CHECKING MODELS AGAINST THE EXPERIMENT AND THE DATA

CLASS 9 · MODEL DIAGNOSTICS

Gavin Simpson

Aarhus University

Wednesday, 4 November 2026

CLASS 9 SCHEDULE

Time	Work
12:30–13:40	Recall, checking tools and shared count example
13:40–13:55	Break
13:55–14:55	Four group cases
14:55–15:10	Break
15:10–16:00	Change, recheck, project check and report

OPENING RECALL: ANIMALS SHARE PENS

Last class, several animals shared each pen, and a diet was assigned to a whole pen.
Discuss with your seated group:

- What must the model say about animals sharing a pen?
- What output would make you doubt that fitting worked?
- What pattern would suggest that the model misses something?

THREE QUESTIONS FOR ANY FITTED MODEL

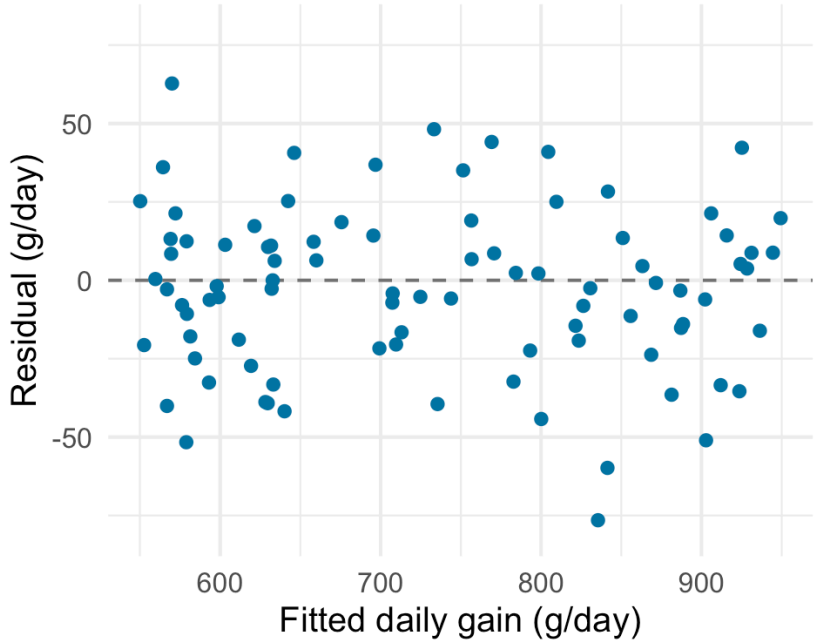
Question	Evidence to inspect
Does it respect the experiment?	Who received each diet; shared conditions
Did the calculation work reliably?	Warnings, extreme estimates or very large standard errors
Does it describe the data?	Patterns left unexplained or differences from simulated data

A standard error describes uncertainty in an estimate under the fitted model.

WHAT THE MODEL LEAVES UNEXPLAINED

For daily gain, a residual is **observed gain minus fitted gain**, in g/day.

In this simulated example, the residuals form a horizontal cloud around zero with a similar spread across fitted values. Look for patterns against starting weight and by pen as well.

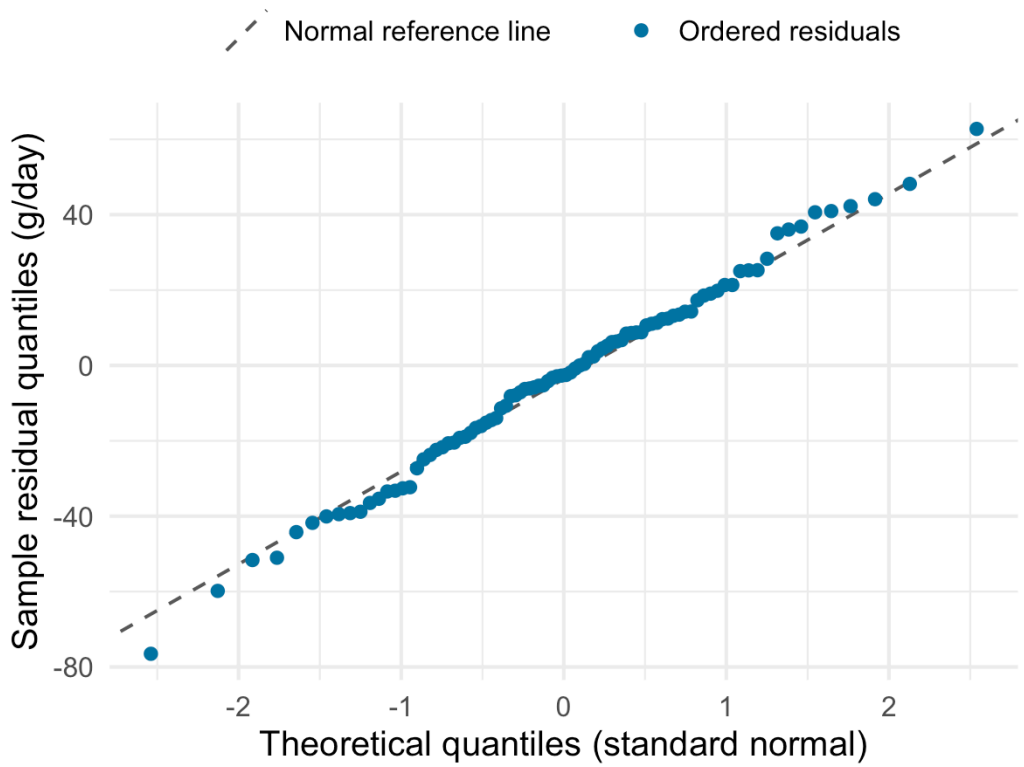


NORMAL QQ PLOT FOR GAIN RESIDUALS

A normal QQ plot checks whether the residuals have the bell-shaped distribution the model assumes.

Each point pairs an ordered residual with the corresponding quantile of a standard normal distribution. The dashed line shows the pattern expected for normal residuals at their observed scale.

Points close to the line are consistent with normality. A bend or a systematic departure at either end suggests a different distribution shape.



COUNT CHECKS COMPARE DATA WITH SIMULATIONS

1. **Simulate counts.** `performance` uses DHARMa to simulate many datasets from the fitted model, keeping each pen's diet and room fixed.
2. **Locate the observed count.** For each pen, find the fractions of simulated counts below and at or below its observed count. The **randomised quantile residual** is a uniform draw between those fractions, or their common value if they agree.
3. **Compare quantiles.** The QQ plot shows theoretical quantiles of a **uniform distribution on 0 to 1** on the x-axis and sample quantiles of the residuals on the y-axis. Points should fall roughly along the diagonal if the residuals follow that reference distribution.

ONE EVENT COUNT PER PEN

Researchers assigned diets to 64 pens in four rooms. Each room had eight pens per diet. Researchers recorded one count of treatment events per pen over the same follow-up time.

Open the guided count worksheet. Set Control as the reference diet and R1 as the reference room.

```
library("dplyr")

d <- readr::read_csv("data/case-1.csv")
d <- d |>
  mutate(
    diet = factor(diet, levels = c("Control", "High fibre")),
    room = factor(room, levels = c("R1", "R2", "R3", "R4"))
  )
head(d, 2)
```

```
# A tibble: 2 × 4
  room  pen    diet    events
  <fct> <chr>  <fct>    <dbl>
1 R1    R1-P01 Control      1
2 R1    R1-P02 Control      1
```

POISSON MODEL FOR EVENT COUNTS

We want to compare expected event counts under High fibre and Control within the same rooms. In `events ~ diet + room`, diet gives that comparison and room lets average counts differ across rooms.

The Poisson model assumes that the variance of counts equals their mean at each diet and room combination.

```
poisson_fit <- glm(
  events ~ diet + room,
  family = poisson(),
  data = d
)
```

POISSON MODEL SUMMARY

```
summary(poisson_fit)
```

```
Call:
glm(formula = events ~ diet + room, family = poisson(), data = d)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)     0.9165     0.1821   5.032 4.86e-07 ***
dietHigh fibre -0.4313     0.1361  -3.168  0.00153 **
roomR2          0.9223     0.2058   4.482 7.40e-06 ***
roomR3          0.6306     0.2155   2.927  0.00343 **
roomR4          0.3747     0.2261   1.657  0.09752 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 250.77  on 63  degrees of freedom
Residual deviance: 216.28  on 59  degrees of freedom
AIC: 380.22

Number of Fisher Scoring iterations: 5
```

Find `dietHigh fibre` and its standard error. The negative coefficient is a log ratio, so this fit estimates fewer expected events under High fibre.

HOW MUCH VARIATION DOES THE POISSON MODEL MISS?

```
library("performance")
check_overdispersion(poisson_fit)
```

```
# Overdispersion test

       dispersion ratio =   3.975
  Pearson's Chi-Squared = 234.526
              p-value = < 0.001
```

The dispersion ratio is about **4.0**, well above one. The observed pen counts vary more than the Poisson model allows after accounting for diet and room.

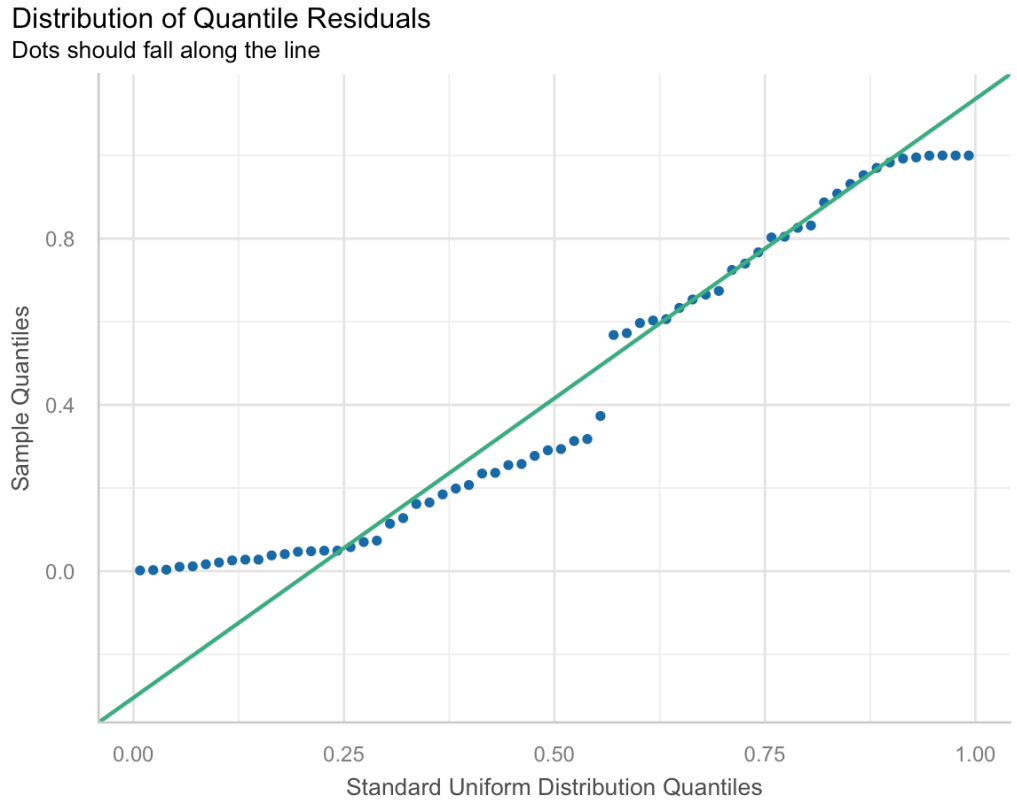
Researchers followed each pen for the same time and assigned diets to pens separately. What else could explain the extra variation in counts? Name two possible causes before changing the model.

POISSON QQ PLOT FOR THE PEN COUNTS

Compare the sample residual quantiles with the theoretical uniform quantiles. Where do the points depart from the diagonal?

The `seed` argument makes the simulation reproducible.

```
poisson_res <- simulate_residuals(
  poisson_fit,
  seed = 9401
)
poisson_qq <- plot(poisson_res)
```



NEGATIVE BINOMIAL MODEL FOR PEN COUNTS

The negative binomial model allows more variation in counts among pens on the same diet in the same room. `method = "ML"` fits the model by maximum likelihood.

```
library("mgcv")

nb_fit <- gam(
  events ~ diet + room,
  family = nb(),
  data = d,
  method = "ML"
)
set.seed(9401)
check_overdispersion(nb_fit, iterations = 500)
```

```
# Overdispersion test (using simulated residuals)

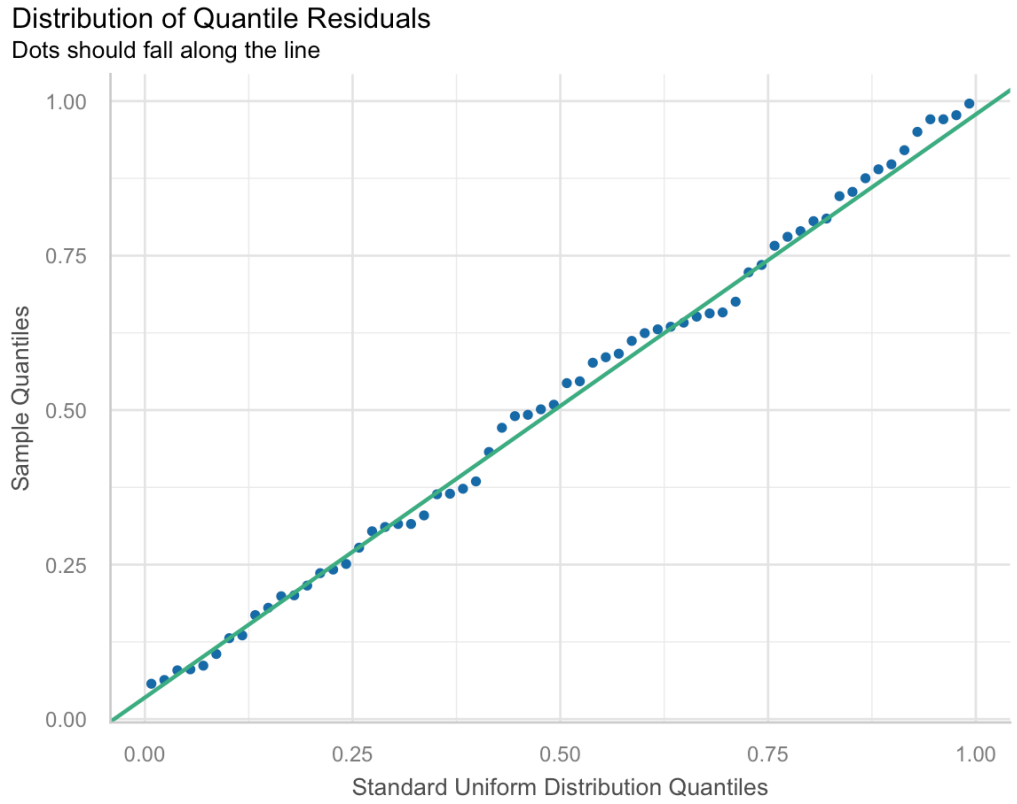
 dispersion ratio = 0.868
          p-value =  0.78
```

In the check with 500 simulated datasets, the dispersion ratio is now close to one.

NEGATIVE BINOMIAL QQ PLOT FOR THE SAME COUNTS

Compare the departures from the diagonal with the Poisson plot. Which part of the pattern improved? Which possible cause would still require information about the pens?

```
nb_res <- simulate_residuals(
  nb_fit,
  seed = 9401
)
nb_qq <- plot(nb_res)
```



CAN THE MODEL RECREATE THE OBSERVED COUNTS?

- **Simulate.** Use each fitted model to generate 500 new sets of pen counts, keeping each pen's diet and room fixed.
- **Compare.** For each possible event count, compare how many pens had that count in the study with how many had it in the simulations.
- **Look for a mismatch.** Can the model recreate how often pens have zero events? What about unusually high counts?

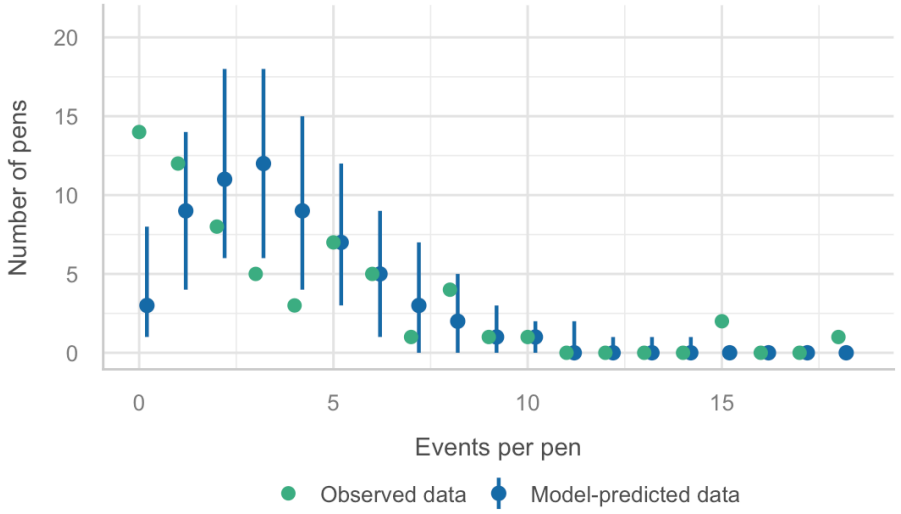
`check_model()` calls this a *posterior predictive check*. Here we simulate at the fitted estimates. A Bayesian posterior predictive check draws parameters from their posterior distribution before simulating new counts.

WHICH MODEL RECREATES THE ZERO COUNTS?

Green points show the observed numbers of pens. Blue points and bars show the simulated medians and intervals. Both panels show the observed range of 0 to 18 events per pen.

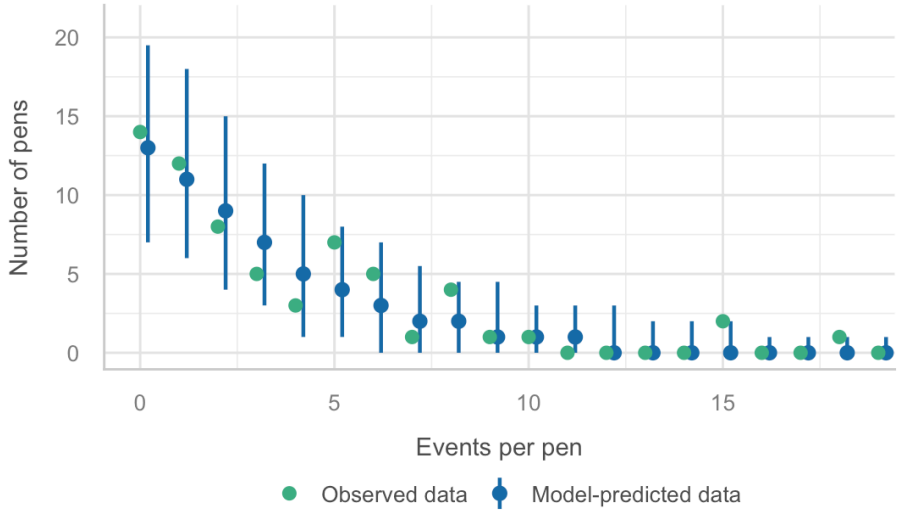
Poisson

```
Ignoring unknown labels:
• size : ""
• alpha : ""
```



Negative binomial

```
Ignoring unknown labels:
• size : ""
• alpha : ""
```



WHICH PENS DOES THE DIET COMPARISON REPRESENT?

`newdata = "balanced"` gives one prediction row for each diet in each of the four study rooms. That gives every room the same weight. Averaging over the observed pens gives the same room weights here because every room has 16 pens.

`comparison = "lnratioavg"` averages the predicted counts for each diet, divides the High fibre mean by the Control mean and takes the log. `transform = "exp"` returns the estimate and interval to the count-ratio scale. **One means equal expected counts.**

POISSON DIET COUNT RATIO

```
library("marginaleffects")

avg_comparisons(
  poisson_fit,
  variables = "diet",
  newdata = "balanced",
  comparison = "lnratioavg",
  transform = "exp"
)
```

```
 Estimate Pr(>|z|)   S 2.5 % 97.5 %
     0.65  0.00153 9.3 0.497  0.848

Term: diet
Type: response
Comparison: ln(mean(High fibre) / mean(Control))
```

The estimate of about **0.65** means 35% fewer expected events under High fibre. The approximate 95% interval stays below one under the Poisson assumption.

NEGATIVE BINOMIAL DIET COUNT RATIO

```
avg_comparisons(
  nb_fit,
  variables = "diet",
  newdata = "balanced",
  comparison = "lnratioavg",
  transform = "exp"
)
```

```
 Estimate Pr(>|z|)   S 2.5 % 97.5 %
    0.673    0.133 2.9 0.401   1.13

Term: diet
Type: response
Comparison: ln(mean(High fibre) / mean(Control))
```

The estimate is about **0.67**, but the wider interval now includes one. Once the model allows extra variation among pen counts, the data are compatible with a reduction, equal expected counts or a small increase.

WHAT CHANGED WHEN THE MODEL ALLOWED EXTRA VARIATION?

For these pen counts, record one concern about the starting model and one reassuring feature of the changed model.

Name another possible cause of the Poisson pattern and the pen records you would need to investigate it. Explain how the wider interval changes what you can say about the two diets.

Break 13:40–13:55. Continue with the workshop deck.