

2.1: Sampling Distribution of the Sample Mean

Inferential testing uses the sample mean (\bar{x}) to estimate the population mean (μ). Typically, we use the data from a single sample, but there are many possible samples of the same size that could be drawn from that population. As we saw in the previous chapter, the sample mean (\bar{x}) is a *random* variable with its own distribution.

- The distribution of the sample mean will have a mean equal to μ .
- It will have a standard deviation (standard error) equal to $\frac{\sigma}{\sqrt{n}}$

Because our inferences about the population mean rely on the sample mean, we focus on the distribution of the sample mean. Is it normal? What if our population is not normally distributed or we don't know anything about the distribution of our population?

The Central Limit Theorem (CLT)

The Central Limit Theorem states that the sampling distribution of the sample means will approach a normal distribution as the sample size increases.

So if we do not have a normal distribution, or know nothing about our distribution, the CLT tells us that the distribution of the sample means (\bar{X}) will become normal distributed as n (sample size) increases. How large does n have to be? A general rule of thumb tells us that $n \geq 30$.

The Central Limit Theorem tells us that regardless of the shape of our population, the sampling distribution of the sample mean will be normal as the sample size increases.

Sampling Distribution of the Sample Proportion

The population proportion (p) is a parameter that is as commonly estimated as the mean. It is just as important to understand the distribution of the sample proportion, as the mean. With proportions, the element either has the characteristic you are interested in or the element does not have the characteristic. The sample proportion (\hat{p}) is calculated by

$$\hat{p} = \frac{x}{n} \quad (2.1.1)$$

where x is the number of elements in your population with the characteristic and n is the sample size.

Example 2.1.1: sample proportion

You are studying the number of cavity trees in the Monongahela National Forest for wildlife habitat. You have a sample size of $n = 950$ trees and, of those trees, $x = 238$ trees with cavities. Calculate the sample proportion.



A naturally formed tree hollow at the base of the tree. (CC BY 2.0; Lauren "Lolly" Weinhold).

Solution

This is a simple application of Equation 2.1.1:

$$\hat{p} = \frac{238}{950} = 0.25$$

The distribution of the sample proportion has a mean of

$$\mu_{\hat{p}} = p \quad (2.1.2)$$

and has a standard deviation of

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}. \quad (2.1.3)$$

The sample proportion is normally distributed if n is very large and \hat{p} is not close to 0 or 1. We can also use the following relationship to assess normality when the parameter being estimated is p , the population proportion:

$$n\hat{p}(1 - \hat{p}) \geq 10 \quad (2.1.4)$$

This page titled [2.1: Sampling Distribution of the Sample Mean](#) is shared under a [CC BY-NC-SA 3.0](#) license and was authored, remixed, and/or curated by [Diane Kiernan \(OpenSUNY\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.