Repeated measures and mixed models

Session 10

MATH 80667A: Experimental Design and Statistical Methods
for Quantitative Research in Management
HEC Montréal

Outline

Why repeated measures?

Repeated measures

Mixed models

Why repeated measures?

Beyond between-designs

Each subject (experimental unit) assigned to a single condition.

- individuals (subjects) are **nested** within condition/treatment.

In many instances, it may be possible to randomly assign multiple conditions to each experimental unit.

Benefits of within-designs

Assign (some or) all treatments to subjects and measure the response.

Benefits:

- Filter out effect due to subject (like blocking):
  - increased precision of effect sizes
  - increased power (tests are based on within-subject variability)
- Each subject (experimental unit) serves as its own control (greater comparability among treatment conditions).

Impact: need smaller sample sizes than between-subjects designs

Drawbacks of within-designs

- Potential sources of bias
  - Period effect (e.g., practice or fatigue)
  - Carryover effects
  - Permanent change in the subject condition after a treatment
  - Loss of subjects over time

Minimizing sources of bias

- Randomize the order of treatment conditions among subjects
- or use a balanced crossover design and include the period and carryover effect in the statistical model (confounding or control variables to better isolate the treatment effect).
- Allow enough time between treatment conditions to reduce or eliminate period or carryover effects.

Repeated measures

Exhaustive or small subsample?

So far, we consider factors (treatment factor, blocking) as **fixed**

- Meaning their effect is constant

Change of scenery

Assume that the levels of a factor form a random sample from a large population

Fixed vs random: no clear definition

Gelman (2005) lists a handful of definitions:

> When a sample exhausts the population, the corresponding variable is fixed; when the sample is a small (i.e., negligible) part of the population the corresponding variable is random [Green and Tukey (1960)].
>
> Effects are fixed if they are interesting in themselves or random if there is interest in the underlying population (e.g., Searle, Casella and McCulloch [(1992), Section 1.4])

One-way ANOVA with a random effect

As before, we have one experimental factor A with a levels, with

$$\underset{\text{response}}{Y_{ij}} = \underset{\text{global mean}}{\mu} + \underset{\text{mean difference}}{\alpha_j} + \underset{\text{random effect for subject}}{S_i} + \underset{\text{error}}{\varepsilon_{ij}}$$

where $S_i \sim \mathsf{No}(0, \sigma_s^2)$ and $\varepsilon_{ij} \sim \mathsf{No}(0, \sigma_e^2)$ are random variables.

The errors and random effects are independent from one another.

The model **parameters** are μ, σ_s^2 and σ_e^2.

Variance components

- The global average is μ.
- The variance of the response Y_{ij} is $\sigma^2_s + \sigma^2_e$.
- The **intra-class correlation** between observations in group i is $\sigma^2_s/(\sigma^2_s + \sigma^2_e)$.

This dependence structure within group is termed **compound symmetry**.

Example: happy fakes

An experiment conducted in a graduate course at HEC gathered electroencephalography (EEG) data.

The response variable is the amplitude of a brain signal measured at 170 ms after the participant has been exposed to different faces.

Repeated measures were collected on 9 participants, but only the average of the 34 replications is provided.

Experimental conditions

The control (`real`) is a true image, whereas the other were generated using a generative adversarial network (GAN) so be slightly smiling (`GAN_S`) or extremely smiling (`GAN_E`, looks more fake).

Research question: do the GAN image trigger different reactions (pairwise difference with control)?

Models for repeated measures

We have $r=1$ replication per participant for each condition. In this specific case, the repeated-measures ANOVA model amounts to a randomized block, i.e.,

- `participant` (blocking factor)
- `condition` (experimental factor)

For balanced designs, we can use `aov` in **R**.

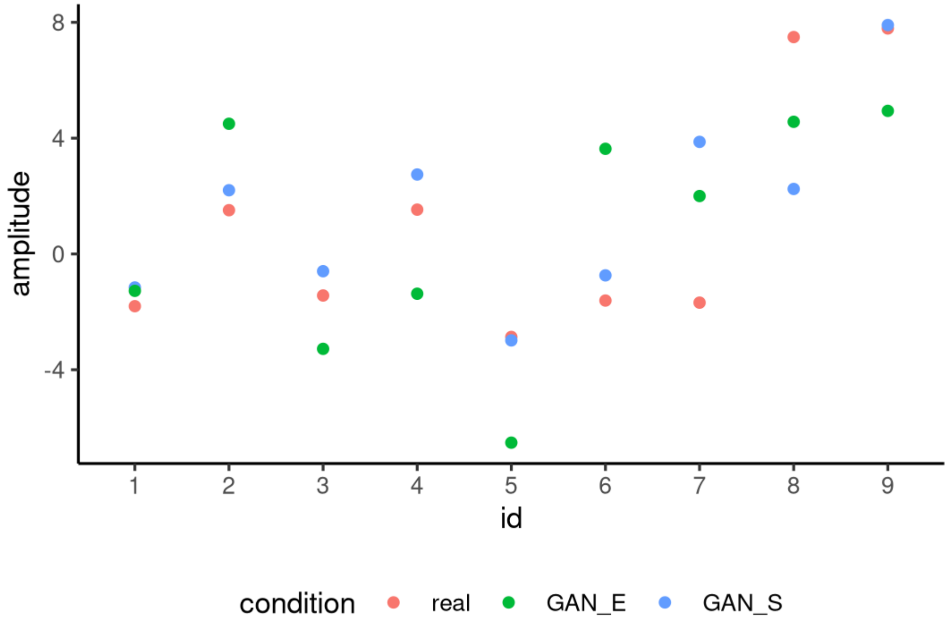
This approach has a drawback: variance estimates can be negative...

Load data | Graph | ANOVA | QQ plots

```r
library(tidyverse)
library(lme4)
library(lmerTest)
options(contrasts = c("contr.sum", "contr.poly"))
url <- "https://edsm.rbind.io/data/faces.csv"
faces <- read.csv(url, header = TRUE,
                  stringsAsFactors = TRUE) %>%
  mutate(id = factor(participant),
         condition = relevel(condition, ref = "real"))
# Declare participant ID as categorical
```

Load data | **Graph** | ANOVA | QQ plots

```
library(tidyverse)
ggplot(data = faces,
       aes(x = id,
           group = condition,
           colour = condition,
           y = amplitude)) +
  geom_point()
```



Load data Graph **ANOVA** QQ plots

```r
anova_model <- aov(amplitude ~ condition + Error(id), data = faces)
# Random intercept for participant
model <- lme4::lmer(amplitude ~ condition + (1 | id),
                    data = faces)
car::Anova(model, test = "F", type = 3)
```

```
## Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)
##
## Response: amplitude
##                  F Df Df.res Pr(>F)
## (Intercept) 0.6051  1   12.5 0.4511
## condition   0.2497  2   16.0 0.7820
```

- No detectable difference between conditions.
- The *p*-value (0.782) for the mixed model is the same as `aov`.
- Residual degrees of freedom is $(a-1) \times (n-1) = 18$ for $n = 9$ subjects and $a = 3$ levels.

Load data Graph ANOVA QQ plots

Model assumptions

The validity of the F null distribution relies on the model having the correct structure.

- Same variance per observation
- equal correlation between measurements of the same subject
- normality of the random effect

- Since we care only about differences in treatment, can get away with a weaker assumption than compound symmetry
  - *sphericity*: variance of difference between treatment is constant

Testing for sphericity

Popular two-stage approach:

- Mauchly's test of sphericity
  - if statistically significant, use a correction
  - if no evidence, proceed with F test as usual

Corrections for sphericity

An idea due to Box is to correct the degrees of freedom from $\mathsf{F}\{a-1,(a-1)(n-1)\}$ to $\mathsf{F}\{\epsilon(a-1),\epsilon(a-1)(n-1)\}$ for $\epsilon<1$

- Since the statistic is a ratio, it is unaffected
- Three widely used corrections:
  - Greenhouse-Geisser
  - Huynh-Feldt (more powerful, but can be larger than 1 - cap)
  - lower bound with $\epsilon=(a-1)^{-1}$, giving $\mathsf{F}(1,n-1)$.

Another option is to go fully multivariate.

Mixed models

Generalization

Using mixed models in place of *old school* ANOVA has benefits in that it's easier to account for complex designs.

In general, things are not obvious

- Estimation via restricted maximum likelihood
- Theory for testing is more complicated
  - F-tests via Kenward-Rogers (best, but costly) or Satterthwaite approximation
  - Determining the degrees of freedom is not always trivial (Hass diagrams)
- For more layers, need replications to estimate variability (estimability/identifiability)

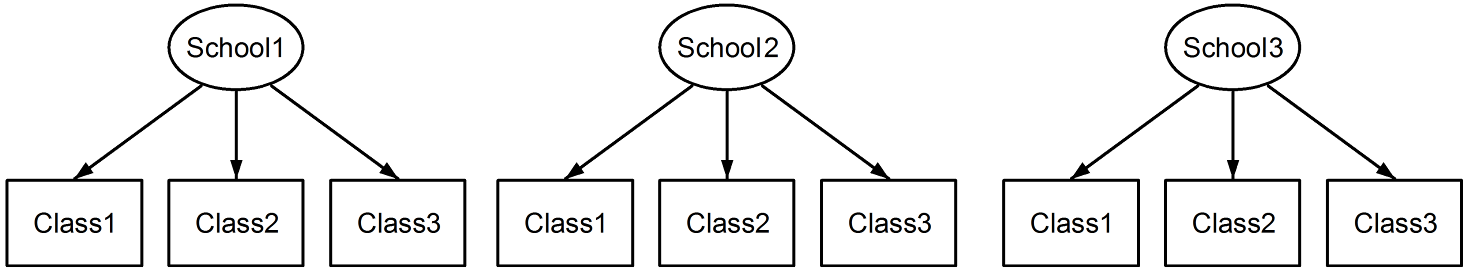
Nested versus crossed

Nested effects if a factor appears only within a particular level of another factor.

Crossed is for everything else (typically combinations of all factors).

Example of nested random effects: class nested within schools

- class 1 is not the same in school 1 than in school 2



Formulae for nested effects

R uses the following notation for nested effect: `group1/group2`, to mean `group2` is nested within `group1`. This formula expands to `group1 + group1:group2`

For crossed effects, use rather `group1*group2` which expands to `group1 + group2 + group1:group2`.

In package `lme4`, a random intercept per group is written `(1 | group1/group2)`.

Demo and examples