

Homework 6

Mari Sanders

2025-04-09

```
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
```

```
library(nlme)
```

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

Problem 1

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

$$= 0 + \sigma_b^2 + \sigma_e^2$$

$$\Rightarrow \text{var}(Y_{ij}) = \sigma_b^2 + \sigma_e^2$$

$$\text{cov}(Y_{ij}, Y_{iu}) = \text{cov}(\mu + b_i + e_{ij}, \mu + b_i + e_{iu})$$

$$= \text{cov}(b_i, b_i) + \underbrace{\text{cov}(b_i, e_{iu})}_{\text{independent}} + \underbrace{\text{cov}(e_{ij}, b_i)}_{\text{independent}} + \underbrace{\text{cov}(e_{ij}, e_{iu})}_{\text{independent}}$$

$$\text{cov}(Y_{ij}, Y_{iu}) = \text{cov}(b_i, b_i) = \text{var}(b_i) = \sigma_b^2$$

$$\text{corr}(Y_{ij}, Y_{iu}) = \frac{\text{cov}(Y_{ij}, Y_{iu})}{\sqrt{\text{var}(Y_{ij}) \cdot \text{var}(Y_{iu})}} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_e^2}$$

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"
  )
```



(b)

$$\begin{aligned}Y_{ij} &= \beta_0 + a_i + b_0 I(\text{sex}_i=0) + b_1 I(\text{sex}_i=1) + \beta_1 \text{age}_{ij} + e_{ij} \\E(Y_{ij}) &= E(\beta_0 + a_i + b_0 I(\text{sex}_i=0) + b_1 I(\text{sex}_i=1) + \beta_1 \text{age}_{ij} + e_{ij}) \\E(a_i) &= 0 \\E(b_i) &= 0 \\E(e_{ij}) &= 0 \\\Rightarrow E(Y_{ij}) &= \beta_0 + \beta_1 \text{age}_{ij} \\\text{Var}(Y_{ij}) &= \text{Var}(\beta_0 + a_i + b_0 I(\text{sex}_i=0) + b_1 I(\text{sex}_i=1) + \beta_1 \text{age}_{ij} + e_{ij}) \\&= \text{Var}(\beta_0) + \text{Var}(a_i) + \text{Var}(b_0 I(\text{sex}_i=0)) + \text{Var}(b_1 I(\text{sex}_i=1)) \\&\quad + \text{Var}(\beta_1 \text{age}_{ij}) + \text{Var}(e_{ij}) \\&= \sigma_a^2 + \sigma_b^2 + \sigma_e^2 \\\text{Var}(Y_{ij}) &= \sigma_a^2 + \sigma_b^2 + \sigma_e^2 \\\text{corr}(Y_{ij}, Y_{iu}) &= \frac{\text{Cov}(Y_{ij}, Y_{iu})}{\sqrt{\text{Var}(Y_{ij}) \text{Var}(Y_{iu})}} \\\text{Cov}(Y_{ij}, Y_{iu}) &= \text{Cov}(\beta_0 + a_i + b_0 I(\text{sex}_i=0) + b_1 I(\text{sex}_i=1) + \beta_1 \text{age}_{ij} + e_{ij}, \\&\quad \beta_0 + a_i + b_0 I(\text{sex}_i=0) + b_1 I(\text{sex}_i=1) + \beta_1 \text{age}_{iu} + e_{iu}) \\&= \text{Cov}(a_i, a_i) + \text{Cov}(a_i, b_u) + \text{Cov}(a_i, e_{iu}) + \text{Cov}(b_u, a_i) + \text{Cov}(b_u, b_u) \\&\quad + \text{Cov}(b_u, e_{iu}) + \text{Cov}(e_{ij}, b_u) + \text{Cov}(e_{ij}, e_{iu}) \\&= \text{Var}(a_i) + \text{Cov}(b_u, b_u) = \sigma_a^2 + \sigma_b^2 \\\text{corr}(Y_{ij}, Y_{iu}) &= \frac{\sigma_a^2 + \sigma_b^2}{\sqrt{(\sigma_a^2 + \sigma_b^2 + \sigma_e^2)^2}} = \frac{\sigma_a^2 + \sigma_b^2}{\sigma_a^2 + \sigma_b^2 + \sigma_e^2}\end{aligned}$$

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')
```

```
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 <- 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
```

```

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.

```
vcov(fit_cs) %>% knitr::kable()
```

	(Intercept)	Gender1	Age
(Intercept)	0.8027888	-0.3435589	-0.0417482
Gender1	-0.3435589	0.5797556	0.0000000
Age	-0.0417482	0.0000000	0.0037953

```
vcov(fit_ar1) %>% knitr::kable()
```

	(Intercept)	Gender1	Age
(Intercept)	0.8027888	-0.3435589	-0.0417482
Gender1	-0.3435589	0.5797556	0.0000000
Age	-0.0417482	0.0000000	0.0037953

```
vcov(fit_exp) %>% knitr::kable()
```

	(Intercept)	Gender1	Age
(Intercept)	0.8298297	-0.3433499	-0.0442254
Gender1	-0.3433499	0.5794030	0.0000000
Age	-0.0442254	0.0000000	0.0040205

```
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

In all the models, age and gender are uncorrelated. The variance of the intercept in the compound structure and the ar 1 structure are slightly lower than the variance in the intercept in the exponential model. The variance in gender is similar among all the modes. The variance of age is larger in the exponential model than in the compound or AR 1 model.

```
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

Based on the AIC, BICm and log likelihood, the exponential covariate structure is the best.