Large Samples

Single proportion:-

$H_0 \Rightarrow P =$

$H_1 \Rightarrow \underbrace{P > \text{ (or) } P <}_{\text{In question they will mention if not think practically}} \quad \text{(or) } P \neq$

↓

1) Dice & coin problems
2) 2 diff cases
   i, wheat & rice
   ii, male & female

Los:- if not mentioned take it as 5%.

1-tail: $Z_{tab} = Z_\alpha$

2-tail: $Z_{tab} = Z_{\alpha/2}$

$$Z_{cal} = \frac{p - P}{\sqrt{\frac{PQ}{n}}}$$

$|Z_{tab}| > |Z_{cal}|$ Ho is accepted

Confidence interval:-

$$p \pm Z_{\alpha/2} \sqrt{\frac{PQ}{n}}$$

Note:- p and P must describe same thing

Normal;-

$$E_{max} = Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow Z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}$$

(or)

$$Z_{\alpha/2} \cdot \sqrt{\frac{PQ}{n}}$$

$\boxed{Z_{area} = \text{point}}$

Maximum error = Sample mean - population mean

Confidence interval

$= (\bar{x} - E_{max}, \bar{x} + E_{max})$

Confidence $= (1-\alpha)100$

$\bar{x} = p \qquad \sigma = \sqrt{PQ}$

Steps:-

Null hypothesis (Ho):-

Alternative hypothesis (H1):-

Level of significance (LOS):-

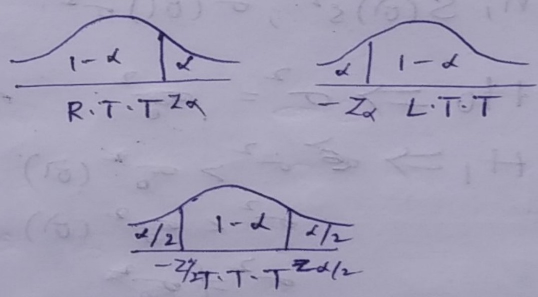
Test-Statistic:-

Conclusion:-

two proportion:-

$H_0 \Rightarrow P_1 = P_2$

$H_1 \Rightarrow \underbrace{P_1 > P_2 \text{ (or) } P_1 < P_2 \text{ (or) } P_1 \neq P_2}_{\text{clarity is provided in question}}$



$$Z_{cal} = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where

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

$Q = 1 - P$

Note:- Some times insted of x, n they will give percentages directly.

Confidence interval:-

$$P_1 \pm Z_{\alpha/2} \sqrt{\frac{P_1 Q_1}{n_1}}$$

Single mean:-

$H_0 \Rightarrow \mu =$

$H_1 \Rightarrow \mu >$ (or) $\mu <$ (or) $\mu \neq$

$$Z_{cal} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Confidence interval:-

$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

two mean:-

$H_0 \Rightarrow \mu_1 = \mu_2$

$H_1 \Rightarrow \mu_1 > \mu_2$ (or) $\mu_1 < \mu_2$ (or) $\mu_1 \neq \mu_2$

(clearly indicated in the question itself)

$$Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

(or)

$$Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Confidence interval:-

$$(\bar{x}_1 - \bar{x}_2) \pm Z_{\alpha/2}(\text{std. error})$$

where std. error $= \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

χ^2 distribution:- (for both small and large samples)

identification:- (1- variance)

inf only contains n, S (or) s^2, σ (or) σ^2.

$H_0 \Rightarrow \sigma^2 =$

$H_1 \Rightarrow \sigma^2 >$ (or)

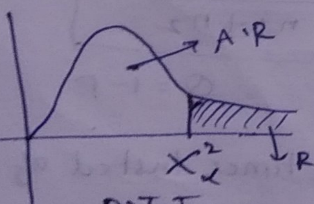
$\sigma^2 <$ (or)

$\sigma^2 \neq$

LOS $\Rightarrow$ 5% or 1%.

Degrees of freedom (Df) $= \nu = n-1$ → $\boxed{\nu, \text{Los} \rightarrow \text{Based table value}}$

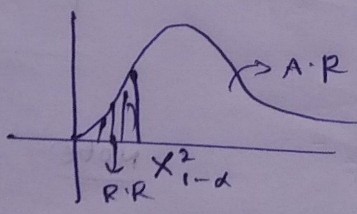
Test statistic: $\chi^2_{cal} = \frac{(n-1)s^2}{\sigma^2}$



R.T.T

If $\chi^2_{cal} < \chi^2_{tab}$

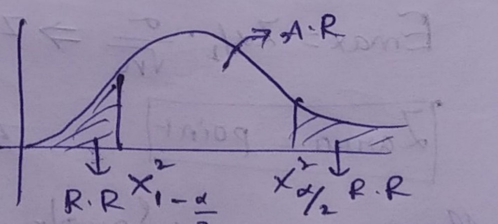
Ho is accepted



L.T.T

If $\chi^2_{cal} > \chi^2_{tab}$

Ho is accepted



$\chi^2_{1-\alpha/2} < \chi^2_{cal} < \chi^2_{\alpha/2}$

Ho is accepted.

Confidence interval:-

identification:-

sample { , , , , } calculate confidence interval for given confidence

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}} , \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}} \right)$$

Note:- In calculator we get (s) only we need to square it.

Special case:-

Here type1 and type2 are compared using number

$H_0 \Rightarrow \mu_1 - \mu_2 = d$ (or) $P_1 - P_2 = d$

$H_1 \Rightarrow \mu_1 - \mu_2 > d$ (or) $< d$ (or) $P_1 - P_2 > d$ (or) $< d$.

2 mean large sample

$$Z_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - d}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

2 mean small sample:-

$$t_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - d}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}}$$

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

2 proportion large sample

$$Z_{cal} = \frac{(P_1 - P_2) - d}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$

$(n \leq 30)$ Small Samples:- (ν), (t tables)

as these are small samples some times we need to calculate all ($\bar{x}$, s etc)

Single mean:-

$H_0 \Rightarrow \mu =$

$H_1 \Rightarrow \mu >$ (or) $\mu <$ (or) $\mu \neq$

Test statistic:- $t_{cal} = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

$|t_{cal}| < |t_{tab}|$ Ho is accepted.

$\boxed{t_{tab} \text{ based on } \alpha \text{ \& } \nu}$

confidence interval:-

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$$

2 - mean

Identification:-

Method A　　Sample A　　Mumbai　　etc ----

Method B　　Sample B　　Delhi

Here

$$\boxed{\nu = n_1 + n_2 - 2}$$

$H_0 \Rightarrow \mu_1 = \mu_2$

$H_1 \Rightarrow \mu_1 > \mu_2$ (or) $\mu_2 > \mu_1$ (or) $\mu_1 \neq \mu_2$

Test statistic :- $t_{cal} = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma^2}{n_1} + \dfrac{\sigma^2}{n_2}}}$　　where $\sigma^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$

Confidence interval:-

$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}$ (std. error)

where std. error $= \sqrt{\dfrac{\sigma^2}{n_1} + \dfrac{\sigma^2}{n_2}}$

paired t test:-

Identification:-

Before __________

After __________

x_i ______

y_i ______

$(x_i - y_i) d_i$ ______

$\bar{d} = \dfrac{1}{n} \sum_{i=1}^{n} d_i$　　$S_d^2 = \dfrac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2$

$H_0 \Rightarrow \bar{d} = 0$

$H_1 \Rightarrow \bar{d} > 0$ (or) $\bar{d} < 0$ (or) $\bar{d} \neq 0$

Test statistic

$t_{cal} = \dfrac{\bar{d}}{\left(\dfrac{s_d}{\sqrt{n}}\right)}$

F test

Identification:-

if only contain n_1, n_2, S_1, S_2
(2-variances)

$H_0 \Rightarrow \sigma_1^2 = \sigma_2^2$

$H_1 \Rightarrow \sigma_1^2 > \sigma_2^2$ (or) $\sigma_1^2 < \sigma_2^2$ (or) $\sigma_1^2 \neq \sigma_2^2$

Test statistic:-

$F_{cal} = \frac{S_1^2}{S_2^2}$ $\quad F_{cal} = \frac{S_2^2}{S_1^2}$ $\quad F_{cal} = \frac{S_2^2}{S_1^2}$ if $(S_2 > S_1)$

$\frac{S_1^2}{S_2^2}$ if $(S_1 > S_2)$

$F_{tab} = F_\alpha(n_1-1, n_2-1)$ $\quad F_{tab} = F_\alpha(n_2-1, n_1-1)$

l.o.s $\longrightarrow$ either 5% or 1%.

$|F_{tab}| > |F_{cal}| \longrightarrow H_0$ is accepted

Chapter-5

Regression line of x on y is

$x - \bar{x} = b_{xy}(y - \bar{y})$

b_{xy} ↓ Regression coefficient of x on y

y on x is

$y - \bar{y} = b_{yx}(x - \bar{x})$

b_{yx} ↓ Regression coefficient on y on x

$b_{xy} = r \frac{\sigma_x}{\sigma_y}$ $\quad b_{yx} = r \frac{\sigma_y}{\sigma_x}$

r → Karl pearsons coefficient (or) Correlation coefficient

where

$r = \frac{E(xy) - E(x)E(y)}{\sigma_x \sigma_y}$

$\boxed{-1 \le r \le 1}$

$r^2 = b_{xy} \cdot b_{yx}$ (if b_{xy} and b_{yx} are -ve r is -ve)

Tip:-

for regression line of x on y $\quad \frac{-b}{a}$ is less than 1 but +ve

y on x $\quad \frac{-b}{a}$ is greater than 1

St line

$$y = a + bx$$

$$\sum_{i=1}^{n} y_i = na + b \sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i y_i = a \sum x_i + b \sum_{i=1}^{n} x_i^2$$

parabola (or) 2nd degree polynomial

$$y = a + bx + cx^2$$

$$\sum_{i=1}^{n} y_i = na + b \sum_{i=1}^{n} x_i + c \sum_{i=1}^{n} x_i^2$$

$$\sum_{i=1}^{n} x_i y_i = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} x_i^2 + c \sum_{i=1}^{n} x_i^3$$

$$\sum_{i=1}^{n} x_i^2 y_i = a \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i^3 + c \sum_{i=1}^{n} x_i^4$$

Non linear curve fitting:-

$$y = ae^{bx}$$

$$\ln y = \ln(ae^{bx}) \Rightarrow \ln y = \ln a + bx$$

$$Y = A + BX$$

$$\sum_{i=1}^{n} Y_i = nA + B \sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} X_i Y_i = A \sum_{i=1}^{n} X_i + B \sum_{i=1}^{n} X_i^2$$

$$\boxed{a = e^{A}} \quad \boxed{b = B}$$