Data Science and Artificial Intelligence

Probability and Statistics

Introduction to Sampling Distribution

Lecture No.- 02

By- Rahul Sir

Topics to be Covered

Topic Law of large numbers

Topic CHI-square distribution

✓ Parent Dis

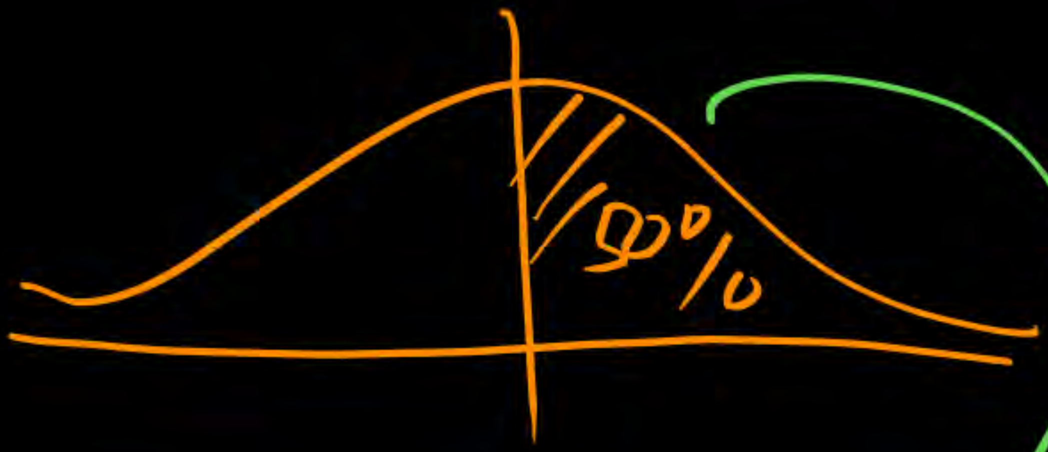
↓

SAMPling Dis

✓ central Limit Theorem

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

Slide 2



- ✓ Inequalities
- ✓ max/min of random variable
- ✓ Last Lecture

02 Hrs.

Introduction to Sampling Distribution

$X_1, X_2, X_3 \ldots X_n$

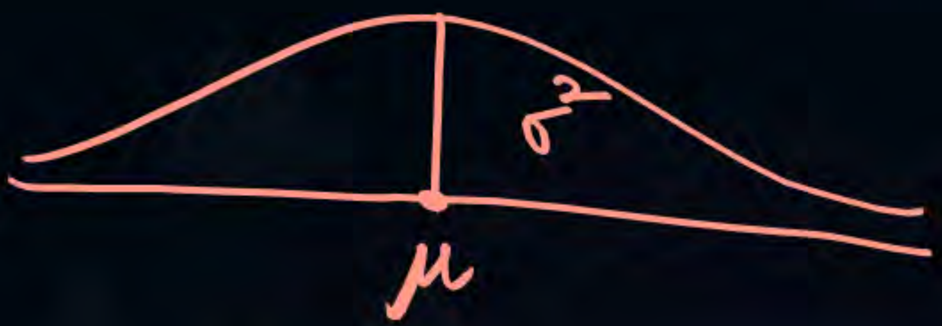
n size — $N(\mu, \sigma^2)$

Q1. If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from normal population with mean μ and variance σ^2 both are unknown then find the statistic in the following:

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

(a) $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)$

(b) $\sum_{i=1}^{n} X_i$



(c) $\sum_{i=1}^{n} X_i / \sigma$

$\to var(X) = \sum_{i=1}^{n} x_i / \sigma$

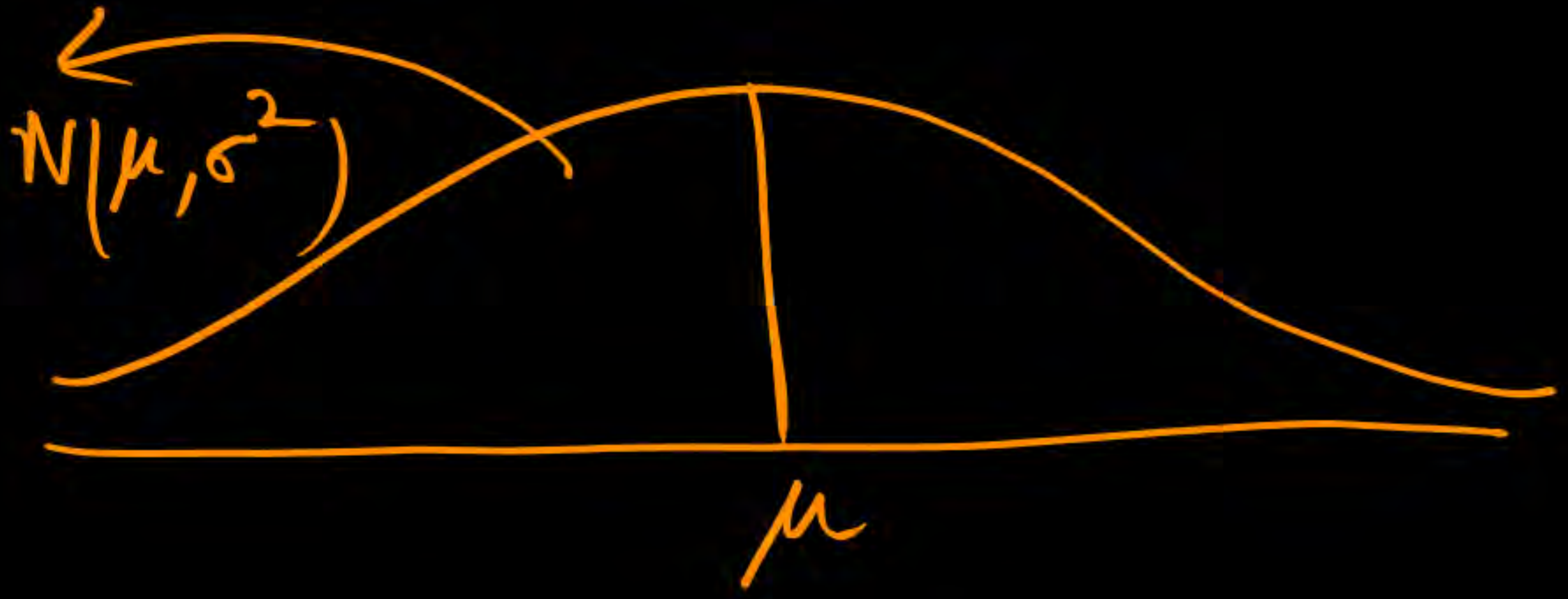
(d) $\sum_{i=1}^{n} X_i^2$

Statistic $\to$

$\sum_{i=1}^{n} x_i^2 = \sigma^2 = E[X^2] - [E(X)]^2$

$$\begin{cases} \text{SAMPLE mean } \bar{X} = \sum_{i=1}^{n} \frac{x_i}{n} = \sum_{i=1}^{n} x_i p_i \\ \sigma^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \mu)^2 \end{cases}$$

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



Introduction to Sampling Distribution

Q2. The weights (in kg.) of 5 workers in a factory are 56, 62, 74, 45 and 50. How many samples of size of 2 are possible with replacement? Also write all possible samples of size 2.

5 workers in a factory

weight 56, 62, 74, 45 and 50

56 62 74 45 50

with replacement

SAMPLE size

$n = 2$

A/B

SAMPLE NO	SAMPLE OBS	SAMPLE NO	SAMPLE observation
1	56, 62	13	
2	56, 74	14	
3	56, 45	15	
4	56, 50	16	
5	56, 56	17	
6	62, 56	18	
7	62, 74	19	
8	62, 45	20	
9	62, 50	21	
10	62, 62	22	
11		23	
12		24	
		25	

56
62
74
45
50

$N^n = 5^2 = 25$ possible outcomes

$\boxed{\text{25 SAMPLE}}$

25 Items

	56	62	74	45	50
56	56, 56	56, 62	56, 74	56, 45	56, 50
62	56, 62	62, 62	62, 74	62, 45	62, 50
74					
45					
50					

Introduction to Sampling Distribution

Q3. If lives of 3 Televisions of certain company are 8, 6 and 10 years then construct the sampling distribution of average life of Televisions by taking all samples of size 2.

3 Television
8, 6, 10 years.

SAMPLE NO	SAMPLE observation	MEAN
1	8,8	8 ✓
2	8,6	7
3	8,10	9
4	6,8	7
5	6,6	6 ✓
6	6,10	8 ✓
7	10,8	9
8	10,10	10
9	10,6	8 ✓

Frequency

✓ 6	1	1/9
✓ 7	2	2/9
✓ 8	3	3/9
✓ 9	2	2/9
✓ 10	1	1/9
		= 1

A, B — SAMPLE size $n = 2$

SAMPLE mean

$\mu = 8$ $S.D = \frac{4}{3}$

Introduction to Sampling Distribution

gate-2024 — 100%

Q4. Diameter of a steel ball bearing produced by a semi-automatic machine is known to be distributed normally with mean 12 cm and standard deviation 0.1 cm. If we take a random sample of size 10 with replacement, then find standard error of sample mean for estimating the population mean of diameter of steel ball bearing for whole population.

mean $\mu = 12$ cm

S.D $\sigma = 0.1$ cm

SAMPLE size $n = 10$

standard error of mean

$S(E) = \frac{\sigma}{\sqrt{n}} = \frac{0.1}{\sqrt{10}} = 0.032$

$N(\mu, \sigma_p^2) \rightarrow {}_S N(\mu, \sigma^2/n)$

Q5. The average weight of certain type of tyres is 200 pounds and standard deviation is 4 pounds. A sample of 50 tyres is selected. Obtain the standard error of sample mean.

$\mu = 200$ pounds.

$\sigma = 4$ Pounds. $n = 50$ tyres

standard Error of sample mean

$$= \frac{\sigma}{\sqrt{n}} = \frac{4}{\sqrt{50}} = \underline{0.57}$$

Introduction to Sampling Distribution

success 0.15   0.85

failure = 0.85 = 0.15

Q6. A machine produces a large number of items of which 15% are found to be defective. If a random sample of 200 items is taken from the population, then find the standard error of sampling distribution of proportion.

SAMPLE Proportion → standard Error of mean $= \sqrt{\dfrac{P(1-p)}{n}}$

P = prob. of success.
$1-P$ = Prob. of failure.
n = sample size.

$p = 0.15 \quad 1-p = 0.85$
$n = 200$ SAMPLE.

standard Error of mean $= \sqrt{\dfrac{0.15(1-0.85)}{200}}$

$= \underline{0.025}$

$\left\{ \begin{array}{l} P(0<z<2.86) = 0.4979 \\ P(0<z<1.43) = 0.4276 \end{array} \right.$

GATE - 100%

Q7. Average height of the students of science group in a college is 65 inches with a standard deviation of 2.2 inches. If a sample of 40 students is selected at random, what is the probability that the average height of these 40 students lies between 64 and 65.5 inches?

$\mu = 65$ inches.
$\sigma = 2.2$ inches.

Parent Distribution

SAMPLE
40 students

$64 \leq \bigcirc \leq 65.5$ inches.

Parent Distribution $X\left[\underline{N}(\mu_X, \sigma_X^2)\right]$ → SAMPling Distri $\bar{X} = N(\mu_{\bar{X}}, \sigma_{\bar{X}}^2)$

$$\Downarrow$$

$$\mu_{\bar{X}} = \mu_X$$

$$\sigma_{\bar{X}} = \frac{\sigma_x}{\sqrt{n}}$$

$$\left[\text{vari} = \frac{\sigma_x^2}{n}\right]$$

$\mu_{\bar{X}} = 65$

$\text{variance} = \frac{(2.2)^2}{40} = 0.12$

$\mu = 65$

$S.D = 0.34$

$$P[64 \le \bar{X} \le 65.5] = \left[\frac{64-\mu}{\sigma} \le \frac{\bar{X}-\mu}{\sigma} \le \frac{65.5-\mu}{\sigma}\right]$$

z score

$$= \left[\frac{64-65}{0.34} \le Z \le \frac{65.5-65}{0.34}\right]$$

$$= P[-2.87 \le Z \le 1.43]$$

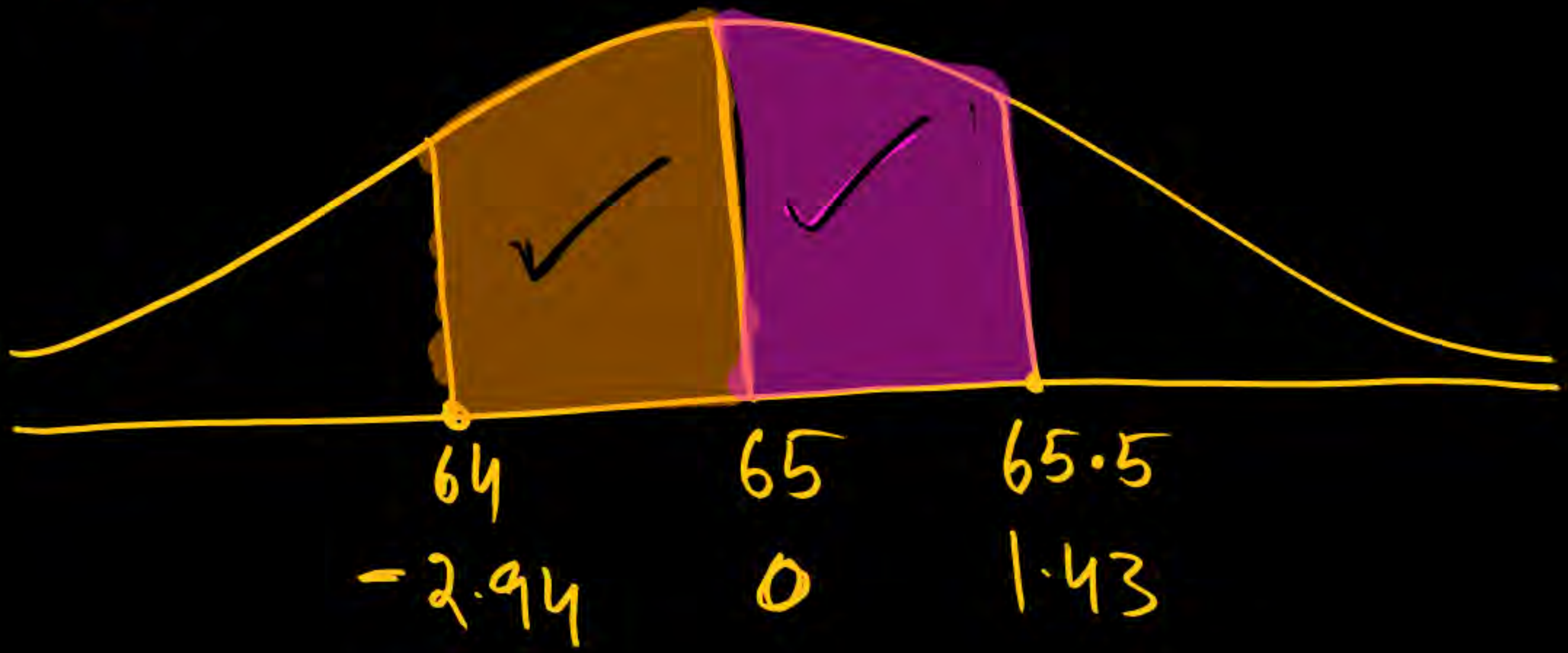
$$\Rightarrow P[-2.94 \le z \le 1.43]$$

draw the Prob.

$$= P(-2.94 \le z \le 0) + P(0 \le z \le 1.43)$$

$$= 0.4976 + 0.4236$$

$$= \underline{\underline{0.92}} \ \underline{\text{Ans}}$$



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

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

$P(\bar{X} \geq 70) = P\left[\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \geq \frac{70-\mu}{\sigma/\sqrt{n}}\right]$

$= P\left[Z \geq \frac{70-60}{25/\sqrt{36}}\right]$

$= P[Z \geq 2.40]$

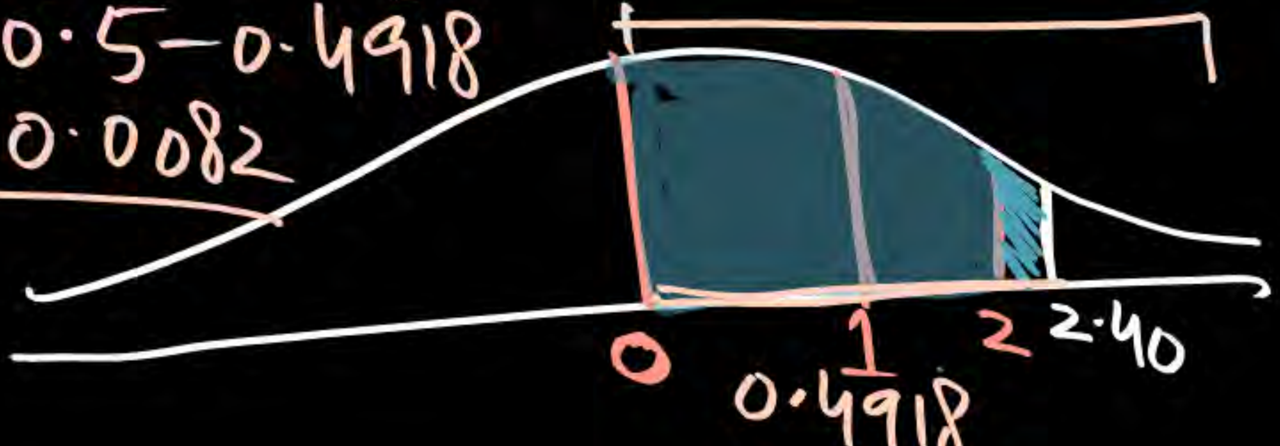
$= P[Z \geq 2.40]$

$= 0.5 - 0.4918$

$= 0.0082$

✓ $P(0 \leq Z \leq 2.40) = 0.4918$

✓ $P(0 \leq Z \leq 1.20) = 0.3849$



Ques-1 SAQ 2 marks

In IIT-K

Average weight male $= 60$ kg

S.D $= 25$ kg

SAMPLE — 36 male

1) find the Prob. Male student

A) more than 70 kg $= 0.0082$

B) less than 55 kg 0.1151

C) between 50 kg and 65 kg. ✓ 0.8767

$$P(\bar{X} < 55) = P\left(\frac{\bar{X}-\mu}{\sigma} < \frac{55-\mu}{\sigma/\sqrt{n}}\right)$$

$$= P\left[Z < \frac{55-60}{25/\sqrt{36}}\right] \quad n=36$$

$$= P[Z < -1.20]$$

$$= 0.5 - 0.3849$$

$$= \underline{0.1155}$$

$$P(Z > 1.20)$$

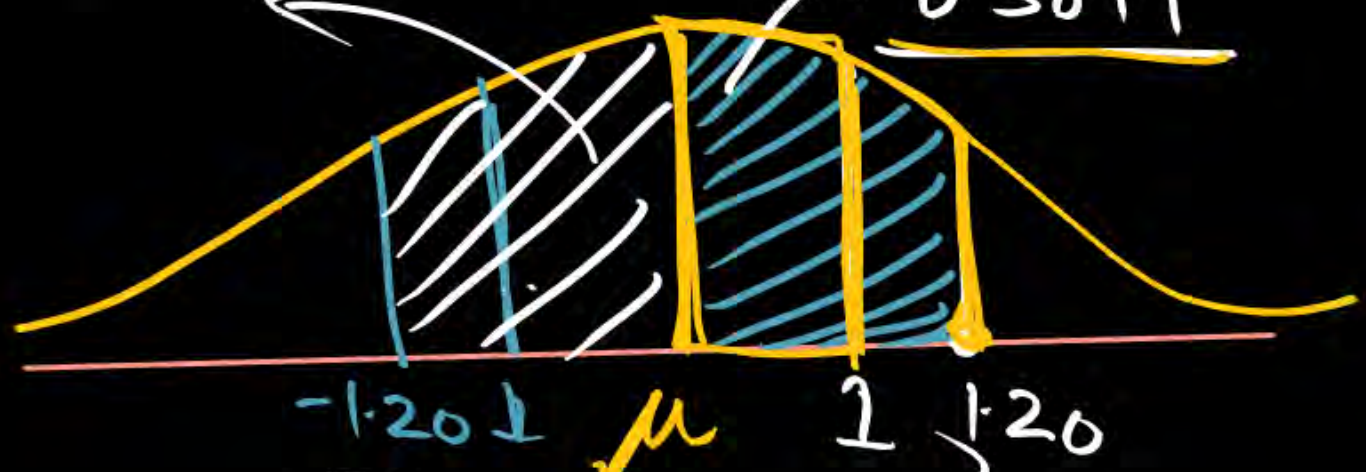
$$= 0.5 - 0.3849$$

$$= \underline{0.1145}$$

0.34

0.3849

0.3849



symmetric

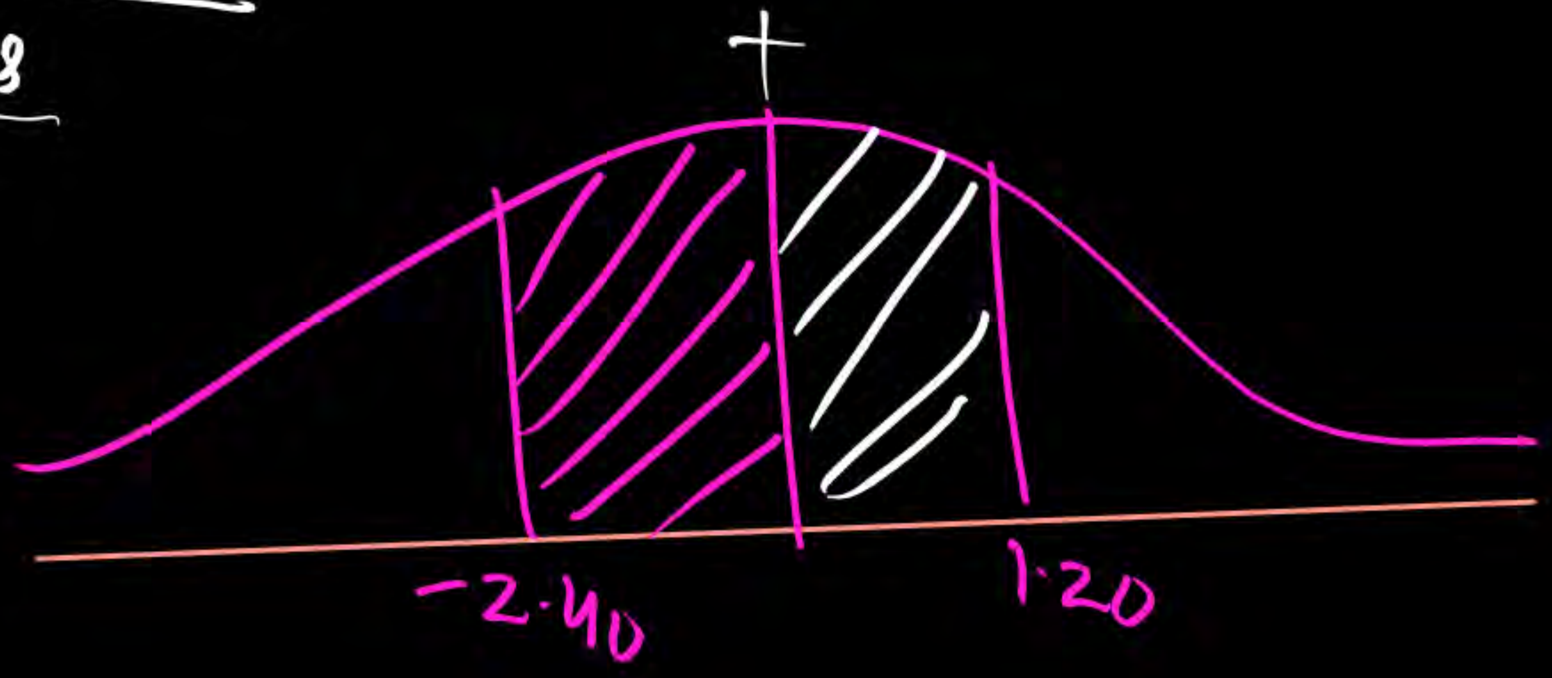
$$P[50 \le \bar{X} \le 65] = P\left[\frac{50-\mu}{\sigma} \le \frac{\bar{X}-\mu}{\sigma} \le \frac{65-\mu}{\sigma}\right]$$

$$= P\left[\frac{50-60}{\frac{25}{\sqrt{36}}} \le Z \le \frac{65-60}{\frac{25}{\sqrt{36}}}\right]$$

$$= P(-2.40 \le Z \le 1.20)$$

$$= \underline{0.8767}$$

<u>Ans</u>



The Law of Large No:

Throwing A Die

X = No. of dots

$P[X=x] = \frac{1}{6}$

If No. of trials Are Increased Population mean equivalent of SAMple mean

3.5 — converge Point

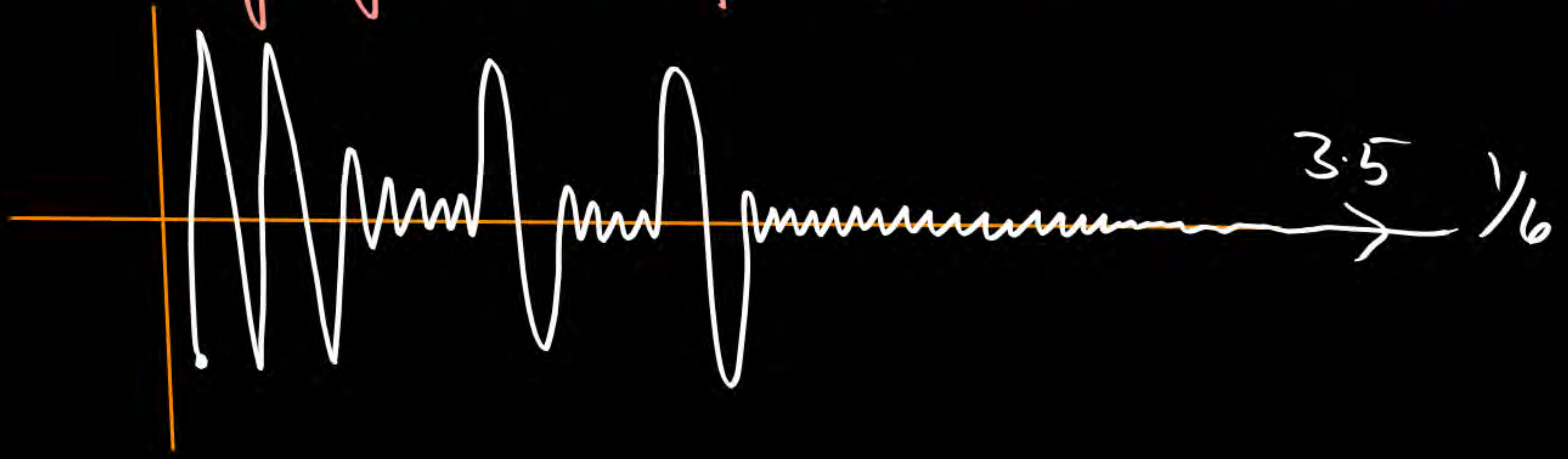
✓ Die Throw $\underline{1, 2, 3, 4, 5, 6}$ Equal Proportion (Balanced)

$P(A) = \frac{1}{6}$

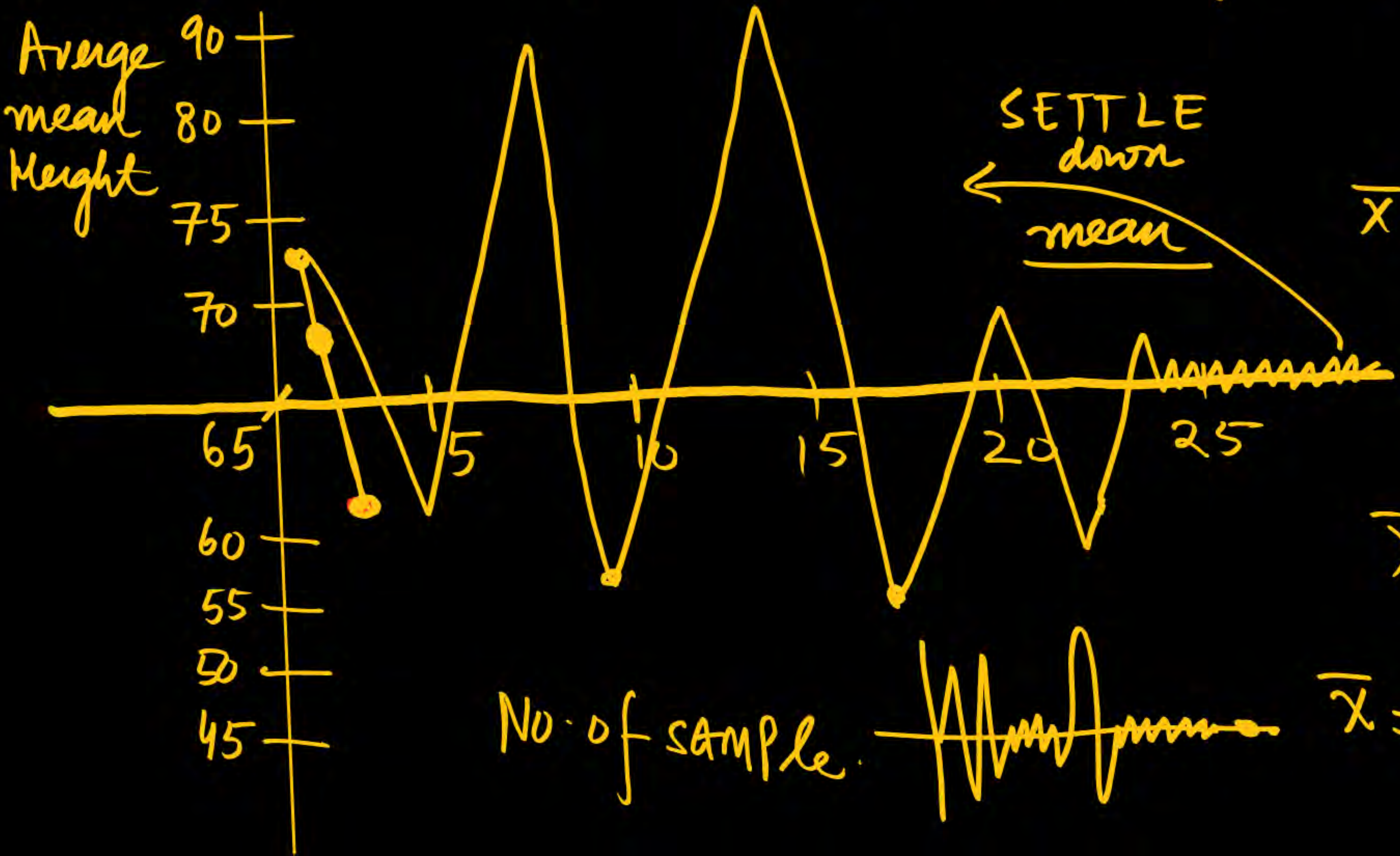
✓ If Die is Thrown randomly and Trials Are Increased

$\underline{5 \quad 4 \quad 3 \quad 2 \quad 1 \quad 3}$ $P(6) = 0$

→ What going on → If No. of trials are increased



If Average weight of male = 65 kg standard dev. 25 kg.



$X_1 = 70\,kg$

$X_2 = 55\,kg$.

$\bar{X} = \frac{X_1 + X_2}{2} = \frac{70+55}{2}$

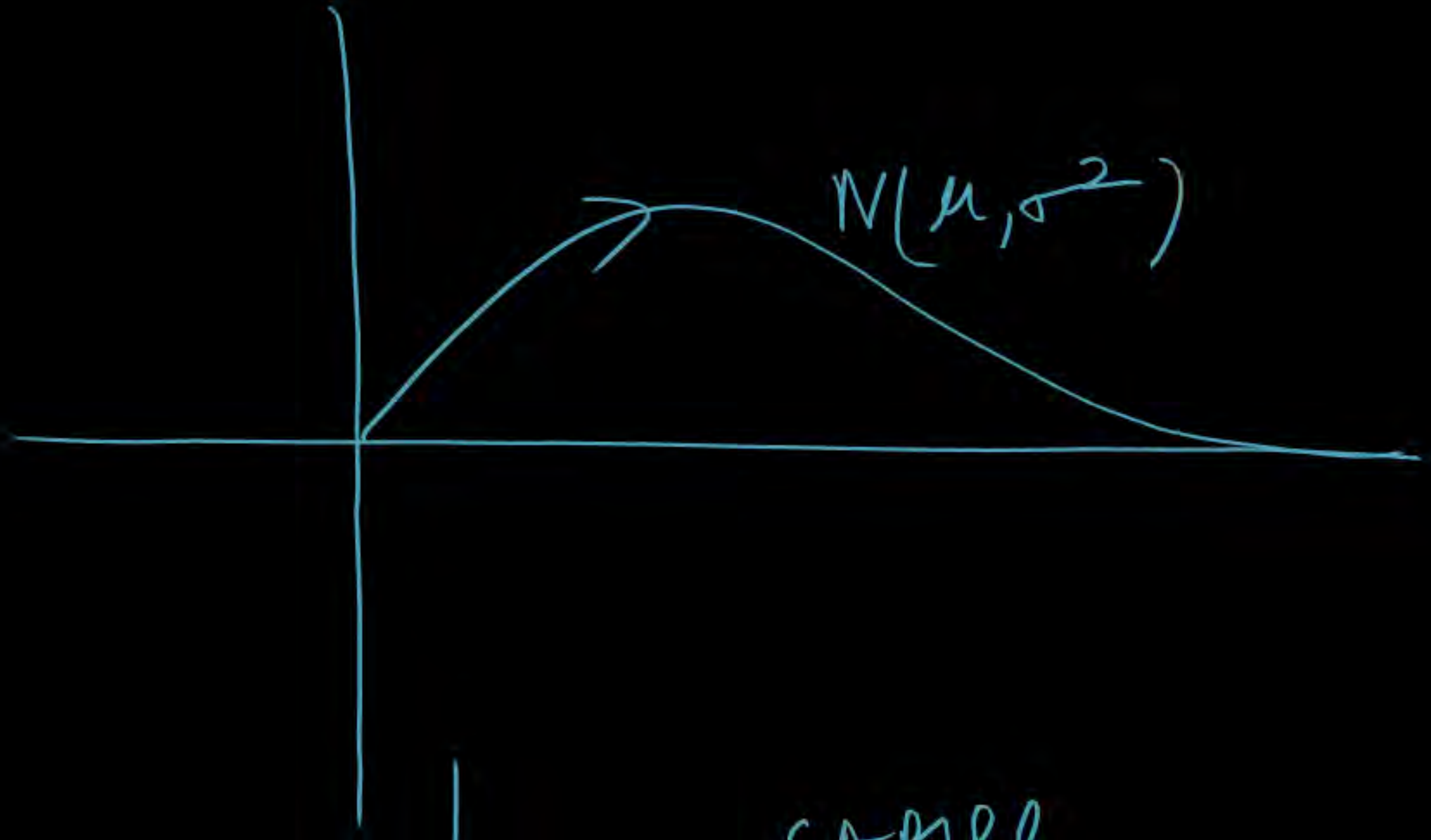
$= 62.5$

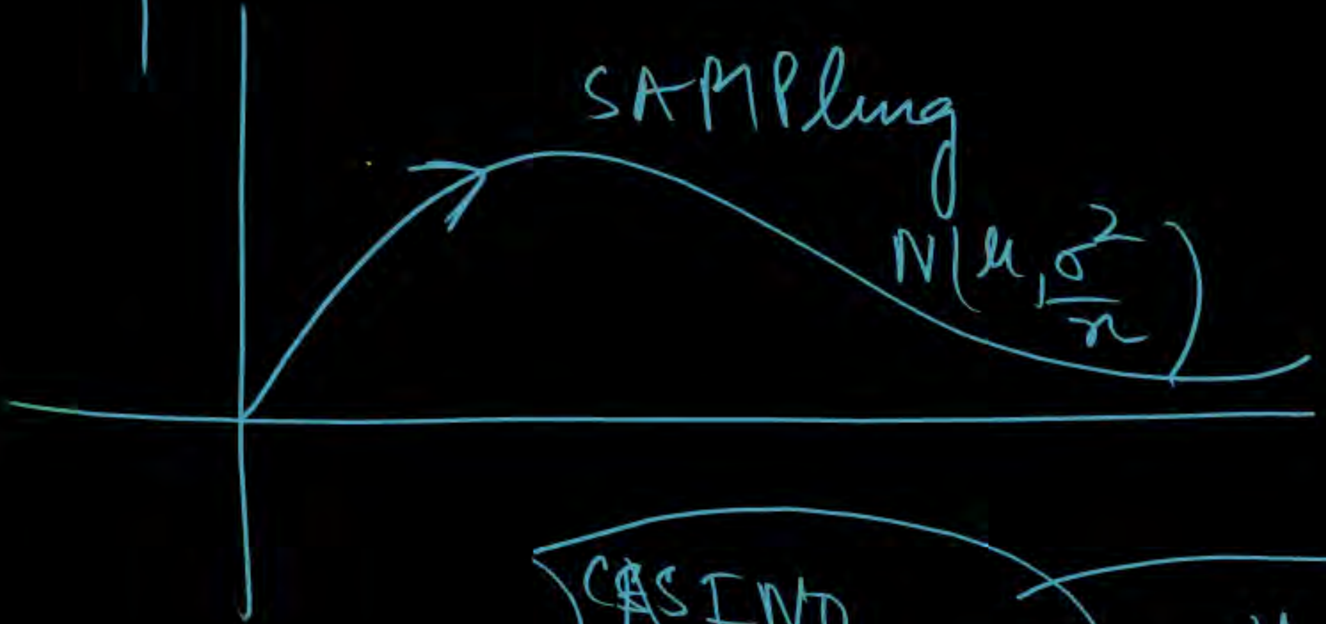
$X_3 = 80\,kg$

$\bar{X} = \frac{80 + 62.5}{2} = 68.33$

$X_4 = 83\,kg$

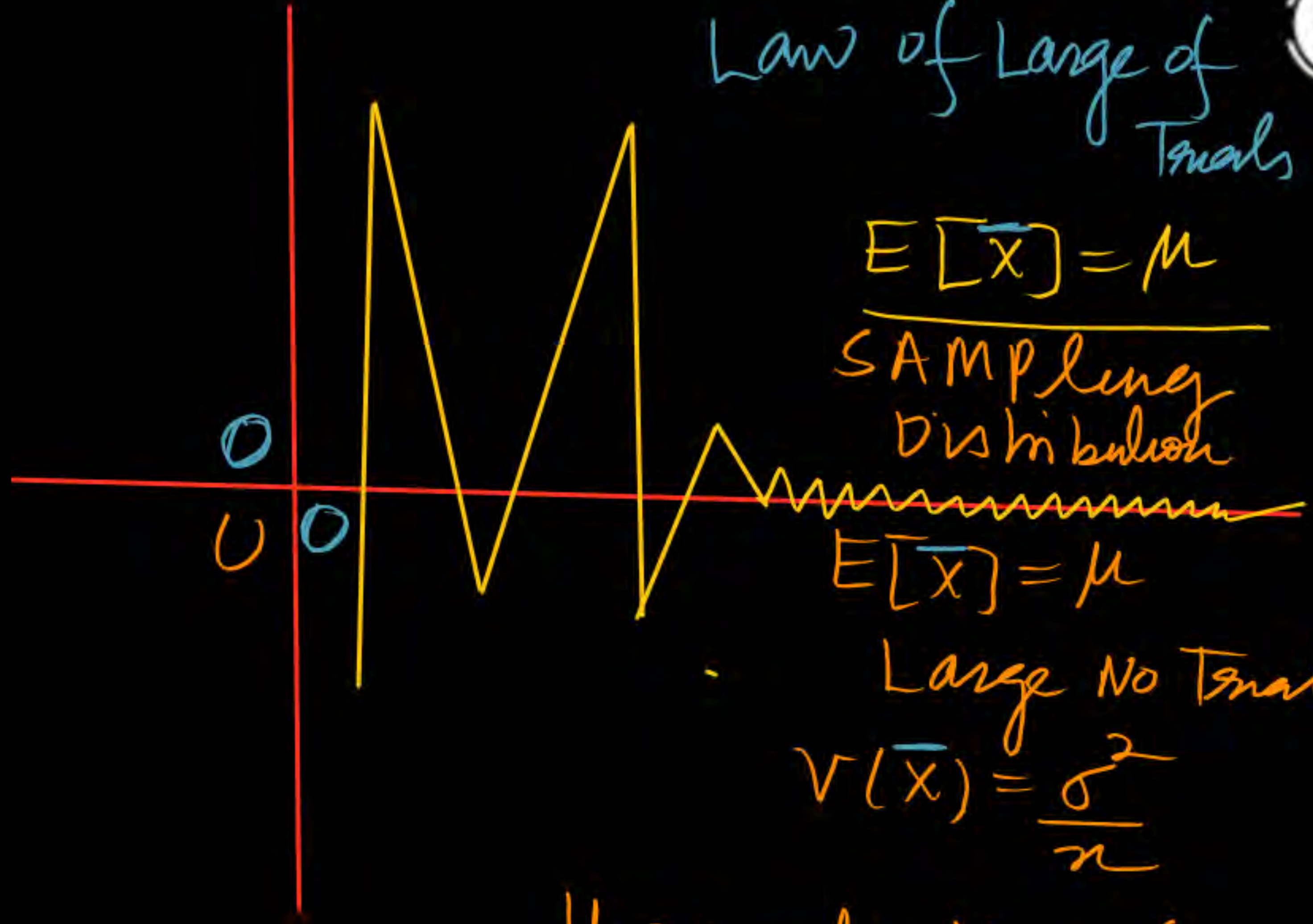
$\bar{X} = \frac{83 + 68.30}{2} = 77.12$





CASINO

How much No. of Trials



Law of Large of Trials

$E[\overline{X}] = \mu$

SAMPling Distribution

$E[\overline{X}] = \mu$

Large No Trials

$V(\overline{X}) = \frac{\sigma^2}{n}$

How much No. of Trials mean settle down

✓ Frequency Are Settle down

THANK - YOU