

Lecture 6:
Sampling Distributions
STAT 310, Spring 2023

- ▶ A **parameter** is a numerical characteristic of the population (fixed number that is usually unknown).
- ▶ A **statistic** is a numerical characteristic of the sample (varies depending on sample).
- ▶ The statistic is also referred to as a **point estimate**, since it is our best guess at the value of a population parameter.

Notation for the proportion:

- ▶ p : population proportion
- ▶ \hat{p} : sample proportion

Notation for the mean:


- ▶ μ : population mean
- ▶ \bar{x} : sample mean

The sample size is denoted as n

Example

A recent Gallup poll found that 63% of Americans are dissatisfied with US gun laws. The survey results were based on a random sample of 1,011 adults.¹

- (a) What is the sample proportion?
- (b) What is the sample size?
- (c) Describe the population proportion?
- (d) Suppose another poll is conducted with a different random sample of adults. Would you expect the sample proportion to be the same, or slightly different?

¹<https://news.gallup.com/poll/470588/dissatisfaction-gun-laws-hits-new-high.aspx> 

- ▶ **Sampling error** refers to how much a statistic, such as the sample proportion, will vary from one random sample to the next.
- ▶ For example, the Gallup poll reported a sampling error of ± 4 percentage points. This means that the population proportion of Americans that are dissatisfied with US gun laws is likely between 59% and 67%.

A **sampling distribution** is the distribution of a statistic when repeatedly taking random samples from a population.

- ▶ In real-world applications, we never actually observe the sampling distribution, since we usually take a single random sample.
- ▶ However, it is useful to always think of a statistic, such as the sample proportion, as coming from such a hypothetical distribution.
- ▶ The concept of a sampling distribution is very important when trying to quantify sampling error.

Central Limit Theorem (CLT)

The sampling distribution for \hat{p} follows an approximate normal distribution centered around the population proportion p , and with standard error $\sqrt{p(1-p)/n}$.

Conditions for CLT

The following conditions should be met to apply the CLT:

- ▶ The data come from a simple random sample. This is called the **independence condition** since it implies that the individuals or cases in the data are unrelated.
- ▶ $np \geq 10$ and $n(1 - p) \geq 10$. This is sometimes called the **success-failure condition** since np can be interpreted as the expected number of successes and $n(1 - p)$ the expected number of failures.

Example

Suppose that the population proportion of Americans who support the expansion of solar energy is $p = 0.88$, and $n = 1000$ Americans are randomly sampled.

- (a) What is the mean, or center, of the sampling distribution for \hat{p} ?
- (b) What is the standard error of the sampling distribution for \hat{p} ?
- (c) What distribution does \hat{p} follow?
- (d) Are the conditions for the CLT satisfied?

Example

Suppose that the population proportion of Americans who support the expansion of solar energy is $p = 0.88$, and $n = 1000$ Americans are randomly sampled. What is the probability that the sample proportion \hat{p} will be greater than 0.9?

