Hypothesis testing

- Hypothesis is an assumption about a parameter. This assumption may or may not be true. Hypothesis testing refers to the formal procedure used by the statisticians to accept or reject this hypothesis.

- There are two types of statistical hypothesis
  1. Null hypothesis (H_0)
  2. Alternative hypothesis (H_1)

  An alternative hypothesis is what the researcher wants to prove.
  A null hypothesis is the inverse of alternative hypothesis.

- Hypothesis testing has 4 steps –

  Step 1: Null hypothesis

  Alternative hypothesis

  Step 2: Test statistic: Test statistic will give a calculated value which will use to take decision either we accept or reject H_0.

  Step 3: Rejection region: If calculated value falls in the rejection region, we reject H_0 (null hypothesis).

  Step 4: Comment. (Since the calculated value falls in the rejection region, so we reject H_0 (null hypothesis) or since the calculated value does not fall in the rejection region, so we can not reject H_0 (null hypothesis)).

- Hypothesis test for the mean (μ)

  Case 1: X has a normal distribution **with known population variance** (σ^2)

  Case 2: X has a normal distribution with unknown population variance (σ^2)

  Case 3: X has a general distribution, but we have a large sample size ($n \geq 30$).

For Case 1, Case 2 and Case 3

$H_0: \mu = \mu_0$

$H_1: \mu > \mu_0$

$or, H_1: \mu < \mu_0$

$or, H_1: \mu \neq \mu_0$

For example, We want to test NSU student's average height is greater than 5 ft.

$H_0: \mu = 5$

$H_1: \mu > 5$

We want to test NSU student's average height is lower than 5 ft.

$H_0: \mu = 5$

$H_1: \mu < 5$

If We want to test NSU student's average height is not equal to 5 ft.

$H_0: \mu = 5$

$H_1: \mu \neq 5$

Case 1: Test statistic is $\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

Here, $\bar{x}$ (Sample mean)

μ_0 (Given)

σ (Population standard deviation)

n (Sample size)

When $H_1: \mu > \mu_0$

The rejection region is $[\,Z_{\alpha}, +\infty[$ [the value of Z_{α} can be found from table page 787, alpha indicates the level of significance]

When $H_1: \mu < \mu_0$

The rejection region is $]-\infty, -Z_{\alpha}]$

When $H_1: \mu \neq \mu_0$

The rejection region is $]-\infty, -Z_{\frac{\alpha}{2}}] \cup [\,Z_{\frac{\alpha}{2}}, +\infty[$

Example: From a long-term experience, a factory owner knows that a worker can produce a product in an average time of 89 min. However, on Monday morning, there is the impression **that takes longer.** To test whether this impression is correct a sample (n=12) is taken with $\bar{x} = 92.2.$ We assume that the production time is normal with $\sigma^2 = 144.$ Verify whether this impression is correct at 5% level of significance.

Solution:

$$H_0: \mu = 89$$

$$H_1: \mu > 89$$

Test statistic is $\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

$$= \frac{92.2-89}{\frac{\sqrt{144}}{\sqrt{12}}}$$

$$= .09237$$

The rejection region is $[\,Z_{\alpha}, +\infty[$

$$= [\,Z_{0.05}, +\infty[$$

$$= [\,1.645, +\infty[$$ [From table page- 787]

Comment: Since test statistic's value (0.9237) does not fall in the rejection region, so we can't reject H_0 at 5% level of significance.

That is the factory owner impression is incorrect.

Example: From a long-term experience, a factory owner knows that a worker can produce a product in an average time of 89 min. However, on Monday morning, there is the impression that **takes lower.** To test whether this impression is correct a sample (n=12) is taken with $\bar{x} = 92.2$. We assume that the production time is normal with $\sigma^2 = 144$. Verify whether this impression is correct at 5% level of significance.

Solution:

$$H_0: \mu = 89$$

$$H_1: \mu < 89$$

Test statistic is $\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

$$= \frac{92.2-89}{\frac{\sqrt{144}}{\sqrt{12}}}$$

$$= .09237$$

The rejection region is $]-\infty, -Z_\alpha]$

$$=]-\infty\ , -Z_{0.05}]$$

$$=]-\infty, -1.645]$$ [From table page- 787]

Comment: Since test statistic's value (0.9237) does not fall in the rejection region, so we can not reject H_0 at 5% level of significance.

That is the factory owner impression is incorrect.

Example: From a long-term experience, a factory owner knows that a worker can produce a product in an average time of 89 min. However, on Monday morning, there is the impression that **it does not take 89 min**. To test whether this impression is correct a sample (n=12) is taken with $\bar{x} = 92.2$. We assume that the production time is normal with $\sigma^2 = 144$. Verify whether this impression is correct at 5% level of significance.

Solution:

$$H_0: \mu = 89$$

$$H_1: \mu \neq 89$$

Test statistic is $\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$

$$= \frac{92.2-89}{\frac{\sqrt{144}}{\sqrt{12}}}$$

$$= .09237$$

The rejection region is $]-\infty, -z_{\frac{\alpha}{2}}] \cup [z_{\frac{\alpha}{2}}, +\infty[$

$$=]-\infty, -z_{\frac{0.05}{2}}] \cup [z_{\frac{0.05}{2}}, +\infty[$$

$$=]-\infty, -Z_{0.025}] \cup [Z_{0.025}, +\infty[$$

$=]-\infty, -1.96] \cup [1.96, +\infty[$ [From table page- 787]

Comment: Since test statistic's value (0.9237) does not fall in the rejection region, so we can not reject H_0 at 5% level of significance.

That is the factory owner impression is incorrect.

Case 2: Test statistic is $\frac{\bar{x}-\mu_0}{\sqrt{\frac{s^2}{n}}} \sim t_{(n-1)}$

Here s^2 indicate sample variance where

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

[Note: 75 min, 78 min, 80 min, 92 min, 93 min

$$\bar{x} = \frac{75+78+80+92+93}{5} = 83.6$$

$$s^2 = \frac{(75-83.6)^2+(78-83.6)^2+(80-83.6)^2+(92-83.6)^2+(93-83.6)^2}{4}$$

$$= 69.3$$]

When H_1: $\mu > \mu_0$

The rejection region is $[\, t_\alpha, +\infty[$

When H_1: $\mu < \mu_0$

The rejection region is $]-\infty, -t_\alpha]$]-∞,-t (α, n-1)]

When H_1: $\mu \neq \mu_0$

The rejection region is $]-\infty, -t_{\frac{\alpha}{2}}] \cup [\, t_{\frac{\alpha}{2}}, +\infty[$

Example: From a long-term experience, a factory owner knows that a worker can produce a product in an average time of 89 min. However, on Monday morning, there is the impression that it takes longer.

To test whether this impression is correct a sample (n=5) is taken

87 89 90 92 88

Verify this impression is correct at significance level 10%.

Solution:

$$H_0: \mu = 89$$

$$H_1: \mu > 89$$

Test statistic is $\frac{\bar{x}-\mu_0}{\sqrt{\frac{s^2}{n}}} \sim t_{(n-1)}$

Here, $\bar{x} = \frac{87+89+90+92+88}{5} = 89.2$

$\mu_0 = 89,$

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

$$= \frac{(87-89.2)^2 + (89-89.2)^2 + (90-89.2)^2 + (92-89.2)^2 + (88-89.2)^2}{5-1}$$

$= 3.7$

$n = 5$

$\therefore$Test statistic is $\frac{89.2-89}{\sqrt{\frac{3.7}{5}}} = 0.2325$

The rejection region is $[\, t_\alpha, +\infty[$

$= [\, 1.533, +\infty[$

Comment: Since the test statistic's value doesn't fall in the rejection region, so we cannot reject H_0 (Null Hypothesis).

Example: From a long-term experience, a factory owner knows that a worker can produce a product in an average time of 89 min. However, on Monday morning, there is the impression that **it takes lower.**

To test whether this impression is correct a sample (n=5) is taken

87 89 90 92 88

Verify this impression is correct at significance level 10%.

Solution:

$$H_0: \mu = 89$$

$$H_1: \mu < 89$$

Test statistic is $\frac{\bar{x}-\mu_0}{\sqrt{\frac{s^2}{n}}} \sim t_{(n-1)}$

Here, $\bar{x} = \frac{87+89+90+92+88}{5} = 89.2$

$\mu_0 = 89,$

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

$$= \frac{(87-89.2)^2 + (89-89.2)^2 + (90-89.2)^2 + (92-89.2)^2 + (88-89.2)^2}{5-1}$$

$= 3.7$

$n = 5$

$\therefore$Test statistic is $\frac{89.2-89}{\sqrt{\frac{3.7}{5}}} = 0.2325$

The rejection region is $]-\infty, \ -t_\alpha \ ]$

$=] -\infty, -1.533]$

Comment: Since the test statistic's value doesn't fall in the rejection region, so we cannot reject H_0 (Null Hypothesis).

Example: From a long-term experience, a factory owner knows that a worker can produce a product in an average time of 89 min. However, on Monday morning, there is the impression that **it does not takes 89 min.**

To test whether this impression is correct a sample (n=5) is taken

87 89 90 92 88

Verify this impression is correct at significance level 10%.

Solution:

$$H_0: \mu = 89$$

$$H_1: \mu \neq 89$$

Test statistic is $\frac{\bar{x}-\mu_0}{\sqrt{\frac{s^2}{n}}} \sim t_{(n-1)}$

Here, $\bar{x} = \frac{87+89+90+92+88}{5} = 89.2$

$\mu_0 = 89,$

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

$$= \frac{(87-89.2)^2 + (89-89.2)^2 + (90-89.2)^2 + (92-89.2)^2 + (88-89.2)^2}{5-1}$$

$= 3.7$

$n = 5$

$\therefore$Test statistic is $\frac{89.2-89}{\sqrt{\frac{3.7}{5}}} = 0.2325$

The rejection region is $]-\infty, -t_{\frac{\alpha}{2}}] \cup [t_{\frac{\alpha}{2}}, +\infty[\; =]-\infty, -t_{\frac{0.1}{2}}] \cup [t_{\frac{0.1}{2}}, +\infty[$

$= \;]-\infty, -t_{0.05}] \cup [t_{0.05}, +\infty[$

$= \;]-\infty, -2.132] \cup [2.132, +\infty[$

Comment: Since the test statistic's value doesn't fall in the rejection region, so we cannot reject H_0 (Null Hypothesis).