Statistical Techniques

Probability:

$$P = \frac{FC}{TC}$$

Let S be a SS and A be an event associated with a Random experiment. Then, the $P(A)$, is defined as a real no satisfying the following axioms.

(I) $0 \le P(A) \le 1$

(II) $P(S) = 1$

(III) If A and B are two MEE, $P(A \cup B) = P(A) + P(B)$

(IV) $\{A_i\}$ of MEE, $P(\bigcup_i A_i) = \sum_i P(A_i)$

Random Variable: A R.V is a function that assign a real no $x \equiv X(s)$ to every element $s \in S$ (sample space).

$$X : S \longrightarrow \mathbb{R} \,/\, R_X \qquad R_X \subseteq \mathbb{R}$$

$\downarrow$
Range space

$\rightarrow \{X \le x\} \equiv \{s : X(s) \le x\}$

Discrete RV's.

X is said to be Discrete if it takes finite no of values.

Continuous RV's

X is said to be Continuous if it takes all values. (infinite values)

(*) Let X be a continuous RV.

X — PDF (Probability Density function)

X — CDF (Cumulative Distribution function)

PDF $f(x)$ is said to be PDF of a RV X, if

(i) $f(x) \geq 0 \quad \forall x \in R_X$

(ii) $\int_{R_X} f(x)\,dx = 1$

$\rightarrow \quad P(X \leq a) = \int_{-\infty}^{a} f(x)\,dx$ → PDF

$\rightarrow \quad P(a \leq X \leq b) = \int_{a}^{b} f(x)\,dx$

$\rightarrow \quad P(X = a) = \int_{a}^{a} f(x)\,dx = 0$

$\rightarrow \quad P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b)$
$= P(a < X < b)$

CDF $F(x)$ is said to be CDF if

$$F(x) = P(X \leq x)$$
$$= \int_{-\infty}^{x} f(x)\,dx$$ → PDF

$$\boxed{F(x) = \int_{-\infty}^{x} f(x)\,dx} \quad \text{--- } (*)$$

$\rightarrow \quad f(x) = \frac{d}{dx} F(x)$

$\rightarrow$ $F(x)$ is non-decreasing function

$\rightarrow$ $F(-\infty) = 0$

$\Rightarrow F(-\infty) = P(X \leq -\infty) = \int_{-\infty}^{-\infty} f(x)\,dx = 0$

$\rightarrow$ $F(\infty) = 1$ (HW).

Moments: (X)

(1) Raw Moments (Moments around zero)

$$\mu_r' = E(X^r)$$

1st moment: $\mu_1' = E(X)$

$$= \int_{R_x} x \, f(x) \, dx \quad f(x) \to \text{PDF}$$

2nd moment $\mu_2' = E(X^2)$

$$= \int_{R_x} x^2 f(x) \, dx \quad f(x) \to \text{PDF}$$

$$Var(X) = \mu_2' - (\mu_1')^2 = \int_{R_x} x^2 f(x) dx - \left(\int_{R_x} x f(x) dx \right)^2$$

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

(*) $Var(X) = E(X - \mu_1')^2 = E(X^2) - (E(X))^2$ (HW)

(2) **Moments**: Moments around mean (μ_1')

$$\mu_r = E[(X-\mu)^r] \quad \mu \text{ is } 1^{st} \text{ raw moment}$$

$$\mu_1 = E(X-\mu)^1$$

$$= \int_{R_x} (x-\mu) f(x) dx \quad f(x) \to \text{PDF}$$

$$\mu_2 = \int_{R_x} (x-\mu)^2 f(x) dx \quad f(x) \to \text{PDF}$$

$E(X-\mu_1')^2$

$\equiv E(X-\mu)^2$

$\equiv \mu_2$

$S \equiv$ Population.

Real World System

$X \to S$

Probability Model

Experiment

Theory

PDF

M_1. Model Prediction

M_2. How to validate ??

Yes

difference

Random Sample

M_i'

No

M_1 is the adequate model

Real World System

PDF

$f(x)$; $f(x) \geq 0$

$\int_{R_x} f(x)\, dx = 1$

1) Uniform Model / Distribution

2) Symmetric distributions

3) Skewed Distributions

4) Exponential Distribution

Normal Distribution

χ^2 Distribution

t Distribution

(1:02)

Normal Distribution / Gaussian Distribution:

A RV X is said to follow ND with parameters μ and σ if its PDF is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \begin{array}{l} x \in (-\infty, \infty) \\ \mu \in (-\infty, \infty) \\ \sigma > 0 \end{array} \quad \text{—(1)}$$

→ $X \sim N(\mu, \sigma)$ ↘ Std. Dev

mean ← $X \sim N(\mu, \sigma^2)$ → Variance

→ $\int_{-\infty}^{\infty} f(x)\,dx = 1$ (H·W)

Standard Normal Distribution: If for a RV X (Z) in eqn (1), we have $\mu = 0$ and $\sigma = \sigma^2 = 1$. Then RV X is called Standard Normal Distributed R.V.

⇒ Z-distribution

From eqn (1), we have the PDF for SND RV,

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad z \in (-\infty, \infty) \quad \text{—(2)}$$

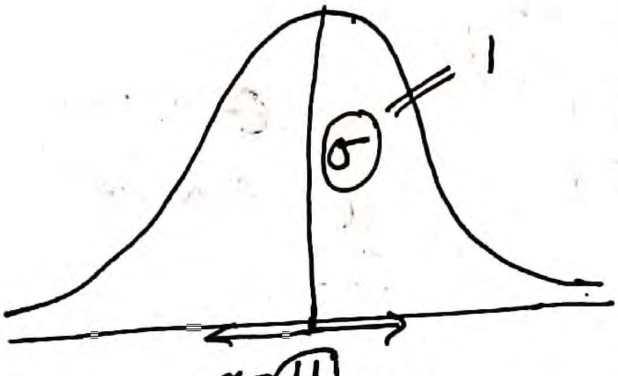
$Z \sim N(0, 1)$

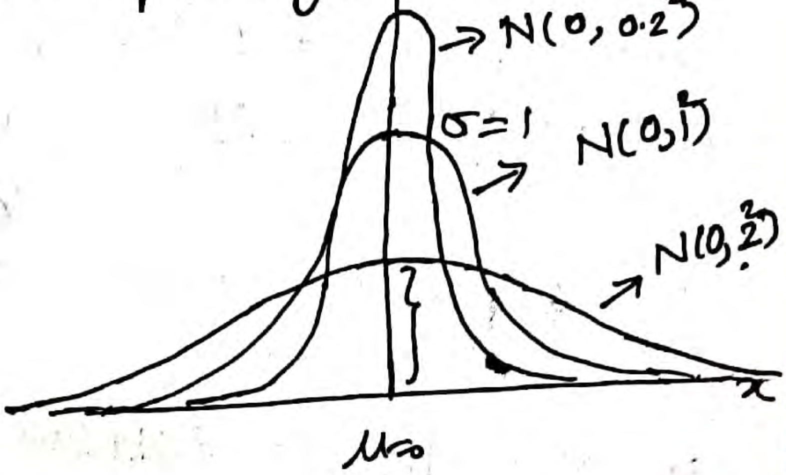
→ Let $f(z)$ be the PDF of SND RV. Then, the CDF

$$\phi(z) = P(Z \le z) = \int_{-\infty}^{z} f(z)\,dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-z^2/2}\,dz$$

→ The PDF of ND is a bell shaped symmetric curve.



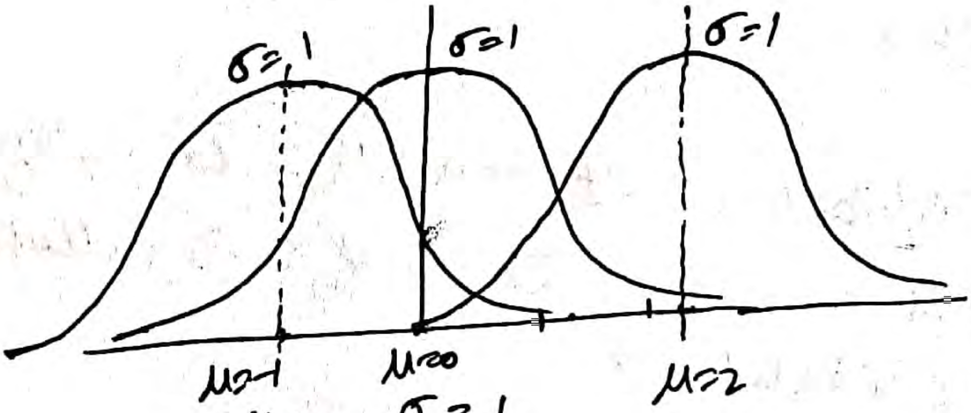


$X \sim N(\mu, \sigma^2)$

$\mu = 0 \quad \sigma = 1$

$\mu = 0 \quad \sigma = 2$

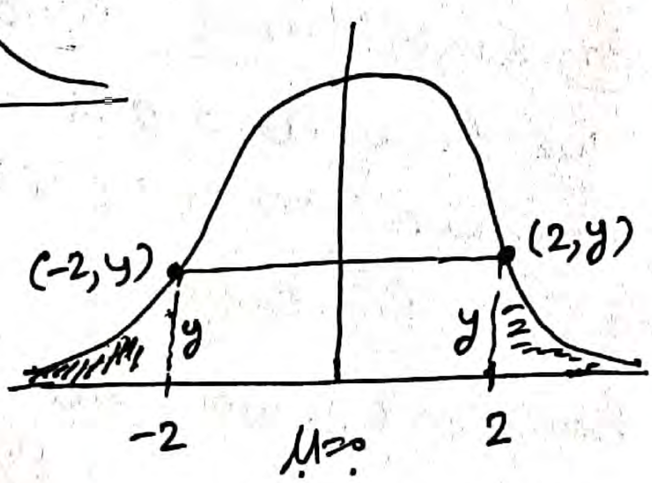
$\mu = 0 \quad \sigma = 0.2$



$\mu = 0 \quad \sigma = 1$

$\mu = 2 \quad \sigma = 1$

$\mu = -1 \quad \sigma = 1$

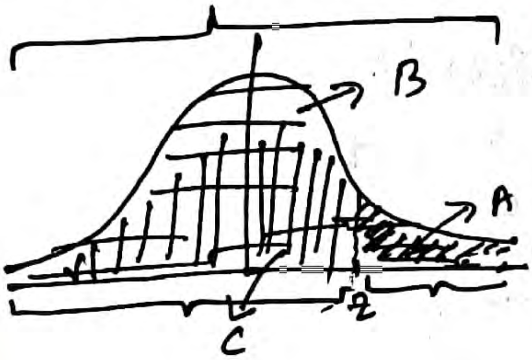


$P(X \le -2) = P(X \ge 2)$

→ $\phi(-z) = 1 - \phi(z)$

Pf $\phi(-z) = P(Z \le -z) = P(Z \ge z) = 1 - P(Z \le z)$

$= 1 - \phi(z)$



$\cancel{A + B = C}$

$A + C = B$

$\Rightarrow A + C = 1$

$P(Z \ge z) + P(Z \le z) = 1$

→ If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

Given $X \sim N(\mu, \sigma^2) \Rightarrow E(X) = \mu \quad Var(X) = \sigma^2$ —— (*)

$\mu_z = E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}[E(X) - E(\mu)]$

$= \frac{1}{\sigma}[\cancel{\mu} - \cancel{\mu}] = 0$

$\begin{cases} E\left(\frac{X}{a}\right) = \frac{1}{a} E(X) \\ E(aX+b) = aE(X) + b \end{cases}$

$Var(Z) = Var\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2} Var(X) = \frac{\sigma^2}{\sigma^2} = 1$

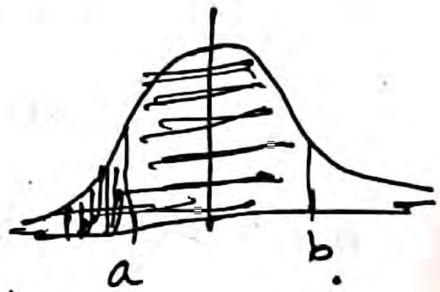
$Var(aX+b) = a^2 Var(X)$

→ $P(a \le X \le b) = \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$

Pf: $P(a \le X \le b) = -P(X \le a) + P(X \le b)$

$= P\left(\frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right)$

$- P\left(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right)$



$= P\left(Z \le \frac{b-\mu}{\sigma}\right) - P\left(Z \le \frac{a-\mu}{\sigma}\right)$

$= \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$

→ Mean, median and Mode are same $X \sim N(\mu, \sigma^2)$

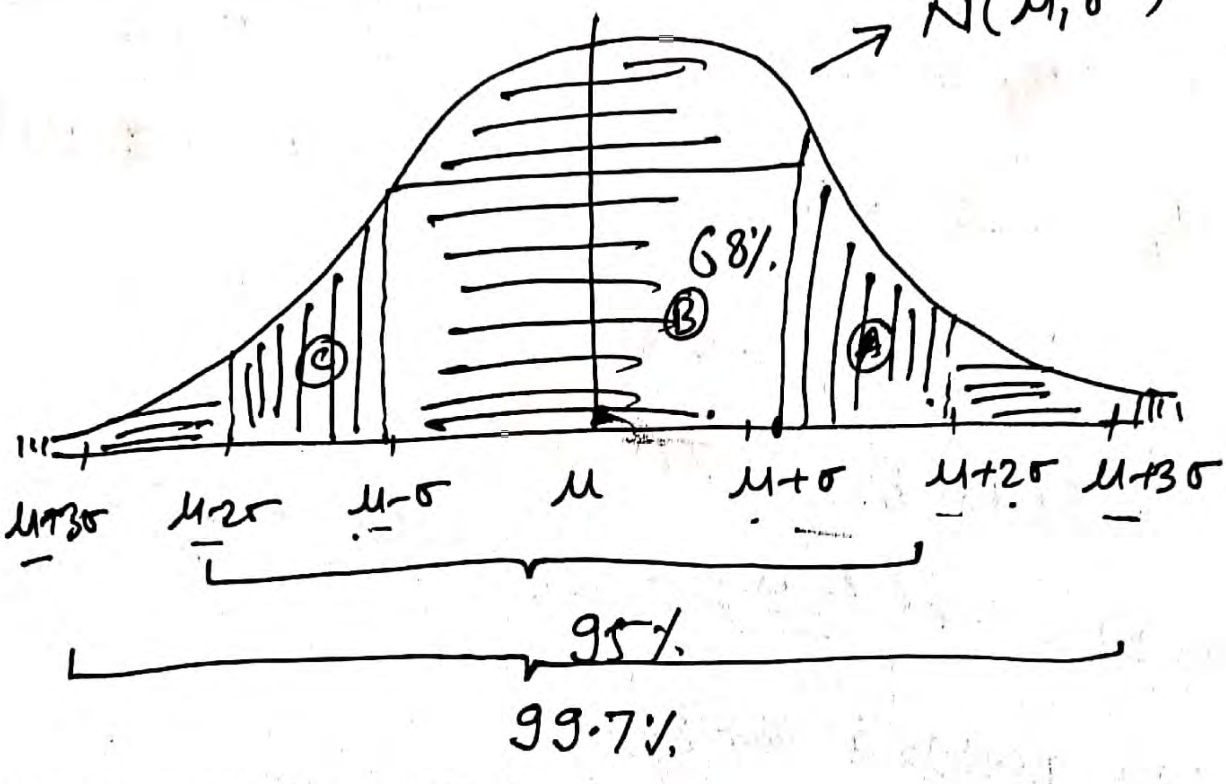
Mean $= \mu_1'$
$= E(X)$
$= \mu$

Median is soln of
$\int_{-\infty}^{M} f(x)\,dx = \frac{1}{2}$
$\Rightarrow \boxed{M = \mu}$
(HW)

Mode is a value for which $f(x)$ is maximum.
$f'(x) = 0$ s.t $f''(x) < 0$
$\Rightarrow \boxed{\text{Mode} = \mu}$

Scanned with CamScanner

→ 68-95-99.7 rule.



ex M.Tech Online (AI)

	Mathematics	ML
	60, 4	79, 2

X → 65/100 (M)

Y → 80/100 (ML) who perform better.??

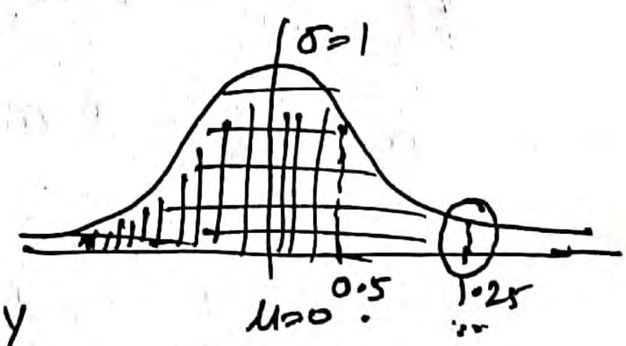
(1) Can not be compared... No

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

$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

X: Z-score $\frac{65-60}{4} = 1.25$

Y: $\frac{80-79}{2} = 0.5$



⇒ X performed better than Y.

χ^2 **Distribution:** Square of a SNV is known as χ^2-variate.

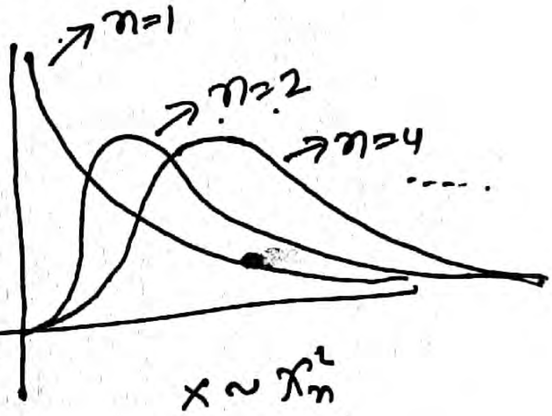
→ Let $Z_i : i = 1, 2, \ldots, n$ be n independent SNV. Then

$$X = \sum_{i=1}^{n} Z_i^2$$

is said to be χ^2-distributed RV with n degree of freedom.

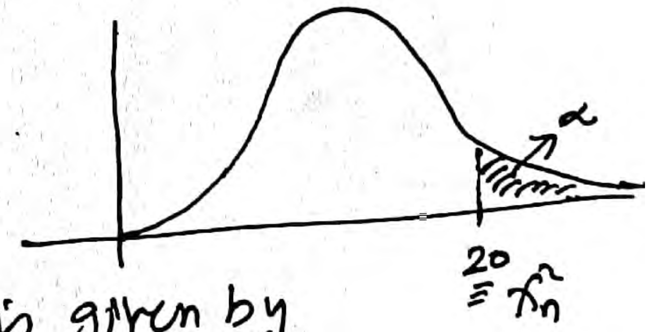
→ Degree of freedom tells how many independent SNV are used to construct the χ^2-distribution.

→ Let $X_i \sim N(\mu_i, \sigma_i^2)$ which are independent in nature. Then, χ^2-distribution can be obtained after converting X_i's into Z_i's.



$X \sim \chi^2_n$

$$\underset{N(\mu_i, \sigma_i^2)}{X_i} \longrightarrow \underset{N(0,1)}{Z_i} \longrightarrow \chi^2$$



$20 \equiv \chi^2_n$

* Let $X \sim \chi^2_n$. Then, its PDF is given by

$$f(x) = \frac{1}{2^{n/2}\,\Gamma\frac{n}{2}} e^{-x/2} x^{(n/2 - 1)} \; ; \; x \in [0, \infty)$$

$$\boxed{P(X \geq \underset{20}{\chi^2_n}) = \alpha}$$

→ $\mu_{\chi^2} = E(X) = n$ (DoF)

→ $\sigma_{\chi^2}^2 = Var(X) = 2n$ (DoF)

⇒ $X \sim \chi^2_n \quad \underline{\underline{n=2}} \quad X \sim \chi^2_2 \quad \underline{\underline{n=3}} \quad X \sim \chi^2_3$

The Student t- Distribution!

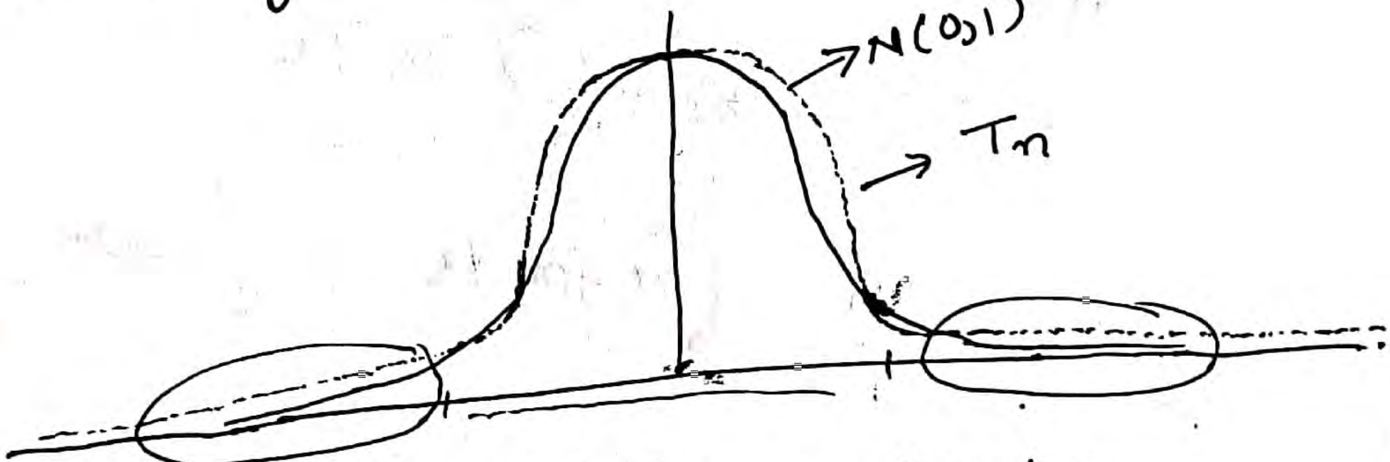
Let $Z \sim N(0,1)$ and $Y \sim \chi^2_n$. If Z and Y are independent, then a RV T_n

$$T_n = \frac{Z}{\sqrt{\frac{Y}{n}}}$$

is said to be t-distributed R.V.

→ T distribution is a bell shaped, symmetric curve.

→ T distribution has thicker tails, indicating greater variability, than that of a Normal distribution



→ The PDF of T-distribution is given by

$$f(x) = \frac{\Gamma\frac{n+1}{2}}{\sqrt{\pi n}\,\Gamma\frac{n}{2}} \cdot \frac{1}{\left(1+\frac{x^2}{n}\right)^{\frac{(n+1)}{2}}} \; ; \; x \in (-\infty, \infty)$$

→ $\underline{\mu_T = E(X)} = 0 \; (n>1)$ $Var(X) = \frac{n}{n-2}$ $(n>2)$

(HW) $\underline{(n>2)}$

Q Let $X \sim T_n$ $(n>1)$. Then prove that $E(X)=0$.

$$E(X) = \int_{R_x} x\, f(x)\, dx$$

$R_x \to$ Range $(-\infty, \infty)$

$f(x)$ – pdf.

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

$$\int_a^b f(x)\,dx = \int_a^c + \int_c^b f(x)\,dx$$

$c \in [a,b]$

$$= \underbrace{\int_{-\infty}^{0} x\, f(x)\, dx}_{(I)} + \underbrace{\int_{0}^{\infty} x\, f(x)\, dx}_{(II)}$$

put $t=-x$ in (I)

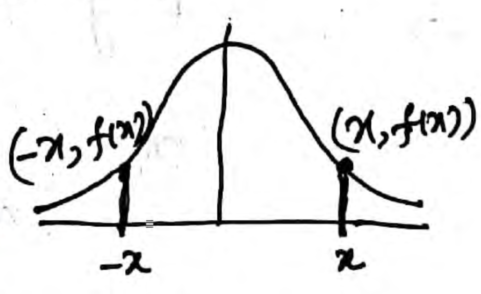
$dt = -dx$

$$= \int_{\infty}^{0} (-t)\, f(-t)\, (-dt) + \int_0^{\infty} x\, f(x)\, dx$$

$$= \int_{\infty}^{0} t\, f(-t)\, dt + \int_0^{\infty} x\, f(x)\, dx$$

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$$

$$= \underbrace{-\int_0^{\infty} t\, f(-t)\, dt}_{(I)} + \underbrace{\int_0^{\infty} x\, f(x)\, dx}_{(II)}$$



put $x=t$ in (I)

$$= -\int_0^{\infty} x\, f(-x)\, dx + \int_0^{\infty} x\, f(x)\, dx$$

$f(x) = f(-x)$

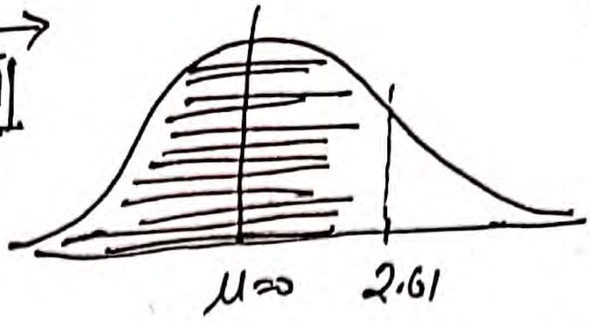
$$= -\int_0^{\infty} x\, \cancel{f(x)\, dx} + \int_0^{\infty} x\, \cancel{f(x)\, dx} = 0.$$

$\mathcal{M}$

Q1 $P(z \leq 2.61)$ $\rightarrow$ $2.61 \rightarrow$

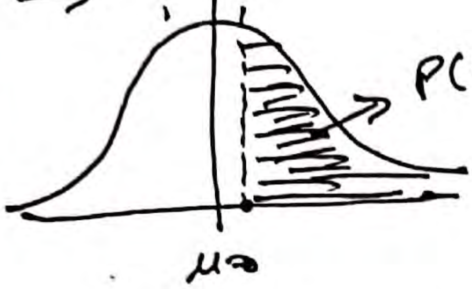
$= 2.6 + \boxed{.01}$

$= .99547$



Q2 $P(z > .05)$

$\rightarrow$



$P(z > .05) = 1 - P(z \leq .05)$

$= 1 - .51994$

$=$

$.05 \rightarrow 0.0 + .05$

Q3 Find the proportion of the population that has their first child before age 27.2?

Let $X \equiv$ Age at first child

$X \sim N(25.4, 1.5^2)$

$P(X < 27.2)$

$= P\left(\frac{X - 25.4}{1.5} < \frac{27.2 - 25.4}{1.5}\right)$

$= P(z < 1.2)$

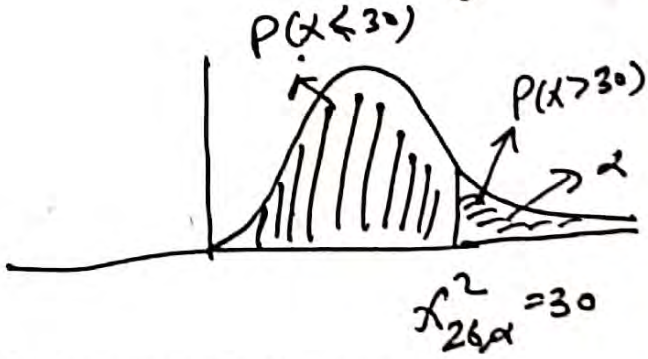
$= 0.88493$

$1.2 \Rightarrow 1.2 + 0.00$

Q4 $P(X_{26} \leq 30)$

$= 1 - P(X_{26} > 30)$

$= 1 -$

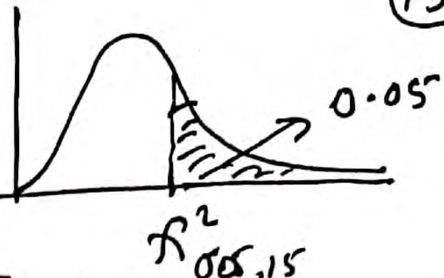


$17.292 \leq P(X_{26} > 30) \leq 35.563$

$1 - 35.563 \leq 1 - P(X_{26} > 30) \leq 1 - 17.292$

$-34.563 \leq P(X_{26} \leq 30) \leq -16.292$

Q5 Find $\chi^2_{0.05,15}$



$\chi^2_{0.05,15} \equiv P(X \geq \chi^2_{0.05,15}) = 0.05$

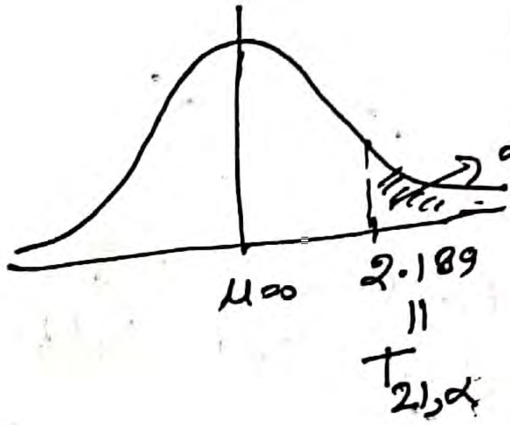
$n = 15$

$\alpha = 0.05$

$\Rightarrow \boxed{\chi^2_{0.05,15} = 24.996}$

Q6 $\cancel{P(T_{21}}$ $P(T_{21} \geq 2.189)$

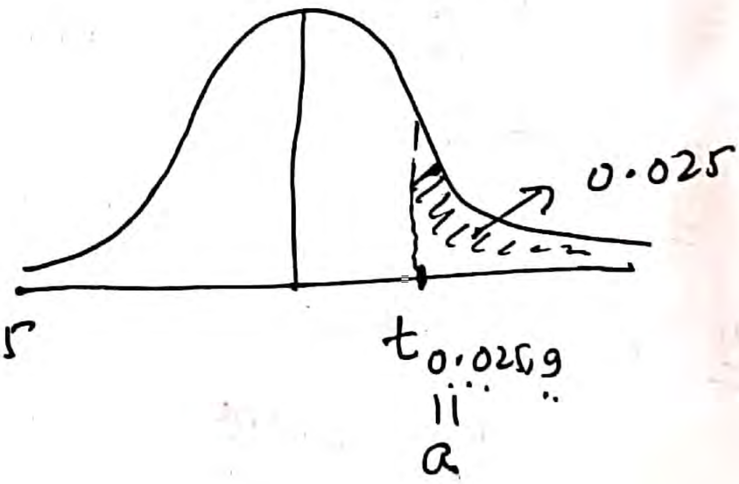
$= 0.02$



Q7 $P(X_{12} \leq 1.4)$

$= 1 - P(X_{12} \geq 1.4)$ (HW)

Q8 Find $t_{0.025,9}$?



$t_{0.025,9} \equiv$

$P(X > a) = 0.025$

$\equiv a = 2.262$

$\Rightarrow t_{0.025,9} = 2.262$

Sampling:

ex We want to perform a study to determine the # km the average person in India drive a car in one day.

⇒ Population: Any person who is the resident of India is a part of population.

⇒ Sample: A part of the population.

10 persons randomly

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

→ Type of Sampling:

(I) Purposive Sampling: Any sample in which the entities are selected with definite purpose in view.

(II) Random Sampling: Select the entities from the population by chance s.t each entity has an equal chance of being included in the sample.

(III) Stratified Sampling: If the population has hetrogeneous entities, then it is a way to construct the sample. Select equal entities from the group of homogeneous entities of the hetrogeneous population.

Systematic Sampling: Selection of entities from the population based on a systematic process in place.

Sample No	Average no of Km
1	25.6
2	50.2
3	15.1
4	43.9
5	36.8
6	60.2
⋮	⋮