# Confidence Intervals

# TERMS

**Point Estimate** is a single value calculated from a sample of data that is used to estimate the value of an unknown parameter of a population.

**Confidence Interval** is a range of values that likely contains the true population parameter with a certain level of confidence.

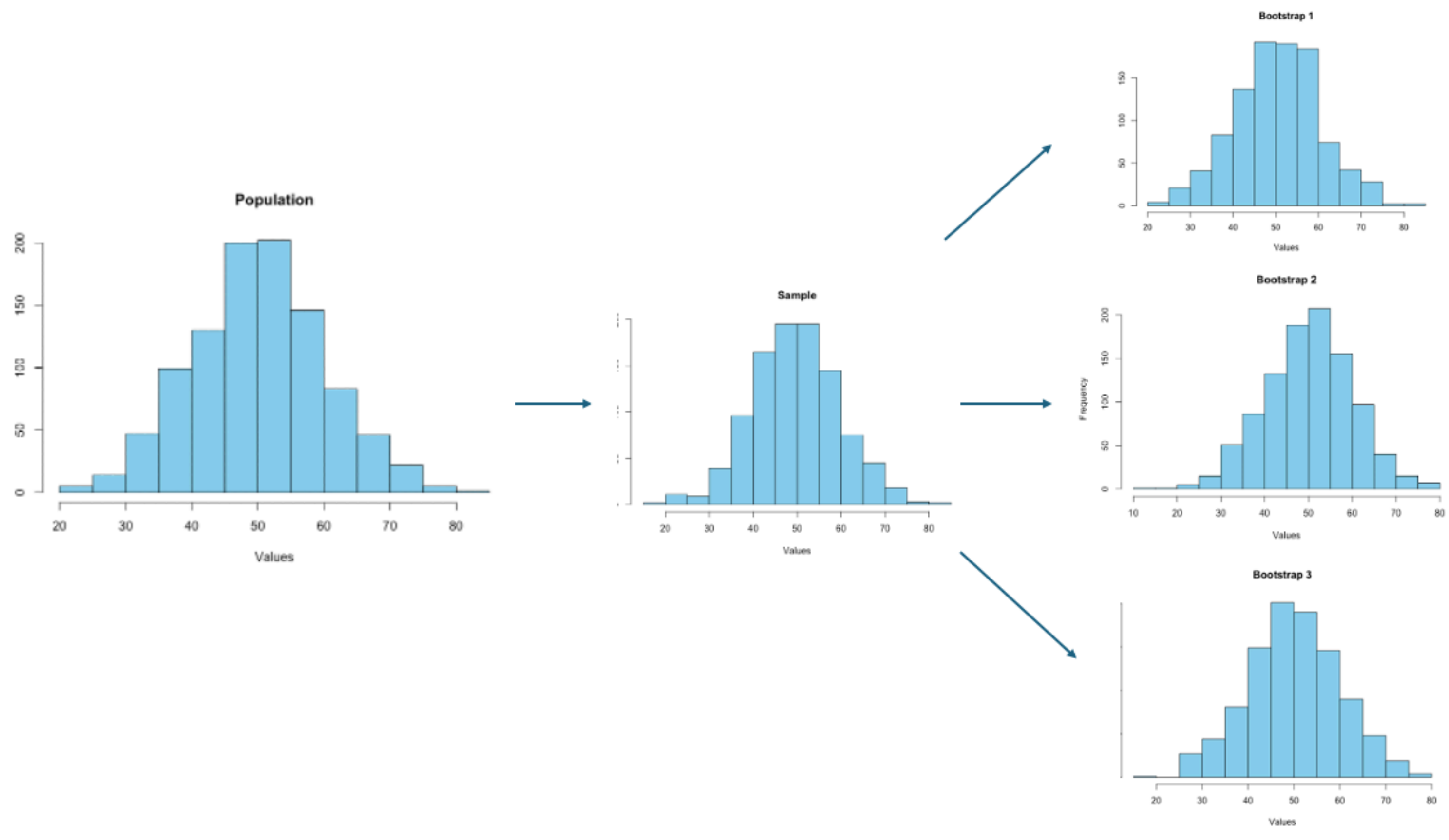
**Level of Confidence** is a probability that the confidence interval actually captures a true value.

# Before we discuss confidence intervals, we need to learn about bootstrapping

🥾

# Bootstrapping

This is a technique used to estimate the sampling distribution of a statistic by **repeatedly re-sampling from the observed data with replacement**



# Bootstrap Example

This is our sample.

$$\overset{\mu}{}(-5,0) * - - * - * - * - * - - - (0,0) - - \overset{\mu}{|} - - * - - * - - * - *$$

Then randomly pull 3 values, which becomes our data.

$$(-5,0) - - - * - - - - - (0,0) \overset{\mu}{|} - - - - * - - - - - * - -(5,0)$$

Plot our mean then throw those values back in (i.e., replacement) and pull another 3.

$$(-5,0) - - - - - - - -(0,0) - - - - * - - *- \overset{\mu}{|} - * - - - - (5,0)$$

Plot that mean and repeat!

# Bootstrap Example

Eventually we will have tons of means (we only have 9 in our example here) but in reality we would have hundreds or even thousands huddled around a similar point on our number line.

$$(-5,0)- \overset{\mu}{|} - - -- \overset{\mu}{|} - - -(0,0) - -- \overset{\mu}{|} - \overset{\mu}{|} - \overset{\mu}{|} - \overset{\mu}{|} - \overset{\mu}{|} - \overset{\mu}{|} - - - \overset{\mu}{|}$$

Once we have these, we have robust estimates of a true mean within a sample.

# Great, but what does this have to do with confidence intervals?

# Confidence Intervals

Well, when we bootstrapped we essentially created an empirical sampling distribution.

$$(-5,0)- \overset{\mu}{|} - - -- \overset{\mu}{|} - - -(0,0) - -- \overset{\mu}{|} - \overset{\mu}{|} - \overset{\mu}{|} - \overset{\mu}{|} - \overset{\mu}{|} - \overset{\mu}{|} - - - \overset{\mu}{|}$$

And as a result we can identify the percentiles that corresponds to your desired confidence level.

These percentiles from the bootstrap replicates become the lower and upper bounds of what we will call a confidence interval.

# Confidence Intervals

Now that we have this sampling distribution we can calculate our 95% confidence interval.

Which simply means, we have a 95% certainty that our population mean will fall between the two red lines.

And 5% of the time, our population mean could fall outside of that zone.

$$(-5,0) - \overset{\mu}{|} -- \overset{\mu}{|} - - -(0,0) - - - {\color{red}|} - \overset{\mu}{|} - \overset{\mu}{|} - \overset{\mu}{|} - \overset{\mu}{|} - \overset{\mu}{|} - \overset{\mu}{|} - {\color{red}|} - - \overset{\mu}{|}$$

# Time for Math! 🧮

# Confidence Interval

$$\text{CI} = \bar{x} \pm z\frac{s}{\sqrt{n}}$$

$\bar{x}$ = sample mean

$z$ = confidence level

$s$ = sample standard deviation

$n$ = sample size

- Large sample are $\geq 30$
- Small samples are $< 30$

# Confidence Levels

Table - Z-Scores for Commonly Used Confidence Intervals

| Desired Confidence Interval | Z Score |
| --- | --- |
| 90% | 1.645 |
| 95% | 1.96 |
| 99% | 2.576 |

# Confidence Level Example

A researcher wants to estimate the average number of hours per week spent by employees in a certain company on remote work activities.

They take a random sample of 80 employees and find that the average number of hours spent on remote work per week in the sample is 12 hours, with a standard deviation of 2 hours.

Calculate the 95% confidence interval for the average number of hours spent on remote work activities per week by all employees in the company.

# Confidence Interval Example

1) Find the z score
2) Add known values to equation
3) Calculate the SE or

$$\frac{s}{\sqrt{n}}$$

4) Multiple value from step 3 by our z value
5) Subtract that from our $\bar{x}$ for our Lower Bound CI
6) Add that from our $\bar{x}$ for our Upper Bound CI

$$\text{CI} = \bar{x} \pm z\frac{s}{\sqrt{n}}$$

- 95% CI: $11.5613 \leq \bar{x} \leq 12.4387$

# Confidence Interval for Small Samples

With smaller samples (n< 30) the Central Limit Theorem does not apply,so we use a different distribution called the t distribution.

In order to locate our t-value, we have to locate the degrees of freedom.

# Degrees of Freedom

The number of independent values that are simple allowed to vary within the calculation of a parameter.

$$df = n - 1$$

n = sample size

But why n-1?

# N-1 with respect to Degrees of Freedom

Let's consider an example where 4 people are asked to pick a number that, when combined with the other members, will equal 50.

- Person 1 chooses 10
- Person 2 chooses 20
- Person 3 chooses 10
- Person 4 chooses...... Oof they can't choose anything! They are no longer free to vary.
- They have to choose 10 because the we want a sum that equals 50.

# Confidence Interval for Small Samples

**Once we have a df and our confidence level, we head to page 373 in our book or head to this link.**

### *t* Table

| cum. prob | $t_{.50}$ | $t_{.75}$ | $t_{.80}$ | $t_{.85}$ | $t_{.90}$ | $t_{.95}$ | $t_{.975}$ | $t_{.99}$ | $t_{.995}$ | $t_{.999}$ | $t_{.9995}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| one-tail | **0.50** | **0.25** | **0.20** | **0.15** | **0.10** | **0.05** | **0.025** | **0.01** | **0.005** | **0.001** | **0.0005** |
| two-tails | **1.00** | **0.50** | **0.40** | **0.30** | **0.20** | **0.10** | **0.05** | **0.02** | **0.01** | **0.002** | **0.001** |
| df | | | | | | | | | | | |
| 1 | 0.000 | 1.000 | 1.376 | 1.963 | 3.078 | 6.314 | 12.71 | 31.82 | 63.66 | 318.31 | 636.62 |
| 2 | 0.000 | 0.816 | 1.061 | 1.386 | 1.886 | 2.920 | 4.303 | 6.965 | 9.925 | 22.327 | 31.599 |
| 3 | 0.000 | 0.765 | 0.978 | 1.250 | 1.638 | 2.353 | 3.182 | 4.541 | 5.841 | 10.215 | 12.924 |
| 4 | 0.000 | 0.741 | 0.941 | 1.190 | 1.533 | 2.132 | 2.776 | 3.747 | 4.604 | 7.173 | 8.610 |
| 5 | 0.000 | 0.727 | 0.920 | 1.156 | 1.476 | 2.015 | 2.571 | 3.365 | 4.032 | 5.893 | 6.869 |
| 6 | 0.000 | 0.718 | 0.906 | 1.134 | 1.440 | 1.943 | 2.447 | 3.143 | 3.707 | 5.208 | 5.959 |
| 7 | 0.000 | 0.711 | 0.896 | 1.119 | 1.415 | 1.895 | 2.365 | 2.998 | 3.499 | 4.785 | 5.408 |
| 8 | 0.000 | 0.706 | 0.889 | 1.108 | 1.397 | 1.860 | 2.306 | 2.896 | 3.355 | 4.501 | 5.041 |
| 9 | 0.000 | 0.703 | 0.883 | 1.100 | 1.383 | 1.833 | 2.262 | 2.821 | 3.250 | 4.297 | 4.781 |
| 10 | 0.000 | 0.700 | 0.879 | 1.093 | 1.372 | 1.812 | 2.228 | 2.764 | 3.169 | 4.144 | 4.587 |
| 11 | 0.000 | 0.697 | 0.876 | 1.088 | 1.363 | 1.796 | 2.201 | 2.718 | 3.106 | 4.025 | 4.437 |
| 12 | 0.000 | 0.695 | 0.873 | 1.083 | 1.356 | 1.782 | 2.179 | 2.681 | 3.055 | 3.930 | 4.318 |
| 13 | 0.000 | 0.694 | 0.870 | 1.079 | 1.350 | 1.771 | 2.160 | 2.650 | 3.012 | 3.852 | 4.221 |
| 14 | 0.000 | 0.692 | 0.868 | 1.076 | 1.345 | 1.761 | 2.145 | 2.624 | 2.977 | 3.787 | 4.140 |
| 15 | 0.000 | 0.691 | 0.866 | 1.074 | 1.341 | 1.753 | 2.131 | 2.602 | 2.947 | 3.733 | 4.073 |
| 16 | 0.000 | 0.690 | 0.865 | 1.071 | 1.337 | 1.746 | 2.120 | 2.583 | 2.921 | 3.686 | 4.015 |
| 17 | 0.000 | 0.689 | 0.863 | 1.069 | 1.333 | 1.740 | 2.110 | 2.567 | 2.898 | 3.646 | 3.965 |
| 18 | 0.000 | 0.688 | 0.862 | 1.067 | 1.330 | 1.734 | 2.101 | 2.552 | 2.878 | 3.610 | 3.922 |
| 19 | 0.000 | 0.688 | 0.861 | 1.066 | 1.328 | 1.729 | 2.093 | 2.539 | 2.861 | 3.579 | 3.883 |
| 20 | 0.000 | 0.687 | 0.860 | 1.064 | 1.325 | 1.725 | 2.086 | 2.528 | 2.845 | 3.552 | 3.850 |
| 21 | 0.000 | 0.686 | 0.859 | 1.063 | 1.323 | 1.721 | 2.080 | 2.518 | 2.831 | 3.527 | 3.819 |
| 22 | 0.000 | 0.686 | 0.858 | 1.061 | 1.321 | 1.717 | 2.074 | 2.508 | 2.819 | 3.505 | 3.792 |
| 23 | 0.000 | 0.685 | 0.858 | 1.060 | 1.319 | 1.714 | 2.069 | 2.500 | 2.807 | 3.485 | 3.768 |
| 24 | 0.000 | 0.685 | 0.857 | 1.059 | 1.318 | 1.711 | 2.064 | 2.492 | 2.797 | 3.467 | 3.745 |
| 25 | 0.000 | 0.684 | 0.856 | 1.058 | 1.316 | 1.708 | 2.060 | 2.485 | 2.787 | 3.450 | 3.725 |
| 26 | 0.000 | 0.684 | 0.856 | 1.058 | 1.315 | 1.706 | 2.056 | 2.479 | 2.779 | 3.435 | 3.707 |
| 27 | 0.000 | 0.684 | 0.855 | 1.057 | 1.314 | 1.703 | 2.052 | 2.473 | 2.771 | 3.421 | 3.690 |
| 28 | 0.000 | 0.683 | 0.855 | 1.056 | 1.313 | 1.701 | 2.048 | 2.467 | 2.763 | 3.408 | 3.674 |

# Confidence Interval for Small Samples Example

A biology class wants to estimate the average length of a certain species of fish in a lake.

They take a random sample of 15 fish and measure their lengths. The sample mean length is 30 cm, and the sample standard deviation is 4 cm.

Calculate the 95% confidence interval for the average length of this species of fish in the lake.

# Confidence Interval Example

1) Find the df and our confidence interval and locate the t value
2) Add known values to equation
3) Calculate the SE or

$$\frac{s}{\sqrt{n}}$$

4) Multiple value from step 3 by our t value
5) Subtract that from our $\bar{x}$ for our Lower Bound CI
6) Add that from our $\bar{x}$ for our Upper Bound CI

$$\text{CI} = \bar{x} \pm t\frac{s}{\sqrt{n}}$$

- 95% CI: $27.781 \leq \bar{x} \leq 32.219$

# That's all folks! Have a great weekend!