Computer Science & DA

Probability and Statistics

SAMPLING THEORY AND DISTRIBUTION

Lecture No. 01

By- Dr. Puneet Sharma Sir

Recap of previous lecture

Topic Joint p.m.f and p. d. f

$P(X+Y=4, X-Y=2) = ?$

$P(X,Y) = P(3,1) = \frac{2}{36}$

if $f(x)$ is pdf then CDF $F(x) = \int_{-\infty}^{x} f(x)\,dx$

if $F(x)$ is cdf then pdf is $f(x) = \frac{d}{dx}F(x)$

$\boxed{|x| < a} \Rightarrow -a < x < a$

$\pm x < a$

$-x < a \qquad +x < a$

$x > -a \qquad x < a$

$-a < x < a$

(ii) $\boxed{|x| > a \Rightarrow x < -a \text{ or } x > a}$

Topics to be Covered

Topic Sampling theory (Basics)

[NAT]

#Q. The joint density function of x and y is given as:

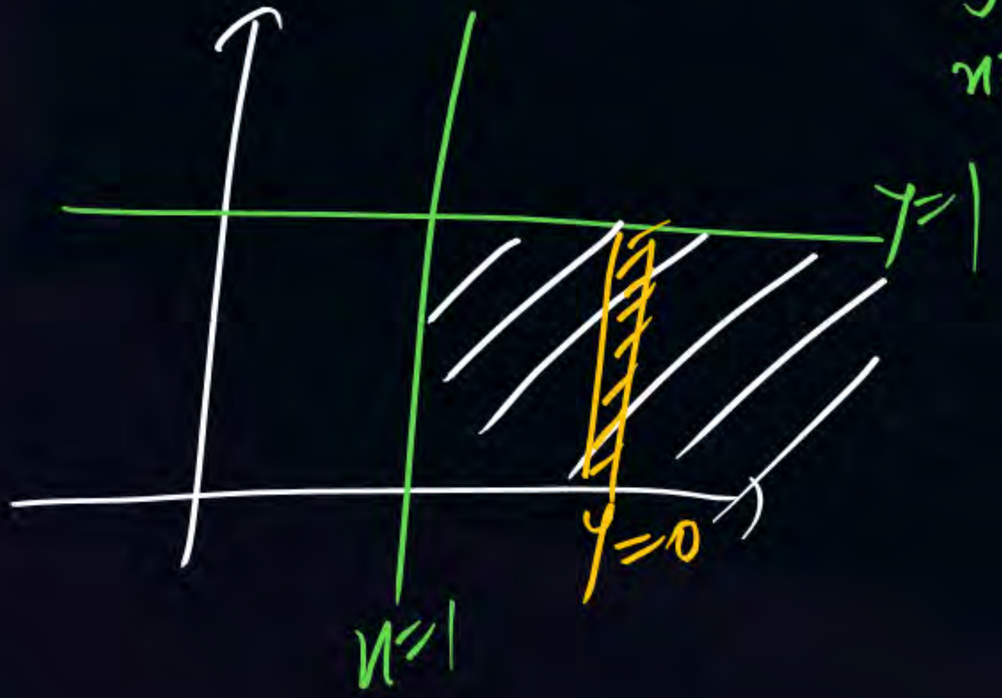
$$f(x,y)=\begin{cases}2e^{-x}\cdot e^{-2y} ; & 0<x<\infty, 0<y<\infty \\ 0 & ; \text{otherwise}\end{cases}$$

Find (i) $P(x > 1, y < 1)$

(ii) $P(x < y)$

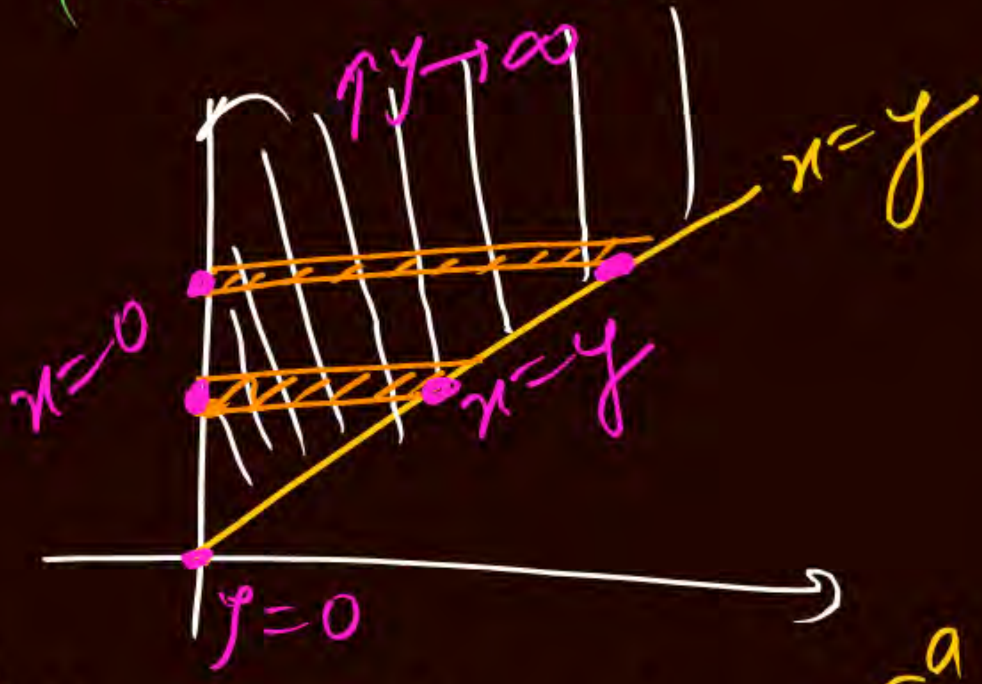
(iii) $P(x < a)$.

(i) $P(x>1, y<1) = ? = \int_{x=1}^{\infty}\int_{y=0}^{1} 2e^{-x}e^{-2y}\,dx\,dy = \frac{1}{e}-\frac{1}{e^3}$

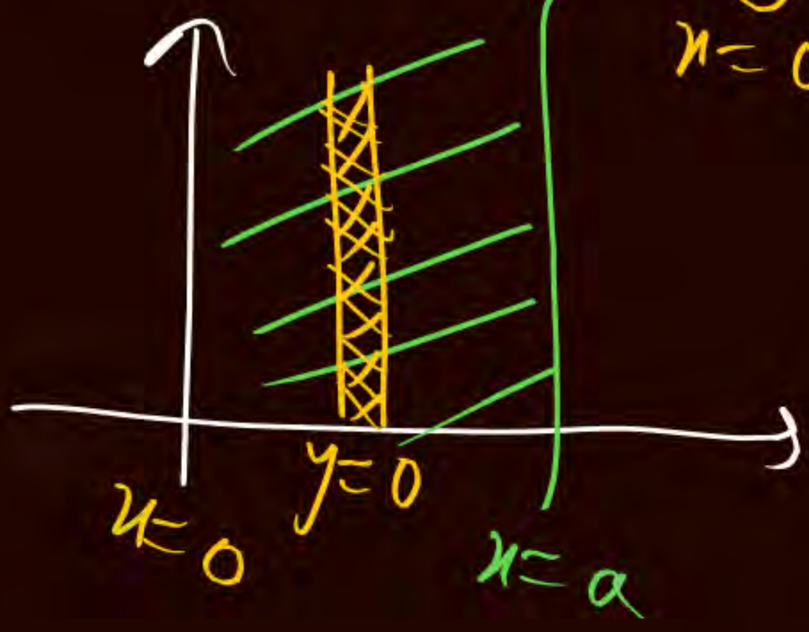


(ii) $P(x<y) = ? = \int_{y=0}^{\infty}\int_{x=0}^{y} \left(2e^{-x}e^{-2y}\right) dx\,dy = \int_{y=0}^{\infty}\left[\frac{e^{-x}}{-1}\right]_{x=0}^{y} 2e^{-2y}\,dy$

$= \int_{y=0}^{\infty}\left(1-e^{-y}\right)2e^{-2y}\,dy = 2\int_{0}^{\infty}\left(e^{-2y}-e^{-3y}\right)dy = \frac{1}{3}$



(iii) $P(x<a) = \int_{x=0}^{a}\int_{y=0}^{\infty}\left(2e^{-x}e^{-2y}\right)dx\,dy = 2\int_{x=0}^{a} e^{-x}\,dx \int_{y=0}^{\infty} e^{-2y}\,dy = 1-\frac{1}{e^{a}}$



[NAT]

joint c.d.f

#Q. What is the joint p.d.f of x and y if their $\boxed{\text{joint distribution function}}$ is given as

$$F(x,y)=\begin{cases}1-e^{-x}-e^{-y}+e^{-(x+y)} & ;x>0,y>0\\ 0 & ;\text{otherwise}\end{cases}$$

(i) Check the independencies of x and y

(ii) Find the marginal densities of x and y

(iii) Evaluate P $(x \le 1 \cap y \le 1)$.

(iv) Find $P(x+y \le 1)$

(v) Also find density function for ~~(x/y)~~? R.V $\left(\frac{x}{y}\right)$

Sol:
$$\boxed{f(x,y)=\frac{\partial^2}{\partial x \partial y}F(x,y)}$$

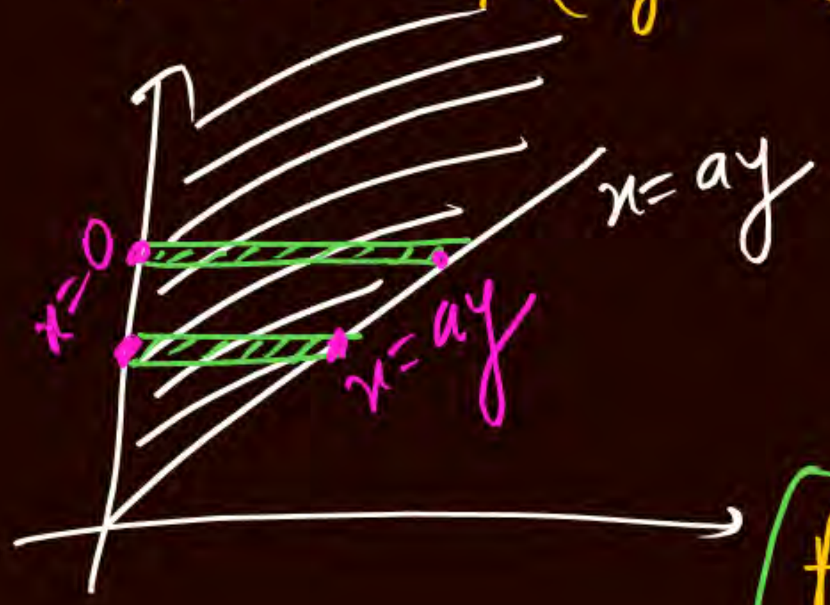
$$=\frac{\partial}{\partial x}\left[0-0+e^{-y}-e^{-x}\cdot e^{-y}\right]$$

$$=0-e^{-y}\left(-e^{-x}\right)$$

$$\boxed{f(x,y)=e^{-(x+y)}\quad ,x>0,y>0}$$

it is given that $\left(\frac{x}{y}\right)$ is also R.Variable so we will try to find it's C.D.F at 'a'

$$F(a) = P\left(\frac{x}{y} < a\right) = \int_{y=0}^{\infty}\int_{x=0}^{ay} \left(e^{-x}\cdot e^{-y}\right) dx\,dy = \int_{y=0}^{\infty} \left\{1 - e^{-ay}\right\} e^{-y}\,dy$$



$$= \int_{y=0}^{\infty} \left[e^{-y} - e^{-(a+1)y}\right] dy = \left[\frac{e^{-y}}{-1} + \frac{e^{-(a+1)y}}{a+1}\right]_0^{\infty} = \left[0 - \left\{-1 + \frac{1}{a+1}\right\}\right]$$

$$\boxed{f(a) = 1 - \frac{1}{a+1}}$$ it is the D.F for R.V $\frac{x}{y}$

So p.d.f for $\frac{x}{y}$ is $f\left(\frac{x}{y}\right) = \frac{d}{da}\left(1 - \frac{1}{a+1}\right) = \frac{1}{(a+1)^2}$, $a > 0$

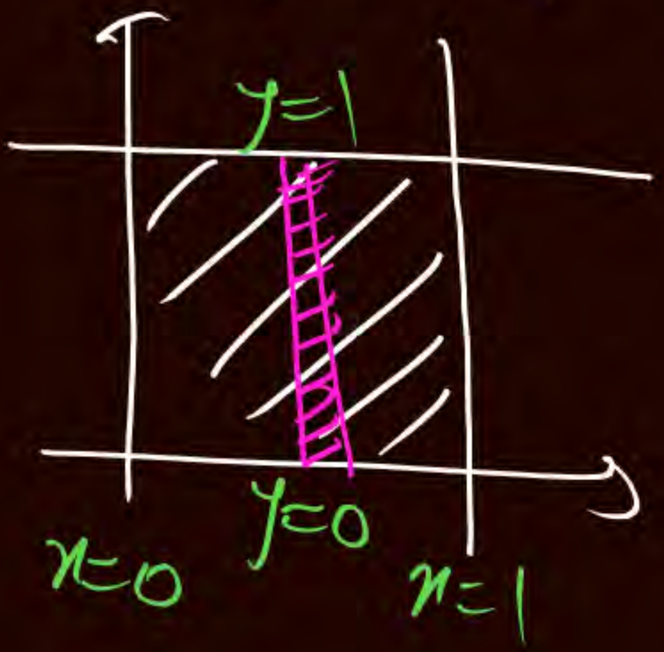
$f(x,y) = e^{-(x+y)}$; $x>0,\ y>0$

M.D.f, $f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_0^{\infty} e^{-(x+y)}\,dy = e^{-x}$, $x>0$

M.D.f, $f(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^{\infty} e^{-(x+y)}\,dx = e^{-y}$, $y>0$

$\because$ $f(x)\cdot f(y) = f(x,y)$ Hence x & y are Ind.

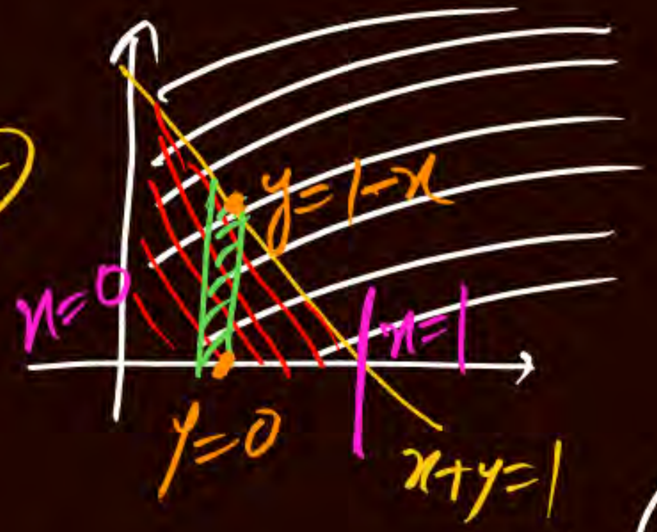
(*) $P(x \le 1 \cap y \le 1) = \int_{x=0}^{1}\int_{y=0}^{1} e^{-(x+y)}\,dy\,dx =$



$= \dots \_\_\_ \dots$

$= \left(1 - \frac{1}{e}\right)^2$

(*)



$P(x+y \le 1) = \int_{x=0}^{1}\int_{y=0}^{(1-x)} e^{-(x+y)}\,dy\,dx$

$= 1 - \frac{2}{e}$

#Q. If joint p.d.f is

$$f(x,y) = \begin{cases} \dfrac{e^{\frac{-x}{y}} \cdot e^{-y}}{y}, & 0 < x < \infty \text{ and } 0 < y < \infty \end{cases}$$

Then find E(x) for given y i.e. E(x/y = y) = ?

$E\{x\} = \int_{-\infty}^{\infty} x.f(x)\,dx$

New $f\left(\frac{x}{y}\right) = \frac{f(x,y)}{f(y)} = \frac{e^{-\frac{x}{y}}.e^{-y}}{y.e^{-y}} = \frac{1}{y}e^{-\frac{x}{y}}$

So $E\left\{x/y=y\right\} = \int_{-\infty}^{\infty} x.f\left(\frac{x}{y}\right)dx = \text{------} = y$

M.D.f for y is $f(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$

$= \int_0^{\infty} \frac{e^{-\frac{x}{y}}.e^{-y}}{y}\,dx = \text{......} = e^{-y}$

(M-II) $f\left(\frac{x}{y}\right) = \frac{1}{y}e^{-\frac{x}{y}}, \; y > 0$

$f(t) = \begin{cases} \mu e^{-\mu t}, & t > 0 \\ 0, & otw \end{cases}$

i.e Condn Dist of x for given y is nothing but E. Dist with mean $= \frac{1}{\mu} = \frac{1}{1/y} = y$

$E\left\{x/y=y\right\} = y$

CENTRAL LIMIT THEOREM (C.L.T)

Let x is R.V {Not Necessarily Normal} then taking some samples (each of size $n \geq 30$) from this population (μ, σ^2) as;

$S_1 = \{x_1, x_2, x_3 \ldots x_{30}\} \Rightarrow$ Mean $= \bar{x}_1$

$S_2 = \{y_1, y_2, y_3 \ldots y_{30}\} \Rightarrow$ Mean $= \bar{x}_2$

$S_3 = \{z_1, z_2, z_3 \ldots z_{30}\} \Rightarrow$ Mean $= \bar{x}_3$

$\vdots$

$S_{1000} = \{\alpha_1, \alpha_2, \alpha_3 \ldots \alpha_{30}\} \Rightarrow$ Mean $= \bar{x}_{1000}$

then $\dfrac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \ldots + \bar{x}_{1000}}{1000} \simeq \mu$

i.e. Average of Sample Means $\simeq$ Pop. Mean

Similarly Var of Sample Mean $= \dfrac{\text{Pop Var}}{\text{S. Size}} = \dfrac{\sigma^2}{n}$

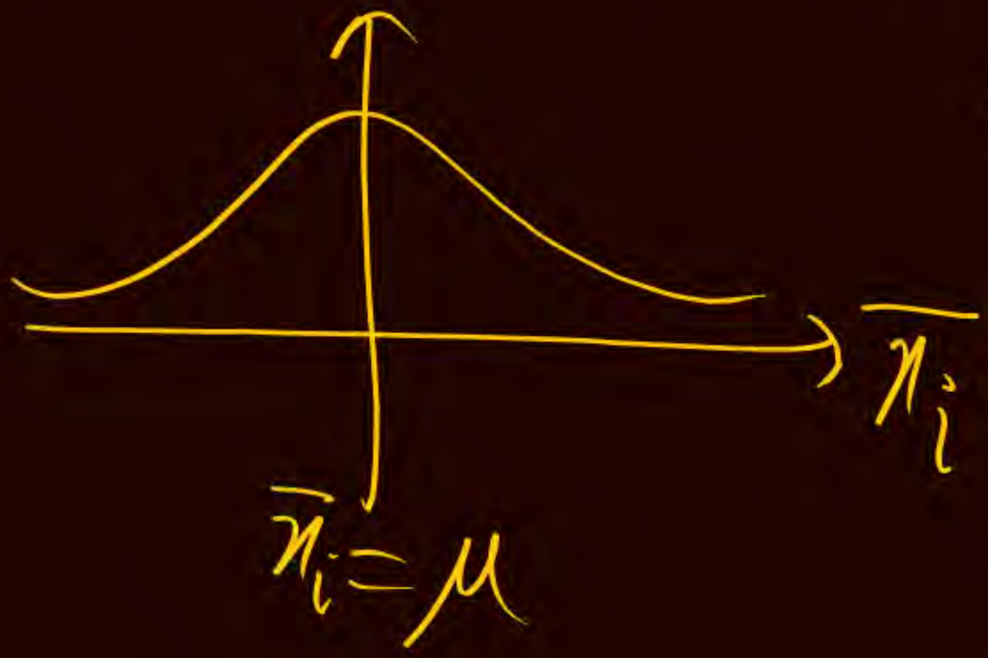
Here $\boxed{\text{New R.V. } \bar{x}_i \text{ is Normally Dist. with mean '}\mu\text{' \& SD} = \frac{\sigma}{\sqrt{n}} \text{ i.e. } \bar{x}_i \sim N\{\mu, \frac{\sigma^2}{n}\}}$

Conclusions :- let x is any R.V (Population) with Mean μ & Var $= \sigma^2$ then

① Av of Distribution formed Sample Means $\cong$ pop. Mean (μ)

② Var of " " " " $= \dfrac{\text{pop Var}}{\text{s. size}} = \left(\dfrac{\sigma^2}{n}\right)$

③ this New R.V $\overline{x}_i$ is also Normally Distributed i.e $\overline{x}_i \sim N\left\{\mu, \dfrac{\sigma^2}{n}\right\}$



N. Condn for C.L.T →

① Sample Members should be selected at Random.

② " " must be Ind.

③ Population Members should be identically Distributed

④ $30 \leq$ Sample size $<$ 10% of Population size

Note :- By C.L.T $\bar{x}_i \sim N\left\{\mu, \frac{\sigma^2}{n}\right\}$

Proof (i) $E\{\bar{x}_i\} = E\left\{\frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \dots + \bar{x}_n}{n}\right\} =$

$$= \frac{1}{n}\left\{E\{\bar{x}_1\} + E\{\bar{x}_2\} + \dots + E\{\bar{x}_n\}\right\}$$

$$= \frac{1}{n}\left[\mu + \mu + \mu + \dots + \mu\right]$$

$$= \frac{1}{n}\{n\mu\} = \mu$$

(ii) $Var\{\bar{x}_i\}$

$$= Var\left\{\frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \dots + \bar{x}_n}{n}\right\}$$

$$= \frac{1}{n^2}\left[Var(\bar{x}_1) + Var(\bar{x}_2) + \dots + Var(\bar{x}_n)\right]$$

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

$$= \frac{1}{n^2}\left[n\sigma^2\right] = \frac{\sigma^2}{n}$$

#Q. A set of samples have been collected from large sample and the sample mean values are 12.8, 10.9, 11.4, 14.2, 12.5, 13.6, 15, 9, 12.6. Find the population mean.

$\bar{x}_1$ $\bar{x}_2$ $\bar{x}_3$ $\bar{x}_9$

By C.L.Th, Pop Mean = Av of Sample Means

$$\mu = \frac{\sum \bar{x}_i}{n} = \frac{12.8+10.9+11.4+\cdots+9+12.6}{9} = 12.44$$

Qe C.L.T does not Hold in which option?

(a) Pop is Normal & $n < 30$

(b) " " & $n \geq 30$

(c) Pop is Not Normal & $n < 30$

(d) " " " & $n > 30$

(Learn)

#Q. In a survey of Lucknow city, it was reported that, Average age of mobile users is 30 years with standard deviation is 12. What is the mean and standard deviation of mobile users in a sample of 100 persons?

$\mu = 30$ yrs
$\sigma = 12$ yrs
$n = 100$ person

By C.L.Th, Sample Mean $(\bar{x}) \cong$ Pop Mean $(\mu) = 30$ yrs

& Sample Var $(\bar{x}) = \dfrac{\text{Pop Var}}{\text{S.size}} = \dfrac{\sigma^2}{n}$

$\sigma_{\bar{x}} = \dfrac{144}{100} = 1.44 \text{ yrs}^2$

So SD $(\bar{x}) = \sqrt{1.44} = 1.2$

#Q. (ii) Also find the probability that Average age of mobile users in a sample of 100 persons is less that 28 years? It is given that area under N curve from z = 0 to 1.6 is 0.4452.

Sol^n:- $\underbrace{\bar{x} = \mu}_{C.L.T} = 30$

$\sigma_{\bar{x}} = 1.2$

$\bar{x} = \{$Age of Mobile use in a sample of 100 persons$\}$

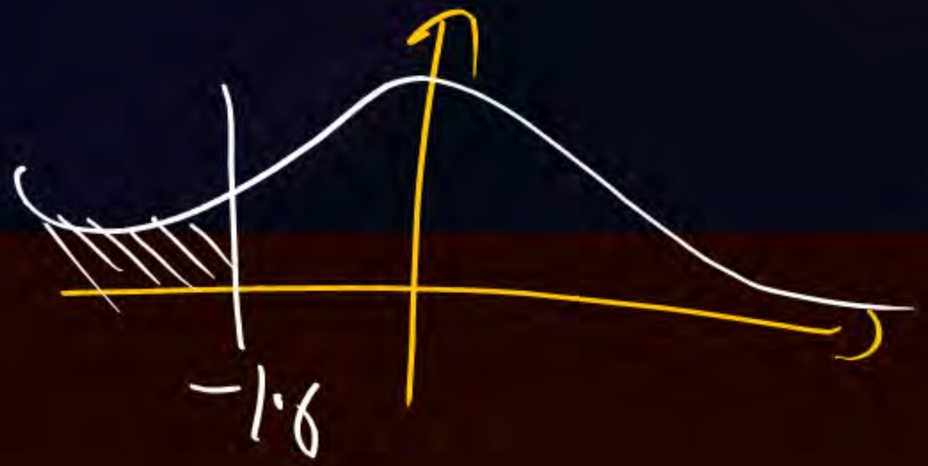
Here $\bar{x}$ is Normally Distributed by C.L.T

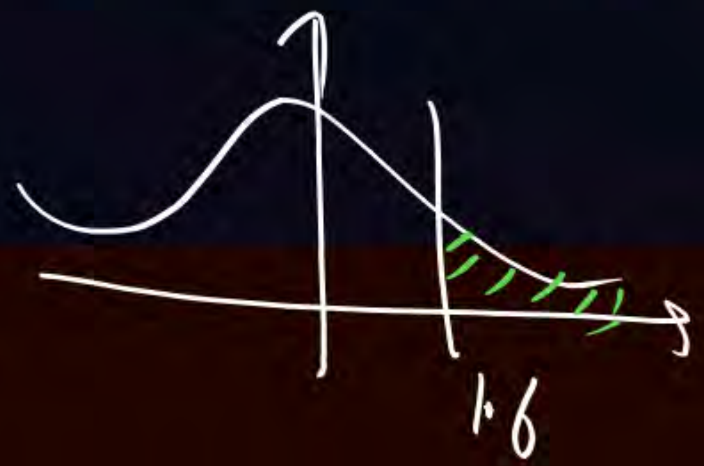
$$Z = \frac{x-\mu}{\sigma} = \frac{28-\bar{x}}{\sigma_{\bar{x}}} = \frac{28-30}{1.2} = -1.6$$

$$P[\bar{x} < 28] = P[z < -1.6] = P(z > 1.6) = \text{Right Half area} - P(0<z<1.6)$$

$$= 0.5 - 0.4452$$

$$= 0.0548 \approx 5.4\%$$





SAMPLING (BASICS)

Population → The group of individual under consideration (whether finite or ∞) is called pop.

Sample → A small set from population is called sample (it is always finite) & this process is called sampling.

Parameters → Numerical quantities from which we can understand population are called parameters

Statistic → " " " " " sample called statistic

for eg μ & σ are the parameters for population
while $\bar{x}$ & S " statistic " sample

with the help of statistic, we will try to understand pop also that's why sampling plays an imp role.

Proportion → The Ratio of Successful Events with Total Events known as proportion

for, eg A Coin is tossed 10 times and we are getting Head exactly 3 times then

$$\text{proportion of } H = ? = \frac{\text{Success}}{\text{Total}} = \frac{3}{10}$$

eg In a sample of 400 Children, There are exactly 210 Boys then

$$\text{Prop of Boys in Sample} = \frac{\text{Success}}{\text{Total}} = \frac{210}{400} = 0.525$$

, while prop of Boys in Population

$$= \frac{500\,Cr}{1000\,Cr} = 0.500$$

i.e $\text{prop in Sample} = \frac{x}{n}$ & $\text{prop in Population} = \frac{X}{N} \simeq \text{Probability}$

i.e $\text{Sample Prop } (\tilde{p}) = \frac{x}{n}$ & $\text{Population Prop} = \frac{X}{N} = p_0$

(*) **Standard Error** → it is the S.D of statistic (in sample)

SE of Mean = ?

$$\boxed{SE(\bar{x}) = \frac{\sigma}{\sqrt{n}}}$$

&

SE of proportion = ?

$$\boxed{SE(\tilde{p}) = \sqrt{\frac{\tilde{p}\tilde{q}}{n}}}$$

where $\tilde{p} = \frac{x}{n}$, $\tilde{q} = 1 - \tilde{p}$

Most probable limits for μ & p_0 →

$$\mu = \bar{x} \pm 3.SE(\bar{x}) \quad \& \quad p_0 = \tilde{p} \pm 3\,SE(\tilde{p})$$

$$\bar{x} - 3SE(\bar{x}) \leq \mu \leq \bar{x} + 3SE(\bar{x}) \quad \& \quad \tilde{p} - 3SE(\tilde{p}) \leq p_0 \leq \tilde{p} + 3SE(\tilde{p})$$

[NAT]

#Q. A sample of 400 members has mean 4.0. If the population is normal with standard deviation 2.6 and it's mean is unknown than find the most probable limits for population mean.

$n = 400$, $\bar{x} = 4$

$\sigma = 2.6$, $\mu = ?$

$$SE(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{2.6}{\sqrt{400}} = \frac{2.6}{20} = \frac{1.3}{10} = 0.13$$

i.e $3\,SE(\bar{x}) = 0.39$

i.e Most probable limits for μ is $\bar{x} - (0.39) \le \mu \le \bar{x} + (0.39)$

$$3.61 \le \mu \le 4.39$$

#Q. In a town, 350 out of 600 persons were found to be vegetarian on the basis of this date can we say that majority of population in the town in vegetarian?

$n = 600$, $x = \{\text{Number of Vegetarian persons}\} \rightarrow$ success

proportion of Veg person (in sample) $= \dfrac{\text{No of success}}{\text{S. Size}} = \dfrac{x}{n} = \dfrac{350}{600} = 0.5833$

i.e Sample prop $(\tilde{p}) = 0.5833$ & $\tilde{q} = 1 - 0.5833 = 0.4167$

S. Error of Sample prop $= SE(\tilde{p}) = \sqrt{\dfrac{\tilde{p}\tilde{q}}{n}} = \sqrt{\dfrac{0.5833 \times 0.4167}{600}} = 0.02$

So Most Probable limits for Population Prop $(p_0) = ?$ $\tilde{p} - 3SE(\tilde{p}) \leq p_0 \leq \tilde{p} + 3SE(\tilde{p})$

$$0.5833 - 0.0600 \le p_0 \le 0.5833 + 0.0600$$

$$0.5233 \le p_0 \le 0.6433$$

ie p_0 lies within 52% to 64%

ie population proportion of Vegetarian person lies b/w 52% & 64%.

Hence Conclusion is "Majority of population in a town is Vegetarian"

[NAT]

#Q. A coin was tossed 400 times and head turned up 210 times. Discuss whether coin is unbiased or not.

THANK - YOU