Tutorial Sheet-01 (Unit-01)

IS41-Statistics, Probability and Linear Programming												
NAME:											Marks:	
USN:												

Curve Fitting, Correlation & Regression:

1.	The normal equations for the best fit straight line of the form $y = a + bx$ for the given data at ***n*** points is: $\boxed{\sum y_i = na + b\sum x_i}$ and $\boxed{\sum x_i y_i = n\sum x_i + b\sum (x_i)^2}$ The normal equations for the best fitting parabola of the form $y = ax^2 + bx + c$ for the given data at ***n*** points is: $\boxed{\sum y_i = a\sum x_i^2 + b\sum x_i + nc}$, $\boxed{\sum x_i y_i = a\sum x_i^3 + b\sum x_i^2 + c\sum x_i}$ and $\boxed{\sum x_i^2 y_i = a\sum x_i^4 + b\sum x_i^3 + c\sum x_i^2}$
2.	**Mean, SD & Variance** $\boxed{\bar{x} = \frac{\sum x_i}{n}}$, $\boxed{\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}}$, $\boxed{\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} \; or \; \sigma^2 = \frac{\sum {x_i}^2}{n} - (\bar{x})^2}$
3.	**Coefficient of correlation:** a) $\boxed{r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x \sigma_y}}$ or $\boxed{r = \frac{\sum X_i Y_i}{n\sigma_x\sigma_y}, where\ X_i = x_i - \bar{x}\ and\ Y_i = y_i - \bar{y}}$ b) $\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum X_i^2}{n}}$ and $\sigma_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}} = \sqrt{\frac{\sum Y_i^2}{n}}$ $\therefore$ $\boxed{r = \frac{\sum X_i Y_i}{\sqrt{(\sum X_i^2)(\sum Y_i^2)}}}$ c) If $z = x - y$ then $\boxed{r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_z^2}{2\sigma_x \sigma_y}}$
4.	**Regression:** a) Equation of the line of regression of y on x is $\boxed{y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})}$ b) Equation of the line of regression of x on y is $\boxed{x - \bar{x} = r\frac{\sigma_x}{\sigma_y}(y - \bar{y})}$ If θ is the acute angle between the two regression lines relating to the variables x and y then $\boxed{\tan\theta = \frac{(1 - r^2)}{r} \cdot \frac{\sigma_x . \sigma_y}{\sigma_x^2 + \sigma_y^2}}$

Tutorial-01

1. Estimate the value of y when $x = 5$ by fitting a least square straight line for the following data

x	2	4	6	8	10	12
y	1.8	1.5	1.4	1.1	1.1	0.9

Solution:

x	y	xy	x^2
2	1.8		
4	1.5		
6	1.4		
8	1.1		
10	1.1		
12	0.9		

2. Find a parabola of the form $y = a + b\,x + c\,x^2$ for the following observations:

x	-3	-2	-1	0	1	2	3
y	4.63	2.11	0.67	0.09	0.63	2.15	4.58

Solution:

3. The voltage V across a capacitor at time t seconds is given by the following table

T	0	2	4	6	8
V	150	63	28	12	56

Use the method of least squares to fit a curve of the form $y = ae^{kt}$ to this data.

Solution:

4. Obtain the coefficient of correlation and the equation of the lines of regression for the following data

x	0	1	2	3	4	5	6
y	14	13	11	9	8	5	3

5. The following table gives the data on rainfall and discharge in a certain river. Obtain the line of regression of y on x.

Rainfall x (inches)	1.53	1.78	2.60	2.95	3.42
Discharge y (1000 cc)	33.5	36.3	40.0	45.8	53.5

6. If $y = 4.686x + 4.927$ and $x = 0.197y - 0.548$ are the equations of regression lines. Find the mean of the variables $\bar{x}$, $\bar{y}$ and the coefficient of correlation r between them.