The usual reason for taking a sample is to infer from the sample data some conclusions about the wider group – population (NOT to learn about the individuals in the sample).

And **Statistical Inference** provides methods for drawing conclusions about a population from sample data.

The main complexity in any inference from a sample is that different samples may lead to different conclusions.

One of the common methods of statistical inference is **Confidence Intervals** (CIs).

**Def. Confidence Intervals** estimate the value of the population parameter.

CIs calculated from the data are usually of the form: \( \text{estimate} \pm \text{margin error} \).

We write the intervals as \((\text{lower bound}, \text{ upper bound})\).

**Def.** A **confidence level** \(C\) gives the probability that the interval will capture the true parameter value in repeated samples. That is, the confidence level is the success rate for the method. (we are usually interested in 90% CI, or 95% CI, or 99% CI).

If we take a single sample, our single confidence interval may or may not include the population parameter.

However if we take many samples of the same size and create a confidence interval from each sample statistic, over the long run 95% of our confidence intervals will contain the true population parameter (if we are using a 95% confidence level).

**The form of interpretation of CIs:**

For example: We are 95% confident that the unknown \( \mu \) lies between \( \text{lower bound} \) and \( \text{upper bound} \).
Illustration of CIs on 100 samples from the Normal population with $\mu=26.8$, $\sigma=7.5$. $n = 654$ (where $n$ is a sample size)

**Simple Conditions for Inference about a mean (assumptions):**

1. We have an SRS from the population of interest. There is no nonresponse or other practical difficulty.
2. The variable we measure has an exactly Normal Distribution $N(\mu, \sigma)$ in the population.
3. We don't know the population mean $\mu$. But we do know the population standard deviation $\sigma$.

The conditions that we have a perfect SRS, the population is exactly Normal, and we know the population $\sigma$ are all unrealistic, but we'll start from this conditions for simplicity. In next chapters we begin to move from the “simple conditions” toward the reality of statistical inference.
Confidence Interval for the mean of a Normal Population

Draw an SRS of size $n$ from a Normal population having unknown mean $\mu$ and known standard deviation $\sigma$. A level $C$ confidence interval for $\mu$ is

$$\bar{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}}$$

(estimate $\pm$ margin error)

The critical value $z^*$ depends on confidence level $C$ and can be found using table A.

Or easily: at the bottom of table C one can find $z^*$ for some popular confidence levels.

<table>
<thead>
<tr>
<th>Confidence level C</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical value $z^*$</td>
<td>1.645</td>
<td>1.960</td>
<td>2.576</td>
</tr>
</tbody>
</table>

Ways to reduce your margin of error:

- Increase sample size; ✔
- Smaller $z^*$, or in other words, use a lower level of confidence (smaller $C$); ✔
- Reduce $\sigma$ (but we can't change the population sd $\sigma$, it's fixed for us due to “simple conditions”)

If you increase your sample size ($n$), you decrease your margin of error:

If you increase your confidence level ($C$), then you increase your margin of error:

A smaller margin of error is good because we get a smaller range of where to expect the true population parameter.

**Sample Size ($n$) for Desired Margin of Error ($m$):**

$$n = \left( \frac{z^* \cdot \sigma}{m} \right)^2$$

**Note** that it is the sample size ($n$) that influences the margin of error. The population size has nothing to do with it.
Problem 1.

An article in The Journal of the American Medical Association examined body temperatures of males and females. The mean body temperature of the 65 female subjects who participated in the study was 98.4 degrees Fahrenheit. Suppose it is known that the body temperature is distributed normally with standard deviation of 0.72 degree.

a) Give a 95% confidence interval for \( \mu \), the population mean body temperature of females.

b) Explain in simple language to someone who knows no statistics what 95% confidence means.

c) What is the value of \( z^* \) for a 94% confidence interval?

d) How many females must be sampled in order to estimate \( \mu \) within ±0.1 degree with 95% confidence?

Problem 2.

A questionnaire of spending habits was given to a random sample of college students. Each student was asked to record and report the amount of money they spent on textbooks in a semester. The sample of 130 students resulted in an average of $422. Suppose it is known that the spent money on textbooks in a semester is distributed normally with standard deviation of $57.

a) Give a 90% confidence interval for the mean amount of money spent by college students on textbooks and interpret the interval.

b) What is the margin of error for the 90% confidence interval?

c) How many students should you sample if you want a margin of error of $5 for a 90% confidence interval?

d) What is the critical value \( z^* \) if the confidence level was changed to 92%?
Problem 3. Confidence level and margin of error

Body mass index (BMI) is used to screen for possible weight problems. It is calculated as weight divided by the square of height, measuring weight in kilograms and height in meters. For data about BMI, we turn to the National Health and Nutrition Examination Survey (NHANES), a continuing government sample survey that monitors the health of the American population. An NHANES report gives data for 654 women aged 20 to 29 years. The mean BMI in the sample was $\bar{x} = 26.8$. We treated these data as an SRS from a Normally distributed population with standard deviation $\sigma = 7.5$.

a) Give three confidence intervals for the mean BMI in this population, using 90%; 95%, and 99% confidence and interpret each of them.

b) What are the margins of error for 90%; 95%, and 99% confidence? How does increasing the confidence level change the margin of error of a confidence interval when the sample size and population standard deviation remain the same?

Problem 4. Sample size and margin of error

The last problem described NHANES survey data on the body mass index (BMI) of 654 young women. The mean BMI in the sample was $\bar{x} = 26.8$. We treated these data as an SRS from a Normally distributed population with standard deviation $\sigma = 7.5$.

a) Suppose that we had an SRS of just 100 young women. What would be the margin of error for 95% confidence?

b) Find the margins of error for 95% confidence based on SRSs of 400 young women and 1600 young women.

c) Compare the three margins of error. How does increasing the sample size change the margin of error of a confidence interval when the confidence level and population standard deviation remain the same?