BHT_Introduction
Friday, 4 July 2025 9:20 PM

What is hypothesis?

The goal of many studies is to check whether the data support certain statements or predictions.

These statements are known as hypothesis about a population. They are usually expressed in terms of population parameters (mean, standard deviation, proportion etc)

Hypothesis usually claim that a population parameter takes a particular numerical value or falls in a certain range of values.

Example:- ① The new website version does not lead to a higher conversion rate than the current version.
② The mean lifetime of light bulbs is 1000 hours.
③ The new drug has no effect on blood pressure (mean variation = 0)

What is hypothesis testing?

It is a fundamental concept in inferential statistics used to determine whether there is enough evidence in data samples to support or reject a hypothesis about a population.

It allows data scientists to decide whether observed patterns in data are statistically significance or due to random chance.

Example:- Let us consider the hypothesis that the mean lifetime of light bulbs is 1000 hours.

The job of hypothesis testing is to determine whether there is enough evidence from data to support this hypothesis or not. Based on the result the hypothesis can be accepted or rejected.

Types of hypothesis:-

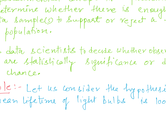
Each test has two hypotheses about a population parameter.

① Null hypothesis:- The null hypothesis is a statement that the parameter takes a particular value. The value in the null hypothesis usually represent no effect.

The symbol H_0 is used to denote the null hypothesis.

② Alternative hypothesis:- The alternative hypothesis states that the parameter falls in some alternative range of values.
The value in the alternative hypothesis represents an effect of some type.
The symbol H_a or H_1 denotes alternative hypothesis.

By done key steps in hypothesis testing.



① How to state/formulate hypothesis

Example 1:-

Scenario: A manufacturer claims that its light bulbs last an average of 1,000 hour. We want to test if there is enough evidence to support this statement.

Null hypothesis (H_0): The average lifetime of the light bulbs is 1,000 hours.

Alternative hypothesis (H_a):- The average life span of the light bulbs is not 1,000 hours.

$H_0: \mu = 1000$ hours
$H_1: \mu \neq 1000$ hours

Two sided hypothesis because $\mu > 1000$ or $\mu < 1000$

Example 2:

Scenario: A school is testing two different teaching methods for improving student test scores. We want to test if one method results in higher score than the other.

Null hypothesis (H_0): There is no difference in the average test scores between the two teaching methods

Alternate hypothesis (H_1): The average test scores are different between the two teaching methods

$H_0: \mu_1 = \mu_2$
$H_1: \mu_1 \neq \mu_2$

μ_1 and μ_2 are the average test scores of method 1 and method 2 respectively (Two-sided hypothesis)

Example 3:

Scenario:- A university claims that at least 80% of students graduate within 4 years. We want to test if the graduation rate is less than 80%

Null hypothesis (H_0): The graduation rate is at least 80%

Alternative hypothesis (H_1): The graduation rate is less than 80%

$H_0: p \geq 0.8$
$H_1: p < 0.8$

p is graduation rate
One-sided hypothesis

Example 4:-

Scenario: A company launches a new marketing campaign and wants to check if increase sales compared to last year's sales.

Null hypothesis (H_0): The marketing campaign has no effect on sales
Alternate hypothesis (H_a): The marketing campaign increases average sales

$H_0: \mu_b = \mu_a$
$H_1: \mu_a > \mu_b$

μ_a and μ_b are the average sales after and before the campaign.
One-sided hypothesis

② Test-statistics:

The parameter (mean, proportion etc) to which the hypothesis refer has a point estimate. A test statistics describes how far the point estimate falls from the parameter value given in the null hypothesis.

It measures the difference between the observed sample statistics (e.g. sample mean, proportion) and the hypothesized population parameter (e.g. population mean, proportion).

Example: Let us consider a Null hypothesis $H_0: p = 1/3$, where p denotes the probability that an actor can correctly predict which of three personality charts belongs to a person.

To test this hypothesis, let we have a sample of 116 predictions. Let 40 out of 116 predictions were correct.
Thus, the point estimate of the probability in the sample $\hat{p} = \frac{40}{116} = 0.345$.

The test statistics compares this point estimate to the value in the null hypothesis ($p = 1/3$).

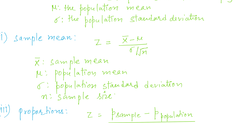
The test statistics compares point estimate to the parameter value in the null hypothesis depends on type of statistical test.

The commonly used test is Z-test.

Z-Test:- This test is used when the sample size is large (typically $n \geq 30$) and the population standard deviation is known (or the sample size is large enough for the central limit theorem to apply, making the sample standard deviation a good approximation).

It is performed by assuming that sample parameter follows the normal prob distribution.

One-tailed and two-tailed Z-test: Based on formulation of an alternate hypothesis, the Z-test results in one-tailed (one-sided) or two-tailed (two-sided) test as shown in the below.



α is the significance level (to be discussed later).

Test-statistics using Z-test computes Z-score to compare the point estimate under the parameter value in the null hypothesis.

Higher values of z-score indicate that the sample statistic is far away from the hypothesized population parameter, suggesting strong evidence against the null hypothesis.

③ Calculation of test/z-statistics:

Z-statistics:- or Z-score measures how many standard deviation away from the population is the sample parameter assuming the data follows normal distribution.
For example, z=2 indicate that a point estimate (sample parameter) is 2 standard deviation away from the hypothesized value in the null hypothesis (population parameter).

The summary of z-statistics formulas is given below

(i) Single data point: $z = \frac{x - \mu}{\sigma}$

x: observed data point (value from sample)
μ: the population mean
σ: the population standard deviation

(ii) Sample mean: $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$\bar{x}$: sample mean
μ: population mean
σ: population standard deviation
n: sample size

(iii) Proportions: $z = \frac{\text{Sample} - \text{Population}}{\sqrt{\frac{\text{Population}(1-\text{Population})}{n}}}$

Sample: sample proportion
Population: population proportion (under null hypothesis)
n: sample size