DISCRETE DISTRIBUTIONS - BINOMIAL DISTRIBUTION

↳ whole Number

∞ trials

N trails

- A distribution where only two outcomes are possible, such as success or failure, gain or loss, win or lose and where the probability of success and failure is same for all the trials is called a Binomial Distribution.
- The outcomes need not be equally likely.
- Each trial is independent. 10 Times 17 heads
- A total number of n identical trials are conducted.
- The probability of success and failure is same for all trials. (Trials are identical.)
- Mathematical Representation

X ⇒ # of successful outcomes in N Trails

$$P(x) = \frac{n!}{(n-x)!\,x!} p^x q^{n-x}$$

This starts the count of number of ways event can occur.

This is the probability of success for x trials.

This ends the count of number of ways event can occur.

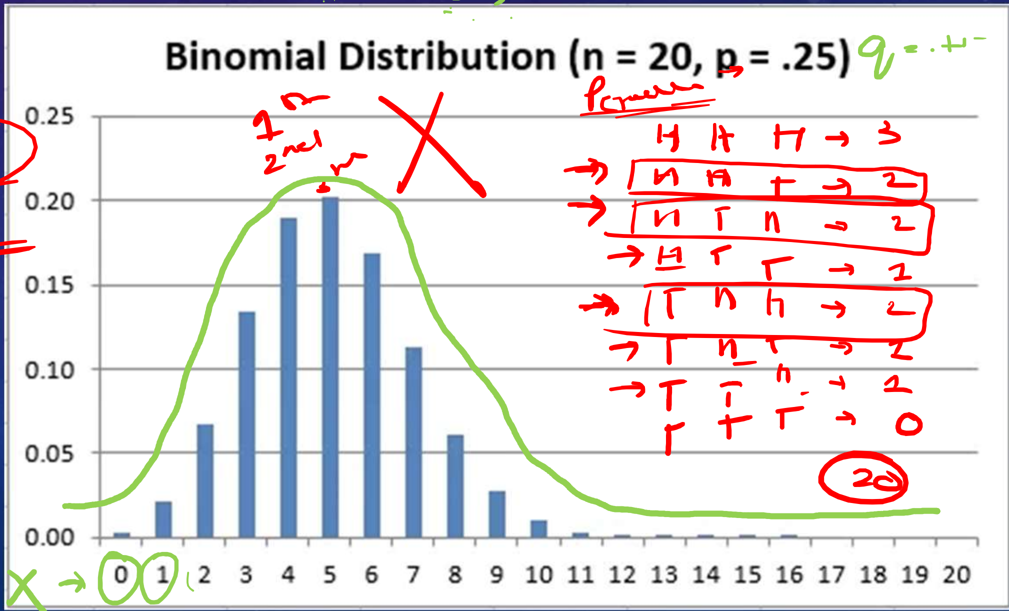
This deletes duplications.

This is the probability of failure for the x trials.

$n! = n * (n-1) * (n-2) * \dots * 2 * 1$

$5! = 5 * 4 * 3 * 2 * 1$

$3! = 3 * 2 * 1$



- If a coin is tossed 5 times, using binomial distribution find the probability of:
  - Exactly 2 heads
  - At least 4 heads

$$P(x) = \frac{n!}{(n-x)!\,x!} p^x q^{n-x}$$

$n = 5$

$p = 1/2 \mid q = 1/2$

$E(x) = \text{Heads}$

$- \text{Tails}$

$p + q = 1$

$q = 1 - p$

$$P(x=2) = \frac{5!}{3!\,2!}\left(\frac{1}{2}\right)^2 * \left(\frac{1}{2}\right)^3$$

$$= \frac{10}{2^5} = \frac{10}{32} = \frac{5}{16}$$

≥ 4, 4 or 5

$$P(x \geq 4) = P(x=4) + P(x=5)$$

$$\Rightarrow \frac{5!}{1!\,4!}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^1 + \frac{5!}{0!\,5!}\left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0$$

$$\Rightarrow \frac{5}{32} + \frac{1}{32} = \frac{6}{32} = \frac{3}{16}$$

- An agent sells life insurance policies to five equally aged, healthy people. According to recent data, the probability of a person living in these conditions for 30 years or more is 2/3. Calculate the probability that after 30 years:

$$P(x) = \frac{n!}{(n-x)!\,x!}\, p^x q^{n-x}$$

$n = 5$

$p = 2/3$

$q = 1/3$

- 1. All five people are still living. $x = 5$
- 2. At least three people are still living. $x \geq 3$
- 3. Exactly two people are still living. $x = 2$

$$P(x=5) = \frac{5!}{0!\,5!}\left(\frac{2}{3}\right)^5 \left(\frac{1}{3}\right)^0 \Rightarrow \frac{32}{243}$$

$$P(x=2) = \frac{5!}{3!\,2!}\left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^3 = \frac{40}{243}$$

$$P(x \geq 3) = P(x=3) + P(x=4) + P(x=5)$$

$$\Rightarrow \frac{5!}{2!\,3!}\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right)^2 + \frac{5!}{1!\,4!}\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^1 + \frac{32}{243}$$

$$\Rightarrow \frac{80}{243} + \frac{80}{243} + \frac{32}{243} = \frac{192}{243}$$

Let’s say that 80% of all business start-ups in the IT industry report that they generate a profit in their first year.

If a sample of 10 new IT business start-ups is selected, find the probability that exactly seven will generate a profit in their first year.

$p = 0.8 \mid q = 0.2$

$n = 10$

$X = 7$

$$P(X=7) \Rightarrow \frac{10!}{3!\,7!}(0.8)^7(0.2)^3$$

$$120 = \frac{10 \times 9 \times 8}{3!}$$

$$\Rightarrow 120 * 0.008 * 0.2097152$$

$$\Rightarrow 0.2013$$

20.13%

According to the latest police reports, 80% of all petty crimes are unresolved, and in your town, at least three of such petty crimes are committed. The three crimes are all independent of each other. From the given data, what is the probability that one of the three crimes will be resolved?

$p = 0.2 \mid q = 0.8$

$n = 3$

$x = 1$

$$P(x=1) = \frac{3!}{2!\,1!}(0.2)^1(0.8)^2$$

$$= 3 \times 0.2 \times 0.64 = 0.384$$

38.4 %

CONTINUOUS DISTRIBUTION – NORMAL DISTRIBUTION

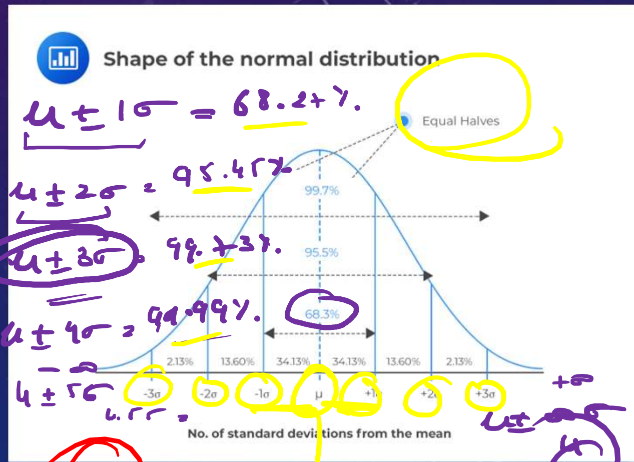
- **Normal distribution** represents the behaviour of most of the situations in the universe (That is why it's called a "normal" distribution.)
- The large sum of (small) random variables often turns out to be normally distributed, contributing to its widespread application.

Characteristics:

- The mean, median and mode of the distribution coincide.
- The curve of the distribution is bell-shaped and symmetrical about the line x=μ.
- The total area under the curve is 1.
- Exactly half of the values are to the left of the center and the other half to the right.

Normal Distribution

Bell Curve



- A normal distribution is highly different from Binomial Distribution. However, if the number of trials approaches infinity then the shapes will be quite similar.

AUC = 1

50% 50%

median, mode

$> \mu + 3\sigma = UL$

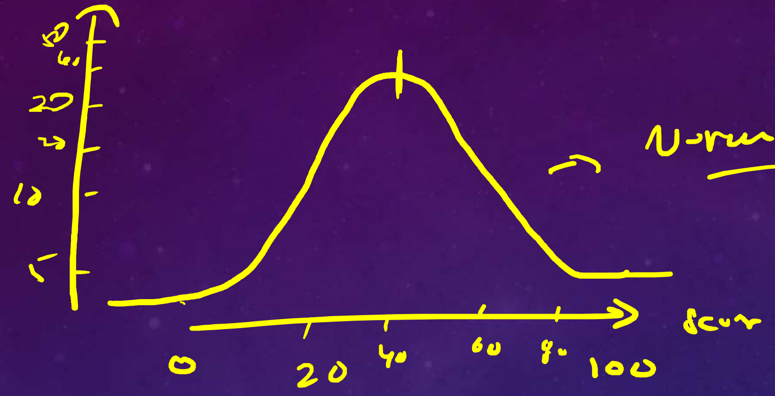
$< \mu - 3\sigma = LL$

→ Outlier

Z-method

Test → /100

Prob



→ Normal

100 students → 5%.

>90

<20

10%

50%

40-60

Approximated →

O_2 N_2

μ = Population mean

$\bar{x}$ = sample mean

σ = std. dev

σ^2 = variance

r = linear corr

ρ = non-linear corr

NORMAL DISTRIBUTION – CENTRAL LIMIT THEOREM

$\sigma_{\bar{x}} = \sigma_p/\sqrt{n}$

converted ANY → Every other type of dist

- The central limit theorem in statistics states that, given a sufficiently large sample size, the sampling distribution of the mean for a variable will approximate a normal distribution regardless of that variable's distribution in the population.
- (In Layman's term – even if the data is not normally distributed, the mean of the distribution are normal distribution provided the sample size is large).
- **Why is it useful**: We can use mean's Normal Distributions

to make confidence Intervals, perform hypothesis testing.

1000 sample → 1000 $\bar{x}$ → $\bar{\bar{x}}$

1 sample → 1 $\bar{x}$

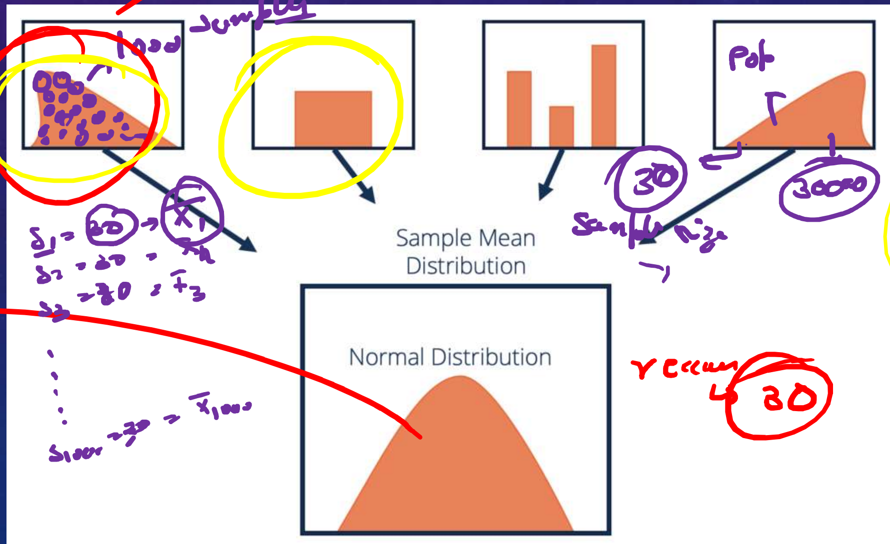
Sample $\bar{x}$ Distribution

Normal Distribution

3 → min

Statistical studies

1 → incident
2 → collect
3 → Pattern



recommend 30

$\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$, …, $\bar{x}_{1000}$

1000

↳ $\bar{\bar{x}}$

↳ $\sigma_{\bar{x}} = \frac{\sigma_p}{\sqrt{n}}$

NORMAL DISTRIBUTION – CENTRAL LIMIT THEOREM

100 1 M

Central Limit Theorem (Key Takeaways)

- The central limit theorem (CLT) states that the distribution of sample means approximates a normal distribution as the sample size gets larger.
- Sample sizes equal to or greater than 30 are considered sufficient for the CLT to hold.
- A key aspect of CLT is that the average of the sample means and standard deviations will equal the population mean and standard deviation.
- A sufficiently large sample size can predict the characteristics of a population accurately.
- Expectation of Sample Mean as a random variable = Population Mean. Symbolically $E(\bar{X}) = \mu$.
- Standard Deviation $(\bar{X}) = \sigma / \sqrt{n}$ (where σ is standard deviation and n is sample size).

$\sigma_{\bar{x}} = \sigma_p / \sqrt{N}$

$\mu = \bar{x}$

NORMAL DISTRIBUTION – EMPIRICAL RULE

The empirical rule states that for a normal distribution, nearly all of the data will fall within three standard deviations of the mean. The empirical rule can be broken down into three parts:

- 68% of data falls within the first standard deviation from the mean.
- 95% fall within two standard deviations.
- 99.7% fall within three standard deviations.

The rule is also called the **68-95-99 7 Rule** or the **Three Sigma Rule**.