# Homework 11

Yonghyun Kwon

4/20/2020

## Problem 1

```r
HeartRate = read.csv("https://dnett.github.io/S510/HeartRate.txt", sep =  "\t")

HeartRate$time <- factor(HeartRate$time)
HeartRate$woman <- factor(HeartRate$woman)
```

### (a)

We can express the model as follows:

$$y = X\beta + Zu + \varepsilon$$

where $Z = I_{15} \otimes \mathbb{1}_4$, $u \sim N(0, \sigma_w^2 I_{15}),\ \varepsilon \sim N(0, \sigma_e^2 I_{60})$. Hence, the variance of $\mathbf{y}$ is

$$\begin{aligned} Var(\mathbf{y}) &= Z\ Var(u) Z^T + Var(\varepsilon) \\ &= \sigma_w^2 I_{15} \otimes (\mathbb{1}_4 \mathbb{1}_4^T) + \sigma_e^2 I_{60} \end{aligned}$$

### (b)

We test the following hypothesis test

$$H_0 : \mu_{ik} - \bar{\mu}_{i\cdot} - \bar{\mu}_{\cdot k} + \bar{\mu}_{\cdot\cdot} = 0 \text{ vs } H_1 : not\ H_o$$

drug-by-time interaction corresponds to the last row of Type III anova table, which can be computed using *lmer* function in *lme4* pakcage.

```r
o.lmer = lmer(y ~ drug * time + (1 | woman), data = HeartRate)
anova(o.lmer)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## drug       22.61  11.306     2    12  1.3518    0.2955
## time      256.33  85.444     3    36 10.2159 5.230e-05 ***
## drug:time 357.07  59.511     6    36  7.1152 4.707e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Hence, F-statistics is 7.1152 with (6, 36) degree of freedom and p-value is 4.707e-05, which implies that drug-by-time interaction effect is significant with significant level 0.05.

## (c)

We test the following hypothesis test

$$H_0 : \mu_{13} = \mu_{23} = \mu_{33} \text{ vs } H_1 : not\ H_o$$

Note that this hypothesis is equivalent to R's parametraiztion

$$\begin{aligned}(Intr) + time10 &= (Intr) + drugB + time10 + drugB : tim10 \\ &= (Intr) + drugC + time10 + drugC : tim10\end{aligned}$$

```
C <- rbind(c(0,1,0,0,0,0,0,0,1,0,0,0),
           c(0,0,1,0,0,0,0,0,0,1,0,0))
contest(o.lmer, L = C, joint = T)
```

```
##     Sum Sq  Mean Sq NumDF    DenDF  F value    Pr(>F)
## 1 30.90492 15.45246     2 17.07708 1.847521 0.1877802
```

The test statistics is 1.847521 with degree of freedom (2, 17.07708) and p-value is 0.1877802. There is no great evidence that mean heart rate 10 minutes after treatment is the same for all three drugs with significant level 0.05.

## (d)

```
C <- rbind(c(0,-1,0,0,0,0,0,0,-1,0,0,0))
contest(o.lmer, L = C, joint = F, confint = T)
```

```
##   Estimate Std. Error       df  t value     lower    upper  Pr(>|t|)
## 1      4.4   3.872554 17.07708 1.136201 -3.767566 12.56757 0.2715664
```

Hence, the confidence interval is (-3.767566, 12.56757)

# Problem 2

## (a)

Set correlation=corCompSymm(form = ~ 1 | woman).

```
o.gls1 = gls(y ~ drug * time, data = HeartRate,
            correlation=corCompSymm(form = ~ 1 | woman),
            method = "REML")
getVarCov(o.gls1)
```

```
## Marginal variance covariance matrix
##        [,1]   [,2]   [,3]   [,4]
## [1,] 37.492 29.128 29.128 29.128
## [2,] 29.128 37.492 29.128 29.128
## [3,] 29.128 29.128 37.492 29.128
## [4,] 29.128 29.128 29.128 37.492
##   Standard Deviations: 6.123 6.123 6.123 6.123
```

### (b)

```r
AIC(o.gls1)
```

```
## [1] 317.9204
```

```r
BIC(o.gls1)
```

```
## [1] 344.1172
```

### (c)

Set correlation=corAR1(form = ~ 1 | woman).

```r
o.gls2 = gls(y ~ drug * time, data = HeartRate,
             correlation=corAR1(form = ~ 1 | woman),
             method = "REML")
getVarCov(o.gls2)
```

```
## Marginal variance covariance matrix
##        [,1]   [,2]   [,3]   [,4]
## [1,] 36.011 29.809 24.675 20.426
## [2,] 29.809 36.011 29.809 24.675
## [3,] 24.675 29.809 36.011 29.809
## [4,] 20.426 24.675 29.809 36.011
##   Standard Deviations: 6.0009 6.0009 6.0009 6.0009
```

### (d)

```r
AIC(o.gls2)
```

```
## [1] 313.9425
```

```r
BIC(o.gls2)
```

```
## [1] 340.1394
```

### (e)

Set correlation=corSymm(form = ~ 1 | woman).

```r
o.gls3 = gls(y ~ drug * time, data = HeartRate,
             correlation=corSymm(form = ~ 1 | woman),
             method = "REML")
getVarCov(o.gls3)
```

```
## Marginal variance covariance matrix
##        [,1]   [,2]   [,3]   [,4]
## [1,] 37.009 30.929 32.728 23.343
## [2,] 30.929 37.009 31.639 23.152
## [3,] 32.728 31.639 37.009 29.893
## [4,] 23.343 23.152 29.893 37.009
##   Standard Deviations: 6.0835 6.0835 6.0835 6.0835
```

### (f)

```r
AIC(o.gls3)
```

```
## [1] 317.4352
```

```r
BIC(o.gls3)
```

```
## [1] 352.9881
```

### (g)

From (b), (d), and (f), second structure for $W$ gives the smallest AIC and BIC. Hence, AR(1) correlation structure is preferred for this dataset.

### (h)

Note that we can compute F-statistics for hypothesis testing of linear combinations of parameters using anova. We can use this F-statistics to find a 95% confidence interval. Observe that

$$\frac{contrast}{MSE} = \sqrt{F}$$

A 95% confidence interval is

$$contrast \pm t_{\alpha/2} \frac{contrast}{\sqrt{F} sign(contrast)}$$

```r
C <- rbind(c(0,-1,0,0,0,0,0,0,-1,0,0,0))
aov <- anova(o.gls2, L = C)
tstat <- sqrt(aov$`F-value`)

contrast <- drop(C %*% o.gls2$coefficients)
margin <- qt(0.025, 48) / tstat * sign(contrast)
contrast * (1  - margin)
```

```
## [1] 12.03095
```

```r
contrast * (1  + margin)
```

```
## [1] -3.230954
```

So 95% confidence interval is (-3.230954, 12.03095)

### (i)

Using the same formula above, we get

```r
C <- rbind(c(0,0,0,-1,1,0,0,0,0,0,0,0))
aov <- anova(o.gls2, L = C)
tstat <- sqrt(aov$`F-value`)

contrast <- drop(C %*% o.gls2$coefficients)
```

```r
margin <- qt(0.025, 48) / tstat * sign(contrast)
contrast * (1 - margin)
```

```
## [1] 2.566787
```

```r
contrast * (1 + margin)
```

```
## [1] -3.766787
```

So 95% confidence interval is (-3.766787, 2.566787)

# Problem 3

```r
ExamScores <- read.csv("https://dnett.github.io/S510/ExamScores.txt", sep = "\t")
ExamScores$student <- factor(ExamScores$student)
ExamScores$exam <- factor(ExamScores$exam)
score <- ExamScores$score
```

## (a)

We can find REML estimates using *getVarCov* function in *nlme* package.

```r
o.lme <- lme(score ~ 0 + exam, random = ~ 1 | student,
    weights = varIdent(form = ~ 1 | exam), data = ExamScores,
    method = "REML")

getVarCov(o.lme, type = "random.effects")[[1]]
```

```
## [1] 180.5058
```

```r
CondCov <- getVarCov(o.lme, type = "conditional", individuals = 2)
diag(unname(CondCov[[1]]))
```

```
## [1] 62.94564 60.29974 17.17003
```

Hence, $\sigma_s^2 = 180.5058, \sigma_1^2 = 62.94564, \sigma_2^2 = 60.29974, \sigma_3^3 = 17.17003$.

## (b)

```r
mu <- unname(fixef(o.lme))
Sigma <- unname(getVarCov(o.lme, type = "marginal", individuals = 2)[[1]])
u <- unname(unlist(ranef(o.lme)))

mu[3] + u[1]
```

```
## [1] 83.73545
```

EBLUP of student 1's exam 3 score is 83.73545.

## (c)

The conditional expectation can be expresses as

$$E(y_{13} \mid y_{11}, y_{12}) = \mu_3 + \begin{pmatrix} \sigma_s^2 & \sigma_s^2 \end{pmatrix} \begin{pmatrix} \sigma_s^2 + \sigma_1^2 & \sigma_s^2 \\ \sigma_s^2 & \sigma_s^2 + \sigma_2^2 \end{pmatrix}^{-1} \begin{pmatrix} y_{11} - \mu_1 \\ y_{12} - \mu_2 \end{pmatrix}$$

### (d)

We can compute an estimate of the conditional expectation by replacing parameters with their estimates using the above formula

```
CondE <- mu[3] + Sigma[3, -3, drop = F] %*%
  solve(Sigma[-3, -3], c(score[1] - mu[1], score[2] - mu[2]))
drop(CondE)
```

```
## [1] 83.73545
```

### (e)

General positive definite structure for variance-covariance matrix can be modeled as

```
o.gls <- gls(score ~ 0 + exam, correlation = corSymm(form =~ 1 | student),
             weights = varIdent(form = ~ 1 | exam), data = ExamScores,
             method = "REML")
```

The conditional expectation can be expressed similarly as

$$E(y_{13} \mid y_{11}, y_{12}) = \mu_3 + \begin{pmatrix} w_{13} & w_{23} \end{pmatrix} \begin{pmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{pmatrix}^{-1} \begin{pmatrix} y_{11} - \mu_1 \\ y_{12} - \mu_2 \end{pmatrix}$$

And the estimated conditional expecation is

```
mu <- unname(coef(o.gls))
Sigma <- unclass(getVarCov(o.gls, type = "marginal", individual = 2))

CondE <- mu[3] + Sigma[3, -3, drop = F] %*%
  solve(Sigma[-3, -3], c(score[1] - mu[1], score[2] - mu[2]))
drop(CondE)
```

```
## [1] 79.9037
```

### (f)

```
ExamScores2 <- ExamScores[-c(1,2),]
res <- tapply(ExamScores2$score, ExamScores2$exam, c)
y <- res[[3]]
X <- cbind(res[[1]], res[[2]])
o <- lm(y ~ X)
```

Coefficients corresponding to the multiple linear regression model are

```
o$coefficients
```

```
## (Intercept)          X1          X2
##  25.6231947   0.2681885   0.4985030
```

So the estimated regression equation is

$$25.6231947 + exam1 \times 0.2681885 + exam2 \times 0.4985030$$

and the a prediction for the exam 3 score based on the estimated regression equation is

```r
t(c(1, score[1], score[2])) %*% o$coefficients
```

```
##          [,1]
## [1,] 79.90372
```