Class Worksheet

April 11 and 16
Math 125 *Kovitz* 2025

The Accuracy of Percentages (Chapter 21)

Inference From the Sample to the Population (the Bootstrap Method).

The bootstrap. When sampling from a zero-one box whose composition is unknown, the SD of the box can be estimated by substituting the fractions of 0's and 1's in the sample for the unknown fractions in the box. The estimate is good when the sample is reasonably large.

Confidence intervals.

A confidence interval for a percentage—with a confidence level specified as a percent—is a range of percentages such that you are that percent confident that the population percentage is in that interval.

A confidence interval is based on the results of a single sample of a particular size.

- the interval "sample percentage ± 1 SE" is an approximate 68%-confidence interval for the population percentage.
- the interval "sample percentage ± 2 SEs" is an approximate 95%-confidence interval for the population percentage.
- the interval "sample percentage ± 3 SEs" is an approximate 99.7%-confidence interval for the population percentage.
- any other confidence level except 100% will lead to a certain number of SEs to go in either direction from the sample percentage in order to get the corresponding confidence interval.

Interpreting a Confidence Interval

The chances are in the sampling procedure, not in the parameter.

A confidence interval is used when estimating an unknown parameter from sample data. The interval gives a range for the parameter—and a confidence level that the range covers the true value.

A sample percentage will be off the population percentage, due to chance error. The SE tells you the likely size of the amount off.

Warning. The formulas for simple random samples may not apply to other kinds of samples.

Problems to think about

In Boston, there are 280,000 voters in a mayoral election. The percentage of votes for Tom Menino is unknown. To estimate it, a simple random sample of 625 voters is drawn, and 475 of them turn out to be Menino voters.

The percentage of Menino voters in Boston is estimated to be _____%, give or take _____% or so.

If possible, find an approximate 95%-confidence interval for the percentage of Menino voters in Boston.

If possible, find an approximate 99%-confidence interval for the percentage of Menino voters in Boston.

If possible, find an approximate 99.7%-confidence interval for the percentage of Menino voters in Boston.

A coin is tossed 2500 times. The percentage of heads among the tosses has an expected value of 50%, and an SE of 1%. Say whether each of the following statements is true or false, and explain briefly.

(a) The 1% measures the likely size of the chance error in the 50%.
(b) The percentage of heads among the 2500 tosses will be around 50%, give or take 1% or so.
(c) An approximate 95%-confidence interval for the percentage of heads in the 2500 tosses is 48% to 52%.

A box contains a large number of red and blue marbles, but the proportions are unknown; 400 marbles are drawn at random and 316 turn out to be red. The percentage of red marbles in the box is estimated as 79%, and the SE for the percentage in the sample is figured as 2%. Say whether each of the following statements is true or false, and explain briefly.

(a) The 2% measures the likely size of the chance error in the 79%.
(b) The 79% is likely to be off the percentage of red marbles in the box by 2% or so.
(c) An approximate 95%-confidence interval for the percentage of red marbles in the box (the population percentage) is 75% to 83%.
(d) An approximate 95%-confidence interval for the percentage of red marbles in the sample is 75% to 83%.
(e) There are about two chances in three for the percentage of red marbles in the box to be in the range $79\% \pm 2\%$.

Another box containing a large number of balls has 82% red and 18% blue. A simple random sample of 900 balls is drawn from this population. The SE for the sample percentage of red balls is figured as 1.3%. True or false: There are about two chances in three for the percentage of red balls in the sample to be in the range $82\% \pm 1.3\%$. Explain.

It rained in Boston on 113 days last year: $113/365 \approx 31\%$. A statistician attaches a standard error to this percentage as follows:

$$\begin{aligned} \text{SE for number} &= \sqrt{365} \times \sqrt{0.31 \times 0.69} \approx 9 \\ \text{SE for percent} &= \frac{9}{365} \times 100\% \approx 2.5\% \end{aligned}$$

Is this the right SE? Answer yes or no, and explain.