

Stats112 HW3

Viraj Vijaywargiya

2023-05-10

```
library( nnet )
library(lattice)
library(nlme)
library(lme4)
```

```
## Loading required package: Matrix
```

```
##
```

```
## Attaching package: 'lme4'
```

```
## The following object is masked from 'package:nlme':
```

```
##
```

```
##      lmList
```

```
library(survival)
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
```

```
## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::collapse() masks nlme::collapse()
## x tidyr::expand()   masks Matrix::expand()
## x dplyr::filter()   masks stats::filter()
## x dplyr::lag()       masks stats::lag()
## x tidyr::pack()      masks Matrix::pack()
## x tidyr::unpack()    masks Matrix::unpack()
```

```
1. birthwt = read.table("/Users/virajvijaywargiya/Downloads/birthwt-data.txt", header=FALSE)
   names(birthwt) = c("MID", "Order", "Wt", "Age", "CID")
```

```
1a)
```

```
mod = lme(Wt ~ 1 + Age, random = ~1 | MID, data=birthwt, method="ML")
summary(mod)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: birthwt
##       AIC      BIC    logLik
##  67079.34 67104.89 -33535.67
##
## Random effects:
##   Formula: ~1 | MID
##       (Intercept) Residual
## StdDev:    354.6307 434.1899
##
## Fixed effects:  Wt ~ 1 + Age
##               Value Std.Error   DF  t-value p-value
## (Intercept) 2785.1495  44.98839 3511  61.90819     0
## Age         17.1431   1.98003 3511   8.65803     0
## Correlation:
##   (Intr)
## Age -0.953
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -6.04734611 -0.46642860  0.05826707  0.56462185  3.13023577
##
## Number of Observations: 4390
## Number of Groups: 878
```

1b) $Y_{ij} = B_0 + b_{0i} + B_1 \text{Age}_{ij} + e_{ij}$

1c)

```
mod0 = lme(Wt ~ 1, random = ~1 | MID, data=birthwt, method="ML")
summary(mod0)
```

```
## Linear mixed-effects model fit by maximum likelihood
##   Data: birthwt
##       AIC      BIC    logLik
##  67150.64 67169.8 -33572.32
##
## Random effects:
##   Formula: ~1 | MID
##       (Intercept) Residual
## StdDev:    368.4007 435.4458
##
## Fixed effects:  Wt ~ 1
##               Value Std.Error   DF  t-value p-value
## (Intercept) 3156.304  14.06466 3512 224.4138     0
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -5.96885118 -0.46691460  0.05425461  0.56494752  3.21491810
##
## Number of Observations: 4390
## Number of Groups: 878
```

```
anova(mod0, mod)
```

```
##      Model df      AIC      BIC    logLik   Test  L.Ratio p-value
## mod0      1  3 67150.64 67169.80 -33572.32
## mod       2  4 67079.34 67104.89 -33535.67 1 vs 2 73.30363 <.0001
```

The LRT test above shows that Model 2 (mod) is better as it has lower AIC value. Therefore, age shouldn't be dropped from the model in part a.

1d)

```
mod1 = lme(Wt ~ 1 + Age, random = ~1+Age | MID, data=birthwt, method="ML")
summary(mod1)
```

```
## Linear mixed-effects model fit by maximum likelihood
## Data: birthwt
##      AIC      BIC    logLik
## 67031.89 67070.21 -33509.94
##
## Random effects:
## Formula: ~1 + Age | MID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev   Corr
## (Intercept) 437.92972 (Intr)
## Age          24.95883 -0.761
## Residual    426.45612
##
## Fixed effects: Wt ~ 1 + Age
##           Value Std.Error   DF  t-value p-value
## (Intercept) 2805.7880  46.68000 3511 60.10685     0
## Age          15.9444   2.23187 3511  7.14394     0
## Correlation:
## (Intr)
## Age -0.959
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -5.96974901 -0.45714542  0.05445112  0.56136671  3.32690930
##
## Number of Observations: 4390
## Number of Groups: 878
```

1e)

```
head(random.effects(mod1))
```

```
##      (Intercept)      Age
## 80 -126.0171098 21.6985269
## 84   0.2130301  9.2238363
## 92  186.0176365 -32.0464361
## 113 135.8221017 -12.7888731
## 199 -39.3357059  0.5286986
## 200 -25.8128943  1.8089241
```

$$Y_{ij} = -126.02 + b_{0i} + 21.7 \text{ Age}_{ij}$$

1f)

```
mod1.reml = lme(Wt ~ 1 + Age, random = ~1+Age | MID, data=birthwt, method="REML")
summary(mod1.reml)
```

```
## Linear mixed-effects model fit by REML
##   Data: birthwt
##       AIC       BIC    logLik
##  67021.44 67059.76 -33504.72
##
## Random effects:
##   Formula: ~1 + Age | MID
##   Structure: General positive-definite, Log-Cholesky parametrization
##              StdDev    Corr
## (Intercept) 440.81930 (Intr)
## Age          25.08002 -0.763
## Residual    426.44113
##
## Fixed effects: Wt ~ 1 + Age
##              Value Std.Error   DF  t-value p-value
## (Intercept) 2805.9041  46.71072 3511  60.06981     0
## Age          15.9389   2.23348 3511   7.13634     0
## Correlation:
##   (Intr)
## Age -0.959
##
## Standardized Within-Group Residuals:
##           Min           Q1           Med           Q3           Max
## -5.96898519 -0.45680972  0.05479743  0.56128622  3.32766608
##
## Number of Observations: 4390
## Number of Groups: 878
```

```
mod.reml = lme(Wt ~ 1 + Age, random = ~1 | MID, data=birthwt, method="REML")
summary(mod.reml)
```

```
## Linear mixed-effects model fit by REML
##   Data: birthwt
##       AIC       BIC    logLik
##  67069.07 67094.62 -33530.53
##
## Random effects:
##   Formula: ~1 | MID
##           (Intercept) Residual
## StdDev:    354.9605 434.2334
##
## Fixed effects: Wt ~ 1 + Age
##              Value Std.Error   DF  t-value p-value
## (Intercept) 2785.2542  44.99276 3511  61.90449     0
## Age          17.1383   1.98014 3511   8.65511     0
## Correlation:
##   (Intr)
```

```
## Age -0.953
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -6.04660152 -0.46606378  0.05847938  0.56450997  3.12988230
##
## Number of Observations: 4390
## Number of Groups: 878
```

```
anova(mod.reml, mod1.reml)
```

```
##      Model df      AIC      BIC    logLik    Test  L.Ratio p-value
## mod.reml    1  4 67069.07 67094.62 -33530.53
## mod1.reml   2  6 67021.44 67059.76 -33504.72 1 vs 2 51.63041  <.0001
```

From the results above, Model 2 (mod1.reml) is better as it has a lower AIC value. Therefore, a random slope and intercept is necessary in the model.

2. In clustered data, measurements within the same cluster (in this case, within the same school) may be correlated, and this correlation needs to be accounted for in the analysis. One common way to model the covariance structure of clustered data is to assume a compound symmetry structure.

A compound symmetry structure assumes that the variances of measurements within a cluster are equal, and that the correlation between any two measurements within a cluster is the same. In other words, it assumes that the within-cluster covariance matrix has a constant value along its diagonal and a constant value off the diagonal.

In the case of this study, it is reasonable to assume that the variance of math achievement scores within a school is roughly the same for all students, regardless of their sex or socioeconomic status. Moreover, it is reasonable to assume that the correlation between any two math achievement scores within a school is the same, regardless of the sex or socioeconomic status of the students.

Therefore, a compound symmetry structure for the variance-covariance of measurements within a school is a reasonable assumption for this study.

3.

```
NCGS = read.table("/Users/virajvijaywargiya/Downloads/cholesterol-data.txt", na.strings=".")
names(NCGS) = c("Trt", "ID", "M0", "M6", "M12", "M20", "M24")

NCGS$Trt = factor(NCGS$Trt, levels=c(2,1), labels=c("Placebo","HighDose"))

NCGS.long = reshape(NCGS, varying=list(3:7), idvar="ID", timevar="Month", times=c(0,6,12,20,24), v.names=c("Chol", "Trt", "M0", "M6", "M12", "M20", "M24"))

NCGS.long$Time = as.numeric(factor(NCGS.long$Month))
```

3a)

```
mod.lme = lme(Chol~Trt+I(Month)+I(Month^2), data=NCGS.long, random=~I(Month)+I(Month^2)|ID, na.action=na.omit)
summary(mod.lme)
```

```
## Linear mixed-effects model fit by maximum likelihood
## Data: NCGS.long
##      AIC      BIC    logLik
## 4364.344 4409.472 -2171.172
##
## Random effects:
```

```
## Formula: ~I(Month) + I(Month^2) | ID
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev      Corr
## (Intercept) 40.87202701 (Intr) I(Mnt)
## I(Month)    1.10827043 -0.679
## I(Month^2)  0.04961848  0.425 -0.770
## Residual    22.33613100
##
## Fixed effects: Chol ~ Trt + I(Month) + I(Month^2)
##           Value Std.Error DF t-value p-value
## (Intercept) 230.04328  6.544938 342 35.14827  0.0000
## TrtHighDose  1.03266  7.794128 101  0.13249  0.8949
## I(Month)     2.29456  0.467767 342  4.90534  0.0000
## I(Month^2)   -0.05230  0.019363 342 -2.70116  0.0073
## Correlation:
##           (Intr) TrtHgD I(Mnt)
## TrtHighDose -0.717
## I(Month)    -0.325  0.001
## I(Month^2)  0.237  0.005 -0.945
##
## Standardized Within-Group Residuals:
##           Min          Q1          Med          Q3          Max
## -2.52995029 -0.59019968 -0.02986775  0.50313472  2.37731578
##
## Number of Observations: 447
## Number of Groups: 103
```

3b) TrtHighDose has a p-value = 0.895 > 0.05. Therefore, we fail to reject the null, and conclude that treatment main effect is not significant.

3c)

```
getVarCov(mod.lme, type = "random.effects")
```

```
## Random effects variance covariance matrix
##           (Intercept)  I(Month) I(Month^2)
## (Intercept)  1670.5000 -30.772000  0.861500
## I(Month)     -30.7720  1.228300  -0.042344
## I(Month^2)    0.8615  -0.042344  0.002462
## Standard Deviations: 40.872 1.1083 0.049618
```

3d) estimated variance of the random intercepts: 1670.5000
 estimated variance of the random slopes: 1.228300
 estimated correlation between the random intercepts and slopes: -0.7720

3e)

```
random.effects(mod.lme)[96,]
```

```
##           (Intercept)  I(Month)  I(Month^2)
## 96    -48.06948  0.6253423  -0.02291203
```

$b(0, 96) = -48.06948$, $b(1, 96) = 0.6253423$, $b(2, 96) = -0.02291203$.

3f)

```
fitted(mod.lme)[names(fitted(mod.lme))=="71"]
```

```
##      71      71      71      71      71
## 201.1366 215.6939 224.7885 228.4169 226.5892
```

estimated response for Month 6: 215.6939

3g)

```
mod.lme.notime = lme(Chol~ Trt+ I(Month)+I(Month^2) , data=NCGS.long, random=~1|ID, na.action=na.omit)
summary(mod.lme.notime)
```

```
## Linear mixed-effects model fit by REML
## Data: NCGS.long
##      AIC      BIC    logLik
## 4358.967 4383.528 -2173.483
##
## Random effects:
## Formula: ~1 | ID
##      (Intercept) Residual
## StdDev:      37.48282 23.80293
##
## Fixed effects: Chol ~ Trt + I(Month) + I(Month^2)
##              Value Std.Error DF t-value p-value
## (Intercept) 231.67412  6.426195 342 36.05152  0.0000
## TrtHighDose -1.68356  7.906953 101 -0.21292  0.8318
## I(Month)      2.29618  0.480814 342  4.77562  0.0000
## I(Month^2)    -0.05248  0.019599 342 -2.67766  0.0078
## Correlation:
##      (Intr) TrtHgD I(Mnt)
## TrtHighDose -0.741
## I(Month)     -0.251  0.000
## I(Month^2)   0.193  0.005 -0.959
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.90778633 -0.59693600 -0.01151234  0.53688000  2.79607708
##
## Number of Observations: 447
## Number of Groups: 103
```

```
mod.rlme.reml = lme(Chol~Trt+I(Month)+I(Month^2), data=NCGS.long, random=~I(Month)+I(Month^2)|ID,
summary(mod.rlme.reml)
```

```
## Linear mixed-effects model fit by REML
## Data: NCGS.long
##      AIC      BIC    logLik
## 4361.85 4406.879 -2169.925
##
## Random effects:
## Formula: ~I(Month) + I(Month^2) | ID
## Structure: General positive-definite, Log-Cholesky parametrization
##      StdDev      Corr
##
```

```
## (Intercept) 41.28519005 (Intr) I(Mnt)
## I(Month)      1.15081783 -0.659
## I(Month^2)    0.05134622  0.412 -0.776
## Residual      22.35220403
##
## Fixed effects: Chol ~ Trt + I(Month) + I(Month^2)
##              Value Std.Error DF  t-value p-value
## (Intercept) 230.04233  6.574626 342 34.98942  0.0000
## TrtHighDose  1.03501  7.836462 101  0.13208  0.8952
## I(Month)      2.29438  0.467087 342  4.91210  0.0000
## I(Month^2)   -0.05230  0.019344 342 -2.70353  0.0072
## Correlation:
##              (Intr) TrtHgD I(Mnt)
## TrtHighDose -0.718
## I(Month)     -0.323  0.001
## I(Month^2)   0.236  0.005 -0.944
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.5108779 -0.5852554 -0.0271818  0.5031683  2.3577561
##
## Number of Observations: 447
## Number of Groups: 103
```

```
anova(mod.lme.notime, mod.rlme.reml)
```

```
##           Model df      AIC      BIC    logLik  Test  L.Ratio p-value
## mod.lme.notime    1  6 4358.967 4383.528 -2173.483
## mod.rlme.reml     2 11 4361.850 4406.879 -2169.925 1 vs 2 7.116648  0.2121
```

From the above results, model 1 (mod.lme.notime) is better as it has lower AIC value. Therefore, we shouldn't have random effects on the time covariates.