8/2/13

# Unit - III

## Chapter - I

**Estimation:** The process of finding out the parameter.

**Estimator:** The statistic, which is used to estimate the parameter.

**Max. Permisible error:** The deviation (max.) that is allowed b/n statistic and the parameter. Represented by E.

Ex. In estimating population mean ($\mu$) where sample mean ($\bar{x}$) is used as an estimator, we get MPE.

$$MPE = E = Z_{\alpha/2}\, SE(\bar{x})$$

$$\Rightarrow \boxed{E = Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)}$$

2. In estimating population proportion, P where sample proportion (p) is used as an estimator, we get,

$$E = Z_{\alpha/2}\, SE(p)$$

$$\boxed{E = Z_{\alpha/2}\sqrt{\frac{PQ}{n}}}$$

**Derivation for E in estimating population mean ($\mu$)**

$$P(\text{acceptance}) = P(\text{acc}) = 1-\alpha.$$

$$P\left\{ |Z_{cal}| \le Z_{tab} \right\} = 1-\alpha$$

$$P\left\{ \left|\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| \le Z_{\alpha/2} \right\} = 1-\alpha$$

$$\left|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right| \le Z_{\alpha/2} \Rightarrow |\bar{x}-\mu| \le Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right).$$

$$\bar{x} - Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{x} + Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right).$$

**Interval estimation:** The $(1-\alpha)$ interval estimation for population mean ($\mu$) $\boxed{\bar{x} - Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \bar{x} + Z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)}$

The interval estimation for population proportion is

$$p - z_{\alpha/2}\left(\sqrt{\frac{PQ}{n}}\right) \leq P \leq p + z_{\alpha/2}\left(\sqrt{\frac{PQ}{n}}\right)$$

**Type-I & Type-II errors:**

**Type-I error:** Reject $H_0$ when $H_0$ is good

ie; Rej $H_0 / H_0$

$P(\text{Type-I}) = \alpha$.

It is known as producer's risk.

**Type-II error:** Accepting $H_0$ when $H_1$ is good

Acc $H_0 / H_1$.

$P(\text{Type-II}) = \beta$

It is known as consumer's risk.

# Testing of Hypothesis

## Large Sample Tests $(n > 30)$

I) Single Population mean.

(i) Null Hypothesis: $\mu = \mu_0 \rightarrow H_0$

(ii) Alternative Hypothesis: $\mu = \mu_0 \rightarrow H_1$

(iii) Fix the LOS, $\alpha$. (LOS: Level of Significance)

(iv) Under $H_0$, compute the test statistic

$$Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$\bar{x}$ = Sample mean

$\mu$ = Population mean under $H_0$

$\sigma$ = population standard deviation

$n$ = Sample size

(v) Compare $Z_{cal}$ with $Z_{tab}$.

If $Z_{cal}$ is in acceptance region, accept $H_0$ or else reject $H_0$.

II Two population mean differences.

(i) Null hypothesis: Two means are equal.

$$H_0 : \mu_1 = \mu_2$$

(ii) Alternative Hypothesis: Two means are not equal.

$$H_1 : \mu_1 \neq \mu_2$$

(iii) Fix, the level of significance (LOS), $\alpha$.

(iv) Under $H_0$, compute the test statistic.

$$Z_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

(v) Compare $Z_{tab}$ and $Z_{cal}$

If $Z_{cal}$ is in accepted region, accept $H_0$, otherwise reject $H_0$.

Note: 1) If $(\sigma_1^2, \sigma_2^2)$ are unknown, use sample variances, $(s_1^2, s_2^2)$.

2) If they have common variance $\sigma_1^2 = \sigma_2^2 = \sigma^2$

then $$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{1}{n_1^2} + \frac{1}{n_2^2}\right)\sigma^2}}$$

3) If the common variance is unknown then

$$\sigma^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$

5/3/13 Testing a numeric value with population proportion.

Null hypothesis: The $pop^n$ $prop^n$ assumes the given value

ie; $H_0 : P_0$

- Alternative hypothesis: The $pop^n$ $prop^n$ is not equal to the given value

ie; $H_1 : P \neq P_0$

- fix the LOS, $\alpha$.
- Under $H_0$, compute the test statistic, $z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}} \sim N(0,1)$

Conclusion: Compare $Z_{cal}$ with $Z_{tab}$.

if $Z_{cal}$ is in the acceptance region, accept $H_0$ otherwise reject $H_0$

7/3/13

## Test for difference b/n 2 pop^n prop^n

Null hypothesis : The two pop^n proportions are equal. ~~H0~~ ie;

$H_0 : P_1 = P_2$

Alternative hypothesis: The two proportions are not equal, ie,

$H_1 : P_1 \neq P_2$ .

fix LOS, $\alpha$. under $H_0$, compute the test statistic.

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}}} \sim N(0,1).$$

Conclusion: Compare $Z_{cal}$ with $Z_{tab}$. If $Z_{cal}$ is in the acceptance region, accept $H_0$ otherwise reject $H_0$.

Note: If we have to test diff. b/n 2 groups drawn from same pop^n, then $Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ where

$$P = \frac{n_1 p_1 + n_2 p_2}{n_2 + n_2}$$

IV Chi-Square Test for Independence of attributes.

(i) <u>Null hypothesis:</u> The attributes are independent

(ii) <u>Alternative hypothesis:</u> They are not independent.

(iii) Fix the ~~exact~~ level of significance, $\alpha$

(iv) Under $H_0$, compute the test statistic,

$$\chi^2 = \sum\sum \left[ \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \right] \sim \chi^2_{(m-1)(n-1)}$$

$O_{ij}$ = observed frequency in $i^{th}$ row and $j^{th}$ column.

$E_{ij}$ = Expected frequency in $i^{th}$ row and $j^{th}$ column.

$= (R_{ij} * C_{ij})/N$.

(v) Compare $\chi^2_{cal}$ with $\chi^2_{tab}$.

If $\chi^2_{cal}$ is in acceptance region, then accept $H_0$ otherwise reject $H_0$.

Note:

If we have $2\times 2$ contingency table

| | |
|---|---|
| a | b |
| c | d |

(i) If all expected frequencies are $>5$.

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi^2_1$$

(ii) If any of Expected frequencies < 5

$$\chi^2 = \frac{N\left[\,|ad-bc| - N/2\right]^2}{(a+b)(c+d)(a+c)(b+d)} \sim \chi^2.$$

This is known as yate's correction.

3. Chi-Square test for single population variance.

(i) <u>Null hypothesis:</u> The population variance assumes the given value.

that is $H_0 : \sigma^2 = \sigma_0^2$.

(ii) <u>Alternative Hypothesis:</u> The population variance doesnot assume the given value

$H_1 : \sigma^2 \neq \sigma_0^2$.

(iii) Fix the level of significance, $\alpha$.

Under $H_0$, compute

$$\chi^2 = \frac{ns^2}{\sigma^2} \sim \chi^2_{n-1}$$

$n$ = sample size.

$s^2$ = sample variance.

(iv) Compare $\chi^2_{cal}$ and $\chi^2_{tab}$

If $\chi^2_{cal}$ is in acceptance region, accept $H_0$ or else reject $H_0$