Interval Estimation

Lesson 11

Digital Finance

Outline

1. Confidence Intervals
2. Factors Affecting CI Width
3. Summary
4. Quiz

Why Intervals?

Point estimates alone are incomplete:

- $\bar{x} = 10.5$ – but how precise?
- Need to quantify uncertainty

Confidence interval:

- Range of plausible values
- Comes with a confidence level
- Example: 95% CI for μ: (9.2, 11.8)

Intervals communicate uncertainty; points don't.

Constructing a CI for μ

Recipe: CI = (sample mean) $\pm$ (margin of error)
Margin of error = (z-value) $\times$ (standard error)

- 95% CI: use z = 1.96 (captures middle 95% of normal curve)
- 99% CI: use z = 2.576 (wider to capture more)

Example: $\bar{x} = 50$, $\sigma = 10$, $n = 100$
95% CI: $50 \pm 1.96 \times \frac{10}{\sqrt{100}} = 50 \pm 1.96 = (48.04, 51.96)$

Higher confidence = wider interval. z-values come from normal distribution tables.

When σ is Unknown

Problem: We usually don't know the true σ – we estimate it from sample (S).
Solution: Use the t-distribution instead of normal.

- t is like normal but with "fatter tails"
- Accounts for extra uncertainty from estimating σ
- Small samples = fatter tails = wider CIs

Degrees of freedom (df) $= n - 1$: how much information we have.
Rule of thumb: For $n \geq 30$, t and z are nearly identical.

Use t-table or software. df = n - 1 because we used 1 parameter ($\bar{X}$) from data.

Interpretation of CI

Correct interpretation:

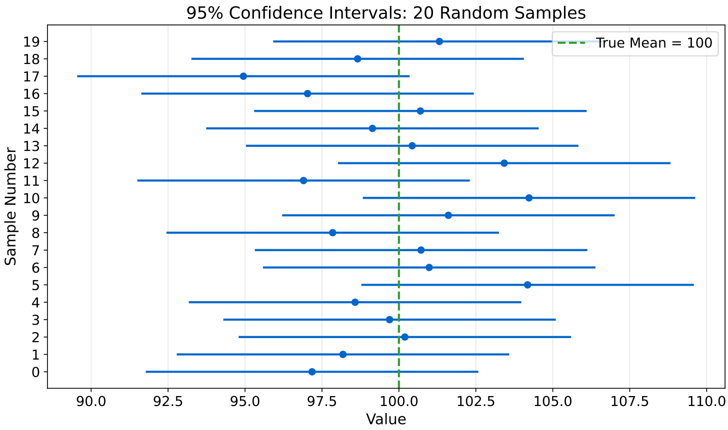
- If we repeat sampling many times
- 95% of the intervals will contain μ
- This specific interval either contains μ or it doesn't

Common misconception:

- NOT "95% probability μ is in this interval"
- μ is fixed; interval is random

Confidence refers to the procedure, not this specific interval.

Confidence Intervals: Visualization



About 95% of 95% CIs capture the true mean.

What Affects CI Width?

CI width $= 2\times$ margin of error

Wider intervals when:

- Higher confidence level (99% vs 95%)
- Smaller sample size
- Higher variability in data

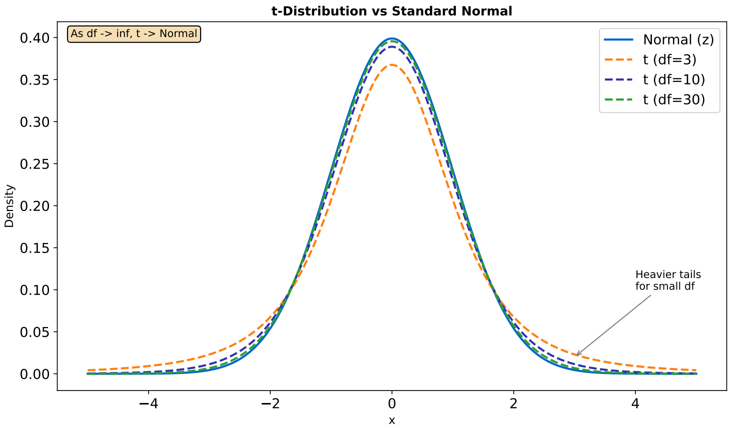
Margin of error:

$$ME = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

To cut ME in half: Quadruple sample size!

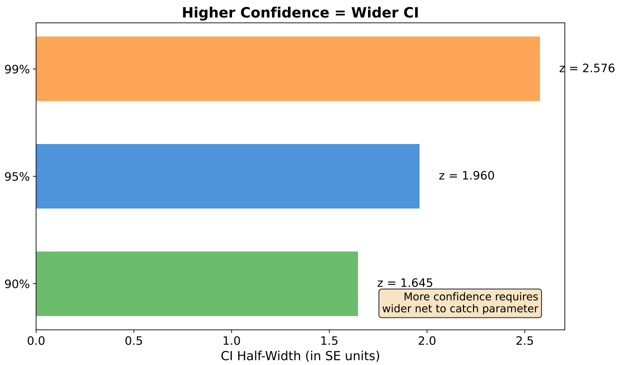
Precision improves slowly with sample size.

t-Distribution vs Normal



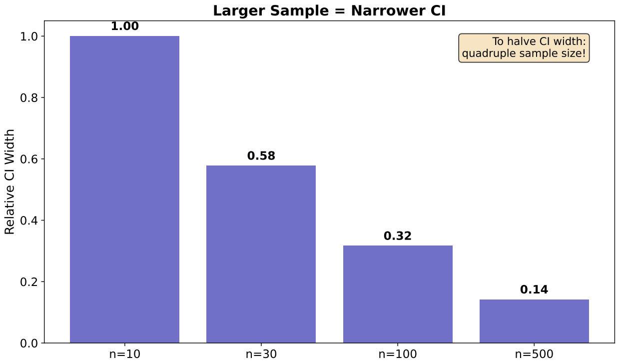
t has heavier tails; approaches normal as df increases.

Effect of Confidence Level



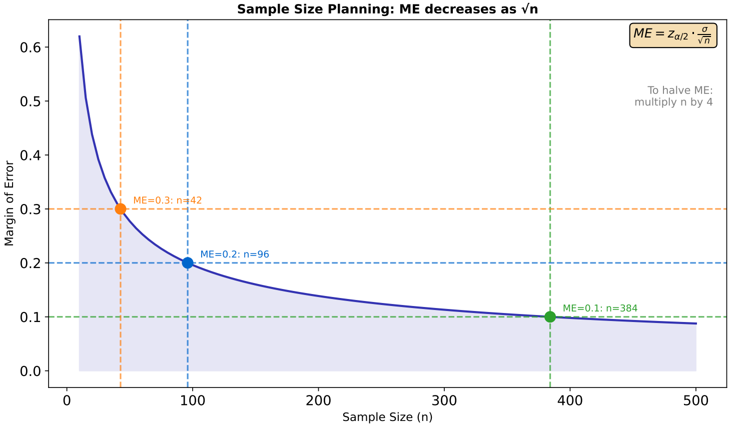
Higher confidence requires wider interval to catch parameter.

Effect of Sample Size



To halve CI width, quadruple sample size.

Sample Size Planning



Larger samples give more precise estimates.

Summary

Confidence intervals:

- Quantify uncertainty in estimates
- Structure: Point estimate $\pm$ Margin of error

Key formulas:

- Known σ: Use z-distribution
- Unknown σ: Use t-distribution

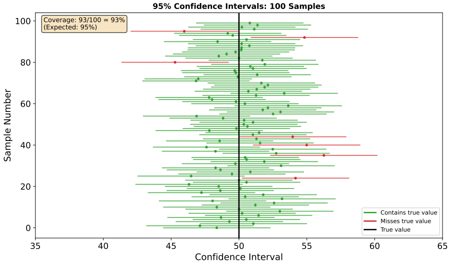
Interpretation:

- Confidence $=$ long-run coverage probability
- Not probability parameter is in interval

Next lesson: Hypothesis Testing

Coverage Simulation

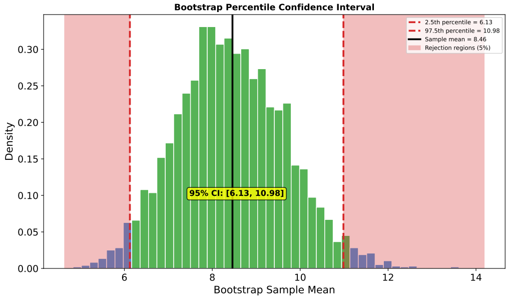
Coverage = fraction of CIs that actually contain the true parameter.
What this chart shows: We took 100 samples, built a 95% CI for each. About 95 of them (green) contain the true μ; about 5 (red) miss it.



"95% confidence" means 95% coverage in repeated sampling.

Bootstrap Confidence Interval

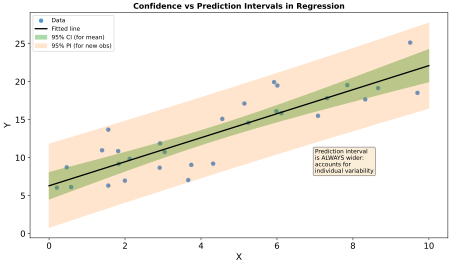
Bootstrap idea: Resample your data (with replacement) many times to approximate the sampling distribution.



CI = middle 95% of bootstrap distribution. Works without assuming normality!

Prediction vs Confidence Interval

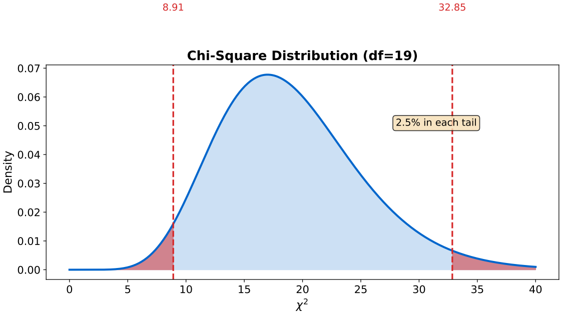
Confidence interval: Where is the true mean? (uncertainty about μ)
Prediction interval: Where will the next observation fall? (includes individual variability)
Key: Prediction intervals are ALWAYS wider than confidence intervals.



CI = uncertainty about mean. PI = uncertainty about individuals.

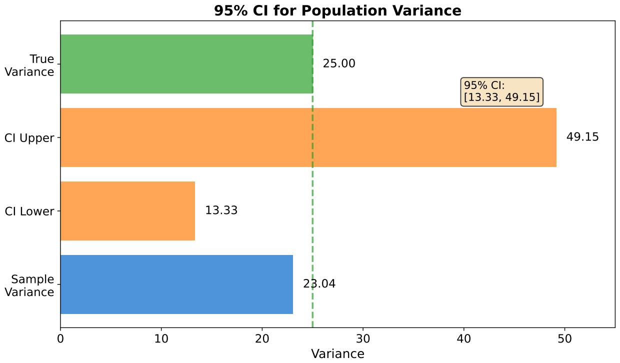
Chi-Square Distribution

Chi-square distribution: Used for CI on variance (not mean).
Why different? Variance is always positive, so we need a distribution that's positive and right-skewed.



$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$

Confidence Interval for Variance



CI bounds computed using chi-square critical values.