254.5526
## 2 235.9268 243.1707 244.7632 257.6000 257.4839
```

```
print(vars)
```

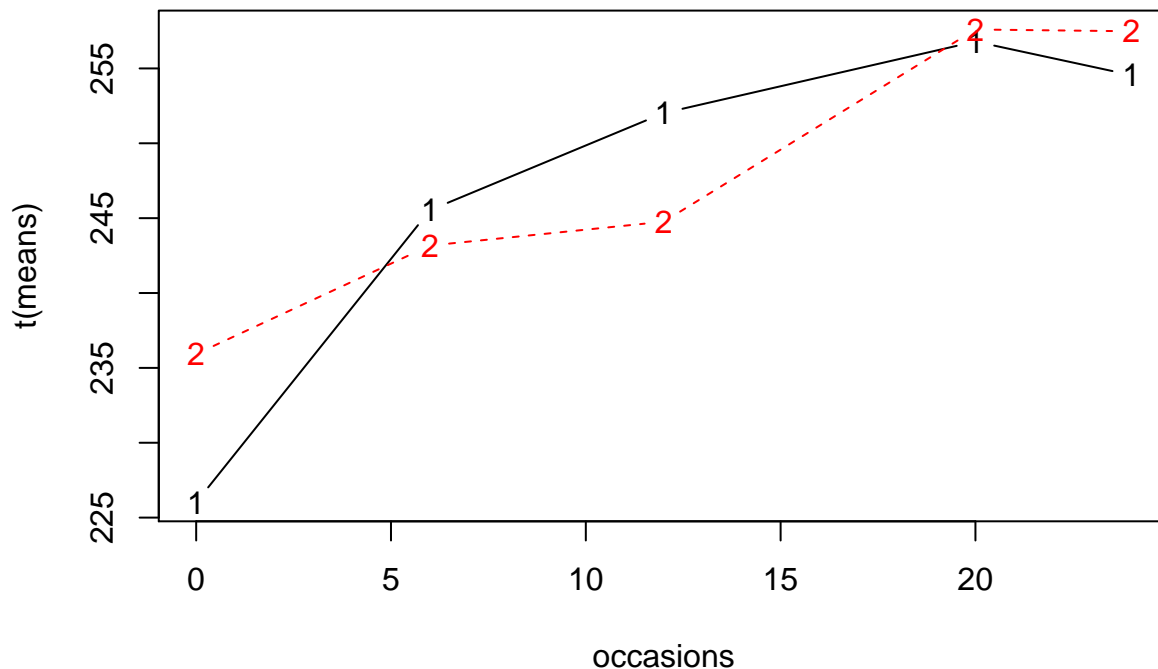
```
##      cho0      cho6      cho12      cho20      cho24
## 1 1573.262 1556.483 1469.129 1189.515 2496.200
## 2 3121.970 2424.545 2126.186 2615.482 2439.191
```

```
print(sds)
```

```
##      cho0      cho6      cho12      cho20      cho24
## 1 39.66437 39.45228 38.32922 34.48935 49.96198
## 2 55.87459 49.23967 46.11058 51.14179 49.38817
```

5.1.3

```
occasions <- c(0, 6, 12, 20, 24)
matplot(occasions, t(means), type="b")
```



Cholesterol levels increase over time for both groups. Compared to that of the placebo group, the cholesterol level of the high-dose group is lower at baseline and the end but higher in the middle.

5.1.4

```
cho.uni <- c()
for(i in 1:5){
  cho.uni.this <- cho[,c(1,2,2+i)]
  colnames(cho.uni.this) <- c("group", "id", "cho")
  cho.uni.this$occasion <- occasions[i]
  cho.uni <- rbind(cho.uni, cho.uni.this)
}
head(cho.uni)
```

```
##      group id cho occasion
## 1      1  1 178          0
## 2      1  2 254          0
## 3      1  3 185          0
## 4      1  4 219          0
## 5      1  5 205          0
## 6      1  6 182          0
```

```
write.table(cho.uni, "biostat653_data_cho.uni", row.names = FALSE, col.names = FALSE)
```

5.1.5

See SAS output. The p-value for the F-statistic for the interaction of group by occasion is 0.0990, so we do not reject the null hypothesis that the patterns of change over time is different in the two groups at $\alpha = 0.05$.

5.1.6

See the SAS output for the covariance matrix (“Estimated R Matrix for ID 1”) and the correlation matrix (“Estimated R Correlation Matrix for ID 1”)

5.1.7

The model is

$$\begin{aligned}
 cholesterol_i = & \beta_0 \\
 & + \beta_1 I(occasion = 6) + \beta_2 I(occasion = 12) \\
 & + \beta_3 I(occasion = 20) + \beta_4 I(occasion = 24) \\
 & + \beta_5 I(group = 1) \\
 & + \beta_6 I(group = 1) I(occasion = 6) + \beta_7 I(group = 1) I(occasion = 12) \\
 & + \beta_8 I(group = 1) I(occasion = 20) + \beta_9 I(group = 1) I(occasion = 24)
 \end{aligned}$$

5.1.8

$$L = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

5.1.9

We can simply add up the effects to get the means. For example,

$$\begin{aligned}
 E[group = 2, occasion = 0] &= 235.93 \\
 E[group = 2, occasion = 6] &= 235.93 + 7.2439 = 243.1739 \\
 E[group = 1, occasion = 0] &= 235.93 + (-9.9107) = 226.0193 \\
 E[group = 1, occasion = 6] &= 235.93 + (-9.9107) + 7.2439 + 12.2722 = 245.5354
 \end{aligned}$$

This gives the same result from the group means calculated earlier.

5.1.10

Interpretations:

β_0 : The mean cholesterol level for an individual in the placebo group at baseline.

β_1 : The mean difference in cholesterol level for an individual in the placebo group from baseline to 6 months after.

β_2 to β_4 : Similar to β_1 .

β_6 : The mean difference in cholesterol level at baseline from an individual in the placebo group to an individual in the high-dose group.

β_7 : The mean group effect on the difference in cholesterol level from baseline to 6 months after.

β_8 and β_9 : Similar to β_7 .