

Activity 12: Statistical reasoning 4: prediction and evaluation

Feb. 18th, 2026, Calvin Munson

Piper Miller & Rowan Crowley

Welcome! This is the fourth statistical reasoning activity. The goals of this activity are to understand how to evaluate and compare the predictive accuracy of multiple models. Specifically, you will:

- Run and interpret multiple models on the same dataset and evaluate them to see which is best supported using WAIC and PSIS.
-

You will submit one output for this activity:

1. A **PDF** of a rendered Quarto document with all of your R code. Please create a new Quarto document (e.g. don't use this `README.qmd`) and include all of the code that appears in this document, your own code, and **answers to all of the questions** in the “Q#” sections. Submit this PDF through Gradescope.

A reminder: **Please label the code** in your final submission in two ways:

1. denote your answers to each question using headers that correspond to the question you're answering, and
 2. thoroughly “comment” your code: remember, this means annotating your code directly by typing descriptions of what each line does after a `#`. This will help future you!
-

1. Practice with model comparison

By making models, we are trying to approximate the processes that affect things in a system. It's important to know how well our models are actually aligning with reality though. If we are not careful, we may “overfit” our models. *Overfitting* is when our model fits the data too closely, usually because we added too many parameters, which explains the data well but makes the model very bad at predicting future data (for instance, the final panel of the XKCD comic below).[XKCD_curve_fitting](#)

It is common in the field of ecology to have multiple candidate models of how a system works. How do we know which is “best” at making predictions? In this activity we will learn two metrics that can help: the Watanabe–Akaike information criterion (**WAIC**) and Pareto Smoothed Importance Sampling (**PSIS**).

Both metrics tell us how well the model will predict data it wasn't trained on, which is important for thinking about how well the model might predict new data that we as scientists have not encountered yet.

Let's start by reading in the relevant packages

```
library(brms) # for statistics
```

Loading required package: Rcpp

Loading 'brms' package (version 2.23.0). Useful instructions can be found by typing `help('brms')`. A more detailed introduction to the package is available through `vignette('brms_overview')`.

Attaching package: 'brms'

The following object is masked from 'package:stats':

ar

```
library(tidyverse) # for data wrangling
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.1      v tibble     3.2.1
v lubridate  1.9.4      v tidyr      1.3.1
v purrr      1.0.2
```

```
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(lterdatasampler)
#installed the packages
```

We are going to work with the fiddler crab and latitude data again:

```
pie_crab <- lterdatasampler::pie_crab
#loaded in data
```

In this section, we will run and interpret three multiple regressions to try and understand what influences crab body width in mm (**size**). These are data from crabs (~30 per site) collected from sites from Florida to Massachusetts. Let's remind ourselves of the columns in the crab data:

```
colnames(pie_crab)
```

```
[1] "date"          "latitude"      "site"          "size"
[5] "air_temp"      "air_temp_sd"  "water_temp"    "water_temp_sd"
[9] "name"
```

We have multiple temperature variables that may be relevant to crab body size here, all measured in degrees Celsius. Mean annual air and water temperature data (**air_temp**, **water_temp**), plus the standard deviations of each (**air_temp_sd** and **water_temp_sd**, representing variability in temperature and perhaps seasonality).

1.1 Create hypotheses for how each variable may affect crab size

Create four separate hypotheses describing how each predictor would be associated with larger or smaller crabs. Why? Please write 2-3 sentences for each predictor.

Q1.1a How might *mean annual water* temperature affect crab size?

A lower mean water temp may have the effect of an increase in crab size. This may function as an adaptation of a lower surface area to body ratio to maintain heat in a colder climate. Water temperature may have a lower effect on fiddler crabs than air temperature because they spend most of their time on land.

Q1.1b How might *mean annual air* temperature affect crab size?

Just like the mean annual water temperature, we think a colder climate and a lower mean annual air temperature might have an effect of an increase in crab size. This might function as an adaptation of a lower surface area to body ratio to maintain heat in a colder climate.

Q1.1c How might the *sd (variability) of water* temperature affect crab size?

If there is more variability in water temperature there might be a wider range of crab sizes within the population. Due to the variable environment which might favor multiple phenotype during different times.

Q1.1d How might the *sd (variability) of air* temperature affect crab size?

If there is high variability in air temperature, there might be a wider range of crab sizes within the population. The sd of water temperature and air temperature are going to follow a similar trend.

1.2 Run, assess, and interpret three multiple regressions with latitude and *mean* temperatures

Let's run three regressions and compare their results. We will start by looking at how body size varies with latitude plus each of the mean temperature values separately, then together. Since these are intertidal estuarine sites, crabs are exposed to water for part of the day and air for another part; air and/or water temperatures may be important. The models will be:

- $\text{size} \sim \text{latitude} + \text{water_temp}$
 - $\text{size} \sim \text{latitude} + \text{air_temp}$
 - $\text{size} \sim \text{latitude} + \text{water_temp} + \text{air_temp}$
-

size ~ latitude + mean water temp

```
# latitude and water model
m.crab.lat.water <-
  brm(data = pie_crab, # Give the model the pie_crab data
      # Choose a gaussian (normal) distribution
      family = gaussian,
      # Specify the model here.
      size ~ latitude + water_temp,
      # Here's where you specify parameters for executing the Markov chains
      # We're using similar to the defaults, except we set cores to 4 so the analysis runs f
      iter = 2000, warmup = 1000, chains = 4, cores = 4,
      # Setting the "seed" determines which random numbers will get sampled.
      # In this case, it makes the randomness of the Markov chain runs reproducible
      # (so that both of us get the exact same results when running the model)
      seed = 4,
      # Save the fitted model object as output - helpful for reloading in the output later
      file = "output/m.crab.lat.water")
```

Now look at the output:

```
summary(m.crab.lat.water)
```

```
Family: gaussian
Links: mu = identity
Formula: size ~ latitude + water_temp
```

Data: pie_crab (Number of observations: 392)
 Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
 total post-warmup draws = 4000

Regression Coefficients:

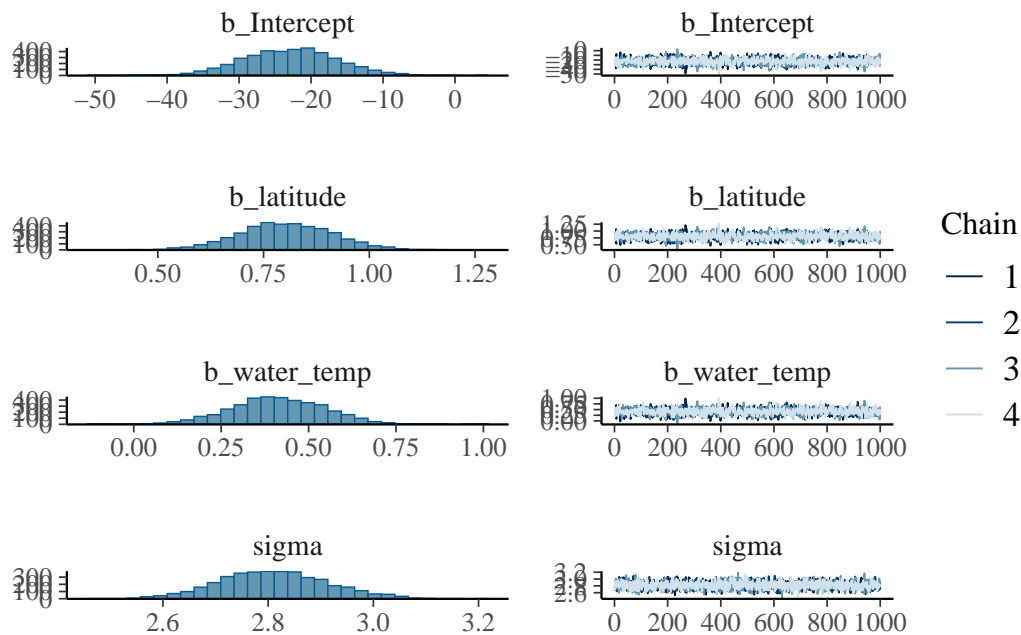
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	-22.90	6.54	-35.47	-10.22	1.00	1580	1739
latitude	0.80	0.11	0.59	1.02	1.00	1626	1724
water_temp	0.41	0.14	0.15	0.67	1.00	1631	2123

Further Distributional Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	2.81	0.10	2.62	3.02	1.00	2863	2575

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS and Tail_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

```
plot(m.crab.lat.water)
```



Q1.2a Assess the output

Assess whether the model ran correctly by looking at \hat{R} , the chains, and the posterior distributions using the `plot()` and `summary()` functions. Describe your thought process about whether the model ran correctly in 1-2 sentences.

We think the model ran correctly due to an \hat{R} of 1 which means the chains converged on a consensus. The posterior distributions look normally distributed with a wide spread and the estimate of slope does not include 0 in the 95% credible interval meaning the slope has an effect we are sure is not zero.

Q1.2b Interpret the output

Interpret your model by answering:

1. What are the effects of your predictors? Remember to describe the effect using the units to make it biologically meaningful.

In our multiple regression model for every 1 degree increase in latitude there is a 0.80 mm increase in crab carapace size. For every 1 degree increase in mean annual water temperature there is a 0.41 mm increase in crab carapace size.

2. Are the effects reasonably different from zero? How do you know?

Yes, both of these effects are reasonably different from zero because they do not include zero in their 95% credible intervals.

size ~ latitude + mean air temp

```
# latitude and air model
m.crab.lat.air <-
  brm(data = pie_crab, # Give the model the pie_crab data
    # Choose a gaussian (normal) distribution
    family = gaussian,
    # Specify the model here.
    size ~ latitude + air_temp,
    # Here's where you specify parameters for executing the Markov chains
    # We're using similar to the defaults, except we set cores to 4 so the analysis runs f
    iter = 2000, warmup = 1000, chains = 4, cores = 4,
```

```
# Setting the "seed" determines which random numbers will get sampled.
# In this case, it makes the randomness of the Markov chain runs reproducible
# (so that both of us get the exact same results when running the model)
seed = 4,
# Save the fitted model object as output - helpful for reloading in the output later
file = "output/m.crab.lat.air")
```

```
summary(m.crab.lat.air)
```

```
Family: gaussian
Links: mu = identity
Formula: size ~ latitude + air_temp
Data: pie_crab (Number of observations: 392)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
       total post-warmup draws = 4000
```

Regression Coefficients:

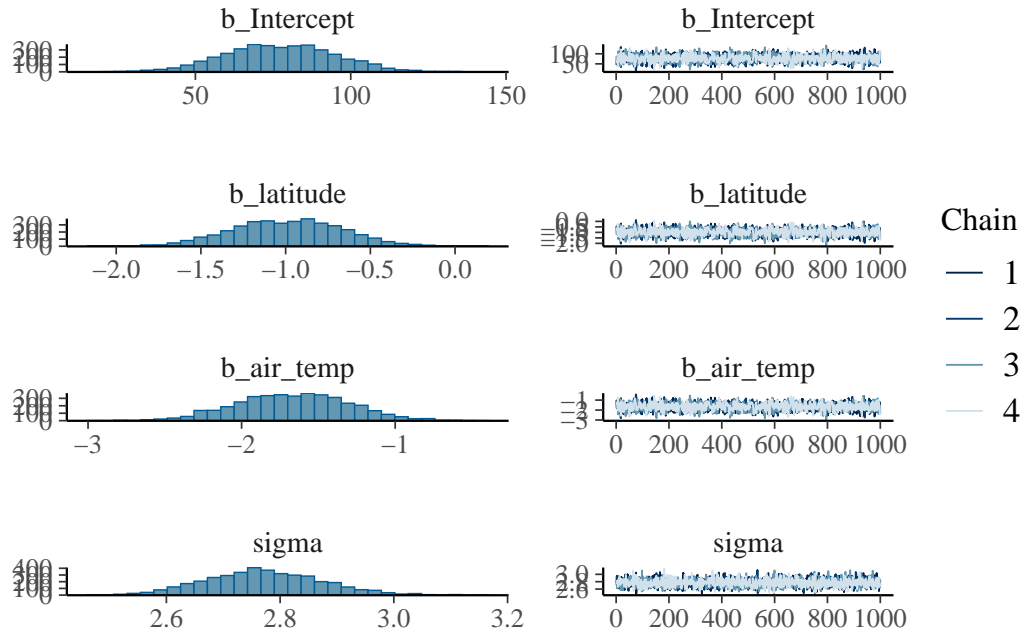
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	77.31	17.75	43.62	111.51	1.00	1730	1898
latitude	-0.99	0.32	-1.61	-0.37	1.00	1732	1977
air_temp	-1.67	0.36	-2.36	-0.97	1.00	1735	1909

Further Distributional Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	2.77	0.10	2.58	2.98	1.00	2400	2104

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS and Tail_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

```
plot(m.crab.lat.air)
```

Q1.3a Assess the output

Assess whether the model ran correctly by looking at \hat{R} , the chains, and the posterior distributions using the `plot()` and `summary()` functions. Describe your thought process about whether the model ran correctly in 1-2 sentences.

\hat{R} is 1, meaning the chains converged on a consensus. The posterior distributions lack a clear peak but looks to be a little bi-modal.

Q1.3b Interpret the output

Interpret your model by answering:

1. What are the effects of your predictors? Remember to describe the effect using the units to make it biologically meaningful.

With every 1 degree increase in latitude, crab size decreases by 0.99mm. With every 1 degree increase in air temperature, crab size decreases by 1.67mm.

2. Are the effects reasonably different from zero? How do you know?

Yes the effects are reasonably different from zero. The 95% CI do not include zero.

size ~ latitude + mean water + mean air temp

```
# latitude and air model
m.crab.lat.air.water <-
  brm(data = pie_crab, # Give the model the pie_crab data
    # Choose a gaussian (normal) distribution
    family = gaussian,
    # Specify the model here.
    size ~ latitude + air_temp + water_temp,
    # Here's where you specify parameters for executing the Markov chains
    # We're using similar to the defaults, except we set cores to 4 so the analysis runs f
    iter = 2000, warmup = 1000, chains = 4, cores = 4,
    # Setting the "seed" determines which random numbers will get sampled.
    # In this case, it makes the randomness of the Markov chain runs reproducible
    # (so that both of us get the exact same results when running the model)
    seed = 4,
    # Save the fitted model object as output - helpful for reloading in the output later
    file = "output/m.crab.lat.air.water")
```

```
summary(m.crab.lat.air.water)
```

```
Family: gaussian
Links: mu = identity
Formula: size ~ latitude + air_temp + water_temp
Data: pie_crab (Number of observations: 392)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
       total post-warmup draws = 4000
```

Regression Coefficients:

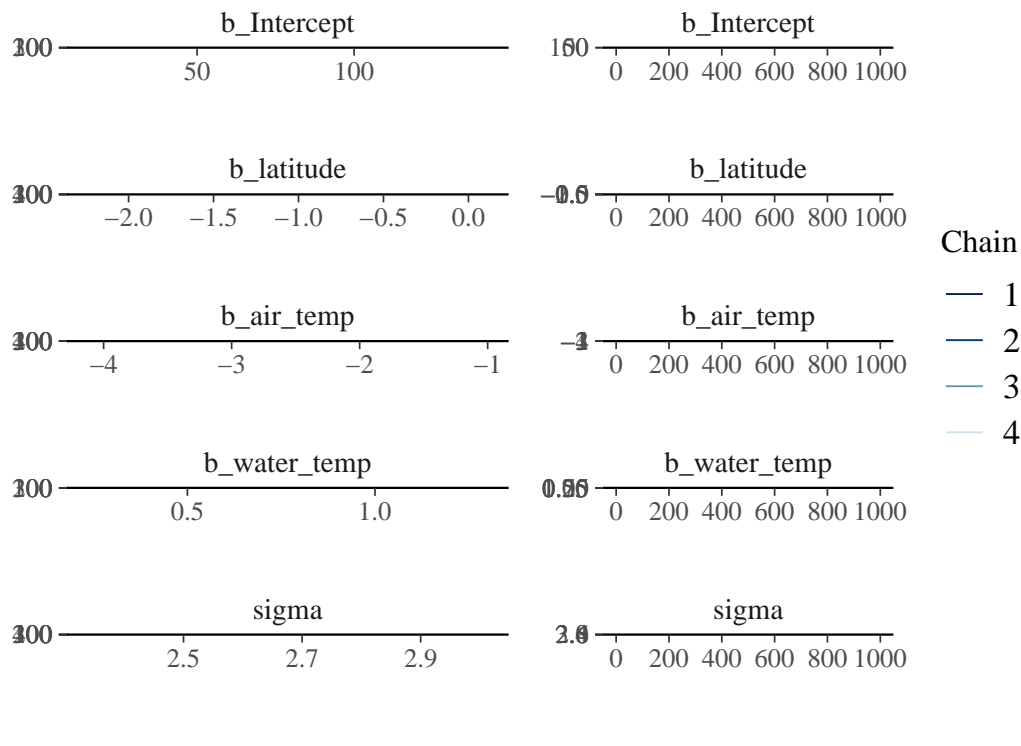
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	77.57	17.57	43.30	112.07	1.00	2624	2189
latitude	-1.06	0.32	-1.69	-0.43	1.00	2578	2242
air_temp	-2.40	0.39	-3.18	-1.65	1.00	2374	2297
water_temp	0.76	0.15	0.46	1.05	1.00	3076	2264

Further Distributional Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	2.68	0.09	2.51	2.87	1.00	3887	2492

Draws were sampled using `sampling(NUTS)`. For each parameter, Bulk_ESS and Tail_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

```
plot(m.crab.lat.air.water)
```



Q1.4a Assess the output

Assess whether the model ran correctly by looking at R hat, the chains, and the posterior distributions using the `plot()` and `summary()` functions. Describe your thought process about whether the model ran correctly in 1-2 sentences.

The rhat is 1 which means the mcmc chains ran correctly. The posterior distributions are normally distributed but very wide spreading.

Q1.4b Interpret the output

Interpret your model by answering:

1. What are the effects of your predictors? Remember to describe the effect using the units to make it biologically meaningful.

In the multiple regression model for every 1 degree increase in latitude there is a decrease of 1.06 mm in crab size. For every 1 degree increase in mean annual air temperature there is a 2.41 mm decrease in crab size. For every 1 degree increase in mean annual water temperature there is a 0.76 mm increase in crab size.

2. Are the effects reasonably different from zero? How do you know?

Yes, these effects are reasonable different from zero because they do not include 0 in their 95% credible interval.

Q1.5 How do the models differ in their estimates?

In 2-4 sentences, compare the three models' estimates of the effect of latitude, water temp, and air temp; did estimates change across different models? Stay the same? Change in whether or not they are different from zero?

In the multiple regression that only included latitude and water temperature the latitude showed an increase in crab size, however in the models which include air temperature the latitude is showing a decrease in crab size. Additionally when all of them air temperature, latitude and water temperature are taken into account the slope estimate for each became larger than in the regressions only including two variables. Whether the effects are different than zero does not change between regressions.

Q1.6 Why do you think a variable's sign changed?

You should have noticed the change in sign for a variable. In 1-2 sentences, and in the context of your knowledge about causal inference from DAGs from last week, describe why you think the variable may have changed signs (hint: remember pipes?).

Latitude may have changed from positive to negative when mean annual air temperature was taken into account because there is a pipe relationship between the three variables where latitude affects air temperature which effects water temperature. This leads to a situation where when the air temperature is not taken into account youre missing an important causal relationship. This caused the effects of latitude to be masked.

1.3 Compare models using WAIC and PSIS

We just compared the models in terms of what values they provided for the estimates of the effects of `latitude`, `water_temp`, and `air_temp`. Now we are going to compare models using the Pareto Smoothed Importance Sampling (**PSIS**) and Watanabe–Akaike information criterion (**WAIC**). Remember, both of these metrics will tell us about a model’s out of sample predictive skill. Lower values = better! A major reason we do this is to avoid *overfitting*, where more complex models with lots of parameters are un-generalizable to out of sample data.

First, let’s look at the PSIS output from the three models. Remember, lower values are better, and more complicated models (models with more parameters) will be “punished”, since more parameters risks overfitting. The PSIS values will be in the `Estimate` column in the third row.

```
# Look at "leave one out" results for all three models
# size ~ lat + mean water
loo(m.crab.lat.water)
```

Computed from 4000 by 392 log-likelihood matrix.

	Estimate	SE
elpd_loo	-962.7	13.6
p_loo	3.5	0.3
looic	1925.5	27.2

MCSE of elpd_loo is 0.0.

MCSE and ESS estimates assume MCMC draws (r_eff in [0.4, 0.9]).

All Pareto k estimates are good (k < 0.7).

See `help('pareto-k-diagnostic')` for details.

```
# size ~ lat + mean air
loo(m.crab.lat.air)
```

Computed from 4000 by 392 log-likelihood matrix.

	Estimate	SE
elpd_loo	-957.3	13.5
p_loo	4.0	0.3
looic	1914.6	26.9

MCSE of elpd_loo is 0.0.

MCSE and ESS estimates assume MCMC draws (r_eff in [0.4, 0.9]).

All Pareto k estimates are good (k < 0.7).

See help('pareto-k-diagnostic') for details.

```
# size ~ lat + mean water + mean air
loo(m.crab.lat.air.water)
```

Computed from 4000 by 392 log-likelihood matrix.

	Estimate	SE
elpd_loo	-945.2	14.2
p_loo	4.9	0.5
looic	1890.5	28.4

MCSE of elpd_loo is 0.0.

MCSE and ESS estimates assume MCMC draws (r_eff in [0.7, 1.0]).

All Pareto k estimates are good (k < 0.7).

See help('pareto-k-diagnostic') for details.

The first thing to look for to assess the leave one out method is the Pareto k estimate. For us, it gives the helpful message: All Pareto k estimates are good (k < 0.7).

The last row in each table's output is our PSIS value. Remember, lower = better. As we can see, the size ~ latitude + mean water + mean air model had the lowest PSIS value, despite having the most parameters. This indicates that the extra parameter made up for the punishment by adding much more predictive power.

Now we do the same for WAIC:

```
# Look at "leave one out" results for all three models
# size ~ lat + mean water
waic(m.crab.lat.water)
```

Computed from 4000 by 392 log-likelihood matrix.

	Estimate	SE
elpd_waic	-962.7	13.6
p_waic	3.5	0.3
waic	1925.5	27.2

```
# size ~ lat + mean air
waic(m.crab.lat.air)
```

Computed from 4000 by 392 log-likelihood matrix.

	Estimate	SE
elpd_waic	-957.3	13.5
p_waic	4.0	0.3
waic	1914.6	26.9

```
# size ~ lat + mean water + mean air
waic(m.crab.lat.air.water)
```

Computed from 4000 by 392 log-likelihood matrix.

	Estimate	SE
elpd_waic	-945.2	14.2
p_waic	4.9	0.4
waic	1890.5	28.4

Q1.7 Which model has the “best” WAIC value?

The model that includes latitude, air temperature and water temperature has the best WAIC value because it is a lower number 1890.5 compared to the WAIC above 1900 from the other models. Both PSIS and WAIC have the same values which is a good sign for our results.

Importantly, we want both the PSIS results and the WAIC results to align. In this case, they do, which is a good sign for our models.

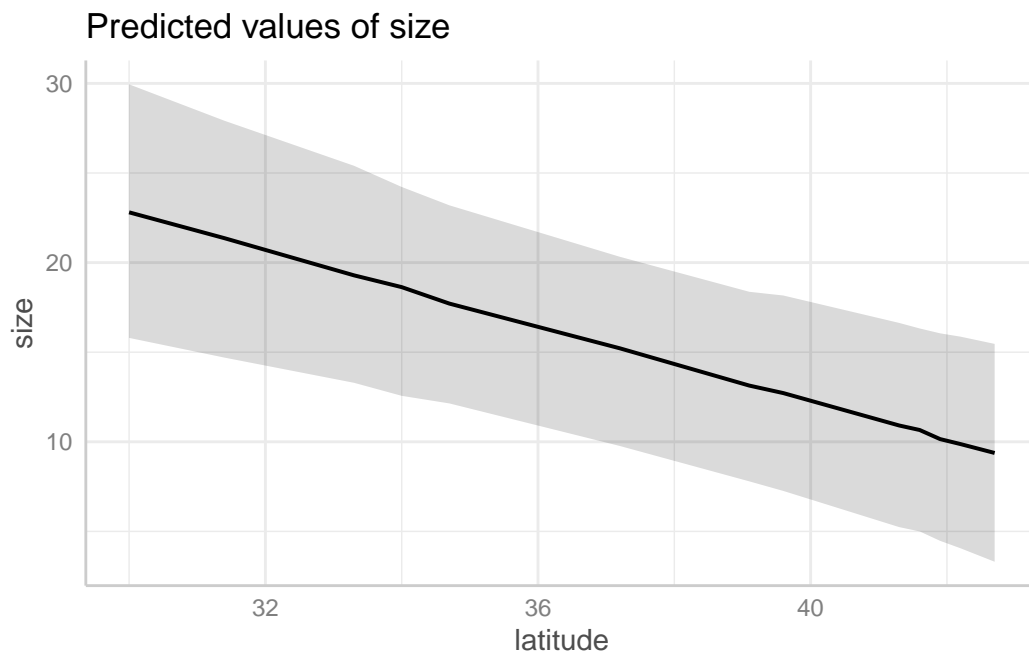
1.4 Look at uncertainty around model predictions

Here we will look at some of the ways we can look at the uncertainty around model predictions from the “model evaluation” lecture using the most complex model with `size ~ latitude + mean water + mean air`.

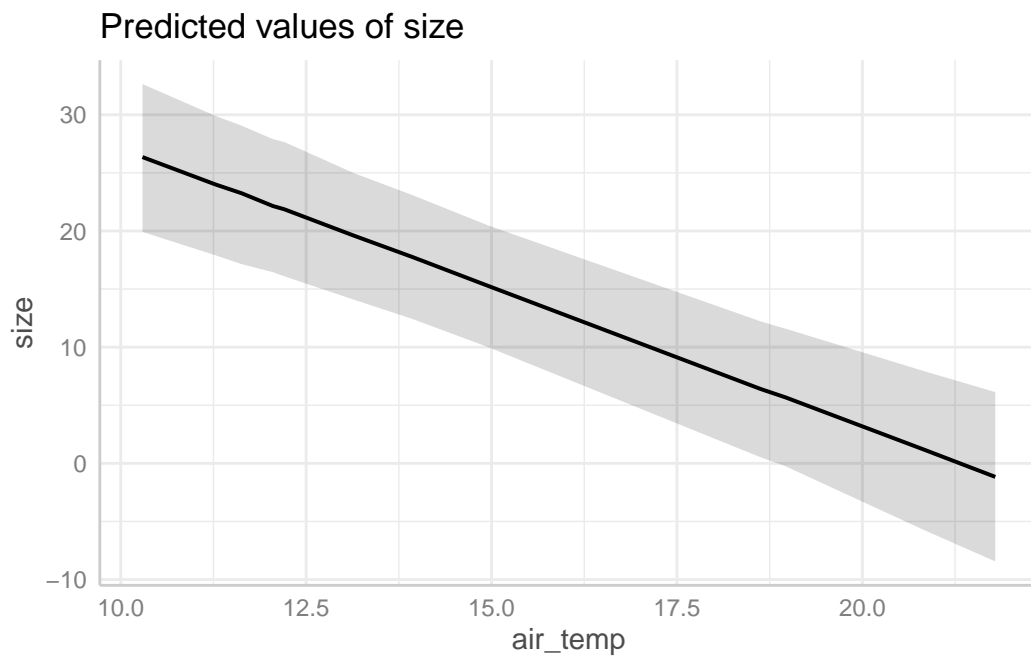
The `predict_response(interval = "prediction")` function plots the 95% prediction intervals separately for each predictor, displaying uncertainty around where the data may lay around the model.

```
preds <-ggeffects::predict_response(m.crab.lat.air.water,  
                                     interval = "prediction")  
plot(preds)
```

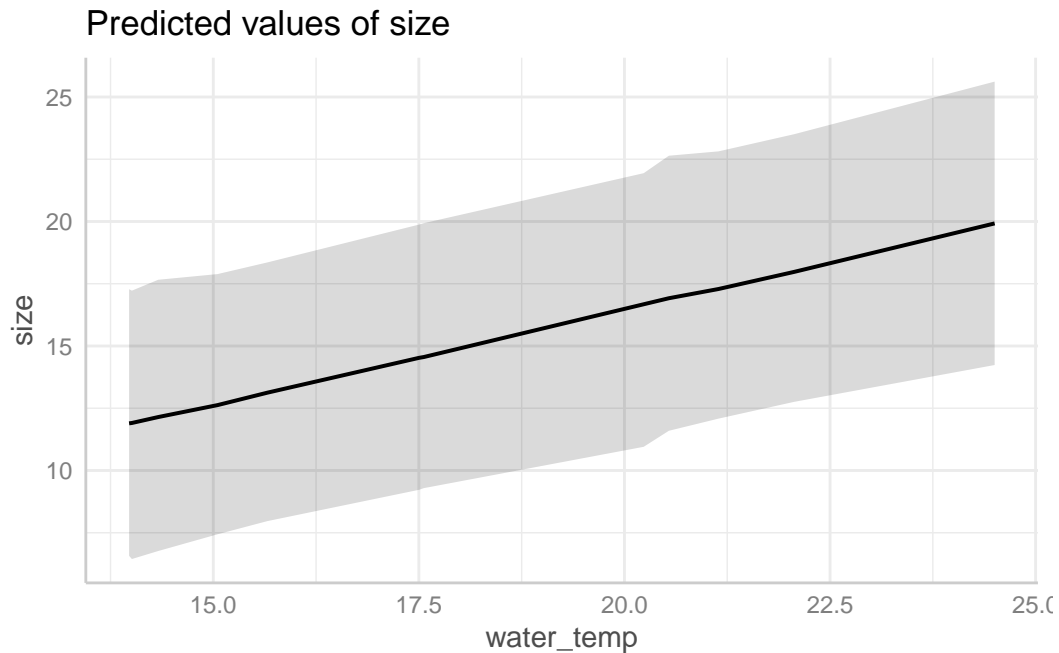
\$latitude



`$air_temp`



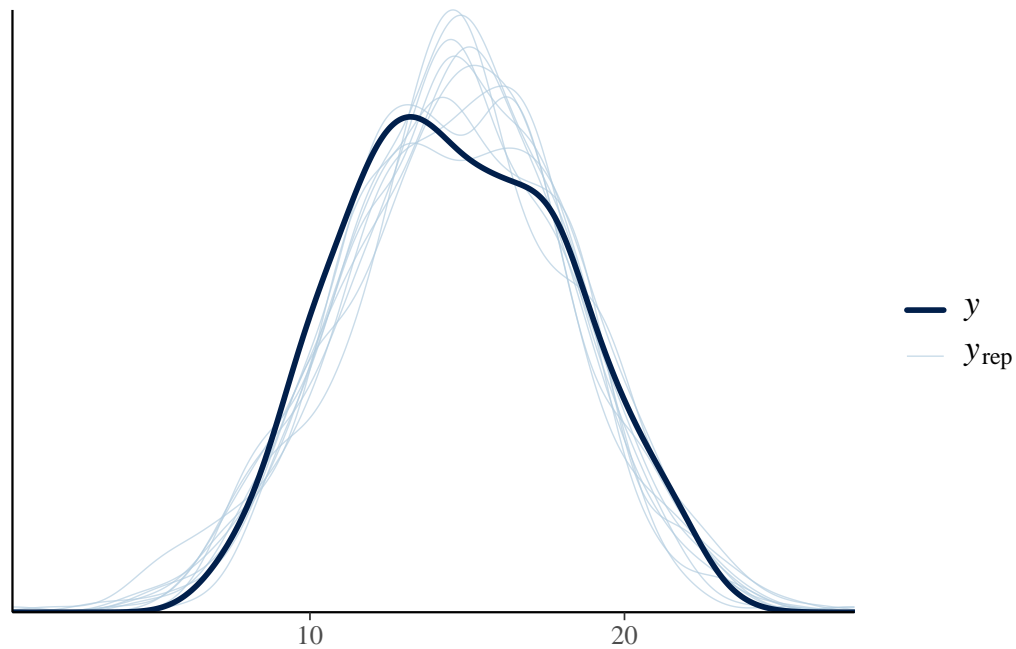
`$water_temp`



There are a few posterior predictive check plots we can look at. For instance, `pp_check(type = "dens_overlay")` shows the probability density of the observed data in a heavy line. The thin blue lines show the range of probability densities that are expected if you simulate from the fitted model's range of estimated posteriors. We want the thin blue lines to align pretty well with the heavy line.

```
pp_check(m.crab.lat.air.water, type = "dens_overlay")
```

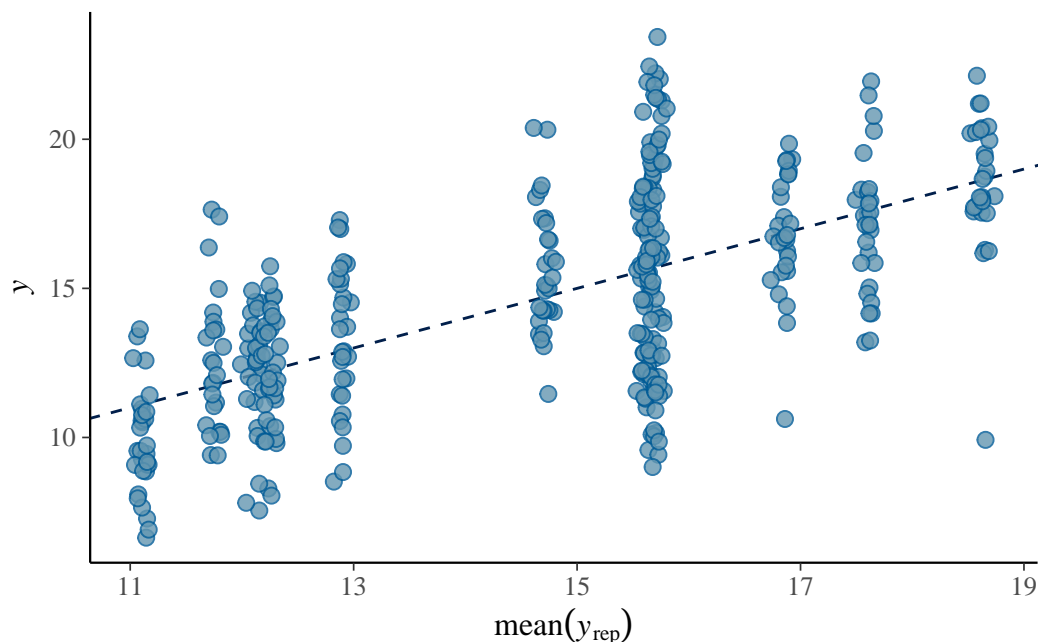
Using 10 posterior draws for ppc type 'dens_overlay' by default.



`pp_check(type = "scatter_avg")` shows the observed values on the y axis and the average of the predicted values on the x axis using a scatterplot. Having all of the points fall along the 1:1 line would indicate good fit. Points that fall outside that line can help us understand whether there are missing predictors.

```
pp_check(m.crab.lat.air.water, type = "scatter_avg")
```

Using all posterior draws for ppc type 'scatter_avg' by default.



Here, it seems that there is a lot of variation at each site: remember, each site (which has one value of latitude) has ~30 crabs. The model line goes through approximately the middle of each cloud of points, which is good, but there is still a lot of unexplained variation at seemingly the site level. This may mean that there are site-specific variables that are causing variation in crab size that are not captured in this model.

2. Repeat with the sd of water and air temp instead of mean temp

Now it's your turn! In this section, repeat what we just did but with the standard deviation (sd) of water and air temperature instead of the mean air and water temperature.

The three models should be:

- $\text{size} \sim \text{latitude} + \text{water_temp_sd}$
 - $\text{size} \sim \text{latitude} + \text{air_temp_sd}$
 - $\text{size} \sim \text{latitude} + \text{water_temp_sd} + \text{air_temp_sd}$
-

Q2.1 Run all three models

```
colnames(pie_crab)
```

```
[1] "date"          "latitude"      "site"          "size"
[5] "air_temp"      "air_temp_sd"   "water_temp"     "water_temp_sd"
[9] "name"
```

Run and store all three models. Remember to change the name of 1) the data that the model output is stored as and 2) the output file name

size ~ latitude + water temp sd

```
# latitude and water temp sd model
m.crab.lat.water.sd <-
  brm(data = pie_crab, # Give the model the pie_crab data
    # Choose a gaussian (normal) distribution
    family = gaussian,
    # Specify the model here.
    size ~ latitude + water_temp_sd,
    # Here's where you specify parameters for executing the Markov chains
    # We're using similar to the defaults, except we set cores to 4 so the analysis runs f
    iter = 2000, warmup = 1000, chains = 4, cores = 4,
    # Setting the "seed" determines which random numbers will get sampled.
    # In this case, it makes the randomness of the Markov chain runs reproducible
    # (so that both of us get the exact same results when running the model)
    seed = 4,
    # Save the fitted model object as output - helpful for reloading in the output later
    file = "m.crab.lat.water.sd")
```

```
summary(m.crab.lat.water.sd)
```

```
Family: gaussian
Links: mu = identity
Formula: size ~ latitude + water_temp_sd
Data: pie_crab (Number of observations: 392)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
       total post-warmup draws = 4000
```

Regression Coefficients:

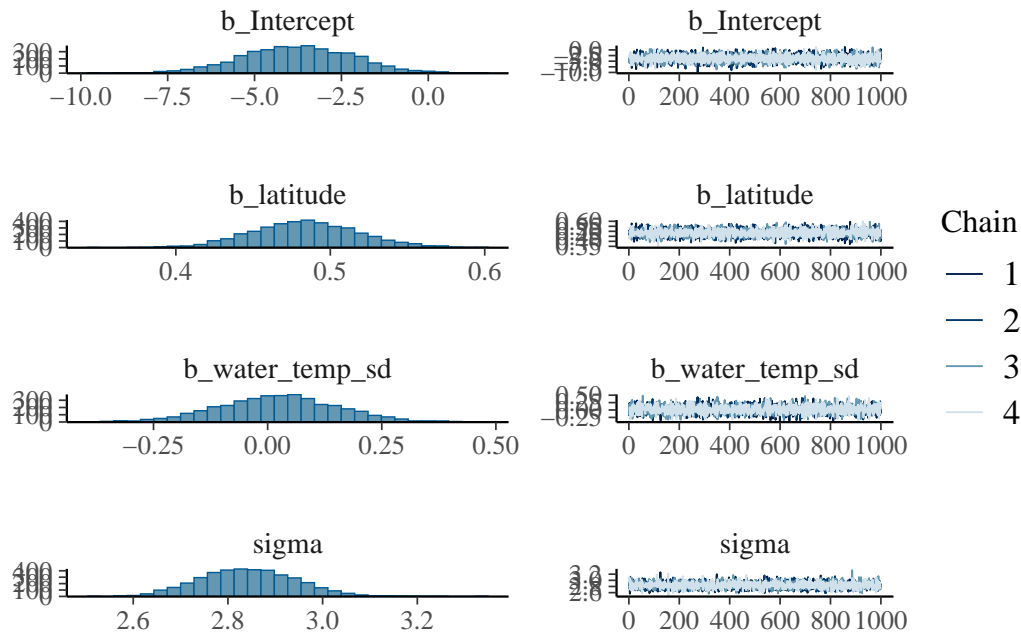
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	-3.76	1.58	-6.89	-0.72	1.00	5059	3408
latitude	0.48	0.03	0.42	0.55	1.00	4836	2873
water_temp_sd	0.02	0.13	-0.23	0.28	1.00	5109	2732

Further Distributional Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	2.84	0.10	2.65	3.04	1.00	5074	3272

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS and Tail_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

```
plot(m.crab.lat.water.sd)
```



The latitude and water temperature variation model looks like it ran correctly with an rhat of 1 and normally distributed posterior distributions. The effect looks to be for every 1 degree increase in latitude there is a 0.48 mm increase in crab size and also for every 1 increase in water temperature variability there is an increase in crab size of 0.02 mm. The latitude does

not include 0 in its 95% credible interval so we can be pretty sure it is different from zero, however this is not the case for the water temperature variability, so it may have no real effect.

size ~ latitude + air temp sd

```
# latitude and air sd model
m.crab.lat.air.sd <-
  brm(data = pie_crab, # Give the model the pie_crab data
    # Choose a gaussian (normal) distribution
    family = gaussian,
    # Specify the model here.
    size ~ latitude + air_temp_sd,
    # Here's where you specify parameters for executing the Markov chains
    # We're using similar to the defaults, except we set cores to 4 so the analysis runs f
    iter = 2000, warmup = 1000, chains = 4, cores = 4,
    # Setting the "seed" determines which random numbers will get sampled.
    # In this case, it makes the randomness of the Markov chain runs reproducible
    # (so that both of us get the exact same results when running the model)
    seed = 4,
    # Save the fitted model object as output - helpful for reloading in the output later
    file = "m.crab.lat.air.sd")

summary(m.crab.lat.air.sd)
```

```
Family: gaussian
Links: mu = identity
Formula: size ~ latitude + air_temp_sd
Data: pie_crab (Number of observations: 392)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
       total post-warmup draws = 4000
```

Regression Coefficients:

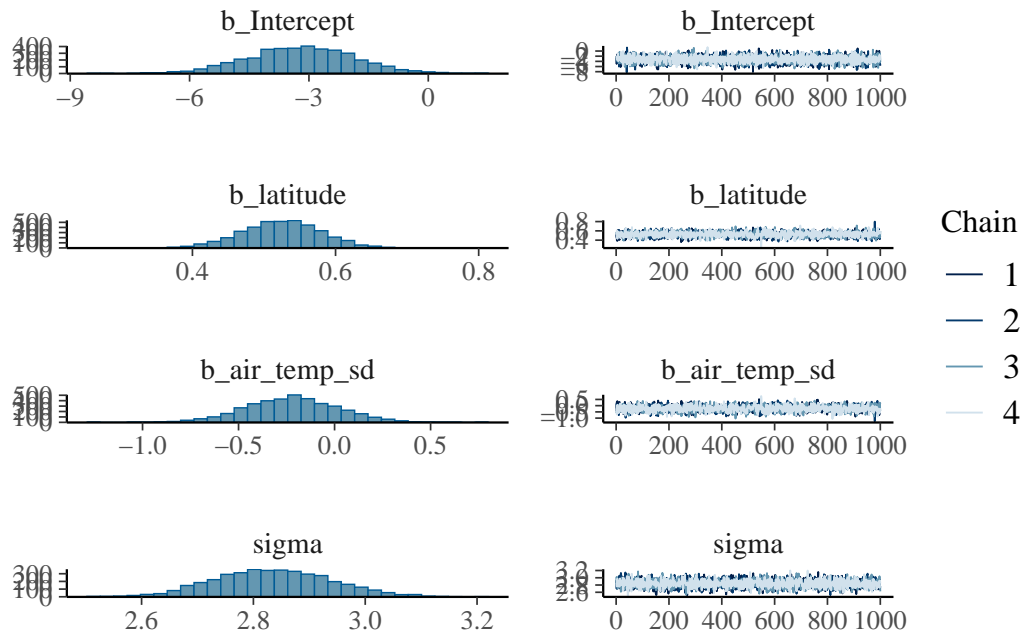
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	-3.20	1.33	-5.81	-0.60	1.00	5650	3329
latitude	0.52	0.05	0.41	0.63	1.00	2156	2482
air_temp_sd	-0.22	0.24	-0.69	0.24	1.00	2060	2401

Further Distributional Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	2.84	0.10	2.64	3.04	1.00	3298	2696

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS and Tail_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

```
plot(m.crab.lat.air.sd)
```



The latitude and air temperature variation model looks like it ran correctly with an rhat of 1 and normally distributed posterior distributions. The effect looks to be for every 1 degree increase in latitude there is a 0.52 mm increase in crab size and also for every 1 increase in air temperature variability there is an decrease in crab size of 0.22 mm. The latitude does not include 0 in its 95% credible interval so we can be pretty sure it is different from zero, however this is not the case for the air temperature variability, so it may have no real effect.

size ~ latitude + water temp sd + air temp sd

```
# latitude, air sd, water temp sd
m.crab.lat.air.water.sd <-
  brm(data = pie_crab, # Give the model the pie_crab data
       # Choose a gaussian (normal) distribution
```



```

family = gaussian,
# Specify the model here.
size ~ latitude + air_temp_sd + water_temp_sd,
# Here's where you specify parameters for executing the Markov chains
# We're using similar to the defaults, except we set cores to 4 so the analysis runs f
iter = 2000, warmup = 1000, chains = 4, cores = 4,
# Setting the "seed" determines which random numbers will get sampled.
# In this case, it makes the randomness of the Markov chain runs reproducible
# (so that both of us get the exact same results when running the model)
seed = 4,
# Save the fitted model object as output - helpful for reloading in the output later
file = "m.crab.lat.air.water.sd")
summary(m.crab.lat.air.water.sd)

```

```

Family: gaussian
Links: mu = identity
Formula: size ~ latitude + air_temp_sd + water_temp_sd
Data: pie_crab (Number of observations: 392)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
       total post-warmup draws = 4000

```

Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	-3.97	1.53	-6.95	-0.95	1.00	3052	2695
latitude	0.56	0.06	0.43	0.69	1.00	1594	1983
air_temp_sd	-0.43	0.31	-1.04	0.18	1.00	1479	2034
water_temp_sd	0.16	0.16	-0.16	0.47	1.00	1856	2617

Further Distributional Parameters:

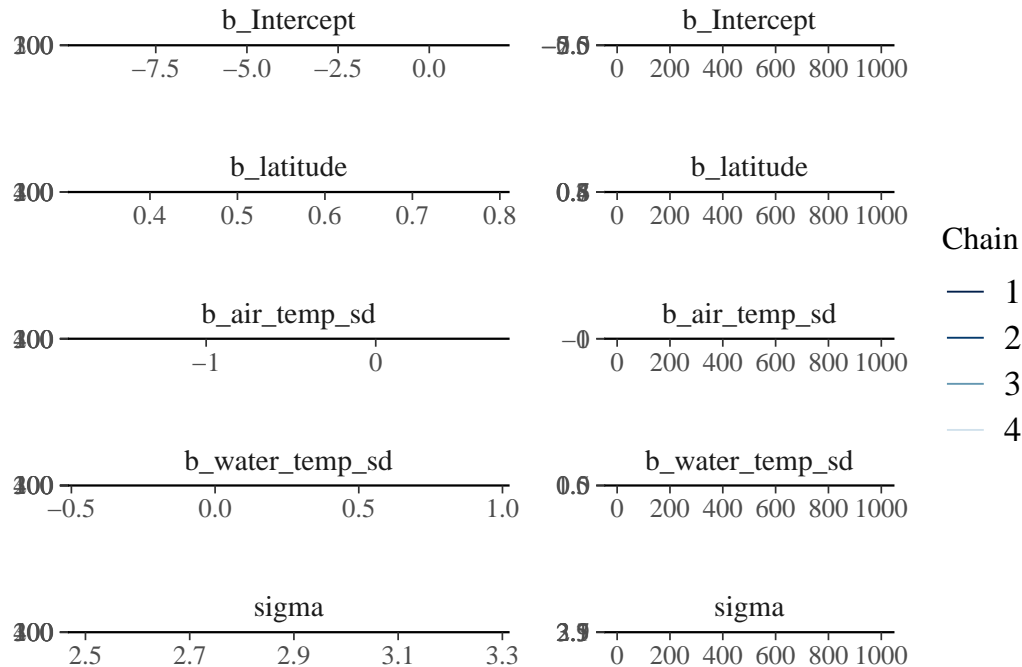
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	2.84	0.10	2.65	3.05	1.00	2760	2554

Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS and Tail_ESS are effective sample size measures, and Rhat is the potential scale reduction factor on split chains (at convergence, Rhat = 1).

```

plot(m.crab.lat.air.water.sd)

```



The latitude, air temperature variation, and water temperature variation model looks like it ran correctly with an \hat{r} of 1 and normally distributed posterior distributions but wide. The effect looks to be for every 1 degree increase in latitude there is a 0.56 mm increase in crab size and also for every 1 increase in air temperature variability there is an decrease in crab size of 0.43 mm. For every one increase in water temperature variability, there is an increase in crab size of 0.16mm. The latitude does not include 0 in its 95% credible interval so we can be pretty sure it is different from zero, however this is not the case for the air temperature variability and water temperature variability, therefore they could have zero effect on crab size.

Q2.2 Assess all three models

Assess whether each model ran correctly by looking at \hat{R} , the chains, and the posterior distributions using the `plot()` and `summary()` functions. Describe your thought process about whether the model ran correctly in 1-2 sentences per model.

Look above \hat{r} .

Q2.3 Interpret all three models

Interpret all three models by answering:

1. What are the effects of your predictors? Remember to describe the effect using the units to make it biologically meaningful.
2. Are the effects reasonably different from zero? How do you know?

Look above $\hat{\cdot}$.

Please write 2-3 sentences for each model

Q2.4 How do the models differ in their parameter estimates?

In 2-4 sentences, compare the three models' estimates of the effect of latitude, water temp sd, and air temp sd; did estimates change across different models? Stay the same? Change in whether or not they are different from zero?

Both effects of water temperature variability and air temperature variability seemed to increase when all three variables were taken to account in the multiple regression compared to the other two regressions accounting for latitude and one of temperature variability variables (water or air). The latitude slope remained relatively the same across models and we can be reasonably sure for every one that the effect is greater than zero because of the credibility intervals. For both the water temperature variability and the air temperature variability in every model they could have an effect of zero due to zero being included in the 95% credible intervals.

Q2.5 Calculate and compare PSIS and AIC values for each model

Calculate and compare the PSIS and AIC values for each model and answer:

1. Are the Pareto k estimates good?
Yes all of them are good with a $k < 0.7$
2. Which model has the lowest PSIS?
The latitude and air temp variation model has the lowest PSIS, but not by much.
3. Which model has the lowest AIC?
The latitude and air temp variation model has the lowest AIC, but not by much.

4. Do PSIS and AIC values agree on which model has the best out of sample prediction?

Yes, both PSIS and AIC values are the exact same for every model, meaning they agree on each models prediction capability.

```
#lat vs water sd checking accuracy of model predictions
loo(m.crab.lat.water.sd)
```

Computed from 4000 by 392 log-likelihood matrix.

	Estimate	SE
elpd_loo	-967.2	13.3
p_loo	3.8	0.3
looic	1934.4	26.6

MCSE of elpd_loo is 0.0.

MCSE and ESS estimates assume MCMC draws (r_eff in [1.1, 1.3]).

All Pareto k estimates are good (k < 0.7).

See help('pareto-k-diagnostic') for details.

```
waic(m.crab.lat.water.sd)
```

Computed from 4000 by 392 log-likelihood matrix.

	Estimate	SE
elpd_waic	-967.2	13.3
p_waic	3.8	0.3
waic	1934.3	26.6

```
#lat and air sd checking accuracy of model predictions
loo(m.crab.lat.air.sd)
```

Computed from 4000 by 392 log-likelihood matrix.

	Estimate	SE
elpd_loo	-966.5	13.5
p_loo	3.5	0.3
looic	1933.0	27.0

MCSE of elpd_loo is 0.0.

MCSE and ESS estimates assume MCMC draws (r_eff in [0.5, 1.2]).

All Pareto k estimates are good ($k < 0.7$).
See `help('pareto-k-diagnostic')` for details.

```
waic(m.crab.lat.air.sd)
```

Computed from 4000 by 392 log-likelihood matrix.

	Estimate	SE
elpd_waic	-966.5	13.5
p_waic	3.5	0.3
waic	1933.0	27.0

```
#lat, air sd and water sd checking accuracy of model predictions  
loo(m.crab.lat.air.water.sd)
```

Computed from 4000 by 392 log-likelihood matrix.

	Estimate	SE
elpd_loo	-967.0	13.2
p_loo	4.5	0.4
looic	1934.0	26.4

MCSE of elpd_loo is 0.0.

MCSE and ESS estimates assume MCMC draws (`r_eff` in `[0.4, 1.0]`).

All Pareto k estimates are good ($k < 0.7$).
See `help('pareto-k-diagnostic')` for details.

```
waic(m.crab.lat.air.water.sd)
```

Computed from 4000 by 392 log-likelihood matrix.

	Estimate	SE
elpd_waic	-967.0	13.2
p_waic	4.4	0.4
waic	1934.0	26.4

Render to PDF

When you have finished, remember to pull, stage, commit, and push with GitHub:

- Pull to check for updates to the remote branch
- Stage your edits (after saving your document!) by checking the documents you'd like to push
- Commit your changes with a commit message
- Push your changes to the remote branch

Then submit the well-labeled PDF on Gradescope. Thanks!