A researcher claimed that there are 10% of the a large population have disease H.

A random sample of 5 people is taken from this population and examined.

If 4 people in this random sample have the disease, what does it mean? **How likely** would this happen if the research is right?
What is probability?

What’s the probability of getting a head on the toss of a single fair coin? Use a scale from 0 (no way) to 1 (sure thing).

So toss a coin twice. Do it! Did you get one head & one tail? What’s it all mean?
A rough definition: (frequentist definition)

Probability of event A is the proportion of times that the event A would occur in a very long series of repetitions of a random experiment.

Common sense?!
Many Repetitions!

Total Heads / Number of Tosses

Number of Tosses

Lecture 2 - 4
Probability Distribution

If a balanced coin is tossed, Head and Tail are equally likely to occur,

\[ P(\text{Head}) = 0.5 = \frac{1}{2} \quad \text{and} \quad P(\text{Tail}) = 0.5 = \frac{1}{2} \]

\[ P(\text{all possible outcomes}) = P(\text{Head or Tail}) = P(\text{Head}) + P(\text{Tail}) = \frac{1}{2} + \frac{1}{2} = 1.0 \]

Total probability is 1.
What is the probability distribution of die?

If outcomes are equally likely to occur, the distribution is

\[ P(1) = \frac{1}{6}, \quad P(2) = \frac{1}{6}, \]
\[ P(3) = \frac{1}{6}, \quad P(4) = \frac{1}{6}, \]
\[ P(5) = \frac{1}{6}, \quad P(6) = \frac{1}{6}, \]

and total probability is 1.
Uniform Distribution

Probability Density (Mass) Distribution

- Bars for values 0, 1, 2, 3, 4, 5
- Y-axis ranges from 0 to 0.2
- X-axis values: 0, 1, 2, 3, 4, 5
- Values on the bars: 0.15, 0.20, 0.15, 0.15, 0.20, 0.20
Properties of Probability

- Probability is always a value between 0 and 1.
- Total probability equals 1.
## Relative Frequency and Probability

**Number of children per household from a sample of 300 households**

<table>
<thead>
<tr>
<th>Class</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>54</td>
<td>.18</td>
</tr>
<tr>
<td>1</td>
<td>117</td>
<td>.39</td>
</tr>
<tr>
<td>2</td>
<td>72</td>
<td>.24</td>
</tr>
<tr>
<td>3</td>
<td>42</td>
<td>.14</td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>.04</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>.01</td>
</tr>
<tr>
<td>Total</td>
<td>300</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Discrete Distribution

Relative Frequency Distribution
Discrete Distribution

If a household is randomly selected from the 300 household, what is the probability that it has more than 3 children?

\[
P(\text{more than 3 children}) = 0.04 + 0.01 = 0.05
\]
Random Variable

A variable that can assume a numerical description of the outcome from a random experiment by chance.

Usually is denoted by a capital letter.
A random variable assumes discrete values.
Discrete Random Variable

Example: (Toss a balanced coin)

X = 1, if Head occurs, and X = 0 if Tail occurs.

P(Head) = P(X=1) = P(1) = .5
P(Tail) = P(X=0) = P(0) = .5
Example: What is probability of getting a number less than 3 when roll a balanced die?

\[ P( X < 3 ) = P( X \leq 2 ) = ? \]

Answer: \( \frac{2}{6} = \frac{1}{3} \)
Why Random Variable?

- A simple mathematical notation to describe an event. e.g.: $X < 3$, $X = 0$, ...
- Mathematical function can be used to model the distribution through the use of random variable. e.g.: Binomial, Poisson, Normal, …
Bernoulli Trial

Definition: Bernoulli trial is a random experiment whose outcomes are classified as one of the two categories. (S, F) or (Success, Failure) or (1, 0)

\[ P(S) = P(X=1) = p, \quad P(F) = P(X=0) = 1 - p. \]

Example: (Head, Tail), (Died, Survived)
Bernoulli Probability

**Example**: In a random experiment of tossing an unbalanced coin, the probability of Head is 0.3, what is the probability distribution?

\[ P(\text{Head}) = P(X=1) = 0.3, \]
\[ P(\text{Tail}) = P(X=0) = 1 - 0.3 = 0.7. \]
Bionomial Experiment

A random experiment involving a sequence of *independent* and *identical* Bernoulli trials.

Example:

- Toss a coin ten times and observing Head or Tail turns up.
- Roll a die 3 times and observing a 6 or not 6 turns up.
In a binomial experiment involving $n$ independent and identical Bernoulli trials each with probability of success $p$, the probability of having $x$ successes can be calculated with the binomial probability mass function, and it is, for $x = 0, 1, \ldots, n$,

$$P(X = x) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x}$$

$$= \frac{n!}{x!(n-x)!} \cdot p^x \cdot (1 - p)^{n-x}$$
Factorial

\[ n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n \]

Example: \[ 3! = 1 \cdot 2 \cdot 3 = 6 \]
**Binomial Distribution**

Parameters of the distribution:

Mean of the distribution, $\mu = n \cdot p$

Variance of the distribution, $\sigma^2 = n \cdot p \cdot (1-p)$

Standard deviation, $\sigma$, is the **square root of variance**.
Binomial Probability

Example: A balanced die is rolled three times (or three balanced dice are rolled), what is the probability to see two 6’s?

\[ \mu = 3 \cdot \frac{1}{6} = \frac{1}{2}, \quad \sigma^2 = 3 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{12} \]

\[ n = 3, \quad p = \frac{1}{6}, \quad x = 2 \]

\[ P(X=2) = \frac{3!}{2!1!} \cdot \left( \frac{1}{6} \right)^2 \cdot \left( \frac{5}{6} \right)^{3-2} \]

\[ = 3 \cdot \left( \frac{1}{6} \right)^2 \cdot \left( \frac{5}{6} \right)^1 \]

\[ = .069 \]
Binomial Probability

**Example:** If there are 10% of the population in a community have a certain disease, what is the probability that 4 people in a random sample of 5 people from this community has the disease?

Identify \( n = 5, \ x = 4, \ p = .10 \)

\[
P(X=4) = \frac{5!}{[4!(5-4)!]} \cdot (.10)^4 \cdot (1-.10)^{5-4}
\]
\[
= 5 \cdot (.10)^4 \cdot (.90)^1
\]
\[
= .00045
\]
Example: In the previous problem, what is the probability that 4 or more people have the disease?

Identify \( n = 5 \), \( x = 4 \) (and also \( x = 5 \)), \( p = .10 \)

\[
P(X \geq 4) = P(X=4) + P(X=5)
\]

\[
= .00045 + \left\{ \frac{5!}{[5!(5-5)!]} \right\} \cdot (.10)^5 \cdot (1-.10)^{5-5}
\]

\[
= .00045 + .00001 = .00046
\]

(What this number is telling us?)
Poisson Distribution

The Poisson distribution is used to model discrete events that occur infrequently in time or space.

Model the number of successes in a given time period or in a given unit space.
Poisson Distribution

Let $X$ represents the number of occurrences of some event of interest over a given interval from a Poisson process, and the $\lambda$ is the mean of the distribution, the probability of $X$ assumes the value $x$ is, for $x = 0, 1, 2, \ldots$,

$$P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Can be used to approximate Binomial prob, with large n.
The probability that a single event occurs within an interval is proportional to the length of the interval.

Within a single interval, an infinite number of occurrences is possible.

The events occurs independently both within the same interval and between consecutive non-overlapping intervals.
A (Simple) **Random Sample** of size \( n \) consists of \( n \) individuals from the population chosen in such a way that every set of \( n \) individuals has an equal chance to be the sample actually selected.
Counting Rule:

**Multiplication Principle:** In a sequence of \( k \) events in which the first one has \( n_1 \) possibilities and the second event has \( n_2 \) and the third has \( n_3 \), and so forth, the total possibilities of the sequence will be 
\[
\times_{k} \prod_{i=1}^{k} n_i
\]
Permutation Rule:

The number of possible permutations of \( r \) objects from a collection of \( n \) distinct objects is

\[
_{n}P_r = \frac{n!}{(n-r)!}
\]
Combination Rule:

The number of possible combinations of \( r \) objects from a collection of \( n \) distinct objects is

\[
_{n}C_r = \left( \begin{array}{c} n \\ r \end{array} \right) = \frac{n!}{(n-r)!r!}
\]
Continuous Random Variables
(Normal Distribution)
Continuous Probability Density Function

1. Mathematical Formula

2. Shows All Values, $x$, & Densities, $f(x)$
   - $f(X)$ Is Not Probability

3. Properties
   \[
   \int_a^b f(x) \, dx = 1
   \]
   All $X$ (Area Under Curve)
   \[
   f(x) \geq 0, \quad a \leq x \leq b
   \]
Continuous Random Variable Probability

Probability is area under curve!

\[ P(c \leq x \leq d) = \int_{c}^{d} f(x) \, dx \]
Normal Distribution
Importance of Normal Distribution

1. Describes Many Random Processes or Continuous Phenomena

2. Can Be Used to Approximate Discrete Probability Distributions
   - Example: Binomial

3. Basis for Classical Statistical Inference
1. ‘Bell-Shaped’ & Symmetrical

2. Mean, Median, Mode Are Equal

3. Random Variable Has Infinite Range
   \[-\infty < X < \infty\]
Normal Probability Density Function

\[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\left(\frac{1}{2}\right) \left(\frac{x-\mu}{\sigma}\right)^2} \]

- \( f(x) \) = Frequency of Random Variable \( x \)
- \( \sigma \) = Population Standard Deviation
- \( \pi \) = 3.14159; \( e \) = 2.71828
- \( x \) = Value of Random Variable (-\( \infty < x < \infty \))
- \( \mu \) = Population Mean
Effect of Varying Parameters ($\mu$ & $\sigma$)
Normal Distribution

Probability

Probability is area under curve!

\[ P(c \leq x \leq d) = \int_{c}^{d} f(x) \, dx \]
Normal distributions differ by mean & standard deviation. Each distribution would require its own table. That’s an infinite number!
Standard Normal Distribution, $\mathcal{N}(\mu = 0, \sigma = 1)$, is a normal distribution with mean 0 and standard deviation 1.

Notation $Z$ is often used to denote Standard Normal random variable.
P(0 < Z < 0.32) = Area between 0 and .32

\[
\begin{array}{c|c|c|c}
Z & .00 & .01 & .02 \\
\hline
0.0 & .500 & .496 & .492 \\
0.1 & .460 & .456 & .452 \\
0.2 & .421 & .417 & .413 \\
0.3 & .382 & .378 & .374 \\
\end{array}
\]

Area = .5 - .374 = .126
Standardize the Normal Distribution

\[ Z = \frac{X - \mu}{\sigma} \]

Normal Distribution

Standardized Normal Distribution

One table!
Standardizing Example

\[ Z = \frac{X - \mu}{\sigma} = \frac{6.2 - 5}{10} = .12 \]

Normal Distribution

\[ \sigma = 10 \]
\[ \mu = 5 \] 6.2  \[ X \]

Standardized Normal Distribution

\[ \sigma = 1 \]
\[ \mu = 0 \]  .12  \[ Z \]
## Obtaining the Probability

### Standardized Normal Probability Table (Portion)

<table>
<thead>
<tr>
<th>$Z$</th>
<th>.00</th>
<th>.01</th>
<th>.02</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>.500</td>
<td>.496</td>
<td>.492</td>
</tr>
<tr>
<td>0.1</td>
<td>.460</td>
<td>.456</td>
<td>.452</td>
</tr>
<tr>
<td>0.2</td>
<td>.421</td>
<td>.417</td>
<td>.413</td>
</tr>
<tr>
<td>0.3</td>
<td>.382</td>
<td>.378</td>
<td>.3745</td>
</tr>
</tbody>
</table>

Area = \( \frac{1}{2} - .452 = .048 \)

\[ \sigma = 1 \]

\( \mu = 0 \)

\( Z = .12 \)
Example

\[ P(3.8 \leq X \leq 5) \]

\[
Z = \frac{X - \mu}{\sigma} = \frac{3.8 - 5}{10} = -.12
\]

Normal Distribution

\[
\sigma = 10
\]

\[
\mu = 5
\]

\[
X
\]

Standardized Normal Distribution

\[
\sigma = 1
\]

\[
\mu = 0
\]

\[
Z
\]

Area = .048
Example

\[ P(2.9 \leq X \leq 7.1) \]

\[ Z = \frac{X - \mu}{\sigma} = \frac{2.9 - 5}{10} = -0.21 \]

\[ Z = \frac{X - \mu}{\sigma} = \frac{7.1 - 5}{10} = 0.21 \]

Area = 0.083 + 0.083 = 0.166
Example

\( P(X > 8) \)

\[ Z = \frac{X - \mu}{\sigma} = \frac{8 - 5}{10} = .30 \]

\[ \sigma = 10 \]

\[ \mu = 5 \]

\[ 8 \]

\[ X \]

\[ \sigma = 1 \]

\[ \mu = 0 \]

Area = .382
Example

\[ P(7.1 \leq X \leq 8) \]

\[
Z = \frac{X - \mu}{\sigma} = \frac{7.1 - 5}{10} = .21
\]

\[
Z = \frac{X - \mu}{\sigma} = \frac{8 - 5}{10} = .30
\]

Normal Distribution

\[ \sigma = 10 \]

\[ \mu = 5 \]

\[ 7.1 \]

\[ 8 \]

\[ X \]

Standardized Normal Distribution

\[ \sigma = 1 \]

\[ \mu = 0 \]

\[ .21 \]

\[ .30 \]

\[ Z \]

Area = .417 - .382 = .035
Normal Distribution
Thinking Challenge

The life time of a medical device has a normal distribution with $\mu = 2000$ hours & $\sigma = 200$ hours. What’s the probability that such a device will last

A. between 2000 & 2400 hours?

B. less than 1470 hours?
Solution*

$P(2000 \leq X \leq 2400)$

$Z = \frac{X - \mu}{\sigma} = \frac{2400 - 2000}{200} = 2.0$

Normal Distribution

$\mu = 2000$  $2400$  $X$

$\sigma = 200$

Standardized Normal Distribution

$\mu = 0$  $2.0$  $Z$

$\sigma = 1$.477
Solution*

\[ P(X \leq 1470) \]

\[ Z = \frac{X - \mu}{\sigma} = \frac{1470 - 2000}{200} = -2.65 \]

Normal Distribution

\( \sigma = 200 \)

\( \mu = 2000 \)

\( X \)

Standardized Normal Distribution

\( \sigma = 1 \)

\( \mu = 0 \)

\( Z \)

\( .004 \)
Finding Z Values for Known Probabilities

What is \( z \) given \( P(Z < z) = 0.622 \)?

Area = 1 - 0.622 = 0.378

\[ z = 0.31 \]

Standardized Normal Probability Table (Portion)

<table>
<thead>
<tr>
<th>Z</th>
<th>.00</th>
<th>.01</th>
<th>.02</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.500</td>
<td>0.496</td>
<td>0.492</td>
</tr>
<tr>
<td>0.1</td>
<td>0.460</td>
<td>0.456</td>
<td>0.452</td>
</tr>
<tr>
<td>0.2</td>
<td>0.421</td>
<td>0.417</td>
<td>0.413</td>
</tr>
<tr>
<td>0.3</td>
<td>0.382</td>
<td>0.378</td>
<td>0.3745</td>
</tr>
</tbody>
</table>

Lecture 2 - 55
Finding X Values for Known Probabilities

Normal Distribution

Standardized Normal Distribution

\[ X = \mu + Z \cdot \sigma = 5 + (0.31)(10) = 8.1 \]
Example: Height of infants is normally distributed with a mean 7 lb and standard deviation of 1.2 lb. Find the 90th percentile.

Area to the left of 90th percentile is 0.100. In the table there is an area value 0.100 corresponding to a z-score of 1.28.

90th percentile = 7 + 1.28 \times 1.2 = 8.536 \text{ lb}
Normal Approximation of Binomial Distribution

1. Not All Binomial Tables Exist

2. Requires Large Sample Size

3. Gives Approximate Probability Only, By Normal Distribution with $\mu = np, \sigma^2 = np(1-p)$

4. Need Correction for Continuity

$n = 10 \; p = 0.50$
Sampling Distributions
Inferential Statistics

1. Involves:
   - Estimation
   - Hypothesis Testing

2. Purpose
   - Make Decisions about Population Characteristics
Inference Process

Estimates & tests
Sample statistic \(X\)
Population
Sample

Lecture 2 - 61
1. Statistics (Random Variables) Used to Estimate a Population Parameter
   - Sample Mean, Sample Proportion, Sample Median
   - Example: Sample Mean $\bar{X}$ is an Estimator of Population Mean $\mu$
     - If $\bar{X} = 3$ then 3 Is the Estimate of $\mu$

2. Theoretical Basis Is Sampling Distribution
Sampling Distribution

Theoretical Probability Distribution of the Sample Statistic.
Standard Error of Mean

1. Standard Deviation of the sampling distribution of the Sample Means, \( \bar{X} \)
   - Measures Scatter in All Sample Means, \( \bar{X} \)

2. Less Than Pop. Standard Deviation

3. Formula (Sampling With Replacement)
   \[
   \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}
   \]
Properties of Sampling Distribution of Mean

1. **Unbiasedness**
   - Mean of Sampling Distribution Equals Population Mean

2. **Efficiency**
   - Sample Mean Comes Closer to Population Mean Than Any Other Unbiased Estimator

3. **Consistency**
   - As Sample Size Increases, Variation of Sample Mean from Population Mean Decreases
Unbiasedness

\[ P(\bar{X}) \]

Unbiased

Biased

A

\[ \mu \]

C

Lecture 2 - 66
Efficiency

Sampling distribution of mean

Sampling distribution of median

$P(\bar{X})$
Consistency

$P(\bar{X})$

Larger sample size

Smaller sample size

$\mu$

A

B

Lecture 2 - 68
Sampling From Normal Population

If a random sample is taken from a normally distributed population that has a mean $\mu$ and a standard deviation $\sigma$, the sampling distribution of the sample means is normal with

$$
\mu_{\bar{x}} = \mu
$$

$$
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}
$$
Sampling from Normal Populations

Central Tendency
\[ \mu_{\bar{X}} = \mu \]

Dispersion
\[ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \]

Population Distribution
- \( \mu = 50 \)
- \( \sigma = 10 \)

Sampling Distribution
- \( n = 4 \)  
  \( \sigma_{\bar{X}} = 5 \)
- \( n = 16 \)  
  \( \sigma_{\bar{X}} = 2.5 \)
Standardizing Sampling Distribution of Mean

$$Z = \frac{X - \mu_x}{\sigma_x} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Sampling Distribution

Standardized Normal Distribution

$\sigma = 1$

$\mu = 0$

$Z$

$Lecture 2 - 71$
Thinking Challenge

Waiting times at a certain type of clinic are normally distributed with $\mu = 8$ min. & $\sigma = 2$ min. If you select random samples of 25 cases, what is the sampling distribution of the mean? What is the probability that the sample mean would be between 7.8 & 8.2 minutes?
The sampling distribution of the mean is normal distribution with mean = 8, and standard deviation = $2/5 = .4$
Sampling Distribution

Solution*

\[
Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{7.8 - 8}{2/\sqrt{25}} = -0.50
\]

\[
Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{8.2 - 8}{2/\sqrt{25}} = 0.50
\]

Sampling Distribution of sample mean

\[\sigma_{\bar{X}} = 0.4\]

Standardized Normal Distribution

\[\sigma = 1\]

\[-0.50 \quad 0 \quad 0.50 \quad Z\]

\[0.383\]
Central Limit Theorem

If a relatively large random sample is taken from a population that has a mean \( \mu \) and a standard deviation \( \sigma \), regardless of the distribution of the population, the distribution of the sample means is approximately normal with

\[
\mu_{\bar{x}} = \mu \\
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}
\]
Central Limit Theorem

As sample size gets large enough \((n \geq 30)\) ... the sampling distribution becomes almost normal.

\[
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}
\]

\(\mu_{\bar{x}} = \mu\)
Sampling from Non-Normal Populations

Central Tendency
\[ \mu_{\bar{X}} = \mu \]

Dispersion
\[ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \]

- Sampling with replacement

Population Distribution
\[ \sigma = 10 \]
\[ \mu = 50 \]
\[ X \]

Sampling Distribution
\[ \sigma_{\bar{X}} = 5 \quad \text{for } n = 4 \]
\[ \sigma_{\bar{X}} = 1.8 \quad \text{for } n = 30 \]

\[ \mu_{\bar{X}} = 50 \]
\[ \bar{X} \]
Example: Consider the distribution of serum cholesterol levels for all 20- to 74-year-old males living in United States has a mean of 211 mg/100 ml, and the standard deviation of 46 mg/100 ml. If a random sample of 100 individuals from the population, what is the probability that the average serum cholesterol level of these 100 individuals is higher than 225?
Solution:

Since n = 100, the sampling distribution of the mean is approximately normal with mean 211 and standard error 4.6 (= 46/10).

\[ P(X > 225) = P(Z > \frac{225 - 211}{4.6}) \]
\[ = P(Z > 3.04) \]
\[ = 0.001 \]
Inferences Based on a Single Sample

Estimation with Confidence Intervals
Estimation Process

Population

Mean, $\mu$, is unknown

Sample

Random Sample

Mean $\bar{X} = 50$

I am 95% confident that $\mu$ is between 40 & 60.
## Unknown Population Parameters Are Estimated

<table>
<thead>
<tr>
<th>Estimate Population Parameter...</th>
<th>with Sample Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ( \mu )</td>
<td>( \bar{x} )</td>
</tr>
<tr>
<td>Proportion ( p )</td>
<td>( \hat{p} )</td>
</tr>
<tr>
<td>Variance ( \sigma^2 )</td>
<td>( s^2 )</td>
</tr>
<tr>
<td>Differences ( \mu_1 - \mu_2 )</td>
<td>( \bar{x}_1 - \bar{x}_2 )</td>
</tr>
</tbody>
</table>
Point Estimation

1. Provides Single Value
   - Based on Observations from 1 Sample

2. Gives No Information about How Close Value Is to the Unknown Population Parameter

3. Example: Sample Mean $\bar{X} = 3$ Is Point Estimate of Unknown Population Mean
Interval Estimation

1. Provides Range of Values
   - Based on Observations from 1 Sample

2. Gives Information about Closeness to Unknown Population Parameter
   - Stated in terms of Probability
     - Knowing Exact Closeness Requires Knowing Unknown Population Parameter

3. Example: Unknown Population Mean Lies Between 50 & 70 with 95% Confidence
Key Elements of
Interval Estimation

A probability that the population parameter falls somewhere within the interval.
Confidence Limits for Population Mean

Parameter = Statistic ± Error

1. \( \mu = \bar{X} \pm \text{Error} \)
2. \( \text{Error} = \bar{X} - \mu \) or \( \bar{X} + \mu \)
3. \( Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{\text{Error}}{\sigma_{\bar{X}}} \)
4. \( \text{Error} = Z\sigma_{\bar{X}} \)
5. \( \mu = \bar{X} \pm Z\sigma_{\bar{X}} \)
Many Samples Have Same Interval

\[ \overline{X} = \mu \pm Z \sigma_{\overline{X}} \]

- 90% Samples: \( \mu - 2.58 \sigma_{\overline{X}} \) to \( \mu + 2.58 \sigma_{\overline{X}} \)
- 95% Samples: \( \mu - 1.96 \sigma_{\overline{X}} \) to \( \mu + 1.96 \sigma_{\overline{X}} \)
- 99% Samples: \( \mu - 1.65 \sigma_{\overline{X}} \) to \( \mu + 1.65 \sigma_{\overline{X}} \)
Confidence Level

1. Probability that the Unknown Population Parameter Falls Within Interval

2. Denoted \((1 - \alpha)\)%
   - \(\alpha\) is Probability That Parameter Is Not Within Interval

3. Typical Values Are 99%, 95%, 90%
Intervals & Confidence Level

Sampling Distribution of Mean

\[
\mu_{\bar{X}} = \mu
\]

Intervals extend from \( \bar{X} - Z\sigma_{\bar{X}} \) to \( \bar{X} + Z\sigma_{\bar{X}} \)

(1 - \( \alpha \)) % of intervals contain \( \mu \).
\( \alpha \) % do not.

Large number of intervals
Factors Affecting Interval Width

1. Data Dispersion
   - Measured by $\sigma$

2. Sample Size
   - $\sigma_{\bar{x}} = \sigma / \sqrt{n}$

3. Level of Confidence
   - $(1 - \alpha)$
   - Affects $Z$

Intervals Extend from $\bar{X} - Z\sigma_{\bar{x}}$ to $\bar{X} + Z\sigma_{\bar{x}}$
Confidence Interval Estimates

Confidence Intervals

Mean

- $\sigma$ Known
- $\sigma$ Unknown

Proportion

Variance
Confidence Interval
Mean (σ Known)

1. Assumptions

- Population Standard Deviation Is Known
- Population Is Normally Distributed
- If Not Normal, Can Be Approximated by Normal Distribution \((n \geq 30)\)
Confidence Interval

Mean (σ Known)

1. Assumptions
   - Population Standard Deviation Is Known
   - Population Is Normally Distributed
   - If Not Normal, Can Be Approximated by Normal Distribution (n ≥ 30)

2. Confidence Interval Estimate
   \[
   X - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq X + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}
   \]
Estimation Example
Mean (σ Known)

The mean of a random sample of $n = 25$ is $\bar{X} = 50$. Set up a 95% confidence interval estimate for $\mu$ if $\sigma = 10$.

\[
\bar{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}
\]

\[
50 - 1.96 \cdot \frac{10}{\sqrt{25}} \leq \mu \leq 50 + 1.96 \cdot \frac{10}{\sqrt{25}}
\]

\[
46.08 \leq \mu \leq 53.92
\]
Thinking Challenge

You’re a Q/C inspector for Gallo. The $\sigma$ for 2-liter bottles is .05 liters. A random sample of 100 bottles showed $\bar{X} = 1.99$ liters. What is the 90% confidence interval estimate of the true mean amount in 2-liter bottles?
Confidence Interval

Solution*

\[
\bar{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}
\]

\[
1.99 - 1.645 \cdot \frac{.05}{\sqrt{100}} \leq \mu \leq 1.99 + 1.645 \cdot \frac{.05}{\sqrt{100}}
\]

\[
1.982 \leq \mu \leq 1.998
\]
Confidence Interval
Mean (σ Unknown)

1. Assumptions
   - Population Standard Deviation Is Unknown
   - Population Must Be Normally Distributed

2. Use Student’s t Distribution
Confidence Interval Mean (σ Unknown)

1. Assumptions
   - Population Standard Deviation Is Unknown
   - Population Must Be Normally Distributed

2. Use Student’s t Distribution

3. Confidence Interval Estimate

   \[
   \bar{X} - t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{\alpha/2,n-1} \cdot \frac{S}{\sqrt{n}}
   \]
Student's $t$ Distribution

- Bell-Shaped
- Symmetric
- 'Fatter' Tails

- Standard Normal
- $t$ ($df = 5$)
- $t$ ($df = 13$)

Lecture 2 - 99
## Student's $t$ Table

<table>
<thead>
<tr>
<th>$v$</th>
<th>$t_{.10}$</th>
<th>$t_{.05}$</th>
<th>$t_{.025}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.078</td>
<td>6.314</td>
<td>12.706</td>
</tr>
<tr>
<td>2</td>
<td>1.886</td>
<td>2.920</td>
<td>4.303</td>
</tr>
<tr>
<td>3</td>
<td>1.638</td>
<td>2.353</td>
<td>3.182</td>
</tr>
</tbody>
</table>

Assume:
- $n = 3$
- $df = n - 1 = 2$
- $\alpha = .10$
- $\alpha/2 = .05$

The $t$ value for $\alpha/2 = .05$ and $df = 2$ is 2.920.
Degrees of Freedom (df)

1. Number of Observations that Are Free to Vary After Sample Statistic Has Been Calculated

2. Example
   - Sum of 3 Numbers Is 6
     \[ X_1 = 1 \text{ (or Any Number)} \]
     \[ X_2 = 2 \text{ (or Any Number)} \]
     \[ X_3 = 3 \text{ (Cannot Vary)} \]
     Sum = 6

   degrees of freedom
   \[ = n - 1 \]
   \[ = 3 - 1 \]
   \[ = 2 \]
Estimation Example
Mean (σ Unknown)

A random sample of \( n = 25 \) has \( \bar{x} = 50 \) & \( s = 8 \). Set up a 95% confidence interval estimate for \( \mu \).

\[
\bar{x} - t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\alpha/2, n-1} \cdot \frac{s}{\sqrt{n}}
\]

\[
50 - 2.0639 \cdot \frac{8}{\sqrt{25}} \leq \mu \leq 50 + 2.0639 \cdot \frac{8}{\sqrt{25}}
\]

\[
46.69 \leq \mu \leq 53.30
\]
Thinking Challenge

You’re a time study analyst in manufacturing. You’ve recorded the following task times (min.): 3.6, 4.2, 4.0, 3.5, 3.8, 3.1.

What is the 90% confidence interval estimate of the population mean task time?
Confidence Interval Solution*

\( \bar{X} = 3.7 \)

\( S = 3.8987 \)

\( n = 6, \quad df = n - 1 = 6 - 1 = 5 \)

\( S / \sqrt{n} = 3.8987 / \sqrt{6} = 1.592 \)

\( t_{0.05,5} = 2.0150 \)

\( 3.7 - (2.015)(1.592) \leq \mu \leq 3.7 + (2.015)(1.592) \)

\( 0.492 \leq \mu \leq 6.908 \)
Confidence Interval

Proportion

1. Assumptions
   - Two Categorical Outcomes
   - Population Follows Binomial Distribution
   - Normal Approximation Can Be Used
     \[ n\hat{p} \pm 3\sqrt{n\hat{p}(1 - \hat{p})} \] Does Not Include 0 or 1
Confidence Interval
Proportion

1. Assumptions
   - Two Categorical Outcomes
   - Population Follows Binomial Distribution
   - Normal Approximation Can Be Used
     - \( n\hat{p} \pm 3\sqrt{n\hat{p}(1 - \hat{p})} \) Does Not Include 0 or 1

2. Confidence Interval Estimate
   \[
   \hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}
   \]
A random sample of 400 graduates showed 32 went to grad school. Set up a 95% confidence interval estimate for \( p \).

\[
\hat{p} - Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \leq p \leq \hat{p} + Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}
\]

\[
.08 - 1.96 \cdot \sqrt{\frac{.08 \cdot (1 - .08)}{400}} \leq p \leq .08 + 1.96 \cdot \sqrt{\frac{.08 \cdot (1 - .08)}{400}}
\]

\[
.053 \leq p \leq .107
\]
Thinking Challenge

You’re a production manager for a newspaper. You want to find the % defective. Of 200 newspapers, 35 had defects. What is the 90% confidence interval estimate of the population proportion defective?
Confidence Interval
Solution*

\[
\hat{p} - z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}
\]

\[
.175 - 1.645 \cdot \sqrt{\frac{.175 \cdot (.825)}{200}} \leq p \leq .175 + 1.645 \cdot \sqrt{\frac{.175 \cdot (.825)}{200}}
\]

\[
.1308 \leq p \leq .2192
\]
Finding Sample Sizes for Estimating $\mu$

(1) $Z = \frac{X - \mu}{\sigma_x} = \frac{\text{Error}}{\sigma_x}$

(2) $\text{Error} = Z \sigma_x = Z \frac{\sigma}{\sqrt{n}}$

(3) $n = \frac{Z^2 \sigma^2}{\text{Error}^2}$

Error Is Also Called Bound, $B$
Sample Size Example

What sample size is needed to be 90% confident of being correct within ± 5? A pilot study suggested that the standard deviation is 45.

\[ n = \frac{Z^2 \sigma^2}{\text{Error}^2} = \frac{(1.645)^2(45)^2}{(5)^2} = 219.2 \approx 220 \]
Thinking Challenge

You work in Human Resources at Merrill Lynch. You plan to survey employees to find their average medical expenses. You want to be 95% confident that the sample mean is within ± $50. A pilot study showed that σ was about $400. What sample size do you use?
Sample Size Solution*

\[ n = \frac{Z^2 \sigma^2}{\text{Error}^2} \]

\[ = \frac{(1.96)^2 (400)^2}{(50)^2} \]

\[ = 245.86 \approx 246 \]