CGS698C Assignment 7

Nikhil Pothuganti

2025-04-03

Load libraries

```
library(brms)
```

```
## Loading required package: Rcpp
```

```
## Loading 'brms' package (version 2.22.0). 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 ──
## ✔ dplyr     1.1.4     ✔ readr     2.1.5
## ✔ forcats   1.0.0     ✔ stringr   1.5.1
## ✔ ggplot2   3.5.1     ✔ tibble    3.2.1
## ✔ lubridate 1.9.4     ✔ tidyr     1.3.1
## ✔ purrr     1.0.4
```

```
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflic
ts to become errors
```

```
library(bayesplot)
```

```
## This is bayesplot version 1.11.1
## - Online documentation and vignettes at mc-stan.org/bayesplot
## - bayesplot theme set to bayesplot::theme_default()
##    * Does _not_ affect other ggplot2 plots
##    * See ?bayesplot_theme_set for details on theme setting
##
## Attaching package: 'bayesplot'
##
## The following object is masked from 'package:brms':
##
##     rhat
```

```
library(truncnorm)
```

Load and preprocess data

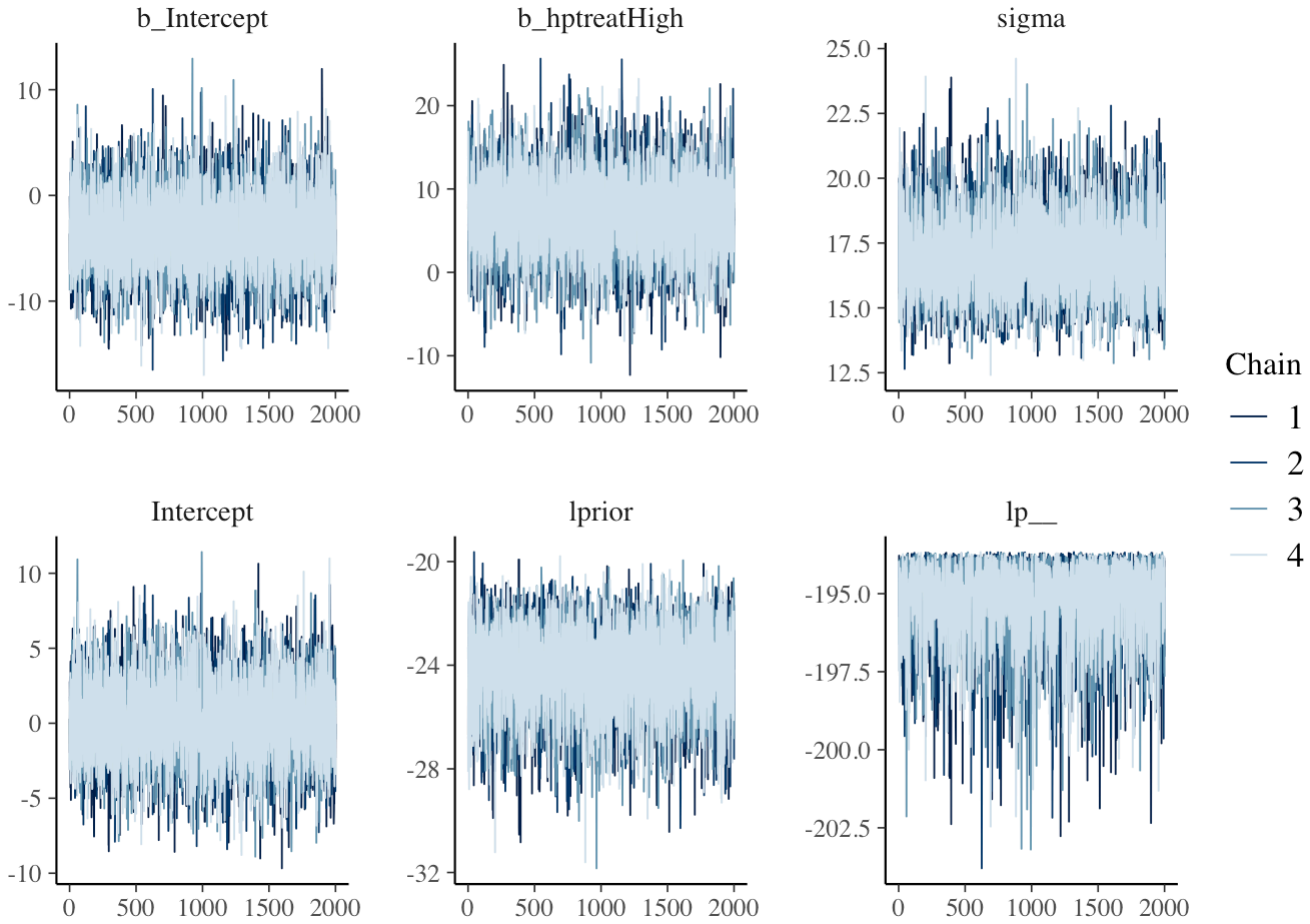
```
df_powerpose <- read.csv("df_powerpose.csv") %>%
  mutate(
    hptreat = factor(hptreat, levels = c("Low", "High")),
    test_diff = testm2 - testm1
  )
```

Bayesian linear regression model

```
fit_powerpose <- brm(
  test_diff ~ hptreat,
  data = df_powerpose,
  family = gaussian(),
  prior = c(
    prior(normal(0, 15), class = "Intercept"),
    prior(normal(0, 10), class = "b"),
    prior(exponential(1), class = "sigma")
  ),
  chains = 4,
  iter = 4000,
  warmup = 2000,
  file = "powerpose_model"
)
```

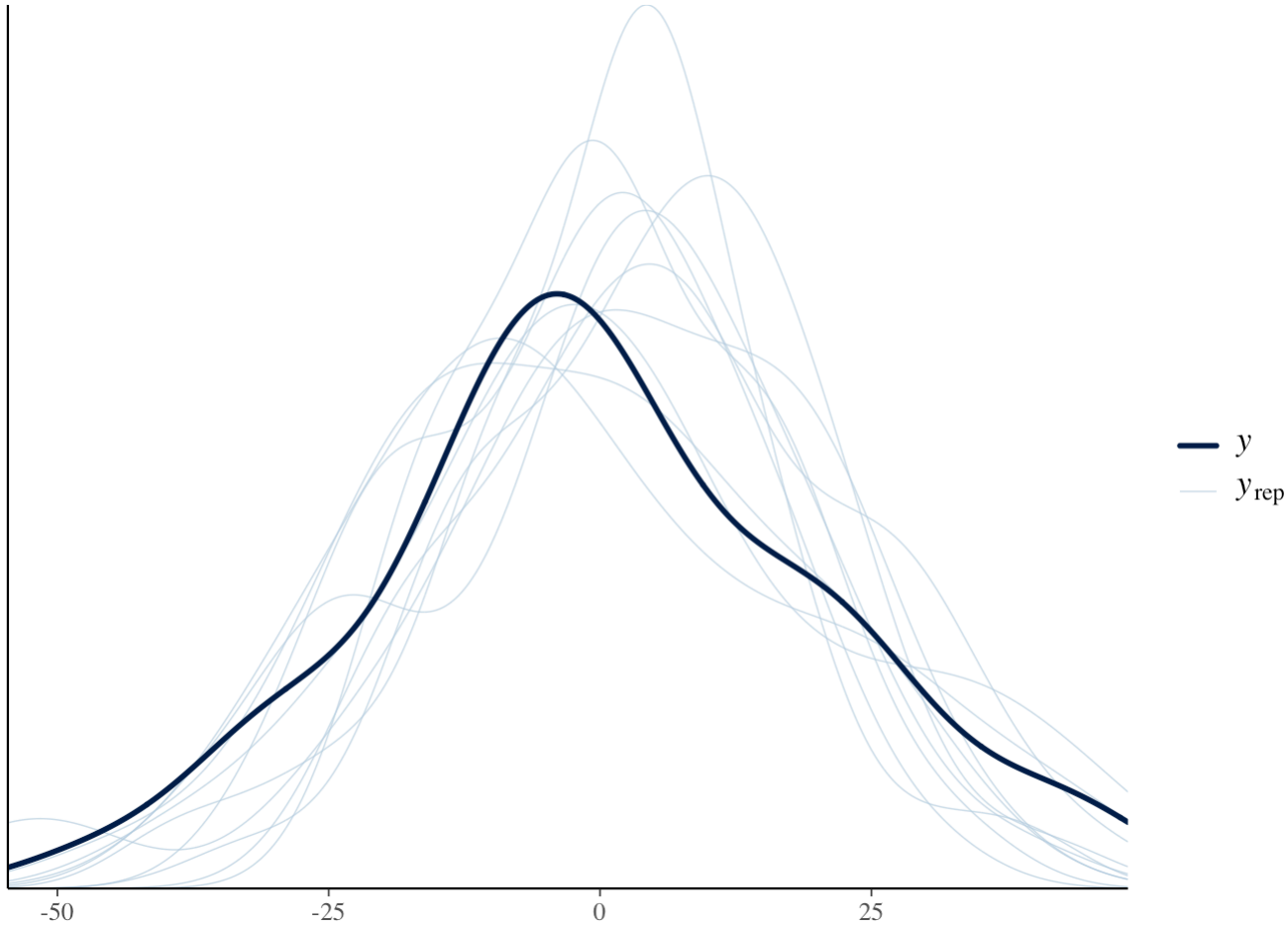
Model diagnostics

```
mcmc_trace(fit_powerpose)
```




```
pp_check(fit_powerpose)
```

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



Results interpretation

```
posterior_summary(fit_powerpose)
```

```
##                   Estimate Est.Error        Q2.5       Q97.5
## b_Intercept     -3.3299163  3.653682  -10.543743    3.806433
## b_hptreatHigh    6.9226464  4.837770   -2.612822   16.497422
## sigma           17.0109824  1.527665   14.315399   20.261977
## Intercept        0.2201588  2.715742   -4.973678    5.528703
## lprior         -24.2326105  1.585233  -27.675733  -21.425456
## lp__          -195.1902545  1.254080 -198.458743 -193.780216
```

```
mean(posterior_samples(fit_powerpose)$b_hptreatHigh > 0)
```

```
## Warning: Method 'posterior_samples' is deprecated. Please see ?as_draws for
## recommended alternatives.
```

```
## [1] 0.92825
```

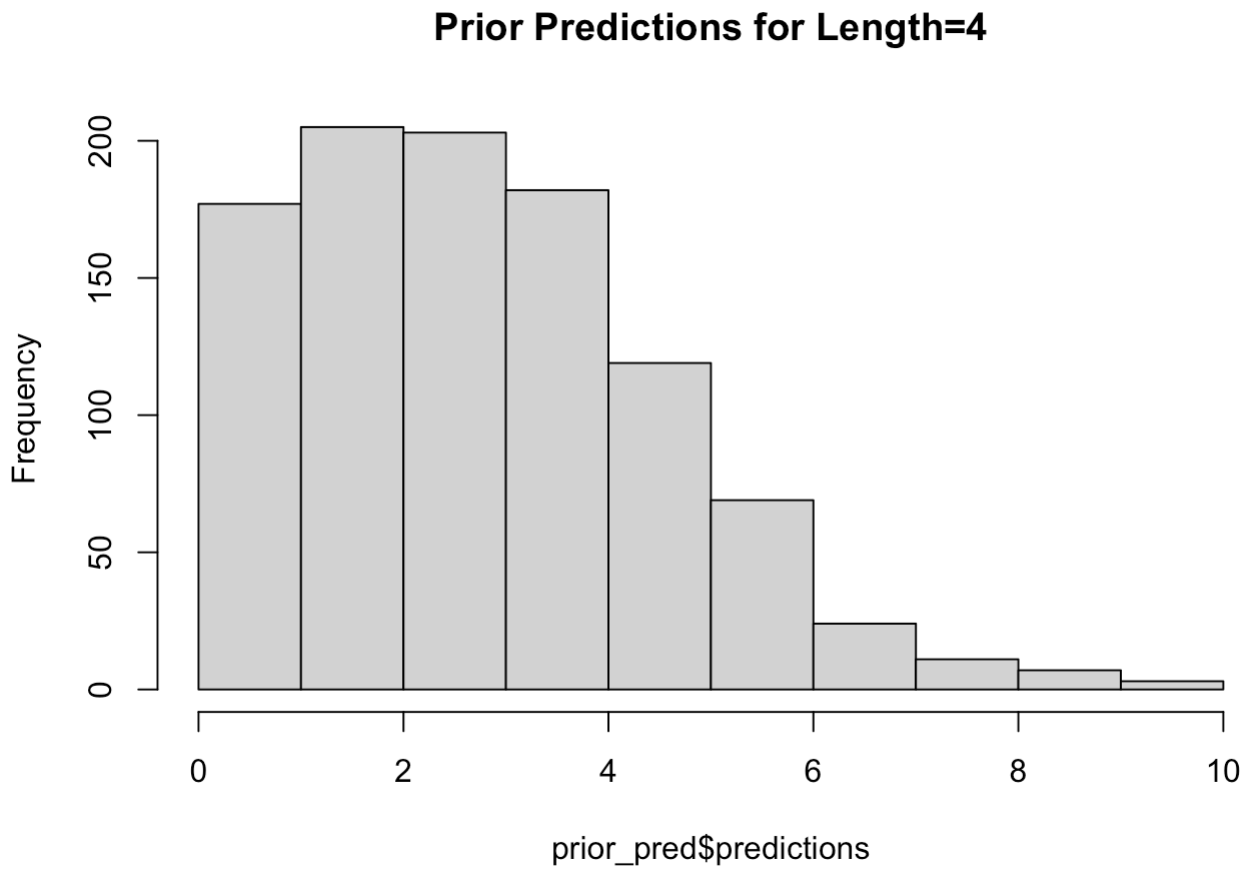
Exercise 2.1: Model implementation

```
generate_crossings <- function(sentence_length, alpha, beta) {
  lambda <- exp(alpha + beta * sentence_length)
  rpois(n = length(sentence_length), lambda = lambda)
}
```

Exercise 2.2: Prior predictions

```
set.seed(123)
prior_pred <- tibble(
  alpha = rtruncnorm(1000, a = 0, mean = 0.15, sd = 0.1),
  beta = rtruncnorm(1000, a = 0, mean = 0.25, sd = 0.05)
) %>%
  mutate(predictions = map2(alpha, beta, ~ generate_crossings(4, .x, .y))) %>%
  unnest(predictions)

hist(prior_pred$predictions, main = "Prior Predictions for Length=4")
```



Exercise 2.3: Model fitting

```
df_crossings <- read.csv("crossings.csv") %>%
  mutate(
    s.length_c = s.length - mean(s.length),
    lang = ifelse(Language == "German", 1, 0)
  )
```

Model M1 (Pooled)

```
m1 <- brm(
  nCross ~ s.length_c,
  data = df_crossings,
  family = poisson(),
  prior = c(
    prior(normal(0.15, 0.1), class = "Intercept"),
    prior(normal(0, 0.15), class = "b")
  ),
  chains = 4,
  iter = 4000,
  file = "model_m1"
)
```

Model M2 (Interaction)

```
m2 <- brm(
  nCross ~ s.length_c * lang,
  data = df_crossings,
  family = poisson(),
  prior = c(
    prior(normal(0.15, 0.1), class = "Intercept"),
    prior(normal(0, 0.15), class = "b")
  ),
  chains = 4,
  iter = 4000,
  file = "model_m2"
)
```

Exercise 2.4: Model comparison

K-fold cross-validation implementation

```
observed <- df_crossings %>%
  mutate(s.length = s.length_c + mean(df_crossings$s.length))

lpds.m1 <- c()
lpds.m2 <- c()
untested <- observed

for(k in 1:5){
  ytest <- sample_n(untested, size = nrow(observed)/5)
  ytrain <- setdiff(observed, ytest)
  untested <- setdiff(untested, ytest)

  fit.m1 <- brm(
    nCross ~ s.length_c,
    data = ytrain,
    family = poisson(),
    prior = c(
      prior(normal(0.15, 0.1), class = "Intercept"),
      prior(normal(0, 0.15), class = "b")
    ),
    cores = 4
  )

  fit.m2 <- brm(
    nCross ~ s.length_c * lang,
    data = ytrain,
    family = poisson(),
    prior = c(
      prior(normal(0.15, 0.1), class = "Intercept"),
      prior(normal(0, 0.15), class = "b")
    ),
    cores = 4
  )

  post.m1 <- posterior_samples(fit.m1)
  post.m2 <- posterior_samples(fit.m2)

  lppd.m1 <- 0
  lppd.m2 <- 0
  for(i in 1:nrow(ytest)){
    lpd_im1 <- log(mean(dpois(ytest$nCross[i],
                              lambda = exp(post.m1[,1] + post.m1[,2]*ytest$s.length_c
[i]))))
    lppd.m1 <- lppd.m1 + lpd_im1

    lpd_im2 <- log(mean(dpois(ytest$nCross[i],
                              lambda = exp(post.m2[,1] + post.m2[,2]*ytest$s.length_c
[i] +
                                           post.m2[,3]*ytest$lang[i] +
                                           post.m2[,4]*ytest$s.length_c[i]*ytest$lang
```

```
[i]))))
    lppd.m2 <- lppd.m2 + lpd_im2
  }
  lpds.m1 <- c(lpds.m1, lppd.m1)
  lpds.m2 <- c(lpds.m2, lppd.m2)
}
```

```
## Compiling Stan program...
```

```
## Start sampling
```

```
## Compiling Stan program...
```

```
## Start sampling
```

```
## Warning: Method 'posterior_samples' is deprecated. Please see ?as_draws for
## recommended alternatives.
## Warning: Method 'posterior_samples' is deprecated. Please see ?as_draws for
## recommended alternatives.
```

```
## Compiling Stan program...
## Start sampling
```

```
## Compiling Stan program...
```

```
## Start sampling
```

```
## Warning: Method 'posterior_samples' is deprecated. Please see ?as_draws for
## recommended alternatives.
## Warning: Method 'posterior_samples' is deprecated. Please see ?as_draws for
## recommended alternatives.
```

```
## Compiling Stan program...
## Start sampling
```

```
## Compiling Stan program...
```

```
## Start sampling
```

```
## Warning: Method 'posterior_samples' is deprecated. Please see ?as_draws for
## recommended alternatives.
## Warning: Method 'posterior_samples' is deprecated. Please see ?as_draws for
## recommended alternatives.
```

```
## Compiling Stan program...
## Start sampling
```

```
## Compiling Stan program...
```

```
## Start sampling
```

```
## Warning: Method 'posterior_samples' is deprecated. Please see ?as_draws for
## recommended alternatives.
## Warning: Method 'posterior_samples' is deprecated. Please see ?as_draws for
## recommended alternatives.
```

```
## Compiling Stan program...
## Start sampling
```

```
## Compiling Stan program...
```

```
## Start sampling
```

```
## Warning: Method 'posterior_samples' is deprecated. Please see ?as_draws for
## recommended alternatives.
## Warning: Method 'posterior_samples' is deprecated. Please see ?as_draws for
## recommended alternatives.
```

```
elpd.m1 <- sum(lpds.m1)
elpd.m2 <- sum(lpds.m2)
difference_elpd <- elpd.m2 - elpd.m1
```

Print results

```
cat("ELPD for M1:", elpd.m1, "\n")
```

```
## ELPD for M1: -2814.493
```

```
cat("ELPD for M2:", elpd.m2, "\n")
```

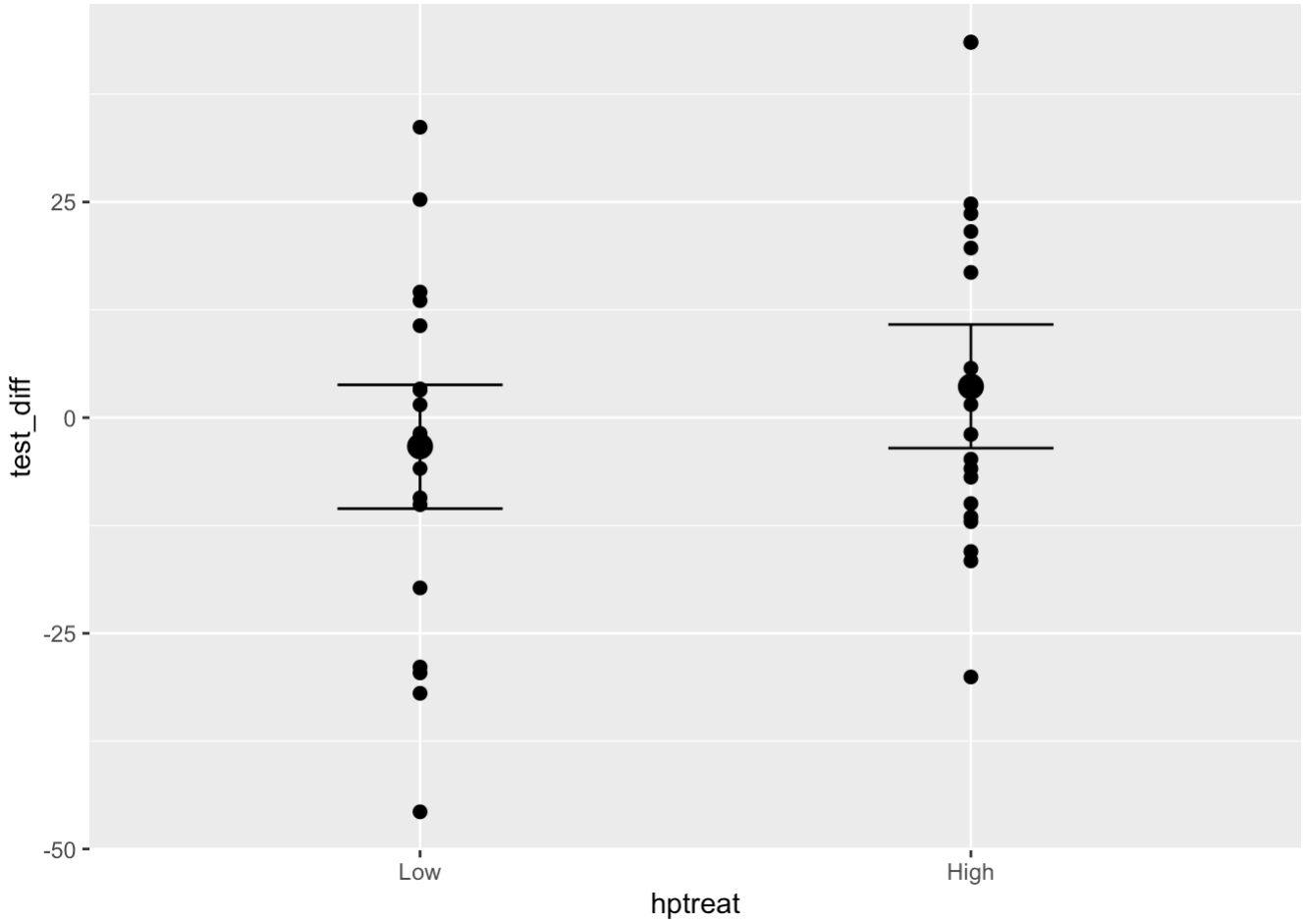
```
## ELPD for M2: -2682.273
```

```
cat("ELPD difference (M2 - M1):", difference_elpd, "\n")
```

```
## ELPD difference (M2 - M1): 132.22
```

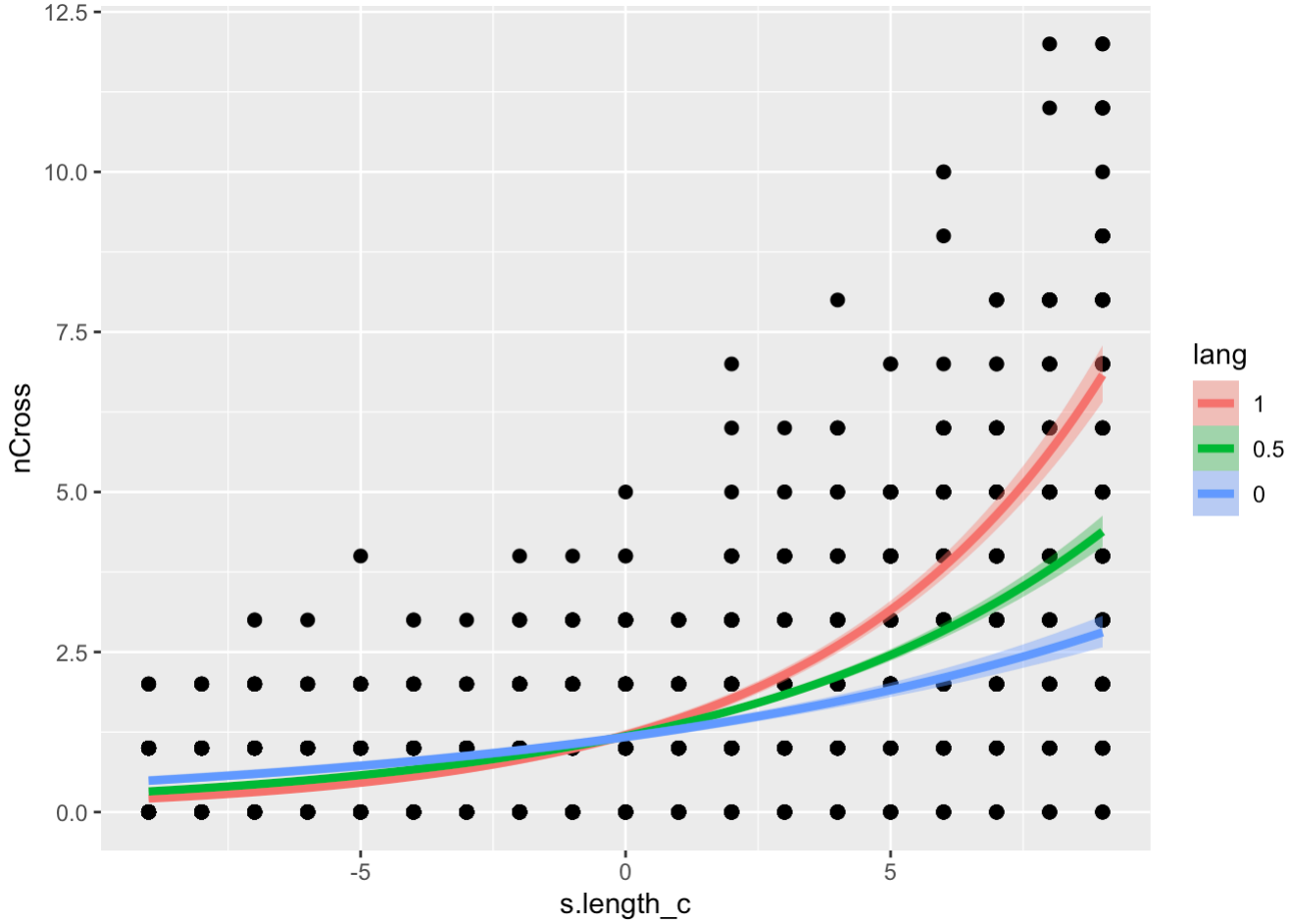
Power pose results

```
plot(conditional_effects(fit_powerpose), points = TRUE)
```



Crossing dependencies effects

```
conditional_effects(m2, effects = "s.length_c:lang") %>%
  plot(points = TRUE, line_args = list(size = 1.5))
```


12.5
10.0
7.5
5.0
2.5
0.0
nCross
-5
0
5
s.length_c
lang
1
0.5
0