Homework 6

Xinyi Lin

3/28/2019

Problem 1

Variance

$$\begin{aligned} Var(Y_{ij}) &= Var(\mu + b_i + e_{ij}) \\ &= Var(\mu) + Var(b_i) + Var(e_{ij}) \\ &= \sigma_b^2 + \sigma_e^2 \end{aligned}$$

Covariance

$$\mu_{Y_{ij}} = \mu_{Y_{ik}} = E[\mu + b_i + e_{ij}] = \mu$$

As e_{ij} and e_{ik} are independent, $E[e_{ij}e_{ik}] = E[e_{ij}]E[e_{ik}] = 0$.

$$\begin{aligned} Cov(Y_{ij}, Y_{ik}) &= E[(Y_{ij} - \mu_{ij})(Y_{ik} - \mu_{ik})] \\ &= E[(b_i + e_{ij})(b_i + e_{ik})] \\ &= E[b_i^2 + b_i(e_{ij} + e_{ik}) + e_{ij}e_{ik}] \\ &= E[b_i^2] + E[b_i(e_{ij} + e_{ik})] + E[e_{ij}e_{ik}] \\ &= Var(b_i) + [E(b_i)]^2 + b_i \times 0 + 0 \\ &= \sigma_b^2 \end{aligned}$$

Correlation

$$cor(Y_{ij}, Y_{ik}) = Cov(Y_{ij}, Y_{ik}) / \sqrt{var(Y_{ij}), var(Y_{ik})} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}$$

According to the correlation result, we can know that correlation between any two Y_{ij} are constant, so this correspond to compound symmetry covariance patterns.

Problem 2

```r
library(ggplot2)
library(patchwork)
library(tidyverse)
library(nlme)
```

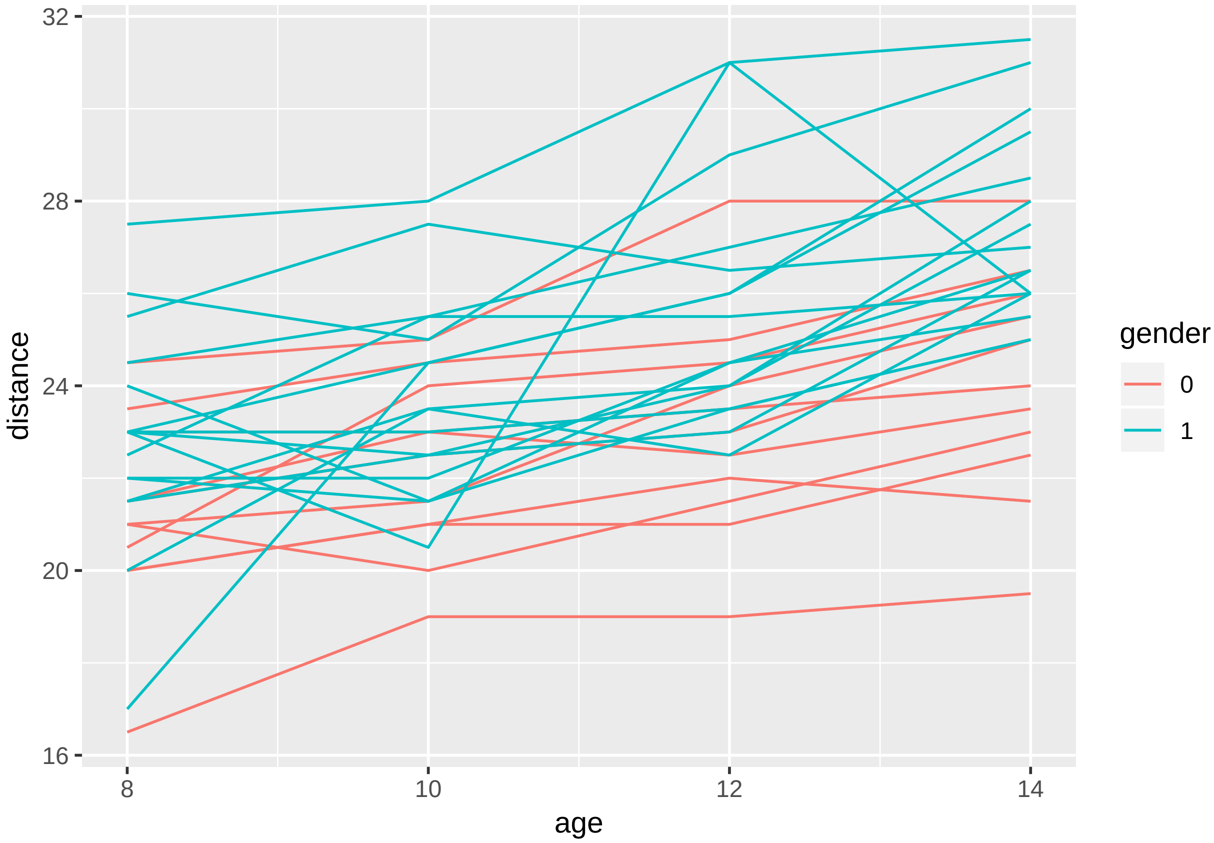
Question 1

Import dental data.

```r
dental_data =
  read.csv("./HW6-dental.txt", sep="") %>%
  janitor::clean_names() %>%
  mutate(gender = as.factor(gender))
```

Make the spaghetti plot.

```r
spag_p = ggplot(dental_data, aes(age, distance, group=child, color=gender)) + geom_line()
spag_p
```



Question 2

$$
\begin{aligned}
var(Yij) &= var[\beta_0 + a_i + b_0 * I(sex_i = 0) + b_1 * I(sex_i = 1) + \beta_1 * age_{ij} + e_{ij}] \\
&= var(a_i) + var(b_k) + var(e_{ij}) \\
&= \sigma_a^2 + \sigma_b^2 + \sigma_e^2
\end{aligned}
$$

$$E(Y_{ij}) = \beta_0 + \beta_1 * age_{ij}$$

For the same individual in different ages:

$$
\begin{aligned}
cov(Y_{ij}, Y_{im}) &= E[(a_i + b_k + e_{ij})(a_i + b_k + e_{im})] \\
&= E[a_i^2 + a_i b_k + a_i e_{im} + a_i b_k + b_k^2 + b_k e_{im} + a_i e_{ij} + b_k e_{ij} + e_{ij} e_{im}] \\
&= E(a_i^2) + E(b_k^2) \\
&= \sigma_a^2 + \sigma_b^2
\end{aligned}
$$

For different individuals:

1) same gender

When measured in different ages,

$$
\begin{aligned}
cov(Y_{hj}, Y_{im}) &= E[(a_h + b_k + e_{hj})(a_i + b_k + e_{im})] \\
&= E[a_i a_h + a_h b_k + a_h e_{im} + a_i b_k + b_k^2 + b_k e_{im} + a_i e_{hj} + b_k e_{hj} + e_{hj} e_{im}] \\
&= E(b_k^2) \\
&= \sigma_b^2
\end{aligned}
$$

When measeured in the same age,

$$
\begin{aligned}
cov(Y_{hj}, Y_{ij}) &= E[(a_h + b_k + e_{hj})(a_i + b_k + e_{ij})] \\
&= E(b_k^2) \\
&= \sigma_b^2
\end{aligned}
$$

2) different genders

$$
\begin{aligned}
cov(Y_{hj}, Y_{im}) &= E[(a_h + b_0 + e_{hj})(a_i + b_1 + e_{im})] \\
&= 0
\end{aligned}
$$

So for same individuals:

$$
M_s = \begin{bmatrix}
\sigma_a^2 + \sigma_b^2 + \sigma_e^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 \\
\sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 + \sigma_e^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 \\
\sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 + \sigma_e^2 & \sigma_a^2 + \sigma_b^2 \\
\sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 + \sigma_e^2
\end{bmatrix}
$$

For different individuals:

$$
M_d = \begin{bmatrix}
\sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\
\sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\
\sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2 \\
\sigma_b^2 & \sigma_b^2 & \sigma_b^2 & \sigma_b^2
\end{bmatrix}
$$

For all grils:

$$
N_g = \begin{bmatrix}
M_s & M_d & \dots & M_d \\
M_d & M_s & \dots & M_d \\
\vdots & \vdots & \ddots & \vdots \\
M_d & M_d & \dots & M_s
\end{bmatrix}_{11 \times 11}
$$

For all boys:

$$
N_b = \begin{bmatrix}
M_s & M_d & \dots & M_d \\
M_d & M_s & \dots & M_d \\
\vdots & \vdots & \ddots & \vdots \\
M_d & M_d & \dots & M_s
\end{bmatrix}_{16 \times 16}
$$

So the model in marginal form is:

$$E(Y_{ij}) = \beta_0 + a_i + b_0 * I(sex_i = 0) + b_1 * I(sex_i = 1) + \beta_1 * age_{ij} + e_{ij}$$

$$var(Y_i) = \begin{bmatrix} N_g & 0 \\ 0 & N_b \end{bmatrix}_{27\times 27}$$

Question 3

Compound symmetry covariance

```
comsym <- gls(distance~gender+age,dental_data,
              correlation=corCompSymm(form = ~ 1|child), method="REML")
summary(comsym)
```

```
## Generalized least squares fit by REML
##   Model: distance ~ gender + age
##   Data: dental_data
##          AIC      BIC    logLik
##   447.5125 460.7823 -218.7563
##
## Correlation Structure: Compound symmetry
##  Formula: ~1 | child
##  Parameter estimate(s):
##       Rho
## 0.6144914
##
## Coefficients:
##                 Value Std.Error   t-value p-value
## (Intercept) 15.385690 0.8959848 17.171820  0.0000
## gender1      2.321023 0.7614169  3.048294  0.0029
## age          0.660185 0.0616059 10.716263  0.0000
##
##  Correlation:
##         (Intr) gendr1
## gender1 -0.504
## age     -0.756  0.000
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -2.59712955 -0.64544226 -0.02540005  0.51680604  2.32947531
##
## Residual standard error: 2.305697
## Degrees of freedom: 108 total; 105 residual
```

According to the result given by R, the model with compound symmetry covariance is:

$$E(y_{ij}) = 15.396 + 2.151 * sex_i + 0.664 * age_{ij}$$

$Var(Y_i) =$ is:

```
corMatrix(comsym$modelStruct$corStruct)[[1]]*(comsym$sigma)^2
```

```
##          [,1]     [,2]     [,3]     [,4]
## [1,] 5.316240 3.266784 3.266784 3.266784
```

```
## [2,] 3.266784 5.316240 3.266784 3.266784
## [3,] 3.266784 3.266784 5.316240 3.266784
## [4,] 3.266784 3.266784 3.266784 5.316240
```

Exponential covariance

```r
expo <- gls(distance~gender+age ,dental_data,
            correlation=corExp(form = ~ 1 |child), method="REML")
summary(expo)
```

```
## Generalized least squares fit by REML
##   Model: distance ~ gender + age
##   Data: dental_data
##        AIC      BIC    logLik
##   455.4483 468.7181 -222.7241
##
## Correlation Structure: Exponential spatial correlation
##  Formula: ~1 | child
##  Parameter estimate(s):
##    range
## 2.133938
##
## Coefficients:
##                 Value Std.Error   t-value p-value
## (Intercept) 15.459995 1.1309319 13.670138   0e+00
## gender1      2.418714 0.6933441  3.488476   7e-04
## age          0.652960 0.0906420  7.203723   0e+00
##
##  Correlation:
##         (Intr) gendr1
## gender1 -0.363
## age     -0.882  0.000
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -2.65148774 -0.69592567 -0.06214639  0.48659340  2.29666951
##
## Residual standard error: 2.301495
## Degrees of freedom: 108 total; 105 residual
```

According to the result given by R, the model with exponential covariance is:

$$E(y_{ij}) = 15.460 + 2.419 * sex_i + 0.653 * age_{ij}$$

$Var(Y_i) =$ is:

```r
corMatrix(expo$modelStruct$corStruct)[[1]]*(expo$sigma)^2
```

```
##          [,1]     [,2]     [,3]     [,4]
## [1,] 5.296881 3.315144 2.074839 1.298574
## [2,] 3.315144 5.296881 3.315144 2.074839
## [3,] 2.074839 3.315144 5.296881 3.315144
## [4,] 1.298574 2.074839 3.315144 5.296881
```

Autoregressive covariance

```
auto1 <- gls(distance~gender+age ,dental_data,
             correlation=corAR1(form = ~ 1 |child), method="REML")
summary(auto1)
```

```
## Generalized least squares fit by REML
##   Model: distance ~ gender + age
##   Data: dental_data
##        AIC      BIC    logLik
##   455.4483 468.7181 -222.7241
##
## Correlation Structure: AR(1)
##  Formula: ~1 | child
##  Parameter estimate(s):
##       Phi
## 0.6258671
##
## Coefficients:
##                  Value Std.Error   t-value p-value
## (Intercept) 15.459995 1.1309319 13.670138   0e+00
## gender1      2.418714 0.6933441  3.488476   7e-04
## age          0.652960 0.0906420  7.203723   0e+00
##
##  Correlation:
##         (Intr) gendr1
## gender1 -0.363
## age     -0.882  0.000
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -2.65148770 -0.69592566 -0.06214639  0.48659339  2.29666947
##
## Residual standard error: 2.301495
## Degrees of freedom: 108 total; 105 residual
```

According to the result given by R, the model with autoregressive covariance is:

$$E(y_{ij}) = 15.460 + 2.419 * sex_i + 0.653 * age_{ij}$$

$Var(Y_i) =$ is:

```
corMatrix(auto1$modelStruct$corStruct)[[1]]*(auto1$sigma)^2
```

```
##          [,1]     [,2]     [,3]     [,4]
## [1,] 5.296881 3.315144 2.074840 1.298574
## [2,] 3.315144 5.296881 3.315144 2.074840
## [3,] 2.074840 3.315144 5.296881 3.315144
## [4,] 1.298574 2.074840 3.315144 5.296881
```

Conclusion:

According to results given by R, models with exponential covariance and autoregressive covariance have same coefficient parameter estimates and covariance estimates as model with exponential covariance is a generalization of autoregressive pattern and in dental data, measurements made at equal intervals of time.

The model with compound symmetry covariance have different coefficient parameter estimates and covariance estimates with other two models but the parameter estimates are closed. Since models with compound symmetry assume correlation between any two visits are constant while other two models do not have such assumption, their covariance estimates are different.