36-617: Applied Linear Models Fall 2022 HW03 – Due Mon Sept 19, 11:59pm

- Please turn the homework in online in our course webspace at canvas.cmu.edu.
 - There is a link to Gradescope in the description of this assignment on Canvas.
 - You should submit a single pdf to Gradescope. If you need help with this, please see https://www.cmu.edu/teaching/gradescope/index.html. Also, allow yourself some extra time to create the pdf & upload it in Gradescope.
 - Don't forget to identify which pages each (part of each) problem appears on in your solitions. Gradescope allows the TA to grade all the problem 1's together, then all the problem 2's, and so forth. This leads to more consistent grading and better comments for you.
 - Remember to list who you worked with, on this and every assignment.
- Reading:
 - For this week: Sheather Ch 6 (supplemental: ISLR 3.3.3; G&H Ch 4)
 - For next week: Sheather Ch 7 (supplemental: ISLR Ch 6; G&H Ch 4)
- There are 4 exercises below. Remember that data sets will be in the "0-books" folder in the files area of our Canvas site (in the "data-sheather" subfolder).

Exercises

- 1. Sheather, Ch 3 (yes!), pp. 109 ff., #5.
- 2. Sheather, Ch 5, p. 147, #2.
- 3. Sheather, Ch 6, pp. 224 ff., #5.
 - For part (b), feel free to transform variables as needed so that (i) the assumptions of the linear regression model are better satisfied; and (ii) the model is still "explainable" to non-statisticians.
 - For part (e): We will discuss variable selection formally next week. For now, remember that the *t* statistic for each column X_j of the *X* matrix tests whether the coefficient β_j is significantly different from zero, *after including all other columns of X in the model*.
- 4. In the folder for this hw assignment you will find a pdf called "COVID breakthrough rates in England". This is a recent article from the medical journal *The Lancet*. Note that the article is exactly in IMRAD format, with sections labelled "Introduction", "Methods", "Results" and "Discussion".

The Abstract for this paper has a special form, called a "structured abstract", in which there is a short, labelled, paragraph corresponding to each section of the paper. We will not be writing structured abstracts in our class, but it is good to see the additional information that a structured abstract can contain. In our class, we will be writing shorter abstracts, consisting of approximately one sentence per section of the paper (so, 4–5 sentences for a paper with 4 sections). Each sentence highlights the main point of each section, possibly with one extra sentence giving the main resul of the paper.

(a) Write a one-paragraph abstract with exactly four sentences, one for each section of the paper: Introduction, Methods, Results and Discussion. Each sentence should highlight the main point of each section, and together the four sentences should tell the story of the paper. The last sentence should include the main result of the paper (or, if you need a fifth sentence to give the main result, that is fine too). (b) Does the paper appropriately address each of the parts of an IMRAD paper as described the "IMRAD: What goes into each section" pdf? (in the hw01 folder in the files area of our Canvas site.)

For each section below, either say "yes this section has the right content", or say "no" and describe what is missing and/or what needs to be moved to another section of the paper or deleted.

- Introduction Results
- Methods Discussion