

Testing the significance of single mean

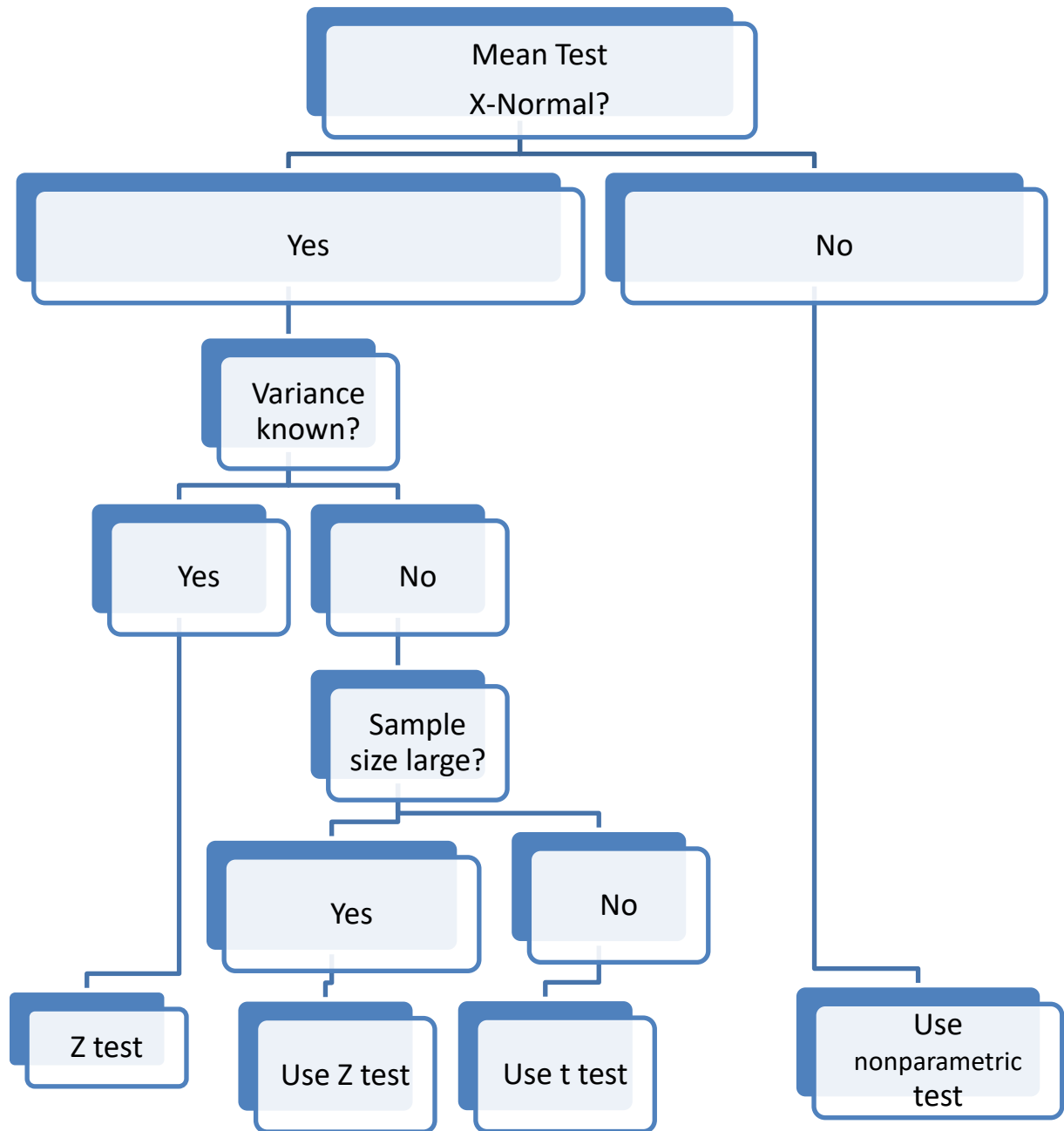
Hypothesis testing is an act in statistics whereby an analyst **tests** an assumption regarding a population parameter.

Hypothesis testing is used to assess the plausibility of a **hypothesis** by using sample data. Such data may come from a larger population, or from a data-generating process.

Testing the significance of single mean

Purpose:

A one sample test of means compares the mean of a sample to a pre-specified value and tests for a deviation from that value.



Population Standard Deviation Known

If the population standard deviation, σ is known, then the population mean has a normal distribution, and you will be using the z-score formula for sample means. The test statistic is the standard formula you've seen before.

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

The critical value is obtained from the normal table.

Population Standard Deviation Unknown and $n < 30$

If the population standard deviation, σ is unknown, then the population mean has a student's t distribution, and you will be using the t-score formula for sample means. The test statistic is very similar to that for the z-score, except that σ has been replaced by s and z has been replaced by t .

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

The critical value is obtained from the t-table. The degrees of freedom for this test is $n-1$.

Steps of a t-Test

Step 1: Set the appropriate Null and Alternative hypothesis.

Step 2: Determine the sample mean, population mean, sample standard deviation and sample size for the data. Calculate any values that are not provided.

Step 3: Calculate the **t-score** for the data using the t-score formula.

$$t = \frac{\bar{X} - \mu}{\sqrt{s^2/n}}$$

Step 4: Identify the **critical t-score**. If it is not provided, it can be found on a *t*-score table, based on the degree of freedom and alpha value. The degree of freedom is one less than the sample size. Unless specified otherwise, 0.05 should be used for the alpha value.

Step 5: Decision- Compare the calculated *t*-score to the critical *t*-score. If the calculated *t*-score is greater than the critical *t*-score, then the sample mean is statistically different than the population mean. Otherwise, there is no statistical difference between the sample and population means.

Decision rule for t-test

Case	Type of test	Reject H_0 , if
1	Two tailed test: $H_a: \mu \neq \mu_0$	$ t > t_{\alpha/2, n-1}$
2	One tailed test: $H_a: \mu > \mu_0$	$t > t_{\alpha, n-1}$
3	One tailed test: $H_a: \mu < \mu_0$	$t < -t_{\alpha, n-1}$

Example-1: A hotel manager claims that they provide breakfast of Tk 300 daily, on average to their guest. An investigator collected sample of last 25 days meal charge and find sample mean Tk 260 with standard deviation Tk 42. Do you think the hotel managers claim is right?.

Solution:

Here Population mean = 300

n=25, Sample mean =260 , S=42

(i) $H_0: \mu = 300$

$H_1: \mu \neq 300$

(ii) Since $n < 30$ and population variance unknown so we have to use

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} = -4.76, \text{ So } |t| = 4.76$$

(iii) At $\alpha = 0.05$ and $df = 25 - 1 = 24$, the critical value of t is 2.06

(iv) Decision: Since calculated t is greater than critical value of t so we reject null hypothesis.

(v) Comment: Hotel managers claim is not significant at $\alpha=0.05$

Example-2 : The oxygen saturation rate of 15 randomly selected Covid patients are recorded as—

65, 70, 68, 55, 90, 75, 82, 70, 66, 58, 92, 85

- (i) Do you think that the average oxygen saturation rate of covid patient is 85?
- (ii) Do you think that the average oxygen saturation rate of covid patient is bellow 85?
- (iii) Do you think that the average oxygen saturation rate of covid patient is more than 85?

Steps of a Z -Test

Applications:

- Test of a single population mean.
- Test of equality of two population mean.
- Test of a single population proportion.
- Test for difference between two population proportion..
- Test of a specified correlation coefficient.
- Test of equality of two population correlation coefficient.

Assumptions:

- All the observations are independent and drawn from normal population.

Cases:

- (i) Population normal, Variance known for any sample size.
- (ii) Population normal , Variance unknown and sample size is large($n > 30$).

Follow the steps::

Step 1: Set the appropriate Null and Alternative hypothesis.

Step 2: Determine the sample mean, population mean, sample standard deviation and sample size for the data. Calculate any values that are not provided.

Step 3: Calculate the **Z-score** for the data using the z-score formula.

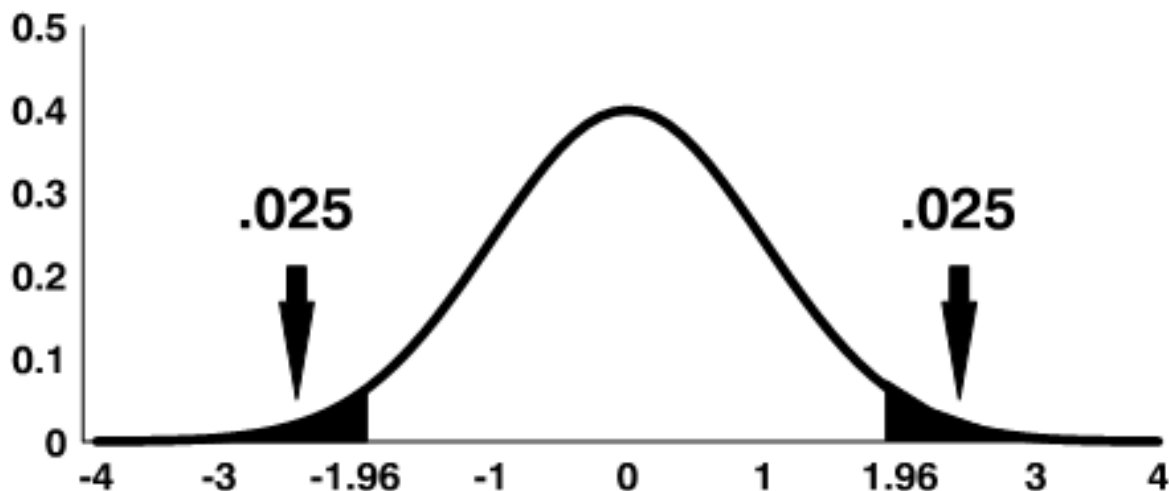
$$Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \quad \text{when population standard deviation is known.}$$

$$Z = \frac{\bar{x} - \mu}{s / \sqrt{n}} \quad \text{when population standard deviation is unknown.}$$

Step 4: Identify the **critical z-score**. If it is not provided, it can be found on a Z-score table, based on the alpha value. Unless specified otherwise, 0.05 should be used for the alpha value.

Step 5: Decision- Compare the calculated Z-score to the critical Z-score.

Cases	Types of test	Decision rule
		Reject H_0 if
1	Two tailed test $H_1: \mu \neq \mu_0$	$ Z > Z_{\frac{\alpha}{2}}$
2	Right tailed test $H_1: \mu > \mu_0$	$Z > z_{\alpha}$
3	Left tail test $H_1: \mu < \mu_0$	$Z < -z_{\alpha}$



Alternative Hypothesis	Decision for Z test		
	Alpha= .01 Reject H_0 if	Alpha= .05 Reject H_0 if	Alpha= .10 Reject H_0 if
$H_a: \mu \neq \mu_0$	$Z > 2.58$ or $Z < -2.58$	$Z > 1.96$ or $Z < -1.96$	$Z > 1.645$ or $Z < -1.645$
$H_a: \mu > \mu_0$	$Z > 2.32$	$Z > 1.645$	$Z > 1.28$
$H_a: \mu < \mu_0$	$Z < -2.32$	$Z < -1.645$	$Z < -1.28$

Confidence interval:

For Z test: $\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ For t test: $\bar{X} \pm t_{n-1, \frac{\alpha}{2}} \frac{s}{\sqrt{n}}$

Examples-1: The managing director of a firm claims that his firm produces 110 items on average daily. A random sample of 15 days gives the following results-

110, 118, 130, 140, 142, 145, 112, 100, 95, 98, 96, 122, 123, 124, 130

It is known that the number of items produced are normally distributed with variance 300.

Can we conclude at 5% level of significance that the average daily production of the firm is

- (a) 110 items
- (b) More than 110 items
- (c) Less than 110 items

Example-2: The producer of a company claims that the selling price of his product is very standard and it is Tk. 1500 per unit with standard deviation Tk. 45. The CAB has some doubt on price and they want to verify the price using statistical technique. A random sample of 100 items of this company on the selling price were collected and found average price Tk. 1510.

Can CAB conclude that the average price of the products is standard?

What happen when Population distribution is unknown???

ANS IS 'USE Z-TEST'

Test of Hypothesis concerning two population mean

The possible null and alternative hypothesis:

- (i) $H_0: \mu_1 = \mu_2$ against $H_a: \mu_1 \neq \mu_2$
- (ii) $H_0: \mu_1 = \mu_2$ against $H_a: \mu_1 > \mu_2$
- (iii) $H_0: \mu_1 = \mu_2$ against $H_a: \mu_1 < \mu_2$

Cases:

- (i) Independent samples with known population variances, sample sizes are large or small.
- (ii) Independent samples with unknown population variances, sample sizes are large.
- (iii) Independent population for small sample sizes with unknown but equal variances.
- (iv) Independent population for small sample sizes with unknown and unequal variances.
- (v) Correlated sample or matched sample (the sample obtained from a bi-variate normal population or paired observations.)

Test statistic

For (i) $Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$

For (ii) $Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim N(0,1)$

For (iii) $t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

which is a student t distribution with $n_1 + n_2 - 2$ df.

where $S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$.

For (iv) $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$

For (v) $t = \frac{\bar{d}}{se(\bar{d})} \sim t_{n-1}$

Where $d = x_i - y_i$, $\bar{d} = \frac{\sum d}{n}$, $s_d^2 = \frac{\sum (d - \bar{d})^2}{n-1}$ and $se(\bar{d}) = \frac{s_d}{\sqrt{n}}$

Example-1: A researcher wanted to investigate the hourly earnings of male and female tea garden workers. He collected data from 60 male and 60 female workers and found the following information:

	Male	Female
Mean	Tk. 48.50	Tk. 61.60
Standard deviation	Tk. 18.00	Tk. 15.00

- (i) Test the equality of mean earnings of male and female workers.
- (ii) Test whether the mean earnings of male worker is less than female worker at 5% and 1% level of significance.

Example-2: A potential buyer of electric bulbs bought 100 bulbs each of two famous brands A and B. Upon testing the bulbs he found that brand A had a mean life of 1500 hours with standard deviation of 50 hours whereas brand B had an average life of 1530 hours with standard deviation of 60 hours. Can it be concluded at 5% level of significance that the bulbs of two brands differ significantly in quality.

Example-3: Manager of a company claims that the average wages of factory I is higher than the wages of factory II. An audit firm conducted a survey on daily wages of workers of two factories and found following results:

	Factory I	Factory II
Sample size	16	11
Mean wage	Tk 290	Tk 250
Sample standard deviation	15	50

Conclude about company manager's claim.

Example-4: Two groups of students are given to solve 50 statistics problems and the time(in minute) the required are recorded as follow:

Students of Dhaka College: 35, 40, 25, 25, 45, 60, 48, 55

Students of Titumir College: 33, 40, 18, 20, 42, 52, 46, 28, 32, 39, 28

Do you think that the performance of the students of both the colleges is equal?

Example-5: A newly invented Covid-19 vaccine is applied to a group of patients. The oxygen saturation rate of before and after vaccination are recorded as –

Before vaccine	88	94	92	82	78	65	72	80	68
After vaccine	85	90	88	90	94	88	95	91	90

Do you think that the vaccine is effective?