Lab 8

Your Name

```
#------------------------------------
#------------------------------------
# Import/Load the rstan library:
library(tidyverse)
library(rstan)
library(loo)
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())
#------------------------------------
#------------------------------------
```

Random effects & Simpson's paradox

In this assignment, we will learn an important lesson about why accounting for group-level differences is essential in your analyses. I have provided you with a data set (`lab8_data.csv`) that has one outcome variable, one input variable, and one group-level identifier. Note that the input variable has been centered and scaled for you.

Your tasks (35 points)

1. Import your data into R.

2. Create two scatterplots (5 points):

(a) A scatterplot with **no** distinction between groups
(b) A scatterplot with data from specific groups that are clearly distinct

3. Fit a linear regression with complete pooling (i.e., no distinction between groups). Write a brief interpretation about the linear relationship between the input and the outcome variable. In other words, interpret the slope parameter. (10 points)

4. Fit a linear regression with partial pooling. Specifically, allow a random intercept per group. Again, write a brief interpretation about this linear regression. How has your conclusion changed compared to the outcome in Task 3? (10 points)

5. Compare your two models using the LOO-IC. Which model provides a more parsimonious explanation of the data? Interpret what this means. (5 points)

6. This assignment has shown an example of Simpson's Paradox. Look up this term and write a brief interpretation of how it applies to this assignment. What have you learned? (5 points)