CHAPTER # 05

"DISCRETE PROBABILITY DISTRIBUTION"

(1) BINOMIAL / BERNOULLI DISTRIBUTION:

▶ Conditions:

- There should be finite no. of trials. (like 2 coins & tossed, 5 coins tossed etc.
- Trials are independent. (like 2 head aane ki probability mein ek head aajye to doosra dependent nhi hai [agr cards ka ~~replacement~~ replacement wala case hai ~~or~~ to wo dependent event hai waha Bernoulli nhi lagega
- Each trial result in an outcome that may be success (p) or failure (q). [agr head aane ki probability chahiye to head aayega (p) ya nhi ayega (q).] [agr p diya hai to q nikal sakta hai ($p = 1-q$ or $q = 1-p$)
- Finite no. of trials must be small. (like 10 coin tossed, 20 coin tossed. 1000 coin tossed will not be good because calculation will be greater.
- Probability of success remains same from trial to trial. (jese two coin tossed mein 2 tail ki probability chahiye to ek ~~to~~ dono mein tail ki same hi rhegi)

- If p & q have near about same values & there are finite small countable trials then best is to use Bernoulli.

$$P(X=x) = b(x;n,p) = {}^nC_x\, p^x q^{n-x} \quad , x = 0,1,2,\ldots$$

where,

n = no. of trials

x = Random variable

p = Prob of Success

q = Prob of Failure

$$\boxed{\sum_{x=0}^{n} b(x;n,p) = 1}$$

Mean $(\mu) = np$

Variance $\sigma^2 = npq$

St. deviate $\sigma = \sqrt{npq}$

$p = q = 1/2$ called Symmetric B.D (like probability of getting head or tail)

Unbaised coin is tossed 10 times. Find the probability ~~using B.D.~~ (i) of Exactly 6 heads (ii) Atleast 6 heads (iii) Atmost 6 heads (iv) Atleast 3 heads.

$n = 10$ $\quad p = 1/2$ (prob of head) $\quad q = 1/2$ (prob of tail)

$X \rightarrow$ no. of hed $\rightarrow 0, 1, 2 — 10$

$P(X=6) = {}^{10}C_6 (1/2)^6 (1/2)^4 = 210/1024$

$P(X \geq 6) = P(X=6) + P(X=7) + P(X=8) + P(X=9) + P(X=10)$

$= {}^{10}C_6 (1/2)^{10} + {}^{10}C_7 (1/2)^{10} + {}^{10}C_8 (1/2)^{10} + {}^{10}C_9 (1/2)^{10} + {}^{10}C_{10} (1/2)^{10}$

p & q dono powers add kardiya

$P(X \geq 6) = 386/1024$

$P(X \leq 6) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=5) + P(X=6)$

$= 210/1024$ or $P(X<6) + P(X \geq 6) = 1 \Rightarrow P(X<6) = 1 - P(X \geq 6)$

$386/1024$

P(

$P(X \leq 6) = P(X<6) + P(X=6)$

$= [1 - P(X \geq 6)] + \frac{210}{}$

(iv) $P(X \geq 1) = P(1) + P(2) - + P(10)$ or $1 - P(0)$

$= 1 - \left({}^{10}C_0 (1/2)^0 (1/2)^{10}\right)$ Ans

Q: The probability that a certain kind of a ~~test~~ component survive a shock test is 3/4. Find the prob that exactly 2 of the next 4 components tested survive

→ Yaha Component survive karega (sucess) ya nhi karega (fail)

$n = 4, \; x = 2, \; p = 3/4, \; q = 1/4$

$b(2; 4, 3/4) = {}^4C_2 (3/4)^2 (1/4)^2 = 27/128$

Q: The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (a) at least 10 survive, (b) from 3 to 8 survive & (c) exactly 5 survive?

$n = 15, \; p = 0.4, \; q = 0.6$

(i) $P(X \geq 10) = P(10) + P(11) + P(12) + P(13) + P(14) + P(15)$

$= {}^{15}C_{10} (0.4)^{10} (0.6)^5 + {}^{15}C_{11} (0.4)^{11} (0.6)^4 + {}^{15}C_{12} (0.4)^{12} (0.6)^3 +$

${}^{15}C_{13} (0.4)^{13} (0.6)^2 + {}^{15}C_{14} (0.4)^{14} (0.6)^1 + {}^{15}C_{15} (0.4)^{15}$

$P(X \geq 10) = 0.338$

(ii) $P(3 \leq X \leq 8) = \sum_{x=3}^{8} b(x; 15, 0.4) = P(3) + P(4) + P(5) + P(6) + P(7) + P(8)$

$= 0.8779$

(iii) $P(X = 5) = b(5; 15, 0.4) = 0.1859$

② MULTINOMIAL DISTRIBUTION:

- Each trial in an experiment has more than 2 outcomes
- The probability each trial remains constant
- The outcomes are independent
- Events must also be mutually exclusive.

eg:
52 playing cards, rolling a dice, a survey might require the responses of "approve", "disapprove" or "no opinion".

$$P(X=x_1, X=x_2, \dots X=x_k) = \frac{n!}{x_1!\, x_2! \dots x_k!} \times P_1^{x_1} P_2^{x_2} \dots P_k^{x_k}$$

$\Rightarrow \sum_{i=1}^{k} x_i = n$ & $\sum_{i=1}^{k} p_i = 1$

→ Mean $= \mu = np_i$

→ Variance $= \sigma^2 = np_i q_i$

Q: The painted light bulbs produced by a company are 50% red, 30% blue & 20% green. In a sample of 5 bulbs, find the probability that 2 red, 1 is green & 2 are blue.

$$P(x_1, x_2, x_3 ; P_1, P_2, P_3, n) = \frac{n!}{x_1!\, x_2!\, x_3!} \cdot P_1^{x_1} \times P_2^{x_2} P_3^{x_3}$$

($x_1 = 2$, $x_2 = 1$, $x_3 = 2$, $P_1 = 0.5$, $P_2 = 0.5$, $P_3 = 0.2$, $n = 5$)

$$= \frac{5}{2! \times 1! \times 2!} \; 0.5^2 \times 0.3^1 \times 0.2^5$$

$= 0.09$

$p(x) = P(X=x) = h(x; N, n, k) = \dfrac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$

X → represents no. of successes

3) HYPERGEOMETRIC DISTRIBUTION:

→ It is some what similar to binomial.

→ ~~The~~ Hypergeometric distribution doesnot req independence & is based on sampling done without replacment.

→ A random sample of size n is selected without replacement from N items. [like 5 cards are drawn from $\underset{N}{52}$]

→ Of the N items, k may be classified as sucess & $N-k$ are classified as failures.

Q: Probability of choosing 2 black & 3 red cards from deck of 52.

$$= \frac{{}^{26}C_3 \; {}^{26}C_2}{{}^{52}C_5} = 0.3251$$

Q: An box contain 6 red balls & 14 yellow balls. 5 balls are randomly drawn without replacement. What is the prob of exactly 4 red balls are drawn.

$N = 20$, $n = 5$, $k = 6$ (red balls) $x = 4$

$$h(4; 20, 5, 6) = \frac{{}^{6}C_4 \times {}^{14}C_1}{{}^{20}C_5}$$

ye bilkul wese hi hai jese hum normally nikaalte hai probability.

$= 0.01354$

⇒ Yaha dependent events hai ∴ we didn't apply binomial.

no. of objects sampled
no. of failures in the population
no. of objects sampled

$\max(0, n-(N-k)) \leq x \leq \min(n, k)$

→ Mean = $\mu = n \times k/N$

→ Variance = $\sigma^2 = \dfrac{N-n}{N-1} \cdot n \cdot \dfrac{k}{N}\left(1 - \dfrac{k}{N}\right)$

Q: Lots of 40 component are deemed unacceptable if they contain 3 or more defective. The procedure for sampling a lot is to select 5 components at random & to reject the lot if a defective is found. What is the prob the exactly 1 def is found in the sample if there are 3 defective in the entire lot.

$N = 40, \; n = 5, \; k = 3, \; x = 1$

$$P(X=1) = \frac{{}^{3}C_1 \, {}^{37}C_4}{{}^{40}C_5} = 0.3011$$

This plan is not desirable ∵ it detects a bad lot (3 defective) only about 30% of time.

Find the mean & variance & then use Chebyshev theorem to interpret the interval $\mu \pm 2\sigma$.

$\mu = (5)(3) \div 40 = 0.375$

$\sigma^2 = \left(\frac{40-5}{39}\right)(5)\left(\frac{3}{40}\right)\left(1 - \frac{3}{40}\right) = 0.3113$

$\sigma = 0.558$

The req interval is $0.375 \pm (2)(0.558)$
$(-0.741, 1.491)$.

" Theorem states that the no. of defectives obtained when 5 components are selected at random from lot of 40 comp of which 3 are def. has a prob of at least 3/4 of falling b/w -0.741 & 1.491. That is at least three-fourth of the time, the 5 components include fewer than 2 def. "

Q: Suppose a large high school has 1100 female students & 900 male. A random sample of 10 students is drawn. What is the probability exactly 7 of the selected students are female.

$N = 2000$, $n = 10$, $k = 1100$, $x = 7$

$$P(x=7) = \frac{{}^{1100}C_7 \; {}^{900}C_3}{{}^{2000}C_{10}} = 0.166490$$

" If we had used binomial distribution "

$$P(7) = {}^{10}C_7 (0.55)^7 (1-0.55)^3 = 0.166478$$

prob of female = $\frac{1100}{2000}$

Note:

The binomial se kabhi kabar answer tagreeban same aata hai. Ye jab hi possible hai jab sample hamare less than equal to 5% ho. In this case from 2000 student we choose 10 so $\frac{10}{2000} = 0.5\%$ { $\frac{n}{N} \leq 5\%$ then binomial ≈ hyper geometric }

(7) MULTIVARIATE HYPERGEOMETRIC DISTRIBUTION:

Q: A group of 10 individuals is used for biological case study. The group contains 3 people with blood type O, 4 with A, 3 with B. What is the prob that a random sample of 5 will contain 1 person with type O, 2 with A & 2 with B?

$$P(x_1, x_2, x_3; a_1, a_2, a_3, N, n) = \frac{\binom{a_1}{x_1}\binom{a_2}{x_2}\binom{a_3}{x_3}}{\binom{N}{n}}$$

$$P(1,2,2;3,4,3,10,5) = \frac{\overset{\text{typeO}}{{}^{3}C_1}\ \overset{\text{typeA}}{{}^{4}C_2}\ \overset{\text{typeB}}{{}^{3}C_2}}{{}^{10}C_5} = \frac{3}{14}$$

* Sir Amjad Slide:

Q: Ten people apply for a job as assistant manager of a restaurant. Five have completed college & five have not. If the manager selects 3 applicants at random, find the prob that all 3 are college students.

$N=10$, $n=3$, $k=5$, $x=3$

$$P(3;10,3,5) = \frac{{}^{5}C_3\ {}^{5}C_0}{{}^{10}C_3} = \frac{1}{12}$$

Q: $N = 10$, $n = 5$, $R = 2$, $x = 1$ (House Insurance)

$$P(1;10,5,2) = \frac{{}^{2}C_{1}\,{}^{8}C_{4}}{{}^{10}C_{5}} = \frac{5}{9}$$

Q: $N = 12$, $n = 3$, $R = 3$, $x = 1$ (Defective Compressor Tank)

N → total; n → # of sample; R → total no. of defective

x → rejected ki prob maangi hai or ag ek bhi defective aagya to reject hi hoga.

$$= \frac{{}^{3}C_{1}\,{}^{9}C_{2}}{{}^{12}C_{3}} = \frac{27}{55} \text{ Ans}$$

5) GEOMETRIC PROBABILITY DISTRIBUTION:

→ The geometric distribution is the distribution of the no. of trials needed to get the first sucess in repeated independent Bernoulli trials.

Condition:

All conditions are same as Bernoulli like independent trials, only two outcomes sucess (p) or failure (q), prob remains constant from trial to trials

Here X represents the no. of trials to get the first success.

→ For the first sucess to occur on x^{th} trial:

↳ The first $x-1$ trials must be failure

↳ The x^{th} trial must be a sucess.

$$\boxed{g(x;p) = p\,q^{x-1}} \quad , \; x = 1, 2, 3 \ldots$$

Q: In a large population of adults, 30% have recieved CPR training. If adults from this population are randomly selected, what is the probability that the 6th person sampled is the first that has recieved CPR training?

$$P(X=6) = (0.3)(0.7)^{6-1} = 0.0504$$

$$\boxed{\begin{aligned} &\rightarrow \text{Mean} = \mu = 1/p \\ &\rightarrow \text{Variance} = \frac{1-p}{p^2} \end{aligned}}$$

iska matlab 1st trial mein sucess
S

iska matlab 2nd trial mein sucess
FS
(0.7)(0.3)

Q: What is tha prob that the first person trained in CPR occurs on or befor the 3rd person sample.

$$P(X \leq 3) = P(1) + P(2) + P(3)$$

→ 3rd trial mein success
FFS
(0.7)(0.7)(0.3)

$$= (0.3)(0.7)^0 + (0.3)(0.7)^1 + \boxed{(0.3)(0.7)^2}$$

$$= 0.657$$

$\Rightarrow P(X > 3) = q^3 = (0.7)^3$

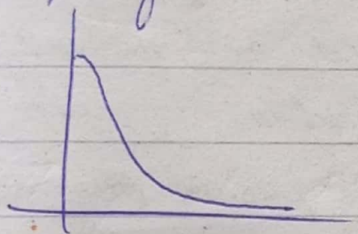
↳ means probability of getting 3 failures in a row.

Q: For a certain manufacturing process it is known that, on the average, 1 in every 100 item is defective. What is the probability that the fifth item inspected is the first defective item found.

→ Yaha humse 5 trial par 1^{st} sucess (means defective) maang rhe hai.

$$= (0.01)(0.99)^4 = 0.0096$$

Graph of Geometric Destribution.



(6) POISSON DISTRIBUTION:

Suppose we are counting the no. of occurence of an event in a given unit of time, distance, area or volume.

For example:

1. No. of car accidents in a day
2. No. of calls recieved by an office per hr.
3. No. of field mice per acre

etc.

Conditions:

1. Events are occuring independently
2. The probability that an event occurs in a given length of time doesnot change through time.

$$p(x;\lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \quad x = 0,1,2,3$$

where λ is the avg no. of outcomes per unit time, area or distance.

→ Mean = $\mu = \lambda t$

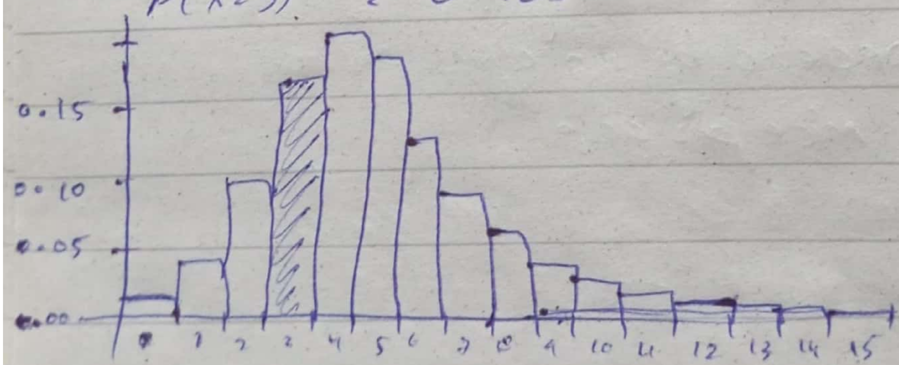
→ Variance = $\sigma^2 = \lambda t$

Q: One nanogram of Plutonium-239 will have on avg of 2.3 radioactive decays per second. and the no. of decays will follow a poisson distribution. What is the prob that in a 2 sec period there are exactly 3 radioactive decays.

→ Let X represents no. of decays in a 2 sec.

$$P(X=3) = \frac{e^{-(2.3)(2)}\left[(2.3)(2)\right]^3}{3!}$$

$$P(X=3) = 0.163$$



→ It is a right skewed but it depends on λt if its large it is close to symmetric when close to zero then its right skewed.

Prob of no more than 3 radioactive decays.

$P(X \leq 3) = P(0) + P(1) + P(2) + P(3)$

$$= \frac{4.6 \times e^{-4.6}}{0!} + \frac{4.6^1 e^{-4.6}}{1!} + \frac{4.6^2 e^{-4.6}}{2!} + \frac{4.6^3 e^{-4.6}}{3!}$$

$= 0.326$

Q: Ten is the avg no. of oil tankers arriving each day at a certain port. The facilities at the port can handle at most 15 tankers/day. What is the prob that on a given day tankers have to be turned away?

Let X be the no. of tankers arriving each day.

$$P(X > 15) = 1 - P(X \leq 15) = 1 - \sum_{x=0}^{15} p(x; 10) = 1 - 0.9513$$

$$= 0.0487$$

Note: The poisson distribution with $\lambda = np$ closely approximates the binomial distribution if n is large & p is small.

Q: In a manufacturing process where glass products are made, defects or bubbles occur, occasionally rendering the piece undesirable for marketing. It is known that on average, 1 in every 1000 of these items produce has one or more bubbles. What is the prob that a random sample of 8000 will yield fewer than 7 items posseing bubbles.

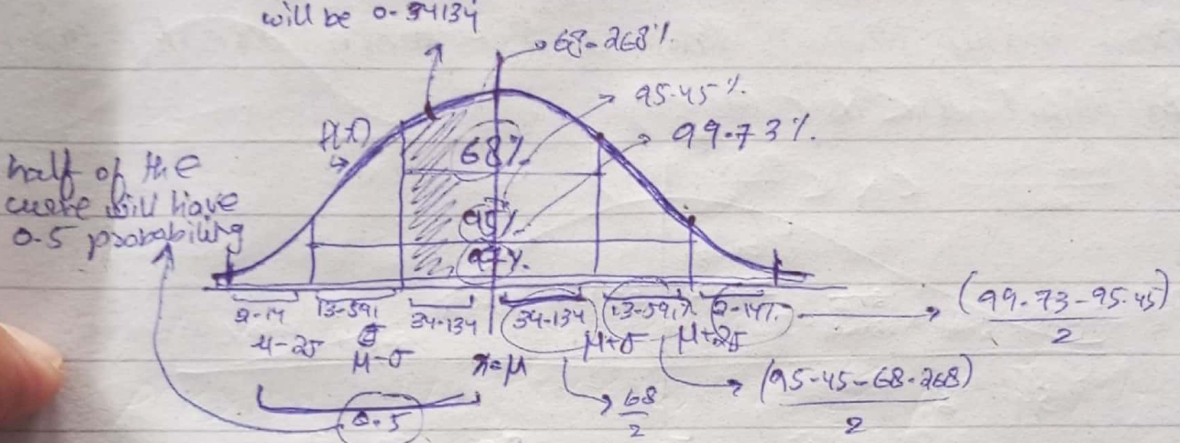
CHAPTER #06

CONTINUOUS PROBABILITY DISTRIBUTION

(1) NORMAL DISTRIBUTION : (Gaussian Distribution)

The PDF (probability Density fn) for normal distribution

$$P(X=x) = n(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty<x<\infty$$



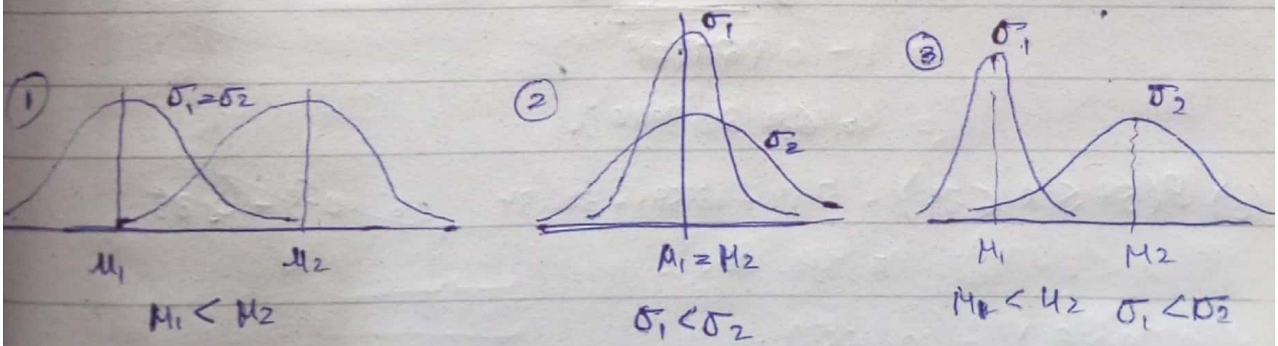
→ $y = f(x)$ is normal curve

→ Curve is bell shaped

→ $\int_a^b f(x)dx$ → Area under the curve

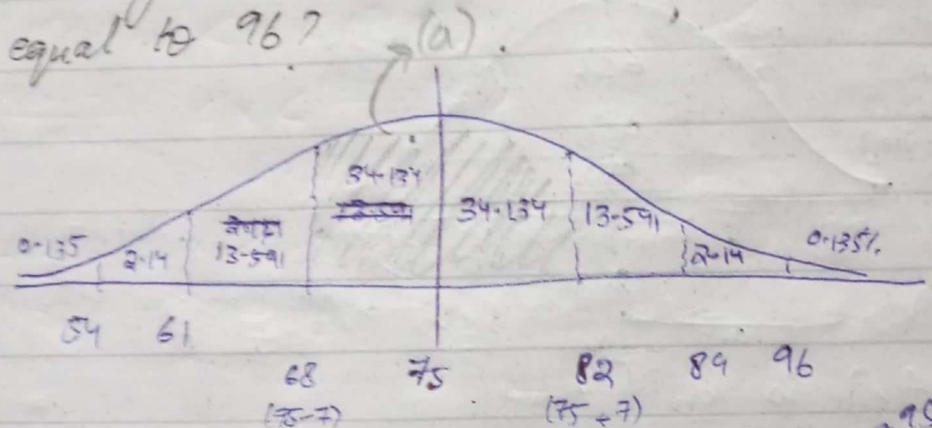
→ Mean = Median = Mode

→ The total area under the normal distribution curve is 100% or 1



Q: The test scores of a physics class with 800 students are distributed normally with a mean of 75 & a standard deviation of 7. (a) What percentage of class has a test score b/w 68 & 82? (b) Approximately how many students have a test score b/w 61 & 89? (c) What is the probability that a student chosen at random has a test score b/w 54 & 75. (d) Approximately how many students have a test score greater than or equal to 96?

Fig



95.45 × 800
= 764 Student

(a) $P(68 \le x \le 82) = 68.268$

(b) $P(61 \le x \le 89) = 13.591 + 34.134 + 34.134 + 13.591 = 95.45\%$

or $P(a \le x \le b) = \int_a^b \frac{1}{\sqrt{2\pi(7)^2}} e^{-(x-75)^2/2\times 7^2}$

$= 68.268949\%$

(c) $P(54 \le x \le 75) = 2.14 + 13.591 + 34.134$ or ~~0.5~~ 0.135.

$= 49.865\%$

(d) $P(x \ge 96) = 0.135\%$.

No. of student $= 0.00135 \times 800 \approx 1$

Standard Normal Distribution:

→ The standard normal distribution is a normal distribution with $\mu = 0$ & $\sigma = 1$

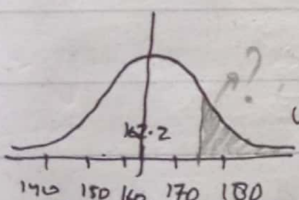
→ Suppose X is a normally distributed random variable with mean μ & st dev σ.

$$x = n(\mu, \sigma^2)$$

$$Z = \frac{X-\mu}{\sigma}$$ ⇒ Z is random variable that has a standard normal distribution.

$$z \sim n(0,1)$$

Q: Suppose the height of adult American female is appr. normally distributed with a mean of 162.2 centimeters. & a standard deviation of 6.8 centimeters.



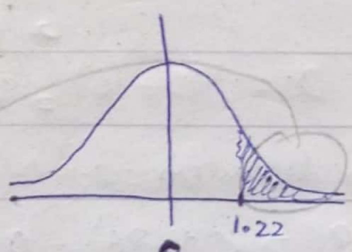
▲ the distribution of height

What is the probability that a randomly selected adult female is taller than 170.5 cm

$P(X > 170.5)$

isko standardized krne ke liye dono sides by minus μ & divide by σ krenge.

$$P\left(\frac{X-\mu}{\sigma} > \frac{170.5-\mu}{\sigma}\right) = P(Z > 1.22)$$
↳ Z



▲ Standard normal distribution

is area ki value we will find by the table. i.e 0.111

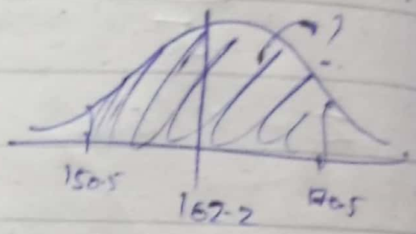
→ ab har "X" ki value ke liye hum corresponding Z find krenge by $Z = \frac{X-\mu}{\sigma}$ & phir us "Z" ki probability table se find krenge.

→ The "z" value tells us that how many standard dev a value is above or below it's mean.
↳ So this z value of "1.22" means that 170.5 is 1.22 above the mean 162.22

Q: What is the probability that a randomly selected adult females has height b/w 150.5 cm & 170.5 cm

$P(150.5\,cm < x < 170.5)$

$P\left(\frac{150.5 - 162.2}{6.8} < x < \frac{170.5 - 162.2}{6.8}\right)$

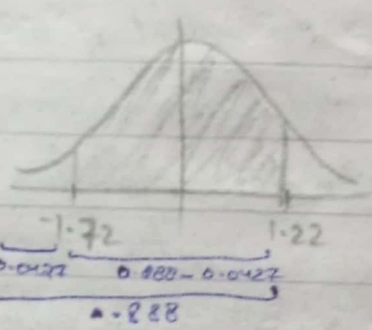


$P(-1.72 < x < 1.22) =$

$P(x < 1.22) = 0.888$

$P(x < -1.72) = 0.0427$

$P(-1.72 < x < 1.22) = 0.8461$



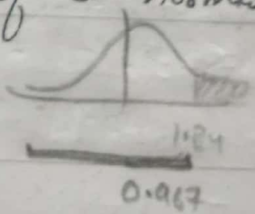
Q: Find area under the curve of st. normal distri

(a) to the right of $z = 1.84$.

$P(z > 1.84) = 1 - P(z \leq 1.84)$

$= 1 - 0.967$

$= 0.0329$

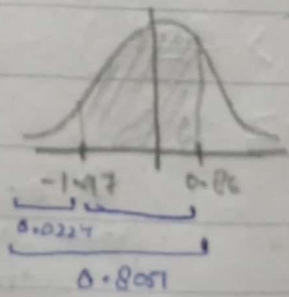


(b) between $z = -1.97$ & $z = 0.86$

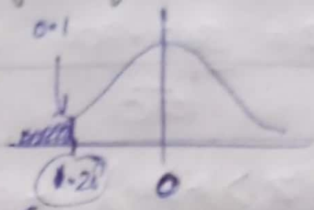
$P(z < 0.86) = 0.8051$

$P(z < -1.97) = 0.0224$

$P(-1.97 < x < 0.86) = 0.8051 - 0.0224 = 0.7807$



Q: What is the 70th percentile of the height of adult American female.



10th percentile means 0.1

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

$$x = 162.2 + 6.8(1.28)$$

$$\boxed{X = 153.5}$$

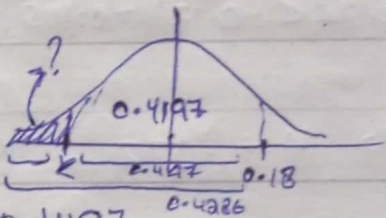
Q. Given a standard normal dis, find k.

(a) $P(Z > k) = 0.3015$ ye right ki hai for left we will $1 - 0.3015 = 0.6985$

now we find this in table

$k = 0.52$

(b) $P(k < Z < -0.18) = 0.4197$



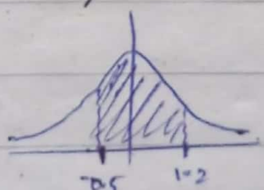
$P(Z < k) = x$

$P(Z < -0.18) = 0.4286$

$0.4286 - x = 0.4197$

$x = 0.0089$ → isko table se dekhenge we get -2.37

Q: Given a random variable X having normal dis with $\mu = 50$ & $\sigma = 10$. Find the prob that X assumes a value b/w 45 & 62.



$$Z_1 = \frac{45-50}{10} = -0.5$$

$$Z_2 = \frac{62-50}{10} = 1.2$$

$$P(-0.5 < X < 1.2) = 0.8849 - 0.3085 = 0.5764$$

CHAPTER # 09

"ESTIMATION"

→ For the purpose of estimating a population parameter we can use various sample statistics
① sample mean ($\bar{X}$) ② Sample median (M)
③ Sample variance (S^2) etc.
→ These are called $\boxed{\text{estimators}}$ and the actual value taken by the estimators are called estimate.

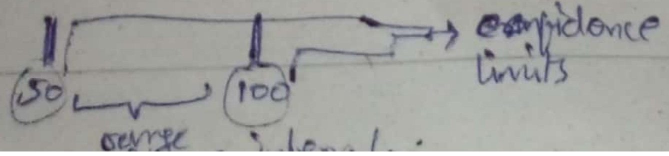
• POINT ESTIMATE:

→ Esi single value jo pure population ko represent kare.
Ex: For example the sample mean ($\bar{X}$) which we use for estimating the population mean μ is a point of estimator of μ. {ek population hai uska mean 30 aya phir mene us popu mein se random sample liya or uska mean 24 aya to ye 24, 30 ko represent krrta hai-

→ Kabhi bhi random sample ka mean, median, mode σ, σ^2 nikale se wo point estimate.

• INTERVAL ESTIMATE:

→ An interval estimate refers to the probable range within which the real value of a parameter is expected to lie.
→ The two extreme limits of such a range are called confidence limits & the range is called a confidence interval.



the interval $\theta_L < \theta < \theta_U$ computed from the selected sample is called $100(1-\alpha)\%$ ... degree of when $\alpha = 0.05 \Rightarrow 100(1-0.05) = 95\%$ confidence interval.

The wider the confidence interval is, the more confident we can be that the interval contains the unknown parameter.

→ It is better to be 95% confident the average life of a transistor is b/w 6 to 7 yr than to be 99% confident that it is b/w 3 & 10.

• PROPERITIES OF GOOD ESTIMATOR:

→ A good estimator is one which is as close to the true value of the parameter as possible.

• A good estimator posses the following properties.

i) UNBIASED ESTIMATOR:

→ An estimator $\hat{\theta}$ is said to be unbaised estimator of the population parameter θ if the mean of the sampling distribution is equal to the corresponding population parameter θ. [jese population ka mean = 50 & sample mean also 50 so it is unbaised estimator]

" A statistic $\hat{\Theta}$ is said to be unbaised estimator of the parameter θ if

$$\mu_{\hat{\Theta}} = E(\hat{\Theta}) = \theta$$

→ S^2 is an unbaised estimator of the parameter σ^2

• But S is a baised estimator of σ.

CONSISTENT ESTIMATOR:–

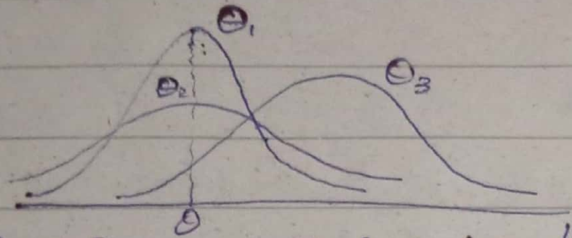
[J]ese jese sample size increase ho estimator ki value [co]mes closed to the population parameter.

$$E(\bar{X}) \to \mu \quad \text{as } n \to \infty \quad (\text{for mean})$$

[E]FFICENT ESTIMATOR:

[Effi]ciency is a relative term. Efficiency of an estimator is [usua]lly defined by comparing it with another [estim]ator.

Let us take two unbaised estimator $\hat{\theta}_1$ & $\hat{\theta}_2$. The estimator $\hat{\theta}_1$ is called an efficent estimator of θ if the variance of $\hat{\theta}_1$ is less than the variance of $\hat{\theta}_2$ i.e Variance $(\hat{\theta}_1)$ < Variance $(\hat{\theta}_2)$.



▸ yaha θ_1 & θ_2 are unbaised bcz their distribution are centered at θ. θ_1 has less variance than θ_2 so it's most efficent.

iv) SUFFICENT ESTIMATOR:

Sufficent estimator utilizes all informations that the given sample can furnish about the population.

▸ CENTRAL LIMIT THEOREM:

If $\bar{X}$ is the mean of a random sample of size "n" taken from a population with mean μ & finite variance σ^2, then the limiting form of the distribution of

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

The formula is generally good for $n \geq 30$.

mean when σ is known: for a
1) The sample is a random sample. 2) Either $n \geq 30$ or the population is normally distributed if $n < 30$

"when σ^2 is known"

- SINGLE SAMPLE : ESTIMATING MEAN

If $\bar{x}$ is the mean of a random sample of size "n" from a population with known variance σ^2, a $100(1-\alpha)\%$ confidence interval for μ is given by

$$\hat{\theta}_L \leftarrow \boxed{\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}} \rightarrow \hat{\theta}_u$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to right

→ for samples selected from non-normal populations we cannot expect our degree of confidence to be accurate. However samples of size $n \geq 30$, with shape of distribution not to skewed, will give good results.

Q: The avg zinc conc. recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 grams/mililetr. Find the 95% & 99% confidence intervals for the mean zinc conc. in the river. Population standard deviation is 0.3

$n \geq 30$ ✓

: $2.6 - 1.96\frac{(0.3)}{\sqrt{36}} < \mu < 2.6 + 1.96\frac{(0.3)}{\sqrt{36}}$

$2.50 < \mu < 2.70$

→ $100(1-\alpha) = 95$

$(1-\alpha) = 0.95$ $\alpha/2 = 0.025$ → $z_{\alpha/2} \cdot z_{0.025} = 1.96$

$(1-\alpha) = 0.99$
$\alpha = 0.01$
$Z_{\alpha/2} = Z_{0.005} = 2.575$

99% : $2.6 - (2.575)\frac{(0.3)}{\sqrt{36}} < \mu < 2.6 + (2.575)\frac{(0.3)}{\sqrt{36}}$

$2.47 < \mu < 2.73$

→ "If $\bar{x}$ is used as an estimate of μ we can be $100(1-\alpha)\%$ confident that the error will not exceed $Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$"

→ "If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error will not exceed a specified amount 'e' when the sample siz is

$$\boxed{n = \left(\frac{Z_{\alpha/2}\,\sigma}{e}\right)^2}$$

↳ This formula only applicable if we new variance of the population from which we select our sample.

Q: How large a sample is req if we want to be 95% conf that our estimate of μ in ~~ex~~ prev Qs is off by less than 0.05?

$\sigma = 0.3$, $n = \left(\frac{(1.96)(0.3)}{0.05}\right)^2$

$\boxed{n = 138.3}$

"Therefore we can be 95% confident that a random sample of 139 will provide an estimate $\bar{x}$ differs from μ by an amount less than 0.05.

One-Side Confidence Bounds on μ:

If $\bar{x}$ is the mean of a random sample of size n from a population with variance σ^2, the one-sided $100(1-\alpha)\%$ confidence bounds for μ are,

upper one-sided bound:	$\bar{x} + (Z_\alpha \sigma) / \sqrt{n}$
lower one-sided bound:	$\bar{x} - (Z_\alpha \sigma) / \sqrt{n}$

Q: In a psychological testing experiment, 25 subj are selected randomly & their reaction time, in seconds, to a particular stimulus is measured. Past experience suggest that the variance in reaction times to these types of stimuli is $4 sec^2$ & that the distribution of reaction time is approx. normal. The avg time for the subject is 6.2 sec. Given an upper 95% bound for mean reaction time.

$$x + (Z_\alpha \sigma)/\sqrt{n} = 6.2 + (1.645)\sqrt{4}/\sqrt{25}$$

$$= 6.858 \text{ sec}$$

we are ~~confi~~ 95% confident that mean reaction time is less than 6.858 seconds.

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

ESTIMATING WHEN "σ^2" is "UNKNOWN":

→ If $\bar{x}$ & s are the mean & the std. dev. of a random sample from a normal population with unknown variance σ^2, a $100(1-\alpha)\%$ confident interval for μ is

$$\boxed{\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}}$$

no. of independent values in a data.

where $t_{\alpha/2}$ is the t-value with $v = n-1$ degrees of freedom, leaving an area of $\alpha/2$ to right.

→ Yaha hum standard normal dist. ki jaga t-distri- use krte hai.

△ Properties of t-distribution:

① It is bell-shaped ② It is symmetric about mean.

③ The mean, median mode are equal to 0 and are located at the center of the distr.

④ curve never touches x-axis.

⑤ $\sigma^2 > 1$

⑥ As the sample size incr., the t-dist. approaches to the normal distr.

One Sided Bound:

$$\boxed{\bar{x} + t_\alpha \frac{s}{\sqrt{n}} \quad \text{and} \quad \bar{x} - t_\alpha \frac{s}{\sqrt{n}}}$$

Q. The contents of seven similar containers of Sulphuric acid are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, & 9.6 liters. Find a 95% confidence interval for the mean of contents of all such container, assuming normal distr.

$$\bar{x} = \frac{9.8+10.2+10.4+9.8+10.0+10.2+9.6}{} = \frac{70}{10} = 10$$

$S = 0.283$ (calculated by us) $S = \frac{(X-\bar{X})^2}{n-1}$

$\alpha = 0.05$

$t_{\alpha/2} = 2.447$ for $\nu = 6 \rightarrow (7-1)$

$$10 - (2.447)\frac{(0.283)}{\sqrt{7}} < \mu < 10 + (2.447)\frac{(0.283)}{\sqrt{7}}$$

$$9.74 < \mu < 10.26$$

→ If normality cannot be assumed, σ is unknown and $n \geq 30$, "S" can replace σ and the confidence interval

$$\bar{x} \pm Z_{\alpha/2}\frac{S}{\sqrt{n}}$$

may be used. This is referred as large sample confidence interval.

Two Sample: Estimating Mean (σ_1^2 & σ_2^2 are known)

- If $\bar{x}_1$ & $\bar{x}_2$ are means of independent random samples of sizes n_1 & n_2 from population with known variances σ_1^2 & σ_2^2 respec, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by, ~~$\mu_1 = \mu_2$~~

$$\boxed{(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}}$$

Example # 9.10:

$n_1 = 50$, $n_2 = 75$

$\bar{x}_1 = 36$ miles/gallon $\bar{x}_2 = 42$ miles/gallon

$\alpha = 1 - 0.96 = 0.04 \Rightarrow z_{0.02} = 2.05$

$\sigma_1 = 6$ $\sigma_2 = 8$

$\mu_B - \mu_A = ?$

$$(42-36) - 2.05\sqrt{\frac{64}{75} + \frac{36}{50}} < \mu_B - \mu_A < (42-36) - 2.05\sqrt{\frac{64}{75} + \frac{36}{50}}$$

$$3.43 < \mu_B - \mu_A < 8.57$$

"σ_1^2 & σ_2^2 are Unkown but Equal"

→ If $\bar{x}_1$ & $\bar{x}_2$ are the means of independent random samples of size n_1 & n_2 respec., from approximately normal populations with unknown but equal variance, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\, S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where S_p is the pooled estimate of the population std. deviation,

- $$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$
- $t_{\alpha/2}$ is t-value with $\nu = n_1 + n_2 - 2$ degree of freedom

Example # 9.11 :

$\bar{x}_1 - \bar{x}_2 = 3.11 - 2.04 = 1.07$

$$S_p^2 = \frac{(11)(0.771^2) + (9)(0.448)^2}{12 + 10 - 2} = 0.417$$

$S_p = 0.646$

$\alpha = 0.1$

$t_{\alpha/2} = 1.725$

$\nu = n_1 + n_2 - 2 = 20$

$$1.07 - (1.725)(0.646)\sqrt{\frac{1}{12} + \frac{1}{10}} < \mu_1 - \mu_2 < 1.07 + (1.725)(0.646)\sqrt{\frac{1}{12} + \frac{1}{10}}$$

[illegible] $< \mu_1 - \mu_2 < 1.547$

" Variances Unknown but UNEQUAL VARIANCE "

$$\boxed{(\bar{x}_1-\bar{x}_2)-t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} < \mu_1-\mu_2 < (\bar{x}_1-\bar{x}_2)+t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

where $t_{\alpha/2}$ is the t-value with,

$$\boxed{v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1-1)]+[(s_2^2/n_2)^2/(n_2-1)]}}$$

↳ value must be round down to nearest integer

Example # 9.12 :

$\bar{x}_1 = 3.84$ $\quad\bar{x}_2 = 1.49$

$S_1 = 3.07$ $\quad S_2 = 0.80$

$n_1 = 15$ $\quad n_2 = 12$

$\alpha = 0.05$

$$v = \frac{(3.07^2/15) + (0.80^2/12)^2}{[(3.07^2/15)^2/14]+[(0.80^2/12)^2/11]} = 16.3 \approx 16$$

for $\mu_1 - \mu_2$,

$\bar{x}_1-\bar{x}_2 = 2.35$

$$2.35 - 2.120\sqrt{\frac{3.07^2}{15}+\frac{0.80^2}{12}} < \mu_1-\mu_2 < 2.35 - 2.120\sqrt{\frac{3.07^2}{15}+\frac{0.80^2}{12}}$$

$$\boxed{0.60 < \mu_1 - \mu_2 < 4.10}$$