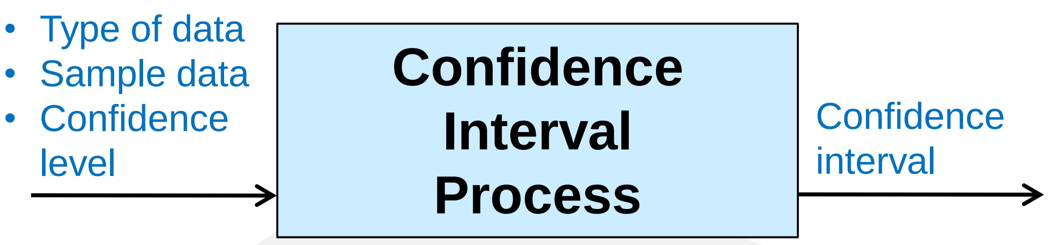
Confidence Intervals

Statistics

Problem
How to estimate the range for a population parameter?

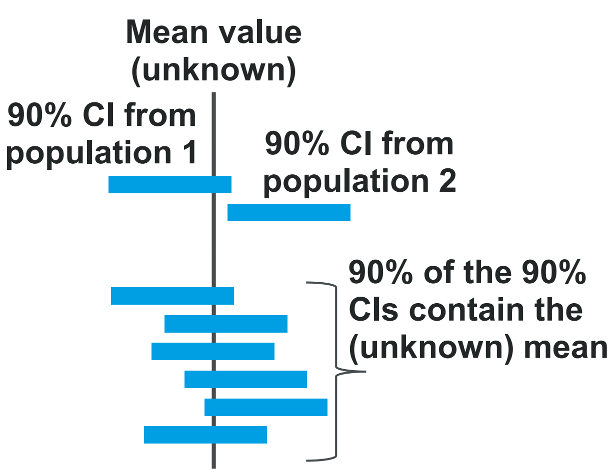
Difficulty
Some training required

- A **confidence interval (CI)** is a range of values likely to contain a population parameter.
- A *confidence level* defines a CI; common values are 90%, 95%, and 99%.
- A higher confidence level or a smaller sample size results in a wider CI.
- Each set of data analyzed creates a different CI.



Two-sided CI for mean of a normal random variable

1. Collect n random sample from the population.
2. Select desired confidence level (e.g., 95%)
3. Look up corresponding Z-score (e.g., $z^* \approx 1.96$)
4. Calculate the sample mean ($\bar{x}$).
5. Calculate the sample standard deviation (s).
6. Calculate the Standard Error ($SE = s/\sqrt{n}$)
7. Calculate the Margin of Error ($ME = z^* \times SE$)
8. Calculate the upper and lower CI bounds: $\bar{x} \pm ME$



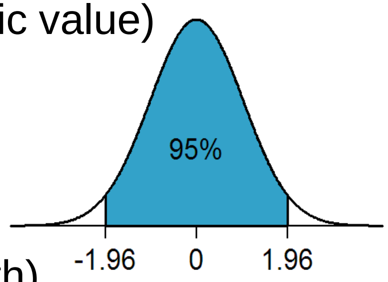
Confidence Intervals – Example

https://commons.wikimedia.org/wiki/File:NormalDist1.96.png

Question: On average, how many hours per month do workers in a specific company spend reading 6in6 presentations?

Solution assumptions

- The data is normally distributed (reasonable, but needs to be checked)
- We will determine a confidence interval for the mean
- Use a two-sided test (a one-sided test could compare the mean to a specific value)
- We collect n=100 data samples



Perform computation

A. Choose a confidence level (say, 95%)
B. From the confidence level, look up the z-score ($z^* \approx 1.96$): see image
C. From the data samples, compute the sample mean (say, $\bar{x}$ = 8 hours/month)
D. From the data samples, compute the sample standard deviation (say, s=2 hours/month)
E. Compute the Standard Error ($SE = s/\sqrt{n} = 2/10 = 0.2$)
F. Compute the Margin of Error ($ME = z^* \times SE = 0.39$)
G. Compute the upper and lower CI bounds:
   ($\bar{x}$ ± ME) = (8 ± 0.39) = [8–0.39, 8+0.39] = [7.61, 8.39] hours/week

State conclusion

A. *A 95% confidence interval is that* workers *spend between 7.6 and 8.4 hours per month reading 6in6 presentations.*
B. The true average number of hours spent per month is unknown.
C. A different data sample, perhaps from the next month, could give a different 95% confidence interval. For example, it might be [7.5, 8.3] hours/month.
D. If you determined 100 different 95% confidence interval (over 100 months, assuming the mean is constant) then, statistically, 95 of the CIs would contain the unknown mean.

Confidence Intervals – Notes

Slide 1

1. CIs have a nuanced interpretation. Many people misunderstand what a CI represents.
2. Much CI information on the web is wrong. For example: a 95% confidence level
   - **does not** mean that the true value has a 95% probability of being within the calculated 95% CI.
   - **does not** mean that 95% of the sample data lie within the confidence interval.
3. Confidence intervals depend on the type of data, the number of data points, and whether a one-sided or two-sided CI is desired.
4. Pros:
   - A CI is easy to interpret.
   - A CI is useful in decision-making.
   - A CI provides a range of possible values for a population parameter
5. Cons:
   - A CI uses a subjective confidence level
   - A CI does not guarantee the parameter lies within the interval.

Slide 2

1. We chose to use a two-sided test. A one-sided test would be appropriate to compare the mean against a single value. For example, this would answer “On average, do workers spend more than 8 hours per month reading 6in6 presentations?
2. If, in fact, you had data for two months (and the mean stayed constant) then you could compute a single confidence interval using all 200 data points.

Recommended web sites for additional information

- https://stattrek.com/estimation/confidence-interval
- https://www.calculator.net/confidence-interval-calculator.html