✳ Probability Density Function:
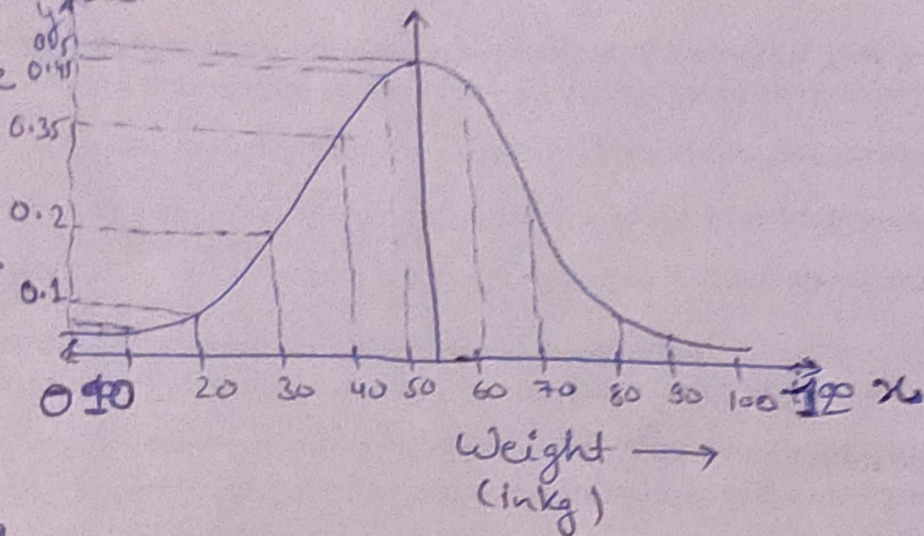
Probability density function, in statistics is a function whose integral is calculated to find probabilities associated with a continuous random variable.

Suppose, that we have a sample, in which we draw a sample distribution for weights of the people in range 0 to 100 kg.

Weight ⟶
(in kg)

Then PDF can help us to find what % of people in the sample lie with in this range.

The probability of people distributed around entire sample coming in some weight range will always be equal to 1, i.e., area under the curve will always be 1.

Suppose, we look at graph, and say probability of people having value of weight less than 20 kg, then we extend the line towards y-axis and see the corresponding probability, and it says 0.1, that means 10% of people in the population will have weight within 20 kg.

Similarly 50% of people in the population has weight within 55 kgs.

→ PDF is really useful function when it comes to understanding the statistical behaviour of a continuous data.
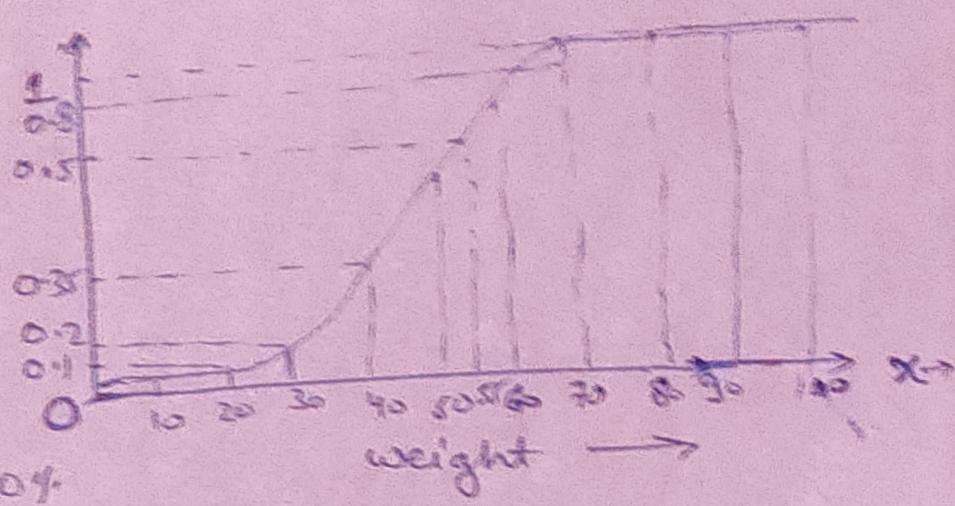
* Cumulative Density Function (CDF):

The cumulative distribution function for a random variable at X gives the probability that the random variable X is less than or equal to that number x.

Here, the y-axis, shows the cumulative sum.

Suppose, if we want to find out no. of people having weight less than 70, we can directly see graph and tell it is 80% of the population.

→ CDF can be used for both discrete and continuous variables.

* Z-score: A z-score gives us an idea of how far from the mean a data point is. But more technically, it's a measure of how many standard deviations below or above the population mean a raw score is.

We can convert a regular normal/ gaussian distribution into a standard normal distribution using the Z-score.
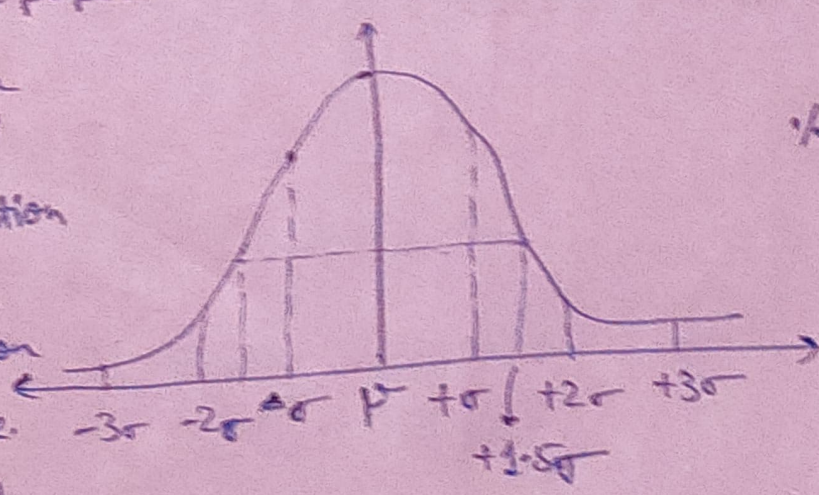
Suppose, we want to find out % people living beyond 1.5σ of mean, we can use the z-score the find the value of people.

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

* We can make use of the Z-score table to find the amount of people that lie in a certain region using the z-score.