

Lab 8

Stephen R. Proulx

Bayesian Statistical Modeling Winter 2024

Lab Exercise, Week 9

When is this lab due? Labs are due on the Thursday after they are assigned. However, in many cases you can complete them during the lab period itself. This assignment is due on Thursday, 3/14/2024.

Reedfrog survival: Models with different predictors

In class we worked with some multilevel models of the reedfrog data where we either allowed each tank to be fit as a “random” effect, or we had a random effect and an effect of predation that was dependent on density. Technically, this is already an interaction model.

Today you will make models that include two of the predictor variables from the dataset, namely tadpole size and predation.

You will create 5 versions of the model:

m1: A model with no factors predicting survivorship, but one that assumes that all tanks have their survivorship drawn at random from a normal distribution where the mean and standard deviation of that normal are parameters. This is model 13.2 from the book.

$$\text{logit}(p) = \bar{a} + \delta a_i * \sigma$$

m2: A model with a random tank effect, as in m1, but also with a direct effect of predation.

$$\text{logit}(p) = \bar{a} + b_P * P_i + \delta a_i * \sigma$$

m3: A model with a random tank effect, as in m1, but also with an effect of tadpole size.

$$\text{logit}(p) = \bar{a} + b_s * ts_i + \delta a_i * \sigma$$

m4: A model with a random tank effect, and with independent (i.e. additive on the logit scale) effects of predation and tadpole size.

$$\text{logit}(p) = \bar{a} + b_P * P_i + b_s * ts_i + \delta a_i * \sigma$$

m5: A model with a random tank effect, and interacting effects of predation and tadpole size. Because both predation and tadpole size have two levels, there are 4 total parameters to describe these effects.

$$\text{logit}(p) = \bar{a} + b_P * P_i + b_s * ts_i + b_{pts} * P_i * ts_i + \delta a_i * \sigma$$

Here's the data- you will need to add some columns to correctly format the predictors for your model.

```
library(rethinking)
data(reedfrogs)

d <- reedfrogs
dat <- list(
  S = d$surv,
  n = d$density,
  tank = 1:nrow(d),
  P = ifelse( d$pred=="no" , 0L , 1L ),
  ts = ifelse( d$size=="small" , 0L , 1L ),
  dtreat = ifelse( d$density=="10" , 1L , ifelse( d$density=="25" , 2L ,3L) )
)
```

Build and sample model 1

Run m1 with 4 chains for 3000 iterations. Get the precis output, check for convergence, and note the range of high posterior probability parameter values.

Use sim_df to simulate the tank specific data and plot it with the original data.

Build and sample model 2

Run m2 with 4 chains for 3000 iterations. Get the precis output, check for convergence, and note the range of high posterior probability parameter values.

Compare m1 and m2 using WAIC.

Compare the values of sigma for m1 and m2.

Use sim_df to simulate the tank specific data and plot it with the original data.

Build and sample model 3

Run m3 with 4 chains for 3000 iterations. Get the precis output, check for convergence, and note the range of high posterior probability parameter values.

Compare m1 and m3 using WAIC.

Compare the values of sigma for m1 and m3.

Use sim_df to simulate the tank specific data and plot it with the original data.

Build and sample model 4

Run m4 with 4 chains for 3000 iterations. Get the precis output, check for convergence, and note the range of high posterior probability parameter values.

Compare models m2 and m3 with m4 using WAIC.

Compare the values of sigma for m2, m3, and m4.

Use sim_df to simulate the tank specific data and plot it with the original data.

Build and sample model 5

Run m4 with 5 chains for 3000 iterations. Get the precis output, check for convergence, and note the range of high posterior probability parameter values.

Compare models m4 with m5 using WAIC.

Compare the values of sigma for m4 and m5.

Use sim_df to simulate the tank specific data and plot it with the original data.

Compare the coefficients from all the models

We can extract all of the mean coefficients from all 5 models and look at them in a single table using `coefstab`. It is in alphabetical order, so `abar` and `sigma` are somewhere in the middle. Note that it included the tank specific `a` values, as well as the `p` values for each row in our data.

```
coefstab( m1.1 , m1.2 , m1.3 , m1.4 , m1.5 )
```

Thought questions: 1. Scan the values for `a[tank]` across models and the values for `p[tank]`. What do you notice about how they vary among models. Why do you think this is? 2. How do the values of `sigma` compare across models? 3. How does the value of `bs` change among models? Why do you think this is?