# Central Limit Theorem

# Central Limit Theorem

**Central Limit Theorem Explanation**

Let us say one has a large samples of observations and each sample is randomly produced and independent of other observations.

Calculate the average of the observations of each sample, thus having a collection of averages of observations.

Now as per Central Limit Theorem, if the sample size was adequately large, then the probability distribution of these sample averages will approximate to a normal distribution.

**Thus, Central limit theorem (CLT) states that**

(1) If one collects samples of size n from a population, the mean of all these samples would follow a normal distribution.

(2) If the sample size is large enough (generally greater than 30), the sample distribution taken from any population distribution would follow a normal distribution.

# Central Limit Theorem

It means that as the sample size increases and its variance is finite, then the distribution of the sample mean approaches normal distribution irrespective of the shape of the population distribution.

**And this is true for any population distribution. For example,**

(1) If population distribution is normal distribution $(\mu, \sigma)$, the sample mean distribution would normal distribution ($\mu_{\bar{X}}$, $\sigma/\sqrt{n}$).

(2) If the population distribution is uniform (a, b), the sample mean distribution would be normal distribution ($\frac{a+b}{2}$, $\frac{b-a}{\sqrt{12n}}$).

(3) Even for exponential distribution or any random population distribution, the sample mean would follow a normal distribution (1/λ, 1/λ$\sqrt{n}$).

# Central Limit Theorem

Let us say that the random variable X ~ N (μ, σ), then

$$P(X < a) = P\left(z < \frac{X - \mu}{\sigma}\right)$$

The sample mean follows the distribution as $\bar{X} \sim N(\mu_{\bar{X}}, \sigma/\sqrt{n})$.

The probability of any value of sample mean ($\bar{X}$) can be calculated as

$$P(\bar{X} < a) = P\left(z_{\bar{X}} < \frac{\bar{X} - \mu_{\bar{X}}}{\sigma/\sqrt{n}}\right)$$

# Central Limit Theorem

If a random variable X ~ U(a, b), then

$$f_X(x) = \frac{1}{b-a}, \forall a \le x \le b$$

with mean $\mu = \frac{a+b}{2}$ and standard deviation $\sigma = \frac{b-a}{\sqrt{12}}$

The sample mean distribution would be normal distribution $\bar{X} \sim N(\frac{a+b}{2}, \frac{b-a}{\sqrt{12n}})$.

So, the probability between two points (c and d) such that both lies between a and b,

$$P(c \le x \le d) = (d-c) \times \left(\frac{1}{b-a}\right)$$

# Central Limit Theorem

Let us consider that X ~ Exponential (λ), then

$$f_X(x) = \lambda e^{-\lambda x}$$

With mean $\mu = 1/ \lambda$ and standard deviation $\sigma = 1/ \lambda$.

The sample mean distribution would be normal distribution $\bar{X} \sim N(1/\lambda,\ 1/\lambda\sqrt{n})$.

$$P(X < x) = 1 - \mathrm{e}^{-\lambda x}$$

Which is actually the cumulative distribution function (CDF) of X.

Similarly,

$$P(X > x) = \mathrm{e}^{-\lambda x}$$

# Central Limit Theorem

## Assumptions of Central Limit Theorem

Central Limit Theorem (CLT) is valid for the following conditions:

- The drawing of the sample from the population should be random.
- The drawing of the sample should be independent of each other.
- The sample size should not exceed ten percent of the total population when sampling is done without replacement.
- Sample size should be adequately large.
- CLT only holds true for population with finite variance.

# Central Limit Theorem

**Some notations to be used**

$\mu$: Population mean

$\sigma$: standard deviation of the population

s: standard deviation of the sample

$\bar{X}$: sample mean

$\mu_{\bar{X}}$: mean of the sample mean distribution

$\sigma_{\bar{X}}$: standard deviation of the sample mean distribution

n: sample size

# Central Limit Theorem

**Law of large numbers:**

(1) If the sample size (n) increases, the sample mean ($\bar{X}$) tends to the population mean (μ).
(2) If the sample size (n) increases, the mean of the mean sample distribution ($\mu_{\bar{X}}$) tends to the population mean (μ).

For population distribution, the Z score is defined as

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

For sample mean distribution, the Z score is defined as

$$Z_{\bar{X}} = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$$

For large value of n, as per CLT

$$z_{\bar{X}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

It implies that as the sample size (n) increase, $\sigma_{\bar{X}}$ (= $\sigma/\sqrt{n}$) decreases. The standard deviation ($\sigma_{\bar{X}}$) of the sample mean distribution is also known as the standard error.

# Central Limit Theorem

## Central Limit Theorem Formula

Let us assume we have a random variable X. Let σ is its standard deviation and μ is the mean of the random variable. Now as per Central Limit Theorem, the sample mean $\bar{X}$ will approximate to the normal distribution which is given as

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

The Z-Score of the random variable $\bar{X}$ is given as

$$Z_{\bar{X}} = \frac{\bar{x} - \mu_{\bar{X}}}{\sigma/\sqrt{n}}$$

Here $\bar{x}$ is the mean of $\bar{X}$.

# Central Limit Theorem

## Central Limit Theorem Proof

Let X is a random variable which follows normal distribution with mean (E(X) = $\mu$) and variance (Var(x) = $\sigma^2$), the probability density function of X is given as

$$N(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

We will use the concept of Moment Generating Function (MGF) to prove the central limit theorem.

To begin with, let us consider first the case of standard normal distribution, i.e., N(0, 1).

The moment generating function for a random variable Z (where Z = $\frac{x-\mu}{\sigma}$) is defined as

$$M_z(t) = E(\mathrm{e}^{tz}) = \int \mathrm{e}^{zt} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}\, \mathrm{d}z$$

$$= \mathrm{e}^{\frac{t^2}{2}}$$

# Central Limit Theorem

## Central Limit Theorem Proof

Now, let us try to determine the MGF of the normal distribution N($\mu$, $\sigma^2$).

$$M_x(t) = E(\mathrm{e}^{xt}) = \int \frac{1}{\sigma\sqrt{2\pi}} e^{tx} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \, \mathrm{d}x$$

Let us define

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

Putting the value of x in the above equation, we get

$$M_x(t) = \int \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{\mu t} \, \mathrm{e}^{z\sigma t} \, \mathrm{e}^{-\frac{1}{2}z^2} \left|\frac{\mathrm{d}x}{dz}\right| \mathrm{d}z$$

# Central Limit Theorem

## Central Limit Theorem Proof

Replacing (dx/dz = σ), we get

$$M_x(t) = \int \frac{1}{\sqrt{2\pi}} e^{\mu t} \; \mathrm{e}^{z\sigma t} e^{-\left(\frac{z}{2}\right)^2} \mathrm{d}z$$

$$= \mathrm{e}^{\mu t} \mathrm{e}^{\frac{1}{2}\sigma^2 t^2}$$

$$= \mathrm{e}^{\mu t \; + \frac{\sigma^2 t^2}{2}}$$

This shows the MGF of N(x, $\mu$, $\sigma^2$).

# Central Limit Theorem

## Central Limit Theorem Proof

Notably, the MGF of a random variable X which follows normal distribution with mean (μ) and variance ($\sigma^2$), is exponential in function. This implies that sum of two identically and independently distributed random variables will have normal distribution.

Thus, Let X ~ N($\mu_x, \sigma_x^2$) and X ~ N($\mu_y, \sigma_y^2$) , then their sum X + Y ~ N($\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2$).

To prove this, let us again use the Moment Generating Function (MGF).

This is to be noted that MGF of sum of two independent variables is the product of their individual MGFs.

That is, if z = x + y, then

$$M_z(t) = M_x(t)M_y(t) = \exp\left\{\mu_{x^t} + \frac{1}{2}\sigma_x^2 t^2\right\} \exp\left\{\mu_y t + \frac{1}{2}\sigma_y^2 t^2\right\}$$
$$= \exp\left\{(\mu_x + \mu_y)t + \frac{1}{2}(\sigma_x^2 + \sigma_y^2)t^2\right\}$$

This is a MGF of X + Y ~ N($\mu_x + \mu_y$, $\sigma_x^2 + \sigma_y^2$). This can be generalized for n random variables.

# Central Limit Theorem

## Steps to Solve Problems on Central Limit Theorem

**Step 1:** First identify the >, < associated with sample size, population size, mean and variance in the problem. Also there can be ‘between; associated with range of two numbers.

**Step 2:** Draw a Graph with Mean as Centre

**Step 3:** Find the Z-Score using the formula

**Step 4:** Refer to the Z table to find the value of Z obtained in the previous step.

**Step 5:** If the problem involves ‘>’ subtract the Z score from 0.5; if the problem involves ‘<‘ add 0.5 to the Z score and if the problem involves ‘between’ then perform only step 3 and 4.

**Step 6:** The Z score value is found along $\bar{X}$.

**Step 7:** Convert the decimal value obtained in all three cases to decimal.

# Central Limit Theorem

**Example 1.**
**The male population's weight data follows a normal distribution. It has a mean of 70 kg and a standard deviation of 15 kg. What would the mean and standard deviation of a sample of 50 guys be if a researcher looked at their records?**

Given: $\mu = 70$ kg, $\sigma = 15$ kg, $n = 50$
As per the Central Limit Theorem, the sample mean is equal to the population mean.

Hence, $\mu_{\bar{x}} = \mu = 70kg$

Now, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{50}} = 2 \cdot 1kg$

# Central Limit Theorem

**Example 2.**
**A distribution has a mean of 69 and a standard deviation of 420. Find the mean and standard deviation if a sample of 80 is drawn from the distribution.**

Given: μ = 69, σ = 420, n = 80
As per the Central Limit Theorem, the sample mean is equal to the population mean.

Hence, $\mu_{\bar{x}} = \mu = 69$

Now, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{420}{\sqrt{80}} = 46.95$

# Central Limit Theorem

**Example 3.**
**The mean age of people in a colony is 34 years. Suppose the standard deviation is 15 years. The sample of size is 50. Find the mean and standard deviation of the sample.**

Given: μ = 34, σ = 15, n = 50
As per the Central Limit Theorem, the sample mean is equal to the population mean.

Hence, $\mu_{\bar{x}} = \mu = 34 \text{ years}$

Now, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{50}} = 2.12$ years

# Central Limit Theorem

**Example 4.**
**The mean age of cigarette smokers is 35 years. Suppose the standard deviation is 10 years. The sample size is 39. Find the mean and standard deviation of the sample.**

Given: μ = 35, σ = 10, n = 39
As per the Central Limit Theorem, the sample mean is equal to the population mean.

Hence, $\mu_{\bar{x}} = \mu = 35 \text{ years}$

Now, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{39}} = 1.601$ years

# Central Limit Theorem

**Example 5.**
**The mean time taken to read a newspaper is 8.2 minutes. Suppose the standard deviation is one minute. Take a sample of size 70. Find its mean and standard deviation.**

Given: $\mu = 8.2$, $\sigma = 1$, $n = 70$
As per the Central Limit Theorem, the sample mean is equal to the population mean.

Hence, $\mu_{\bar{x}} = \mu = 8.2$ minutes

Now, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1}{\sqrt{70}} = 0.11$ minutes

# Central Limit Theorem

**Example 6.**
**A distribution has a mean of 12 and a standard deviation of 3. Find the mean and standard deviation if a sample of 36 is drawn from the distribution.**

Given: μ = 12, σ = 3, n = 36
As per the Central Limit Theorem, the sample mean is equal to the population mean.

Hence, $\mu_{\bar{x}} = \mu = 12$

Now, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.5$

# Central Limit Theorem

**Example 7.**

**A distribution has a mean of 4 and a standard deviation of 5. Find the mean and standard deviation if a sample of 25 is drawn from the distribution.**

Given: $\mu = 4$, $\sigma = 5$, $n = 25$

As per the Central Limit Theorem, the sample mean is equal to the population mean.

Hence, $\mu_{\bar{x}} = \mu = 4$

Now, $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{25}} = 1$

# Central Limit Theorem

**Example 8. Let the JEE score follows a normal distribution with mean = 74 and standard deviation = 6.8. (a) If a student is selected at random, what is the probability that his/her exam score is less than 65? (b) If a sample of 50 students is selected at random, what is the probability that the mean exam score of this group is greater than 75? (c) What is the distribution for the mean exam score of 50 students? (d) Find the 80$^{th}$ percentile for the mean exam score of the 50 students.**

(a) Given that mean (μ) = 74 and standard deviation (σ) = 6.8 and n = 1. We need to find

$$P(X < 65) = P\left(Z < \frac{65-74}{6\cdot 8}\right) = P(Z < -1.32) = 0.09342 \Rightarrow 9.342\%$$

(b) Given that mean (μ) = 74 and standard deviation (σ) = 6.8 and n = 50.

$$P(\bar{X} > 75) = 1 - P(\bar{X} < 75) = 1 - P\left(Z_{\bar{X}} < \frac{75-74}{6\cdot 8/\sqrt{50}}\right) = 1 - P(z_{\bar{X}} < 1.04)$$

$$= 1 - 0.85083 = 0.14917 \Rightarrow 14.917\%$$

(c) Mean exam score follows normal distribution

$$\bar{X} \sim N\left(\mu_{\bar{X}}, \frac{\sigma}{\sqrt{n}}\right) \sim N(74, 0.9617)$$

(d) Given that area = 0.80, then

$$\bar{X} = \mu_{\bar{X}} + Z_{\bar{x}} \times \frac{\sigma}{\sqrt{n}} = 74 + 0.84 \times 0.9617 = 74.81$$

**Thank You**