BHT_Introduction

Friday, 4 July 2025 9:18 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 hypotheses about a population. They are usually expressed in terms of population parameters (mean, standard deviation, proportion etc).

Hypotheses 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 reduction = 0).

What is hypothesis testing?

It is a fundamental concept in inferential statistics used to determine whether there is enough evidence in data sample(s) 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 their 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.

Fig shows key steps in hypothesis testing.

STEPS IN HYPOTHESIS TESTING

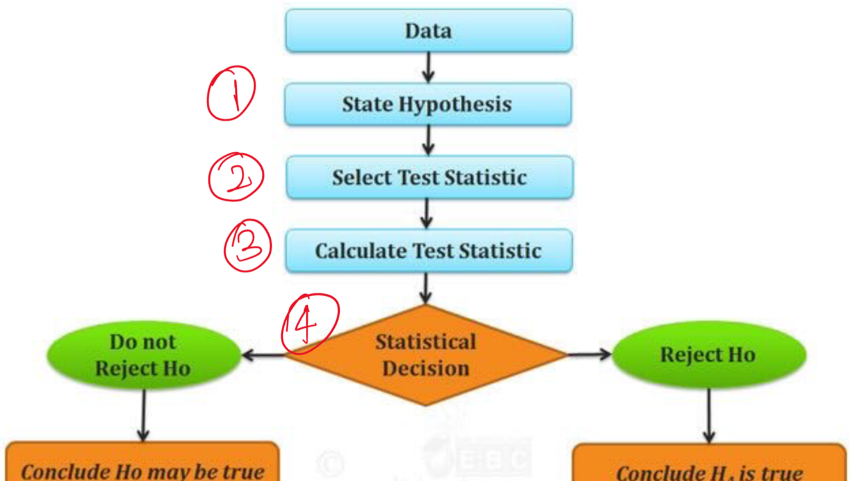
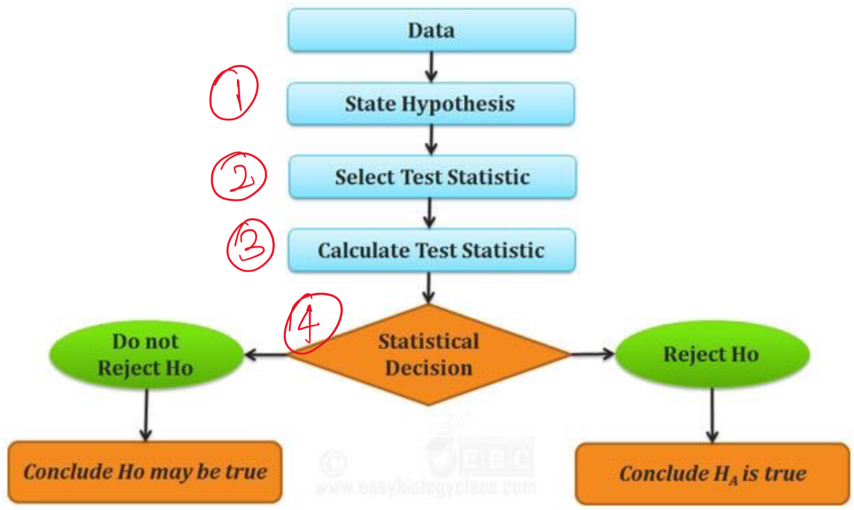


Fig shows key steps in hypothesis testing.

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 lifespan 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.

$$\boxed{\begin{aligned} H_0 &: \mu = 1{,}000 \text{ hours} \\ H_1 &: \mu \neq 1{,}000 \text{ hours} \end{aligned}}$$

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

$$\boxed{\begin{array}{l} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{array}}$$

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 with in 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%.

$$\boxed{\begin{array}{l} H_0: p \geq 0.8 \\ H_1: p < 0.8 \end{array}}$$

p is graduation rate one-sided hypothesis.

Example 4:-

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

Null hypothesis (H_0): The marketing campaign has no effect on sales

Alternate hypothesis (H_1): The marketing campaign increases average sales.

$$\boxed{\begin{array}{l} H_0: \mu_b = \mu_a \\ H_1: \mu_a > \mu_b \end{array}}$$

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 astrologer 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 prediction 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 ($p = 1/3$).

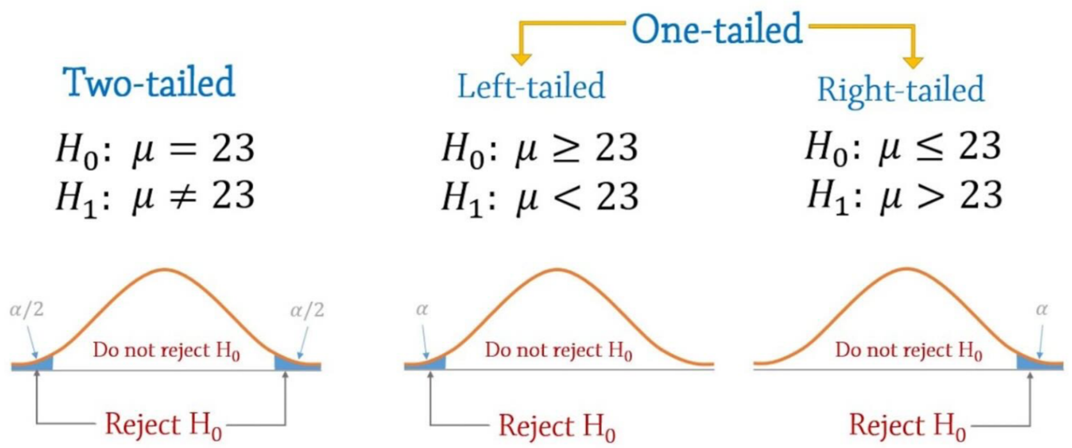
How test-statistics compares point estimate to the parameter value in the null hypothesis depends on type of statistical test.

The commonly used test is Z-test.

Z-Test:- This test is used when the sample size is large (Typically $n > 30$) 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 Fig below.



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

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

Higher values of z-score indicate that the sample statistics is farther 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, $z=2$ indicates that a point estimate (sample parameter) is 2 standard deviation away

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, $z=2$ indicates that a point estimate (sample parameter) is 2 standard deviation away from the parameter 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: $$Z = \frac{p_{sample} - p_{population}}{\sqrt{\frac{p_{population}(1-p_{population})}{n}}}$$

p_{sample}: sample proportion
$p_{population}$: population proportion (under null hypothesis)
n: sample size.

④ Statistical decision:

In hypothesis testing, the p-value and critical value are used to make decision whether to reject the null hypothesis.

Probability (p) - Value:—

It is the probability that the test statistics equals the observed value or a value even more extreme, assuming that the null hypothesis is true.

It helps determine the strength of the evidence against the null hypothesis.

$$\boxed{\text{p-value} = P\left(\begin{array}{r}\text{test statistics takes observed value} \\ \text{or beyond in tails}\end{array} \middle| H_0 \text{ true}\right)}$$

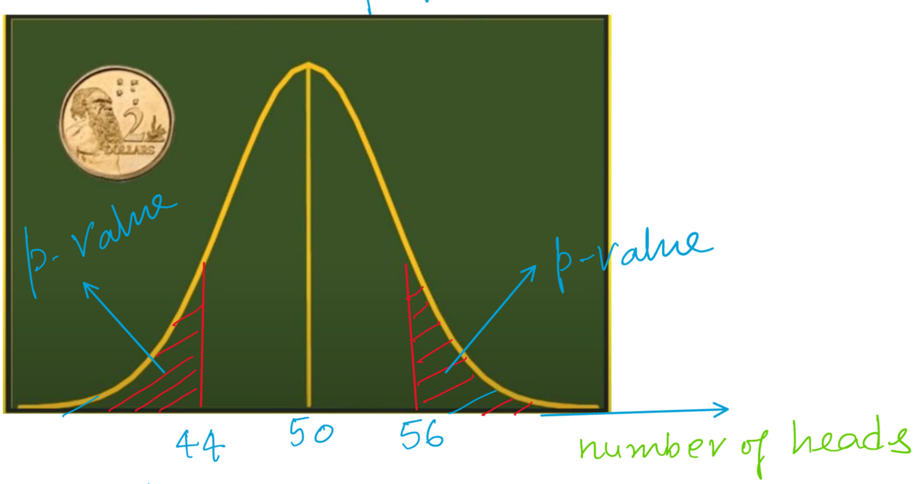
Example:- Suppose that we want to test wether a coin is fair. So, we have

H_0: the coin is fair $(P(H) = P(T) = 0.5)$

H_1: The coin is not fair

For this let we have 100 coin flips.

Fig shows the prob distribution of the number of heads.

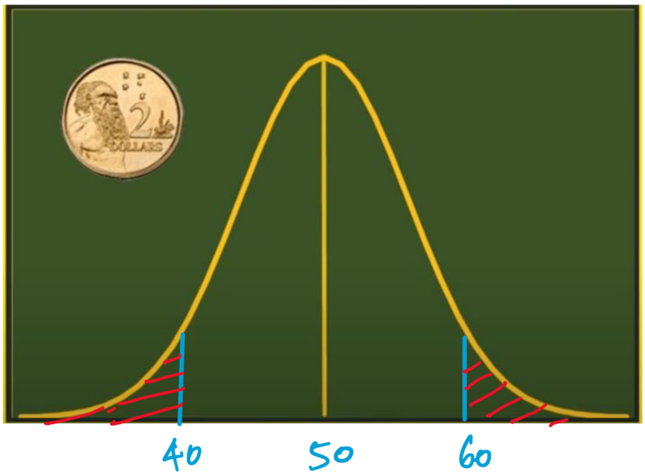


Let we observe 56 heads in the 100 coin flips. Then, the p-values is the probability of getting a sample (test statistics) equals the observed value (56) or a value even more extreme.

The p-value in the above example is 0.193.
This p-value is calculated using z-score (see later how to calculate it.

The above p-value means that if the coin is fair (Ho is true), the prob of getting 56 heads (or a sample more extreme) is 0.193.

Let us now assume that we get 60 heads in 100 coin flips.



p-value now is 0.032.

The value 0.032 of p-value indicates that if the coin is fair, the prob of getting 60 heads (or a sample more extreme) is 0.032.
We now start thinking that the null hypothesis may not be true because we are getting 60 heads out of 100 coin flips.

p-value (90 heads in 100 coin flips) = 0.0001 (almost zero)

Thus, smaller the p-value, stronger is the evidence against the null hypothesis.

To put it in other words, larger is the difference between sample statistics and hypothesized population parameter, smaller is the p-value and stronger is the evidence against the null hypothesis.

The obvious question now is what should be the p-value to reject a null hypothesis?

The answer is the significance level (α).

Significance level (α) :- The significance level α is the threshold for rejecting the null hypothesis when it is actualy true.

Commonly used value of α is 0.05, meaning there is 5% chance of rejecting a null hypothesis when it is actually true. It is set by the researchers.

Interpretation :

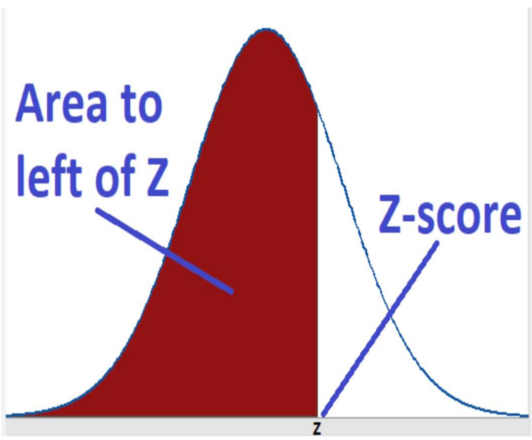
① If $p \leq \alpha$ (significance level), we reject the null hypothesis, suggesting that the observed data is statistically significant.

② If $p > \alpha$, we fail to reject the null hypo-thesis, meaning that there is not enough evidence to support the alternative hypothesis.

How to map z-score to probability & vice-versa :-

It is calculated using z-score table. A part of this table is shown below. Table entry for z is the area under the standard normal curve to the left of z.

Second Decimal Place of z										
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
...										
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9139	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441



For example, let us find the value of Z for cumulative prob of 0.909. From the table, this corresponds to Z = 1.3 + 0.04 = 1.34. Thus, to get a cummulative prob of 0.909, we have to move 1.34σ away from the mean.

$$\int_{\mu-1.34\sigma}^{\mu+1.34\sigma} f_X(x)\,dx = 0.909$$

where $f_X(x)$ is the normal distribution.

Critical value :- A critical value is a threshold or cut off point that helps in deciding whether to reject the null hypothesis. It is determine by the chosen significance level (α) and the distribution of the test statistics.

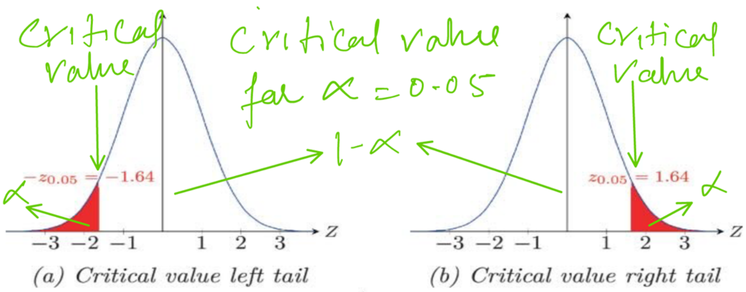
The critical value marks the boundary for the rigion of rejection (also called the critical region). If the test statistics falls beyond this critical value, the null hypothesis is rejected.

Critical value for a two-tailed test :-

Critical value for a two-tailed (two-sided) test for a given α is the Z-score at the value $\alpha/2$.

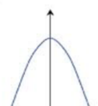
For example, for $\alpha = 0.05$, the critical value is the Z-score for at $\alpha/2 = 0.025$, which is ± 1.96, as show in Fig.

The Z-values in the red portion give the region of rejection for null

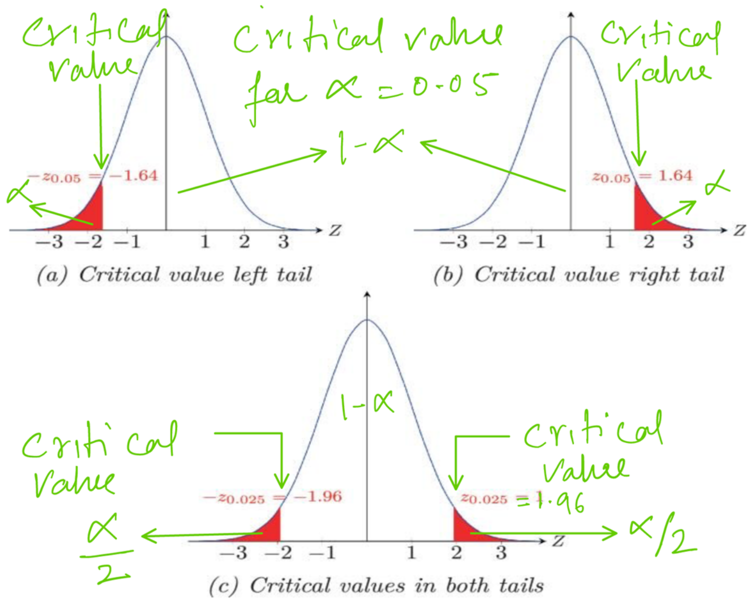


(a) Critical value left tail (b) Critical value right tail

y-axis is Z-score.



The Z-values in the red portion give the region of rejection for null hypothesis



y-axis is z-score.

Critical value for a one-tailed test:-

The critical value for a one-sided (one-tailed) test for a given α is the z-score at the value α.

For the above example with $\alpha = 0.05$, the critical value for one-side test is the z-score at $\alpha = 0.05$, which is 1.64 for right tail and -1.64 for the left tail.

Example 1:

Scenario: We want to test if the average score on test for a sample of 15 students is significantly different from 80, with the population standard deviation of 12 and sample mean of 85.

Solution:

Null hypothesis (H_0): $\mu = 80$

Alternative hypothesis (H_1): $\mu \neq 80$ (two-tailed test)

Recall, the z-statistics for sample is

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{85-80}{12/\sqrt{15}}$$

$$\boxed{Z = 1.61}$$

$\boxed{Z = 1.61}$

We next need to find p-value from the z-score

from the z-score table, $Z = 1.61$ corresponds to the cummulative prob of 0.9463.

The area under the upper tail beyond the z-value

$$1 - 0.9463 = 0.0537$$

Since it is two-tailed test, the total p-value

$$p\text{-value} = 2 \times 0.0537 = 0.1074$$

the critical value with $\alpha = 0.05$ for two tail test is 0.025, which leads to critical z-value of (using the z-table)

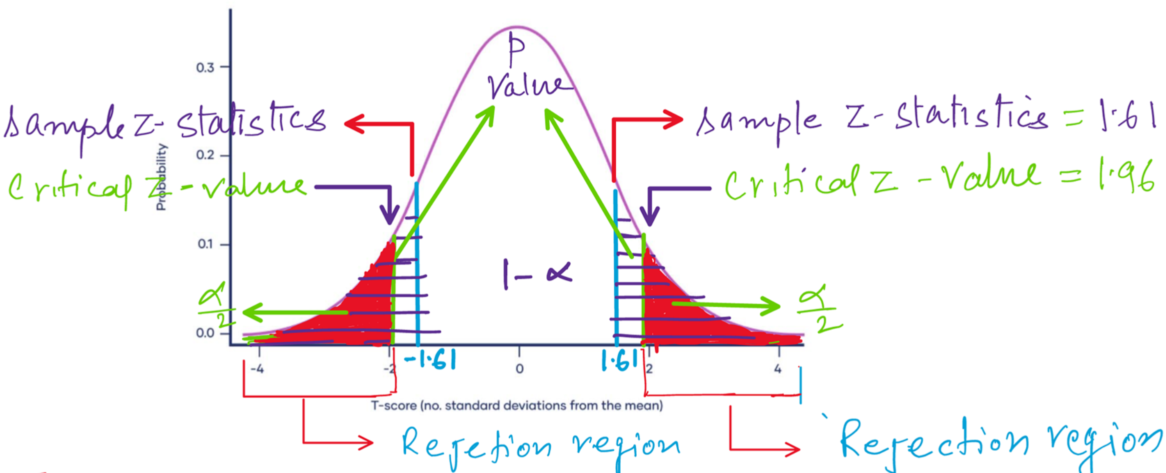
$$\text{Critical Z-value} = \pm 1.96$$

We see that p-value $(0.1) > \alpha\ (0.05)$, so we fail to reject the null hypothesis.

In other words, there is not enough evidence to conclude that the average score is significantly different from 80.

Or

The critical Z-value (1.96) is higher than 1.61, meaning that the z-statislice does not fall in the rejection region. Thus, there is not enough evidence to reject the null hypothesis.



Example 2:

→ Rejection region → Rejection region

Example 2:

Scenario:- A company claims that 60% of its customers are satisfied with their services. A survey of 200 customers shows that 130 are satisfied. We want to test if the proportion of satisfied customers is different from the claimed 60% at $\alpha = 0.05$.

Solution:

Null hypothesis (H_0): $P = 0.60$

Alternative hypothesis (H_1): $P \neq 0.60$

The sample size $n = 200$

Recall that the Z-statistics for proportion is

$$Z = \frac{\hat{p}_0 - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, \qquad \hat{p}_0 = \frac{130}{200} = 0.65$$

$$p_0 = 0.60$$

$$n = 200$$

$$Z = \frac{0.65 - 0.60}{\sqrt{\frac{0.6(1-0.6)}{200}}} = \frac{0.05}{0.03464} \approx 1.44$$

From the z-table, the cummulative prob for $Z = 1.44$ is approximately 0.9251.

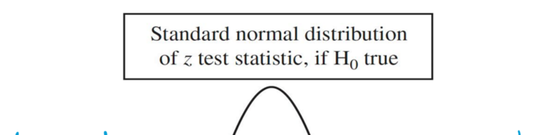
The area in the upper tail beyond this z-value is $1 - 0.9251 = 0.0749$.

Since its a two-tailed test, the total p-value

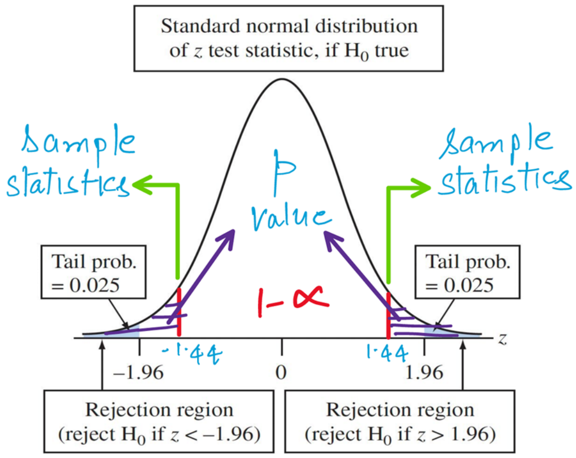
$$p\text{-value} = 2 \times 0.0749 = 0.1498$$

the Critical value with $\alpha = 0.05$ for two tail test is 0.025, which leads to critical z-value of (using the z-table)

Critical z-value $= \pm 1.96$



Critical Z-value = ± 1·96



Since p-value (0·1498) > α (0·05), there is not enough evidence to conclude that the proportion of satisfied customers is different 60%.

or

Since the critical Z-value (1·96) is greater than the Z-statistics (1·44), there is not enough evidence to reject the null hypothesis.

Summary of hypothesis testing :-

① State the hypotheses:

② Choose a significance level (α)

③ Collect and analyze data :-

Collect a sample from the population and Calculate the relevent sampl statistics (e.g; sample mean, sample standard deviation)

④ Calculate Test statistics

⑤ Determine the p-value or/and critical value

⑥ Take a decision

Eample 1:

Scenario:- A factory claims that the average weight of a product is 500 grams. We want to test if the average weight is significantly different from 500 grams.

Solution: Step 1- State hypotheses

Null hypothesis (H_0) : $\mu = 500$ (the population mean)

Alternative hypothesis (H_1): $\mu \neq 500$ (two tailed test)

Two tailed test because H_1 includes all other values, bothe below and above the value 500 grams in H_0

Step: 2 choose a significance level

Let the significance level α be 0.05.

Step 3: Collect and analyse data

Let we select 30 sample and find the average weight to be 505 grams. Let the population standard deviation is 10.

Step 4 : Calculate test statistics:

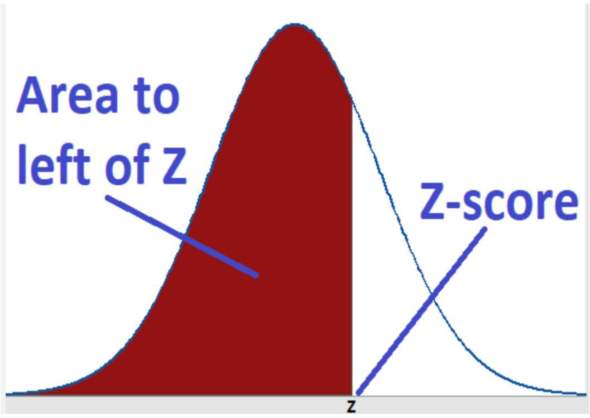
Step 4: Calculate test statistics:

The Z-statistics for mean

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{505 - 500}{10/\sqrt{30}}$$

$$\boxed{Z \approx 2.74}$$

Step 5: Determine p-value and critical value



From the z-score table, the cummulative probability for $z = 2.74 = 0.9969$

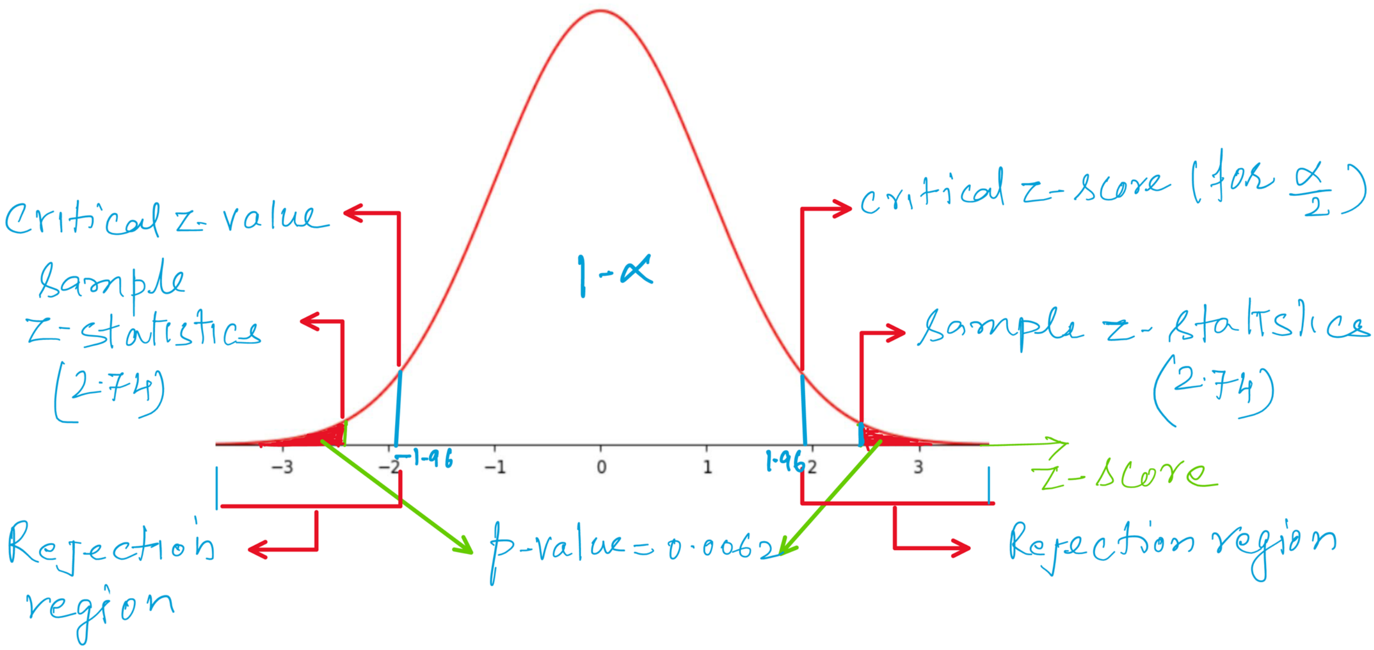
The area in the upper tail beyond this z-value is

$$1 - 0.9969 = 0.0031$$

Since it is a two-tailed test, the total p-value is

$$2 \times 0.0031 = 0.0062$$

For $\alpha = 0.05$, the critical value $= \alpha/2 = 0.025$ of the total probability of both side. The critical z-value corresponds to 0.025 is is ± 1.96 (from z-score table)



Step 6: Descision:

compare the p-value with the significance level ($\alpha = 0.05$)

Step6 : Descision:

compare the p-value with the significance level ($\alpha = 0.05$)

$0.0062 < 0.05$, so reject the null hypothesis, meaning that there is significant evidence to conclude that the mean weight is different from 500 gram.

OR

The calculated z-statistics (2.74) is greater than the critical z-value, the null hypothesis is rejected.

BHT_Errors

Friday, 4 July 2025 10:02 PM

When we need to decide if the evidence is strong enough to reject H_0, we have seen that the key is whether p-value falls below a prespecified significance level α.

We reject H_0 if p-value $\leq \alpha$

We do not reject H_0 if p-value $> \alpha$

The smaller α is, the stronger the evidence must be reject H_0.

Because of sampling variability, decisions in hypothesis testings always have some uncertainty. A decision can be in error.

Hypothesis testing has two types of potential errors

1. Type I error
2. Type II error

Type I error:- Type I error occurs when H_0 is true, but it is rejected.

One can think Type-I error as a false positive, because a positive decision is reached by rejecting H_0, yet the decision is false.

Type II error:- Type II error occurs when H_0 is false, but it is not rejected.

One can think Type-II error as false negative, because a negative decision is reached by not rejecting H_0, yet the decision is false.

Example: Consider a decision in a legal trial.

H_0: defendent is innocent

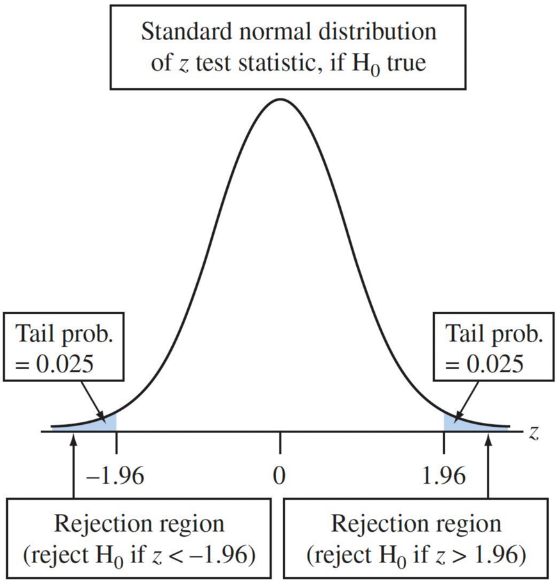
H_1: defendent is guilty

A Type-I error, rejecting null hypothesis, occurs in convicting a defendent who is actually innocent.
A Type-II error, not rejecting H_0 even though it is false, occurs in acquitting a defendent who is actually guilty

A potential consequence of a Type-I error is sending an innocent person to Jail, whereas a potential consequence of a Type-II error is setting free a guilty person.

The significance level is the probability of Type-I error.

Let us look at a two-sided test about a proportion to see how we get this result.
Recall for a two-sided test, the two-tail probability that forms p-value ≤ 0.05 corresponds to test statistics z satisfies $|Z| \geq 1.96$, which forms rejection region.



The probability of rejecting H_0 is the probability of of observing the z-statistics in the rejecting region.
This prob would be

$$P(Z < -1.96) + P(Z > 1.96) = 0.05$$

We see that the summation of probability is equal to the significance level.

$$\text{Prob}(\text{Type I error}) = \text{significance level} (\alpha)$$

The good news is that we can control the prob of Type-I error by our choice of the significance level. The more serious the consequences of a Type-I error, the smaller α should be.

Trade-off between Type-I and Type-II errors:-

Why don't we make extreamly small prob of Type-I error by setting α to be small such as $\alpha = 0.000001$. For example, why don't we make it almost impossible to convict someone who is really innocent.

When we make α smaller, we need a smaller p-value to reject Ho. It then becomes harder to reject Ho. Thus, it will also be harder to reject Ho even if Ho is false.

In other words, the smaller we make the prob of Type-I error, the larger the prob of Type-II error becomes, that is failing to reject Ho even though it is false

$$\boxed{\text{As } P(\text{Type-I error}) \text{ goes down, } P(\text{Type-II error}) \text{ goes up}}$$

Calculation of P(Type-II error) can be complex. In practice, to make decision, we need to set only P(Type-I error), which is the significance level.

$\boxed{\text{As } P(\text{Type-I error}) \text{ goes down, } P(\text{Type-II error}) \text{ goes up}}$

Calculation of P(Type-II error) can be complex. In practice, to make decision, we need to set only P(Type-I error), which is the significance level.