Estimating Proportions from Samples

Let's estimate population proportions with some data.

10.1 Notice and Wonder: Children's Movies

Five students want to see how many children's movies have human lead characters. Each student takes a sample of children's movies, finds the proportion of movies that have human lead characters, then uses their own result to simulate 100 additional samples. The table shows some of the findings based on the original sample and the simulations.

What do you notice? What do you wonder?

student	number of movies used in the original random sample	estimated proportion	margin of error
A	20	0.3	0.204
B	20	0.2	0.189
C	30	0.3	0.180
D	30	0.233	0.157
E	50	0.2	0.115

 CC BY NC 2024 Illustrative Mathematics®

10.2 Fly Memory

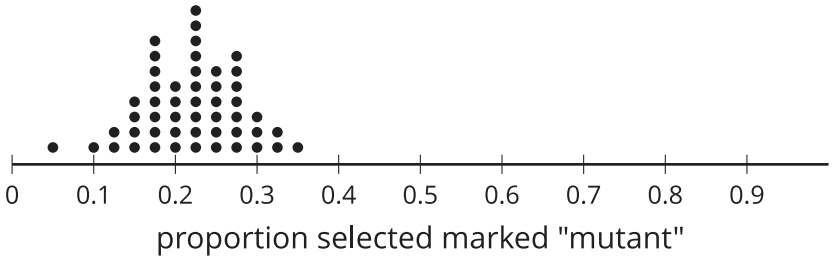
A biologist is breeding fruit flies to include a specific genetic mutation that will be useful in understanding memory in humans. To check whether a fly has the mutation, a DNA sequence is analyzed in a way that kills the fly, so the biologist wants to test only a sample of the flies to estimate the proportion of flies that have the mutation.

The biologist selects 40 flies to sequence at random and finds that 9 of them have the genetic mutation.

1. Based on this sample, estimate the proportion of flies in this group that has the genetic mutation.

2. The scientist is worried that having only one sample may not be reliable for estimating the proportion of flies with the mutation, but does not want to sacrifice more flies to get a larger sample. The proportion from the sample is a good estimate for the population proportion, but it is difficult to understand the possible variability from a single value. Andre has a suggestion for how to better understand the variability:
   a. Assume the sample is representative of the population of flies, and create a simulation that mimics what the scientist found. Andre gets 200 pieces of paper, marks 45 of them as “mutant,” and puts them all in a bag. Because Andre decided to use 200 pieces of paper, why should 45 of them be marked “mutant”? What are some other combinations of total number of pieces of paper and number of papers marked “mutant” that he could use?

b. Andre then simulates the scientist's sample by drawing a slip of paper from the bag and noting whether it is mutant or not, then replacing the paper into the bag and drawing another paper until he has a sample of 40. He repeats this process for 50 trials and creates a dot plot showing the proportion that are mutant from each trial. Use the dot plot to estimate a range of proportions that include about 95% of the proportions from the trials.



c. Andre then finds the standard deviation of this sampling distribution model to be 0.06. What should he report as the margin of error associated with his point estimate?

d. Does your estimate for the range of proportions fit with the values Andre should report?

Are you ready for more?

Suppose the biologist breeds 600 flies. What is the minimum number of flies the biologist should expect to have the mutation, according to Andre's margin of error? What is the maximum that should be expected?

10.3 Finding a Job

Elena and Clare are each working on a project about how high school students are having trouble finding jobs. They each find the proportion of students without jobs from a random sample, use a computer to do 1,000 simulations with the proportion they found, and then report the results.

Elena says, “The proportion of high school students without jobs is about 0.70 with a margin of error of 0.280.”

Clare says, “The proportion of high school students without jobs is about 0.75 with a margin of error of 0.138.”

1. Both students reported the margin of error based on 2 standard deviations from their simulations. What are the mean of the sample and standard deviation of the sampling distribution each student found? For at least 1 student, show your reasoning.

2. Clare and Elena try to figure out why Clare had such a smaller range of values in her report.
   a. First, they consider the proportion they used in the simulations. Elena says, “My simulation used 0.7 as the proportion since I found that proportion in my original sample.” Clare says, “My simulation used 0.75 as the proportion since I found that proportion in my original sample.” The students used different proportions in their simulations. Do you think this is why Clare has a lesser margin of error? Explain your reasoning.

b. They look for more differences in their samples and simulation and discover that Elena surveyed 10 people in her initial sample and Clare surveyed 40 people. Then Elena used samples of size 10 in her simulation and Clare used samples of size 40 in her simulation. Do you think this is why Clare has a lesser margin of error? Explain your reasoning.

Lesson 10 Summary

Although reality doesn't always match up with our estimates, using sample data to estimate a characteristic for a larger group can be very useful, especially when we attach a margin of error to the estimate. A point estimate together with its associated margin of error gives a range of values in which we can expect the actual characteristic for the larger group to be with a certain amount of confidence.

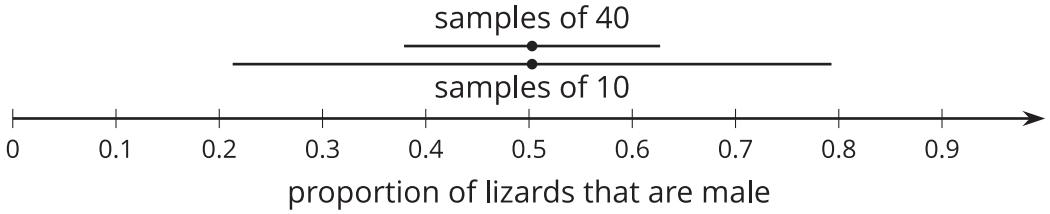
One way to determine a margin of error is by running a simulation based on the data from a sample to create a sampling distribution. When we double the standard deviation of the sampling distribution, we can be about 95% sure that the value for the actual population characteristic will be at most that far away from our point estimate.

For example, a group goes to an island and collects a random sample of 10 lizards, finding that 5 of them are male. This random sample has a proportion of 0.5 males in the group. How close is this likely to be to the actual proportion of lizards that are male on the island? To investigate, we can simulate taking many samples from a population in which the proportion is 0.5 and see how far away from 0.5 the sample proportions tend to be. The sampling distribution of proportions from the simulated samples will give us an idea of how far off our sample estimate of 0.5 might be from the actual population value.

Suppose we simulate taking 30 random samples of 10 lizards from a population with a 0.5 probability of each one being male, and this results in a sampling distribution that has a mean of 0.503 and a standard deviation of 0.145. The sampling distribution is approximately normal in shape, so it is reasonable to think that simulated proportions from the population should be within about 2 standard deviations, or 0.290 ($2 \cdot 0.145 = 0.290$) of the actual population mean.

Based on the simulations and analysis, we expect that the original estimate of 0.5 for the proportion of the population that is male is likely to be within 0.290 of the actual value of the population proportion of lizards that are male. The researchers should report an estimate of 0.5 for the population proportion with an associated margin of error of 0.290.

Later, another group goes to the island and collects a sample of 40 lizards, finding that 20 of them are male. After simulating 30 samples of 40 lizards with a 0.5 probability of each one being male, the sampling distribution has a mean proportion of 0.503 again, but the standard deviation is 0.062. This group should report an estimated proportion of lizards on the island that are male of 0.5 with a margin of error of 0.124 ($2 \cdot 0.062 = 0.124$). This means that they believe their estimate of 0.5 is within 0.124 of the actual population proportion.



In general, the standard deviation of sampling distributions tends to be less with larger samples, so the margin of error reported is less.