Behavior of Sample Proportion: Introduction

Learning Objective: Apply the sampling distribution of the sample proportion (when appropriate). In particular, be able to identify unusual samples from a given population.

The first step to drawing conclusions about parameters based on the accompanying statistics is to understand how sample statistics behave relative to the parameter that summarizes the entire population. We begin with the behavior of sample proportion relative to population proportion (when the variable of interest is categorical). After that, we will explore the behavior of sample mean relative to population mean (when the variable of interest is quantitative).

Behavior of Sample Proportion $\hat{p}$

Example

Approximately 60% of all part-time college students in the United States are female. (In other words, the population proportion of females among part-time college students is p = 0.6.) What would you expect to see in terms of the behavior of a sample proportion of females ($\hat{p}$) if random samples of size 100 were taken from the population of all part-time college students?

As we saw before, due to sampling variability, sample proportion in random samples of size 100 will take numerical values which vary according to the laws of chance: in other words, sample proportion is a **random variable**. To summarize the behavior of any random variable, we focus on three features of its distribution: the center, the spread, and the shape.

Based only on our intuition, we would expect the following:

Center: Some sample proportions will be on the low side—say, 0.55 or 0.58—while others will be on the high side—say, 0.61 or 0.66. It is reasonable to expect all the sample proportions in repeated random samples to average out to the underlying population proportion, .6. In other words, the mean of the distribution of $\hat{p}$ should be p.

Spread: For samples of 100, we would expect sample proportions of females not to stray too far from the population proportion 0.6. Sample proportions lower than 0.5 or higher than 0.7 would be rather surprising. On the other hand, if we were only taking samples of size 10, we would not be at all surprised by a sample proportion of females even as low as 4/10 = 0.4, or as high as 8/10 = 0.8. Thus, sample size plays a role in the spread of the distribution of sample proportion: there should be less spread for larger samples, more spread for smaller samples.

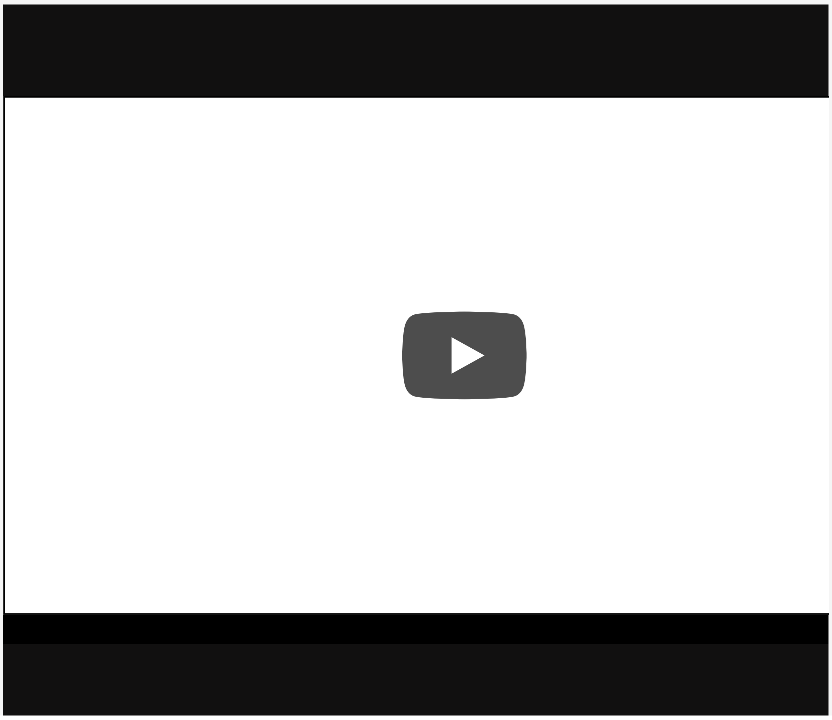
Shape: Sample proportions closest to 0.6 would be most common, and sample proportions far from 0.6 in either direction would be progressively less likely. In other words, the shape of the distribution of sample proportion should bulge in the middle and taper at the ends: it should be somewhat **normal.**

Comment

The **distribution** of the values of the sample proportions ($\hat{p}$) in repeated **samples** is called the **sampling distribution of $\hat{p}$**.

The purpose of the next activity is to check whether our intuition about the center, spread and shape of the sampling distribution of $\hat{p}$ was right via simulations.

Behavior of Sample Proportion 1



Start of transcript. Skip to the end.

In this movie we're going to discuss the behavior of sample proportions by

investigating these two questions: when we collect random samples what patterns

emerge? More specifically, what is the shape, center, and spread of the

0:00 / 4:14 1.50x

distribution of sample proportions? To investigate these questions we're going

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Did I Get This

1/1 point (graded)

In the simulation, when we are building a sampling distribution, what does each dot represent in the graph?

- ○ An individual part-time college student
- ● A random sample of 25 part-time college students ✔
- ○ A specific college's part-time college students

Answer

Correct: Each sample is represented in the sampling distribution with a dot at its $\hat{p}$ value.

Submit

Did I Get This

1/1 point (graded)

In the simulation, we collected thousands of random samples from a population of part-time college students. 60% of this population is female.

What is the mean of the sample proportions? Type in a whole number. Do NOT include a symbol for percentage (%).

60 ✔

60

Answer

Correct: The mean of the sample proportions is the population proportion.

Submit

At this point, we have a good sense of what happens as we take random samples from a population. Our simulation suggests that our initial intuition about the shape and center of the sampling distribution is correct. If the population has a proportion of p, then random samples of the same size drawn from the population will have sample proportions close to p. More specifically, the distribution of sample proportions will have a mean of p. We also observed that for this situation, the sample proportions are approximately normal. We will see later that this is not always the case. But if sample proportions are normally distributed, then the distribution is centered at p. Now we want to use simulation to help us think more about the variability we expect to see in the sample proportions. Our intuition tells us that larger samples will better approximate the population, so we might expect less variability in large samples. In the next walk-through we will use simulations to investigate this idea. After that walk-through, we will tie these ideas to more formal theory.

Behavior of Sample Proportion 2

Start of transcript. Skip to the end.

in this movie we will continue our discussion of the behavior of sample

proportions. In particular, we're going to investigate this question: How does the

size of the sample impact the spread and the sample proportions? To investigate

this question, we're going to return to the familiar context of the

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

Did I Get This

1/1 point (graded)

Compared to small samples, do large samples have more variability, less variability, or about the same?

- less ✔
- more
- about the same

Answer
Correct:
In the simulation, the sample proportions for n = 100 were more tightly grouped about the population proportion.

Submit

Did I Get This

1/1 point (graded)
We collect random samples of 25 students at a time and calculate the proportion of females in each sample. The standard deviation of $\hat{p}$s is approximately 0.10. Which of the following is a plausible standard deviation for samples of 100?

- 0.40
- 0.10
- 0.05 ✔

Answer
Correct:
If the sample size is increased, the standard deviation will decrease because larger samples have less variability.

Submit