STATISTICS FOR DATA SCIENCE

Normal Distribution

Prof. Uma D

Department of Computer Science and Engineering

STATISTICS FOR DATA SCIENCE

Normal Distribution

Prof. Uma D

STATISTICS FOR DATA SCIENCE

Topics to be covered...

- Normal Distribution
- Probability Density Function
- Standard Normal Distribution

STATISTICS FOR DATA SCIENCE

Continuous Probability Distribution

- There are many different types of continuous random variables

- We try to pick a model that
  - Fits the data well
  - Allows us to make the best possible inferences using the data.

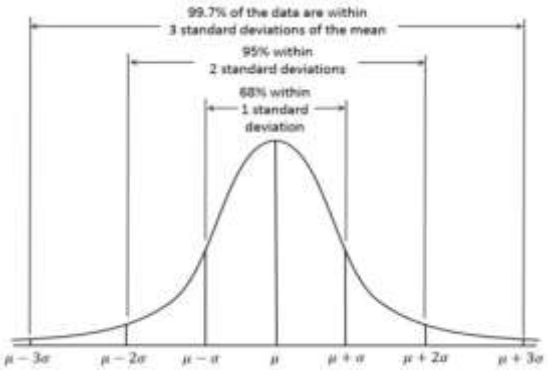
- One important continuous random variable is the normal random variable and the corresponding distribution is Normal distribution.

- Most commonly used continuous distribution is Normal Distribution.

STATISTICS FOR DATA SCIENCE

Normal Distribution

- German mathematician and scientist
- Contributions in many fields of mathematics and science
- Referred to as the “Prince of Mathematicians”
- Credited with the use of the probability distribution now known as the normal or Gaussian distribution (bell curve)



Carl Friedrich Gauss
(1777–1855)

Source:https://en.wikipedia.org/wiki/Carl_Friedrich_Gauss

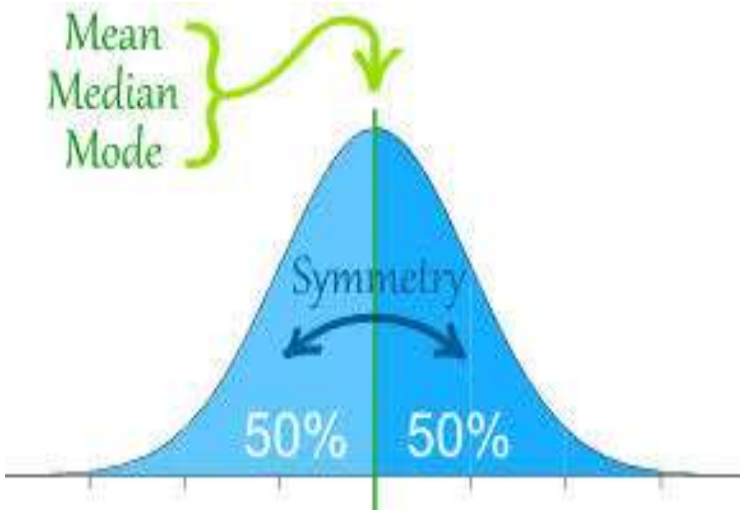
STATISTICS FOR DATA SCIENCE

Normal Distribution

Extremely important continuous probability distribution.

Rises frequently in theory and practice.

We say that a random variable is normally distributed with mean μ and standard deviation σ if the **probability density function** is given by



Source Image: dietassessmentprimer.cancer.gov

Probability Density Function

The formula that generates the normal probability distribution is :

Examples of Normal Distribution

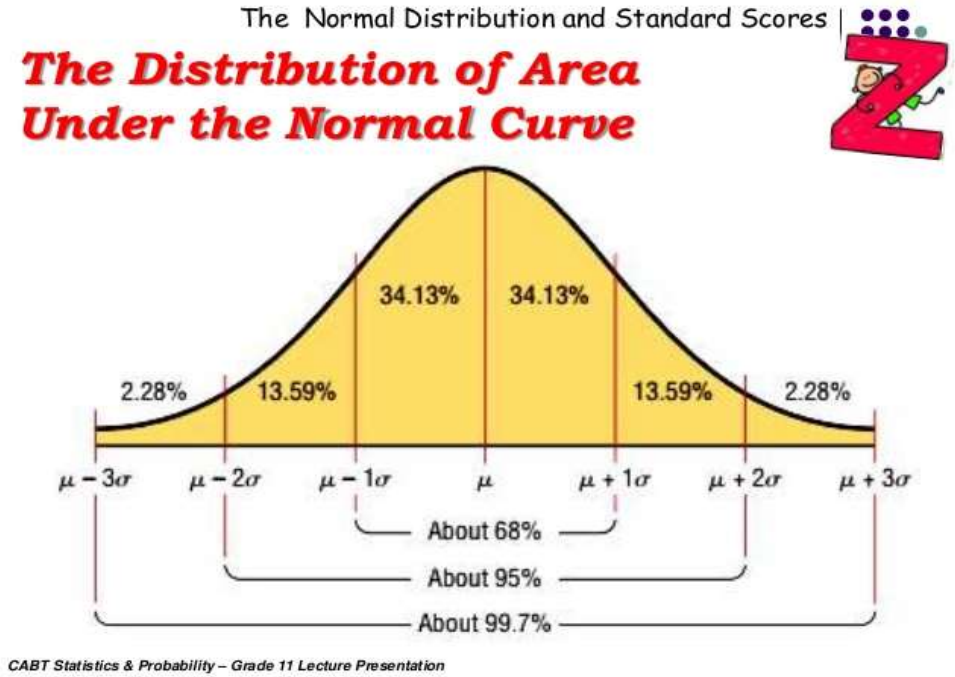
- Heights of People
- Test Scores
- Errors in measurements
- Blood Pressure
- Size of things produced by machines

Normal Distribution

Why to know Standard Deviation?

Any value is

- **Likely** to be within **1** standard deviation of the mean.
- **Very Likely** to be within **2** standard deviations.
- **Almost certainly** within **3** standard deviations.



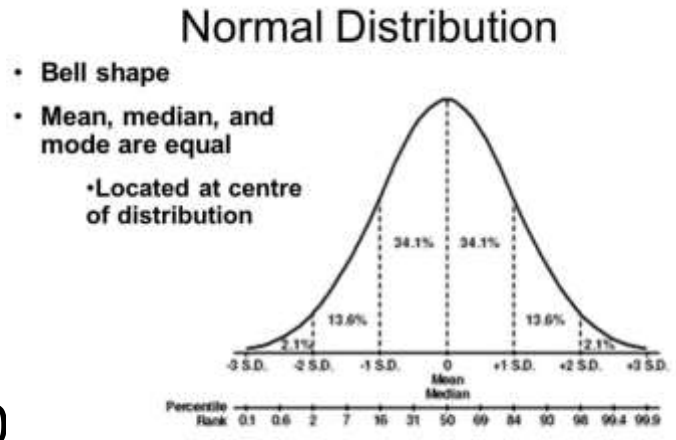
Source Image: www.slideshare.net

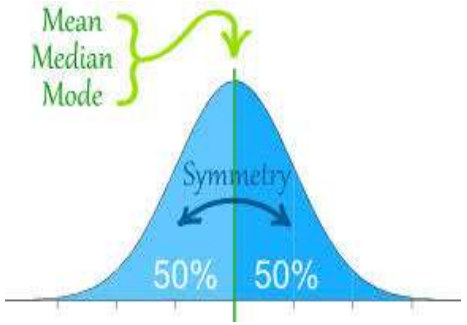
STATISTICS FOR DATA SCIENCE

Normal Distribution

The Properties of a Normal Distribution

- Mean=Median=Mode
- Symmetry about the center(mean) μ.
- 50% of the values less than the mean and 50 greater than the mean.
- Changing **μ *shifts*** the ***distribution*** **left or right**.
- Changing **σ *increases or decreases*** the **spread.**





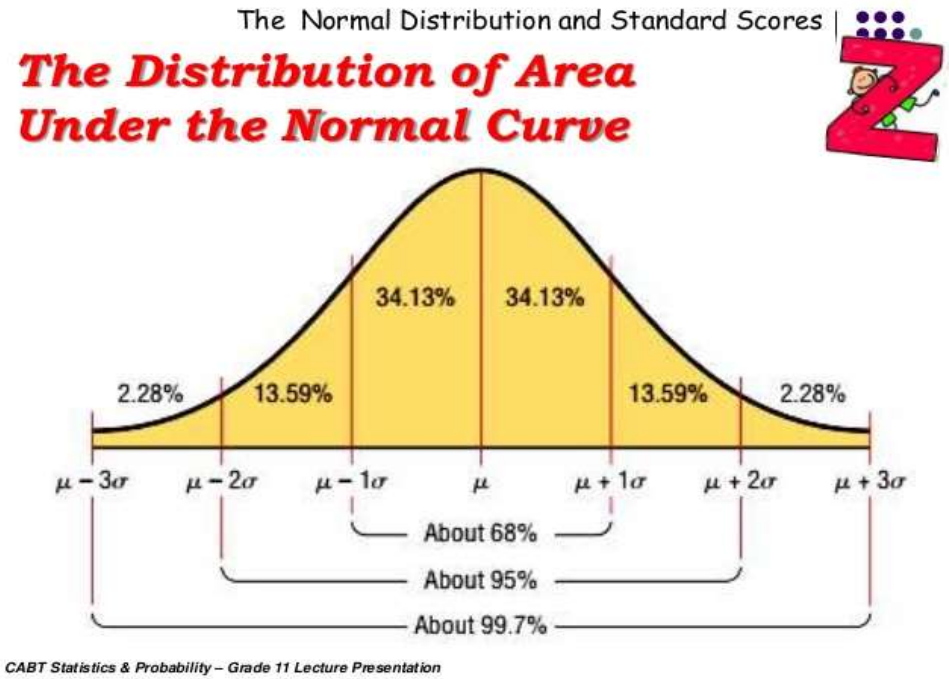
Source Image: dietassessmentprimer.cancer.gov, www.google.com

STATISTICS FOR DATA SCIENCE

Normal Distribution

The Properties of a Normal Distribution

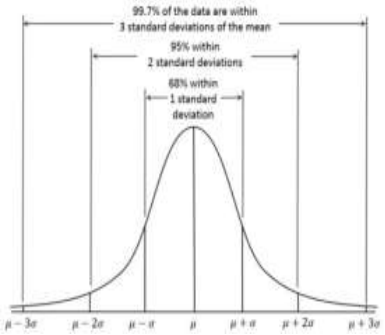
- Approximately **68%** of the area is within **1** standard deviation.
- Approximately **95%** of the area is within **2** standard deviations.
- Approximately **99.7%** of the area is within **3** standard deviations.



Source Image: www.slideshare.net

Normal Distribution

How good is rule for real data?



Suppose SAT scores roughly follow a normal distribution(with range restricted to 200-800), and the average math SAT is 500 with a standard deviation of 50, then:

- 68% of students will have scores between 450 and 550.
- 95% will be between 400 and 600.
- 99.7% will be between 350 and 650.

What if you wanted to know the math SAT score corresponding to the 95th percentile(=95% of students are lower)?

Source Image: www.slideshare.net

The Standard Normal(Z): "Universal Currency"

The formula for the Standardized Probability Density Function is

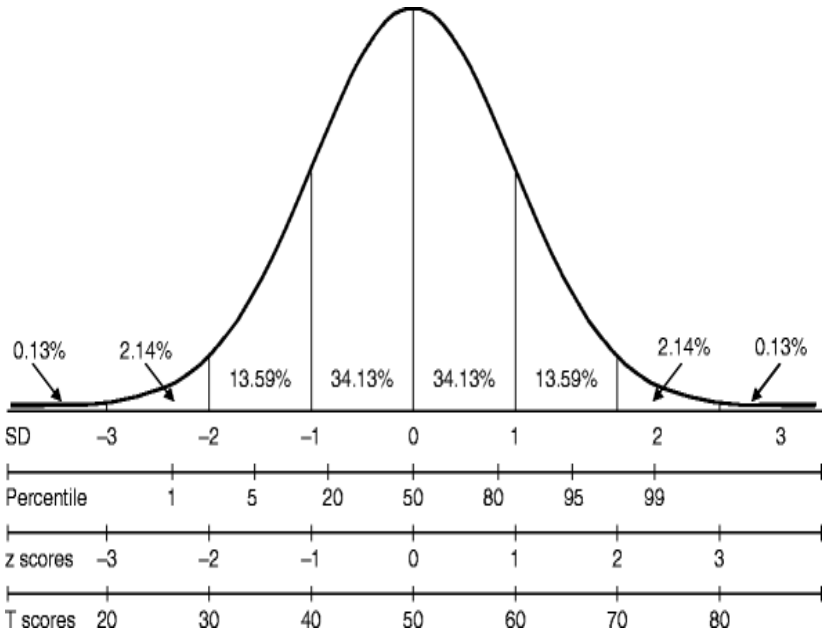
STATISTICS FOR DATA SCIENCE

Standard Normal Distribution

- Is a Normal distribution with mean 0 and variance 1.
- Random Variable that has standard normal distribution is referred using letter Z.

$$Z \sim N(0, 1)$$

- Probabilities associated with Normal Variates can be calculated by using transformations to the Standard Normal Variate (z) – using z-table.



Source Image: SpringerLink

STATISTICS FOR DATA SCIENCE

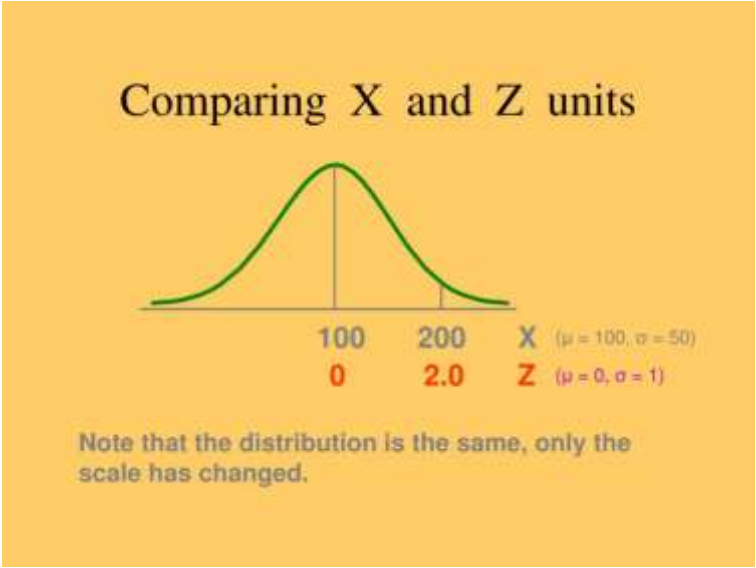
Standardizing Normally Distributed Random Variables

We can convert a Random Variable X having a Normal distribution with any mean and Standard deviation in to the Random variable that has a Standard Normal Distribution.

$$X \sim N(\mu, \sigma^2)$$

Standardizing X : using a basic linear transformation:

$$z = (x - \mu) / \sigma$$



Source Image: www.slideshare.net

THANK YOU

Prof. Uma D

Department of Computer Science and Engineering