# PRACTICALS ON STATISTICS

BY

TANUJIT CHAKRABORTY

Indian Statistical Institute

Mail : tanujitisi@gmail.com

# PROBLEMS ON LINEAR ALGEBRA

1)(a) Determine k so that the set S is linearly independent in $E^3$

(i) $S = \{(1,2,1), (k,3,1), (2,k,0)\}$

(ii) $S = \{(k,1,1), (1,k,1), (1,1,k)\}$

(b) For what values of $\alpha$ coill the vectors $(1,5,7), (4,0,\alpha)$, $(1,0,0)$ form a basis for $E^3$? [C.U.2009]

ANS:- (a) (i) Construct a matrix with the given vectors :

$$A = \begin{pmatrix} 1 & 2 & 1 \\ k & 3 & 1 \\ 2 & k & 0 \end{pmatrix}$$

As this vectors are linearly independent, so, $\text{rank}(A) = 3$.

$\therefore |A| \neq 0$

$$\Rightarrow \begin{vmatrix} 1 & 2 & 1 \\ k & 3 & 1 \\ 2 & k & 0 \end{vmatrix} \neq 0$$

$\Rightarrow (k^2-6) - (k-4) \neq 0$

$\Rightarrow (k^2 - k - 2) \neq 0$

$\Rightarrow (k-2)(k+1) \neq 0 \Rightarrow k \neq 2, -1$.

(ii) Construct a matrix with the given vectors :

$$A = \begin{pmatrix} k & 1 & 1 \\ 1 & k & 1 \\ 1 & 1 & k \end{pmatrix}$$

As this vectors are Linearly independent, so, $\text{rank}(A) = 3$

$\therefore |A| \neq 0$

$\Rightarrow (k-1)^2 (k+2) \neq 0$

$\Rightarrow k \neq -2, 1$.

(b) As the vectors $(1,5,7), (4,0,\alpha), (1,0,0)$ are linearly independent, so we construct a matrix A with $\text{rank}(A) = 3$.

$$A = \begin{pmatrix} 1 & 5 & 7 \\ 4 & 0 & \alpha \\ 1 & 0 & 0 \end{pmatrix}$$

$\Rightarrow |A| \neq 0$

$\Rightarrow 5\alpha \neq 0$

$\Rightarrow \alpha \neq 0$.

2) Extend the set $\{(2,3,-1), (1,-2,-9)\}$ of vectors to a basis for $E^3$. Also find an orthonormal basis for $E^3$.

**Ans:-**

$(0,0,1) \in E^3$ and the set $\{(2,3,-1), (1,-2,-9), (0,0,1)\}$ of vectors form a basis of $E^3$, as the vectors $(0,0,1), (2,3,-1), (1,-2,-9)$ are LIN and they span in $E^3$.

$$\underset{\sim}{v}_1 = (0,0,1)$$

$$\therefore \underset{\sim}{u}_1 = \frac{\underset{\sim}{v}_1}{\|\underset{\sim}{v}_1\|} = \frac{(0,0,1)}{\sqrt{0^2+0^2+1^2}} = (0,0,1)$$

$$\therefore \underset{\sim}{V}_2 = (2,3,-1) - \{(2,3,-1)'(0,0,1)\}(0,0,1)$$
$$= (2,3,-1) + (0,0,1)$$
$$= (2,3,0)$$

$$\therefore \underset{\sim}{u}_2 = \frac{\underset{\sim}{V}_2}{|\underset{\sim}{V}_2|} = \frac{(2,3,0)}{\sqrt{4+9}} = \frac{1}{\sqrt{13}}(2,3,0)$$

$$\therefore \underset{\sim}{v}_3 = (1,-2,-9) - (\underset{\sim}{u}_1'\underset{\sim}{a}_3)\underset{\sim}{u}_1 - (\underset{\sim}{u}_2'\underset{\sim}{a}_3)\underset{\sim}{u}_2$$
$$= (1,-2,-9) + 9(0,0,1) + \frac{4}{\sqrt{13}}\left(\frac{1}{\sqrt{13}}(2,3,0)\right)$$
$$= (1,-2,-9) + (0,0,9) + \left(\frac{8}{13}, \frac{12}{13}, 0\right)$$

$$= \left(\frac{21}{13}, -\frac{14}{13}, 0\right)$$

$$= \frac{7}{13}(3,-2,0)$$

$$\therefore \underset{\sim}{u}_3 = \frac{\underset{\sim}{V}_3}{\|\underset{\sim}{V}_3\|} = \frac{7/13\,(3,-2,0)}{\sqrt{\left(\frac{21}{13}\right)^2 + \left(\frac{14}{13}\right)^2}} = \frac{7}{13}(3,-2,0) \times \frac{\sqrt{13}}{7}$$

$$= \frac{1}{\sqrt{13}}(3,-2,0)$$

We know, $\{\underset{\sim}{u}_1, \underset{\sim}{u}_2, \underset{\sim}{u}_3\}$ forms orthonormal basis for $E^3$.
$\{\underset{\sim}{v}_1, \underset{\sim}{v}_2, \underset{\sim}{v}_3\}$ forms orthogonal basis for $E^3$.

In general, $\underset{\sim}{V}_k = \underset{\sim}{a}_k - \sum_{i=1}^{k-1} (\underset{\sim}{u}_i'\underset{\sim}{a}_k)\,\underset{\sim}{u}_i$ & $\underset{\sim}{u}_k = \frac{\underset{\sim}{V}_k}{\|\underset{\sim}{V}_k\|}$.

$\therefore$ the orthonormal basis for $E^3$ is

$$\left\{(0,0,1), \frac{1}{\sqrt{3}}(2,3,0), \frac{1}{\sqrt{13}}(3,-2,0)\right\}$$

3) Find the dimension of the vector space generated by the vectors: $\underset{\sim}{\alpha}_1' = (0\ 1\ 2\ 3)$
$\underset{\sim}{\alpha}_2' = (2\ -1\ 5\ 4)$
$\underset{\sim}{\alpha}_3' = (4\ 0\ 6\ 1)$
$\underset{\sim}{\alpha}_4' = (0\ -2\ 4\ 7)$
Find a vector in the space orthogonal to the vector space spanned by $\underset{\sim}{\alpha}_1', \underset{\sim}{\alpha}_2', \underset{\sim}{\alpha}_3', \underset{\sim}{\alpha}_4'$. [C.U. 2001]

ANS:-
$$A = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 2 & -1 & 5 & 4 \\ 4 & 0 & 6 & 1 \\ 0 & -2 & 4 & 7 \end{pmatrix}$$

Now, to find Rank (A), we will reduce this matrix into its echelon form (E) and the solution of $E\underset{\sim}{x} = \underset{\sim}{0}$ is a vector which is orthogonal to the vector space spanned by $\underset{\sim}{\alpha}_1, \underset{\sim}{\alpha}_2, \underset{\sim}{\alpha}_3, \underset{\sim}{\alpha}_4$.

$$A = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 2 & -1 & 5 & 4 \\ 4 & 0 & 6 & 1 \\ 0 & -2 & 4 & 7 \end{pmatrix} \xrightarrow{R_2 \leftrightarrow R_1} \begin{pmatrix} 2 & -1 & 5 & 4 \\ 0 & 1 & 2 & 3 \\ 4 & 0 & 6 & 1 \\ 0 & -2 & 4 & 7 \end{pmatrix}$$

$$\xrightarrow{R_3' = R_3 - 2R_1} \begin{pmatrix} 2 & -1 & 5 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 2 & -4 & -7 \\ 0 & -2 & 4 & 7 \end{pmatrix} \xrightarrow{R_4' = R_3 + R_4} \begin{pmatrix} 2 & -1 & 5 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 2 & -4 & -7 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\xrightarrow[R_3' = R_3 - 2R_2]{R_1' = R_1 + R_2} \begin{pmatrix} 2 & 0 & 7 & 7 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & -8 & -13 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow[R_3' = \frac{R_3}{-8}]{R_1' = R_1/2} \begin{pmatrix} 1 & 0 & 7/2 & 7/2 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & \frac{13}{8} \\ 0 & 0 & 0 & 0 \end{pmatrix} = E$$

$\therefore$ Rank (A) = 3.
$\therefore$ dim (S4) = 3.

Now, $E\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \underset{\sim}{0}$ gives —

$$\Rightarrow \quad \underset{\sim}{x}_1 + \frac{7}{2}\underset{\sim}{x}_3 + \frac{7}{2}\underset{\sim}{x}_4 = \underset{\sim}{0}$$
$$\underset{\sim}{x}_2 + 2\underset{\sim}{x}_3 + 3\underset{\sim}{x}_4 = \underset{\sim}{0}$$
$$\underset{\sim}{x}_3 + \frac{13}{8}\underset{\sim}{x}_4 = \underset{\sim}{0}$$

Let, $x_4 = t$, then $x_3 = -\frac{13}{8}t$, $x_2 = \frac{1}{4}t$, $x_1 = \frac{35}{18}t$.

$\therefore \underset{\sim}{x} = t\left(\frac{35}{18}, \frac{1}{4}, -\frac{13}{8}, 1\right)$, $t \in \mathbb{R}$.

4) The following vectors from a spanning set for a vector space $S_4$:

$\underset{\sim}{\alpha_1}' = (1.000 \quad 0.313 \quad 0.280 \quad 0.156)$
$\underset{\sim}{\alpha_2}' = (0.333 \quad 1.313 \quad 0.628 \quad 0.315)$
$\underset{\sim}{\alpha_3}' = (0.309 \quad 0.553 \quad 0.480 \quad 0.165)$
$\underset{\sim}{\alpha_4}' = (1.024 \quad 1.073 \quad 0.428 \quad 0.306)$

i) Find the $\dim(S_4)$?
ii) Find a basis for $S_4$?
iii) If $\dim(S_4) < 4$, find a basis for $O(S_4)$? [C.U.1996]

ANS:- Dimension of the vector space (S) spaned by $\alpha_i$'s, $i = 1(1)4$,
= the rank of $A = \begin{pmatrix} \underset{\sim}{\alpha_1}' \\ \underset{\sim}{\alpha_2}' \\ \underset{\sim}{\alpha_3}' \\ \underset{\sim}{\alpha_4}' \end{pmatrix}$

Now, $A = \begin{pmatrix} 1.000 & 0.313 & 0.280 & 0.156 \\ 0.333 & 1.313 & 0.628 & 0.315 \\ 0.309 & 0.553 & 0.480 & 0.165 \\ 1.024 & 1.073 & 0.428 & 0.306 \end{pmatrix}$

$R_4' = R_3 + R_4 - R_2 - R_1$ $\begin{pmatrix} 1.000 & 0.313 & 0.280 & 0.156 \\ 0.333 & 1.313 & 0.628 & 0.315 \\ 0.309 & 0.553 & 0.480 & 0.165 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

$R_2' \to R_2 - 0.333 R_1$, $R_3' = R_3 - 0.309 R_1$ $\begin{pmatrix} 1.000 & 0.313 & 0.280 & 0.156 \\ 0 & 1.2088 & 0.5348 & 0.2630 \\ 0 & 0.4563 & 0.3985 & 0.1168 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

$R_2' = R_2 / 1.2088$, $R_3' = R_3 - .4563 R_2$ $\begin{pmatrix} 1.000 & 0.313 & 0.280 & 0.156 \\ 0 & 1.000 & 0.4424 & 0.2176 \\ 0 & 0 & 0.1918 & 0.0175 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

$R_3' = R_3 / .1916$ $\begin{pmatrix} 1.000 & 0.313 & 0.280 & 0.156 \\ 0 & 1.000 & 0.4424 & 0.2176 \\ 0 & 0 & 1.000 & 0.0913 \\ 0 & 0 & 0 & 0 \end{pmatrix} = E$

$\therefore$ Rank (A) = No. of non-null rows in it's echelon form
= 3
$\therefore \dim(S) = 3$

ii) The basis of S is $\{(1.000\ \ 0.313\ \ 0.280\ \ 0.156), (0\ \ 1.000\ \ 0.4424\ \ 0.2176), (0\ \ 0\ \ 1.000\ \ 0.0913)\}$

iii) $E\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \underset{\sim}{0}$

$$\Rightarrow \begin{aligned} 1.000x_1 + 0.313x_2 + 0.280x_3 + 0.156x_4 &= 0 \\ 1.000x_2 + 0.4424x_3 + 0.2176x_4 &= 0 \\ 1.000x_3 + 0.0918x_4 &= 0 \end{aligned}$$

Let, $x_4 = t, t \in \mathbb{R}$.

$$\begin{cases} \therefore x_3 = -0.0913\ t \\ \therefore x_2 = -0.1772\ t \\ \therefore x_1 = -0.0749\ t \end{cases}$$

$\therefore O(S) = t(-0.0749, -0.1772, -0.0913, 1), t \in \mathbb{R}$.

5) S and T are subspaces of $V_4$ given by

$S = \{(x_1, x_2, x_3, x_4) : 2x_1 + x_2 + 3x_3 + x_4 = 0\}$

$T = \{(x_1, x_4, x_3, x_4) : x_1 + 2x_2 + x_3 + 3x_4 = 0\}$

Find a basis and the dimension of (i) $S \cap T$, (ii) $S + T$.

ANS:- Here, $S \cap T = \{(x_1, x_2, x_3, x_4) : \begin{matrix} 2x_1 + x_2 + 3x_3 + x_4 = 0 \\ x_1 + 2x_2 + x_3 + 3x_4 = 0 \end{matrix}\}$

Let, $\underset{\sim}{x} \in S \cap T$, then

$2x_1 + x_2 + 3x_3 + x_4 = 0 \Rightarrow 2x_1 + x_2 = -3x_3 - x_4$ —①

$x_1 + 2x_2 + x_3 + 3x_4 = 0 \Rightarrow x_1 + 2x_2 = -x_3 - 3x_4$ —②

Solving ① & ②, we get, $x_1 = \dfrac{-5x_3 + x_4}{3}$

$$x_2 = \frac{x_3 - 5x_4}{3}$$

Here, $\underset{\sim}{x} = \left(\dfrac{-5x_3 + x_4}{3}, \dfrac{x_3 - 5x_4}{3}, x_3, x_4\right)$

$$= x_3\left(-\frac{5}{3}, \frac{1}{3}, 1, 0\right) + x_4\left(\frac{1}{3}, -\frac{5}{3}, 0, 1\right)$$

Here, $\{(-\frac{5}{3}, \frac{1}{3}, 1, 0), (\frac{1}{3}, -\frac{5}{3}, 0, 1)\}$ forms a basis.

$\therefore \dim(S \cap T) = 2$.

Now, $x \in S$

$\therefore \underset{\sim}{x} = (x_1, x_2, x_3, x_4)$ , $2x_1 + x_2 + 3x_3 + x_4 = 0$

$= (x_1, x_2, x_3, -2x_1 - x_2 - 3x_3)$

$= x_1(1, 0, 0, -2) + x_2(0, 1, 0, -1) + x_3(0, 0, 1, -3)$

Clearly, $\{(1,0,0,-2), (0,1,0,-1), (0,0,1,-3)\}$ forms a basis for S.

$\therefore \dim(S) = 3$.

Now, $\underset{\sim}{x} \in T$

$\therefore \underset{\sim}{x} = (x_1, x_2, x_3, x_4)$ ; $x_1 + 2x_2 + x_3 + 3x_4 = 0$

$= (-2x_2 - x_3 - 3x_4, x_2, x_3, x_4)$

$= x_2(-2, 1, 0, 0) + x_3(-1, 0, 1, 0) + x_4(-3, 0, 0, 1)$

Clearly, $\{(-2,1,0,0), (-1,0,1,0), (-3,0,0,1)\}$ forms a basis for T.

$\therefore \dim(T) = 3$.

$\therefore \dim(S+T) = \dim(S) + \dim(T) - \dim(S \cap T)$

$= 3 + 3 - 2 = 4$

Clearly, $S+T \subseteq V_4$

And $\dim(S+T) = \dim(V_4)$

$\Rightarrow S+T = V_4$.

$\Rightarrow \{\underset{\sim}{e}_1, \underset{\sim}{e}_2, \underset{\sim}{e}_3, \underset{\sim}{e}_4\}$ is a basis of $(S+T)$.

6) If $U = \text{span}\{(1,2,1), (2,1,3)\}$, $W = \text{span}\{(1,0,0), (0,1,0)\}$ Show that U and W are subspaces of $R^3$. Determine a basis and dimension of $U \cap W$ and $U+W$.

ANS:- Let $\underset{\sim}{x} \in U \cap W$

$$\therefore \underset{\sim}{x} = l_1(1,2,1) + l_2(2,1,3)$$
$$= l_1(1,0,0) + l_2(0,1,0)$$

$$\therefore l_1 + 2l_2 = l_1$$
$$2l_1 + l_2 = l_2$$
$$l_1 + 3l_2 = 0$$

$$\therefore l_1 = -3l_2$$

$$\therefore \underset{\sim}{x} = l_2\{(-3)(1,2,1) + (2,1,3)\} = l_2(-1,-5,0)$$
$$= t(1,5,0), \quad t \in \mathbb{R}$$

$$\therefore U \cap W = \{t(1,5,0) : t \in \mathbb{R}\}$$

$\therefore$ Basis of $U \cap W = \{(1,5,0)\}$

$\therefore \dim(U \cap W) = 1$.

$(1,2,1), (2,1,3) \in U$ and they are linearly independent, so therefore they form a basis of U.

$\therefore \dim(U) = 2$

Similarly, $\dim(W) = 2$

$$\therefore \dim(U+W) = \dim(U) + \dim(W) - \dim(U+W)$$
$$= 2+2-1 = 3.$$

Clearly, $U+W \subseteq R^3$

$\therefore \dim(U+W) = \dim(R^3)$

$\therefore U+W = R^3$

$\therefore \{\underset{\sim}{e}_1, \underset{\sim}{e}_2, \underset{\sim}{e}_3\}$ is a basis of $(U+W)$.

7) Given $S_1 = \{(1,2,3), (0,1,2), (3,2,1)\}$ and $S_2 = \{(1,-2,3), (-1,1,-1), (1,-3,4)\}$, Determine the dimension and a basis for

i) $[S_1] \cap [S_2]$, ii) $[S_1] + [S_2]$,

where $[S]$ denotes the span of S. [C.U. 2004]

i)

$$S_1 \cap S_2 = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 2 \\ 3 & 2 & 1 \\ -1 & -2 & 3 \\ -1 & -1 & -2 \\ 1 & -3 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 2 \\ 0 & 5 & -5 \\ 0 & -1 & 1 \\ 0 & -2 & 2 \\ 0 & -5 & 1 \end{bmatrix}$$

$R_3' = R_3 + 3R_1$
$R_4' = R_4 + R_2$
$R_5' = R_5 + R_6$
$R_6' = R_6 - R_1$

$$\sim \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 2 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \\ 0 & -5 & 1 \end{bmatrix}$$

$R_3' = R_3/5$
$R_5' = R_4 - 2R_3$

$$\sim \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 2 \\ 0 & 0 & -3 \\ 0 & 0 & 0 \\ 0 & -5 & 1 \end{bmatrix}$$

$R_4' = R_4 + R_3$
$R_3' = R_3 - R_2$

$$\sim \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & -5 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$R_3' = R_3/-3$
$R_4 \leftrightarrow R_6$

$$\sim \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$R_2' = R_2 - 2R_3$
$R_4' = R_4 + 5R_2' - R_3$

Here, $\dim(S_1 \cap S_2) = 3$

Now, $\dim(S) = \dim(S_1) + \dim(S_2) - \dim(S_1 \cap S_2)$

$$S_1 = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 3 & 2 & 1 \end{pmatrix} \xrightarrow{R_3 \to R_3 - 3R_1} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & -4 & -8 \end{pmatrix}$$

$$\xrightarrow{R_3' = R_3 + 4R_2} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

$\therefore \dim(S_1) = 2$

$$S_2 = \begin{pmatrix} 1 & -2 & 3 \\ -1 & 1 & -2 \\ 1 & -3 & 4 \end{pmatrix} \xrightarrow[R_3' \to R_3 - R_1]{R_2' \to R_1 + R_2} \begin{pmatrix} 1 & -2 & 3 \\ 0 & -1 & 1 \\ 0 & -1 & 1 \end{pmatrix}$$

$$\xrightarrow[R_2' \to R_2/-1]{R_3' \to R_3 - R_2} \begin{pmatrix} 1 & -2 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

$\dim(S_2) = 2$

$\therefore \dim(S_1 + S_2) = \dim(S_1) + \dim(S_2) - \dim(S_1 \cap S_2)$
$= 2 + 2 - 3$
$= 1.$

8) Construct an orthogonal matrix $A^{4\times4}$ whose first row is $(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2})$. find $A^{-1}$. The matrix B is obtained by replacing the third column of A by (2 × 3rd column of A). Using $A^{-1}$, find $B^{-1}$. [C.U. 2002]

**ANS:-** Note that $\{e_1, e_2, e_3, e_4\}$ form a basis for $E^4$.

Now, $(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2}) = \frac{1}{2}e_1 + \frac{1}{2}e_2 + \frac{1}{2}e_3 + \frac{1}{2}e_4$

$\therefore \{(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2}), e_2, e_3, e_4\}$ forms a basis for $E^4$.

Taking, $v_1 = (\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2})$

$\therefore u_1 = \frac{v_1}{\|v_1\|} = (\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2})$

$v_2 = (0,1,0,0) - \{(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2})(0,1,0,0)\}\frac{1}{2}(1,1,1,1)$

$= (0,1,0,0) - \frac{1}{4}(1,1,1,1)$

$= (-\frac{1}{4},\frac{3}{4},-\frac{1}{4},-\frac{1}{4})$

$\therefore u_2 = \frac{v_2}{\|v_2\|} = \frac{(-\frac{1}{4},\frac{3}{4},-\frac{1}{4},-\frac{1}{4})}{\sqrt{\frac{1}{16}+\frac{9}{16}+\frac{1}{16}+\frac{1}{16}}} = -\frac{1}{\sqrt{12}}(1,-3,1,1)$

$v_3 = e_3 - \{e_3 \cdot u_1\}u_1 - \{e_3 \cdot u_2\}u_2$

$= (0,0,1,0) - \{(0,0,1,0)(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2})\}(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2})$

$- \{(0,0,1,0)(-\frac{1}{\sqrt{12}},\frac{3}{\sqrt{12}},-\frac{1}{\sqrt{12}},-\frac{1}{\sqrt{12}})\}(-\frac{1}{\sqrt{12}},\frac{3}{\sqrt{12}},-\frac{1}{\sqrt{12}},-\frac{1}{\sqrt{12}})$

$= (-\frac{1}{3},0,\frac{2}{3},-\frac{1}{3})$

$\therefore u_3 = \frac{v_3}{\|v_3\|} = \frac{(-\frac{1}{3},0,\frac{2}{3},-\frac{1}{3})}{\sqrt{\frac{1}{9}+0+\frac{4}{9}+\frac{1}{9}}} = \frac{1}{\sqrt{6}}(-1,0,2,-1).$

$v_4 = (0,0,0,1) - \{(0,0,0,1)(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2})\}(\frac{1}{2},\frac{1}{2},\frac{1}{2},\frac{1}{2})$

$- \{(0,0,0,1)(-\frac{1}{\sqrt{12}})(1,-3,1,1)\}(-\frac{1}{\sqrt{12}})(1,-3,1,1)$

$- \{(0,0,0,1)(\frac{1}{\sqrt{6}})(-1,0,2,-1)\}(\frac{1}{\sqrt{6}})(-1,0,2,-1)$

$= (-\frac{1}{2},0,0,\frac{1}{2})$

$\therefore u_4 = \frac{v_4}{\|v_4\|} = \frac{(-\frac{1}{2},0,0,\frac{1}{2})}{\sqrt{\frac{1}{4}+0+0+\frac{1}{4}}} = \frac{1}{\sqrt{2}}(-1,0,0,1)$

Hence, $\{u_1, u_2, u_3, u_4\}$ is a set of 4 orthonormal vectors.

Hence, $A = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{pmatrix}$ is an orthonormal matrix with first

now, $u_1 = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$.

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{\sqrt{12}} & \frac{3}{\sqrt{12}} & \frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}$$

As A is an orthogonal matrix, so $A' = A^{-1}$.

$$\therefore A^{-1} = A' = \begin{bmatrix} \frac{1}{2} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{3}{\sqrt{12}} & 0 & 0 \\ \frac{1}{2} & -\frac{1}{\sqrt{12}} & \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{2} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Here, $B = AE_3(2)$, where $E_3(2) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

Now, $B^{-1} = [A E_3(2)]^{-1}$

$= E_3^{-1}(2) A^{-1}$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{3}{\sqrt{12}} & 0 & 0 \\ \frac{1}{2} & -\frac{1}{\sqrt{12}} & \frac{2}{\sqrt{6}} & 0 \\ \frac{1}{2} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{2} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{3}{\sqrt{12}} & 0 & 0 \\ \frac{1}{2} & -\frac{1}{2\sqrt{12}} & \frac{1}{\sqrt{6}} & 0 \\ \frac{1}{2} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{6}} & \frac{1}{2} \end{pmatrix}$$

Q) Let A and B be defined by

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ -1 & 1 & -3 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 2 & 1 \\ 0 & 4 & 2 \\ 0 & -2 & -1 \end{pmatrix}$$

(i) Show that the column space of B is a subspace of A.
(ii) Find a matrix C such that $AC = B$. [C.U. 2000]

**ANS:-** (i) Column space of $A = C(A) = \left\{ l_1 \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} + l_2 \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + l_3 \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}; \; l_1, l_2, l_3 \in R \right\}$

Column space of $B = C(B) = \left\{ l \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}; \; l \in R \right\}$

Now, for any vector $x \in C(B)$

we have, $x = l \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$

$= l_1 \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} + l_2 \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + l_3 \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$, where $l = l_1$ & $l_2 = l_3 = 0$.

which belongs to $C(A)$.

$\therefore x \in C(A) \Rightarrow x \in C(B)$.

$\therefore C(B) \subseteq C(A)$

(ii) Here, note that, $\rho(A) = 2$, i.e. $\rho(A) < 3$,

Since 1st & 2nd rows are LD.

$\Rightarrow$ A is singular.

$\Rightarrow A^{-1}$ does not exist.

[ If $A^{-1}$ exists, then $AC = B \Rightarrow C = A^{-1}B$ and it is unique ]

Hence, choice of C is not unique.

Note that, $\underset{(A)}{\begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ -1 & 1 & 1 \end{pmatrix}} \underset{(C)}{\begin{pmatrix} 0 & 2 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}} = \underset{(B)}{\begin{pmatrix} 0 & 2 & 1 \\ 0 & 4 & 2 \\ 0 & -2 & -1 \end{pmatrix}}$

Alternative choice of $C = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$.

10) Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 5 & 7 \\ 9 & 16 & 23 \end{pmatrix}$. Suppose $\rho(A) = r$.

i) Find $r$. ii) Write A as the sum of $r$ matrices each of rank unity. iii) Find an orthonormal basis of the row-space of A. [C.U. 2008]

ANS:-

Note that, $(9, 16, 23) = 3(1,2,3) + 2(3,5,7)$

i) $A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 5 & 7 \\ 3\times1+2\times1 & 3\times2+2\times5 & 3\times3+2\times7 \end{pmatrix}$

clearly, Rank (A) = 2, as there have two LIN rows.

ii) $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 3 & 6 & 9 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 3 & 5 & 7 \\ 6 & 10 & 14 \end{pmatrix}$

$= \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 3\times1 & 3\times2 & 3\times3 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 3 & 5 & 7 \\ 3\times2 & 5\times2 & 7\times2 \end{pmatrix}$

$= A_1 + A_2$, where Rank $(A_i) = 1$, $\forall i = 1, 2$.

iii) $\{(1,2,3), (3,5,7)\}$ form a basis of the row space of A.

$\therefore \underset{\sim}{v}_1 = (1,2,3)$

$\therefore \underset{\sim}{u}_1 = \frac{\underset{\sim}{v}_1}{\|\underset{\sim}{v}_1\|} = \frac{(1,2,3)}{\sqrt{1+4+9}} = \frac{1}{\sqrt{14}}(1,2,3)$

$\therefore \underset{\sim}{v}_2 = (3,5,7) - \{(3,5,7)\frac{1}{\sqrt{14}}(1,2,3)\}\frac{1}{\sqrt{14}}(1,2,3)$

$= \left(\frac{4}{7}, \frac{1}{7}, -\frac{2}{7}\right)$

$\therefore \underset{\sim}{u}_2 = \frac{\underset{\sim}{v}_2}{\|\underset{\sim}{v}_2\|} = \frac{7}{\sqrt{21}}\left(\frac{4}{7}, \frac{1}{7}, -\frac{2}{7}\right)$

$= \frac{1}{\sqrt{21}}(4, 1, -2)$

11) Suppose $A = \begin{pmatrix} 7 & -12 & 4 \\ 3 & -2 & 5 \\ -6 & 11 & 8 \end{pmatrix}$.

(a) Find $A^{-1}$.

(b) Find a matrix B such that $AB = \begin{pmatrix} 8 & -12 & 4 \\ 3 & -1 & 5 \\ -6 & 11 & 9 \end{pmatrix}$.

(c) Write A as a sum of symmetric and a skew-symmetric matrices.

(d) Let V be a vector space generated by the first two column vectors of A. Then write the third column vector of A as the sum of the two non-zero vectors such that one is a member of V and the other is orthogonal to V. [C.U. 1998]

ANS:-

(d) $V = \left\{ l_1 \begin{pmatrix} 7 \\ 3 \\ -6 \end{pmatrix} + l_2 \begin{pmatrix} -12 \\ -2 \\ 11 \end{pmatrix} : l_1, l_2 \in R \right\}$

Let, $\begin{pmatrix} 4 \\ 5 \\ 8 \end{pmatrix} = \underset{\sim}{x} + \underset{\sim}{y}$ where $\underset{\sim}{x} \in V$ and $\underset{\sim}{y}$ is $\perp$ to V.

Now, $\underset{\sim}{x} \in V$, $\underset{\sim}{x} = l_1 \begin{pmatrix} 7 \\ 3 \\ -6 \end{pmatrix} + l_2 \begin{pmatrix} -12 \\ -2 \\ 11 \end{pmatrix}$

Here, $\underset{\sim}{y}$ is $\perp$ to $V \Rightarrow 7y_1 + 3y_2 - 6y_3 = 0$

$-12y_1 - 2y_2 + 11y_3 = 0$

Let, $y_3 = t$, now, $y_1 = \frac{21t}{22}$, $y_2 = -\frac{5t}{22}$ ; $y = t \begin{pmatrix} 21/22 \\ -5/22 \\ 1 \end{pmatrix}$

$\therefore \underset{\sim}{x} = l_1 \begin{pmatrix} 7 \\ 3 \\ -6 \end{pmatrix} + l_2 \begin{pmatrix} -12 \\ -2 \\ 11 \end{pmatrix}$ ; $l_1, l_2 \in R$ $= 22t \begin{pmatrix} 21 \\ -5 \\ 22 \end{pmatrix}$

$= l_3 \begin{pmatrix} 21 \\ -5 \\ 22 \end{pmatrix}$

$\therefore \begin{pmatrix} 4 \\ 5 \\ 8 \end{pmatrix} = l_1 \begin{pmatrix} 7 \\ 3 \\ -6 \end{pmatrix} + l_2 \begin{pmatrix} -12 \\ -2 \\ 11 \end{pmatrix} + l_3 \begin{pmatrix} 21 \\ -5 \\ 22 \end{pmatrix}$ ;

$l_1, l_2, l_3 \in R$.

12) (a) Suppose $A = \begin{pmatrix} 4 & 3 & 6 \\ 5 & 9 & 8 \\ 2 & 7 & 1 \end{pmatrix}$. Find two non-singular matrices P and Q such that $PAQ = I_3$. [C.U. 1999]

(b) Find inverse of $A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 3 & -1 \\ -1 & 2 & 0 \end{pmatrix}$, by Pivotal Condensation or sweeping-out method.

(c) Obtain the fully reduced normal form of $A = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 3 \\ 1 & 1 & 2 \end{pmatrix}$

(d) Express $A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 2 \end{pmatrix}$ as the product of elementary matrices.

ANS:-

(i) By problem, $PAQ = I_3$

$\therefore$ Rank $(PAQ)$ = Rank $(I_3)$

$\therefore$ rank $(A) = 3$, as P and Q are non-singular matrices and rank is not altered by non-singular matrix multiplication.

$\Rightarrow$ A is non-singular.

$\Rightarrow$ $A^{-1}$ exists.

$\therefore$ $A^{-1}AI_3 = I_3$

$\Rightarrow PAQ = I_3$, where $P = A^{-1}$, $Q = I_3$.

(ii) $(A \mid I_3) = \left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 1 & 0 & 0 \\ 2 & 3 & -1 & 0 & 1 & 0 \\ -1 & 2 & 0 & 0 & 0 & 1 \end{array}\right)$

$R \sim$ [Row operations]

$\sim$

$\sim$

$\sim$

$\sim$

$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 7 & -1 & 5 \end{array}\right)$

$\therefore A^{-1} = \begin{pmatrix} 2 & 0 & 1 \\ 1 & 0 & 1 \\ 7 & -1 & 5 \end{pmatrix}$

(c) $[I_3 \mid A^{3\times3} \mid I_3]$

$$= \left[\begin{array}{ccc|ccc|c} 1 & 0 & 0 & 1 & 0 & 1 & \\ 0 & 1 & 0 & 2 & 1 & 3 & I_3 \\ 0 & 0 & 1 & 1 & 1 & 2 & \end{array}\right]$$

$R_2' \to R_2 - 2R_1$, $R_3' \to R_3 - R_1$

$$\left[\begin{array}{ccc|ccc|c} 1 & 0 & 0 & 1 & 0 & 1 & \\ -2 & 1 & 0 & 0 & 1 & 1 & I_3 \\ -1 & 0 & 1 & 0 & 1 & 1 & \end{array}\right]$$

$R_3' \to R_3 - R_2$

$$\left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ -2 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

$C_3' \to C_3 - C_1 - C_2$

$$\left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & -1 \\ -2 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & -1 \\ 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$= \left[\begin{array}{c|cc|c} P & I_2 & 0 & Q \\ & 0 & 0 & \end{array}\right]$$

Hence, $P = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}$, $Q = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$

(d) $[I_3 \mid A^{3\times3} \mid I_3]$

$$= \left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 2 & 0 & 0 & 1 \end{array}\right]$$

$$\sim \sim \sim \left[\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & -1 \\ -1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & -1 & 1 & 0 & 0 & 1 & -1 & 0 & 1 \end{array}\right]; \quad P = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix}, \quad Q = \begin{pmatrix} 2 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

$\therefore A = P^{-1} I_3 Q^{-1} = P^{-1} Q^{-1}$.

$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} P$$

$$\Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} P \quad [R_3' \to R_3 + R_2]$$

$$\Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} P \quad [R_2' \to R_1 + R_2]$$

$$\therefore P^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} Q$$

$$\Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} Q$$

$$\Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix} Q$$

$$\Rightarrow Q^{-1} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 2 \end{pmatrix}$$

12) (a) Solve the following system of linear equations using a numerical method:

$$x_1 - 2x_2 + 3x_3 + 4x_4 = 4.5$$
$$3x_1 - x_2 + 2x_3 + 5x_4 = 9.5$$
$$2x_1 + 4x_2 + 5x_3 + x_4 = 15$$
$$4x_1 + 2x_2 - x_3 + 3x_4 = 12$$

[C.U. 2007]

ANS:-

$$\left(\begin{array}{cccc|c} 1 & -2 & 3 & 4 & 4.5 \\ 3 & -1 & 2 & 5 & 9.5 \\ 2 & 4 & +5 & 1 & 15 \\ 4 & 2 & -1 & 3 & 12 \end{array}\right)$$

$R_2' = R_2 - 3R_1$

$R_3' = R_3 - 2R_1$

$R_4' = R_4 - 4R_1$

$$\left(\begin{array}{cccc|c} 1 & -2 & 3 & 4 & 4.5 \\ 0 & 5 & -7 & -7 & \\ 0 & & & & \\ 0 & & & & \end{array}\right)$$

$$\sim \left(\begin{array}{cccc|c} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \end{array}\right)$$

$$\therefore \begin{pmatrix} & & & \\ & & & \\ & & & \\ & & & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} & \\ & \\ & \\ & \end{pmatrix}$$

$\therefore x_1$

$x_2$

$x_3$

$x_4$

(b) For what value of $k$ the planes

$x - 4y + 5z = k$

$x - y + 2z = 3$ (i) intersect in a line ?,

$2x + y + z = 0$, (ii) intersect in a point ?

ANS:-

Augmented matrix $= [A : b]$

$$= \begin{bmatrix} 1 & -4 & 5 & k \\ 1 & -1 & 2 & 3 \\ 2 & 1 & 1 & 0 \end{bmatrix}$$

$R_2' \to R_2 - R_1$, $R_3' \to R_3 - 2R_1$

$$\begin{bmatrix} 1 & -4 & 5 & k \\ 0 & 3 & -3 & 3-k \\ 0 & 9 & -9 & -2k \end{bmatrix}$$

$R_3' \to R_3 - 3R_2$, $R_2' \to R_2/3$

$$\begin{bmatrix} 1 & -4 & 5 & k \\ 0 & 1 & -1 & 3-k/3 \\ 0 & 0 & 0 & k-9 \end{bmatrix}$$

(i) If $k = 9$, then $\text{Rank}(A:b) = 2 = \text{Rank}(A)$

$\Rightarrow$ the system is consistent and has infinitely many solutions and intersect in a line.

(ii) $R(A:b) = R(A) = 3$ is not satisfied by no value of $k$.

13) (i) Identify the definiteness of the following quadratic forms :

$x_1^2 + 2x_2^2 + 3x_3^2 + 2x_1x_2 + 4x_2x_3 + 2x_1x_3$.

(ii) $2x_1^2 + 2x_2^2 + 5x_3^2 - 4x_1x_2 - 2x_1x_3 + 2x_2x_3$.

ANS:- (i) $A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}$ ; $a_{11} = 1 > 0$, $\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} = 1 > 0$,

$\begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{vmatrix} = 1 > 0$

$\therefore$ A is a p.d. matrix.

(ii) $A = \begin{pmatrix} 2 & -2 & -1 \\ -2 & 2 & 1 \\ -1 & 1 & 5 \end{pmatrix}$

$a_{11} = 2 > 0$, $\begin{vmatrix} 2 & -2 \\ -2 & 2 \end{vmatrix} = 0$, $\begin{vmatrix} 2 & -2 & -1 \\ -2 & 2 & 1 \\ -1 & 1 & 5 \end{vmatrix} = 0$

$\therefore$ A is a p.s.d. matrix.

14) (i) S.T. $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \\ 3 & 7 & 11 \end{pmatrix}$ is p.d. and find the matrix B such that $A = BB^T$.

**ANS:-** $a_{11} = 1 > 0$, $\begin{vmatrix} 1 & 2 \\ 2 & 5 \end{vmatrix} = 1 > 0$, $\begin{vmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \\ 3 & 7 & 11 \end{vmatrix} = 1 > 0$

∴ All the principal minors of A is positive.

∴ A is p.d. matrix.

$$\left(\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 0 & 2 & 5 & 7 & 0 & 1 & 0 \\ 0 & 0 & 1 & 3 & 7 & 11 & 0 & 0 & 1 \end{array}\right)$$

$R_2' \to R_2 - 2R_1$, $C_2' \to C_2 - 2C_1$

$$\left(\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 1 & 0 & 3 & 1 & -2 & 0 \\ -2 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 3 & 1 & 11 & 0 & 0 & 1 \end{array}\right)$$

$R_3' \to R_3 - 3R_1$, $C_3' \to C_3 - 3C_1$

$$\left(\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 & 1 & -2 & -3 \\ -2 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ -3 & 0 & 1 & 0 & 1 & 2 & 0 & 0 & 1 \end{array}\right)$$

$R_3' \to R_3 - R_2$, $C_3' \to C_3 - C_2$

$$\left(\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 & 1 & -2 & -1 \\ -2 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & -1 \\ -1 & -1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right)$$

$\therefore CAC' = I_3 \Rightarrow A = C^{-1} I_3 (C^{-1})'$

$= (C^{-1})(C^{-1})'$

$= BB^T$

where $B = C^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}^{-1}$

$= \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{pmatrix}$

(ii) Solve the equations:

$$5x_1 - 2x_2 - x_3 = 4z$$
$$5x_2 - 2x_3 - x_1 + 3z = 0$$
$$5x_3 - 2x_1 - x_2 = 3z$$

to express $x_i$'s in terms of $z$. For what value of $z$, do $x_1, x_2, x_3$ further satisfy $x_1 + x_2 + x_3 = 1$?

[C.U. 1991]

ANS:- Augmented matrix is $[A : b]$

$$= \left(\begin{array}{ccc|c} 5 & -2 & -1 & 4z \\ -1 & 5 & -2 & -3z \\ -2 & -1 & 5 & 3z \end{array}\right)$$

$R_2 \leftrightarrow R_1$

$$\sim \left(\begin{array}{ccc|c} -1 & 5 & -2 & -3z \\ 5 & -2 & -1 & 4z \\ -2 & -1 & 5 & 3z \end{array}\right)$$

$R_2' \leftrightarrow R_2 + 5R_1$, $R_3' \to R_3 - 2R_1$

$$\left(\begin{array}{ccc|c} -1 & 5 & -2 & -3z \\ 0 & 23 & -11 & -11z \\ 0 & -11 & 9 & 9z \end{array}\right)$$

$R_1' \to R_1(-1)$, $R_2' \to R_2/23$, $R_3' \to R_3/11$

$$\left(\begin{array}{ccc|c} 1 & -5 & 2 & 3z \\ 0 & 1 & -11/23 & -11z/23 \\ 0 & 1 & -9/11 & -\frac{9z}{11} \end{array}\right)$$

$R_3' \to R_3 - R_2$

$$\left(\begin{array}{ccc|c} 1 & -5 & 2 & 3z \\ 0 & 1 & -11/23 & -11/23z \\ 0 & 0 & -86/253 & -\frac{86}{253}z \end{array}\right)$$

$R_3' = R_3 / -\frac{86}{253}$

$$\left(\begin{array}{ccc|c} 1 & -5 & 2 & 3z \\ 0 & 1 & -11/23 & -11/23z \\ 0 & 0 & 1 & z \end{array}\right)$$

$\therefore x_3 = t$, $x_2 = -\frac{11}{23}t$,

$\Rightarrow x_2 = 0$

$\Rightarrow x_1 = t$

$\therefore (x_1, x_2, x_3) = t(1, 0, 1)$ is a solution.

Again $x_1 + x_2 + x_3 = 1 \Rightarrow t + 0 + t = 1 \Rightarrow t = 1/2$

15) i) Determine the null space of $A = \begin{bmatrix} 1 & 1 & -1 & 2 \\ 2 & 2 & -3 & 1 \\ -1 & -1 & 0 & -5 \end{bmatrix}$.

Also find dim (N(A)) and rank (A).

Ans:- To find dim(N(A)), we use,

$$\dim(N(A)) = 4 - \text{rank}(A).$$

We know, $\{\underset{\sim}{x} : A\underset{\sim}{x} = 0\} = [N(A)]$

Here, $\begin{pmatrix} 1 & 1 & -1 & 2 \\ 2 & 2 & -3 & 1 \\ -1 & -1 & 0 & -5 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & 0 & -1 & -3 \\ 0 & 0 & -1 & -3 \end{pmatrix}$

$\sim \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

$\sim \begin{pmatrix} 1 & 1 & 0 & 5 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix} = H$

Rank(A) = 2.

$H\underset{\sim}{x} = \underset{\sim}{0}$

$\Rightarrow x_1 + x_2 + 5x_4 = 0$

$\Rightarrow x_3 + 3x_4 = 0 \Rightarrow x_3 = -3x_4$

$\therefore x_4 = t,\ x_3 = -3t,\ x_2 = -3t,\ x_1 = -2t$

$[N(A)] = t\begin{pmatrix} 1 \\ -3 \\ -3 \\ -2 \end{pmatrix}$

$\therefore \dim(N(A)) = 4 - 2 = 2.$

(ii) Find an orthogonal matrix which diagonalizes $A=\begin{pmatrix} 6 & -2 & 2 \\ -2 & 3 & -1 \\ 2 & -1 & 3 \end{pmatrix}$. Also, find $A^8$.

ANS:- The characteristic equation of A is

$$0=|A-\lambda I_3|=\begin{vmatrix} 6-\lambda & -2 & 2 \\ -2 & 3-\lambda & -1 \\ 2 & -1 & 3-\lambda \end{vmatrix}$$

$$=\begin{vmatrix} 6-\lambda & -2 & 2 \\ 0 & 2-\lambda & 2-\lambda \\ 2 & -1 & 3-\lambda \end{vmatrix} \quad R_2' \to R_2+R_3$$

$$=(2-\lambda)\begin{vmatrix} 6-\lambda & -2 & 2 \\ 0 & 1 & 1 \\ 2 & -1 & 3-\lambda \end{vmatrix}$$

$$=(2-\lambda)\{(6-\lambda)(3-\lambda+1)+2(-2-2)\}$$

$$=(2-\lambda)(\lambda-2)(\lambda-8)$$

$\therefore \lambda=2,2,8$ are the eigen values of A.

For $\lambda=8$,

$$(A-8I_3)\underset{\sim}{x}=\underset{\sim}{0}$$

$$\Rightarrow \begin{pmatrix} -2 & -2 & 2 \\ -2 & -5 & -1 \\ 2 & -1 & -5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\Rightarrow \begin{cases} x_1+x_2-x_3=0 \\ 2x_1+5x_2+x_3=0 \\ 2x_1-x_2-5x_3=0 \end{cases}$$

Let, $x_3=t$, $\Rightarrow x_1+x_2=t$.

$\Rightarrow x_2=t-x_1$

$\therefore x_2=-t$, & $x_1=2t$.

$$\therefore \underset{\sim}{x}=t\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}, \quad \underset{\sim}{q_3}=\frac{\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}}{\sqrt{4+1+1}}=\frac{1}{\sqrt{6}}\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$$

For $\lambda=2$, $(A-2I_3)\underset{\sim}{x}=0$

$$\Rightarrow \begin{pmatrix} 4 & -2 & 2 \\ -2 & 1 & -1 \\ 2 & -1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$\Rightarrow 2x_1-x_2+x_3=0$

$x_2=t_1$, $x_3=t_2$, $x_1=\frac{t_1-t_2}{2}$.

$$\underset{\sim}{x}=t_1\begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \end{pmatrix}+t_2\begin{pmatrix} -\frac{1}{2} \\ 0 \\ 1 \end{pmatrix}$$

$\underset{\sim}{q_1}=\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$, $\underset{\sim}{q_2}=\begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}$ are LIN eigen vectors when $\lambda=2$.

$$x_1 = a_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$$

$$u_1 = \frac{a_1}{\|a_1\|} = \frac{\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}}{\sqrt{5}} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$$

$$x_2 = a_2 - (u_1 \cdot a_2)\, u_1$$

$$= \begin{pmatrix} -4/5 \\ 2/5 \\ 2 \end{pmatrix}$$

$$u_2 = \frac{x_2}{\|x_2\|} = \frac{1}{\sqrt{30}} \begin{pmatrix} -2 \\ 1 \\ 5 \end{pmatrix}$$

$$x_3 = a_3 - (a_3 u_1) u_1 - (a_3 u_2) u_2$$

$$= \frac{1}{\sqrt{6}} \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$$

$$\therefore u_3 = \frac{x_3}{\|x_3\|} = \frac{\frac{1}{\sqrt{6}} \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}}{1} = \frac{1}{\sqrt{6}} \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$$

$$\therefore Q = (u_1 \; u_2 \; u_3)$$

$$= \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{30}} & \frac{2}{\sqrt{6}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{30}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{5}{\sqrt{30}} & \frac{1}{\sqrt{6}} \end{pmatrix}$$ = orthogonal matrix.

$$\therefore Q' A Q' = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{pmatrix}$$

$\lambda_i^m$ is the eigen value of the matrix $A^m$.

So, $Q' A^8 Q = \begin{pmatrix} 2^8 & 0 & 0 \\ 0 & 2^8 & 0 \\ 0 & 0 & 8^8 \end{pmatrix}$

$$\Rightarrow A^8 = Q' \begin{pmatrix} 256 & 0 & 0 \\ 0 & 256 & 0 \\ 0 & 0 & 16777216 \end{pmatrix} Q$$

as $Q^{-1} = Q'$

# PRACTICAL PROBLEMS ON DESCRIPTIVE STATISTICS

**Q.No.1.** [CU'03] The sound intensity levels measured in decibels at 50 construction sites are

68, 63, 59, 77, 60, 57, 63, 62, 64, 73, 63, 70, 71, 65, 68, 67,
67, 62, 56, 61, 69, 64, 58, 73, 68, 66, 65, 64, 68, 67, 69, 68,
67, 70, 69, 61, 65, 62, 68, 62, 67, 64, 82, 86, 65, 69, 68, 70, 66.

(i) Prepare a frequency table for considering 6 classes of equal width, (ii) Calculate the exact mean and variance of sample from the raw data. (iii) Estimate the mean and variance from the grouped frequency distribution by assuming each value is equal to the mid-point of the class to which it belongs. (iv) Compare the answers in (ii) and (iii) and in this connection mention about the hole of Sheppard's correction.

**Q.No.2.** [CU'96] A Chemical compound contains 12.5% of iron was given to two technicians A and B for chemical analysis. A made of 15 determinations, and B 10 determinations, of the percentage of iron. Their results are given in the following table

| Determinations by A | | | Determinations by B | |
|---|---|---|---|---|
| 12.46 | 12.11 | 12.47 | 12.23 | 12.11 |
| 11.80 | 12.44 | 12.11 | 11.91 | 12.45 |
| 12.40 | 11.85 | 12.56 | 12.45 | 12.39 |
| 11.95 | 12.12 | 12.65 | 12.22 | 12.37 |
| 12.77 | 12.43 | 12.72 | 12.05 | 12.65 |

(i) Find seperately for A and B various measures of central tendency and dispersion. Also, find their respective coefficients of variation.
(ii) Based on the above measures prepare a report comparing the accuracy and consistency of the two technicians.

**Q.No.3.** The following tables relate to the weights of new born babies recored at the different clinics. One of the clinics is located in a locality where the average family income is also homogeneous throughout the locality where as the other clinic caters to the same category but also has a considerable number of citizens from a significant lower income group. Examine the locations, dispersions and the shapes of the two distributions and interpret your findings.

[CU'1995]

| Weights in (kgs.) | Clinic I | Clinic II |
|---|---|---|
| 0 - 1 | 0 | 0 |
| 2 - 3 | 1 | 43 |
| 3 - 4 | 60 | 230 |
| 4 - 5 | 304 | 372 |
| 5 - 6 | 318 | 320 |
| 6 - 7 | 56 | 54 |
| > 7 | 3 | 2 |
| | 0 | 0 |

Q.No.4. [CU'2001] The scores in English of 250 candidates appearing in an Examination have

$$m_1' = 39.7213 \;,\; \delta = 9.8894$$

$$\frac{m_3}{m_2^{3/2}} = -0.1182 \;,\; b_2 = 2.9719.$$

It is later found on scrutiny that score 61 has been wrongly recorded as 51. Obtain the correct values of $b_1$ and $b_2$.

Q.No.5. Particulars relating to the monthly wage distributions of two manufacturing firms are given below:

| Measures | Firm A | Firm B |
|---|---|---|
| Mean wage | Rs. 1477 | 1495 |
| Median wage | Rs. 1389 | 1354 |
| Modal wage | Rs. 1350 | 1312 |
| Quartiles | Rs. 1278, 1422 | 1262, 1435 |
| S.D. | Rs. 87 | Rs. 99 |

Compare the two distributions w.r.t. the characteristic central tendency, dispersion and skewness, kurtosis.

Q.No.6. The weights (in grams) of 25 indicator housing used on gauges are as follows:

| | | | | |
|---|---|---|---|---|
| 102.0 | 106.3 | 106.6 | 108.8 | 107.7 |
| 106.1 | 105.9 | 106.7 | 106.8 | 110.2 |
| 101.7 | 106.6 | 106.3 | 110.2 | 109.9 |
| 102.0 | 105.8 | 106.1 | 106.7 | 108.3 |
| 102.0 | 106.8 | 110.0 | 107.9 | 109.3 |

(a) Construct an ordered stem-leaf display using integers as the stems and tenths as the leaves.
(b) Find the five-number summary of the data and draw a box-plot.
(c) Are there any suspected outliers?

Q.No.7. The following table gives the yearly corn yield (X) in bushels per acre, in six Corn Belt states (Iowa, Illinois, Nebraska, Missouri, Indiana, Ohio) and rain fall measurements in inches (Y) in six states from 1915 to 1927.

| Year: | 1915 | 1916 | 1917 | 1918 | 1919 | 1920 | 1921 | 1922 | 1923 | 1924 | 1925 | 1926 | 1927 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X : | 33·3 | 29·7 | 35·0 | 29·9 | 35·2 | 38·3 | 35·2 | 35·5 | 36·7 | 26·8 | 38·0 | 31·7 | 32·6 |
| Y : | 16·5 | 9·3 | 9·4 | 9·0 | 9·5 | 11·6 | 12·1 | 8·0 | 10·7 | 13·9 | 11·8 | 11·6 | 10·1 |

Fit a linear regression model : $X = \alpha + \beta Y + \epsilon$ to the data by the method of least squares, making the usual assumptions.

Q.No.8. [CU'1997] In an experiment on certain fertilizers were applied at various levels (in appropriate units) with resulting yields (in appropriate unit) as follows:

| Fertilizer level ($x$) : | 0 | 5 | 10 | 15 | 20 | 25 | 30 |
|---|---|---|---|---|---|---|---|
| Yield ($y$) : | 27·1 | 32·1 | 35·0 | 36·2 | 36·9 | 36·1 | 35·2 |

(i) Fit an appropriate polynomial to the given data.
(ii) Obtain a suitable measure of association between $x$ and $y$ and comment
(iii) Obtain the optimum fertilizer level which maximizes yield.

Q.No.9. [CU'2001] The following data relate to the height ($x$) and weight ($y$) of 15 students

| $x$ | 5'6" | 5'8" | 5'3" | 5'9" | 5'6" | 5'6" | 5'9" | 5'3" | 5'6" | 5'6" | 5'9" | 5'6" | 5'3" | 5'9" | 5'3" |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 55 | 52 | 60 | 61 | 57 | 54 | 60 | 51 | 58 | 50 | 59 | 55 | 44 | 65 | 47 |

Compute $r_{yx}$ and $e_{yx}$. Is the regression linear? If not, compute a measure of deviation from linearity.

Q.No.10. Two supervisors ranked 12 workers working under them in order of efficiency as follows. Compute Spearman's rank correlation coefficient between the two rankings. Also compute kendall's $\gamma$.

| Worker | A | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Supervison I : | 5 | 6 | 1 | 2 | 3 | 8 | 8 | 4 | 7 | 11 | 10 | 12 |
| Supervison II : | 5 | 5 | 1 | 1 | 1 | 9 | 7 | 4 | 8 | 10 | 12 | 10 |

Q.No.11. [CU'2008]

The following table gives data on income in thousand dollars($x$), the number of families ($N$) at income $x$ and the number of families owing a house ($n$).

| $x$ : | 10 | 13 | 15 | 20 | 25 | 30 | 35 | 40 |
|---|---|---|---|---|---|---|---|---|
| $N$ : | 60 | 80 | 100 | 70 | 65 | 50 | 40 | 25 |
| $n$ : | 18 | 28 | 45 | 36 | 39 | 33 | 30 | 20 |

Suggest an appropriate regression equation to explain the effect of income on owning a house. Also estimate the parameters of this equation using the above data and predict there from the proportion of families of income 32 thousands dollars who own a house.

Solutions:→ PROBLEMS ON DESCRIPTIVE STATS.

1. Minimum value = 56, Maximum value = 86

Range = 30.

(i) Considering 6 classes of width 5:

| Class-interval limit | Class-boundary | Frequency ($f_i$) | Tally marks | Cls. mark ($x_i'$) |
|---|---|---|---|---|
| 53-58 | 52.5 - 58.5 | 2 | // | 55.5 |
| 59-64 | 58.5 - 64.5 | 16 | 𝍸 𝍸 𝍸 / | 61.5 |
| 65-70 | 64.5 - 70.5 | 25 | 𝍸 𝍸 𝍸 𝍸 𝍸 | 67.5 |
| 71-76 | 70.5 - 76.5 | 4 | /// | 73.5 |
| 77-82 | 76.5 - 82.5 | 2 | // | 79.5 |
| 83-88 | 82.5 - 88.5 | 1 | / | 85.5 |
| TOTAL | | 50 | | |

(i) From the raw data,

$$\text{mean}(\bar{x}) = \frac{1}{50}\sum_{i=1}^{50} x_i$$

$$\text{Variance}(s^2) = \frac{1}{50}\sum_{i=1}^{50} x_i^2 - \bar{x}^2$$

(ii) From the grouped data,

$$\text{mean}(\bar{x}') = \frac{1}{\sum_{i=1}^{6} f_i}\sum_{i=1}^{6} x_i' f_i$$

$$\text{Variance}(s'^2) = \frac{1}{50}\sum_{i=1}^{6} x_i' f_i - \bar{x}'^2$$

(iv) The mean and variance computed from the raw data and from grouped data are different.

In general, frequency distribution, we assume that all the values in a class are equal to the mid-point which is true if values are uniformly distributed over the class. So mean and variance computed from the raw data and grouped data are different due to the error due to grouping, the correction for the error due to grouping is given by Sheppard's correction.

$m_1'$ (corrected) $= m_1'$

$m_2$ (corrected) $= m_2 - \frac{c^2}{12}$ , $c$ = class width

2. (i) Exact mean of A $= \frac{1}{15}\sum_{i=1}^{15} x_{iA} =$

Exact mean of B $= \frac{1}{10}\sum_{i=1}^{10} x_{iB} =$

Measure of Dispersion, $SD(A) = \sqrt{\frac{1}{15}\sum_{i=1}^{15} x_{iA}^2 - \bar{x}_A^2}$

$=$

$SD(B) = \sqrt{\frac{1}{10}\sum_{i=1}^{10} x_{iB}^2 - \bar{x}_B^2} =$

$RMSD_A\ (12.5) = \sqrt{\frac{1}{15}\sum_{i=1}^{15} (x_{iA} - 12.5)^2}$

$RMSD_B\ (12.5) = \sqrt{\frac{1}{10}\sum_{i=1}^{10} (x_{iB} - 12.5)^2}$

Co-efficient of variation, $CV(A) = \frac{SD(A)}{\bar{x}_A} \times 100\% =$

$CV(B) = \frac{SD(B)}{\bar{x}_B} \times 100\% =$

(ii) As a measure of accuracy, we use RMSD about 12.5, The smaller the RMSD, is more accuracy.

As a measure of consistency, we use C.V. The smaller C.V. is more consistency in Data set.

3) The frequency distribution of clinic-I is highly positively skewed and that of clinic-II is near about symmetric.
[To get the measure of distribution we may draw histogram of the frequency distribution].

(i) As a measure of location, we may use median.

∴ Median of Clinic I is

∴ Median of Clinic II is

(ii) For Dispersion, we use $Q.D. = \frac{Q_3 - Q_1}{2}$.

∴ Q.D. of Clinic I is

& Q.D. of clinic II is

(iii) For shape, we use, $S_k = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$

∴ skewness of clinic I is

∴ Skewness of clinic II is

iv) For peackedness, we use, $k_p = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$

∴ kurtosis of Clinic I is

∴ kurtosis of Clinic II is

The smaller the value of $k_p$, the higher the kurtosis.

| Weights in (kgs) | Clinic I | Clinic II | ≤ I | ≤ II |
|---|---|---|---|---|
| 0-1 | 0 | 0 | 0 | 0 |
| 1-2 | 1 | 43 | 1 | 43 |
| 2-3 | 60 | 230 | 61 | 273 |
| 3-4 | 304 | 372 | 365 | 645 |
| 4-5 | 318 | 320 | 683 | 965 |
| 5-6 | 56 | 54 | 739 | 1019 |
| 6-7 | 3 | 2 | 742 | 1021 |

$P_{90}^I = 4 + \frac{667.8 - 365}{318} \times 1$

$= 4.95$

$P_{10}^I = 3 + \frac{74.2 - 61}{304}$

$= 3.04$

$Q_1^I = 3 + \frac{185.5 - 61}{304} \times 1$

$= 3.41$

$Q_2^I = 4 + \frac{371 - 365}{318} \times 1$

$= 4.02$

$Q_3^I = 4 + \frac{556.5 - 365}{318} \times 1$

$= 4.60$

$n = 250$

$m_1'$ (incorrected) $= 39.7213$

$\Rightarrow \frac{1}{250}\sum_{i=1}^{50} x_i$ (inconnected) $= 39.7213$

$\Rightarrow \sum_{i=1}^{50} x_i$ (inconnected) $= 9930.325$

$\therefore \sum_i x_i$ (connected) $= (9930.325 - 51 + 61)$

$= 9940.325$

$\therefore m_1'$ (connected) $= 39.7613$

$$s\,(\text{inconnected}) = \sqrt{\frac{1}{250}\sum x_i^2(\text{inconnect}) - \bar{x}^2(\text{inconnect})}$$

$= 9.8894$

$\therefore \sum x_i^2$ (inconnected) $= 418895.4765$

$\therefore \sum x_i^2$ (connected) $= 418895.4765 - 51^2 + 61^2$

$= 420015.4765$

$$\therefore s\,(\text{connected}) = \sqrt{\frac{1}{250}\sum x_i^2(\text{connected}) - \bar{x}^2(\text{connected})}$$

$= 9.9549$

$$b_1(\text{inconnected}) = \frac{m_3(\text{inconnected})}{m_2^{3/2}(\text{inconnected})} = -0.1182$$

$\Rightarrow m_3$ (inconnected) $= -114.32134$.

5) For Firm A: Mean > Median > Mode and the distribution is +vely skewed.

For Firm B: Mean > Median > Mode & the distn. is +vely skewed.

(For skewed distn. measures should be based on quantiles)

Measures:-

(i) Location: Median ($Q_2$)

(a) Median of the Firm A is 1389

(b) " " " B is 1354

(ii) Dispersion: $Q.D. = \frac{Q_3 - Q_1}{2}$.

(a) Q.D. of Firm A is $\frac{1422-1278}{2} = 72$

(b) Q.D of Firm B is $\frac{1435-1262}{2} = 86.5$

(iii) Skewness: $S_k = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$

(a) $S_k$ of firm A is $(1422+1278-2\times1350)/144 = 0$

(b) $S_k$ of firm B is $(1262+1435-2\times1312)/173 = 0.42$

(iv) kurtosis: $K_p = \frac{Q_3 - Q_1}{2 \times SD}$

(a) $K_p$ of firm A is $(1422-1278)/2\times87 = 0.8276$

(b) $K_p$ of firm B is $(1435-1262)/2\times99 = 0.8739$

[ $(Q_3 - Q_1)$ represents the length within which we have central 50% value. The smaller the length $(Q_3-Q_1)/2$, the higher the kurtosis. To set a unit free measure, we divide $\frac{Q_3-Q_1}{2}$ by S.D.]

6) (a) Consider the integer as stems and decimals as leaves.

| Stem | Leaf |
|---|---|
| 101 | 7 |
| 102 | 0 0 0 |
| 105 | 8 9 |
| 106 | 1 3 3 6 6 7 7 8 8 |
| 107 | 3 7 9 |
| 108 | 8 |
| 109 | 1 3 9 |
| 110 | 0 2 2 |

Comment: The frequency distribution is located at 106 (median or mode) & is +vely skewed.

(b) Minimum value = 101.7, maximum value = 110.2

$Q_1 = \frac{25}{4}$ th ordered value = 6th ordered value + $\frac{1}{4}$ (7th ordered value − 6th ordered value)

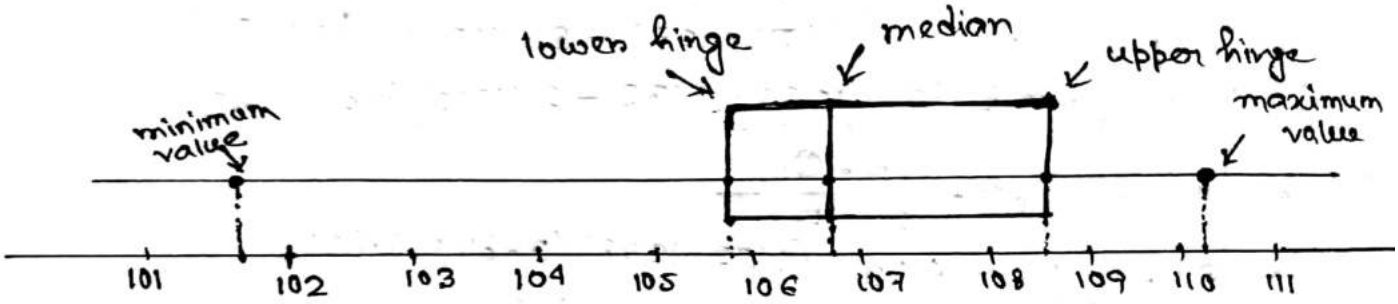
$= 105.9 + \frac{1}{4}(.2) = 105.95$ gram

$Q_2 = \frac{25}{2}$ th ordered value = 12th ordered value + $\frac{1}{2}$ (13th ordered value − 12th ordered value)

$= 106.7 + \frac{1}{2} \times 0 = 106.7$ gm.

$Q_3 = \frac{25 \times 3}{4}$ th ordered value = 18th ordered value + $\frac{3}{4}$ (19th ordered value − 18th ordered value)

$= 107.9 + \frac{3}{4}(.9) = 108.575$ gm.




Comment:— The frequency distribution is located at (near about) 107. The H-~~spread~~ spread is small. Hence, the dispersion of the data is not high. The distance between upper hinge and median is greater than that of lower hinge and median. The distribution is +vely skewed. The length of H-spread is small w.r.t. the range, the kurtosis is high.

(c) The data values that are smaller than $L_H - \frac{3}{2}H$ or greater than $L_U + \frac{3}{2}H$ are called outliers.

$L_H - \frac{3}{2}H = 105.95 - \frac{3}{2}(108.575 - 105.95)$

$= 102.0125$

$L_U + \frac{3}{2}H = 108.575 + \frac{3}{2}(108.575 - 105.95)$

$= 112.5125$.

In the data, 101.7, 102.0, 102.0, 102.0 are the outliers of the data set.

17) Linear Regression model:

$$X = \alpha + \beta Y + \epsilon$$

The constant $\alpha, \beta$ are determined by minimising error sum of square or residual sum of square.

$$S = \sum_{i=1}^{n} (x_i - \alpha - \beta y_i)^2$$

The normal equations are:

$$\sum_{i=1}^{n} x_i = n\alpha + \beta \sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n} x_i y_i = \alpha \sum_{i=1}^{n} y_i + \beta \sum_{i=1}^{n} y_i^2$$

From data, $n = 13$, $\sum x_i = 437.9$, $\sum y_i = 143.3$,

$\sum x_i y_i = 4419.35$, $\sum y_i^2 = 1543.67$.

The normal equations are:

$$437.9 = 13\alpha + (\beta \times 143.3) \quad \text{——} ①$$

$$4419.35 = 143.3\alpha + 1543.67\beta \quad \text{——} ②$$

① $\times 143.3$ − ② $\times 13$, we get

$$\beta = 11.34.$$

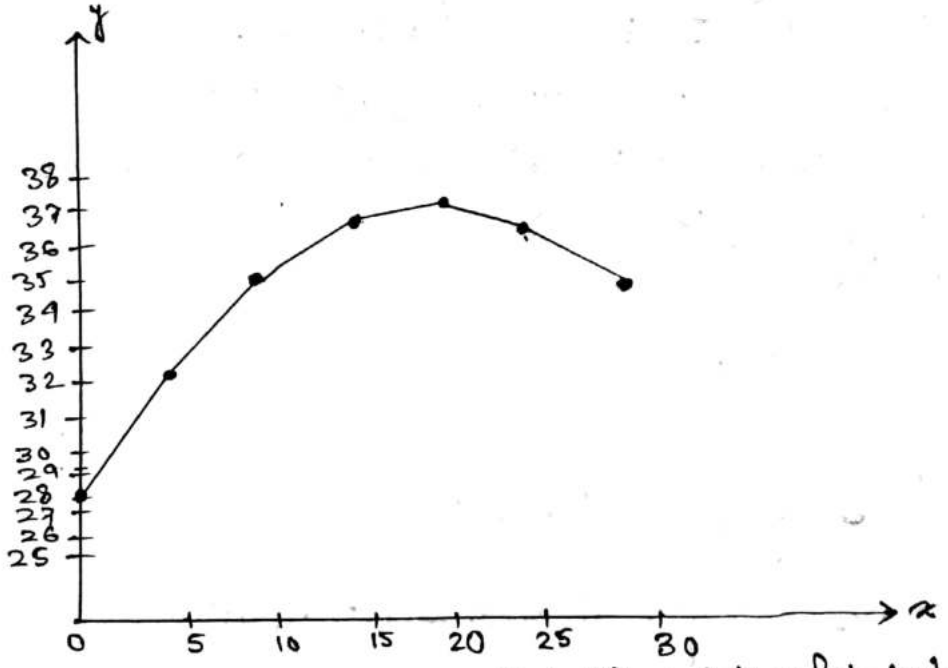
$$\therefore \alpha = -91.36$$

$$\therefore \hat{\alpha} = -91.36, \quad \hat{\beta} = 11.34$$

Hence, the predicting formula is —

$$X = -91.36 + 11.34Y.$$

18)




From scatter plot, it is clear that the relationship between x & y is approximately a quadratic.

(ai) A measure of association between $x$ and $y$ is the measure of usefulness of the 2nd degree polynomial (least square) regression as a predicting formula, i.e.

$$r_2^2 = \frac{\Sigma(Y_{pi} - \bar{Y}_p)^2}{\Sigma(y_i - \bar{y})^2} = 1 - \frac{\Sigma(y_i - \hat{Y}_{pi})^2}{\Sigma(Y_i - \bar{y})^2}$$

| $x$ | $y$ | $u_i = \frac{x_i - 15}{5}$ | $u_i^2$ | $u_i^3$ | $u_i^4$ | $u_i y_i$ | $u_i^2 y_i$ | $y_i^2$ |
|---|---|---|---|---|---|---|---|---|
| 0 | 27.1 | −3 | | | | | | |
| 5 | 32.1 | −2 | | | | | | |
| 10 | 35.0 | −1 | | | | | | |
| 15 | 36.2 | 0 | | | | | | |
| 20 | 36.9 | 1 | | | | | | |
| 25 | 36.1 | 2 | | | | | | |
| 30 | 35.2 | 3 | | | | | | |

Let, $Y = a_0 + a_1 u + a_2 u^2$

the constants $a_0, a_1, a_2$ are determined by minimizing

$$\sum_{i=1}^{n} (y_i - a_0 - a_1 u_i - a_2 u_i^2)$$

∴ Normal eqns are:-

$$\Sigma y_i = n a_0 + a_1 \Sigma u_i + a_2 \Sigma u_i^2$$

$$\Sigma u_i y_i = a_0 \Sigma u_i + a_1 \Sigma u_i^2 + a_2 \Sigma u_i^3$$

$$\Sigma u_i^2 y_i = a_0 \Sigma u_i^2 + a_1 \Sigma u_i^3 + a_2 \Sigma u_i^4$$

Here, $n = 7$, $\Sigma u_i = 0$, $\Sigma u_i^2 = 28$, etc.

Now, the total variability $= \Sigma y_i^2 - n\bar{y}^2$

& unexplained variability or Residual Sum of squares (RSS)

$$= \sum_{i=1}^{n} (y_i - Y_{pi})^2$$

$$= \sum_i (y_i - Y_{pi}) e_i$$

$$= \sum_i y_i e_i$$

$$= \sum_i y_i (y_i - \hat{a}_0 - \hat{a}_1 u_i - \hat{a}_2 u_i^2)$$

$$= \sum_i y_i^2 - \hat{a}_0 \sum_i y_i - \hat{a}_1 \sum_i u_i y_i - \hat{a}_2 \sum_i y_i u_i^2$$

$\therefore R^2 = 1 - \frac{\text{Unexplained variability}}{\text{Total variability}}$.

(iii) $y = \hat{a}_0 + \hat{a}_1 u + \hat{a}_2 u^2$, $u = \frac{x-15}{5}$

$\frac{dy}{du} = \hat{a}_1 + 2\hat{a}_2 u$

$\Rightarrow u = -\frac{\hat{a}_1}{2\hat{a}_2}$, since $\frac{dy}{dx} = 0$

& $\frac{d^2y}{du^2} = 2\hat{a}_2 < 0$

Hence, $y$ is maximum when $u_0 = -\frac{\hat{a}_1}{2\hat{a}_2}$

$\Rightarrow x = 15 + 5u_0$

9)

| $x_i$ | $y_{ij}$ | $\bar{y}_i$ | $n_i$ | $n_i \bar{y}_i^2$ |
|---|---|---|---|---|
| $x_1 = 5'3'' = 63''$ | 52, 60, 51, 44, 47 | 50.8 | 5 | 12903 |
| $x_2 = 5'6'' = 66''$ | 55, 57, 54, 58, 50, 55 | 54.8 | 6 | 18018 |
| $x_3 = 5'9'' = 69''$ | 61, 60, 58, 65 | 61 | 4 | 14884 |

Now, $e^2_{yx} = \frac{\sum_{i=1}^{k} n_i (\bar{y}_i - \bar{y})^2}{\sum_i \sum_j (y_{ij} - \bar{y})^2} = \frac{\sum n_i \bar{y}_i^2 - n\bar{y}^2}{\sum_i \sum_j y_{ij}^2 - n\bar{y}^2}$, where

$\bar{y} = \frac{\sum_{i=1}^{k} n_i \bar{y}_i}{\sum_{i=1}^{k} n_i}$, $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$, $i = 1(1)k$

$\bar{x} = \frac{1}{\sum n_i} \sum n_i x_i$, $\bar{x} = 65.8$, $\bar{y} = \frac{\sum_i \sum_j y_{ij}}{n} = 55.13$

$$\therefore \eta^2_{yx} = \frac{45805 - 45590}{46039 - 45590} = 0.4788$$

$$\therefore \eta_{yx} = 0.692$$

$$r^2_{yx} = \frac{\sum_{i=1}^{k} n_i x_i \bar{y}_i - n\bar{x}\bar{y}}{\sqrt{\sum_i n_i x_i^2 - n\bar{x}^2}\sqrt{\sum_i \sum_j y_{ij}^2 - n\bar{y}^2}}$$

$$= \frac{54538.8 - 54413.31}{8.97 \times 21.20}$$

$$= 0.1658$$

$$\therefore r_{yx} = 0.2585$$

The regression equation is linear iff $\eta^2_{yx} = r^2_{yx}$. A measure of deviation of the (true) regression equation from linearity is $d^2_{yx} = \eta^2_{yx} - r^2_{yx} = 0.4788 - 0.1658$

$= 0.413$.

10) Both series of efficiency have some tie. So, we replace the tied rank by their average of those ...

| Workers | A | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I | 5 | 6 | 1 | 2 | 3 | 8.5 | 8.5 | 4 | 7 | 11 | 10 | 12 |
| II | 5.5 | 5.5 | 2 | 2 | 2 | 9 | 7 | 4 | 8 | 10.5 | 12 | 10.5 |

$$r_R = \frac{\frac{n^2-1}{12} - \frac{T_u + T_v}{2} - \frac{1}{2n}\sum d_i^2}{\sqrt{\frac{n^2-1}{12} - T_u}\sqrt{\frac{n^2-1}{12} - T_v}}$$

11)

Explanatory variable ($x$): income,
Response Variable ($y$): ocoing a house,
Hence $y$ is Binary.
Hence, a logistic regression may be appropriate.

In logistic regression,

$$\bar{y}_x = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$$

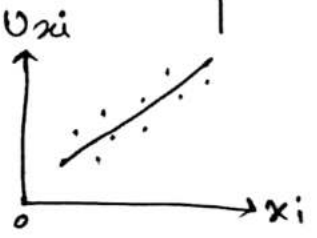
$$\Rightarrow \ln\left(\frac{\bar{y}_n}{1-\bar{y}_n}\right) = \alpha+\beta x.$$

$$\Rightarrow u_x = \alpha+\beta x$$

$\Rightarrow x$ & $u_x$ are linearly related,

| $x$ | $y$-values | Array total $= \sum_{j=1}^{N_i} y_{ij}$ | $\bar{y}_{xi} = \frac{n_i}{N_i}$ | $v_{xi}$ |
|---|---|---|---|---|
| $x_1=10$ | $y_{11} \ldots y_{1N_1}$ | | | |
| $x_2=13$ | $y_{21} \ldots y_{2N_2}$ | | | |
| $x_3=15$ | . . . . | | | |
| ⋮ | | | | |
| $x_8=40$ | $y_{81} \ldots y_{8N_8}$ | | | |

Plot the points $(x_i, v_{xi})$, $i=1(1)8$.
From graph the points $(x_i, v_{xi})$ lie near a straight line.




Hence logistic regression is appropriate.
Hence, $\alpha$ & $\beta$ can be determined by minimising $\sum w_i (v_{xi} - \alpha - \beta x_i)^2$
where $w_i = N_i \bar{y}_{xi}(1-\bar{y}_{xi})$

Normal equn., $\sum w_i v_{xi} = \alpha \sum w_i + \beta \sum x_i w_i$

$$\sum w_i x_i v_{xi} = \alpha \sum x_i w_i + \beta \sum x_i^2 w_i$$

| $v_{xi}$ | $w_i$ | $x_i w_i$ | $x_i^2 w_i$ | $w_i v_{xi}$ | $w_i x_i v_{xi}$ |
|---|---|---|---|---|---|
| | | | | | |

Fitted logistic regression is

$$Y = \frac{e^{\hat{\alpha} + \hat{\beta} x}}{1 + e^{\hat{\alpha} + \hat{\beta} x}}$$

# PROBLEMS ON CATEGORICAL DATA ANALYSIS

1. In one phase of a study regarding the effectiveness of several drugs on post-operating nausea 167 patients were assigned at random, 30 to Drug-P, 67 to Drug-C and the remaining 70 to Placebo. The no. of patients suffering from severe, moderate, slight or no nausea is shown below:

| | SEVERE | MODERATE | SLIGHT | NO NAUSEA | TOTAL |
|---|---|---|---|---|---|
| PLACEBO | 8 | 8 | 19 | 35 | 70 |
| DRUG-P | 2 | 3 | 5 | 20 | 30 |
| DRUG-C | 3 | 4 | 15 | 45 | 67 |
| TOTAL | 13 | 15 | 39 | 100 | 167 |

Calculate a suitable measure of association.

2. A drug, supposed to have some effect in curing diabetis was treated on 100 patients in a certain hospital and their records were compared with 100 other patients not treated with drug. Study the efficacy of the drug in curing diabetis.

| | Cured | Not cured |
|---|---|---|
| Treated | 54 | 46 |
| Untreated | 24 | 76 |

3. The following table gives the data on the results of a visual test and a Balance test performed on 413 male college students.

| | Left eyed | Ambiocular | Right-eyed | TOTAL |
|---|---|---|---|---|
| Left handed | 48 | 25 | 52 | 125 |
| Ambidestrous | 32 | 13 | 25 | 70 |
| Right handed | 94 | 33 | 91 | 218 |
| TOTAL | 174 | 71 | 168 | 413 |

Compute at least two measures of association and comment.

4. A study was conducted to find out what people of different educational level feel about the role of the caste in life.

| Role / Edu. level | No Role | Little Role | Large Role | Complete Role |
|---|---|---|---|---|
| Upto cls-8 (Grp-I) | 2 | 8 | 25 | 36 |
| Cls-9-H.S. (Grp-II) | 10 | 15 | 14 | 12 |
| Graduate (Grp-III) | 29 | 24 | 2 | 1 |

(a) Compute a suitable measure to find if there is any association between educational level and the individual's perception towards the role of region or caste in social life. Interpret your finding.

(b) Merge the columns 'No role' and 'little role' and rename it as 'minor role', similarly merge, the last two columns 'Large role' and 'complete Role' and rename it as 'Major Role'. Find the odds ratio of

i) Group I with Group II

ii) Group III cons.t Group I, cons.t. the Minor and Major roles and interpret your findings.

# PROBLEMS ON CATEGORICAL DATA ANALYSIS

**Solutions:-**

1. We now calculate $\sum_i \sum_j \frac{(n_{ij})^2}{n_i n_j}$

$$= \frac{8^2}{13\times70} + \frac{2^2}{13\times30} + \frac{3^2}{13\times67} + \frac{8^2}{70\times15} + \frac{3^2}{15\times30} + \frac{4^2}{15\times67} + \frac{19^2}{39\times70} + \frac{5^2}{39\times30} + \frac{15^2}{39\times67} + \frac{35^2}{100\times70} + \frac{20^2}{100\times30} + \frac{45^2}{100\times67}$$

$$= 1.0381.$$

**Mean-square contingency:-**

$$\chi^2 = N\left\{\sum_{i=1}^{3}\sum_{j=1}^{4} \frac{(n_{ij})^2}{n_i n_j} - 1\right\}$$

$$= 167\,(1.0381 - 1)$$

$$= 6.3827.$$

**Karl Pearson's coefficient of mean-square contingency :-**

$$C = \sqrt{\frac{\chi^2}{N+\chi^2}} = 0.1916$$

**Tschuprow's coefficient :-**

$$T^2 = \frac{\chi^2}{N\sqrt{(s-1)(t-1)}}, \quad s=4, t=3$$

$$= 0.1247.$$

2.

| | Cured (B) | Not cured ($\beta$) | Total |
|---|---|---|---|
| Treated (A) | $(AB) = 54$ | $(A\beta) = 46$ | 100 |
| Untreated ($\alpha$) | $(\alpha B) = 24$ | $(\alpha\beta) = 76$ | 100 |

A measure of association between A and B in a 2X2 table is:

(i) Yule's coefficient of association:-

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= 0.5760$$

(ii) Yule's coefficient of colligation:-

$$Y = \frac{\sqrt{(AB)(\alpha\beta)} - \sqrt{(A\beta)(\alpha B)}}{\sqrt{(AB)(\alpha\beta)} + \sqrt{(A\beta)(\alpha B)}}$$

$$= 0.3169.$$

Hence, A and B have moderate positive association.
i.e. Drug has moderate effect on Diabetis.

3. Here both the characters 'Balance Test' (A) and 'Visual test' (B) are in normal scale.

| A \ B | $B_1$ | $B_2$ | $B_3$ | TOTAL |
|---|---|---|---|---|
| $A_1$ | $f_{11} = 48$ | $f_{12} = 25$ | $f_{13} = 52$ | $f_{10} = 125$ |
| $A_2$ | $f_{21} = 32$ | $f_{22} = 13$ | $f_{23} = 25$ | $f_{20} = 70$ |
| $A_3$ | $f_{31} = 94$ | $f_{32} = 33$ | $f_{33} = 91$ | $f_{30} = 218$ |
| TOTAL | $f_{01} = 174$ | $f_{02} = 71$ | $f_{03} = 168$ | 413 |

Here, we shall use measures based on $\chi^2$.

$$\chi^2 = N \left\{ \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{f_{ij}^2}{f_{oj} f_{io}} - 1 \right\}$$

$$= 2.3738$$

(i) Karl Pearson's coefficient:- $C_{AB} = \sqrt{\frac{\chi^2}{N + \chi^2}}$

$$= \sqrt{\frac{2.3738}{413 + 2.3738}}$$

$$= 0.0756.$$

(ii) Tschuprow's coefficient :- $T_{AB} = \sqrt{\frac{\chi^2}{N \sqrt{(r-1)(s-1)}}}$

$$= 0.05361.$$

4. (a) Here both the characters → educational level (X) and Role of religion (Y) are in ordinal scale.
We shall use measures based on the no. of concordant pairs (C) and no. of discordant pairs (D).

A pair (i, j) of individuals with scores $(x_i, y_i)$ and $(x_j, y_j)$ is in concordency if

$$\{x_i > x_j, y_i > y_j\} \text{ or } \{x_i < x_j, y_i < y_j\}$$

and discordent if

$$\{x_i > x_j, y_i < y_j\} \text{ or } \{x_i < x_j, y_i > y_j\}.$$

a pair (i, j) is on tie w.r.t. X if $x_i = x_j$.

Here, C = No. of concordent pairs

$$= 2(15+14+12+24+2+1) + 8(14+12+2+1) + 25(2+1) + 36(0) + 10(24+12+1) + 15(2+1) + 14 \times 1$$

$$= 1022.$$

D = No. of discordent pairs

$$= 8(10+29) + 25(10+15+29+24) + 36(10+15+14+29+24+2) + 15 \times 29 + 14(29+24) + 12(29+24+2)$$

$$= 7483.$$

Goodman-Kruskal ($\gamma$) measures of association:—

$$\gamma = \frac{C - D}{C + D} = -.7597.$$

So, we can say educational level and the individual's perception towards the role of religion is negatively associated.

(b)

| | Minor Role | Major Role |
|---|---|---|
| Grp I | $f_{11} = 10$ | $f_{12} = 61$ |
| Grp II | $f_{21} = 25$ | $f_{22} = 26$ |
| Grp III | $f_{31} = 53$ | $f_{32} = 3$ |

(i) Sample odds ratio of Grp. I with Grp. II $= \frac{f_{11} f_{22}}{f_{21} f_{12}}$

$\simeq 0.17$.

Hence the success in 'minor role' in Grp I is less likely than Grp II, i.e., Hence, more people feel that religion has 'Major Role' from Grp I compare to Grp II.

(ii) Sample odds ratio of Grp III w.r.t. Grp I is

$$\hat{\theta} = \frac{f_{11} f_{22}}{f_{21} f_{22}} = 107.76$$

| | Minor | Major |
|---|---|---|
| Grp III | $f_{11} = 53$ | $f_{12} = 3$ |
| Grp I | $f_{21} = 10$ | $f_{22} = 61$ |

The success 'minor role' in Grp III is more likely than Grp I.

# PRACTICAL PROBLEMS FROM VITAL STATISTICS

**Q.1.** [CU '92, 09] Consider the following data set for two countries

Country I

| Age Group | Population size (in '000) $P_x$ | No. of Males (in '000) $mP_x$ | No. of Deaths $D_x$ | Number of death in males $mD_x$ |
|---|---|---|---|---|
| < 1 | 100 | 50 | 750 | 400 |
| 1-4 | 400 | 200 | 3000 | 1600 |
| 5-20 | 1500 | 750 | 12000 | 6500 |
| 21-100 | 4000 | 1500 | 31000 | 12000 |

Country II

| Age Group | Population size (in '000) | No. of Males (in '000) | No. of Deaths | Number of deaths in males |
|---|---|---|---|---|
| < 1 | 15 | 7.5 | 12 | 6 |
| 1-4 | 60 | 30 | 45 | 24 |
| 5-20 | 200 | 100 | 160 | 85 |
| 21-100 | 440 | 170 | 350 | 200 |

(i) Compute the CDR for both the countries.
(ii) Compute the ASDR for both the countries seperately for male, female and total population.
(iii) In order to compare the mortality situation of the two countries propose a good measure and calculate its value for both the countries.

**Q.2.** [CU '05]
Fill in the blanks of the following table which are marked with question marks:

| Age(x) | $l_x$ | $d_x$ | $q_x$ | $p_x$ | $L_x$ | $T_x$ | $e_x^\circ$ |
|---|---|---|---|---|---|---|---|
| 20 | 6,93,435 | ? | ? | ? | ? | 35081126 | ? |
| 21 | 6,90,673 | – | – | – | – | – | ? |

**Q.3.** The number of persons dying at age 75 is 476 and the complete expectation of life at 75 and 76 years are 3.92 and 3.66 years. Find the numbers living at ages 75 and 76.

Q.4. [CU'08,'05] From the following data relating to a particular community compute the GRR and the NRR. Interpret your results.

| Age of mother | No. of Women | No. of Births | Survival factor |
|---|---|---|---|
| 15-19 | 9000 | 140 | 0.920 |
| 20-24 | 9200 | 1312 | 0.914 |
| 25-29 | 8900 | 1067 | 0.908 |
| 30-34 | 8600 | 771 | 0.891 |
| 35-39 | 8400 | 468 | 0.879 |
| 40-44 | 8500 | 160 | 0.869 |

Assume that 48.7 % of the total are female births. Survival factor gives the Rate of survival from Birth to the mid point of the corresponding age-group.

[CU'07]

Q.5. The following table provides data of female population and the number of live births in different age groups in the USA in 2004.

| Age Group (years) | Female Population ('000) | Live Births ('000) |
|---|---|---|
| 14-19 | 20,724 | 422 |
| 20-24 | 20,973 | 1034 |
| 25-29 | 19,555 | 1104 |
| 30-34 | 20,467 | 966 |
| 35-39 | 21,050 | 476 |
| 40-44 | 23,055 | 104 |
| 45-49 | 22,121 | 6 |

Given that in 2004, the total population of the USA was 293,657,00 and the sex-ratio at birth was 105 male births per 100 female births. Compare, for the year 2004,
(a) the CBR, (b) the GFR (c) the ASFRs (d) the TFR
(e) the GRR.

Q.6. In 1951, the total number of live births in W.B. was estimated [CU'06] as 399680. The table below records the number of females in the child bearing age intervals and the survival rates for WB and also the number of female live births in India during the year mentioned. Can you gather any idea about the GRR and NRR for WB from the given data?

| Age | Female Population in WB (100) | Survival rate (100000) | Total Number of female live-births in India (100) |
|---|---|---|---|
| 15-19 | 12652 | 59753 | 4632 |
| 20-24 | 11493 | 56924 | 14443 |
| 25-29 | 10001 | 54032 | 14058 |
| 30-34 | 8386 | 50252 | 8329 |
| 35-39 | 6897 | 45921 | 4036 |
| 40-44 | 5695 | 40856 | 2158 |
| 45-49 | 4728 | 36205 | 689 |

Q.7. [CU'1997] The following informations are obtained from a sample survey

| Age-group | Number of women | Proportion of Marriage | No. of births | Proportion surviving from birth to mid point of age group among married women |
|---|---|---|---|---|
| 15-19 | 16592 | 0.82 | 2692 | 0.902 |
| 20-24 | 19137 | 0.83 | 4272 | 0.891 |
| 25-29 | 10860 | 0.84 | 2179 | 0.878 |
| 30-34 | 4990 | 0.85 | 790 | 0.865 |
| 35-39 | 2468 | 0.74 | 203 | 0.849 |
| 40-44 | 928 | 0.69 | 47 | 0.830 |

(a) Calculate the GRR and NRR, assuming that there was no illegimate birth and sex ratio at birth is 1000:942 in favour of male.

(b) Determine the probability of a female children dying (after marriage at 32 age.

Q.8. The following table gives the female ASFR for 1993 and the life table for females of India with cohort $l_0 = 1,000$.

| Age of Years | ASFR | Female life-table stationary popln. |
|---|---|---|
| 15-19 | 0.0696 | 4180 |
| 20-24 | 0.2346 | 4123 |
| 25-29 | 0.1897 | 4063 |
| 30-34 | 0.1143 | 4001 |
| 35-39 | 0.0611 | 3934 |
| 40-44 | 0.0285 | 3860 |
| 45-49 | 0.0101 | 3763 |

Compute TFR, GRR and NRR for India assuming sex ratio at birth is 1.05:1.

Q.10. Consider a population on July 1, 1985 equal to 10,00,000 and growing at 2% per year as a continuous instantaneous rate, the crude rate of natural increase for 1995 was 3% and the crude rate was 1%. Determine the number of live births in 1995.

Q.11. [CU'07] For a stationary population with radix, $l_0 = 1,00,000$, out of the children born in 1980, number of deceased was 20,000 and the number of deceased in 1981 was 5000. Given the ASDRs of this population for the following ages (L.B.D)

| x | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $m_x$ | ·4073 | ·021 | ·0023 | ·0017 | ·0012 |

(a) Compute complete expectation of life at age 4, given the same at birth is 65.89 years.
(b) What is the chance that two new born babies will survive 4 years after their birth?

Q.9. [CU'03] The following table gives the ten decennial census popln. of two countries, say, A and B.

| Year : | 1901 | 1911 | 1921 | 1931 | 1941 | 1951 | 1961 | 1971 |
|---|---|---|---|---|---|---|---|---|
| Popln. of Country A (in million) : | 283·3 | 252·0 | 251·2 | 278·7 | 318·5 | 361·0 | 439·1 | 547·0 |
| Popln. of country B (in million) : | 5·3 | 7·2 | 9·6 | 12·9 | 17·1 | 23·2 | 31·4 | 38·5 |

| Year : | 1981 | 1991 |
|---|---|---|
| Popln. of C-A (in million) : | 653·8 | 823·8 |
| Popln. of C-B (in million) : | 50·2 | 62·9 |

Justify by some graphical procedure which of these countries population growth follows approximately Logistic model and comment on it. Fit a logistic model to the population that fits as well.

# PROBLEMS ON VITAL STATS.

1) (iii) To compare the mortality situation of country I and country II, we consider STDR based on age specific death rates,

$$STDR_I = \frac{\sum {}^{m}_{n}m_x \, {}^{m}_{n}p_x^s + \sum {}^{f}_{n}m_x \, {}^{f}_{n}p_x^s}{\sum {}^{m}_{n}p_x^s + \sum {}^{f}_{n}p_x^s}.$$

Here, we select ${}^{m}_{n}p_x^s = \frac{{}^{m}_{n}p_x^I + {}^{m}_{n}p_x^{II}}{2}$, ${}^{f}_{n}p_x^s = \frac{{}^{f}_{n}p_x^I + {}^{f}_{n}p_x^{II}}{2}$

2) $d_x = l_x - l_{x+1}$

$\therefore d_{20} = 6,93,435 - 6,90,673 = 2762$

$\therefore q_x = \frac{d_x}{l_x} \Rightarrow q_{20} = \frac{d_{20}}{l_{20}} = 3.98 \times 10^{-3}$

$\therefore p_x = 1 - q_x \Rightarrow p_{20} = 1 - q_{20} = 0.996$

$L_x = \frac{l_x + l_{x+1}}{2}$, assuming that deaths are uniformly distributed.

$\therefore L_{20} = \frac{1}{2}(l_{20} + l_{21}) = 692054.$

$e_x^\circ = \frac{T_x}{l_x} = \frac{T_{20}}{l_{20}} = e_{20}^\circ = 50.59$

$\therefore T_x = L_x + T_{x+1}$

$\Rightarrow T_{x+1} = T_x - L_x$

$\Rightarrow T_{21} = T_{20} - L_{20} = 3438907 2,$

$e_{21}^\circ = \frac{T_{21}}{l_{21}}$

3)

$e_{75}^{\circ} = 3.92$, $e_{76}^{\circ} = 3.66$

$d_{75} = 476$

$\Rightarrow l_{75} - l_{76} = 476$

$\therefore e_{75}^{\circ} \simeq e_{75} + \frac{1}{2}$

$\Rightarrow e_{75} = e_{75}^{\circ} - \frac{1}{2} = 3.92 - .5 = 3.42$

$\therefore e_{76} = 3.66 - .5 = 3.16$

$$\therefore \frac{e_{76}}{e_{75}} = \frac{\left(\sum_{t=1}^{\infty} l_{76+t}\right)/l_{76}}{\left(\sum_{t=1}^{\infty} l_{75+t}\right)/l_{75}}$$

$$= \frac{l_{75}}{l_{76}} \times \frac{\sum_{t=1}^{\infty} l_{76+t}}{l_{76} + \sum_{t=1}^{\infty} l_{76+t}}$$

$$= \frac{l_{75}}{l_{76}} \times \frac{1}{\frac{l_{76}}{\sum_{t=1}^{\infty} l_{76+t}} + 1}$$

$$= \frac{l_{75}}{l_{76}} \times \frac{1}{1 + \frac{1}{e_{76}}}$$

$$\Rightarrow \frac{l_{76}}{l_{75}} = \frac{e_{75}}{1 + e_{76}}$$

$$\Rightarrow l_{76} = l_{75} \times \left(\frac{e_{75}}{1 + e_{76}}\right)$$

$$= (476 + l_{76}) \times \frac{e_{75}}{1 + e_{76}}$$

$$\Rightarrow l_{76}\left(1 - \frac{e_{75}}{1 + e_{76}}\right) = 476 \times \frac{e_{75}}{1 + e_{76}}$$

$$\Rightarrow l_{76} = \frac{476 \times e_{75}}{1 + e_{76} - e_{75}} = 2200$$

$\therefore l_{75} = 476 + 2200 = 2676$.

4. Age-specific fertility rate of age group $[x, x+5)$ is

4)
$$ {}_5^1 i_x = \frac{{}_5^1 B_x}{{}_5^1 P_x} $$

$$ \therefore GRR = 5 \times \sum_x \frac{{}_5^1 B_x}{{}_5^1 P_x} = 5 \times \sum_x {}_5^1 i_x $$

$$ NRR = 5 \sum_x {}_5^1 i_x \times {}_5^1 R_x $$

where ${}_5^1 R_x$ is the survivors factor per year in the age-group $[x, x+5)$.

Here, $\frac{{}_5^1 B_x}{{}_5 B_x} = 48.7\% = \frac{48.7}{100} \; \forall \, x$.

| Age Group | No. of women ${}_5^1 P_x$ | No. of births ${}_5 B_x$ | ${}_5^1 B_x = \frac{48.7}{100} \times {}_5 B_x$ | ${}_5^1 i_x = \frac{{}_5^1 B_x}{{}_5^1 P_x}$ | Survival Rate ${}_5^1 R_x$ | ${}_5^1 i_x \times {}_5^1 R_x$ |
|---|---|---|---|---|---|---|
| 15-19 | 9000 | 140 | | | 0.920 | |
| | | | | .2153679 | | 0.19414 |

$$ GRR = 5 \times \sum_x {}_5^1 i_x = 5 \times .2153679 = 1.0768 $$

$$ \therefore NRR = 5 \times \sum_x {}_5^1 i_x \times {}_5^1 R_x = 5 \times .19414 = 0.9707 $$

"GRR = 1.0768" indicates the no. 1.0768 of daughters would be born on the average, to each of a group of females beginning life together, supposing none of them died before reaching the end of the reproductive period and all of them experience, throughout the reproductive period.

"NRR .9707" implies a group of 10000 females is expected to be replaced by 9707 females in the next generation under the given rates of fertility and mortality and the population will shows a tendency to decrease. Hence the population will ultimately decrease and will ultimately die out, unless the fertility and mortality change.

5) P = 293,857,000

$$\frac{m_B}{f_B} = \frac{105}{100}$$

(a) $CBR = \frac{B}{P} \times 1000$

$$= \frac{4112,000}{293,657,000} \simeq 14$$

(b) $GFR = \frac{B}{\sum_{\alpha} {}_5f_x} \times 1000$

(c) $ASFR = \frac{5B_x}{{}_5f_x} \times 1000 = 5i_x$

(d) $TFR = 5 \times \sum_{\alpha} 5i_x = 5 \times$ sum of ASFR's.

(e) $GRR = \left(TFR \times \frac{f_B}{B}\right)$ ; $GRR = \frac{100}{205} \times \frac{TFR}{1000}$.

8. c) $B = 399680$

To compute GRR & NRR of W.B., one should have information regarding ${}_5^fB_x$.

Assuming, $\frac{{}_5^fB_x}{{}^fB} = \frac{{}_5^fB_x^I}{{}^fB^I} \quad \forall x$

$$\Rightarrow {}_5^fB_x = \frac{{}_5^fB_x^I}{{}^fB^I} \times {}^fB$$

Assume, $\frac{{}^fB}{B} = \frac{100}{205} \Rightarrow {}^fB = \frac{100}{205} \times B$

$$= \frac{100}{205} \times 399680$$

$$= 194966$$

| Age | Female Popln. ${}_5^fP_x$ | Survival Rate (${}_5^fR_x$) | Total no. of female live births in india ${}_5^fB_x^I$ | ${}_5^fB_x$ | ${}_5^fi_x$ | ${}_5^fi_x \times {}_5^fR_x$ |
|---|---|---|---|---|---|---|
| 15-19 | 1265200 | .59753 | 463200 | 18680 | .01476 | .00882 |
| 20-24 | 1140300 | .56924 | 1449300 | 58246 | .05108 | .02908 |
| 25-29 | 1000100 | .54082 | 1405800 | 56693 | .05669 | .03063 |
| 30-34 | 886600 | .50352 | 832900 | 83589 | .04015 | .02022 |
| 35-39 | 684700 | .45921 | 403600 | 16276 | .02377 | .01091 |
| 40-44 | 569600 | .40356 | 215800 | 8703 | .01528 | .00617 |
| 45-49 | 472800 | .36205 | 68900 | 2779 | .00588 | .00213 |
| TOTAL | 5969800 | | 4834500 | 194966 | .20761 | .10797 |

$${}_5^fB_2 = \frac{{}_5^fB_x^I}{{}^fB^I} \times 194966, \qquad {}_5^fi_x = \frac{{}_5^fB_x}{{}_5^fP_x}$$

$$\therefore GRR = 5 \times \sum_x {}_5^fi_x$$

$$= 5 \times 0.20761 = 1.03805$$

$$\therefore NRR = 5 \times \sum_x {}_5^fi_x \times {}_5^fR_x$$

$$= 5 \times 0.10797$$

$$= 0.53985.$$

7) (a) there was no illegimate birth. We consider

${}_5P_x$ = No. of women in age group $[x, x+5)x$ proportion of marriage.

$$\frac{{}_5B}{mB} = \frac{942}{1000} \Rightarrow \frac{{}_5B}{B} = \frac{942}{1942}$$

$$\therefore GRR = 5 \times \frac{{}_5B}{B} \times \sum_x \frac{{}_5B_x}{{}_5P_x} = 5 \times \frac{{}_5B}{B} \times \sum_x {}_5i_x$$

| Age Group | No. of women | Proportion of marriage | ${}_5P_x$ | ${}_5B_x$ | ${}_5i_x$ | ${}_5R_x$ | ${}_5R_x \times {}_5i_x$ |
|---|---|---|---|---|---|---|---|
| 15-19 | 16592 | 0.82 | 13605 | 2642 | 0.19419 | 0.902 | 0.17516 |
| 20-24 | 19137 | 0.83 | 15884 | 4272 | 0.26895 | 0.891 | 0.28963 |
| 25-29 | 10860 | 0.84 | 9122 | 2179 | 0.23887 | 0.878 | 0.20973 |
| 30-34 | 4990 | 0.85 | 4241 | 790 | 0.18628 | 0.865 | 0.16113 |
| 35-39 | 2463 | 0.74 | 1823 | 203 | 0.11138 | 0.849 | 0.09456 |
| 40-44 | 925 | 0.64 | 592 | 47 | 0.07939 | 0.830 | 0.06589 |
| TOTAL | — | — | — | — | 1.07906 | — | 0.57328 |

$$\therefore GRR = 5 \times \frac{942}{1942} \times 1.07906 = 2.61708$$

$$\therefore NRR = 5 \times \frac{942}{1942} \times 0.57328 = 1.3904.$$

(b) Probability of the female children dying after marriage at 32 age = {1 - Probability of surviving at age 32} x {Prob. of married}

$$= (1 - 0.865) \times 0.85$$

$$= 0.11475$$

8) $NRR = \sum_\alpha {}_5i_x \times \frac{{}_5L_x}{{}_1l_0}$, where ${}_5L_x = {}_1L_x + {}_1L_{x+1} + \cdots + {}_1L_{x+4}$

| Age Group | ASFR ($5i_x$) | ${}_5L_x$ | ${}_5R_0 = \frac{{}_5L_x}{{}_1l_0[=1000]}$ | ${}_5i_x \times {}_5R_0$ |
|---|---|---|---|---|
| 15-19 | 0.0696 | 4180 | 4.180 | .29023 |
| 20-24 | 0.2346 | 4123 | 4.123 | .96726 |
| 25-29 | 0.1897 | 4063 | 4.063 | .77075 |
| 30-34 | 0.1143 | 4001 | 4.001 | .45731 |
| 35-39 | 0.0611 | 3934 | 3.934 | .24037 |
| 40-44 | 0.0289 | 3860 | 3.860 | .11001 |
| 45-49 | 0.0101 | 3763 | 3.763 | .03801 |
| TOTAL | 0.7079 | 27924 | 27.924 | 2.87394 |

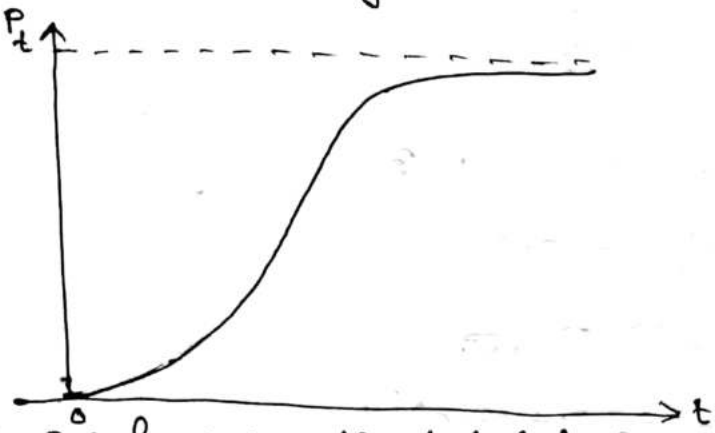
$$\therefore TFR = 5 \sum_x {}_5i_x = 5 \times 0.7079 = 3.5395$$

$$\therefore GRR = \frac{{}_5B}{B} \times TFR = \frac{100}{205} \times 3.5395 = 1.7266$$

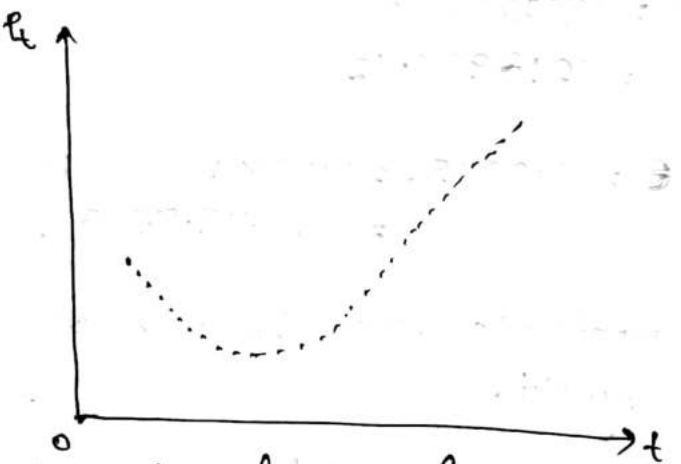
$$\therefore NRR = \frac{{}_5B}{B} \times \sum_x {}_5i_x \times {}_5R_0 = \frac{100}{205} \times 2.87394 = 1.40192$$

9)

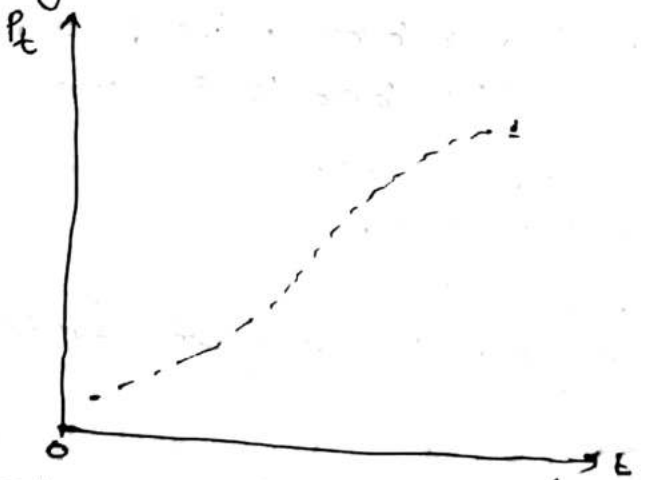
→ The shape of the Logistic curve is given by:




Plot $(t, P_t)$ for both the population on graph paper.




Clearly the population figures for the country I is not following the logistic law.




Clearly, II population figures are following logistic law approximately.

10) CBR − CDR = Crude rate of natural increase

$= \left(\frac{B}{P} - \frac{D}{P}\right) 100\%$.

$= 3\%$.

$\Rightarrow \frac{B}{P} - \frac{D}{P} = 0.03$ $\quad \left[\frac{D}{P} \times 100 = 1\right]$.

$\Rightarrow \frac{B}{P} = 0.04$

$P_0 = 10,000,00$

$P_{10} = P_0 (1+0.02)^{10}$

$= 10,000,000 (1+0.02)^{10}$

$= 1218994.42$

$\therefore B = 0.04 \times P_{10} = 0.04 \times 1218994.42$

$= 48759.7768$.

11) (a) To calculate the value $q_0$, we use the formula given by kuczynski.

$$\hat{q}_0 = 1 - \left(1 - \frac{D'''}{B_0}\right)\left(1 - \frac{D''}{B_{-1} - D'}\right)$$

$B_{-1}$ = No. of children born in the preceding calender year.

$B_0$ = No. of children born in the current calender year.

$D'$ = No. of children born and deceased in the preceding calender year.

$D''$ = No. of children born in the preceding calender year & deceased in the current calender year before reaching age 1 and

$D'''$ = No. of children born and deceased in the current calender year.

$B_{-1} = 1,00,000$

$B_0 = 1,00,000$

$D' = 20,000$

$D'' = 5,000$

$D''' = 25,000$

$\hat{q}_0 = 1 - \left(1 - \frac{20000}{100000}\right)\left(1 - \frac{5000}{80000}\right)$

$= 0.25$

$$m_x = \frac{d_x}{L_x}$$

$$\Rightarrow m_x = \frac{d_x}{l_x - (1-a_x)d_x}$$

$$= \frac{q_x}{1-(1-a_x)q_x}$$

$a_1 = 0.43$
$a_2 = 0.45$
$a_3 = 0.47$
$a_4 = 0.49$

(b) Required probability $= \left(\frac{l_4}{l_0}\right)^2$.

# Problems On Index Number — For Prac. Exam

▷ The following table shows the quantities consumed and the values (price × quantity) of 5 commodities for 3 successive years.

| Commodities | 1990 Quantity | 1990 Value | 1991 Quantity | 1991 Value | 1992 Quantity | 1992 Value |
|---|---|---|---|---|---|---|
| I | 50 | 350 | 60 | 420 | 70 | 490 |
| II | 120 | 600 | 140 | 700 | 160 | 800 |
| III | 30 | 330 | 20 | 200 | 15 | 225 |
| IV | 20 | 360 | 15 | 300 | 10 | 220 |
| V | 5 | 40 | 5 | 50 | 5 | 60 |

Calculate the price index number for 1992 taking 1990 as the base period adopting chain base formula and using Paasches formula at each stage. Also verify whether the circular test is satisfied by the Paasches formula or not on the basis of the above data. [C.U. 1996, 2010]

Soln → The chain index number for 1992 w.r.t. base 1990 is

$$P'_{90,92} = P_{90,91} \times P_{91,92}$$

$$= \frac{\Sigma p_{91} q_{91}}{\Sigma p_{90} q_{91}} \times \frac{\Sigma p_{92} q_{92}}{\Sigma p_{91} q_{92}} = \text{, by Paasches formula.}$$

Note that, $\text{Price}(p) = \frac{\text{Value}}{\text{Quantity}}$.

For circular test, $P_{90,91} \cdot P_{91,92} \cdot P_{92,90} = 1$.

$$\Leftrightarrow P_{90,91} \cdot P_{91,92} = P_{90,92}$$

$$\Leftrightarrow P'_{90,92} = P_{90,92}$$

i.e. chain index = Fixed base index.

Hence, $P_{90,92} = \frac{\Sigma p_{92} \cdot q_{92}}{\Sigma p_{90} q_{92}} =$

Hence, $P'_{90,92} \neq P_{90,92}$

⇒ The circular test is not satisfied by the Paasches formula.

2. The following table shows the average per capita consumption of cereals and prices of cereals in rural India for four different periods $T_1, T_2, T_3, T_4$.

| Commodities | Consumptions (in kg) per month | | | | Price (Rs. per kg) | | | |
|---|---|---|---|---|---|---|---|---|
| | $T_1$ | $T_2$ | $T_3$ | $T_4$ | $T_1$ | $T_2$ | $T_3$ | $T_4$ |
| Rice | 4.69 | 3.94 | 3.58 | 4.28 | 5.37 | 5.75 | 6.15 | 6.30 |
| Wheat | 5.51 | 6.93 | 6.43 | 6.78 | 4.94 | 5.15 | 5.02 | 5.10 |
| Others | 7.66 | 8.19 | 7.75 | 7.71 | 3.32 | 2.94 | 3.25 | 3.52 |

Calculate the price index number of cereals for period $T_4$ taking $T_1$ as the base adopting chain-base formula and using any uniformly suitable formula at each stage. If after calculations it is found that consumption figures in $T_4$ are in error by 5%, do you think that your index is going to be affected by this? Give reason for your answer. [C.U. 1992]

Soln. → If we use Laspeyre's formula, then

$$P_{14}' = P_{12} \cdot P_{23} \cdot P_{34}$$

$$= \frac{\Sigma p_2 q_1}{\Sigma p_1 q_1} \times \frac{\Sigma p_3 q_2}{\Sigma p_2 q_2} \times \frac{\Sigma p_4 q_3}{\Sigma p_3 q_3} = \text{———}$$

is independent of $q_4$ (the consumption figure for $T_4$). Hence, it will not affect the calculation of chain index.

If we use Paasche's formula, then

$$P_{14}' = P_{12} \cdot P_{23} \cdot P_{34}$$

$$= \frac{\Sigma p_2 q_2}{\Sigma p_1 q_2} \times \frac{\Sigma p_3 q_3}{\Sigma p_2 q_3} \times \frac{\Sigma p_4 q_4}{\Sigma p_3 q_4} = \text{- - -}$$

which depends on $q_4$.

Hence, all the consumption figures in $T_4$ are in error by 5%, then the corrected figures are

$$q_{4i}' = q_{4i} \times \frac{105}{100} \text{ or } q_{4i} \times \frac{95}{100}$$

and hence, $$\frac{\Sigma p_{4i} q_{4i}'}{\Sigma p_{3i} q_{4i}'} = \frac{\Sigma p_{4i} \frac{105}{100} q_{4i}}{\Sigma p_{3i} \frac{105}{100} q_{4i}} = \frac{\Sigma p_{4i} q_{4i}}{\Sigma p_{3i} q_{4i}}$$

Ultimately, it will not affect the calculation of the index.

3. The data below show the percentage increases in price of a few tabulated food items and weights attached to each of them. Calculate the index number for the food group.

| Food items : | Rice | Wheat | Dal | Ghee | Oil | Spices | Milk | Fish | Others |
|---|---|---|---|---|---|---|---|---|---|
| Weights : | 30 | 11 | 8 | 5 | 5 | 3 | 7 | 9 | 19 |
| Percentage increase in price : | 180 | 202 | 115 | 212 | 175 | 517 | 260 | 426 | 332 |

Using the above Food Index and information given below. Calculate CLI.

| Group : | Food | Clothing | Fuel and Light | Rents and Rates | Miscellaneous |
|---|---|---|---|---|---|
| Index : | — | 310 | 220 | 150 | 300 |
| Weight : | 60 | 5 | 8 | 9 | 18 |

[C.U. 2000]

Soln. :→ The percentage increase in price for the ith item

$$(\gamma) = \frac{p_{ii} - p_{oi}}{p_{oi}} \times 100$$

$$\Rightarrow \frac{p_{ii}}{p_{oi}} \times 100 = (100 + \gamma)$$

The food index is computed by

$$I_F = \frac{\sum \frac{p_{ii}}{p_{oi}} \times w_i}{\sum w_i} \times 100 = \frac{\sum \left(\frac{p_{ii}}{p_{oi}} \times 100\right) \times w_i}{\sum w_i}$$

| food item | weights ($w_i$) | $\frac{p_{ii}}{p_{oi}} \times 100$ | $\left(\frac{p_{ii}}{p_{oi}} \times 100\right) w_i$ |
|---|---|---|---|
| | | | |
| | $\sum w_i =$ | | $\sum \left(\frac{p_{ii}}{p_{oi}} \times 100\right) w_i =$ |

Hence, $I_F = \frac{\text{Column 4 total}}{\text{Column 2 total}} = \dots$

The CLI $= \frac{\sum I_i w_i}{\sum w_i}$

4. The following table shows the group indices and the corresponding weights for the year 1995 with 1981 as the base year of a given community.

| Group | Group index | Weight |
|---|---|---|
| Food | 212.45 | 65.3 |
| Clothing | 328.06 | 4.8 |
| Fuel and light | 345.89 | 8.5 |
| House rent | 173.841 | 7.6 |
| Miscellaneous | 201.35 | 13.8 |

(a) Find the CLI for the year 1995.
(b) What is the purchasing power in 1995 as compared to 1981.
(c) If Mr. Dasgupta's salary increased from 2400 Rs in 1981 to Rs. 4950 in 1995, how has his economic status change?
(d) The weights are proportional to the consumption expenditure of each group. Suppose Mr. Dasgupta has to maintain the same status for each of the 1st four groups and can only adjust his spending on Miscellaneous items to come to terms with his income changes.
Find his spendings for each of the groups in 1995.
(e) What could be Mr. Dasgupta's weights for each of the groups in 1995? [C.U. 2001, 2009]

Soln. :→

(a) $C.L.I = \frac{\sum I_i w_i}{\sum w_i} = 224.842$

(b) The purchasing power of money in 1995 w.r.t. 1981 is $\frac{1}{(C.L.I/100)} = \frac{100}{CLI} = 0.44956$.

(c) CLI for 1995 w.r.t. 1981 as base year is 224.842 means that if some individual spends Rs. 100 in 1981, in 1995 he/she has to spend Rs. 224.842 to maintain the same standard of living as she/he has in 1981. For Mr. Dasgupta's expenditure on 1981 was Rs. 2400. In 1995 she has to spend Rs. $2400 \times 2.24842 = 5396.2$ to maintain the same standard of living as he has in 1981. But his salary in 1995 was 4950. She has to adjust Rs. $(5396.2 - 4950) = 446.2$ Rs. i.e. his economic status falls in 1995 as compared to 1981.

(d)

| Group | Index | Weights | In 1981 Expenditure | In 1995 to maintain the same standard of living |
|---|---|---|---|---|
| Food | 212.45 | 65.3 | 2400 × 65.3/100 = 1567.8 | 1567.8 × 212.45/100 = 3329.51 |
| Clothing | 328.06 | 4.8 | 115.2 | 377.925 |
| Fuel & light | 345.89 | 8.5 | 204 | 705.615 |
| House rent | 178.41 | 7.6 | 182.4 | 316.299 |
| Miscellaneous | 201.35 | 13.8 | 326.4 | 666.206 |
| TOTAL = | | 100 | 2400 | 5396.52 |

In 1981, expenditure on a group = Income × $\frac{w_i}{\Sigma w_i}$

Hence his spendings on the first four groups remain same and the deficit of Rs. 446.27 will be adjusted on Miscellaneous group and the spending on Miscellaneous is Rs. (666.86 − 446.27)

= 220.59 Rs. in 1995.

(e) In 1995, the weight of a group = $\frac{\text{Expenditure on the Group}}{\text{Salary (4950)}} \times 100$

| Group | Expenditure | Weights |
|---|---|---|
| Food | 3329.51 | 3329.51 × 100/4950 = 67.2 |
| Clothing | 377.925 | 377.925 × 100/4950 = ... |
| ⋮ | ⋮ | ⋮ |
| Miscellaneous | 220.59 | 220.59 × 100/4950 = ... |
| TOTAL = | 4950 | 100 |

— x —

Q. 5. The following data relate to the urban middle-class people of a particular region in the years 2003 and 2005.

| Group | % of total expenditure | Group Index (Base: 2000) 2003 | 2005 |
|---|---|---|---|
| Food | 35 | 118 | 122 |
| Clothing | 15 | 112 | 118 |
| Housing | 20 | 113 | 115 |
| Transport | 10 | 112 | 117 |
| Durable Goods | 8 | 105 | 110 |
| Others | 12 | 120 | 125 |

(a) Compute the CLI for 2003 and 2005 with Base year 2000.
(b) If a family saved 15% of its monthly income in 2003, find the relative change in its average savings over the years 2003-2005, assuming that it maintained the same standard of living as in 2003. (C.U. - 2008)

Soln.

(a) $$CLI_{2003} = \frac{\sum I_{1i} w_i}{\sum w_i} = \frac{11470}{100} = 114.70$$

$$CLI_{2005} = \frac{\sum I_{2i} w_i}{\sum w_i} = \frac{11890}{100} = 118.90$$

∵ Salary is constant, i.e. Rs. 100,

So, weights of the group = Expenditure of the group.

(b) Let Rs. 100 be the salary in 2003. Saving, Rs. = 15.

Expenditure = Rs. 85

| C.I. | 2003 | 2005 |
|---|---|---|
| | 114.70 | 118.90 |

If some individual spent Rs. 114.70 in 2003, in 2005 he has spent Rs. $\frac{118.90}{114.70} \times 85$

$= 88.11$ Rs.

In 2005, his salary is 100 Rs.

Exp. = 88.11 Rs.

Saving = 11.89 Rs. ∴ Saving % = $\frac{11.89}{100} \times 100$

$= 11.89\%$

———×———

6. The following table gives (with two missing values) the overall and group-wise CLI (with 2000 as the base year) with six different expenditure groups and their respective weights, for the urban middle class people of a particular city, in 2004 and 2005.

| Group | Weight | Group Index 2004 | Group Index 2005 |
|---|---|---|---|
| Food | 350 | 117 | 120 |
| Clothing | 156 | 113 | 118 |
| Housing | 187 | 118 | – |
| Transport | 108 | 112 | 117 |
| Durable Goods | 76 | 102 | 111 |
| Others | 123 | 121 | 125 |
| CLI | TOTAL = 1000 | = 115.4 | = 119.5 |

(a) If Mr. X saved 20% of his salary in 2004, Determine the relative change in his average savings, relative to 2004, in 2005 if his salary increased by 10% and he maintained the same standard of living as in 2004. [C.U. 2007]

Soln. (a) Let Rs. 100 be the salary in 2004. Saving = Rs. 20

Exp. = Rs. 80,

| CLI. | 2004 | 2005 |
|---|---|---|
| | 115.4 | 119.5 |

If X spends Rs. 115.4 in 2004, in 2005, he has spent.

Rs. $\frac{119.5}{115.4} \times 80 = 82.84$ Rs.

In 2005, his salary is 110.

Exp. = Rs. 82.84.

Saving = Rs. 27.16

Saving % = $\frac{27.16}{110} \times 100 = 24.69\%$

— x —

# PRACTICALS ON PROBABILITY DISTRIBUTION

1. (a) Given that $\sum_{1}^{10} f(x) = 500420$, $\sum_{4}^{10} f(x) = 329240$, $\sum_{7}^{10} f(x) = 195212$. $f(10) = 40365$. Find $f(1)$.

(b)

1. (a) The no. of females in each of 100 queues of length 10 at a metro railway station in Kolkata. The data are shown below:

| Count | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Freq | 1 | 3 | 4 | 23 | 25 | 19 | 18 | 5 | 1 | 1 | 0 |

Propose an appropriate theoretical distn and fit it to the above data. Also comment on the fitting.

Solution:- Let X : 'No of females in a queue of length 10

Consider "getting a female in queue" as a success.

Assuming probability of successes in each queue is $p$ (constant).

Then, Binomial model is appropriate to the given RV X.

Then PMF of X is

$$p(x) = \begin{cases} \binom{10}{x} p^x (1-p)^{10-x} & ; x = 0(1)10 \\ 0 & ; \text{ow} \end{cases}$$

By method of moments, $\mu_1' = m_1'$.

$$\Rightarrow 10p = \bar{x} = \sum_{i=1}^{10} x_i f_i \Big/ \sum_{i=1}^{10} f_i = 435/100 = 4.35.$$

$$\Rightarrow \hat{p} = \frac{\bar{x}}{10} = 0.435.$$

Hence the fitted distn is given by:

$$p(x) = \begin{cases} \binom{10}{x} (\hat{p})^x (1-\hat{p})^{10-x} & ; x = 0(1)10 \\ 0 & ; \text{ow} \end{cases}$$

Comment on Fitting :- Expected frequency of the value 'X' is

$$N \times P[X = x]$$
$$= 100 \times p(x).$$

[ Consider getting a value 'x' as a success. Then

$f_x$ = the frequency of 'x' in N observation
= the no. of successes in N Bernoulli trials
$\sim$ Bin $(N, P[X=x])$
$\therefore E(f_x) = N \times P[X=x]$ ]

| x | p(x) | Exp. freq = N.p(x) | Observed freq. |
|---|---|---|---|
| | | | |

(b) When the first proof of 200 pages of an encyclopaedia of 500 pages were read, the distn. of printing mistakes were found to be as shown in the table:

| No. of misprints on Page: | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Frequency : | 112 | 63 | 20 | 3 | 1 | 1 |

Fit a suitable probability distribution to the data.

Estabish the total cost of correcting the first proof of the whole encyclopaedia by using the information given below:

| No. of misprints on page: | 0 | 1 | 2 | 3 | 4 | 5 or more |
|---|---|---|---|---|---|---|
| Cost of detection and correction (dollars) per page: | 0.10 | 0.18 | 0.23 | 0.29 | 0.34 | 0.36 |

**Solution:-** Let $X$ denotes the number of misprints on a page. Getting a misprint on a page is a rare event. Hence $X$ is expected to follow Poisson distribution.

Assume that $X \sim P(\lambda)$.

The PMF of $X$ is

$$p(x) = \begin{cases} e^{-\lambda} \cdot \frac{\lambda^x}{x!} & , x = 0, 1, 2, 3, \ldots \\ 0 & , \text{ow} \end{cases}$$

where $\lambda > 0$.

By method of moments, $\mu_1' = m_1'$

$$\Rightarrow \hat{\lambda} = \bar{x} = \sum_{i=0}^{5} x_i f_i / \Sigma f_i = \text{_____}$$

Hence the fitted distn. is

$$p(x) = \begin{cases} e^{-\hat{\lambda}} \cdot \frac{(\hat{\lambda})^x}{x!} & , x = 0, 1, 2, \ldots \\ 0 & , \text{ow} \end{cases}$$

The expected frequency of the value '$x$' is

$$N \cdot P[X = x] = 5000 \times p(x)$$

Table showing expected freq and cost of correction:-

| $x$ | $p(x)$ | Exp. freq | Cost per page | Cost of correction = Exp. freq × Cost per page |
|---|---|---|---|---|
| 0 | | | 0.10 | |
| 1 | | | 0.16 | |
| 2 | | | 0.23 | |
| 3 | | | 0.29 | |
| 4 | | | 0.34 | |
| $\geq 5$ | | | 136 | |
| | 1 | N | | |

2. (a) A farmer sells bean seeds in packets of 100 and agrees to refund the price if the number of seeds germinating from a packet is less than 75. He knows from past experience that on an average 80% of the seeds germinate and it costs him Rs. 200 for a packet of seeds. How should he fix the price of a packet so as to ensure an average profit of 25% ?

Solution:- $X$ : The no. of seeds germinated in packet.

Here, $X \sim \text{Bin}(100, p = 0.8)$.

Let $x$ be the selling price.

Define, $Z = \begin{cases} (x-2) & \text{if } x \geq 75 \\ -2 & \text{if } x < 75 \end{cases}$

Now, $E(Z) = \frac{x}{4} = (x-2)\,P[X \geq 75] + (-2)\,P[X < 75]$

$= (x-2)\left[1 - P[X < 75]\right] - 2P[X < 75]$

By CLT, $\frac{x - 80}{\sqrt{16}} \overset{a}{\sim} N(0,1)$

$P[X \geq 75] = 1 - \Phi\left(\frac{5}{4}\right)$

$\Rightarrow P[X < 75] = \Phi\left(\frac{5}{4}\right)$

(b) Let $X \sim \text{Bin}(2, p)$, $Y \sim \text{Pois}(\lambda = 1)$ and $X$ and $Y$ are independently distributed. It is known that $P[Y > X] = 1 - 2e^{-1}$. What is the probability that $X$ is strictly positive?

Solution:-

$$1 - 2e^{-1} = P[Y > X] = 1 - P[Y \leq X]$$

$$= 1 - \sum_{x=0}^{2} P[Y \leq x, X = x]$$

$$= 1 - \{P[Y \leq 0]P[X = 0] + P[Y \leq 1]P[X = 1] + P[Y \leq 2]P[X = 2]\}$$

$$= 1 - \left[e^{-1}\binom{2}{0}p^0(1-p)^0 + \left\{e^{-1} + \frac{e^{-1}}{1!}\right\}\binom{2}{1}p^0(1-p)^2 + \left\{e^{-1} + \frac{e^{-1}}{1!} + \frac{e^{-1}}{2!}\right\}\binom{2}{2}p^2\right]$$

$$\Rightarrow 1 + 2p - \frac{p^2}{2} = 2$$

$$\Rightarrow p^2 - 4p + 2 = 0$$

$$\Rightarrow p = \frac{4 \pm \sqrt{4^2 - 4.1.2}}{2}$$

$$\Rightarrow p = 2 \pm \sqrt{2}.$$

* $P[Y \leq 0] = P[Y = 0]$

$P[Y \leq 1] = P[Y = 0] + P[Y = 1]$.

Now, required probability is $= P[X>0]$

$= 1 - P[X=0]$

$= 1 - (1-p)^2$

$= [1 - (\sqrt{2}-1)^2]$

$= 2(\sqrt{2}-1)$.

(c) Suppose that the number of accidents per week at an industrial plant is a poisson random variable with mean four. Suppose also that the number of workers injured in different accidents are independent Poisson RVs with a common mean 2. Assume that the number of workers injured in each accident is independent of the number of accidents that occur. What are the mean and variance of the number of injured during a week?

Solution:- Let N denotes the no. of accidents per week and X denotes the no. of workers injured in the $i^{th}$ accident.

Here $N \sim P(\lambda=4)$ and $X \sim P(\lambda=2)$, independently.

Hence, the number of injured during a week is

$$S_N = X_1 + X_2 + \cdots + X_N$$

$$E(S_N) = E(N)\,E(X_1) = 2 \times 4 = 8.$$

(d) A sample of size 10 is drawn from a normal population with mean and variance both equal to $\theta\ (>0)$ and the first quartile equal to 2.65. Find the probability that

(i) the first four observations are negative

(ii) four of the observations are negative.

Solution:- Let $X_1, X_2, \ldots, X_n$ be a r.v. from $N(0,\theta)$; $\theta>0$.

By definition of 1st quartile,

$$P[X_1 \le Q_1] = 0.25$$

$$\Rightarrow P\left[\frac{X_1-\theta}{\sqrt{\theta}} \le \frac{Q-\theta}{\sqrt{\theta}}\right] = 0.25$$

$$\Rightarrow \Phi\left(\frac{Q-\theta}{\sqrt{\theta}}\right) = 0.25$$

$$\Rightarrow \Phi\left(-\frac{Q_1-\theta}{\sqrt{\theta}}\right) = 0.75$$

From Biometrica,

| $x$ | $\Phi(x)$ |
|---|---|
| 0.67 | 0.74857 |
| 0.68 | 0.75175 |

$$\therefore \quad -\frac{Q_1-\theta}{\sqrt{\theta}} = 0.6745$$

$$\Rightarrow \quad \frac{2.65-\theta}{\sqrt{\theta}} = 0.6745$$

$$\Rightarrow \quad \sqrt{\theta} = \frac{0.6745 + \sqrt{(0.6745)^2 + 4 \times 2.65}}{2}$$

as $\sqrt{\theta} > 0$.

$\therefore \sqrt{\theta} = 1.99 \simeq 2$.

$\therefore \theta = 4$.

Hence, $X \sim N(4, 2^2)$.

(i) The probability that first four observations are negative,

$$P\left[X_1 < 0,\ X_2 < 0,\ X_3 < 0,\ X_4 < 0\right]$$

$$= \left\{P\left[X_1 < 0\right]\right\}^4 \quad , \text{as } X_i\text{'s are i.i.d.}$$

$$= \left\{\Phi\left[\frac{0-\theta}{\sqrt{\theta}}\right]\right\}^4$$

$$= \left\{\Phi\left(-\sqrt{\theta}\right)\right\}^4$$

$$= \Phi^4(-2) = \text{\_\_\_\_\_}.$$

(ii) Let $Y$ denotes the no. of negative values in $X_1, X_2, \ldots, X_{10}$

Then $Y \sim \text{Bin}(10, p)$, where, $p = P\left[X_1 < 0\right] = \Phi(-2)$

$$= \text{\_\_\_\_\_}.$$

$\therefore$ Required Probability $= P[Y=4]$

$$= \binom{10}{4} p^4 (1-p)^6 = \text{\_\_\_\_\_}.$$

3. (i) Twenty five leaves were selected at random from each of six similar apple trees. The number of adult female European red mites of each leaf was counted, the resulting information is summarized in the table below:

| No. of Mites per leaf : | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Frequency : | 70 | 38 | 17 | 10 | 9 | 3 | 2 | 1 |

(a) Fit a negative binomial distribution to the data.
(b) Plot the observed and fitted values on the same graph paper, and comment on the goodness of fit after visual inspection.

Solution:- (i) Let $X$ denotes the no. of adult female European red mites on a leaf.

(a) Assume that $X \sim NB(r, p)$

The PMF of $X$ is $f(x) = \begin{cases} \binom{x+r-1}{x} p^r q^x ; & x = 0, 1, 2, \ldots \\ 0 ; & \text{OW} \end{cases}$

where $0 < p < 1$ and $p + q = 1$ and $r > 0$.

By method of moments,

$$\mu_1' = \bar{x}, \mu_2 = s^2$$

$$\Rightarrow \frac{rq}{p} = \bar{x}, \frac{rq}{p^2} = s^2$$

$$\Rightarrow \hat{p} = \frac{\bar{x}}{s^2} \text{ and } \hat{r} = \bar{x} \cdot \frac{\hat{p}}{1-\hat{p}}.$$

From data, $\bar{x} = \frac{\sum_{i=0}^{7} i f_i}{\sum f_i} = \_\_\_\_\_.$

$$s^2 = \left(\sum_{i=0}^{7} i^2 f_i / \sum f_i\right) - \bar{x}^2 = \_\_\_\_\_.$$

$\hat{p} = \_\_\_\_\_$

$\hat{r} = \_\_\_\_\_.$

The fitted NB distribution is

$$p(x) = \begin{cases} \binom{x+\hat{r}-1}{x} (\hat{p})^r (\hat{q})^x , & x = 0, 1, 2, \ldots \\ 0 , & \text{OW} \end{cases}$$

(b)

| $x$ | $p(x)$ | Exp. freq. $N \times p(x)$ | Obs. freq. |
|---|---|---|---|
| 0 | | | |
| 1 | | | |
| 2 | | | |
| 3 | | | |
| 4 | | | |
| 5 | | | |
| 6 | | | |
| $\geq 7$ | (by subtraction) | | |

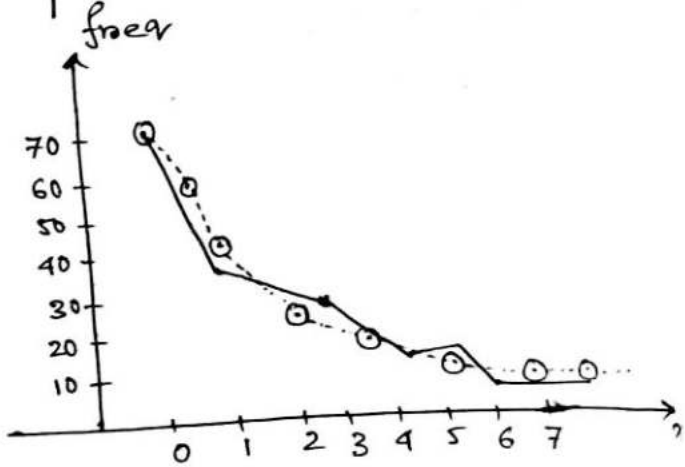
Here $p(x+1) = \left(\frac{x+r}{x+1}\right)\hat{q} \cdot p(x)$.

$p(1) = r \cdot \hat{q} \cdot p(0)$

$\therefore p(0) = \hat{p}^r = $ _____

$p(1) = \left(\frac{r}{1} \cdot \hat{q}\right) p(0) = $ _____

$p(2) = \left(\frac{r+1}{2} \hat{q}\right) p(1) = $ _____




If the both frequency graphs are very near to each other throughout range, then the fitting is satisfactory.

Remark:- It is observed that the red mites on a leaf form a cluster, therefore it seems like that NB model is appropriate for X.

Alt:- For a poisson distribution, we should have $\bar{x} \simeq s^2$.
For a negative binomial distn., we should have $\bar{x} << s^2$.

(ii) The following table gives the freq. distn of the number of albino children in families of five children including at least one albino child (Pearson's data)

| No. of albino in family | 1 | 2 | 3 | 4 | 5 | TOTAL |
|---|---|---|---|---|---|---|
| No. of families | 25 | 23 | 10 | 1 | 1 | 60 |

Fit an appropriate probability distribution to the data. Also, compare the observed and expected frequencies.

Solution:- Let X denotes the no. of albino children in families of five children. Then, considering "getting an albino children" as a success,

$$X \sim \text{Bin}(n=5, p).$$

The PMF of X is $p(x) = \begin{cases} \binom{5}{x} p^x q^{5-x}, & x = 0(1)5 \\ 0, & \text{OW} \end{cases}$

Let Y denotes the no. of albino children in a family of five having at least one albino. The data is given on the R.V. Y.

Note that Y is the RV "X is truncated at $x=0$".

The PDF of Y is $p(y) = \begin{cases} \dfrac{\binom{5}{y} p^y (1-p)^{5-y}}{1-(1-p)^5}, & y = 1(1)5 \\ 0, & \text{OW} \end{cases}$

By method of moments, $E(Y) = \bar{y}$,

$$\Rightarrow \frac{5p}{1-(1-p)^5} = \bar{y} \qquad \text{---------- (*)}$$

$$\bar{y} = \sum_{i=1}^{5} i f_i \Big/ \sum_{i=1}^{5} f_i = 1.833.$$

Solving (*) by Iteration method:-

$$p = \frac{\bar{y}}{5}\left\{1-(1-p)^5\right\} = \phi(p), \text{ say}$$

Trial root:- $p_0 = \dfrac{\bar{y}}{5} =$ _____

Condition of Convergence:- $|\phi'(p_0)| =$ _____ $< 1$.

Successive approximation, $p_1 = \phi(p_0)$

$p_2 = \phi(p_1)$

$\vdots$

$\hat{p} = \text{——}$

$\therefore$ The fitted distribution is $p'(y) = \begin{cases} \dfrac{\binom{5}{y}\hat{p}^y(1-\hat{p})^{5-y}}{1-(1-\hat{p})^5}, & y = 1(1)5 \\ 0, & \text{ow} \end{cases}$

(iii) The table gives the frequency distribution of the number of dust nuclei in a small volume of air that fell onto a stage in a chamber containing moisture and filtered air. It is suspected that a number of zero counts were wrongly rejected on the ground that the apparatus was not working.

| No. of dust nuclei | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Frequency | 23 | 56 | 88 | 95 | 73 | 40 | 17 | 5 | 3 | 400 |

Fit an appropriate distribution to the frequency distribution, omitting the zero counts.

Solution:- Let X denotes the number of dust nuclei in a small volume of air.

Here, the poisson distribution is appropriate for X, let,

$X \sim P(\lambda)$.

Let, Y be the RV of X truncated at $x=0$.

Hence, the PMF of Y is,

$$p(y) = \begin{cases} \dfrac{e^{-\lambda}\cdot\frac{\lambda^y}{y!}}{1-e^{-\lambda}}, & \text{if } y = 1, 2, 3, \ldots \\ 0, & \text{ow} \end{cases}$$

By method of moments,

$$E(Y) = \bar{y}$$

$$\Rightarrow \frac{\lambda}{1-e^{-\lambda}} = \bar{y} = \frac{\sum_{i=1}^{8} i \cdot f_i}{\sum_{i=1}^{8} f_i} = \text{_____}.$$

$$\Rightarrow \lambda = \bar{y}(1-e^{-\lambda}) = \phi(\lambda).$$

Trial root :- $\lambda_0 = \bar{y} = \text{_____}$.

Condition of Convergence :-

$|\phi'(\lambda_0)| = \text{_____} < 1.$

Successive improvements are :

$\lambda_1 = \phi(\lambda_0)$

$\lambda_2 = \phi(\lambda_1)$

$\vdots$

$\hat{\lambda} = \text{_____}$, correct to two decimal places.

Hence, fitted PMF of Y is,

$$p(y) = \begin{cases} e^{-\hat{\lambda}} \cdot \frac{(\hat{\lambda})^y}{y!}, & y = 1, 2, 3, \ldots \\ 0, & \text{ow} \end{cases}$$

4. The life time (T) in hours of an electron tube manufactured in a company is a r.v. with the d.f.

$$F(t) = 1 - e^{-t/\theta}, \quad t > 0.$$

A sample of 337 electron tubes give the following frequency distribution of T.

| Life (in hours) : | 0-50 | 50-100 | 100-150 | 150-200 | 200-300 | 300-400 | 400-6 |
|---|---|---|---|---|---|---|---|
| Frequency : | 100 | 68 | 48 | 31 | 12 | 21 | 27 |

(i) Estimate $\theta$ from the data and compare the observed and expected frequencies.

(ii) Estimate the probability that a supply of 20 tubes will not last more than 1900 hours, if they used one at a time successively.

(iii) Find the number of electronic tubes likely to burn away within 80 hours of their life and also the average life of these tubes.

Solution:-

The DF of T is

$$F(t) = 1 - e^{-t/\theta}, \quad t > 0$$

PDF of T is

$$f(t) = \frac{1}{\theta} e^{-t/\theta}, \text{ if } t > 0$$

(i) By method of moments,

$$\mu_1'(T) = \bar{t}$$

$$\Rightarrow \theta = \bar{t} = \frac{\sum_{i=1}^{7} t_i \cdot f_i}{\sum_{i=1}^{7} f_i} = \hat{\theta}$$

where, $t_i, f_i$ are mid point and frequency of the $i^{th}$ class.

The fitted distribution is given by,

$$F(t) = 1 - e^{-t/\hat{\theta}}, \quad t > 0.$$

∴ Expected frequency of the class interval $(a, b)$ is

$$N.P[a < T < b] = N\{F(b) - F(a)\}$$

$$= N\{e^{-a/\hat{\theta}} - e^{-b/\hat{\theta}}\}$$

## Computation of Expected Frequencies :-

| Class interval $(a,b)$ | $P[a < T < b]$ | Exp. Freq. $N.P[a<T<b]$ | Observed Freq. |
|---|---|---|---|
| 0-50 | | | |
| 50-100 | | | |
| 100-150 | | | |
| ⋮ | | | |
| . | | | |

(ii) Let $T_i$ denotes the life-time of the $i^{th}$ tube, $i = 1(1)20$.

Here, $T_i \overset{iid}{\sim} Exp(\theta)$, $i = 1(1)20$.

Required Probability $= P\left[\sum_{i=1}^{20} T_i < 1900\right]$

$= P[T < 1900]$, $T \sim Gamma(20, \theta)$.

$$= \int_0^{1900} \frac{e^{-t/\theta} \cdot t^{20-1}}{\theta^{20} \cdot \lceil 20} \, dt$$

$$= \int_0^{1900/\theta} \frac{e^{-x} \cdot x^{20-1}}{\lceil 20} \, dx \quad \left[\because \frac{2T_i}{\theta} \overset{iid}{\sim} \chi^2_2, \Rightarrow \frac{2T}{\theta} = \sum_{i=1}^{20} \frac{2T_i}{\theta} \Rightarrow \frac{2T}{\theta} \sim \chi^2_{40}\right]$$

$$= \lceil_{\frac{1900}{\theta}}(20)$$

So, estimated probability $= \lceil_{\frac{1900}{\hat{\theta}}}(20)$.

[ Use Pearson Table for Incomplete Gamma]

(iii) The no. of tubes which burns away within 80 hours.

$= N.P[T < 80]$

$= 337. \left\{1 - e^{-\frac{80}{\theta}}\right\}$

$= \_\_\_\_\_.$

The distribution of lifetime of a tube which burns away with 80 hours is

$$f'(t) = \begin{cases} \frac{\frac{1}{\theta} e^{-t/\theta}}{1 - e^{-80/\theta}} & , 0 < t < 80 \\ 0 & , \text{ow} \end{cases}$$

$$E(T') = \int_0^{80} \frac{t \cdot \frac{1}{\theta} e^{-t/\theta}}{1 - e^{-80/\theta}} dt = \theta - 80 \left(\frac{e^{-80/\theta}}{1 - e^{-80/\theta}}\right).$$

$$\hat{E}(T') = \hat{\theta} - 80. \frac{e^{-80/\hat{\theta}}}{1 - e^{-80/\theta}}.$$

5. (a) The following table gives the frequency distribution of the length of the left middle finger of 3000 criminals, obtained in the course of a study in clinical anthropometry.

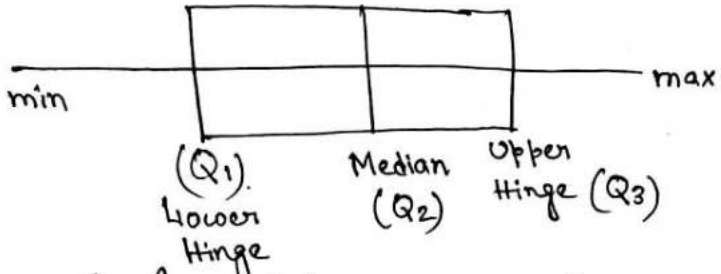
| Length (cm): (Mid-point of class interval) | 9.5 | 9.8 | 10.1 | 10.4 | 10.7 | 10.0 | 11.3 | 11.6 | 11.9 | 12.2 | 12.5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Frequency: | 1 | 4 | 24 | 67 | 198 | 417 | 575 | 691 | 509 | 306 | 131 |

| Length (cm): | 12.8 | 13.1 | 13.4 |
|---|---|---|---|
| Frequency: | 63 | 18 | 3 |

(i) Display the data graphically be means of box-plot.
(ii) Fit an appropriate probability distribution to the data.
(iii) Plot observed and estimated frequencies on a graph paper and comment on the fit.

**Solution:-**

(i)




(ii) Here the freq. distn. is more on less symmetric and the variable given continuous. Let X denotes the length of the left middle finger of a criminal.

Assume that $X \sim N(\mu, \sigma^2)$.

The PDF of X is $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$; $x \in \mathbb{R}$.

By method of moments,

$$\mu_1' = m_1', \quad \mu_2 = m_2.$$

$$\Rightarrow \hat{\mu} = \bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \text{_____}$$

$$\hat{\sigma}^2 = s^2 = \frac{\sum x_i^2 f_i}{\sum f_i} = \text{_____} = \bar{x}^2.$$

Hence, the fitted PDF of X is

$$f(x) = \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right)^2}; \quad x \in \mathbb{R}.$$

(iii)

| Cls Mark | Class Boundaries |
|---|---|
| 9.5 | 9.35 – 9.65 |
| 9.8 | 9.65 – 9.95 |
| ⋮ | ⋮ |

Exp. freq. of the class $(a, b)$ is

$= N.P[a < X < b] = N\left\{\Phi\left(\frac{b-\hat{\mu}}{\hat{\sigma}}\right) - \Phi\left(\frac{a-\hat{\mu}}{\hat{\sigma}}\right)\right\}$

| (x) Class boundaries | $\Phi\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right)$ | $\Phi\left(\frac{x-\hat{\mu}}{\hat{\sigma}}\right)$ | Exp. Freq $N.\Delta\Phi$ | Observed Freq. |
|---|---|---|---|---|
| $-\infty$ | 0 | - | | |
| 9.35 | - | | | |
| 9.65 | - | | | |
| ⋮ | ⋮ | | | |
| 13.65 | - | | | |
| $\infty$ | 1 | ... | | |

**Comment:-** If the graph of observed and expected frequencies are very close to each other then the fitting is good.

(b) The following table gives the frequency distribution of annual salaries of individuals in rupees obtained from Indian Income Tax Returns (1995)

Statistics of Individual salaries Assessed, 1955

| Annual Salary (rupees) | Frequency |
|---|---|
| below 5000 | 27,000 |
| 5001 – 10000 | 60,000 |
| 10001 – 25000 | 90,000 |
| 25001 – 40000 | 26,032 |
| 40001 – 80000 | 13,000 |
| 80001 – 100000 | 60000 |
| 100001 – 200000 | 1178 |
| 200001 and above | 312 |

Plot the data, using a suitable scale, and find out whether Pareto's law will give an adequate fit over the entire range of the observed distn., if not, try to fit the data in a suitable range.

Solution:- Let $X$ denotes the salary of an individual.

Assume that the distribution of $X$ is given. Then the DF of $X$ is

$$F(x) = 1 - \left(\frac{x_0}{x}\right)^{\gamma} \quad \text{if } x > x_0$$

$$\Rightarrow 1 - F(x) = \left(\frac{x_0}{x}\right)^{\gamma}$$

$$\Rightarrow P[X \leq x] = \left(\frac{x_0}{x}\right)^{\gamma}$$

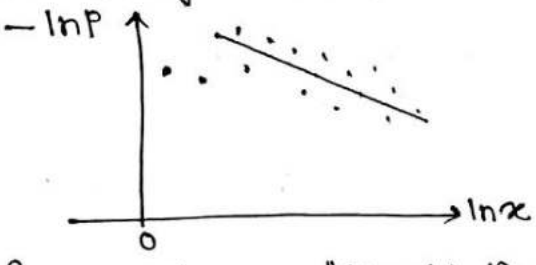
$$\Rightarrow \ln p = \ln x_0^{\gamma} - \gamma \ln x$$

$\Rightarrow \ln p = \alpha - \gamma \ln x$, where, $p$ is the proportion of individuals whose income is above $x$.

Note that, $-\ln p = -\alpha + \gamma \ln x$.

| Class | Freq. | $p = \frac{\text{c.f. of } \geq \text{ type}}{N}$ | $\ln p$ | $\ln x$ |
|---|---|---|---|---|
| | | | | |

$x$ = upper bound of the class.




From graph, we get — "Except the first two points, all the points are near about a straight line".
Hence, the Pareto law is appropriate after including 3rd class.

**Fitting of Pareto Distribution:-** Here $-\ln p = (-\alpha) + \ln x$

$\Rightarrow p' = -\alpha + \gamma x'$ $\quad (p' = \ln p,\ x' = \ln x)$

Hence $(p', x')$ are linearly related.
The parameters $\alpha, \gamma$ are estimated by method of least squares based on the values $(p_i', x_i')$, from 3rd class and above.

Normal Equations are ——

$$\sum p_i' = N'(-\alpha) + \gamma \sum x_i'$$

$$\sum x_i' p_i' = -\alpha \sum x_i' + \gamma \sum x_i'^2$$

$\Rightarrow \hat{\alpha} =$ _____ , $\hat{\gamma} =$ _____ .

Also $\hat{x}_0 =$ _____ , $\hat{\gamma} =$ _____ .

Hence, the fitted Pareto law is given by the D.F.

$$F(x) = 1 - \left(\frac{\hat{x}_0}{x}\right)^{\hat{\gamma}}, \text{ if } x > x_0 .$$

(c) The differences (in mins) of the actual arrival times and the scheduled arrival are tabulated below:

| Difference (in mins) | No. of days |
|---|---|
| $<4$ | 12 |
| 4-7 | 23 |
| 7-55 | 141 |
| 55-90 | 14 |
| $>90$ | 11 |

(i) fit a log-normal distribution to the data.
(ii) Use the fitted distribution to estimate the probability that the train will arrive at 1hr. 15mins late on a typical day.

**Solution:-** Let X denotes the differences (in mins) of the actual arrival times and the scheduled arrival time.

(i) Here $X \sim \Lambda(\mu, \sigma^2)$.

Note that —
$$p = P[X \le \xi_p] = P[\ln X \le \ln \xi_p]$$
$$= P\left[\frac{\ln X - \mu}{\sigma} \le \frac{\ln \xi_p - \mu}{\sigma}\right]$$
$$= \Phi\left(\frac{\ln \xi_p - \mu}{\sigma}\right)$$
$$= \Phi(z_p) \quad \left[\because \ln X \sim N(\mu, \sigma^2)\right]$$
$$\Rightarrow \xi_p = e^{\mu + \sigma z_p} \text{ ; where } \Phi(z_p) = p.$$

To product the tails and to cover up the most of the values, we shall use $P_{10}$ and $P_{90}$.

Now, equating the sample and population percentile, we get —
$$P_{10} = \hat{P}_{10} = 5.5 \Rightarrow e^{\mu + \sigma(z_{0.1})} = \hat{P}_{10}$$
$$P_{90} = \hat{P}_{90} = 72.5 \Rightarrow e^{\mu + \sigma(z_{0.9})} = \hat{P}_{90}$$
$$\Rightarrow \ln \hat{P}_{10} = \mu + \sigma(z_{0.1}) \quad \& \quad \ln \hat{P}_{90} = \mu + \sigma(z_{0.9})$$
$$\Rightarrow \hat{\mu} = \text{\_\_\_\_\_}, \quad \hat{\sigma} = \text{\_\_\_\_\_}.$$

(ii) Required Probability $= P[74.5 < X < 75.5]$
$$= \Phi\left(\frac{\ln 74.5 - \hat{\mu}}{\hat{\sigma}}\right) - \Phi\left(\frac{\ln 74.5 - \hat{\mu}}{\hat{\sigma}}\right).$$

# PRACTICALS ON STATISTICAL INFERENCE I

## ESTIMATION

1. The length of life recorded in hours for 10 electron tubes were:
980, 1020, 995, 1015, 990, 1030, 975, 950, 1050, 870
Assume that life-times are distributed in the form:

$$f(t) = \frac{1}{\theta} e^{-t/\theta}, \text{ if } t > 0, \text{ where } \theta > 0$$

(i) Obtain an estimate of $\theta$ and the estimated S.E. of this estimate.
(ii) Estimate also the probability that an electron tube will survive at least 100 hours.
(iii) Determine the lower confidence limit with confidence coefficient 0.95 to $\theta$ and to the probability of survival for 100 hours on the one.

**Solution:-** (i) By method of moments :-

$$E(T) = \bar{t}$$

$$\Rightarrow \hat{\theta} = \bar{t} = \frac{1}{10}\sum_{i=1}^{10} t_i = \text{:} \_\_\_\_\_ .$$

$$\text{S.E. of } \hat{\theta} = SE(\hat{\theta}) = \sqrt{Var(T)} = \sqrt{\frac{\theta^2}{n}} = \frac{\theta}{\sqrt{n}} .$$

$$\therefore \text{Estimated S.E. is } \widehat{SE}(\hat{\theta}) = \frac{\hat{\theta}}{\sqrt{n}} = \frac{\bar{t}}{\sqrt{10}} = \_\_\_\_\_ .$$

(ii) Probability that an electron tube will survive at least 100 hours

$$= P[T_i > 100]$$

$$= e^{-100/\theta}$$

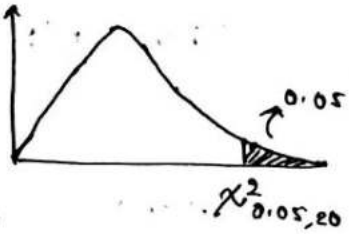
$$\therefore \hat{p} = e^{-100/\hat{\theta}} = e^{-100/\bar{t}} = \_\_\_\_\_ .$$

(iii) Note that $\frac{2T_i}{\theta} \sim \chi^2_2$, $i = 1(1)10$

$$\Rightarrow 2\sum_{i=1}^{10} \frac{T_i}{\theta} \sim \chi^2_{20}$$

Now, $P\left[2\frac{\sum_{i=1}^{10} T_i}{\theta} \le \chi^2_{0.05,20}\right] = 0.95$

$$\Rightarrow P\left[\frac{20.\bar{T}}{\chi^2_{0.05,20}} \le \theta < \infty\right] = 0.95$$




Hence, the observed lower confidence limit is

$$\left(\frac{20\bar{t}}{\chi^2_{0.05,20}}, \infty\right) = \left(\_\_, \infty\right)$$

2. Let $X_1, X_2, \ldots, X_{10}$ be ten independent and identically distributed random variables where each $X_i$ is normally distributed with mean 2.08. Further suppose 9.68% of its value is negative. If $X_{(1)} < X_{(2)} < \cdots < X_{(10)}$ be the ordered arrangement of $X_i$'s, then calculate $P\left[X_{(3)} \geq 5.28\right]$.

Solution:- The pdf of $X_{(r)}$ is

$$f_{X_{(r)}}(x) = \frac{\lfloor n}{\lfloor r-1 \; \lfloor n-r} \{F(x)\}^{r-1} \cdot f(x) \{1-F(x)\}^{n-r}$$

$$f_{X_{(3)}}(x) = \frac{\lfloor 10}{\lfloor 2 \; \lfloor 7} \{F(x)\}^2 \{1-F(x)\}^7 f(x)$$

and $$P\left[X_{(3)} \geq 5.28\right] = \int_{5.28}^{\infty} \frac{1}{\beta(3,8)} \{F(x)\}^2 \{1-F(x)\}^7 f(x)\, dx$$

$$= \int_{F(5.28)}^{1} \frac{1}{\beta(3,8)} z^2 (1-z)^7 dz$$

Now, $X_i \overset{iid}{\sim} N(2.08, \sigma^2)$

and $P[X_i < 0] = \frac{9.68}{100}$

$$\Rightarrow P\left[\frac{X_i - 2.08}{\sigma} < \frac{0 - 2.08}{\sigma}\right] = 0.0968$$

$$\Rightarrow \Phi\left(-\frac{2.08}{\sigma}\right) = 0.0968 \; ; \; X_i \sim N(2.08, \sigma^2)$$

$$\Rightarrow \Phi\left(\frac{2.08}{\sigma}\right) = 0.9132$$

$$\Rightarrow \sigma = \text{\_\_\_\_\_}.$$

Then $F(5.28) = P[X_i < 5.28] = P\left[\frac{X_i - 2.08}{\sigma} < \frac{5.28 - 2.08}{\sigma}\right]$

$$= \Phi\left(\frac{5.28 - 2.08}{\sigma}\right) = \text{\_\_\_\_\_} = x_0, \text{ say.}$$

Hence, $P\left[X_{(3)} \geq 5.28\right] = 1 - P\left[X_{(3)} \leq 5.28\right]$

$$= 1 - \int_0^{x_0} \frac{z^{3-1}(1-z)^{8-1}}{\beta(3,8)} dz$$

$$= 1 - I_{x_0}(3,8) \quad \left[\text{Use Pearson table for incomplete beta}\right]$$

## * TESTING OF HYPOTHESIS *

1. Let X be equal to the thickness of spearment gum manufactured for bending machine. Assume that the distn. of X is normal. The target thickness is 7.5 hundred of an age. The following 10 thickness of hundred of an age for pieces of gum that were selected randomly from the production line are

7.65, 7.60, 7.65, 7.70, 7.55, 7.40, 7.40, 7.50, 7.50, 7.55.

i) At $\alpha = 0.05$ significance level, was the company successful to meet the target thickness?

ii) What is the appropriate p- value of your test?

iii) Is $\mu = 7.50$ contained in a 95% confidence interval for $\mu$?

**Solution:-** Let $X \sim N(\mu, \sigma^2)$

From the sample, $\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 7.55$

and $s^2 = \frac{1}{10}\sum_{i=1}^{10} x_i^2 - \bar{x}^2 = 0.097$

(i) To test $H_0: \mu = 7.5$ Vs. $H_1: \mu \neq 7.5$

Test statistic is $T = \frac{(\bar{X} - \mu_0)\sqrt{n}}{S} \sim t_{n-1}$, under $H_0$.

Here $T = \frac{(\bar{X} - 7.5)\sqrt{10}}{S} \sim t_9$, under $H_0$.

**Critical region:-** If the observed value of $|T| > t_{\alpha/2, 9}$; we shall reject $H_0$ at $\alpha$ level of significance.

i.e. $\left| \left( \frac{\bar{x} - 7.5}{s} \right) \sqrt{10} \right| > t_{0.025, 9}$ ; at $\alpha = 0.05$

From table, $t_{0.025, 9} = 2.262$.

Observed value, $\frac{(\bar{x} - 7.5)\sqrt{10}}{s} = \left( \frac{7.55 - 7.5}{0.097} \right) \sqrt{10}$

$= 1.63 < t_{0.025, 9} = 2.262$

Hence, there is no reason to reject $H_0$ at $\alpha = 0.05$ level of significance, i.e., the company was successful to meet the target value.

(ii) p-value $= P_{H_0}[|T| \geq |t_0|]$

$= 2P_{H_0}[T \geq |t_0|]$ , $t_0$ is the observed value of T.

$= 2 . P_{H_0}[T \geq 1.63]$ [From Biometrica table]

$= 2 \times 0.1$

$= 0.2$.

As p-value is quite large, the observed value is a likely value under $H_0$.

(iii) 95% confidence interval for $\mu$ is

$\left(\bar{x} - t_{0.025,9} \cdot \frac{s}{\sqrt{n}} , \bar{x} + t_{0.025,9} \cdot \frac{s}{\sqrt{n}}\right)$

$= \left(7.55 - 2.62 \times \frac{0.097}{\sqrt{10}}, \text{ ——— }\right) = (7.48, 7.62)$ when $\mu = 7.5$.

2. The heart weights in groups of 12 females and 15 male cats are given below:

Male: 12.7, 15.6, 9.1, 12.1, 8.3, 11.2, 9.4, 8.0, 14.9, 10.7, 13.6, 9.6, 11.7, 9.3, 7.6

Female: 7.4, 7.3, 7.1, 9.0, 7.6, 9.5, 10.1, 10.2, 10.1, 9.5, 8.7, 7.2

Does the heart of a male cat on an average weight for male that of a female cat. State clearly any assumption you use.

Solutions:- Let X and Y denote the heart weights of a male and a female.

We assume that, $X \sim N(\mu, \sigma_1^2)$ independently
$Y \sim N(\mu, \sigma_2^2)$

To test $H_0: \mu_1 = \mu_2$ Vs. $H_1: \mu_1 > \mu_2$

We also assume that $\sigma_1 = \sigma_2 = \sigma$ (unknown)

Let $x_{11}, x_{12}, \ldots, x_{1n}$ be a r.s. from $N(\mu, \sigma^2)$

Let $x_{21}, x_{22}, \ldots, x_{2n}$ be a r.s. from $N(\mu_2, \sigma^2)$.

Here $n_1 = 15$, $n_2 = 12$

$\bar{x}_1 = \frac{1}{15} \sum_{i=1}^{15} x_{1i} =$ ———

$\bar{x}_2 = \frac{1}{12} \sum_{i=1}^{12} x_{2i} =$ ———

$s_1^2 = \frac{1}{n_1 - 1} \left\{ \sum x_{1i}^2 - n_1 \bar{x}_1^2 \right\} =$ ———

$s_2^2 = \frac{1}{n_2 - 1} \left\{ \sum x_{2i}^2 - n_2 \bar{x}_2^2 \right\} =$ ———.

We know, $s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} =$ ——— is an UE of $\sigma^2$.

Test statistic:-

$$T = \frac{\bar{X}_1 - \bar{X}_2}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}, \text{ under } H_0.$$

Critical Region:-

$$\frac{(\bar{x}_1 - \bar{x}_2)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} > t_{\alpha; n_1+n_2-2}$$

So, observed value of T is $\frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{15}+\frac{1}{12}}} =$ ———

From table, $t_{0.05,\ 15+12-2} = t_{0.05,25} = 1.708$.

Conclusion:-

3) In an experiment to investigate, the effect of light on root growth in mustors. sidlings, to growth of siblings, these grocon in identical condition, except that one coas kept in the dark, find the other coas explodes in sunlight during the day. After a certain period of time, the root lengths in mm of all the sidlings are measured. The following table gives the data obtained performed appropriate statistical list of significance to assist cohether

i) light effects, root growth, & to the variances of the root length in the two population to be equal but unknocon.

ii) Light effects in both growths in root lengths is the same.

The table makes two rows first for light and second for darkness.

L: 21, 39, 31, 30, 52, 39, 55, 50, 29, 17

D: 22, 16, 20, 14, 32, 28, 36, 41, 17, 22

Solution:-

i) To test:- $H_0: \mu_1 = \mu_2$ against $H_1: \mu_1 > \mu_2$.

If $H_1$ is accepted, then light affects the root growth.

(same as previous problem)

ii) To test $H_0: \sigma_1 = \sigma_2$ vs. $H_1: \sigma_1 \neq \sigma_2$

Test statistic:- $F = \frac{S_1^2}{S_2^2} \sim F_{n_1-1, n_2-1}$

Critical Region:- $\frac{s_1^2}{s_2^2} > F_{\alpha/2;\, n_1-1, n_2-1}$ or $< F_{1-\alpha/2;\, n_1-1, n_2-1}$.

cohere, $n_1 = 10$, $n_2 = 10$

Observed value of F is $\frac{s_1^2}{s_2^2} =$ _____.

$\alpha = 0.1$.

From table, $F_{0.05, 9.9} = 3.18$.

$$F_{0.95, 9.9} = \frac{1}{F_{0.05, 9.9}} = \frac{1}{3.18}$$

Conclusion:-

4) [CU '2006]

The following table contains observations on the supline systolic and diastolic blood pressures (in mm of Hg) for 15 patients with moderate hypertension, immediately before and two hours after taking a drug, captopril.

(i) Perform an appropriate test for the hypothesis that the drug is successful in reducing the dia stolic blood pressure.

(ii) Obtain a 95% confidence interval for the mean difference in the systolic blood pressure before and after treatment.:

| Patient No. | Systolic BP | | Diastolic BP | |
|---|---|---|---|---|
| | before | After | Before | After |
| 1 | 210 | 201 | 130 | 125 |
| 2 | 169 | 165 | 122 | 121 |
| 3 | 187 | 166 | 124 | 121 |
| 4 | 160 | 157 | 104 | 106 |
| 5 | 167 | 147 | 112 | 101 |
| 6 | 176 | 145 | 101 | 85 |
| 7 | 185 | 168 | 121 | 98 |
| 8 | 206 | 180 | 124 | 105 |
| 9 | 173 | 147 | 115 | 103 |
| 10 | 146 | 136 | 102 | 98 |
| 11 | 174 | 151 | 98 | 90 |
| 12 | 201 | 168 | 119 | 98 |
| 13 | 198 | 179 | 106 | 110 |
| 14 | 148 | 129 | 107 | 103 |
| 15 | 154 | 131 | 100 | 82 |

Solution:- (i) Let X and Y denote the diastolic BP of a patient before and after taking the drug.

Assume that, $(X, Y) \sim BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$.

To test whether the drug is successful in reducing the diastolic BP, i.e. to test:

$H_0: \mu_x = \mu_y$ Vs. $H_1: \mu_x > \mu_y$.

Let $(x_i, y_i)$, $i = 1(1)15$, be 15 paired samples on $(X, Y)$.

Define, $D_i = X_i - Y_i$, $i = 1(1)n$.

$\Rightarrow D_i \overset{iid}{\sim} N(\mu_0, \sigma_0^2)$, where $\mu_D = \mu_x - \mu_y$.

Here to test, $H_0: \mu_0 = 0$ Vs. $H_1: \mu_0 > 0$,

Test statistic:- $\frac{\sqrt{n}(\bar{D}-0)}{S_D} \sim t_{n-1}$

Critical region:- $\frac{\sqrt{n}.\bar{d}}{S_d} > t_{\alpha;n-1}$.

From the data:- for $n=15$,

| Patient No. | 1 | 2 | 3 | ... | 15 |
|---|---|---|---|---|---|
| d : | | | | | |

$\bar{d} = \frac{1}{15}\sum d_i =$ _____

$S_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i-\bar{d})^2 = \frac{1}{14}\{\sum d_i^2 - 15\bar{d}^2\} =$ _____.

Observed value of the test statistic is $\frac{\sqrt{n}.\bar{d}}{S_d} =$ _____.

Tabulated value: $t_{0.05,14} =$ _____.

Conclusion:-

(ii) Let U and V denote the systolic BP of a patient before and after taking drug.

Assume that, $(U,V) \sim BN(\mu_U, \mu_V, \sigma_U^2, \sigma_V^2, \rho)$

Let $(U_i, V_i)$, $i=1(1)15$ be the 15 given paired sample on $(U,V)$.

Define, $D_i = U_i - V_i \overset{iid}{\sim} N(\mu_D^2, \sigma_D^2)$, $\mu_D = \mu_U - \mu_V$.

$\Rightarrow \frac{\sqrt{n}(\bar{D}-\mu_0)}{S_D} \sim t_{n-1}$.

Now, 95% C.I. for $\mu_D$ is $\left(\bar{d} - t_{0.05,14}\cdot\frac{S_d}{\sqrt{15}}, \bar{d} + t_{0.05,14}\cdot\frac{S_d}{\sqrt{15}}\right)$

Here,

| Patient No. | 1 | 2 | ... | 15 |
|---|---|---|---|---|
| d = | | | | |

$\bar{d} =$ _____

$S_d =$ _____

Hence the observed 95% C.I. for $\mu_D$ is

( , ).

5) [CU'2005]

The members of a team of nine men were asked to load and fire naval guns by one of two methods $M_1$ and $M_2$, attempting to get off as many rounds per minute as possible. Two sets of such firing were made per method. The following table gives the outcome of this experiment.

| Participate No. | Number of rounds fired per minute in first set | | second set | |
|---|---|---|---|---|
| | by $M_1$ | by $M_2$ | by $M_1$ | by $M_2$ |
| 1 | 20.2 | 14.2 | 24.1 | 16.2 |
| 2 | 22.0 | 14.1 | 23.5 | 16.1 |
| 3 | 23.1 | 14.1 | 22.9 | 16.1 |
| 4 | 26.2 | 18.0 | 26.9 | 19.1 |
| 5 | 22.6 | 14.0 | 24.6 | 18.1 |
| 6 | 22.9 | 12.2 | 23.7 | 13.8 |
| 7 | 23.8 | 12.5 | 24.9 | 15.4 |
| 8 | 22.9 | 13.7 | 25.0 | 16.0 |
| 9 | 21.8 | 12.7 | 23.5 | 15.1 |

In the light of this data, are we justified in inferring that method $M_1$ is significantly better than $M_2$? Obtain a 95% C.I. for the difference in the mean performance by the two methods.

Solution:- Let $(X_i, Y_i)$ be the paired samples in the 1st set by $M_1$ and $M_2$, $i=1(1)9$.

Let $(X_i', Y_i')$ be the paired samples in the 2nd set by $M_1$ and $M_2$ for $i=1(1)9$.

We assume that, $(X_i, Y_i) \overset{iid}{\sim} BN$, $i=1(1)9$ $\Big\rangle$ independent.
$(X_i', Y_i') \overset{iid}{\sim} BN$, $i=1(1)9$

As $(X_i, Y_i)$, $(X_i', Y_i')$ are observed on the $i^{th}$ gun man.

Note that, $X_i \overset{iid}{\sim}$ Normal
$X_i' \overset{iid}{\sim}$ Normal , $i=1(1)9$.

$$\Rightarrow \frac{X_i + X_i'}{2} \overset{iid}{\sim} N\left(\mu_{M_1}, \sigma^2_{M_1}\right)$$

Similarly, $\frac{Y_i + Y_i'}{2} \overset{iid}{\sim} N\left(\mu_{M_2}, \sigma^2_{M_2}\right)$.

But $X_i + X_i'$, $Y_i + Y_i'$ are observed on the same $i^{th}$ gun man, $i = 1(1)9$.

So, $\left( \frac{X_i + X_i'}{2}, \frac{Y_i + Y_i'}{2} \right) \overset{iid}{\sim} BN\left( \mu_{M_1}, \mu_{M_2}, \sigma^2_{M_1}, \sigma^2_{M_2}, \rho \right)$

To test $H_0: \mu_{M_1} = \mu_{M_2}$ Vs. $H_1: \mu_{M_1} > \mu_{M_2}$.

Define, $D_i = \frac{(X_i + X_i') - (Y_i + Y_i')}{2} \overset{iid}{\sim} N(\mu_D, \sigma_D^2)$, $i = 1(1)9$.

where, $\mu_D = \mu_{M_1} - \mu_{M_2}$.

Now, To test, $H_0: \mu_D = 0$ Vs. $H_1: \mu_D > 0$

(Use Paired t-test).

6. (a) The correlation coefficient between head-length and stature for a sample of 36 members of an Indian tribe has been found to be 0.4339. Is it reasonable to assume that in the population of the characters are uncorrelated?

(b) Examine on the basis of the data given in the table below whether the variances of muscle weights of right and left legs of rabbits are equal.

| Sample no. of rabit | Weight (gms) of anterior muscle of | |
|---|---|---|
| | left leg | Right leg |
| 1 | 5.0 | 4.9 |
| 2 | 4.8 | 5.0 |
| 3 | 4.3 | 4.3 |
| 4 | 5.1 | 5.3 |
| 5 | 4.1 | 4.1 |
| 6 | 4.0 | 4.0 |
| 7 | 7.1 | 6.9 |
| 8 | 5.9 | 6.3 |
| 9 | 5.3 | 5.2 |
| 10 | 5.3 | 5.5 |
| 11 | 5.3 | 5.5 |
| 12 | 5.9 | 5.9 |
| 13 | 6.5 | 6.8 |
| 14 | 6.3 | 6.3 |
| 15 | 6.6 | 6.6 |
| 16 | 5.2 | 6.3 |

(c) Given in the table are the means, the s.d.s and the correlation coefficient of scores $x, y$ on two halves of a psychological test on 20 days.

(i) Examine whether the scores on the two halves are uncorrelated.

(ii) Examine whether the scores on the two halves are equally variable.

| Score | Mean | S.d. | Correlation |
|---|---|---|---|
| $x$ | 45.5 | 9.51 | 0.76 |
| $y$ | 50.2 | 5.25 | |

(a) Let $(X_i, Y_i)$, $i = 1(1)36$, be the paired values on head-length and stature for 36 indian tribe. Assume that $(X_i, Y_i) \sim BN$ with correlation coefficient $\rho$.

To test $H_0: \rho = 0$ Vs. $H_1: \rho \neq 0$

Test statistic:-

$$\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}, \text{ under } H_0.$$

Critical region:-

$$\left| \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \right| > t_{\alpha/2, n-2}$$

Here $n = 36$, $r = 0.4339$

Observed value of the test statistic is $\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.4339 \times \sqrt{34}}{\sqrt{1-(0.4339)^2}}$

$= 2.808$.

From Biometrica, Vol-I, for $\alpha = 0.05$

$t_{0.025, 30} = 2.042$

$t_{0.025, 40} = 2.021$

Here, $\left| \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \right| = 2.808 > t_{0.025, 34}$

$\Rightarrow$ $H_0$ is rejected at 5% level based on the given information.

$\Rightarrow$ The population characters are not uncorrelated.

(b) Let $(X_i, Y_i)$ denote the weights of right and left legs of the $i^{th}$ rabit, $i = 1(1)6$.

Assume that $(X_i, Y_i) \overset{iid}{\sim} BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$, $i = 1(1)n$, $n = 16$.

To test $H_0: \sigma_x = \sigma_y$ Vs. $H_1: \sigma_x \neq \sigma_y$

Define, $U_i = X_i + Y_i$

$V_i = X_i - Y_i$ , $i = 1(1)16$.

Note that $\rho_{UV} = 0$, under $H_0$.

To test, $H_0: \sigma_x = \sigma_y$

$\Leftrightarrow$ $H_0': \rho_{UV} = 0$

To test $H_0: \rho_{UV} = 0$, based on the paired sample $(U_i, V_i)$, $i = 1(1)16$; compute

$$r_{UV} = \frac{\frac{1}{n}\sum u_i v_i - \bar{u}\bar{v}}{\sqrt{\left\{\frac{1}{n}\sum u_i^2 - \bar{u}^2\right\}}\sqrt{\left\{\frac{1}{n}\sum v_i^2 - \bar{v}^2\right\}}}.$$

| Serial no | 1 2 . . . . . 16 |
|---|---|
| $U_i$ | |
| $V_i$ | |

Test statistic: $\frac{r_{UV}\sqrt{n-2}}{\sqrt{1-r_{UV}^2}} \sim t_{n-2}$, under $H_0'$.

(c) (i) Here we are to test whether $\rho = 0$ or not.

(ii) To test $\sigma_1 = \sigma_2$

If $\rho = 0$, then it becomes a testing of equality of variances in uncorrelated test.

If $\rho \neq 0$, it becomes a testing of equality of variances in correlated test.

To test $\mu_1 = \mu_2$

If $\rho = 0$, & $\sigma_1 = \sigma_2$ are accepted, then it becomes a Fisher's t-test.

If $\rho \neq 0$ then it becomes a paired t-test.

7. (a) In an excavation 10 fossils were discovered, of which 6 could be definitely classified as male and 3 as a female. The sex of the last fossil could not be precisely determined. Are these findings compatible with a sex ratio?

**Solution:-** Let $X$ denotes the no. of male fossils out of 9 fossils obtained.

Assuming probability of getting a male fossils $= p$,

$X \sim Bin(9, p)$.

To test $H_0 : p = 1/2$ against $H_1 : p \neq 1/2$.

Here the observed value of $X$ is $x_0 = 6$.

$$\text{The p-value} = 2 \min \{ P_{H_0}[X \geq x_0], P_{H_0}[X \leq x_0] \}$$

$$= 2 \times \min \left\{ \sum_{x=6}^{9} \binom{9}{x} \left(\frac{1}{2}\right)^9, \sum_{x=0}^{6} \binom{9}{x} \left(\frac{1}{2}\right)^9 \right\}$$

$$= 2 \times \frac{\binom{9}{6} + \binom{9}{7} + \binom{9}{8} + 1}{2^9}$$

$$= 0.50$$

Let $\alpha = 0.05$ be the chosen level of significance.

Clearly, p-value $> 0.05 = \alpha$.

$\Rightarrow$ $H_0$ is accepted.

$\Rightarrow$ The sex-ratio is 1:1 is supported by the data.

(c) The following table gives the result of an experiment to compare the effect of newly discovered medicine on a certain disease with that of the prevailing treatment (control). Do the data confirm the superiority of the new drug?

| Treatment | Cured | Not cured | Total |
|---|---|---|---|
| Control | 3 | 5 | 8 |
| New medicine | 5 | 3 | 8 |
| Total | 8 | 8 | 8 |

**Solution:-** Let $X_1$ denotes the no. of cured out of 8 patients treated under control.

Let $X_2$ " " " " " " " "

under new medicine.

Let $X_1 \sim Bin(8, p_1)$
$X_2 \sim Bin(8, p_2)$ > independently ; $n_1 = 8, n_2 = 8$

To test whether neco medicine is superior.

$\Rightarrow$ To test $H_0 : p_1 = p_2$ Vs. $H_1 : p_1 < p_2$.

From data, the observed value of $X_1$ and $X_2$ are $x_{10} = 3, x_{20} = 5$

Then $X_1 + X_2 = X$ has the observed value $x_0 = 8$.

In the testing problem,

p-value is $= P_{H_0}\left[ X_1 \le x_{10} \mid X = x_0 \right]$

$$= \sum_{x_1=0}^{x_{10}} \frac{\binom{n_1}{x_1}\binom{n_2}{x_0 - x_1}}{\binom{n_1+n_2}{x_0}} = \sum_{x_1=0}^{3} \frac{\binom{8}{x_1}\binom{8}{8-x_1}}{\binom{16}{8}}$$

$= \_\_\_\_\_$.

Conclusion :-

8) (a) The number of deaths from dracoring in a certain river in two consecutive months were 8 and 5. Are these fluctuations due to chance?

(b) A newspaper in a certain city observed that driving conditions have much improved because the number of fatal automobile accidents in last year was 9 where as the average number per year over the past several year was 15. Is the statement justified? If further, the number of fatal automobile accidents in the first six month of the current year is given to be 3, could you modify your earlier conclusion?

(c) A 5 cm specimen of a new type of fibre is found to have 13 defects while the manufacturer claims that there are no more than 150 defects per 100 cm. Do the above data support this claim?

Solution:- (a) Let $X_1$ and $X_2$ denote the number of deaths of dracoring in a certain river in two consecutive months.

Let $X_1 \sim P(\lambda_1)$ $\Big\rangle$ independently
$X_2 \sim P(\lambda_2)$

To test whether the observed values are the values of the same population or not.

$\Leftrightarrow$ To test $H_0: \lambda_1 = \lambda_2$ vs. $H_1: \lambda_1 \neq \lambda_2$.

Define, $X = X_1 + X_2 \sim P(\lambda)$, under $H_0: \lambda_1 = \lambda_2 = \lambda$.

The observed values of $X_1$ and $X$ are $x_{10} = 8$, $x_0 = 8+5 = 13$.

$\therefore$ p-value $= 2 \times \min \left\{ P_{H_0}[X_1 \geq x_{10} / X = x_0], P_{H_0}[X_1 \leq x_{10} / X = x_0] \right\}$

$$= 2 \times \min \left\{ \sum_{x=x_{10}}^{x_0} \binom{x_0}{x} \frac{1}{2^{x_0}}, \sum_{x=0}^{x_{10}} \binom{x_0}{x} \frac{1}{2^{x_0}} \right\}$$

$$= 2 \times \min \left\{ \sum_{x=8}^{13} \binom{13}{x} \frac{1}{2^{13}}, \sum_{x=0}^{8} \binom{13}{x} \frac{1}{2^{13}} \right\}$$

$=$ _____.

Conclusion:-

(b) Let $X$ denotes the number of accidents per year.

We assume that $X \sim P(\lambda)$

To test $H_0: \lambda = 15$ Vs. $H_1: \lambda < 15$

Observed value of $X$ is $x_0 = 9$

$p$-value $= P_{H_0}[X \le x_0]$

$$= \sum_{x=0}^{9} e^{-15} \cdot \frac{15^x}{x!} = 1 - e^{-15} \sum_{x=0}^{15} \frac{15^x}{x!} = \text{_____}.$$

Conclusion:- Let $\alpha = 0.05$, If $p$-value $< \alpha = 0.05$ then $H_0$ is rejected.

Let $X'$ denotes the no. of accidents per six month. Then

$$X' \sim \text{Poi}(\lambda/2).$$

The observed value of $X'$ is $x_0' = 8$.

$$\therefore p\text{-value} = P_{H_0}[X' \le x_0'] = \sum_{x=0}^{8} e^{-7.5} \frac{(7.5)^x}{x!} = \text{_____}.$$

Conclusion:-

(c) Let $X$ denotes the no. of defects in a 5 cm specimen. We assume that $X \sim P(\lambda)$.

Manufacturer claims:

There are not more than 150 defects in 100 cm.

$\Rightarrow$ " " " " " 7.5 " " 5 cm.

To test $H_0: \lambda_0 = 7.5$ Vs. $H_1: \lambda > 7.5$

The observed value of $X$ is $x_0 = 13$

$$\therefore p\text{-value} = P_{H_0}[X \ge 13]$$

$$= \sum_{x=13}^{\infty} e^{-7.5} \frac{(7.5)^x}{x!}$$

$$= 1 - \sum_{x=0}^{13} e^{-7.5} \frac{(7.5)^x}{x!}$$

$$= \text{_____}.$$

9. For 20 pairs of fathers and sons, the regression equation of height of son ($y$) on height of father ($x$), both measured in cm, was found to be

$$y = 9.29 + 0.932x$$

For 20 pairs, $\bar{x} = 168.17$, $\sum(x_i - \bar{x})^2 = 777.80$ and $\sum(y_i - \bar{y})^2 = 939.42$. Test whether the regression coefficient differs significantly from unity. Find the 95% confidence limits to the conditional mean of $y$ given $x = 177$. Also, the prediction limits of the height of a son when the height of the father is known to be 177 cm.

Solution:- Let $\eta_x = \alpha + \beta x$ be the regression equation of $y$ on $x$.

The least square linear regression equation of $y$ on $x$ is

$$y = a + bx$$

i.e. $y = 9.29 + 0.932x$.

To test $H_0: \beta = 1$ Vs. $H_1: \beta \neq 1$

Test statistic:- $$\frac{(b-1)\sqrt{S_{xx}}}{S_{y.x}} \sim t_{n-2}, \text{ under } H_0.$$

From data, $b = 0.932$ ; $n = 20$

$$S_{xx} = \sum(x_i - \bar{x})^2 =$$

$$s^2_{y.x} = \frac{1}{n-2}\sum_{i=1}^{n}\{y_i - \bar{y} - b(x_i - \bar{x})\}^2$$

$$= \frac{\sum(y_i - \bar{y})^2 - b^2\sum(x_i - \bar{x})^2}{n-2}.$$

$$= \frac{S_{yy} - b^2 S_{xx}}{n-2} = \text{_____}.$$

If the observed $\left|\frac{(b-1)\sqrt{S_{xx}}}{S_{y.x}}\right| > t_{\alpha/2, n-2}$.

We shall reject $H_0$ at $\alpha$-level.

Now, observed $\left| \frac{(b-1)\sqrt{S_{xx}}}{S_{y.x}} \right| =$ _____.

and from table, for $\alpha = 0.05$, $t_{0.025,18} =$ _____.

Conclusion:-

- To find 95% C.I. for $\eta_x = \alpha + \beta x$, when $x = 177$.

Let $Y_x = a + bx$ be the predicton value when $x$ is given.

$$\frac{Y_x - \eta_x}{S_{y.x}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}} \sim t_{n-2}.$$

$\left[ Y_x = a + bx = \text{—} \right.$
$\left. \therefore \bar{x} = 168.17 \right]$

The 95% C.I. for $\eta_x$ is $\left( Y_x - t_{0.025,18} \cdot S_{yx}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}, \right.$

where, $n = 20$, $x = 177$, $\left. Y_x + t_{0.025,18} \cdot S_{y.x}\sqrt{\frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}} \right)$

- To find 95% prediction limit for $y$ when $x = 177$

$$\frac{y - Y_x}{S_{y.x}\sqrt{1 + \frac{1}{n} + \frac{(x-\bar{x})^2}{S_{xx}}}} \sim t_{n-2}.$$

The 95% prediction limit for $y$ when $x = 177$ is

$$\left( Y_{177} \mp t_{0.025,18} \cdot S_{yx}\sqrt{1 + \frac{1}{20} + \frac{(177-\bar{x})^2}{S_{xx}}} \right).$$

# Problems on Large Sample [Only CU problems]

1. In a sample of size 100 from a bivariate popln., the correlation coefficient is found to be 0.25. Test whether it is (i) significant (ii) significantly less than 0.5.

**Solution:-** Here $n = 100$

$r = 0.25$

(i) To test $H_0: \rho = 0$ Vs $H_1: \rho > 0$.

[exact test: $\frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$]

Using Fisher's Z-transformation,

$$\sqrt{n-3}\,(Z - \xi) \overset{a}{\sim} N(0,1)$$

Here $Z = \frac{1}{2}\log\left(\frac{1+r}{1-r}\right) =$ ——.

$$\xi_0 = \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right) = 0.$$

$\therefore$ Under $H_0$,

$$\tau = \sqrt{n-3}\,(Z - 0) \overset{a}{\sim} N(0,1).$$

**Critical region:-** Observed $\tau > \tau_\alpha$.

(ii) To test $H_0: \rho = 0.5$ Vs $H_1: \rho < 0.5$

Here $\xi_0 = \frac{1}{2}\log\left(\frac{1+0.5}{1-0.5}\right) =$

Under $H_0$,

$$\tau = \sqrt{n-3}\,(T - \xi_0) \overset{a}{\sim} N(0,1).$$

**Critical region:-** Observed $T < -\tau_\alpha$.

2. To examine a manufacturer claim that not more than 1·9% of the products are defectives, then 830 items are put to inspection and 25 found defected can be consider the claim to be justified.

Sol. To $H_0: p = 0.019$ vs. $H_1: p > 0.019$

Here $n = 830$

$$\hat{p} = \frac{25}{830}.$$

Using $\sin^{-1}$ transformation,

Under $H_0$,

$$\gamma = \sqrt{4n}\left(\sin^{-1}\sqrt{\hat{p}} - \sin^{-1}\sqrt{0.019}\right) \sim N(0,1)$$

Critical region: observed $\gamma > \gamma_\alpha$.

3. A proponent of inovative teaching methods wishes to compare the effectiveness of teaching english by the traditional classroom lecture system and by the extensive use of audio-visual aid. To do so, 100 students are selected at random from a class of 250 and assign to audio-visual instruction. the remaining 150 students are taught english in classroom lecture. At the end of the term, all 250 students are given a test; the no. of students from each group to pass the test is recored in the following table.

| Medium | Pass | Fail |
|---|---|---|
| Audio-visual instruction | 63 | 37 |
| classroom lecture | 107 | 43 |

(a) Find the 95% C.I. for the difference between success rate for the two methods of instructions.

(b) Do the data strongly support that a better passing rate is achieved using the classroom lecture that is acheived using A.V. method.

(c) Explain whether your inferential procedure crucial depends on the term 'random'.

Sol. Let $p_1$ and $p_2$ be the passing rates in Audio-visual and classroom lecture.

Here $\hat{p}_1 = \frac{63}{100}$, $\hat{p}_2 = \frac{107}{150}$.

(a) Here $P\left[\sin^2\left(\sin^{-1}\sqrt{\hat{p}_i} - \frac{\gamma_{\alpha/4}}{\sqrt{4n_i}}\right) \le p_i \le \sin^2\left(\sin^{-1}\sqrt{\hat{p}_i} + \frac{\gamma_{\alpha/4}}{\sqrt{4n_i}}\right)\right]$

$= 1-\alpha$, $i = 1,2$.

$\Rightarrow P[L_i \le p_i \le U_i] = 1-\alpha$, $i = 1,2$.

Then $P[L_1 \le p_1 \le U_1, L_2 \le p_2 \le U_2]$

$= P[A_1 \cap A_2] \ge P[A_1] + P[A_2] - 1 = 1 - \alpha/2 + 1 - \alpha/2 - 1$

$= 1 - \alpha$.

$\Rightarrow P[L_1 - U_2 \le p_1 - p_2 \le U_1 - L_2] \ge 1 - \alpha$,

Here $L_1 =$

$L_2 =$

$U_1 =$

$U_2 =$

and $\alpha = 0.05$.

Hence $(L_1 - U_2, U_1 - L_2) = (\quad , \quad)$ is an observed C.I for $(p_1 - p_2)$ with confidence leve 0.95.

(b) To test $H_0: p_1 = p_2$ against $H_1: p_1 < p_2$

[ Here $\sqrt{4n_i}\left(\sin^{-1}\sqrt{\hat{p}_i} - \sin^{-1}\sqrt{p_i}\right) \overset{a}{\sim} N(0,1)$, $i = 1,2$, independently as the groups are selected randomly. ]

Here, $\sin^{-1}\sqrt{\hat{p}_i} \overset{a}{\sim} N\left(\sin^{-1}\sqrt{p_i}, \frac{1}{4n_i}\right)$, $i = 1,2$ independently.

Now, $\sin^{-1}\sqrt{\hat{p}_1} - \sin^{-1}\sqrt{\hat{p}_2} \overset{a}{\sim} N\left(\sin^{-1}\sqrt{p_1} - \sin^{-1}\sqrt{p_2}, \frac{1}{4n_1} + \frac{1}{4n_2}\right)$

Under $H_0$, $\tau = \dfrac{\sin^{-1}\sqrt{\hat{p}_1} - \sin^{-1}\sqrt{\hat{p}_2}}{\sqrt{\frac{1}{4n_1} + \frac{1}{4n_2}}} \overset{a}{\sim} N(0,1)$.

Critical region:- observed $\tau < -\tau_\alpha$.

(c) As the groups are selected randomly, the data obtained from the two groups then constitute two independent random samples. Therefore the estimates $\hat{p}_1$ and $\hat{p}_2$ are independently distributed and accordingly we obtain the test statistic.

4. For 600 beans of particular variety, the frequency distn. of breath (in m.m) has $\gamma_1 = 0.093$, $\gamma_2 = -0.125$, where $\gamma_1$ and $\gamma_2$ are sample measures of skewness and kurtosis. Examine if the popln. can be supposed to be normal.

Sol. Here to test $H_0$: the data is a r.s. from normal popln., As the measures skewness $g_1$ and kurtosis $g_2$ are given, then $H_0$ reduces to $H_{01}: \gamma_1 = 0$ and $H_{02}: \gamma_2 = 0$ as far as given information is concerned.

If both $H_{01}$ and $H_{02}$ are accepted, then we accept $H_0$.

Under $H_0$, $g_1 \overset{a}{\sim} N\left(0, \frac{6}{n}\right)$, $g_2 \overset{a}{\sim} N\left(0, \frac{24}{n}\right)$.

Test statistic: $T_1 = \sqrt{\frac{n}{6}} g_1 \overset{a}{\sim} N(0,1)$

$T_2 = \sqrt{\frac{n}{24}} g_2 \overset{a}{\sim} N(0,1)$.

If observed $|T_1| > T_{\alpha/2}$, we reject $H_{01}$.
If observed $|T_2| > T_{\alpha/2}$, we reject $H_{02}$.

[ Prob. of accepting $H_0 = P[|T_1| < T_{\alpha/2}, |T_2| < T_{\alpha/2}]$

$\geq (1-\alpha) + (1-\alpha) - 1 = 1 - 2\alpha$.

$\Rightarrow$ prob. of rejecting $H_0 \leq 2\alpha$. ]

[ Calculation & Conclusion: suggestion: use $\alpha = 1\%$. ]

5. The corr. coeff. between stature (cm) and nasal height (cm) of a group of 106 males is 0.672 and that for a female group of 117 females is 0.725. Can you treat the corr. coefficients to differ significantly.

Sol. Let $\rho_1, \rho_2$ be the correlation coefficient between stature and nasal height of male and female, respectively.

Assuming the popln.s are bivariate.

To test $H_0: \rho_1 = \rho_2$ vs. $H_1: \rho_1 \neq \rho_2$.

Here $Z_i = \frac{1}{2} \log\left(\frac{1+r_i}{1-r_i}\right)$

$\xi_i = \frac{1}{2} \log\left(\frac{1+\rho_i}{1-\rho_i}\right)$

Here, $Z_i \overset{a}{\sim} N\left(\xi_i + \frac{\rho_i}{2(n-1)}, \frac{1}{n_i - 3}\right)$, $i = 1, 2$, independently.

Under $H_0: \rho_1 = \rho_2 = \rho$

$Z_1 - Z_2 \overset{a}{\sim} N\left(0, \frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}\right)$.

where $E(Z_1 - Z_2) = \frac{\rho}{2}\left\{\frac{1}{n_1 - 1} - \frac{1}{n_2 - 1}\right\} = \frac{\rho}{2} \cdot \frac{n_1 - n_2}{(n_1 - 1)(n_2 - 1)} \simeq 0$.

Test statistic:-

$$T = \frac{(Z_1 - Z_2)}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \overset{a}{\sim} N(0,1), \text{ under } H_0.$$

6. Consider the following set $A_1 = \{x: -\infty < x \le 0\}$

$A_i = \{x: i-2 \le x \le i-1\}$, $i = 2(1)7$.

$A_8 = \{x: 6 \le x < \infty\}$

A certain Hypothesis Ho assign probabilities $p_{i0}$ to this sets $A_i$ is accordence coith

$$p_{i0} = \int_{A_i} \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i - 3}{2}\right)^2} dx,$$

the hypothesis can be tested on the basis of the observed freaeencies of the sets $A_i$, $i = 2(1)8$, which are respectively

80, 96, 140, 210, 172, 160, 88, 74.

Would Ho is accepted at the 5% level of significance?

**Sol.** To test $H_0: p_{i0} = \int_{A_1} \frac{1}{2\sqrt{2\pi}} e^{-\frac{1}{8}(x-3)^2} dx$, $i = 1(1)6$.

$\Leftrightarrow$ $H_0: X \sim N(3, 2^2)$.

Here $p_{10} = P[X \le 0] = \Phi\left(-\frac{3}{2}\right)$

$p_{20} = P[0 < X \le 1] = \Phi(-1) - \Phi\left(-\frac{3}{2}\right)$

$\vdots$

$p_{80} = P[6 < X < \infty] = 1 - \Phi\left(\frac{3}{2}\right)$

| Class | Observed freq. ($O_i$) | Exp. Freq. $e_i = np_{i0}$ |
|---|---|---|
| $A_1$ | | |
| $A_2$ | | |
| $\vdots$ | | |
| $A_8$ | | |

This is a test of Goodness of fit, $n = \sum_{i=1}^{8} f_i$.

Test statistic

$$\chi^2 = \sum \frac{O_i^2}{e_i} - n \sim \chi^2_{8-1}, \text{ under Ho.}$$

Critical region:- Observed $\chi^2 > \chi^2_{\alpha, 7}$.

7. A survey of drivers was taken to see if they have been in an accident during the previous year, and if so was it the minor and major accident. The results are tabulated by age group

| | Accident type | | |
|---|---|---|---|
| Age | None | minor | major |
| Under 18 | 67 | 10 | 5 |
| 18-25 | 42 | 6 | 5 |
| 26-40 | 75 | 8 | 4 |
| 40-65 | 56 | 4 | 6 |
| over 65 | 57 | 15 | 1 |

Use an appropriate test to check if the data suggest that the nature of the accidents depends on age.

Sol. Let A: Nature of Accident
B: Age
To test A and are independent.

The test statistic is : $\chi^2 = n\left\{\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{f_{ij}^2}{f_{io}f_{oj}} - 1\right\} \sim \chi^2_{(r-1)(s-1)}$

Critical region:-

observed $\chi^2 > \chi^2_{\alpha,(r-1)(s-1)}$.

8. To compare the results of college 1 and 2 the following data on the final examination results have been obtained.

| College | Excellent | Results Good | Mediocore | Bad |
|---|---|---|---|---|
| 1 | 62 | 173 | 28 | 8 |
| 2 | 51 | 126 | 70 | 17 |

State the null hypothesis and test it to determine if the results vary between the colleges.

Sol.

$$\sum_{i=1}^{k}\frac{(f_{i1}-n_1p_{i1})^2}{n_1p_{i1}} + \sum_{i}\frac{(f_{i2}-n_2p_{i2})^2}{n_2p_{i2}} + \cdots\cdots$$

$$= \sum_{j=1}^{l}\sum_{i=1}^{k}\frac{(f_{ij}-n_ip_{ij})^2}{n_ip_{ij}} \sim \chi^2_{l(k-1)}.$$

Under $H_0$: $\sum\sum\frac{\left(f_{ij}-n_j\frac{f_{io}}{n}\right)^2}{n_j\cdot\frac{f_{io}}{n}}$ ; $\hat{p}_{io} = \frac{f_{io}}{n}$.

$$\chi^2 = n\left\{\sum\sum\frac{f_{ij}^2}{n_jf_{io}} - 1\right\} \sim \chi^2_{(l-1)(k-1)}.$$

Let A denotes the result of a college.

| | A | | | |
|---|---|---|---|---|
| | $A_1$ | $A_2$ | $A_3$ | $A_4$ |
| College I (sample I) | $f_{11}$ = 62 | $f_{21}$ = 173 | $f_{31}$ = 28 | $f_{41}$ = 8 |
| College II sample II | $f_{12}$ | $f_{22}$ | $f_{32}$ | $f_{42}$ |
| | $f_{10}$ | $f_{20}$ | $f_{30}$ | $f_{40}$ |

$n_1$ = ____

$n_2$ = ____

$f_{ij}$, $i = 1(1)4$, $j = 1(1)2$.

To test $H_0$: the homogeneity of the result in the colleges.

Test statistic, under $H_0$,

$$\chi^2 = n\left\{\sum_{j=1}^{2}\sum_{i=1}^{4}\frac{f_{ij}^2}{n_j f_{i0}} - 1\right\} \sim \chi^2_{(2-1)(4-1)}$$

Critical region:- observed $\chi^2 > \chi^2_{\alpha,3}$.

9. The no. of occurences of a word [redacted] in 48 essays per thousand essays examined, written by Hamilton and 50 essays written by Maddisson. The following table is obtained.

| Author | No. of essays according to rate of occurence of 'to' words | |
|---|---|---|
| | upto 45 | above 45 |
| Hamilton | 37 | 11 |
| Madison | 48 | 2 |

Test whether the data occurence of 'to' vary significantly between the two authors by applying the $\chi^2$ test, and also by applying the formulae for exact probability.

Sol.:→ Let A denotes the rate of occurence of [redacted] 'to'.

Let $B_1, B_2$ denote the Authors : Hamilton and Madison.

| | $A_1$ | $A_2$ | |
|---|---|---|---|
| $B_1$ | a = 37 | c = 11 | 48 |
| $B_2$ | b = 48 | d = 2 | 50 |
| | 85 | 13 | N = 98 |

To test $H_0$: A and B are independent.

Applying Yates continuity correction to Pearsonian $\chi^2$, we get

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

Critical region:- Observed $\chi^2 > \chi^2_{\alpha,1}$.

Exact Prob. Test:- Under Ho, given marginals, the prob. of obtaining the cell frequency $\begin{bmatrix} a & c \\ b & d \end{bmatrix}$ is

$$p_d = \frac{\binom{a+c}{a}\binom{b+d}{b}}{\binom{N}{a+b}} = \frac{\lfloor a+c \rfloor \lfloor b+d \rfloor \lfloor a+b \rfloor \lfloor c+d \rfloor}{\lfloor N \rfloor a \lfloor b \rfloor c \lfloor d}$$

p-value $= p_2 + p_1 + p_0$ [since Here $d = 2$]

p value $< \alpha \Rightarrow$ reject Ho.

$p_2 \rightarrow$

| | | |
|---|---|---|
| 37 | 11 | 48 |
| 48 | 2 | 50 |
| 85 | 13 | 98 |

$p_1 \rightarrow$

| | | |
|---|---|---|
| 36 | 12 | 48 |
| 49 | 1 | 50 |
| 85 | 13 | 98 |

$p_0 \rightarrow$

| | | |
|---|---|---|
| 35 | 13 | 48 |
| 50 | 0 | 50 |
| 85 | 13 | 98 |

10. The variance of the stature in cm for males in 8 different capital through out the world are given below. Examine whether the variance differs from capital to capital. If not, find the pooled estimates of the variance.

| Capital | Sample size | Variance of the stature (in $cm^2$) |
|---|---|---|
| 1 | 299 | 38.8306 |
| 2 | 77 | 24.8924 |
| 3 | 131 | 39.5068 |
| 4 | 59 | 24.4685 |
| 5 | 124 | 24.3490 |
| 6 | 170 | 32.4325 |
| 7 | 387 | 39.2262 |
| 8 | 139 | 33.4936 |

Sol. Let $\sigma_i$ be the popln s.d. of the $i^{th}$ capital.
Assuming the popln distributions are normal,

$$\ln s_i \sim N\left(\ln \sigma_i, \frac{1}{2n_i}\right), \quad i = 1(1)8.$$

$$\Rightarrow \sum_{i=1}^{8} \sqrt{2n_i}\left(\ln s_i - \ln \hat{\sigma}\right)^2 \sim \chi^2_7, \text{ under Ho: } \sigma_1 = \sigma_2 = \cdots = \sigma_8 = \sigma.$$

where, $\ln \hat{\sigma} = \frac{\sum 2n_i \ln s_i}{\sum 2n_i}$ = pooled estimate

Critical region: Observed $\chi^2 > \chi^2_{\alpha,7}$.

11. The correlation between height and weight coas formed to be 0·7, 0·8, 0·95 for samples of thousands object each for three different ethenic groups of children aged between 1 to 3 years. Test whether there is significant evidence for the dependence of the correlation on ethencity. If not, find the pool estimate of the common correlation.

**Sol:** Let $\rho_i$ be the correlation coefficient between height and weight of the $i^{th}$ ethnic group, $i=1,2,3$.

Assume, the popln. distns are BN.

To test $H_0: \rho_1=\rho_2=\rho_3$ ag. $H_1$: not $H_0$.

Here, $\sqrt{n_i-3}\,(z_i-\xi_i) \sim N(0,1)$, $i=1,2,3$, independently.

where, $z_i = \frac{1}{2}\log\left(\frac{1+r_i}{1-r_i}\right) =$

$\xi_i = \frac{1}{2}\log\left(\frac{1+\rho_i}{1-\rho_i}\right) =$

Under $H_0: \rho_1=\rho_2=\rho_3=\rho$,

$\sum_{i=1}^{3}(n_i-3)(z-\hat{\xi})^2 \sim \chi^2_{3-1}$, where $\hat{\xi} = \frac{\sum(n_i-3)z_i}{\sum(n_i-3)} = \bar{z}$.

Critical region:- Observed $\chi^2 > \chi^2_{\alpha;2}$

□ If $H_0: \rho_1=\rho_2=\rho_3=\rho$, the pooled estimate of the common correlation $\rho$ is given by

$\hat{\xi} = \bar{z}$

$\Rightarrow \frac{1}{2}\log\left(\frac{1+\hat{\rho}}{1-\hat{\rho}}\right) = \bar{z}$

$\Rightarrow \hat{\rho} = \frac{e^{2\bar{z}}-1}{e^{2\bar{z}}+1} =$ _____ .

12. Sample of sizes 75, 125, 100, 150, 130 from 5 poisson popln. give the mean values as 115/75, 197/125, 141/100, 237/150, 185/130, respectively. Do the popln have the same mean value?

Solution:-

To test $H_0: \lambda_1=\lambda_2=\lambda_3=\lambda_4=\lambda_5$

Under $H_0$: $\lambda_i=\lambda$, $i=1(1)5$.

$4n_i\left(\sqrt{\bar{x}_i}-\sqrt{\lambda}\right)^2 \sim \chi^2_{5-1}$.

where $\sqrt{\hat{\lambda}} = \sum 4n_i\sqrt{\bar{x}_i} \Big/ \sum 4n_i$

# Problems on Statistical Inference

1. The following observations ($Y$) are drawn from $N(\theta, \theta^K)$. Obtain the MLE of $\theta$ in each case $K=0,1$ and 2. Are they unique?

20·2, 22·9, 23·3, 20·0, 19·4, 22·0, 22·1, 22·0, 21·9, 21·5, 19·7, 21·5, 20·9.

**Hint:-** Let $y_1, y_2, \ldots, y_n$ be a r.s. from $N(\theta, \theta^K)$, $K=0,1,2$.

**K=0** The MLE of $\theta$ is $\hat{\theta} = \bar{y} =$ ——.

**K=1** $N(\theta, \theta)$

$$\hat{\theta} = -1 + \frac{1+\sqrt{\frac{4}{n}\sum y_i^2}}{2} = \text{——}$$

**K=2** $N(\theta, \theta^2)$, $\theta \neq 0$

$$L(\theta | y_1, \ldots, y_n) = \left(\frac{1}{\sqrt{2\theta^2\pi}}\right)^n \cdot e^{-\frac{1}{2\theta^2}\sum_{i=1}^{n}(y_i-\theta)^2}, \quad \theta \neq 0$$

$$0 = \frac{\partial \ln L}{\partial \theta}$$

$$0 = -\frac{n}{\theta} + \frac{\sum y_i^2}{\theta^3} - \frac{\sum y_i}{\theta^2}.$$

$$\Rightarrow \theta^2 + \theta \cdot \bar{y} - \frac{1}{n}\sum y_i^2 = 0$$

$$\Rightarrow \hat{\theta} = \frac{-\bar{y} \pm \sqrt{\bar{y}^2 + \frac{4}{n}\sum y_i^2}}{2} = \text{——}, \text{——}$$

[MLE is not unique]

2. The general experiment on linsead, the observed frequencies for petal and stigma colour are given in the following table along with probabilities of the cell in terms of a parameter $\theta$.

| Stigma Colour | Petal colour: lilac | Petal colour: deep lilac |
|---|---|---|
| White | $357\left(\frac{2+\theta}{4}\right)$ | $33\left(\frac{1-\theta}{4}\right)$ |
| Purple | $37\left(\frac{1-\theta}{4}\right)$ | $94\left(\frac{\theta}{4}\right)$ |

Estimate $\theta$ by the method of MLE, [ ( ) → Probability]
Also estimate the variance of this estimate of $\theta$.

**Sol.**

Likelihood function,

$$L(\theta | n_1, n_2, n_3, n_4) = \frac{\lfloor n}{\prod_{i=1}^{4} \lfloor n_i} \left(\frac{2+\theta}{4}\right)^{n_1} \left(\frac{1-\theta}{4}\right)^{n_2} \left(\frac{1-\theta}{4}\right)^{n_3} \left(\frac{\theta}{4}\right)^{n_4}$$

**Likelihood equation:**

$$\theta = \frac{\partial \ln L}{\partial \theta} = \frac{n_1}{2+\theta} - \frac{n_2+n_3}{1-\theta} + \frac{n_4}{\theta}.$$

$$\Rightarrow n\theta^2 - \{n_1 - 2(n_2+n_3) - n_4\}\theta - 2n_4 = 0, \quad n = \sum_{i=1}^{4} n_i$$

$$\Rightarrow \hat{\theta} = \text{——}.$$

For large $n$,

$\hat{\theta} \sim N\left(\theta, \frac{1}{I_n(\theta)}\right)$, where $I_n(\theta) = E\left(-\frac{\partial^2}{\partial\theta^2} \ln L\right)$

$\therefore I_n(\theta) = E\left\{\frac{n_1}{(2+\theta)^2} + \frac{n_2+n_3}{(1-\theta)^2} + \frac{n_4}{\theta^2}\right\}$

$= n\left\{\frac{\frac{2+\theta}{4}}{(2+\theta)^2} + \frac{2\left(\frac{1-\theta}{4}\right)}{(1-\theta)^2} + \frac{\theta/4}{\theta^2}\right\}$

$= \frac{n}{4}\left\{\frac{1}{2+\theta} + \frac{2}{1-\theta} + \frac{1}{\theta}\right\}$

Asymptotic variance is $\text{Var}(\hat{\theta}) = \frac{1}{I_n(\theta)}$.

$\widehat{\text{Var}}(\hat{\theta}) \simeq \frac{1}{I_n(\hat{\theta})} = \frac{n}{4}\left\{\frac{1}{2+\hat{\theta}} + \frac{2}{1-\hat{\theta}} + \frac{1}{\hat{\theta}}\right\}$

3) The length of life recorded in hours for 10 electron tubes were

980, 1020, 995, 1015, 990, 1030, 975, 950, 1050, 870.

Assume that life times are distributed in the form:

$$f(t,\theta) = \frac{1}{\theta} e^{-(t-\theta)}, \quad \theta > 0, \quad 0 < t < \infty.$$

Obtain the MLE of $\theta$ and the estimated standard error of this estimate. Estimate also the probability that an electron tube will survive at least 100 hours. Given estimate of the large sample standard error of this estimated probability. Determine the lower confidence limit. Confidence coefficient = 0.05, to the true prob. of survivor for 100 hours or more.

(i) By using the exact distn of MLE of $\theta$,

(ii) Assuming some approximate distn for the estimated prob. of survivor.

Sol. The likelihood function is

$$L(\theta \mid t_1, t_2, \ldots, t_n) = \frac{1}{\theta^n} e^{-\sum_{i=1}^{n} t_i/\theta}, \quad \theta > 0.$$

Here $n = 10$.

Likelihood equation is $0 = \frac{\partial}{\partial\theta} \ln L = -\frac{n}{\theta} + \frac{\sum t_i}{\theta^2}$

$\Rightarrow \hat{\theta} = \bar{t} =$ _____ is the MLE of $\theta$.

Now, $\text{S.E.}(\hat{\theta}) = \sqrt{V(\bar{t})} = \sqrt{\frac{\theta^2}{n}} = \frac{\theta}{\sqrt{n}}$.

and $\widehat{\text{S.E.}}(\hat{\theta}) = \frac{\hat{\theta}}{\sqrt{n}} = \frac{\bar{t}}{\sqrt{n}} =$ ____.

To estimate $p = P[T \geq 100] = e^{-100/\theta}$

MLE of $p$ is $\hat{p} = e^{-100/\hat{\theta}} =$ _____.

For large sample,

$$\hat{\theta} \overset{a}{\sim} N\left(\theta, \frac{1}{I_n(\theta)}\right), \text{ where } I_n(\theta) = \frac{n}{\theta^2}.$$

$$\Rightarrow \hat{p} = e^{-100/\theta} \overset{a}{\sim} N\left(p, \frac{\left\{\frac{\partial}{\partial\theta} p\right\}^2}{I_n(\theta)}\right)$$

For large $n$,

$$Var(\hat{p}) = \frac{\left\{p \cdot \frac{100}{\theta^2}\right\}^2}{\frac{n}{\theta^2}}.$$

$$\Rightarrow S.\hat{E}.(\hat{p}) = \frac{p \cdot \frac{100}{\theta^2}}{\sqrt{\frac{n}{\theta^2}}} = \frac{100p}{\sqrt{n}\,\theta}.$$

Hence, $\hat{SE}(\hat{p}) \simeq \frac{100\hat{p}}{\hat{\theta}\sqrt{n}} = \text{——}$.

Confidence limit of $p$:-

① Exact C.I.:- $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} t_i$

Here $y_i = \frac{2t_i}{\theta} \overset{iid}{\sim} \chi^2_2$, $i = 1(1)n$.

$$\Rightarrow 2\frac{\Sigma t_i}{\theta} \sim \chi^2_{2n}$$

$$\Rightarrow \frac{2n\hat{\theta}}{\theta} \sim \chi^2_{2n}$$

$$\therefore P\left[\frac{2n\hat{\theta}}{\theta} < \chi^2_{\alpha,2n}\right] = 1-\alpha$$

$$\Rightarrow P\left[\theta > \frac{2n\hat{\theta}}{\chi^2_{\alpha,2n}}\right] = 1-\alpha,$$

$$\Rightarrow P\left[e^{-100/\theta} > e^{-\frac{\chi^2_{\alpha,2n}}{2n\hat{\theta}} \cdot 100}\right] = 1-\alpha,$$

$$\Rightarrow P\left[p > e^{-100 \cdot \frac{\chi^2_{\alpha,2n}}{2n\hat{\theta}}}\right] = 1-\alpha,$$

Lower confidence limit with confidence coefficien 0.95 is

$$p_L = e^{-\frac{100 \cdot \chi^2_{0.05,20}}{20 \times 2 \times \hat{\theta}}} = \text{——}.$$

(ii) Approximate confidence limit:-

$$\hat{p} \overset{a}{\sim} N\left(p, SE^2(\hat{p})\right)$$

$$\Rightarrow \frac{\hat{p}-p}{\hat{SE}(\hat{p})} \overset{a}{\sim} N(0,1).$$

Hence for large $n$,

$$P\left[\left|\frac{\hat{p}-p}{S.\hat{E}(\hat{p})}\right| < \tau_{\alpha/2}\right] = 1-\alpha.$$

$$\Rightarrow P\left[\hat{p} - \tau_{\alpha/2} \cdot S.\hat{E}(\hat{p}) < p < \hat{p} + \tau_{\alpha/2} \cdot S.\hat{E}.(\hat{p})\right] = 1-\alpha.$$

4) At time $t=0$, 20 identical components are put on test. The life distn. of each is exponential with mean $\theta$, after 24 hours. We were found that 15 of the 20 components were still working. Derive the MLE of $\theta$. Also give an estimate of the S.E. of the estimator.

Hint:- Here $T \sim$ Exp with mean $\theta$

Let $Y$ = the no. of bulbs survived upto 24 hours out of 20 bulbs.

Clearly, $Y \sim \text{Bin}(n=20, p)$; where $p = P[T>24] = e^{-24/\theta}$,

MLE of $p$ is $\hat{p} = \frac{y}{n} = \frac{15}{20} = 0.75$

$$\Rightarrow e^{-24/\hat{\theta}} = \hat{p}$$

$$\Rightarrow \hat{\theta} = \frac{-24}{\ln(.75)} = \text{_____}$$

5) In a life testing experiment, 10 electric lamp are put to test. The lamps are burnt at a streach for 20 hours. 2 of this survive time to termination (in hours) of life. For the remaining lamps, are observed and given below

9.8, 15.8, 17.2, 11.2, 13.8, 18.9, 14.6, 19.6

find an estimate of the mean life $\theta$ assuming the life distn. to be exponential. If all the bulbs survive, what could have been your estimate of $\theta$.

Sol. $T \sim$ Exp. with mean $\theta$.

$$p = P[T>20] = e^{-20/\theta}$$

Let $x_1, \ldots, x_8$ be the lifetime of 8 lamps.

Likelihood function is $L(\theta) = \prod_{i=1}^{8} \left\{ \frac{1}{\theta} \cdot e^{-x_i/\theta} \right\} \cdot p^2$

$$= \frac{1}{\theta^8} \cdot e^{-\sum_{i=1}^{8} x_i/\theta} \cdot e^{-40/\theta} \cdot 2$$

$$= \frac{1}{\theta^8} \cdot e^{\frac{-8(\bar{x}+5)}{\theta}}$$

Likelihood equation is:- $0 = \frac{\partial}{\partial\theta} \ln L(\theta) = -\frac{8}{\theta} + \frac{8(\bar{x}+5)}{\theta^2}$

$$\Rightarrow \hat{\theta} = \bar{x}+5 = \text{_____}.$$

If all 10 bulbs survived, the likelihood equation is

$$L(\theta) = \left\{ e^{-20/\theta} \right\}^{10} = e^{-200/\theta}$$

$L(\theta)$ is max. iff $200/\theta$ is minimum.

iff $\theta$ is max.

Hence, MLE does not exist.

6) The following in the freq. distn. of 154 obsn drawn at random from a multinomial popln with 6 classes

| Class no | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| No. of observation in the sample | 79 | 32 | 5 | 6 | 17 | 15 |

denoting the probabilities of 6 classes by $\pi_1, \pi_2, \ldots, \pi_6$, respectively. Find the MLE of $\pi_1 - 2\pi_2 + 6\pi_5$.

Sol.→ Let $f_i$, $i=1(1)6$ be the frequency of the $i^{th}$ class in a r.s. of size $n=154$.

Likelihood function:-

$$L(\pi_i | f_i) = \frac{\lfloor n}{\prod_{i=1}^{n} \lfloor f_i} \cdot \prod_{i=1}^{6} \pi_i^{f_i}, \text{ where } \sum_{i=1}^{6} \pi_i = 1.$$

and $0 < \pi_i < 1$, $i=1(1)6$.

To maximise $\ln L$ subject to $\sum_{i=1}^{6} \pi_i = 1$

Let, $F = \ln L + \lambda \left( \sum_{i=1}^{n} \pi_i - 1 \right)$

$$= \text{constant} + \sum_{i=1}^{6} f_i \cdot \ln \pi_i + \lambda \left( \sum_{i=1}^{n} \pi_i - 1 \right)$$

Solve:- $0 = \frac{\partial F}{\partial \pi_i} = \frac{f_i}{\pi_i} + \lambda$

$$\Rightarrow \pi_i = -\frac{f_i}{\lambda}.$$

and $1 = \sum \pi_i = \frac{-\sum f_i}{\lambda} = -\frac{n}{\lambda}.$

$$\Rightarrow -\frac{1}{\lambda} = \frac{1}{n}$$

$$\Rightarrow \hat{\pi}_i = \frac{f_i}{n}, \text{ are the MLE's.}$$

$$\Rightarrow \underset{\sim}{\hat{\pi}} = \left( \frac{f_1}{n}, \frac{f_2}{n}, \ldots, \frac{f_6}{n} \right).$$

The MLE of $(\pi_1 - 2\pi_2 + 6\pi_5)$

$$= (1, -2, 0, 0, 6, 0) \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \\ \pi_6 \end{pmatrix}$$

$$= \underset{\sim}{\ell}' \underset{\sim}{\pi}$$

is $\underset{\sim}{\ell}' \underset{\sim}{\hat{\pi}} = \hat{\pi}_1 - 2\hat{\pi}_2 + 6\hat{\pi}_5$.

7) The IQ's of 10 teenagers are
98, 114, 105, 101, 123, 117, 106, 92, 110, 108, where as those of 6 teenagers belonging to other ethnic group are
122, 105, 95, 126, 114, 108.
Assuming that the data can be looked upon as a independent r.s. from Normal popln with mean $\mu_1$ and $\mu_2$ and common variance $\sigma^2$. Estimate the parameters by the method of MLE.

**Sol.** Let $X_1, X_2, \ldots, X_{10} \overset{iid}{\sim} N(\mu_1, \sigma^2)$
$Y_1, Y_2, \ldots, Y_6 \overset{iid}{\sim} N(\mu_2, \sigma^2)$ } independent.

Likelihood Equation:-

$$L(\mu_1, \mu_2, \sigma | x_i, y_j) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{16} \cdot \exp\left[-\frac{\Sigma(x_i-\mu_1)^2+\Sigma(y_j-\mu_2)^2}{2\sigma^2}\right]$$

Likelihood Equation:—

$$① \rightarrow 0 = \frac{\partial \ln L}{\partial \mu_1} = -\frac{2\Sigma(x_i-\mu_1)(-1)}{2\sigma^2}$$

$$\Rightarrow \mu_1 = \bar{x}.$$

$$② \rightarrow 0 = \frac{\partial \ln L}{\partial \mu_2} \Rightarrow \mu_2 = \bar{y}.$$

$$③ \rightarrow 0 = \frac{\partial \ln L}{\partial \sigma} = \frac{-16}{\sigma} + \frac{\Sigma(x_i-\mu_1)^2+\Sigma(y_j-\mu_2)^2}{\sigma^3}$$

$$\Rightarrow \hat{\sigma}^2 = \frac{\Sigma(x_i-\hat{\mu}_1)^2+\Sigma(y_j-\hat{\mu}_2)^2}{16} = \text{_____}.$$

8) Score obtained from 10 students in two parallel forms of a test of a test are recorded below:
Assuming that score in the two sets arise from a bivariate normal popln. with the common mean $\mu$, common s.d. $\sigma$ and the corrln. coeff. $\rho$, find the MLE of $\mu, \sigma, \rho$.

| 1st form | 2nd form |
|---|---|
| 41 | 47 |
| 52 | 54 |
| 49 | 50 |
| 44 | 48 |
| 32 | 40 |
| 62 | 65 |
| 40 | 38 |
| 45 | 42 |
| 39 | 45 |
| 48 | 45 |

**Sol.** Let $(X_i, Y_i)$, $i=1(1)10$, be a r.s. from $BN(\mu_x, \mu, \sigma^2, \sigma^2, \rho)$

Let $U_i = X_i + Y_i \overset{iid}{\sim} N(2\mu, 2\sigma^2(1+\rho)) = N(\mu_U, \sigma_U^2)$

$V_i = X_i - Y_i \overset{iid}{\sim} N(0, 2\sigma^2(1-\rho))$, $i=1(1)n$, indep.

$\equiv N(0, \sigma_V^2)$

**MLE's:-** (i) $\begin{cases} \hat{\mu}_U = \bar{u} \\ 2\hat{\mu} = \bar{u} = \bar{x} + \bar{y} \\ \Rightarrow \hat{\mu} = \dfrac{\bar{x}+\bar{y}}{2}. \end{cases}$

(ii) $\hat{\sigma}_U^2 = s_u^2 = \frac{1}{n}\sum (u_i - \bar{u})^2 = \frac{1}{n}\sum (x_i - \bar{x} + y_i - \bar{y})^2$

$= \frac{1}{n}\sum_i \left[ (x_i-\bar{x})^2 + (y_i-\bar{y})^2 + 2(x_i-\bar{x})(y_i-\bar{y}) \right]$

$\Rightarrow 2\hat{\sigma}^2 (1+\hat{\rho}) = s_x^2 + s_y^2 + 2s_{xy}.$

(iii) $\hat{\sigma}_V = s_v^2 = \frac{1}{n}\sum (v_i - \bar{v})^2 = \frac{1}{n}\sum_i \left( [x_i-\bar{x}]^2 + [y_i-\bar{y}]^2 - 2(x_i-\bar{x})(y_i-\bar{y}) \right)$

$\Rightarrow 2\hat{\sigma}^2(1-\rho) = s_x^2 + s_y^2 - 2s_{xy}$

(ii) + (iii) gives $4\hat{\sigma}^2 = 2(s_x^2 + s_y^2)$

$\Rightarrow \hat{\sigma}^2 = \dfrac{s_x^2 + s_y^2}{2} = \text{———}.$

(ii) − (iii) gives

$4\hat{\sigma}^2\hat{\rho} = 4s_{xy}$

$\Rightarrow \hat{\rho} = \dfrac{2s_{xy}}{s_x^2 + s_y^2} = \text{———}$

9. Suppose heights of 6 pairs of identical adult male bengali films, there observed that

5'6"; 5'5"; 6'1"; 5'11"; 5'7"; 5'8"; 6'2"; 6'; 5'4"; 5'3"; 5'5"; 5'3";

Suppose it is known that average height of adult bengali is 5'5" and s.d. is 3'3". Find the MLE of $\rho$.

Assuming that pairs of height follow a BN distn., with unknown corr. coeff. $\rho$.

**Sol.** → Assume that, $(X, Y) \sim BN(\mu, \mu, \sigma, \sigma, \rho)$

$\mu = 5'5"$

$\sigma = 3'3"$

and $\rho$ is unknown.

$\Rightarrow (U, V) = \left( \dfrac{x-\mu}{\sigma}, \dfrac{y-\mu}{\sigma} \right) \sim BN(0, 0, 1, 1, \rho)$

Likelihood function is :—

$$L(\rho/(u_i,v_i)) = \left\{\frac{1}{2\pi\sqrt{1-\rho^2}}\right\}^n \cdot e^{-\frac{1}{2(1-\rho^2)}\{\sum u_i^2+\sum v_i^2-2\rho\sum u_iv_i\}}$$

Likelihood equation:-

$$0 = \frac{\partial}{\partial\rho}\ln L = -\frac{n(-2\rho)}{2(1-\rho^2)} - \frac{1}{2(1-\rho^2)}\{-2\sum u_iv_i\} + \frac{\sum u_i^2+\sum v_i^2-2\rho\sum u_iv_i}{2(1-\rho^2)^2}(-2\rho)$$

$$\Rightarrow \frac{\rho}{(1-\rho^2)}\left\{n+\frac{1}{\rho}\sum u_iv_i - \frac{1}{1-\rho^2}(\sum u_i^2+\sum v_i^2-2\rho\sum u_iv_i)\right\}=0$$

$$\Rightarrow n\rho^3 - \rho^2\sum u_iv_i + \rho(\sum u_i^2+\sum v_i^2-n) - \sum u_iv_i = 0$$

$$\Rightarrow \rho^3 + a\rho^2 + b\rho + c = 0, \text{ say.}$$

Use some numerical method of solution.

10. Estimate the value of $\theta$ by the ML method and find out 95% C.I. on the basis of this sample

3·58, 2·86, 3·16, 2·46, 6·33, 8·31, 15·17, 9·59

drawn from the cauchy population —

$$DF(x) = \frac{1}{\pi\{1+(x_i-\theta)^2\}}, \quad \theta\in\mathbb{R}, \quad -\infty<x_i<\infty.$$

Sol. $$L(\theta/x_1,\ldots,x_n) = \frac{1}{\pi^n\cdot\prod_{i=1}^{n}\{1+(x_i-\theta)^2\}}, \quad \theta\in\mathbb{R}.$$

Likelihood equation:-

$$0 = \frac{\partial}{\partial\theta}\ln L = \sum_{i=1}^{n}\frac{2(x_i-\theta)}{1+(x_i-\theta)^2} = f(\theta), \text{ say.}$$

By NR method,

$$\theta_{n+1} = \theta_n - \left\{\frac{f(\theta)}{f'(\theta)}\right\}_{\theta=\theta_n}$$

$$= \theta_n + \left\{\frac{\frac{\partial\ln L}{\partial\theta}}{-\frac{\partial^2\ln L}{\partial\theta^2}}\right\}_{\theta=\theta_n}$$

$$\simeq \theta_n + \left\{\frac{\frac{\partial\ln L}{\partial\theta}}{I_n(\theta)}\right\}_{\theta=\theta_n}$$ , replacing $-\frac{\partial^2\ln L}{\partial\theta^2}$ by its expectation $I_n(\theta)$.

Here $I_n(\theta) = \frac{n}{2}$.

$$\theta_{n+1} = \theta_n + \frac{4}{n}\sum_{i=1}^{4}\frac{(x_i-\theta_n)}{1+(x_i-\theta_n)^2}$$

Trial root $\theta_1 = \tilde{x}$ = the sample median = _____.

$\theta_2$ = _____

$\theta_3$ = _____

etc. $\hat{\theta}$ = _____.

For large sample,

$$\hat{\theta} \overset{a}{\sim} N\left(\theta, \frac{1}{I_n(\theta)}\right)$$

$$\Rightarrow \hat{\theta} \overset{a}{\sim} N\left(\theta, \frac{2}{n}\right)$$

$$\Rightarrow 1-\alpha = P\left[\left|\frac{\hat{\theta}-\theta}{\sqrt{\frac{2}{n}}}\right| < \tau_{\alpha/2}\right]$$

$$= P\left[\hat{\theta} - \sqrt{\frac{2}{n}}\tau_{\alpha/2} < \theta < \hat{\theta} + \sqrt{\frac{2}{n}}\tau_{\alpha/2}\right]$$

11. To test a null hypothesis $H_0: p = 0.6$ of a popln. $\sim B(m=5, p)$ vs. $H_1: p > 0.6$. Suggest an UMP test with exact size $\alpha$, the choice of $\alpha$ is left to you. If the observed no. of success is 4, construct a randomised test of exact size 0.1 and conclude.

Sol. To test $H_0: p = 0.6$ vs. $H_1: p > 0.6$

Let $x$ be an observation from $B(m=5, p)$

MP test of its size of testing $H_0: p = p_0$ Vs. $H_1: p = p_1$, $p_1 > p_0 = 0.6$ is

$$\phi(x) = \begin{cases} 1 & \text{if } \frac{f(x, p_1)}{f(x, p_0)} > c \\ 0 & \text{ow} \end{cases}$$

$$= \begin{cases} 1 & , x > k \\ 0 & , \text{ow} \end{cases}$$

where, $P_{H_0}[\phi(x)] = \alpha$ (say)

and the test $\phi(x)$ is independent of $p_1$ $(> p_0)$.

Hence, the test $\phi(x)$ is UMP of $H_0: p = 0.6$ vs. $H_1: p > 0.6$ of its size.

Let $k = 4$,

Then $\phi(x) = \begin{cases} 1 & , x > 4 \\ 0 & , \text{ow} \end{cases}$

is a UMP test of size $= E_{H_0}[\phi(X)] = P_{H_0}[X > 4]$

$$= \binom{5}{5}(0.6)^5(0.4)^0$$

$$= 0.07776$$

If $\phi(x) = \begin{cases} 1 & , x > 3 \\ 0 & , \text{OW} \end{cases}$

then size $= P_{H_0}[X > 3]$

$= \binom{5}{4}(0.6)^4(0.4)^1 + (0.6)^5$

$= 0.2592 + 0.07776 > 0.1$

To get the exact size 0.1, it is required to randomize at $x = 4$.

Let $\phi(x) = \begin{cases} 1 & , x > 4 \\ \gamma & , x = 4 \\ 0 & , \text{OW} \end{cases}$

where $\gamma$ is such that $0.1 = E_{H_0}[\phi(X)]$

$= 1 \cdot P_{H_0}[X > 4] + \gamma \cdot P_{H_0}[X = 4]$

$= 0.07776 + \gamma(0.2592)$

$\Rightarrow \gamma = 0.086$

$\therefore \phi(x) = \begin{cases} 1, & x > 4 \\ 0.086, & x = 4 \\ 0, & x < 4 \end{cases}$

is the UMP test at level $\alpha = 0.1$.

Here $x = 4$ is the observed value.

Hence, we reject $H_0$ with probability $\gamma = 0.086$ and accept $H_0$ with prob. $1 - \gamma$.

Draco a 3-digited random number, then the prob. of $R = \{$the selected no. $\leq 0.85\}$ is $P(R) = \frac{86}{1000}$.

Let the selected no. be 126.

Then we accept $H_0: p = 0.6$ at level 0.1.

12. Suppose the waiting time denoted by $x$ in minutes for a bus is uniformly distributed over $U[0, \theta]$. To test the hypothesis $H_0: \theta = 10$ vs. $H_1: \theta > 10$ on the basis of a sample of size 6 to decision rules are proposed. Reject $H_0$ if (i) $\max\{x_1, \ldots, x_6\} > c_1$

(ii) (no. of $\{x_1, \ldots, x_6\} > 8) > c_2$

Find the values of $c_1$ and $c_2$ taking level of significance is 0.05. Also, draw power curve for both these procedure and comment on the relative performance.

**Sol.** Let $X_1, \ldots, X_6 \overset{iid}{\sim} R(0, \theta)$

To test $H_0: \theta = 10$ Vs. $H_1: \theta \neq 10$

(a) $\phi_1(x) = \begin{cases} 1 & \text{if } x_{(6)} > c_1 \\ 0 & \text{ow} \end{cases}$

(b) $\phi_2(x) = \begin{cases} 1 & \text{if } y > c_2 \\ 0 & \text{ow} \end{cases}$

where $Y$ = the no. of $x_i$'s which are greater than 8.

$\therefore Y \sim Bin(6, p)$.

$\therefore p = \int_8^{\theta} \frac{1}{\theta} d\theta = \left(1 - \frac{8}{\theta}\right) = P[X_i > 8]$

**Power function:-**

(a) $\beta_1(\theta) = P_\theta [X_{(6)} > c_1]$

$= 1 - P_\theta [X_{(6)} \leq c_1]$

$= 1 - \left(\frac{c_1}{\theta}\right)^6$.

(b) $\beta_2(\theta) = P_\theta [Y > c_2]$

$= \sum_{y=c_2+1}^{6} \binom{6}{y} \left(1 - \frac{8}{\theta}\right)^y \left(\frac{8}{\theta}\right)^{6-y}$

Here $\alpha = 0.05$,

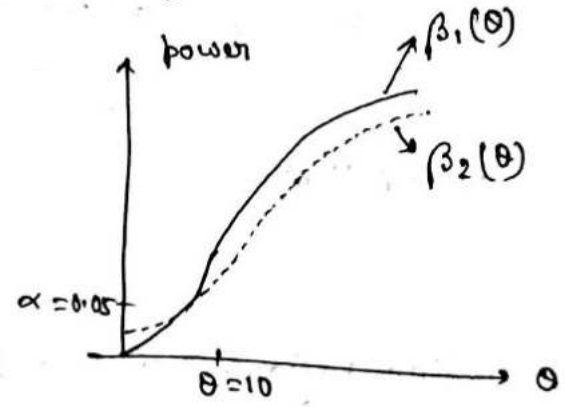
$\therefore 0.05 = \beta_1(\theta)\Big|_{\theta=10} = 1 - \left(\frac{c_1}{10}\right)^6$

$\Rightarrow c_1 = 10 \cdot (0.95)^{1/6}$.

and $0.05 = \beta_2(\theta)\Big|_{\theta=10}$

$= \sum_{y=c_2+1}^{6} \binom{6}{y} \left(\frac{1}{5}\right)^y \left(\frac{4}{5}\right)^{6-y}$

**Comment:-** the power of the test (a) which is based on exact $x_i$'s is greater than that of the test (b) which is based on Bernoulli trial.



## — : Non-Parametric Inference :-

13. The following are 15 measurements of the octane rating of a certain kind of Gasoline

97.5, 95.2, 97.3, 96.0, 96.8, 100.3, 97.4; 95.3, 93.2, 99.1, 96.1, 97.6, 98.2, 98.5, 94.9.

Use an exact non-parametric test and also an approximate test to examine whether or not the average octane rating of the given kind of Gasoline is 98.5.

Sol. X : the measurement of the octane of gasoline.

Let $x_1, x_2, \ldots, x_n$, $n = 15$ be given random samples.

To test $H_0: \xi_{1/2} = 98.5$ Vs. $H_1: \xi_{1/2} \neq 98.5$

(a) Exact Non-Parametric Test :- count the no. of +ve signs among $x_1 - \xi_0, \ldots, x_n - \xi_0$.

We have two +ve sign and 12 -ve sign and one observation $x_{14}$ is $= 98.5$. We ignore the value $x_{14}$ from the sample.

Now, we have a r.s. of size $n = 14$.

Let $Y$ = the no. of +ve signs among $(x_i - \xi_{y_0})$'s, $i = 1(1)14$.

Under $H_0$, $Y \sim \text{Bin}(n = 14, p = \frac{1}{2})$.

The observed value of $Y$ is $y_0 = 2$.

The p-value $= 2 \cdot \min\{P_{H_0}[Y \geq y_0], P_{H_0}[Y \leq y_0]\}$

$= 2 \cdot P_{H_0}[Y \leq y_0]$

$= 2 \cdot \sum_{y=0}^{2} \binom{14}{y} \cdot \frac{1}{2^{14}} =$ ———.

If p-value $< 0.05$, we reject $H_0$, at 5% level of significance.

(b) Assume that the distn. of X is $N(\mu, \sigma^2)$, whether the actual distn. is normal or not.

Here, $\xi_{1/2} = \mu$.

To test $H_0: \mu = 98.5$ Vs. $H_1: \mu \neq 98.5$

t-test :- $t = \frac{\sqrt{n}(\bar{x} - 98.5)}{s} \sim t_{n-1}$, under $H_0$.

$n = 15$.

Computation :-

Critical region :- Observed $|t| > t_{\alpha/2, 14}$.

The test procedure described is an approximate test procedure, as the actual distn. of x may not be normal.

14. The following table shows Hamilton depression scale factor measurements in 9 patients suffering from depression, taken before (X) and after (Y) a visit to therapist :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| X | 1·83 | 0·50 | 1·62 | 2·48 | 1·68 | 1·88 | 1·55 | 3·06 | 1·3 |
| Y | ·878 | ·647 | ·598 | 2·05 | 1·06 | 1·29 | 1·06 | 3·14 | 1·29 |

Perform a suitable (a) Parametric (b) non parametric test to judge whether the therapy can be considered to be effective.

Sol. (a) Parametric:-

Assume that $(X,Y) \sim BN(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$

To judge whether the therapy is considered to be effective, i.e.

$H_0: \mu_x = \mu_y$ alternative $H_1: \mu_x > \mu_y$.

This is a paired-t testing problem.

Define $d_i = x_i - y_i \overset{iid}{\sim} N(\mu_d, \sigma_d^2)$,

where $\mu_d = \mu_x - \mu_y$.

Test statistic $\frac{\sqrt{n}(\bar{d} - 0)}{s_d} \sim t_{n-1}$, under $H_0$.

Computation:-

| | 1 | 2 | ... | 9 |
|---|---|---|---|---|
| $d_i$ | | | | |

$\bar{d} =$ _____

$s_d^2 = \frac{1}{n-1}\{\sum d_i^2 - n\bar{d}^2\}$

= _____.

Critical region:- Observed $t > t_{\alpha, 8}$.

(b) Non-Parametric:-

Here $d_i = x_i - y_i$, $i = 1(1)9$.

And we are interested in testing $H_0$: median$(d) = 0$ Vs. $H_1$: median$(d) > 0$

Test:- $H_0: \xi_{1/2}(d) = 0$ Vs. $H_1: \xi_{1/2}(d) > 0$

We shall use sign test based on the values $d_1, d_2, \ldots, d_9$ as a r.s. from an absolutely continuous univariate distn.

Let $Z$ = the no. of +ve $d_i$'s

$\sim Bin(n=9, p=1/2)$, under $H_0$.

The observed value of $Z$ is $Z_0 =$ _____.

p-value = $P_{H_0}[Z \geq Z_0]$ = _____.

15. Conduct the parametric and non-parametric test for differences of location for the data given below;

| | | | | | |
|---|---|---|---|---|---|
| loop I : | 9 | 7 | 12 | 16 | 14 |
| loop II : | 5 | 2 | 8 | 4 | 20 |

**Sol.** Let $X_1, \ldots, X_m$, $m=5$ and $Y_1, \ldots, Y_n$, $n=5$ be the given r.s. from gr-I, gr-II, respectively.

**Parametric test:-** (Fisher's t-test)

**Assumption:-** $x_1, \ldots, x_m \overset{iid}{\sim} N(\mu_x, \sigma_x^2)$, $y_1, \ldots, y_n \overset{iid}{\sim} N(\mu_y, \sigma_y^2)$ } independently distributed

Assume that $\sigma_x = \sigma_y$

To test: $H_0: \mu_x = \mu_y$ Vs. $H_1: \mu_x \neq \mu_y$

**Fisher's t-statistic:-**

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim t_{m+n-2}.$$

**Non-parametric test:-** (Median Test)

Assume that the distn. of the two groups are absolutely continuous and independent.

To test $H_0: \xi_{1/2}(x) = \xi_{1/2}(y)$ Vs. $H_1: \xi_{1/2}(x) \neq \xi_{1/2}(y)$

Let $\tilde{Z}$ be the median in the combined data set.

Define, $V$ = the no. of $x_i$'s which are $\leq \tilde{Z}$

Under $H_0$, $P[V=v] = \dfrac{\binom{m}{v}\binom{n}{p-v}}{\binom{m+n}{p}}$, $v = 0(1)n$

$m+n = 10 = 2p$

$\therefore p = 5$.

Combine the data in an increasing order, we have

2, 4, 5, 7, 9, 12, 14, 16, 20.

$\tilde{Z} = \dfrac{8+9}{2} = 8.5$.

The observed of $V = 1 = v$

$\therefore$ p-value $= 2 \min\{P_{H_0}[V \geq v_0], P_{H_0}[V \leq v_0]\}$

$= 2\, P_{H_0}[V \leq v_0 = 1]$

$= 2 \cdot \dfrac{\binom{5}{v}\binom{n}{5-v}}{\binom{10}{5}}$

# ANOVA

1. The following measurements refer to the no. of hours in which 9 patients are free from pain after taking placebo, a new drug and aspirin:

| Group | Observations | | | |
|---|---|---|---|---|
| Placebo | 0.0 | 1.0 | | |
| New drug | 2.3 | 3.5 | 2.8 | 2.5 |
| Aspirin | 3.1 | 2.7 | 3.8 | |

(i) Test whether the new drug is more effective than the others
(ii) Test if the effects of the new drug is the same as the average effect of the other two.

Sol. The factor considered here is drug (A) with levels Placebo ($A_1$), New drug ($A_2$) and Aspirin ($A_3$). The data given is one-way classified data. Let $y_{ij}$ denotes the $j^{th}$ observations in the $i^{th}$ level or group, $i=1,2,3$, $j=1(1)n_i$; $n_1=2$, $n_2=4$, $n_3=3$.

Model [Fixed Effects] :-

$y_{ij} = \mu + \alpha_i + e_{ij}$ with $\sum n_i \alpha_i = 0$.

where $e_{ij} \overset{iid}{\sim} N(0, \sigma_e^2)$, $i=1(1)3$, $j=1(1)n_i$.

Here $\mu_i = \mu + \alpha_i$ is mean-effect (fixed) of the $i^{th}$ level.

To test $H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$.

Computation:-

$$G_1 = \sum_{i=1}^{3} \sum_{j=1}^{n_i} y_{ij} = \text{_____}.$$

$$C.F. = \frac{G^2}{n} = \text{_____} \quad \text{with } n = 9.$$

$$SS(\text{Total}) = \sum\sum y_{ij}^2 - CF = \text{_____}.$$

$$SS(\text{between}) = \sum_{i=1}^{3} \frac{T_{i0}^2}{n_i} - CF = \text{_____}$$

$$= \frac{T_{10}^2}{n_1} + \frac{T_{20}^2}{n_2} + \frac{T_{30}^2}{n_3} - CF = \text{_____}.$$

where $T_{i0} = \sum_{j=1}^{n_i} y_{ij}$

ANOVA Table:-

| Source of Variation | d.f. | SS | MS | F | |
|---|---|---|---|---|---|
| Between groups. | $3-1=2$ | $SS$ (between) $=$ | MSB | $F = \frac{MSB}{MSE}$ | $F_{0.05;\,2,6}$ $=$ _____ |
| Within groups [Error] | $9-3=6$ | $SSE =$ ___ | MSE | | |
| Total = | $9-1=8$ | $SS(\text{total})$ $=$ | | | |

If observed $F > F_{0.05;2,6}$, reject $H_0$.

(i) To test $H_{01}: \mu_2 = \mu_1$ Vs. $H_{01}': \mu_2 > \mu_1$

and $H_{02}: \mu_2 = \mu_3$ Vs $H_{02}': \mu_2 > \mu_3$.

To test $H_{01}$, the test statistic

$$\frac{\bar{y}_{20} - \bar{y}_{10}}{\sqrt{MSE\left(\frac{1}{n_2} + \frac{1}{n_1}\right)}} \sim t_6, \text{ under } H_{01}.$$

Reject $H_{01}$, if observed $t > t_{\alpha;6}$.

To test $H_{02}$,

the test statistic is $t_1 = \dfrac{\bar{y}_{20} - \bar{y}_{30}}{\sqrt{MSE\left(\frac{1}{n_2} + \frac{1}{n_3}\right)}}$

(ii) To test $H_{03}: \mu_2 = \dfrac{\mu_1 + \mu_3}{2}$.

$\Leftrightarrow H_{03}: \mu_1 - 2\mu_2 + \mu_3 = 0$

Vs. $H_{03}': \mu_1 - 2\mu_2 + \mu_3 \neq 0$.

Test statistic $(t) = \dfrac{\bar{y}_{10} - 2\bar{y}_{20} + \bar{y}_{30}}{\sqrt{MSE\left(\frac{1}{n_1} + \frac{4}{n_2} + \frac{1}{n_3}\right)}} \sim t_6$, under $H_{03}$.

Critical region: observed $|t| > t_{\alpha/2,6}$.

2. 4 randomly selected typiest work on each of 4 given type writers of different make and speeds were recorded coise of typiest and typewritter.

| Type-writter | Typiest | | | |
|---|---|---|---|---|
| 1 | 30, 32 | 40, 30 | 25, 40 | 20, 30 |
| 2 | 40, 35 | 32, 25 | 20, 30 | 31, 21 |
| 3 | 20, 31 | 25, 25 | 26, 30 | 25, 21 |
| 4 | 27, 28 | 31, 24 | 30, | 19, 25 |

Analyse the data and give estimate of the parameters.

Sol. Type writter [factor A] : $A_1, A_2, A_3, A_4$

Typists [factor B] : $B_1, B_2, B_3, B_4$

The effects of different levels of A are fixed. As 4 typists are selected randomly, there effects are random, the data given is a two-way classification data with 2 obs.p. per cell. Let $Y_{ijk}$ be the $k^{th}$ obsn. corresponding to the $i^{th}$ level of A and $j^{th}$ level of B.

$k = 1(1)m = 2$, $i = 1(1)p = 4$; $j = 1(1)q = 4$.

**Model:-** [Two way classified data with mixed effects]

$$y_{ijk} = \mu + \alpha_i + b_j + c_{ij} + e_{ijk}$$

(fixed) (random)

where $\sum_{i=1}^{p} \alpha_i = 0$, $\sum_{i=1}^{p} c_{ij} = 0$, $j = 1(1)q$

and $\{b_j\}$, $\{c_{ij}\}$ and $\{e_{ijk}\}$ are jointly normal,

$$e_{ijk} \overset{iid}{\sim} N(0, \sigma_e^2).$$

Define, $\sigma_A^2 = \frac{1}{p-1} \sum_{i=1}^{p} \alpha_i^2$

$$\sigma_B^2 = Var(b_j)$$

$$\sigma_{AB}^2 = \frac{1}{p-1} \sum_{i=1}^{p} V(c_{ij})$$

To test $H_0: \sigma_A^2 = 0$ and to estimate the parameters

$$\sigma_e^2, \sigma_B^2, \sigma_{AB}^2.$$

**Computation:-** $G_1 = \sum_i \sum_j \sum_k y_{ijk} =$ _____

C.F. $G_1^2/n =$ _____

$n = pqm = 32.$

$$SS(Total) = \sum_i \sum_j \sum_k y_{ijk}^2 - C.F.$$

Defined, $T_{i00} = \sum_j \sum_k y_{ijk}$

$$T_{0j0} = \sum_i \sum_k y_{ijk}$$

$$T_{ij0} = \sum_k y_{ijk}.$$

| A \ B | $B_1$ | $B_2$ | $B_3$ | $B_4$ | Total |
|---|---|---|---|---|---|
| $A_1$ | $T_{110} =$ ___ | $T_{120} =$ ___ | $T_{130} =$ ___ | $T_{140} =$ ___ | $T_{100} =$ ___ |
| $A_2$ | $T_{210} =$ ___ | $T_{220} =$ ___ | | | $T_{200} =$ ___ |
| $A_3$ | | | | | |
| $A_4$ | | | | | |
| | $T_{010} =$ | $T_{020} =$ | | | $G_1 =$ ___ |

$$SSA = \sum_{i=1}^{p} \frac{T_{i00}^2}{qm} - CF = \text{_____}$$

$$SSB = \sum_{j} \frac{T_{0j0}^2}{pm} - CF = \text{_____}$$

$$SS(AB) = \sum_i \sum_j \frac{T_{ij0}^2}{m} - \sum_i \frac{T_{i00}^2}{qm} - \sum_j \frac{T_{0j0}^2}{pm} + C.F.$$

$$= \left( \sum_i \sum_j \frac{T_{ij0}^2}{m} - CF \right) - SSA - SSB.$$

ANOVA Table

| Source of Variation | d.f. | SS | MS | E(MS) |
|---|---|---|---|---|
| Due to A (fixed) | $p-1=3$ | SSA | $MSA = \frac{SSA}{3}$ | $\sigma_e^2 + m\sigma_{AB}^2 + qm\sigma_A^2$ |
| Due to B (random | $q-1=3$ | SSB | $MSB = \frac{SSB}{3}$ | $\sigma_e^2 + pm\sigma_B^2$ |
| Due to AXB | $(p-1)(q-1)=9$ | SS(AB) | $MS(AB) = \frac{SS(AB)}{9}$ | $\sigma_e^2 + m\sigma_{AB}^2$ |
| Error | $pq(m-1)=16$ | SSE | $MSE = \frac{SSE}{16}$ | $\sigma_e^2 =$ ____ |
| Total | $pqm-1=31$ | SST | | |

$$F_A = \frac{MSA}{MS(AXB)}.$$

If observed $F > F\alpha; 3, 9$, reject $H_A: \sigma_A^2 = 0$.

[ If $H_A$ is rejected, give $\hat{\alpha}_i = \bar{y}_{i00} - \bar{y}_{000} =$ _____ ]

The estimates of the parameters $\hat{\sigma}_e^2 = MSE =$ _____

$$\hat{\sigma}_{AB}^2 = \frac{MS(AB) - MSE}{m} = \text{_____}$$

$$\hat{\sigma}_B^2 = \frac{MSB - MSE}{pm} = \text{_____}.$$

9. In an experiment on yield of sugar beets (tons/acre) there were two levels of irrigation treatment and three of fertilizer treatment and each combination of treatments was carried out in 5 replicants. The following values of sum of squares (SS) were obtained : SS(Irrigation) = 120.0 ; SS(Fertilizer) = 221.7 ; SS(Interaction) = 35.0 ; SS(Error) = 108.0 ; Assuming that the irrigation & fertilizer effects are random, estimate the components of variance. Also, test whether the variance of the irrigation component is zero.

Sol. Here two effects are : Irrigation (A) : $A_1$ $A_2$

Fertilizer (B) : $B_1$ $B_2$ $B_3$

Both the effects are random.

It is two way classified data with m=5 obsn. per cell.

$p = 2, q = 3, m = 5.$

Model:- $y_{ijk} = \mu + a_i + b_j + c_{ij} + e_{ijk}$, where

$a_i \overset{iid}{\sim} N(0, \sigma_A^2)$

$b_j \overset{iid}{\sim} N(0, \sigma_B^2)$

$c_{ij} \overset{iid}{\sim} N(0, \sigma_{AB}^2)$

$e_{ijk} \overset{iid}{\sim} N(0, \sigma_e^2)$ are independent.

Note that, $\sigma^2 = V(y_{ijk}) = \sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma_e^2$.

ANOVA MODEL:-

| Source of variation | Df | SS | MS | E(MS) |
|---|---|---|---|---|
| due to A | 2−1 = 1 | 120 | 120 | $\sigma_e^2 + m\sigma_{AB}^2 + qm\sigma_A^2$ |
| due to B | 3−1 = 2 | 221.7 | 110.85 | $\sigma_e^2 + m\sigma_{AB}^2 + pm\sigma_B^2$ |
| due to AB | 2 | 35 | 17.5 | $\sigma_e^2 + m\sigma_{AB}^2$ |
| Error | 24 = pq(m−1) | 108 | 4.5 | $\sigma_e^2 = 4.5$ |
| Total | 29 = pqm−1 | | | |

Here $\hat{\sigma}_e^2 = MSE = 4.5$

$\hat{\sigma}_{AB} = \frac{MS(AB) - MSE}{m} = \frac{17.5 - 4.5}{5} = 2.6$

$\hat{\sigma}_A^2 = \frac{MSA - MS(AB)}{qm} = \frac{120 - 17.5}{15} = $ ____

$\sigma_B^2 = \frac{MSB - MS(AB)}{pm} = \frac{110.85 - 17.5}{10} = $ ____.

Reject $H_0: \sigma_A^2 = 0$ if observed $F = \frac{MSA}{MS(AB)} > F_{\alpha;1,2}$.

## DESIGN

1. In the experiment described below 4 materials were tested in each of 4 runs on a machine with 4 different position. The letters A to D refer to the 4 materials. The layout of the expt. is given below. Here the figures denote the loss in weight in a run of standard length.

| Run | Position in machine 4 | 2 | 1 | 3 |
|---|---|---|---|---|
| 2 | A(251) | B(241) | D(227) | C(229) |
| 3 | D(234) | C(237) | A(274) | B(226) |
| 1 | C(235) | D(236) | B(218) | A(268) |
| 4 | B(195) | A(270) | C(230) | D(225) |

(a) Analyse the data and comment.
(b) If the variation due to the different position of the machine is ignored, will you modify your conclusion?

**Sol.** Factor A (Row): Factor runs are the four levels A
Factor B (Column): Four positions of the machine are the four levels of B.
Treatment: A, B, C, D.

(a) Experiment is conducted according to the LSD with four treatments A, B, C, D.

Let $y_{ijk}$ be the obsn. of the $k^{th}$ treatment in the $(i,j)^{th}$ cell.

**Model:-**
$$y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + e_{ijk}$$
with $\sum \alpha_i = \sum \beta_j = \sum \tau_k = 0$,
$$e_{ijk} \overset{iid}{\sim} N(0, \sigma_e^2)$$

To test: $H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$.

**Computation:-**
$$G = \sum_i \sum_j \sum_k y_{ijk} = \text{_____}$$
$$C.F. = G^2/n, \quad n = 4.$$

$T_{i00}$ → row totals
$T_{0j0}$ → column totals
$T_{00k}$ → treatment totals

$$SS(\text{row}) = \sum_{i=1}^{m} \frac{T_{i00}^2}{m} = \text{_____}. - CF$$
$$SS(\text{column}) = \sum_{j=1}^{m} \frac{T_{0j0}^2}{m} - CF$$
$$SS(\text{treatment}) = \sum_k \frac{T_{00k}^2}{m} - CF$$

ANOVA Table

| Source of Variation | d.f. | SS | MS | |
|---|---|---|---|---|
| row | $m-1=3$ | SSR | | $F = \frac{MS(tr)}{MSE}$ |
| column | $=3$ | SSC | | |
| Treatment | $=3$ | SS(tr) | MS(tr) | |
| Error | $6$ | — | MSE | |
| Total | $= m^2-1=15$ | SS(Total) | | |

If observed $F > F_{\alpha;3,6}$, we reject $H_0$.
If $H_0$ is rejected, then, the different treatment has different effects in general.

(b) If the variation due to different position of the machine is ignored that is columns are ignored, then corresponding 4 rows as 4 blocks, the design of experiment reduced to RBD.

Let $y_{ik}$ be the observation on the $k^{th}$ treatment in the $i^{th}$ block (row).

Model:- $y_{ik} = \mu + \alpha_i + \tau_k + e_{ik'}$

where, $e_{ik'} \overset{iid}{\sim} N(0, \sigma_{e'}^2)$.

$$\sum_i \alpha_i = 0 = \sum_k \tau_k$$

To test $H_0: \tau_k = 0$, $k = 1(1)4$.

Computation:- SS(Block) = SS(row)

SS(treatment) in RBD is same as the SS(tr) in LSD.

But SS(column) is added to SSE of LSD to get the SSE* in RBD.

ANOVA table

| Source of Variation | d.f. | SS. | |
|---|---|---|---|
| Row | $m-1=3$ | SSR | $F^* = \frac{MS(tr)}{MSE^*}$ |
| Treatment | 3 | SS(tr) | |
| Error | $3+6=9$ | $SSE^* = SSE + SS(column)$ | |
| Total | $m^2-1=15$ | SS(total) | |

If observed $F^* > F_{0.05;3,9}$, reject $H_0$.

2) In order to compare the hardness of alloy three furnaces (F) and three levels of moulds (M) were tried. The lay-out as well as the hardness (in suitable unit) are shown below:

| | Rep-I | | | Rep II | | | Rep III | | |
|---|---|---|---|---|---|---|---|---|---|
| $F_1$ | $M_3$ 156 | $M_2$ 118 | $M_1$ 140 | $M_2$ 104 | $M_1$ 89 | $M_3$ 117 | $M_1$ 103 | $M_2$ 126 | $M_3$ 149 |
| $F_2$ | $M_1$ 130 | $M_3$ 174 | $M_2$ 157 | $M_3$ 112 | $M_1$ 89 | $M_2$ 81 | $M_3$ 149 | $M_1$ 124 | $M_2$ 129 |
| $F_3$ | $M_1$ 114 | $M_2$ 161 | $M_3$ 141 | $M_1$ 103 | $M_2$ 132 | $M_3$ 133 | $M_3$ 100 | $M_1$ 91 | $M_2$ 97 |

Analyse the data.

**Sol.** **Split Plot Design:-**

**factor A:** $(A_1, A_2, \ldots, A_p)$

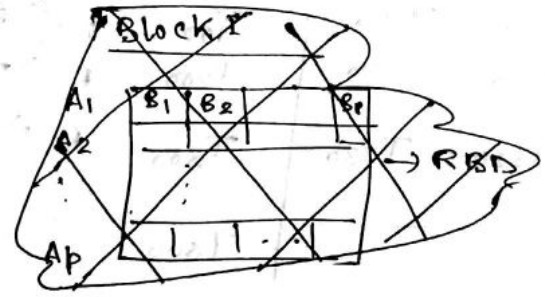
↳ they exclude use of small plots

i.e. they are applied on large plots.

→ Their effects are known to be different / effect of $A_i$'s are to be tested with less precision.

**Another factor:-** $B_1, B_2, \ldots, B_q$.

→ they are applicable small plots and we want test B and AXB more accurately than A.




F denoted by $F_1, F_2, F_3$, M denoted by $M_1, M_2, M_3$.
Here $p=3$, levels of F arranged is a R.B.D. using $r=3$ blocks or replicable and the factor M at $q=3$ levels, are applied to the plots of a block after subdividing each plot into $q=3$ subplots. This is a split plot design.

**Model:-** $$y_{ijk} = \underbrace{\{\mu + b_i + \tau_j + e_{ij}\}}_{\text{model for whole plot treatment F by RBD}} + \underbrace{\gamma_k + \delta_{jk} + e_{ijk}'}_{\text{model for subplot treatment in whole plot.}}$$

$i=1(1)r=3$, $j=1(1)p=3$, $k=1(1)q=3$.

$$\sum \tau_j = \sum \gamma_k = \sum_k \delta_{jk} = \sum_j \delta_{jk} = 0$$

I

| | | | |
|---|---|---|---|
| $F_3$ | $M_3$ | $M_2$ | $M_1$ |
| $F_1$ | $M_1$ | $M_3$ | $M_2$ |
| $F_2$ | | | |

where, $b_i, e_{ij}, e_{ijk}$ are independently normal $(0, \sigma_b^2; \sigma_e^2; \sigma_{e'}^2)$.

To test $H_{01}: \tau_j = 0$

$H_{02}: \gamma_k = 0$

$H_{03}: \delta_{jk} = 0$

Computation:—

$G_1 = \sum_i \sum_j \sum_k y_{ijk} =$ ——

$C.F. = G^2 / rpq$

$SS(total) = \sum\sum\sum y_{ijk} - C.F. =$ ——

Whole Plot Analysis:—

$$SS(Block) = \sum_{i=1}^{r} \frac{T_{i00}^2}{pq} - C.F.$$

$$SS(F) = \sum \frac{T_{0j0}^2}{rq} - C.F.$$

$$SSE_I = \left\{ \sum_i \sum_j \frac{T_{ij0}^2}{pr} - C.F. \right\} - SSB - SSF$$

Whole Plot analysis Table:—

| Replicate \ F | $F_1$ | $F_2$ | $F_3$ | Total | |
|---|---|---|---|---|---|
| I | $T_{110} = 461$ | $T_{120}$ | $T_{130}$ | $T_{100}$ | R-I) $F_1(M_1+M_2+M_3) = T_{110}$ |
| II | $T_{210}$ | $T_{220}$ | $T_{230}$ | $T_{200}$ | R-II) $F_1(M_1+M_2+M_3) = T_{210}$ |
| III | $T_{310}$ | $T_{320}$ | $T_{330}$ | $T_{300}$ | R-I) $F_2(M_1+M_2+M_3) = T_{120}$ |
| | | | | $T_{000}$ | |

Split plot Analysis:—

$$SS(M) = \sum \frac{T_{00k}^2}{rp} - CF$$

$$SS(FXM) = \left( \sum \frac{T_{0jk}^2}{r} - CF \right) - SSF - SSM$$

Table for SS (FXM)

| F \ M | $M_1$ | $M_2$ | $M_3$ | |
|---|---|---|---|---|
| $F_1$ | $T_{011}$ | $T_{012}$ | $T_{013}$ | $T_{010}$ |
| $F_2$ | $T_{021}$ | $T_{022}$ | $T_{023}$ | $T_{020}$ |
| $F_3$ | $T_{031}$ | $T_{032}$ | $T_{033}$ | $T_{030}$ |
| | $T_{001}$ | $T_{002}$ | $T_{003}$ | $T_{000}$ |

$F_1(M_1+M_1+M_1) = T_{011}$ under $R_1$-I, II, III

$F_1(M_2+M_2+M_2) = T_{012}$ under R-I, II, III.

## ANOVA Table

| Source of Variation | df | SS | MS | F |
|---|---|---|---|---|
| Replicates | $n-1=2$ | | | |
| whole plot treat (F) | $p-1=2$ | | $MS(F)$ | $F_1=\frac{MSF}{MSE_I}$ |
| Error (I) | $(n-1)(p-1)=9$ | | $MS(E_I)$ | |
| Subplot treat (M) | $q-1=2$ | | $MS(M)$ | $F_2=\frac{MSM}{MSE_{II}}$ |
| Interaction (FXM) | $(p-1)(q-1)=4$ | | $MS(FXM)$ | $F_3=\frac{MS(FXM)}{MSE_{II}}$ |
| Error (II) | $p(q-1)(n-1)=12$ | | $MS(E_{II})$ | |
| Total | $df=npq-1=26$ | | | |

$F_1 \leftrightarrow F\ \alpha; 2, 4$

$F_2 \leftrightarrow F\ \alpha; 2, 12$

$F_3 \leftrightarrow F\ \alpha; 4, 12$

# SAMPLE SURVEY

(1). (a) Draw a random sample of size 7 from an exp. popln with mean 2.345.

(b) Draw a r.s. of size 5 from a Cauchy popln. with median 0 and scale 2.

(c) Draw a r.s. of size 6 from a univariate normal popln. with mean 17.95 and s.d. 6.23.

(d) Draw a r.s. of size 5 from the distn

$P[X=0] = \frac{1}{5}$, $P[X=1] = \frac{2}{5}$, $P[X=2] = \frac{2}{5}$.

**Solution:-** (a) Let $X \sim Exp(\theta = 2.345)$

$$\left[ f_X(x) = \frac{1}{\theta} e^{-x/\theta} \right.$$

$$\left. F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = 1 - e^{-x/\theta} \right]$$

D.F. of X is

$$F(x) = 1 - e^{-x/\theta}, \quad x > 0$$

By probability integral transformation,

$$U = F(X) \sim R(0,1).$$

If u is an observed sample from R(0,1), then $U = F(x)$

$$\Rightarrow u = 1 - e^{-x/\theta}.$$

$$\Rightarrow x = -\ln(1-u)$$

$= -2.345 \ln(1-u)$ is an observed sample from Exp. with mean $\theta = 2.345$.

We take seven 3-digited random number from Fisher-Yates table.

Page No: 125, row = 5, column = 4

260, 573, 375, 204, 056, 930, 001.

We place decimal point before the selected nos.

Then 0.260, 0.573, 0.375, 0.204, 0.056, 0.930, 0.001 are 7 r.s. from R(0,1).

| Serial No. | $u_i$ | $x_i = -2.345 \ln(1-u_i)$ |
|---|---|---|
| 1 | | |
| 2 | | |
| 3 | | |
| 4 | | |
| 5 | | |
| 6 | | |
| 7 | | |

(b) $\left[ f_X(x) = \frac{\sigma}{\pi\{\sigma^2 + (x-\mu)^2\}} \right.$

$\Rightarrow F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$

$= \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x-\mu}{\sigma}\right)$. Here $\mu = 0, \sigma = 2.$ $\left.\right]$

$F(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x}{2}\right)$

where $X \sim R(0,1)$

Here $u = F(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x}{2}\right)$.

$\Rightarrow x = 2\tan\{\pi(u - \frac{1}{2})\}$

$= 2\tan x\{\pi u - \pi/2\}$

$= -2\cot \pi u$.

(c) $X \sim N(17.95, (6.23)^2)$

$\Rightarrow Z = \frac{X - 17.95}{6.23} \sim N(0,1)$

Hence, $U = \Phi\left(\frac{x - 17.95}{6.23}\right) \sim R(0,1)$

$\Rightarrow u = \Phi\left(\frac{x - 17.95}{6.23}\right)$

$\Rightarrow \Phi^{-1}(u) = \frac{x - 17.95}{6.23}$ → From Biometrica.

(d) Draw a r.s. of size 5 from the distn

$P[X=0] = \frac{1}{5}$, $P[X=1] = \frac{2}{5}$, $P[X=2] = \frac{2}{5}$.

Let us define the r.v. $Y \sim U(0,1)$.

So, $P[0 < Y < .2] = .2 = P[X=0]$

$P[.2 \le Y < .6] = .4 = P[X=1]$

$P[.6 \le Y < 1] = .4 = P[X=2]$

So, we take 5 3-digited random no. If the no. selected is in between (000 − 199) then we consider is equivalent to choosing $X=0$, if the no. is in between (200 − 599), we equivalently choose $X=1$, and finally, if the selected no. is in between (600-999), then we choose $X=2$.

(2). For the cost function a) $C = C_0 + \sum C_h n_h$

b) $C = C_0 + \sum C_h \sqrt{n_h}$.

where $C_0$ and $C_1$ are known constants. Find the optimum values of $n_h$'s by minimizing $\text{Var}(\bar{y}_{st})$ for fixed total cost with total sample size being 60, given the following data:

| Stratum | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $N_h$ | 30 | 40 | 60 | 70 |
| $S_h$ | 1.5 | 2.0 | 3.5 | 4.0 |
| $C_h$ | 1 | 2 | 3 | 4 |

Also, compute the optimum allocation due to Neyman & compare the efficiency of the optimum allocations with that of proportional allocation to estimate the population mean $\bar{Y}$.

Solution:- (a) Define, $F = \frac{1}{N^2} \sum N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) S_h^2 + \lambda\left(C_0 + \sum_{h=1}^{L} C_h n_h - C\right)$

Now, $\frac{\partial F}{\partial n_h} = 0$

$\Rightarrow \frac{1}{N^2} N_h^2 S_h^2 \left(-\frac{1}{n_h^2}\right) + \lambda \cdot C_h = 0$

$\Rightarrow n_h = \lambda_1 \frac{N_h \cdot S_h}{\sqrt{C_h}}$

Given $\sum_h n_h = 60$

$\Rightarrow \lambda_1 = \frac{60}{\sum_h \frac{N_h S_h}{\sqrt{C_h}}} = \text{——}$.

| Stratum | $N_h$ | $S_h$ | $C_h$ | $\frac{N_h S_h}{\sqrt{C_h}}$ | $n_h = \lambda_1 \cdot \frac{N_h S_h}{\sqrt{C_h}}$ |
|---|---|---|---|---|---|
| 1 | | | | | |
| 2 | | | | | |
| 3 | | | | | |
| 4 | | | | | |

(b) Define $F = \frac{1}{N} \sum_{h=1}^{L} N_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right) \cdot S_h^2 + \lambda\left(C_0 + \sum_{h=1}^{L} C_h n_h^{1/2} - C\right)$

Now, $0 = \frac{\partial F}{\partial n_h}$

$\Rightarrow \frac{1}{N^2} \cdot N_h^2 \cdot S_h^2 \left(-\frac{1}{n_h^2}\right) + \lambda \cdot \frac{C_h}{2\sqrt{n_h}} = 0$

$\Rightarrow n_h = \lambda_1^* \left(\frac{N_h S_h}{\sqrt{C_h}}\right)^{4/3}$

Given $\sum_h n_h = 60 \Rightarrow \lambda_1^* = \frac{60}{\sum_h \left(\frac{N_h S_h}{\sqrt{C_h}}\right)^{4/3}} = \text{——}$.

| Stratum | $N_h$ | $S_h$ | $C_h$ | $\frac{N_h - S_h}{\sqrt{C_h}}$ | $n_h = \lambda \cdot \frac{N_h S_h}{\sqrt{C_h}}$ | $\left(\frac{N_h S_h}{\sqrt{C_h}}\right)^{4/3}$ | $n_h = \lambda_1^* \left(\frac{N_h S_h}{\sqrt{C_h}}\right)$ |
|---|---|---|---|---|---|---|---|
| | | | | | | | |

## Neyman's optimum and proportional allocations:-

Here $n_h \propto \frac{N_h S_h}{\sqrt{C_h}}$.

(1) If $C_h$ = constant, then $n_h = \lambda_2 N_h S_h$ and $60 = \sum n_h$.

$\Rightarrow \lambda_2 = \frac{60}{\sum N_h S_h}$.

$\therefore n_h = \left(\frac{60}{\sum N_h S_h}\right) N_h S_h$.

$\longrightarrow$ optimum allocation.

(2) Here $C_h$ = constant

$S_h$ = constant

$n_h = \lambda_3 \cdot N_h$

and $\lambda_3 = \frac{60}{\sum N_h} = \frac{60}{N}$.

$\therefore n_h = \left(\frac{60}{N}\right) N_h \rightarrow$ proportional allocation.

(3). Ques:- Using the information given below, find the relative standard error that one could expect if a sample of 1% of the villagers group of district the steps be selected by SRSWOR for estimating the total population Y of the region.

| District Serial No. | No. of Villages ($N_h$) | Average popln for village ($Y_h$) | Standard deviation ($S_h$) |
|---|---|---|---|
| 1 | 1953 | 487 | 564 |
| 2 | 1864 | 829 | 931 |
| 3 | 1381 | 822 | 996 |
| 4 | 1174 | 1083 | 1167 |
| 5 | 531 | 1956 | 1940 |
| 6 | 1391 | 664 | 625 |
| 7 | 1996 | 456 | 779 |
| 8 | 1951 | 372 | 856 |
| 9 | 3369 | 339 | 591 |

Solution:- Relative standard Error = $RSE(\hat{\bar{Y}}_{st})$

$= \frac{S.E(\hat{\bar{Y}}_{st})}{\hat{\bar{Y}}_{st}} = \frac{N \cdot SE(\hat{\bar{Y}}_{st})}{N \cdot \hat{\bar{Y}}_{st}}$

$= \frac{\left\{\frac{1}{N^2} \sum N_h^2 S_h^2 \left(\frac{1}{n_h} - \frac{1}{N_h}\right)\right\}^{1/2}}{\frac{\sum N_h \cdot \bar{Y}_h}{N}}$.

$$= \frac{\sqrt{\sum N_h \left(N_h - \frac{N_h}{100}\right) \cdot \frac{S_h^2}{\frac{N_h}{100}}}}{\sum N_h \bar{Y}_h} \quad \left[\text{Here } \frac{n_h}{N_h} = \frac{1}{100}\right]$$

$$= \frac{\sqrt{99 \sum N_h S_h^2}}{\sum N_h \bar{Y}_h} = \text{——}$$

| District | $N_h$ | $\bar{Y}_h$ | $S_h$ | $N_h \bar{Y}_h$ | $N_h S_h^2$ |
|---|---|---|---|---|---|
| 1 | | | | | |
| 2 | | | | | |
| . | | | | | |
| . | | | | | |
| . | | | | | |
| 9 | | | | | |

(3) All the farms of a country have been stratified according to size and the following data have been obtained.

| Farm size | No. of Farm ($N_h$) | Mean ($\bar{Y}_h$) | S.D. ($S_h$) |
|---|---|---|---|
| 0-40 | 394 | 5.4 | 8.3 |
| 41-80 | 464 | 16.3 | 13.3 |
| 81-120 | 391 | 24.3 | 15.1 |
| 121-160 | 334 | 34.5 | 19.8 |
| 161-200 | 169 | 42.1 | 24.5 |
| 201-240 | 113 | 50.1 | 26.0 |
| 241-280 | 148 | 63.8 | 35.2 |

If the first 3 stratum combine into a 1 stratum and last 4 into another. Find the relative loss in efficiency as compared to 7 strata situation in estimating the popⁿ mean by stratified Random sampling (a) Proportional allocation (b) optimum allocation that the total no. of farms are selected 100.

Sol. In the new stratified popln

| Farm Size | No. of firms | $\bar{Y}_h'$ | $S_h'^2$ |
|---|---|---|---|
| 0-120 | $N_1' = \sum_{h=1}^{3} h_n$ | | |
| 121-280 | $N_2' = \sum_{h=4}^{7} h$ | | |

where $\bar{Y}_1' = \frac{\sum_{h=1}^{3} N_h \bar{Y}_h}{\sum_{h=1}^{3} N_h}$, $\bar{Y}_2' = \frac{\sum_{h=4}^{7} N_h \bar{Y}_h}{\sum_{h=4}^{7} N_h}$

and $S_1'^2 = \frac{1}{\sum_{h=1}^{3} N_h - 1} \left\{ \sum_{h=1}^{3} (N_h - 1) S_h^2 + \sum_{h=1}^{3} N_h (\bar{Y}_h - \bar{Y}_1')^2 \right\}$

and $S_2'^2 = \frac{1}{\sum_{h=4}^{7} N_h - 1} \left\{ \sum_{h=4}^{7} (N_h - 1) S_h^2 + \sum_{h=4}^{7} N_h (\bar{Y}_h - \bar{Y}_2')^2 \right\}$

(a) $V_{prop} = \frac{1-f}{n} \sum_{h=1}^{7} \frac{N_h}{N} \cdot S_h^2$ (Given stratification)

and $V_{prop}' = \frac{1-f}{n} \sum_{h=1}^{2} \frac{N_h'}{N} S_h'^2$ (New stratification)

Loss in eff $= \frac{V_{prop}' - V_{prop}}{V_{prop}} =$ ____.

(b) $V_{opt} = \frac{\left( \sum_{h=1}^{7} \frac{N_h}{N} \cdot S_h \right)^2}{n} - \frac{\sum_{h=1}^{7} \frac{N_h}{N} \cdot S_h^2}{N}$

$V_{opt}' = \frac{\sum_{h=1}^{2} \frac{N_h'}{N} \cdot S_h'^2}{n} - \frac{\sum_{h=1}^{2} \frac{N_h'}{N} \cdot S_h'^2}{N}$.

Loss in eff $= \frac{V_{opt}' - V_{opt}}{V_{opt}} =$ ____.

(4) (Ratio - Regression Estimator)

An experimenter makes an farmer's eye-estimate of the weight of peaches on each tree in a orchard of 200 trees. He finds a total weight of 11600 lbs and weight for a SRS of 10 trees which yield the following result;

| Serial No. of tree : | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Actual weight : | 61 | 42 | 50 | 58 | 67 | 45 | 39 | 57 | 71 | 53 |
| Estimated is : | 59 | 47 | 52 | 60 | 67 | 48 | 44 | 58 | 76 | 58 |

Compute the ratio and regression estimates of the total actual weight of Peaches of all the 200 trees in the orchard and compare the precisions of 2 estimates.

Solution:- $Y$ = total weight, $X$ = an eye estimate.

To estimate the "the weight of Peaches ($y$)" of all the 200 trees. Here auxiliary information " an eye estimate of weight of Peaches ($x$) " is given.

$$X = 11600 \text{ lbs.}$$

Data: $(x_i, y_i)$, $i = 1(1)n$.

(a) Ratio estimation:- Ratio estimate $\hat{Y}_R = \frac{\bar{y}}{\bar{x}} \cdot X = \frac{\bar{y}}{\bar{x}} \cdot X$

where $\bar{y} = \frac{1}{n}\sum y_i = $ ____, $\bar{x} = \frac{1}{n}\sum x_i = $ ____.

$\Rightarrow \hat{Y}_R =$ ____.

$$MSE(\hat{Y}_R) \simeq N^2 \left(\frac{1}{n} - \frac{1}{N}\right)\left(s_y^2 + s_x^2 - 2\hat{R}s_{xy}\right)$$

where $s_x^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2$, $s_y^2 = \frac{1}{n-1}\sum (y_i - \bar{y})^2$,

$s_{xy} = \frac{1}{n-1}\sum (x_i - \bar{x})(y_i - \bar{y})$, $\hat{R} = \frac{\bar{y}}{\bar{x}} =$ ____.

(b) Regression Estimation:-

Regression estimate, $\hat{Y}_{lr} = N.\hat{\bar{Y}}_{lr} = N\{\bar{y} + b(\bar{X} - \bar{x})\}$

$= N\bar{y} + b(X - N\bar{x})$

$\therefore =$ ____.

$b = \frac{s_{xy}}{s_x^2} =$ ____,

$MSE(\hat{Y}_{lr}) \simeq N^2\left(\frac{1-f}{n}\right)s_e^2$; $s_e^2 = s_y^2(1-\rho^2) = s_{y.x}^2$

where $s_{y.x}^2 = \frac{1}{n-2}\sum_i (y_i - \bar{y} - b(x_i - \bar{x}))^2$

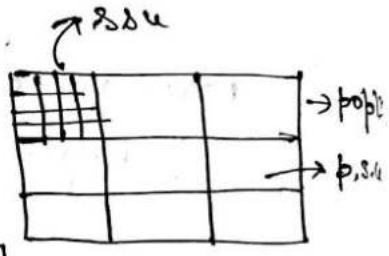
$= \frac{1}{n-2}\{\sum (y_i - \bar{y})^2 - b^2 \sum (x_i - \bar{x})^2\}$

$= \frac{n-1}{n-2}\{s_y^2 - b^2 s_x^2\}$

$= s_R^2$.

Comparison:- $Eff(\text{regression ratio}) = \frac{\hat{MSE}(\hat{Y}_R)}{\hat{MSE}(\hat{Y}_{lr})} =$ ____.

(5). Two-stage-cluster :→ In an experimental investigation, 100 fields each consisting of 16 plots of equal size, were sown with wheat. Out of 100 fields, 10 fields are selected by SRSWOR and out of each field so selected 4 plots are selected by WOR to observe the yield. From the given observation following values are estimated. Sample mean (in kg) for the selected fields are:
4·290, 4·255, 3·795, 4·220, 4·070, 3·636, 4·550, 4·285, 4·375, 3·790. Sample average with variance = ·018215 $(kg)^2$.
Estimate the total yield of wheat in the experimental station along with its standard error.
Compare the efficiency of this estimate with on that would have been obtained by selecting an SRSWOR of 40 plots out of 1600 plots in the station.




Sol.: [ First we select n p.s.u from N. p.s.u. From selected n p.s.u., with in each p.s.u. have M units, we select m units. ]

N = 100 fields (p.s.u.), each consisting of M = 16 plots (s.s.u.). Out of N p.s.u., n = 10 fields and out of each selected (p.s.u.) fields m = 4 plots (s.s.u.) are selected under SRSWOR.

Let $y_{ij}$ be the value obtained for the $j^{th}$ selected s.s.u. in the $i^{th}$ selected p.s.u., $i = 1(1)10 = n$; $j = 1(1)4 = m$.

An unbiased estimate of Y (total yield) is

$$\hat{Y} = N.\bar{\bar{y}} = N\left(\frac{1}{n}\sum_{i=1}^{n}\bar{y}_i\right), \quad \bar{y}_i = \frac{1}{m}\sum_{j=1}^{m} y_{ij}.$$

$$= \underline{\qquad\qquad}.$$

$$V(\hat{Y}) = (MN)^2\left\{\frac{1-f_1}{n}.S_n^2 + \frac{1-f_2}{mn}.S_w^2\right\}$$

and $$\hat{V}(\hat{\bar{y}}) = (MN)^2\left\{\frac{1-f_1}{n}.S_b^2 + \frac{f_1(1-f_2)}{mn}.S_w^2\right\}$$

where $$S_b^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{y}_i - \bar{\bar{y}})^2 = \underline{\qquad\qquad}$$

and $$S_w^2 = \frac{1}{n(m-1)}\sum_{i=1}^{n}\sum_{j=1}^{m}(y_{ij} - \bar{y}_i)^2 = \underline{\qquad}$$

$$\hat{V}(\hat{Y}) = \underline{\qquad\qquad}.$$

In case of SRSWOR, with sample of size
$mn = 40 = n'$. from $MN = 1600 = N'$ plots.

$$V_{SRS}(\hat{\bar{Y}}) = N'^2 \cdot \frac{N'-n'}{N'n'} \cdot S_y^2,$$

$$S_y^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} (y_{ij} - \bar{\bar{y}})^2}{nm-1}.$$

$$= \frac{\sum_i \sum_j (y_{ij} - \bar{y}_i)^2 + m\sum_{i=1}^{n} (\bar{y}_i - \bar{\bar{y}})^2}{nm-1}.$$

$$= \frac{n(m-1)\cdot S_w^2 + m(n-1)\cdot S_b^2}{mn-1}.$$

$$= \text{\_\_\_\_\_}.$$