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)

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

This is compound symmetric structure.

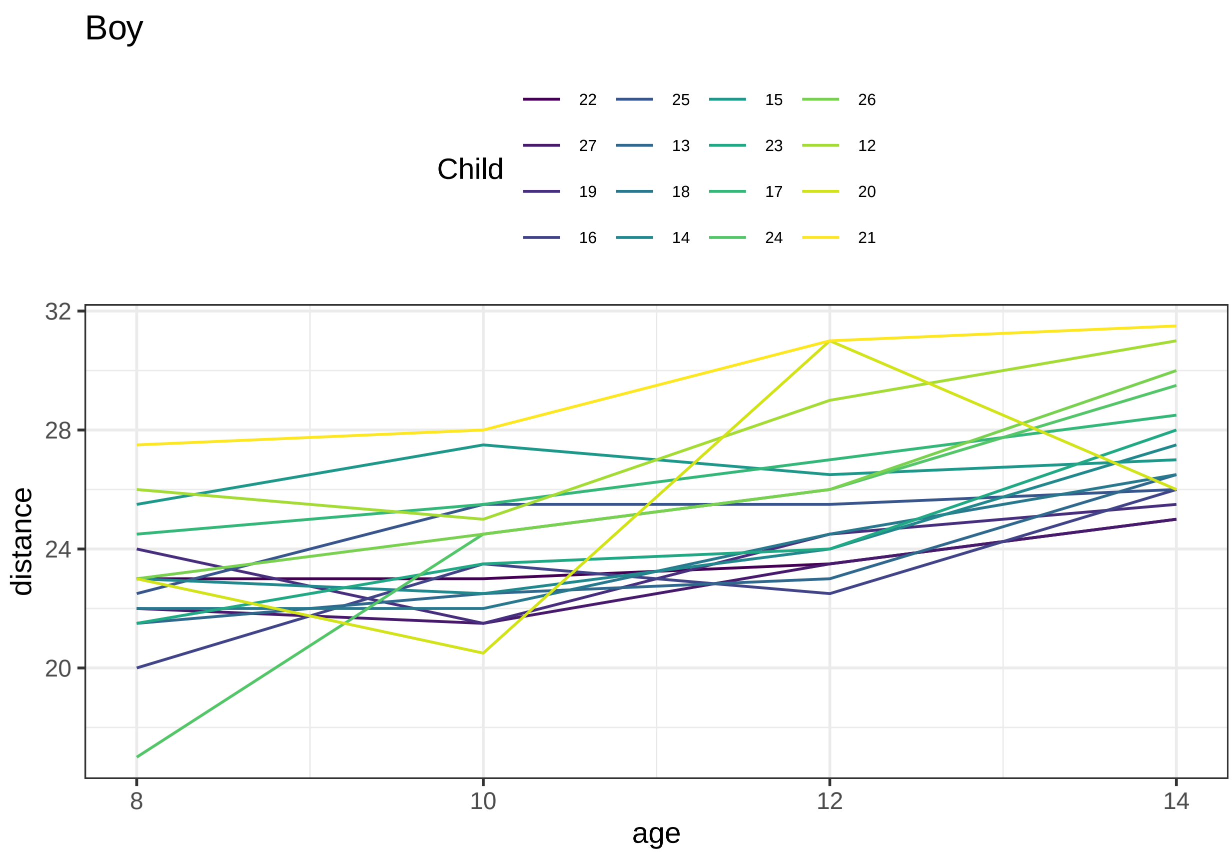
2

```
# 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

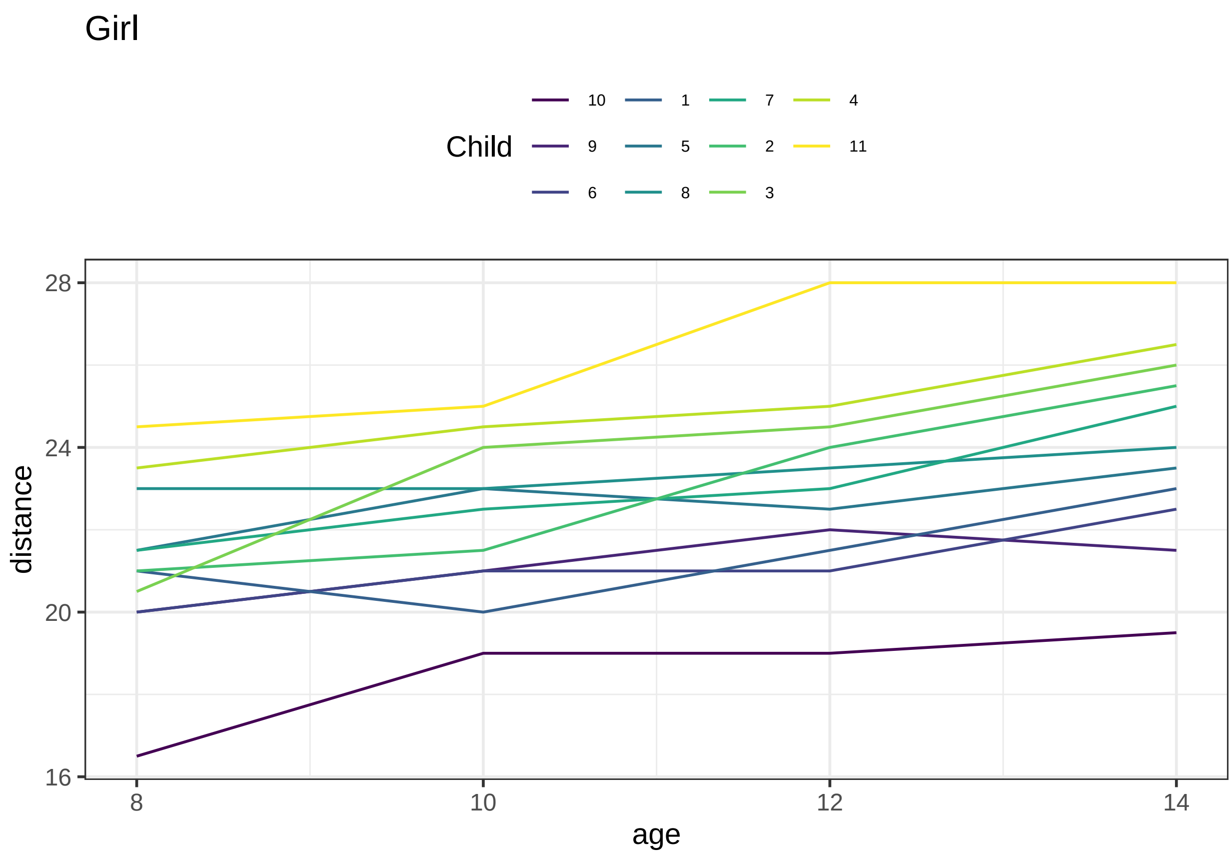
```r
# grouped data
df_new = groupedData (Distance ~ Age | Child, data = as.data.frame(df))

# 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

The marginal form is

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

c

Compound symmetry covariance

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

```
## Generalized least squares fit by REML
##   Model: Distance ~ Gender + Age
##   Data: df
##        AIC      BIC    logLik
##   494.5239 507.7937 -242.2619
##
## Correlation Structure: Compound symmetry
##  Formula: ~1 | Gender
##  Parameter estimate(s):
##          Rho
## 3.415237e-18
##
## Coefficients:
##                 Value Std.Error   t-value p-value
## (Intercept) 17.706713 1.1122095 15.920304       0
## Gendergirl  -2.321023 0.4448862 -5.217115       0
## Age          0.660185 0.0977589  6.753194       0
##
##  Correlation:
##            (Intr) Gndrgr
## Gendergirl -0.163
## Age        -0.967  0.000
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -2.63598228 -0.65509800 -0.02578003  0.52453740  2.36432398
##
## Residual standard error: 2.271713
## Degrees of freedom: 108 total; 105 residual
```

Exponential covariance

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

```
## Generalized least squares fit by REML
##   Model: Distance ~ Gender + Age
##   Data: df
##        AIC      BIC    logLik
##   471.9098 485.1796 -230.9549
##
## Correlation Structure: Exponential spatial correlation
##  Formula: ~1 | Gender
##  Parameter estimate(s):
##    range
## 1.260024
##
```

```
## Coefficients:
##                 Value Std.Error   t-value p-value
## (Intercept) 17.742220 0.9591615 18.497636  0.0000
## Gendergirl  -2.267395 0.7230427 -3.135908  0.0022
## Age          0.658172 0.0763618  8.619124  0.0000
##
##  Correlation:
##            (Intr) Gndrgr
## Gendergirl -0.309
## Age        -0.876  0.000
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -2.60701505 -0.65422544 -0.03208556  0.51035600  2.32586731
##
## Residual standard error: 2.304396
## Degrees of freedom: 108 total; 105 residual
```

Autoregressive covariance

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

```
## Generalized least squares fit by REML
##   Model: Distance ~ Gender + Age
##   Data: df
##        AIC      BIC    logLik
##   471.9098 485.1796 -230.9549
##
## Correlation Structure: AR(1)
##  Formula: ~1 | Gender
##  Parameter estimate(s):
##       Phi
## 0.4521979
##
## Coefficients:
##                 Value Std.Error   t-value p-value
## (Intercept) 17.742220 0.9591615 18.497636  0.0000
## Gendergirl  -2.267395 0.7230427 -3.135908  0.0022
## Age          0.658172 0.0763618  8.619124  0.0000
##
##  Correlation:
##            (Intr) Gndrgr
## Gendergirl -0.309
## Age        -0.876  0.000
##
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -2.60701502 -0.65422543 -0.03208556  0.51035600  2.32586729
##
## Residual standard error: 2.304396
## Degrees of freedom: 108 total; 105 residual
```

Coefficient Parameter Estimates:
Intercept - Similar across all correlation structures, with values around 17.74.
Gender (girl) - Coefficient estimates vary slightly.
Compound symmetry: -2.321
Exponential: -2.267
Autoregresive: -2.267
P-values are significant across all structures, indicating gender has a significant influence on the distance.
Age - Coefficient estimates are similar across all structures, approximately 0.658. Age also shows significance in predicting distance across all correlation structures.

Covariance Estimates:
Compound Symmetry
Parameter estimate (Rho): 3.42e-18
Exponential Correlation
Parameter estimate (range): 1.26
Autoregressive covariance
Parameter estimate (Phi): 0.45

The covariance estimates vary significantly across correlation structures.