Assignment6_yf2555

Yihan Feng

2021/3/21

Question 1

(Just in case the image cannot show properly, I also attached a jpg file of question 1 in submission.)

- $Var(Y_{i,j}) = Var(\mu + b_i + e_{i,j})$

$= Var(b_i) + Var(e_{i,j})$

$= \sigma^2_b + \sigma^2_e$

- $Cov(Y_{i,j}, Y_{i,k}) = E[Y_{i,j} \cdot Y_{ik}] - E[Y_{ij}] \cdot E[Y_{ik}]$

$= E[(\mu + b_i + e_{i,j})(\mu + b_i + e_{i,k})] - E[Y_{ij}] \cdot E[Y_{i,k}]$

$= E[\mu^2 + b_i^2 + 2\mu b_i + \mu e_{i,j} + \mu e_{i,k} + b_i \cdot e_{i,j} + b_i \cdot e_{i,k} + e_{i,j} \cdot e_{i,k}) - \mu^2.$

as $b_i \sim N(0, \sigma_b^2)$, $\left(\frac{b_i}{\sigma_b}\right)^2 \sim \chi^2_1$, $\therefore E[b_i]^2 = \sigma^2_b$
$e_{i,j}, e_{i,k}$ are independent, $E(e_{i,j}, e_{i,k}) = E(e_{i,j}) \cdot E(e_{ik})$

$Cov(Y_{i,j}, Y_{i,k}) = \mu^2 + \sigma^2_b - \mu^2$
$= \sigma^2_b$

- $Corr(Y_{i,j}, Y_{i,k}) = \frac{cov(e_{i,j}, e_{i,k})}{\sqrt{Var(Y_{i,j}) \cdot Var(Y_{i,k})}}$

$= \frac{\sigma^2_b}{\sigma^2_b \cdot \sigma^2_e}$

Figure 1: question 1 answer

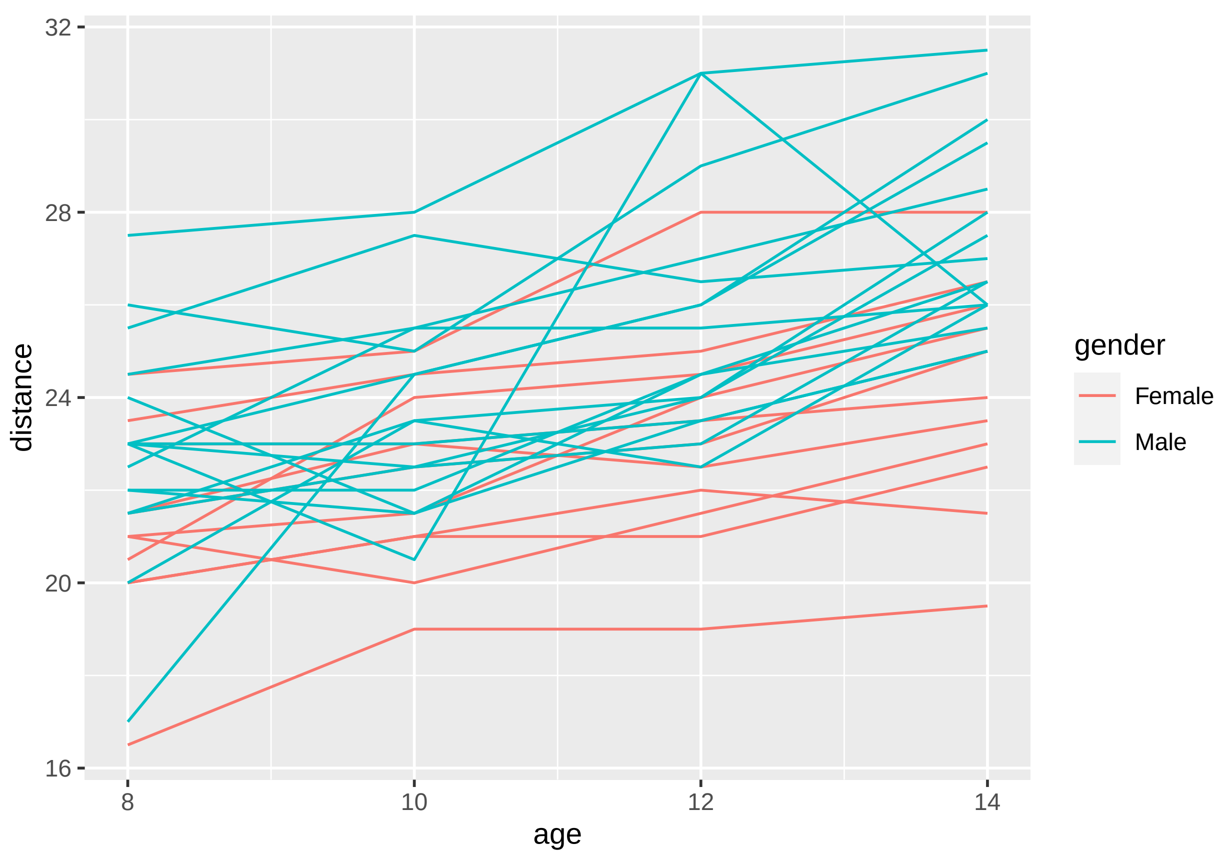
Question 2

A study was conducted involving 27 children, 16 boys and 11 girls. On each child, the distance (mm) from the center of the pituitary to the pterygomaxillary fissure was made at ages 8, 10, 12, and 14. The goal was to study how distance is affected by age and gender. The data file has 5 columns: (1) observation number, (2) child number (1-27), (3) age, (4) distance measurement, and (5) indicator of gender (0 =girl, 1 =boy).

```
dental.df = read.csv("./HW6-dental.txt", sep = "") %>%
  janitor::clean_names() %>%
  mutate(gender = as.factor(gender),
         gender = recode(gender, "0" = "Female", "1" = "Male")) %>%
  dplyr::select(-index)
```

```
ggplot(dental.df, aes(x = age, y = distance, col = gender, group = child)) +
  geom_line()
```

(a) Make a spaghetti plot to infer how distance is affected by age and gender.



(b) (Just in case the image cannot show properly, I also attached a jpg file of question 2-b in submission.)

(b) Mean: $a_i \sim N(0, \sigma_a^2)$

$b_k \sim N(0, \sigma_b^2)$

$e_{i,j} \sim N(0, \sigma_e^2)$

$E[Y_i] = \beta_0 + \beta_1 \cdot age_{i,j}$.

Variance:
$$
\begin{aligned}
Var(Y_{ij}) &= Var(\beta_0 + a_i + b_0 \cdot I + b_1 \cdot I + \beta_1 \cdot age_{ij} + e_{ij}) \\
&= Var(\beta_0) + Var(a_i) + Var(b_0) + Var(b_1) + Var(\beta_1 \cdot age_{ij}) \\
&\quad + Var(e_{i,j}) + 2Cov(a_i, b_k) + 2cov(a_i, e_{ij}) + 2cov(b_k, e_{ij}) \\
&= \emptyset + \sigma_a^2 + \sigma_b^2 + \sigma_e^2 + \emptyset \\
&= \sigma_a^2 + \sigma_b^2 + \sigma_e^2
\end{aligned}
$$

$$
\begin{aligned}
Cov(Y_{ij}, Y_{ik}) &= Cov\{(a_i + b_0 \cdot I + b_1 \cdot I + e_{ij}), (a_i + b_0 \cdot I + b_1 \cdot I + e_{ik})\} \\
&= Cov(a_i + b_i \cdot I + e_{ij},\ a_i + b_i \cdot I + e_{ik}) \\
&= cov(a_i + b_i + e_{ij},\ a_i + b_i + e_{ik}) \\
&= E[a_i + b_i + e_{ij}] \cdot E[a_i + b_i + e_{ik}] \\
&= E[a_i^2] \cdot E[b_i^2] \\
&= Var(a_i) + Var(b_i) \\
&= \sigma_a^2 + \sigma_b^2
\end{aligned}
$$

$$
\therefore Var(Y_i) = \begin{pmatrix}
\sigma_a^2 + \sigma_b^2 + \sigma_e^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 \\
\sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 + \sigma_e^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 \\
\sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 + \sigma_e^2 & \sigma_a^2 + \sigma_b^2 \\
\sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 & \sigma_a^2 + \sigma_b^2 + \sigma_e^2
\end{pmatrix}
$$

$$
Corr(Y_i) = \begin{pmatrix}
1 & \frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2} & \frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2} & \frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2} \\
\frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2} & 1 & \frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2} & \frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2} \\
\frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2} & \frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2} & 1 & \frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2} \\
\frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2} & \frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2} & \frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2} & 1
\end{pmatrix}
$$

Figure 2: question 2-b answer

(c) Fit a model with (a) compound symmetry covariance, (b) exponential covariance, (c) autoregressive covariance. Compare the coefficient parameter estimates and the covariance estimates.

1. Compound symmetry covariance:

```
comsym <- gls(distance ~ age + gender, dental.df, correlation = corCompSymm(form = ~ 1| child), method =
summary(comsym)
```

```
## Generalized least squares fit by REML
##   Model: distance ~ age + gender
##   Data: dental.df
##        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
## age          0.660185 0.0616059 10.716263  0.0000
## genderMale   2.321023 0.7614169  3.048294  0.0029
##
##  Correlation:
##            (Intr) age
## age        -0.756
## genderMale -0.504  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
```

2. exponential covariance:

```
exp.fit <- gls(distance ~ age + gender, dental.df, correlation = corExp(form = ~ 1 |child), method = "RE
summary(exp.fit)
```

```
## Generalized least squares fit by REML
##   Model: distance ~ age + gender
##   Data: dental.df
##        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
## age          0.652960 0.0906420  7.203723   0e+00
## genderMale   2.418714 0.6933441  3.488476   7e-04
##
##  Correlation:
##            (Intr) age
## age        -0.882
## genderMale -0.363  0.000
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -2.65148775 -0.69592567 -0.06214639  0.48659340  2.29666951
##
## Residual standard error: 2.301495
## Degrees of freedom: 108 total; 105 residual
```

3. Autoregressive covariance:

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

```
## Generalized least squares fit by REML
##   Model: distance ~ age + gender
##   Data: dental.df
##        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
## age          0.652960 0.0906420  7.203723   0e+00
## genderMale   2.418714 0.6933441  3.488476   7e-04
##
##  Correlation:
##            (Intr) age
## age        -0.882
## genderMale -0.363  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
```

```r
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
```

Compare coefficient parameter estimates:

```
coeff = rbind(comsym$coefficients, exp.fit$coefficients, auto1$coefficients)
rownames(coeff) = c("Comp Symmetry", "Exponential", "Autoregressive")
coeff %>%
  knitr::kable()
```

	(Intercept)	age	genderMale
Comp Symmetry	15.38569	0.6601852	2.321023
Exponential	15.45999	0.6529597	2.418714
Autoregressive	15.45999	0.6529597	2.418714

According to the table above, the three methods give similar coefficient parameters.

Compare covariance estimates:

```
getVarCov(comsym)
```

```
## Marginal variance covariance matrix
##        [,1]   [,2]   [,3]   [,4]
## [1,] 5.3162 3.2668 3.2668 3.2668
## [2,] 3.2668 5.3162 3.2668 3.2668
## [3,] 3.2668 3.2668 5.3162 3.2668
## [4,] 3.2668 3.2668 3.2668 5.3162
##   Standard Deviations: 2.3057 2.3057 2.3057 2.3057
```

```
getVarCov(exp.fit)
```

```
## Marginal variance covariance matrix
##        [,1]   [,2]   [,3]   [,4]
## [1,] 5.2969 3.3151 2.0748 1.2986
## [2,] 3.3151 5.2969 3.3151 2.0748
## [3,] 2.0748 3.3151 5.2969 3.3151
## [4,] 1.2986 2.0748 3.3151 5.2969
##   Standard Deviations: 2.3015 2.3015 2.3015 2.3015
```

```
getVarCov(auto1)
```

```
## Marginal variance covariance matrix
##        [,1]   [,2]   [,3]   [,4]
## [1,] 5.2969 3.3151 2.0748 1.2986
## [2,] 3.3151 5.2969 3.3151 2.0748
## [3,] 2.0748 3.3151 5.2969 3.3151
## [4,] 1.2986 2.0748 3.3151 5.2969
##   Standard Deviations: 2.3015 2.3015 2.3015 2.3015
```

According to the tables above, the exponential and autoregressive method have similar covariance matrices.