2.06 Hypothesis Testing

What is Hypothesis Testing?

- Hypothesis testing is a statistical method that is used in making statistical decisions using sample data.

- Hypothesis testing is basically an assumption that we make about the population parameter.
  - Example: Average number of students in a DSI class is 20

- For hypothesis testing, we need a statistical way to prove and arrive at some mathematical conclusion whatever we are assuming is true.

What is Hypothesis Testing?

- A hypothesis test evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sample data.

- When we say that a finding is statistically significant, it’s thanks to a hypothesis test.

Where do we use Hypothesis Testing?

- Hypothesis testing tend to be used by researchers and data professionals in academia, healthcare and management consulting

- Sample Use Case
  - Facilitate business decision making (e.g. determine which strategy will lead to higher ROI)

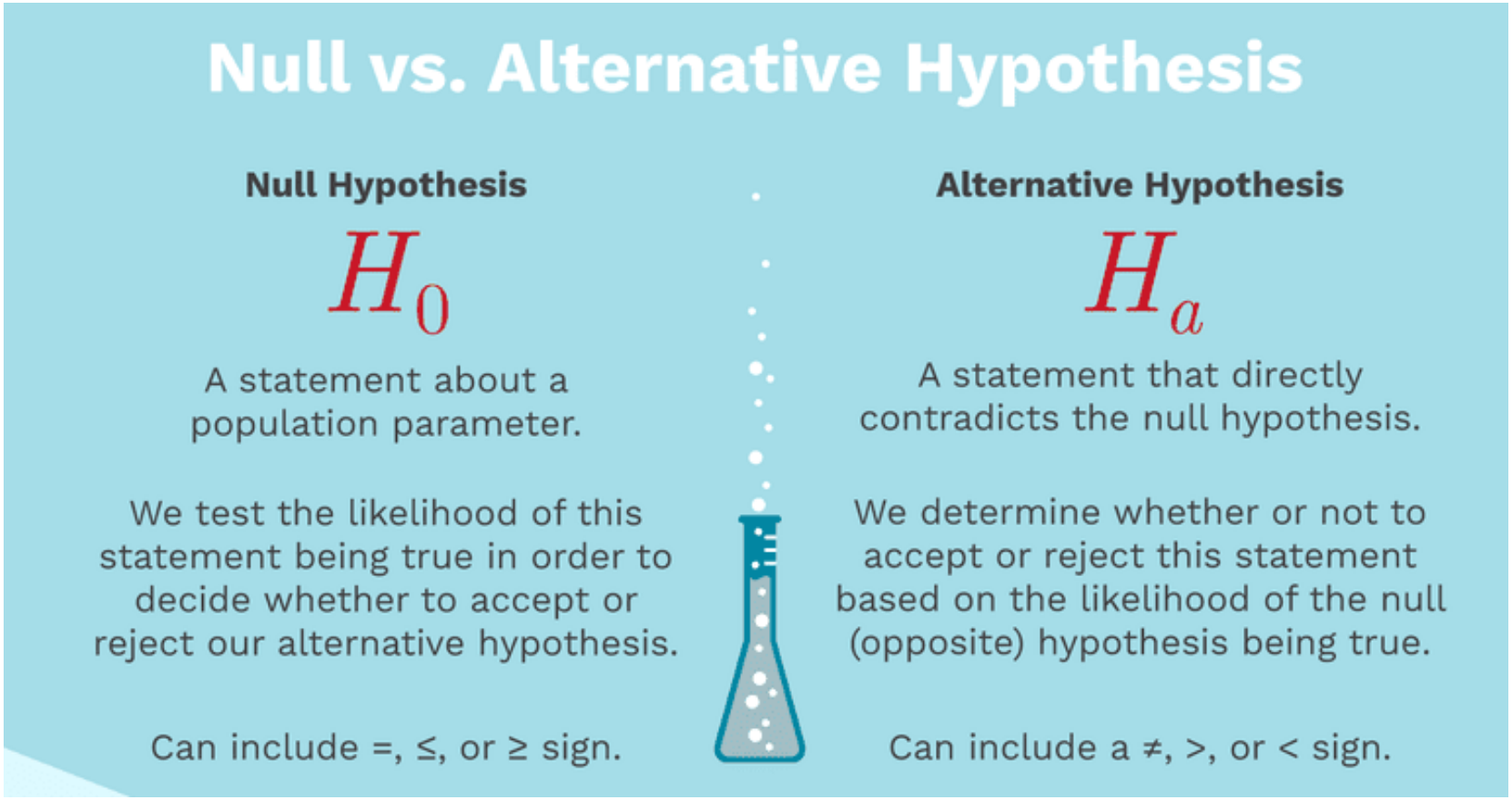
What are the key parameters of hypothesis testing?

- Null Hypothesis
  - General statement or default position that there is no relationship between two measured phenomena, or no association among groups
  - In other words, it is a basic assumption or made based on domain / knowledge of problem statement.
  - Example : company production = 50 unit/per day

What are the key parameters of hypothesis testing?

- Alternative hypothesis
  - Hypothesis used in hypothesis testing that is contrary to the null hypothesis.
  - It is usually taken to be that the observations are the result of a real effect (with some amount of chance variation superposed)
  - Example : company production is not equal to 50 unit/per day etc.

Null Hypothesis vs. Alternative Hypothesis



Level of significance

- Refers to the degree of significance in which we accept or reject the null-hypothesis.

- 100% accuracy is not possible for accepting or rejecting a hypothesis, so we therefore select a level of significance that is usually 5%

- This is normally denoted with alpha (maths symbol) and generally it is 0.05 or 5%, which means our output should be 95% confident to give similar kind of result in each sample.

Hypothesis Tests ➔ T-Test

- A t-test is a type of inferential statistic which is used to determine if there is a significant difference between the means of two groups which may be related in certain features.

- It is mostly used when the data sets, like the set of data recorded as outcome from flipping a coin a 100 times, would follow a normal distribution and may have unknown variances.

- T test is used as a hypothesis testing tool, which allows testing of an assumption applicable to a population.

Hypothesis Tests ➔ Z-Test

- You would use a Z test if:
  - Your sample size is greater than 30. Otherwise, use a t test.
  - Data points should be independent from each other. In other words, one data point isn't related or doesn't affect another data point.
  - Your data should be normally distributed. However, for large sample sizes (over 30) this doesn't always matter.
  - Your data should be randomly selected from a population, where each item has an equal chance of being selected.
  - Sample sizes should be equal if possible.
- Example: We use z-test for blood pressure with mean = 156 for a one-sample Z test.