Unit 5:- Sampling and statistical inference

1) Define the following:-

i) **Population**:- The process of considering the entire group of studies and analise the data is called population.

eg:- a) The height of men and women in a city
b) Literacy in a tumkur city.

ii) **Sampling distribution**:-

Suppose that we have different sample of size 'n' drawn from population for each and every sample of size 'n', we can compute mean, standard deviation etc

Suppose we group these ~~characters~~ characteristics according to their frequencies, the frequency distribution generated is called sampling distribution.

iii) **Standard Error**:- The standard deviation of a sampling distribution is called standard Error.

iv) **Confidence intervals**:- (Confidence limits):-

Suppose that sampling distribution of a ~~Stasti~~ Statistics under normal variable with mean μ and standard deviation σ etc.

The sample statistics can be expected to lie in the interval $(\mu-1.96\sigma, \mu+1.96\sigma)$ for 95% times is the confident of finding μ in interval $(S-1.96\sigma, S+1.96\sigma)$, confidence interval for estimation of μ. The end of this interval $S\pm1.96\sigma$ are called 95% confidence limits. IIIly $S\pm2.58\sigma$ are called 99% confidence limits.

The numbers 1.96, 2.58 are called confidence coefficients.

v) **Type I, Type II errors**:- If the hypothesis (testing) is rejected, while it should have been accepted we conclude that type I error is occured, on the other hand if the hypothesis is accepted, while it should have been rejected, we conclude that type II error is occured.

vi) **Level of significance**:- The probability level which leads to rejection of hypothesis is called level of significance.

The level of significance is fixed at 0.05 or 5% and 0.01 or 1%.

vii) **One Tailed & Two Tailed testing**:-

Test	Critical value of z	
one tailed	-1.645 or 1.645	-2.33 or 2.33
	5% level, 95% confidence	1% level, 99% confidence
Two tailed	-1.96 & 1.96	-2.58 & 2.58

Note:-

1) If $|z| > 1.96$, the difference b/w $\bar{x}$ & μ are significant (accepted).
2) If $|z| < 1.96$, the difference b/w $\bar{x}$ & μ is not significant (rejected).
3) If $|z| > 2.58$, the difference b/w $\bar{x}$ & μ is are significant (accepted).

1) A survey was conducted in a slum area of 2000 families by selecting a sample of size 800. It was revealed that 180 families were illiterates. Find the probable limits of illiterates families in the population of 2000.

Solns:-

$$p = \frac{180}{800}$$

Confident limits $= p \pm (2.58)\sqrt{\frac{pq}{n}}$

Here $n = 800$, $p = \frac{180}{800} = 0.23$, $q = 1-p = 1-0.23$

$q = 0.77$

Thus the ~~probable~~ probable limits of illiterates families are $= p \pm (2.58)\sqrt{\frac{pq}{n}}$

$$= 0.23 \pm (2.58)\sqrt{\frac{(0.23)\times(0.77)}{800}}$$

$= 0.23 \pm 0.04$, $0.23 - 0.04$

$= 0.27, 0.19$

Thus for 2000 families illiteraties are

$0.27 \times 2000 = 540$

$0.19 \times 2000 = 380$

$\Rightarrow (380, 540)$

2) Ten individuals are chosen at random from a population & their heights in inches are 63, 63, 66, 67, 68, 69, 70, 70, 71, 71. Test the hypothesis for mean height of universe is 66 inches ($t_{0.05} = 2.262$ for 9 d.f.)

Soln:-

$$t = \frac{\bar{x}-\mu}{s}\sqrt{n}$$

$$s^2 = \frac{1}{n-1}\Sigma(x_i - \bar{x})^2$$

By t-distribution,

$$t = \frac{\bar{x}-\mu}{s}\sqrt{n} \rightarrow (1)$$

Here $n = 10$,

$$\bar{x} = \frac{\Sigma x}{n} = \frac{678}{10} = 67.8$$

$\mu = 66$

$$S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

$$= \frac{1}{10-1} \{(63-67.8)^2 + (63-67.8)^2 + (66-67.8)^2 + (67-67.8)^2 + (68-67.8)^2 + (69-67.8)^2 + (70-67.8)^2 + (70-67.8)^2 + (71-67.8)^2 + (71-67.8)^2\}$$

$$= \frac{1}{10-1} \{23.04 + 23.04 + 3.24 + 0.64 + 0.04 + 1.44 + 4.84 + 4.84 + 10.24 + 10.24\}$$

$$= \frac{1}{9} (81.60) = 9.07$$

$\therefore S = \sqrt{9.07}$

$$\boxed{S = 3.01}$$

$① \Rightarrow \quad t = \frac{67.8 - 66}{3.01} \sqrt{10}$

$t = 0.60 \times 3.16$

$\boxed{t = 1.90} \; // < 2.262$

$\therefore$ ~~thus~~ hypothesis is accepted.

3] A dice is thrown 264 times & numbers appearing on face(x) follows frequencies

x	1	2	3	4	5	6
f	40	32	28	58	54	60

calculate χ^2

Soln:-

$$\chi^2 = \frac{(O_i - E_i)^2}{E_i}$$

Here, $\chi^2 = \frac{(O_i - E_i)^2}{E_i}$, $O_i \to$ observed frequency, $E_i \to$ excepted frequency

$O_i = 40, 32, 28, 58, 54, 60$

$E_i = \frac{264}{6} = 44$

$$\therefore \chi^2 = \frac{(40-44)^2}{44} + \frac{(32-44)^2}{44} + \frac{(28-44)^2}{44} + \frac{(58-44)^2}{44} + \frac{(54-44)^2}{44} + \frac{(60-44)^2}{44}$$

$$= \frac{1}{44}(16 + 144 + 256 + 196 + 100 + 256)$$

$$\chi^2 = \frac{1}{44} \times 968 = 22 \; //$$

4] 4 coins are tossed 100 times, the following frequencies are obtained.
Fit a binomial distribution for the data and test goodness of fit ($\chi^2_{0.05} = 9.49$ for 4 d.f)

No. of heads	0	1	2	3	4
Frequency	5	29	36	25	5

Soln:-

The binomial distribution is given by

$P(x) = {}^nC_x \, p^x q^{n-x} \to ①$

Here $n = 4$, ~~$\mu = \frac{\sum fx}{\sum f}$~~ = ~~5 + 29 + 36 + 25~~

~~$\mu = np$~~

$$\mu = \frac{\Sigma fx}{\Sigma f} = \frac{0 + 29 + 72 + 75 + 20}{5 + 29 + 36 + 25 + 5} = 1.96$$

$\mu = np$

$1.96 = 4p$

$p = \frac{1.96}{4} = 0.49$ // $\quad q = 1 - p = 1 - 0.49 = 0.51$ //

$$\therefore \boxed{F(x) = 100 \times {}^{n}_{4}C_x (0.49)^x (0.51)^{4-x}} \rightarrow ②$$

put $x = 0$, $F(0) = 100 \times 0.07 = 7$

$x = 1$, $F(1) = 100 \times 0.2548 = 25.48$

$x = 2$, $F(2) = 100 \times 0.37 = 37$

$x = 3$, $F(3) = 100 \times 0.24 = 24$

$x = 4$, $F(4) = 100 \times 0.06 = 6$

$\therefore O_i = 5, 29, 36, 25, 5$

& $E_i = 7, 26, 37, 24, 6$

$$\therefore \chi^2 = \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(5-7)^2}{7} + \frac{(29-26)^2}{26} + \frac{(36-37)^2}{37} + \frac{(25-24)^2}{24} + \frac{(5-6)^2}{6}$$

$$= 0.57 + 0.35 + 0.03 + 0.04 + 0.17$$

$\chi^2 = 1.16$ // < 9.49

$\therefore$ Hypothesis is accepted.

5] A population consist of five members 2, 3, 6, 8, 11. Consider all possible sample of size 2 which can be drawn with replacement from the population.

Find the following

i] mean and SD of the population

ii] mean and SD of the sampling distribution

iii] mean and SD of the sampling distribution without replacement.

Soln:

By data :- 2, 3, 6, 8, 11

$$\mu = \frac{\Sigma x}{n} = \frac{2+3+6+8+11}{5} = 6 //$$

$$SD,\ \sigma^2 = \frac{\Sigma (x_i - \mu)^2}{n}$$

$$= \frac{(2-6)^2}{5} + \frac{(3-6)^2}{5} + \frac{(6-6)^2}{5} + \frac{(8-6)^2}{5} + \frac{(11-6)^2}{5}$$

$$= 10.80$$

~~$\sigma = \sqrt{10.80} = 3.29$~~ //

$\sigma = \sqrt{10.80} = 3.29$ //

b] Sample of size 2 with replacement.

$N = 5$, $n = 2$, $N^n = 5^2 = 25$

(2,2), (2,3), (2,6), (2,8), (2,11) → (2, 2.5, 4, 5, 6.5)

(3,2), (3,3), (3,6), (3,8), (3,11) → (2.5, 3, 4.5, 5.5, 7)

(6,2), (6,3), (6,6), (6,8), (6,11) → (4, 4.5, 6, 7, 8.5)

(8,2), (8,3), (8,6), (8,8), (8,11) → (5, 5.5, 7, 8, 9.5)

(11,2), (11,3), (11,6), (11,8), (11,11) → (6.5, 7, 8.5, 9.5, 11)

x	2	2.5	3	4	4.5	5	5.5	6	6.5	7	8	8.5	9.5	11
y	1	2	1	2	2	2	2	1	2	4	1	2	2	1

$\mu = \frac{\Sigma xf}{\Sigma f} = \frac{150}{25} = 6$

$\sigma^2 = \frac{\Sigma x^2 f}{\Sigma f} - \mu^2 = \frac{1035}{25} - 6^2 = 5.4$

$\sigma = \sqrt{5.4} = 2.32$

iii) (2, 3) (2, 6) (2, 8) (2, 11) → 2.5, 4.5, 6.5

(3, 6) (3, 8) (3, 11) → 4.5, 5.5, 7

(6, 8) (6, 11) → 7, 8.5

(8, 11) → 9.5

$\mu = \frac{\Sigma xf}{\Sigma f} = \frac{2.5+4+5+6.5+4.5+5.5+7+7+8.5+9.5}{10}$

$= \frac{60}{10} = 6$

$\sigma^2 = \frac{1}{n} \Sigma (x_i - \mu)^2$

$= \frac{1}{10} \{(40.5)^2$

$= \frac{1}{10} \times 40.5$

$\sigma = \sqrt{4.05} = 2.01$

~~problem~~ unique fixed

1] Find ~~transition~~ probability matrix.

$$\begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/4 & 1/4 \end{bmatrix}$$

Soln: $VP = V$

$x+y+z=1$

$\boxed{x = 1-y-z}$

$$[x\ y\ z]\begin{bmatrix} 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/4 & 1/4 \end{bmatrix} = [x\ y\ z]$$

$$\begin{bmatrix} \frac{y}{2}+\frac{z}{2} & x+\frac{z}{4} & \frac{y}{2}+\frac{z}{4} \end{bmatrix} = [x\ y\ z]$$

equating

$\frac{y}{2}+\frac{z}{2} = x$

x'ly by 2

$y+z = 2x$

$2x - y - z = 0$

$x + \frac{z}{4} = y$

x'ly by 4

$4x + z = 4y$

$4x - 4y + z = 0$

$\frac{y}{2}+\frac{z}{4} = z$

$2(1-y-z) - y - z = 0$

$-3y - 3z = -2$

$4(1-y-z) - 4y + z = 0$

$-8y - 3z = -4$

$-3y - 3z = -2$

$-8y - 3z = -4$

$y = 0.4$

$z = 0.3$

$x = 1 - y - z$

$x = 1 - 0.4 - 0.3$

$x = 0.3$ //

$V = [0.3 \quad 0.4 \quad 0.3]$