Data Science and Artificial Intelligence

Probability and Statistics

Introduction to Sampling Distribution

Lecture No.- 01

By- Rahul Sir

Topics to be Covered

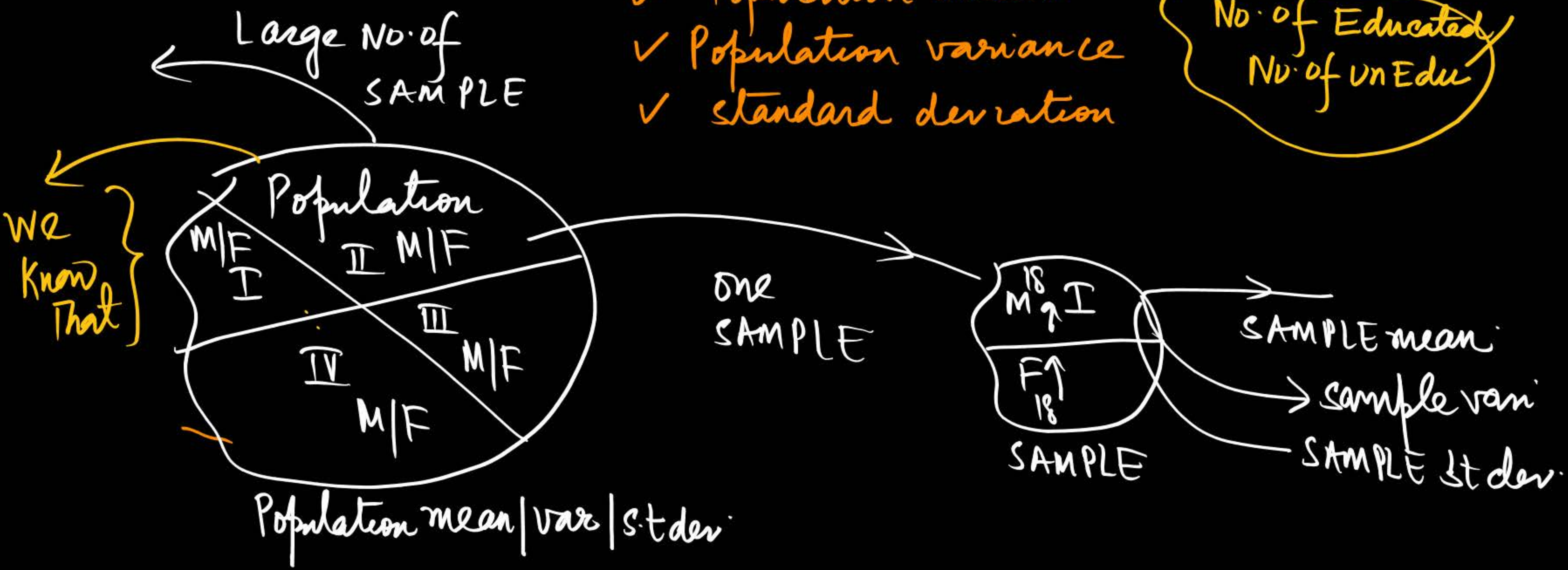
- **Topic** — Introduction to Sampling Distribution
- **Topic** — Standard Error
- **Topic** — Central Limit Theorem → already done
  Z-SCORE
- **Topic** — Law of Large Numbers

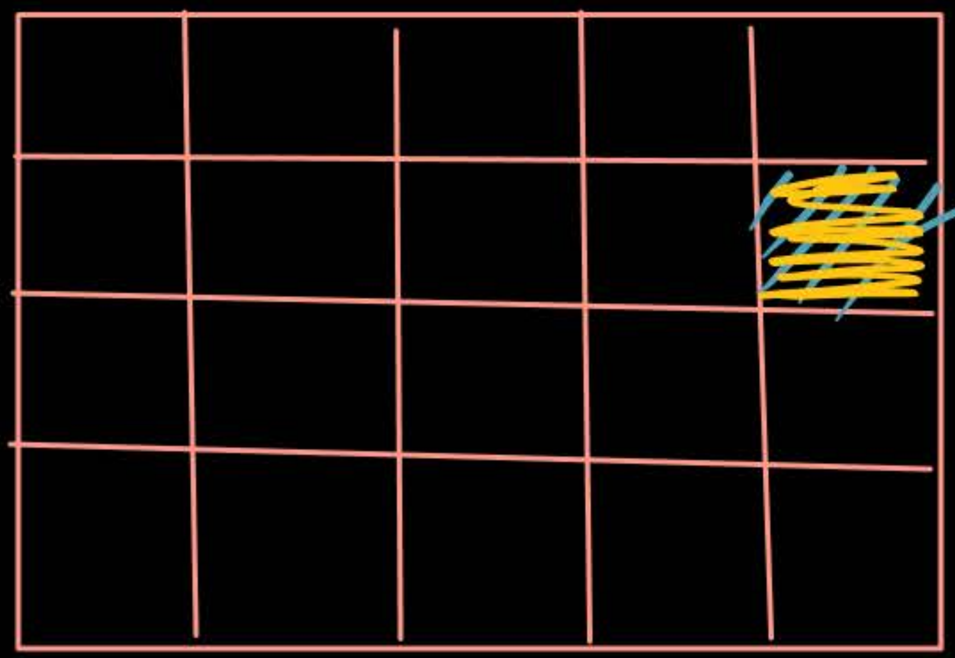
DESCRIPTIVE STATISTICS: Population

No. of Males / females.

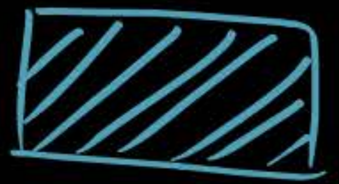
No. of Educated / No. of UnEdu.

✓ Population mean
✓ Population variance
✓ standard deviation





Inferntial statistics



Analysis

SAMPLE
(Inferential statistics)

Predicative statistics
(Using machine learning)

Population
(Descriptive statistics)

- ✓ Mean
- ✓ variance
- ✓ Standard deviation

Single sample.

SAMPLE mean
variance.
st dev.

SEA

salt Nature

Throwing A Die: 30 times (SAMPLE)

30 SAMPLE

4	3	2	5	6	3	2	6	4	3	2	5	3	1	5	2	1	4	3
1	2	1	1	2	5	2	5	4	3	6								

✓ for $n = 30$ times

Expected value $E[X] = \mu = \dfrac{\sum_{i=1}^{n} x_i}{n} \Rightarrow 3.2 = \dfrac{96}{30} = 3.2$

Theortical way.

Frequency	1	2	3	4	5	6
	5	7	6	4	5	3

multiply with frequency

$= 1\times5 + 2\times7 + 3\times6 + 4\times4 + 5\times5 + 6\times3$

$= 5 + 14 + 18 + 16 + 25 + 18 = \underline{\underline{96}}$

Total frequency = 96

$$\text{mean} = \frac{96}{30} = 3.2$$

X	1	2	3	4	5	6
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E[X] = \sum_{i=1}^{n} x_i p_i$$

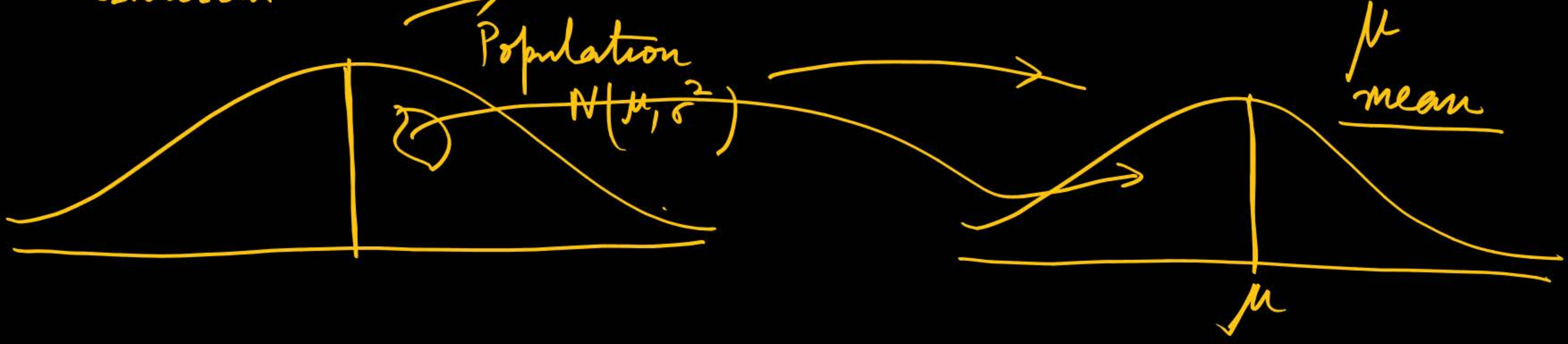
$$= \frac{1+2+3+4+5+6}{6} = \underline{\underline{3.5}}$$

If $n \uparrow$ increased

$\uparrow$ Mean

Population mean μ $\longrightarrow$ SAMPLE mean μ

Gaussian



Population n

SAMPLE $(n-1)$

$$\text{variance}_{\text{SAMPLE}} = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)^2$$

$\bar{x}$ mean

$\mu = 3.2$ (SAMPLE mean)

$$\text{SAMPLE Vari} = \frac{1}{(30-1)} \sum_{i=1}^{30} (x_i - 3.2)^2 \Rightarrow (\approx 2.65) \text{ approximate}$$

X	1	2	3	4	5	6
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

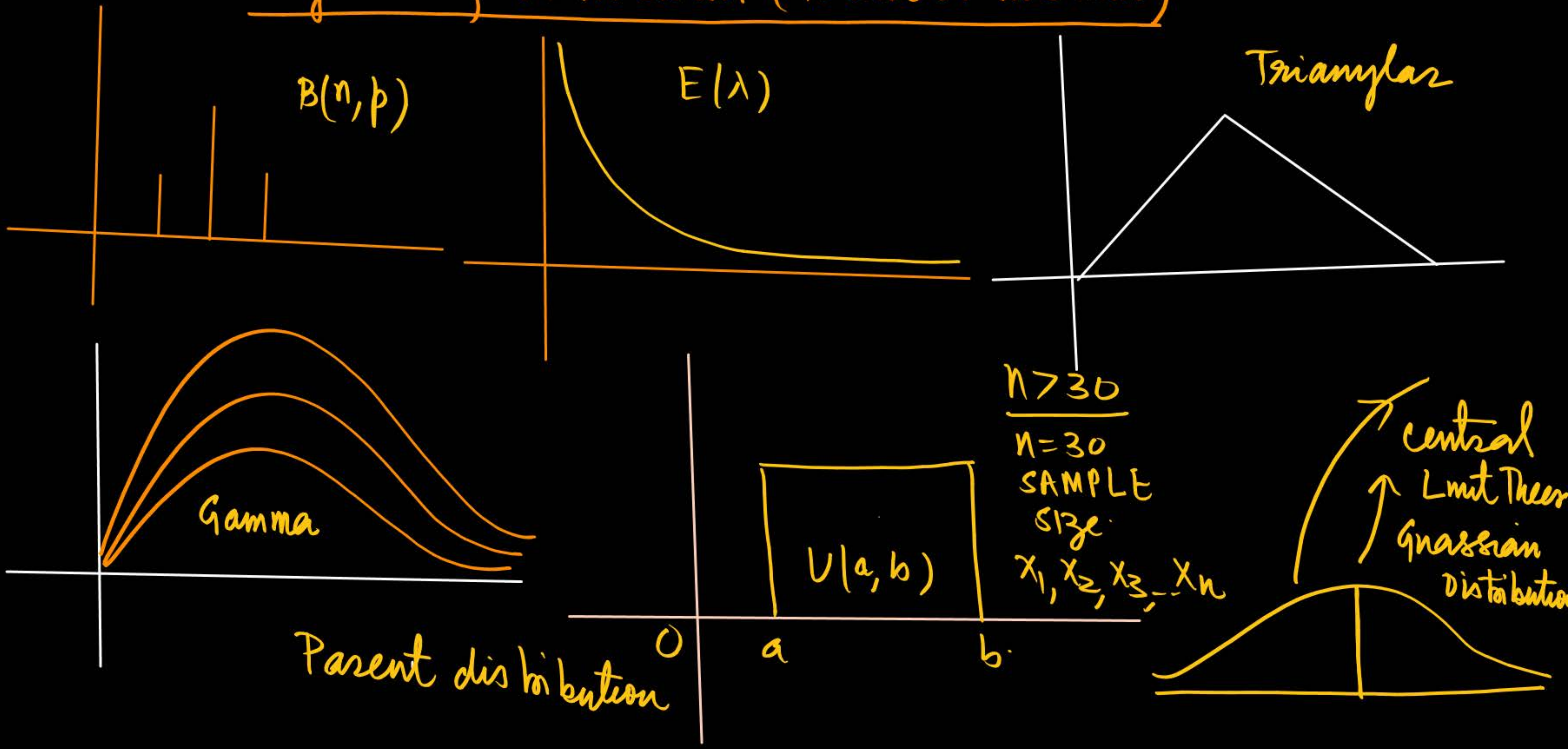
$$V(X) = E[x^2] - [E[X]]^2$$

$$V(X) = (1)^2 \times \frac{1}{6} + (2)^2 \times \frac{1}{6} + (3)^2 \times \frac{1}{6} + (4)^2 \times \frac{1}{6} + (5)^2 \frac{1}{6} + (6)^2 \frac{1}{6} - (3.5)^2$$

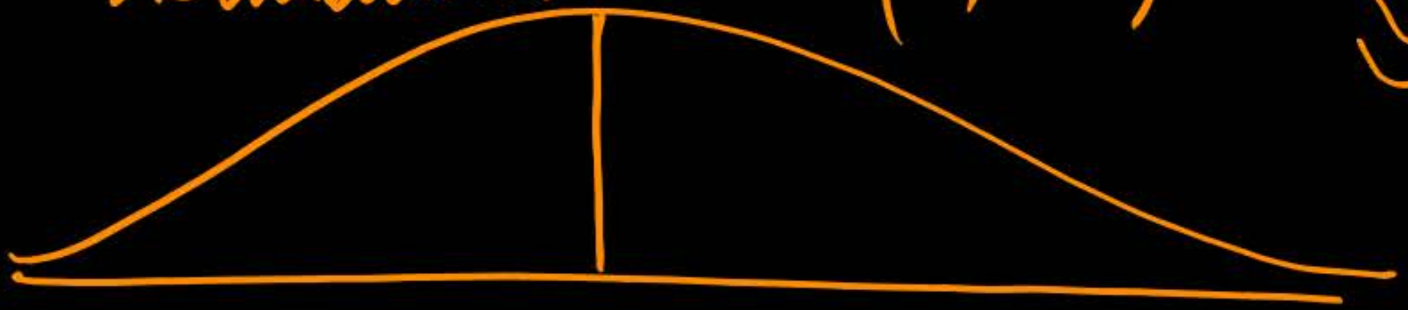
$$V(X) = 2.92$$

✓ $S.D = \sqrt{2.92}$

Any Kind of Distribution (Parent Distribution)



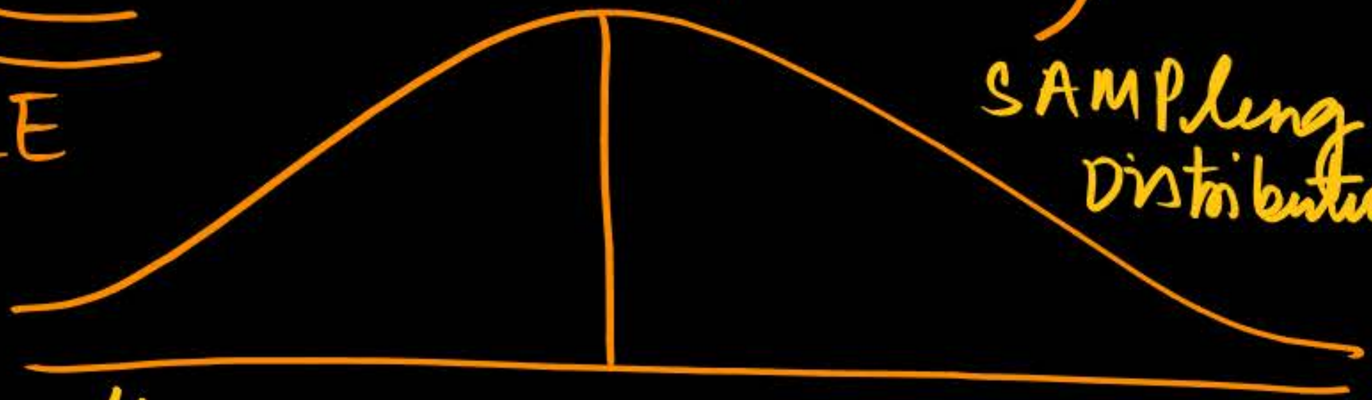
Parent Distribution $N(\mu, \sigma^2)$



SAMPLE

$N\left(\mu, \frac{\sigma^2}{n}\right)$

SAMPLing Distribution



$$N(\mu, \sigma^2) \xrightarrow[\text{Dist}]{\text{SAMPLing}} \text{mean} = \mu$$

$$\text{variance} = \frac{\sigma^2}{n} \qquad S.D = \boxed{\frac{\sigma}{\sqrt{n}}}$$

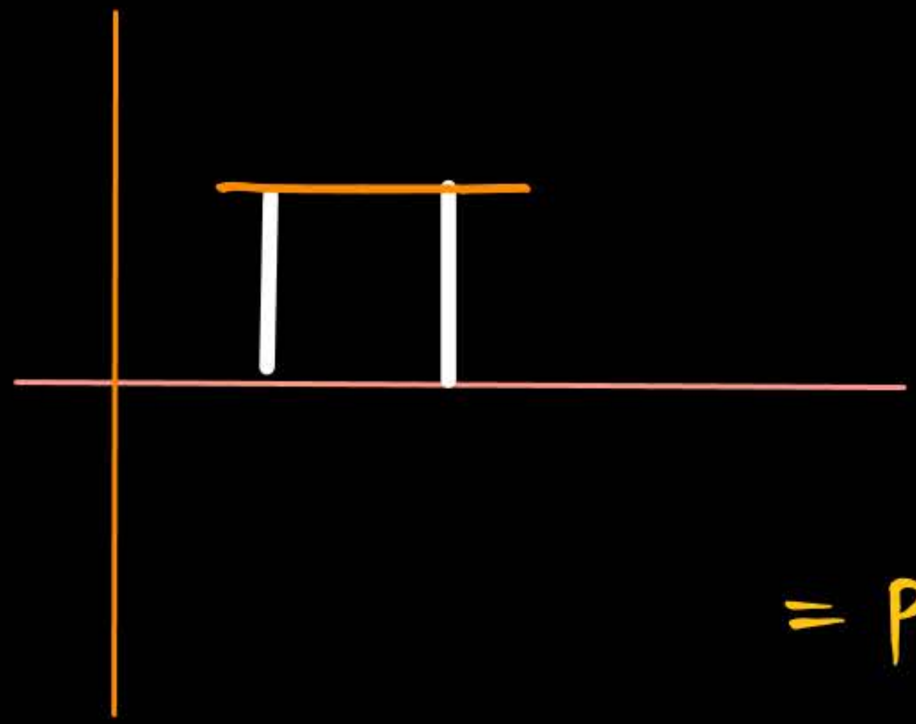
Standard Error of mean $= \frac{\sigma}{\sqrt{n}}$

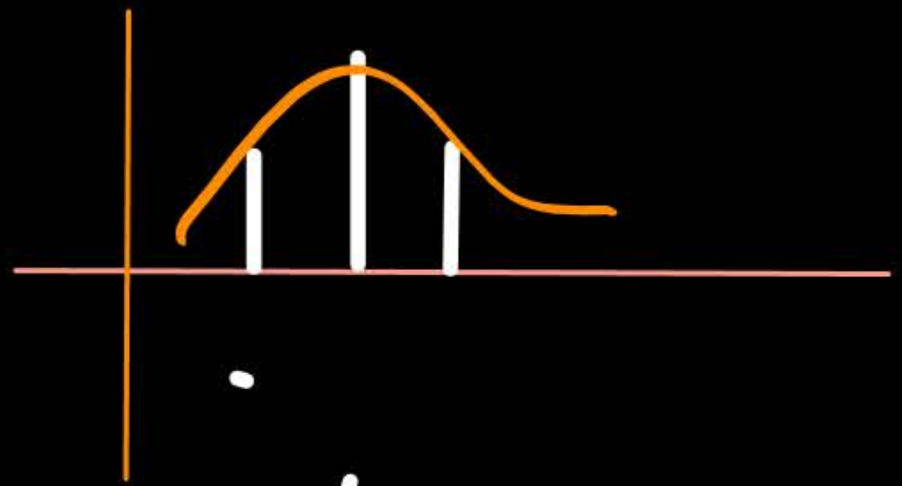
Parent Distribution

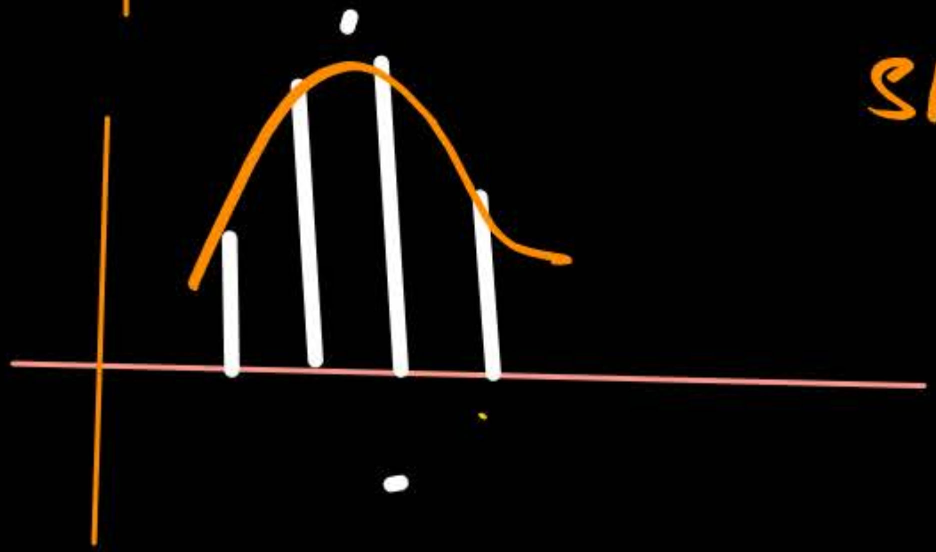
$$P(X \geq a) = P\left[\frac{X-\mu}{\sigma} \geq \frac{a-\mu}{\sigma}\right] = P\left[Z \geq \frac{a-\mu}{\sigma}\right]$$

$$P[X \geq a] = P\left[\frac{X-\mu}{\sigma} \geq \frac{a-\mu}{\sigma/\sqrt{n}}\right] = P\left[Z \geq \frac{a-\mu}{\sigma/\sqrt{n}}\right]$$

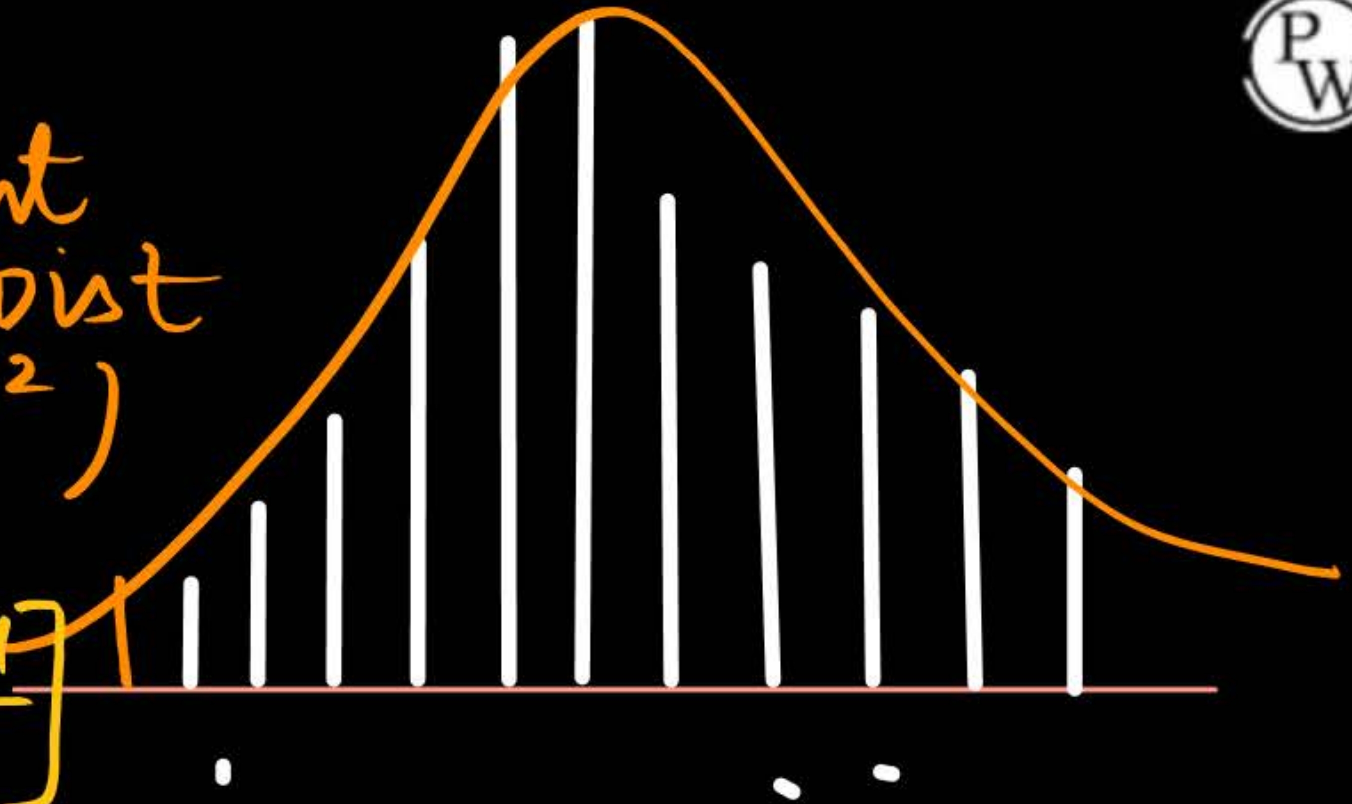
$\boxed{\text{ZSCORE} \rightarrow \text{Table}}$







Parent Dist $N(\mu, \sigma^2)$



$$P[a \le X \le b] = P\left[\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right]$$

→ SAMPLing Distribution $\mu = \mu$, $\sigma = \frac{\sigma}{\sqrt{n}}$

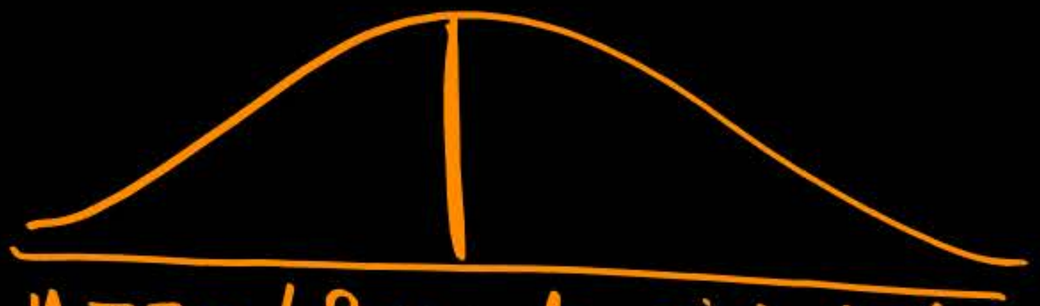
$$P[a \le X \le b]$$

SAME $P\left[\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right]$

$\mu = \mu$

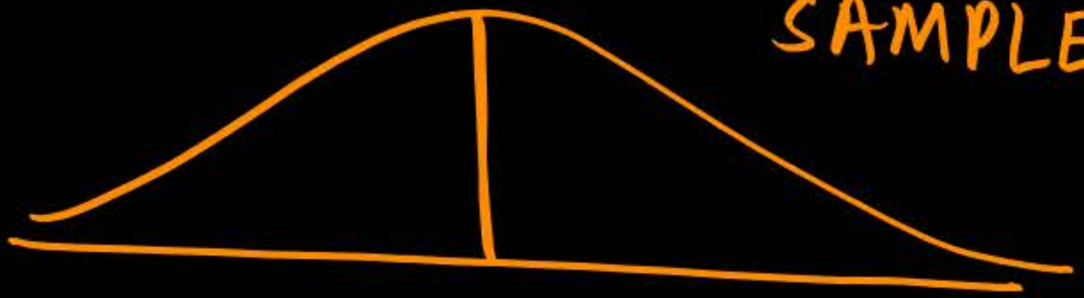
$\left.\sigma = \frac{\sigma}{\sqrt{n}}\right\}$ S.D

$var = \frac{\sigma^2}{n}$



n>30 (Parent Distribution)

SAMPLE



$$\mu = \mu$$

$$\sigma^2 = \frac{\sigma^2}{n}$$

$$\sigma = \frac{\sigma}{\sqrt{n}}$$

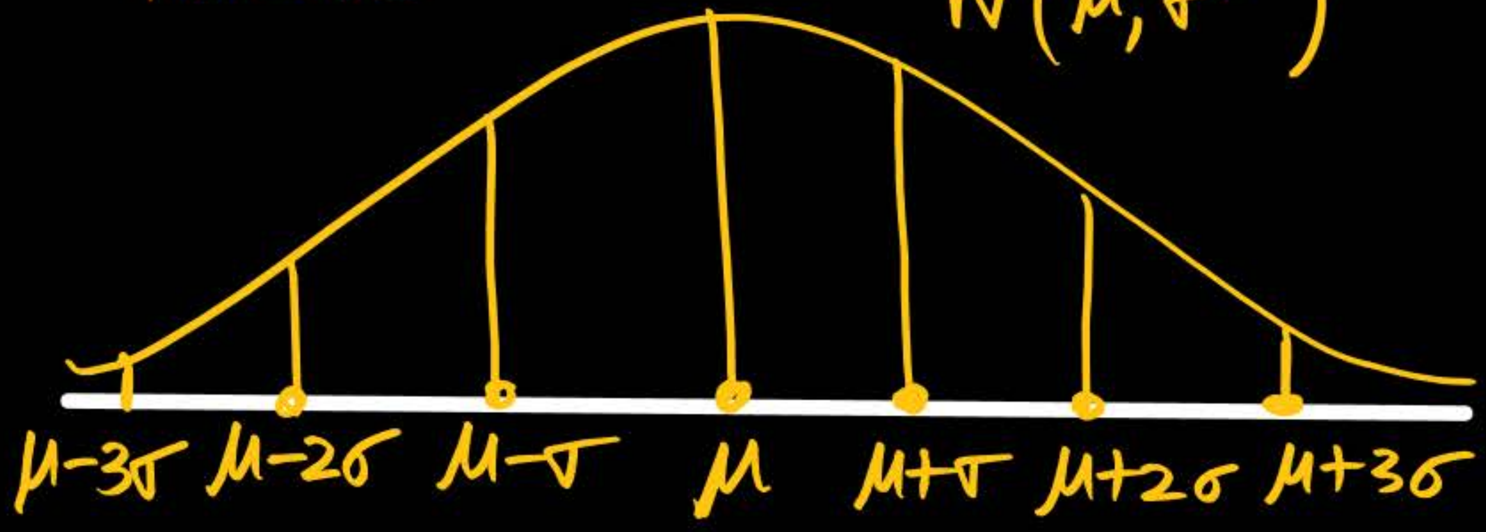
Standard Error of mean

$$= \frac{\sigma}{\sqrt{n}}$$

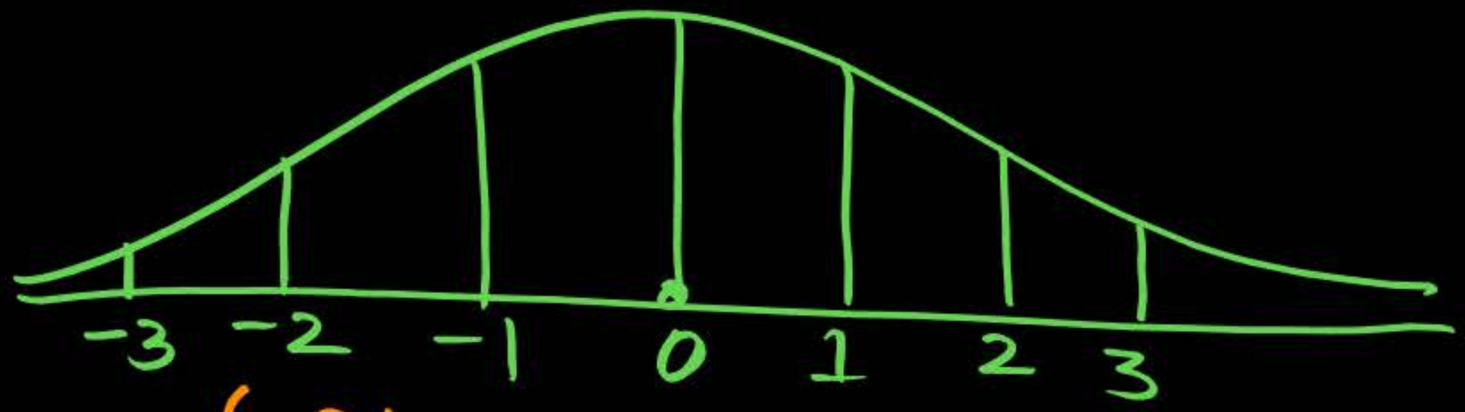
n = No. of Sample Size

Parent

$N(\mu, \sigma^2)$



$N(0,1)$ standard



$P(-1 \le z \le 1) = 0.6834$

$P(-2 \le z \le 1) = 0.9545$

$P(-3 \le z \le 3) = 0.9971$

Parent Distribution

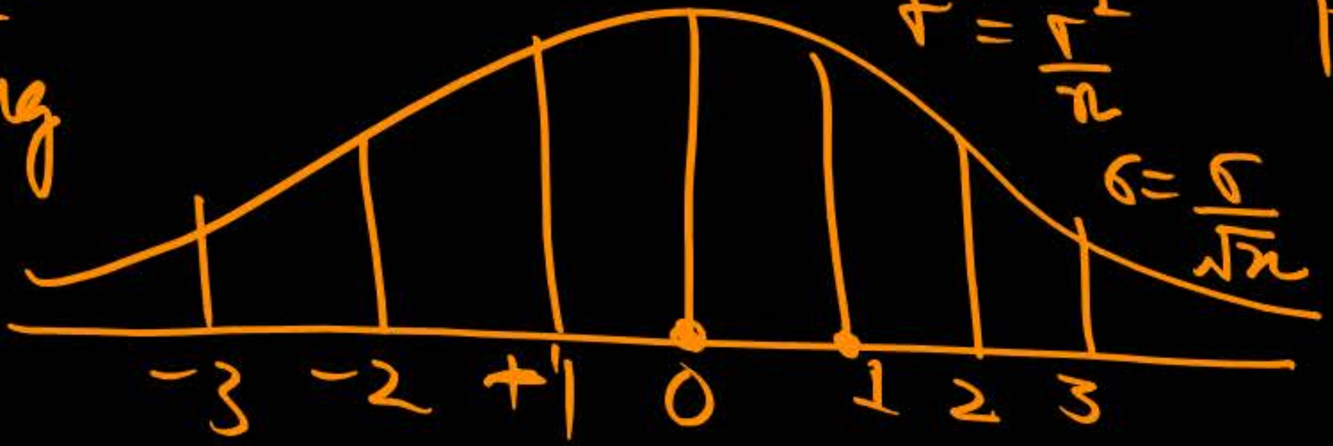
Table.

SAMPLING Dist

$\mu = \mu$

$\sigma^2 = \frac{\sigma^2}{n}$

$\sigma = \frac{\sigma}{\sqrt{n}}$



$P(-1 \le z \le 1)$ $P(-3 \le z \le 3)$

$P(-2 \le z \le 1)$

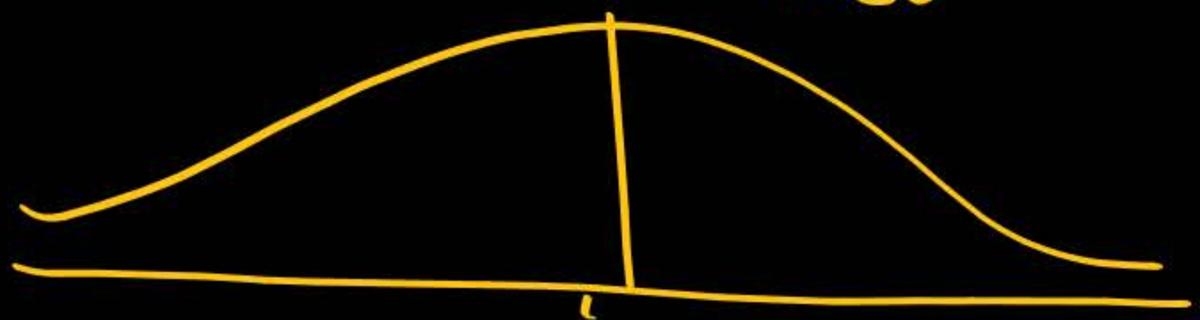
$\frac{\sqrt{2.65}}{\sqrt{30}}\Big]$ $\frac{2.65}{(\_\_\_)}$

X If the Parent Distribution has mean μ variance σ^2

$\bar{X}$ then the Sampling Distribution has mean μ and variance $\frac{\sigma^2}{n}$

Sampling Distribution $\frac{X-\mu}{\frac{\sigma}{\sqrt{n}}}$ approaches to ∞.

$n > 30$



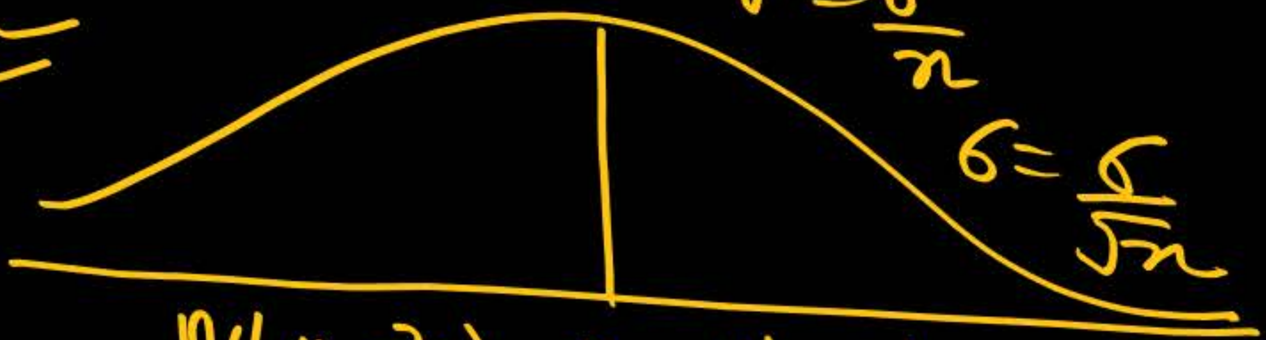
$N(\mu, \sigma^2) \to N(0,1)$

$\approx$

$\mu = \mu$

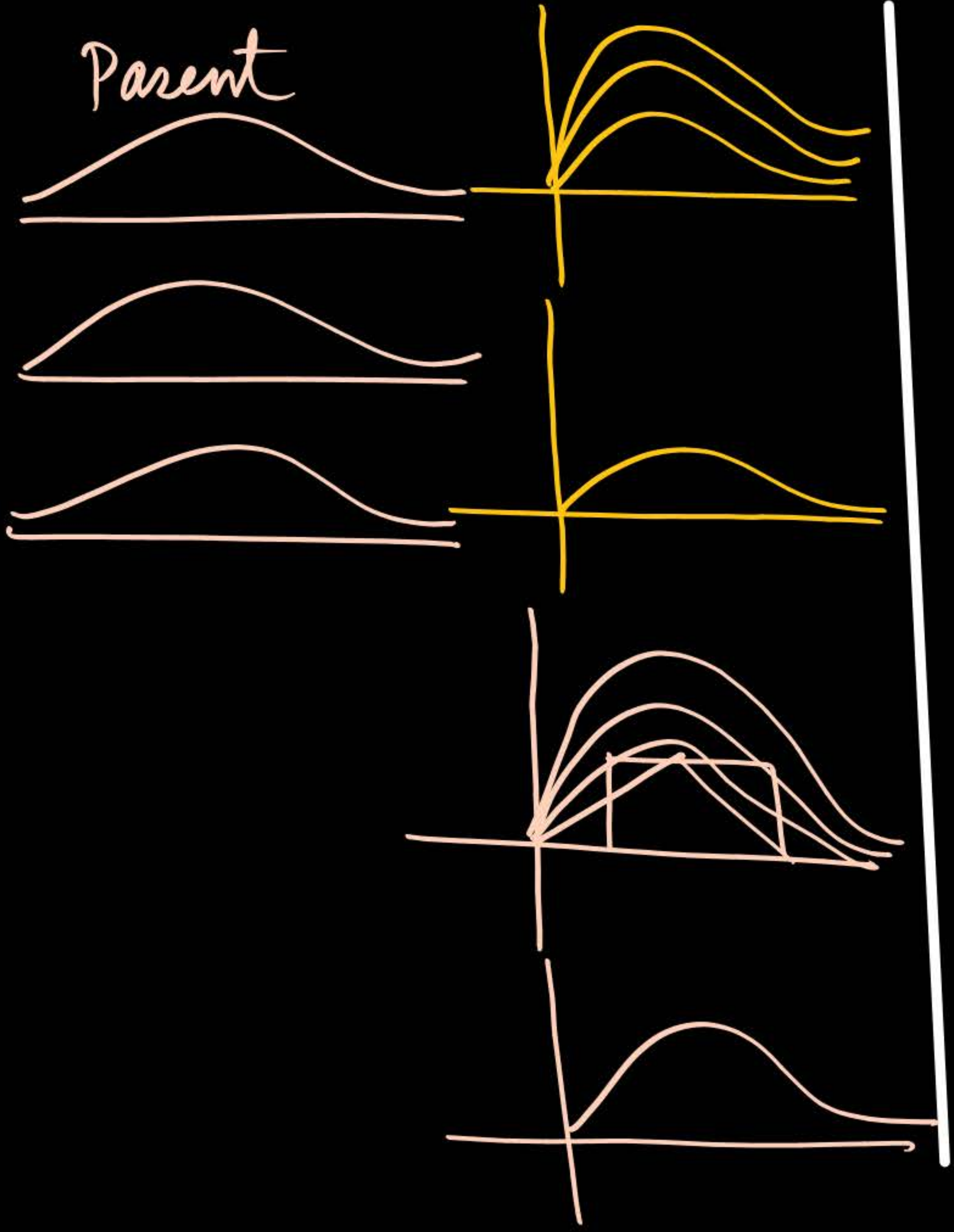
$\sigma^2 = \frac{\sigma^2}{n}$

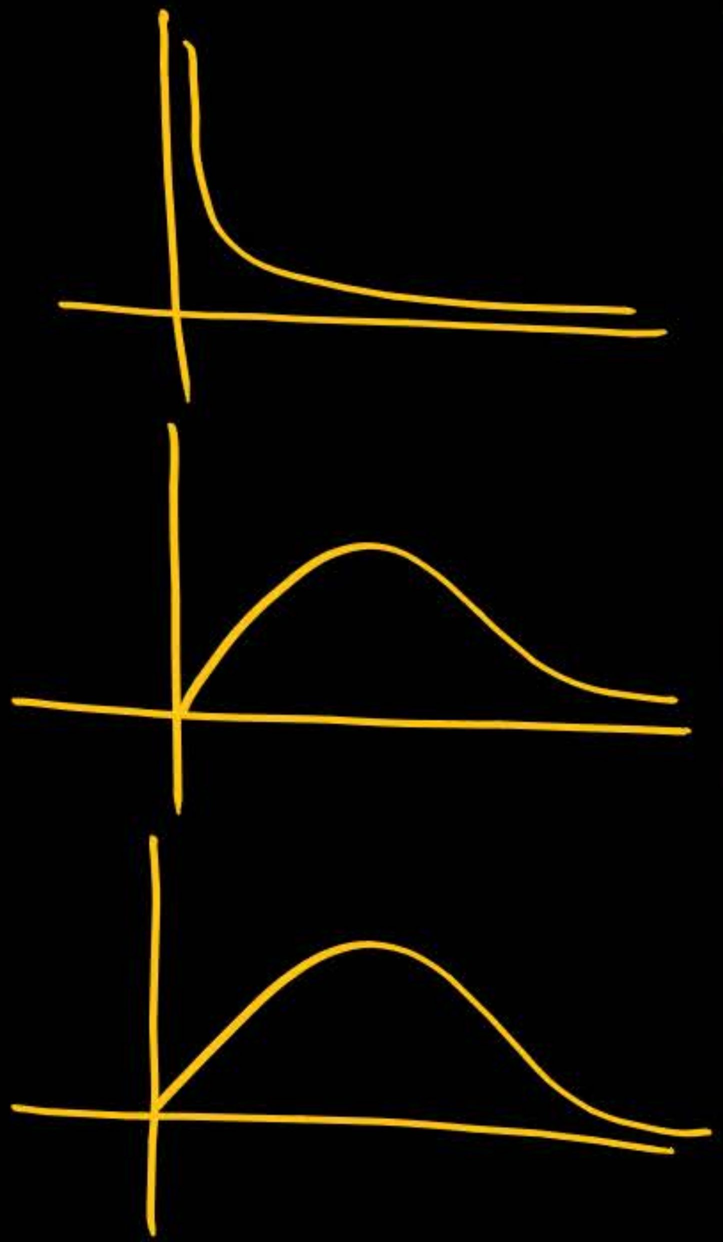
$\sigma = \frac{\sigma}{\sqrt{n}}$



$N\left(\mu, \frac{\sigma^2}{n}\right) \to N(0,1)$

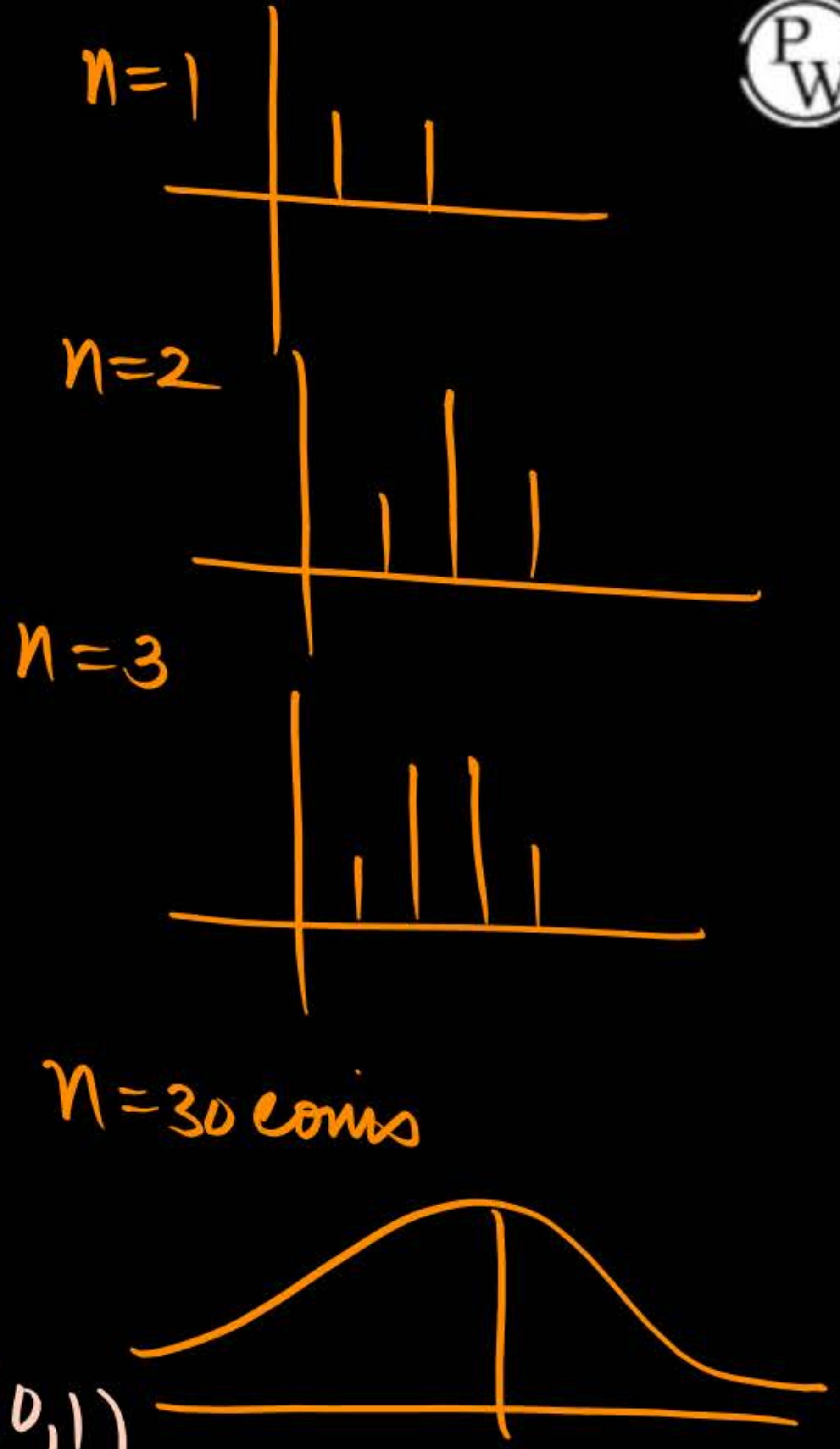
Parent

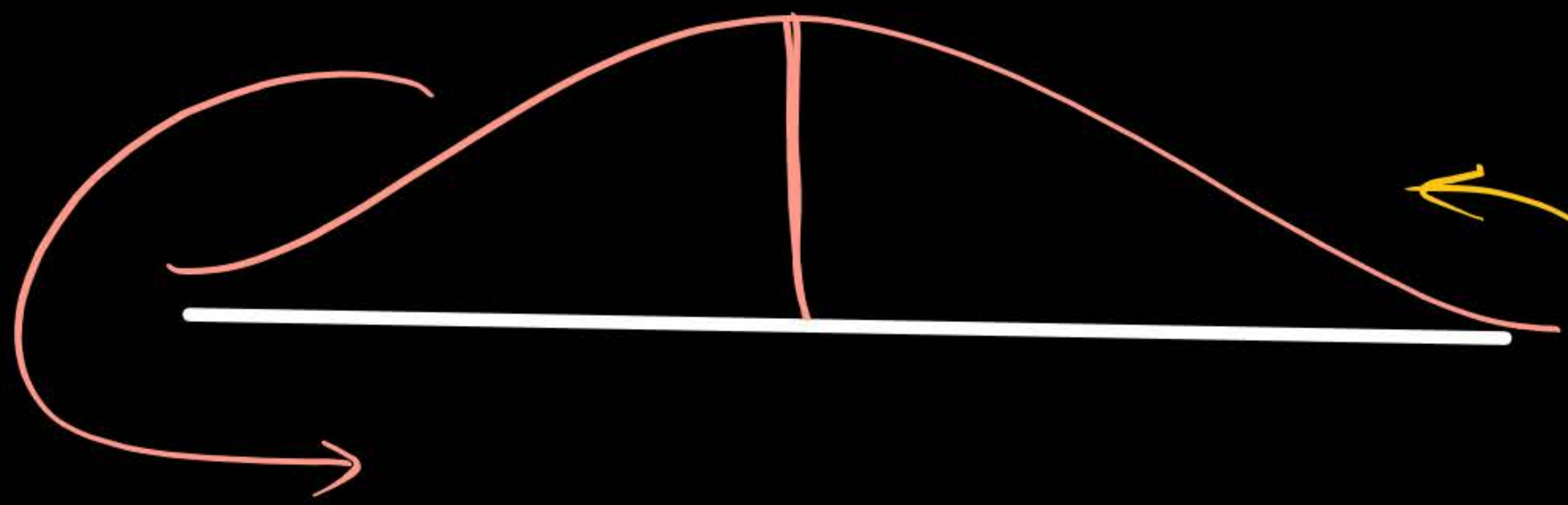




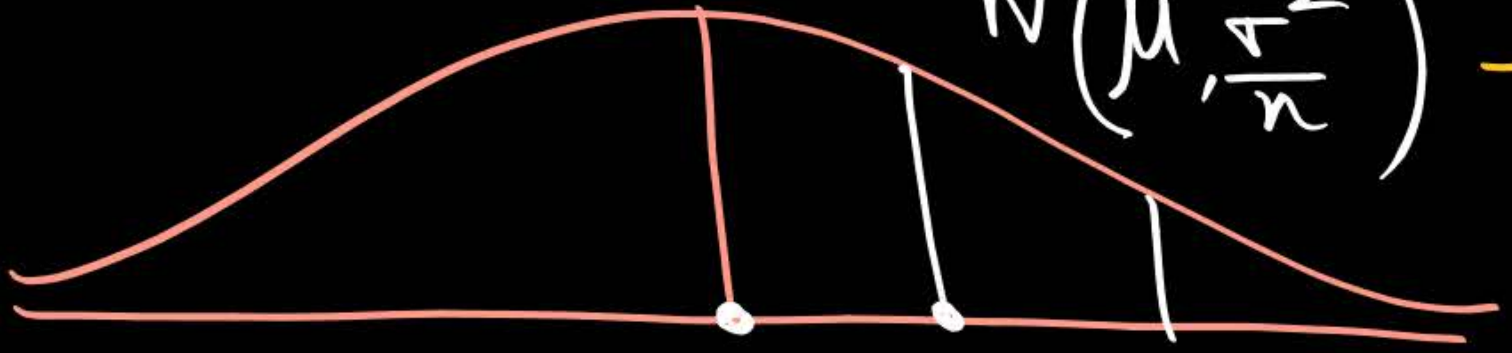
$N(\mu, \sigma^2)$ $\longrightarrow$ $N(0,1)$

$n > 30$





Distribution



$N\left(\mu, \frac{\sigma^2}{n}\right)$

$\mu = \mu$

var: $\sigma^2 = \frac{\sigma^2}{n}$

S.D $= \sigma = \frac{\sigma}{\sqrt{n}}$

Inferential statistics

$\rightarrow N\left(\mu, \frac{\sigma^2}{n}\right) \rightarrow N(0,1)$

Z SCORE

THANK - YOU