Part 1: Introduction to Confidence Intervals

The **confidence interval** is the plus-or-minus figure usually reported in newspaper or television opinion poll results. For example, if you calculate a confidence interval of 4. If 47% percent of your sample picks an answer you can be "95% confident" that if you ask the question of the entire population, then between 43% (47 - 4) and 51% (47 + 4) would have picked that answer.

The **confidence level** tells you how sure you can be. It is expressed as a percentage and represents how often the true percentage of the population who would pick an answer lies within the confidence interval. The 95% confidence level means you can be 95% certain; the 99% confidence level means you can be 99% certain. Most researchers use the 95% confidence level.

When you put the confidence level and the confidence interval together, you can say that you are 95% confident that the true percentage of the population is between 43% and 51%.

A Note on Sample Size

The larger your sample, the more sure you can be that their answers truly reflect the population. This indicates that for a given **confidence level**, the larger your sample size, the smaller your **confidence interval**.

Part 2: Calculating a Confidence Interval When You Have Sample Data

Charlie Baker is running for governor of Massachusetts and is interested in how he is doing in a target population of voters, women. His organization obtained a random sample of 200 female voters and asked them to choose between Baker and Martha Coakley. The sample reveals that Coakley is leading among female votes 55% to 45%. [Note: This is a point estimate of voter preference]. However, we know that this is just a sample of the population. And, we know that samples are not perfect indicators of the population.

Step 1: Determine the Sample Mean and the Standard Deviation of the Sample Mean

→ Sample Mean = .55

→ Standard Deviation of the Sample Mean

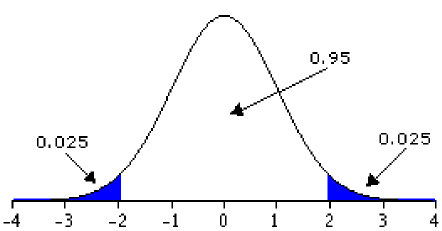
$$= \sqrt{\frac{(Proportion\ favoring\ Coakley)*(1-Proportion\ favoring\ Coakley)}{N}} \quad : \quad \sqrt{\frac{(p)*(q)}{N}}$$

$$= \sqrt{\frac{(.55)*(.45)}{200}}$$

$$= .035707$$

Step 2: Decide your confidence interval and determine the corresponding z-score. *Note*: 95% is standard in psychological research, and should be used for this problem.

→ 90% confidence interval: 1.645
→ 95% confidence interval: 1.96
→ 98% confidence interval: 2.326
→ 99% confidence interval: 2.576



Step 3: Calculate the Upper Limit and Lower Limit for the 95% Confidence Interval.

$$\boldsymbol{C.I. = Mean + (z\_score)(Standard\ Deviation\ of\ the\ Sample\ Mean)}$$

$C.I. for\ Upper\ Limit = .55 + (+1.96)(.035707)$

$C.I. for\ Upper\ Limit = .62$

$C.I. for\ Lower\ Limit = .55 + (-1.96)(.035707)$

$C.I. for\ Lower\ Limit = .48$

Step 4: Make a conclusion statement about the mean of your sample (n=25).
-There is a .95 probability that the true population value falls between .48 and .62.
-Said another way: We are 95% confident that the true population mean falls between .48 and .62.
-Notice that the 95% confidence interval includes 50%, so in other words, the race among women voters is too close to call based on a sample size of N = 200
-The difference between the candidates is within the Margin of Error.

Now, using the same data from the problem above of SAT scores (Mean=.55), you now sample 1000 women voters. Determine the 95% confidence interval for the sample mean.

Step 1: Calculate the standard deviation of the sample mean.

Step 2: Decide your confidence interval and determine the corresponding z-score.

Step 3: Calculate the Upper Limit and Lower Limit for the 95% Confidence Interval.

Step 4: Make a conclusion statement about the mean of your sample (n=1000).

Thought question: What happens to the confidence interval when you increase the sample size? Explain.

Part 3: Calculating a Confidence Interval When You Have Population Parameters

For example, SAT scores are normalized in a way that the population mean is 500 and the population standard deviation is 100. From the entire population of SAT test-takers, we sample 25 test-takers. Using confidence intervals, we can estimate the range in which the sample mean (n=25) will fall with 95% confidence.

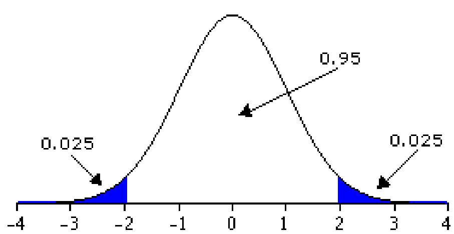
Step 1: Turn the standard deviation of the population mean ($\theta = 100$) into a standard deviation of the sample mean.

$$\mathbf{s} = \frac{\boldsymbol{\theta}}{\sqrt{\boldsymbol{N}}} \qquad \text{Standard Deviation of Sample Mean} = \frac{\textit{Population Standard Deviation}}{\sqrt{\# \textit{ of Sample}}}$$

$$s = \frac{100}{\sqrt{25}} = 20$$

Step 2: Decide your confidence interval and determine the corresponding z-score. *Note*: 95% is most standard.

- → 90% confidence interval: 1.645
- → 95% confidence interval: 1.96
- → 98% confidence interval: 2.326
- → 99% confidence interval: 2.576



Step 3: Calculate the Upper Limit and Lower Limit for the 95% Confidence Interval.

$$\boldsymbol{C.I.} = \boldsymbol{Mean} + (\boldsymbol{z\_score})(\boldsymbol{Standard\ Deviation\ of\ the\ Sample\ Mean})$$

$$C.I. for\ Upper\ Limit = 500 + (+1.96)(20)$$
$$C.I. for\ Upper\ Limit = 539.2$$

$$C.I. for\ Lower\ Limit = 500 + (-1.96)(20)$$
$$C.I. for\ Lower\ Limit = 460.8$$

Step 4: Make a conclusion statement about the mean of your sample (n=25).

-There is a 95% probability that the sample mean will fall within the boundaries of 460.8 and 539.2

-We are 95% confident that the sample mean will fall between 460.8 and 539.2

Now, using the same data on SAT scores (Mean = 500, S.D. = 100), you a sample 100 test-takers. Your task is to determine the 95% confidence interval for the sample mean.

Step 1: Turn the standard deviation of the population mean ($\theta = 100$) into a standard deviation of the sample mean.

Step 2: Decide your confidence interval and determine the corresponding z-score.

Step 3: Calculate the Upper Limit and Lower Limit for the 95% Confidence Interval.

Step 4: Make a conclusion statement about the mean of your sample (n=100).

Thought question: What happens to the confidence interval when you increase the sample size? Explain.

Part 4: Practice Problems

Be careful when choosing the formula to calculate the Standard Deviation of the Sample Mean.

(1) The mean income of the entire population in the state of Massachusetts is $60,000, with a population standard deviation of $5,000. You sample 100 BC alumni. Based on the population parameters, caulculate the 95% confidence interval, and make a conclusion statement.

(2) Which dining hall do BC students prefer, Hillside or McElroy? You survey 100 students at random and find that 63% prefer Hillside and 37% prefer McElroy. Based on the known proportions of this sample, calculate the 95% confidence interval.

(3) The mean weight for the entire population of 5th graders in Boston is 70 lbs, with a population standard deviation of 9lbs. You go to the Children's Museum and randomly weigh a sample of 400 5th graders from Boston. Based on the population parameters, calculate the 95% confidence interval, and make a conclusion statement.