Homework 6

Yuki Joyama

2024-03-24

1

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

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} ($j \neq k$)

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

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

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

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

$$= Var(b_i) = \sigma_b^2 \qquad (\text{Given ① and ②})$$

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

This is compound symmetric structure.

2

```r
# 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
  )
```
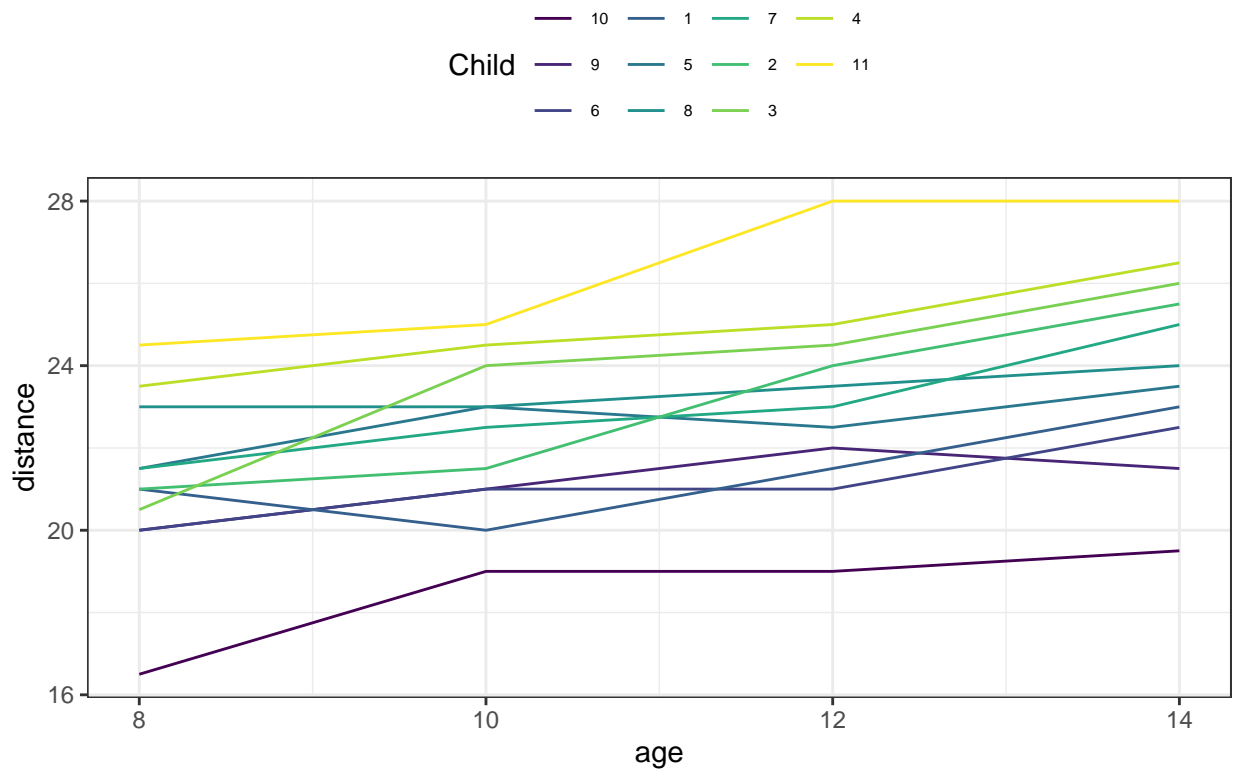

Boy
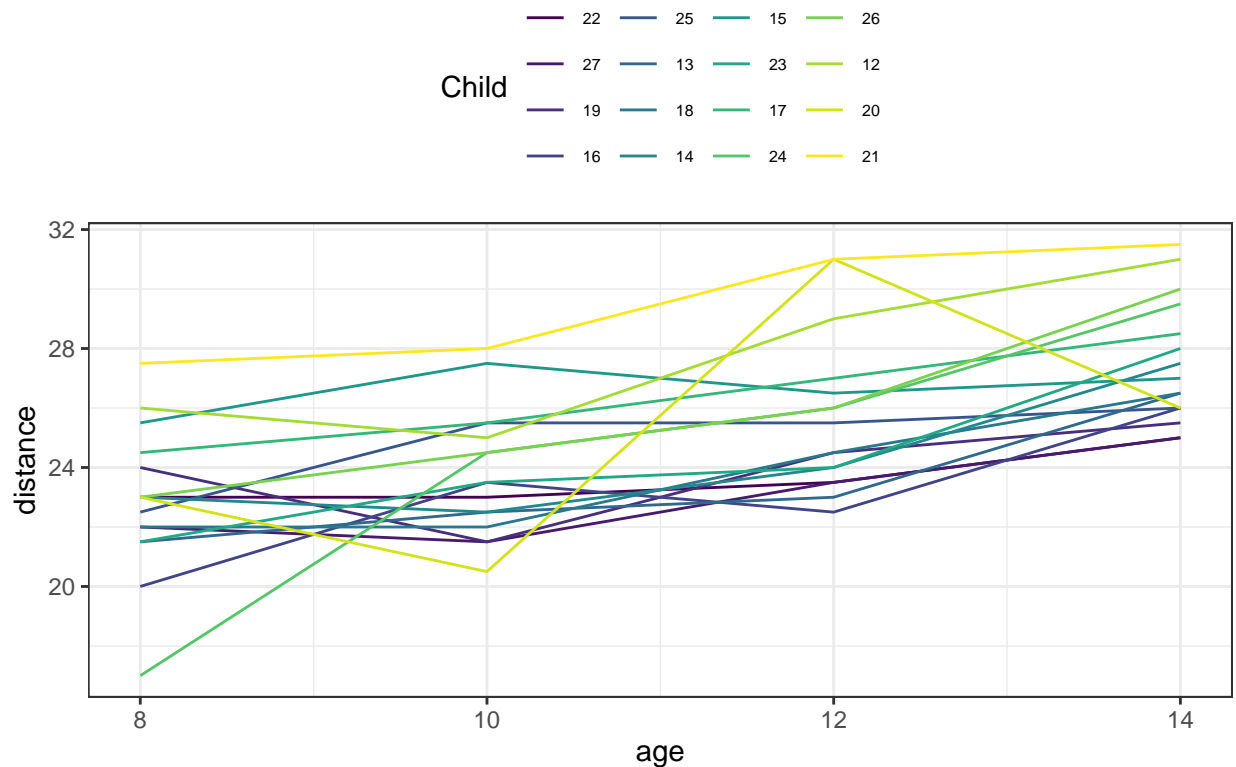
```r
# girl
df_new |>
  filter(Gender == "girl") |>
  ggplot(aes(x = Age, y = Distance, group = Child, color = Child)) +
```

```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
)
```

Girl

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

b

c

Compound symmetry covariance

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

Exponential covariance

Autoregressive covariance