MEANINGFUL ESTIMATES AND SCIENTIFIC INTERPRETATION

WEEK 12 · ESSENTIAL TEACHING IN CLASS

Gavin Simpson
Aarhus University

Wednesday, 25 November 2026

TODAY: SCIENTIFIC TARGETS AND COMMUNICATION

Time	Work
12:30–13:15	Opening recall and teaching
13:15–13:40	Separate cress teaching example
13:40–13:55	Break
13:55–14:55	Figure, paragraph and peer review
14:55–15:10	Break
15:10–16:00	Your project decisions and exit

OPENING RECALL: WHOSE OUTCOME?

Recall the calf study: researchers assign a diet to each calf and weigh it five times. Then consider the study in your own project.

- To which animals or groups did researchers assign treatments? If they only observed animals, which animals did they follow?
- What did researchers measure, and which measurements came from the same animal, pen or tray?
- What outcome and comparison would answer the question?
- What information is still needed?

COMPLETE THE TARGET SENTENCE

For example: we compare **mist minus pulse watering** in the **chance that a seed sprouts by day 10**, giving **each sowing batch equal weight** and **setting the tray deviation to zero**.

For your study, name the comparison, units, time, animals or groups, and what you hold fixed or average over.

Choose the scientific target before choosing the analysis.

DIFFERENCES, RATIOS AND RISK UNITS

Imagine that 40% of seeds sprout with pulse watering and 55% with mist watering. These illustrative values are separate from the fitted cress results.

- Difference: **15 percentage points**.
- Ratio: **1.375**.
- Relative increase: **37.5%**.

If the two methods have the same chance of sprouting, their difference is **0** and their ratio is **1**.

Odds compare the chance of sprouting with the chance of not sprouting. An odds ratio compares those odds between methods.

AN INVENTED CRESS EXPERIMENT

Researchers sow **25 seeds in each of 48 trays**, in six batches.

In each batch, they randomly choose four trays for pulse watering and four for mist watering. They count sprouted seeds on **day 10** and record one count for each tray.

Synthetic teaching data; no published or biological finding implied.

STEP 1: INSPECT THE GROUPED COUNTS

```
library("dplyr")

library("lme4")
library("performance")
library("marginaleffects", warn.conflicts = FALSE)
trays <- readr::read_csv("../assets/week-12/data/cress-trays.csv")
trays <- trays |>
  mutate(
    watering = factor(watering, levels = c("pulse", "mist")),
    batch = factor(batch),
    tray_id = factor(tray_id)
  )
head(trays |> select("tray_id", "watering", "seeds_sown", "seeds_sprouted"), 3)
```

One row is one tray. Pulse is the reference watering condition.

STEP 1: INSPECT THE GROUPED COUNTS

```
# A tibble: 3 × 4
  tray_id watering seeds_sown seeds_sprouted
  <fct>   <fct>         <dbl>          <dbl>
1 T01     mist             25             11
2 T02     mist             25             17
3 T03     mist             25             12
```

STEP 2: FIT THE TRAY COUNT MODEL

Each tray can have its own tendency to sprout, around the watering and batch pattern.

```
fit <- glmer(
  cbind(seeds_sprouted, seeds_sown - seeds_sprouted) ~
    watering + batch + (1 | tray_id),
  data = trays,
  family = binomial()
)
summary(fit)
```

The binomial model describes sprouts out of 25 seeds. `(1 | tray_id)` allows a shared tray difference, called a random intercept. The coefficient compares log-odds, not probabilities: odds are the chance of sprouting divided by the chance of not sprouting.

STEP 2: FIT THE TRAY COUNT MODEL

```
Generalized linear mixed model fit by maximum likelihood (Laplace
  Approximation) [glmerMod]
 Family: binomial  ( logit )
Formula: cbind(seeds_sprouted, seeds_sown - seeds_sprouted) ~ watering +
    batch + (1 | tray_id)
   Data: trays

      AIC       BIC    logLik -2*log(L)  df.resid
    287.6     302.6    -135.8     271.6        40

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.43101 -0.29504 -0.05139  0.49570  1.04567

Random effects:
 Groups  Name        Variance Std.Dev.
 tray_id (Intercept) 0.3581   0.5984
Number of obs: 48, groups:  tray_id, 48

Fixed effects:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.477441   0.279414  -1.709   0.0875 .
wateringmist  0.435295   0.211707   2.056   0.0398 *
batchB2       0.068796   0.364614   0.189   0.8503
batchB3       0.435833   0.365244   1.193   0.2328
batchB4       0.001746   0.365056   0.005   0.9962
```

CHECK THE COUNT MODEL

```
set.seed(1212)
check_overdispersion(fit)
```

```
# Overdispersion test (using simulated residuals)

 dispersion ratio = 1.203
          p-value = 0.168
```

No overdispersion detected.

```
check_singularity(fit)
```

```
[1] FALSE
```

Do counts vary more than the fitted model expects? Has the estimated tray variation collapsed to zero?

These checks do not establish that every model assumption is correct.

CHOOSE WHICH TRAYS AND BATCHES

Trays can have higher or lower chances of sprouting than their watering-and-batch pattern predicts. We call this a tray deviation. Set it to zero, predict the chance in each batch, then average the **six probabilities equally**.

```
grid <- tidyr::crossing(
  watering = trays$watering,
  batch = trays$batch
) |>
  mutate(tray_id = trays$tray_id[1])
head(grid, 3)
```

```
# A tibble: 3 × 3
  watering batch tray_id
  <fct>    <fct> <fct>
1 pulse    B1    T01
2 pulse    B2    T01
3 pulse    B3    T01
```

PREDICT THE CHANCE OF SPROUTING

Use every batch for each watering method. Average its predicted probabilities equally across the six batches.

```
probabilities <- avg_predictions(
  fit,
  newdata = grid,
  by = "watering",
  type = "response",
  re.form = NA
)
probabilities
```

`type = "response"` asks for probabilities; `re.form = NA` leaves out the fitted tray differences. Zero means no additional tray departure, not zero sprouts. We do not average over trays whose departures vary.

PREDICT THE CHANCE OF SPROUTING

```
 watering Estimate Std. Error    z Pr(>|z|)     S 2.5 % 97.5 %
    pulse    0.436     0.0362 12.0   <0.001 108.6 0.365  0.506
    mist     0.542     0.0365 14.9   <0.001 163.7 0.471  0.613

Type: response
```

STEP 4: MIST MINUS PULSE

```
library("dplyr")

difference <- avg_comparisons(
  fit,
  newdata = grid,
  variables = list(watering = c("pulse", "mist")),
  type = "response",
  re.form = NA
)
print(difference)
```

Percentage points, with an approximate 95% interval. Second condition minus first.

STEP 4: MIST MINUS PULSE

```
 Estimate Std. Error    z Pr(>|z|)   S   2.5 % 97.5 %
    0.106     0.0514 2.07   0.0383 4.7 0.00573  0.207

Term: watering
Type: response
Comparison: mist - pulse
```

```
# Express the same estimate and interval in percentage points for the figure.
comparison <- as.data.frame(difference)
comparison <- comparison |>
  mutate(across(c(estimate, std.error, conf.low, conf.high), ~ 100 * .x))
comparison
```

```
      term     contrast estimate std.error statistic    p.value  s.value
1 watering mist - pulse 10.63834  5.135468  2.071542 0.03830816 4.706205
   conf.low conf.high
1 0.5730062  20.70367
```

MEAN UNCERTAINTY IS NOT INDIVIDUAL VARIABILITY

A probability interval concerns **estimation of a probability**.

A seed either sprouts or does not.

A future tray count also varies with its tray environment and binomial sampling.

A probability interval does not contain 95% of seeds.

EVIDENCE AND PRACTICAL IMPORTANCE

Before seeing the results, imagine the grower decided that **15 more sprouted seeds per 100 seeds** would make a change in watering worthwhile. That is a change of **15 percentage points**.

- Crossing zero does not prove equality.
- Excluding zero does not establish useful benefit.
- A precise association does not establish causation.

Teaching threshold: **15 percentage points**, explicitly hypothetical.

MEANINGFUL OUTCOMES: EXPLAIN THE TEACHING RESULT

Use the worksheet and the separate tray dataset.

1. Count the trays researchers assigned to each watering method. Explain why each count is out of 25 seeds.
2. State target, units, estimate and uncertainty.
3. Explain why setting tray deviation to zero differs from averaging over varying trays.
4. Name a limitation and the practical-threshold question.

Break at 13:40.