Biostatistics 140.655

Lab 3

**Topics:**

* Fitting linear mixed effects models
* Understand equivalence between random intercept and marginal model with exchangeable correlation
* Interpret all elements from random intercept and random slope mixed models
* Explore the variance structure induced by random intercept and random slope (over time) models

**Learning Objectives:**

Students who successfully complete this lab will be able to:

* Specify and fit a linear mixed effects model
* Interpret the regression coefficients and random effect variance terms within linear mixed models
* Apply a model-selection criteria to select a “best” fitting model among a set of candidate models
* Describe that a random intercept and random slope (for time) model induces a variance model for the outcome that is non-constant with time.

**Scientific Background**:

You will be extending the previous analyses of the Childhood Asthma Management Program (CAMP) trial. Recall that the CAMP trial was a multicenter, masked, placebo-controlled, randomized trial designed to determine the long-term effects of three treatments (budesonide, nedocromil, or placebo) on pulmonary function among children with asthma.

In CAMP trial, children with asthma aged 5-12 years were enrolled between 1993 and 1995. The primary outcome of the trial was Forced Expiratory Volume at 1 second (FEV1) after the administration of a bronchodilator. After a baseline assessment, children were randomized to receive one of the three treatments and then followed for 4 years. The primary outcome was measured at baseline and then at 2, 4, 12, 16, 24, 28, 36, 40 and 48 months after randomization.

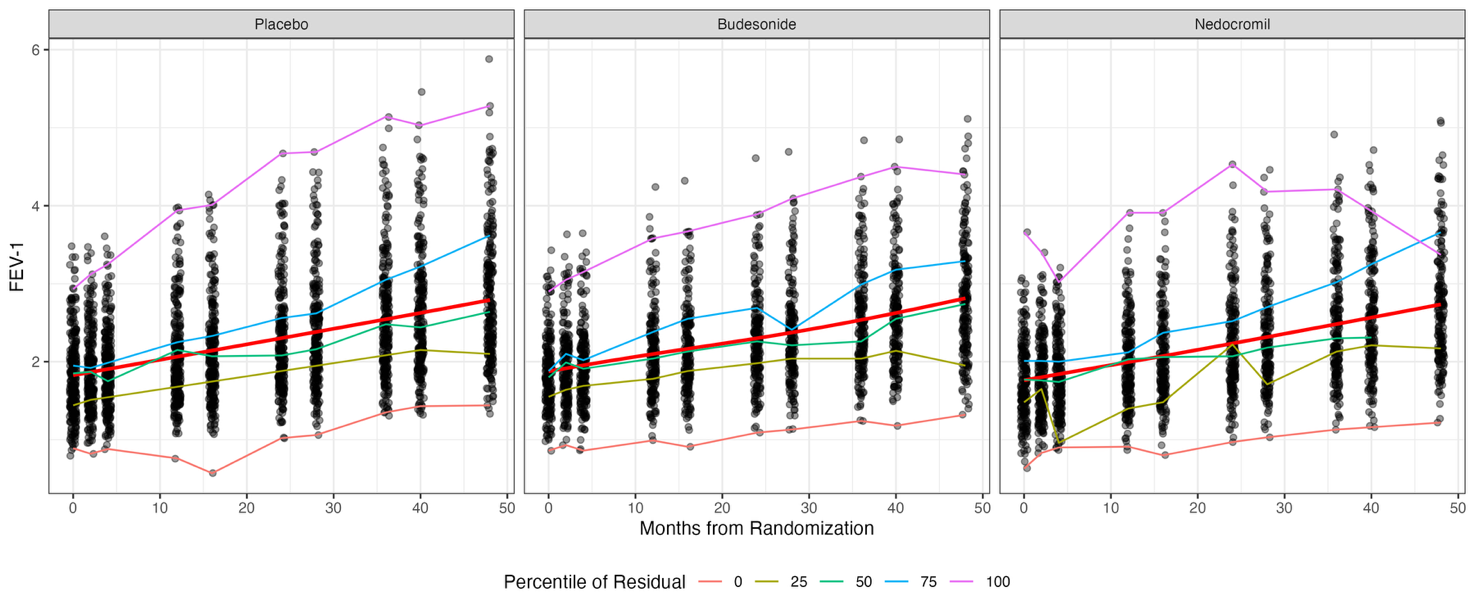
**We will only use data from follow-up times 0, 2, 4, 12, 24 in this Lab.**

In your exploratory analysis of the CAMP trial data, you found that the average FEV1 increased roughly linearly over time during the 4-year follow-up AND that the variance in FEV1 increased over time (see Table 1). In addition, you made spaghetti plots (see Figure 1 below), where you found variation in FEV1 among children at baseline and then also heterogeneity in the linear slopes over the 4-year follow-up.

Table 1: Estimates of the variance of FEV1 at each assessment time (in months). The variance estimates were obtained by taking the average of the squared residuals from a two-way ANOVA regressing FEV1 on visit and treatment arm model (ANOVA means fit both treatment and time as factors (i.\_\_\_ in Stata; factor(\_\_\_\_) in R).

|  |  |
| --- | --- |
| Follow-up (Months) | Estimated Variance |
| 0 | .2592463 |
| 2 | .2567264 |
| 4 | .2641309 |
| 12 | .3362065 |
| 16 | .3588506 |
| 24 | .4221124 |
| 28 | .4578405 |
| 36 | .5249448 |
| 40 | .5525803 |
| 48 | .6293908 |

Figure 1: Spaghetti plot of FEV1 over time, separately among the three treatment arms. In addition to an average line (in red), trajectories of individual children are highlighted. Children were selected who have the 0, 25, 50, 75, and 100th percentile (rounded to the second decimal place) of average residual (across visits) from a one-way ANOVA model regressing FEV on visit, separately by treatment group.



Reference:

1. The Childhood Asthma Management Program (CAMP): Design, Rationale, and Methods. Controlled Clinical Trials 1999; 20:91-120.

2. Long-Term Effects of Budesonide or Nedocromil in Children with Asthma. New England Journal of Medicine 2000;343:1054-63.

Data Description

The “camp\_primary.csv” dataset contains a random sample of 695 children from the CAMP trial. The data consist of the primary outcome (post-bronchodilator FEV1), treatment group, as well as baseline characteristics of the children.

Variable List:

* id (participant ID)
* trt (treatment group: 0 “placebo” 1 “nedocromil”)
* age\_rz (age at randomization in years)
* gender (0 “male” 1 “female”)
* ethnicity (0 “White” 1 “Black” 2 “Hispanic” 3 “Other”)
* POSFEV (primary outcome: post-bronchodilator FEV1measured in liters)
* visit (the nth visit)
* visitc (months since randomization)
* fdays (days since randomization)

Objectives

The goal of CAMP was to determine if there are greater improvements in pulmonary function over time with the use of active treatment compared to placebo in children with asthma.

**In addition,** from your initial analysis of the data, you discovered that the variation in FEV1 increased over the two years students were followed. Some of the modeling approaches we applied previously assumed that the variance in FEV1 constant over time (e.g. exchangeable, AR-1, Toeplitz).

In this lab, you will explore using **linear mixed effects models** as one method to address the non-constant variance assumption. We will start with demonstrating the equivalence of the random intercept model and the exchangeable within-subject correlation assumption; followed by extending the random intercept model to also include a random slope for time.

**Exercise:**

1. Two models are defined below: Model 1: marginal linear model assuming an exchangeable correlation structure and Model 2: random intercept linear mixed effects model. NOTE: I have rescaled the time variable to represent years; . This is out of convenience and to make the random intercept variance on a larger scale for ease of interpretation. (Some might ask in frustration why we divided month by 24 before in lecture and by 12 now? Answer: so you understand that you can divide by any constant and get the same fitted model and to urge you to tailor the model to the problem at hand, making interpretation easier in different contexts.)

Review the notation of each model and answer the questions below.

Model 1: Marginal linear assuming exchangeable correlation structure.

where and , for all *j not equal to k* and ,0 for all *i not equal to k and any j,l.*

Model 2: Random intercept for child linear mixed model

where and , for all *j not equal to k*, and , for all *i not equal to j*, and , for all *i,j,k.*

* 1. In Model 1, is often referred to as the total residual. Define this quantity in the context of the CAMP trial.
  2. In Model 2, is often referred to as the within-subject residual. Define this quantity in the context of the CAMP trial.
  3. In Model 2, is referred to as the random intercept for subject. Define this quantity in the context of the CAMP trial.

1. Use the code provided below to fit Model 1 and 2 and answer the questions that follow:

**STATA:**

drop if posfev == .

gen year = visitc/12

\* Exchangeable marginal linear model

mixed posfev year i.trt#c.year || id: , nocons residuals(exch) ml stddev

\* Random intercept linear mixed model

mixed posfev year i.trt#c.year || id: , ml stddev

**R:**

library(nlme)

dat$year = dat$visitc/12

## Model 1: fit exponential model

fit1 <- gls(POSFEV ~ year + as.factor(trt):year, data=dat, na.action=na.omit,

correlation=corCompSymm(form=~1|id), method="ML")

summary(fit1)

## Model 2: random intercept for child model

fit2 <- lme(POSFEV ~ year + as.factor(trt):year, data=dat,

random=~1|id, method="ML",na.action=na.omit)

summary(fit2)

* 1. From Model 1 and 2, . Use the model results to confirm this.
  2. In Model 2, ,. Use the results of Model 2 to compute this correlation and compare it the value of obtained from Model 1.
  3. Compare the estimates of the population intercept and slopes for Model 1 and 2. Do you expect that these should be the same or different?
  4. Using the results of Model 2, provide an interval that includes roughly 95% of children’s expected FEV1 values at baseline.

1. The model below expands Model 2 to allow for a random intercept for each child and random slope for time for each child. Refer to this model as Model 3.

where and , for all *j not equal to k*, ,, and ,. In addition, the children *i* are independent and the within subject residuals ( are independent of the child specific effects (

Review this model and answer the following:

1. Define the population mean yearly change in FEV1 among children receiving budesonide.
2. Define a child’s mean yearly change in FEV1 if the child received budesonide.
3. Interpret within the context of the CAMP trial.
4. Suppose is positive (i.e. > 0). Interpret what this means in the context of the CAMP trial.
5. Fit the model proposed above and answer the following questions.

**STATA:**

\* Fit Model 3: random intercept and random slope for time

\* linear mixed model

mixed posfev year i.trt#c.year || id: year, cov(uns) ml

**R:**

## Model 3: random intercept and random slope for time

fit3 = lme(POSFEV ~ year+as.factor(trt):year,data=dat,na.action=na.omit,

random = ~ 1 + year | id,method=”ML”)

summary(fit3)

* 1. Provide an interval that contains roughly 95% of children’s expected FEV1 values at baseline.
  2. Provide an interval that contains roughly 95% of the expected yearly change in FEV1 among children receiving the placebo.
  3. Provide intervals that contains roughly 95% of the expected yearly change in FEV1 among children receiving budesonide and nedocromil, respectively

BUDESONIDE:

NEDOCROMIL

1. Compare the fit of Model 3 to the models you considered in Homework 1. Namely, fit the following parametric correlation models:
2. Independence
3. Exchangeable (or random intercept model)
4. Exponential
5. Random intercept + Exponential correlation structure
6. Random intercept + random slope for time

Based on the AIC statistics, which of these models is most consistent with the data?

**STATA:**

\* Fit the independence model

quietly mixed posfev year i.trt#c.year || id: , nocons residuals(ind, t(year)) ml

est store A

\* Fit the exchangeable model

quietly mixed posfev year i.trt#c.year || id: , nocons residuals(exch, t(year)) ml

est store B

\* Fit the exponential model

quietly mixed posfev year i.trt#c.year || id: , nocons residuals(exp, t(year)) ml

est store C

\* Fit the random intercept + exponential within subject correlation structure

quietly mixed posfev year i.trt#c.year || id: , residuals(exp, t(year)) ml

est store D

\* Fit the random effect models

quietly mixed posfev year i.trt#c.year || id: year, cov(uns) ml

est store E

est stats A B C D E

**R:**

# Fit the independence model

mA = gls(POSFEV~year+as.factor(trt):year,data=dat,na.action=na.omit,method="ML")

# Fit the exchangeable model

mB = gls(POSFEV~year+as.factor(trt):year,data=dat,na.action=na.omit,correlation=corCompSymm(form=~1|id),method="ML")

# Fit the exponential model

mC = gls(POSFEV~year+as.factor(trt):year,data=dat,na.action=na.omit,correlation=corExp(form=~year|id),method="ML")

# Fit the random intercept + exponential within subject correlation structure

mD = lme(POSFEV~year+as.factor(trt):year,data=dat,na.action=na.omit,

random = ~ 1 | id, correlation = corExp(form=~year|id),method="ML")

# Fit the random intercept + slope model

mE = lme(POSFEV~year+as.factor(trt):year,data=dat,na.action=na.omit,

random = ~ 1 + year | id,method="ML")

c(AIC(mA),AIC(mB),AIC(mC),AIC(mD),AIC(mE))

1. Use the formula below to compute the estimate of at 2, 12, and 48 months using the fit of Model 3. Compare these values to those obtained from the exploratory analysis conducted in Homework 1 (see Table 1).