23.08.2023
Wednesday

Date: Page:

(iii) Area = 0.4 40th

Sol: 0.65542

Heart Beat Rate

Class Interval	f	x_i	$f x_i$	$f x_i^2$
50.5 - 55.5	4	53	212	
55.5 - 60.5	14	58		
60.5 - 65.5	24	63		
65.5 - 70.5	40	68		
70.5 - 75.5	54	73		
75.5 - 80.5	43	78		
80.5 - 85.5	36	83		
85.5 - 90.5	20	88		
90.5 - 95.5	11	93		
95.5 - 100.5	3	98		
100.5 - 105.5	1	103		

$\mu = 1701.0818$

$\Sigma f_i x_i^2 = 1425370$

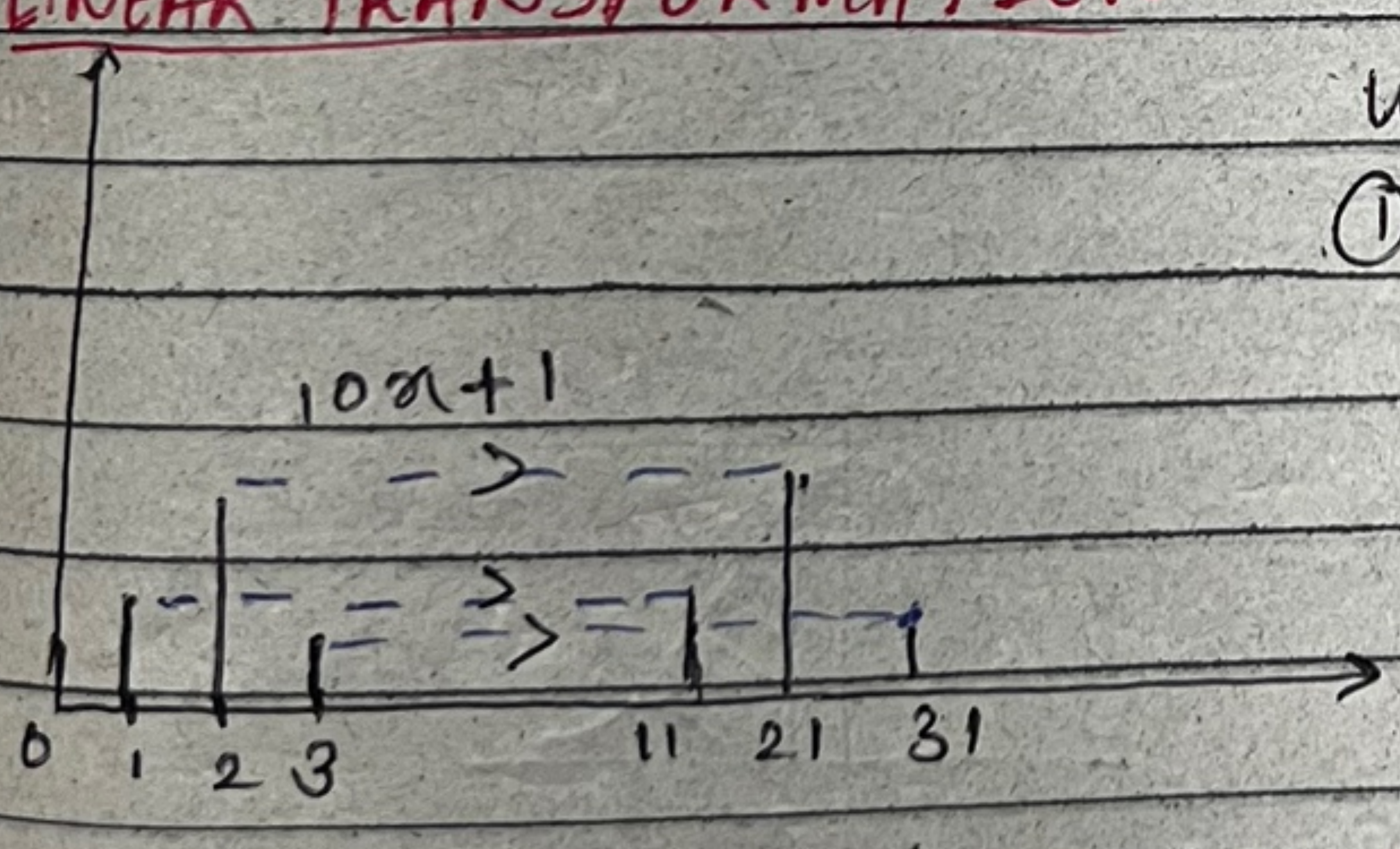
18720

$\mu = 74.88$

$\boxed{\mu \approx 75}$

$\boxed{\sigma = 10}$

LINEAR TRANSFORMATION



~~No. of~~

What does not changes?

① No. of instanced before transformation = No. of instances after trans^n

② Probability Dist^n remains same

Outcome of Random Var. = Sample Space

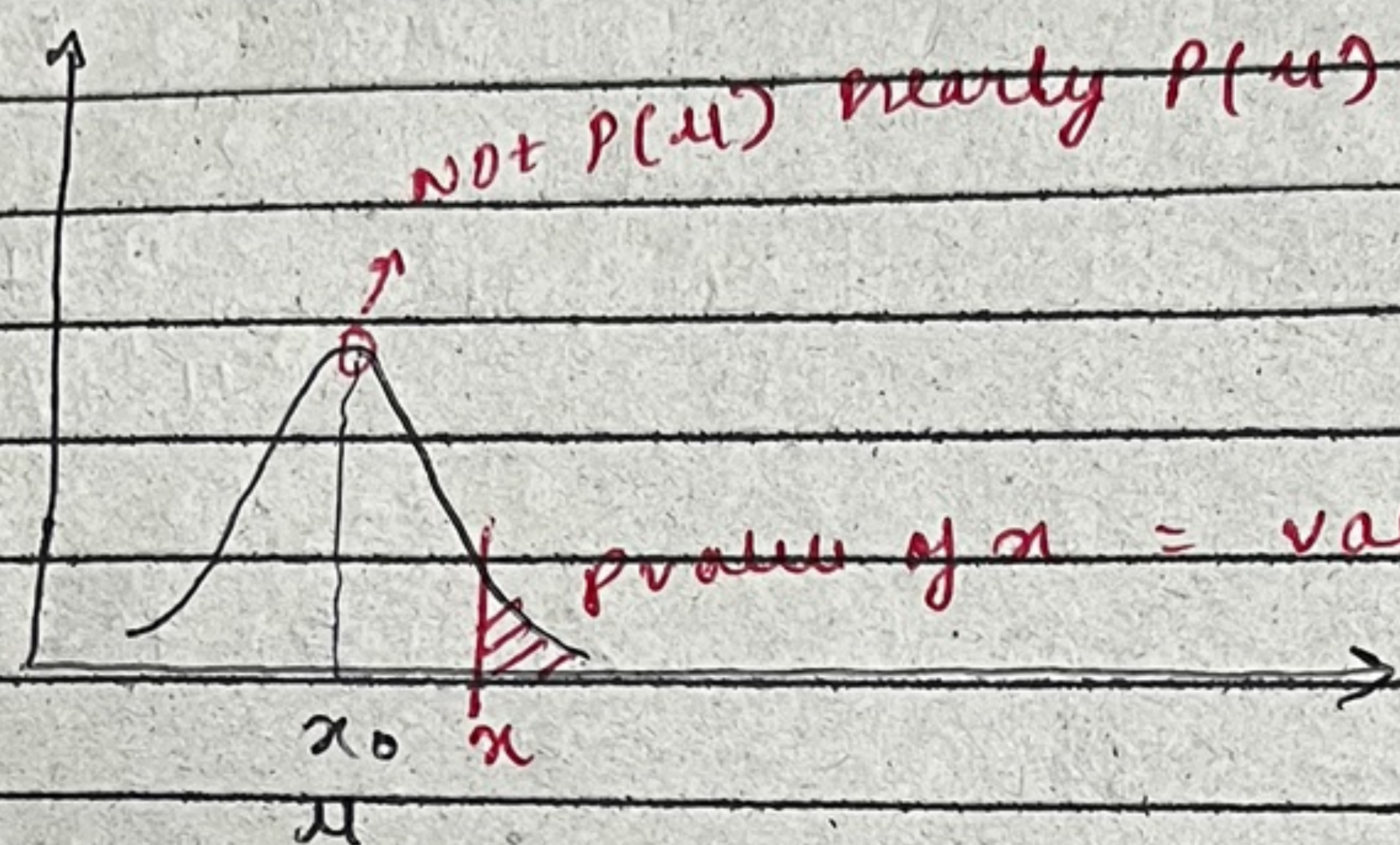
What changes?

① Mean ② Standard deviation

③ Variation

ON Normal Distribution (Linear Transf^n)

Not $P(\mu)$ nearly $P(\mu)$ interval basis



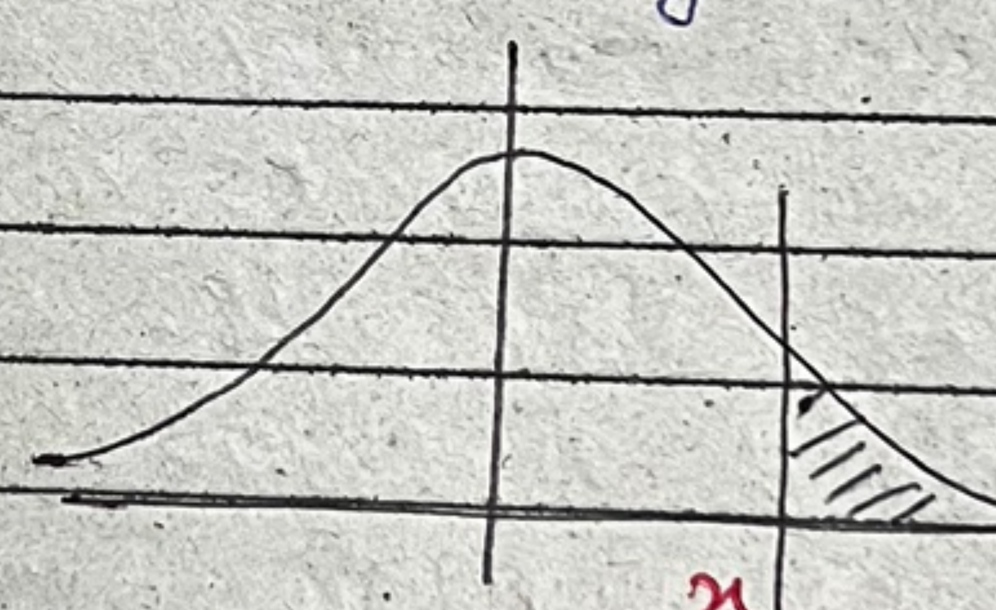
pvalue of x = value of area after pt. x

① Find

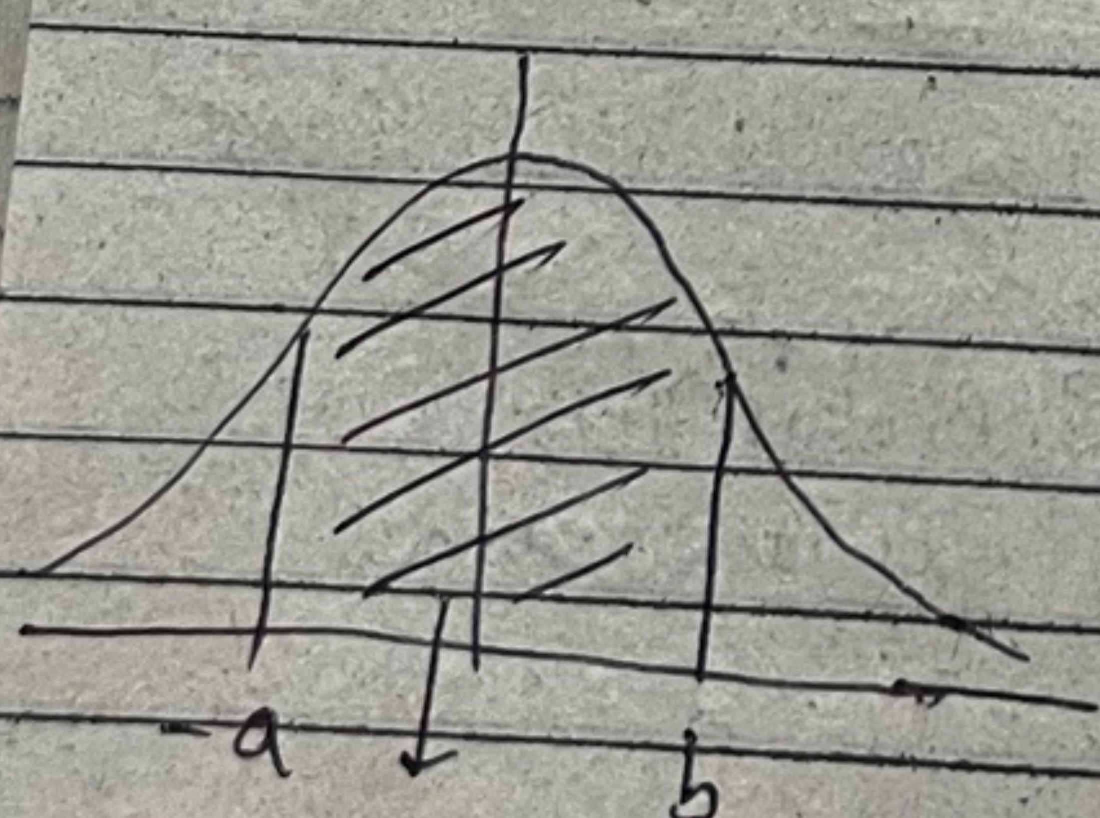
Q) 2% of whole population is behaving abnormally.

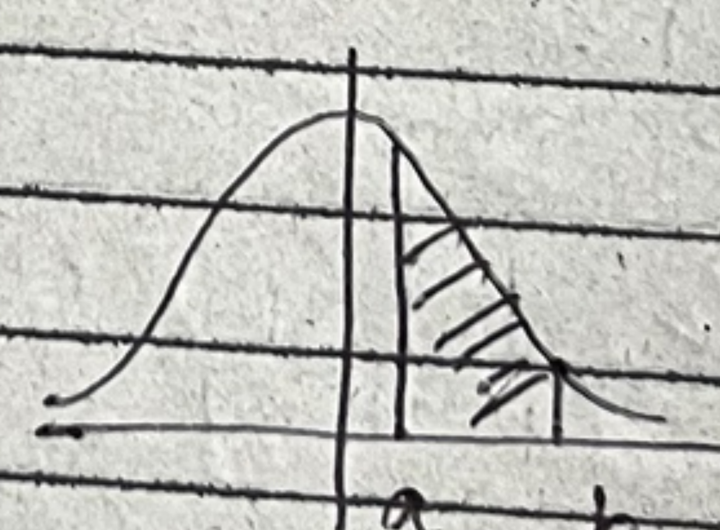
Sol. 0.98] 0.02
↳ Outlier

x: After which outlier exists | Abnormal



Area after x ; p value of pt. x





Area = Pvalue at a - Pvalue at b

b = a

Area = 1 - Pvalue of b - Pvalue of a

Mean
Median
Mode
Std deviation

Total Prob. ✓
Bayes Rules ✓
Binomial ✓
Poisson
Normal

P value

$$P(Z \geq z) = \frac{1}{2} e^{(bz + az^2)}$$

$$z = \frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

$a = -0.416$
$b = -0.717$

$c = -\log_2(2p)$

Pvalue of $z = 0.02$

$c = 1.3979$

$$\frac{-0.717 - \sqrt{(0.717)^2 + 4 \times (0.416)(1.3979)}}{-(2 \times 0.416)}$$

$= \cancel{+2.887}\ 2.06$

$$z = \frac{x - \mu}{\sigma} \Rightarrow \cancel{2.88}$$

$z\sigma + \mu = x$

$x = \cancel{2.887}\ 2.06 \times 10 + 75$

$= \cancel{103.87}\ 95.6$

≈ 96

After 96 $\rightarrow$ heart is outlier abnormal behaviours.

Q) % of population b/w 68 to 77 beats per min

Sol:
$x = 68$ to 77
$z = -0.7$ to 0.2

Area $= 1 -$ Pvalue (0.7) $-$ Pvalue (0.2)

$bz + az^2$

25.08.2023
Friday

Date: Page:

1$ for ticket

X_1	0	5	10	15	20	
$X = X_1 - 1$	-1	4	9	14	19	
$5X_1$	0	25	50	75	100	→ New ~~Prize~~ Prize

$Y = 5X_1 - 2$
$= 5X + 3$

No. of instances will be same after linear transformatn

Mean of New - changes
variance does not changes

When Variance changes -
① If new instances are added ⇒ Var ↑ses
② Interval size change ⇒ Var ↑ses if IS ↑

$Var[X+Y] = Var[X] + Var[Y]$
$Var[X-Y] = Var[X] + Var[Y]$

$Var[X+X] \neq Var[2X]$
↳ Not linear Transfn

2 times sample space

X x_1 x_2 x_3
P_1 P_2 P_3

Find combn for $X + X$

25.08.2023
Friday

Date: Page:

1$ for ticket

X_1	0	5	10	15	20
$X = X_1 - 1$	-1	4	9	14	19
$5X_1$	0	25	50	75	100

→ New ~~Pros~~ Prize

$Y = 5X_1 - 2$
$= 5X + 3$

No. of instances will be same after linear transformatⁿ

Mean of New · changes
variance does not changes

When Variance Changes –
① If new instances are added ⇒ Var ↑ses
② Interval size change ⇒ Var ↑ses if IS ↑

$Var[X+Y] = Var[X] + Var[Y]$
$Var[X-Y] = Var[X] + Var[Y]$

$Var[X+X] \neq Var[2X]$

↳ Not linear Transfⁿ

2 times sample space

#	X	x_1	x_2	x_3
		P_1	P_2	P_3

Find combⁿ for $X + X$

Date: Page:

x	$\cancel{x}_1$	x_2	$\cdots$ x_n
x_1	P_1P_1	P_1P_2	
x_2			
$\vdots$			
x_n			

$x_i + x_j$ $\quad$ $x_j + x_i$

P_iP_j $\quad$ P_jP_i

Same

merge

Unnecessary to write 2 times.

Instance : $x_i + x_j$

Probability : $2P_iP_j$

$E[X] = \sum x_i P_i$

↳ Expected value of X

$E[aX+b] = \sum (ax_i + b) P_i$

$= a \sum x_i P_i + b \sum P_i$

$E[aX+b] = a\, E[X] + b \quad [\because \sum P_i = 1]$

$i=1 \quad j=1, j=2, \ldots, j=n \qquad i=2 \quad j=1, \ldots, j=n$

$E[X+X] = \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i + x_j) P_i P_j$

$= \sum_{i=1}^{n} \left[x_i P_i \sum_{j=1}^{n} P_j + \right.$

$= \sum_{i=1}^{n} \left[x_i P_i + P_i E[X] \right]$

$\Rightarrow E[X+X] = \cancel{2}\ 2\, E[X]$

Date: Page:

$$Var[X+X] = \cancel{E[(X+X)^2]}\; E[(X+X)^2] - \{E(X+X)\}^2$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i+x_j)^2 p_i p_j - \{2E[X]\}^2$$

(Ranlin)

$$= \sum_i\sum_j (x_i^2 + x_j^2 + 2x_i x_j) p_i p_j - \text{''}$$

$$= \sum_i \sum_j [x_i^2 p_i p_j + x_j^2 p_i p_j + 2x_i x_j p_i p_j]$$

consider i constant

$$\sum_i [x_i^2 p_i + p_i E[X^2] + 2x_i p_i E[X]]$$

$$E[X^2] + E[X^2] + 2(E[X])^2$$

$$2E[X^2] - 2(E[X])^2$$

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

$$\boxed{Var[X+X] = 2\,Var[X]}$$

$E[X + Y]$

n trials of a same game $\equiv$ n machines पर एक साथ खेलूं।

Exp. is same.

↳ Sample space is same

↳ X

$X' = x_1, x_2, x_3 \ldots x_n$

$E[X'] = n E[X]$

$Var[X'] = n\, var[X]$

Why sometimes $n^2 var[X]$ & sometimes $n\, var[X]$

Linear Transⁿ — increasing no. of observation

$$\boxed{\begin{aligned} E[X+Y] &= E[X] + E[Y] \\ E[X-Y] &= E[X] - E[Y] \end{aligned}}$$

$Var[X-Y] = E[(X-Y)^2] - [E(X-Y)]^2$

Proof

$= \sum_{i=1}^{n}\sum_{j=1}^{m} (x_i - y_j)^2 p_i p_j - (E(X-Y))^2$

$= \sum_{i=1}^{n}\sum_{j=1}^{m} (x_i^2 + y_j^2 - 2x_i y_j) p_i p_j$ — ‖ —

$= \sum (x_i^2 p_i p_j + y_j^2 p_i p_j - 2x_i y_j p_i p_j)$ — ‖ —

$= \sum_{i=1}^{n} (x_i^2 p_i + E(Y^2) p_i - 2x_i p_i E(Y))$ — ‖ —

$= E[X^2] + E[Y^2] - 2E[X]E[Y] - (E[X-Y])^2$

$E[X^2] + E[Y^2] - 2E[X]E[Y] - (E[X])^2 - (E[Y])^2 + 2E[X]E[Y]$

$= (E[X^2] - (E[X])^2) + (E[Y^2] - (E[Y])^2)$

$= var[X] + var[Y]$

$$\boxed{var[X-Y] = Var[X] + var[Y]}$$

LSS ICT

Date: Page:

x 1 2 3

$(x-\bar{x}) = -2$ -1 0 1 2 → 5 instances

Var. ↑ se

$(x+\bar{x}) =$ 1 2 3 4 5 6

Var ↑ ses

No. of instances are increased.

(interval same)

Systolic & Diastolic

BP

↓

E[SBP] + E[DBP]

E[XY]

V[X Y]

06.09.2023
Wednesday

Date:________ Page:________

SAMPLING :-

$a_1, a_2, \ldots, a_n$ $(a_i \neq a_j)$

① Pair of dice → 36 individuals (different)
② For measuring characteristic we define Random Variable
③ Characteristics may or may not be same.

Statistical Constants for knowing characteristics
mean, variance, standard deviation ✓
Proportion (Qualitative characteristics)
↓
① Characteristics : Smoker

Proportion of smoker = $\dfrac{\text{No of students} \in \text{smoker class}}{\text{Total No. of Students}}$

[same formula as probability]

② Characteristics : age < 20 → returns Boolean

Probability Distribution ~~& no~~ must be known.
x_i & P_i must be known. before knowing Statistical constants.

Problem :- (for whole population analysis)
① Population is infinite (Time ↑ses)
② Object may be destroyed.
Car - Air bag example (time to open airbag)
- needs to destroy objects
③ Money problem. Survey - costly (budget)

$x_1, x_2, \ldots, x_N$ → Population

Date: Page:

Solution: Observe some of the individuals of population

Census :- study of whole population
Sampling :- study of group of population

(Statistical constants) of sampling = Parameters

Population Parameters

$X \rightarrow \mu_x \quad \sigma_x^2 \quad \sigma_x \quad P$

Random variable

N : Population size
n : ~~Population size~~ Sample size

(x_i → i^{th} unit of all different sample)

$X_1 \quad X_2 \quad X_3 \quad \ldots \quad X_n$ → Samples

$\boxed{\text{Sampling}}$

$\boxed{\text{Random Sampling}}$

Random Sampling with replacement
Repetition will be same
No. of samples = n^N

Random Sampling without replacement
No two individuals will be same
No. of samples = ${}^N C_n$

Variance सिर्फ Pt. का Possible नहीं है।

(a) N is very very large ⇒ N^n or $^N C_n$ can't be calculated

Sample's unit X_i → can be can any value of Population

∴ Expected value of Sample = Expected value of Population = μ_x

(Deviation of 2nd unit) population $= x_2 - \mu_x$

Variance can only be of Random Variable.

$Var(X_2)_{sample} = \sigma_x^2$

$\bar{X} \rightarrow \begin{cases} \text{Individuals} \rightarrow \text{different samples} \\ \text{Characteristics} \rightarrow \text{mean of samples} \end{cases}$

Defination : $\bar{X} = \frac{1}{n}[X_1 + X_2 + X_3 + \dots + X_n]$

↳ Expectation variance same as population

(sampling mean) / Mean of different samples — Random Variable for sample's mean (Sampling Distribution) $= \bar{x}_1$

NOTE $\cancel{(\bar{x}_1)}$ $(\bar{x})_1$

$(\bar{x})_1 = \frac{X_1 + X_2 + \dots X_n}{n}$

$\bar{X} = [\bar{x}_1, \bar{x}_2, \dots\dots, \bar{x}_{^N C_n \text{ or } N^n}]$

Date:______ Page:______

$$x_1 \quad x_2 \quad \ldots \quad x_n$$
$$\downarrow \quad \downarrow \quad \quad \downarrow$$

Expected value μ_x μ_x μ_x

Expected value of mean $= \frac{n\mu_x}{n} = \mu_x$

1, 2, 3 $\mu_x = 2$ (Population)

1, 2   1, 3 (samples)

~~$\mu_x = 1.5$~~   ~~$\mu_x = 2$~~

$\bar{X} = 1.5$ $\bar{X} = 2$

2, 3

$\bar{X} = 2.5$

$$\frac{\Sigma \bar{X}}{3} = \frac{6}{2} = 2$$

$P(\text{correct estimation of mean}) = \frac{1}{3}$

Sample Statics (μ(sample) var(sample))
↳ determined from individual sample

Sample may or may not be representative of population.
(Prob. Distribution should be same)

Mean may or may not be valid instance.

Statistical constants → Sampling Statistics

When we consider all the samples & consider their mean & calculate its mean
= μ_x (Population mean)

Sampling Error

Diff. Sample statistics and Population mean

Stand. Error of sampling mean ↓

$$\frac{1}{M}\left[\Sigma(\bar{x}-\underset{\text{pop. mean}}{\mu})^2\right]$$

Point estimation — mean = 5 (exact)
किसी एक value से

Interval estimation — mean = 5 ± 0.2 (interval mean)
↳ एक sample से ~~ज्ञात~~ बता
पाया जाता है।

Interval estimation of Population mean में Probability of correct estimation increase हो जायेगी as compared to Point estimation.

09.09.2023

Player	Height
A	76
B	78
C	79
D	81
E	86

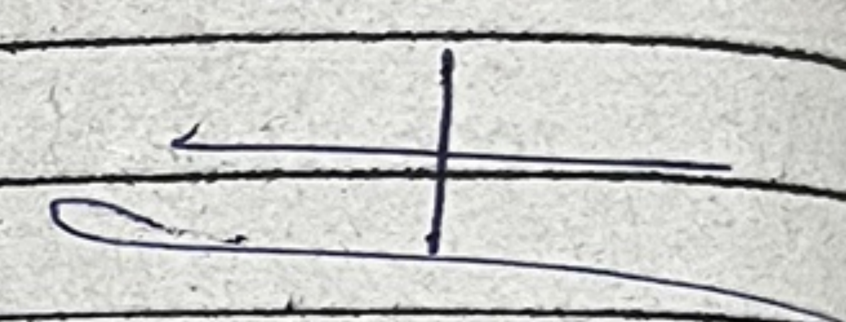
$\mu = 80$

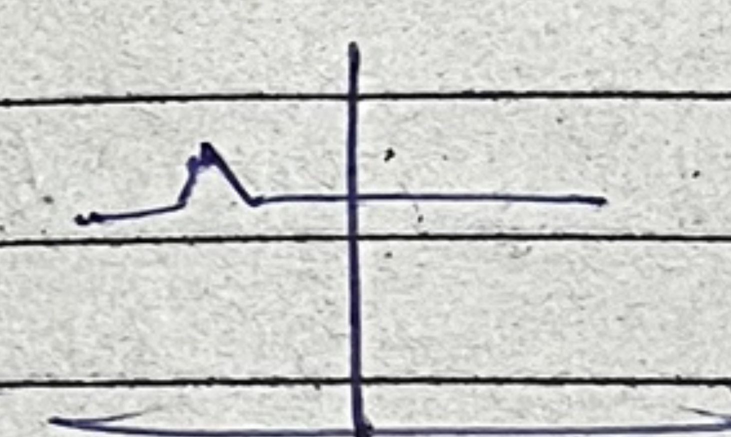
${}^5C_2 = \frac{5 \times 4^2}{2} = 10$

(samples of 2)

$P = \frac{1}{5}$ uniform constant line

(A, B, C)	(A, B)	77
(A, B, D)	(A, C)	77.5
(A, B, E)	(A, D)	78.5
(A, C, D)	(A, E)	81
(A, C, E)	(B, C)	78.5
(A, D, E)	(B, D)	79.5
(B, C, D)	(B, E)	82
(B, C, E)	(C, D)	80
	(C, E)	82.5
	(D, E)	83.5





Point Estimation

P(Correct estimation) $= \frac{1}{10}$

Error of approximation $= 1 - \frac{1}{10} = \frac{9}{10}$ (No margin)

Interval estimation

margin $= 0.5$ Error of app. $= 1 - \frac{2}{10} = \frac{8}{10}$

margin $= 1$ Error of appr. $= 1 - \frac{3}{10} = \frac{7}{10}$

Prob of error

Prob. of error ~~reduces~~ increase as interval size increases.

(B, C, D, E) = 81
(A, C, D, E) = 80.5
(A, B, C, D) = 78.5
(A, B, C, E) = 79.75
(A, B, D, E) = 80.25

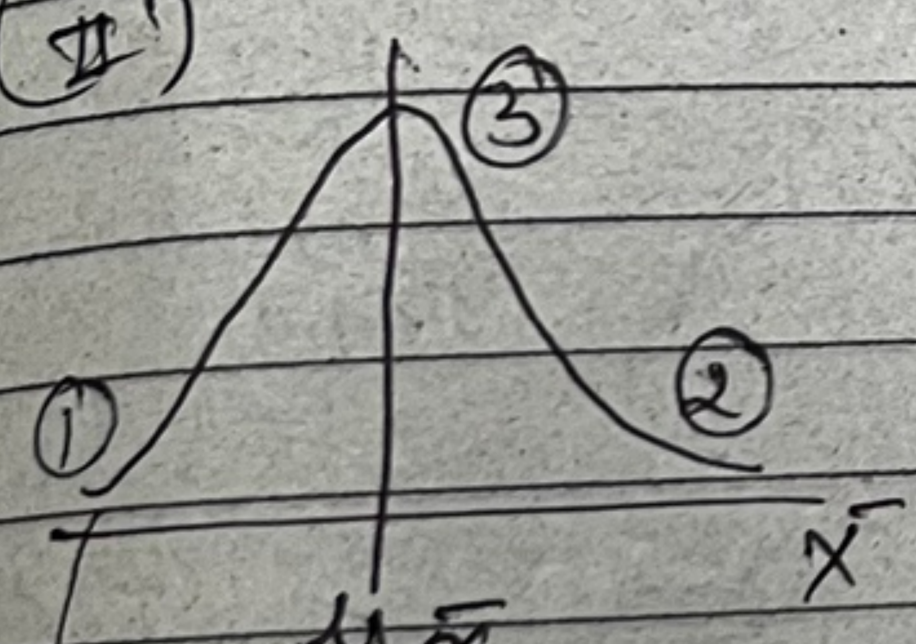
P(correct estimation) = 1

P(correct estimation) {margin = 0.25} $= \frac{2}{10}$

As we increase sample size ⇒ P(correct estimation) ↑

$$\bar{X} = \frac{1}{n}[x_1 + x_2 + x_3 + \dots x_n]$$

(II)



(I)

Sample from ①$_I$ & ②$_I$ → ③$_{II}$

①$_I$ ③$_I$ ③$_I$ → ③$_{II}$

Individual → samples

Samples having mean very far from expectation

Sample ①$_I$ ①$_I$ → ③$_I$ ①$_{II}$

Sample ②$_I$ ②$_I$ → ②$_{II}$

(Variance) sampling distribution < (Variance ~~Sampling~~) Random Variance

But mean is sam

$\bar{X}$: Sampling

Date: Page:

$$\bar{X} = \frac{1}{n}[x_1 + x_2 + \dots x_n]$$

$$E[\bar{X}] = \frac{1}{n}(E[x_1] + E[x_2] + \dots E[x_n])$$

$$= \frac{1}{n} \cdot \mu_x n$$

$$E[\bar{X}] = \mu_x$$

$\rightarrow$ Scaling

$$Var[\bar{X}] = \frac{1}{n}x_1 + \frac{1}{n}x_2 + \dots + \frac{1}{n}x_n$$

$$= \frac{1}{n^2}[\sigma^2 + \sigma^2 + \dots \sigma^2]$$

$$\boxed{Var[\bar{X}] = \frac{\sigma^2}{n}} \qquad \sigma[\bar{X}] = \frac{\sigma}{\sqrt{n}}$$

$\therefore$ Scatteredness is less

True only till $n \neq 1$

Only true for smaller value

$(1, 2, 3)$

$x_1\ x_2\ x_3$

$n=1 \qquad Var[X] = \sigma^2$

$n=n \qquad Var[X] = \frac{\sigma^2}{n} \qquad$ n is large

$\therefore$ This formula is saying variance ≈ 0 but it should exactly 0

$\therefore$ We use correction factor on next page

13.09.2023 Wednesday.

$E[z] = (E[\bar{x}] - E[\mu])\frac{\sqrt{n}}{\sigma} = (\mu-\mu)\frac{\sqrt{n}}{\sigma} = 0$

Date: Page:

$z = \left(\frac{\bar{x}-\mu}{\sigma}\right)\sqrt{n} \rightarrow Var(z) = \frac{n}{\sigma^2}\{\sigma^2[\bar{x}] - \sigma^2[\mu]\} = \frac{n}{\sigma^2}\left(\frac{\sigma^2}{n} - 0\right) = 1$

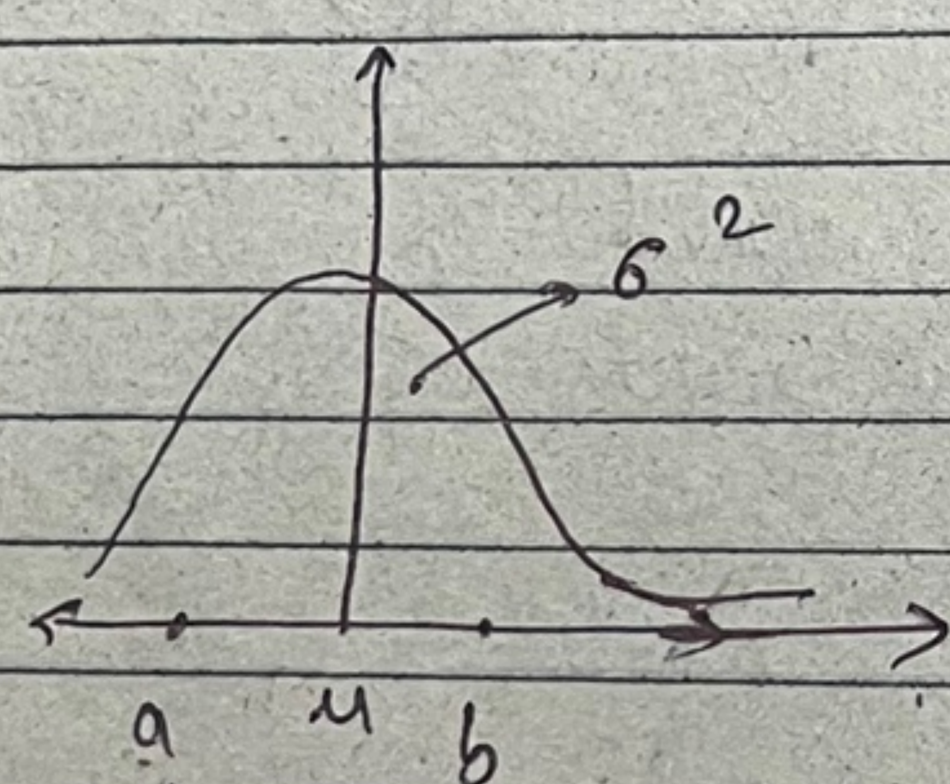
{Earlier formula}

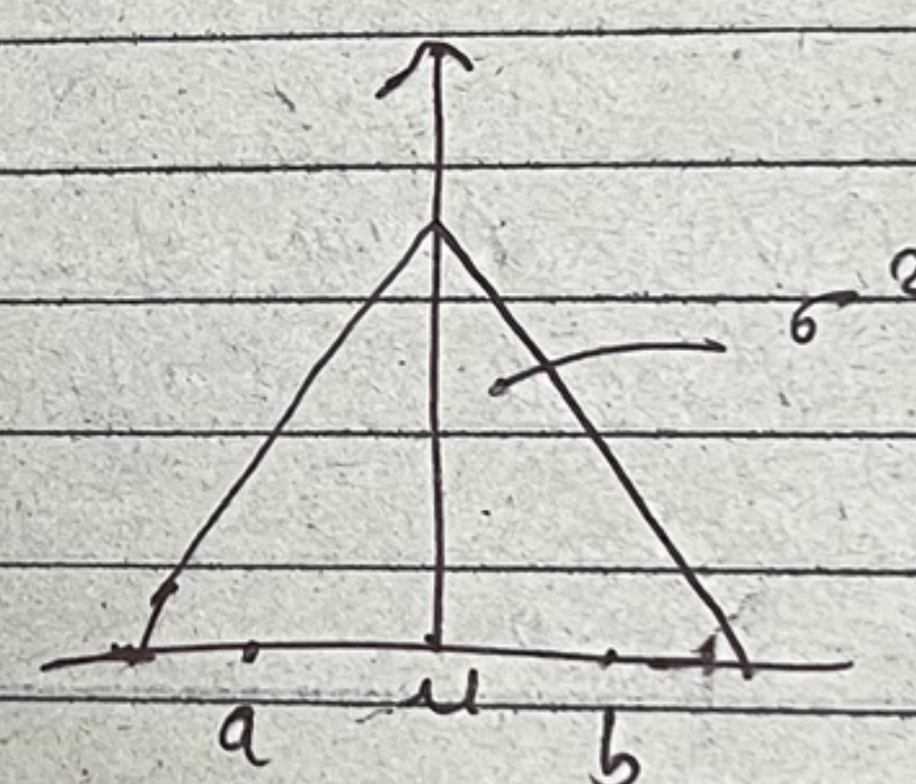
$\sigma^2(\bar{x} \pm \bar{y}) = \sigma^2(\bar{x}) + \sigma^2(\bar{y})$

sample size n_1 n_2

$= \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}$

$$\boxed{\sigma(\bar{x} \pm \bar{y}) = \sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$





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

Changing only scale & shift Shrink

(No change in shape, only scatteredness will change)

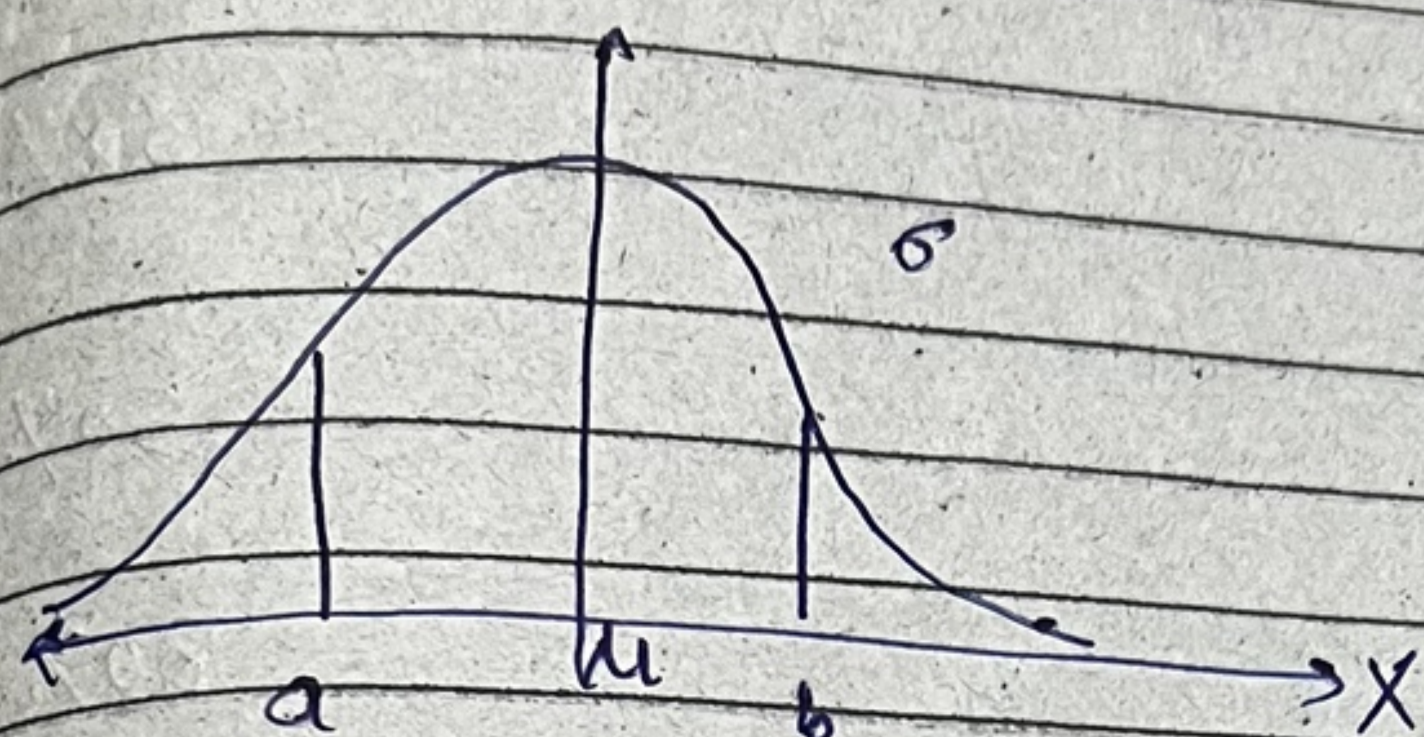
Normal नहीं तो पहले ⇒ $\frac{x-\mu}{\sigma}$ करके भी Normal नहीं होगा।

Can't use z-table if not Normal distn is not there

After standardization of this curves we can't use Z-table

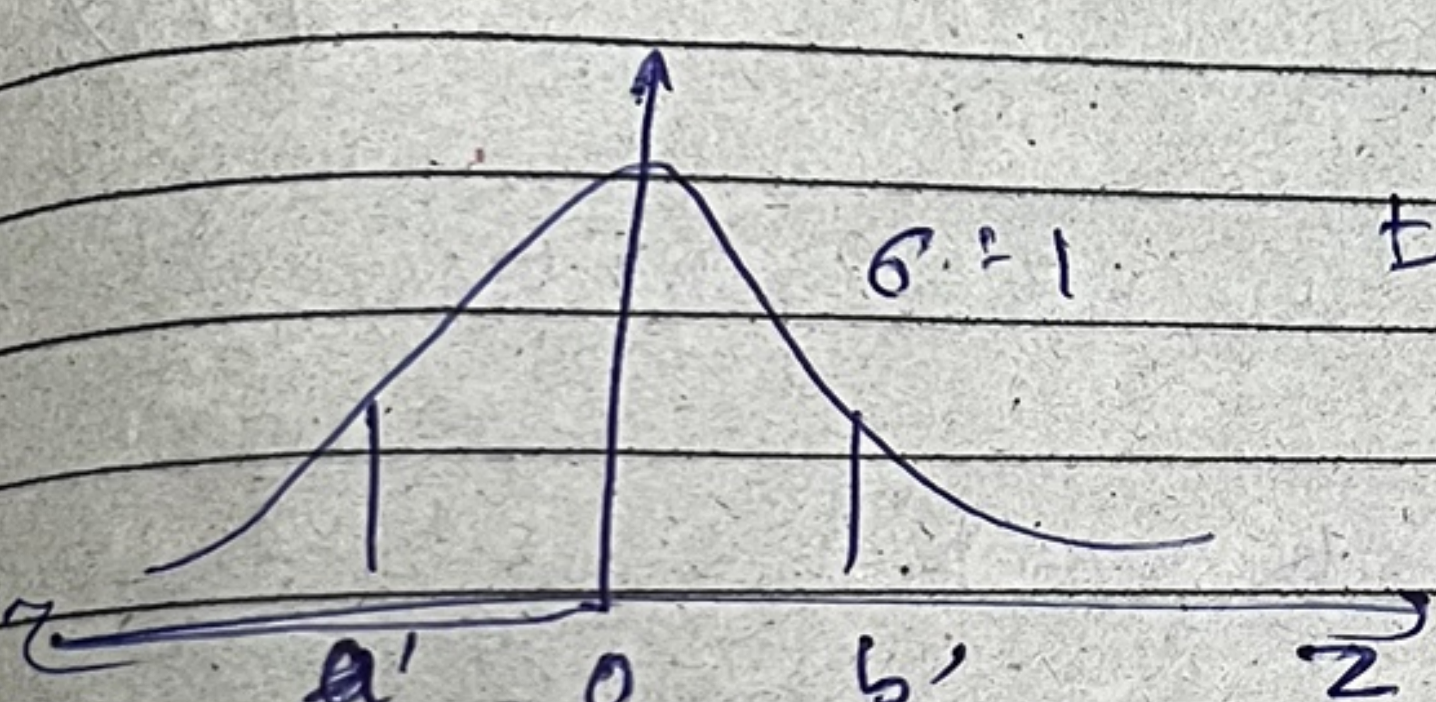
$\cancel{x} = z = \frac{x-\mu}{\sigma}$

(Not a Normal distribution)



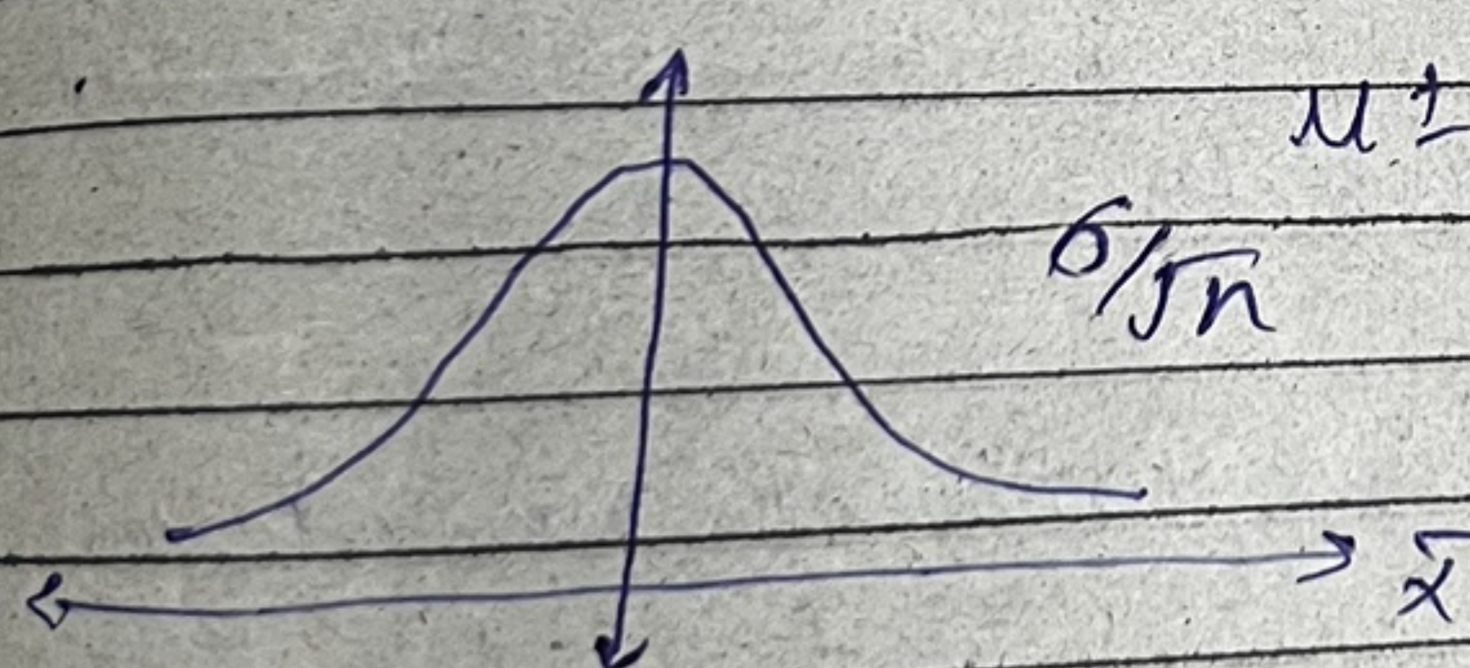
$\mu \pm k\sigma$
↓
Standard pts.

Normal Dist^n of population



$\pm k$ → standard pts.

Standard Normal distribution of population



$\mu \pm \frac{k\sigma}{\sqrt{n}}$ → Standard points

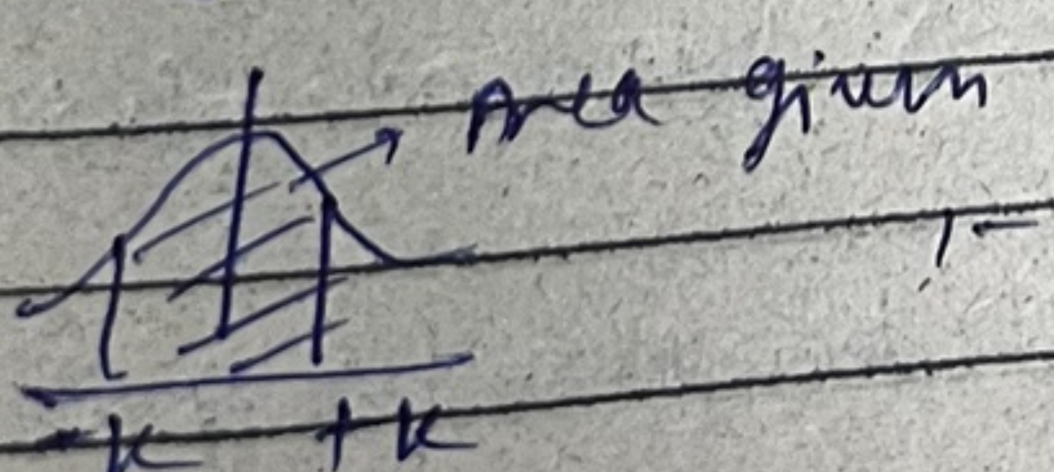
sampling mean Normal Dist^n.

For single pt., P (of finding pt) $\simeq 0$

find k ← 0.90

0.95

0.997



Sample size	No. of possible samples	No. of samples correctly estimated with margin 1	%age of correctly estimated
1	5	2	40%
2	10	3	30%
3	10	5	50%
4	5	4	80%
5	1	1	100%

Sampling error reduces as we increase

(I) If curve is not normal distribution

⇓

Then if take larger sample size (>30) & ~~more than 30~~ has more number of samples.

⇓

Then it will become Normal distⁿ (Sample mean)

(II) If curve is normal distribution

⇓

Then even if we take small size & less no. of samples

⇓

Then too Sample mean curve will be Normally Distributed

%age of correctly estimated

40%

30%

50%

80%

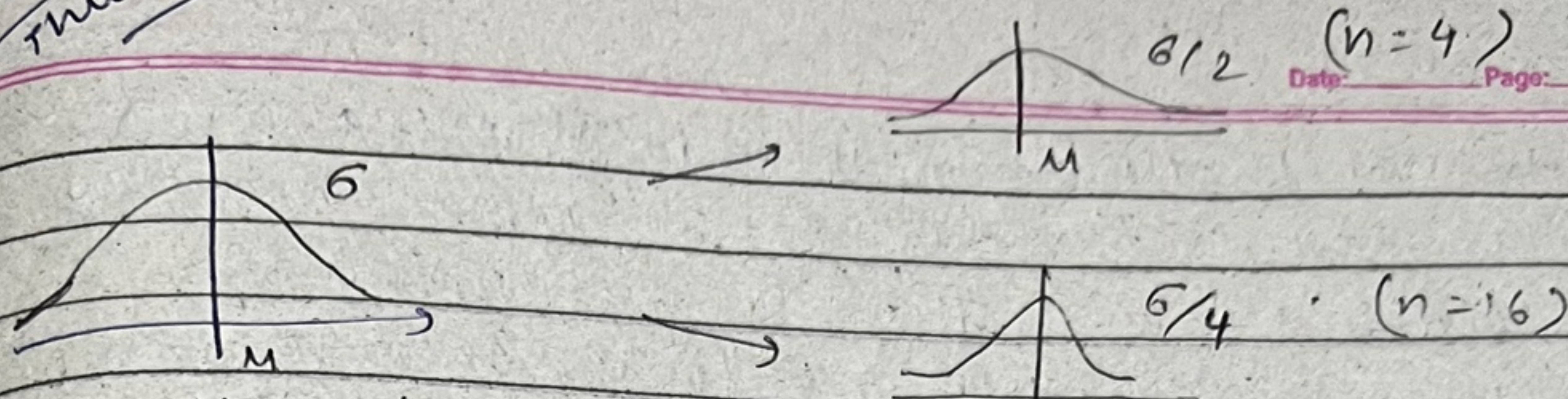
100%

~~n 30~~

mean)

distributed

14.09.2023 Thursday



sampling with replacement (better to observe)

N^n $\quad$ ${}^{N}C_n$

$n\uparrow$ $\quad$ $N\downarrow$

But if population $= \infty$

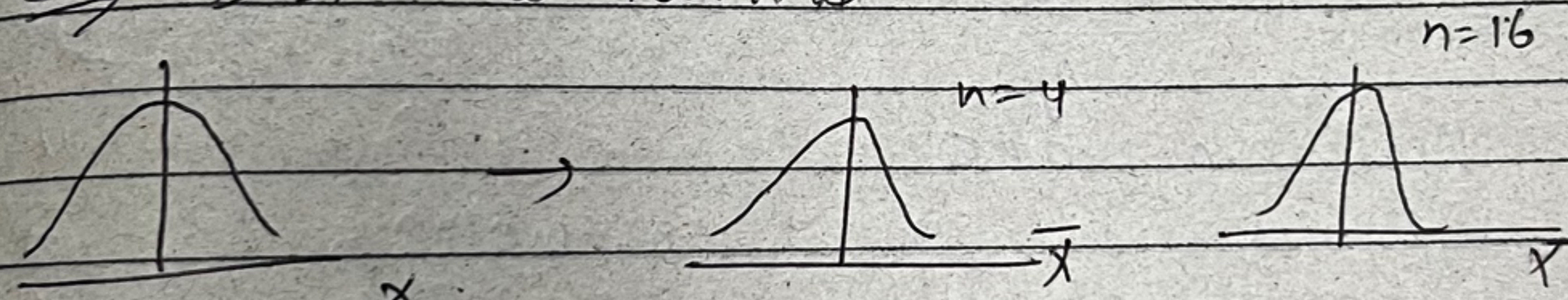
→ larger No. of samples → resulting

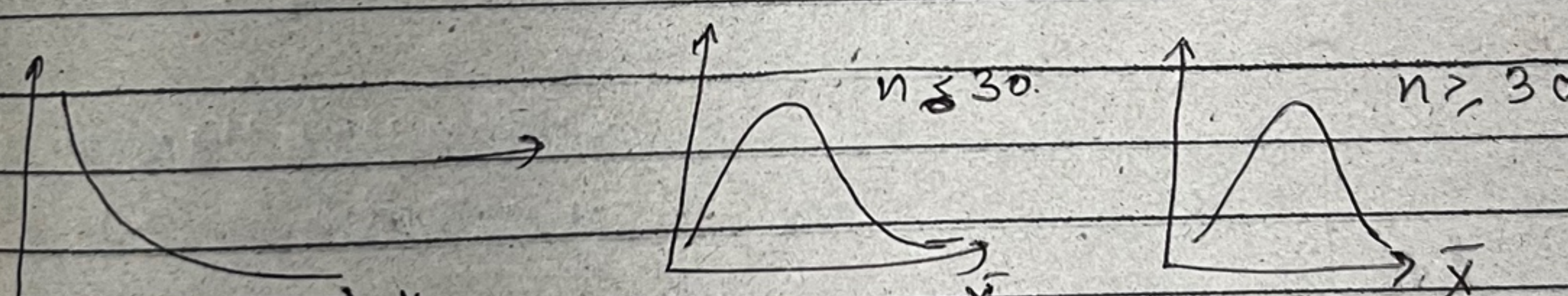
Central Limit Theorem

Pop^l $(x) \neq$ Normal && sample size ≥ 30

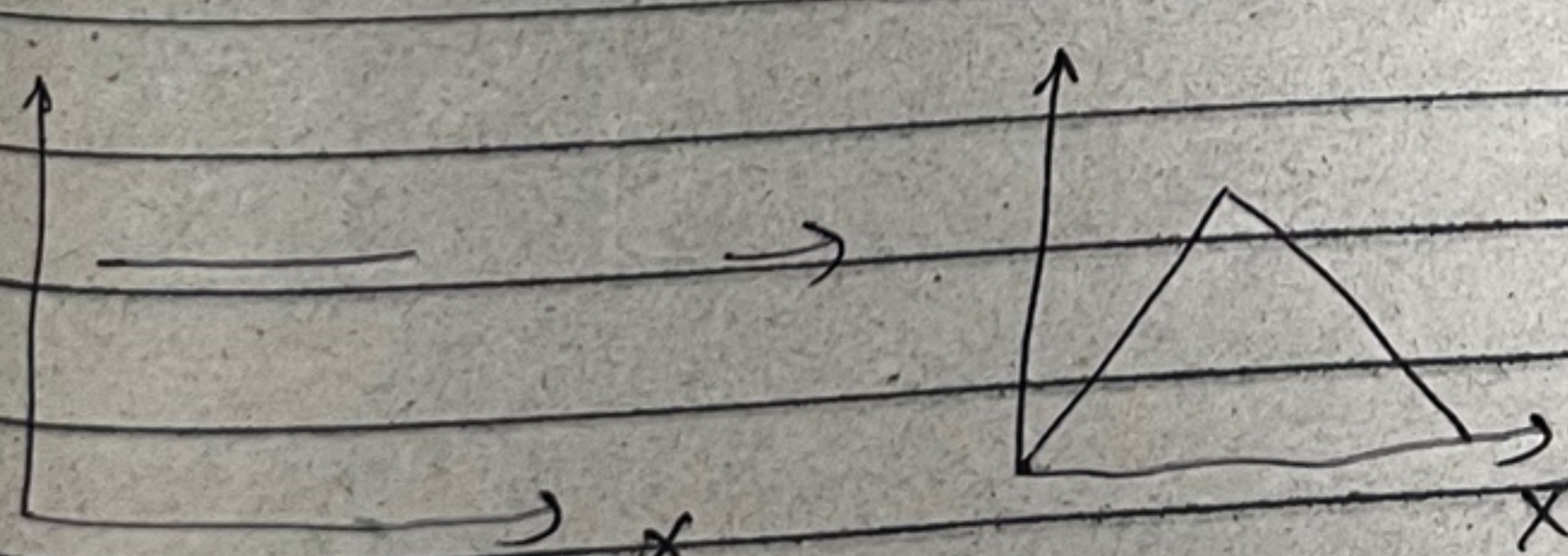
&& Taking samples as much as possible

$\Rightarrow$ Dist^r is Normal





exponential $\quad$ Right skewed $\quad$ Normal



(ii) $n \geq 30$ Apply z-table

(i) If Pop. is not Normal distr. → then can't use z-table

Date: Page:

Q) The no. of calories in the order of pizza in a certain fast food restaurant is approx. Normally distributed mean = 740 cal & SD = 20

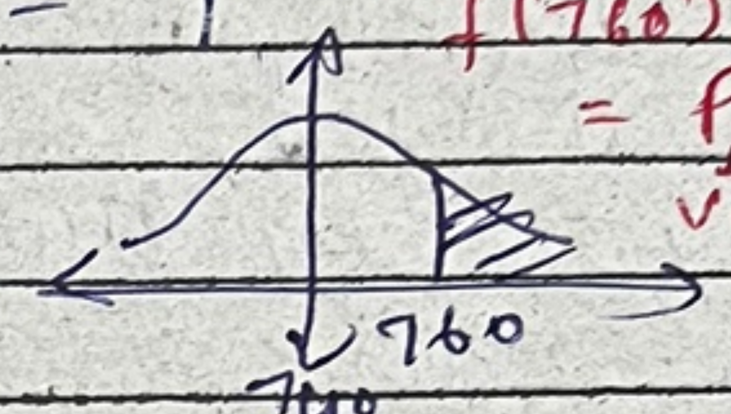
(i) What is the Prob. that random order of pizza has atleast 760 cal

Sol: (i) $z = \frac{x - \mu}{\sigma} = \frac{760 - 740}{20} = 1$

$f(760) = P(760)$ value

$x = 0.84134$

$P(i) = 0.15866$



(ii) What P(order of randomly selected 9 pizzas have atleast 760 cal on an avg.)

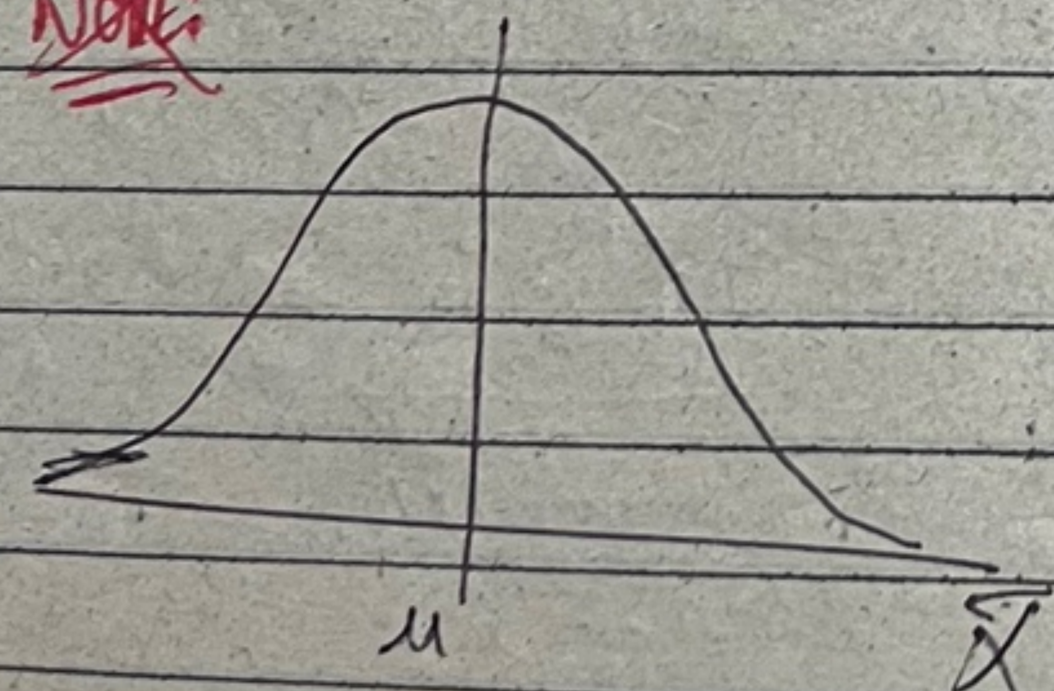
Sampling distribution $\mu = \mu$, $\sigma = \frac{\sigma}{\sqrt{n}}$

(ii) $z = \frac{x - \mu}{\sigma / \sqrt{9}} = \frac{20 \times 3}{20} = 3$

$P(z) = 0.99865$

$P(ii) = 1 - P(z) = 0.00135$

NOTE:



Expected value of mean of samples = μ

$\bar{x}_1$

$\bar{x}_2$ [μ]

$\bar{x}_3$

$\vdots$

$\bar{x}_N$

Two ways –

① ~~sampling error~~ → point /

~~Sampling~~ = Prob. of correct estimation

② Standard error = $\dfrac{\sum_{i=1}^{N} (\bar{x}_i - \mu)^2}{N}$

(Not known as std. deviation)

$X \rightarrow x_1, x_2, \ldots x_n \longrightarrow$ Population

$\bar{X} \Rightarrow X_1, X_2, \ldots X_n \longrightarrow$ Sample units

$dist^n(X_1) = dist^n(X)$
$dist^n(X_2) = dist^n(X)$
$\therefore\ dist^n(X_1) = dist^n(X_2)$

Independent & Identical

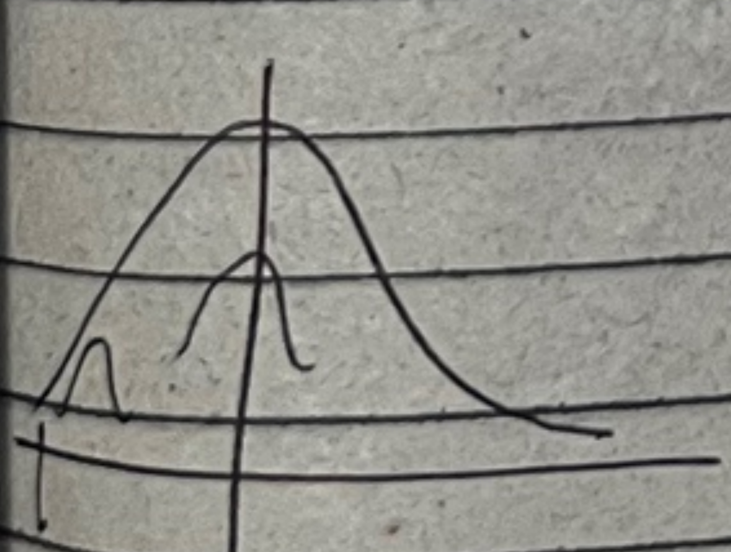
Sum of the Gaussian variable is Gaussian.
(Normal Dist^n)

" " Non " " is Gaussian for $(n > 30)$

Estimation of $\bar{X} = \mu$

$$S'^2 = \frac{1}{n-1}\left[\sum (x_i - \bar{x})^2\right]$$

$$S^2 = \frac{1}{n}\left[\sum (x_i - x)^2\right]$$



outlier

σ depends on [illegible] sample 30|2|1

15.09.2023
Friday

M.1

$E(s^2) = E\ \frac{1}{n}\left[\sum (x_i - \bar{x})^2\right]$

$= \frac{1}{n} E\left[\sum (x_i^2 + \bar{x}^2 - 2x_i\bar{x})\right]$

$= \frac{1}{n} \left(\sum x_i^2 + \sum \bar{x}^2 - 2\sum x_i \bar{x}\right)$

$= \frac{1}{n} \left(\sum x_i^2 + \cancel{\sum \bar{x}} n u^2 - 2u \sum x_i\right)$

$\sum x_i^2 - \sum$

$= \frac{1}{n}\sum\left(x_i^2 + (\bar{x})^2 - 2x_i\bar{x} + u^2 + 2x_iu - u^2 - 2x_iu\right)$

$= \frac{1}{n}\sum\left((x_i - u)^2 + (\bar{x})^2 - 2x_i\bar{x} + 2x_iu - u^2\right)$

$\frac{1}{n}\sum E\left[(x_i - u)^2\right] + \left\{(\bar{x})^2 - 2(\bar{x})^2 + 2\bar{x}u - u^2\right.$

$-(\bar{x})^2 - u^2 + 2\bar{x}u$

mean of sq deviation

$\frac{1}{n}\sum \sigma^2$ $\quad -E[\bar{x} - u]^2$ $\to$ variance in

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

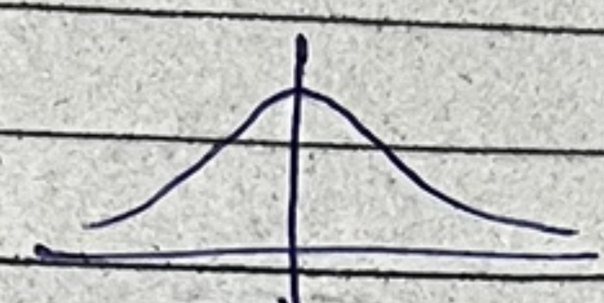
$\Rightarrow \boxed{E(S^2) = \left(\frac{n-1}{n}\right)\sigma^2}$

Date:______ Page:______

$$\boxed{s'^2 = \frac{n}{n-1} s^2}$$ from formula.

$$E[s'^2] = \frac{n}{n-1} E[s^2]$$
$$= \sigma^2$$

For class
$\mu = 70$ kg
SD $= 10$ kg



$\bar{x} = \left(\frac{x - \mu}{\sigma}\right)$

0

$P(wt = \mu) = 0$ (Nearer to zero)

P(randomly selected student will be in the range of 10 kg from mean) =

$\mu \pm \sigma k \rightarrow \pm k \quad f(\pm 1) \rightarrow 0.682$

M.2.
$$S^2 = \frac{1}{n} \Sigma [x_i - \bar{x}]^2$$

$$S^2 = \cancel{\frac{1}{n}} E[x_i^2] - E[\bar{x}]$$

Shifting

Date: Page:

Q) A coin is tossed 200 times. Find the approx. prob. that the ~~to~~ no. of heads obtained is betn ~~Φ~~. 80 to 120

Sol: M. ~~bi~~ $p = \frac{1}{2}$

$$P(x=80) + P(x=81) + \cdots P(x=120)$$

(Approx.)

M.2. BD → ND

mean $= np$ $\quad$ SD $= \sqrt{npq}$

$= 100$ $\quad$ $= \sqrt{50}$

$x_1 = \frac{80-100}{\sqrt{50}}$ $\quad$ $x_2 = \frac{120-100}{\sqrt{50}}$ $\quad$ $-2.'$

use table

$P(80<X<120) = 0.9952$

There is no sample.

∴ Don't use $\frac{6}{\sqrt{n}}$

Q) A random sample of size 100 is taken from a population whose $\mu = 60$ & variance is 400, Using Central Limit Theorem, with what prob. can we claim that the mean of the sample will not differ from $\mu = 60$ by more than 4.

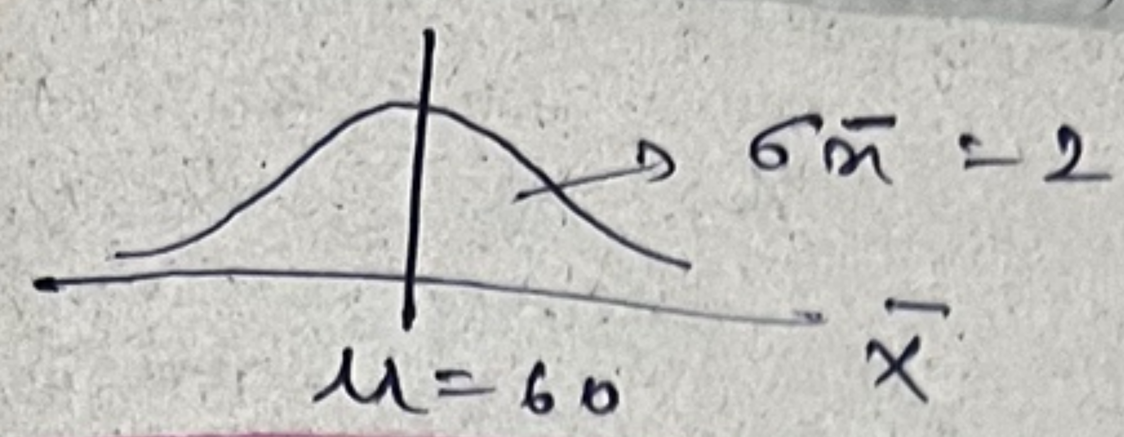
Sol: 100 sample $\quad$ Population

$\mu = 60$ $\quad$ $\mu = 60$ $\quad$ $6^2 = 400$

$6 = 20$

$6\bar{x}^2 = \frac{6^2}{n} = \frac{400}{100} = 4$

$6\bar{x} = \sqrt{4} = 2$



$|\bar{x} - \mu| \le 4$

$[\quad]$

$60 - 4 \qquad 60 + 4$

$\bar{x} \in (56, 64)$

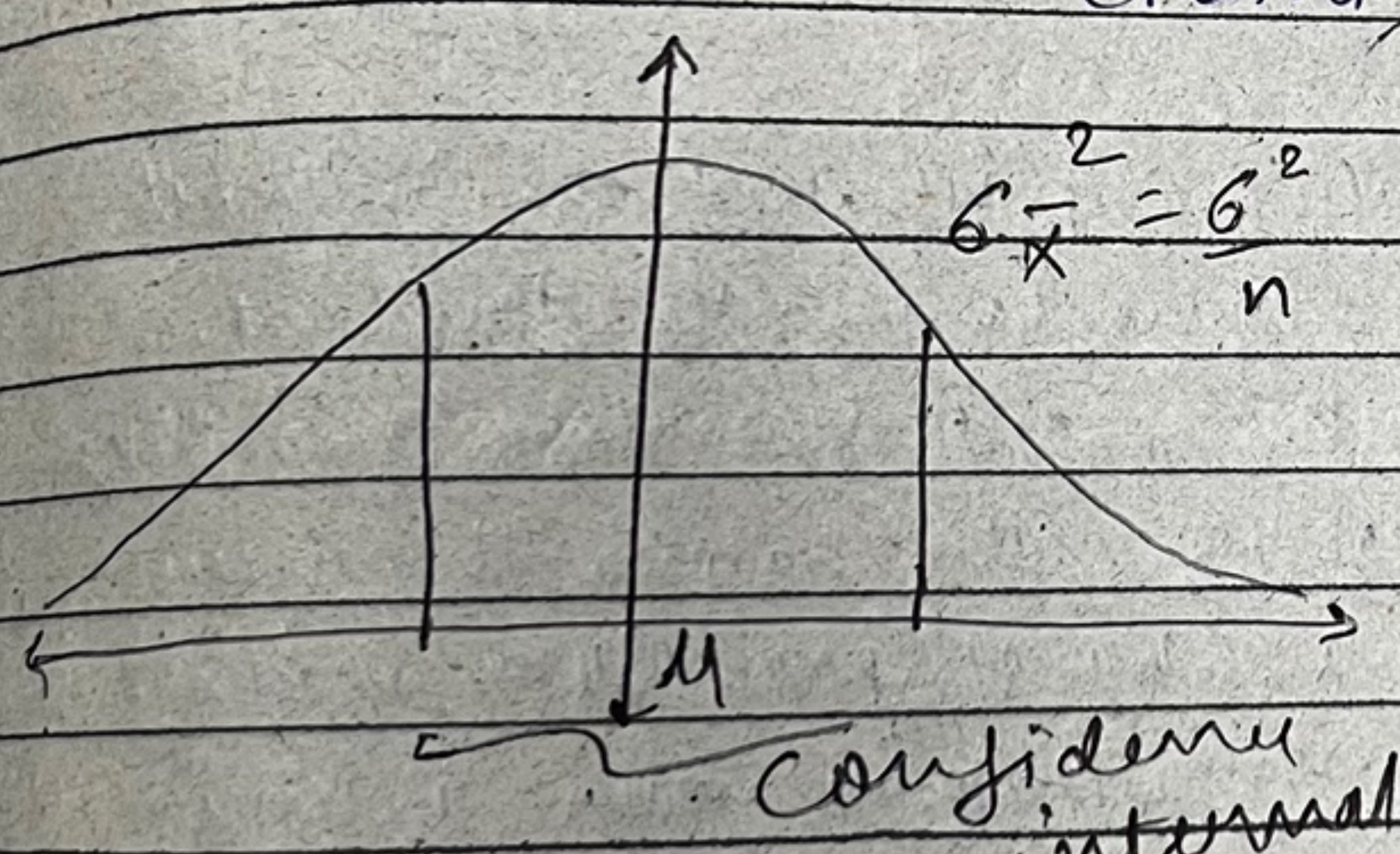
$P(56 \le \bar{x} \le 64)$ $\boxed{\cancel{z} = \frac{\bar{x} - \mu}{\sigma}}$

$P(-2 \le \bar{x} \le 2) = 0.954$

$\hookrightarrow$ 2P(Z ≤ 2) − 1

P(mean of the sample is μ) ≈ 0 (nearer to 0)

[Point estimation]



Interval Estimation

$\mu \pm$ Marginal value

$= \mu \pm k\sigma_{\bar{x}}$

~~(we are taking~~ marginal value as standard deviation)

$\mu \pm k\sigma_{\bar{x}}$

$\in [\mu]$

Confidence in finding estimating the particular value

If sample mean = Population mean

⇒ Exact Representative of popl.

K	Area
1.645	90%
2	95%
2.58	99%
3	99.9%

k = 1, 68%

90% confident that selected sample is representation

(Interval Estimation)

σ known — $\mu \pm K\,\sigma/\sqrt{n}$

σ known & n not known — $\mu \pm K\,\sigma/\sqrt{n}$ $(n \geq 30)$

σ unknown — $\mu \pm K\,\frac{S}{\sqrt{n}}$ → popl ... $\frac{1}{n-1}\Sigma(x_i - \bar{x})^2$

$\sigma_{\bar{X}}$

Q) National centre for health statistics published the birth rate of newly born male babies. According to the document, birth wt. of 81 male babies have std. deviation of 1.33kg. Determine % all samples of 400 male babies that have mean birth rate within 0.125kg of the population mean birth rate of all male babies. Interpret your answer in terms of sampling error.

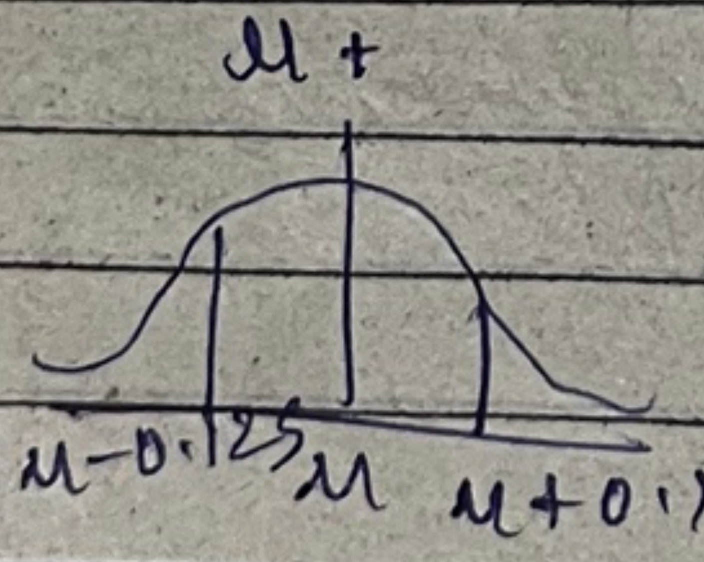
Sol: $\sigma = 1.33$kg   margin = 0.125 kg

$n = 400$

~~Range~~ = $\sigma' = \sigma/\sqrt{n} = \frac{1.33}{\sqrt{400}} = 0.0665$

$z_a = K = 1.879$ $z_b = K = -1.879$

Area = 94% (~93.98%)



$a = \mu - 0.125$ $b = \mu + 0.125$

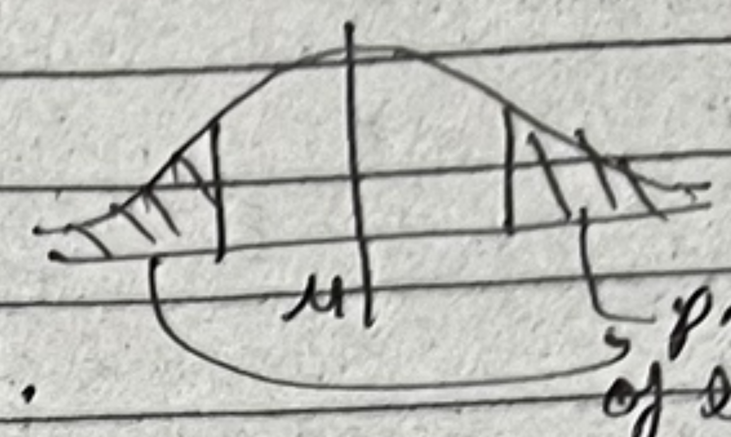
$z_a = \frac{x-\mu}{\sigma/\sqrt{n}} = \frac{-0.125}{0.0665}$

gives standard error

→ Two endpoints of confidence interval are critical value

→ After critical value area will be critical area

$$\mu - k\frac{\sigma}{\sqrt{n}} <= \mu <= \mu + k\frac{\sigma}{\sqrt{n}}$$



Interval should be minimized but prob. should be maximized.
↳ Area under the curve

Reduce the interval size Δ reduce the area of error (error estimation) reduces Δ
Thus increase confidence level (area b/w critical intervals)

(Need to find interval of μ)

Subtract $\bar{x}$ and add μ

$$-\bar{x} - k\frac{\sigma}{\sqrt{n}} \le -\mu \le -\bar{x} + k\frac{\sigma}{\sqrt{n}}$$

$$\Rightarrow \bar{x} - k\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + k\frac{\sigma}{\sqrt{n}}$$

[Multiply by -1]

Interval of μ

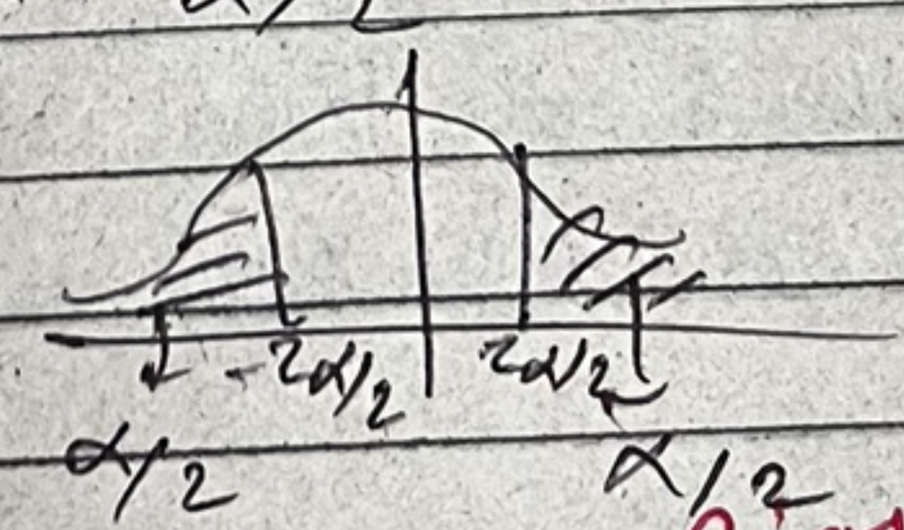
k value can be calculated {Sign changes}

confidence interval

$k = z_{\alpha/2}$ = Critical points of z-scale when critical area of one side $\alpha/2$ & other $\alpha/2$

Critical Area = α

Confidence Level = 1-α



p value of = right tail area = $\frac{\alpha}{2}$

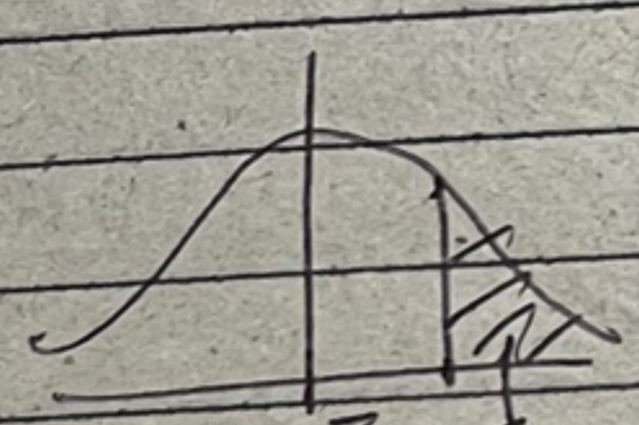
2nd (of $z_{\alpha/2}$ =) or prob. of after it

Always calculated for Right tail

If the confidence level is 90% value of $k = 1.645$

Confidence level	K value
90%	1.64
95%	1.96
99%	2.58

Pop.^{n} is Normally distr.
Std. dev. Pop. known
$k = z_{\alpha/2}$



z_α error

Q) Past experience with labourers in a certain industry indicates that the time required for randomly selected labourer to complete a job is approx. normal with the std. dev. 2 hrs. A random sample of 35 labourers indicated the mean time required to complete the job was 6.5 hrs. Find out 95% & 90% confidence interval for μ.

Date: Page:

Q1: $\sigma = 2$ hrs $n = 35$ $\bar{x} = 6.5$ hrs.

$\bar{x} - k\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + k\frac{\sigma}{\sqrt{n}}$

95% $k = 1.96$

$5.837 \le \mu \le 7.162$

90% $k = 1.64$

$5.94 \le \mu \le 7.054$

Popl is normally distributed

~~forms~~ sample size of size > 30

Then ~~σ (sample) =~~ σ. (Population) = σ (Sample)

S (sample)

Q) $X_1, X_2, \ldots\ldots, X_N$ are independent (iidrv) identically distributed random variables with mean = 3 & variance = 1/2

Use central limit theorem to estimate $P(340 \le S_n \le 370)$

Sol: $S_n = X_1 + X_2 + \ldots X_n$ and $n = 120$

$x_1 = 340$ $x_2 = 370$

S_n = sum of iidrv

$E(X_1 + X_2 + \ldots X_n) = n\,E(X_i)$

$E(S_n) = 3 \times 120 = 360$

$Var(S_n) = var(X_1 + X_2 + \ldots X_n)$

$= n \times Var(X_1) = 120 \times \frac{1}{2} = 60$

$\sigma = \sqrt{60}$ $\mu = 360$ $(-2.58 \le z \le 1.29)$

$x_1 = 340$ $x_2 = 370$

$z_1 = \frac{-20}{\sqrt{60}} = -2.58$ $z_2 = \frac{10}{\sqrt{60}} = 1.29$

0.90147
−0.00494
0.89653

89.65%

drv
crv
Sampling
Estimation

X_1, X_2

iid rv ; with $\lambda =$
↳ poisson