we use observed data in a sample to draw conclusions about unobserved data

**Implication of hypothesis testing:**

* clinical research: widely used in psychology, biology and healthcare research to examine the effectiveness of clinical trials
* A/B testing: can be applied in business context like campaign incentives, website designs etc
* feature selection in machine learning

**Null hypothesis (H0)** the claim that no difference, no relationship or no dependency exists between two or more variables.

**Alternative hypothesis (H1)**is contradictory to the null hypothesis and it claims that relationships exist.

**T-test:**compare two groups/categories of numeric variables with small sample size

**Z-test:** compare two groups/categories of numeric variables with large sample size

**ANOVA test:** compare the difference between two or more groups/categories of numeric variables

**Chi-Squared test:** examine the relationship between two categorical variables

[7 stories about Practical Guides to Machine Learning curated by Destin Gong - Medium](https://destingong.medium.com/list/practical-guides-to-machine-learning-a877c2a39884)

[8 stories about Get Started in Data Science curated by Destin Gong - Medium](https://destingong.medium.com/list/get-started-in-data-science-8006bb4ba3ad)

Semantic similarity is usually calculated between a pair of text segments.  Distance can be cosine distance, L1 distance, exponential negative Manhattan distance and any other distance function