Lesson 17: Margin of Error when Estimating a Population Proportion

Classwork

In this lesson, you will find and interpret the standard deviation of a simulated distribution for a sample proportion and use this information to calculate a margin of error for estimating the population proportion

Exercises 1-6: Standard Deviations for Proportions

In the previous lesson, you used simulated sampling distributions to learn about sampling variability in the sample proportion and the margin of error when using a random sample to estimate a population proportion. However, finding a margin of error using simulation can be cumbersome and take a long time for each situation. Fortunately, given the consistent behavior of the sampling distribution of the sample proportion for random samples, statisticians have developed a formula that will allow you to find the margin of error quickly and without simulation.

1. 30% of students participating in sports at Union High School are female (a proportion of 0.30).

   a. If you took many random samples of 50 students play sports and made a dot plot of the proportion of females in each sample, where do you think this distribution will be centered? Explain your thinking.

   b. In general, for any sample size, where do you think the center of a simulated distribution of the sample proportion of females in sports at Union High School will be?

2. Below are two simulated sampling distributions for the sample proportion of females in random samples from all the students at Union High School.

Simulated Sampling Distribution 2	Simulated Sampling Distribution 2
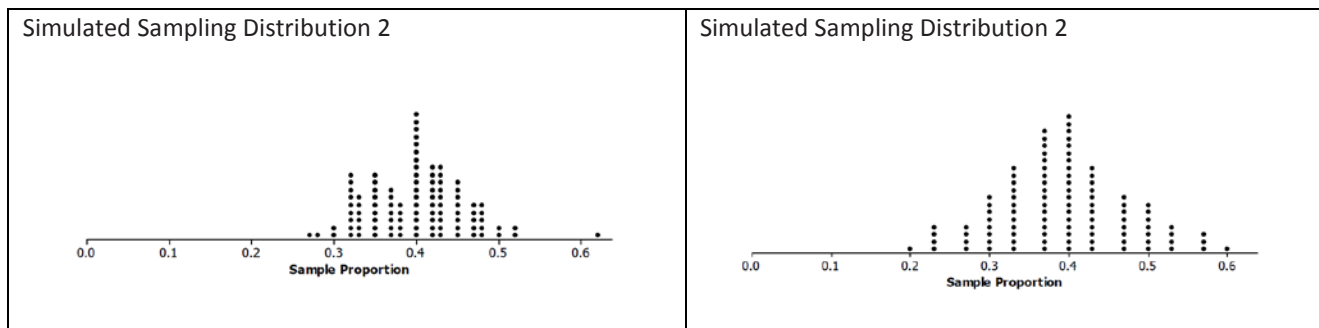
   a. Based on the two sampling distributions above, what do you think is the population proportion of females?

   b. One of the sampling distributions above is based on random samples of size 30, and the other is based on random samples of size 60. Which sampling distribution corresponds to the sample size of 30? Explain your choice.

3. Remember from your earlier work in statistics that distributions were described using shape, center, and spread. How was spread measured?

4. In previous lessons, you saw a formula for the standard deviation of the sampling distribution of the sample mean. There is also a formula for the standard deviation of the sampling distribution of the sample proportion. For random samples of size n, the standard deviation can be calculated using the following formula:

    standard deviation $= \sqrt{\frac{p(1-p)}{n}}$, where p is the value of the population proportion and n is the sample size.

    a. If the proportion of females at Union High School is 0.4, what is the standard deviation of the distribution of the sample proportions of females for random samples of size 50? Round your answer to three decimal places.

    b. The proportion of males at Union High School is 0.6. What is the standard deviation of the distribution of the sample proportions of males for samples of size 50? Round your answer to three decimal places.

    c. Think about the graphs of the two distributions in parts (a) and (b). Explain the relationship between your answers using the center and spread of the distributions.

5. Think about the simulations that your class performed in the previous lesson and the simulations in Exercise 2 above.

    a. Was the sampling variability in the sample proportion greater for samples of size 30 or for samples of size 50? In other words, does the sample proportion tend to vary more from one random sample to another when the sample size is 30 or 50?

    b. Explain how the observation that the variability in the sample proportions decreases as the sample size increases is supported by the formula for the standard deviation of the sample proportion.

6. Consider the two simulated sampling distributions of the proportion of females in Exercise 2 where the population proportion was 0.4. Recall that you found $n = 60$ for Distribution 1, and $n = 30$ for Distribution 2.

   a. Find the standard deviation for each distribution. Round your answer to three decimal places.

   b. Make a sketch and mark off the intervals one standard deviation from the mean for each of the two distributions. Interpret the intervals in terms of the proportion of females in a sample.

Simulated Sampling Distribution 1: $n = 60$
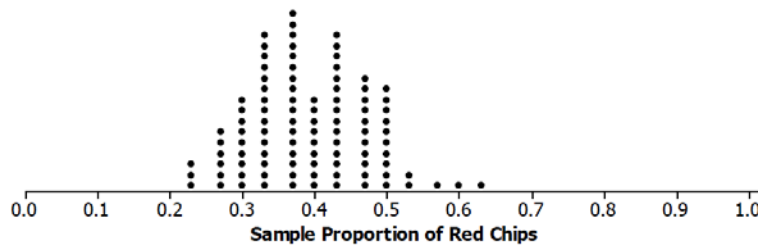
Simulated Sampling Distribution 2: $n = 30$
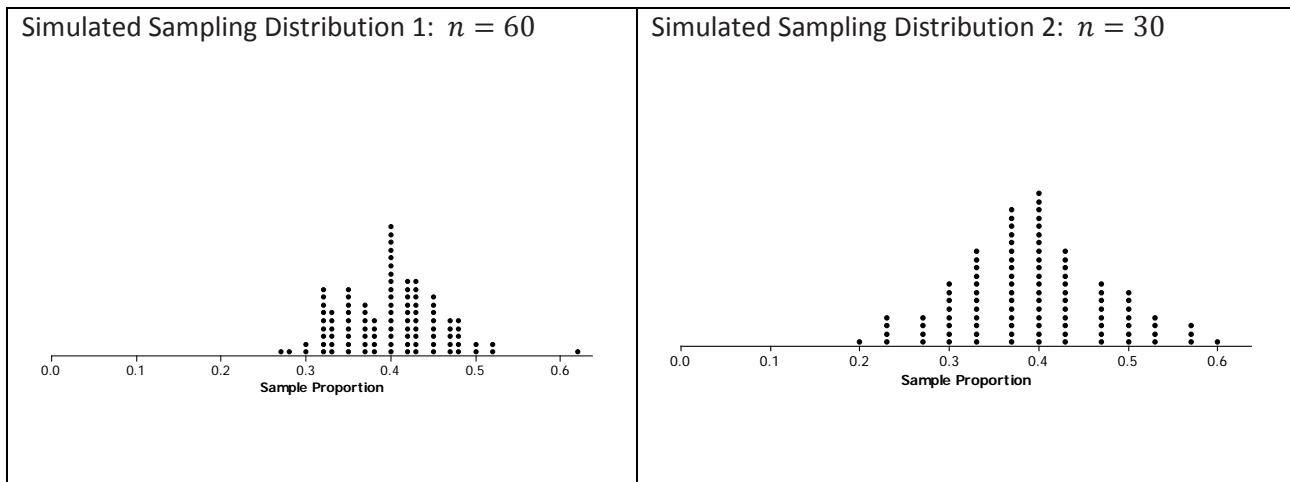
In general, three results about the sampling distribution of the sample proportion are known:

- The sampling distribution of the sample proportion is centered at the actual value of the population proportion, p.

- The sampling distribution of the sample proportion is less variable for larger samples than for smaller samples. The variability in the sampling distribution is described by the standard deviation of the distribution, and the standard deviation of the sampling distribution for random samples of size n is $\sqrt{\frac{p(1-p)}{n}}$, where p is the value of the population proportion. This standard deviation is usually estimated using the sample proportion, which is denoted by $\hat{p}$ (read as p-hat), to distinguish it from the population proportion. The formula for the estimated standard deviation of the distribution of sample proportions is $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

- As long as the sample size is large enough that the sample includes at least 10 successes and failures, the sampling distribution is approximately normal in shape. That is, a normal distribution would be a reasonable model for the sampling distribution.

**Exercises 7–12: Using the Standard Deviation with Margin of Error**

7. In the work above, you investigated a simulated sampling distribution of the proportion of females in a sample of size 30 drawn from a population with a known proportion of 0.4 females. The simulated distribution of the proportion of red chips in a sample of size 30 drawn from a population with a known proportion of 0.4 is displayed below.

a. Use the formula for the standard deviation of the sample proportion to calculate the standard deviation of the sampling distribution. Round your answer to three decimal places.

b. The distribution from Exercise 2 for a sample of size 30 is below. How do the two distributions compare?
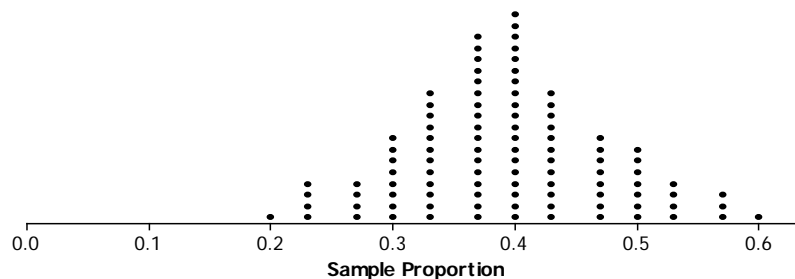
c.   How many of the values of the sample proportions are within one standard deviation of 0.4? How many are within two standard deviations of 0.4?

Sample Proportion of Red Chips
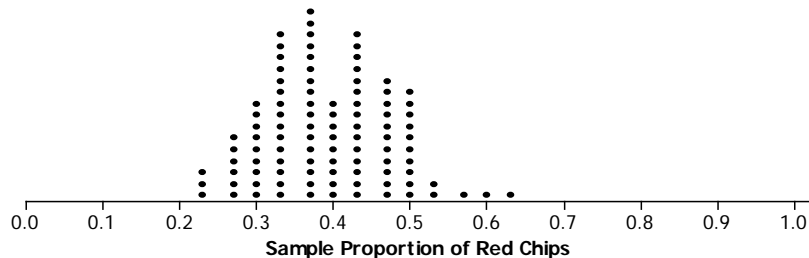
In general, for a known population proportion, about 95% of the outcomes of a simulated sampling distribution of a sample proportion will fall within two standard deviations of the population proportion. One caution is that if the proportion is close to 1 or 0, this general rule may not hold unless the sample size is very large. You can build from this to estimate a proportion of successes for an unknown population proportion and calculate a margin of error without having to carry out a simulation.

If the sample is large enough to have at least 10 of each of the two possible outcomes in the sample, but small enough to be no more than 10% of the population, the following formula (based on an observed sample proportion $\hat{p}$) can be used to calculate the margin of error. The standard deviation involves the parameter p that is being estimated. Because p is often not known, statisticians replace p with its estimate $\hat{p}$ in the standard deviation formula. This estimated standard deviation is called the standard error of the sample proportion.

8.

a.   Suppose you draw a random sample of 36 chips from a mystery bag and find 20 red chips. Find $\hat{p}$, the sample proportion of red chips, and the standard error.

b.   Interpret the standard error.

When estimating a population proportion, **margin of error** can be defined as the **maximum expected difference** between the value of the population proportion and a sample estimate of that proportion (the farthest away from the actual population value that you think your estimate is likely to be).

> If $\hat{p}$ is the sample proportion for a random sample of size n from some population, and if the sample size is large enough:
>
> $$\text{estimated margin of error} = 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

9. Henri and Terence drew samples of size 50 from a mystery bag. Henri drew 42 red chips, and Terence drew 40 red chips. Find the margins of error for each student.

10. Divide the problems below among your group, and find the sample proportion of successes and the estimated margin of error in each situation:

   a. Sample of size 20, 5 red chips

   b. Sample of size 40, 10 red chips

   c. Sample of size 80, 20 red chips

   d. Sample of size 100, 25 red chips

11. Look at your answers to Exercise 2.

    a. What conjecture can you make about the relation between sample size and margin of error? Explain why your conjecture makes sense.

    b. Think about the formula for a margin of error. How does this support or refute your conjecture?

12. Suppose that a random sample of size 100 will be used to estimate a population proportion.

    a. Would the estimated margin of error be greater if $\hat{p} = 0.4$ or $\hat{p} = 0.5$? Support your answer with appropriate calculations.

    b. Would the estimated margin of error be greater if $\hat{p} = 0.5$ or $\hat{p} = 0.8$? Support your answer with appropriate calculations.

    c. For what value of $\hat{p}$ do you think the estimated margin of error will be greatest? (Hint: Draw a graph of $\hat{p}(1 - \hat{p})$ as $\hat{p}$ ranges from 0 to 1.)

Lesson Summary

- Because random samples behave in a consistent way, a large enough sample size allows you to find a formula for the standard deviation of the sampling distribution of a sample proportion. This can be used to calculate the margin of error: $M = 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, where $\hat{p}$ is the proportion of successes in a random sample of size n.

- The sample size is large enough to use this result for estimated margin of error if there are at least 10 of each of the two outcomes.

- The sample size should not exceed 10% of the population.

- As the sample size increases, the margin of error decreases.

Problem Set

1. Different students drew random samples of size 50 from the mystery bag. The number of red chips each drew is given below. In each case, find the margin of error for the proportions of the red chips in the mystery bag.

   a. 10 red chips

   b. 28 red chips

   c. 40 red chips

2. The school newspaper at a large high school reported that 120 out of 200 randomly selected students favor assigned parking spaces. Compute the margin of error. Interpret the resulting interval in context.

3. A newspaper in a large city asked 500 women the following: "Do you use organic food products (such as milk, meats, vegetables, etc.)?" 280 women answered "yes." Compute the margin of error. Interpret the resulting interval in context.

4. The results of testing a new drug on 1000 people with a certain disease found that 510 of them improved when they used the drug. Assume these 1000 people can be regarded as a random sample from the population of all people with this disease. Based on these results, would it be reasonable to think that more than half of the people with this disease would improve if they used the new drug? Why or why not?

5. A newspaper in New York took a random sample of 500 registered voters from New York City and found that 300 favored a certain candidate for governor of the state. A second newspaper polled 1000 registered voters in upstate New York and found that 550 people favored this candidate. Explain how you would interpret the results.

---

6. In a random sample of 1,500 students in a large suburban school, 1,125 reported having a pet, resulting in the interval 0.75 ± 0.022. While in a large urban school, 840 out of 1,200 students reported having a pet, resulting in the interval 0.7 ± 0.026. Because these two intervals do not overlap, there appears to be a difference in the proportion of suburban students owning a pet and the proportion of urban students owning a pet. Suppose the sample size of the suburban school was only 500 but 75% still reported having a pet. Also, suppose the sample size of the urban school was 600 and 70% still reported having a pet. Is there still a difference in the proportion of students owning a pet in suburban schools and urban schools? Why does this occur?

7. Find an article in the media that uses a margin of error. Describe the situation (an experiment, an observational study), and interpret the margin of error for the context.