Central Limit Theorem

&

Normal Distribution
What is a **Sampling Distribution**?

It is a probability distribution of a statistic obtained through a large number of samples drawn from a specific population.

What is **Statistic**?

**Statistic** is any numerical measurement related to a sample.

Here are a couple of examples of statistics:

- Sample mean $\bar{X}$.
- Sample proportion $\hat{p}$. 
What is the **Central Limit Theorem**?

It is the conclusion of the sampling distribution of $\bar{X}$ from any population with mean $\mu$ and variance $\sigma^2$ when random samples of size $n$ are drawn from.

The sampling distribution of $\bar{X}$

- is approximately normally distributed

  with

- mean $\mu_{\bar{X}} = \mu$,

- variance $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$, and

- standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$. 
Example:

Consider a discrete population consisting of values 2, 4, 6, 8 and 10.

- Find \( \mu \) and \( \sigma^2 \).
- List all possible samples of size 2 with replacement.
- Find the mean of each samples.
- Construct a table that contains the mean of each samples and the probability of each mean.
- Draw the probability histogram using the mean of each sample and the probability of each mean.
- Show that the probability histogram has a shape of a normal curve.
Solution:

- Find $\mu$ and $\sigma^2$.
  We can simply enter these values in $L_1$ and perform basic statistical computations.

  $\Rightarrow \mu = 6$, $\sigma = 2.828$, and $\sigma^2 = 8$

- List all possible samples of size 2 with replacement.

  | 2,2 | 4,2 | 6,2 | 8,2 | 10,2 |
  | 2,4 | 4,4 | 6,4 | 8,4 | 10,4 |
  | 2,6 | 4,6 | 6,6 | 8,6 | 10,6 |
  | 2,8 | 4,8 | 6,8 | 8,8 | 10,8 |
  | 2,10| 4,10| 6,10| 8,10| 10,10|
Solution Continued:

- Find the mean of each samples.
  
<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

- Construct a table that contains the mean of each samples and the probability of each mean.

<table>
<thead>
<tr>
<th>$\bar{x}$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(\bar{x})$</td>
<td>$\frac{1}{25}$</td>
<td>$\frac{2}{25}$</td>
<td>$\frac{3}{25}$</td>
<td>$\frac{4}{25}$</td>
<td>$\frac{5}{25}$</td>
<td>$\frac{4}{25}$</td>
<td>$\frac{3}{25}$</td>
<td>$\frac{2}{25}$</td>
<td>$\frac{1}{25}$</td>
</tr>
</tbody>
</table>
Solution Continued:

- Draw the probability histogram using $\bar{x}$ and $P(\bar{x})$. 

![Histogram showing the sampling distribution of $\bar{x}$]
Solution Continued:

Show that the probability histogram has a shape of a normal curve.
Example:
The probability distribution chart below displays sampling distribution of $\bar{x}$ with samples of size 2 from our last example.

<table>
<thead>
<tr>
<th>$\bar{x}$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(\bar{x})$</td>
<td>$\frac{1}{25}$</td>
<td>$\frac{2}{25}$</td>
<td>$\frac{3}{25}$</td>
<td>$\frac{4}{25}$</td>
<td>$\frac{5}{25}$</td>
<td>$\frac{4}{25}$</td>
<td>$\frac{3}{25}$</td>
<td>$\frac{2}{25}$</td>
<td>$\frac{1}{25}$</td>
</tr>
</tbody>
</table>

Use the discrete probability distribution

- to find $\mu_{\bar{x}}$, $\sigma_{\bar{x}}$,
- the exact value of $\sigma_{\bar{x}}^2$, and
- use these results to verify the conclusion of the Central Limit Theorem.
Solution:

- Using $L_1$ and $L_2$ for $\bar{x}$ and $P(\bar{x})$ respectively. Now we can perform basic statistical computation, we get

  \[ \Rightarrow \mu_{\bar{x}} = 6 \quad \text{and} \quad \sigma_{\bar{x}} = 2 \]

- Now we simply use the formula $\sigma = \sqrt{\sigma^2}$.

  \[ \Rightarrow \sigma^2_{\bar{x}} = 2^2 = 4 \]

- Use these results to verify the conclusion of the Central Limit Theorem.

  We can verify that $\mu_{\bar{x}} = \mu = 6$, and $\sigma^2_{\bar{x}} = \frac{\sigma^2}{n} = \frac{8}{2} = 4$. 
**Example:**

Use sampling distribution of $\bar{x}$ when samples of size 16 are selected at random from a normally distributed population with mean 375 and variance 100.

- Find $\mu_{\bar{x}}$.
- Find $\sigma^2_{\bar{x}}$.

**Solution:**

Using the Central Limit Theorem,

- Find $\mu_{\bar{x}}$. $\Rightarrow \mu_{\bar{x}} = \mu = 375$
- Find $\sigma^2_{\bar{x}}$. $\Rightarrow \sigma^2_{\bar{x}} = \frac{\sigma^2}{n} = \frac{100}{16} = 6.25$
Example:
Use sampling distribution of $\bar{x}$ when samples of size 10 are selected at random from a normally distributed population with mean 82 and standard deviation 7.5.

- Find $\mu_{\bar{x}}$.
- Find $\sigma_{\bar{x}}$.

Solution:
Using the Central Limit Theorem,

- Find $\mu_{\bar{x}}$.  $\Rightarrow \mu_{\bar{x}} = \mu = 82$
- Find $\sigma_{\bar{x}}$.  $\Rightarrow \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{7.5}{\sqrt{10}} \approx 2.372$. 
we know that $z = \frac{x - \mu}{\sigma}$, now we can replace $x$ with $\bar{x}$, $\mu$ with $\mu_{\bar{x}}$, $\sigma$ with $\sigma_{\bar{x}}$, and simplify using the central limit theorem.

$$z = \frac{x - \mu}{\sigma}$$

$$= \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$$= \frac{\bar{x} - \mu}{\sigma} \frac{1}{\sqrt{n}}$$
Example:

Use sampling distribution of $\bar{x}$ when samples of size 36 are selected at random from a normally distributed population with mean 6250 and standard deviation 275.

- Find the $z$ score for $\bar{x} = 6450$.
- Find the $z$ score for $\bar{x} = 6200$.

Solution:

Using the formula $z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n}$,

- Find the $z$ score for $\bar{x} = 6450$. $\Rightarrow z = \frac{6450 - 6250}{275} \frac{275}{\sqrt{36}} \approx 4.364$
- Find the $z$ score for $\bar{x} = 5820$. $\Rightarrow z = \frac{6200 - 6250}{275} \frac{275}{\sqrt{36}} \approx -1.091$
Example:

The average life of a certain blender is 5.1 years with a standard deviation of 1.2 years. Assume the lives of these blenders are normally distributed.

- Find the probability that a mean life of a random sample of 9 such blenders fall between 4.5 and 5.5 years.
- Find the value of $\bar{x}$ that separates the top 10% from the rest of the means computed from random samples of size 9.

Solution:

We have a normal probability distribution with $\mu = 5.1$, $\sigma = 1.2$, and random sample of size 9. We can use the central limit theorem to compute $\mu_{\bar{x}} = \mu = 5.1$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.2}{\sqrt{9}} = 0.4$. 

Find the probability that a mean life of a random sample of 9 such blenders fall between 4.5 and 5.5 years.

\[ P(4.5 < \bar{x} < 5.5) \]

\[ P(4.5 < \bar{x} < 5.5) = \text{normaldf}(4.5, 5.5, 5.1, 0.4) = 0.7745 \]
Solution Continued:

- Find the value of $\bar{x}$ that separates the top 10% from the rest of the means computed from random samples of size 9.

$\Rightarrow P(\bar{x} > k) = 0.1$

$\bar{x} = k = P_{90} = \text{invNorm}(0.9, 5.1, 0.4) \approx 5.6$
Example:

Suppose the hourly wages of all workers in a manufacturer company have a normal distribution with a mean of $15.50 and a standard deviation of $2.75. If we randomly select a sample of 10 workers from this company, find the probability that their mean hourly wages is

- less than $14.25.
- more than $16.50.

Solution:

We have a normal probability distribution with $\mu = 15.50$, $\sigma = 2.75$, and random sample of size 10. We can use the central limit theorem to compute $\mu_{\bar{x}} = \mu = 15.50$ and

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.75}{\sqrt{10}} \approx 0.87$$
Solution Continued:

- less than $14.25.

\[ P(\bar{x} < 14.25) \]

\[ P(\bar{x} < 14.25) = \text{normaldf}(\mu_{\bar{x}} = 15.50, \sigma_{\bar{x}} = 0.87) = 0.0754 \]
Solution Continued:

- more than $16.50.

\[ P(\bar{x} > 16.50) \]

\[ P(\bar{x} = 16.50) = \text{normalcdf}(16.50, E99, 15.50, 0.87) \approx 0.125 \]