Random Variable(X):

Random variable means outcome of the event is more than one value and each value of the variable has equal chance to occur.

Random variable is nothing but if an event (tossing coin, rolling dices) takes multiple values, its associate likelihood or chance of each of that value to occur. A variable remains random unit it occurred. If it occurs, we know the outcome and no more random.

Random variable should take more than one value and each value of the variable there should be associate probability (equal chance to occur for all the values). All the futuristic events are random variables except universal truths.

- Example whether it rains today in Hyderabad or not? As possibilities of outcome is Yes and No. there is the possibility to occur of either of the value.
- Tossing the coin landing with Heads. Outcome of the value stores as Random Variable

Random variables are divided into 2 types:

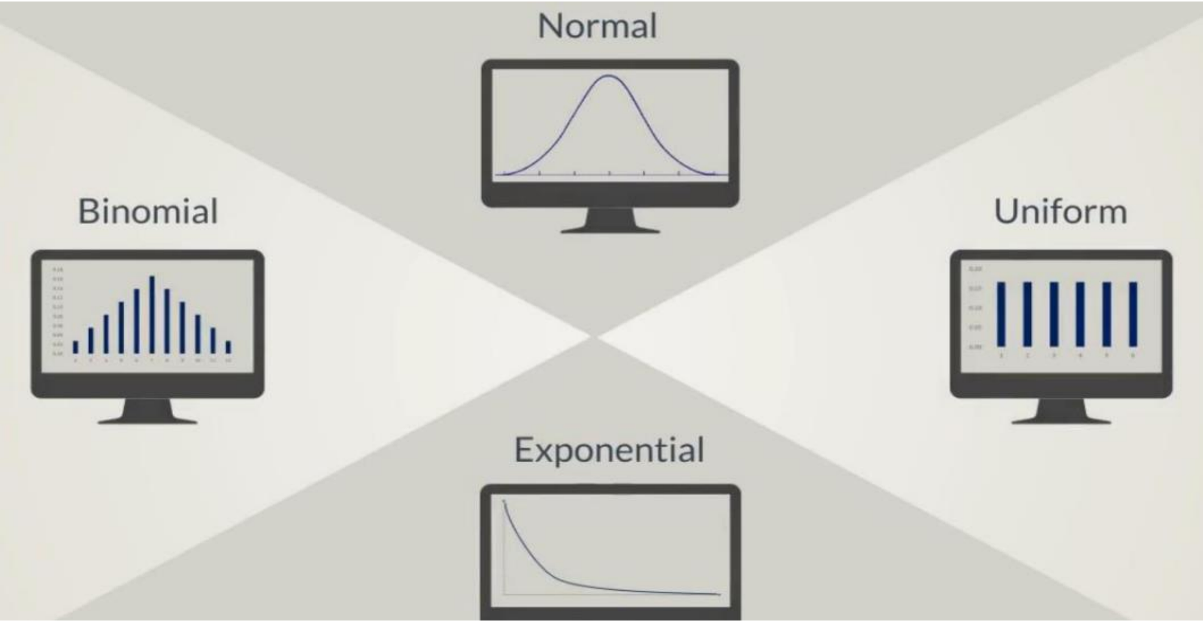
A. **Discrete Random Variable:** Random variable that can assume a countable number of values. Random variable that takes finite number of values.

   EX: number of defect light bulbs in the box. Number of children in the family.

B. **Continuous Random Variable:** Random variable that can assume any quantity (value) that can assume all possible values on numerical scale. These are infinite numbers.

   Ex: time it takes to complete the race? It can be 3 min or 5 min or 3.2 min……

Distribution: It is nothing but framework or set of rules or a function. Any data follows those set of rules defined by a distribution, then call that data follows distribution.



Probability Distribution: Possible values for a variable and how often they occur.

Ex: Rolling a die and probability of getting 1 is 1/6

Probability Distribution has 2 types.

A. Discrete Probability Distribution: Probabilities associated with discrete variables is called discrete probability distribution. It can be expressed in terms of graph or table. Example is Binomial Distribution

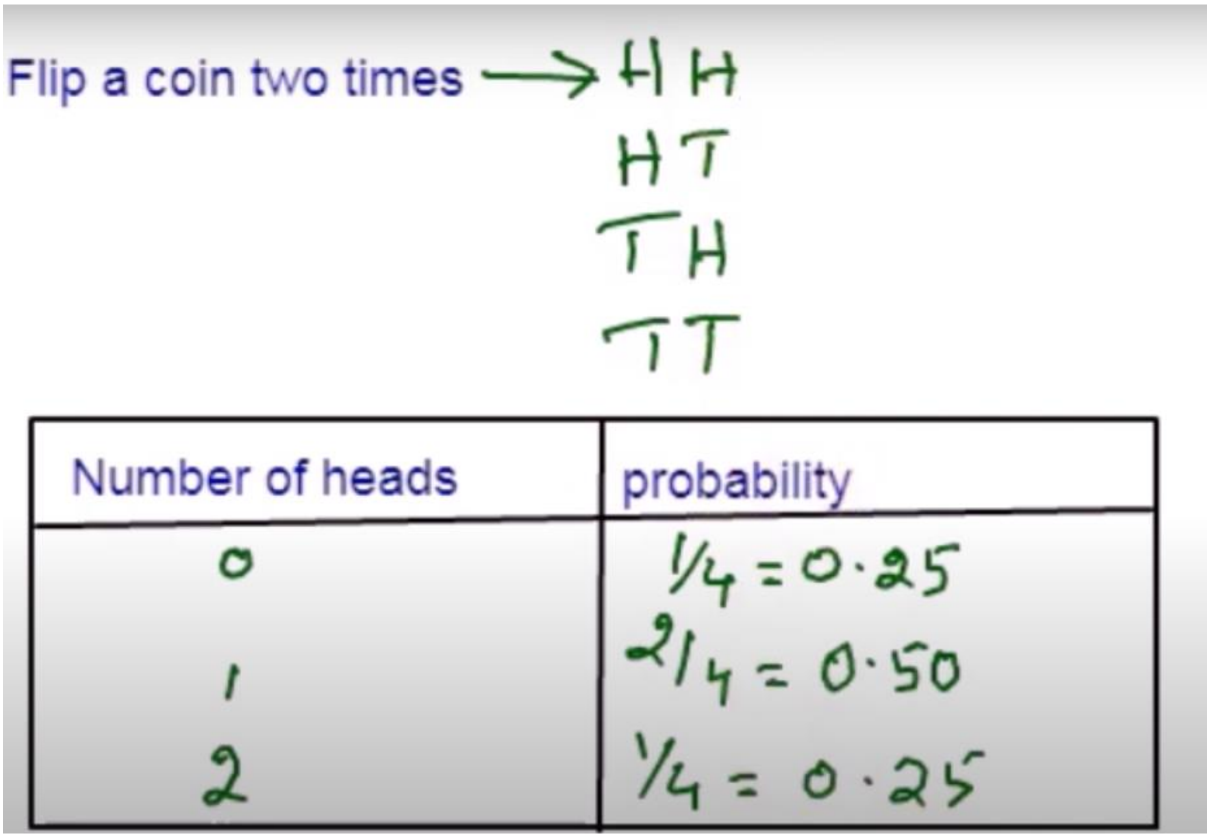
Binomial Distribution: Binomial distribution has its own rules, any data satisfies these rules, we can call that dataset follows Binomial distribution.

Following are the set of rules:

- The experiment should consist of “n” repeated trials (more than 2 times). That means it can 100 time, 50 times… finite number of times.
- Each trial can result in just two possible outcomes.
- The Probability of Success, denoted by P, is the same on every trail. That means always P(success = 0.5)
- The trails are independent; that is, the outcome on one trail does not affect the outcome on others trail.

Examples: Tossing a coin then what is the probability of landing Head. P(X=Head) =0 or P(X=Head)=1. Here 0 or 1 is the Binomial distribution value.

What is the probability of passing the exam? P(X=Pass)=0 or P(X=Pass)=1



Flip a coin two times $\longrightarrow$ H H
H T
T H
T T

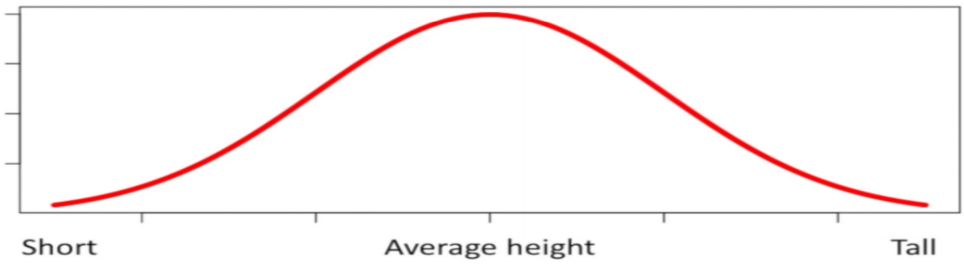
Number of heads	probability
0	1/4 = 0·25
1	2/4 = 0·50
2	1/4 = 0·25

Probability is not for single time/event, occurs for many repetitions/longer run.

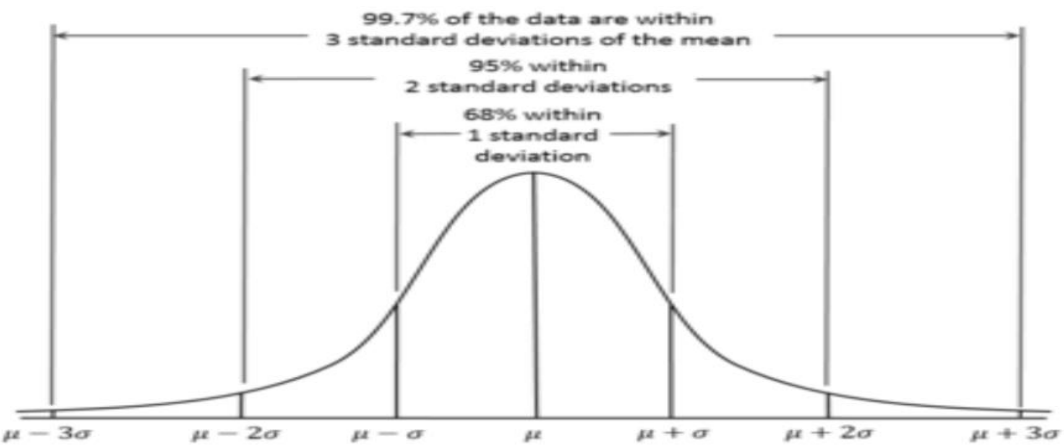
B. **Continuous Probability Distribution:** Probabilities associated with continuous variables is called continuous probability distribution. Example is Normal Distribution

Normal Distribution: Following are the rules/properties of Normal Distribution.

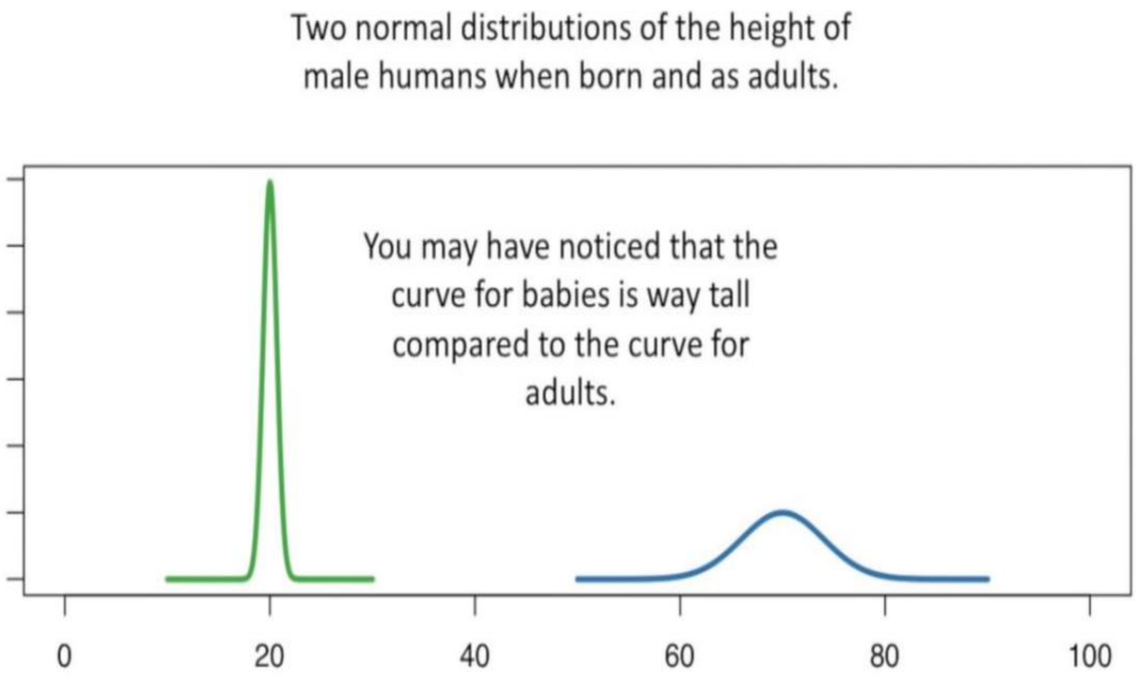
In our universe everything follows normal distribution examples height, weight, intelligence, lifestyle and so on. If we consider the height short persons and tall persons heights are less when compare to average height persons

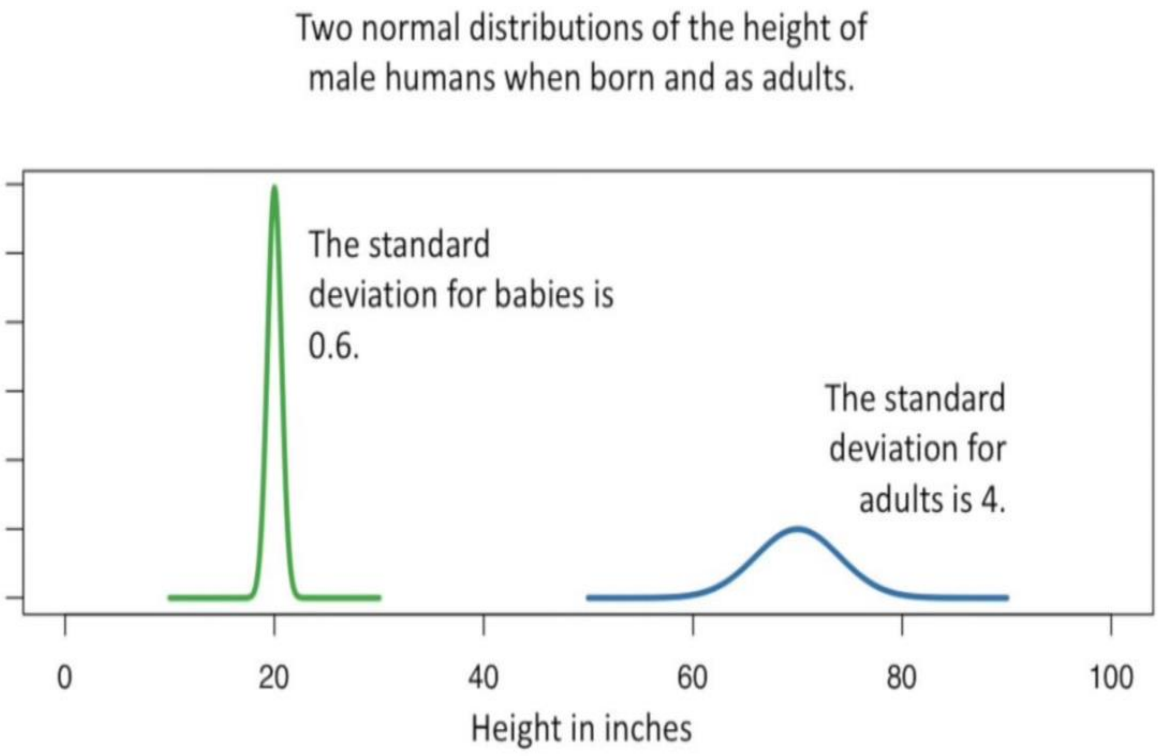


- Data points falls in bell shared curve.
- Shape will be symmetric around the mean. It means if we cut the center point(median), divide number of data points equally.
- Always Mean is 0 and Standard Deviation is 1
- X axis in Normal Distribution is Range.
- Y axis in Normal Distribution is Probabilities.
- Mean, Median and Mode of the data points are equal
- Empirical formula:
  - Probability of the variable that falls between range $\mu - \sigma$ and $\mu + \sigma$ which is basically range of first standard deviation will approximately equals 68%
  - Probability of the variable that falls between range $\mu - 2\sigma$ and $\mu + 2\sigma$ which is basically range of second standard deviation will approximately equals 95%
  - Probability of the variable that falls between range $\mu - 3\sigma$ and $\mu + 3\sigma$ which is basically range of third standard deviation will approximately equals 99.7%

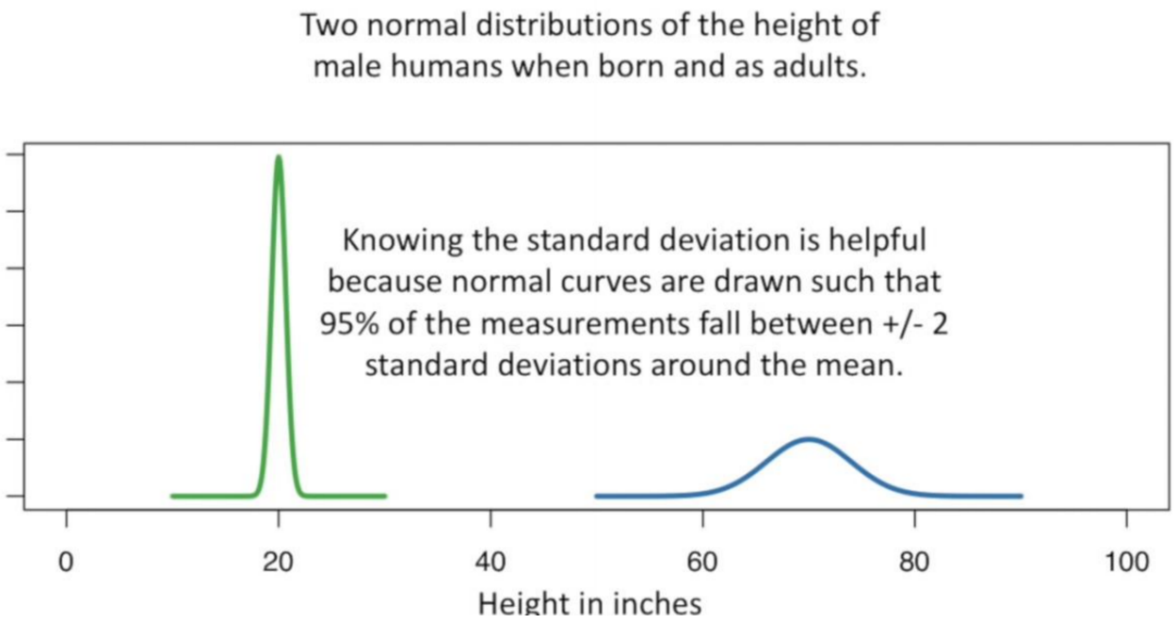


Interpretation of Normal Distribution:





- In the above examples, for the babies(green) bell curve is tall means there are less possible values and standard deviation is less. More data in the middle and thinner tails
- For adults(blue) bell curve is flatten means there are high possible values and standard deviation is high. Less data in the middle and flatten tails



Calculation of range of value based on Mean and standard deviation

- For Babies height Mean is 20 and Standard Deviation is 0.6, to estimate the 95% of the range of the value that is from -2 standard deviation to +2 standard deviation is = 20 - 2*std to 20+ 2*std
  = 20-1.2 to 20+1.2
  = 18.8 to 21.2

  That means 95% range of the newborn babies' height falls between 18.8 inches to 21.2 inches

- For Adult height Mean is 70 and Standard Deviation is 4, to estimate the 95% of the range of the value that is from -2 standard deviation to +2 standard deviation is = 70 - 2*std to 70+ 2*std
  = 70-18 to 70+8
  = 62 to 78

  That means 95% range of the Adults height falls between 62 inches to 78 inches

Why we are converting from Normal Distribution (ND) to Standard Normal Distribution (SND)

As the area proportions are same in Normal Distribution and Standard Normal Distribution. In the Standard Normal Distribution, we have standardized tables (Z distribution). With the help of this we can easily find the probability value for the given Random variable. Characteristic of Standard Normal Distribution Mean is 0 and Standard Deviation is 1.

$$\text{Standardized variable} = \frac{\text{Original variable} - \text{Mean}}{\text{Standard deviation}}$$

N~(μ,σ)

N → Normal

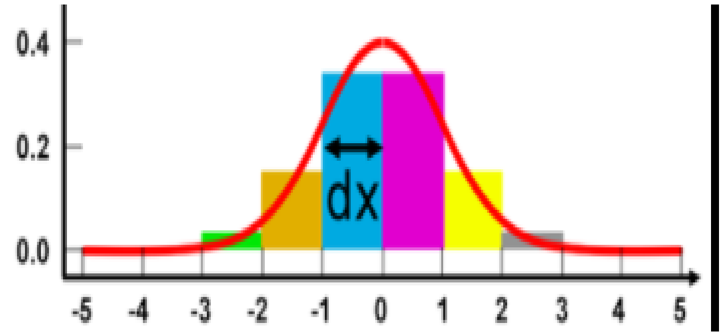
~ → Distribution

μ → Mean

σ → Standard Deviation

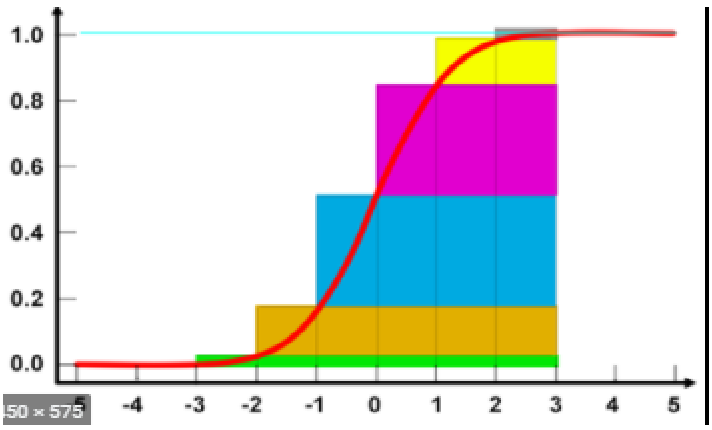
Probability Density Function: When we smoothen the histogram, it creates bell curve (Normal distribution). Total (sum of) area under the bell curve is always 1. The PDF is used to specify the probability value of the X value or random variable falling within a particular range of values.

Example: With the probability density function we can find the probability value of random variable value falls marks between 70 to 75



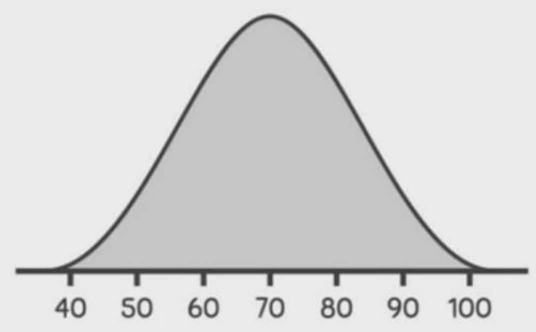
Cumulative Distribution Function: sum-up the x axis probability values. Here we can find the probability value up to the x axis value.

Example: With this cumulative distribution/density function we can find the probability value of random value up to 75 marks.

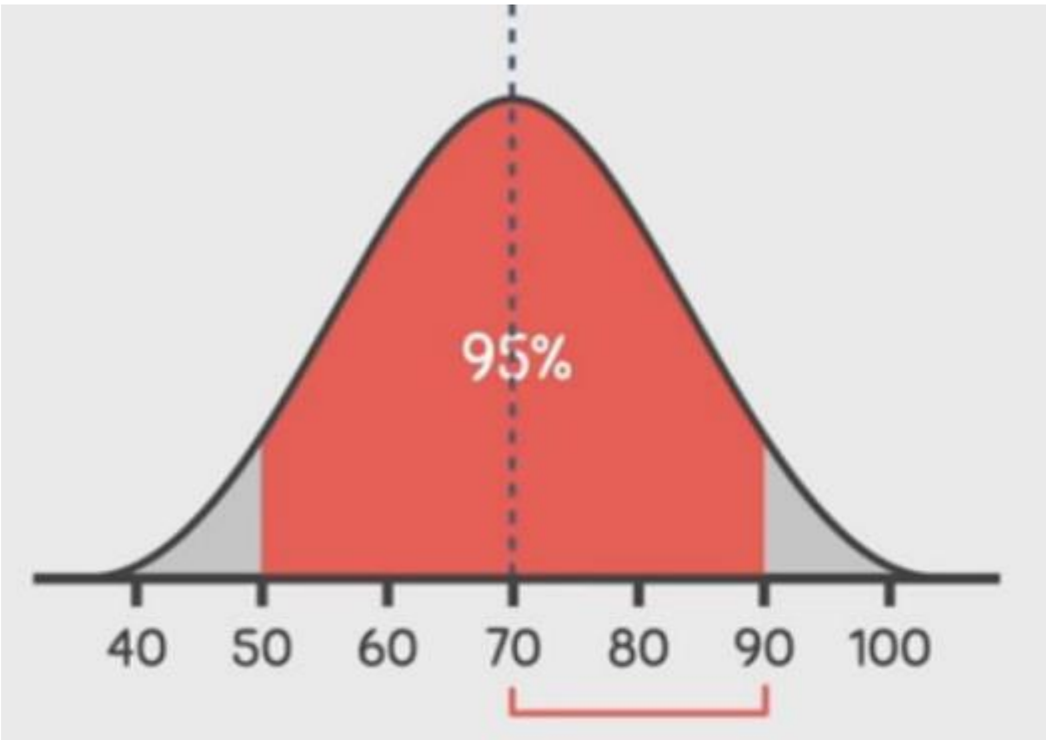


Practical Questions: 1

The normal distribution below has a standard deviation of 10. Approximately what area is contained between 70 and 90?

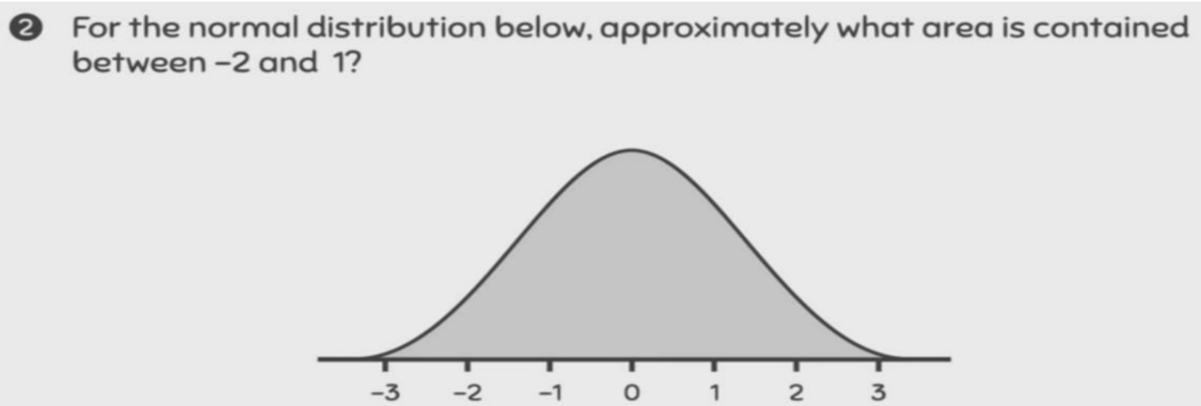


Solution: As per Empirical Formulae for 2 standard deviation ie area from -2 σ to 2 σ is 95%.

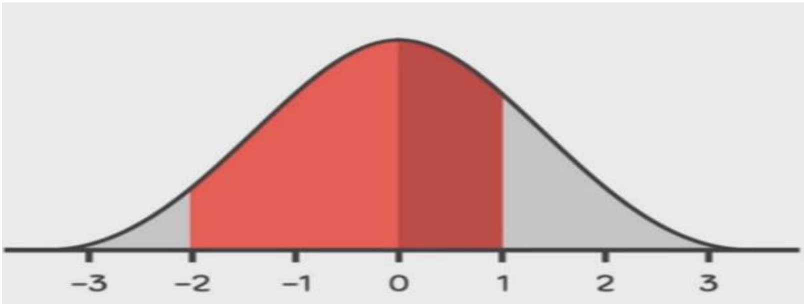


That means area from 50 to 90 is 95%. Hence Area from 70 to 90 is 95/2 = 47.5%

Practical Questions: 2



Need to find the following area

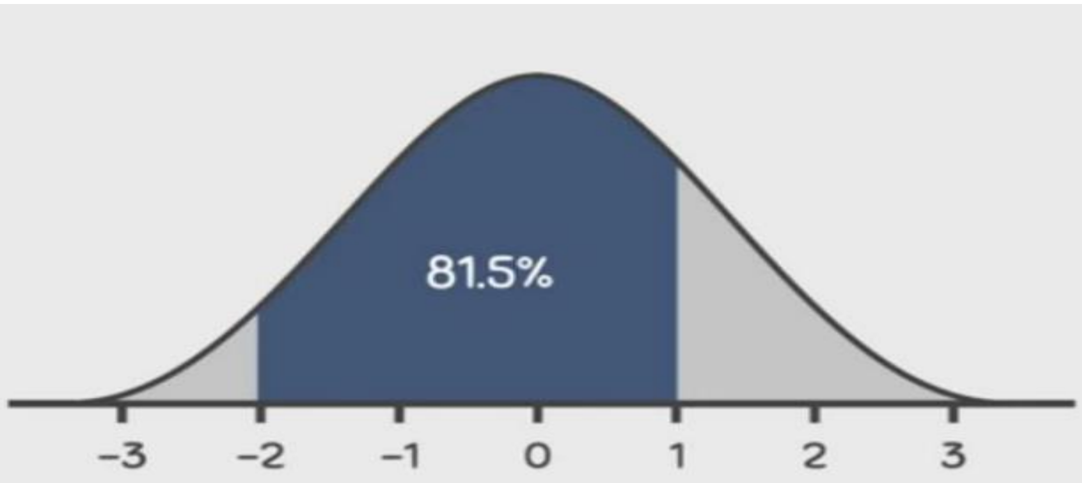


Solution: As per Empirical Formulae for 2 standard deviation ie area from -2 σ to 2 σ is 95% (that is from -2 to +2) and 1 standard deviation ie area from -1 σ to 1 σ is 68%% (that is from -1 to +1)

Area from -2 to +2 is 95% hence from -2 to 0 is 95/2 = 47.5%

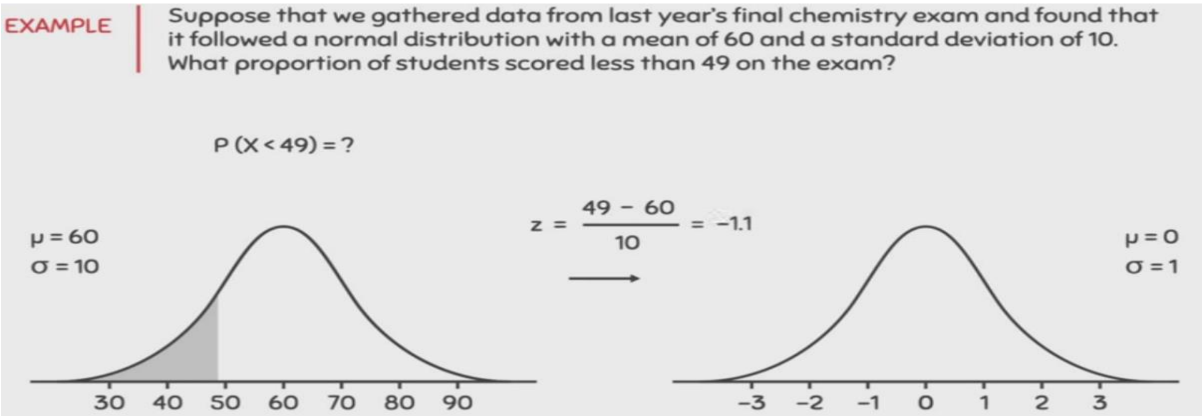
Area from -1 to +1 is 68% hence from -0 to 1 is 68/2 = 34 %

Total Area from -2 to 1 is 47.5+34 = 81.4%



In the above examples, as the area follows within Empirical Rules we found with the Empirical rules percentages. But assume as need to find area up to -1.1 σ or 49 marks, it is difficult to calculate manual. With the help of Z tables, we can find easily.

Practical Problem 1:



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

$$\text{Standardized variable} = \frac{\text{Original variable} - \text{Mean}}{\text{Standard deviation}}$$

$$z = \frac{49 - 60}{10} = -1.1$$

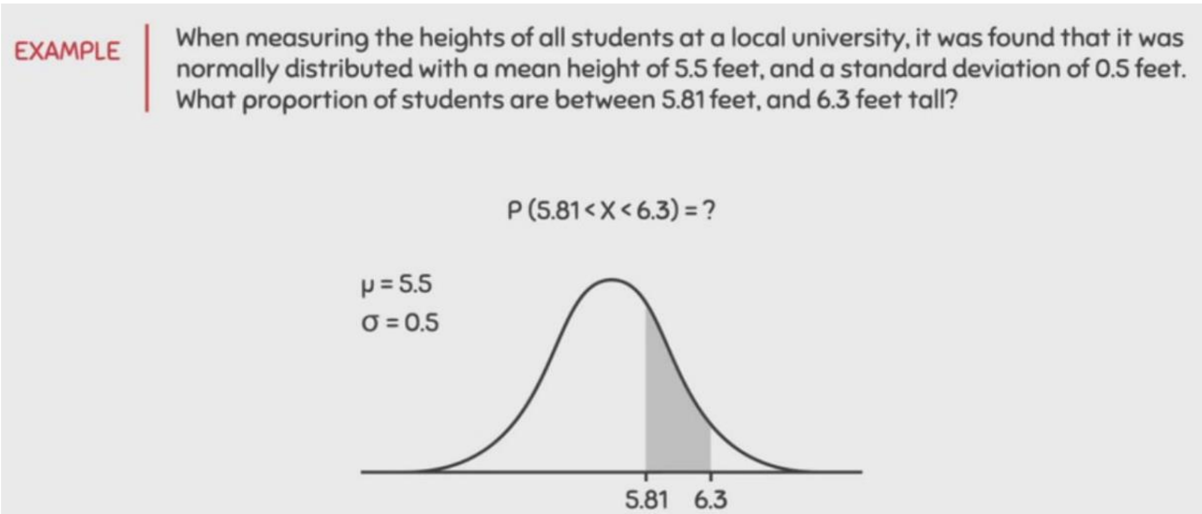
Need to check Z value -1.1 in Z table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379

$$P(X < 49) = 0.1357$$

That means 13.57% of the students scored less than 49 marks

Practical Problem 2:



Find the proportion of the students height up to 5.81 feet

$$z = \frac{5.81 - 5.5}{0.5} = 0.62$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
↓										
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633

Z value of 0.62 in Z table is 0.7324

Proportion of the students height up to 5.81 feet is 0.7324

Find the proportion of the students height up to 6.3 feet

$$z = \frac{6.3 - 5.5}{0.5} = 1.6$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
↓										
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633

Z value of 1.6 in Z table is 0.9452

Proportion of the students height up to 6.3 feet is 0.9452

$P(5.81 < X < 6.3)$ → Proportion of the students height between 5.81 to 6.3

= P(X<6.3) – P(X<5.81)

= 0.9452 - 0.7324

= 0.2128 Proportion or 21.28%