KIN 610: Quantitative Methods in Kinesiology

Chapter 8: Probability and Sampling Error

Ovande Furtado Jr., PhD.

2026-02-11

1 FYI

This presentation is based on the following books. The references are coming from these books unless otherwise specified.

Main sources:

- Moore, D. S., Notz, W. I., & Fligner, M. (2021). *The basic practice of statistics* (9th ed.). W.H. Freeman.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.
- Furtado, O., Jr. (2026). *Statistics for movement science: A hands-on guide with SPSS* (1st ed.). https://drfurtado.github.io/sms

ClassShare App

You may be asked in class to go to the ClassShare App to answer questions.

- https://classsharedrfurtado.netlify.app/

SPSS Tutorial

- SPSS Tutorial: Sampling and the Central Limit Theorem

2 Intro Question

- If you took 100 different random samples from the same population, would you expect all the sample means to be exactly the same? Why or why not?

Click to reveal answer

No! Each sample would include different individuals, so sample means would naturally vary. This variability is called **sampling error**, and understanding it is the foundation of statistical inference.

- In this chapter, we'll explore how sample statistics vary from sample to sample, introduce the **Central Limit Theorem**, and learn how **confidence intervals** and **probability** help us make inferences about populations from sample data.

3 Learning Objectives

By the end of this chapter, you should be able to:

- Define probability and explain its role in statistical inference
- Explain sampling error and why sample statistics vary
- Describe the sampling distribution of the mean
- State the Central Limit Theorem and explain its importance
- Calculate and interpret the standard error of the mean
- Construct and interpret confidence intervals
- Distinguish between point estimates and interval estimates
- Apply probability concepts to movement science research

4 Symbols

Symbol	Name	Pronunciation	Definition
μ	Population mean	"myoo"	True average of the population
$\bar{x}$	Sample mean	"x bar"	Average of the sample
σ	Population standard deviation	"sigma"	Population variability
s	Sample standard deviation	"s"	Sample variability
n	Sample size	"n"	Number of observations in a sample
SE	Standard error	"standard error"	Standard deviation of the sampling distribution

Symbol	Name	Pronunciation	Definition
CI	Confidence interval	"C.I."	Range of plausible values for a parameter
$P(A)$	Probability of event A	"probability of A"	Likelihood of event A occurring
z	Z-score	"zee"	Number of standard deviations from the mean

5 Introduction: From Samples to Populations

Statistical inference is the process of drawing conclusions about a **population** based on information from a **sample**[1].

- We rarely have access to entire populations
- Instead, we collect a sample and use it to **estimate** population parameters
- The key challenge: How confident can we be in our estimates?
- This chapter introduces the foundational concepts that make inference possible

- Opening: We almost never study entire populations—we rely on samples.
- Key question: How do we go from what we observe (sample) to what we want to know (population)?
- The answer involves understanding probability, sampling distributions, and the Central Limit Theorem.

6 What is Probability?

Probability quantifies the likelihood that an event will occur, expressed as a number between 0 and 1[1].

Key concepts:

- $P(A) = 0$: Event A is **impossible**
- $P(A) = 1$: Event A is **certain**
- $P(A) = 0.5$: Event A has a **50-50 chance**

Types of probability:

1. **Classical**: Based on equally likely outcomes (e.g., coin flip: $P(\text{heads}) = 0.5$)

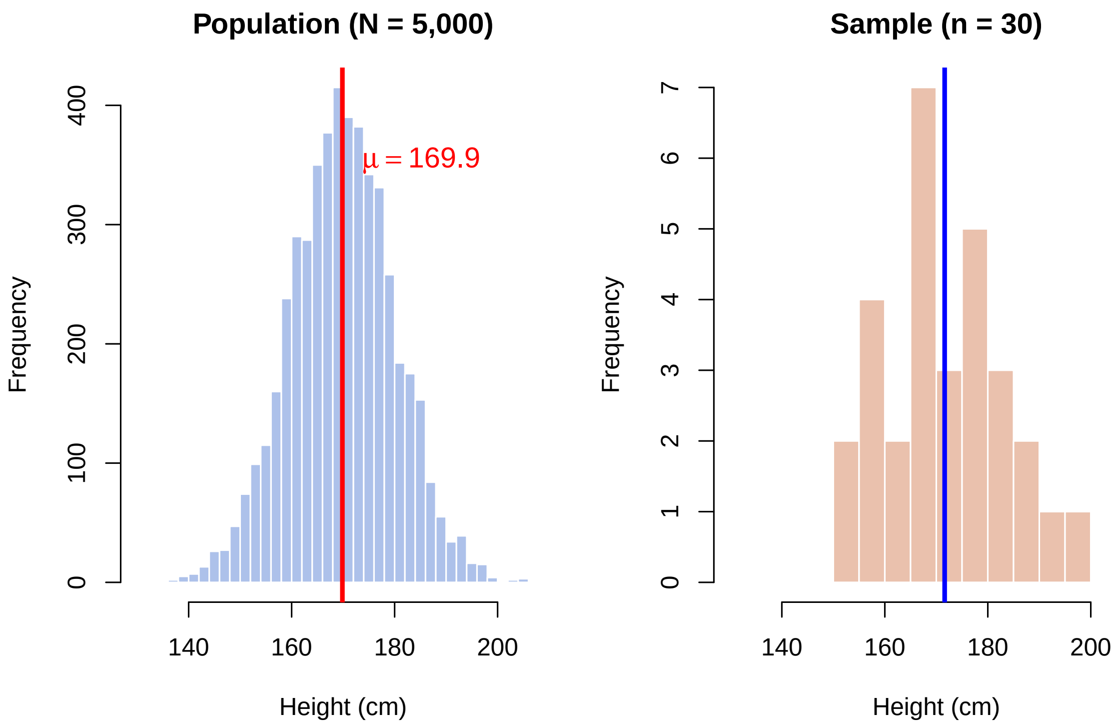

Figure 1: Relationship between population and sample

2. **Relative frequency**: Based on long-run proportions from data (e.g., 68% of athletes jump above 40 cm)
3. **Subjective**: Based on personal judgment or expertise

Movement Science example:

If vertical jump heights are normally distributed with $\mu = 50$ cm and $\sigma = 8$ cm, what is the probability of jumping higher than 58 cm?

$$z = \frac{58 - 50}{8} = 1.0$$

$P(X > 58) = 1 - 0.8413 = 0.1587 \rightarrow$ **15.87%**

Real-World Context

In a study of 200 college athletes, approximately 32 would be expected to jump higher than 58 cm (15.87% $\times$ 200   32).

- Emphasize: Probability is the mathematical foundation for all statistical inference.
- Connection to z-scores: Students already learned z-scores in Ch. 6—now we use them for probability.

7 Sampling Error

Sampling error is the natural, unavoidable difference between a sample statistic and the true population parameter[1,2].

Key points:

- Even with perfect random sampling, sample means will **differ from** the population mean
- This is **not a mistake** — it's a natural consequence of randomness
- Different samples from the same population will produce **different statistics**
- Sampling error decreases as sample size increases

Formula:

$$\text{Sampling Error} = \bar{x} - \mu$$

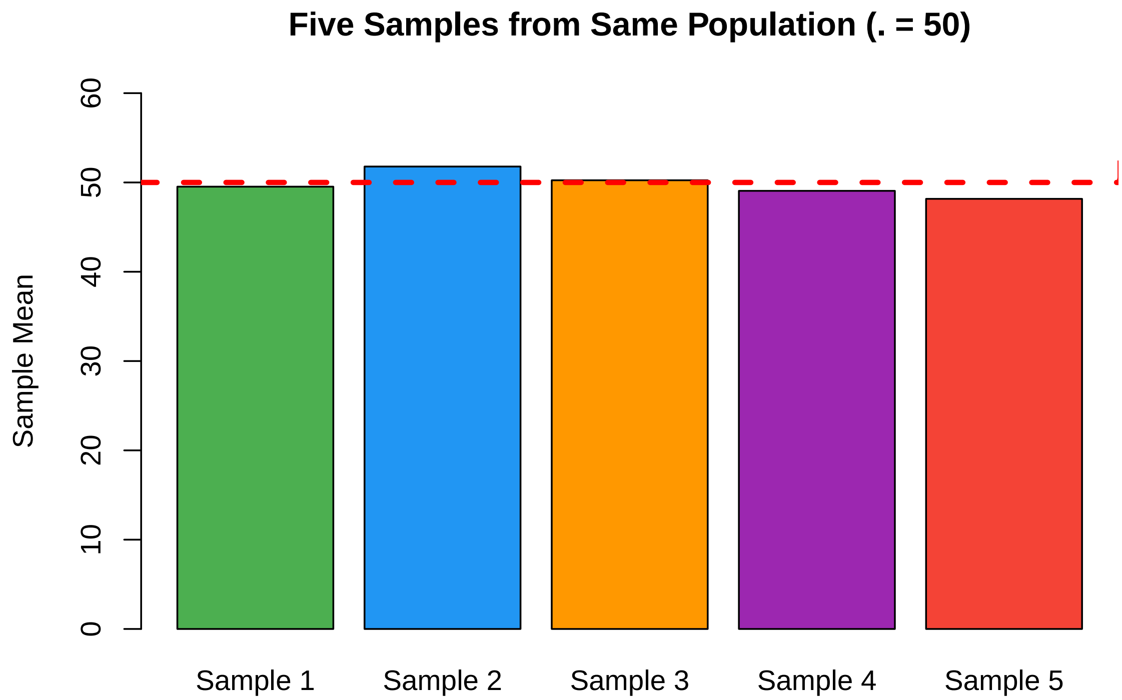

Figure 2: Five random samples from the same population show different means

> **Important**
>
> Sampling error is **not an error** in the colloquial sense — it is the expected variability that arises from using a sample to estimate a population parameter.

- Emphasize: "Sampling error" is a misleading name — it's not a mistake, it's natural variability.
- Example: If the population mean vertical jump is 50 cm, one sample might give 48.5 cm, another 51.2 cm.
- Key question: "How much variability should we expect?" → This leads to the sampling distribution.

8 The Sampling Distribution

The **sampling distribution** is the distribution of a statistic (like the sample mean) across **all possible samples** of the same size from a population[1].

Key properties:

1. **Center**: The mean of the sampling distribution equals the population mean ($\mu_{\bar{x}} = \mu$)
2. **Spread**: The standard deviation is the **standard error** ($SE = \sigma/\sqrt{n}$)
3. **Shape**: Becomes approximately normal as n increases (Central Limit Theorem)

> **Think of it this way**
>
> If you could take an infinite number of samples (each of size n) from the same population and calculate the mean of each, the distribution of those means would form the **sampling distribution**.

- This is the most important concept in this chapter.
- The sampling distribution is theoretical — we don't actually take infinite samples, but the concept allows us to quantify uncertainty.

9 Standard Error of the Mean

The **standard error (SE)** is the standard deviation of the sampling distribution — it tells us how much sample means vary from sample to sample[1,2].

Key insights:

- SE **decreases** as n increases (larger samples → more precise estimates)

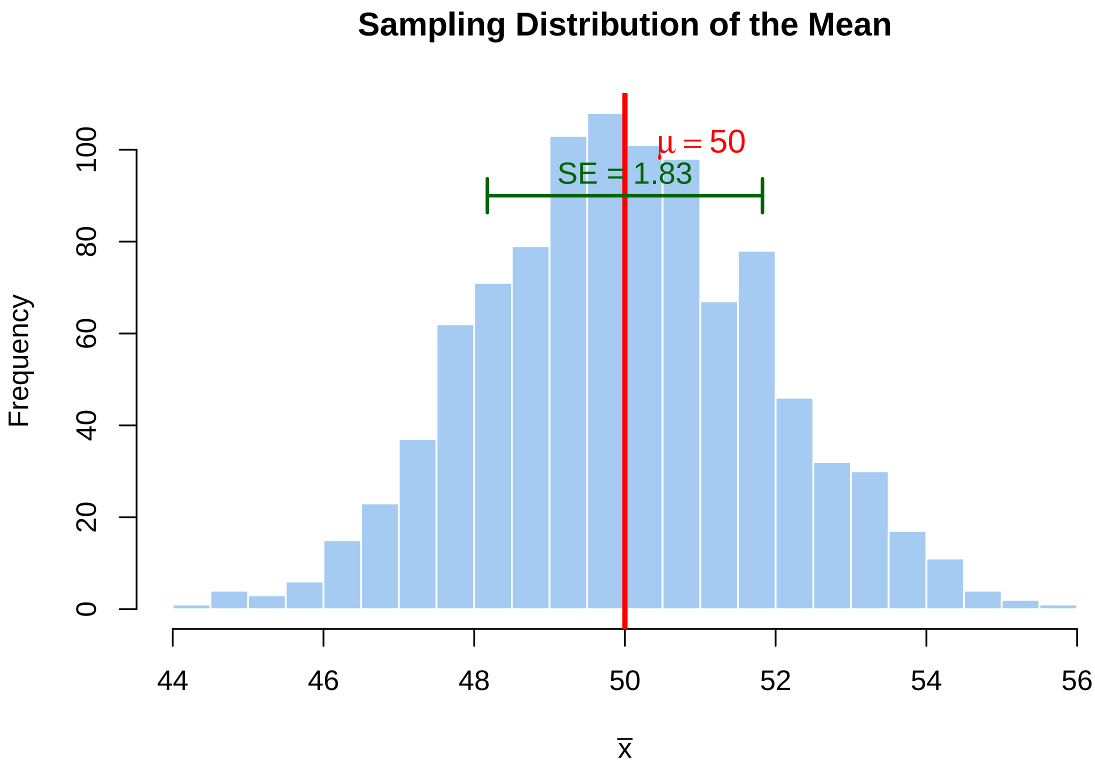

Figure 3: Sampling distribution of the mean (1000 samples of n = 30)

$$SE = \frac{\sigma}{\sqrt{n}}$$

Equation 1: Standard error formula

- SE depends on **population variability** (σ) and **sample size** (n)
- When σ is unknown, we estimate it with s: $SE = s/\sqrt{n}$

Example: If $\sigma = 10$ cm:

n	SE
10	3.16 cm
30	1.83 cm
100	1.00 cm
400	0.50 cm

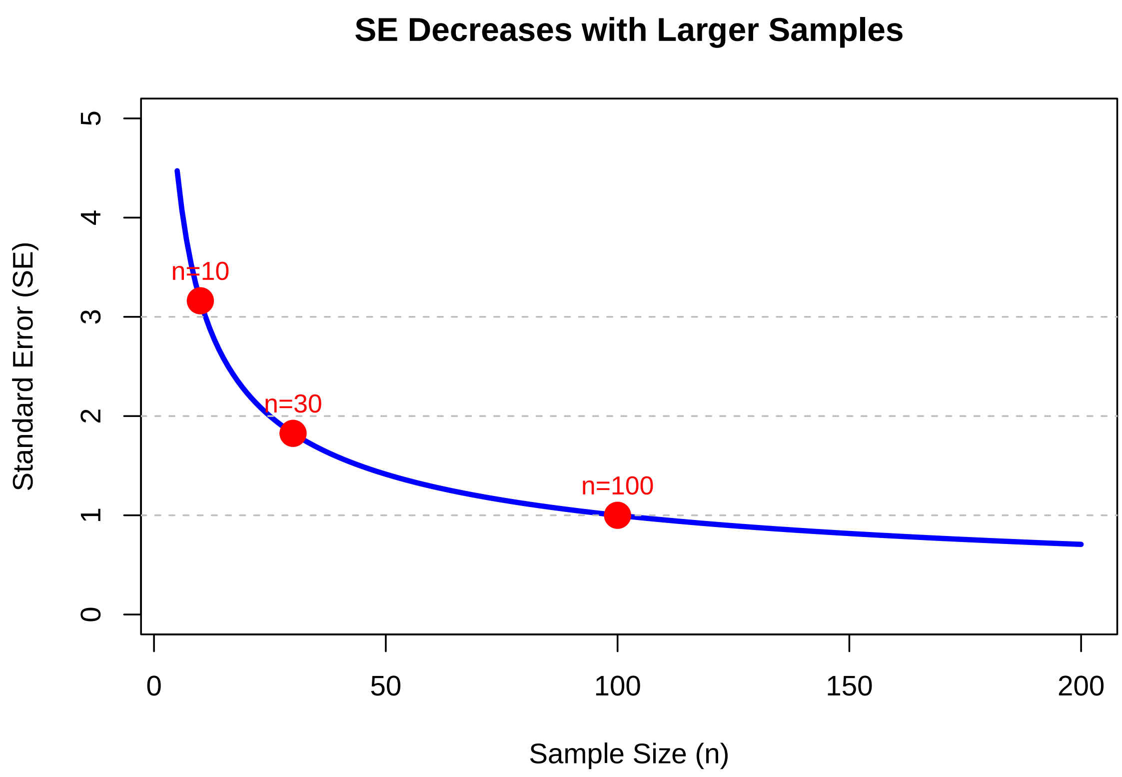

Figure 4: Standard error decreases as sample size increases

> **!** Important
>
> **Diminishing returns**: Doubling the sample size does NOT halve the SE — it reduces it by a factor of $\sqrt{2} \approx 1.41$. Going from $n = 25$ to $n = 100$ ($4\times$) only halves the SE.

- Teaching tip: Ask students "If you want to halve the SE, how many times larger does your sample need to be?" (Answer: 4 times)
- This has practical implications for study design and cost-effectiveness.

10 The Central Limit Theorem

The **Central Limit Theorem (CLT)** is one of the most important theorems in statistics[1,2].

> **CLT**: Regardless of the shape of the population distribution, the sampling distribution of the mean approaches a **normal distribution** as the sample size increases.

Conditions:

1. Random sampling from the population
2. Sample size is sufficiently large ($n \geq 30$ as a common rule of thumb)
3. Observations are independent

Implications:

- Even if the population is skewed, the distribution of **sample means** will be approximately normal
- This justifies using normal-based methods (z-tests, confidence intervals) for inference

> **!** Important
>
> The CLT explains **why statistics works**: Even when we don't know the shape of the population, we can make valid inferences about the mean as long as our sample is large enough.

- This is the theoretical justification for most parametric statistics.
- The population can be any shape — uniform, skewed, bimodal — and the CLT still applies.
- Rule of thumb: n 30 is usually sufficient, but more skewed populations may need larger samples.

11 Confidence Intervals

A **confidence interval (CI)** is a range of plausible values for a population parameter, constructed from sample data[1,2].

Formula for 95% CI:

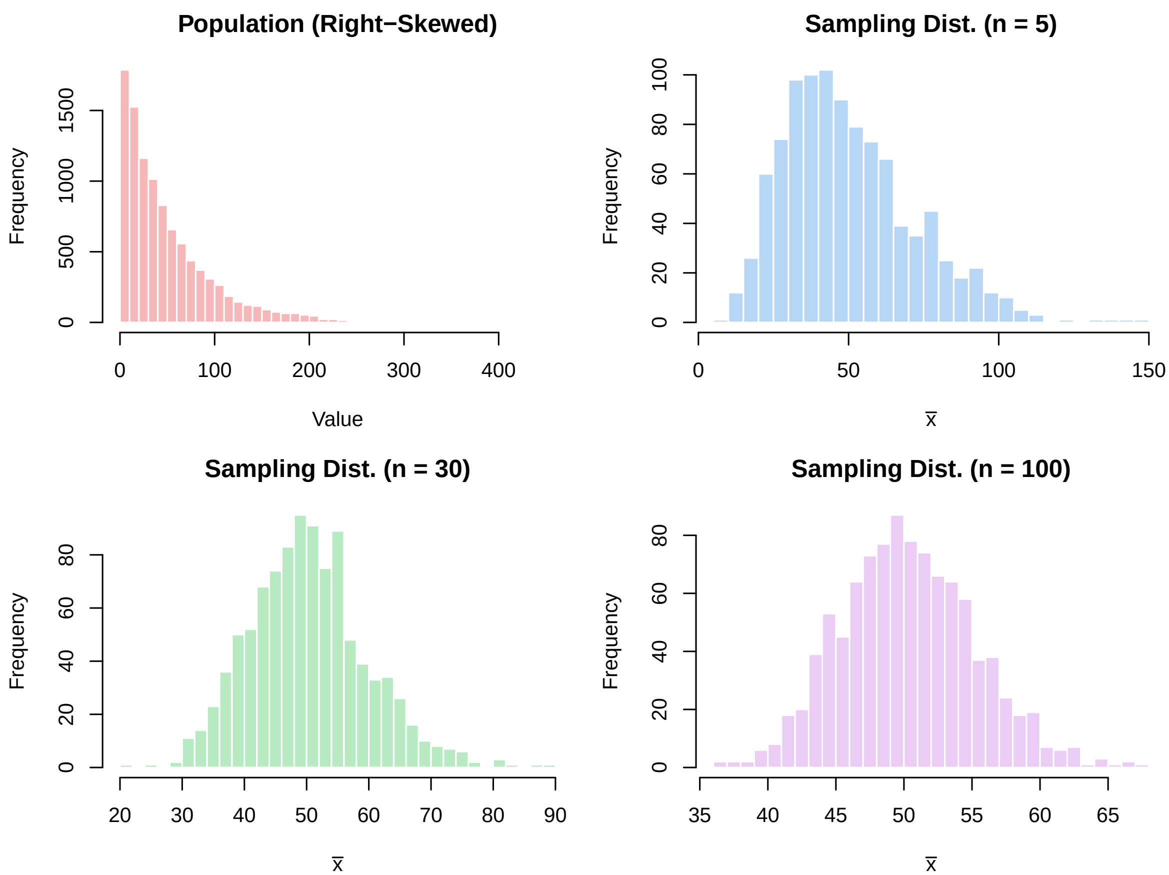

Figure 5: CLT: Population is right-skewed, but sampling distribution of means becomes normal

$$CI_{95\%} = \bar{x} \pm 1.96 \times SE$$

Equation 2: 95% confidence interval formula

Components:

- $\bar{x}$: Sample mean (our best estimate)
- 1.96: Z-value for 95% confidence
- SE: Standard error ($\sigma/\sqrt{n}$ or $s/\sqrt{n}$)

Example: $\bar{x} = 53$ cm, $SE = 1.1$ cm

$$CI_{95\%} = 53 \pm 1.96(1.1) = [50.84, 55.16]$$

> ❗ Important
>
> **Interpretation**: "We are 95% confident that the true population mean falls between 50.84 and 55.16 cm." This means that if we repeated this process many times, about 95% of our intervals would capture the true .

- Common misconception: It does NOT mean "there's a 95% probability   is in this interval."
- The parameter is fixed — it's the intervals that vary.
- The plot shows that 1 of 20 intervals (5%) misses , consistent with 95% coverage.

12 Factors Affecting Confidence Interval Width

The width of a confidence interval is influenced by three factors[1]:

1. Sample size (n):

- Larger $n \to$ narrower intervals (more precision)
- $SE = \sigma/\sqrt{n}$, so larger n reduces SE

2. Variability (σ or s):

- More variable data $\to$ wider intervals
- Less control over this factor

3. Confidence level:

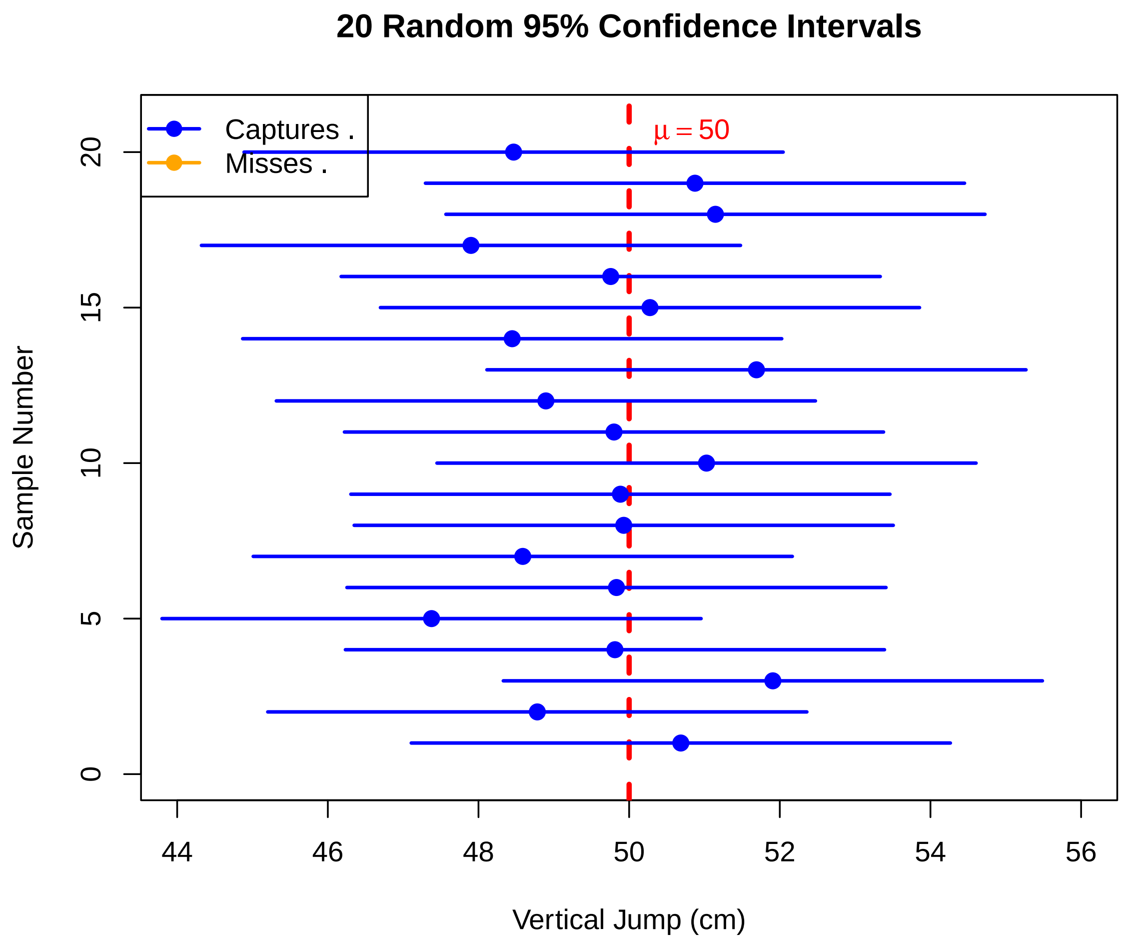

Figure 6: 20 confidence intervals: Most capture , but some miss it

- Higher confidence → wider intervals
- 99% CI is wider than 95% CI (z = 2.576 vs. 1.96)

Confidence Level	Z-value	Interval Width
90%	1.645	Narrowest
95%	1.960	Moderate
99%	2.576	Widest

Trade-off

Higher confidence requires wider intervals (less precision). The 95% level is a convention that balances confidence and precision.

- Teaching tip: "You can always get 100% confidence — just say the mean could be anything! But that's useless."
- The key insight: Increasing confidence means accepting less precision.

13 Point Estimates vs. Interval Estimates

Point estimate: A single value used to estimate a population parameter

- $\bar{x} = 53$ cm
- Simple but provides **no information about precision**
- Could be very close to or far from it

Interval estimate: A range of plausible values

- 95% CI: [50.84, 55.16] cm
- Communicates **both the estimate and its uncertainty**
- Preferred in modern statistical reporting

APA Recommendation

The American Psychological Association (APA) recommends reporting confidence intervals alongside or instead of p-values[?].
Why? Confidence intervals provide more information:

- The point estimate (center of the CI)
- The precision of the estimate (width of the CI)
- The direction and magnitude of the effect

- Modern best practice: Always report CIs, not just point estimates or p-values.
- CIs tell a more complete story than p-values alone.

14 Summary: Key Takeaways

1. **Probability** quantifies uncertainty and is the foundation of statistical inference
2. **Sampling error** is natural variability, not a mistake — it's expected when using samples
3. **The sampling distribution** shows how sample statistics vary across all possible samples
4. **Standard error** measures the precision of sample estimates ($SE = \sigma/\sqrt{n}$)
5. **Central Limit Theorem**: Sampling distribution of the mean is approximately normal for large n, regardless of population shape
6. **Confidence intervals** provide a range of plausible values for population parameters
7. **Larger samples** lead to smaller standard errors and narrower confidence intervals
8. **95% confidence** means that 95% of intervals from repeated sampling would capture the true parameter

> **!** Important
>
> These concepts form the foundation for **hypothesis testing** (Chapter 10), where we'll use probability to make formal decisions about whether observed differences are real or due to chance.

15 Practice Questions

1. What is the difference between a population parameter and a sample statistic?
2. If $\sigma = 15$ and $n = 25$, what is the standard error?
3. Explain the Central Limit Theorem in your own words. Why is it so important?
4. A 95% CI is [42, 58] cm. What does this mean?
5. How does increasing sample size affect the confidence interval width?
6. If you took 200 random 95% CIs, how many would you expect to miss the true ?
7. What is the relationship between probability and statistical inference?
8. Why is the sample mean considered a good estimator of the population mean?

16 References

1. Moore, D. S., McCabe, G. P., & Craig, B. A. (2021). *Introduction to the practice of statistics* (10th ed.). W. H. Freeman; Company.
2. Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

3. Furtado, O., Jr. (2026). *Statistics for movement science: A hands-on guide with SPSS* (1st ed.). https://drfurtado.github.io/sms/