Behavior of Sample Proportion: The Sampling Distribution

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

Again, the simulations on the previous page reinforced what makes sense to our intuition. Larger random samples will better approximate the population proportion. When the sample size is large, sample proportions will be closer to p. In other words, the sampling distribution for large samples has less variability. Advanced probability theory confirms our observations and gives a more precise way to describe the standard deviation of the sample proportions. This is described next.

The Sampling Distribution of the Sample Proportion

If repeated random samples of a given size n are taken from a population of values for a categorical variable, where the proportion in the category of interest is p, then the mean of all sample proportions ($\hat{p}$) is the population proportion (p). As for the spread of all sample proportions, theory dictates the behavior much more precisely than saying that there is less spread for larger samples. In fact, the standard deviation of all sample proportions ($\hat{p}$) is exactly $\sqrt{\frac{p(1-p)}{n}}$.

Since sample size n appears in the denominator of the square root, the standard deviation does decrease as sample size increases. Finally, the shape of the distribution of $\hat{p}$ will be approximately normal as long as the sample size n is large enough. The convention is to require both np and n(1 - p) to be at least 10.

We can summarize all of the above by the following:

$\hat{p}$ has a normal distribution with a mean of $\mu_{\hat{p}} = p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ (and as long as np and n(1 - p) are at least 10).

Let's apply this result to our example and see how it compares with our simulation.

In our example, n = 25 (sample size) and p = 0.6. Note that np = 15 ≥ 10 and n(1 - p) = 10 ≥ 10. Therefore we can conclude that $\hat{p}$ is approximately a normal distribution with mean p = 0.6 and standard deviation $\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.6(1-0.6)}{25}} = 0.097$ (which is very close to what we saw in our simulation).

Scenario: Student Loans

According to the National Postsecondary Student Aid Study conducted by the U.S. Department of Education in 2008, 62% of graduates from public universities had student loans.

Learn By Doing

1/1 point (graded)

We randomly sample college graduates from public universities and determine the proportion in the sample with student loans.

For which of the following sample sizes is a normal model a good fit for the sampling distribution of sample proportions? Check all that apply.

- [ ] 10
- [ ] 20
- [x] 30

✔

Answer

Correct:

Both conditions are met when n = 30. np = (30)(0.62) = 18.6 and n(1 - p) = (30)(0.38) = 11.4. Both are greater than 10. So a normal model is a good fit for the sampling distribution of sample proportions when n = 30.

Submit

Learn By Doing

1/1 point (graded)

If we randomly sample 50 students at a time, what will be the mean of the distribution of sample proportions? Make sure to provide your answer as a proportion. Report your answer to TWO decimal places.

0.62

0.62

Submit

Learn By Doing

1/1 point (graded)

If we randomly sample 50 students at a time, what is the standard deviation of the distribution of sample proportions? Make sure to provide your answer as a proportion. Round your answer to TWO decimal places.

0.07

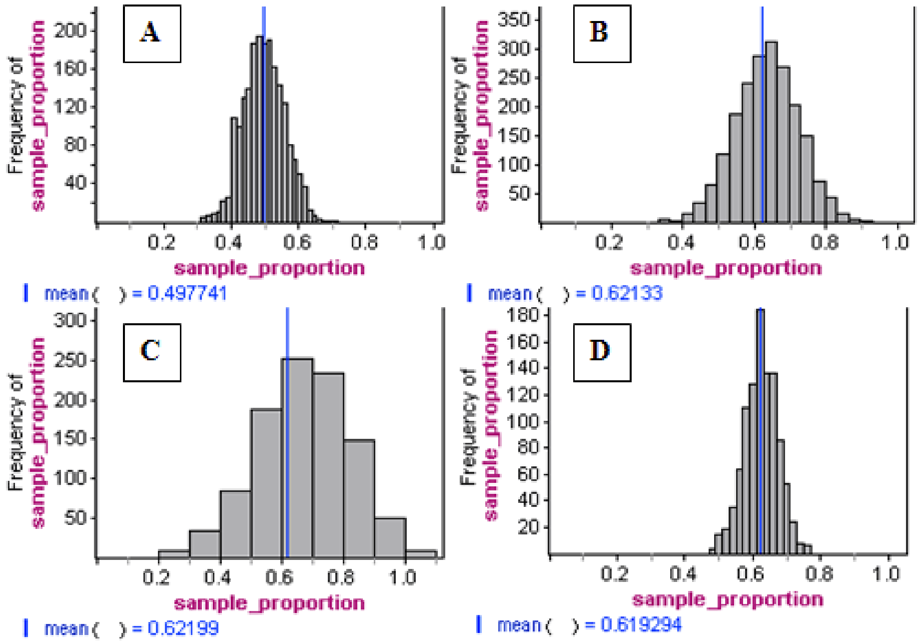
0.07

Submit

Learn By Doing

1/1 point (graded)

Which distribution is a plausible representation of the sampling distribution for random samples of 30 students?



- A
- B ✔
- C
- D

Answer
Correct:
The mean of the sampling distribution should be p = 0.62 with standard deviation sqrt(p(1 - p) / n) = sqrt(0.62(1 - 0.62) / 30) = 0.09. Typical values should fall within one standard deviation of the mean, from about 0.53 to 0.71. This distribution fits this description, as shown in the graph.

Submit

Learn By Doing

1/1 point (graded)

According to the official M&M website, 24% of the plain milk chocolate M&M's produced by Mars Corporation are blue. Annie buys a large family-size bag of M&M's. Sarah buys a small fun-size bag. Which bag is more likely to have more than 40% blue M&M's?

- Annie, because there are more M&M's in her bag, so she will have more blue ones.
- Annie, because there is more variability in the proportion of blues among larger samples.
- Sarah, because there is more variability in the proportion of blues among smaller samples. ✔
- Both have the same chance because they are both random samples.

Answer
Correct:
There is more variability in small samples, so it is more likely to get sample results further from p = 0.24 with a small bag.

Submit

Learn By Doing

1/1 point (graded)
Imagine that you have a very large barrel that contains tens of thousands of M&M's. According to the official M&M website, 20% of the M&M's produced by the Mars Corporation are orange. 5 students each take a random sample of 50 M&M's and record the percentage of orange in each sample. Which sequence is the most plausible for the percentage of orange candies obtained in these 5 samples?

- 20%, 20%, 20%, 20%, 20%
- 15%, 25%, 22%, 20%, 28% ✔
- 5%, 80%, 8%, 65%, 70%
- Each of the sequences is equally plausible.

Answer
Correct:
We expect $\hat{p}$s within 1 standard deviation of p = 0.20 to be most common. The standard deviation is about 0.06. 4 of the 5 $\hat{p}$s in this sequence are within 0.06 of p = 0.20.

Submit