With a point estimate, we used a single number to estimate a parameter. 
We can also use a set of numbers to serve as “reasonable” estimates for the parameter.

Example: Assume we have a sample of size 100 from a population with \( \sigma = 0.1 \).

From CLT:

Empirical Rule: If we take many samples, calculating \( \bar{X} \) each time, then about 95% of the values of \( \bar{X} \) will be between:

Therefore:

This interval is called an approximate 95% “confidence interval” for \( \mu \).
Confidence Interval: An interval (along with a level of confidence) used to estimate a parameter.

- Values in the interval are considered “reasonable” values for the parameter.

Confidence level: The percentage of all CIs (if we took many samples, each time computing the CI) that contain the true parameter.

Note: The endpoints of the CI are statistics, calculated from sample data. (The endpoints are random, not the parameter!)

In general, if $\bar{X}$ is normally distributed, then in $100(1 - \alpha)\%$ of samples, the interval will contain $\mu$.

Note: $z_{\alpha/2}$ = the z-value with $\alpha/2$ area to the right:

$$100(1 - \alpha)\% \text{ CI for } \mu: \quad \bar{X} \pm z_{\alpha/2}(\sigma / \sqrt{n})$$
Problem: We typically do not know the parameter $\sigma$. We must use its estimate $s$ instead.

**Formula**: CI for $\mu$ (when $\sigma$ is unknown)

$$\bar{X} - \mu$$

Since $\frac{\bar{X} - \mu}{s / \sqrt{n}}$ has a t-distribution with $n - 1$ d.f., our $100(1 - \alpha)\%$ CI for $\mu$ is:

where $t_{\alpha/2} =$ the value in the t-distribution ($n - 1$ d.f.) with $\alpha/2$ area to the right:

- This is valid if the data come from a normal distribution.

Example: We want to estimate the mean weight $\mu$ of trout in a lake. We catch a sample of 9 trout. Sample mean $\bar{X} = 3.5$ pounds, $s = 0.9$ pounds. $95\%$ CI for $\mu$?
**Question:** What does 95% confidence mean here, exactly?

- If we took many samples and computed many 95% CIs, then about 95% of them would contain $\mu$.

The fact that contains $\mu$ “with 95% confidence” implies the method used would capture $\mu$ 95% of the time, if we did this over many samples.

**Picture:**

---

**A WRONG statement:** “There is .95 probability that $\mu$ is between 2.81 and 4.19.” Wrong! $\mu$ is not random – $\mu$ doesn’t change from sample to sample. It’s either between 2.81 and 4.19 or it’s not.
Level of Confidence

Recall example: 95% CI for μ was (2.81, 4.19).
• For a 90% CI, we use t₀.05 (8 d.f.) = 1.86.
• For a 99% CI, we use t₀.005 (8 d.f.) = 3.355.

90% CI:

99% CI:

Note tradeoff: If we want a higher confidence level, then the interval gets wider (less precise).

Confidence Interval for a Proportion

• We want to know how much of a population has a certain characteristic.
• The proportion (always between 0 and 1) of individuals with a characteristic is the same as the probability of a random individual having the characteristic.

Estimating proportion is equivalent to estimating the binomial probability p.

Point estimate of p is the sample proportion:
Note $\hat{p} = \frac{x}{n}$ is a type of sample average (of 0’s and 1’s), so CLT tells us that when sample size is large, sampling distribution of $\hat{p}$ is approximately normal.

For large $n$:

$100(1 – \alpha)\%$ CI for $p$ is:

How large does $n$ need to be?

Example 1: A student government candidate wants to know the proportion of students who support her. She takes a random sample of 93 students, and 47 of those support her. Find a 90% CI for the true proportion.

Check:
Example 2: We wish to estimate the probability that a randomly selected part in a shipment will be defective. Take a random sample of 79 parts, and find 4 defective parts. Find a 95% CI for $p$. 
Confidence Interval for the Variance $\sigma^2$ (or for s.d. $\sigma$)

Recall that if the data are normally distributed,

$$\frac{(n-1)s^2}{\sigma^2}$$

has a $\chi^2$ sampling distribution with $(n - 1)$ d.f.

This can be used to develop a $(1 - \alpha)100\%$ CI for $\sigma^2$:

Example: Trout data example (assume data are normal – how to check this?) $s = 0.9$ pounds, so $s^2 = n = 9$. Find 95% CI for $\sigma^2$.

95% CI for $\sigma$:
Also, a CI for the ratio of two variances, $\frac{\sigma_1^2}{\sigma_2^2}$, can be found by the formula:

Example: If we have a second sample of 13 trout with sample variance $s_2^2 = 0.7$, then a 95% CI for $\frac{\sigma_1^2}{\sigma_2^2}$ is:
Sample Size Determination

Note that the bound (or margin of error) $B$ of a CI equals half its width.

For the CI for the mean (with $\sigma$ known), this is:

For the CI for the proportion, this is:

Note: When the sample size $n$ is bigger, the CI is narrower (more precise).

We often want to determine what sample size we need to achieve a pre-specified margin of error and level of confidence. Solving for $n$:

CI for mean:

CI for proportion:
Note: Always round \( n \) up to the next largest integer.

These formulas involve \( \sigma, p \) and \( q \), which are usually unknown in practice. We typically guess them based on prior knowledge – often we use \( p = 0.5, q = 0.5 \).

Example 1: How many patients do we need for a blood pressure study? We want a 90% CI for mean systolic blood pressure reduction, with a margin of error of 5 \( mmHg \). We believe that \( \sigma = 10 \ mmHg \).

Example 2: Pollsters want a 95% CI for the proportion of voters supporting President Bush. They want a 3% margin of error \( (B = .03) \). What sample size do they need?