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 μ and variance σ^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 $\sigma_{\bar{x}}^2 = \dfrac{\sigma^2}{n}$, and
- standard deviation $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$.

Example:

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

- Find μ and σ^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 μ and σ^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.

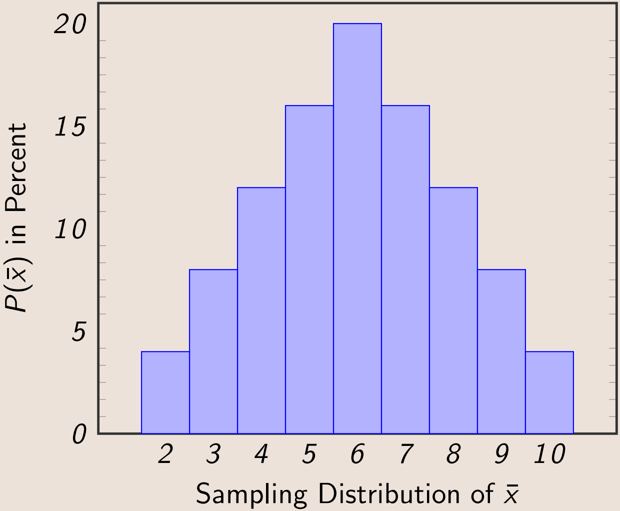
2	3	4	5	6
3	4	5	6	7
4	5	6	7	8
5	6	7	8	9
6	7	8	9	10

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

$\bar{x}$	2	3	4	5	6	7	8	9	10
$P(\bar{x})$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{4}{25}$	$\frac{5}{25}$	$\frac{4}{25}$	$\frac{3}{25}$	$\frac{2}{25}$	$\frac{1}{25}$

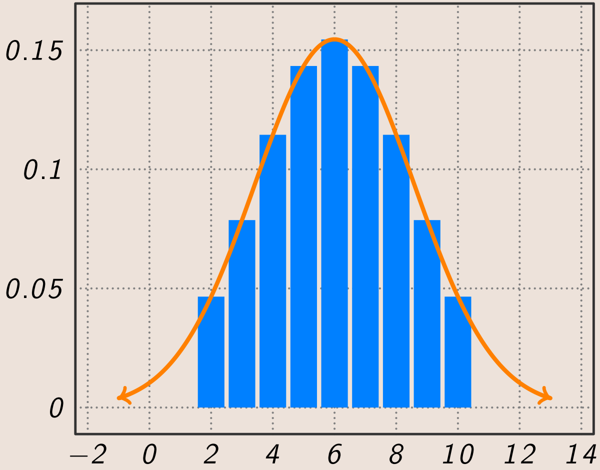
Solution Continued:

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



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.

$\bar{x}$	2	3	4	5	6	7	8	9	10
$P(\bar{x})$	$\frac{1}{25}$	$\frac{2}{25}$	$\frac{3}{25}$	$\frac{4}{25}$	$\frac{5}{25}$	$\frac{4}{25}$	$\frac{3}{25}$	$\frac{2}{25}$	$\frac{1}{25}$

Use the discrete probability distribution

- to find $\mu_{\bar{x}}$, $\sigma_{\bar{x}}$,
- the exact value of $\sigma^2_{\bar{x}}$, 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

  $\Rightarrow \mu_{\bar{x}} = 6$ & $\sigma_{\bar{x}} = 2$

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

  $\Rightarrow \sigma_{\bar{x}}^2 = 2^2 = 4$

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

  We can verify that $\mu_{\bar{x}} = \mu = 6$, and $\sigma_{\bar{x}}^2 = \dfrac{\sigma^2}{n} = \dfrac{8}{2} = 4$.

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 $\mu_{\bar{x}}$. $\Rightarrow \mu_{\bar{x}} = \mu = \boxed{375}$
- Find $\sigma^2_{\bar{x}}$. $\Rightarrow \sigma^2_{\bar{x}} = \dfrac{\sigma^2}{n} = \dfrac{100}{16} = \boxed{6.25}$

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 $\mu_{\bar{x}}$. $\Rightarrow \mu_{\bar{x}} = \mu = \boxed{82}$
- Find $\sigma_{\bar{x}}$. $\Rightarrow \sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{7.5}{\sqrt{10}} \approx \boxed{2.372}$.

Z score & $\bar{x}$ Sampling Distribution

we know that $z = \dfrac{x-\mu}{\sigma}$, now we can replace x with $\bar{x}$, μ with $\mu_{\bar{x}}$, σ with $\sigma_{\bar{x}}$, and simplify using the central limit theorem.

$$z = \frac{x-\mu}{\sigma} = \frac{\bar{x}-\mu_{\bar{x}}}{\sigma_{\bar{x}}} = \boxed{\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}}$$

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 $z = \dfrac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$,

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

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.

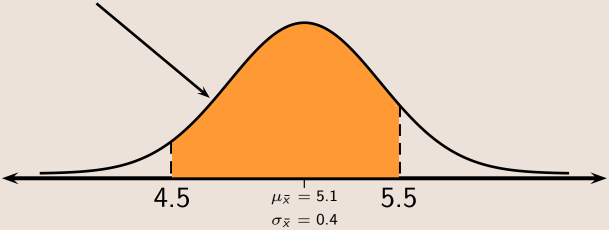
Solution:

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 $\mu_{\bar{x}} = \mu = \boxed{5.1}$ and $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{1.2}{\sqrt{9}} = \boxed{0.4}$

Solution Continued:

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

  $\Rightarrow P(4.5 < \bar{x} < 5.5)$

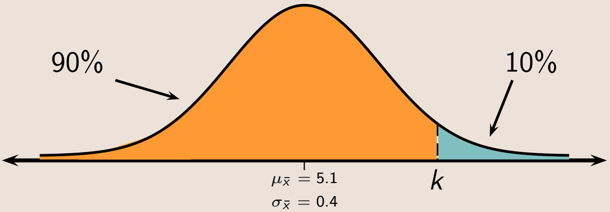


$P(4.5 < \bar{x} < 5.5) = \text{normaldf}(4.5, 5.5, 5.1, 0.4) = \boxed{0.7745}$

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.

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



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

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:

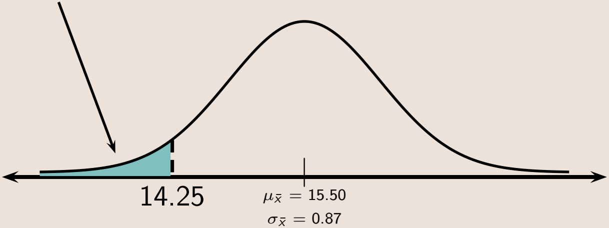
We have a normal probability distribution with $\mu = 15.50$, $\sigma = 2.75$, and random sample of size 10. We can use the central limit theorem to compute $\mu_{\bar{x}} = \mu = \boxed{15.50}$ and

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

Solution Continued:

- less than \$14.25.

  $\Rightarrow P(\bar{x} < 14.25)$

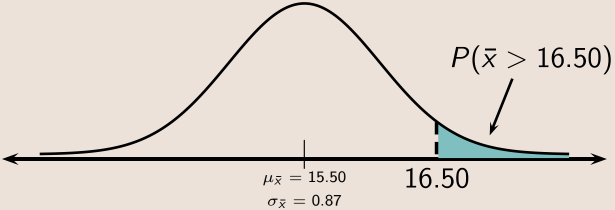


$P(\bar{x} < 14.25) = \text{normaldf}(-E99, 14.25, 15.5, 0.87) = \boxed{0.0754}$

Solution Continued:

- more than \$16.50.

$\Rightarrow P(\bar{x} > 16.50)$



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