5.2 Sampling theory

Introduction:

- Entire group of individuals under study is called **population**. Quantity associated with population like mean(μ), SD(σ) is called **parameter**.

- A small part of the population is called **sample**. Quantity associated with sample like mean($\bar{x}$), SD(s) is called **statistic**.

- The number of units in the sample is called sample size. It is denoted by n. If $n < 30$, the sample is called **small sample**. If $n \geq 30$ the sample is called **large sample**.

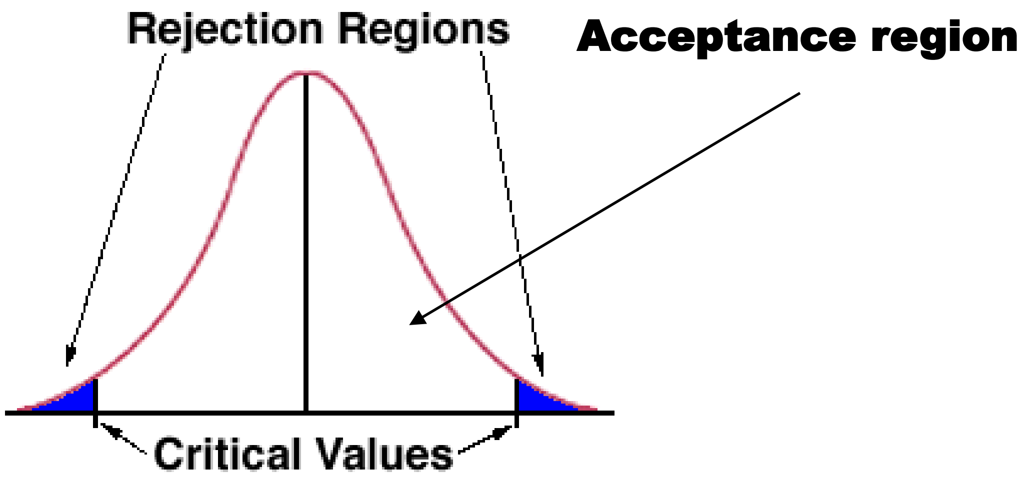
- The distribution of values of the statistic for different samples is called **sampling distribution** of the statistic. The SD of sampling distribution of a statistic is called **standard error** of a statistic.

Test of significance

- Some assumption about the population is called **statistical hypothesis**.
- A statistical hypothesis which we formulate to check whether it can be rejected is called **null hypothesis** (H_0). A hypothesis which differs from the null hypothesis is called **alternative hypothesis** (H_1).
- Procedure which enables to decide whether to accept or reject the null hypothesis is called **test of hypothesis** or **test of significance**.

Acceptance region and critical region:

- The limits of the critical region are called **critical values**.
- Critical value splits the region in to **acceptance region** and **critical region**.



Critical value of z

If $H_1 : \bar{x} \neq \mu_0$ then the test is two tailed. Use $z_{\alpha/2}$.

If $H_1 : \bar{x} < (>) \mu_0$ then the test is singled tailed. Use z_{α}.

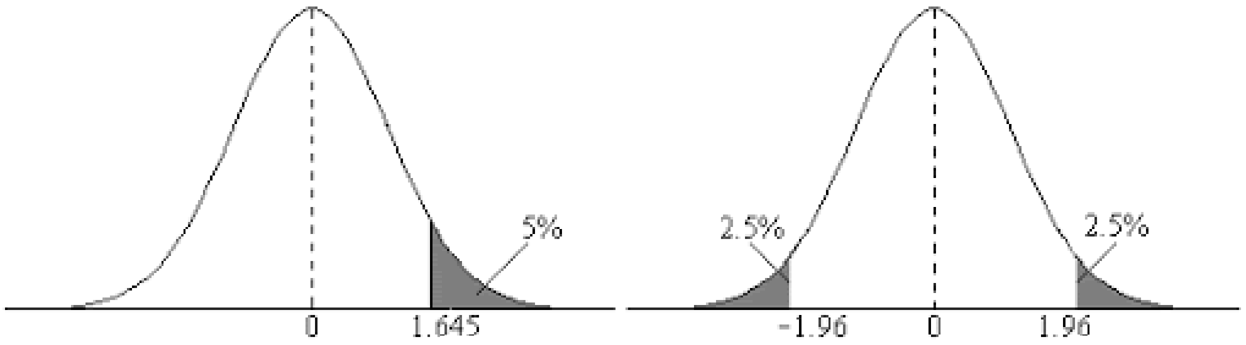
At 1% level of significance

$z_{\alpha} = z_{0.01}$ $R(x) = 0.01$ $x = 2.33$	$z_{\alpha/2} = z_{0.05}$ $R(x) = 0.005$ $x = 2.58$

At 5% level of significance

$z_{\alpha} = z_{0.05}$ $R(x) = 0.05$ $x = 1.65$	$z_{\alpha/2} = z_{0.025}$ $R(x) = 0.025$ $x = 1.96$

	$z_{\alpha/2}$	z_{α}
$\alpha = 0.05$	1.96	1.65
$\alpha = 0.01$	2.58	2.33



(a) One-tailed test z_{α} (b) Two-tailed test $z_{\alpha/2}$

Calculated value of z

$$|z| = \left|\frac{x-\mu}{\sigma}\right| (or) \left|\frac{\bar{x}-\mu}{S.E\,(\bar{x})}\right| (or) \left|\frac{\overline{x_1}-\overline{x_2}}{SE(\overline{x_1}-\overline{x_2})}\right|$$

Use the 1st result if x, μ are known.

Use the 2nd result if $\bar{x}, \mu$ are known.

Use the 3rd result if $\bar{x}_1, \bar{x}_2$ are known.

where $S.E(\bar{x}) = \sqrt{\frac{s^2}{n}}$, $S.E(\overline{x_1}-\overline{x_2}) = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$

	Mean	**S.D**
Sample	$\bar{x}$	s
Population	μ	σ

Type 1 error and Type 2 error

- Rejecting H_0 when it is true is called **Type I error**. P (Type I error) is called **level of significance**. It is denoted by α.

- Accepting H_0 when it is false is called **Type II error**. P (Type II error) is called **power of the test**. It is denoted by β.

	True	False
Accept H_0	Correct decision	Type II error
Reject H_0	Type I error	Correct decision

5.1 Test of significance - z test

Working rule:

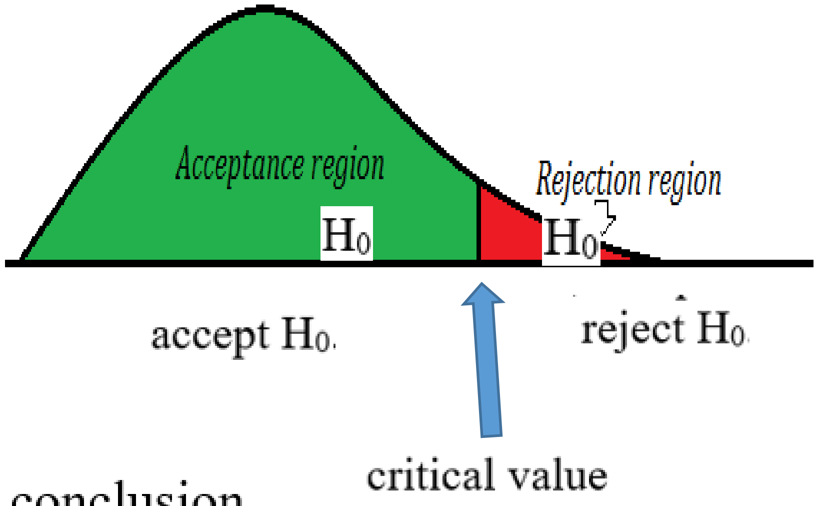
- ❖ Write the null hypothesis H_0.
- ❖ Find the calculated value using

$$|z| = \left|\frac{x-\mu}{\sigma}\right| \; (or) \; \left|\frac{\bar{x}-\mu}{S.E\,(\bar{x})}\right| \; (or) \; \left|\frac{\overline{x_1}-\overline{x_2}}{SE(\overline{x_1}-\overline{x_2})}\right|$$

where $S.E(\bar{x}) = \sqrt{\frac{s^2}{n}}$, $S.E(\overline{x_1}-\overline{x_2}) = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$

- ❖ Find the critical value using the table:

	Two tailed	Single tailed
$\alpha = 0.05$	1.96	1.65
$\alpha = 0.01$	2.58	2.33



- ❖ If calculated value < critical value, accept H_0. H_0 is the conclusion.
- ❖ If calculated value > critical value reject H_0. H_1 is the conclusion.

1. **A coin was tossed 400 times and the head turned up 216 times. Test the hypothesis that the coin is unbiased at 5% level of significance.** $[z_{\frac{\alpha}{2}} = 1.96]$

Since $n = 400$, apply z test.

By data, $x = 216, p = \frac{1}{2}, \ \mu = np = 200$

$H_0: \mu = 200$, The coin is unbiased.

$$|z| = \left|\frac{x - \mu}{\sigma}\right|$$

$$= \left|\frac{x - \mu}{\sqrt{npq}}\right|$$

$$= \left|\frac{216 - 200}{10}\right|$$

$$= 1.6$$

Therefore, calculated value of $z = 1.6$

At $\alpha = 0.05$, critical value of $z = 1.96$

Since calculated value < critical value,

Accept H_0.

Therefore, the coin is unbiased at 5% level of significance.

2. **A die was thrown 9000 times and a throw of 5 or 6 was obtained 3240 times. On the assumption of random throwing, do the data indicate an unbiased die? ($\alpha = 0.01$)**

$[z_{\frac{\alpha}{2}} = 2.58]$

Since $n = 9000$, apply z test.

By data, $x = 3240,\ p = \frac{1}{3},\ \mu = np = 3000$

$H_0: \mu = 3000$, The die is unbiased.

$$|z| = \left|\frac{x-\mu}{\sigma}\right|$$

$$= \left|\frac{x-\mu}{\sqrt{npq}}\right|$$

$$= \left|\frac{3240-3000}{\sqrt{2000}}\right|$$

$$= 5.4$$

Therefore, calculated value of $z = 5.4$

At $\alpha = 0.01$, critical value of $z = 2.58$

Since calculated value> critical value,

Reject H_0.

Therefore, the die is biased

at 1% level of significance.

3. **In 324 throws of a die, an odd number turned up 181 times. Is it reasonable to think that ay 1% level of significance the die is an unbiased one?**$[z_{\frac{\alpha}{2}} = 2.58]$

Since $n = 324$, apply z test.

By data, $x = 181,\ p = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2},$

$\mu = np = 162$

$H_0: \mu = 162$,The die is unbiased.

$$|z| = \left|\frac{x-\mu}{\sigma}\right|$$

$$= \left|\frac{x-\mu}{\sqrt{npq}}\right|$$

$$= \left|\frac{181-162}{\sqrt{162 \times \frac{1}{2}}}\right| = 2.11$$

Therefore, calculated value of $z = 2.11$

At $\alpha = 0.01$, critical value of $z = 2.58$

Since calculated value < critical value,

accept H_0.

Therefore, the die is unbiased

at 1% level of significance.

4. **A sample of 900 members is found to have a mean of 3.4 cm. Can it be reasonably regarded as a truly random sample from a large population with mean 3.25 cm and SD 1.61 cm.**

By data, $\bar{x} = 3.4, n = 900,\ \mu = 3.25, \sigma = 1.61$

This is a large sample. Apply z test.

$$S.E(\bar{x}) = \sqrt{\frac{1.61^2}{900}} = 0.0537$$

$H_0: \mu = 3.25,$

Sample is taken from the population with mean 3.25

$$|z| = \left|\frac{\bar{x} - \mu}{S.E(\bar{x})}\right|$$

$$= \left|\frac{3.4 - 3.25}{0.0537}\right| = 2.8$$

Therefore, calculated value of $z = 2.8$

At $\alpha = 0.05$, critical value of $z = 1.96$

Since calculated value > critical value,

Reject H_0.

Therefore, sample is not taken from the population with mean 3.25

5. **If a mean breaking strength of copper wire is 575 lbs with a standard deviation 8.3 lbs. How large a sample must be used in order that there be one chance in 100 that the mean breaking strength of the sample is less than 572 lbs. ($Z_\alpha = 2.33$)**

By data, $\bar{x} = 572,\ \mu = 575, \sigma = 8.3$.

This is a large sample. Apply z test.

$$S.E(\bar{x}) = \sqrt{\frac{8.3^2}{n}}$$

$H_0: \mu = 575$,

mean breaking strength of copper wire is 575 lbs.

$$|z| = \left|\frac{\bar{x}-\mu}{S.E(\bar{x})}\right|$$

$$= \left|\frac{572-575}{\sqrt{\frac{8.3^2}{n}}}\right|$$

To find: n such that $\mu > 575$.

Suppose $\mu > 575$

Calculated value > Critical value.

$$\left|\frac{572-575}{\sqrt{\frac{8.3^2}{n}}}\right| > 2.33$$

$$n > 41.56$$

Therefore, $n = 42$.

6. **The means of samples of sizes 1000 and 2000 are 67.5 and 68.0 cm respectively. Can the samples be regarded as drawn from the same population of SD 2.5 cm?**$[z_{\frac{\alpha}{2}}(0.05) = 1.96]$

Since $n_1 = 1000,\ n_2 = 2000$, apply z test.

By data, $\overline{x_1} = 67.5, \overline{x_2} = 68.0$

$\sigma_1 = \sigma_2 = 2.5, \alpha = 0.05$

H0: $\mu_1 = \mu_2$ Both the samples are drawn from the same population.

$$|z| = \left|\frac{\overline{x_1} - \overline{x_2}}{SE(\overline{x_1} - \overline{x_2})}\right|$$

$$= \left|\frac{\overline{x_1} - \overline{x_2}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}\right|$$

$$= \left|\frac{67.5 - 68}{\sqrt{2.5/1000 + 2.5/2000}}\right|$$

$$= 5.1$$

Therefore, calculated value of $z = 5.15$

At $\alpha = 0.05$, critical value of $z = 1.96$

Since calculated value > critical value,

Reject H0.

Therefore, Both the samples are not drawn from the same population.

7. **A sample of height of 6400 soldiers has a mean of 67.85 inches and a standard deviation of 2.56 inches while a sample of height of 1600 sailors has a mean of 68.55 inches and a SD of 2.52 inches. Does the data indicate that the sailors are on an average taller than soldiers? Use 0.05 level of significance. $[z_\alpha = 1.65]$**

Since $n_1 = 6400, n_2 = 1600$, apply z test.

By data, $\overline{x_1} = 67.85$, $\overline{x_2} = 68.55$,

$s_1 = 2.56$, $s_2 = 2.52$, $\alpha = 0.05$

H$_0$: $\mu_1 = \mu_2$, The sailors are not taller than soldiers.

$$|z| = \left|\frac{\overline{x_1} - \overline{x_2}}{SE(\overline{x_1} - \overline{x_2})}\right| = \left|\frac{\overline{x_1} - \overline{x_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}\right|$$

$$= \left|\frac{67.85 - 68.55}{\sqrt{2.56^2/6400 + 2.52^2/1600}}\right| = \frac{0.7}{0.005} = 140$$

Therefore, calculated value = 140

At $\alpha = 0.05$, critical value = 1.65

Since calculated value > critical value,

Reject H$_0$.

Therefore, the sailors are taller than soldiers

at 0.05 level of significance.