Chapter 9: Confidence Intervals
STAT 2601 – Business Statistics

Esam Mahdi

School of Mathematics and Statistics
Carleton University

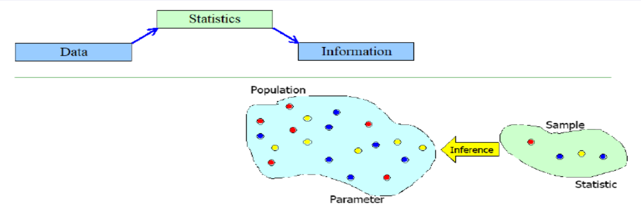
January 21, 2026

Table of Contents

Statistical Inference: Revisited

Statistical Inference

Statistical inference is the process by which we acquire information and draw conclusions about populations from samples.



Example

The mean of the sample of 60 customers waiting times using a new system is $\bar{x} = 7.16$ minutes.
We'd like to use this sample mean to say something about population mean, μ, the mean of all possible customer waiting times using the new system.

Statistical Inference

Estimation: There are two types of inference: estimation and hypothesis testing; estimation is introduced first. The objective of estimation is to determine the approximate value of a population parameter on the basis of a sample statistic.

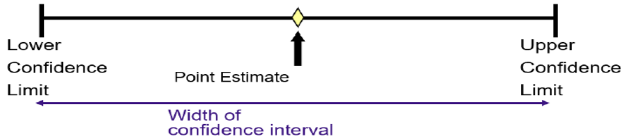
1. **Point Estimate:** A point estimate is a single number, used to estimate an unknown population parameter. The point estimate is not likely to exactly equal the population parameter. For example, sample mean,

$$\bar{x} = \frac{\sum x}{n}$$

   is a point estimate of population mean,

$$\mu = \frac{\sum x}{N}$$

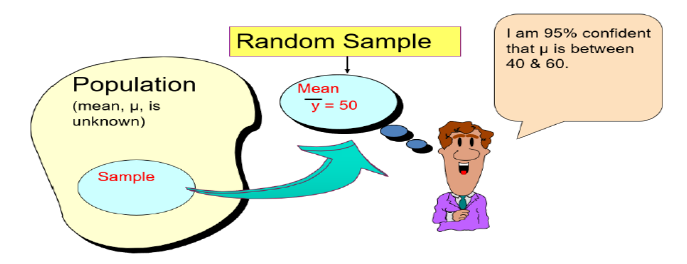
2. **Interval Estimate:** How much uncertainty is associated with a point estimate of a population parameter? An interval estimate (i.e., confidence interval (CI)) provides more information about a population characteristic by incorporating the sampling error.



Confidence Intervals for unknown μ

Estimate of μ

1. Point Estimate: A single statistic, determined from a sample, that is used to estimate the corresponding population parameter.
   - $\bar{x}$ is a point estimator of μ.
   - S^2 and S are point estimators of σ^2 and σ, respectively.
   - $\hat{p}$ is a point estimator of p.
2. Interval Estimate: An interval estimate provides more information about a population characteristic than does a point estimate.



Confidence Interval for Population μ (σ is Known)

Generic Formula:

The general confidence interval can be expressed in terms of the margin of error as follows:

$$\begin{aligned} &\text{Point Estimate } \pm \text{ Margin of Error} \\ &= \text{Point Estimate } \pm \underbrace{\text{Critical Value} \times \text{Standard Error}}_{\text{Margin of Error}} \end{aligned}$$

Interval Estimate of μ (when σ is known):

- Use a normal curve as a model of the sampling distribution of the sample means.
- Exactly, because the population is Normal.
- Approximately, by the Central Limit Theorem for large samples ($n \geq 30$) (when population is non-Normal).

Confidence Intervals:

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = \left(\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

Critical Value for Means

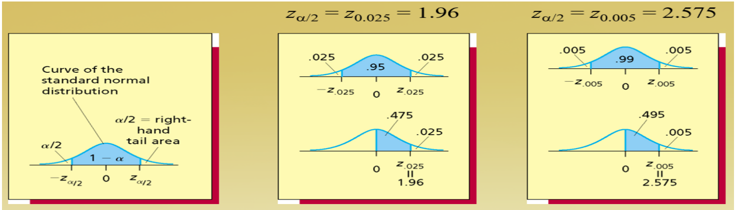
- Critical Value: To change the confidence level, we will need to change the critical value to correspond to the new level. For any confidence level the number of SEs we must stretch out on either side of $\bar{x}$ is called the critical value (CV).
- Working Approach:
  1. For a given confidence level (e.g., 90%, 95%), find $\alpha = 1 -$ confidence level.
  2. For two-sided confidence interval, divide α by 2 and find the critical value corresponding to $\dfrac{\alpha}{2}$ from Standard Normal (z) table.

Confidence Interval for Population μ (σ is Known)

- Margin of Error (E): The extent of the interval on either side of $\bar{x}$ is called the margin of error (ME), a measure of uncertainty.

$$\overline{x} \pm \underbrace{z_{\frac{\alpha}{2}} \times SE(\overline{x})}_{\text{Margin of Error}}$$

- Factors Affecting Width of Confidence Interval:
    - If margin of error $\uparrow$ (due to higher confidence level $(z_{\alpha/2} \uparrow)$ or standard error $(SE(\bar{x}) \uparrow)$), confidence interval becomes wider and less precise.
    - If margin of error $\downarrow$ (due to lower confidence level $(z_{\alpha/2} \downarrow)$ or standard error $(SE(\bar{x}) \downarrow)$), confidence interval shrinks and becomes more precise.
- **Note:** Every confidence interval is a balance between certainty and precision.



Example

A management consulting firm has installed a new computer-based electronic billing system and uses a sample of $n = 65$ payment times to estimate the mean payment time with 99% confidence. The mean of the 65 payment times is 18.1077 days. Here we assume that the population standard deviation for the new billing system is 4.2 days.

Solution:

1. Step 1: Check Conditions

$$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \quad \text{using CLT, } n > 30$$

2. Step 2: Construction of Confidence Interval for μ

$$\bar{x} \pm z_{\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} = \underbrace{18.1077}_{\text{Point Estimate}} \pm \underbrace{2.575}_{\text{Critical Value}} \times \underbrace{\frac{4.2}{\sqrt{65}}}_{\text{Standard Error}}$$

$$\rightarrow 18.1077 \pm \underbrace{1.3414}_{\text{Margin of Error}} = (16.8, 19.4)$$

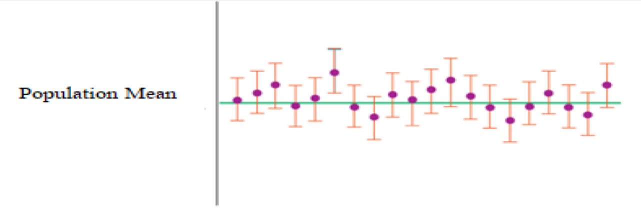
We are 95% confident that the true population mean μ lies between 16.8 and 19.4.

What does 95% really mean? (Long-run Approach)

Interpretation: Long-run Approach

Below we see the confidence intervals produced by simulating 20 samples.

- The purple dots are the simulated mean payment times of the new billing system.
- The orange segments show each sample's confidence intervals.
- The green line represents the true population mean.



Most CIs capture the true population mean (μ), but one misses (i.e., $1/20 = 0.05$, consistent with a 95% confidence level).

Interpretation: Long-run Approach: If we were to repeatedly take samples and construct 95% confidence intervals in the same way, about 95% of those intervals would contain the true value of μ.

Interpretation of the $(1-\alpha)100\%$ Confidence Interval for Population Mean

- Interpretation 1: In case of repeated sampling, $(1-\alpha)100\%$ of the confidence intervals constructed in such a manner is expected to contain the population mean, μ.

- Interpretation 2: It can be said with $(1-\alpha)100\%$ confidence that the confidence interval captures/contains population mean, μ.

Imaginary Framework

Note that we just draw one sample in reality but the interpretation is based on an imaginary framework (repeated samples). Our uncertainty is about whether the particular sample we have at hand is one of the successful ones or one of the $\alpha\%$ that fail to produce an interval that captures the true value.

Selecting the Sample Size: When Estimating μ

Sample Size

Solving for n we produce the required sample size to estimate μ and where E is the margin of error or bound on the error of estimation.

1. Step 1: Calculate n using the following formula:

$$n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{\mathrm{E}} \right)^2$$

2. Step 2: Always round up the value of n calculated in step 1 (if the result is not an integer)

Example: Determining Sample Size

Suppose $\sigma = 45$. What sample size is needed to be 90% confident of being correct within ± 5 of the population mean?

Solution:

1. Step 1: Margin of error (E) = 5.
2. Step 2:
$$n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{\text{E}}\right)^2$$
$$n = \left(\frac{1.645 \times 45}{5}\right)^2 = 219.19$$
3. Step 4: Round up the calculated value of n
$$n = 219.19 \approx 220$$
4. Interpretation: The required sample size at least 220 to keep the margin of error as small as 5 with confidence level 90%.

Standard Error and the Problem with z

- Standard error $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ describes how much difference is reasonable to expect between $\bar{x}$ and μ.
- **Problem with z Statistic:** Requires knowledge of σ, which is usually unknown.

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

- **Solution:** Replace σ with s (sample standard deviation).

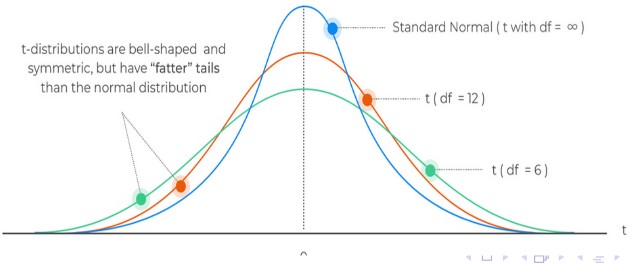
$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \quad \text{is NOT Normal!}$$

- This statistic follows the **Student's t distribution** with $n - 1$ degrees of freedom.
- Estimated standard error:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

The t Statistic: An Alternative to z

- **Student's t Distribution:** Discovered by William S. Gosset (pseudonym: Student).
- **Assumptions:**
  1. Independent observations.
  2. Population must be Normally distributed.
- **Properties of t Distribution:**
  - Mound-shaped and symmetric about 0.
  - More variable (flatter, heavier tails) than z.
  - Shape depends on degrees of freedom $(n-1)$.
  - As n increases, t approaches z (as a rule of thumb t and z distributions become almost identical for $n \geq 30$ (df ≥ 29)).



Degrees of Freedom

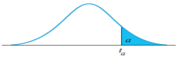
- **Degrees of Freedom (df):** The number of independent data values available to estimate a parameter.
- If k parameters must be estimated before calculating σ, then:

$$\text{df} = n - k$$

- For estimating μ with sample mean:
  - Sample mean uses 1 parameter.
  - Then $\text{df} = n - 1$.
- Once $\bar{x}$ is known, only $n - 1$ pieces of data are free to vary.

The t Table

- t **Table Structure:**
  - Top row: Area/probability in the right tail.
  - Left/right column: Degrees of freedom.
  - Intersection: t-value.
- When df is large (∞), t-values $\approx$ z-values.



df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	df
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
∞	1.282	1.645	1.960	2.326	2.576	∞

t-based Confidence Interval for Population μ (σ unknown)

Use Student's t distribution as the model for $\bar{x}$.

Two-sided Confidence Interval:

$$\underbrace{\bar{x}}_{\text{Point Estimate}} \pm \overbrace{\underbrace{t_{\frac{\alpha}{2}}}_{\text{Critical Value}} \times \underbrace{\frac{s}{\sqrt{n}}}_{\text{Standard Error}}}^{\text{Margin of Error}} = \left(\bar{x} - t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right)$$

Interpretation of the $(1-\alpha)100\%$ Confidence Interval for μ

- Interpretation 1: In case of repeated sampling, $(1-\alpha)100\%$ of the confidence intervals constructed in such a manner is expected to contain the population mean, μ.
- Interpretation 2: It can be said with $(1-\alpha)100\%$ confidence that the confidence interval captures/contains population mean, μ.

Example: Parking Garage Revenue

Hoping to lure more shoppers downtown, a city builds a new public parking garage in the central business district. The city plans to pay for the structure through parking fees. For a random sample of 44 week-days, daily fees collected averaged \$126, with a standard deviation of \$15.

Questions:

1. Find a 90% confidence interval for the mean daily income this parking garage will generate.
2. Interpret the 90% confidence interval.
3. Consultant predicted that parking revenues would average \$128 per day. Is this plausible?

Example: Solutions

Here $\alpha = 1 - 0.9 = 0.1 \Rightarrow \alpha/2 = 0.05$. Thus, $t_{\alpha/2} = t_{0.05}$ with degrees of freedom $n - 1 = 44 - 1 = 43$ is approximately identical to $z_{0.05} = 1.645$.

1. $$\bar{x} \pm t_{0.05,43} \cdot \frac{s}{\sqrt{n}} = 126 \pm 1.645 \cdot \frac{15}{\sqrt{44}}$$
$$= 126 \pm 3.72 = (122.28, 129.72)$$
2. **Interpretation:** We are 90% confident the true mean daily income of parking garage is between \$122.28 and \$129.72.
3. \$128 is within the interval $\rightarrow$ plausible prediction value.

Sample Size Determination (σ unknown)

- **Sample size formula:**

$$n = \left(\frac{t_{n-1} \cdot s}{E} \right)^2$$

  where E is the margin of error.
- Always round n up to the nearest whole number (if the results is not an integer).
- **Problem:** Need s before collecting data.
- **Solution:**
  - Use a "good guess" for s.
  - Or run a pilot study.
  - Or adopt the value of s from previous studies that are similar to the present study.

Confidence Interval for a Population Proportion

Generic formula: The general confidence interval (CI) can be expressed in terms of the margin of error (E) as follows:

$$\begin{aligned} &\text{Point Estimate} \pm \text{Margin of Error} \\ &= \text{Point Estimate} \pm \underbrace{\text{Critical Value} \times \text{Standard Error}}_{\text{Margin of Error}} \end{aligned}$$

Two-sided CI for Population Proportion (p):

$$\begin{aligned} &\hat{p} \pm z_{\frac{\alpha}{2}} \times SE(\hat{p}) \\ &= \underbrace{\hat{p}}_{\text{Point Estimate}} \pm \underbrace{z_{\frac{\alpha}{2}}}_{\text{Critical value}} \times \underbrace{\sqrt{\frac{\hat{p}\hat{q}}{n}}}_{\text{Standard Error}} \end{aligned}$$

Width of CI:

$$\text{Width} = 2 \cdot z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(\hat{q})}{n}} = 2 \times E, \qquad \text{where } \hat{q} = 1 - \hat{p}$$

Example: Gallup Poll on Economy

A Gallop Poll found that 1495 out of 3559 respondents thought economic conditions were getting better. We want to construct a 95% confidence interval for p (true proportion).

Solution: The sample proportion:

$$\hat{p} = \frac{1495}{3559} \approx 0.42$$

and the 95% confidence interval for p:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \times \sqrt{\frac{\hat{p}\hat{q}}{n}} = \underbrace{0.42}_{\text{Point Estimate}} \pm \underbrace{1.96}_{\text{Critical Value}} \times \underbrace{\sqrt{\frac{(0.42)(0.58)}{3559}}}_{\text{Standard Error}}$$

$$= 0.42 \pm \underbrace{0.016}_{\text{Margin of Error}} = (\underbrace{0.404}_{\text{Lower Limit}}, \underbrace{0.436}_{\text{Upper Limit}})$$

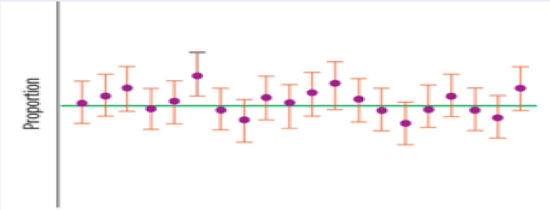
We are 95% confident that the true population proportion p lies between 0.404 and 0.436.

What Does 95% Really Mean? (Long-run Approach)

Interpretation: Long-run Approach

Below we see the confidence intervals produced by simulating 20 samples.

- The purple dots are the simulated proportions of adults who thought the economy was improving.
- The orange segments show each sample's confidence intervals.
- The green line represents the true proportion of the entire population.



Most CIs capture p, but one misses (i.e., $1/20 = 0.05$, consistent with a 95% confidence level).

Interpretation: Long-run Approach: If we were to repeatedly take samples and construct 95% confidence intervals in the same way, about 95% of those intervals would contain the true value of p.

Interpretation of the $(1-\alpha)100\%$ Confidence Interval for p

- Interpretation 1: In case of repeated sampling, $(1-\alpha)100\%$ of the confidence intervals constructed in such a manner is expected to contain the population proportion, p.

- Interpretation 2: It can be said with $(1-\alpha)100\%$ confidence that the confidence interval captures/contains population proportion, p.

Critical Value for Proportions

- Critical Value (CV): To adjust the confidence level, the critical value $(z_{\alpha/2})$ must be updated to correspond to the new level. For any given confidence level, the critical value represents the number of Standard Errors (SE) that must be stretched out on either side of the point estimate $(\hat{p})$ to form the interval.
- Working Approach:
  1. For given confidence level (e.g., 90%, 95%), find $\alpha = 1 -$ confidence level.
  2. For two-sided confidence interval, divide α by 2 and find the critical value corresponding to $\dfrac{\alpha}{2}$ from Standard Normal (z) table.

Assumptions and Conditions for Proportion CI

- **Independence:** The sampled values must be independent of each other.
- **Randomization:** Data must be collected randomly.
- **5% Condition:** When the sample is drawn without replacement, the sample size, n, should be no more than 5% of the population (i.e., $n \leq 0.05N$, where N is the population size)
- **Success/Failure Condition:**

$$n\hat{p} \geq 5, \quad n(1-\hat{p}) \geq 5$$

Example: College Soccer Players

Of a random sample of $n = 150$ college students, 104 of the students said that they had played on a soccer team during their K-12 years. Estimate the proportion of college students who played soccer in their youth with a 98% confidence interval.

Solution:

- Check conditions:

$$n\hat{p} = 150 \times \frac{104}{150} = 104 > 5 \qquad \text{and} \qquad n\hat{q} = 150 \times \frac{46}{150} = 46 > 5$$

$$\text{Thus, } \hat{p} \sim N\left(p, \sqrt{\frac{\hat{p}\hat{q}}{n}}\right) \text{ using CLT}$$

- The 98% Two-sided Confidence Interval for p:

$$\begin{aligned} \hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}} &\approx 0.69 \pm 2.33 \cdot \sqrt{\frac{0.69 \cdot 0.31}{150}} \\ &= 0.693 \pm 0.089 = (0.604, 0.782) \end{aligned}$$

where
$\hat{p} = \frac{104}{150} \approx 0.69, \quad \hat{q} = \frac{46}{150} \approx 0.31, \quad \alpha = 1 - 0.98 = 0.02 \Rightarrow z_{\alpha/2} = z_{0.01} = 2.33.$

Sample Size for Proportion Estimation

- Formula:

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p^\star q^\star$$

  where E is the margin of error.
- Always round n up.
- **How to get $p^\star$ before sampling?**
  - Method 1 (no prior information): Use $p^\star = 0.5$.
  - Method 2: Use prior estimate from pilot/previous study.

Sample Size Example 1: New Service

Suppose a company wants to offer a new service and wants to estimate, to within 3%, the proportion of customers who are likely to purchase this new service with 95% confidence. How large a sample do they need?

Solution:

- Step 1: Margin of error (E) $= 0.03$.
- Step 2: No prior info about $p^\star \rightarrow$ use $p^\star = 0.5$.
- Step 3: 95% confidence interval $\rightarrow \alpha = 0.05$ and $z_{\alpha/2} = z_{0.025} = 1.96$.
- Step 4: Plug in the numbers using the formula

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p^\star q^\star = \left(\frac{1.96}{0.03}\right)^2 0.5 \times 0.5 = 1067.1$$

- Step 5: Round up the calculated value of n

$$n \approx 1068$$

- Interpretation: The company will need at least 1068 respondents to keep the margin of error as small as 3% with confidence level 95%.

Sample Size Example 2: PVC Pipe Producer

A producer of PVC pipe wants to survey wholesalers who buy his product in order to estimate the proportion who plan to increase their purchases next year. What sample size is required if he wants his estimate to be within 0.04 of the actual proportion with probability equal to 0.95?

Solution:

- Step 1: Margin of error (E) $= 0.04$.
- Step 2: No prior info about $p^\star \to$ use $p^\star = 0.5$.
- Step 3: 95% confidence interval $\to \alpha = 0.05$ and $z_{\alpha/2} = z_{0.025} = 1.96$.
- Step 4: Plug in the numbers using the formula

$$n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p^\star q^\star = \left(\frac{1.96}{0.04}\right)^2 0.5 \times 0.5 = 600.25$$

- Step 5: Round up the calculated value of n

$$n \approx 601$$

- Interpretation: At least 601 respondents should be surveyed.

Summary: Confidence Interval Formulas and Rules

1. Confidence Interval for Population Mean μ

- **When σ is known:**

$$\boxed{\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

  - Use z-distribution (Normal) - Requires: $n \geq 30$ or population Normal

- **When σ is unknown:**

$$\boxed{\bar{x} \pm t_{\alpha/2, n-1} \frac{s}{\sqrt{n}}}$$

  - Use t-distribution with $df = n - 1$ - Requires: population Normal (or $n \geq 30$ for approximation)

2. Confidence Interval for Population Proportion p

$$\boxed{\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}}$$

Assumptions:

- Random sample
- Independence
- Success/Failure: $n\hat{p} \geq 5$ and $n(1 - \hat{p}) \geq 5$

Summary: Key Concepts and Rules (Continued)

3. Margin of Error (E) and Sample Size Determination

- **Margin of Error:**

$$E = \text{Critical Value} \times \text{Standard Error}$$

- **Sample Size for Mean (σ known):**

$$\boxed{n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2}$$

- **Sample Size for Proportion:**

$$\boxed{n = \left(\frac{z_{\alpha/2}}{E}\right)^2 p^* q^*}$$

- Use $p^* = 0.5$ when no prior information - Always round n UP to nearest integer

4. General Rules

- **Formula:** Point Estimate $\pm$ Margin of Error
- **Interpretation:** "We are $(1 - \alpha)100\%$ confident that the true parameter lies within the interval"
- **Critical Values:** $z_{0.025} = 1.96$ (for 95% CI); $z_{0.05} = 1.645$ (for 90% CI); t-values depend on degrees of freedom
- **Width vs. Precision:** Higher confidence $\Rightarrow$ wider interval $\Rightarrow$ less precision