## Quiz #3
STAA 565: QUANTITATIVE REASONING
problems equally weighted

1. The figure below shows a data set with response y and predictors x and d, where $d_i = 0$ if the ith response is in Group 1 and $d_i = 1$ if the ith response is in Group 2. The data are plotted below. The model is

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 d_i + \beta_3 x_i d_i + \varepsilon_i,$$

for $i = 1, \ldots, n$. The fitted model is shown superimposed on the data. What are the signs (positive or negative) of the estimated coefficients in the fitted model, for these data?
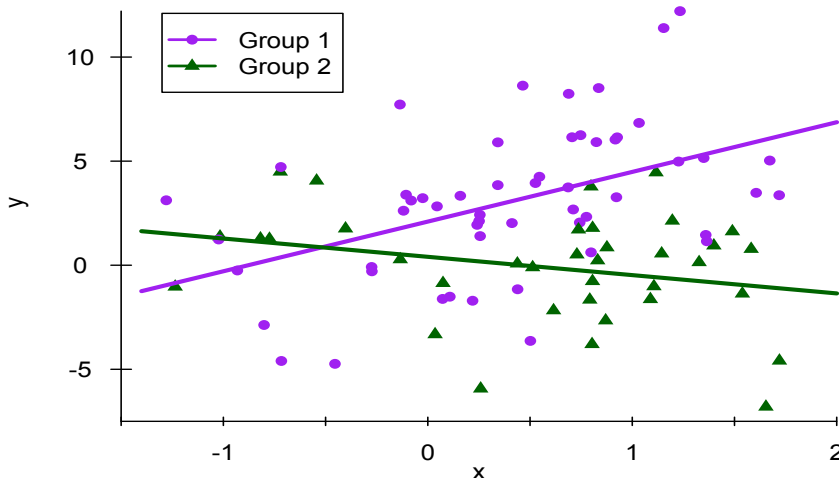
2. An amateur cheese maker is interested in finding out what makes the best cheddar cheese. She attends a cheese exposition, and gets samples of cheddars from thirty artisan cheese makers in a taste competition. For each of the thirty samples, she determines the concentrations of acetic acid and lactic acid. She also has the taste scores from the competition, so she can model the taste score (response) as a function of the two concentrations (predictors).

   (a) Describe how these two predictors might be confounding, and give a specific example of *how* they are confounding.

   (b) Give a scenario in which the two predictors interact, and explain how the interaction occurs.

3. An environmental agency is considering different diesel particulate filters (DPFs) on diesel trucks, to determine their effectiveness in reducing the emission of diesel particulate matter (PM), a known carcinogen. They identify three popular brands: A, B, and C, of the DPFs, and collect data by randomly selecting diesel delivery trucks arriving at the agency for state emissions testing. For each truck, they determine the brand of DPF and the level of PM emissions. They also record the size of the engine in liters. The data are in `diesel1.txt`.

   (a) Ignoring the engine size covariate, determine if there are any significant differences in average emissions across brand of DPF. If 'yes,' describe the differences.

   (b) Now include the covariate in your model. Determine if there are any significant differences in average emissions across brand of DPF. If 'yes,' describe the differences.

   (c) Get a plot of the data and superimpose your fit in part (b).

   (d) What can you tell the agency about the effectiveness of the three brands of DPF in terms of having the smallest PM emissions?

   (e) Explain why it was necessary to put the covariate (size of engine) in the model.

4. This is a repeat of question 3, but use the data in `diesel2.txt`.