IDS 702: Modeling and Representation of Data
Methods and Data Analysis 2

This assignment involves multiple regression. The data can be found in the Assignments folder on Sakai. Useful R commands are in the R scripts folder on Sakai.

Due Date: Thursday, September 20.

1 Methods and Data Analysis

1. **Brain Weights** (Problem 12). For part a, just turn in the plot. For part b, just report the output from the regresssion model described in that part (include a table with coefficients and SEs, their associated confidence intervals, and somewhere in the text or table the estimated regression standard deviation and R^2). In part c, answer the questions, but you don't need to turn in the plot.

2. **Brain Weights** additional problems D, E, and F.

   - PART D: Fit the regression model with log(brain size) on log(body weight), log(gestation) and litter size on its natural scale. Report the output of the model (include a table with coefficients and SEs, their associated confidence intervals, and somewhere in the text or table the estimated regression standard deviation and R^2).
   - PART E: Provide an interpretation of each of the coefficients from the regression in Part D on the natural scale of brain weight and each predictor. Include 95% confidence intervals in your interpretations.
   - PART F: Based on the quality of the residual plots and the value of R^2, which model do you prefer: the one in Part B or the one in Part D? If you decide that you are ambivalent between the two models based on these measures of model fit, explain which one you would prefer for interpretability.

3. **Kentucky Derby** (Problem 20). See instructions below

4. **Old Faithful** (Problem 15). Report the value of the F statistic, the p-value, and your conclusion.

5. **Wages and Race** (Problem 29). See instructions below

Instructions for Problem 20 and Problem 29: Submit a data analysis write-up for your answers. Be sure to include the following in your story: the model you ultimately decided to use, a justification for that model (e.g., why you chose certain transformations and why you decided the final model is reasonable to use based on residual diagnostics), the relevant regression output (includes: a table with coefficients and SEs, and p-values or confidence intervals; and somewhere in the text or table the estimated regression standard deviation and R^2), your interpretation of the results in the context of the questions of interest, and any potential limitations of the analysis. Make sure that you explain any interaction effects or quadratic effects included in the model.