SIMULATION-BASED POWER

WEEK 6: THE ESSENTIALS FOR TODAY'S INVESTIGATION

Gavin Simpson

Aarhus University

Wednesday, 21 October 2026

TODAY'S INVESTIGATION

How often would our planned experiment detect a specified diet effect?

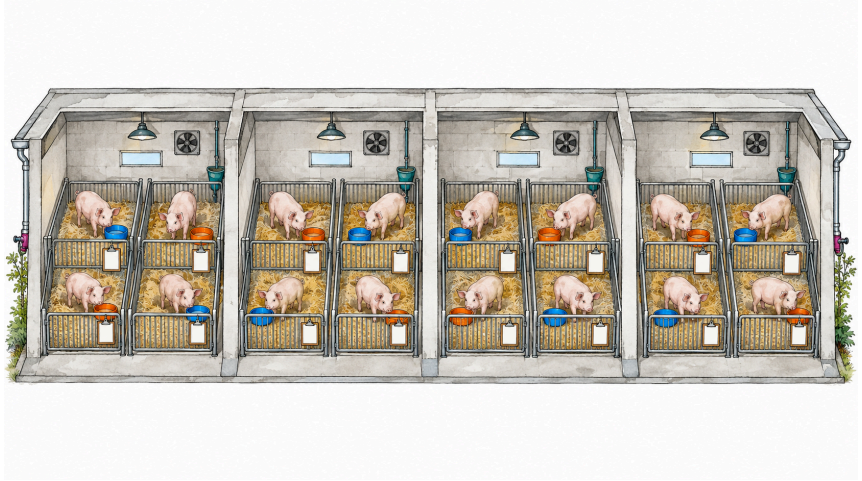
Time	Work
12:30–13:00	Reconstruct the Week 5 simulation
13:00–13:40	Power and one complete analysis
13:40–13:55	Break
13:55–14:55	Estimate power and build a curve
14:55–15:10	Break
15:10–16:00	Reports, recommendation, project launch and exit

THE SHARED DESIGN

Four rooms, initially four pens per room, and six pigs per pen.

Half the pens in each room receive each diet. Follow-up time is equal. We analyse one response per pen.

Room 1 is the reference room; Control is the reference diet. Room and diet effects are specified assumptions.



BUILDING THE DESIGN TABLE

Load the table packages. Create `design`: one row per pen, with four pens in each room.

```
library("dplyr")
library("tibble")
design <- tibble(
  pen_id = sprintf("P%03d", 1:16),
  room = factor(rep(paste0("Room ", 1:4), each = 4)),
  position = rep(1:4, 4)
)
```

`pen_id` identifies a pen; `room` identifies its block.

ALLOCATING THE DIETS

Add a `diet` column to `design`. Randomise two pens to each diet within every room.

```
set.seed(505)
design <- design |>
  group_by(room) |>
  mutate(diet = sample(rep(c("Control", "High fibre"), each = 2))) |>
  ungroup() |>
  mutate(diet = factor(diet, levels = c("Control", "High fibre")))
```

The allocation seed makes this example reproducible. Keep this design fixed when drawing new responses.

ASSUMPTIONS BEFORE OBSERVATIONS

For daily gain, assume:

- Room 1 control expectation: **695 g/day**
- high-fibre diet effect: **+45 g/day** in every room
- room offsets relative to Room 1: **0, +25, +45, +35 g/day**
- variation among pen responses: **SD 60 g/day**

Store these assumptions before generating observations:

```
beta_0 <- 695
beta_diet <- 45
room_effect <- c(0, 25, 45, 35)
sd_response <- 60
```

EXPECTED RESPONSES

For this Gaussian identity-link model, $\eta = \mu$.

$$\mu = 695 + 45\, I(\text{High fibre}) + \text{room offset}$$

Room	Control mean	High-fibre mean
Room 1	695	740
Room 2	720	765

All values are g/day. Each pen has an expected response; we have not drawn an observation yet.

ADDING EXPECTED RESPONSES

Create `expected` from `design`, adding the model score `eta` and expected response `mu` to each pen.

```
expected <- design |>
  mutate(eta = beta_0 + beta_diet * (diet == "High fibre") +
           room_effect[as.integer(room)],
         mu = eta)
```

`room_effect[as.integer(room)]` selects that room's offset. For this identity-link model, `mu` equals `eta`.

DRAWING OBSERVATIONS

Copy `expected` to `first`, then add a random `response` for every pen.

```
first <- expected
set.seed(6070)
first$response <- rnorm(nrow(first),
                        mean = first$mu, sd = sd_response)
```

```
first |> filter(room == "Room 1") |>
  select(pen_id, mu, response)
```

```
# A tibble: 4 × 3
  pen_id    mu response
  <chr>  <dbl>    <dbl>
1 P001     695     737.
2 P002     740     673.
3 P003     695     659.
4 P004     740     685.
```

Why do pens with the same expected response differ?

FITTING TO THE OBSERVATIONS

```
fit_first <- glm(response ~ diet + room,
                 family = gaussian(), data = first)
coef(fit_first)["dietHigh fibre"]
```

```
dietHigh fibre
      50.15963
```

Assumed diet effect: +45 g/day
Estimated diet effect in this dataset: +50.2 g/day

The fitting call uses observations, diet and room. It is not given the assumed effect or the true expected responses.

A SECOND POSSIBLE EXPERIMENT

Draw fresh observations and fit again. Keep the design, expected responses and SD fixed.

```
second <- expected
second$response <- rnorm(nrow(second), second$mu, sd_response)
fit_second <- glm(response ~ diet + room,
                  family = gaussian(), data = second)
coef(fit_second)["dietHigh fibre"]
```

```
dietHigh fibre
     -21.42496
```

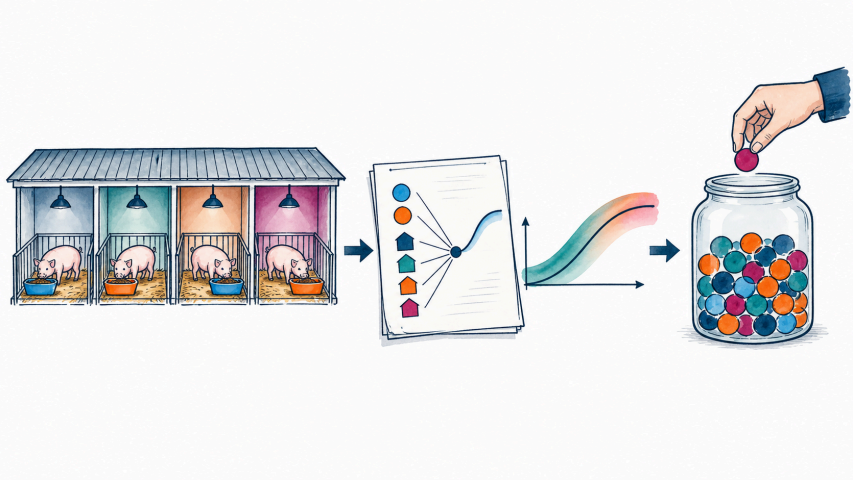
The assumed effect remains **+45 g/day**. Compare this output with the first estimate of **+50.2 g/day**. Even the estimated direction can change.

Across many such experiments, how often would the planned analysis identify a statistically interesting effect?

RECONSTRUCT THE WORKFLOW

With your group, put these stages in order:

- fitted model
- design
- observation
- decision
- expected response (μ)
- model score (η)



Which quantities stay fixed across repeated experiments?

PENS AND REPETITIONS

We simulate **200 possible experiments**, each with **16 pens**.

A larger real experiment would contain more independently treated pens. More simulated experiments give a more precise estimate of the existing design's performance.

Discuss: does the proposed real experiment now have 3,200 pens (16 x 200)? What has actually increased?

FOUR EFFECTS, FOUR INTERPRETATIONS

Group	Response	Assumed diet effect
1	Daily gain	+45 g/day
2	Any pig treated	Odds ratio 0.60
3	Treatment events	Count ratio 0.65
4	Feed conversion ratio	Expected FCR ratio 0.92

Groups 2–4 use the logarithm of the ratio as their model coefficient. The same coefficient value would not mean the same scientific effect across families.

THE DESIGN QUESTION

Before simulation, state **the effect worth detecting** and the assumptions that make the responses plausible.

A statistically detectable difference may be too small to matter from a **biological perspective**. An effect that matters biologically may still be missed by a noisy experiment.

We investigate a planned design under specified effects.

We do not calculate “observed power” from a realised estimate.

POWER UNDER AN ASSUMED EFFECT

Imagine repeating the experiment under the same nonzero diet effect and the same assumptions.

Power: the probability that the specified test/model **rejects** the no-effect hypothesis (H_0)

It is conditional on the data generating process, covariate / treatment effect, baseline, variation among experimental units, experimental / study design and analysis.

It is not the probability that the alternative hypothesis is true — we assume H_0 is true when conducting the test!

WHY USE 80% POWER?

At a specified relevant effect, **80% power** means detecting it in about **4 out of 5 repetitions**; about **1 in 5** would miss it.

0.80 is a common planning convention, not a statistical law. It balances the risk of missing an effect against the resources needed for a larger experiment.

A higher target, such as 0.90, may be appropriate when missing an effect is particularly costly. Choose and justify the target **before comparing designs**.

Power does not mean an 80% probability that the hypothesis is true. The 5% false-positive threshold is a separate choice.

WHEN DO WE DECLARE WE DETECTED AN EFFECT

Compare:

```
full_formula <- response ~ diet + room
reduced_formula <- response ~ room
```

A **statistically interesting effect** requires a usable comparison with $p < \alpha$.

α is the Type I error rate; commonly $\alpha = 0.05$

Both models retain `room`

Record estimated direction separately; i.e. does `diet` lead to an increase or decrease in the `response`.

Rejection does not imply biological importance

UNRESOLVED ANALYSES

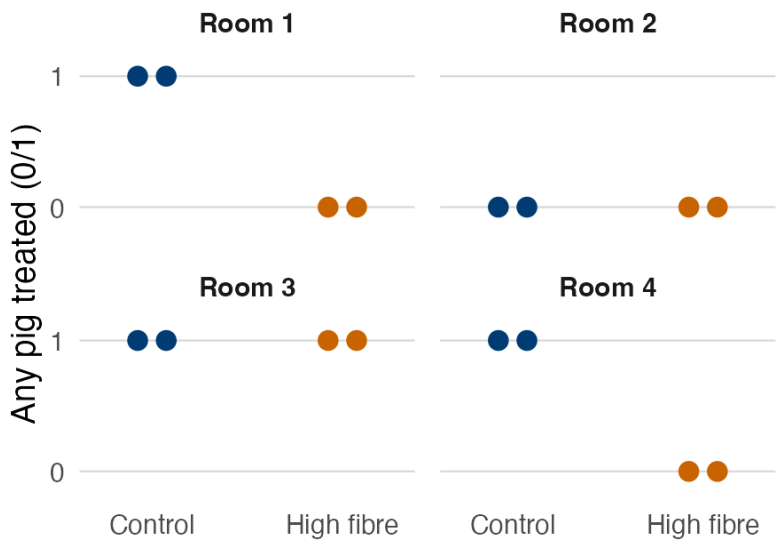
Sparse binary responses can produce unreliable fits even when R returns a coefficient.

Our declared procedure flags warnings, convergence or boundary problems, invalid results and extreme binary fits.

Keep every attempt. Report **detections / all attempts** and **unresolved / all attempts** separately. Do not keep drawing until a fit looks favourable.

COMPLETE SEPARATION

Binomial logit model: `response ~ diet + room`



16 pens: two per diet in each room. Diet **and room together** perfectly distinguish 0 from 1.

`summary(fit)` — **coefficient table**

Term	Estimate	SE	z	p
Intercept	25.36	119,300	≈0	≈1
High fibre	−50.72	137,700	≈0	≈1
Room 2	−50.72	182,200	≈0	≈1
Room 3	50.72	182,200	≈0	≈1
Room 4	≈0	137,700	≈0	≈1

Residual deviance ≈ 0; 25 iterations.
`fit$converged` is **TRUE**.
Estimates and test statistics rounded.

Look for: extreme log-odds coefficients, huge SEs and fitted probabilities near **0 or 1**. The Wald p-values are unreliable—not evidence of no effect.

R warns: `fitted probabilities numerically 0 or 1 occurred`.

POWER ACROSS CANDIDATE DESIGNS

Return to Gaussian daily gain: **+45 g/day**, SD **60 g/day**, four fixed rooms. Read the supplied results into `power_reference`; select our baseline scenario as `gain_curve`.

```
power_reference <- read.csv("reference/power-summary.csv")
gain_curve <- power_reference |>
  filter(group == 1, scenario == "Baseline", !null)
gain_curve |>
  select(total_pens, attempted, detections, detection, unresolved_rate)
```

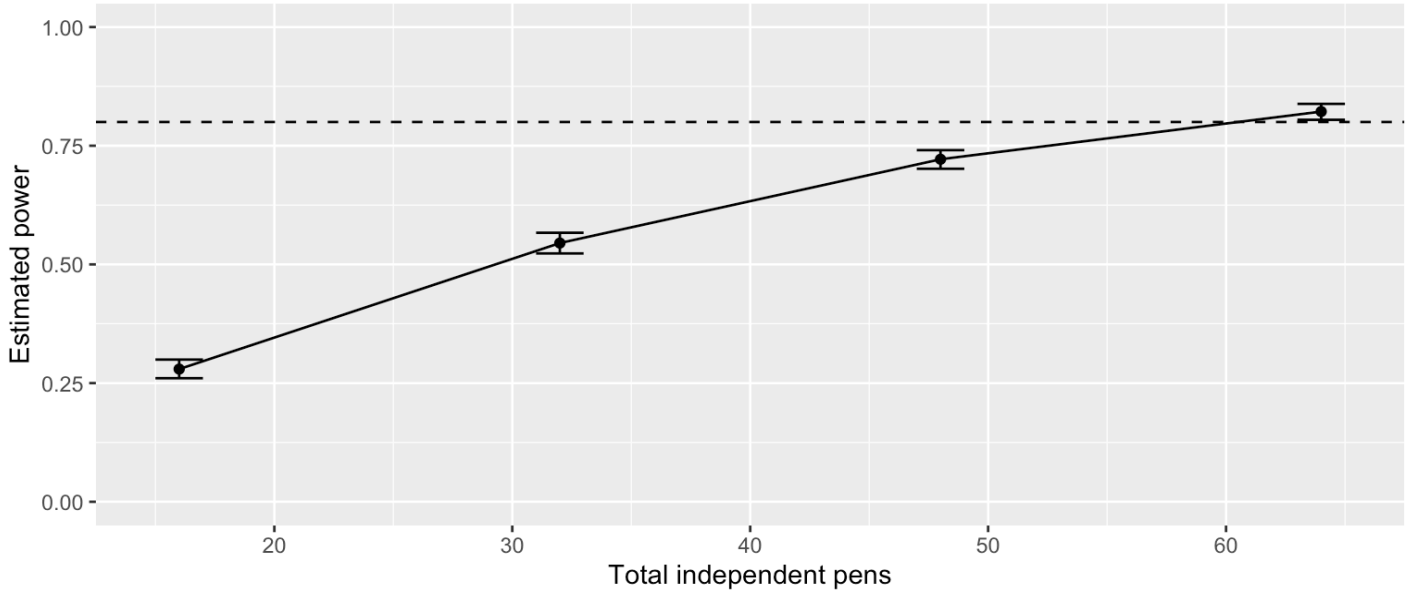
```
  total_pens attempted detections detection unresolved_rate
1         16      2000        559    0.2795               0
2         32      2000       1090    0.5450               0
3         48      2000       1443    0.7215               0
4         64      2000       1644    0.8220               0
```

Each row summarises **2,000 complete simulated experiments** at that sample size. `detection` is detections divided by attempts.

THE POWER CURVE

Plot the estimates against total pens. The dashed line is our chosen **0.80 target**; vertical bars show **95% Monte Carlo intervals**.

```
library("ggplot2")
ggplot(gain_curve, aes(total_pens, detection)) +
  geom_line() + geom_point() +
  geom_errorbar(aes(ymin = lower, ymax = upper), width = 2) +
  geom_hline(yintercept = 0.80, linetype = "dashed") +
  scale_y_continuous(limits = c(0, 1)) +
  labs(x = "Total independent pens", y = "Estimated power")
```



READING THE CURVE FOR A DECISION

At **48 pens**, estimated power is **72.2%**; at **64 pens**, it is **82.2%**.

Of the four sizes tested, **64 pens** is the first to exceed our 80% target under these assumptions.

This does **not** establish the exact minimum sample size. We have not tested the sizes between 48 and 64, and the assumed effect and variation may be wrong.

More simulation repetitions refine the plotted estimates. More independently treated pens change the experiment's power.

Would 64 pens still be convincing if the responses were more variable than we assumed?

PACKAGE CHOICE AND PROJECT

We use **dplyr, tibble, ggplot2 and readr** for the table grammar already used in Week 5 and for ggplot2. R's supplied **stats** package handles generators and GLMs.

Open your assigned `group-N-student.R` in the `06-power-analysis` project. Run from that project folder. `source()` defines the supplied helpers; `scenario(3)` creates `s`, the Poisson assumptions for this demonstration.

```
source("power-helpers.R")
s <- scenario(3) # use your assigned group
```

ONE COMPLETE REPETITION

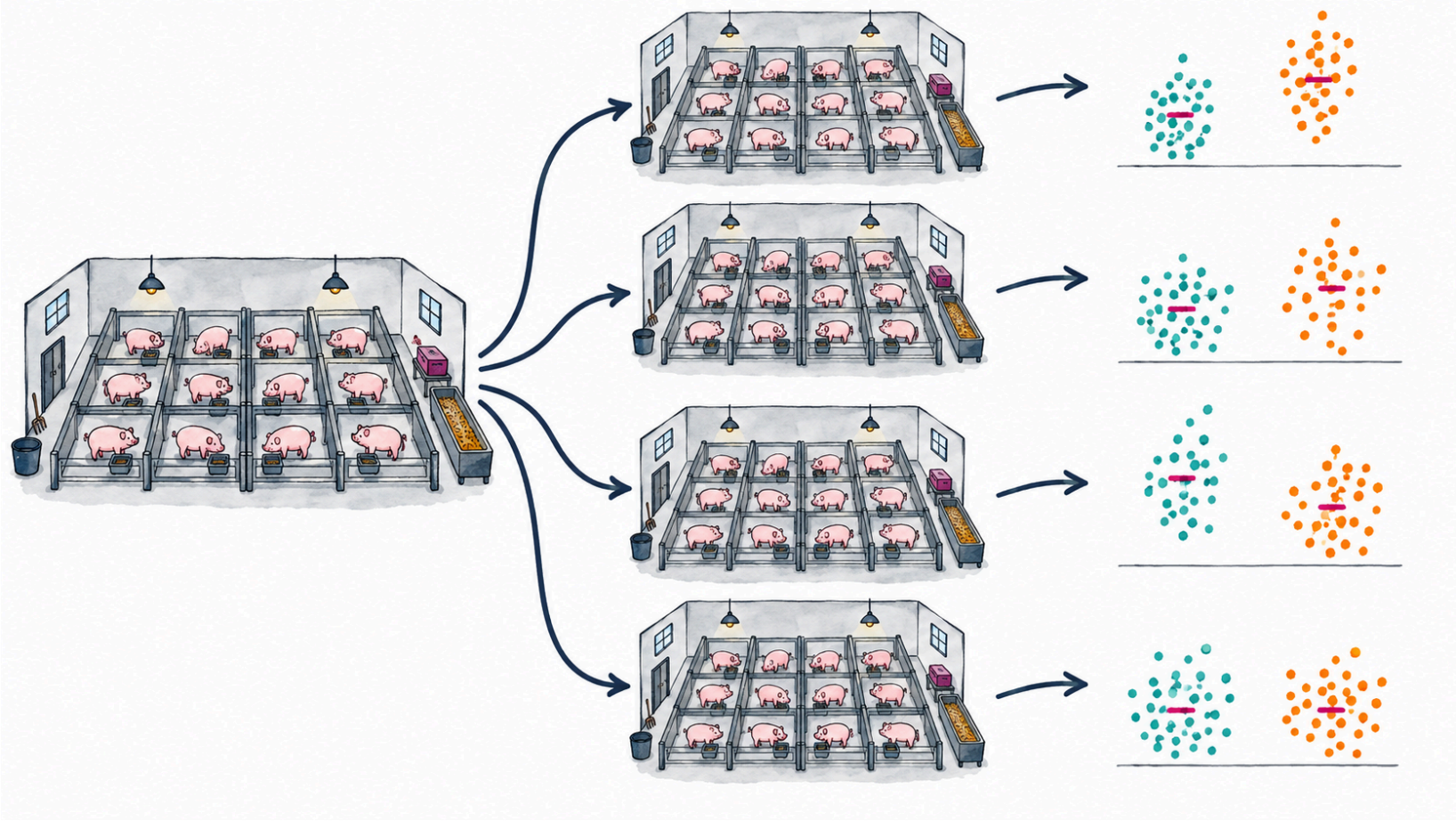
`make_design()` builds the balanced table we constructed earlier. `generate_response()` adds the responses; `analyse_one()` fits, tests and returns the result.

```
design <- make_design(16)
set.seed(6103)
one <- generate_response(design, s)
result <- analyse_one(one, s)
result |> select(beta, p, unresolved, detected)
```

```
# A tibble: 1 × 4
    beta     p unresolved detected
   <dbl> <dbl> <lgl>      <lgl>
1 -0.182 0.546 FALSE      FALSE
```

Identify the assumed effect, fitted effect, p-value and decision. Read `unresolved` before interpreting the result; inspect `result$warnings` in your script if an analysis is unresolved.

ONE DESIGN, MANY POSSIBLE DATASETS



The design and parameters stay fixed. The response draw, fitted estimate and decision can change.

SOURCES

- R Core Team: GLMs, model comparisons, families.
- Wickham et al. (2019): Welcome to the Tidyverse.
- Colegrave and Ruxton (2020): *Power Analysis: An Introduction for the Life Sciences*. Instructor background.

All biological scenarios and parameter values are teaching assumptions.