

Homework 6

Yuki Joyama

2024-03-24

1

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

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 $j, k=1, \dots, n, j \neq k$ — ②

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

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

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

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

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

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

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

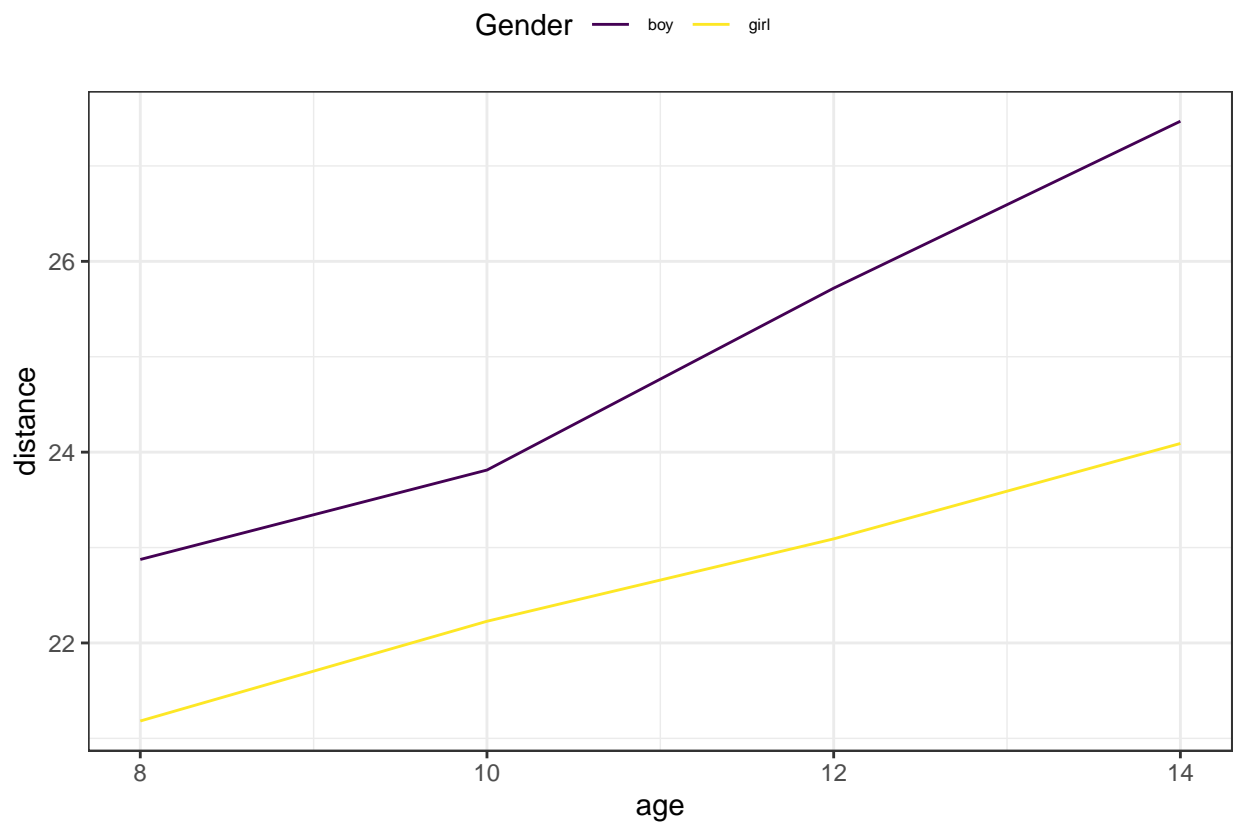
This is compound symmetric structure.

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

```
# create a spaghetti plot
df |>
  group_by(Age, Gender) |>
  summarise(value_mean = mean(Distance)) |>
  ggplot(aes(x = Age, y = value_mean, group = Gender, color = Gender)) +
  geom_line() + # spaghetti plot
  theme(legend.text = element_text(size = 6)) + # changed legend text size
  guides(color = guide_legend(ncol = 3)) + # changed the number of legend columns
  labs(
    x = "age",
    y = "distance"
  ) +
  viridis::scale_color_viridis(
    discrete = TRUE
  )
)
```



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

b

c