

**INSTRUCTIONS:**

- A. *For each question*, include a:
  - I. summary output table and/or graphs, as appropriate
  - II. narrative answer that directly and clearly answers the question, based on the results.
- B. Provide your code in an Appendix, organized so that we can relate it to questions.
- C. Submit a pdf (with your name in the file name).

**Background**

We return to baseball data we analyzed earlier (on the course web site; [MLBbatting2022](#)), where each player is listed for their team, position, and data on:

G: games played	AB: at bats	R: runs scored	H: hits
AVG: batting average	HR: home runs	RBI: runs batted in	OBP: on base percentage

Last time we worked with these data, you may have noticed that many data (except for OBP and AVG) are counts, which *are notoriously non-normal*. Now you use GLMs to tackle that problem. Also, you must still ensure your analyses and answers below:

- represent the most plausible model (showing AICc table results).
- prove you used an appropriate distribution (showing model assumptions diagnostics),
- do not include strongly collinear predictors (showing VIF),
- include a comparison among scaled coefficients of the model.

**Questions**

1. Some complain about the attention on home runs, whereas others love the spectacle. Develop a legitimate, most-plausible, non-collinear model to predict the number of home runs scored in a season by any given batter. Explain your model structure (e.g., *why* do you have interactions?) and your results, including which variable(s) are most important. [4 pts]
2. Produce and explain a CART to for your most plausible and legit model in #1. [1 pt]
3. Beyond home runs, batters who often walk or hit singles, doubles or triples are valuable because they are in scoring position if there is another hit. OBP (on base percentage) represents those batters. Develop and explain a legitimate, most-plausible, non-collinear multiple regression model for OPB. Explain your model structure (e.g., any interactions) and your results, including which variable(s) are most important. [4 pts]
4. Produce and explain a CART to for your most plausible and legit model in #2. [1 pt]