

statisticalreasoning_v2

Yeonu&Sam

```
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)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
v dplyr      1.1.4      v readr      2.1.6
v forcats    1.0.1      v stringr    1.6.0
v ggplot2    4.0.1      v tibble     3.3.1
v lubridate  1.9.4      v tidyr      1.3.2
v purrr      1.2.0
```

```
-- 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(ggeffects) # for the prediction plot
library(lterdatasampler) # for built-in datasets
```

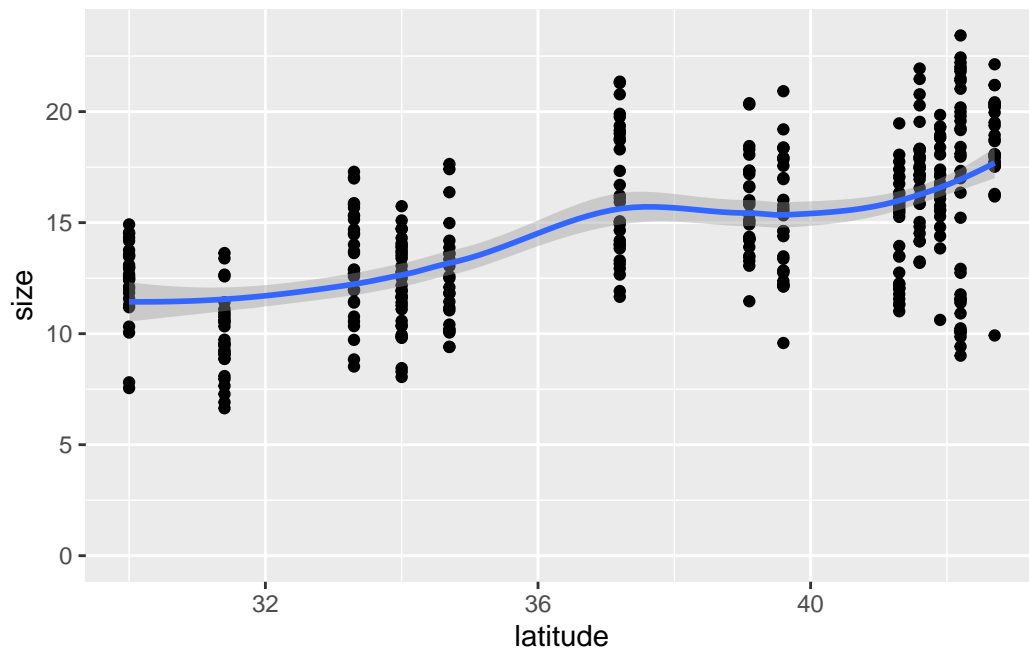
Q1.1

It looks like size increases with latitude. I'm not confident with it.

Q1.2

```
pie_crab %>%
  ggplot(aes(x = latitude, y = size)) +
  geom_point() +
  # Make the y-axis include 0
  ylim(0, NA)+
  geom_smooth()
```

`geom_smooth()` using method = 'loess' and formula = 'y ~ x'



Q1.3

iter=

Number of total iterations per chain (including warmup; defaults to 2000).

```
?brm
```

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

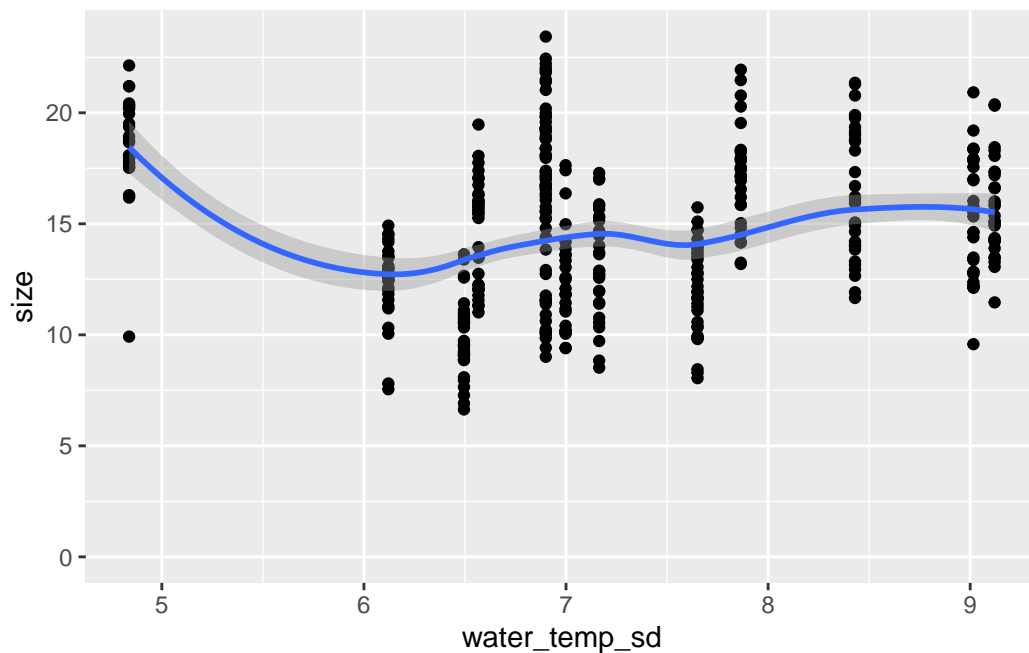
Q1.4

I think higher variability would be associated with larger crabs.

Q1.5

```
pie_crab %>%  
  ggplot(aes(x = water_temp_sd, y = size)) +  
  geom_point() +  
  # Make the y-axis include 0  
  ylim(0, NA)+  
  geom_smooth()
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



Q1.6

It looks like size doesn't change with the sd of water temperature. I'm decently confident with it.

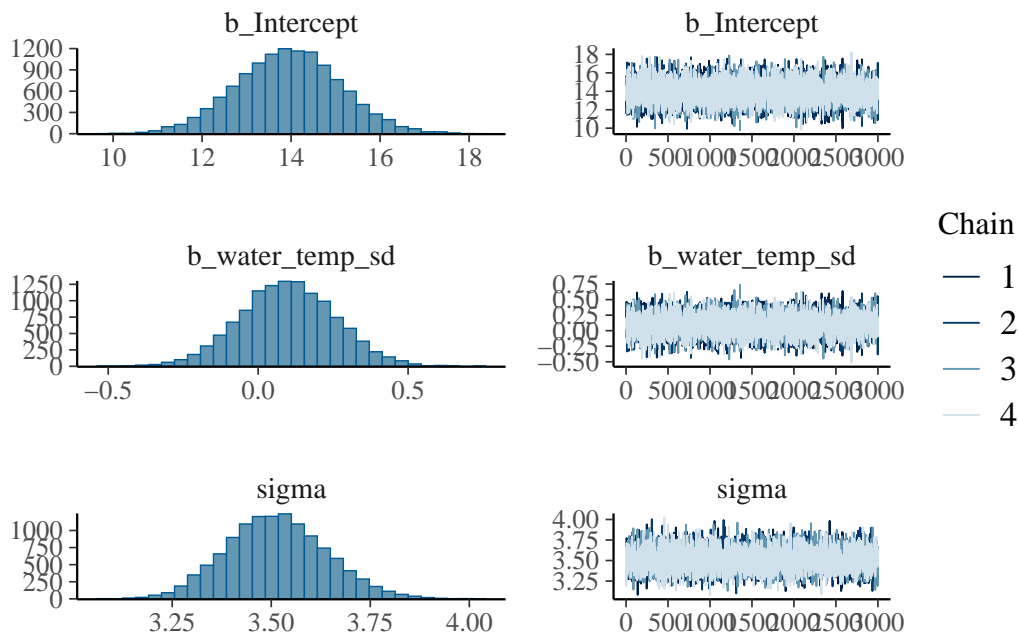
Q1.7

```
# water temp sd 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 f
    iter = 4000, 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")
```

Q1.8

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



```
# show summary, including rhat
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 = 4000; warmup = 1000; thin = 1;
       total post-warmup draws = 12000
```

Regression Coefficients:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	13.95	1.16	11.66	16.24	1.00	12105	8951
water_temp_sd	0.10	0.16	-0.21	0.41	1.00	12186	9187

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	12527	8641

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).

The model ran correctly. Because, posterior samples have smooth distribution with one clean peak and four chains on the overlapping each other and flat.

Q1.9

```
?pie_crab
```

Estimate for water temp sd is 0.1. Every 1 degree water temperature sd, carapace width increases by 0.1mm.

The slope estimate ranges from -0.21 to 0.41 which include 0. Therefore, the effect isn't reasonably different from zero. Because a slope of zero is compatible with the data, we can say that predictor has any effect on our response.

Back to pikas!

```
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>
```

Q2.1

Does day of year correlated with stress?

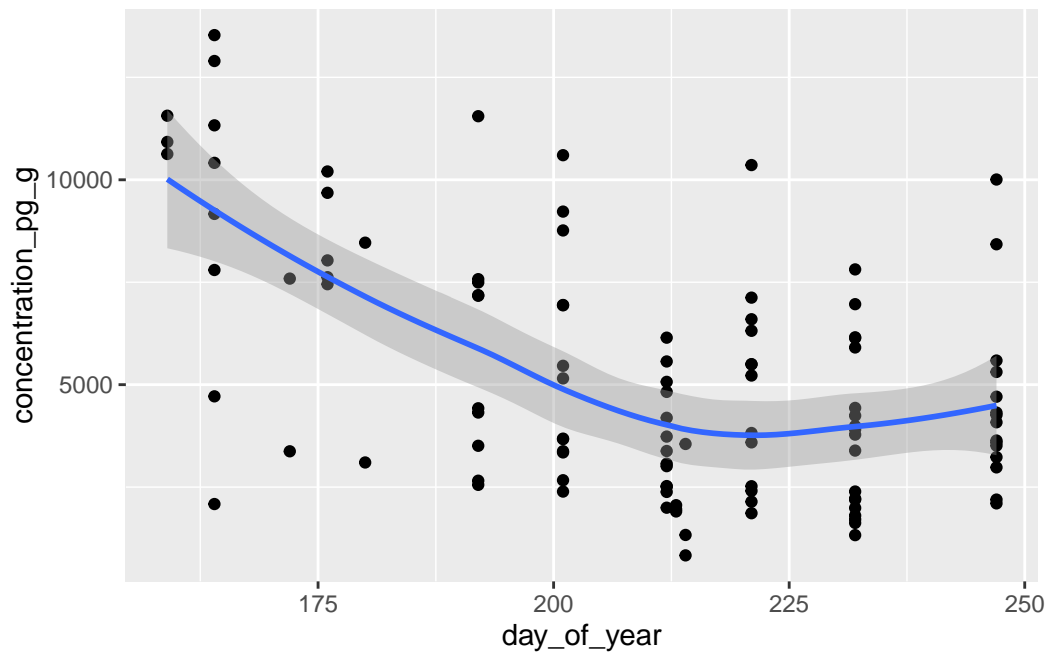
Q2.2

Later of year would be more stressful. Larger value of the predictor with more stressed pika.

Q2.3

```
nwt_pikas_doy %>%
  ggplot(aes(x = day_of_year, y = concentration_pg_g )) +
  geom_point() +
  geom_smooth()
```

``geom_smooth()`` using method = 'loess' and formula = 'y ~ x'

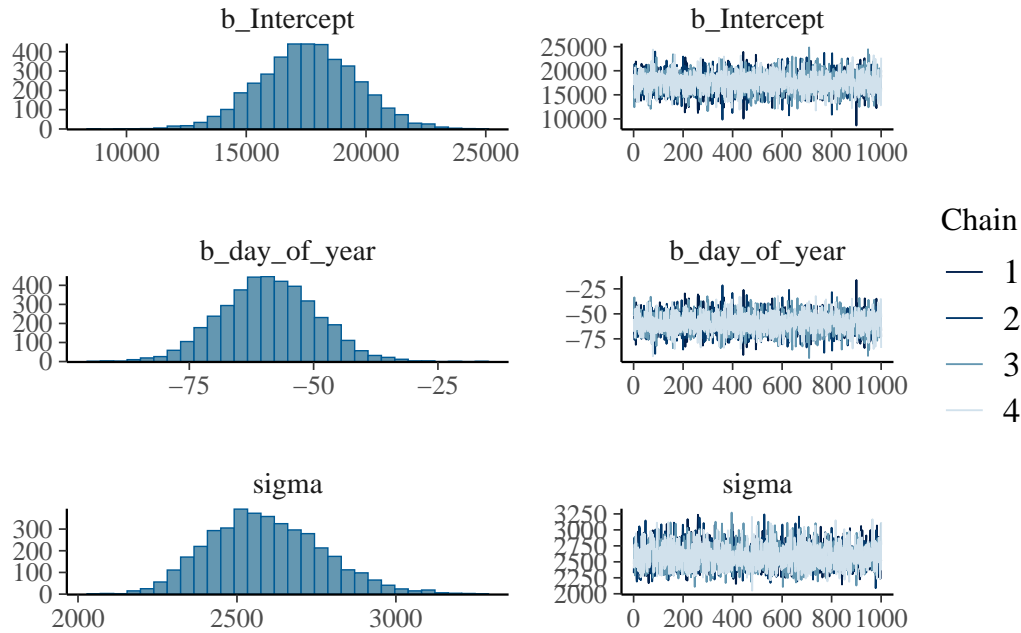


Q2.4

```
m.pikas.yod <-  
  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 ~ day_of_year,  
    # 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.pikas.yod")
```


Q2.5

```
# show posteriors and chains
plot(m.pikas.yod)
```



```
# show summary, including rhat
summary(m.pikas.yod)
```

```
Family: gaussian
Links: mu = identity
Formula: concentration_pg_g ~ day_of_year
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	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	17586.49	2030.08	13617.13	21521.81	1.00	4116	2851
day_of_year	-59.06	9.55	-77.57	-40.48	1.00	4153	2866

Further Distributional Parameters:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	2591.46	178.65	2279.77	2971.27	1.00	4248	3034

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).

Rhat is 1.0 which is good! Posterior samples have smooth distribution with one clean peak and four chains on the overlapping each other and flat which is good, too!

Q2.6

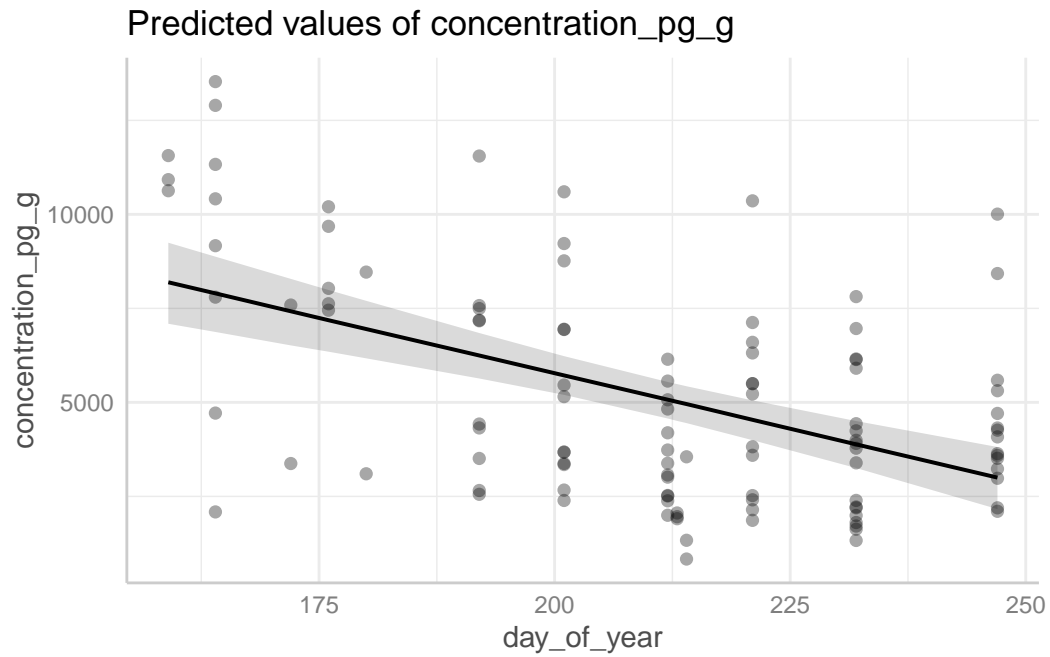
Estimate for year of day is -59.06. Every 1 year of day, GCM concentration in dry pika feces decreases by 59.06 concentration in picogram.

The slope estimate ranges from -77.57 to -40.48 which not include 0. Therefore, the effect is reasonably different from zero. We can say that predictor has effect on our response.

Q2.7

```
# compatibility interval. the shows uncertainty in the average response.
confm.pikas.yod <- predict_response(m.pikas.yod)
plot(confm.pikas.yod, 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.



Q2.8

In conclusion we found that for every day of the year, the concentration of glucocorticoid metabolite decreased by 59.06 pg/g. Meaning as the summer progresses toward fall and winter, pika stress decreases.