FIT AND EXPLAIN A GLM

WEEK 7: COMPARE DIETS, EXPLAIN RESULTS AND CHECK ANSWERS

Gavin Simpson

Aarhus University

Wednesday, 21 October 2026

DOES THE DIET DIFFERENCE CHANGE WITH HOUSING?

This is a **separate study** with 48 pens in four rooms. Pens receive one of two diets and one of two housing conditions, Standard or Enriched.

Every room contains all four combinations, with three pens per combination. Both treatments are assigned to pens.

We ask whether the High fibre minus Control difference in average gain is the same in both housing conditions.

READ THE DATA FOR THE DIET-AND-HOUSING STUDY

The supplied `read_factorial_data()` function reads `data/factorial-pens.csv` into `f`. Its reference categories are Control diet, Standard housing and Room 1.

```
source("glm-helpers.R")
f <- read_factorial_data()
head(f[c("diet", "housing", "room", "gain")], 2)
```

```
     diet  housing   room     gain
1 Control Enriched Room 4 723.8040
2 Control Standard Room 1 705.8254
```

`gain` is the pen's average daily gain in g/day. `f` is different from the group datasets we analyse later.

TWO WAYS TO DESCRIBE THE DIET DIFFERENCE

An **additive model** requires the same High fibre minus Control gain difference in Standard and Enriched housing.

A model with an **interaction** allows that difference to change with housing. In R, `diet * housing` includes both variables and their interaction.

```
additive_formula <- gain ~ diet + housing + room
interaction_formula <- gain ~ diet * housing + room
additive_formula
```

```
gain ~ diet + housing + room
```

```
interaction_formula
```

```
gain ~ diet * housing + room
```

FIT THE TWO MODELS TO THE SAME OBSERVATIONS

We use the same gain data, rooms and Gaussian identity family/link for both models. Only the treatment terms change.

```
additive <- glm(additive_formula, data = f,
                family = gaussian("identity"))
interaction <- glm(interaction_formula, data = f,
                   family = gaussian("identity"))
# The two diet-related coefficients; other terms are omitted here.
coef(interaction)[
  c("dietHigh fibre", "dietHigh fibre:housingEnriched")]
```

```
             dietHigh fibre dietHigh fibre:housingEnriched
                   43.03714                     35.67392
```

In the interaction model, `dietHigh fibre` describes the diet difference in Standard housing. The `dietHigh fibre:housingEnriched` coefficient tells us how much larger or smaller that difference is in Enriched housing.

LIST THE FOUR COMBINATIONS TO COMPARE

To compare diets fairly, keep room fixed. We choose Room 1 and list both diets under both housing conditions.

```
factorial_grid <- expand.grid(
  diet = levels(f$diet), housing = levels(f$housing),
  room = "Room 1")
factorial_grid
```

```
        diet  housing   room
1    Control Standard Room 1
2 High fibre Standard Room 1
3    Control Enriched Room 1
4 High fibre Enriched Room 1
```

The four rows describe the conditions for predictions, not four new measurements.

CALCULATE THE PREDICTED AVERAGE GAINS

`predict_means()` uses each fitted model and the four combinations above. We label the results so the plot can show both models. The printed table shows the interaction model's expected gains, in g/day.

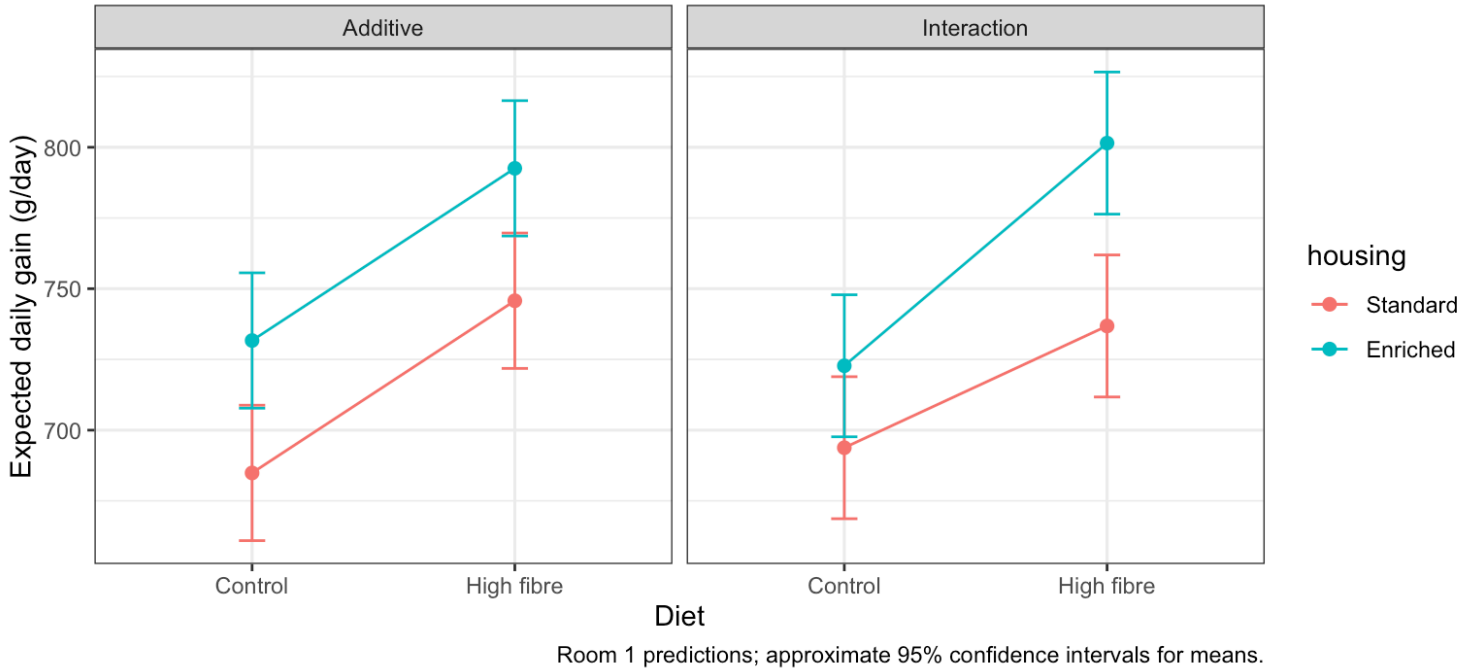
```
pred_add <- predict_means(additive, factorial_grid)
pred_add$model <- "Additive"
pred_int <- predict_means(interaction, factorial_grid)
pred_int$model <- "Interaction"
factorial_predictions <- rbind(pred_add, pred_int)
pred_int[c("diet", "housing", "expected")]
```

```
        diet  housing expected
1    Control Standard 693.7933
2 High fibre Standard 736.8305
3    Control Enriched 722.7639
4 High fibre Enriched 801.4749
```

READ THE TWO PLOTS

In each panel, compare Control and High fibre within Standard housing, then make the same comparison within Enriched housing.

```
plot_factorial_predictions(factorial_predictions)
```



Which model gives the same increase from Control to High fibre in both housing conditions?

CALCULATE HOW THE DIET DIFFERENCE CHANGES

First subtract Control from High fibre within each housing condition. Then subtract the Standard-housing difference from the Enriched-housing difference.

```
diet_difference <- with(pred_int,
  expected[diet == "High fibre"] - expected[diet == "Control"])
names(diet_difference) <- levels(f$housing)
diet_difference
```

```
Standard Enriched
43.03714 78.71106
```

```
c(difference_of_differences = unname(diff(diet_difference)))
```

```
difference_of_differences
                 35.67392
```

The final number is in g/day. A positive value means the estimated diet difference is larger in Enriched housing.

NOW OPEN YOUR GROUP'S SCRIPT

Use `group-1-student.R`, `group-2-student.R`, `group-3-student.R` or `group-4-student.R` in the supplied Posit Cloud project.

One person runs R, a second explains each result, and a third checks the explanation against the study description. Swap jobs at the next break.

Run sections 1 and 2 first. Check that there are eight pens on each diet in every room. All groups fit `response ~ diet + room`; their measured outcomes and families differ.

WORKED EXAMPLE: TREATMENT-EVENT COUNTS

Here we demonstrate Group 3. `response_spec(3)` supplies its family, link and labels; `spec$family` selects the family/link settings from that list. `read_pen_data(3)` reads the counts into `d`.

```
spec <- response_spec(3) # worked example: Group 3
# These d rows count treatment events per pen.
d <- read_pen_data(3)
fit <- glm(response ~ diet + room,
           family = spec$family, data = d)
coef(summary(fit))[1:2, c("Estimate", "Std. Error")]
```

```
                Estimate Std. Error
(Intercept)     0.8776216  0.1891301
dietHigh fibre -0.4924765  0.1562395
```

With a log link, the printed diet estimate is a log ratio. The next step converts it to an easier-to-read ratio and shows expected counts.

EXPECTED COUNTS UNDER THE TWO DIETS

The first table below gives expected treatment-event counts for pens in Room 1 over 28 days. The second gives the High fibre / Control expected-count ratio and its approximate 95% confidence interval.

```
grid <- prediction_grid(d)
predictions <- predict_means(fit, grid)
subset(predictions, room == "Room 1",
       select = c(diet, expected, lower, upper))
```

```
       diet expected    lower    upper
1   Control 2.405172 1.660192 3.484449
2 High fibre 1.469828 0.985200 2.192847
```

```
diet_comparison(fit)
```

```
   estimate     lower     upper
1 0.6111111 0.4499134 0.8300638
```

`estimate` in the second table is the High fibre / Control expected-count ratio; `lower` and `upper` are its confidence limits. Which diet has the lower expected count? Explain what the Control count is being multiplied by.

WRITE AN EXPLANATION OF YOUR OWN RESULTS

Use Room 1 for your first explanation. Write:

1. What you measured, including units, and that diet was assigned to pens.
2. Which response, explanatory variables, family and link you used.
3. The estimated outcome for each diet, and the difference or ratio between them.
4. The interval showing uncertainty, and one assumption that still needs checking.

Use the worksheet's examples and recap if you need help with the wording.

APPLY THE SAME STEPS TO A CALF STUDY

There are 40 pens in five barns, with four calves in every pen. In each barn, four pens receive Standard feed and four receive Supplemented feed, assigned at random.

The outcome `any_treated` is 1 if at least one calf in a pen needed treatment during 21 days, and 0 otherwise.

Use the worksheet to describe a model comparing the two feeds while allowing for differences between barns. No new dataset or R fit is needed for this question.

CHECK ANOTHER GROUP'S ANSWER

Exchange answers. For each point below, explain whether you agree and why:

- Does one row describe a pen, with feed assigned to that pen?
- Is the outcome correctly described as 0 or 1 for any calf treated?
- Does the formula include both feed and barn?
- Are the family and link stated?
- Do the proposed predictions compare both feeds in the same barn?

Give the group one specific suggestion, then revise your own answer.

DISCUSS YOUR GROUP PROJECT

Bring out the short project description from the week 42 task.

Tell us the question you want to answer, what you will measure, and which groups or conditions you will compare. Which dataset will you use, and can you access it?

Write down any information you still need about how the data were collected. If your preferred data are unavailable, choose one of the provided studies to work with.

A SHORT QUESTION BEFORE YOU LEAVE

For the pen study, a student writes only `response ~ diet + room` as their description of the model.

What else must they state about the measured outcome, the family and the link so that someone else could fit the same model?

Then use your own output to write one sentence comparing the two diets in Room 1. Include units, and mention uncertainty.