

Expected lifetime

K.A. Flagg

Intro

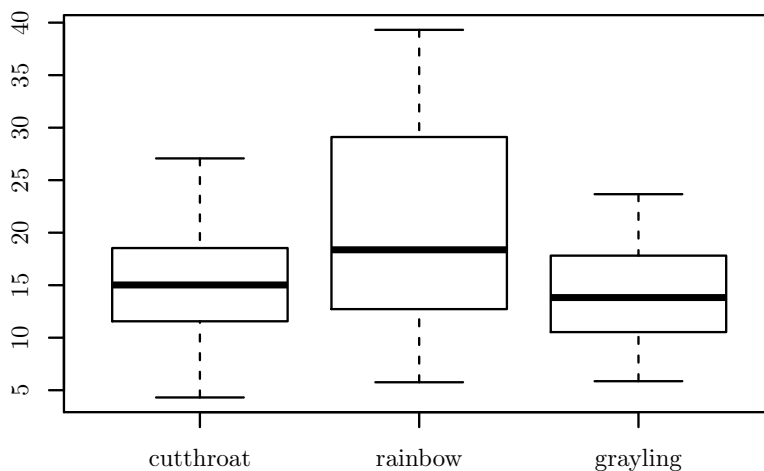
A collection of 37 cutthroat trout, 44 rainbow trout, and 21 grayling were exposed to a toxic biological agent over the course of a 50 day study with objective to determine the typical length of survival for these species of fish after being exposed to the agent. However, some fish survived the entire duration of the study and, to complicate matters, each fish was exposed at a different time during the study. So, we have a tricky censored data problem.

For fish that died, the length of time until death from exposure is recorded. For fish that did not die during the study, the length of time from exposure to completion of the study (i.e. last known time alive) was recorded. Time measurements are given in hours.

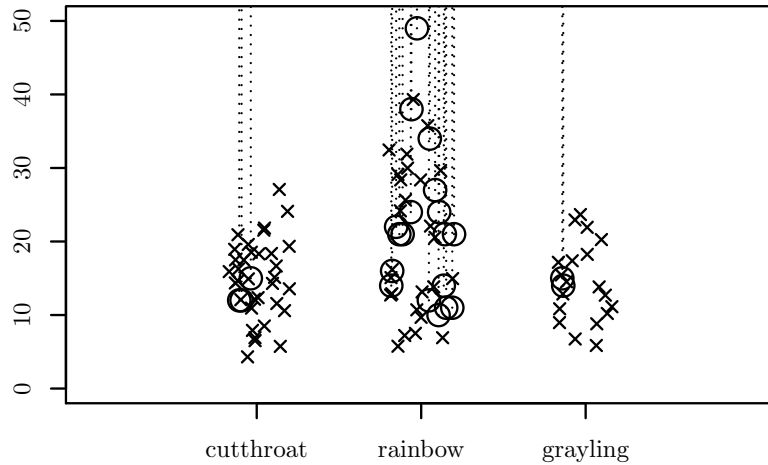
The researchers would like to estimate the average time until death for each species and also in contrasting the average time until death of the cutthroat and rainbow trouts.

Data

First, we take a look at the time until death for the fish that did die during the study. The boxplots below give the time until death by species.



We must be careful of looking too closely at the boxplot. Trying to make inference from the boxplot and the observed data alone might cause us to underestimate the average lifetime. Below is a plot that marks known time until death with an “x” and last known time alive with a “o” followed by a dashed line to indicate that death happens at some point beyond the mark. However, I suspect there are better ways to visualize this information.



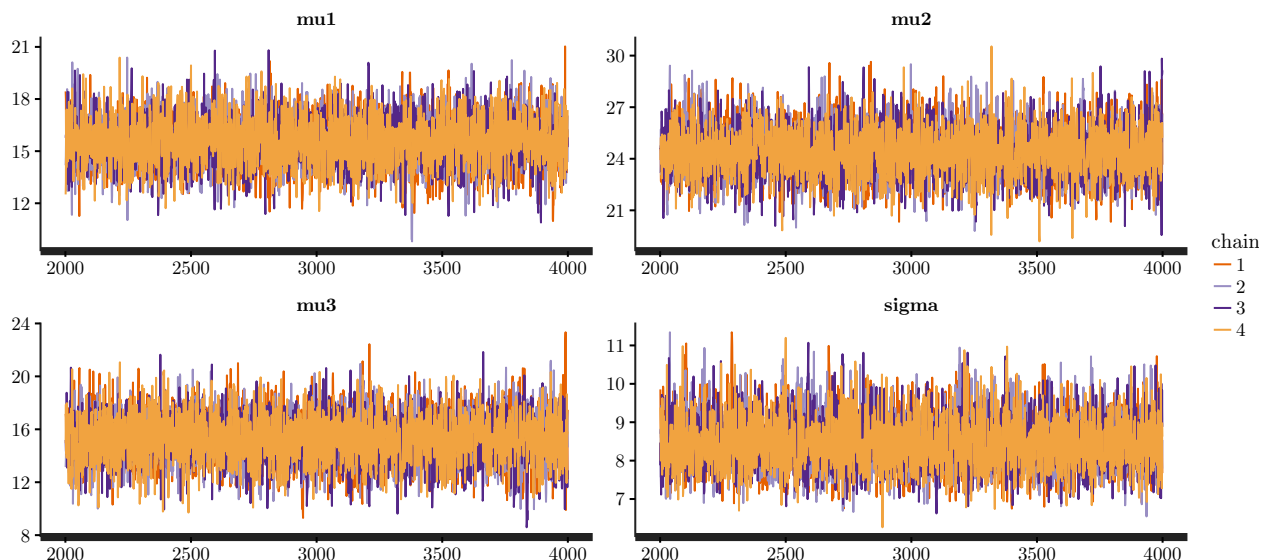
Fit model

Now that we've played around with the dataset a little bit, we are going to fit a Bayesian model that takes into account both the observed times until death as well as the last known time alive. We are using Rstan and the model code is in the file `lifetime.stan`. The simple model assumes that the likelihood that fish i of species $j[i]$ perishes after y time is $y_i \sim N(\mu_{j[i]}, \sigma^2)$. To spell it out, we assume that there is a mean time until death for each species, but all species share the same variability of time until death for individual observations. (The model is fairly simple and you are welcome to investigate and tweak. In a future R meetup we will figure out exactly what is going on with the code and maybe adjust the model to something more realistic.)

So, we have four parameters of interest, the mean time until death for each species (μ_1 , μ_2 , and μ_3) and the common standard deviation (σ) of time until death.

Evaluate sampler

Before making inference on the model fit, we should investigate to make sure it seems like the sampler has converged. First let's take a look at traceplots for each variable to make sure our 4 chains are mixing.



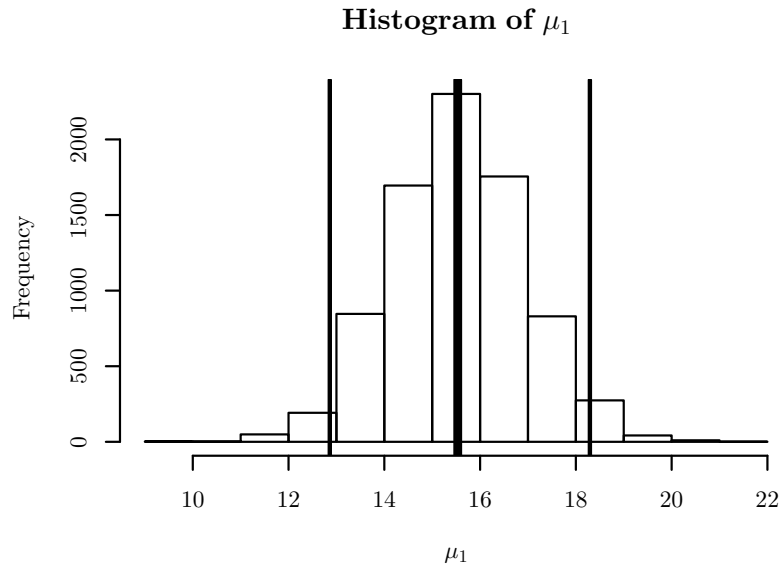
Looks reasonable. And the summary of our sampler suggests that \hat{R} is essentially 1.

Table 1: Summary of posterior draws.

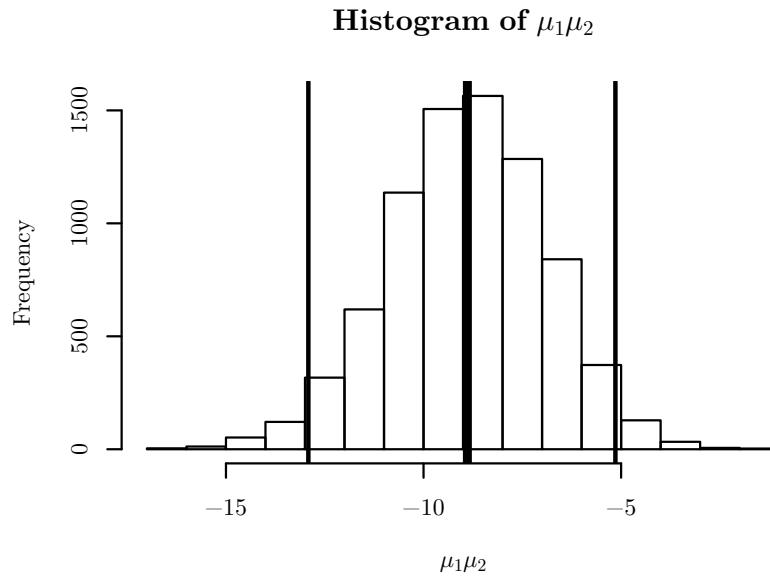
	mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
mu1	15.53	1.398	12.86	14.6	15.52	16.44	18.29	5587	1
mu2	24.42	1.435	21.67	23.47	24.4	25.35	27.31	5242	0.9999
mu3	15.22	1.842	11.63	13.96	15.2	16.45	18.8	5137	1.001
sigma	8.456	0.6974	7.238	7.963	8.405	8.894	9.977	4592	1
cut_rain	-8.89	1.995	-12.91	-10.2	-8.843	-7.535	-5.144	5392	1
lp__	-211.1	1.46	-214.9	-211.8	-210.8	-210.1	-209.3	2274	1.001

Results

Below is a histogram of the posterior of μ_1 , the average lifetime for cutthroat. The posterior mean and a 95% credible interval are marked.



Below is a histogram for the difference in mean time until death between cutthroat and rainbow.



(I've gone pretty light on the results here. Maybe others can spruce this up!)

Stan Model

```

data {
  int<lower=0> n1obs;
  int<lower=0> n2obs;
  int<lower=0> n3obs;
  int<lower=0> n1cens;
  int<lower=0> n2cens;
  int<lower=0> n3cens;
  real y1obs[n1obs];
  real y2obs[n2obs];
  real y3obs[n3obs];
  real y1cens[n1cens];
  real y2cens[n2cens];
  real y3cens[n3cens];
}

parameters {
  real<lower=0> mu1;
  real<lower=0> mu2;
  real<lower=0> mu3;
  real<lower=0> sigma;
}

model {
  mu1 ~ normal(20, 10);
  mu2 ~ normal(20, 10);
  mu3 ~ normal(20, 10);
  sigma ~ normal(0, 10);
  y1obs ~ normal(mu1, sigma);
  y2obs ~ normal(mu2, sigma);

```

```

y3obs ~ normal(mu3, sigma);
increment_log_prob(normal_ccdf_log(y1cens, mu1, sigma));
increment_log_prob(normal_ccdf_log(y2cens, mu2, sigma));
increment_log_prob(normal_ccdf_log(y3cens, mu3, sigma));
}

generated quantities {
  real cut_rain;

  cut_rain <- mu1 - mu2;
}

```