Practice in the Accuracy of Sample Averages

1. A box of tickets has an average of 250, and an SD of 130. Sixteen hundred draws will be made at random with replacement from this box.

   (a) Estimate the chance that the average of the draws will be in the range 120 to 380.

   (b) Estimate the chance that the average of the draws will be in the range 245 to 255.

2. One hundred draws will be made at random with replacement from box of tickets. The average of the numbers in the box is 500. The SE for the average of the draws is computed and turns out to be 20. True or false:

   (a) About 68% of the tickets in the box are in the range 480 to 520.

   (b) There is about a 68% chance for the average of the hundred draws to be in the range 480 to 520.

3. You are drawing at random with replacement from a box of numbered tickets. Fill in the blanks.

   (a) The expected value for the average of the _______ equals the average of the _______.
   Options: box, draws.

   (b) As the number of draws goes up, this SE for the _______ of the draws goes up but the SE for the _______ of the draws goes down. *Options:* sum, average.

4. Two hundred draws are made at random with replacement from
   | 1 | 2 | 2 | 3 | . True or false and explain.

   (a) The expected value for the average of the draws is exactly 2.

   (b) The expected value for the average of the draws is around 2, give or take 0.05 or so.

   (c) The average of the draws will be around 2, give or take 0.05 or so.

   (d) The average of the draws will be exactly 2.

   (e) The average of the box is exactly 2.

   (f) The average of the box is around 2, give or take 0.05 or so.

5. A university has 15,000 registered students. As part of a survey, 784 of these students were chosen at random. The average age of the sample students turned out to be 24.6 and the SD was 3.5 years.

   (a) The average age of all 15,000 students is estimated as _____. This estimate is likely to be off by _____ or so.

   (b) Find a 95%-confidence interval for the average age of all 15,000 registered students.

6. There are about 200,000 households in Boston. A researcher took a simple random sample of 2,500 of these households. The average income in the 2,500 sample households was \$59,000, with an SD of \$30,000. A histogram for the incomes was plotted and did not follow the normal curve. However, the average income of all the 200,000 households was estimated to be around \$59,000, give or take \$600 or so. Say whether each of the following statements is true or false, and explain why.

   (a) It is estimated that 95% of the 200,000 households in Boston have incomes between \$59,000 − \$1200 = \$57,800 and \$59,000 + \$1200 = \$60,200.

   (b) An approximate 95%-confidence interval for the average income of all households in Boston runs from \$57,800 to \$60,200.

   (c) If someone takes a simple random sample of 2,500 households in Boston, and goes two SEs either way from the average income of the 2,500 sample households, there is about a 95% chance that this interval will cover the average income of all 200,000 households in Boston.

   (d) The normal curve can't be used to figure confidence levels here at all, because the data don't follow the normal curve.

7. A survey organization takes a simple random sample of 625 households from a city of 2,000,000 households. On the average, there are 0.42 dogs per sample household, and the SD is 0.50. Say whether each of the following statements is true or false, and explain.

   (a) The SE for the sample average is 0.02.

   (b) The 0.42 is 0.02 or so off the average number of dogs per household in the whole city.

   (c) Since only one sample was taken, it would be improper to use a confidence interval to estimate the average number of dogs per household in the city.

   (d) A 95%-confidence interval for the average number of dogs per household in the sample is 0.38 to 0.46.

   (e) A 95%-confidence interval for the average number of dogs per household in the city is 0.38 to 0.46.

   (f) 95% of the households in the city have between 0.38 and 0.46 dogs.

   (g) The 95%-confidence level is about right because the number of dogs per household follows the normal curve.

   (h) The 95%-confidence level is about right because, with 625 draws from the box, the probability histogram for the average of the draws follows the normal curve.

Answers follow on the next page.

Answers.

1. The SE for the average is 3.25. The answer to (a) is almost 100%. The answer to (b) is about 88%. Don't confuse the SE for the average of the draws with the SD of the box.

2. (a) False. (b) True.

   To repeat, do not mix up the SE for the average of the draws with the SD of the box.

3. (a) draws, box
   (b) sum, average

4. (a), (c), (e) are true; (b), (c), (f) are false. You know the contents of the box; you can compute the expected value for the average of the draws without error; however, there is chance error in the average of the draws. See exercise 6 on p. 294, exercises 4–6 on pp. 386–387.

5. The box has 15,000 tickets, one for each registered student, showing his or her age. The data are like 784 draws from the box; the sample average is like the average of the draws. The SD of the box is estimated at 3.5 years, the SE for the sum of the draws is $\sqrt{784} \times 3.5 = 98$ years, the SE for the average is $98/784 = 0.125$ years (or calculate as $\frac{3.5\,\text{years}}{\sqrt{784}} = 0.125$ years).

   (a) Estimate is 24.6 years, off by 0.125 years or so.
   (b) The interval is 24.6 ± 0.25 years. (sample avg $\pm 2 \times$ standard error of the average)

6. (a) False: $600 is not the SD. (The data aren't normal, which is another problem.)
   (b) True: the interval is "average ± 2 SE."
   (c) True: that's what confidence intervals are all about (section 21.3).
   (d) False. The normal curve is being used on the probability histogram for the sample average, not the data (pp. 411 and 418–19).

7. (a) True: the SD of the box is estimated from the sample data and the SE is obtained by dividing it by the square root of the number of draws, as on page 416.
   (b) True.
   (c) False. As long as the sample was randomly chosen with replacement from a box and the sample size is reasonably large, confidence interval methods are appropriate.
   A confidence interval is based on the results of a single sample of a particular size.
   (d) False. There is no such thing as a 95%-confidence interval for the *sample average*; you *know* the sample average. It's the population average that you have to worry about (pages 385–386).
   (e) True. The sample average plus or minus 2 SEs *is* a 95%-confidence interval by definition.
   (f) False, This confuses the SD with the SE. And it's ridiculous in the first place, because a household must have a whole number of dogs (0, or 1, or 2, or 3, and so forth). The range 0.38 to 0.46 is impossible for any particular household, let alone 95% of them; although this range is fine for the average of all the households.
   (g) False. For instance, if household size followed the normal curve, there would be many households with a negative number of dogs; we're not ready for that. The distribution must have a long right tail.
   (h) True. See pages 411 and 418–419. Even though household size does not follow the normal curve, you can still use the normal curve to approximate the probability histogram for the sample average.