Ch. 23: The Accuracy of Averages

Looking at the average of a sample is similar to looking at percentages in a sample.

Ex: Suppose we roll a die 100 times. The box model is:

The average of the box is 3.5, and the SD is 1.7.

The sum of the 100 rolls will be about __________, give or take about __________.

If the sum is 350, the average will be: ___________

If the sum is 1 SE low, 350 - 17 = 333, the average will be: ___________

If the sum is 1 SE high, 350 + 17 = 367, the average will be: ___________

The average of the rolls will be about 3.5, give or take about 0.17.

Note:

The EV of the average of draws is:

$$EV_{avg} =$$

The SE of the average of draws is:

$$SE_{avg} =$$

Normal Approximation for Averages

Recall: When drawing at random with replacement, the sum of the draws will approximately follow the normal curve for a large number of draws.

Also, the percentage of 1's in the draws (from a 0-1 box) will approximately follow the normal curve for a large number of draws.

Finally, the average of the draws will also approximately follow the normal curve for a large number of draws.

Ex: If we make 100 rolls of a fair die, what is the chance that the average of the rolls will be greater than 3.6?

And what is the chance that the average of 400 rolls is greater than 3.6?

Note:
As with percentages, increasing the number of draws by a factor will decrease the SE of the average by the square root of the factor.

The Sample Average

Usually, we don't know the box, just the sample.

As in the percentages case, we assume the box is well-represented by the sample.

Ex: A telephone company wishes to estimate the average length of weekend long-distance calls. A random sample of 50 calls gives an average length of 15 minutes, and an SD of 7 minutes.

Find the average length of all weekend long-distance calls and the corresponding SE:

CI's for the Average

We can calculate confidence intervals in the same way as for percentages – with a large sample, we can be 95% confident that the true population (box) average is within 2 SE's of the sample average.

Ex: For the telephone call example, a 95% confidence interval for the average length of a call is:

A 90% confidence interval for the average length of a call is:

What do these CIs mean?

Ex: Educational level is measured for a sample of 400 people, and the average is found to be 11.6 years, with an SD of 4.1 years. What is a 95% confidence interval for average educational level?

Is the normal approximation reasonable?

Which SE?

- SE for sum $= \sqrt{\text{number of draws}} \times$ box SD

- SE for average $= \dfrac{\text{SE for sums}}{\text{number of draws}}$

- SE for count = SE for a sum from a 0-1 box

- SE for percentage $= \dfrac{\text{SE for count}}{\text{number of draws}} \times 100\%$

- If we know the box, we reason forward. The chance quantity will be near the EV, but probably off by an SE or so.

- If we don't know the box, we reason backward. We estimate the EV by the chance quantity, but recognize that we are probably off by an SE or so, which we estimate by estimating the box SD from the sample.

Note:

The formulas for simple random samples don't apply in other cases!