Homework 6

Yuki Joyama

2024-03-24

1

$Y_{ij} = \mu + b_i + e_{ij} \quad i=1,\cdots,m, \ j=1\cdots n, \ b_i \sim N(0, \sigma_b^2), e_{ij} \sim N(0, \sigma_e^2)$

b_i and e_{ij} are statistically independent for each i and j — ①

e_{ij} and e_{ik} are statistically independent for any two values $j,k = 1,\cdots n$, $j \neq k$ — ②

? variance of Y_{ij}, covariance and correlation between any values Y_{ij} and Y_{ik} $(j \neq k)$

$var(Y_{ij}) = var(\mu + b_i + e_{ij}) = \sigma_b^2 + \sigma_e^2$ (Given ①)

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

$= Cov(\mu,\mu) + Cov(\mu, b_i) + cov(\mu, e_{ik}) + Cov(b_i, \mu) + Cov(b_i, b_i)$

$+ Cov(b_i, e_{ik}) + Cov(e_{ij}, \mu) + Cov(e_{ij}, b_i) + Cov(e_{ij}, e_{ik})$

$= Var(b_i) = \sigma_b^2$ (Given ① and ②)

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

This is compound symmetric structure.

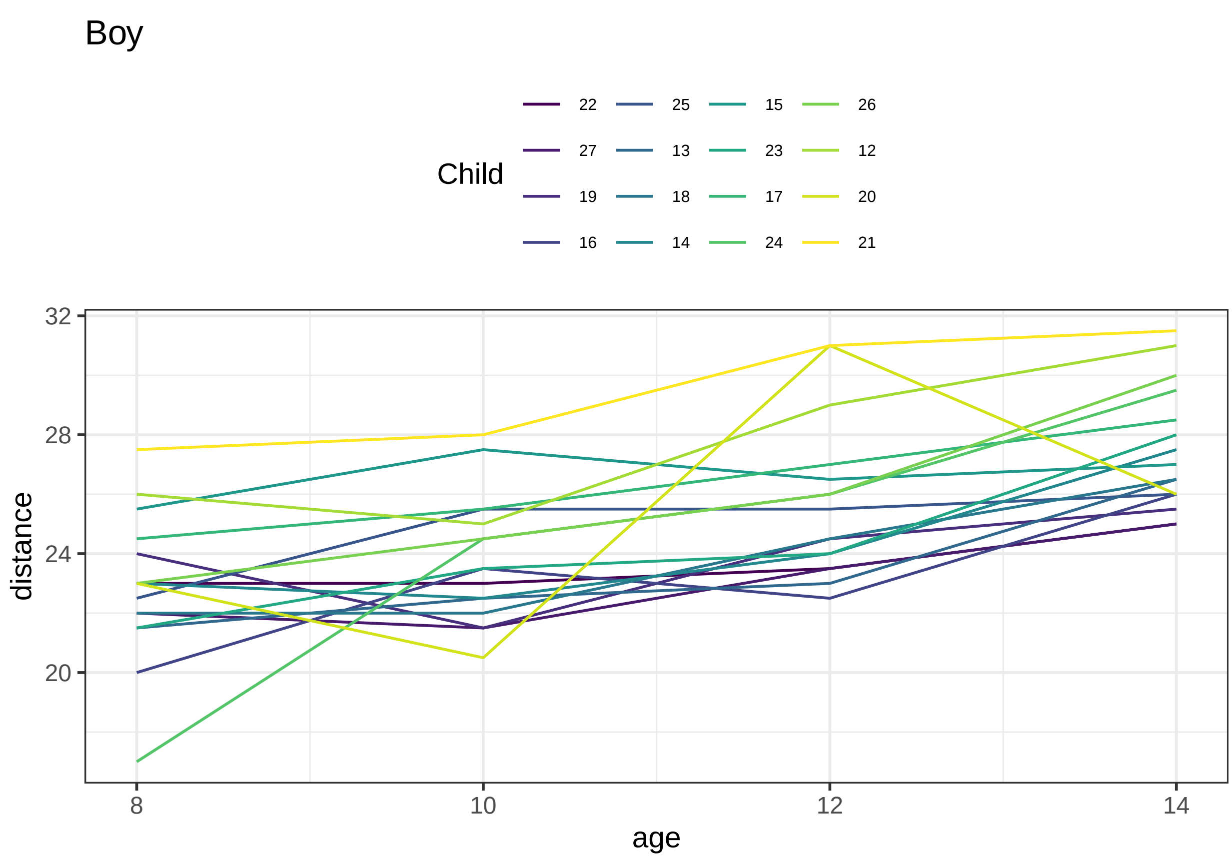
2

```
# data prep
df = read.table("/Users/yukijoyama/Library/CloudStorage/GoogleDrive-jikeyu1995@gmail.com/My Drive/versi
  mutate(
    Gender = as.factor(
      case_when(
        Gender == 0 ~ "girl",
        Gender == 1 ~ "boy"
      )
    )
  )
```

a

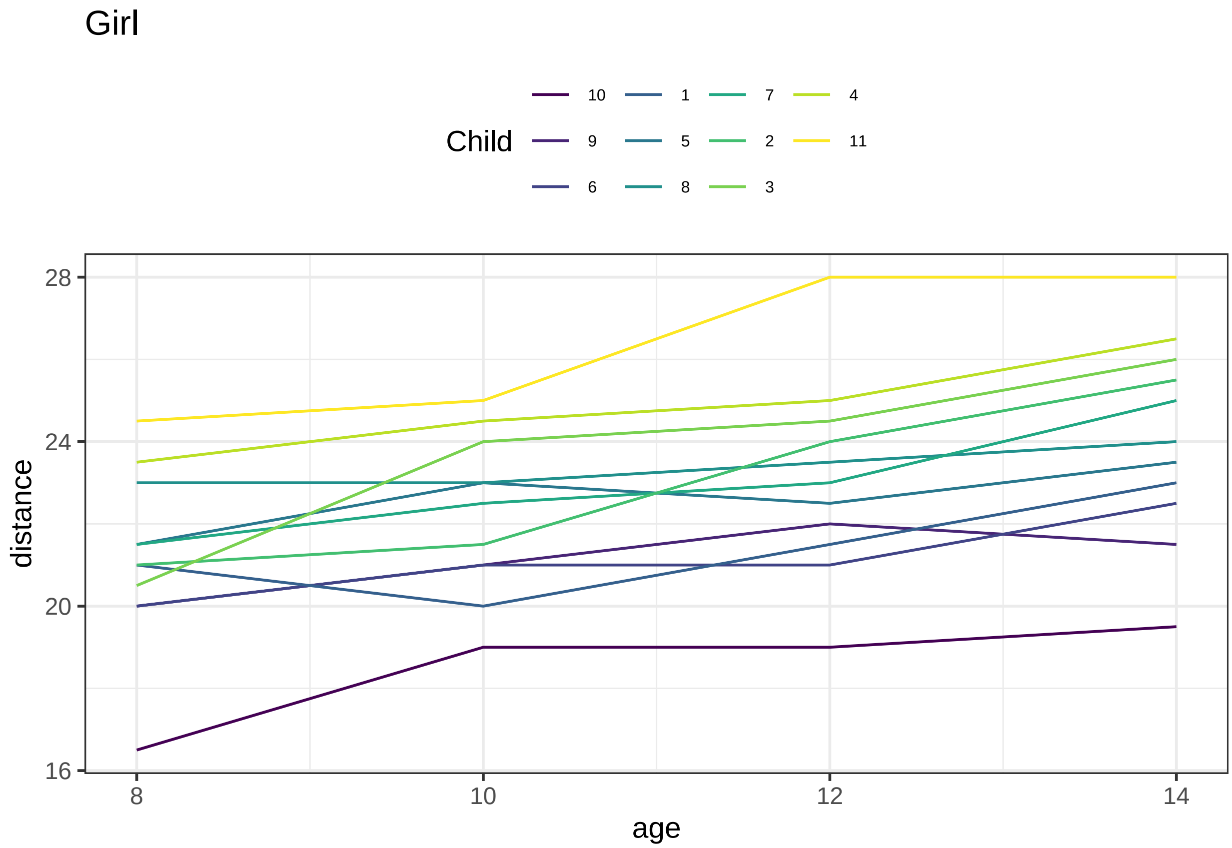
```r
# grouped data
df_new = groupedData (Distance ~ Age | Child, data = as.data.frame(df))

# create a spaghetti plot
# boy
df_new |>
  filter(Gender == "boy") |>
  ggplot(aes(x = Age, y = Distance, group = Child, color = Child)) +
  geom_line() + # spaghetti plot
  theme(legend.text = element_text(size = 6)) + # changed legend text size
  labs(
    title = "Boy",
    x = "age",
    y = "distance"
  ) +
  viridis::scale_color_viridis(
    discrete = TRUE
  )
```




```r
# girl
df_new |>
  filter(Gender == "girl") |>
  ggplot(aes(x = Age, y = Distance, group = Child, color = Child)) +
```

```r
  geom_line() + # spaghetti plot
  theme(legend.text = element_text(size = 6)) + # changed legend text size
  labs(
    title = "Girl",
    x = "age",
    y = "distance"
  ) +
  viridis::scale_color_viridis(
    discrete = TRUE
  )
```



Distance tends to increase with age, with boys having relatively higher distance values than girls.

b

Marginal form:

$$E[Y_{ij}] = \beta_0 + \beta_1 \cdot age_{ij}$$

$$Var(Y_{ij}) = \sigma_a^2 + 2 \cdot sex_{ij} \cdot Cov(a_i, b_k) + (sex_{ij})^2 \cdot \sigma_b^2 + \sigma_e^2$$

$$Cov(Y_{ij}, Y_{ik}) = \sigma_a^2 + (I_{(sex=girl)} + I_{(sex=boy)}) \cdot Cov(a_i, b_k) + I_{(sex=girl)} \cdot I_{(sex=boy)} \cdot \sigma_b^2$$

c

Compound symmetry covariance

```r
summary(gls(Distance ~ Gender + Age, data = df,
            correlation = corCompSymm(form = ~ 1 | Child), method = "REML"))
```

```
## Generalized least squares fit by REML
##   Model: Distance ~ Gender + Age
##   Data: 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) 17.706713 0.8339225 21.233044  0.0000
## Gendergirl  -2.321023 0.7614169 -3.048294  0.0029
## Age          0.660185 0.0616059 10.716263  0.0000
##
##  Correlation:
##            (Intr) Gndrgr
## Gendergirl -0.372
## Age        -0.813  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
```

Exponential covariance

```r
summary(gls(Distance ~ Gender + Age, data = df,
            correlation = corExp(form = ~ 1 | Child), method = "REML"))
```

```
## Generalized least squares fit by REML
##   Model: Distance ~ Gender + Age
##   Data: 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) 17.878709 1.0908637 16.389499   0e+00
## Gendergirl  -2.418714 0.6933441 -3.488476   7e-04
## Age          0.652960 0.0906420  7.203723   0e+00
##
##  Correlation:
##            (Intr) Gndrgr
## Gendergirl -0.259
## Age        -0.914  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
```

Autoregressive covariance

```r
summary(gls(Distance ~ Gender + Age, data = df,
            correlation = corAR1(form = ~ 1 | Child), method = "REML"))
```

```
## Generalized least squares fit by REML
##   Model: Distance ~ Gender + Age
##   Data: 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) 17.878709 1.0908637 16.389499   0e+00
## Gendergirl  -2.418714 0.6933441 -3.488476   7e-04
## Age          0.652960 0.0906420  7.203723   0e+00
##
##  Correlation:
##            (Intr) Gndrgr
## Gendergirl -0.259
## Age        -0.914  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
```

Coefficient Parameter Estimates:
Intercept - Similar across all correlation structures.
Gender (girl) - Coefficient estimates vary slightly.
P-values are significant across all structures, indicating gender has a significant influence on the distance.
Age - Coefficient estimates are similar across all structures. Age also shows significance in predicting distance across all correlation structures.

Covariance Estimates:
Compound Symmetry
Parameter estimate (Rho): 0.614
Exponential Correlation
Parameter estimate (range): 2.134
Autoregressive covariance
Parameter estimate (Phi): 0.626

The covariance estimates vary significantly across correlation structures.