Homework 6

Mari Sanders

2025-04-09

```r
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(nlme)
```

```
##
## Attaching package: 'nlme'
##
## The following object is masked from 'package:dplyr':
##
##     collapse
```

Problem 1

$$Var(Y_{ij}) = Var(\mu) + Var(b_i) + Var(e_{ij})$$
$$= 0 + \sigma_b^2 + \sigma_e^2$$
$$\Rightarrow Var(Y_{ij}) = \sigma_b^2 + \sigma_e^2$$

$$Cov(Y_{ij}, Y_{ik}) = Cov(\mu + b_i + e_{ij}, \mu + b_i + e_{ik})$$
$$= Cov(b_i, b_i) + \underbrace{Cov(b_i, e_{ik})}_{\text{independent}} + \underbrace{Cov(e_{ij}, b_i)}_{\text{independent}} + \underbrace{Cov(e_{ij}, e_{ik})}_{\text{independent}}$$

$$Cov(Y_{ij}, Y_{ik}) = Cov(b_i, b_i) = Var(b_i) = \sigma_b^2$$

$$Corr(Y_{ij}, Y_{ik}) = \frac{Cov(Y_{ij}, Y_{ik})}{\sqrt{Var(Y_{ij}) \cdot Var(Y_{ik})}} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}$$

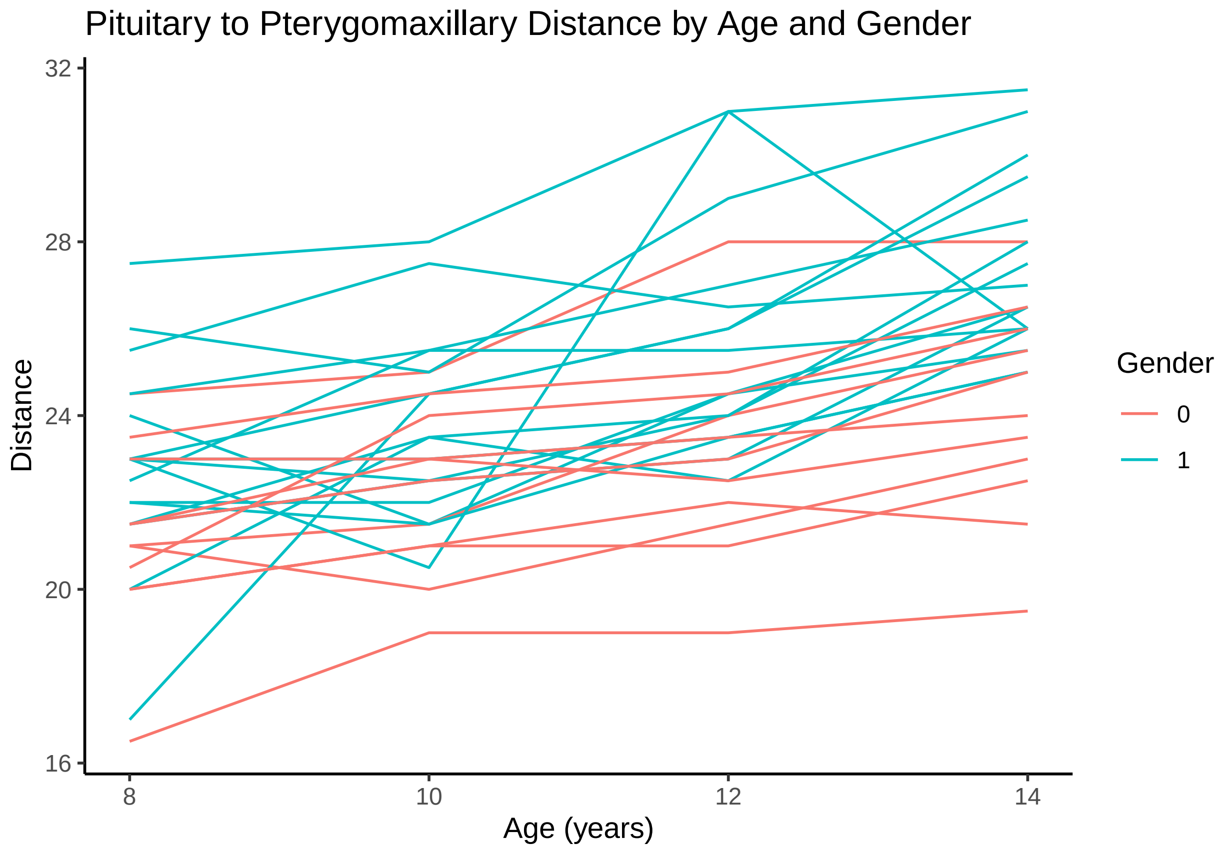
Structure is a compound symmetry covariance structure because it has the same correlation regardless of the time points in each group.

Problem 2

(a)

```
data <- read.table("HW6-dental.txt") %>% drop_na() %>% as.data.frame() %>% slice(-1) %>% rename(Index =

ggplot(data) +
  geom_path(aes(x = Age, y = Distance, group = Child, color = factor(Gender, levels = c(0, 1)))) +
  ggtitle("Score change after intervention") +
  theme_classic() +
  labs(
    title = "Pituitary to Pterygomaxillary Distance by Age and Gender",
    x = "Age (years)",
    y = "Distance",
    color = "Gender"
  )
```



Pituitary to Pterygomaxillary Distance by Age and Gender
Distance
32
28
24
20
16
8
10
12
14
Age (years)
Gender
0
1

(b)

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

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

$$E(a_i) = 0$$
$$E(b_i) = 0$$
$$E(e_{ij}) = 0$$

$$\Rightarrow E(y_{ij}) = \beta_0 + \beta_1 age_{ij}$$

$$Var(y_{ij}) = Var(\beta_0 + a_i + b_0 I(sex_i=0) + b_1 I(sex_i=1) + \beta_1 age_{ij} + e_{ij})$$
$$= Var(\beta_0) + Var(a_i) + Var(b_0 I(sex_i=0)) + Var(b_1 I(sex_i=1))$$
$$+ Var(\beta_1 age_{ij}) + Var(e_{ij})$$
$$= \sigma_a^2 + \sigma_b^2 + \sigma_e^2$$

$$var(y_{ij}) = \sigma_a^2 + \sigma_b^2 + \sigma_e^2$$

$$corr(y_{ij}, y_{iu}) = \frac{Cov(y_{ij}, y_{iu})}{\sqrt{Var(y_{ij})Var(y_{iu})}}$$

$$cov(y_{ij}, y_{iu}) = cov(\beta_0 + a_i + b_0 I(sex_i=0) + b_1 I(sex_i=1) + \beta_1 age_{ij} + e_{ij},$$
$$\beta_0 + a_i + b_0(I sex_i=0) + b_1(I sex=1) + \beta_1 age_{iu} + e_{iu})$$

$$= cov(a_i, a_i) + cov(a_i, b_u) + cov(a_i, e_{iu}) + cov(b_u, a_i) + cov(b_u, b_u)$$
$$+ cov(b_u, e_{iu}) + cov(e_{ij}, b_u) + cov(e_{ij}, e_{iu})$$
$$= Var(a_i) + cov(b_u, b_u) = \sigma_a^2 + \sigma_b^2$$

$$corr(y_{ij}, y_{iu}) = \frac{\sigma_a^2 + \sigma_b^2}{\sqrt{(\sigma_a^2+\sigma_b^2+\sigma_e^2)^2}} = \frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2+\sigma_b^2+\sigma_e^2}$$

Compound symmetry covariance structure because covariance is the same among those in the same group

(c)

```
fit_cs <- lme(Distance ~ Gender + Age, random = ~1 | Child, data = data,
              correlation = corCompSymm(form = ~1 | Child),  method='REML')
```

```r
summary(fit_cs)
```

```
## Linear mixed-effects model fit by REML
##   Data: data
##        AIC      BIC    logLik
##   449.5125 465.4363 -218.7563
##
## Random effects:
##  Formula: ~1 | Child
##         (Intercept) Residual
## StdDev:    1.807425 1.431592
##
## Correlation Structure: Compound symmetry
##  Formula: ~1 | Child
##  Parameter estimate(s):
## Rho
##   0
## Fixed effects:  Distance ~ Gender + Age
##                  Value Std.Error DF   t-value p-value
## (Intercept) 15.385690 0.8959848 80 17.171820  0.0000
## Gender1      2.321023 0.7614168 25  3.048294  0.0054
## Age          0.660185 0.0616059 80 10.716263  0.0000
##  Correlation:
##         (Intr) Gendr1
## Gender1 -0.504
## Age     -0.756  0.000
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.74889609 -0.55034466 -0.02516628  0.45341781  3.65746539
##
## Number of Observations: 108
## Number of Groups: 27
```

```r
fit_exp <- lme(Distance ~ Gender + Age, random = ~1 | Child, data = data,
                 correlation = corExp(form = ~ Age | Child),  method='REML')
summary(fit_exp)
```

```
## Linear mixed-effects model fit by REML
##   Data: data
##        AIC      BIC    logLik
##   449.3968 465.3206 -218.6984
##
## Random effects:
##  Formula: ~1 | Child
##         (Intercept) Residual
## StdDev:    1.788899 1.454494
##
## Correlation Structure: Exponential spatial correlation
##  Formula: ~Age | Child
##  Parameter estimate(s):
##     range
```

```
## 0.7045117
## Fixed effects:  Distance ~ Gender + Age
##                 Value Std.Error DF   t-value p-value
## (Intercept) 15.393931 0.9109499 80 16.898768  0.0000
## Gender1      2.327485 0.7611852 25  3.057711  0.0053
## Age          0.659405 0.0634074 80 10.399499  0.0000
##  Correlation:
##         (Intr) Gendr1
## Gender1 -0.495
## Age     -0.766  0.000
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -3.683026667 -0.540915318 -0.008097445  0.461167542  3.612579065
##
## Number of Observations: 108
## Number of Groups: 27
```

```
fit_ar1 <- lme(Distance ~ Gender + Age, random = ~1 | Child, data = data,
                 correlation = corAR1(form = ~ Age | Child), method='REML')
summary(fit_ar1)
```

```
## Linear mixed-effects model fit by REML
##   Data: data
##        AIC      BIC    logLik
##   449.5125 465.4363 -218.7563
##
## Random effects:
##  Formula: ~1 | Child
##         (Intercept) Residual
## StdDev:    1.807425 1.431592
##
## Correlation Structure: ARMA(1,0)
##  Formula: ~Age | Child
##  Parameter estimate(s):
## Phi1
##    0
## Fixed effects:  Distance ~ Gender + Age
##                 Value Std.Error DF   t-value p-value
## (Intercept) 15.385690 0.8959848 80 17.171820  0.0000
## Gender1      2.321023 0.7614168 25  3.048294  0.0054
## Age          0.660185 0.0616059 80 10.716263  0.0000
##  Correlation:
##         (Intr) Gendr1
## Gender1 -0.504
## Age     -0.756  0.000
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.74889609 -0.55034466 -0.02516628  0.45341781  3.65746539
##
## Number of Observations: 108
## Number of Groups: 27
```

```r
sm_cs <- summary(fit_cs)$tTable
sm_ar1 <- summary(fit_ar1)$tTable
sm_exp <- summary(fit_exp)$tTable

bind_rows(
  data.frame(
    Model = "Compound Symmetry",
    Intercept = sm_cs["(Intercept)", "Value"],
    Gender = sm_cs["Gender1", "Value"],
    Age = sm_cs["Age", "Value"]
  ),
  data.frame(
    Model = "AR(1)",
    Intercept = sm_ar1["(Intercept)", "Value"],
    Gender = sm_ar1["Gender1", "Value"],
    Age = sm_ar1["Age", "Value"]
  ),
  data.frame(
    Model = "Exponential",
    Intercept = sm_exp["(Intercept)", "Value"],
    Gender = sm_exp["Gender1", "Value"],
    Age = sm_exp["Age", "Value"]
  )
)
```

```
##               Model Intercept   Gender       Age
## 1 Compound Symmetry  15.38569 2.321023 0.6601852
## 2             AR(1)  15.38569 2.321023 0.6601852
## 3       Exponential  15.39393 2.327485 0.6594049
```

The intercept, gender and age coefficients for the compound symmetry and autoregressive covariance are the same. For the exponential covariance model, the intercept, gender, and age are similar to the others. Boys consistently have about a 2.3 mm greater distance than girls. Distance increases by about 0.65–0.66 mm per year of age.

```r
summary(fit_cs)$corFixed %>% knitr::kable()
```

	(Intercept)	Gender1	Age
(Intercept)	1.0000000	-0.5035911	-0.7563354
Gender1	-0.5035911	1.0000000	0.0000000
Age	-0.7563354	0.0000000	1.0000000

```r
summary(fit_ar1)$corFixed %>% knitr::kable()
```

	(Intercept)	Gender1	Age
(Intercept)	1.0000000	-0.5035911	-0.7563354
Gender1	-0.5035911	1.0000000	0.0000000
Age	-0.7563354	0.0000000	1.0000000

```r
summary(fit_exp)$corFixed %>% knitr::kable()
```

	(Intercept)	Gender1	Age
(Intercept)	1.0000000	-0.4951674	-0.7656635
Gender1	-0.4951674	1.0000000	0.0000000
Age	-0.7656635	0.0000000	1.0000000

```r
data.frame(
  Model = c("Compound Symmetry", "AR(1)", "Exponential"),
  AIC = c(
    as.numeric(summary(fit_cs)$AIC),
    as.numeric(summary(fit_ar1)$AIC),
    as.numeric(summary(fit_exp)$AIC)
  )
) %>% knitr::kable()
```

Model	AIC
Compound Symmetry	449.5125
AR(1)	449.5125
Exponential	449.3968

Compound symmetry assumes constant correlation across all time points: $\rho - 1.0986123$. This means that these might be weak intra-subject correlation because the value is a fairly large negative value. Autoregressive covariance fit failed to detect autocorrelation: $\phi = 0$ implies no correlation over time. Exponential has a $\rho = -0.3502403$, meaning that there may be a decaying correlation with increasing time difference, but it is not a very strong correlation. Exponential seems to be the best model with the lowest AIC, but only slightly. It is marginally better at explaining the data than constant or no correlation.

```r
fit_cs_comp <- lme(Distance ~ Gender + Age, random = ~1 | Child, data = data,
                   correlation = corCompSymm(form = ~1 | Child),  method='ML')
summary(fit_cs)
```

```
## Linear mixed-effects model fit by REML
##   Data: data
##        AIC      BIC    logLik
##   449.5125 465.4363 -218.7563
##
## Random effects:
##  Formula: ~1 | Child
##         (Intercept) Residual
## StdDev:    1.807425 1.431592
##
## Correlation Structure: Compound symmetry
##  Formula: ~1 | Child
##  Parameter estimate(s):
## Rho
##   0
## Fixed effects:  Distance ~ Gender + Age
##                 Value Std.Error DF   t-value p-value
```

```
## (Intercept) 15.385690 0.8959848 80 17.171820  0.0000
## Gender1      2.321023 0.7614168 25  3.048294  0.0054
## Age          0.660185 0.0616059 80 10.716263  0.0000
##  Correlation:
##         (Intr) Gendr1
## Gender1 -0.504
## Age     -0.756  0.000
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.74889609 -0.55034466 -0.02516628  0.45341781  3.65746539
##
## Number of Observations: 108
## Number of Groups: 27
```

```
fit_exp_comp <- lme(Distance ~ Gender + Age, random = ~1 | Child, data = data,
                 correlation = corExp(form = ~ Age | Child),  method='ML')
summary(fit_exp)
```

```
## Linear mixed-effects model fit by REML
##   Data: data
##        AIC      BIC    logLik
##   449.3968 465.3206 -218.6984
##
## Random effects:
##  Formula: ~1 | Child
##         (Intercept) Residual
## StdDev:    1.788899 1.454494
##
## Correlation Structure: Exponential spatial correlation
##  Formula: ~Age | Child
##  Parameter estimate(s):
##     range
## 0.7045117
## Fixed effects:  Distance ~ Gender + Age
##                 Value Std.Error DF   t-value p-value
## (Intercept) 15.393931 0.9109499 80 16.898768  0.0000
## Gender1      2.327485 0.7611852 25  3.057711  0.0053
## Age          0.659405 0.0634074 80 10.399499  0.0000
##  Correlation:
##         (Intr) Gendr1
## Gender1 -0.495
## Age     -0.766  0.000
##
## Standardized Within-Group Residuals:
##          Min           Q1          Med           Q3          Max
## -3.683026667 -0.540915318 -0.008097445  0.461167542  3.612579065
##
## Number of Observations: 108
## Number of Groups: 27
```

```
fit_ar1_comp <- lme(Distance ~ Gender + Age, random = ~1 | Child, data = data,
                 correlation = corAR1(form = ~ Age | Child), method='ML')
anova(fit_cs_comp, fit_exp_comp, fit_ar1_comp)
```

```
##              Model df      AIC      BIC    logLik
## fit_cs_comp      1  6 446.8565 462.9493 -217.4282
## fit_exp_comp     2  6 446.7899 462.8827 -217.3949
## fit_ar1_comp     3  6 446.8565 462.9493 -217.4282
```