Week 12 Activity: Regression

On Thursday we will use historical data to predict NCAA basketball games. For this activity we will generate (simulate) synthetic data.

1.

Simulate synthetic data that represents historical NCAA games. In particular, let's consider the model

$$pointdiff = \beta_0 + \beta_1 x_{seeddiff} + \epsilon; \epsilon \sim N(0, \sigma^2)$$

where:

- $\beta_0 = 0$
- $\beta_1 = 2$
- $\sigma = 12$

```
set.seed(04032023)
num_games <- 960 # approximately 15 years
seeddiff <- rep(0:15, each = 60)


beta1 <- 2
sigma <- 12
```

2.

Interpret the three coefficients in the model specified in part 1.

- $\beta_0 =$
- $\beta_1 =$
- $\sigma =$

3.

Create a visualization of your point differential versus seed differential.

4.

Specify prior distributions on your parameters.

- $\beta_0 \sim$
- $\beta_1 \sim$
- $\sigma \sim$

5.

Write JAGS code to fit this model. Output the results.

6.

Summarize your findings. Write a paragraph so that a college basketball coach could understand.

7.

Now we will construct a posterior predictive distribution. This will enable us answers questions like "What is the probability that a one seed is upset by a 16 seed (15 point difference in seeds.) In particular, construct a posterior predictive distribution, conditional on the following scenarios:

- Seed difference = 15
- Seed difference = 7 (most commonly 1 vs. 8)
- Seed difference = 1

Use those distributions to compute the probability of an upset occuring.