# SDS 439 - Homework 05

Due Mar 31, 1:00 pm

## Pinewood Derby Revisited

We return to the pinewood derby data, but this time with more complex models at our disposal.

1. To refamiliarize yourself with the data, make one plot of the data, showing some of the major features. This may be a plot from the previous assignment, but see if you can improve it in some way.
2. The official winner of the competition was determined by which racer had the lowest average time over their six runs. Use both `tapply` and `dplyr`'s `group_by` and `summarize` functions to calculate the average times for each racer.
3. Convert the racer column to a factor, and set its reference level to be the racer with the fastest average time. Then fit the one-factor model in racer and print the results. If you do it correctly, all of the coefficients should be positive. Explain why.
4. Show that the estimated expected time for each racer is equal to the sample averages that you computed in a previous question.
5. Now fit a two-factor model to the times, with racer and lane as the two factors. Keep the fastest racer as the reference level and print the summary. How do you interpret the intercept in this model? Note that the intercept is a slightly different number and has a slightly different interpretation than that in the previous model.
6. Show that the coefficients for the racers in the two-factor model are exactly the same as they were in the one-factor model. This means that, after accounting for lane, the relative differences among the racers is estimated to be the same. This is because the structure of the pinewood derby is set up so that each racer appears in each lane exactly once. If you take an experimental design course, you will learn that this is an example of a balanced experiment with respect to racer and lane. In a balanced experiment, the estimated effects in a two-factor model can be calculated by taking differences of simple averages.
7. Use the `anova` function to perform an F-test for whether `lane` significantly improves the model and interpret the results. The easiest way to do this is to put both the reduced model and the full model into the `anova` function as arguments.
8. Now fit a three-factor additive model with racer, lane, and heat as the three factors and show that the coefficients for the racers change. Why do you think this happens?