

Assignment 4

Julia Haaf & Nicole Cruz

2025-08-26

Packages

```
library(tidyverse)    # Graphs with ggplot()
library(knitr)         # Tables with kable()
library(ggokabeito)   # colorblind friendly colors
library(viridis)      # colorblind friendly colors
library(brms)         # Hierarchical Bayesian models
library(rstan)        # Interface to Stan
library(shinystan)    # Model output graphs
library(bayesplot)    # Model output graphs

# General settings
options(mc.cores = parallel::detectCores())
rstan_options(auto_write = TRUE)
options(contrasts = c("contr.equalprior_pairs", "contr.poly"))
set.seed(1234) # For reproducibility
```

Preparation

Data: penguins

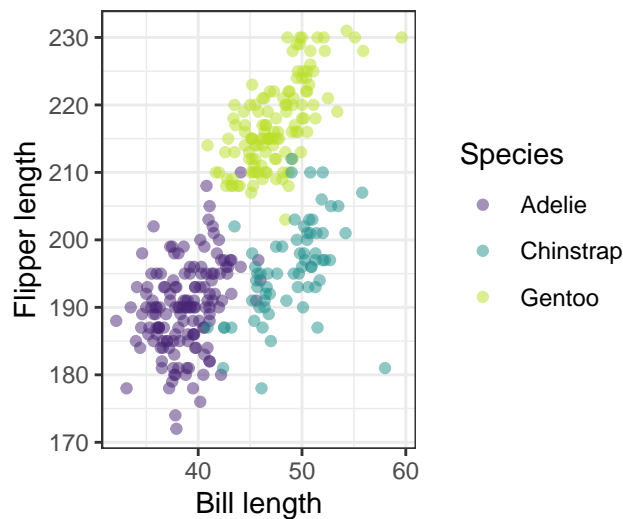
```
head(penguins)
```

```
##   species   island bill_len bill_dep flipper_len body_mass  sex year
## 1  Adelie Torgersen   39.1    18.7        181      3750  male 2007
## 2  Adelie Torgersen   39.5    17.4        186      3800 female 2007
## 3  Adelie Torgersen   40.3    18.0        195      3250 female 2007
## 4  Adelie Torgersen    NA      NA         NA         NA   <NA> 2007
## 5  Adelie Torgersen   36.7    19.3        193      3450 female 2007
## 6  Adelie Torgersen   39.3    20.6        190      3650  male 2007
```

The data consist of several measurements of penguins of the Palmer archipel in Antarctica.

Variable	Description
species	Factor: Adelie, Chinstrap, Gentoo
island	Factor: Biscoe, Dream, Torgersen
bill_len	Numeric: Mean = 43.91, range [32.10, 59.6]
bill_dep	Numeric: Mean = 17.15, range [13.10, 21.5]
flipper_len	Numeric: Mean = 200.9, range [172, 231]
body_mass	Numeric: Mean = 4202, range [2700, 6300]
sex	Factor: female, male
year	Numeric: 2007, 2008, 2009

```
g1 <- ggplot(penguins, aes(x = bill_len, y = flipper_len, color = species))+
  geom_point(alpha = .5)+
  scale_color_viridis_d(begin = .1, end = .9)+
  labs(x = "Bill length", y = "Flipper length", color = "Species")+
  theme_bw()
g1
```



Exercise 1

Analyse the penguins dataset to investigate the hypothesis that flipper length is influenced by bill length and / or bill depth. Do so following a series of steps in the Bayesian workflow:

1. Translate the hypothesis into a Bayesian mathematical model. What is the general structure of this model?
2. Define prior distributions for the model.
3. Conduct prior predictive checks.
4. Apply the model to the data with `brm()`. What convergence parameters would you look at? Briefly interpret the output.
5. Plot the posterior densities for the parameters and compute their credible intervals.
6. Define two alternative models and compare them to your initial model via Bayes factors. Briefly interpret the results.
7. Conduct posterior predictive checks.
8. Consider the evidence for your model as a whole after applying it to the data. What implications do the outcomes have for the hypothesis, if any? What can we learn from the analysis results?