HW07: Inference

name and student ID

Today's date

- Due date: July 31, 10:00pm (make sure to provide enough time for Gradescope submission to be uploaded).
- Each answer must be on a separate page. Don't delete or add any `\newpage` tags.

Remember: This homework does not involve autograder checking! The submission process is different than usual

- Submission process: Once finished, export your PDF and upload directly to Gradescope. Make sure to not remove or add any `\newpage` tags.

Helpful hints:

- Every function you need to use was taught during lecture! So you may need to revisit the lecture code to help you along by opening the relevant files on Datahub. Alternatively, you may wish to view the code in the condensed PDFs posted on the course website. Good luck!

- Knit your file early and often to minimize knitting errors! If you copy and paste code for the slides, you are bound to get an error that is hard to diagnose. Typing out the code is the way to smooth knitting! We recommend knitting your file each time after you write a few sentences/add a new code chunk, so you can detect the source of knitting errors more easily. This will save you and the GSIs from frustration! **You must knit correctly before submitting.**

- If your code runs off the page of the knitted PDF then you will LOSE POINTS! To avoid this, have a look at your knitted PDF and ensure all the code fits in the file (you can easily view it on Gradescope via the provided link after submitting). If it doesn't look right, go back to your .Rmd file and add spaces (new lines) using the return or enter key so that the code runs onto the next line.

- Useful mathematical notation in markdown:

$$\mu$$

$$\sigma$$

- When this file is completed and knit, your pdf should have 7 pages. Each answer should be on a separate page. Please leave the pagebreaks statements in place to ensure that this is the case.

Question 1 [1 point]

1. After a vaccine is created for SARS-CoV-2, the next important step would be understanding how many Americans will actually get the vaccine. Suppose we want to estimate the proportion of Americans who would get the vaccine if it were available. We interview a random sample of 100 Americans about whether they would choose to be vaccinated if it were an option. Unknown to us, the true population proportion who would be vaccinated is 0.50. What is the expected value and the standard error of the sample proportion?

Note: This sample proportion is only an estimate but reflects the proportion of Americans willing to accept the hypothetical. vaccine in a recent study.

[YOUR ANSWER HERE]

Question 2 [2 points]

Which of the following is an appropriate statement of the Law of Large Numbers? Select just one.

Your answer: **(replace this with a numeric choice, 1-6)**

(1) The law of large numbers states that if you take a large random sample from a population and the data in the population are normally distributed, the data in your sample will be normally distributed.

(2) The law of large numbers states that if you take a large random sample from a population, the data in your sample will be normally distributed.

(3) The law of large numbers states that as the sampe size of a random sample from a population increases, the sample mean approaches the population mean.

(4) The law of large numbers states that as the sample size of a random sample from a population increases, the sample mean will become lower than the population mean.

(5) The law of large numbers states that as the sample size of a random sample from a population increases, the sample mean will become higher than the population mean.

(6) The law of large numbers states that if you take many large random samples from a population, the data from the pooled samples will be normally distributed.

Question 3 [1 point]

Fill in the blanks:

With a couple of assumptions, the time until you receive your next email can be approximated with an exponential population distribution. Say $\bar{x}$ is the sample mean, which in this case is the average amount of time you wait until your next email in the sample of n hours. What are the implications for the sampling distribution of $\bar{x}$?

As the sample size n increases, the sampling distribution of $\bar{x}$ gets closer to a _____ distribution.

Question 4 [2 points]

Read the article from the Scientist times on the Santa Clara antibody Study for SARS-CoV-2.

https://www.the-scientist.com/news-opinion/how-not-to-do-an-antibody-survey-for-sars-cov-2-67488

Name 2 sources of error mentioned in the article other than the error related to sampling variability.

[YOUR ANSWER HERE]

Question 5 [3 points]

Please watch this short video about shifting the population distribution.

https://www.youtube.com/watch?v=8BJNzH6_JpU

Read the 2001 reprint of the 1985 article "Sick Individuals and Sick Populations" by Geoffrey Rose.

Some things to think about from this article. What is the issue Rose highlights with exposures that are very common in a population?

[YOUR ANSWER HERE]

What are the differences in how the high risk vs population strategies affect the distribution?

[YOUR ANSWER HERE]

The ban on smoking in public places (restaurants, bars etc) was argued legally based on the rights of staff in these locations to be free of second hand smoke. The impact has been a shift in the curve of tobacco exposure, smoking and smoking related health outcomes. Name another type or example of an intervention that has been promoted recently based on this idea of shifting the curve in the population?

[YOUR ANSWER HERE]