Lecture 1

You've seen fixed effect models
- we'll now study random effect models
& mixed effect (random + fixed) models.

Fixed effects are used for effects of direct interest

Random effects are used for effects not of direct interest
↳ interest is in the underlying population.

Crop data

Y_{ij} = yield of crop on farm i in year j

farm \ year	1	2	3	4	5
A	Y_{A1}	Y_{A2}	...		Y_{A5}
B	Y_{B1}				
C	Y_{C1}				Y_{C5}

What model would you fit? lm(Yield ~ farm - 1)

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

where $\mathbb{E}\,\varepsilon_{ij} = 0$
$Var(\varepsilon_{ij}) = \sigma^2$
ε_{ij} are uncorrelated

One way ANOVA

μ_i = average yield on farm i $\overset{\shortparallel}{\alpha+\mu_i}$

Equivalent models are lm(Yield ~ farm)

$$Y_{ij} = \begin{cases} \alpha + \varepsilon_{ij} & \text{if farm } A \\ \alpha + \mu_i + \varepsilon_{ij} & \text{if farm} \neq A \end{cases}$$

Or $Y_{ij} = \alpha + \mu_i + \varepsilon_{ij}$ $\alpha + \mu_i$ = av. yield for farm i

where $\sum_i \mu_i = 0$ lm(Yield ~ farm, contrasts = contr.sum)

This is a fixed effects model

↳ it is not capable of predicting the yield on new farms. It can only tell us about farms A to F.

Random effect models

$$Y_{ij} = \overset{\text{fixed effect}}{\mu} + b_i + \varepsilon_{ij}$$

where $b_i \sim N(0, \sigma_b^2)$ & $\varepsilon_{ij} \sim N(0, \sigma^2)$

random effect

between farm variance i.e. differences between farms

within farm variance i.e. differences between years on a single farm.

Use lme4 package in R

lmer(Yield ~ 1 + (1 | Farm))

fixed effect just an intercept for this model

random effect part in (brackets)

(model structure | Grouping)

(intercept | Farm)