Textbooks -

① Introduction to probability and statistics for engineers & scientists.

- S. M. Ross

② Probability & statistics in engineering

- D. Montogomery, & ---

Wiley India

③ Fundamental of methematical statistics.

- S. C. Gupta & V. K. Kapoor.

Topics -

①

- Concepts of frequency distribution, measures of central tendency - mean, median, mode and quantiles.
- Measures of dispersion - range mean absolute deviation variation and standard deviation coefficient of variation.
- Moments about (i) origin (ii) any point (iii) mean. Relation b/w the moments about the mean and the moment about any points.

Pearson's β and γ coefficients

Skewness & kurtosis & their measures.

② Theory of least square and curve fitting, a second degree parabola, exponential type curves

③ Correlation: Definitions its limits, effects of change of origin and scale on correlation coefficient. ⑦

Regression and regression lines, point of intersection of the two lines of regression. Regression coefficients and their properties Definition of multiple and partial correlation coefficients in a Trivariate distribution.

④ Probability: classical, emperical (statistical), subjective and axiomatic approaches. Addition and multiplication theorem of probability (with proof) Statement & proof of Bayes's theorem simple numerical problems.

⑤ Random variable, probability mass function and probability density function, distribution function and their properties. Mathematical expectation and their properties (without proof) Moment generating function and its properties.

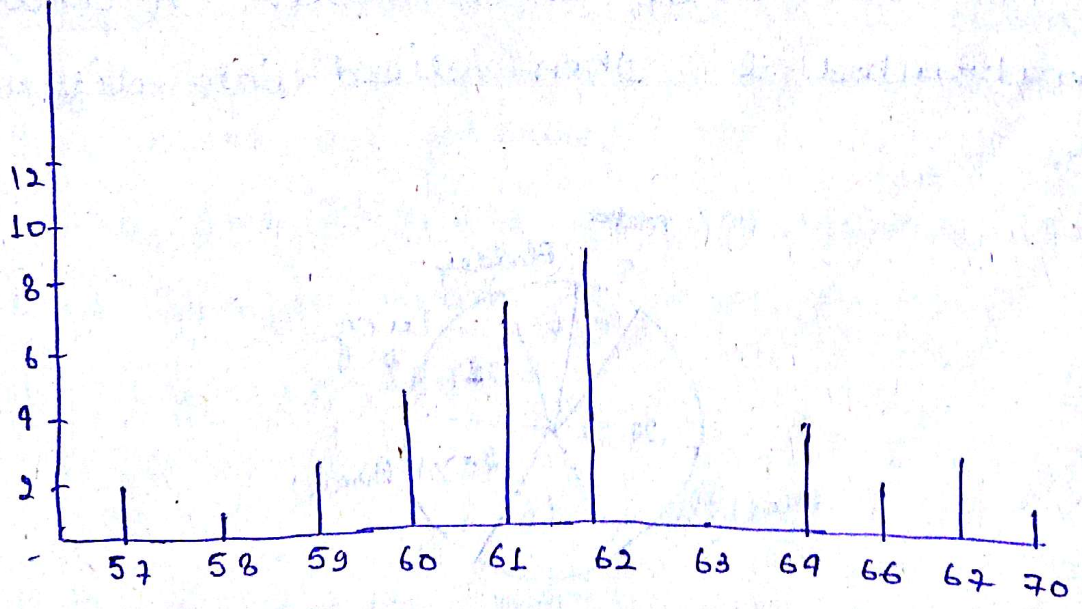
⑥ Probability distribution. Definition, mean and variance of Bernoulli, Binomial, negative binomial and poission distribution Normal distributions and its characteristics (without proof), its m.g.f and area property

(1) Tests of significance: Definitions of various terms, normal test, t-test, χ^2-test, F-test.

Frequency tables and Graphs:

Consider a frequency table for a data consisting of the starting yearly salaries (to nearest thousand dollars) of 42 recently graduated student in EE

Starting Salary	Frequency
57	4
58	1
59	3
60	5
61	8
62	10
63	0
64	5
66	2
67	3
70	1

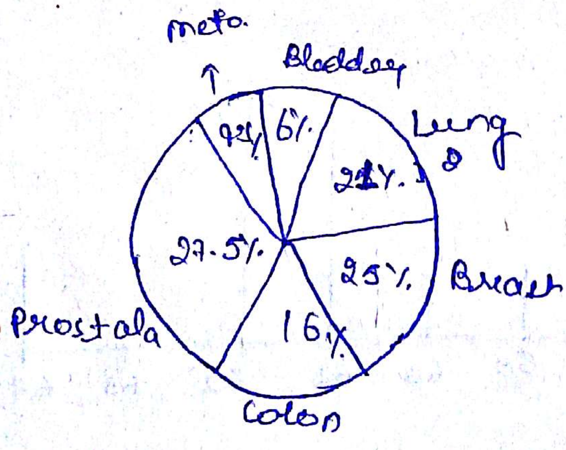


- **Relative Frequency table and Graphs:**

Consider a data set consisting of n values. If f is the frequency of a particular value then the ratio f/n is called its relative frequency. That is, relative.

Starting Salary	Frequency
57	4/42
58	1/42
59	3/42
60	5/42
61	8/42
62	10/42
63	0
64	5/42
66	2/42
67	3/42
70	1/42

Pie Chart- A pie chart is used to indicate relative frequencies when the data are not numerical in nature. A circle is constructed & then sliced into different sectors.



Different types of cancers affecting that values.

• **Summarising Data sets:**

Sample mean - Suppose we have a data set consisting of n natural values $x_1, x_2, \dots x_n$

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

$$= \sum_{i=1}^{n} x_i / n = \frac{1}{n} \sum_{i=1}^{n} x_i$$

• The computation of the sample mean can be simplified:

$$y_i = a x_i + b \quad , \quad i = 1, 2, \dots, n$$

then,
$$\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{n} (a x_i + b)$$

$$= \left\{ \frac{a}{n} \cdot \sum_{i=1}^{\infty} x_i + \frac{1}{n} \cdot \sum_{i=1}^{\infty} b \right\}$$

$$\Rightarrow \boxed{\bar{y} = a \cdot \bar{x} + b}$$

Example: The winning scores in a certain golf tournament in the years from 2004 to 2013 were as follows,

280, 278, 272, 276, 281, 279, 276, 281, 289, 280

Find sample mean of these scores.

Sol^n: $y_i = x_i - 280$

$0, -2, -8, -4, 1, -1, -4, 1, 9, 0$

$\bar{y} = -8/10 = -0.8 \quad ; \quad \bar{y} = \bar{x} - 280$

$\Rightarrow \bar{x} = \bar{y} + 280 = 279.2$

- To determine the sample mean of a data set that is presented in a frequency table listing k disting values. $v_1, v_2, \ldots v_k$ having corresponding frequency/ies $f_1, f_2, \ldots f_k$

& $n = \sum_{i=1}^{n} f_i$

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{k} v_i f_i$$

$$\Rightarrow \bar{x} = \frac{f_1}{n} \cdot v_1 + \frac{f_2}{n} \cdot v_2 + \ldots + \frac{f_k}{n} \cdot v_k$$

__Example__: The following is a frequency table giving the ages of members of a symphony orchestra for young adults:

Age	Frequency
15	2
16	5
17	11
18	9
19	14
20	13

Ans:- $18.24 = \bar{x}$

__Sample median__- $x_1, x_2, \ldots, x_n$

Order the values from smallest to largest.

If n is odd, the sample median is the value in position $\frac{(n+1)}{2}$

If n is even, it is the average of the values in positions $(\frac{n}{2})$ & $(\frac{n}{2}+1)$

$$\text{sample median} = \begin{cases} x\left(\frac{n+1}{2}\right) & \text{if } n \text{ is odd} \\ \dfrac{x(\frac{n}{2}) + x(\frac{n}{2}+1)}{2} & \text{if } n \text{ is even} \end{cases}$$

Sample variance:- $x_1, x_2, \ldots, x_n$

The sample variance, call it s^2, of the data set, is defined by

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

Example:- Find the sample variances of the data sets A and B

A : 3, 4, 6, 7, 10

B : -20, 5, 15, 24

Ans:-

$\bar{x} = 6$, $s^2 = 7.5$

$\bar{x} = 6$, $s^2 = 360.67$

- Measures of Dispersion -
  - → Sample Variance
  - → Sample standard deviation
  - → Range
  - → Coefficient of Variance (CV)
  - → mean absolute deviation.

① Sample Variance- $x_1, x_2, \ldots x_n$; $\bar{x}_n = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$

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

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

② Sample standard deviation-

The quantity is defined by

$$s = \sqrt{\frac{1}{(n-1)} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$$ is called the sample standard deviation.

③ Range- $R = \max(x_i) - \min(x_i)$

The sample variance (or the sample standard deviation) provides a "better measure of variability."

④ The coefficient of variation (CV) - is used to express variation as a fraction of the mean. A measure of relative variation is called the sample coefficient of variation of and it is defined as.

$$CV = \frac{s}{\bar{x}}$$

⑤ mean absolute deviation -

$$\rightarrow \frac{1}{n} \cdot \sum_{i=1}^{n} |x_i - \bar{x}|$$

Example - show that,

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

Soln:

$$LHS = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} \{x_i^2 - 2x_i.\bar{x} + \bar{x}^2\}$$

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

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

$$= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

• Grouped data :

If the data are in a frequency distribution; the formula for the measure of central tendancy and despersion; suppose for each of p distincts values of x, say $x_1, x_2, --- x_p$, the observed frequencies are $f_1, f_2, --- f_p$ respectively.

$$\bar{x} = \frac{\sum_{j=1}^{p} f_j \cdot x_j}{\sum_{j=1}^{p} f_j} = \frac{1}{n} \cdot \sum_{j=1}^{p} f_j \cdot x_j$$

$$s^2 = \frac{1}{(n-1)} \cdot \left\{ \sum_{j=1}^{p} f_j \cdot x_j^2 - \frac{1}{n} \left(\sum_{j=1}^{p} f_j x_j \right)^2 \right\}$$

• Measuring Association :

(Person or sample correlation coefficient)

When data set contain a number of variables denoted by x and y. The simple correlation coefficient between the variable x and variable y.

The correlation coefficient is,

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

• $|r_{xy}| \leq 1$

Or, $r = \dfrac{S_{xy}}{(S_{xx} \cdot S_{yy})^{1/2}}$ where, $S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x}) \cdot (y_i - \bar{y})$

$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$

$S_{yy} = \sum (y_i - \bar{y})^2$

On simplifying,

$$r = \frac{\sum_{i=1}^{n} x_i . y_i - n\bar{x}.\bar{y}}{\sqrt{\left\{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right\}\left\{\sum_{i=1}^{n} y_i^2 - n.\bar{y}^2\right\}}}$$

Properties -

① $-1 \leq r \leq 1$

$$\sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{\sqrt{S_{xx}}} - \frac{y_i - \bar{y}}{\sqrt{S_{yy}}}\right) \geq 0$$

or $$\sum_{i=1}^{n} \left\{\frac{(x_i - \bar{x})^2}{S_{xx}} + \frac{(y_i - \bar{y})^2}{S_{yy}} - 2\frac{(x_i - \bar{x}).(y_i - \bar{y})}{\sqrt{S_{xx}.S_{yy}}}\right\} \geq 0$$

or $$\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{S_{xx}} + \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{S_{yy}} \mp \frac{2\sum_{i=1}^{n} (x_i - \bar{x}).(y_i - \bar{y})}{\sqrt{S_{xx}.S_{yy}}} \geq 0$$

or $1 + 1 \mp 2r \geq 0$

$2 \mp 2r \geq 0$

or $-1 \leq r \leq 1$

② If for constants a and b with $b > 0$,
$y_i = a + bx_i$, $i = 1, 2, \ldots n$, then $r = 1$

Also if $b < 0$

$y_i = a + bx_i$, $i = 1, 2, \ldots n$, then $r = -1$

(3) If r is the sample correlation coefficient for the data pairs x_i, y_i, $i = 1, 2, \ldots, n$, then it is also the sample correlation coefficient for the data ~~points~~ pairs $a + bx_i$, $c + dy_i$, $i = 1, 2, \ldots n$ provided that b and d are both positive or both negative.

*Imp

- Mean and ~~Variation~~, Variance of combined Sample:

$x_{11}, x_{12}, x_{13}, \ldots x_{1n_1}$: $\bar{x}_1 = \frac{1}{n_1} \cdot \sum_{i=1}^{n_1} x_{1i}$,

$$S_1^2 = \frac{1}{n_1} \cdot \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2$$

$x_{21}, x_{22}, \ldots \ldots x_{2n_2}$: $\bar{x}_2 = \frac{1}{n_2} \cdot \sum_{i=1}^{n_2} x_{2i}$

$$S_2^2 = \frac{1}{n_2} \cdot \sum_{i=1}^{n_2} (x_{2i} - \bar{x}_2)^2$$

Now,

combined sample mean,

$$\bar{x} = \frac{\sum_{i=1}^{n_1} x_{1i} + \sum_{i=1}^{n_2} x_{2i}}{n_1 + n_2} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

and combined sample variance,

$$S^2 = \frac{1}{(n_1 + n_2)} \cdot \left\{ \sum_{i=1}^{n_1} (x_{1i} - \bar{x})^2 + \sum_{i=1}^{n_2} (x_{2i} - \bar{x})^2 \right\}$$

$$\Rightarrow S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2} + \frac{n_1 n_2 (\bar{x}_1 - \bar{x}_2)}{(n_1 + n_2)^2}$$

- **Moments:-**

For a sample $\{x_1, x_2, \ldots, x_n\}$ of size n, the r^{th} moment about origin, also called the r^{th} raw moment denoted by m_r', is defined by:

$$\boxed{m_r' = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^r}$$

the r^{th} moment about the mean $\bar{x}$, also called the r^{th} order central moment denoted by m_r, is defined by,

$$\boxed{m_r = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^r}$$

the r^{th} order moment about the point a is

$$\boxed{m_r' = \frac{1}{n} \sum_{i=1}^{n} (x_i - a)^r}$$

- Relationship between central moments and raw moments:

$$m_1 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

$$m_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n} \cdot \left\{ \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \right\}$$

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

$$\boxed{m_2 = m_2' - m_1'^2}$$

$$m_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{x_i^3 - 3x_i^2\bar{x} + 3x_i\bar{x}^2 - \bar{x}^3\right\}$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{n}x_i^3 - 3\bar{x}\sum_{i=1}^{n}x_i^2 + 3\bar{x}^2\sum_{i=1}^{n}x_i - \sum_{i=1}^{n}\bar{x}^3\right\}$$

$$= \frac{1}{n}\left\{\sum_{i=1}^{n}x_i^3 - 3\bar{x}\sum_{i=1}^{n}x_i^2 + 3n\bar{x}^3 - n\bar{x}^3\right\}$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^3 - 3\bar{x}\,\frac{1}{n}\sum_{i=1}^{n}x_i^2 + 2\bar{x}^3$$

$$\boxed{m_3 = m_3' - 3m_1' m_2' + 2m_1'^3}$$

$$m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{x_i^4 - 4x_i^3\bar{x} + 6x_i^2\bar{x}^2 - 4x_i\bar{x}^3 + \bar{x}^4\right\}$$

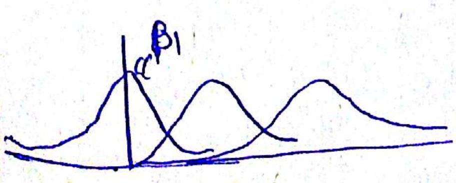
$$= \frac{1}{n}\left\{\sum_{i=1}^{n}x_i^4 - 4\bar{x}\sum_{i=1}^{n}x_i^3 + 6\bar{x}^2\sum_{i=1}^{n}x_i^2 - 4\bar{x}^3\sum_{i=1}^{n}x_i + n\bar{x}^4\right\}$$

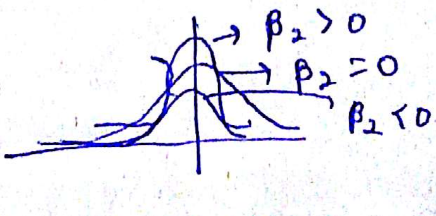
$$= \frac{1}{n}\sum_{i=1}^{n}x_i^4 - 4\bar{x}\,\frac{1}{n}\sum_{i=1}^{n}x_i^3 + 6\bar{x}^2\,\frac{1}{n}\sum_{i=1}^{n}x_i^2 - 3\bar{x}^4$$

$$\boxed{m_4 = m_4' - 4m_3' m_1' + 6m_2' m_1'^2 - 3m_1'^4}$$

Measure of skewness : $\boxed{\beta_1 = \frac{m_3}{(m_2)^{3/2}}}$

measure of kurtosis : $\boxed{\beta_2 = \frac{m_4}{(m_2)^2} - 3}$





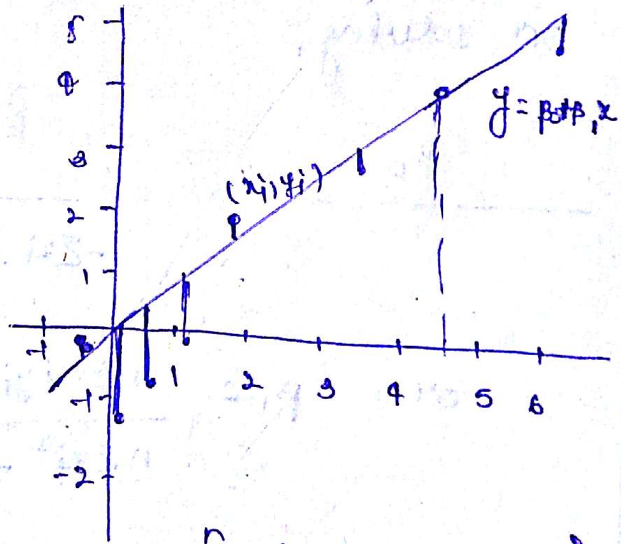
→ Skewness reflects the degree of symmetry about the mean; and negative skew results from an asymmetric tail toward smaller values of the variable while positive skew results from an asymmetric tail extending toward the larger values of the variable.

→ Kurtosis describes the relative peakness of a distribution as compared to the normal distribution, where a negative value is associated with a relatively flat distribution and a positive value is associated with relatively peaked distributions.

- The method of least squares:

Suppose that each of 10 patients is treated with the same amount of two different drugs that can affect blood pressure. For $i = 1, 2, \ldots 10$, let x_i denote the ~~relation~~ reaction ~~p~~ measured in appropriate units, of the i^{th} patient to drug A & y_i denote the reaction to drug B.

i	x_i	y_i
1	1.9	0.7
2	0.8	-1.0
3	1.1	-0.2
4	0.1	-1.2
5	-0.1	-0.1
6	4.4	3.4
7	4.6	0.0
8	1.6	0.8
9	5.5	3.7
10	3.4	2.0



$$Q = \sum_{i=1}^{n} \{y_i - (\beta_0 + \beta_1 x_i)\}^2$$

$$\Rightarrow \frac{\partial Q}{\partial \beta_0} = 0 \Rightarrow -2 \sum_{i=1}^{n} \{y_i - (\beta_0 + \beta_1 x_i)\} = 0 \quad \text{--- (1)}$$

$$\text{and } \frac{\partial Q}{\partial \beta_1} = 0 \Rightarrow -2 \sum_{i=1}^{n} \{y_i - (\beta_0 + \beta_1 x_i)\} x_i = 0 \quad \text{--- (2)}$$

eq^n ① $\Rightarrow$ $\sum_{i=1}^{n} \{y_i - \beta_0 - \beta_1 . x_i\} = 0$

or $\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \beta_0 - \beta_1 \sum_{i=1}^{n} x_i = 0$

or $\sum_{i=1}^{n} y_i - n\beta_0 - \beta_1 . \sum_{i=1}^{n} x_i = 0$

eq^n ② $\Rightarrow$ $\sum_{i=1}^{n} \{y_i - \beta_0 - \beta_1 x_i\} x_i = 0$

or $\sum_{i=1}^{n} \{x_i y_i - \beta_0 . x_i - \beta_1 x_i^2\} = 0$

or $\sum_{i=1}^{n} x_i . y_i - \beta_0 . \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0$

$$\Rightarrow \left.\begin{aligned} & n\beta_0 + \beta_1 . \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i = 0 \\ & \beta_0 . \sum_{i=1}^{n} x_i + \beta_1 . \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i y_i = 0 \end{aligned}\right\} — ③$$

let $\left\{ \sum_{i=1}^{n} x_i = \sum x_i \right\}$

On solving,

$$\frac{\beta_0}{} = \frac{\beta_1}{-\sum x_i . \sum y_i + n \sum x_i y_i} = \frac{1}{n . \sum x_i^2 - (\sum x_i)^2}$$

or $$\beta_1 = \frac{n . \sum x_i y_i - \sum x_i . \sum y_i}{n \sum x_i^2 - (n\bar{x})^2}$$

$$\Rightarrow \beta_1 = \frac{n \sum x_i y_i - (n\bar{x}).(n\bar{y})}{n \sum x_i^2 - (n\bar{x})^2}$$

or $$\boxed{\beta_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}}$$

Scanned by CamScanner

from ③,

$n\beta_0 + n\beta_1\bar{x} - n\bar{y} = 0$

or $\beta_0 + \beta_1\bar{x} - \bar{y} = 0 \Rightarrow \boxed{\beta_0 = \bar{y} - \beta_1 . \bar{x}}$

we can use a higher order curve (like parabola) rather than straight line to get better result.

Theorem :- Let $(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)$ be a set of n points. The straight line that minimizes the sum of squares of the vertical deviations of all the points from the line has the following slope and intercept:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x}).(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

Defn :- Let $\hat{\beta}_0$ and $\hat{\beta}_1$ be as defined above the line defined by the eqn $y = \hat{\beta}_0 + \hat{\beta}_1 x$ is called the least square line.

Ex $n = 10$; $\hat{\beta}_0 = -0.786$; $\hat{\beta}_1 = 0.685$

$y = -0.786 + 0.685x$

Solution :-

i	x_i	y_i	$x_i y_i$	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})^2$
1	1.9	0.7				
2	0.8	-1.0				
3	1.1	-0.2				
4	0.1	-1.2				
5	-0.1	-0.1				
6	4.4	3.4				
7	4.6	0.0				
8	1.6	0.8				
9	5.5	3.7				
10	3.4	2.0				

Remarks -

① Experimental data may not be always linear, One may be interested in fitting either a curve of the form ⓐ $y = ax^b$ or ⓑ $y = a e^{bx}$

Case ⓐ - $y = ax^b$

$\log_{10} y = \log_{10} a + b \log_{10} x$

or $\boxed{Y = A + bX}$ —① where $Y = \log_{10} y$

$A = \log_{10} a$

$X = \log_{10} x$

Eg:- By the method of least square, fit a curve of the form $y = ax^b$ to the following data

x :	2	3	4	5
y :	27.8	62.1	110	161

Solⁿ:- Data in modified variable X, Y is

X :	0.3010	0.4771	0.6021	0.6990
Y :	1.4440	1.7931	2.0414	2.2068

X_i	Y_i	X_iY_i	X_i^2
0.3010	1.4440	0.4346	0.0906
0.4771	1.7931	0.8555	0.2274
0.6021	2.0414	1.2291	0.3625
0.6990	2.2068	1.5426	0.4886
Σ 2.0792	7.4833	4.0618	1.1693

Now from eqⁿ ①

$nA + b \sum_{i=1}^{4} X_i = \sum_{i=1}^{4} Y_i$

$A \sum_{i=1}^{4} X_i + b \sum_{i=1}^{4} X_i^2 = \sum X_i Y_i$

i.e. $4A + 2.0792 \cdot b = 7.4853$

$2.0792A + 1.1693b = 4.0618$

On solving,

$A = 0.8678$ & $b = 1.9311$

i.e $\log_{10} a = 0.8678$

$\Rightarrow a = 7.375$

The required curve is,

$$\boxed{y = 7.375 \cdot x^{1.9311}}$$

<u>Example</u>- The pressure of the gas corresponding to various volumes V is measured, given by the following data:

V(cm^3):	50	60	70	90	100
P($kg\ cm^{-2}$):	64.7	51.3	40.5	25.9	78

Fit the data to the eqⁿ $\underline{PV^{\gamma} = C}$

<u>Solⁿ</u>:- $P \cdot V^{\gamma} = C$

or $P = C \cdot V^{-\gamma}$

or $\log_{10} P = \log_{10} C - \gamma \cdot \log_{10} V$

$Y = A + BX$

$B = -\gamma$

where $Y = \log_{10} P$,

$A = \log_{10} C$

$X = \log_{10} V$

<u>Ans</u>: $PV^{0.28997} = 167.7876$

case (b): $y = a.e^{bx}$

$\log_{10} y = \log_{10} a + bx \log_{10} e$

or $Y = A + BX$, where $Y = \log_{10} y$

$A = \log_{10} a$

$B = b\log_{10} e$

Least Square Fit (Parabola)

Given a data set of n observations (x_i, y_i); $i = 1, 2, \ldots n$ of an experiment, we want to fit a best possible parabola,

$$y = ax^2 + bx + c$$

using the principle of least square

The sum of the squares of the residual's $(e_i, i = 1, 2, \ldots n)$ is given by

$$E = \sum_{i=1} e_i^2 = \sum_{i=1}^{n} [y_i - (ax_i^2 + bx_i + c)]^2$$

$$\frac{\partial E}{\partial a} = \frac{\partial E}{\partial b} = \frac{\partial E}{\partial c} = 0$$

$\Rightarrow$

$$\left.\begin{aligned} \frac{\partial E}{\partial a} &= -2.\sum_{i=1}^{n} [y_i - (ax_i^2 + bx_i + c)]x_i^2 = 0 \\ \frac{\partial E}{\partial b} &= -2\sum_{i=1}^{n} [y_i - (ax_i^2 + bx_i + c)]x_i = 0 \\ \frac{\partial E}{\partial c} &= -2\sum_{i=1}^{n} [y_i - (ax_i^2 + bx_i + c)] = 0 \end{aligned}\right\}$$

$$\left\{\begin{aligned} &a\sum_{i=1}^{n} x_i^4 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2 \cdot y_i \\ &a \cdot \sum_{i=1}^{n} x_i^3 + b\sum_{i=1}^{n} x_i^2 + c \cdot \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i \cdot y_i \\ &a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i + n \cdot c = \sum_{i=1}^{n} y_i \end{aligned}\right. \quad (2)$$

Example: Fit a parabola in the form $y = ax^2 + bx + c$.

Following the principles of least square to given data

x :	1.0	1.6	2.5	4.0	6.0
y :	9.4	11.8	14.7	18.0	23.0

Sol^n: $n = 5$.

x	y	x^2	x^3	x^4	xy	x^2y
1.0	9.4	1.0	1.000	1.00	9.40	9.400
1.6	11.8	2.56	4.096	6.554	18.88	30.208
2.5	14.7	6.25	15.625	39.0625	36.75	91.8
4.0	18.0	16.00	64.00	256.000	72.00	288.0
6.0	23.0	36.00	216.00	1296.00	138.00	828.0
15.1	76.9	61.81	110.721	1598.6161	275.03	1241

put these in eqⁿ ③,

$$1598.6161\,a + 110.7216\,b + 61.81\,c = 1247.783$$

$$10.721\,a + 61.81\,b + 15.1\,c = 275.03$$

$$61.81\,a + 15.1\,b + 5\,c = 76.9$$

on solving

$a = 1.0399$; $b = 7.5088$; $c = -20.1811$