

Hypothesis Testing

1- formulate a hypothesis

A statistical hypothesis $\rightarrow$ is a statement about numerical value of population parameter

In each test $\rightarrow$ Null (H_0) , $\mu = 5.5$
 $\rightarrow$ Alternative (H_1) , $\mu \neq 5.5$

Simple vs. composite

$\mu = 5.5$ $\mu > 5.5$

$\mu \neq 5.5$ $\mu < 5.5$

2- Calculate value of test statistics

Test procedure partitions possible value of test stat into two subsets $\rightarrow$ acceptance

$\rightarrow$ non-rejection region for H_0

$\rightarrow$ rejection region for H_0

3- Decide on the result of hypo. test

Rejection of H_0 when it is true $\rightarrow$ Type I error

$\hookrightarrow$ Probability of committing Type I error is denoted by α

Non-rejection of H_0 when it is false $\rightarrow$ Type II error

$\hookrightarrow$ Prob. of committing Type II error is denoted by β

Size of critical (rejection) region $\Rightarrow$ is prob. α of committing Type I error

$\hookrightarrow$ This prob. is called as level of significance

	H_0 True	H_0 False
Not Reject H_0	correct decision Prob $\rightarrow 1 - \alpha$	Type II error Prob $\rightarrow \beta$
Reject H_0	Type I error prob $\rightarrow \alpha$	correct decision Prob $\rightarrow 1 - \beta$
		$\xrightarrow{\quad}$ Power test

To perform a hypothesis test

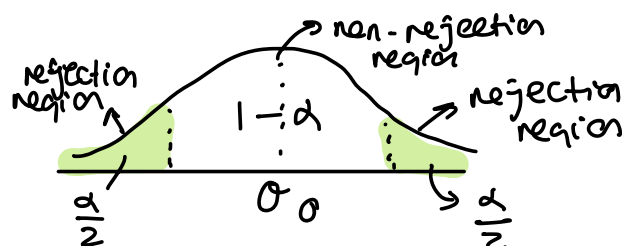
- 1- State the hypothesis
- 2- State Decision rule
- 3- Compute value of test stats
- 4- Decide $\rightarrow$ Reject H_0
 $\rightarrow$ not reject H_0

Suppose we want to test $H_0: \theta = \theta_0$

Alternative hypothesis $\left\{ \begin{array}{l} H_1: \theta \neq \theta_0 \rightarrow \text{two tailed test} \\ \text{or} \\ H_1: \theta < \theta_0 \text{ or } H_1: \theta > \theta_0 \rightarrow \text{one tailed test} \end{array} \right.$

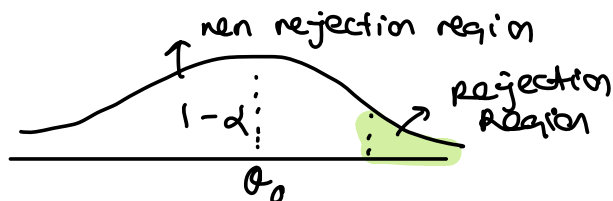
$$H_0: \theta = \theta_0$$

$$H_1: \theta \neq \theta_0$$



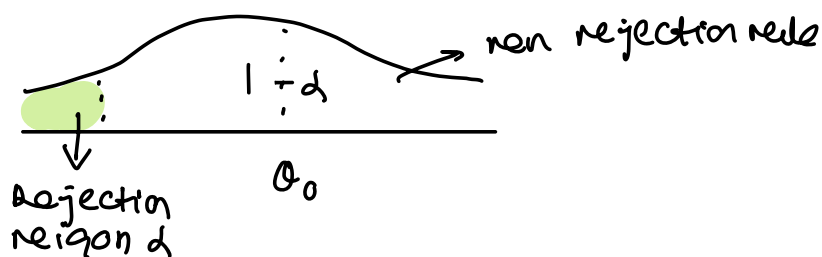
$$H_0: \theta = \theta_0$$

$$H_1: \theta > \theta_0$$



$$H_0: \theta = \theta_0$$

$$H_1: \theta < \theta_0$$



Tests of mean

$$n \rightarrow \begin{array}{l} \geq 30 \\ \rightarrow \infty \end{array}$$

$$\sigma^2 \rightarrow \begin{array}{l} \text{known} \\ \rightarrow \text{unknown} \end{array}$$

$$H_0: \mu = \mu_0$$

$$H_1: \mu > \mu_0$$

$$H_0: \mu \neq \mu_0$$

or

$$H_1: \mu < \mu_0$$

By using a random sample of size n from a normal population with known variance, σ^2

Calculate ^①
$$Z = \frac{\bar{X} - M_0}{\sigma / \sqrt{n}}$$

Critical regions for alternative hypothesis H_0 is rejected

$$\left\{ \begin{array}{l} * H_1: \mu \neq M_0 \\ |Z| \geq Z_{\alpha/2} \\ * H_1: \mu > M_0 \\ Z \geq Z_{\alpha} \\ * H_1: \mu < M_0 \\ Z < -Z_{\alpha} \end{array} \right.$$

We can also use prob. statements

$$\left. \begin{array}{l} P(Z \geq Z_{\alpha/2}) < \alpha/2 \\ P(Z > Z_{\alpha}) > \alpha \\ P(Z < -Z_{\alpha}) < \alpha \end{array} \right\} \text{reject } H_0$$

Ex: A firm claims that contents of boxes of cement sold weight in average 16 kgs. Distribution of weights is known to be normal with s.t. dev. of 0.4 kg

A random sample 16 boxes yielded a sample mean weight of 15.84 kg. Test at 10% significance level that population mean weight is 16 kgs

$$H_0: \mu = 16$$

$$H_1: \mu \neq 16$$

$$\begin{aligned} M_0 &= 16 \\ \bar{X} &= 15.84 \\ \sigma &= 0.4 \\ n &= 16 \end{aligned}$$

$$Z = \frac{15.84 - 16}{0.4 / \sqrt{16}} = -1.60$$

$$|Z| < Z_{\alpha/2}$$

↳ not reject H_0

$$H_0: \mu = 16$$

$$H_1: \mu < 16$$

$$\begin{aligned} \alpha &= 0.10 \\ \alpha/2 &= 0.05 \\ Z_{0.05} &= 1.645 \end{aligned}$$

When we have $n \geq 30$ from a non-normal population, we can use large sample test

$$\textcircled{2} \quad Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

If σ^2 is unknown replace by S^2 if $n \geq 30$

$$\textcircled{3} \quad Z = \frac{\bar{X} - \mu}{S / \sqrt{n}} \sim N(0,1)$$

When $n < 30$, σ^2 is unknown

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1} \longrightarrow \begin{aligned} |t| &\geq t_{\alpha/2, n-1} \quad (H_1: \mu \neq \mu_0) \\ t &\geq t_{\alpha, n-1} \quad (H_1: \mu > \mu_0) \\ t &\leq -t_{\alpha, n-1} \quad (H_1: \mu < \mu_0) \end{aligned}$$

Ex: In a collective bargaining a firm claims average weekly earnings is \$400 on the other hand, trade union takes a random sample of 15 workers and calculate their weekly earnings. have an average of \$381.35 with s.t. dev of \$48.60. Test firm's claim at 5% significance

$$\begin{aligned} n &= 15 \\ \bar{x} &= 381.35 \\ \mu_0 &= 400 \end{aligned}$$

$$t = \frac{381.35 - 400}{48.60 / \sqrt{15}} = -3.189$$

$$\begin{aligned} S &= 48.60 \\ \alpha &= 0.05 \end{aligned}$$

$$t_{\alpha/2, n-1} = t_{0.025, 14} = 2.145$$

$$|t| > t_{\alpha/2, n-1} \Rightarrow \text{Reject } H_0$$

$$\left. \begin{aligned} H_0: \mu &= 400 \\ H_1: \mu &\neq 400 \end{aligned} \right\} \text{hypothesis}$$

Tests for Differences Between Means

Suppose that we have independent random samples of sizes n_1 & n_2 from two normal populations. Having means μ_1 & μ_2 & known variances σ_1^2 & σ_2^2 we want to test

$$H_0: \mu_1 - \mu_2 = \delta$$

$$H_1: \mu_1 - \mu_2 \neq \delta \text{ or } H_1: \mu_1 - \mu_2 > \delta \text{ or } H_1: \mu_1 - \mu_2 < \delta$$

Calculate

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

reject H_0 :

$$H_1: \mu_1 - \mu_2 \neq \delta \rightarrow |Z| \geq z_{\alpha/2}$$

$$H_1: \mu_1 - \mu_2 > \delta \rightarrow Z \geq z_{\alpha}$$

$$H_1: \mu_1 - \mu_2 < \delta \rightarrow Z \leq -z_{\alpha}$$

When n_1 & n_2 are small, σ_1^2 & σ_2^2 are unknown (we assume $\sigma_1^2 = \sigma_2^2$)

$$t = \frac{\bar{X}_1 - \bar{X}_2 - \delta}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Pooled variance is

$$s_p^2 = \frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}$$

Reject H_0 :

$$\star H_0: \mu_1 - \mu_2 \neq \delta \rightarrow |t| > t_{\alpha/2, n_1 + n_2 - 1}$$

$$\star H_1: \mu_1 - \mu_2 > \delta \rightarrow t > t_{\alpha, n_1 + n_2 - 1}$$

$$\star H_1: \mu_1 - \mu_2 < \delta \rightarrow t < -t_{\alpha, n_1 + n_2 - 1}$$

Ex: It is claimed in a certain industry that there is no difference between average weekly wages paid to male & female workers. According to a survey for males, $\bar{X}_1 = 78$, $s_1 = 24.4$, $n_1 = 4$, for females $\bar{X}_2 = 63.5$, $s_2 = 20.2$, $n_2 = 4$. Assume that population distributions are normal with equal variances & samples are independent. For 5% significance level, test the claim that there is no difference between average weekly wages of male & female workers

$$\begin{array}{lll} \bar{X}_1 = 78 & \bar{X}_2 = 63.5 & H_0: \mu_1 - \mu_2 = 0 \\ s_1 = 24.4 & s_2 = 20.2 & H_1: \mu_1 - \mu_2 \neq 0 \\ n_1 = 4 & n_2 = 4 & \\ \delta = 0 & \alpha = 0.05 & \end{array}$$

$$s_p^2 = \frac{3 \cdot (24.4)^2 + (3) (20.2)^2}{4 + 4 - 2} = 501.7$$

$$s_p = \sqrt{501.7} = 22.4 \quad t = \frac{78.0 - 63.5 - 0}{22.4 \sqrt{\frac{1}{4} + \frac{1}{4}}} = 0.915$$

$$t_{0.025, 16} = 2.1197$$

$$t \leq t_{0.025, 16} \rightarrow \text{Not reject } H_0$$