HW07: Inference

name and student ID

Today's date

- Solutions posted on: March 16th

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:

$$\mu$$

$$\sigma$$

Question 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.

$\mathbb{E}[Vaccination] = 0.50$

Standard Error = 0.05

Question 2

Which of the following is an appropriate statement of the central limit theorem? Select just one.

(1) The central limit theorem 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 central limit theorem states that if you take a large random sample from a population, the data in your sample will be normally distributed.

(3) The central limit theorem states that if you take many large random samples from a population and the data in the population are normally distributed, the sample means will be normally distributed.

(4) The central limit theorem states that if you take many large random samples from a population, the sample means will be normally distributed.

(5) The central limit theorem states that if you take many large random samples from a population and the data in the population are normally distributed, the data from the pooled samples will be normally distributed.

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

4

Question 3

Fill in the blanks:

As n increases the estimate $\bar{x}$ gets closer to _____

μ

Question 4

Read the article from the New York times on margins of error in polling

https://www.nytimes.com/2016/10/06/upshot/when-you-hear-the-margin-of-error-is-plus-or-minus-3-percent-think-7-instead.html

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

sampling bias (sampling non-voters) measurement/analysis error

Question 5 [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. 5a) What is the issue Rose highlights with exposures that are very common in a population?

Very common exposures cannot explain much regarding the distribution of a disease within a population (because almost everyone is exposed).

5b) What are the differences in how the high risk vs population strategies affect the distribution?

High risk intervention strategies truncate the upper tail of the risk distribution by focusing on worst individual cases. The population strategy, on the other hand, causes a downward shift in the entire distribution of risk.

5c) 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?

Many examples. Recently, handwashing, mask wearing, getting vaccinated in relation to COVID-19.