A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

CSE DATA SCIENCE

Semester : **V**   Subject **Statistics for AI&DS**   Academic Year: 20**23** 2024

NORMAL DISTRIBUTION :

(Major Distribution).

34% 34%

$M-3\sigma$ $M-2\sigma$ $M-\sigma$ M $M+\sigma$ $M+2\sigma$ $M+3\sigma$

← 68% within →
1 standard
Deviation.

← 95% within 2 SD →

← 99.7% of the data within →
3 SD of the Mean
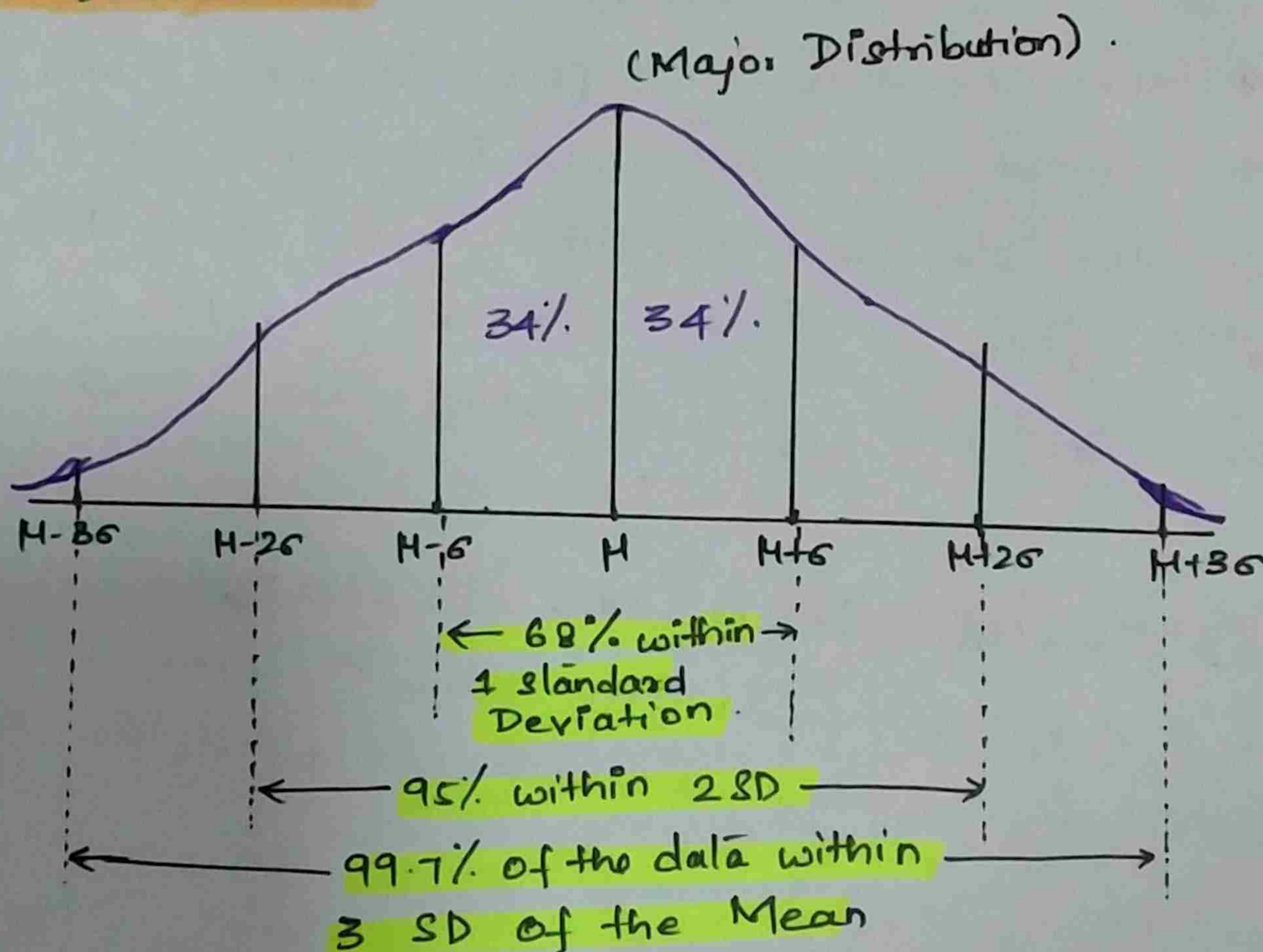
* The Normal Distribution is called as Gaussian Distribution.

 * It is also called as bell-curve.

 * The major distribution lies in Mean.
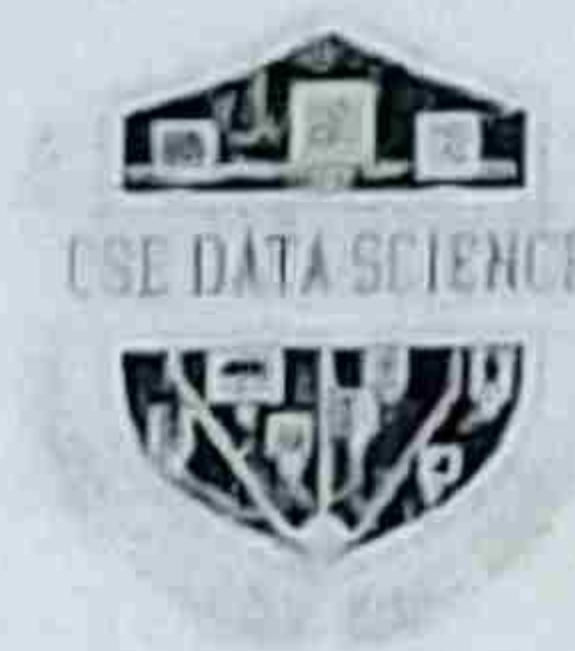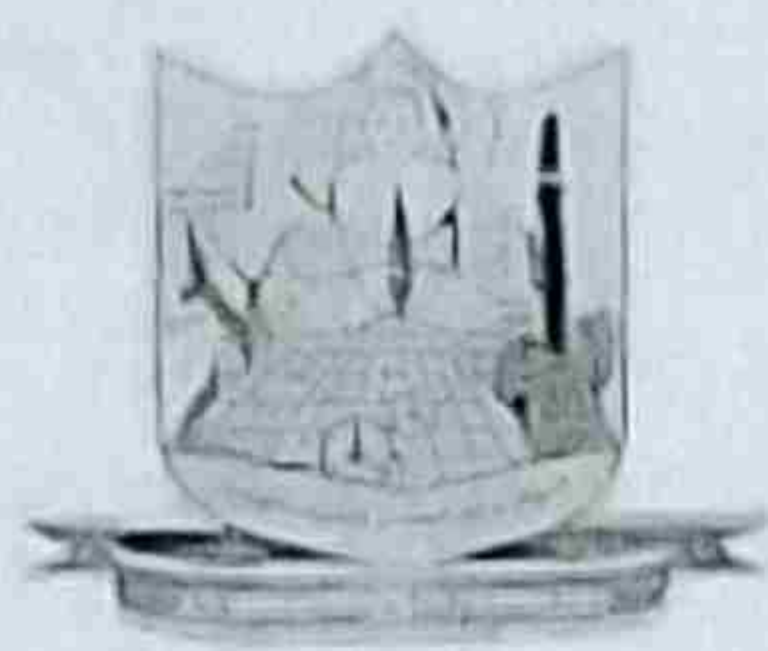
 * In normal distribution   Mean = Median = Mode.

 * The Normal Curve is symmetric about the Mean.

 * The Mean parameter serves as a location parameter.

 * The Standard Deviation is a scale parameter
    → Different curve gives different scale.

Semester : ⇧   Subject : Statistics for AI&DS   Academic Year: 2023 2024

* From the diagram (M+σ) is one standard deviation away from the mean in right side. M+2σ is two standard deviation away from the right side. M+3σ is three standard deviation away from the right side. It is called as influence point.

STANDARD NORMAL FORM AND QQ-PLOTS :

* A Q-Q plot, or quantile - quantile plot is a scatter plot created by plotting two sets of quintiles against one - another.

* If both sets of quantiles come from the some distribution, we see that the points forming a line is roughly a straight line.

* We mention here an example of a normal Q-Q plot when both sets of quantiles truely comes from normal distribution.
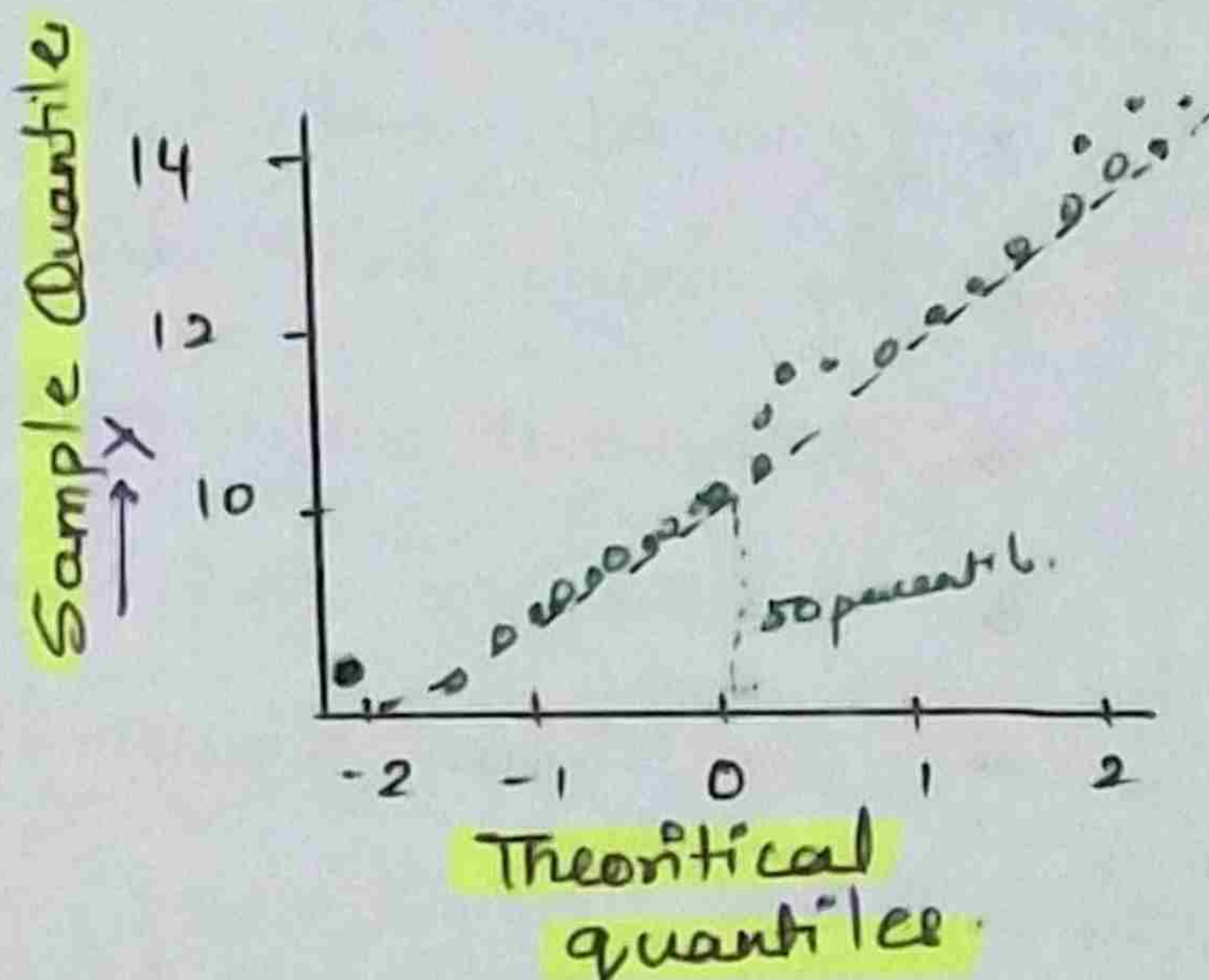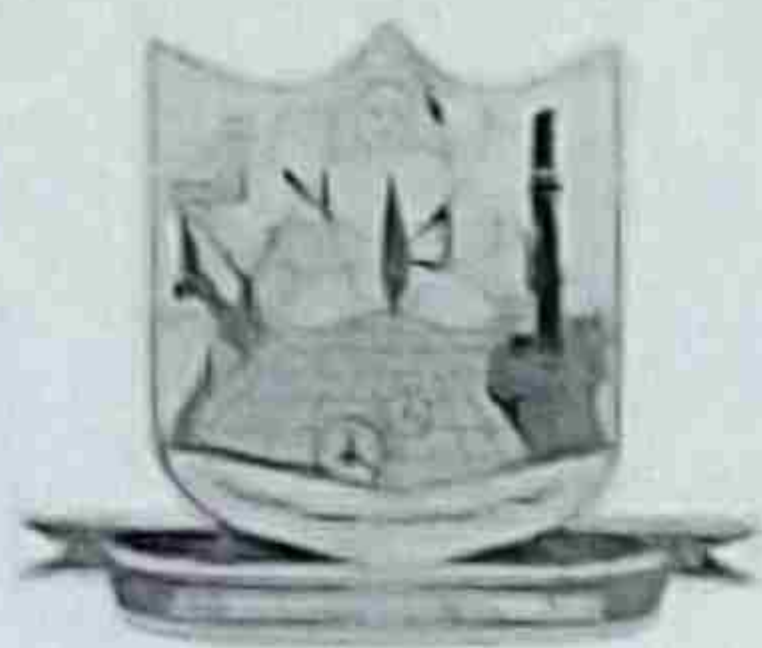

Fig. 11 For Normal Dist.
Example 1 :→ QQ-Plot.

PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

CSE DATA SCIENCE

Semester ____1____   Subject **Statistics for AIPDS**   Academic Year 20 23 20 24.

* Quantiles are often referred to as 'percentiles'. These are points in the data set below which a certain proportion of data fall.
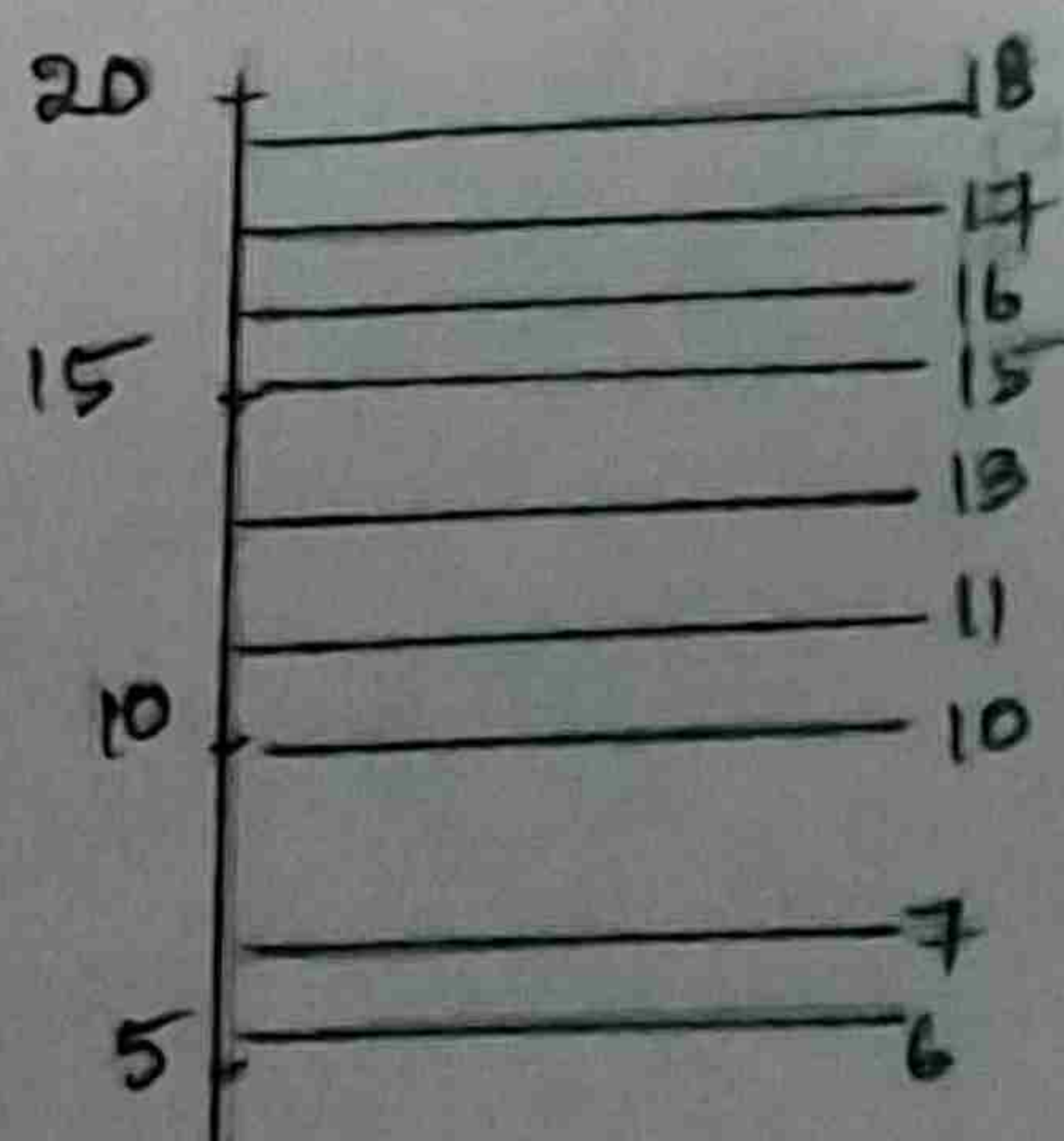
* For example, consider the bell-shaped standard normal distribution with mean of 0.

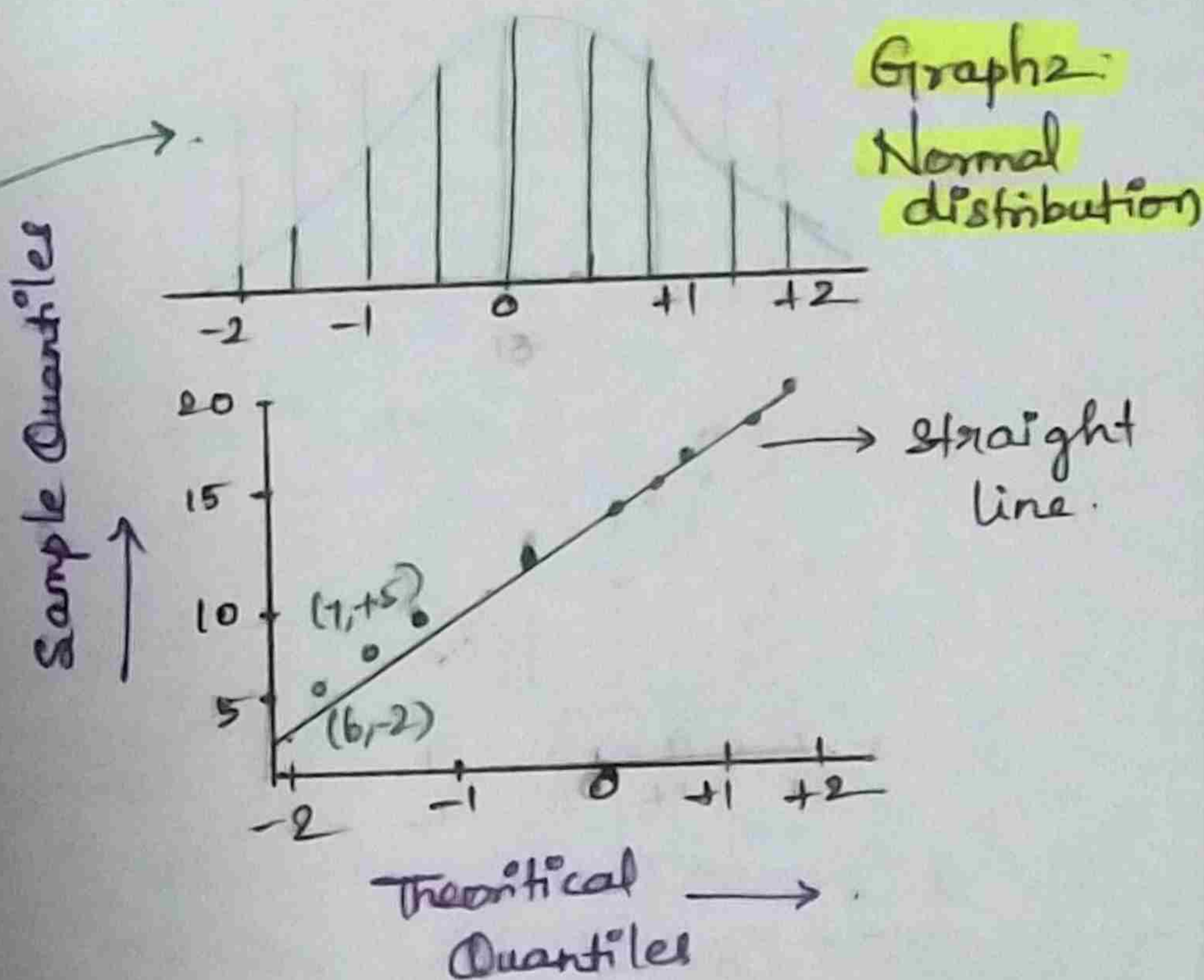* The 50th percentile is 0.

* Half of the data lie below 0.

* The 95% of data lie below 1.64 from the given example.

* Since the data is normally distributed, the QQ-plot generates a straight line.
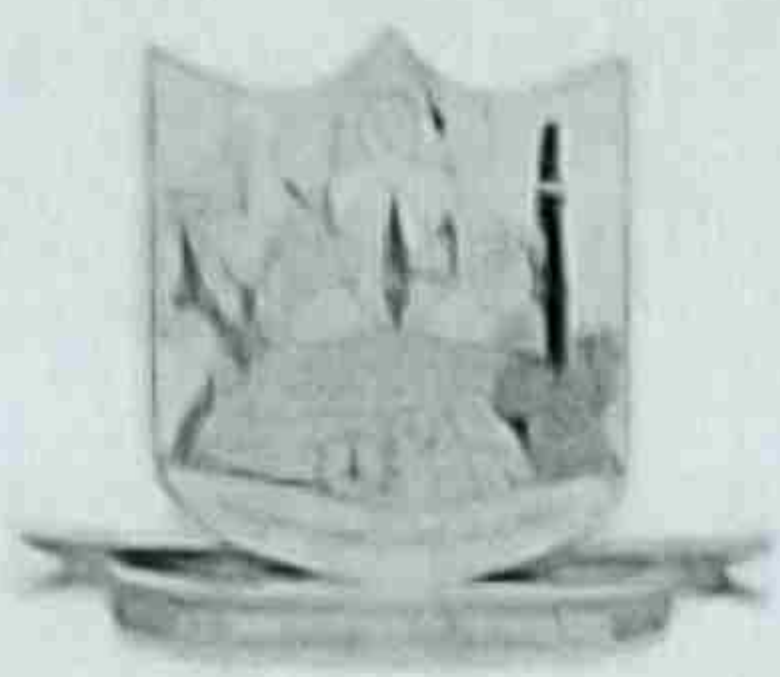

Graph1:
Sample Data

Sample Quantiles

Graph2:
Normal distribution

→ straight line.

(1,+5)

(6,-2)

Theoritical
Quantiles

QQ-Plot – Graph3.

Example2:

PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science

CSE DATA SCIENCE

Semester ___I___   Subject _Statistics for AIDS_   Academic Year: 2023 2024

  *The example 2 clearly describes how the QQ plot is generated.

   * Y axis takes the sample data and x-axis takes the normal Distribution Data

   * They are together ploted in QQ-Plot If it generates a straight line then it is normal distribution.

<mark>LONG TAILED DISTRIBUTION</mark>

   *In statistics, a long tail of some distributions of numbers is the portion of the distribution having many occurences, and they are far from the "head" or central part of the distribution.

<mark>Long-Tail Distribution.</mark>
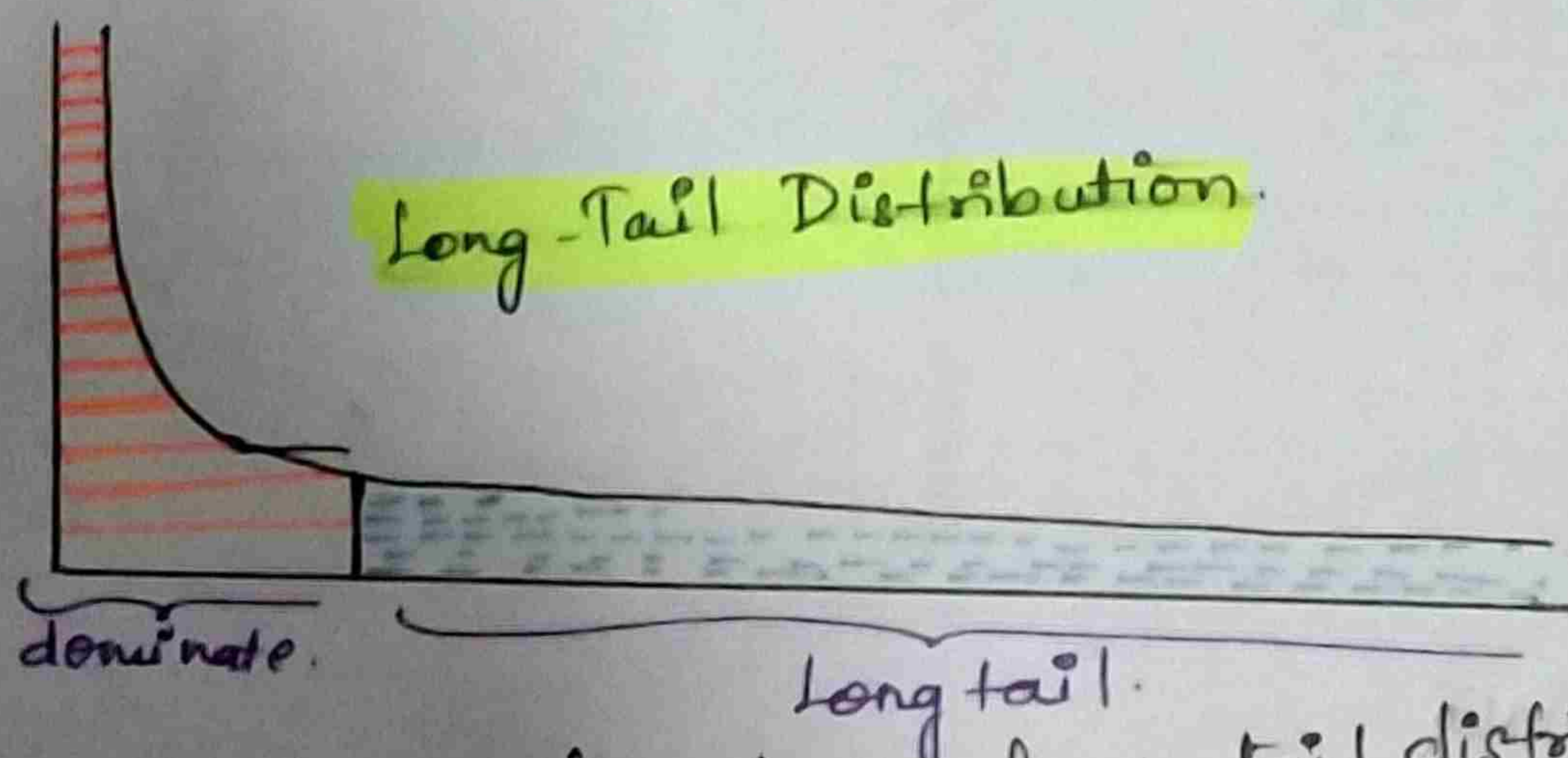
dominate.

Long tail.

   *.The above graph shows long-tail distribution. To the right (dotted part) is the long tail and to the left (lineal segment) are the few that dominate