Hypothesis Testing Explained

Confidence Interval: Here we try to find population mean using sample mean.

Hypothesis Testing: Here we try to find whether the claimed population mean is correct or not.

To understand Hypothesis testing process, let's start with a simple example:

Court & The Crime Suspect

Some people claimed that the person has committed a crime.

(So the person was arrested, now the court has to decide whether this person is guilty).

Step 1: Define Null & Alternate Hypothesis

Null Hypothesis: Person is innocent.

Alternate Hypothesis: Person is guilty.

(2) Our main objective : We always try to reject the Null Hypothesis.

(3) Collect the evidence/proof to be able to decide whether we can reject the NH or not.

→ For a business use case, we will collect a sample data related to a problem statement and then run hypothesis tests on the sample data to get the evidence.

(4) We need significantly strong evidence not just any amount of evidence →

(Let's see some evidence we got)

(i) During arrest, the suspect was carrying a knife in their pocket

(ii) Knife had blood stains on it.

(iii) Blood has matching DNA of the victim.

(iv) Suspect's t-shirt had fingerprints of the victim.

(v) CCTV footage shows that the person was in house during the incident for 45 mins.

Now we got all these evidences, it seems that these evidences are strong enough to prove that person is guilty.

verdict by court → Reject the Null Hypothesis (Person is not innocent).

(5) For business problem statement, the proof will be in the form of "p-value".

(6) But what is p-value?

p-value is the probability of getting such evidence, if null hypothesis is actually true.

In simple words,

p-value is the :

P(Null Hypothesis being actually true)

↓

if we have such evidence

So, as we can understand :

The probability of getting such type of evidence is very low for an innocent person. So in this case the p-value is very low.

And then we can also say :

It we have got such evidences, it means only 1 thing, which is, the person is not innocent.

↓

And hence the evidence strongly suggests us to "reject the Null Hypothesis".

So in summary:

If p-value is very low → Reject the Null Hypothesis.

But how low is very low. For that, we will significance level next.

One more thing before we discuss the significance level.

Types of Error in Hypothesis Testing:

(i) Type 1 Error (False Conviction):

Declaring the suspect to be guilty, when they are innocent. Also, rejecting NH when it should have been accepted.

(2) Type II Error: Failing to punish the suspect when they were guilty. Failing to reject a NH when it should have been rejected.

	Null Hypothesis is TRUE	Null Hypothesis is FALSE
Reject null hypothesis	Type I Error (False positive)	Correct Outcome! (True positive)
Fail to reject null hypothesis	Correct Outcome! (True negative)	Type II Error (False negative)

Note: We never say that we accept Null Hypothesis. Instead we say, "we failed to reject Null Hypothesis due to lack of evidence".

Significance Level (α):

Imagine you're a doctor testing a new medicine. You want to know if it really works better than the old one. To decide, you run an experiment with patients and collect data.

The Courtroom Analogy : Think of your experiment like a courtroom trial:

- **Null Hypothesis (H_0):** The new medicine is no better than the old one (the defendant is innocent).

- **Alternative Hypothesis (H_1):** The new medicine is better (the defendant is guilty).

In court, you don't want to wrongly convict an innocent person. In science, you don't want to wrongly claim your medicine works when it doesn't. This mistake is called a Type I error.

Enter the Significance Level (α)

The significance level, often written as α, is like the judge setting a threshold for evidence. It's the maximum probability you're willing to accept for making a Type I error—declaring the medicine works when it actually doesn't.

If you set $\alpha = 0.05$, you're saying: "I'm willing to accept a 5% chance of being wrong if I claim the new medicine is better."

In simple words,

The maximum amount of error you are ready to accept, when you reject Null Hypothesis and accept the alternate.

Commonly used values for SL:

5% (0.05), 1% (0.01),

10% (0.1)

Most commonly we use SL = 5%.

But why are we talking about it?

Let's see

When we run a Hypothesis Test on the collected sample data, we get p-value as the output.

Hypothesis Test
↓
Sample data
↓
p-value

For example →

if p-value is 0.03 (3%), then it means that the probability of Null Hypothesis being true is 3%.

And here if we decide to reject NH, then there is 3% chance that we are wrongly rejecting Null Hypothesis.

If you remember, we have already decided that we are ready to make 5% error.

Hence we can say →

if

$$\boxed{\text{p-value} < \text{Significance Level}}$$

↓

Reject Null Hypothesis

Selecting the appropriate test:

We will talk about these major tests:

Z-test, t-test, ANOVA test, chi-squared test

Every test has its own assumptions or requirements and once these assumption are met we can use it on the sample test.

(i) Z-test

→ Sample should be numerical.

→ Sample should be normally distributed.

→ Sample size > 30.

→ Select the sample randomly.

→ Population standard deviation must be given.

if we apply Z-test on the data

↓

Then we calculate Z-statistic using:

$$Z\text{-stat} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad \left(\frac{PM - SM}{PSD / \sqrt{SS}} \right)$$

↓

Then we calculate p-value using Z-statistic.

↓

And if,

p-value < Significance Level

↓

Reject Null Hypothesis.

(ii) t-test

Like z-test but more flexible

→ Use sample standard deviation.

→ Works with ss > 30 or ss < 30.

→ But the sample has to be normally distributed.

Here we calculate t-statistic

$$t\_statistic = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

↓

Then the p-value.

In both, z-test & t-test, we have two versions:

(i) One-sample z-test & two sample z-test.

(ii) One sample t-test & two-sample t-test.

(iii) ANOVA test:

Analysis of Variance (ANOVA) is a test used to compare the means of 3 or more groups. It analyzes the variance within the group and between groups.

The main is to assess whether the observed variance b/w group means is more significant than within the groups.

Assumptions:

(i) → Each group should be normally distributed.
(ii) → Variance within each group should be roughly the same.

(iii) Observations must be independent of each other amongst the groups. One group values should not affect the other group.

We calculate F-statistic here:

$$\text{F-statistic} = \frac{\text{B/w group variance}}{\text{within group variance}}$$

ANOVA test can be thought of as an exten-
-sion of two-sample t-test.

(iv) Chi-squared test:

Used to determine if there is a signi-
- ficant association b/w categorical variables.

Ex: Association b/w cart outcome (pur-
-chased vs abondoned) with device type (mobile, tablet, desktop)

→ Asses whether customer churn is associ-
-ated with last customer interaction type (email, phone, chat, no-contact).

We calculate chi-statistic in this case.

$$\text{chi-statistic} = \sum \frac{(O_i - E_i)^2}{E_i}$$

Now, let's see our first problem statement to understand the calculations of test-statistic & p-value:

Ex: PepsiCo claims that their INR 20 packet contains 50 grams of chips in it with a population standard deviation of 5 grams.

The quality control department wants to verify this claim by testing 50 packets.

→ Sample size = 50 packets

→ Sample data:

(i) 49.8 gms

(ii) 50.6 gms

(iii) 54.2 gms

⋮

(50) 53 gms

→ Sample Mean : 53 gms (suppose)

→ Population Mean : 50 gms (given)

So now we have to compare sample mean with population mean.

When we test the sample, we are expecting the sample mean to be similar to very close to population mean. If it happens, we don't reject the Null Hypothesis. Else we reject Null Hypothesis.

In simple words,

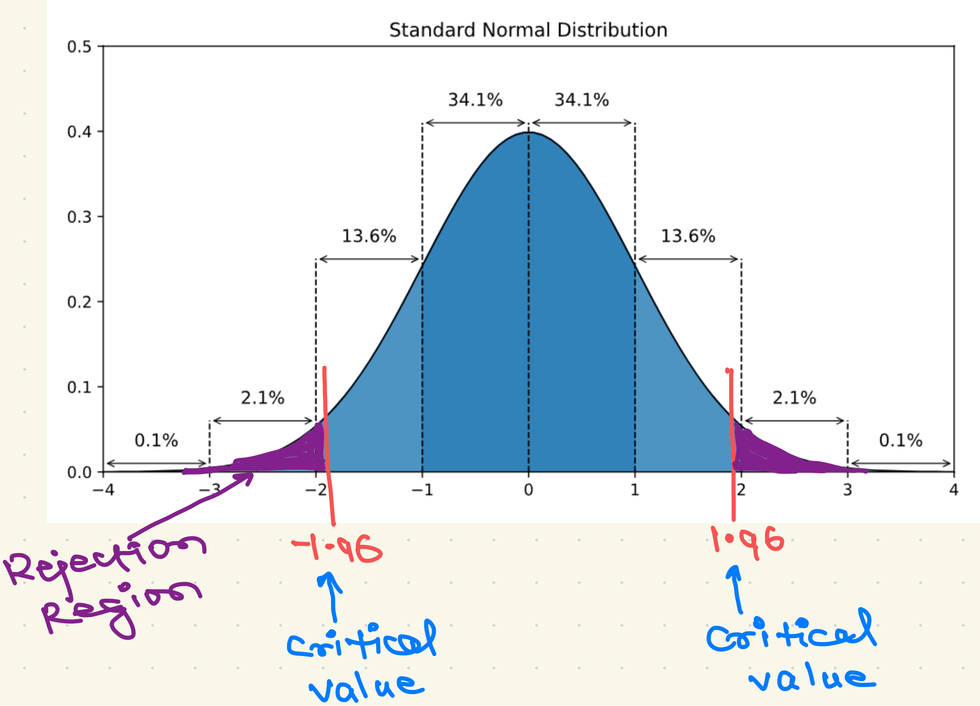
if PM = SM → Do not reject NH

else if PM ≠ SM → Reject Null Hypothesis

Hypothesis statements:

Null Hyp: The average weight of chips packets = 50 gms.

Alternate Hyp: The average weight of chips packet ≠ 50 gms.



Here, we pick significance level to be

$\alpha = 5\%$ or 0.05

hence, our rejection region is 5%

and the remaining area b/w two critical values is 95%.

Now, we calculate z-statistic & then p-value to be able to decide:

$$z\text{-stat} = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

$$= \frac{53 - 50}{5 / \sqrt{50}}$$

$$= 4.22$$