FIT, PREDICT AND CHECK LONGITUDINAL MODELS

WEEK 11 · SUPPORTED WORKSHOP

Gavin Simpson
Aarhus University

Wednesday, 18 November 2026

READ THE SAME CALF STUDY

Researchers assign Control or Enhanced feed to each of 32 calves, balancing the two diets within each of four rooms. They weigh each calf at weeks 0, 2, 4, 6 and 8. These invented records illustrate longitudinal data: measurements collected from the same animals over time.

```
library("dplyr")

library("lme4")
library("performance")
library("marginaleffects", warn.conflicts = FALSE)
library("ggplot2")
d <- readr::read_csv("data/calf-growth.csv")
d <- d |>
  mutate(
    diet = factor(diet, levels = c("Control", "Enhanced")),
    room = factor(room),
    calf = factor(calf),
    week = time_week
  )
head(d |> select("calf", "diet", "week", "weight_kg"), 3)
```

`week` counts weeks since researchers assigned the diets. They record week 0 before feeding starts.

READ THE SAME CALF STUDY

```
# A tibble: 3 × 4
  calf  diet     week weight_kg
  <fct> <fct>   <dbl>     <dbl>
1 C01   Control     0      47.9
2 C01   Control     2      58.6
3 C01   Control     4      65.8
```

WORKSHOP: USE THE PACKAGES DIRECTLY

Open `week-11.Rproj` and run `longitudinal-student.R` one section at a time.

- `lme4`: fit the model.
- `performance`: inspect model checks.
- `marginaleffects`: expected weights and diet differences.
- `ggplot2`: plot the evidence.

Use supplied results if computing fails. Rotate driver, translator and design auditor.

FIT THE TWO GROWTH RELATIONSHIPS

```
linear <- lmer(
  weight_kg ~ diet * week + room + (1 | calf),
  data = d,
  REML = TRUE
)
curved <- lmer(
  weight_kg ~ diet * week + I(week^2) + room + (1 | calf),
  data = d,
  REML = TRUE
)
# Compare mean relationships by ML; retain REML fits for estimation.
linear_ml <- update(linear, REML = FALSE)
curved_ml <- update(curved, REML = FALSE)
AIC(linear_ml, curved_ml)
```

```
          df      AIC
linear_ml  9 774.4759
curved_ml 10 747.5685
```

Both fits use the same calf records and shared calf differences. Which fit allows growth to slow, and how does its AIC compare?

ASK PERFORMANCE TO INSPECT THE FIT

```
check_singularity(curved)
```

```
[1] FALSE
```

```
check_normality(curved)
```

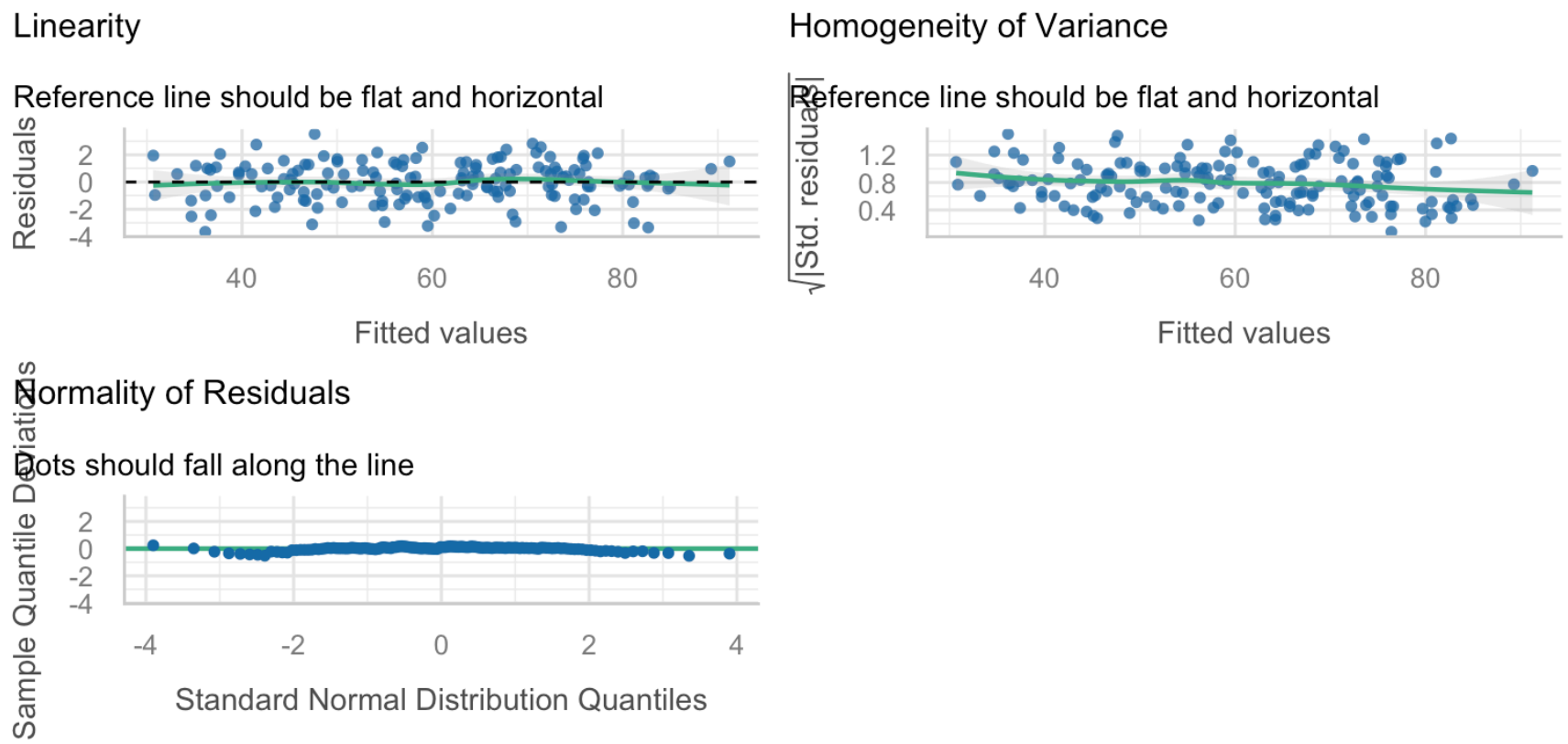
```
OK: residuals appear as normally distributed (p = 0.268).
```

The first check asks whether the estimated spread of calf differences is nearly zero; `FALSE` means it is not. The second checks the shape of the remaining errors: observed weights minus fitted weights.

Neither result proves that errors at nearby visits are unrelated.

LOOK AT THE CHECKS, NOT ONLY THEIR LABELS

```
check_model(curved, check = c("qq", "linearity", "homogeneity"))
```



Errors are observed minus fitted weights, also called residuals. These panels examine their shape, patterns against fitted weight and changing spread.

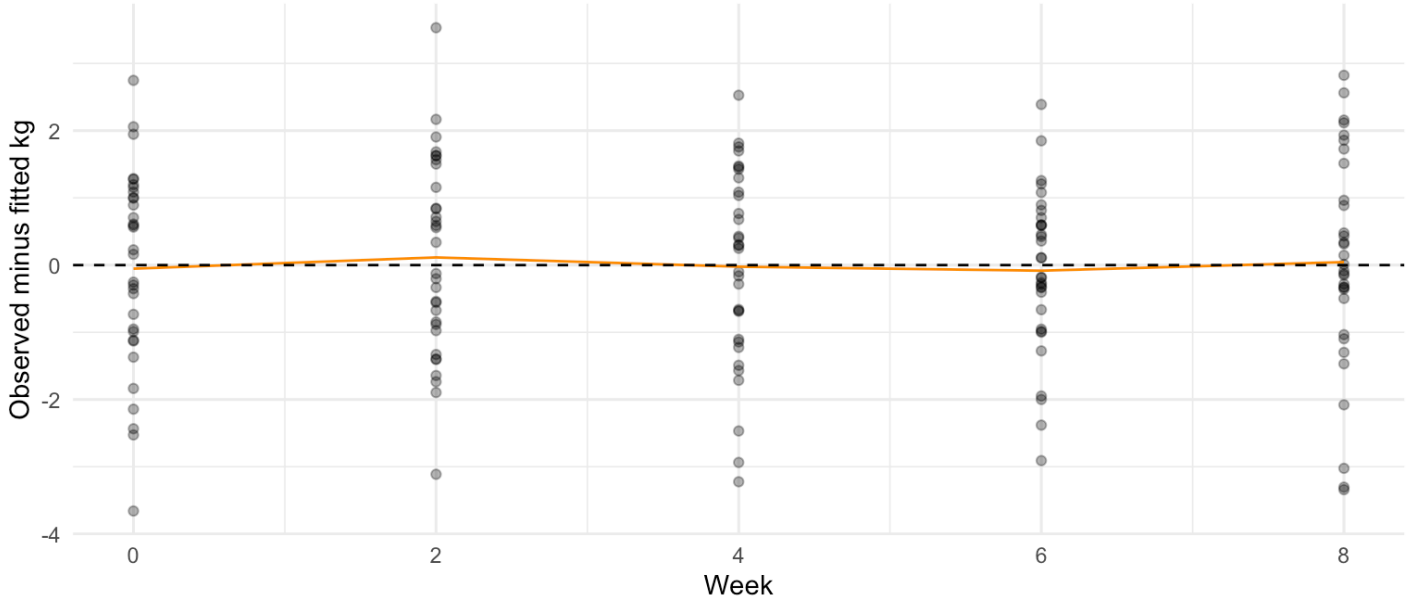
ADD THE QUESTION SPECIFIC TO GROWTH

```
library("dplyr")

d <- d |>
  mutate(
    error_kg = residuals(curved, type = "response")
  )
ggplot(d, aes(week, error_kg)) +
  geom_point(alpha = .35) +
  stat_summary(fun = mean, geom = "line", colour = "darkorange") +
  geom_hline(yintercept = 0, linetype = 2) +
  labs(x = "Week", y = "Observed minus fitted kg") +
  theme_minimal()
```

Does the orange mean-error line still bend with time?

ADD THE QUESTION SPECIFIC TO GROWTH



CHOOSE THE PREDICTION CONDITIONS

We want to compare the diets for new calves, giving each of the four rooms equal weight. We leave out the fitted weight difference for any particular calf; this difference is called its deviation.

```
grid <- tidyr::crossing(
  week = c(0, 4, 8),
  diet = d$diet,
  room = d$room
) |>
  mutate(calf = d$calf[1])
head(grid, 3)
```

```
# A tibble: 3 × 4
   week diet    room  calf
  <dbl> <fct>   <fct> <fct>
1     0 Control R1    C01
2     0 Control R2    C01
3     0 Control R3    C01
```

`re.form = NA` in the next calls sets calf deviations to zero. With this normal-outcome model, that also equals averaging over new calves.

SHOW EXPECTED WEIGHTS

```
means <- avg_predictions(
  curved,
  newdata = grid,
  by = c("week", "diet"),
  re.form = NA
)
means
```

Weights and interval limits are in kg. The intervals concern expected weights, not the spread of individual calves.

SHOW EXPECTED WEIGHTS

```
 week     diet Estimate Std. Error    z Pr(>|z|)     S 2.5 % 97.5 %
    0 Control      40.0       1.39 28.7   <0.001 601.2  37.3   42.7
    0 Enhanced     40.2       1.39 28.9   <0.001 607.4  37.5   43.0
    4 Control      59.9       1.37 43.7   <0.001   Inf  57.2   62.5
    4 Enhanced     61.8       1.37 45.1   <0.001   Inf  59.1   64.5
    8 Control      76.3       1.39 54.8   <0.001   Inf  73.5   79.0
    8 Enhanced     79.9       1.39 57.4   <0.001   Inf  77.2   82.6

Type: response
```

ASK FOR THE DIET DIFFERENCE DIRECTLY

```
differences <- avg_comparisons(
  curved,
  newdata = grid,
  variables = list(diet = c("Control", "Enhanced")),
  by = "week",
  re.form = NA
)
differences
```

Enhanced minus Control, kg, with approximate 95% intervals.

ASK FOR THE DIET DIFFERENCE DIRECTLY

```
 week Estimate Std. Error     z Pr(>|z|)   S  2.5 % 97.5 %
    0    0.208       1.96 0.106   0.9154 0.1 -3.628   4.04
    4    1.918       1.92 0.997   0.3188 1.6 -1.852   5.69
    8    3.628       1.96 1.854   0.0638 4.0 -0.208   7.46

Term: diet
Type: response
Comparison: Enhanced - Control
```

OPTIONAL: NEARBY VISITS CAN SHARE EXTRA VARIATION

A calf may be heavier than predicted at one visit and again at the next. This extra relationship between errors is called correlation. Here visits are two weeks apart; `nlme::corAR1()` uses visit numbers 1–5 in that order. lme4 cannot represent this residual correlation, so the optional extension uses nlme explicitly.

```
library("dplyr")

d <- d |>
  mutate(
    visit = factor(week, levels = c(0, 2, 4, 6, 8))
  )
# lme4 cannot fit residual AR(1) correlation. This optional nlme model
# retains the calf intercept and one correlated residual process.
# visit_index counts two-week visits, so rho describes adjacent visits.
d <- d |>
  mutate(visit_index = match(week, c(0, 2, 4, 6, 8)))
library("nlme")
ar1 <- nlme::lme(
  weight_kg ~ diet * week + I(week^2) + room,
  random = ~ 1 | calf,
  correlation = nlme::corAR1(form = ~ visit_index | calf),
  data = d,
```

OPTIONAL: NEARBY VISITS CAN SHARE EXTRA VARIATION

```
Linear mixed-effects model fit by REML
  Data: d
       AIC      BIC    logLik
  737.8064 771.0691 -357.9032

Random effects:
 Formula: ~1 | calf
        (Intercept) Residual
StdDev:    5.324848  1.85135

Correlation Structure: AR(1)
 Formula: ~visit_index | calf
 Parameter estimate(s):
      Phi
0.4160279
Fixed effects:  weight_kg ~ diet * week + I(week^2) + room
                    Value Std.Error  DF   t-value p-value
(Intercept)       38.54645 2.1790112 125 17.689882  0.0000
dietEnhanced       0.19987 1.9768356  27  0.101104  0.9202
week               5.40378 0.1748254 125 30.909576  0.0000
I(week^2)         -0.10741 0.0194670 125 -5.517599  0.0000
roomR2            -0.10715 2.7224800  27 -0.039356  0.9689
roomR3             2.07494 2.7224800  27  0.762152  0.4526
```

SAME QUESTION, DIFFERENT DEPENDENCE ASSUMPTION

```
library("dplyr")

week8 <- grid |> filter(week == 8)
extension_difference <- avg_comparisons(
  ar1,
  newdata = week8,
  variables = list(diet = c("Control", "Enhanced")),
  level = 0
)
extension_difference
```

Compare with the earlier week-8 row. Keep the scientific question fixed while examining assumptions.

Break 10:25–10:40.

SAME QUESTION, DIFFERENT DEPENDENCE ASSUMPTION

```
 Estimate Std. Error    z Pr(>|z|)   S  2.5 % 97.5 %
     3.51       1.98 1.77   0.0762 3.7 -0.369   7.38

Term: diet
Type: response
Comparison: Enhanced - Control
```

```
curved_ml <- update(curved, REML = FALSE)
ar1_ml <- update(ar1, method = "ML")
AIC(curved_ml, ar1_ml)
```

```
          df      AIC
curved_ml 10 747.5685
ar1_ml    11 740.5392
```

YOUR OWN PROJECT: 10:40–11:15

Retrieve your saved question, design and analysis plan.

- Explain the outcome type and grouping before fitting.
- Use the packages directly to fit, check and estimate a useful quantity.
- Record one result, one concern and one justified alternative.

Your project may have no repeated measurements.

REPORTS AND EXIT: 11:15–11:30

Each group gives a preliminary result and one concern.

Individually explain:

1. Why 160 weights are not 160 diet assignments.
2. Why a better growth curve may leave dependence unresolved.

Save your work for the Unit 12 workshop at 12:30.