HW10 - Confidence Intervals

Stat 20 & 131A, Spring 2017, Prof. Sanchez

Due Apr-15

1) Suppose there is a box of 100,000 tickets, each marked 0 or 1. Suppose that in fact, 20% of the tickets in the box are 1's. Calculate the standard error for the percentage of 1's in 400 draws from the box. Show your work. $_{0.5pts}$

2) Three different people take simple random samples of size 400 from the box in the previous exercise, without knowing its contents. The number of 1's in the first sample is 72. In the second, it is 84. In the third, it is 98. Each person estimates the SE by the bootstrap method. Show your work. $_{0.75pts}$

a. The first person estimates the percentage of 1's in the box as _____, and figures this estimate is likely to be off by _____ or so.

b. The second person estimates the percentage of 1's in the box as _____, and figures this estimate is likely to be off by _____ or so.

c. The third person estimates the percentage of 1's in the box as _____, and figures this estimate is likely to be off by _____ or so.

3) A box contains 10,000 marbles, of which some are orange and the others magenta. To estimate the percentage of orange marbles in the box, 100 are drawn at random without replacement. Among the draws, 3 turned out to be orange. The percentage of orange marbles in the box is estimated as 3%, with an SE of 1.7%. True or False: a 95%-confidence interval for the percentage of orange marbles in the box is 3% $\pm$ 3.4%. Explain. $_{0.5pts}$

4) In a certain town, there are 30,000 people aged 16 and over. To estimate the percentage of them who use a certain social network website, a statistician chooses a simple random sample of size 1,500. As it turns out, 510 of the sample people did use the website. Complete the following table; the first 3 lines refer to the sample percentage who saw the show. (N/A = not applicable.) $_{0.6pt}$

	Known to be	Estimated from data
Observed value	34%	N/A
Expected value	N/A	34%
SE		
SD of box		
Number of draws		

5) A simple random sample of 400 persons is taken to estimate the percentage of Republicans in a large population. It turns out that 214 of the people in the sample are Republicans. True or False and explain. $_{1.5pts}$

a. The sample percentage is 54.3%; the SE for the sample percentage is 1.6%.

b. The sample percentage is 53.5%; the SE for the sample percentage is 2.5%.

c. 53.5% ± 4.12% is a 90%-confidence interval for the population percentage.

d. 53.5% ± 5% is a 95%-confidence interval for the population percentage.

e. 53.5% ± 5% is a 95%-confidence interval for the sample percentage.

f. There is about a 95% probability for the percentage of Republicans in the population to be in the range 53.5% ± 5%.

6) The Residential Energy Consumption Survey found in 2001 that 47% of American households had internet access. A market survey organization repeated this study in a certain town with 25,000 households, using a simple random sample of 500 households: 239 of the sample hoseholds had internet access. Show your work. $_{0.5pts}$

a. The percentage of households in the town with internet access is estimated as _____; this estimate is likely to be off by _____ or so.

b. If possible, find a 95%-confidence interval for the percentage of all 25,000 households with internet access. If this is not possible, explain why not.

7) One year, there were 252 trading days on the New York Stock Exchange, and IBM commoon stock went up 131 of them: $131/252 \approx 52\%$. A statistician attaches a standard error to this percentage as follows:

- SE for number $= \sqrt{252} \times \sqrt{0.52 \times 0.48} \approx 8$
- SE for percent $= (8/252) \times 100\% \approx 3\%$

Is this the right SE? Answer yes or no, and explain. $_{1pt}$

8) A box of tickets averages out to 75, and the SD is 10. One hundred draws are made at random with replacement from this box. Show your work. $_{0.5pts}$

a. Find the chance (approximately) that the average of the draws will be in the range 65 to 85.

b. Repeat, for the range 74 to 76.

9) A box contains 10,000 tickets. The numbers on these tickets average out to 50, and the SD is 20. Show your work. $_{0.6pts}$

a. One hundred tickets are drawn at random with replacement. The average of these draws will be around _____, give or take _____ or so.

b. What are if 100 draws are made without replacement?

c. What if 100 draws are made without replacement, and there are only 100 tickets in the box?

10) One hundred draws are made at random with replacement from a box. The average of the box is 3.1. $_{0.6pts}$

a. True or False: the expected value for the average of the draws is exactly equal to 3.1. If this cannot be determined from the information given, what else do you need to know, and why?

b. What is the SE for the average of the draws? If this cannot be determined from the information given, what else do you need to know, and why?

11) A real state office wants to make a survey in a certain town, which has 50,000 households, to determine average commute distance to work. A simple random sample of 1,000 households is chosen. The real state office interviewed all persons age 16 and over in the sample of households; there were 2,500 such persons. On the average, these 2,500 people commuted 7.1 miles to work, and the SD of the distances was 10.2 miles. If possible, find a 95%-confidence interval for the average commute distance for all people age 16 and over in this town. If this isn't possible, explain why not. $_{1pt}$

12) A survey organization takes a simple random sample of 625 households from a city of 80,000 households. On the average, there are 2.30 persons per sample household, and the SD is 1.75. Say whether each of the following statements is ture or false, and explain. $_{1.5pts}$

a. The SE for the sample average is 0.07.

b. A 95%-confidence interval for the average household size in the sample is 2.16 to 2.44.

c. A 95%-confidence interval for the average household size in the city is 2.16 to 2.44.

d. 95% of the households in the city contain between 2.16 and 2.44 persons.

e. The 95%-confidence level is about right because household size follows the normal curve.

f. 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.

13) One term at the University of California, Berkeley, 400 students took the final in a Statistics course. Their scores averaged 65.3 out of 100, and the SD was 25. Now

$$\sqrt{400} \times 25 = 500, \qquad \frac{500}{400} = 1.25$$

Is 65.3 ± 2.5 a 95%-confidence interval? If so, for what? If not, why not? $_{0.45pts}$