ISN Methods Meeting 11.5.2021

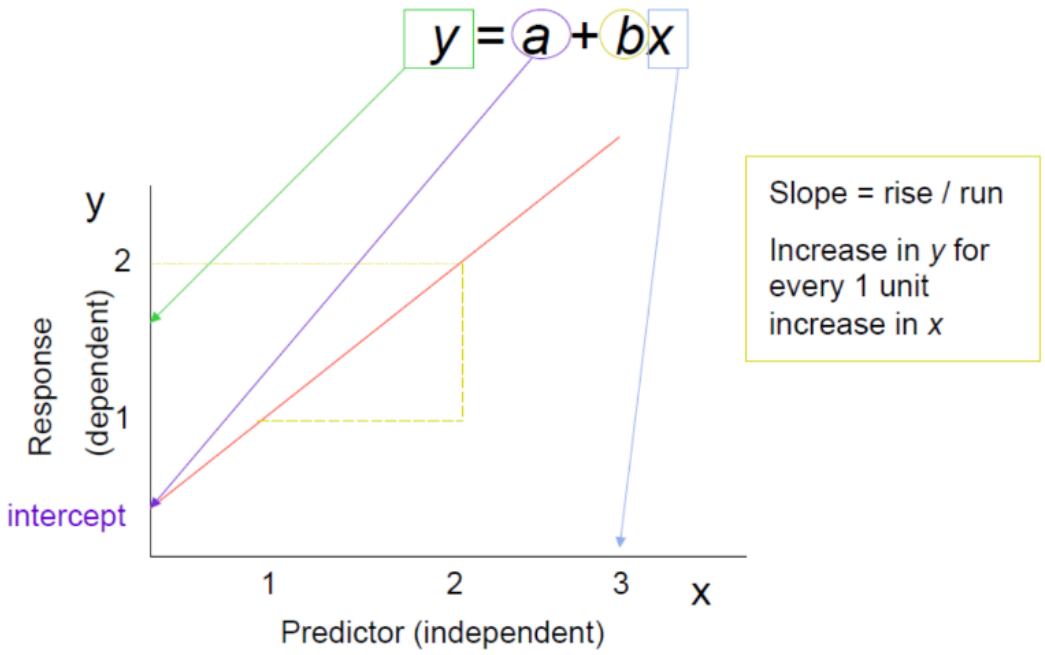
Linear Mixed Effects Models (LMEs)

Karita Ojala

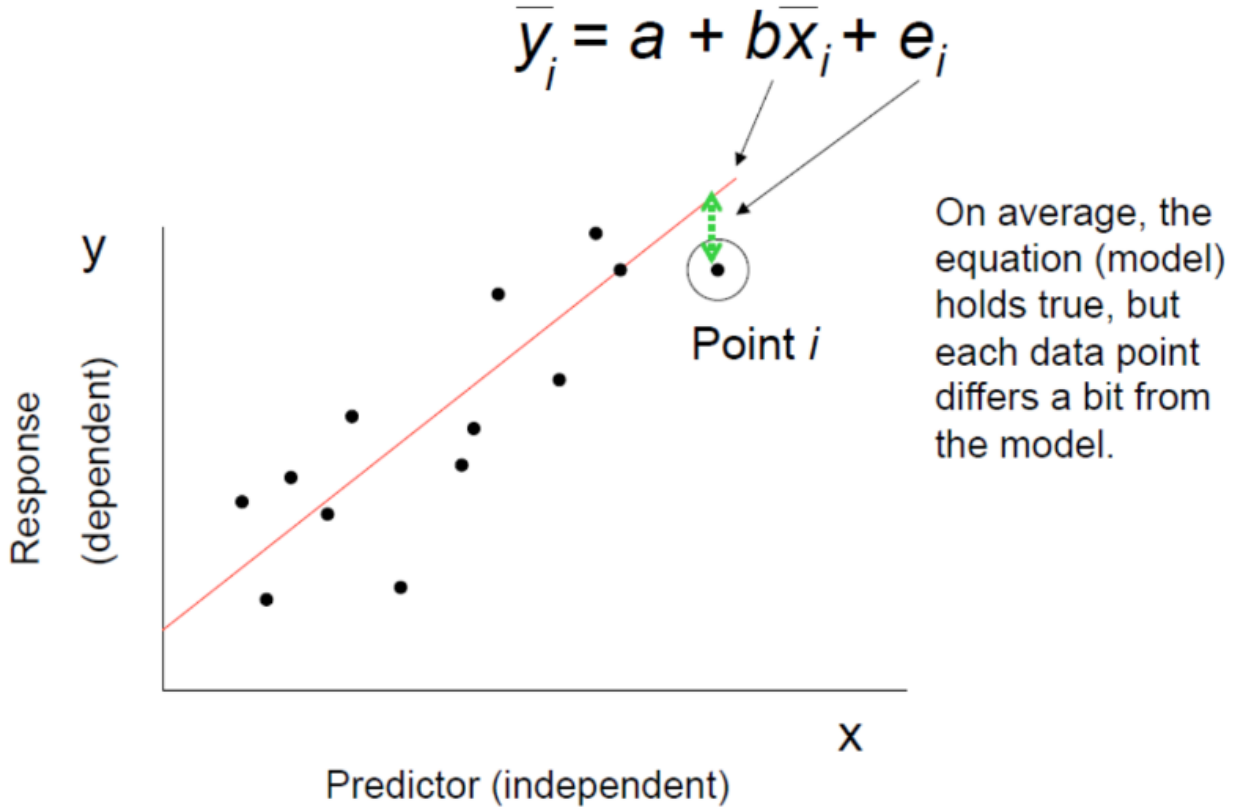
General Linear Model

- Linear mixed effect models (LMEs) are special cases of the general linear model, as are linear regression and ANOVAs
- Simplified GLM formula:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_i X_i + \epsilon$$



General Linear Model



https://github.com/RoseannaGG/LinearMixedEffectsModels

Why use LMEs?

Fixed and random effects

Mixed effects

- We want the best explanation (model) for the data
- That is, in statistical terms, to explain the most variance in the data with our statistical model

- In general, the data we deal with comes from experimental paradigms that have both **fixed** and **random** effects:

  - **Fixed effects**: predetermined categories of a variable / fixed (non-random) group means

    e.g. sex, age, treatment group

  - **Random effects**: randomly sampled categories of a variable / random group means

    e.g. participant, repeated measures (time/day, drug/dose), trial, stimulus, also nested/hierarchical: class in a school in a district

  - **Mixed effects = fixed + random effects**

Why use LMEs?

Fixed and random effects

Mixed effects

- Statistics are calculated differently for fixed and random effects (e.g. F-value)
  - Repeated measures within an individual, or other random factors e.g. pupils in a class in a school in a district are not independent!
  - For fixed effects, residual error variance (sampling error) is the only source of random variation
  - For random effects, there is an additional source of random variation from the interaction of the random sample (e.g. patients/participants) with the effect of interest (e.g. treatment)
- Both fixed and random effects can be either of interest or something to "explain away" (e.g. age/sex, or for random effects e.g. study site)
- In any case, relevant factors should be included in the statistical model
- But: overcomplicated models may not converge especially if there is little data per each level of a factor
  - e.g. few stimuli or trials

Why use LMEs?

Fixed and random effects

Mixed effects

- Linear mixed effect models contain both fixed and random effects
- LMEs can handle unequal sample sizes
- Easy to implement in R
- Can obtain ANOVA-like results tables

- Assumptions of LMEs:

Core assumption: residuals and random effect coefficients are independent and identically distributed (i.i.d.)

"Overall, our results show remarkable robustness of mixed-effects models that should allow researchers to use mixed-effects models even if the distributional assumptions are objectively violated. However, this does not free researchers from careful evaluation of the model. Estimates that are based on data that show clear violations of key assumptions should be treated with caution because individual datasets might give highly imprecise estimates, even if they will be unbiased on average across datasets."

Schielzeth et al., 2020. Methods Ecol Evol.

R packages for LMEs

nlme
lme4

- Essentially, there are two choices for good R packages for LMEs: *lme4* and *nlme*
- Some differences how models are constructed (syntax), in estimation, and compatibility with other packages
- lme4 is older and perhaps more widely used (and may have some compatibility with other packages that nlme doesn't) but I have chosen to use nlme as it's newer and it worked better with my data/paradigm

Practical example with *nlme*

Defining the model in *nlme*

nlme::lme(y ~ x * w, random = ~ 1|a, data = dataframe)

x, w = fixed effects (two different ones)

1|random effect = random intercept

0+x|random effect = random slope, no random intercept

1+x|random effect = random slope, random intercept

x|random effect = random slope, random intercept

x-1|random effect = random slope and intercept, uncorrelated

(1|random effect a) + (1|random effect b) = random intercepts for two different random effects a and b

Extensive list see:
https://github.com/RoseannaGG/LinearMixedEffectsModels
BUT take care, some of this syntax works for lme4 but maybe not for nlme – google exact syntax for each package

Practical example with *nlme*

Example dataset

Experiment: TMS manipulation to interfere with fear memory

Fixed effects:

- Group (TMS hemisphere): experimental or control
- CS type: CS+ (predicts shock) or CS− (predicts safety)
- CS complexity: simple or complex
- Trial: factor with values 1 to 24

  Each trial is of either CS type and either CS complexity and there are only 6 trials of each of the 4 types

Random effects:

- Subject: number 1 to N where N is sample size

  Sample sizes of the groups are unequal

Bayesian Model Comparison

for frequentist models in R

How to select the best model out of all alternatives?

- The good old Bayes' theorem:

$$\Pr(M|D) = \frac{\Pr(D|M)\Pr(M)}{\Pr(D)}$$

- Bayes Factor for comparing models (or hypotheses):

$$K = \frac{\Pr(D|M_1)}{\Pr(D|M_2)} = \frac{\Pr(M_1|D)}{\Pr(M_2|D)}\frac{\Pr(M_2)}{\Pr(M_1)}$$

Penalty for model complexity

K	dHart	bits	Strength of evidence
$< 10^0$	< 0	< 0	Negative (supports M_2)
10^0 to $10^{1/2}$	0 to 5	0 to 1.6	Barely worth mentioning
$10^{1/2}$ to 10^1	5 to 10	1.6 to 3.3	Substantial
10^1 to $10^{3/2}$	10 to 15	3.3 to 5.0	Strong
$10^{3/2}$ to 10^2	15 to 20	5.0 to 6.6	Very strong
$> 10^2$	> 20	> 6.6	Decisive

Bayesian Model Comparison

for frequentist models in R

(including LMEs)

- BayestestR package in R for both Bayesian and frequentist models (approximation)

- Importantly, do not blindly follow model comparison! I only used it to have a tool to think about which random effects structure I should choose as my main analysis

- Rather think first: what is the most reasonable model based on my research questions, paradigm and type of data?

Materials and references

- **Extensive tutorials into LMEs:**
  https://m-clark.github.io/mixed-models-with-R/introduction.html

- Roseanna Gamlen-Greene, Linear Mixed Effects Models: Intro & examples in R with lmer
https://github.com/RoseannaGG/LinearMixedEffectsModels

- Schielzeth, H, Dingemanse, NJ, Nakagawa, S, et al. Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods Ecol Evol*. 2020; 11: 1141–1152. https://doi.org/10.1111/2041-210X.13434