THE CHANCE ERROR IN A SAMPLE PERCENTAGE

• A certain town has a population of 100,000 people age 18 and over. 20% of these people are well-educated, that is, have college degrees. A simple random sample of 1,600 people will be drawn from this population.

- Simple random sampling means drawing at random *without* replacement.
    - At each draw, all the people not already selected have an equal chance of being chosen.
- The percentage of well-educated people in the sample will be around ______%, give or take ______% or so.
- The chance that between 19% and 21% of the people in the sample are well-educated is about ______ .

How do you go about filling in the blanks?

• Step 1. Realize that

$$\text{Percentage of well-educated people in the sample} = \frac{\text{Number of well-educated people in the sample}}{\text{size of the sample}} \times 100\%$$

- Any statement about the chance variability in the *number* of well-educated people in the sample can be converted to a similar statement about the chance variability in the *percentage* of such people, by dividing by the ____________ and multiplying by ______ .

• Step 2. Set up a box model. The number of well-educated people in the sample is like the sum of ______ draws made at random ______ replacement from the box

| □'s □'s |

- We want the process of selecting people from the population to be like drawing tickets from the box. So there should be:
    - ______ ticket in the box for each person in the population, and
    - ______ draw for each person in the sample.
- So there should be ______ tickets in the box, and ______ draws.
- The tickets for the well-educated people should be marked □, and the other tickets as □.
    - We want the number of well-educated people in the sample to be like the sum of the draws from the box. This can be achieved by classifying and counting. To illustrate:

The sample people	Tom	Dick	Harry	Number who are well-educated =
	☺	☺	☺	
Their tickets	□	□	□	Sum of the draws =

- Since 20% of the population is well-educated, there should be ______ □'s in the box, and ______ □'s.

• Step 3. Realize that for drawing from the box

| 20,000 $\boxed{1}$'s 80,000 $\boxed{0}$'s |

there isn't much difference between drawing 1600 times *without* replacement and drawing 1600 times *with* replacement.

- When drawing with replacement, the chance of getting a $\boxed{1}$ is ____________ in ____________, or 1 in 5, for each draw.
- When drawing without replacement 1600 times, the chance of getting a $\boxed{1}$ changes from draw to draw, but stays close to 1 in 5, because the percentage of $\boxed{1}$'s in the box can't change much during the course of the draws.
  - The number of draws is ________ compared to the numbers of $\boxed{1}$'s and $\boxed{0}$'s in the box.
- So, the number of well-educated people in the sample is almost like the sum of 1600 draws made at random *with* replacement from the box.

• Step 4. Find the expected value and SE for the sum of 1600 draws (made with replacement) from the box.

- The expected value for the sum of the draws from the box equals (number of draws) × (average of box); that's

$$\_\_\_\_\_\_\_\_ \times \_\_\_\_\_\_\_\_ = \_\_\_\_\_\_\_\_.$$

- The SE for the sum of the draws from the box equals $\sqrt{\text{number of draws}} \times$ (SD of box); that's

$$\_\_\_\_\_\_\_\_ \times \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ = \_\_\_\_\_\_\_\_.$$

- The sum of the draws from the box will be around ________, give or take ________ or so.

• Step 5. Interpret the results of Step 4 in terms of the sample.

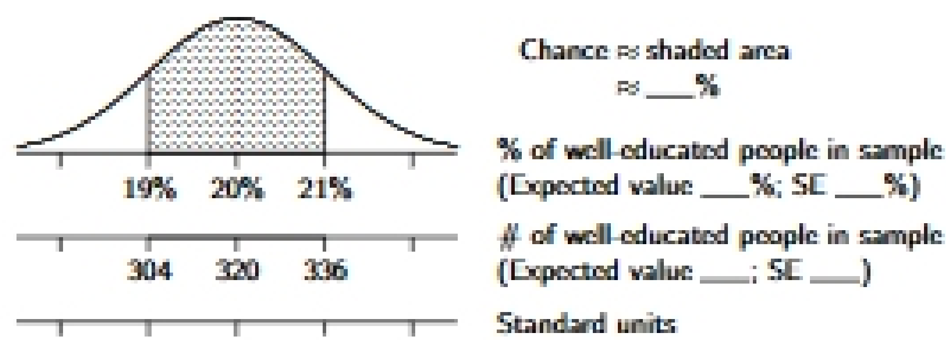
- The number of well-educated people in the sample will be about ________, give or take ________ or so.

• Step 6. Convert to percents, relative to the size of the sample.

- 16 is ________% of 1600.
- 320 is ________% of 1600.
- So the percentage of well-educated people in the sample will be around ________% (i.e., the percentage of well-educated people in the population), give or take ________% or so.

• Step 7. Use the normal approximation to estimate chances about the percentage of well-educated people in the sample.

- The chance that between 19% and 21% of the people in the sample are well-educated is the chance that the number of well-educated people in the sample lies between 320 − 16 = 304 and 320 + 16 = 336.



- The normal approximation can be used, because the sample size (and therefore the number of draws from the box) is reasonably ________.

THE STANDARD ERROR FOR A SAMPLE PERCENTAGE

• Consider taking a simple random sample from a box of tickets, some of which are marked 1, the others 0:

$$\boxed{\text{some } \boxed{1}\text{'s} \qquad \text{some } \boxed{0}\text{'s}} \leftarrow \text{The "population", in abstract form}$$

Suppose the number of draws is large in absolute terms, but small relative to the number of tickets in the box.

• The number of $\boxed{1}$'s in the sample is expected to be about

$$\begin{aligned} &(\text{number of draws}) \times (\text{average of the box}) \\ &= (\text{number of draws}) \times (\text{fraction of } \boxed{1}\text{'s in the box}), \end{aligned}$$

give or take

$$\begin{aligned} &\text{SE for the number of } \boxed{1}\text{'s in the sample} \\ &= \sqrt{\text{number of draws}} \times (\text{SD of the box}) \end{aligned}$$

or so.

• Now convert to percents relative to the number of draws, by dividing by the number of draws and multiplying by 100%. The percentage of $\boxed{1}$'s in the sample is expected to be about

$$\begin{aligned} &\frac{\text{expected number of } \boxed{1}\text{'s in the sample}}{\text{number of draws}} \times 100\% \\ &= \boxed{\text{percentage of } \boxed{1}\text{'s in the box,}} \end{aligned}$$

give or take

$$\begin{aligned} &\frac{\text{SE for the number of } \boxed{1}\text{'s in the sample}}{\text{number of draws}} \times 100\% \\ &= \boxed{\frac{\text{SD of the box}}{\sqrt{\text{number of draws}}} \times 100\%} \end{aligned}$$

or so.

• To put it another way, as an estimate of the percentage of $\boxed{1}$'s in the box, the percentage of $\boxed{1}$'s in the sample will be off by a chance amount similar in size to the quantity

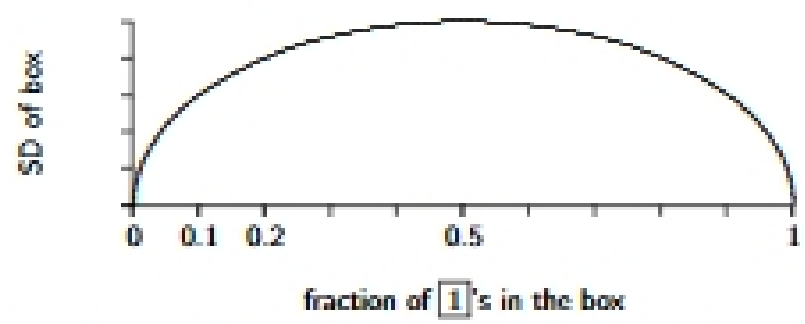
$$\frac{\text{SD of the box}}{\sqrt{\text{number of draws}}} \times 100\%;$$

this quantity is called the standard error (SE) for the sample percentage.

• How does the SE for the sample percentage depend on the box?

  • Only through the variability of the tickets in the box, as measured by their ________, which equals

SD of a box containing $\boxed{0}$'s and $\boxed{1}$'s, as a function of the fraction of $\boxed{1}$'s



• How does the SE depend on the number of draws?

  • The SE is inversely proportional to the square root of the size of the sample. In particular, large samples are more accurate than small ones. Multiplying the size of the sample by some factor divides the SE for the sample percentage by the square root of that factor.