MATH 145: Biostatistics

Assignment #11

- Read the rest of Chapter 7 in the Ellenberg book, beginning with the section called "The Insignificance of Significance." When you have done so, send me an email message **with the subject line** "Ellenberg Chapter 7". In your message, tell me the thing from the reading you found most interesting. If you can relate it to work we have done in the course, please do. Do this by noon on Monday, May 4.

- Look over the following exercises, and write up your solutions. (In the case of exercises coming from the Lock 5 text, Section 7.2, write up only the underlined ones.) As before, do not submit hand-written answers, but produce a .pdf document with well-written, supported (by commands, graphs, explanations) responses. Submit them as before to `https://www.gradescope.com` by 5 pm Fri., May 1.

  - Lock 5 exercises: 7.30, 7.31, 7.32, 7.34, 7.35, 7.36, 7.37, 7.39, 7.41, 7.44, 7.46, 7.47, 7.48, 7.49, 7.50, 7.53, and 7.54.

  - **Call this one Problem E1**. From a heart disease study, we have record that 63 healthy men with no history of arterial disease died suddenly of stroke. Here is a breakdown of the days of the week on which they died.

Day of the week	Su	Mo	Tu	We	Th	Fr	Sa
No. of deaths	6	22	7	6	13	5	4

    Test the hypothesis that these men were just as likely to die on one day as on any other. Compute the P-value, and state whether the result is significant at level $\alpha = 0.05$.

  - **E2**. Suppose you have a theory that, when you ask subjects to sort single-phrase descriptions of people (such as "eats too fast") into five categories

    not at all like me (10%),
    not much like me (20%),
    occasionally like me (40%),
    often like me (20%),
    very much like me (10%),

    the proportion of sentences that will make it into each category follows the percentages in parentheses. You ask a friend to sort 50 such sentences, and the resulting (observed) frequencies for these 5 categories are 8, 9, 21, 8 and 4, respectively. Does it seem your friend has data that is consistent with your

hypothesized model distribution? Carry out a goodness-of-fit test, compute the P-value, and state a conclusion at level $\alpha = 0.05$.

– **E3**. Suppose you flip a coin 200 times and get 92 heads. Flips only have two categories, "Heads" and "Tails", so a frequency table is only going to have 2 "cells"

Value	Head	Tail
Frequency	92	108

Carry out a hypothesis test where the null hypothesis is that the coin is "fair", and the alternative hypothesis is that the coin is not fair (i.e., has a distribution of values that is different the the 50-50 split that defines fair coins). Do this in *two* different ways and compare your results:

1. Use a goodness-of-fit test.
2. Use a 2-proportion test, like those we carried out in Chapter 6. Note that this 2^{nd} method is available as an option only because the categorical variable *has just two values*: Head and Tail.