Linear Mixed-Effects Models

Dale Barr
University of Glasgow

Model specification

- specifying fixed effects
- specifying random effects

Categorical predictors

- Recommendation: Make your own, don't rely on R defaults. Why?
  - model comparison doesn't work with variables of type `factor`
  - defaults don't support ANOVA-style interpretation

LMEM versus t-test

> You have run a study looking at the effects of alcohol consumption on simple reaction time. Data is stored in in the tables `subjects` and `simple_rt`. Subjects (`sub`) were randomly assigned to one of two groups (`cond`). One group drank alcohol before performing the task, while the other had a placebo drink.

> As a dependent variable, you measured how quickly each subject pressed a button in response to a flashing light (`RT`, in milliseconds). Each subject provided 8 measurements. Remove data from subjects S01 and S11 before analysis.

t-test on subject means (1)

simple_rt.zip

```
library("tidyverse")

subjects <- read_csv("simple_rt/subjects.csv",
                     col_types = "icc",
                     progress = FALSE)

simple_rt <- read_csv("simple_rt/simple_rt.csv",
                      col_types = "icci",
                      progress = FALSE)

combined <- subjects %>%
  filter(sub != "S01",
         sub != "S11") %>%
  inner_join(simple_rt, "sub") %>%
  select(sub, cond, RT)

subj_means <- combined %>%
  group_by(sub, cond) %>%
  summarise(mean_RT = mean(RT),
            .groups = "drop")

subj_means
```

```
# A tibble: 14 × 3
   sub   cond    mean_RT
   <chr> <chr>     <dbl>
 1 S02   placebo    514.
 2 S03   placebo    528.
 3 S04   alcohol    507
 4 S05   placebo    476.
 5 S06   alcohol    450.
 6 S07   placebo    488.
 7 S08   placebo    411.
 8 S09   alcohol    430.
 9 S10   alcohol    458.
10 S12   alcohol    537.
11 S13   alcohol    500
12 S14   placebo    434.
13 S15   placebo    393.
14 S16   alcohol    425
```

t-test on subject means (2)

```
t.test(mean_RT ~ cond,
       subj_means, var.equal = TRUE)
```

```

	Two Sample t-test

data:  mean_RT by cond
t = 0.35278, df = 12, p-value = 0.7304
alternative hypothesis: true difference in means
between group alcohol and group placebo is not equal
to 0
95 percent confidence interval:
 -46.21515  64.07230
sample estimates:
mean in group alcohol mean in group placebo 
             472.2500              463.3214 
```

Random-intercepts LMEM

Level 1:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + e_{ij}$$

Level 2:

$$\beta_0 = \gamma_{00} + S_{0i}$$

$$\beta_1 = \gamma_{10}$$

Variance Components

$$S_{0i} \sim N\left(0, {\tau_{00}}^2\right)$$

$$e_{ij} \sim N\left(0, \sigma^2\right)$$

Random intercepts with ICC

γ_{00}, fixed intercept

325 400 475

325 355 385 415 445 475

γ_{10}, fixed slope

-70 0 70

-70 -40 -10 5 20 35 50 65 70

τ_{00}, rint SD

0 15 30

0 3 6 9 12 18 24 30

σ, error SD

0 40 80

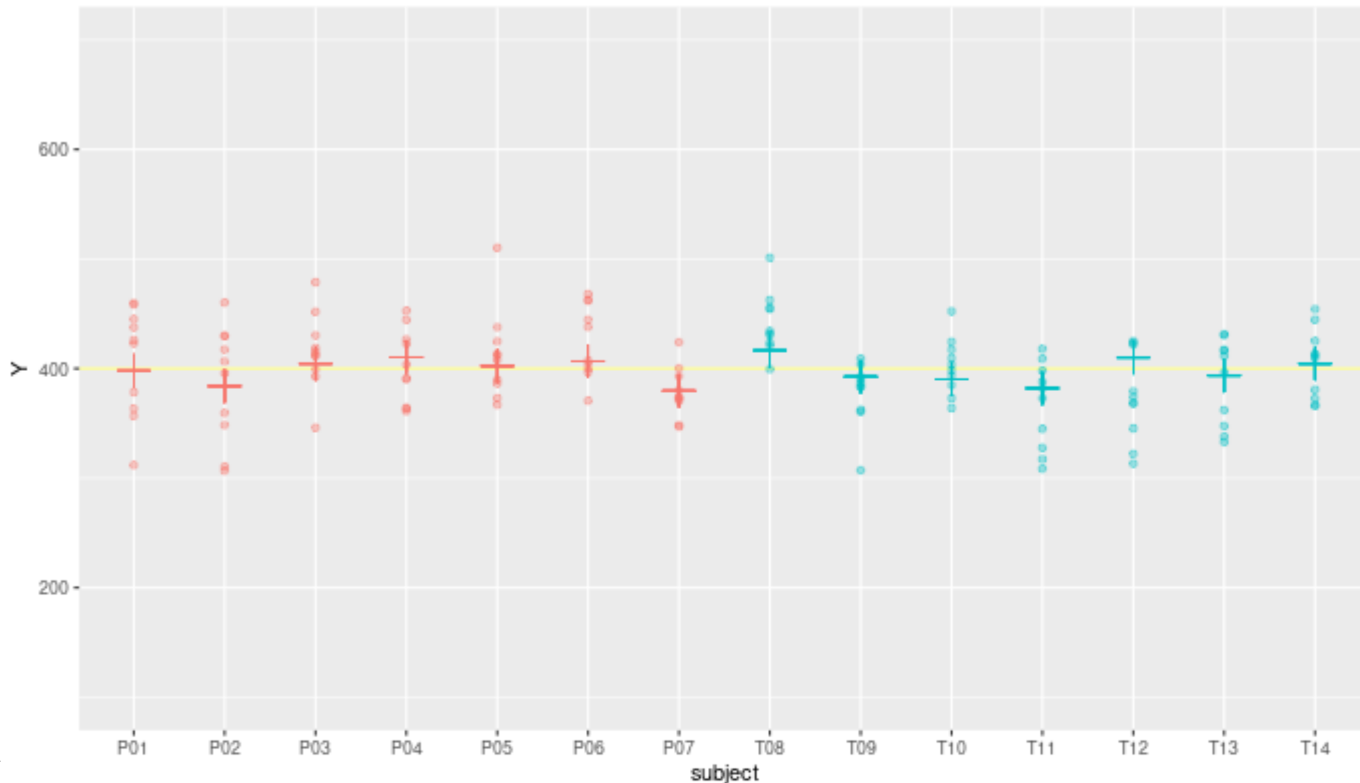
0 8 16 32 48 64 80

New Sample

Download Data (session/d9

plots | lmer output

intraclass correlation (ICC) $= \frac{\tau_{00}{}^2}{\tau_{00}{}^2+\sigma^2} = 0.123$



For multi-level data, random-intercepts linear mixed-effects modeling can replace these analyses:

between-subjects

- one-sample t-test
- independent samples t-test
- one-way ANOVA

within/mixed designs

- paired samples t-test
- repeated-measures one-way ANOVA
- fully-within factorial ANOVA
- mixed-design ANOVA

NB: one obs per factor/cell

Rules for random effects (1)

Always include random intercepts for any random factor (e.g., subjects) where you have multiple observations on the DV.

```
Y ~ (1 | subject)
```

Do I also need a random slope for factor `A`?

1. `A` is within-subjects
2. multiple observations per level of `A`

```
Y ~ A + (1 + A | subject)
```

Rules for random effects (2)

What random slopes do I need for interaction ABC?

- identify highest-order combination of within factors
- if you have multiple observations per level of that factor / per cell of those factors, then you need a random slope for that factor / interaction of factors

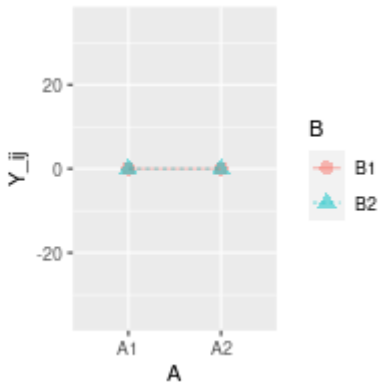
Coding factorial predictors

Scheme	A_1	A_2
Treatment (dummy)	0	1
Sum	-1	1
Deviation	$-\frac{1}{2}$	$\frac{1}{2}$

Choice of a coding scheme impacts interpretation of:

1. the intercept term; and
2. the interpretation of the tests for all but the highest-order effects and interactions in a factorial design.

Main Effects and Interactions: 2x2 Factorial



Intercept:

-10 0 10

-10 -8 -6 -4 -2 0 2 4 6 8 10

Main Effect of A:

-20 0 20

-20 -16 -12 -8 -4 0 4 8 12 16 20

Main Effect of B:

-20 0 20

-20 -16 -12 -8 -4 0 4 8 12 16 20

AB Interaction:

-80 0 80

-80 -64 -48 -32 -16 0 16 32 48 64 80

Reset

ANOVA

Cell and Marginal Means

	B1	B2	
A1	0	0	0
A2	0	0	0
	0	0	NA

Simple Effects

eff	
A@B1	0
A@B2	0

eff	
B@A1	0
B@A2	0

Decomposition of Cell Means

cell	Y_ij	int	A_i	B_j	AB_ij
A1B1	0	0	0	0	0
A1B2	0	0	0	0	0
A2B1	0	0	0	0	0
A2B2	0	0	0	0	0

Regression

Coding of A

Dummy (A1=0, A2=1)

Coding of B

Dummy (B1=0, B2=1)

$$Y_{ijk} = \beta_0 + \beta_1 A_i + \beta_2 B_j + \beta_3 AB_{ij} + e_{ijk}$$

$$Y_{ijk} = 0 + 0A_i + 0B_j + 0AB_{ij} + e_{ijk}$$

cell	Y_ij	b0	b1	A_i	b2	B_j
A1B1	0	-0.00	0.00	0.00	-0.00	0.00
A1B2	0	-0.00	0.00	0.00	-0.00	1.00
A2B1	0	-0.00	0.00	1.00	-0.00	0.00
A2B2	0	-0.00	0.00	1.00	-0.00	1.00