

EMA1002 Engineering Mathematics 2

Hypothesis Testing for Mean

Basic Definitions

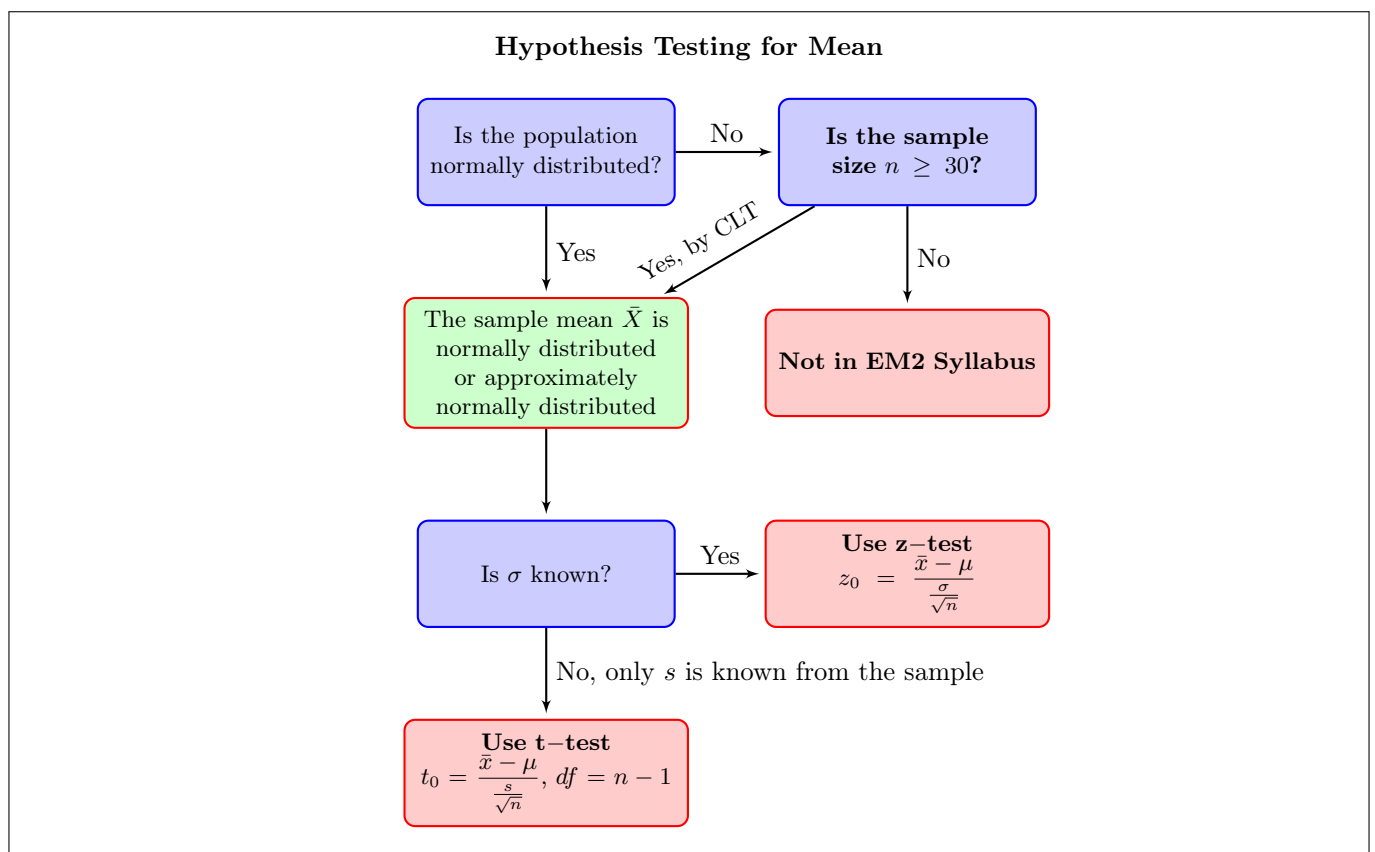
- A hypothesis is a proposed explanation for a phenomenon.
- The null hypothesis, denoted by H_0 , is a statistical hypothesis that states that there is no difference between a parameter and a specific value, or that there is no difference between the two parameters.
- The alternative hypothesis, denoted by H_1 , is a statistical hypothesis that states the existence of a difference between a parameter and a specific value, or states that there is a difference between two parameters.

Null Hypothesis	Alternative Hypothesis
$H_0 : \mu = \mu_0$	$H_1 : \mu < \mu_0$ left-tailed test (one-tailed) $H_1 : \mu > \mu_0$ right-tailed test (one-tailed) $H_1 : \mu \neq \mu_0$ two-tailed test

where μ_0 is used to denote a number. Hypothesis testing can help us answer questions such as whether:

- a new medication helps change a person's heart rate.
- a genetically modified food contains different levels of potassium.

Test statistics are sample statistics or values based on the sample data. They help us determine if we should reject or fail to reject the null hypothesis. Depending on the conditions, we have to use different hypothesis tests (z-test or t-test), as shown in the flowchart below.



Excel Formula

The formula to calculate the p-value for two-tailed z-test is: =2*(1-NORM.S.DIST(ABS(z0),TRUE))

The formula to calculate the p-value for two-tailed t-test is: =2*(T.DIST.RT(ABS(t0),n-1))

Adjust the formula accordingly when performing a one-tailed test (either left-tailed or right-tailed).

Example 1

A pharmaceutical company claims that the mean dose of a brand of drugs is 0.3 mg per pill. An independent consumer group selected a random sample of 31 pills and recorded the data. The mean of the sample is 0.317 mg and the standard deviation is 0.032 mg. Assume that the population standard deviation is 0.03 mg. Test the company's claim at a 0.5% significance level using a two-tailed test.

Step 1. **Write down the null and alternative hypotheses and the level of significance**

$$H_0 : \mu = 0.3$$

$$H_1 : \mu \neq 0.3 \text{ (two-tailed test)}$$

$$\alpha = \frac{0.5}{100} = 0.005$$

Step 2. **Compute test statistic**

Since the population standard deviation is known, use z-test.

$$z_0 = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = 3.155$$

Step 3. **Find the p value using Excel**

$$P\text{-value} = 0.00160$$

Step 4. **Compare the p value with α**

$$P\text{-value} = 0.00160 < 0.005 = \alpha$$

Step 5. **Conclusion**

There is sufficient evidence at the 0.5% significance level to reject the claim that the mean dosage is 0.3 mg.

The evidence suggests that the mean dosage differs from 0.3 mg.

Note:

The Excel formula to calculate the p-value for a two-tailed z-test is: =2*(1-NORM.S.DIST(ABS(z0),TRUE))

Example 2

A delivery company claims that its mean delivery time is 30 minutes. A customer service manager suspects that the mean delivery time exceeds 30 minutes. A random sample of 40 deliveries has a mean delivery time of 31 minutes and a sample standard deviation of 5 minutes. The population standard deviation is unknown. Test the manager's suspicion at the 5% significance level.

Step 1. **Write down the null and alternative hypotheses and the level of significance**

$$H_0 : \mu = 30$$

$$H_1 : \mu > 30 \text{ (right-tailed test)}$$

$$\alpha = \frac{5}{100} = 0.05$$

Step 2. **Compute test statistic**

Since the population standard deviation is unknown and the sample size is large, use a t-test.

$$t_0 = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{31 - 30}{\frac{5}{\sqrt{40}}} \approx 1.265, df = n - 1 = 39$$

Step 3. **Find the p value using Excel**

$$P\text{-value} = 0.10670$$

Step 4. **Compare the p value with α**

$$P\text{-value} = 0.10670 > 0.05 = \alpha$$

Step 5. **Conclusion**

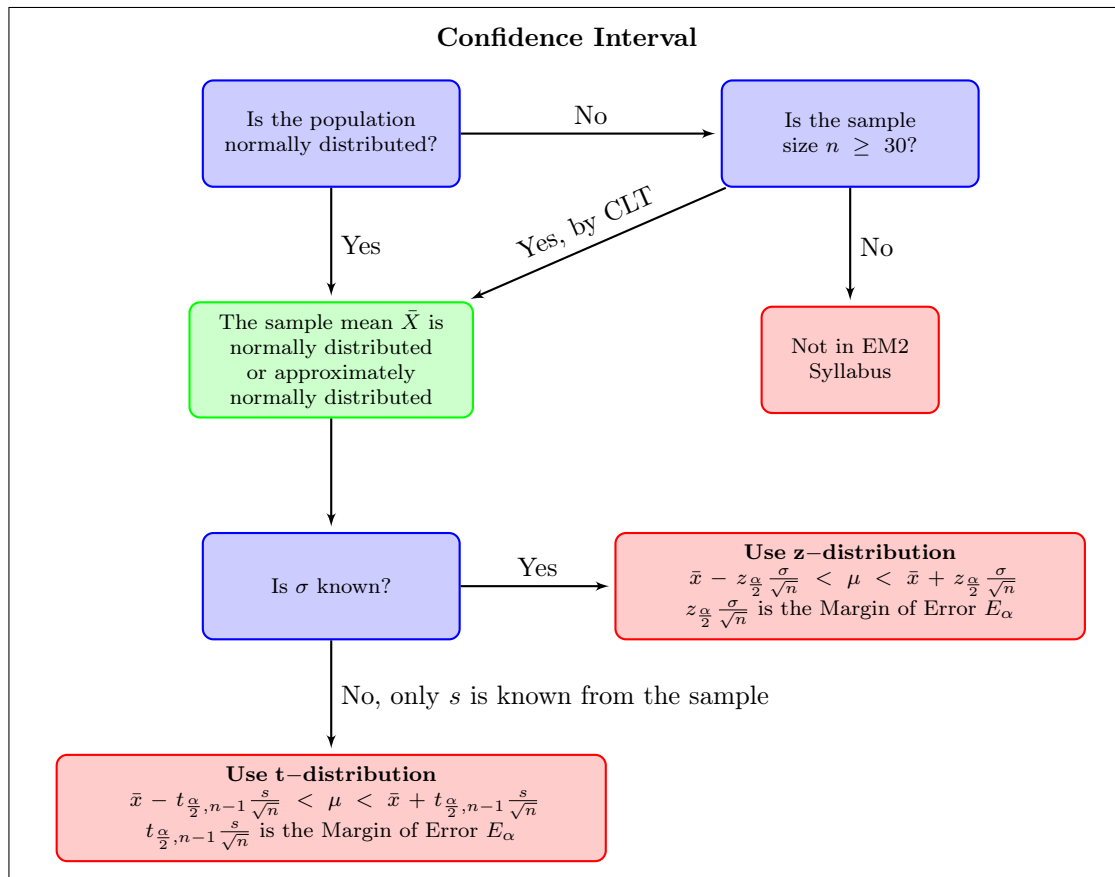
There is insufficient evidence at the 5% significance level to reject the claim that the mean delivery time is 30 minutes.

Note:

The Excel formula to calculate the p-value for a right-tailed t-test is: =T.DIST.RT(t0,n-1).

Confidence Interval for Mean

- The confidence level is $100(1 - \alpha)\%$, where α is the significance level. It refers to the reliability of the estimation process. If the sampling and estimation process were repeated many times, $100(1 - \alpha)\%$ of the confidence intervals constructed would contain the true population parameter.
- A confidence interval is a range of values calculated using sample data to estimate a population parameter at a specified confidence level.



Example 3

A random sample of 64 polytechnic students has a mean travelling time to campus of 47.5 minutes. The population standard deviation is known to be 12 minutes. Find the 95% confidence interval for the population mean travelling time to campus.

Step 1. Find the Margin of Error using Excel

Since the population standard deviation is known, use the z-distribution.

$$\alpha = 1 - 0.95 = 0.05$$

$$E_{\alpha} = 2.93995 \quad \rightarrow \quad =\text{CONFIDENCE.NORM}(0.05, 12, 64)$$

Step 2. Construct the Confidence Interval

$$\bar{x} - E_{\alpha} < \mu < \bar{x} + E_{\alpha}$$

$$47.5 - 2.93995 < \mu < 47.5 + 2.93995$$

$$44.560 < \mu < 50.440$$

Example 4

A random sample of 50 households has a mean internet download speed of 1.82 Gbps and a sample standard deviation of 0.36 Gbps. The population standard deviation is unknown. Find the 99% confidence interval for the population mean internet download speed.

Step 1. Find the Margin of Error using Excel

Since the population standard deviation is unknown, use the t-distribution.

$$\alpha = 1 - 0.99 = 0.01$$

$$E_{\alpha} = 0.13644 \quad \rightarrow \quad =\text{CONFIDENCE.T}(0.01, 0.36, 50)$$

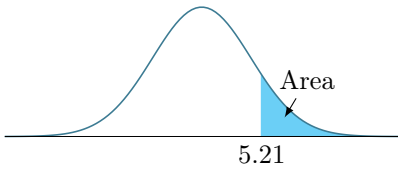
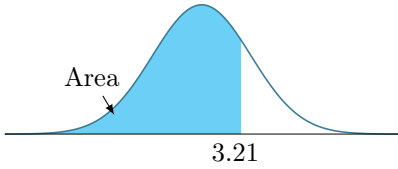
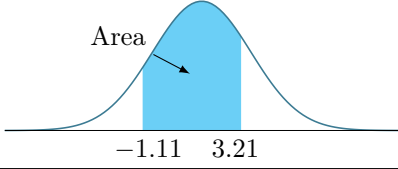
Step 2. Construct the Confidence Interval

$$\bar{x} - E_{\alpha} < \mu < \bar{x} + E_{\alpha}$$

$$1.82 - 0.13644 < \mu < 1.82 + 0.13644$$

$$1.684 < \mu < 1.956$$

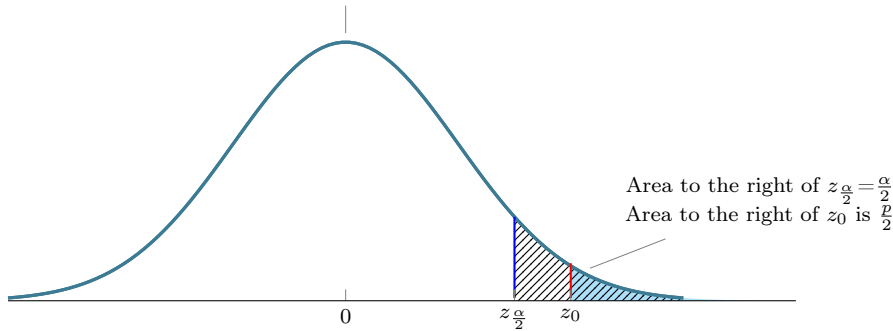
The Relationship between Probabilities and Area under Normal Curve

$\mathbb{P}(X > 5.21)$		This is the Area to the right of 5.21.
$\mathbb{P}(X < 3.21)$		This is the Area to the left of 3.21
$\mathbb{P}(-1.11 < X < 3.21)$		$\mathbb{P}(X < 3.21) - \mathbb{P}(X < -1.11)$

p-value Approach

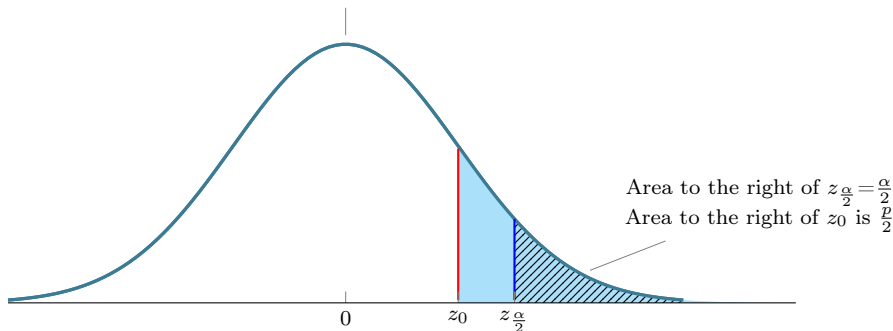
The diagrams below illustrate a two-tailed test with a positive observed test statistic z_0 . Here, p denotes the total two-tailed p-value.

This is the distribution assuming the null hypothesis H_0 is correct



Sufficient evidence to reject the null hypothesis (H_0) since $p \leq \alpha$

This is the distribution assuming the null hypothesis H_0 is correct



Insufficient evidence to reject the null hypothesis since $p > \alpha$

The above explanation works for the t -test as well. Take note that z_0 needs to be replaced with t_0 and $z_{\frac{\alpha}{2}}$ needs to be replaced with $t_{\frac{\alpha}{2}, n-1}$. z_0 and t_0 are known as the test statistics whereas $z_{\frac{\alpha}{2}}$ and $t_{\frac{\alpha}{2}, n-1}$ are known as the critical values.