Statistical Hypothesis

This is a claim made about the statistical measure of an unknown population parameter value. Statistical hypothesis is aimed at the following:

i. To know the meaning and techniques of formulating statistical hypothesis
ii. To learn the rudiments of accepting/rejecting Null and Alternative hypotheses
iii. To reject or accept decision based on sampling distribution of statistic.
iv. To have cleared understanding of level of significance
v. To be able to distinguish between One-tail and Two-tail tests.

Basic Concepts in Hypothesis Testing

Level of significance: is the probability level at which the decision maker concludes that observed difference between the values of the test statistic and hypothesized parameter value cannot be due to chance.

Null Hypothesis (H_0): It is the hypothesis being tested and it is usually refers to as the hypothesis of no significance or no change. .

Alternative Hypothesis (H_1): It is the hypothesis of practical interest which is expected to disprove the null hypothesis. It is the hypothesis of change or hypothesis that shows difference.

One-tailed test: It is also known as Directional hypothesis. A test is one-tailed if the alternative hypothesis states that the population parameter is greater than or less than the one being tested such that it leads to the following hypothetical statements

$H_o: \mu = \mu_o$ or $H_o: \mu = \mu_o$

$H_1: \mu > \mu_o$ $H_1: \mu < \mu_o$

$at\ \alpha\ level\ of\ significance$

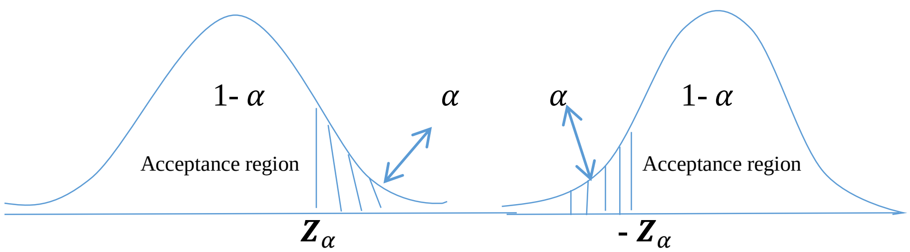
Two-tailed test: It is also known as Non-Directional hypothesis. A test is two-tailed if the alternative hypothesis H_1 states that the population parameter is different from the one being tested such that it leads to the following hypothetical statements

$H_o: \mu = \mu_o$

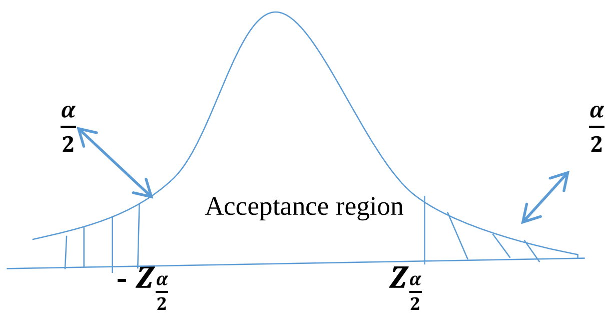
$H_1: \mu \neq \mu_o$

$at\ \alpha\ level\ of\ significance$

Critical Region: Is that region such that the null hypothesis is rejected if the test statistic computed from a particular sample takes on a value in the region. The critical region for one-tail hypothesis occupies only one side of the normal curve tails while that of two-tail lies on the two tails. This can be illustrated in the figures below



Two-tail test



Acceptance Region: is the complement of critical region and they are two mutually exclusive regions.

Type I Error: Error of rejecting null hypothesis when it is true (α). [P(rejecting $H_0|H_0$ is true)]

Type II Error: Error of accepting null hypothesis when it is false (β). [P(accepting $H_0|H_0$ is false)]

Power of a Test: Probability of rejecting H_0 when it is false ($1-\beta$) or when H_1 is true.

Critical Value: is the value of the sample statistic that separate the regions of acceptance and rejection.

Note, ($1 - \alpha$) refers to as confidence level, measures the probability level of not rejecting a true null hypothesis.

Assumptions of hypothesis testing

- Independence
- Randomness

- Normality
- Homogeneity.

Basic Steps in Hypothesis Testing

The following steps are involved in hypothesis testing

1. State the hypothesis to be tested in terms of null and alternative hypotheses.
2. Choose the appropriate test statistic to be used.
3. Choose a significance level, α based on the seriousness of rejecting H_0 when true.
4. Derive a decision rule based on α which imply finding the critical points and acceptance regions.
5. Take a decision to accept or reject H_0.

Hypothesis Test for Mean, μ

Case 1: When σ is known and n ≥ 30 (Large)

The hypothesis can be stated as follows

$H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$ or $H_1: \mu > \mu_0$ or $H_1: \mu < \mu_0$

$at\ \alpha\ level\ of\ significance$

The test statistic is given as

$$Z = \frac{\bar{X}-\mu_0}{\sigma_{\bar{X}}} = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{(\bar{X}-\mu_0)\sqrt{n}}{\sigma} \qquad (12.1)$$

Example 1

With reference to example 1 in estimation theory, test the hypothesis that the average age of the entire registered voters is significantly different from 27years at 5% level of significance.

Solution: Given $\bar{X} = 25$, $\sigma = 5$, $n = 40$, $\alpha = 0.05$, $Z_{0.025} = 1.96$, $\mu_0 = 27$

$H_0: \mu = 27$ vs $H_1: \mu \neq 27$

$$Z = \frac{(\bar{X}-\mu_0)\sqrt{n}}{\sigma} = \frac{(25-27)\sqrt{40}}{5} = -2.5298$$

$|Z| = 2.5298$

Decision: since $|Z| > 1.96$, H_0 is rejected and we conclude that the average age of voters is significantly different from 27years.

Case 2: When σ is unknown and n < 30 (Small)
The test statistic is given as

$$t = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} = \frac{(\bar{X}-\mu_0)\sqrt{n}}{S} \qquad (12.2)$$

Example 2. With reference to example 2 in estimation theory, test hypothesis against the claim that average scores of the entire students in Statistics methods exams is different from 72 at 5% level of significance.

Solution: Given n = 10, $\alpha = 0.05$, $\bar{X} = 56.8$, $S = 14.9874$, $t_{0.025}^{(9)} = 2.26.\, \mu_0 = 72$
H_0: μ = 72 vs H_1: μ ≠ 72

$$t = \frac{(\bar{X}-\mu_0)\sqrt{n}}{S} = \frac{(56.8-72)\sqrt{10}}{14.9874} = -3.2071$$

$$|t| = 3.2071$$

Decision: since $|t| > 2.26$, H_0 is rejected and we conclude that the average scores of entire students in statistics methods is significantly different from 72.

Hypothesis Test for Difference of two Means, $\mu_1 - \mu_2$

Case 1: when σ_1^2 and σ_2^2 are known and $n_1, n_2 \geq 30$
H_0: $\mu_1 = \mu_2$ vs H_1: $\mu_1 \neq \mu_2$ or H_1: $\mu_1 > \mu_2$ or H_1: $\mu_1 < \mu_2$

The test statistic is given as

$$Z = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{X}_1-\bar{X}_2}} \qquad (12.3)$$

Where $\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\sigma_{\bar{X}_1}^2 + \sigma_{\bar{X}_2}^2} = \sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}$

Case 2: when σ_1^2 and σ_2^2 are unknown and $n_1, n_2 < 30$

Given the same set of hypothesis as in case 1, the test statistic becomes

$$t = \frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_{\bar{X}_1-\bar{X}_2}} \qquad (12.4)$$

Where $S_{\bar{X}_1-\bar{X}_2} = Sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

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

$$S_1^2 = \frac{1}{n_1 - 1}\left[\sum_{i=1}^{n_1} X_{1i}^2 - \frac{\left(\sum_{i=1}^{n} X_{1i}\right)^2}{n_1}\right] S_2^2 = \frac{1}{n_2 - 1}\left[\sum_{i=1}^{n_2} X_{2i}^2 - \frac{\left(\sum_{i=1}^{n} X_{2i}\right)^2}{n_2}\right]$$

When $\sigma_1^2 \neq \sigma_2^2$ (Fisher Pschren's),

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Example 3: With reference to example 3 in estimation theory , test the hypothesis that the average cost of conducting seminars for both non-academic and academic staffs is significantly different from #900,000 at 10% level of significance.

Solution: **Given** $\boldsymbol{n_1 = 50}, \boldsymbol{n_2} = 40,\ \sigma_1 = 10; \sigma_2 = 5; \bar{X}_1 = 2{,}000{,}000; \bar{X}_2 = 1{,}000{,}000$
$\alpha = 0.10,\ \ Z_{0.05} = 1.645$, $\sigma_{\bar{X}_1 - \bar{X}_2} = 1.6202,\ \mu_2 = 900{,}000$

H_0: $\mu_1 - \mu_2$ = 900.000 vs H_1: $\mu_1 - \mu_2 \neq 900{,}000$

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1 - \bar{X}_2}} = \frac{(2{,}000{,}000 - 1{,}000{,}000) - 900{,}000}{1.6202}$$

$= 61720.7752$

Decision: since $|Z| > 1.645$, H_0 is rejected and we conclude that the average cost of conducting seminars for both non-academic and academic staffs is significantly different from #900,000.

Example 4: With reference to example 4 in estimation theory, test the hypothesis against a claim that there is no significant difference in the performance of the two production lines.

Solution: Given $n_1 = 10, n_2 = 8, \alpha = 0.10, \bar{X}_1 = 9.78, \bar{X}_2 = 7.95, v = 16, t_{0.025}^{(16)} = 2.120$

$S_{\bar{x}_1 - \bar{x}_2}$ = 0.0407563238

H_0: $\mu_1 - \mu_2$ = 0 vs H_1: $\mu_1 - \mu_2 \neq 0$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{\bar{X}_1 - \bar{X}_2}} = \frac{(9.78 - 7.95) - 0}{0.0407563238} = 44.90$$

Decision: since $|t| > 2.120$, H_0 is rejected and we conclude that there is significant difference in the performance of the two production lines.

Hypothesis Test for Proportion

H_0: $P = P_0$ vs H_1: $P \neq P_0$

The test statistic becomes

$$Z = \frac{\hat{P} - P_0}{\sigma_{\bar{P}}} = \frac{\hat{P} - P_0}{\sqrt{\hat{P}(1-\hat{P})/n}} \qquad (12.5)$$

Example 5: With reference to example 5 in estimation theory, test hypothesis against the claim that the proportion of registered voters who cast their votes for APC is significantly different from 0.30.

Solution: Given $n = 800, x = 400, \alpha = 0.05,\ \hat{P} = 0.5,\ \sigma_P = 0.025, Z_{0.025} = 1.96$

H_0: $P = 0.3$ vs H_1: $P \neq 0.3$

$$Z = \frac{\hat{P} - P_0}{\sigma_{\bar{P}}} = \frac{0.5 - 0.3}{0.025} = 8$$

Decision: since $|Z| > 1.96$, H_0 is rejected and we conclude that the proportion of registered voters who cast their votes for APC is significantly different from 0.30.

Hypothesis Test for Difference of two Proportions, $P_1 - P_2$

H_0: $P_1 = P_2$ vs H_1: $P_1 \neq P_2$ or H_1: $P_1 > P_2$ or H_1: $P_1 < P_2$

Given the above set of hypothesis, the test statistic becomes

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sigma_{\hat{P}_1 - \hat{P}_2}} \qquad (12.6)$$

Where $\sigma_{\hat{P}_1 - \hat{P}_2} = \sqrt{\sigma^2_{\hat{P}_1} + \sigma^2_{\hat{P}_2}} = \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{n_1} + \frac{\hat{P}_2(1-\hat{P}_2)}{n_2}}$

Example 6: With reference to example 6 in estimation theory, test the hypothesis against the claim that there is significant difference between the proportions of students who passed business statistics in the two departments.

Solution: Given $\boldsymbol{n_1 = 50},\ \boldsymbol{x_1} = 40, \boldsymbol{n_2} = 40,\ x_2 = 20,\ \hat{P}_1 = 0.8, \hat{P}_2 = 0.5,$

$\sigma_{\hat{P}_1 - \hat{P}_2} = 0.0972111105,$ $Z_{0.025} = 1.96$

H_0: $P_1 = P_2$ vs H_1: $P_1 \neq P_2$

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sigma_{\hat{P}_1 - \hat{P}_2}} = \frac{(0.8 - 0.5) - 0}{0.0972111105} = 3.0861$$

Decision: since $|Z| > 1.96$, H_0 is rejected and we conclude that there is significant difference between the proportions of students who passed business statistics in the two departments.