

Statistical Reasoning 1

Ellie and Paige

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

Warning: package 'brms' was built under R version 4.4.3

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)
```

Warning: package 'ggplot2' was built under R version 4.4.3

Warning: package 'tibble' was built under R version 4.4.3

Warning: package 'purrr' was built under R version 4.4.3

```
-- 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   4.0.1     v tibble    3.3.1
v lubridate 1.9.4     v tidyverse 1.3.1
v purrr     1.2.1

-- 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 non-conflicting
```

```
library(ggeffects) # for the prediction plot
```

```
Warning: package 'ggeffects' was built under R version 4.4.3
```

```
library(lterdatasampler) # for built-in datasets
```

```
Warning: package 'lterdatasampler' was built under R version 4.4.3
```

Load in data

```
head(pie_crab)
```

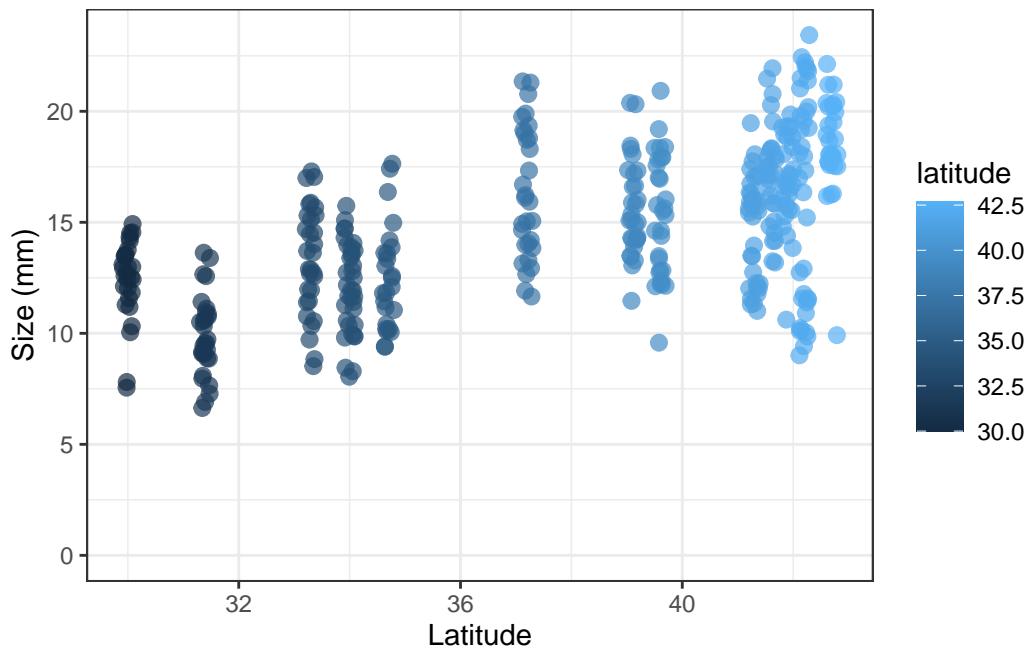
```
# A tibble: 6 x 9
  date      latitude site    size air_temp air_temp_sd water_temp water_temp_sd
  <date>        <dbl> <chr>    <dbl>     <dbl>       <dbl>       <dbl>       <dbl>
1 2016-07-24      30 GTM     12.4     21.8      6.39      24.5      6.12
2 2016-07-24      30 GTM     14.2     21.8      6.39      24.5      6.12
3 2016-07-24      30 GTM     14.5     21.8      6.39      24.5      6.12
4 2016-07-24      30 GTM     12.9     21.8      6.39      24.5      6.12
5 2016-07-24      30 GTM     12.4     21.8      6.39      24.5      6.12
6 2016-07-24      30 GTM     13.0     21.8      6.39      24.5      6.12
# i 1 more variable: name <chr>
```

Q1.1 Interpret the graph

Answer: It appears that crabs may on average get larger as latitude increases, but variance in size also increases.

Q1.2 Beautify this graph

```
pie_crab %>%
  ggplot(aes(x = latitude, y = size, color = latitude)) +
  geom_jitter(size = 2.5, alpha = 0.7, width = .1) +
  scale_color_continuous()+
  # Make the y-axis include 0
  ylim(0, NA)+
  labs(x= "Latitude",
       y = "Size (mm)")+
  theme_bw()
```



Q1.3

The `iter=` argument specifies the number of times to run the model, or the number of iterations.

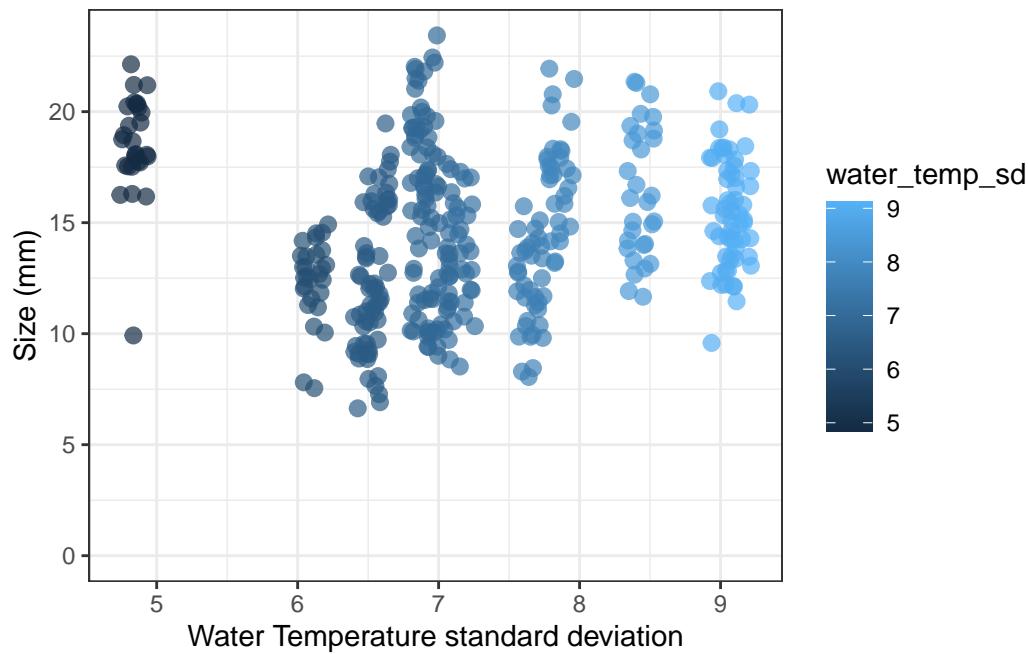
Q1.4 Make a hypothesis

We hypothesize that higher variability in water temperature will increase crab size because there will be more nutrients in the water and increased metabolism, thereby increasing growth

rate.

Q1.5 Graph the data

```
pie_crab %>%
  ggplot(aes(x = water_temp_sd, y = size, color = water_temp_sd)) +
  geom_jitter(size = 2.5, alpha = 0.7, width = .1) +
  scale_color_continuous()+
  # Make the y-axis include 0
  ylim(0, NA) +
  labs(x= "Water Temperature standard deviation",
       y = "Size (mm)")+
  theme_bw()
```



Q1.6 Interpret the graph

Answer: there does not appear to be much variation in crab size explained by water temperature variation. We are fairly confident in this, as there is a lot of overlap between all groups.

Q1.7

```
# water temp variation model
m.crab.watersd <-
  brm(data = pie_crab, # Give the model the pie_crab data
       # Choose a gaussian (normal) distribution
       family = gaussian,
       # Specify the model here.
       size ~ 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 faster
       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.watersd")

summary(m.crab.watersd)
```

```
Family: gaussian
Links: mu = identity
Formula: size ~ 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	13.94	1.15	11.66	16.22	1.00	4011	2715		
water_temp_sd	0.10	0.16	-0.21	0.41	1.00	3995	2685		

Further Distributional Parameters:

	Estimate	Est.Error	l-95%	CI	u-95%	CI	Rhat	Bulk_ESS	Tail_ESS
sigma	3.51	0.13	3.28	3.77	1.00	4093	2770		

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).

Q1.8 Assess the model

```
# show posteriors and chains
plot(m.crab.watersd)

# show summary, including rhat
summary(m.crab.watersd)
```

Answer: The Rhat value = 1, the posterior distributions have a clear hump, and the chains overlap. Therefore, we can conclude that the models ran correctly.

Q1.9

Answer: The linear regression shows that carapace width increases by 0.10 mm per increase in water temperature standard deviation. However, the 95% credible intervals were between -0.21 and .41. Since this range includes 0, our model shows that we cannot conclude that there is any effect of water temp variation on crab size.

Q2.1 Make a question

```
head(nwt_pikas)

# A tibble: 6 x 8
  date      site    station utm_easting utm_northing sex concentration_pg_g
  <date>    <fct>   <fct>        <dbl>          <dbl> <fct>            <dbl>
1 2018-06-08 Cable Ga~ Cable ~       451373        4432963 male             11563.
2 2018-06-08 Cable Ga~ Cable ~       451411        4432985 male             10629.
3 2018-06-08 Cable Ga~ Cable ~       451462        4432991 male             10924.
4 2018-06-13 West Kno~ West K~       449317        4434093 male             10414.
5 2018-06-13 West Kno~ West K~       449342        4434141 male             13531.
6 2018-06-13 West Kno~ West K~       449323        4434273 <NA>              7799.
# i 1 more variable: elev_m <dbl>

nwt_pikas_doy <- nwt_pikas %>%
  # Add a new column called day_of_year
  # yday extracts the day of year from the date column
  mutate(day_of_year = yday(date)) %>%
  # relocate the day_of_year column after the date column
```

```

relocate(day_of_year, .after = date)

head(nwt_pikas_doy)

# A tibble: 6 x 9
  date      day_of_year site       station      utm_easting utm_northing sex
  <date>        <dbl> <fct>       <fct>        <dbl>          <dbl> <fct>
1 2018-06-08      159 Cable Gate Cable Gate 1     451373        4432963 male
2 2018-06-08      159 Cable Gate Cable Gate 2     451411        4432985 male
3 2018-06-08      159 Cable Gate Cable Gate 3     451462        4432991 male
4 2018-06-13      164 West Knoll West Knoll 3    449317        4434093 male
5 2018-06-13      164 West Knoll West Knoll 4    449342        4434141 male
6 2018-06-13      164 West Knoll West Knoll 5    449323        4434273 <NA>
# i 2 more variables: concentration_pg_g <dbl>, elev_m <dbl>

```

Question: How does the stress-gradient hypothesis apply to pikas? More clearly, how does stress in pikas vary as a function of elevation?

Q2.2 Make a hypothesis

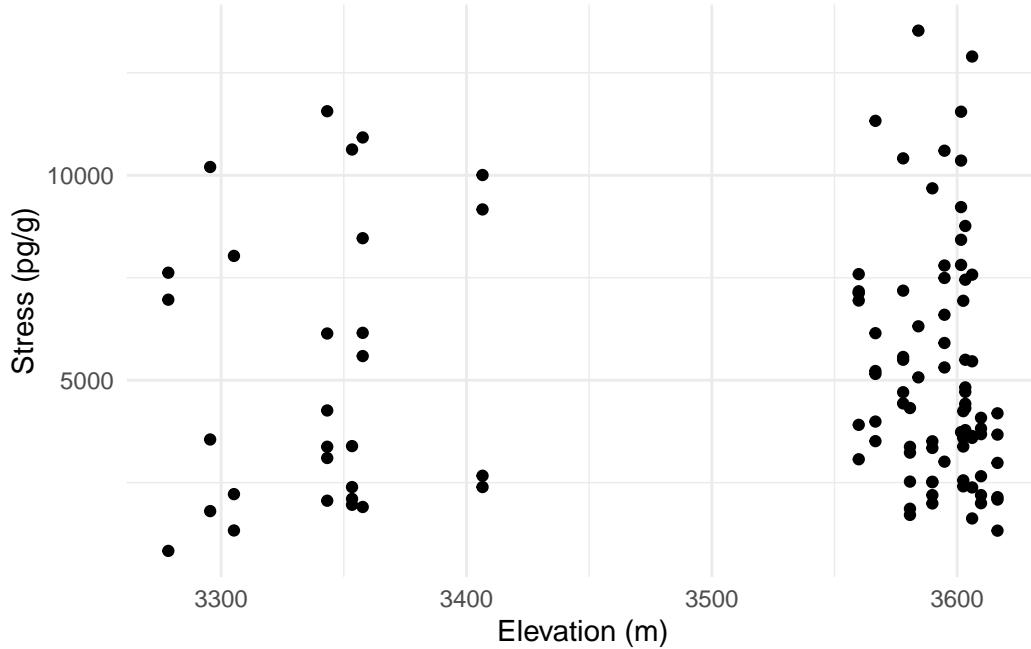
Because (we assume) there are less predators at higher elevations, pikas at high elevation tend to be less stressed than those at low elevations. Therefore, a larger value of elevation should show lower values of stress.

Q2.3 Graph the data

```

ggplot(nwt_pikas_doy, aes(x= elev_m, y = concentration_pg_g))+
  geom_point()+
  labs(x= "Elevation (m)",
       y= "Stress (pg/g)")+
  theme_minimal()

```



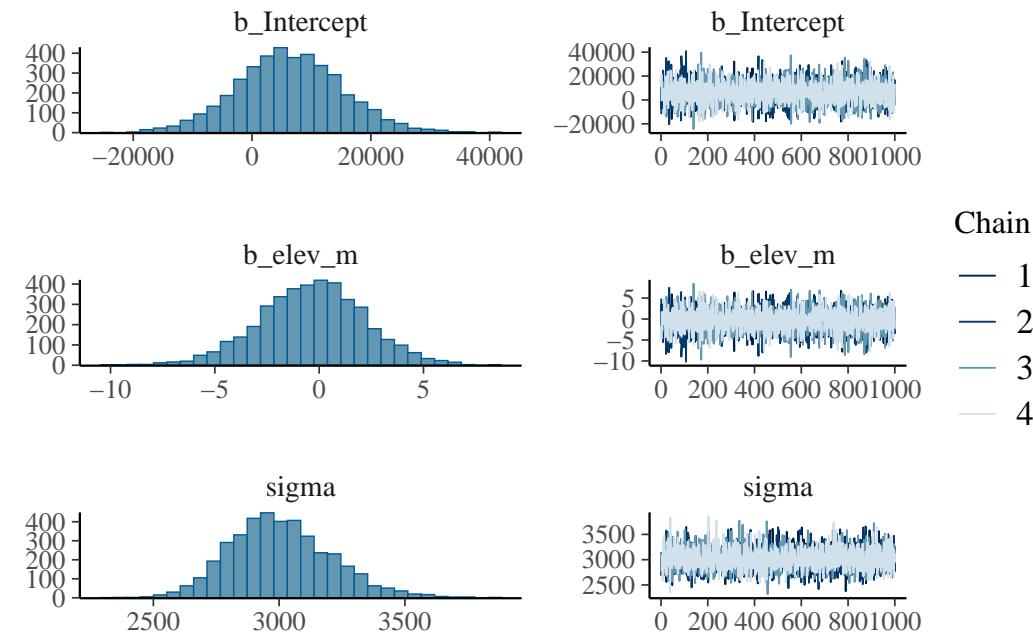
Q2.4 Set up and run a model

```
m.pika_elev <-
  brm(data = nwt_pikas_doy, # Give the model the pie_crab data
       # Choose a gaussian (normal) distribution
       family = gaussian,
       # Specify the model here.
       concentration_pg_g ~ elev_m,
       # 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 faster
       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.pika_elev")
```

Q2.5 Assess the model

Answer: The model ran correctly as we have an Rhat of 1, the chains overlap, and there is a clear peak in the posterior distributions.

```
plot(m.pika_elev)
```



```
summary(m.pika_elev)
```

```
Family: gaussian
Links: mu = identity
Formula: concentration_pg_g ~ elev_m
Data: nwt_pikas_doy (Number of observations: 109)
Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
       total post-warmup draws = 4000

Regression Coefficients:
Estimate  Est.Error  1-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept  6363.53   8891.62 -10910.37 24425.53 1.00    3877    2576
elev_m     -0.34      2.52     -5.45    4.52 1.00    3871    2354
```

Further Distributional Parameters:

```
Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma  3007.30    202.67  2652.16  3446.91  1.00      4082     2528
```

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).

Q2.6 Interpret the model

Interpret your model by answering:

1. What is the effect of your predictor? Remember to describe the effect using the units to make it biologically meaningful.
2. Is the effect reasonably different from zero? How do you know?

Answer: pg/g in the pika poop decreased by 0.34 for every increase 1 meter in elevation. The effect is not reasonably different from 0 because our credible interval includes 0.

Q2.7 Plot the model on the data

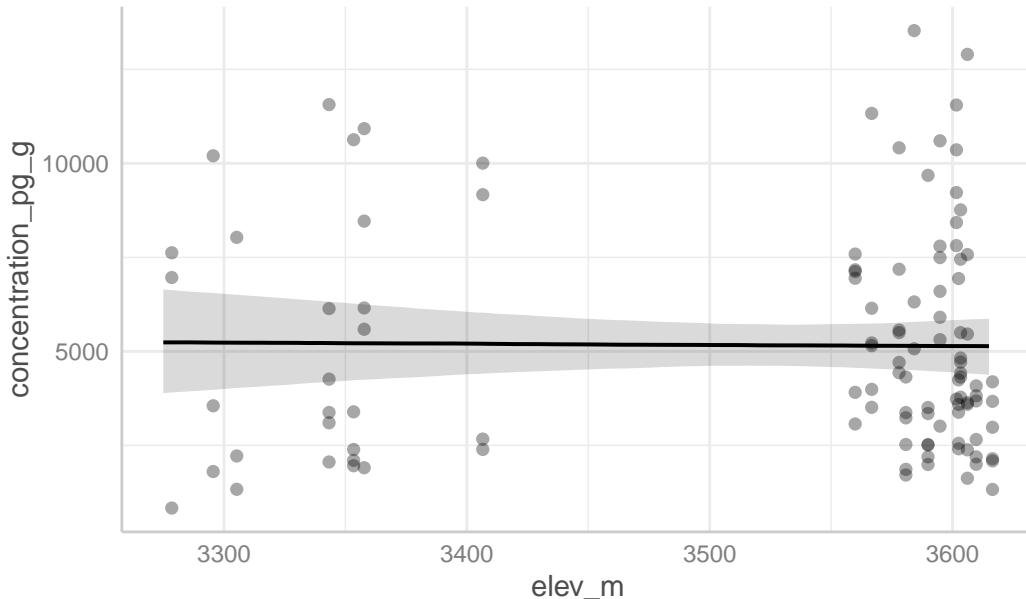
```
# compatibility interval. the shows uncertainty in the average response.
confm.pika.elev <- predict_response(m.pika_elev)
```

```
Warning in check_dep_version(dep_pkg = "TMB"): package version mismatch:
glmmTMB was built with TMB package version 1.9.18
Current TMB package version is 1.9.16
Please re-install glmmTMB from source or restore original 'TMB' package (see '?reinstalling'
```

```
plot(confm.pika.elev, show_data = TRUE)
```

Data points may overlap. Use the `jitter` argument to add some amount of random variation to the location of data points and avoid overplotting.

Predicted values of concentration_pg_g



```
?nwt_pikas
```

```
starting httpd help server ...
```

```
done
```

Q2.8 Write a small results paragraph

Answer: A linear regression showed little to no effect of elevation on pika stress. Glucocorticoid metabolite (GCM) concentration in the pika poop decreased by 0.34 pg/g for every 1 meter in elevation. However, the effect is not reasonably different from 0 because our credible interval includes 0 (95%CI 5.45 - 4.52).