Probability & Statistics :-

Random variable :-

ex:- rolling dice → 6 sides = {1, 2, 3, 4, 5, 6}

when rolled → any one of these equal outcome

Random Experiment

random variable X = {1, 2, 3, 4, 5, 6} → Sample Space

tossing a coin → Y = {H, T}

P(X = 1) = 1/6 P(X = 2) = 1/6 ....

P(X is even) = 3/6 = 1/2

(probability of X being even)

(P(X=2) + P(X=4) + P(X=6) = 1/6 + 1/6 + 1/6)

P(X is odd) = 1/2

$P(x = x_1) \rightarrow P(x_i)$ same thing diff notations.

Finite set of values → Discrete random value

→ Height of randomly picked student :

y could be 162, 180, 120, 140, ... → infinite values → Continuous Random Variable

Outliers :-

Y : height of student

{122.2, 146.4, 132.5, ... 12.2, 156.3, 92.7 ...}

92.7 → could be an outlier

12.2 → outlies → could be human error (or) actual height

→ Outlies can corrupt data.

→ A discrete value is obtained by counting

→ A continuous value is obtained by measuring.

Sample Space :- Set of all possible outcomes of an experiment.

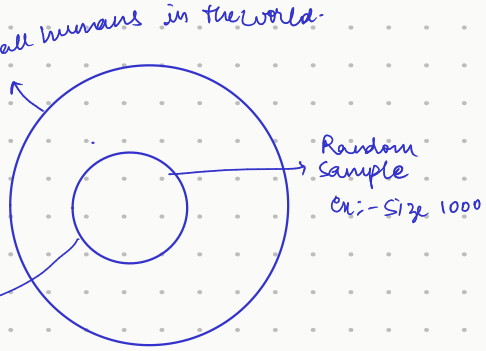
→ A random variable value depends on the outcome of a random phenomenon.

Population & Sample :-

→ Estimating the average height of human

↳ $\mu = \frac{1}{Pop} \sum_{i=1}^{Pop} h_i$ (IMPOSSIBLE)

So we estimate.



often represented by $\bar{h}$ ← $\bar{x} = \frac{1}{1000} \sum_{i=1}^{1000} h_{is}$

→ As sample size increases $\bar{x} = \mu$

($\bar{x}$ → Sample mean; μ → original mean)

Sampling is of two types:-

(i) Simple Sampling

(ii) Stratified Sampling → Unbiased sampling & more accurate results.

ex:- 1000 people ········→ original sample

- 400 have cars
- 300 have bikes
- 200 have cycles
- 100 have nothing

Sample Size → 250

Simple Random Sampling

250 could have cars

(or)

250 could have bikes

(or)

100 bikes + 150 cars.

Stratified Random Sampling

cars → 100

bikes → 75

cycles → 50

nothing → 25

} There are random but equal imp to all classes