P8131_hw6

Shan Jiang

Problem 1

(1) Variance of Y_{ij}.

Since $b_i \sim N(0, \sigma_b^2), e_{ij} \sim N(0, \sigma_e^2)$, and b_i, e_{ij} are independent, so we have:

$$Var(Y_{ij}) = Var(\mu + b_i + e_{ij})$$

As b_i, e_{ij} are independent, $Cov(b_i, e_{ij}) = 0$, so $Var(Y_{ij}) = Var(\mu + b_i + e_{ij}) = Var(b_i) + Var(e_{ij}) = \sigma_b^2 + \sigma_e^2$.

(2) Covariance between any two values Y_ij and Y_ik, j $j \neq k$

$$E(Y_{ij}) = E(\mu + b_i + e_{ij}) = \mu + 0 + 0 = \mu$$
$$E(Y_{ik}) = E(\mu + b_i + e_{ik}) = \mu + 0 + 0 = \mu$$

$$Cov(Y_{ij}, Y_{ik}) = E[(Y_{ij} - \bar{Y})(Y_{ik} - \bar{Y})] = E[(b_i + e_{ij})(b_i + e_{ik})] = E[(b_i)^2 + e_{ij} * e_{ik} + b_i * (e_{ij} + e_{ik})]$$

$$= E[(b_i)^2] + E(e_{ij}) * E(e_{ik}) + 0 = E[(b_i)^2] = Var(b_i) + [E(b_i)]^2 = \sigma_b^2$$

(3) correlation between any two values Y_ij and Y_ik

$$Var(Y_{ij}) = Var(b_i + e_{ij}) = \sigma_b^2 + \sigma_e^2$$
$$Var(Y_{ik}) = Var(b_i + e_{ik}) = \sigma_b^2 + \sigma_e^2$$

The correlation coefficient:

$$Corr(Y_{ij}, Y_{ik}) = \frac{Cov(Y_{ij}, Y_{ik})}{\sigma Y_{ij} * \sigma Y_{ik}} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}$$

(4) Covariance Pattern

This one has a constant ovariance between any two values Y_{ij} and Y_{ik} and the correlation between any two Y also being constant, so this corresponds to Compound Symmetry covariance pattern.

Problem 2

1). Speghetti plot

1. read in data

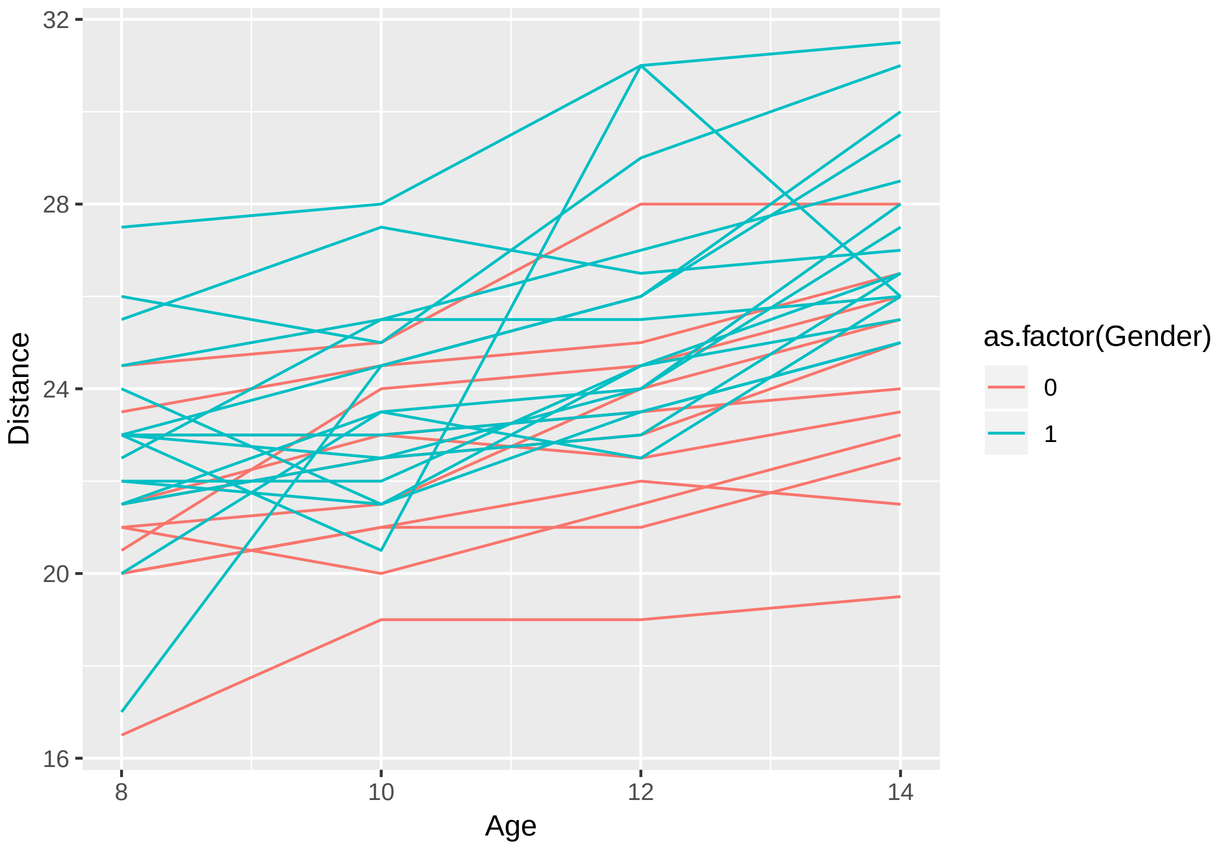
```r
# import in data
distance <- read.delim("./HW6-dental.txt",sep = '', header = T)
head(distance)
```

```
##   Index Child Age Distance Gender
## 1     1     1   8     21.0      0
## 2     2     1  10     20.0      0
## 3     3     1  12     21.5      0
## 4     4     1  14     23.0      0
## 5     5     2   8     21.0      0
## 6     6     2  10     21.5      0
```

2. group by color

```r
p = ggplot(distance, aes(x = Age, y = Distance, group = Child)) +
  geom_line(aes(color = as.factor(Gender)))

print(p)
```



2). Marginal model form

Marginal Models: There are several cases for different covariances:

1) Same gender

(1) Same individual at different ages

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

(2) For different individuals: same gender

When measured in different ages,

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

When measured at 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

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

Finaly, the model in marginal form is:

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

a_i, b_k, e_{ij} are mutually independent,

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

where $N_g$and N_b are matrix for different genders.

3). Model fitting

(1) Compound symmetry covariance with RMLE fit

```
## Compound symmetry
comsym <- gls(Distance ~ Gender + Age, distance, correlation = corCompSymm(form = ~ 1 | Child),
   method = "REML")
summary(comsym)
```

```
## Generalized least squares fit by REML
##   Model: Distance ~ Gender + Age
##   Data: distance
##        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
## Gender       2.321023 0.7614169  3.048294  0.0029
## Age          0.660185 0.0616059 10.716263  0.0000
##
##  Correlation:
##        (Intr) Gender
## Gender -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
```

```
corMatrix(comsym$modelStruct$corStruct)[[1]]
```

```
##           [,1]      [,2]      [,3]      [,4]
## [1,] 1.0000000 0.6144914 0.6144914 0.6144914
## [2,] 0.6144914 1.0000000 0.6144914 0.6144914
## [3,] 0.6144914 0.6144914 1.0000000 0.6144914
## [4,] 0.6144914 0.6144914 0.6144914 1.0000000
```

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

```
##          [,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
```

$$E(Y) = 15.385 + 2.321 * Gender - 0.6601 * Age$$

Compound symmetry:

- 2 covariance parameters, σ^2 and ρ, regardless of n;
- constant variance, and correlation does not decay.

(2) Exponential covariance

```
exp1 <- gls(Distance ~ Gender + Age, distance, correlation = corExp(form = ~ 1 |Child), method = "REML")
summary(exp1)
```

```
## Generalized least squares fit by REML
##   Model: Distance ~ Gender + Age
##   Data: distance
##        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
## Gender       2.418714 0.6933441  3.488476   7e-04
## Age          0.652960 0.0906420  7.203723   0e+00
##
##  Correlation:
##        (Intr) Gender
## Gender -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
```

```r
corMatrix(exp1$modelStruct$corStruct)[[1]]
```

```
##           [,1]      [,2]      [,3]      [,4]
## [1,] 1.0000000 0.6258671 0.3917097 0.2451582
## [2,] 0.6258671 1.0000000 0.6258671 0.3917097
## [3,] 0.3917097 0.6258671 1.0000000 0.6258671
## [4,] 0.2451582 0.3917097 0.6258671 1.0000000
```

$$E(Y_{ij}) = 15.4599 + 2.418714 * Gender - 0.6529 * Age$$

Variance:

```r
exp1cov = corMatrix(exp1$modelStruct$corStruct)[[1]]  * (exp1$sigma)^2
exp1cov
```

```
##          [,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
```

Exponential:

- Unbalanced study, ith subject has response times $t_{i,1}, t_{i,2}, ..., t_{i,ni}$;
- Reduces to AR(1) when all response times are the same.

(3) Autoregrassive covariance

```r
# AR(1)
auto1 <- gls(Distance ~ Gender + Age, distance,
             correlation = corAR1(form = ~ 1 |Child), method = "REML")
summary(auto1)
```

```
## Generalized least squares fit by REML
##   Model: Distance ~ Gender + Age
##   Data: distance
##        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
## Gender       2.418714 0.6933441  3.488476   7e-04
## Age          0.652960 0.0906420  7.203723   0e+00
##
##  Correlation:
##        (Intr) Gender
## Gender -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
```

```
corMatrix(auto1$modelStruct$corStruct)[[1]]
```

```
##           [,1]      [,2]      [,3]      [,4]
## [1,] 1.0000000 0.6258671 0.3917097 0.2451582
## [2,] 0.6258671 1.0000000 0.6258671 0.3917097
## [3,] 0.3917097 0.6258671 1.0000000 0.6258671
## [4,] 0.2451582 0.3917097 0.6258671 1.0000000
```

$$E(Y_{ij}) = 15.4599 + 2.418714 * Gender - 0.6529 * Age$$

Autoregressive (order 1) Variance:

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

```
##          [,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
```

- special case of Toeplitz;
- 2 covariance parameters, σ^2 and ρ, regardless of n;
- constant variance, correlation decays geometrically (exponentially)

Model comparison

parameter estimates

```r
sumtable <- matrix(c(2.045166,  2.418714,   2.418714,
                     0.674651,  0.65296,    0.65296), ncol = 3,byrow = TRUE)
colnames(sumtable) <- c("Compound", "Exponential", "AutoRegression")
rownames(sumtable) <- c("Gender","Age")
sumtable <- as.table(sumtable)
sumtable
```

```
##        Compound Exponential AutoRegression
## Gender 2.045166    2.418714       2.418714
## Age    0.674651    0.652960       0.652960
```

Conclusion: here coefficient parameter estimates are the same for Exponential and AutoRegression, while different from Compound model. The covariance estimates of compund symmetry model proves that all covariance are the same without weight. Since models with compound symmetry assume correlation between any two visits are constant, their covariance estimates are different.