MID

Zirui Zhang

2023-10-25

```r
library(tidyverse)
library(ggplot2)
library(patchwork)
library(nlme)
library(lme4)
library(knitr)
```

```r
load("~/Documents/2023Fall/P8157/P8157/Six Cities.RData")
data = topeka |> group_by(id) |> filter(n() >= 5) |> ungroup()
length(unique(data$id))
```

```
## [1] 196
```

```r
data = data |>
  mutate(y = exp(log.FEV1)/(height^2),
         age.2 = age^2,
         age.3 = age^3)
```
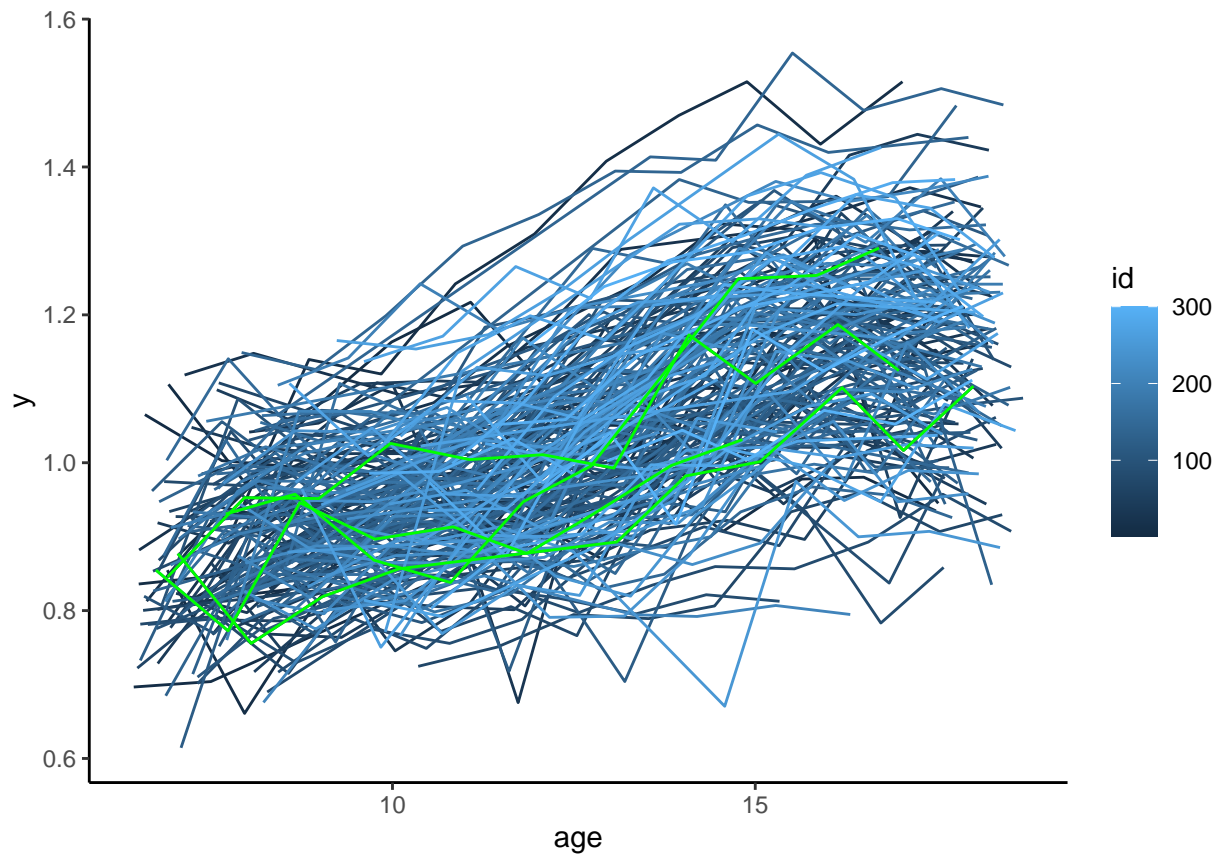
Question (a)

Produce a figure of the response, Yki as a function of age. On the figure indicate the individual trajectories for a random sample of 4 girls.

```r
set.seed(200324)
sample = data |>
  filter(id %in% sample(unique(data$id), 4))

ggplot(data, aes(x = age, y = y, group = id, color = id)) +
  geom_line() +
  geom_line(data = sample, color = "blue") +
  scale_color_distiller(type = "seq",
                        direction = -1,
                        palette = "Greys") +
  theme_classic()
```
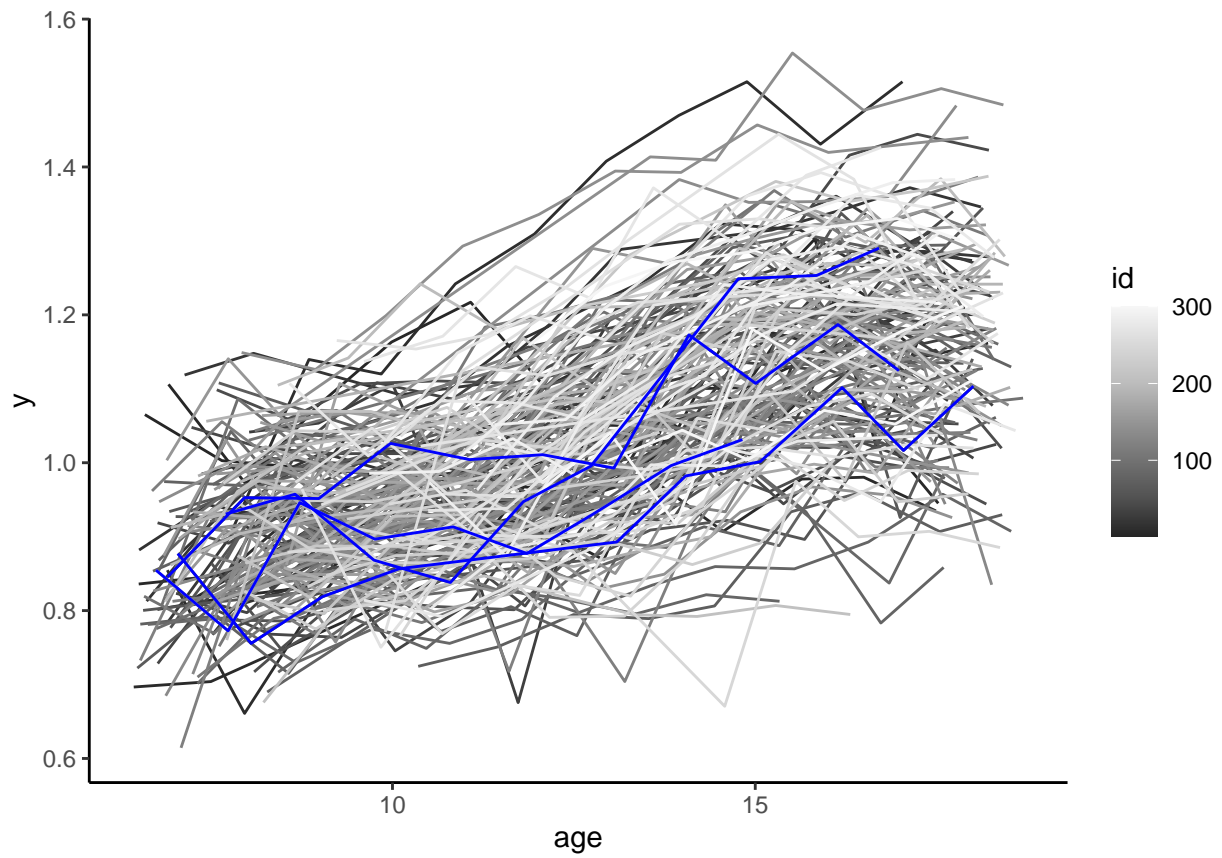
```
ggplot(data, aes(x = age, y = y, group = id, color = id)) +
  geom_line() +
  geom_line(data = sample, color = "green") +
  theme_classic()
```

Question (b)

```r
# 1 naivee
fit1.ML = glm(y ~ age + age.2 + age.3, data, family=gaussian)

# 2 randon intercept + independent error
fit2.ML = lme(fixed=y ~ age + age.2 + age.3, random=reStruct(~ 1 | id), data=data, method="ML")

# 3 random intercept/slope + independent error
fit3.ML = lme(fixed=y ~ age + age.2 + age.3, random=reStruct(~ age | id, pdClass="pdDiag"), data=data,

# 4. random intercept + auto_regressive error
fit4.ML = lme(fixed=y ~ age + age.2 + age.3, random=reStruct(~ 1 | id), correlation=corAR1(form= ~ age|

# 5 random intercept + exponential spatial error
fit5.ML = lme(fixed=y ~ age + age.2 + age.3, random=reStruct(~ 1 | id), correlation=corExp(form= ~ age|

# 6 random intercept + exponential spatial error + independent homo error
fit6.ML = lme(fixed=y ~ age + age.2 + age.3, random=reStruct(~ 1 | id), correlation=corExp(form= ~ age|

# 7 random intercept + independent hetero error
data_cat = data |>
  dplyr::mutate(age.cat = floor(age/2))
fit7.ML = lme(fixed=y ~ age + age.2 + age.3, random=reStruct(~ 1 | id), weights=varIdent(form= ~1 | age
```

```r
# 8 random intercept/slope + independent hetero error
fit8.ML = lme(fixed=y ~ age + age.2 + age.3, random=reStruct(~ age | id), weights=varIdent(form= ~1 | ag
```

```r
sum = (data.frame(
  logLik = c(logLik(fit1.ML), logLik(fit2.ML), logLik(fit3.ML),logLik(fit4.ML),
             logLik(fit5.ML), logLik(fit6.ML), logLik(fit7.ML), logLik(fit8.ML)),
  AIC = c(AIC(fit1.ML),AIC(fit2.ML),AIC(fit3.ML),AIC(fit4.ML),
          AIC(fit5.ML),AIC(fit6.ML),AIC(fit7.ML),AIC(fit8.ML))
))

colnames(sum) = c("log-Like", "AIC")
rownames(sum) = c("0. Independence", "1. Random intercept + inde. errors",
                  "2. Random intercept/slope + inde. errors", "3. Random intercept + AR errors",
                  "4. Random intercept + ES errors", "5. Random intercept + ES with a 'nugget'",
                  "6. Random intercept + heteroske inde. errors",
                  "7. Random intercept/slope + heteroske inde. errors")
kable(sum, format = "markdown")
```

	log-Like	AIC
0. Independence	1291.696	-2573.393
1. Random intercept + inde. errors	2073.425	-4134.850
2. Random intercept/slope + inde. errors	2138.977	-4263.954
3. Random intercept + AR errors	2156.186	-4298.373
4. Random intercept + ES errors	2168.018	-4322.035
5. Random intercept + ES with a 'nugget'	2175.572	-4335.145
6. Random intercept + heteroske inde. errors	2092.402	-4160.803
7. Random intercept/slope + heteroske inde. errors	2162.089	-4296.178

Model 4 and 5 give the largest loglikelihood and lowest AIC, provide best fit of the data.

```r
summary(fit5.ML)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data
##         AIC       BIC    logLik
##   -4322.035 -4283.609 2168.018
##
## Random effects:
##  Formula: ~1 | id
##         (Intercept)  Residual
## StdDev:    0.094101 0.0696275
##
## Correlation Structure: Exponential spatial correlation
##  Formula: ~age | id
##  Parameter estimate(s):
##     range
## 1.185422
## Fixed effects:  y ~ age + age.2 + age.3
##                 Value  Std.Error   DF   t-value p-value
## (Intercept)  1.4009980 0.10895317 1590 12.858718       0
```

```
## age          -0.1741504 0.02769370 1590 -6.288450       0
## age.2         0.0175949 0.00225466 1590  7.803767       0
## age.3        -0.0004801 0.00005899 1590 -8.138419       0
##  Correlation:
##       (Intr) age    age.2
## age   -0.992
## age.2  0.977 -0.995
## age.3 -0.957  0.983 -0.996
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -4.65923585 -0.53300785  0.07268422  0.60047532  2.60157186
##
## Number of Observations: 1789
## Number of Groups: 196
```

```r
summary(fit6.ML)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: data
##         AIC       BIC   logLik
##   -4335.145 -4291.23 2175.572
##
## Random effects:
##  Formula: ~1 | id
##         (Intercept)   Residual
## StdDev:  0.08951122 0.07477168
##
## Correlation Structure: Exponential spatial correlation
##  Formula: ~age | id
##  Parameter estimate(s):
##     range     nugget
## 2.8975885 0.2822974
## Fixed effects:  y ~ age + age.2 + age.3
##                  Value  Std.Error   DF   t-value p-value
## (Intercept)  1.4342926 0.10525523 1590 13.626806       0
## age         -0.1825535 0.02672730 1590 -6.830227       0
## age.2        0.0182797 0.00217316 1590  8.411544       0
## age.3       -0.0004980 0.00005678 1590 -8.771950       0
##  Correlation:
##       (Intr) age    age.2
## age   -0.991
## age.2  0.975 -0.995
## age.3 -0.953  0.981 -0.996
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -4.39817065 -0.51464710  0.04853461  0.57968727  2.41673358
##
## Number of Observations: 1789
## Number of Groups: 196
```

```
coef = t((data.frame(
  fit5 = c(summary(fit5.ML)$coefficients$fixed, sqrt(diag(summary(fit5.ML)$varFix))),
  fit6  = c(summary(fit6.ML)$coefficients$fixed, sqrt(diag(summary(fit6.ML)$varFix)))
)))
rownames(coef) = c("Model.4", "Model.5")
colnames(coef) = c("b0", "b1", "b2", "b3",
                   "sd(b0)", "sd(b1)", "sd(b2)", "sd(b3)")
kable(coef, format = "markdown")
```

	b0	b1	b2	b3	sd(b0)	sd(b1)	sd(b2)	sd(b3)
Model.4	1.400998	-0.1741504	0.0175949	-0.0004801	0.1088313	0.0276627	0.0022521	5.89e-05
Model.5	1.434293	-0.1825535	0.0182797	-0.0004980	0.1051375	0.0266974	0.0021707	5.67e-05

Question (c)

Question (d)

Question (d)

Question (e)

Question (f)