Ch. 20: Chance Errors in Sampling

When the box contains $\boxed{0}$'s and $\boxed{1}$'s, the percentage of $\boxed{1}$'s in the draws will be around

$$\text{EV}_{\%} = \%\text{ of } \boxed{1} \text{ 's in the box}$$

The corresponding SE is

$$\text{SE}_{\%} = \frac{\text{SE}_{\text{sum}}}{\#\text{draws}} \times 100\%$$

i.e., $\text{SE}_{\%} = \text{SE}_{\text{sum}}$ converted to a percentage.

Ex: A certain university has 4,000 male students and 6,000 female students. If we sample 100 students without replacement, what percentage of our sample do we expect to be female? How much do we expect this to vary by?

Our sample is like drawing 100 times without replacement from the box:

The box average is:

The box SD is:

For a sample of size 100, the $EV_{\%}$ is:

The $SE_{\%}$ is:

If we sample 400 students, what will be the $EV_{\%}$ and $SE_{\%}$ of women?

Note:
Multiplying the sample size by some factor divides the $SE_{\%}$ by the square root of the factor.

Ex: In a population of 100,000 telephone subscribers, 20% earn more than $50,000. In a random sample of 400 of these subscribers, how many do you expect to earn over $50,000? What is the corresponding SE?

In the sample of 400, what percentage do you expect to earn over $50,000. What is the corresponding $SE_{\%}$?

And what is the $SE_{\%}$ if the sample size is 1,600? And for 3,600?

Using the Normal Curve

Provided the sample size is large, the normal curve can be used to figure out chances.

Ex: 25% of college students own an automobile. If we take a random sample of 400 of these students, what is the chance that less than 20% own an automobile?

Accuracy of Samples

When we estimate a percentage based on a sample survey, we find:

- The sample size makes a big difference in the SE.

- Provided the sample size is small compared to the population size (10% or less), the accuracy of samples is determined by the sample size itself (and not the size relative to the population).

Ex: The Gallup Poll can use a sample of a few thousand people to estimate percentages for the entire country (200 million voters or so).

Ex: If we do simple random samples of size 2,500 of voters in every state, will the SE of the percentage of Republicans in the Utah sample be much smaller than that of California? (You may assume the SD of the "box" in each case is essentially 0.5.)

Ex: (Old Exam question):
A local politician is interested in estimating the percentage of voters who are opposed to the marriage tax. She can only afford to sample 1000 people. Other things being equal, to get equal accuracy in Logan and Salt Lake City, she would sample:

1. 500 people in Logan and 500 people in Salt Lake City.

2. more people in Salt Lake City than in Logan.

3. more people in Logan than in Salt Lake City.