# UNIT - 2

## ② CURVE FITTING & REGRESSION.

* Curve fitting:

① Straight line: $y = a + bx$

$y = a + bx$ [Straight line]

2 normal Equations:

$$\sum_{i=1}^{n} y_i = na + b\sum_{i=1}^{n} x_i \quad \text{—— (i)}$$

$$\sum_{i=1}^{n} x_i y_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 \quad \text{—— (ii)}$$

Solving (i) & (ii) we get values of a & b. substitute these values in $y = a + bx$ to get best fitting curve in least square sense.

② Parabola & a quadratic curve:

$y = a + bx + cx^2$

normal equations:

$$\sum_{i=1}^{\infty} y_i = na + b\sum_{i=1}^{\infty} x_i + c\sum_{i=1}^{\infty} x_i^2 \quad \text{—— (i)}$$

$$\sum_{i=1}^{\infty} x_i y_i = a\sum_{i=1}^{\infty} x_i + b\sum_{i=1}^{\infty} x_i^2 + c\sum_{i=1}^{\infty} x_i^3 \quad \text{—— (ii)}$$

$$\sum_{i=1}^{n} x_i^2 y_i = a\sum_{i=1}^{\infty} x_i^2 + b\sum_{i=1}^{\infty} x_i^3 + c\sum_{i=1}^{\infty} x_i^4 \quad \text{—— (iii)}$$

Solving (i), (ii) & (iii) we get values of a, b & c. substitute these values in $y = a + bx + cx^2$ to get best fitting curve in least square sense.

### Problems:

① Obtain a straight line $y = a + bx$ fitting the given data in the least square sense.

| y | 12 | 15 | 21 | 25 |
|---|---|---|---|---|
| x | 50 | 70 | 100 | 120 |

Sol: The normal eq are:

$$\sum_{i=1}^{4} y_i = na + b\sum_{i=1}^{4} x_i$$

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

here $n = 4$.

$$\sum y_i = 4a + b\sum_{i=1}^{4} x_i$$

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

| | $x_i$ | $y_i$ | $x_i y_i$ | $x_i^2$ |
|---|---|---|---|---|
| | 50 | 12 | 600 | 2500 |
| | 70 | 15 | 1050 | 4900 |
| | 100 | 21 | 2100 | 10000 |
| | 120 | 25 | 3000 | 14400 |
| Total | 340 | 73 | 6750 | 31800 |

$\sum y_i = 4a + b(x_i)$

$\Rightarrow 73 = 4a + b(340)$

$\Rightarrow 4a + 340b = 73$ —— (ii)

$\sum x_i y_i = a\sum x_i + b\sum x_i^2$

$6750 = a(340) + b(31800)$

$\Rightarrow 340a + 31800b = 6750$ —— (iii)

Solving (i) & (iii)

$\boxed{a = 2.2758}$ $\boxed{b = 0.1879}$

$\therefore$ Best fitting straight line is

$y = 2.2758 + 0.1879x$.

$(y = a + bx)$.

---

(2) If P is the pull required to lift a load W. Find a linear law of the form $P = a + bW$ connecting P & W using the following data.

| P | 12 | 15 | 21 | 25 |
|---|---|---|---|---|
| W | 50 | 70 | 100 | 120 |

Soln Same solution as (1).

(3) ~~Fo~~ Fit the best parabola in the least square sense for the following data

| x | 2 | 4 | 6 | 8 | 10 |
|---|---|---|---|---|---|
| y | 8.07 | 12.84 | 31.47 | 57.38 | 91.29 |

Soln Here, $n = 5$.

To fit the curve, $y = a + bx + cx^2$

normal equations,

$$\sum_{i=1}^{5} y_i = 5a + b\sum_{i=1}^{5} x_i + c\sum_{i=1}^{5} x_i^2$$

$$\sum_{i=1}^{5} x_i y_i = a\sum_{i=1}^{5} x_i + b\sum_{i=1}^{5} x_i^2 + c\sum_{i=1}^{5} x_i^3$$

$$\sum_{i=1}^{5} x_i^2 y_i = a\sum_{i=1}^{5} x_i^2 + b\sum_{i=1}^{5} x_i^3 + c\sum_{i=1}^{5} x_i^4$$

| | $x_i$ | $y_i$ | $x_i y_i$ | $x_i^2$ | $x_i^3$ | $x_i^4$ |
|---|---|---|---|---|---|---|
| | 2 | 3.07 | 6.14 | 4 | 8 | 16 |
| | 4 | 12.84 | 51.36 | 16 | 64 | 256 |
| | 6 | 31.47 | 188.82 | 36 | 216 | 1296 |
| | 8 | 57.38 | 459.04 | 64 | 512 | 4096 |
| | 10 | 91.29 | 912.9 | 100 | 1000 | 10000 |
| Total | 30 | 196.05 | 1618.26 | 220 | 1800 | 15664 |

| $x_i^2 y_i$ |
|---|
| 12.28 |
| 205.44 |
| 1132.92 |
| 3672.32 |
| 9129 |
| Total 14151.96 |

$\Rightarrow 196.05 = 5a + b(30) + c(220)$

$\Rightarrow 5a + 30b + 220c = 196.05$ —— (i)

$1618.26 = a(30) + b(220) + c(1800)$

$\Rightarrow 30a + 220b + 1800c = 1618.26$ —— (ii)

$14151.96 = a(220) + b(1800) + c(15664)$

$\Rightarrow 220a + 1800b + 15664c = 14151.96$ —— (iii)

Equating eq (i), (ii) & (iii) we get

$a = 0.696$

$b = 0.8567$

$c = 0.9921$

∴ Best fitting curve:

$y = a + bx + cx^2$

$= 0.696 + (-0.8567)x + 0.9921x^2$

$y = 0.696 - 0.8567x + 0.9921x^2$

(4) Fit in 2nd degree curve for the following data.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $x$ | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |
| $y$ | 1.1 | 1.3 | 1.6 | 2.0 | 2.7 | 3.4 | 4.1 |

Sol$^n$ Here $n = 7$

To fit the curve, $y = a + bx + cx^2$

normal equations,

$$\sum_{i=1}^{7} y_i = na + b\sum_{i=1}^{7} x_i + c\sum_{i=1}^{7} x_i^2$$

$$\sum_{i=1}^{7} x_i y_i = a\sum_{i=1}^{7} x_i + b\sum_{i=1}^{7} x_i^2 + c\sum_{i=1}^{7} x_i^3$$

$$\sum_{i=1}^{7} x_i^2 y_i = a\sum_{i=1}^{7} x_i^2 + b\sum_{i=1}^{7} x_i^3 + c\sum_{i=1}^{7} x_i^4.$$

| $x_i$ | $y_i$ | $x_i^2$ | $x_i^3$ | $x_i^4$ | $x_i y_i$ | $x_i^2 y_i$ |
|---|---|---|---|---|---|---|
| 1 | 1.1 | 1 | 1 | 1 | 1.1 | 1.1 |
| 1.5 | 1.3 | 2.25 | 3.375 | 5.0625 | 1.95 | 2.925 |
| 2 | 1.6 | 4 | 8 | 16. | 3.2 | 6.4 |
| 2.5 | 2.0 | 6.25 | 15.625 | 39.0625 | 5. | 12.5 |
| 3. | 2.7 | 9 | 27 | 81. | 8.1 | 24.3 |
| 3.5 | 3.4 | 12.25 | 42.875 | 150.0625 | 11.9 | 41.65 |
| 4. | 4.1 | 16 | 64 | 256. | 16.4 | 65.6 |
| 17.5 | 16.2 | 50.75 | 161.875 | 548.1875 | 47.65 | 154.475 |

$\Rightarrow 16.2 = 7a + b(17.5) + c(50.75)$

$47.65 = a(17.5) + b(50.75) + c(161.875)$

$154.475 = a(50.75) + b(161.875) + c(548.1875)$

equating (i), (ii) & (iii)

$a = 1.0357$

$b = -0.1928$

$c = 0.2428$

$\therefore$ Best fitting curve:

$y = a + bx + cx^2$

$y = 1.0357 - 0.1928x + 0.2428x^2$

③ If velocity & resistance related by $R = a + bV^2$ find the values of a & b with method of least squares.

| V | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| R | 8 | 10 | 15 | 21 | 30 |

Sol$^n$ Here $n = 5$.

To fit the curve $R = a + bV^2$

normal equations are, $x_i = V_i$ $y_i = R_i$

$$\sum_{i=1}^{5} y_i = na + \underbrace{b}_{c} \sum_{i=1}^{5} x_i^2 + 0 \quad b\sum_{i=1}^{5} x_i$$

$$\sum_{i=1}^{5} x_i^2 y_i = a\sum_{i=1}^{5} x_i^2 + \underbrace{b}_{c} \sum_{i=1}^{5} x_i^4 + 0 \quad b\sum_{i=1}^{5} x_i^3$$

| $x_i$ | $y_i$ | $x_i^2$ | $x_i^3$ | $x_i^4$ | $x_i^2 y_i$ |
|---|---|---|---|---|---|
| 10 | 8 | 100. | 1000. | 10000. | 800. |
| 20 | 10 | 400 | 8000. | 160000. | 4000 |
| 30 | 15 | 900 | 27000 | 810000. | 135000. |
| 40 | 21 | 1600 | 64000. | 2560000. | 336000. |
| 50 | 30. | 2500 | 125000 | 6250000. | 75000. |
| 150 | 84. | 5500. | | 9790000. | 126900. |

$\sum R_i = na + b\sum V_i^2 \Rightarrow 84 = 5a + b5500.$

$\sum V_i^2 R_i = a\sum V_i^2 + b\sum V_i^4 \Rightarrow 126900 = a5500 + b(9790000)$

equating (ii) equations;

$a = 6.6529$

$b = 0.0092$

$\therefore$ Best fitting curve,

$P = a + bv^2$

$\Rightarrow P = 6.6529 + 0.0092 v^2$

(6) Fit a parabola to the 2nd degree curve

| x | 5 | 10 | 15 | 20 | 25 |
|---|---|---|---|---|---|
| y | 21 | 24 | 33 | 42 | 38 |

Soln here $n = 5$

To fit the curve $y = a + bx + cx^2$

normal equations are

$\sum y_i = na + b\sum x_i + c\sum x_i^2$

$\sum x_i y_i = a\sum x_i + b\sum x_i^2 + c\sum x_i^3$

$\sum x_i^2 y_i = a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4$

| $x_i$ | $y_i$ | $x_i^2$ | $x_i^3$ | $x_i^4$ | $x_i y_i$ | $x_i^2 y_i$ |
|---|---|---|---|---|---|---|
| 5 | 21 | 25 | 125 | 625 | 105 | 525 |
| 10 | 24 | 100 | 1000 | 10000 | 240 | 2400 |
| 15 | 33 | 225 | 3375 | 50625 | 495 | 7425 |
| 20 | 42 | 400 | 8000 | 160000 | 840 | 16800 |
| 25 | 38 | 625 | 15625 | 390625 | 950 | 23750 |
| tot 75 | 158 | 1375 | 28125 | 611875 | 2630 | 50900 |

equation (i) (ii) & (iii).

$\Rightarrow 158 = 5a + (75)b + c(1375)$

$158 = 5a + 75b + 1375c$

$\Rightarrow 2630 = a(75) + b(1375) + c(28125)$

$2630 = 75a + 1375b + 28125c$

$\Rightarrow 50900 = a(1375) + b(28125) + c(611875)$

$50900 = 1375a + 28125b + 611875c$

$a = 9$

$b = 2.24$

$c = -0.04$

$\therefore$ best fitting curve,

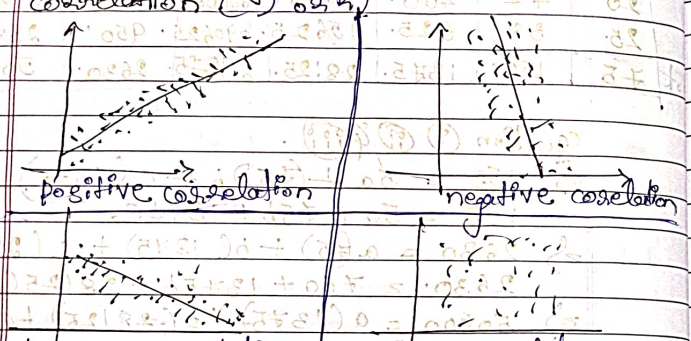
$y = a + bx + cx^2$

$\Rightarrow y = 9 + 2.24x + (-0.04)x^2$

$\therefore y = 9 + 2.24x - 0.04x^2$

## ② CORRELATION & REGRESSION

* The degree of variation is called correlation (CR or R)




$$\boxed{\rho = r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}} \text{ where,}$$

$X = x - \bar{x}$ & $Y = y - \bar{y}$

$$\boxed{\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}}$$

$$\boxed{\bar{y} = \frac{y_1 + y_2 + \cdots + y_n}{n}}$$

* Problems:-

① Psycological test of intelligiences & of engineering ability were applied to 10 students. Here is a record of ungrouped data showing intelligence ratio (IR) & engineering ratio (ER). calculate coefficent of relation.

| Student | IR | ER. |
|---|---|---|
| A | 105 | 101 |
| B | 104 | 103 |
| C | 102 | 100 |
| D | 101 | 98 |
| E | 100 | 95 |
| F | 99 | 96 |
| G | 98 | 104 |
| H | 96 | 92 |
| I | 93 | 97 |
| J | 92 | 94 |

Sol: WKT $\rho = r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 \Sigma Y^2}}$ where $X = x - \bar{x}$, $Y = y - \bar{y}$.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots x_n}{n}$$

$$\bar{y} = \frac{y_1 + y_2 + y_3 + \cdots y_n}{n}$$

(last column)

| xy | Student | TR(x) | ER(y) | $X = x - \bar{x}$ | $y = y - \bar{y}$ | $x^2$ | $y^2$ |
|---|---|---|---|---|---|---|---|
| 18 | A | 105 | 101 | 6 | 3 | 36 | 9 |
| 25 | B | 104 | 103 | 5 | 5 | 25 | 25 |
| 6 | C | 102 | 100 | 3 | 2 | 9 | 4 |
| 0 | D | 101 | 98 | 2 | 0 | 4 | 0 |
| -3 | E | 100 | 95 | 1 | -3 | 1 | 9 |
| 0 | F | 99 | 96 | 0 | -2 | 0 | 4 |
| -6 | G | 98 | 104 | -1 | 6 | 1 | 36 |
| 18 | H | 96 | 92 | -3 | -6 | 9 | 36 |
| 6 | I | 93 | 97 | -6 | -1 | 36 | 1 |
| 28 | J | 92 | 94 | -7 | -4 | 49 | 16 |
| Total | 10 | 990 | 980 | | | 170 | 140 |
| 92 | | | | | | | |

$\bar{x} = \frac{990}{10} = 99.$ $\boxed{\bar{x} = 99}$

$\bar{y} = \frac{980}{10} = 98$ $\boxed{\bar{y} = 98}$

$x = x - \bar{x}$ for each column:
$105 - 99 = 6$. (do same till last (J))

$y = y - \bar{y}$ for each column:
$101 - 98 = 3$ (do same till last (J))

$xy = 6 \times 3 = 18$ (Same till J)

$\rho = \frac{92}{\sqrt{170 \times 140}} = 0.5963.$

$\therefore \boxed{\rho = 0.5963.}$

(1) Note: r is always $\boxed{-1 \le r \le 1}$

(2) Find the coefficient relation of industrial production & export using following data & comment on the result.

| Production | 55 | 56 | 58 | 59 | 60 | 60 | 62 |
|---|---|---|---|---|---|---|---|
| Export | 35 | 38 | 38 | 39 | 44 | 43 | 45 |

Soln

| Production (x) | Export (y) | X | y | $x^2$ | $y^2$ | xy |
|---|---|---|---|---|---|---|
| 55 | 35 | -3.5714 | -5.2857 | 12.7548 | 27.9386 | 18.8773 |
| 56 | 38 | -2.5714 | -2.2857 | 6.6120 | 5.2244 | 5.8771 |
| 58 | 38 | -0.5714 | -2.2857 | 0.3264 | 5.2244 | 1.3060 |
| 59 | 39 | 0.4286 | -1.2857 | 0.1836 | 1.6530 | -0.5510 |
| 60 | 44 | 1.4286 | 3.7143 | 2.0408 | 13.7960 | 5.3062 |
| 60 | 43 | 1.4286 | 2.7143 | 2.0408 | 7.3674 | 3.8776 |
| 62 | 45 | 3.4286 | 4.7143 | 11.7552 | 22.2246 | 16.1634 |
| Total 410 | 282 | | | 35.7136 | 83.4284 | 50.8569 |

$\bar{x} = \frac{410}{7} = 58.5714.$

$\bar{y} = \frac{282}{7} = 40.2857$

$\rho = \frac{50.8569}{\sqrt{35.7136 \times 83.4284}} = 0.9317.$

$\therefore \boxed{\rho = 0.9317.}$

## Lines of regression:

If the dots in the scattered diagram tend to cluster around a line, then the relationship between the variable is linear & such line is called line of regression.

(i) Line of regression of $y$ on $x$ is,

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

→ Where $r$ is the correlation coefficient

(ii) Line of regression of $x$ on $y$ is

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

$\frac{r\sigma_y}{\sigma_x}$ & $\frac{r\sigma_x}{\sigma_y}$ are regression coefficient of $y$ on $x$ & $x$ on $y$ respectively.

(*) Remark:- the correlation coefficient $r$ is geometric mean between 2 regression coefficient since $\boxed{\frac{r\sigma_x}{\sigma_y} \times \frac{r\sigma_y}{\sigma_x} = r^2}$

$$\Rightarrow \boxed{r = \sqrt{\frac{r\sigma_x}{\sigma_y} \times \frac{r\sigma_y}{\sigma_x}}}$$

(*) Problems:

(1) The regression equation of 2 variables $x$ & $y$ are $x = 0.7y + 5.2$ & $y = 0.3x + 2.8$. Find the means of the variable and the coefficient of variables.

Soln: Since the two lines of regression pass to the point $(\bar{x}, \bar{y})$ it satisfies both the equation.

$$\Rightarrow \left.\begin{aligned} \bar{x} &= 0.7\bar{y} + 5.2 \\ \bar{y} &= 0.3\bar{x} + 2.8 \end{aligned}\right\} \text{—} (*)$$

i.e.

$$\left.\begin{aligned} \bar{x} - 0.7\bar{y} &= 5.2 \\ -0.3\bar{x} + \bar{y} &= 2.8 \end{aligned}\right\} \text{—} (i)$$

Solving (i)

$\bar{x} = 9.0632$ & $\bar{y} = 5.5189$.

So mean of $x$ is 9.0632 & $y$ is 5.5189.

from (*), $\frac{r\sigma_y}{\sigma_x} = 0.3$ & $\frac{r\sigma_x}{\sigma_y} = 0.7$.

Now $\frac{r\sigma_y}{\sigma_x} \times \frac{r\sigma_x}{\sigma_y} = 0.3 \times 0.7$

$r^2 = 0.21$

coefficient correlation. $r = \sqrt{0.21} = 0.4532$.

(2) Two lines of regression is given by $8x - 10y + 66 = 0$ & $40x - 18y = 214$. Find mean values of $x$ & $y$. also find coefficient of coilation between $x$ & $y$.

(iii) Find Standard divation of $y$ if $\sigma x = 3y$.

Soln: If we consider first equation as the line of reggression of $x$ on $y$. we can rewrite it as $8x = 10y - 66$.

$\Rightarrow x = \frac{10}{8}y - \frac{66}{8}$ —— (i)

If we consider second equation as the line of reggression of $y$ on $x$.

$-18y = -40x + 214$.

$y = \frac{40}{18}x - \frac{214}{18}$ —— (ii)

Solving (i) & (ii) $\bar{x}$ & $\bar{y}$

$\bar{x} = 13.01$ $\bar{y} = 17.01$.

$\frac{r\sigma x}{\sigma y} = \frac{10}{8}$ $\frac{r\sigma y}{\sigma x} = \frac{40}{18}$

$r^2 = \frac{10}{8} \times \frac{40}{18} = \frac{25}{9} = 2.27$.

$r = 1.656$

which is not possible because $-1 \le r \le 1$. So consider in the other way.

(i) i.e Consider first equation as the line of reggression of $y$ on $x$.

we can rewrite it as $8x - 10y + 66 = 0$.

$\Rightarrow 10y = +8x + 66$

or $y = \frac{8}{10}x + \frac{66}{10}$ —— (i)

Consider 2nd equation as the line of reggression of $x$ on $y$. we can rewrite as $40x - 18y = 214$.

$\Rightarrow$ ~~$-18y = -40x + 214$~~ $40x = 18y + 214$.

~~$18y = 40x - 214$~~ or $x = \frac{18}{40}y + \frac{214}{40}$. —— (ii)

eq (i) & (ii) $\Rightarrow \frac{8}{10}x - y + \frac{66}{10} = 0$.

& $x - \frac{18}{40}y - \frac{214}{40} = 0$.

solving this we get $\boxed{\bar{x} = 1.75}$ & $\boxed{\bar{y} = 8}$ mean.

$\frac{r\sigma x}{\sigma y} = \frac{18}{40}$ $\frac{r\sigma y}{\sigma x} = \frac{8}{10}$.

$r^2 = \frac{18}{40} \times \frac{8}{10} = \frac{144}{400} = 0.36$.

(i) $\therefore r = \sqrt{0.36}$. $\therefore \boxed{r = 0.6}$ coefficient of coilation between $x$ & $y$.

(ii) Standard deviation of y

Given $\sigma_x = 3y$

regression $= \frac{\sigma_y}{\sigma_x}$

where $\sigma_y$ is standard deviation of y

& $\sigma_x$ is standard diviation of x

$\Rightarrow 0.6 = \frac{\sigma_y}{3y}$

$\Rightarrow \sigma_y = 0.6(3y)$

$\Rightarrow \sigma_y = 0.6\{3(3)\}$

$= 0.6 \times 18$

$\boxed{\sigma_y = 10.8}$

* Line of Reggression :-

$$\left.\begin{aligned} \bar{x} &= \frac{\Sigma x}{n} \\ \bar{y} &= \frac{\Sigma y}{n} \end{aligned}\right\} \text{mean}$$

$$\left.\begin{aligned} \sigma_x^2 &= \frac{1}{n}\Sigma(x-\bar{x})^2 \\ \sigma_y^2 &= \frac{1}{n}\Sigma(y-\bar{y})^2 \end{aligned}\right\} \text{Varience}$$

$$r = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma(x-\bar{x})^2}\cdot\sqrt{\Sigma(y-\bar{y})^2}} = \frac{\Sigma xy}{\sqrt{\Sigma x^2}\cdot\sqrt{\Sigma y^2}}$$

* The line of Reggression y on x.

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

* The line of Reggression x on y.

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

* Problems:

(*) Find the coefficient of corelation and find lines of reggression.

(1)

| x | y | $(x-\bar{x})$ | $(x-\bar{x})^2$ | $(y-\bar{y})$ | $(y-\bar{y})^2$ | $(x-\bar{x})(y-\bar{y})$ |
|---|---|---|---|---|---|---|
| 1 | 9 | −4 | 16 | −3 | 9 | 12 |
| 2 | 8 | −3 | 9 | −4 | 16 | 12 |
| 3 | 10 | −2 | 4 | −2 | 4 | 4 |
| 4 | 12 | −1 | 1 | 0 | 0 | 0 |
| 5 | 11 | 0 | 0 | −1 | 1 | 0 |
| 6 | 13 | 1 | 1 | 1 | 1 | 1 |
| 7 | 14 | 2 | 4 | 2 | 4 | 4 |
| 8 | 15 | 3 | 9 | 4 | 16 | 12 |
| 9 | 15 | 4 | 16 | 3 | 9 | 12 |
| total 45 | 108 | | 60 | | 60 | 57 |

$\bar{x} = \frac{45}{9} = 5$ $\quad \bar{y} = \frac{108}{9} = 12$ $\quad \Sigma(x-\bar{x})^2 = 60$

$\Sigma(y-\bar{y})^2 = 60$

$$r = \frac{57}{\sqrt{60}\sqrt{60}} = \frac{57}{(\sqrt{60})^2}$$

$$= \frac{57}{60}$$

(i) $\therefore \boxed{r = 0.95}$

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

$$= \sqrt{\frac{1}{9}\times 60} = \sqrt{\frac{60}{9}}$$

$$\boxed{\sigma_x = 2.5819}$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum(y-\bar{y})^2}$$

$$= \sqrt{\frac{1}{9}\times 60} = \sqrt{\frac{60}{9}}$$

$$\boxed{\sigma_y = 2.5819}$$

The line of regression will be:

$$(y-\bar{y}) = (0.95)\left\{\frac{\cancel{2.5819}}{\cancel{2.5819}}\right\}(x-\bar{x})$$

$$\therefore \boxed{(y-\cancel{12}) = 0.95(x-\cancel{5})}$$

$$(x-\bar{x}) = 0.95(1)(y-\bar{y})$$

$$\therefore \boxed{(x-5) = 0.95(y-12)}$$

(2)

| $x$ | $y$ | $(x-\bar{x})$ | $(x-\bar{x})^2$ | $(y-\bar{y})$ | $(y-\bar{y})^2$ | $(x-\bar{x})(y-\bar{y})$ |
|---|---|---|---|---|---|---|
| 1 | 20 | −4.5 | 20.25 | −11.7 | 136.89 | |
| 2 | 12 | −3.5 | 12.25 | −19.7 | 388.09 | |
| 3 | 16 | −2.5 | 6.25 | −15.7 | 246.49 | |
| 4 | 28 | −1.5 | 2.25 | −3.7 | 13.69 | |
| 5 | 25 | −0.5 | 0.25 | −6.7 | 44.89 | |
| 6 | 36 | 0.5 | 0.25 | 4.3 | 18.49 | |
| 7 | 41 | 1.5 | 2.25 | 9.3 | 86.49 | |
| 8 | 49 | 2.5 | 6.25 | 17.3 | 299.29 | |
| 9 | 40 | 3.5 | 12.25 | 8.3 | 68.89 | |
| 10 | 50 | 4.5 | 20.25 | 18.3 | 334.89 | |
| $\frac{55}{10}$ | $\frac{317}{10}$ | | 82.5 | | 1638.1 | 340.5 |
| $\boxed{5.5}$ | $\boxed{31.7}$ | | | | | |

$$r = \frac{\cancel{1638.1}\ 340.5}{\sqrt{82.5}\sqrt{1638.1}}$$

$$= \frac{340.5}{367.6183}$$

$$\boxed{r = 0.9262}$$

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

$$= \sqrt{\frac{1}{10}(82.5)}$$

$$\boxed{\sigma_x = 2.8722}$$

$$\sigma_y = \sqrt{\frac{1}{n}\sum(y-\bar{y})^2}$$

$$= \sqrt{\frac{1}{10}(1638.1)}$$

$$\boxed{\sigma_y = 12.7988 \approx 12.8}$$

Line of regression $x$ on $y$.

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

$$x - 5.5 = 0.9262\left\{\frac{2.8722}{2.8}\right\}(y-31.7)$$

$$(x-5.5) = 0.9262(1.02)(y-31.7)$$

$$x - 5.5 = 0.9447(y-31.7)$$

(3) Given two lines $y = 0.3x + 2.2$ & $x = 0.7y + 5.2$. Find means of variables & correlation coefficient.

Sol$^n$ Let $\bar{x}$ & $\bar{y}$ be the points on the line, that implies

$$\bar{y} = 0.3\bar{x} + 2.2$$

$$\bar{x} = 0.7\bar{y} + 5.2$$

$$\bar{x} = 9.0632 \Rightarrow \boxed{\bar{x} = 9.06} \text{ (i)}$$

$$\bar{y} = 5.5189 \Rightarrow \boxed{\bar{y} = 5.51}$$

$\therefore$ means of variables,

$$\bar{x} = 9.06 \ \& \ \bar{y} = 5.51$$

$$r = \sqrt{0.3 \times 0.7} = \sqrt{0.21}$$

(ii) $\therefore$ correlation coefficient $= \boxed{r = 0.45}$

* Note: (Not Required for Exam).

(i) To find error:

$$\delta_y = \sigma_y \cdot \sqrt{1-r^2}$$

$$\delta_x = \sigma_x \sqrt{1-r^2}$$

(ii) Correlation coefficient is geometric mean ($\sqrt{r.r}$) of regression coefficient.

$$r = \sqrt{r.\frac{\sigma_x}{\sigma_y} \times r.\frac{\sigma_y}{\sigma_x}} = \sqrt{r.r} = r.$$

(iii) Angle between the lines:

$$\tan\theta = \left(\frac{1-r^2}{r}\right)\left(\frac{\sigma_x \sigma_y}{\sigma_x^2+\sigma_y^2}\right)$$

$\tan\theta = 0$ lines are parallel.

$\tan\theta = \infty$ lines are perpendicular.

(iv) Correlation coefficient,

$$r = \frac{\sigma_x^2+\sigma_y^2-\sigma_{x-y}^2}{2\sigma_x\sigma_y}.$$

(4) Find the line of regression: (y on x)

| | | | | | |
|---|---|---|---|---|---|
| x : | 0 | 1 | 3 | 6 | 8 |
| y : | 1 | 3 | 2 | 5 | 4 |

(5) Find the line of regression: (x on y)

| | | | | | | |
|---|---|---|---|---|---|---|
| x : | 1 | 2 | 3 | 4 | 6 | 8 |
| y : | 2.1 | 3 | 3.6 | 4 | 5 | 6 |

# Unit-II

## 3) Sampling Theory :-

→ The set of all outcomes of the random experiment is called 'population'

→ The random variables $X_1, X_2 \cdots X_n$ is called random sample if :

(i) They are independent

(ii) The joint pdf is $f(x_1) f(x_2) \cdots f(x_n)$

→ The sample mean is

$$\bar{x} = \frac{\sum x_i}{n}.$$

→ The sample varience is

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

⇒ NOTE : $\bar{x}$ & $s^2$ are always independent.

(Imp) (*) With usual notation prove that

(i) $E(\bar{x}) = \mu$ &

(ii) $V(\bar{x}) = \frac{\sigma^2}{n}$.

Proof: WKT $E(\bar{x}) = E\left\{\frac{x_1 + x_2 + \cdots x_n}{n}\right\}$

$$= \frac{1}{n}\{E(x_1) + E(x_2) + \cdots E(x_n)\}$$

$$= \frac{1}{n}\{\mu + \mu + \cdots \mu\}$$

$$= \frac{n\mu}{n} \quad (\because E[x_i] = \mu \ \forall i)$$

(i) $\therefore \boxed{E(\bar{x}) = \mu}$

WKT $V(\bar{x}) = V\left\{\frac{x_1 + x_2 + \cdots + x_n}{n}\right\}$

$$= \frac{1}{n^2}\{V(x_1) + V(x_2) + \cdots V(x_n)\}$$

$$= \frac{1}{n^2}\{\sigma^2 + \sigma^2 + \cdots \sigma^2\} \quad (\because V[x_i] = \sigma^2 \ \forall i)$$

$$= \frac{n\sigma^2}{n^2}$$

(ii) $\therefore \boxed{V(\bar{x}) = \frac{\sigma^2}{n}}$

(*) NOTE:

Suppose $X$ has normal distribution $X \sim N(\mu, \sigma^2)$

then $Z = \frac{X - \mu}{\sigma} \sim N(0,1)$

If $\bar{x}$ has normal distribution $\bar{x} \sim N(\mu, \frac{\sigma^2}{n})$

then $\boxed{Z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \sim N(0,1)}$

(iv) Problems:

(1) Let $\bar{x}$ denote a random sample of size '25' from a distribution $N(3, 100)$. Evaluate $P(\bar{x} < 6)$.

Sol$^n$ $\mu = 3$ $\sigma^2 = 100$ $n = 25$. $\sqrt{n} = 5$.

$\sigma = 10$.

$$P\left[\frac{\bar{x}-3}{\left(\frac{10}{5}\right)} < \frac{6-3}{\left(\frac{10}{5}\right)}\right]$$

$$= P\left(Z < \frac{3}{2}\right)$$

$$= P(Z < 1.50)$$

$$= \Phi(1.50)$$

$$= 0.9332$$

(2) Let $\bar{x}$ be the sample mean of size '5' from $N(0, 1.25)$ distribution. Find $C$ such that $P(\bar{x} < C) = 0.9$.

Sol$^n$ $n = 5$ $\mu = 0$ $\sigma^2 = 1.25 \Rightarrow \sigma = 2.5$.

$$P(\bar{x} < C) = P\left[\frac{\bar{x}-0}{\frac{2.5}{\sqrt{5}}} < \frac{C-0}{\frac{2.5}{\sqrt{5}}}\right]$$

$$= P\left[\frac{\bar{x}}{\frac{5\sqrt{5}}{\sqrt{5}}} < \frac{C}{\frac{5\sqrt{5}}{\sqrt{5}}}\right]$$

$$= P\left(\frac{\bar{x}}{5} < \frac{C}{5}\right) = P\left(Z < \frac{C}{5}\right)$$

$$= \Phi\left[\frac{C}{5}\right] = 0.9$$

$$\Rightarrow \frac{C}{5} = 1.28. \quad (1.2 + 0.08) = 1.28$$

from distribution table.

$$\Rightarrow C = (1.28)5.$$

$$\therefore C = 6.40$$

(3) Let $\bar{x}$ be the sample mean of size 'n' from normal distribution with $\mu$, $\sigma^2 = 100$. Find $n$ such that $P(\mu - 5 < \bar{x} < \mu + 5) = 0.954$

Sol$^n$ $n = n$, $\mu = \mu$, $\sigma^2 = 100$, $\sigma = 10$.

$$P(\mu - 5 < \bar{x} < \mu + 5) = P\left(\frac{\mu - 5 - \mu}{\frac{10}{\sqrt{n}}} < \frac{\bar{x} - \mu}{\frac{10}{\sqrt{n}}} < \frac{\mu + 5 - \mu}{\frac{10}{\sqrt{n}}}\right)$$

$$\Rightarrow 0.954 = P\left[\frac{-5}{\left(\frac{10}{\sqrt{n}}\right)} < Z < \frac{5}{\left(\frac{10}{\sqrt{n}}\right)}\right]$$

$$\Rightarrow 0.954 = \Phi\left[\frac{5}{\left(\frac{10}{\sqrt{n}}\right)}\right] - \Phi\left[\frac{-5}{\left(\frac{10}{\sqrt{n}}\right)}\right]$$

$$= \Phi\left[\frac{5}{\left(\frac{10}{\sqrt{n}}\right)}\right] - 1 + \Phi\left[\frac{5}{\left(\frac{10}{\sqrt{n}}\right)}\right]$$

$$0.954 = 2\Phi\left(\frac{5}{\frac{10}{\sqrt{n}}}\right)$$

$$\frac{0.954}{2} = \Phi\left(\frac{5}{\frac{10}{\sqrt{n}}}\right)$$

$\Rightarrow$ ~~[illegible]~~ $= \Phi$

$$\Rightarrow 0.977 = \phi\left(\frac{5}{\frac{10}{\sqrt{n}}}\right)$$

$$\Rightarrow \frac{5}{\frac{10}{\sqrt{n}}} = 2$$

$$\Rightarrow \frac{5\sqrt{n}}{10} = 2$$

$$\Rightarrow \sqrt{n} = 4$$

$$\therefore \boxed{n = 16}$$

(4) Find the probability of the sample mean with size 15 from a distribution which lies between 3/5 & 4/5 with pdf $f(x) = 3x^2$, $0<x<1$

Soln $E(x) = \int_0^1 x \cdot 3x^2\, dx = 3\left\{\frac{x^4}{4}\right\}_0^1$

$$E(x) = \frac{3}{4}$$

$$\therefore \boxed{\mu = \frac{3}{4}}$$

$$E(x^2) = \int_0^1 x^2\, 3x^2\, dx = 3\left\{\frac{x^5}{5}\right\}_0^1$$

$$E(x^2) = \frac{3}{5}$$

$$V[x] = E(x^2) - (E(x))^2$$

$$= \frac{3}{5} - \left[\frac{3}{4}\right]^2$$

$$= \frac{3}{5} - \frac{9}{16}$$

$$V[x] = \frac{3}{80}$$

$$\therefore \boxed{\sigma^2 = \frac{3}{80}}$$

$n = 15$ given.

$$\Rightarrow P\left(\frac{3}{5} < \bar{x} < \frac{4}{5}\right)$$

$$= P\left[\frac{\left(\frac{3}{5} - \frac{3}{4}\right)}{\left(\sqrt{\frac{3}{80}}/\sqrt{15}\right)} < \frac{\bar{x} - 3/4}{\left(\sqrt{\frac{3}{80}}/\sqrt{15}\right)} < \frac{\left(\frac{4}{5} - \frac{3}{4}\right)}{\left(\sqrt{3/80}/\sqrt{15}\right)}\right]$$

$$= P(-3 < z < 1)$$

$$= \phi(1.00) - \phi(1 - \phi(3.00))$$

$$= 0.8413 - (1 - 0.9987)$$

$$= 0.84$$

Not Required

V- Find the mean & varience of $\chi^2$ distribution:

WKT $f(x) = \dfrac{e^{-x/2} x^{n/2-1}}{2^{n/2}\sqrt{\frac{n}{2}}}$

$E[x] = \int_0^\infty x\, f(x)\, dx$

$= \int_0^\infty \dfrac{x \cdot e^{-x/2} x^{n/2-1}}{2^{n/2}\sqrt{\frac{n}{2}}}\, dx.$

(gama) $\sqrt{n} = \int_0^\infty e^{-x} x^{n-1}\, dx$ —— (i)

$= \dfrac{1}{2^{n/2}\sqrt{n/2}} \int_0^\infty e^{-x/2} x^{n/2}\, dx$

Put $x/2 = t$. $\Rightarrow x = 2t$ $dx = 2dt$

$\Rightarrow = \dfrac{1}{2^{n/2}\sqrt{n/2}} \int_0^\infty e^{-t} \cdot (2t)^{n/2}\, 2dt$

$= \dfrac{2 \cdot 1 (2^{n/2})}{2^{n/2}\sqrt{n/2}} \int_0^\infty e^{-t} t^{n/2}\, dt.$

$= \dfrac{2}{\sqrt{n/2}} \left(\sqrt{\frac{n}{2}+1}\right)$ from (i).

$= \dfrac{2}{\sqrt{n/2}} \times \dfrac{n}{2}\sqrt{n/2} = n.$

$\boxed{V[x] = 2n}$

Mean: $\boxed{E[x] = n.}$

Varience:

## * Central Limit Theorem :-

Let $x_1, x_2, x_3 \ldots x_n$ denote a random variable sample of size 'n' from a distribution with mean $= \mu$, varience $= \sigma^2$ then,

$$Z_n = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ \& } n \to \infty$$

i.e $$Z_n = \frac{\frac{\sum x_i}{n} - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{(\sum x_i - n\mu)}{\sqrt{n}\,\sigma}$$

$$\therefore \boxed{Z_n = \frac{S - n\mu}{\sqrt{n}\,\sigma}}$$

### Problem :

(1) A random sample of size 100 is taken from infinite population with $\mu = 53$ & $\sigma^2 = 100$ find $P(50 < \bar{x} < 56)$.

Sol: $n = 100$

$\mu = 53$

$\sigma^2 = 100 \Rightarrow \sigma = 10$.

By Theorem:

$$P\left(\frac{50-53}{1} < \frac{\bar{x}-53}{1} < \frac{56-53}{1}\right)$$

$$= P(\underset{\text{lower}}{-3} < Z < \underset{\text{(upper)}}{3})$$

$$\phi(\text{upper}) - \phi(\text{lower}).$$

$$= \phi(3) - \phi(-3).$$

$$= \phi(3) - (1 - \phi(3))$$

$$= 2\phi(3) - 1$$

$$= 2(0.9987) - 1.$$

$$= 1.9974 - 1$$

$$= 0.9974.$$

(ch2).

## * $\chi^2$-distribution :-

Let $X$ be a continious random variable & pdf $f(x) = \dfrac{e^{-x/2} \cdot x^{n/2 - 1}}{2^{n/2}\sqrt{\frac{n}{2}}}$ $x \geq 0$.

then we say that $X$ has $\chi^2$ distribution.

n: degree of freedom.

$$\boxed{\begin{array}{l} E[x] = n. \text{ , mean.} \\ V[x] = 2n. \text{ , varience.} \end{array}}$$

### Problems :-

(1) Let $\bar{x}$ denote sample mean of random sample of size n from $\chi^2(50)$. Find probability of $P(49 < \bar{x} < 51)$.

Sol$^n$ $X \sim \chi^2(50) \Rightarrow n = 50$.

mean, $\mu = 50$ $(\because \mu = n)$

Varience, $\sigma^2 = 100$ $(\sigma^2 = 2n = 2 \times 50)$

$\sigma = 10$

$P(49 < \bar{x} < 51)$

$= P\left(\frac{49-50}{10/\sqrt{50}} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < \frac{51-50}{10/\sqrt{50}}\right)$

$= P(-0.7071 < Z < 0.7071)$

$= \Phi(0.7071) - \Phi(-0.707)$

$= \phi(0.7071) - (1 - \phi 0.707)$

$= \cancel{2\phi(0.7071) - 1}$ $2\phi(0.7071) - 1$

$= 2(0.7580) - 1$

$= \cancel{0.5588}$ $0.516$.

(2) Let $\bar{x}$ & $s^2$ be the sample mean & varience from a random sample of size 25 from a distribution $N(3, 100)$. Evaluate $P(\bar{x} < 6, 55.2 < s^2 < 145.6)$.

Sol$^n$ $P(\bar{x} < 6)\ 55.2 < s^2 < 145.6)$

$= P(\bar{x} < 6) \times P(55.2 < s^2 < 145.6)$

$n = 25$, $\mu = 3$, $\sigma = 10$ $(\because \sigma^2 = 100)$

$P(\bar{x} < 6) = P\left(\frac{\bar{x}-3}{10/\sqrt{25}} < \frac{6-3}{10/\sqrt{25}}\right)$

$= P(Z < 1.5)$.

$= \phi(1.50)$.

$\boxed{P(\bar{x} < 6) = 0.9332}$

Standardisation process for $s^2$:

$y = \frac{ns^2}{\sigma^2} \sim \chi^2(n-1)$

$P(55.2 < s^2 < 145.6) = P\left(\frac{55.2 \times 25}{100} < \frac{25 s^2}{100} < \frac{145.6 \times 25}{100}\right)$

$= P(13.8 < y < 36.4)$

$= 0.95 - 0.05$

$= 0.9$.

$n = 25$
$n-1 = 24$
so
10 to
24 show

$\therefore P(\bar{x} < 6, 55.2 < s^2 < 145.6) = 0.9332 \times 0.9$

$= 0.8398$

③ A computer adding numbers rounds off to the nearest integer. Suppose all roundings are independent & uniformly distributed over $(-0.5, 0.5)$. If 1500 numbers are taken & added, what is P that magnitude of total error exceeds 15.

$n = 1500$ $\quad$ $P(|S| \geq 15)$ $\quad$ so $|S|$ (mod)

Sol$^n$ Let $X_i$ be the error & $S$ be the total error.

$$P(|S| \geq 15) = 1 - P(|S| < 15)$$
$$= 1 - P(-15 < S < 15)$$

$\text{mean} = \frac{a+b}{2} \Rightarrow \text{mean} = 0 = \mu$

$\text{var} = \frac{(b-a)^2}{12}$ $\quad$ $\text{variance} = \frac{[0.5-(-0.5)]^2}{12}$

$$\sigma^2 = \frac{1}{12}$$

$$\Rightarrow P(|S| \geq 15) = 1 - P\left(\frac{-15}{\frac{1}{\sqrt{12}} \times \sqrt{1500}} < \frac{S}{\sigma\sqrt{n}} < \frac{15}{\frac{1}{\sqrt{12}} \times \sqrt{1500}}\right)$$
$$= 1 - P(-1.341 < Z < +1.341)$$
$$= 1 - \{\phi(1.341) - (1 - \phi(1.341))\}$$
$$= 2 - 2\phi(1.341)$$
$$= 2 - 2(0.9099)$$
$$= 2\{1 - 0.9099\}$$
$$= 0.1802$$

$\therefore$ the probability that magnitude of total error is 0.1802


edu GAMES

LEARN WHILE YOU PLAY

IMPROVE YOUR VOCABULARY WITH THIS ACTIVITY.

List down 4 names for each category in 60 seconds.

Get. Set. Go.

| | | | | |
|---|---|---|---|---|
| Things That Fly | Leaves. | Kite. | Aeroplane. | Chopper. |
| Things That Are Yellow | | | | |
| Things Found In A Classroom | Chalk. | Blackboard | Bench | Desk. |
| Things To Pack For A Trip | | | | |
| Things That Ring | | | | |
| Things With A Screen | | | | |
| Things That Are Naturally Red | | | | |

EDG0922111A

classmate

| S. No. | Date | Title | Page No. | Teacher's Sign / Remarks |
|---|---|---|---|---|
| | | | | |

# Unit - II

## t-Distribution:

If $X$ has $N(0,1)$ distribution & $y$ has $\chi^2(n)$ distribution where '$n$' is degree of freedom & $x$ & $y$ are independent

then, $\boxed{T = \frac{X}{\sqrt{y/n}}}$ & $\boxed{g(t) = \frac{\sqrt{\frac{n+1}{2}}\left(1+\frac{t^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}}{\sqrt{\pi n}\,\sqrt{n/2}}}$

{Γ gamma}

; $-\infty < t < \infty$

Then we say $T$ has t-distribution

## Problems:

(1) Let $X$ has t-distribution with 14 Degree of Freedom. Determine $b$ so that $P(-b < x < b) = 0.9$.

Soln

$n = 14$

$P(-b < x < b) = 0.9$

$\Rightarrow F(b) - F(-b) = 0.9$

$\Rightarrow F(b) - \{1 - F(b)\} = 0.9$

$\Rightarrow 2F(b) - 1 = 0.9$

$F(b) = \frac{1.9}{2} = 0.95$

$\therefore \boxed{b = 1.76}$ {refering the t-distribution table}

② If a random variable X has t-distribution with 10 degree of freedom. Find $P(|T| \geq 2.228)$.

Soln $P(|T| > 2.228) = 1 - P(|T| < 2.228)$

$= 1 - P(-2.228 < T < 2.228)$

$= 1 - \{F(2.228) - F(-2.228)\}$

$= 1 - \{F(2.228) - \{1 - F(2.228)\}\}$

$= 2 - 2F(2.228)$

$= 2 - 2\{0.975\}$ $(t_{.975})$

$= 2\{0.025\}$

$= 0.05$

# Unit - II

## ESTIMATION OF PARAMETER.

Unbaised: If the mathematical Expectation of any statistic is same as the parameter $\theta$ then we say that the statistic is

$\Rightarrow$ NOTE: $E(\bar{x}) = \mu$, then $\bar{x}$ is unbaised.

Show that $s^2$ {Sample Varience} is not an unbaised estimator for $\sigma^2$ {population varience}

proof: $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$= \frac{1}{n}\left\{\sum_{i=1}^{n} x_i^2 - 2x_i\bar{x} + \bar{x}^2\right\}$

$\boxed{s^2 = \frac{1}{n}\sum x_i^2 - \bar{x}^2}$

$\left(\because \frac{1}{n}\sum 2x_i\bar{x} = \frac{2\bar{x}}{n}\sum x_i \text{ where } \bar{x} = \frac{\sum x_i}{n}\right.$

$\Rightarrow 2x_i\bar{x} = (2\bar{x})^2$

$E(s^2) = E\left\{\frac{1}{n}\sum x_i^2 - (\bar{x})^2\right\}$

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

$= \frac{1}{n}\sum\left(V(x_i) + [E(x_i)]^2\right) - \left(V(\bar{x}) + [E(\bar{x})]^2\right)$

$\left(\because V(x_i) = E(x_i^2) - E([x_i])^2\right)$

$$= \frac{1}{n^2}\left\{n\sigma^2 + n(\mu)^2 - \left(\frac{\sigma^2}{n} + \mu^2\right)\right\}$$

($\because$ WKT $V(x) = \sigma^2$ & $E(x) = \mu$.)

$$= \sigma^2 + \cancel{\mu^2} - \frac{\sigma^2}{n} - \cancel{\mu^2}$$

$$= \sigma^2\left\{1 - \frac{1}{n}\right\} \neq \sigma^2$$

$\therefore$ $\hat{\sigma}$ is not biased.

NOTE:- $E(\bar{x}) = \mu$ (unbiased).

$V(\bar{x}) = \lim_{n\to\infty} \frac{\sigma^2}{n}$ (consistent).

## Point Estimation:

Maximum Likely Wood Estimator method.
(MLE method).

Steps:

Let $x_1, x_2 \ldots x_n$ denote a random sample of size n from a distribution with pdf $f(x, \theta)$ then ① $L(\theta, x_1, x_2 \ldots x_n) = f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta)$.

② $\frac{dL}{d\theta} = 0$.

③ Solve for $\theta$.
& $\theta$ is the MLE that we get.

## Problems:

① Let $x_1, x_2 \ldots x_n$ denote random sample of size n from $f(x, \theta) = \frac{1}{\theta} e^{-x/\theta}$; $0 < x < \infty$, $0 < \theta < \infty$

Find MLE for $\theta$.

Sol:

Step 1: $L = \frac{1}{\theta} e^{-x_1/\theta} \cdot \frac{1}{\theta} e^{-x_2/\theta} \cdots \frac{1}{\theta} e^{-x_n/\theta}$.

$$L = \left(\frac{1}{\theta}\right)^n e^{-\left(\sum x_i/\theta\right)}$$

take log

$$\Rightarrow \log L = -n\log\theta - \left(\frac{\sum x_i}{\theta}\right)$$

$\left\{\because \log\frac{1}{\theta^n} = \log 1 - n\log\theta \;\&\; \log_e e^{-\sum x_i/\theta} = 0 - \frac{\sum x_i}{\theta}\right\}$

Step 2: Diff WRT $\theta$.

$$\Rightarrow \frac{dL}{d\theta} \Rightarrow \frac{1}{L}\frac{dL}{d\theta} = \frac{-n}{\theta} + \frac{\sum x_i}{\theta^2}$$

Step 3:

$$\Rightarrow 0 = \frac{-n}{\cancel{\theta}} + \frac{\sum x_i}{\theta^{\cancel{2}}}$$

$$\Rightarrow n = \frac{\sum x_i}{\theta}.$$

$$\therefore \theta = \frac{\sum x_i}{n}.$$

$$\therefore \boxed{\theta = \bar{x}}$$

(2) Let $x_1, x_2 \cdots x_n$ denote a random variable sample of size 'n' from a distribution

$$f(x,\theta) = \frac{e^{-\theta}\theta^{x}}{\lfloor x}$$

. Find MLE for $\theta$.

Soln

$$L = f(x_1,\theta)\cdot f(x_2,\theta)\cdots f(x_n,\theta).$$

$$= \frac{e^{-\theta}\theta^{x_1}}{\lfloor x_1} \cdots \frac{e^{-\theta}\theta^{x_2}}{\lfloor x_2} \cdots \frac{e^{-\theta}\theta^{x_n}}{\lfloor x_n}$$

$$= \frac{e^{-n\theta}\cdot\theta^{x_1+x_2+\cdots+x_n}}{\lfloor x_1 \cdot \lfloor x_2 \cdots \lfloor x_n}.$$

Step 1:

$$L = \frac{e^{-n\theta}\theta^{\Sigma x_i}}{\lfloor x_1 \lfloor x_2 \cdots \lfloor x_n}$$

Take log.

Step 2: $\log L = -n\theta + \Sigma x_i \log\theta - \log(\lfloor x_1 \lfloor x_2 \cdots \lfloor x_n)$

Diff WRT $\theta$.

$$\Rightarrow \frac{1}{L}\frac{dl}{d\theta} = -n + \frac{\Sigma x_i}{\theta}$$

Step 3: $\Rightarrow 0 = -n + \frac{\Sigma x_i}{\theta}$

$$n = \frac{\Sigma x_i}{\theta}$$

$$\therefore \boxed{\theta = \frac{\Sigma x_i}{n}}$$

(3) Let $x_1, x_2, x_n$ denote a random sample of size 'n' from a distribution $N(\theta_1, \theta_2)$. Where $-\infty < \theta_1 < \infty$ & $0 < \theta_2 < \infty$. Find MLE for $\theta_1$ & $\theta_2$

Soln

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-1/2\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sqrt{\theta_2}} e^{-1/2\left(\frac{x-\theta_1}{\theta_2}\right)^2}$$

$$L = \left[\frac{1}{\sqrt{2\pi}\sqrt{\theta_2}} e^{-1/2\theta_2(x_1-\theta_1)^2}\right]\cdots\left[\frac{1}{\sqrt{2\pi}\sqrt{\theta_2}} e^{-1/2\theta_2(x_n-\theta_1)^2}\right]$$

$$= \frac{1}{(2\pi\theta_2)^{n/2}} e^{-1/2\theta_2 \Sigma(x_i-\theta_1)^2}$$

take log.

$$\Rightarrow \log L = \frac{-n}{2}\log 2\pi\theta_2 - \frac{1}{2\theta_2}\Sigma(x_i-\theta_1)^2$$

$$\frac{1}{L}\frac{\partial l}{\partial \theta_1} = \frac{-1}{2\theta_2}\Sigma 2(x_i-\theta_1)(-1)$$

$$0 = \sum_{i=1}^{n}(x_i-\theta_1)$$

$$0 = \Sigma x_i - n\theta_1$$

$$\boxed{\theta_1 = \frac{\Sigma x_i}{n}}$$

## * Interval Estimation :-

(1) Confidence Interval:

Given $P(-c < z < c)$ = given propbability.

$P\left(-c < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < c\right)$ = giv probability.

$\Rightarrow P\left(\frac{-c\sigma}{\sqrt{n}} < \bar{x}-\mu < \frac{c\sigma}{\sqrt{n}}\right)$ = given probability

$\Rightarrow P\left(\frac{-c\sigma}{\sqrt{n}} - \bar{x} < -\mu < \frac{c\sigma}{\sqrt{n}} - \bar{x}\right)$ = given probability

$\boxed{P\left(\bar{x} - \frac{c\sigma}{\sqrt{n}} < \mu < \bar{x} + \frac{c\sigma}{\sqrt{n}}\right)}$ = given probability

$\therefore$ Interval is $\left(\bar{x} - \frac{c\sigma}{\sqrt{n}}, \bar{x} + \frac{c\sigma}{\sqrt{n}}\right)$

* Problem:

(1) If the observed values are 8.6, 7.9, 8.3, 6.4, 8.4, 9.8, 7.2, 7.8, 7.5. The are the observed samples of size '9' from a $N(\mu, 8)$ distribution. Find 90% confidence interval.

Sol: $P\left(\bar{x} - \frac{c\sigma}{\sqrt{n}}, \bar{x} + \frac{c\sigma}{\sqrt{n}}\right)$.

$n = 9 \quad \sigma^2 = 8 \quad \sigma = \sqrt{8}$.

$\bar{x} = \frac{\Sigma x_i}{n} = 7.9889$.

$\boxed{2\phi(c) - 1 = \text{give probability.}}$ to find to c.

$$2\phi(c) - 1 = 0.9$$

$$\phi(c) = \frac{1.9}{2}$$

$$\phi(c) = 0.95.$$