Start Up Screen

Introduction to Business Statistics – Part 2

Take Aways

Canopus Business Management Group - Chennai – India | +91 44 42851080

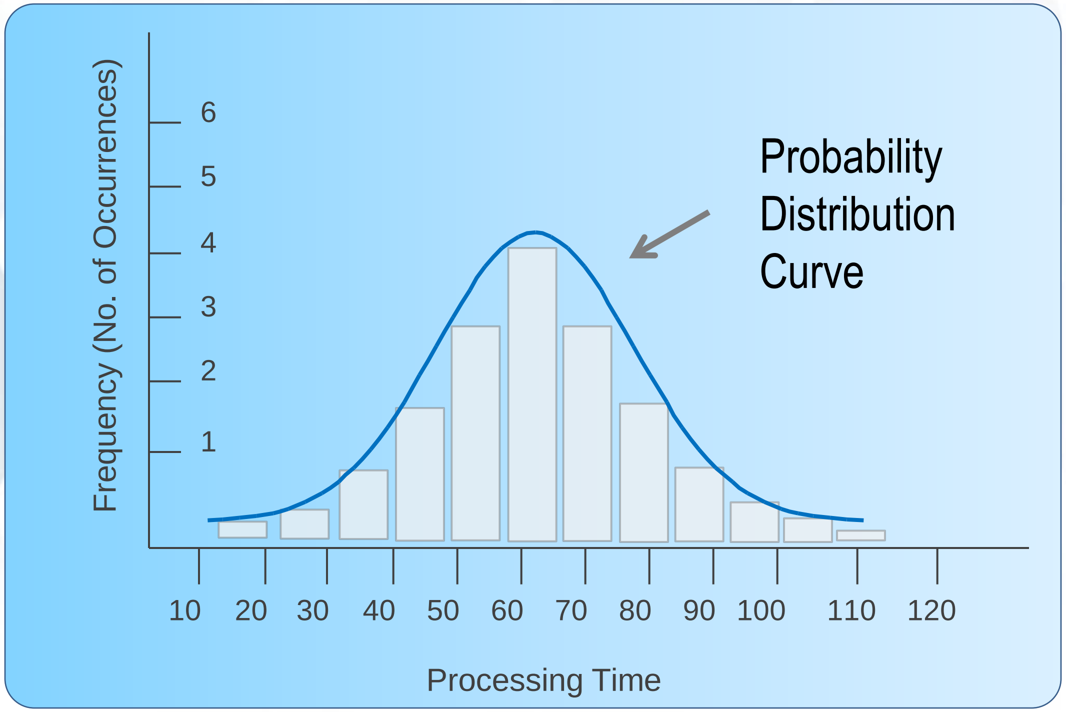
Definition of Probability Distribution

Canopus Business Management Group

- The relation between a given descriptive measure & its probability is shown using a probability distribution
- Graphically represented by a Histogram
- Mathematically by a formula
- Indicates that there is a probability associated to each descriptive measure

For example, the processing time in a bakery or an other process can be represented by a probability distribution

Illustration of Probability Distribution



- The probability distribution is obtained by drawing a smooth curve over the histogram
- Curve fitting method is used here
- Can be created in statistical software

Types of Distribution

There are various types of distributions such as:

- Normal Distribution
- Binomial Distribution
- Weibull Distribution
- Poisson Distribution
- Chi-Square Distribution
- Student's t-Distribution
- F Distribution

Each distribution has specific characteristics that are different from the others

Importance of Distribution

Certain processes & their metrics are known to follow pre-defined distributions

Example:

- A natural phenomenon like human population, height, etc follow normal distribution
- At the same time, arrival rates of people in ticketing counters follow Poisson distribution

Importance of Distribution

Canopus Business Management Group

Knowledge of various distributions can help us in the following:

- Study process performance
- Fix performance issues quickly
- Plan for capacity & performance more effectively
- Predict process performance

Definition of Mean

Canopus Business Management Group

- Arithmetic average of a set of values
- Reflects the influence of all values
- Extreme values can gravely influence mean
- Represented by the symbols "μ" & "x̄" (X-bar)

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{X_1 + X_2 + \ldots + X_N}{N}$$

Population

$$\overline{X} = \hat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Sample

Definition of Median

Canopus Business Management Group

- Reflects the 50% rank
- Centre number after a set of numbers have been sorted from low to high
- Is never affected by extreme values or outliers
- For even number of data points, use average of two middle numbers

Definition of Mode

- The most frequently occurring value in a data set
- It is possible to have more than one mode

Relationship Between Mean Median & Mode

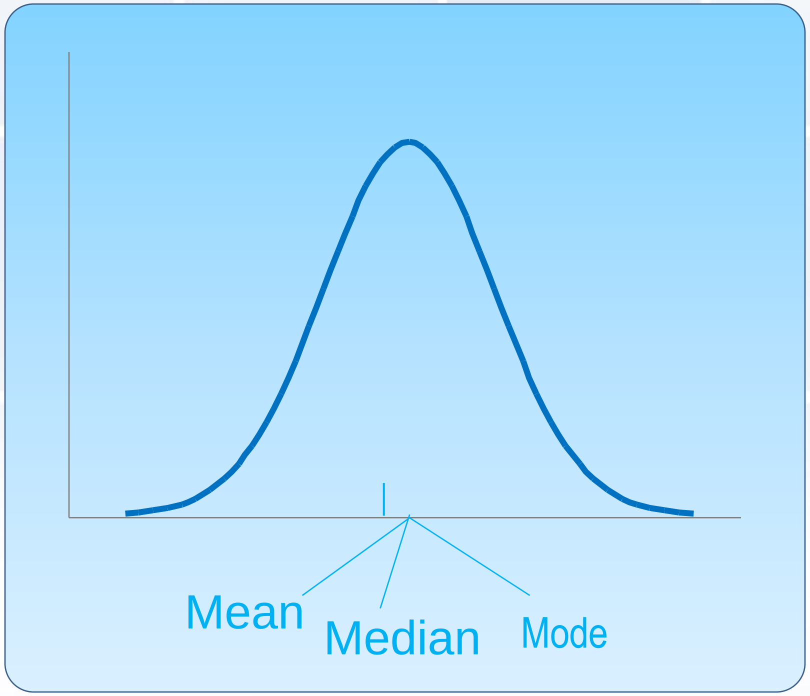
Canopus Business Management Group

- Let's understand relationship between Mean, Median & Mode using following scenarios:
  - Un-Skewed Distribution
  - Skewed Distributions
    - Right-Skewed Distribution
    - Left-Skewed Distribution
  - Bi-Modal Distribution

Definition of Un-Skewed Distribution

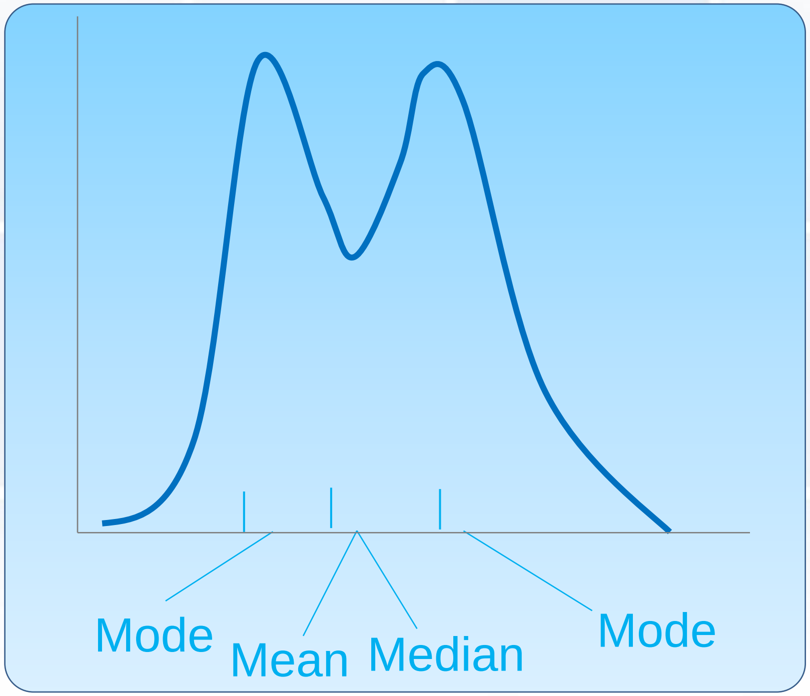
Canopus Business Management Group

- Evenly distributed points surrounding a central value or location
- Mean, median, and mode are all close to the central location
- Coefficient of skewness, a measure of skewness, is close to zero



Definition of Bi-Modal Distribution

- The data actually reflects two distinct processes with different centers
- Either data is mixed up or process has two distinct behaviour

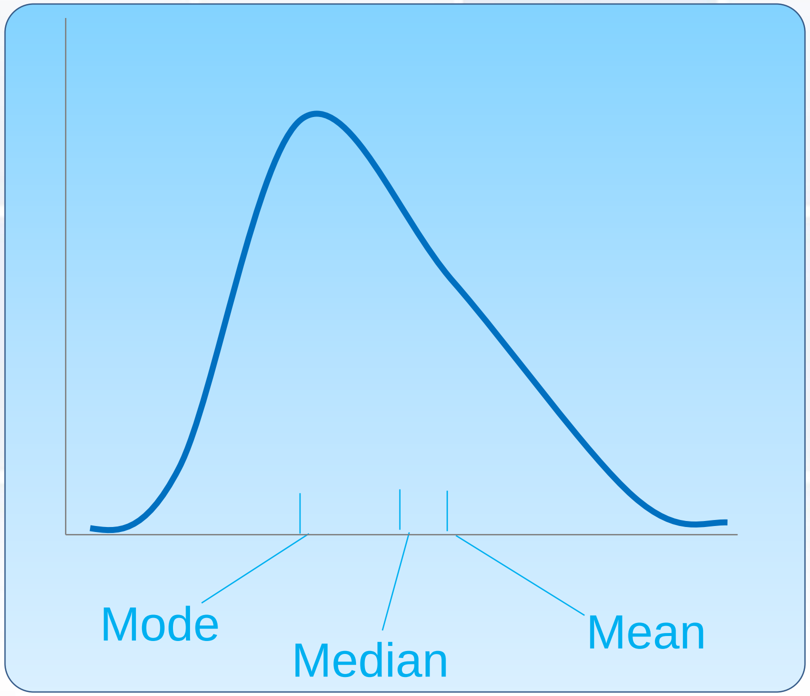


Example:

- Surface finish of components obtained from two different machines
- Processing times of two employees with different levels of experience

Definition of Right-Skewed Distribution

- Aka Positive distribution
- Distribution of data leans more towards the right side
- Caused by a process with a lower boundary

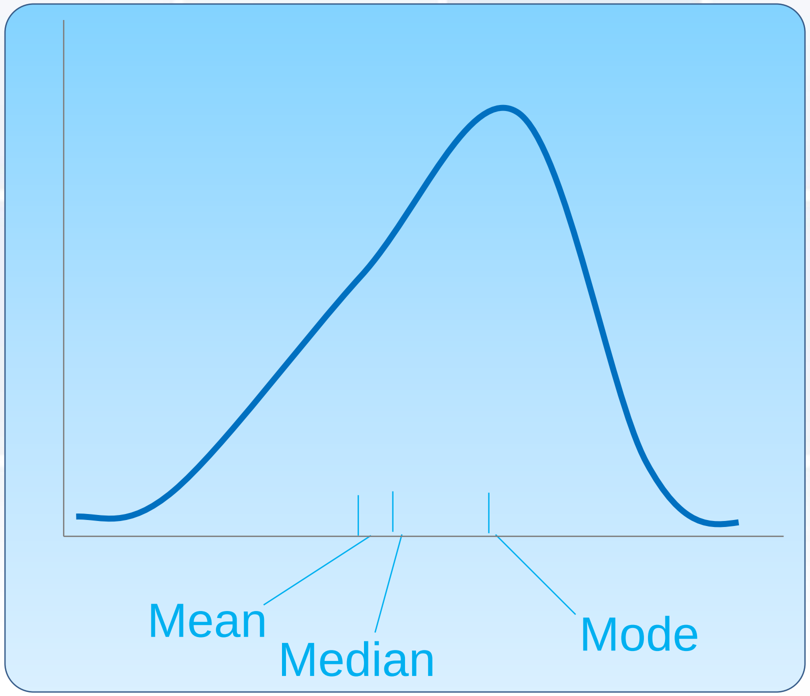


Example:

TAT or Lead time of processes such as the time taken for us to wait in a queue and buy a ticket, at the cinemas

Definition of Left-Skewed Distribution

- Aka Negative distribution
- Distribution of the data exists more on the right side including most common values
- Data extends more to the left side due to an upper specification limit



Example :

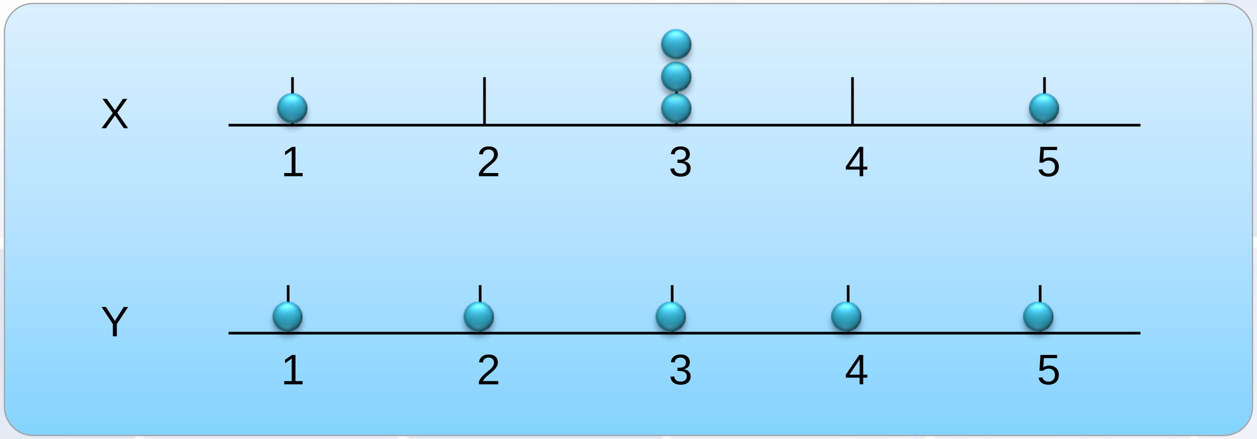
The tenure or life of customers in a given telecom operator's network

Definition of Range

- Simplest measure of dispersion
- Shows the difference between the maximum and minimum values present in a set
- Usually represented by the symbol “R”

Limitations of Range

Consider the following scenarios:



- Both process X and Y, have same range (4 = 5 - 1)
- Range is not a very robust measure of dispersion
- An easy & quick way to assess the dispersion
- Mostly used for screening purpose

Definition of Span

Canopus Business Management Group

- A measure of dispersion similar to 'range' in which extreme 5% or 1% of the observations on either side are omitted
- It is the difference between the highest & lowest normal value being analyzed
- It is more robust & eradicates the impact of freak values

Example of Span

Age of employees in a team

28 32 29 15 34 37 47

Extreme Values: 15, 47

Span = 37 - 28 = 9

37 = Highest Normal Value

28 = Lowest Normal Value

Definition of Variance

- Average squared difference (deviation) of each data point from the mean
- Numerically, the variance is the square of the standard deviation
- Represented by the symbols "σ^2" or "s^2"

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

Population

$$s^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}$$

Sample

Definition of Standard Deviation

Canopus Business Management Group

- A statistical measure of the variation from the mean in a distribution or set of data
- The square root of the variance provides a measure of the average difference (deviation) of each data point from the mean
- Represented by the symbols "σ" or "s"

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}}$$

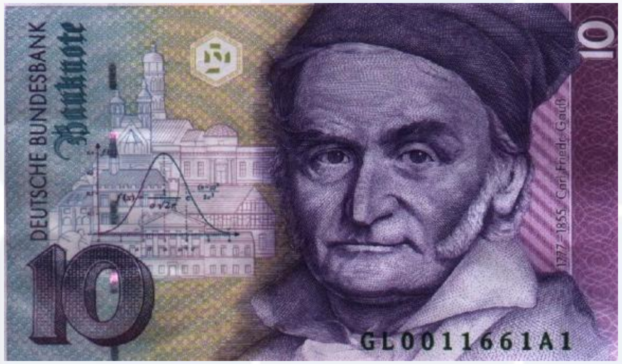
Population

$$\hat{\sigma} = s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$$

Sample

Definition of a Normal Distribution

- Aka Gaussian Distribution or Bell Shaped Curve
- Defined by two parameters - location and scale:
  - Mean/average(μ)
  - Variance /standard deviation squared (σ^2)
- Natural phenomenon follow Normal Distribution



Example:

- Height of a group of students of same age &race
- Performance of employees of a given function within an organization
- Dimensions of machined components within lower & upper specification limits

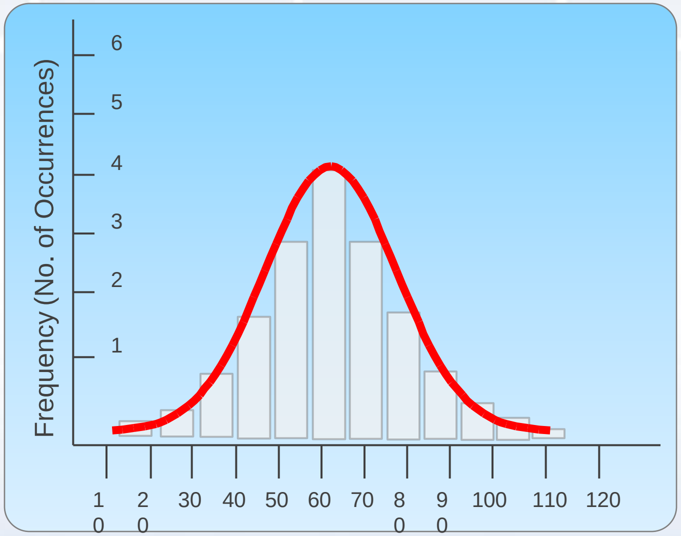
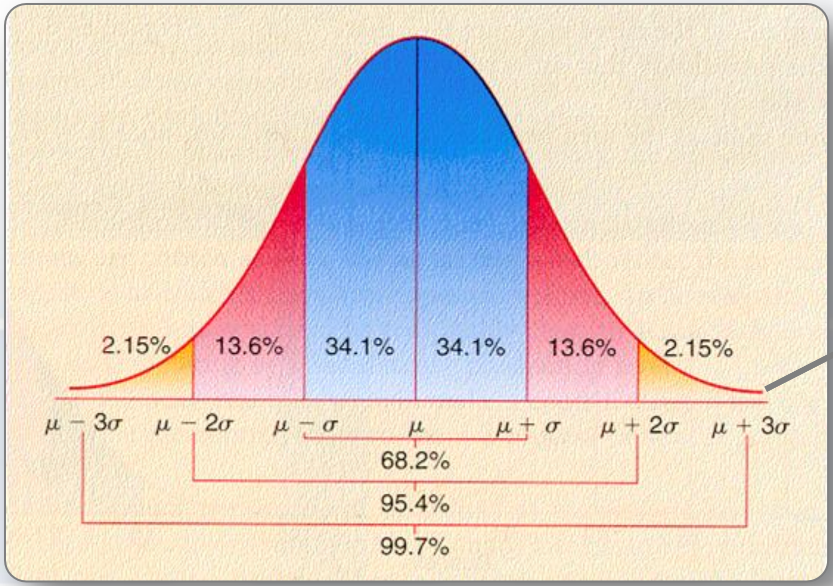


Illustration of a Normal Distribution

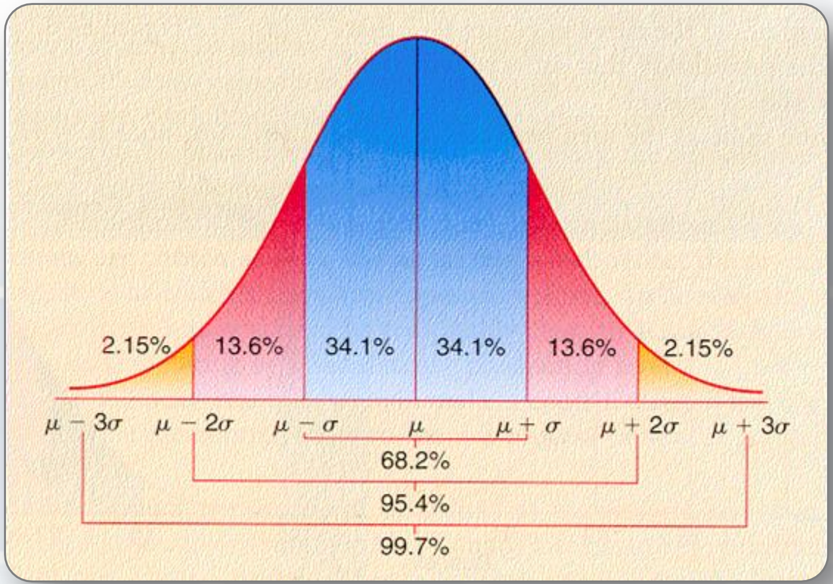


Percentage of Data getting covered under the curve

- Probability distribution curve above, is from a process which is normally distributed, with a mean (μ) and standard deviation (σ)
- Mean of the process occurs at the center of distribution
- Probability of occurrence :
  - Represented by the curve
  - Represented in percentages across the distribution
  - Diminishes on both sides

Illustration of a Normal Distribution

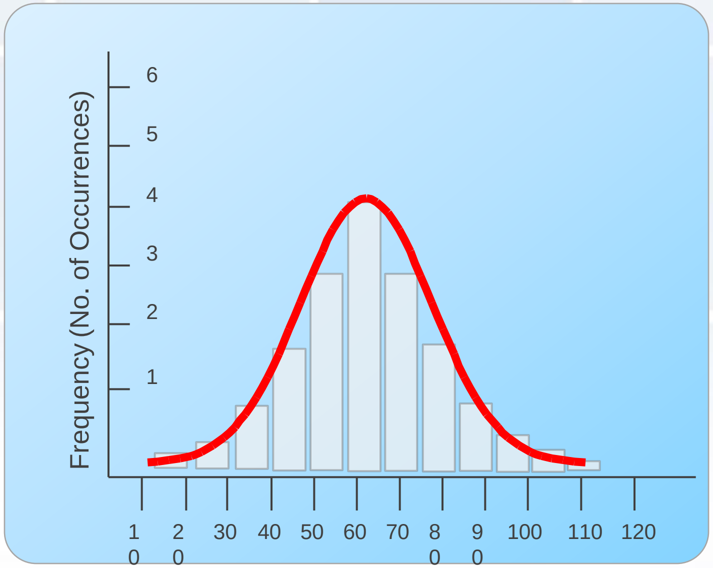
Canopus Business Management Group



- Probability shown for multiples of standard deviation
- For example, between one standard deviation (1σ) on either side of mean, probability is 68.2%
- 68 data points of out 100 fall between the one standard deviation limit on either side of mean

Properties of a Normal Distribution

- Normal Curve is asymptotic, i.e., it never touches X axis on both sides
- Mean, Median and Mode nearly coincide
- Curve is symmetrical, i.e., it can be divided in half with equal pieces falling on either side of the mean
- Peak of the curve represents the center of the process
- Area under curve virtually represents 100% of data



Outlier

Canopus Business Management Group

- Outliers are extreme values and occur in few numbers
- They indicate that a process has ‘special’ causes of variation
- In other-words it is not natural that you will find an outlier in a Normal Distribution

Example: If a vendor’s delivery of components to their customer’s warehouse is around 7.00am every morning

Delivery Timings :

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
07.00	06.50	06.55	07.10	07.05	07.15	10.00

Outlier

Application of Normal Distribution

Can be used for computing a value that would occur at an given probability

Example:

- If we want to be 95 % sure, what will be the thickness of a metal component that is being machined
- If we want to offer 95 % service level to our customers, then what should be process lead time

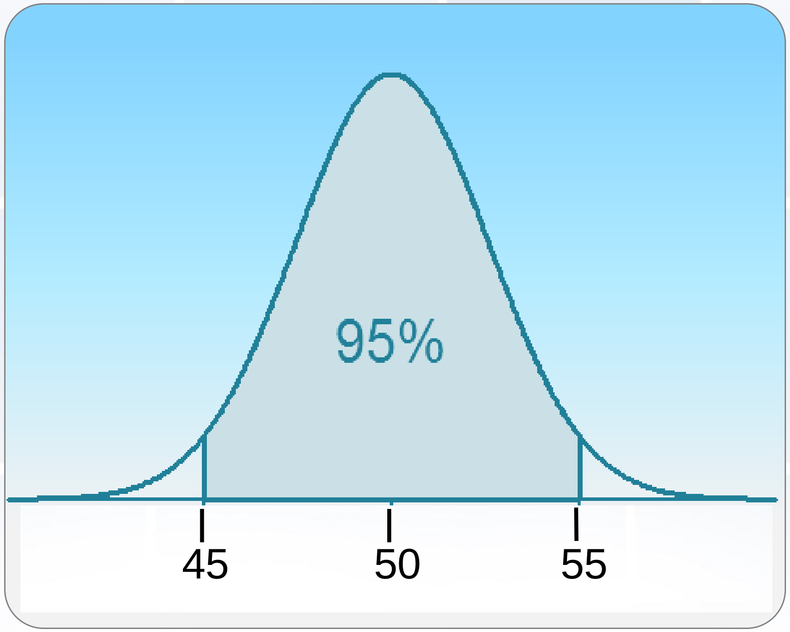
Use normal distribution to draw inferences about data of other distributions using Central Limit Theorem

Definition of Confidence Level

- Degree of certainty in estimating the characteristics of population
- Probability percentage indicates how frequently an expected range of the population will exhibit certain characteristics
- Expressed as a percentage or proportion
- Set at different percentages based on the priority of the tests
- Most analyses are conducted at a defined level of confidence which is usually 95% or (.95)
- Safety & health situations such as automotive brakes, space missions, surgeries, etc., use 99% confidence

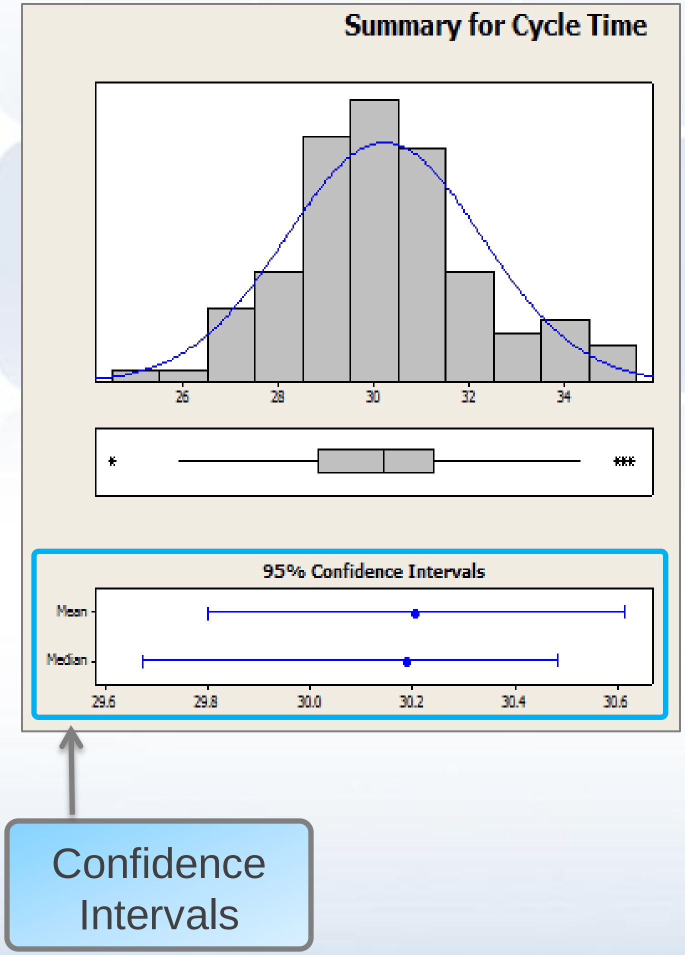
Definition of Confidence Intervals

- Confidence Intervals are nothing but upper and lower limits of a band within which we find the actual value (true mean)
- The limits are established at a given probability or certainty expressed in %



- Most of the time, we calculate 95% confidence intervals (CIs) for a parameter (occasionally 90% or 99%)
- In this illustration the mean will lie anywhere between 45 and 55

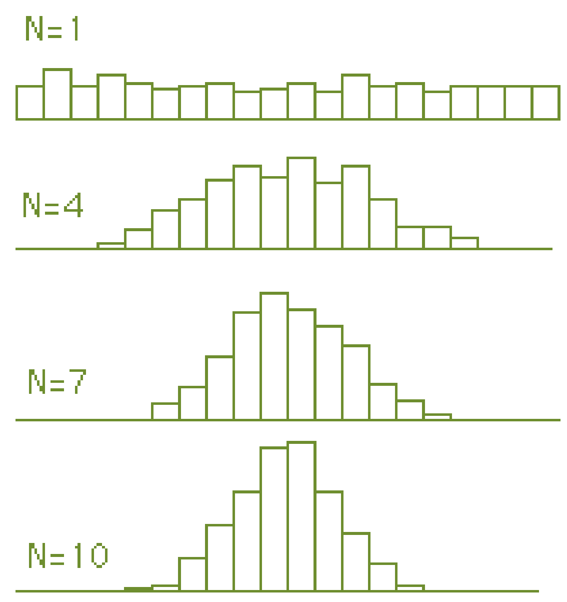
Illustration of Confidence Interval



- It implies that the population mean can appear anywhere between two determined limits
- Also if we draw another set of sample, its mean is likely to appear between these two limits

Definition of Central Limit Theorem

- The central limit theorem states that for a given distribution with a mean μ & variance σ^2, the sampling distribution of the mean approaches a normal distribution with a mean (μ) & a variance σ^2/N, as N the sample size, increases



- No matter what the shape of the original distribution, the sampling distribution of the mean approaches a normal distribution

About Us

Canopus
Business Management Group

We are a Consulting & Management firm & help clients in:

- Deriving & executing **Business & Functional Strategies**
- Making **Customer Experience** a competitive advantage
- Improving profitability by leading **Business Transformation** initiatives

Consulting Clientele

- BARCLAYS
- vodafone
- PEPSICO
- Barry-Wehmiller International Resources
- TAFE
- MacLellan
- CALIBER POINT
- TEMENOS
- element k
- LUMIS PARTNERS
- karROX Technologies Ltd.
- Capgemini