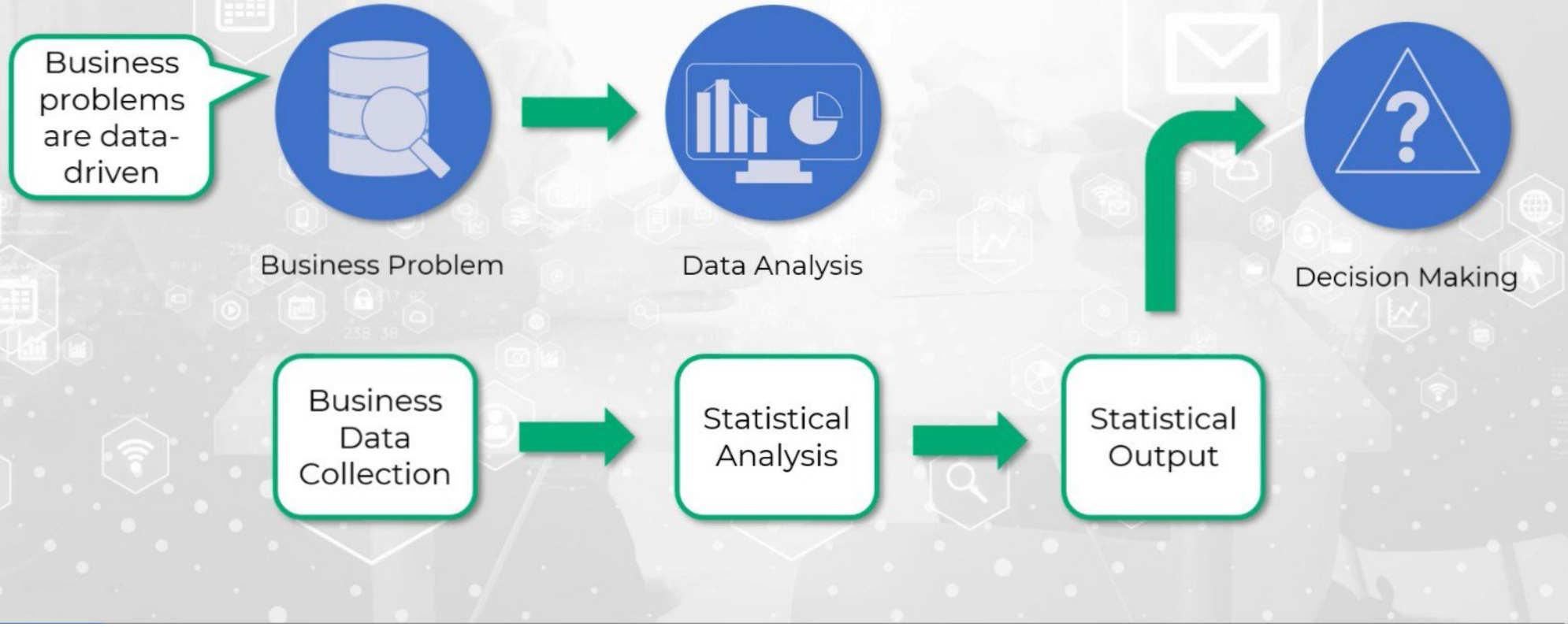
What's in Here

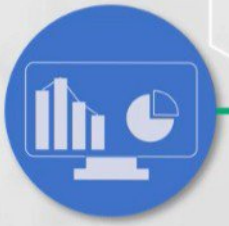
What is Statistics

The practice or science of collecting and analysing numerical data in large quantities, especially for the purpose of inferring proportions in a whole from those in a representative sample.



What is Statistics

The practice or science of collecting and analysing numerical data in large quantities, especially for the purpose of inferring proportions in a whole from those in a representative sample.



Statistics

Statistical Analysis

- E.g. – Average sales of a product is X units per month

Non - Statistical Analysis

- E.g. – Packaging of a product is beautiful

Categories In Statistics

There are two major categories in Statistics.

Statistical Analysis

Descriptive Statistics

Describe the basic features of the data in a study. They provide simple summaries about the sample and the measures

Min | Max | Mean

Numerical Analysis/Output

Inferential Statistics

Helps in generalising about the population by using various analytical tests and tools

Tall | Short

Relative Analysis/Output

Statistical Measures

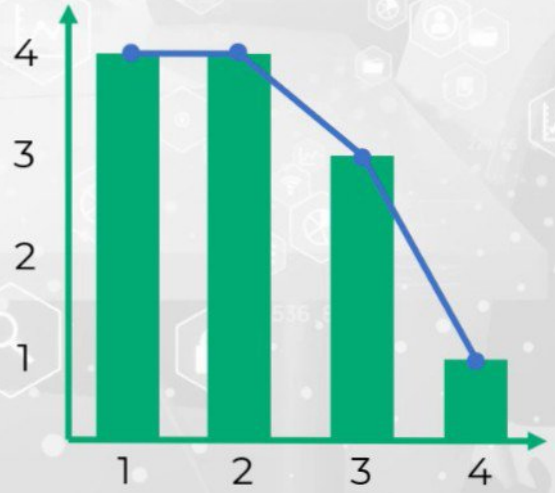
1. Measure of Frequency

Describe the basic features of the data in a study. They provide simple summaries about the sample and the measures

- Frequency Table
- Histogram
- Frequency Distribution

Data
1, 2, 3, 4
1, 2, 3
1, 1, 2, 2, 3

N	Freq.
1	4
2	4
3	3
4	1



Statistical Measures

2. Measure of Central Tendency

Helps in generalising about the population by using various analytical tests and tools

- Mean
- Median
- Mode

N	Freq.
1	4
2	4
3	3
4	1
5	2
6	1

Mode = 1, 2

Mode is the element(s) with the highest frequency

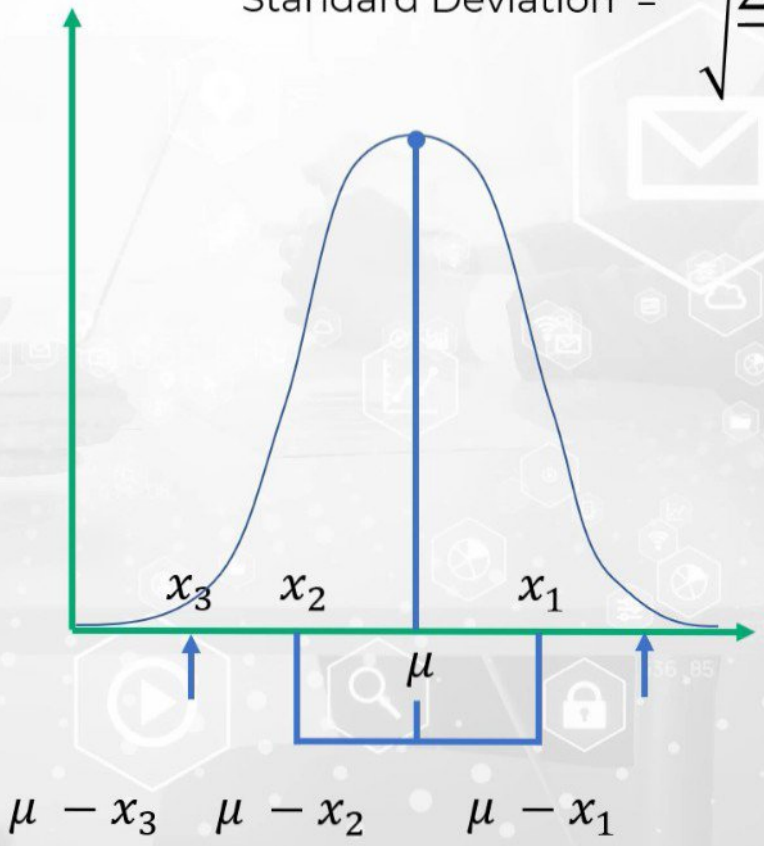
Statistical Measures

3. Measure of Spread

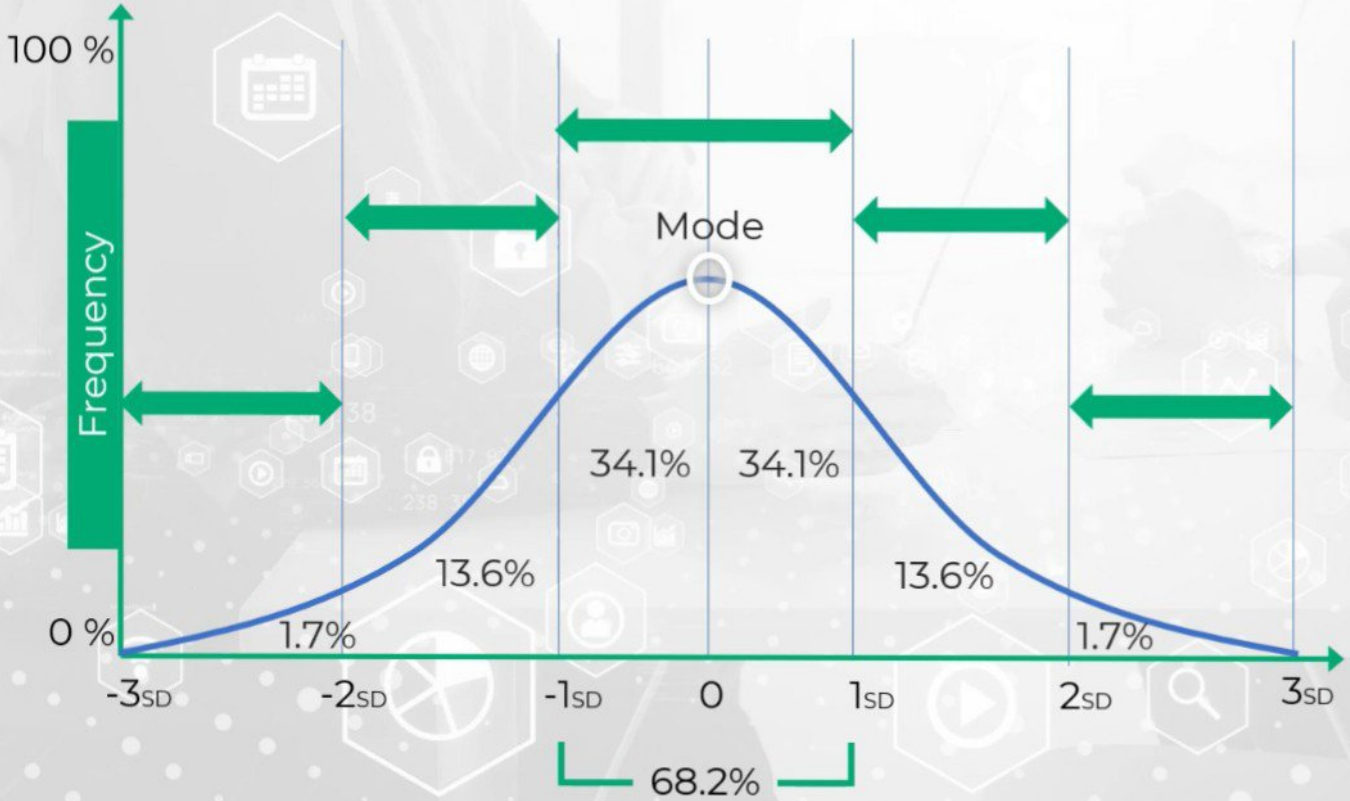
Helps in generalising about the population by using various analytical tests and tools

- Standard Deviation
- Variance

$$\text{Standard Deviation} = \sqrt{\frac{\sum(\mu - x_i)^2}{N}}$$



Distribution Curve



Range : The Distance between Minimum and Maximum values.

Frequency: Number of repetitions of the values in X - Axis.

Central Tendency: Refers to the accumulation of data points towards the center.

What is Null Hypothesis

Null Hypothesis (H_0) is the commonly accepted fact. In ML, it usually represents the conclusion that the model will not work. It is the opposite of the Alternate Hypothesis (H_A). Researchers attempt to reject the Null Hypothesis.

Claim: Model Accuracy = 80%

H_A – Model is good — **Alternate Hypothesis** ✓

H_0 – Model is bad — **Null Hypothesis** ✗

H_0 Probability Score (p value) < 0.05

H_A Probability Score (p value) > 0.95