Confidence Interval Practice

1. A random sample of 18,000 people in the United States was selected, and each person was asked if he felt supportive of the Occupy USA movement. In this sample, 10,811 felt supportive.

   Find a 95%-confidence interval for the percentage of people in the United States who feel supportive of the Occupy USA movement. (Assume that the sample is sufficiently large that a normal approximation is applicable.)

2. True or false: with a well-designed sample survey, the sample percentage is very likely to equal the population percentage. Explain.

3. True or false: with a very large sample, the sample percentage and the population wpercentage will be the same.

4. A monthy opinion survey is based on a sample of 1,500 persons, "scientifically chosen as a representative cross section of the American public." The press release warns that the estimates are subject to chance error, but guarantees that they are "reliable to within two percentage points." The word "reliable" is ambiguous. According to statistical theory, the guarantee should be interpreted as follows:

   (i) In virtually all these surveys, the estimates will be within two percentage points of the parameters.

   (ii) In most such surveys, the estimates will be within two percentage points of the parameters, but in some definite percentage of the time larger errors are expected.

   Choose the appropriate interpretation and explain.

Answers follow on the next page.

Answers.

1. The box has millions of tickets, one for each person in the United States. Tickets are marked 1 for supporter, and 0 for non-supporter. The data are like 18,000 draws from the box, and the number of supporters in the sample like the sum of the draws.

   The fraction of 1's in the box can be estimated from the sample as 0.600611.

   On this basis, the SD of the box is estimated as $\sqrt{0.600611 \times 0.399389} \approx 0.489773$.

   The SE for the number of supporters in the sample is estimated as $\sqrt{18,000} \times 0.489773 \approx 65.71$.

   The SE for the percentage is $65.71/18,000 \times 100\%$, which is about 0.365%. The percentage of supporters in the population is estimated as 60.06%, give or take 0.365% or so.

   <div align="center">The 95%-confidence interval is 60.06% ± 0.73%.</div>

   Calculation Details The sample percentage is $(10,811/18,000) \times 100\% = 60.06\%$. Using the alternate formula, the SE for the sample percentage of Democrats is figured as $\dfrac{\sqrt{0.600611 \times 0.399389}}{\sqrt{18,000}} \times 100\% = 0.365\%$.

   The formula for a 95%-confidence interval is:

   sample percentage plus or minus twice the standard error for the percentage.

   So 60.06% ± 0.73% is a 95%-confidence interval for the percentage of supporters in the whole country.

   It could be restated as: between 59.33% and 60.79%.

2. False, because of chance error. The sample percentage is likely to be close to the population percentage, but not exactly equal. The SE for the percentage says how far off you can expect to be.

3. False. They will be close if the sample is very large, but there is still chance error present.

   Chance error can be tamed, but not eliminated.

4. Option (ii) is it. For example, about 95% of the estimates will be right to within 2 SEs, about 99.7% of them will be right to within 3 SEs, and so forth.