

# Lab 6

## Part 1. Logistic Regression Separation

For this question, consider the dataset constructed with the following code:

```
x <- seq(-3, 3, length.out = 100)
sep_data <- tibble(x = x) |>
  mutate(y = case_when(
    x < 0 ~ 0,
    x >= 0 ~ 1))
```

### 1. Data Vizualization (4 points)

Create a visualization the displays the data - and a smoothed fit.

Comment on this figure.

### 2. Bayes (4 points)

Fit the model using `stan_glm`. Comment on the results and any potential warnings.

### 3. MLE (4 points)

Fit the model using `glm`. Comment on the results and any potential warnings.

### 4. Bayesian Visualization (4 points)

Let's use the Bayesian estimation to plot an estimated (posterior) mean fit line.

## Part 2. Detective Work Part 2

For this question, consider the dataset constructed with the following code:

I've created a synthetic dataset for you to explore. The goal is to try and recover the true model. Note: that in practice we don't know a "true model."

```
set.seed(10012025)
n <- 200
x1 <- seq(-3, 3, length.out = n)
x2 <- rnorm(n)
x3 <- rnorm(n)
x4 <- sample(c('A','B','C'), size = n, replace = T)

beta0 <- .4
beta1 <- .8
beta2 <- -.1
beta3 <- 0
betaB <- 0
betaC <- 1.5
beta1B <- 0
beta1C <- -1

dat_tibble <- tibble(x1, x2, x3, x4)
X_mat <- model.matrix(~ x1 + x2 + x3 + x4 + x1:x4)
beta_vec <- c(beta0, beta1, beta2, beta3, betaB, betaC, beta1B, beta1C)

dim(X_mat)
```

```
[1] 200    8
```

```
pi <- invlogit(X_mat %*% beta_vec)

y <- rbinom(n, 10, pi)
secret_data2 <- tibble(y = y, x1, x2, x3, x4)
```

## 5. Data Visualization (8 points)

Similar to this week's activity, explore how the  $x_1$  variables relate to  $y$ . Note, you should look for interactions between  $x_4$  and the other variables. Create a set of 4 figures and summarize your findings.