Stat 22000 Autumn 2017 Homework 5

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

Reading: Section 3.4, 4.1, 4.4

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

- Exercise 3.27, 3.29 on p.164 and Exercise 4.33, 4.35, 4.37, 4.39, 4.41 on p.214-217
- Answers can be found at the end of the book (p.413-415).

Problems to Turn In: due 3 pm on Friday, November 3, on Canvas.

1. (Revision of Exercise 3.28 on p.164) The National Vaccine Information Center estimates that 90% of Americans have had chickenpox by the time they reach adulthood. We now consider a random sample of 120 American adults. We want to find out the probability that 110 or more people in this sample have had chickenpox in their childhood.

   (a) Write down an expression for the exact probability that 110 or more people in this sample have had chickenpox in their childhood in terms of binomial probabilities. You don’t have to evaluate the expression.

   (b) Use normal approximation to approximate the probability that 110 or more people in this sample have had chickenpox in their childhood. Do not use continuity correction.

   (c) Use normal approximation with continuity correction to approximate the probability that 110 or more people in this sample have had chickenpox in their childhood.

   (d) Find the exact probability in (a) in R using the command `sum(dbinom(110:120, size=120, p=0.9))`. Compare the exact probability with the approximate probabilities in (b) and (c).

2. 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 `price`, which shows the population distribution of the sale price of all 2930 homes in the data set. Comment on the shape of the histogram.

   (b) Find the population mean of `price`, which is the mean sale price of all 2930 homes in the data set.

   (c) Take a random sample of size 50 from `price`. Find the sample mean, and compare it with the population mean you found in (a).

   (d) Since you have access to the population, simulate the sampling distribution for $\bar{x}_{price}$ by taking 5000 samples from the population of size 50 and computing 5000 sample means. Make a histogram of the 5000 sample means, then describe the shape of this histogram. Is the center of the histogram close to the population mean?

   (e) Change your sample size from 50 to 150, then compute the sampling distribution using the same method as above in part (d). Make a histogram of the 5000 sample means. Describe the shape of this sampling distribution, and compare it to the sampling distribution for a sample size of 50. Is the center of the histogram close to the population mean?

   (f) Of the sampling distributions from (d) and (e), which has a smaller spread? If we’re concerned with making estimates that are more often close to the true value, would we prefer a distribution with a large or small spread?
3. Exercise 4.36 on p. 215

4. 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?

5. Exercise 4.42 on p. 217

6. Suppose if an introductory statistics textbook was given to all potential readers for review in 1-to-5 stars rating system, with 5 stars being the best, the population distribution of the ratings would be as follows:

50% of the readers give 5 stars, 25% give 4 stars, 25% of the readers give 2 stars.

Of course it is impossible to ask all potential readers to rate the textbook. So the book seller samples potential readers from the population and ask them to rate the book.

The following graphs are

(1) Histogram of the ratings given by all potential readers
(2) Histogram of the ratings given by 100 randomly sampled readers
(3) Histogram of 5000 sample means of the ratings from random samples of each size 10
(4) Histogram of 5000 sample means of the ratings from random samples of each size 100

Determine which graph (A, B, C, or D) is which and explain your reasoning.

\(^1\)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