This problem set covers material from Week 6, dates 10/14 – 10/17.

Instructions: Write or type complete solutions to the following problems and submit answers to the corresponding Canvas assignment. Your solutions should be neatly-written, show all work and computations, include figures or graphs where appropriate, and include some written explanation of your method or process (enough that I can understand your reasoning without having to guess or make assumptions). A general rubric for homework problems appears on the final page of this assignment.

Monday 10/14

1. People providing an organ for donation sometimes seek the help of a special medical consultant. These consultants assist the patient in all aspects of the surgery, with the goal of reducing the possibility of complications during the medical procedure and recovery. Patients might choose a consultant based in part on the historical complication rate of the consultant's clients.

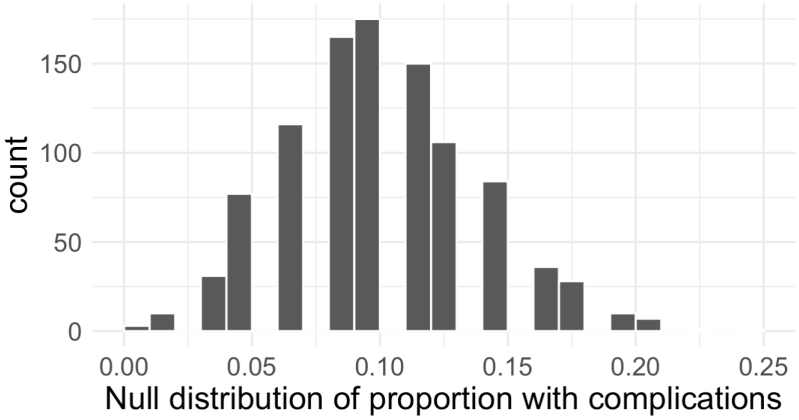
   One consultant tried to attract patients by noting that the average complication rate for liver donor surgeries in the US is about 10%, but her clients have only had 3 complications in the 62 liver donor surgeries she has facilitated. She claims this is strong evidence that her work meaningfully contributes to reducing complications (and therefore she should be hired!).

   (a) The consultant's claim is causal. Is it possible to assess the consultant's claim using the data?

   (b) Using proper notation and defining quantities where necessary, state the hypotheses to test the consultant's claim.

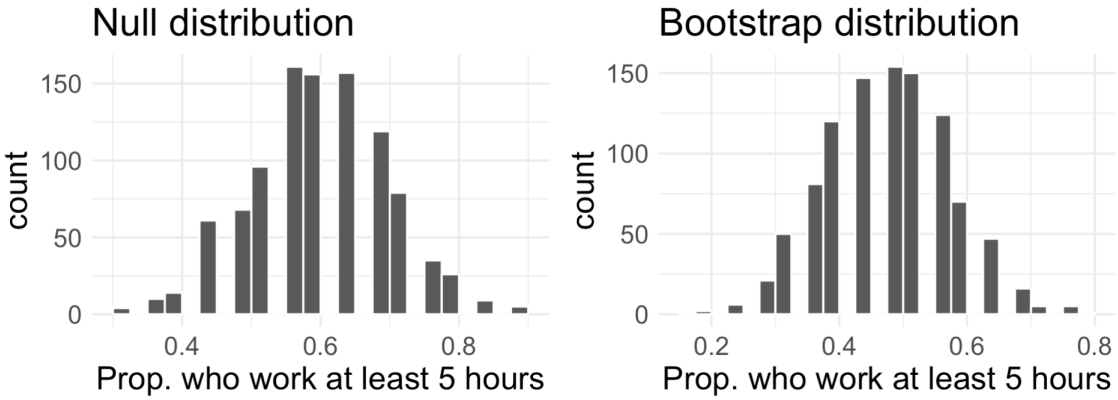
   (c) In words, describe how you would conduct a simulation scheme to obtain a null distribution for the sample statistic. Also describe how you would use the null distribution to calculate the p-value.

   (d) A histogram of the null distribution is shown below, obtain from 1000 simulations. Using the histogram, make an appropriate conclusion in response to the consultant's claim. Use a significance level of 0.05 .



2. In a large university where 60% of the full-time students are employed at least 5 hours per week, the members of the Statistics Department faculty wonder if the same proportion of their students work at least 5 hours per week. They randomly sample 25 of their majors and find that 12 of the students work 5 or more hours per week.

   Two sampling distributions were created to describe the variability in the proportion of statistics majors who work at least 5 hours per week: a null distribution and a bootstrap distribution. In both cases, $B = 1000$ simulations were generated.



   (a) Which distribution(s) was/were obtained by sampling with replacement, and which distribution(s) was/were obtained by sampling without replacement?

   (b) Estimate the standard error of the simulated proportions based on each distribution. Are the two standard errors you estimated roughly equal?

   (c) Using the appropriate histogram, conduct a hypothesis test to answer the faculty's question. State the hypotheses, find the p-value, and conclude in the context of the problem. Use a significance level of 0.10.

   (d) Using the appropriate histogram, find a 90% bootstrap confidence interval for the true proportions of statistics majors who work at least 5 hours per week. Interpret the confidence interval in the context of the problem.

   (e) Briefly comment on how your conclusions in (c) and (d) compare.

Wednesday 10/16

Problems 1-3 in the associated `.Rmd`.

Thursday 10/17

Finish remaining problems in the associated `.Rmd`.

See rubric below for how I will grade coding style!

General rubric

Points	Criteria
5	The solution is correct *and* well-written. The author leaves no doubt as to why the solution is valid.
4.5	The solution is well-written, and is correct except for some minor arithmetic or calculation mistake.
4	The solution is technically correct, but author has omitted some key justification for why the solution is valid. Alternatively, the solution is well-written, but is missing a small, but essential component.
3	The solution is well-written, but either overlooks a significant component of the problem or makes a significant mistake. Alternatively, in a multi-part problem, a majority of the solutions are correct and well-written, but one part is missing or is significantly incorrect.
2	The solution is either correct but not adequately written, or it is adequately written but overlooks a significant component of the problem or makes a significant mistake.
1	The solution is rudimentary, but contains some relevant ideas. Alternatively, the solution briefly indicates the correct answer, but provides no further justification.
0	Either the solution is missing entirely, or the author makes no non-trivial progress toward a solution (i.e. just writes the statement of the problem and/or restates given information).
Notes:	For problems with multiple parts, the score represents a holistic review of the entire problem. Additionally, half-points may be used if the solution falls between two point values above.
Notes:	For problems with code, well-written means only having lines of code that are necessary to solving the problem, as well as presenting the solution for the reader to easily see. It might also be worth adding comments to your code.

There will be an additional section of the rubric for good coding style. Each of the following is worth one point:

- Only including lines of code that are necessary to solving the problem. Comments in code may remain in your submission.
- Having meaningful (or at the very least, not misleading) variable names for objects in `R`.
- Having reproducible code where values are stored as variables for future use, and are actually *used*.
- Setting a seed in code chunks where random sampling occurs.