# Mixed Models for Clustered & Repeated Data

## Separate subject-specific variation from population-level effects

Rverse Analytics

**Dependence belongs in the model.** Repeated visits within patients and patients within centres violate ordinary regression independence; mixed models represent that structure explicitly.

## Model map

| Outcome | Model | R function | Typical effect |
|---|---|---|---|
| Continuous | linear mixed model | `lme4::lmer()` | adjusted mean difference |
| Binary | logistic mixed model | `lme4::glmer(..., binomial)` | conditional OR |
| Count | Poisson / NB mixed model | `glmer(..., poisson)` / `glmmTMB::glmmTMB()` | conditional rate ratio |
| Correlated residual pattern | generalized least squares | `nlme::lme()` | mean difference + residual structure |
| Population-average target | GEE | `geepack::geeglm()` | marginal effect |

## Long data first

```
library(dplyr)
library(lme4)
library(emmeans)

long <- d |>
  mutate(
    id = factor(id),
    group = relevel(factor(group), ref = "Control"),
    visit = factor(visit, levels = c("Baseline", "Week 6", "Week 12")),
    time = c(0, 6, 12)[visit]
  )

stopifnot(!anyDuplicated(long[c("id", "visit")]))
with(long, table(group, visit, useNA = "ifany"))
```

## Random intercept and random slope

```
# Each patient gets a personal baseline
m0 <- lmer(score ~ group * visit + (1 | id),
           data = long, REML = TRUE, na.action = na.exclude)

# Each patient gets a baseline and a time trajectory
m1 <- lmer(score ~ group * time + (time | id),
           data = long, REML = FALSE, na.action = na.exclude)

# Compare fixed random-effects structures with ML, not REML
anova(update(m0, REML = FALSE), m1)

summary(m1)
performance::icc(m1)
performance::check_singularity(m1)
```

| Syntax | Meaning |
|---|---|
| `(1 \| id)` | patient-specific intercept |
| `(time \| id)` | correlated intercept and time slope |
| `(time \|\| id)` | uncorrelated intercept and slope |
| `(1 \| center/id)` | patients nested within centres |
| `(1 \| patient) + (1 \| clinician)` | crossed random intercepts |

Do not delete a random slope only because its variance is small. Check design support, convergence, singularity, the estimand and sensitivity of fixed-effect inference.

## Marginal means and planned contrasts

```r
emm <- emmeans(m0, ~ group | visit)
emm
pairs(emm, adjust = "holm")

# Change from baseline by group and difference-in-differences
emm_change <- emmeans(m0, ~ group * visit)
contrast(emm_change, interaction = "revpairwise", adjust = "holm")

# Response-scale probability from a logistic mixed model
g1 <- glmer(event ~ group * visit + (1 | id),
            data = long, family = binomial)
emmeans(g1, ~ group | visit, type = "response")
```

The group × visit term tests whether trajectories differ. It is not automatically the treatment effect at the final visit; use a planned contrast on the fitted marginal means.

## Diagnostics and inference

```r
plot(m0)                           # residuals vs fitted
qqnorm(resid(m0)); qqline(resid(m0))
qqnorm(ranef(m0)$id[[1]])
performance::check_model(m0)

# Profile or bootstrap CIs are preferable for variance components
confint(m0, method = "profile")

# Parametric bootstrap comparison when asymptotics are doubtful
pb <- pbkrtest::PBmodcomp(full_model, reduced_model, nsim = 1000)
```

## Common design decisions

| Decision | Ask first | Practical consequence |
|---|---|---|
| Visit as factor or numeric | nonlinear visits or linear trend? | separate visit effects vs one slope |
| Random slope | enough repeated time points per unit? | models trajectory heterogeneity |
| REML or ML | estimating final model or comparing fixed effects? | REML final; ML fixed-effect comparison |
| Mixed model or GEE | subject-specific or population-average effect? | different estimands, especially logistic |
| Residual covariance | serial correlation beyond random effects? | consider `nlme::lme()` |

## Reporting checklist

- Describe levels, cluster counts, observations per cluster and timing.
- State fixed effects, random effects, covariance structure, estimator and software.
- Report adjusted marginal means/contrasts with 95% CIs, not interaction coefficients alone.
- Document convergence, singularity, residual diagnostics and sensitivity models.
- Explain the missing-outcome assumption; mixed models do not automatically fix MNAR data.

---