Stat 22000 Summer 2020 Homework 8

All page, section, and exercise numbers below refer to the course text (*OpenIntro Statistics*, 3rd edition, by Diez, Barr, and Cetinkaya-Rundel.).

Reading: Section 4.1, 4.4 and 4.2 (Please read in this order)

Problems for Self-Study : (Do Not Turn In)

- Exercise 4.33, 4.35, 4.37, 4.39, 4.41, 4.13 on p.207-217
- Answers can be found at the end of the book (p.413-415).

Problems to Turn In: due **midnight of Tuesday, July 14, on Canvas**.

1. This problem is essentially the "On Your Own" part in Lab #6.

   `http://www.stat.uchicago.edu/~yibi/s220/labs/lab06.html`

   Please complete the lab and submit answers to the following questions.

   (a) Make a histogram of the `price`. Comment on the shape of the histogram.

   (b) Let's treat the 2930 homes in the data set as the population. Find the mean μ and the SD σ of the sale price of the population.

   (c) Take a random sample of size 25 from `price`. Find the sample mean, and compare it with the population mean μ you found in the previous part.

   (d) Since we have access to the population, simulate the sampling distribution for the sample mean by taking 5000 samples each of size 25 from the population and computing 5000 sample means. Store these means in a vector called `sample_means25`. Make a histogram of the 5000 sample means.

   (e) Repeat the previous part but change the sample size from 25 to 4. Store the 5000 sample means in a vector called `sample_means4`.

   (f) Repeat the previous part again but change the sample size to 100. Store the 5000 sample means in a vector called `sample_means100`

   (g) The 3 histograms made above show (roughly) the sampling distributions of the sample mean for a sample of size 4, 25, and 100, respectively. Compare the center, spread, and the shape of the 3 sampling distributions. How do the center, spread, and shape of the sampling distributions change with the sample size?

   (h) Use the CLT to find the (approximate) probability of getting a sample mean below \$170,000 for a sample of size 100.

   (i) What percentage of the 5000 sample means in `sample_means100$mean` are below \$170,000? Is the percentage close to the probability computed in the previous part using CLT?

   (j) Use the CLT to find the (approximate) probability of getting a sample mean between \$130,000, and \$190,000 for a sample of size 4.

   (k) What percentage of the 5000 sample means in `sample_means4$mean` are between \$130,000 and \$190,000? Hint: Use the following R commands.

   ```
   table(sample_means4 < 190000 & sample_means4 > 130000)
   table(sample_means4 < 190000 & sample_means4 > 130000)/5000
   ```

   Is the percentage close to the probability computed in the previous part using CLT? Does the CLT work well when the sample size is only 4?

2. In the Southern Ocean food web, the krill species Euphausia superba is the most important prey species for many marine predators, from seabirds to the largest whales. Body lengths of the species are normally distributed with a mean of 40 mm and a standard deviation of 12 mm[1].

   (a) What is the probability that a randomly selected krill is longer than 46 mm?

   (b) Describe the distribution of the mean length of a sample of four krill.

   (c) What is the probability that the mean length of a sample of four krill is more than 46 mm?

   (d) Could you estimate the probabilities from parts (a) and (c) if the lengths of krill had a skewed distribution?

3. A study of rush-hour traffic in San Francisco counts the number of people in each car entering a freeway at a suburban interchange. Suppose that this count has mean 1.55 and standard deviation 0.85 in the population of all cars that enter at this interchange during 7-9 am.

   (a) Could the exact distribution of the count be normal? Why or why not?

   (b) Find the (approximate) probability that 800 randomly selected cars at this freeway interchange between 7-9 am will carry a total of 1200 people or more. Show your work. (Hint: Restate this event in terms of the mean number of people $\overline{x}$ per car.)

   (c) Is it possible to find the (approximate) probability that 4 randomly selected cars at this freeway interchange between 7-9 am will carry more than 5 people? Why or why not?

[1]Source : K. Reid et al., "Krill Population Dynamics at South Georgia 1991-1997 Based on Data From Predators and Nets",*Marine Ecology Progress Series*, Vol. 177, pp. 103-14