LECTURE 9

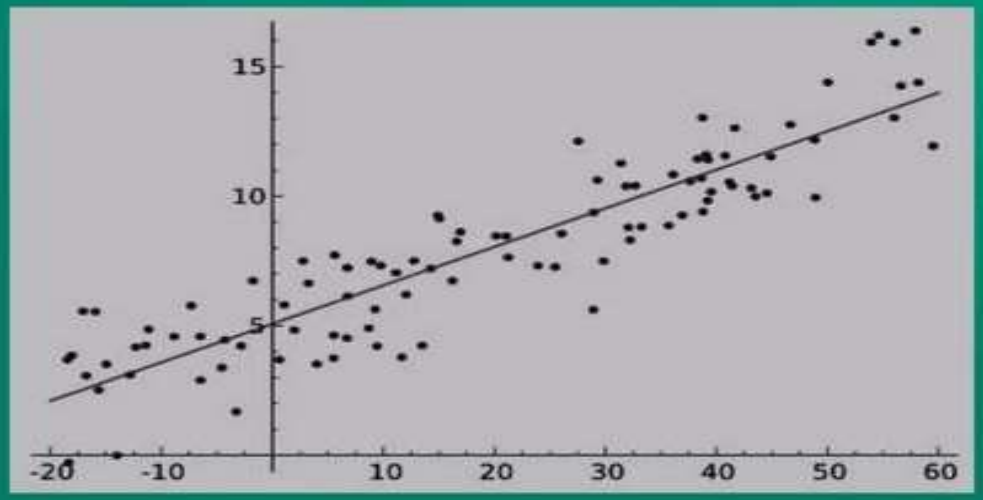
CURVE FITTING

Definition

- **Curve fitting:** is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points, possibly subject to constraints.
- It is a statistical technique use to drive coefficient values for equations that express the value of one(dependent) variable as a function of another (independent variable)

What is curve fitting

Curve fitting is the process of constructing a curve, or mathematical functions, which possess closest proximity to the series of data. By the curve fitting we can mathematically construct the functional relationship between the observed fact and parameter values, etc. It is highly effective in mathematical modelling some natural processes.

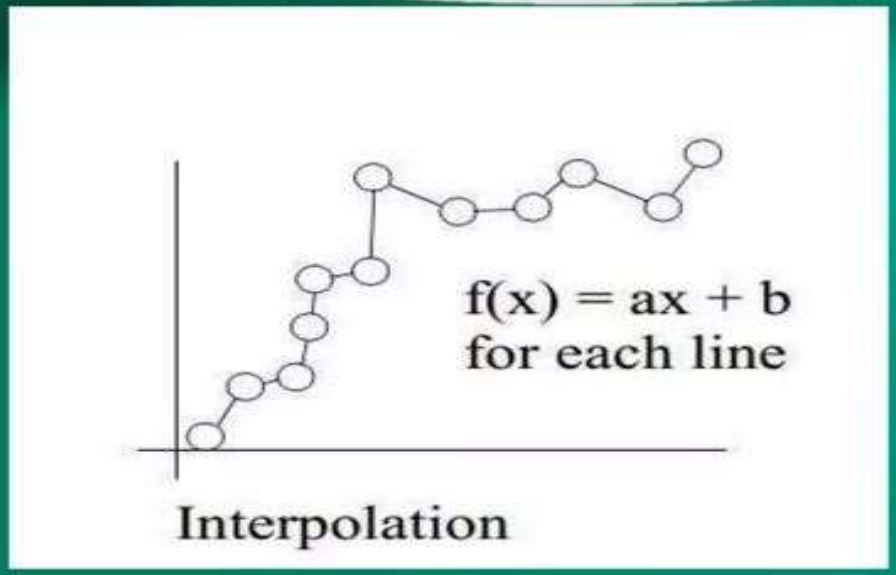


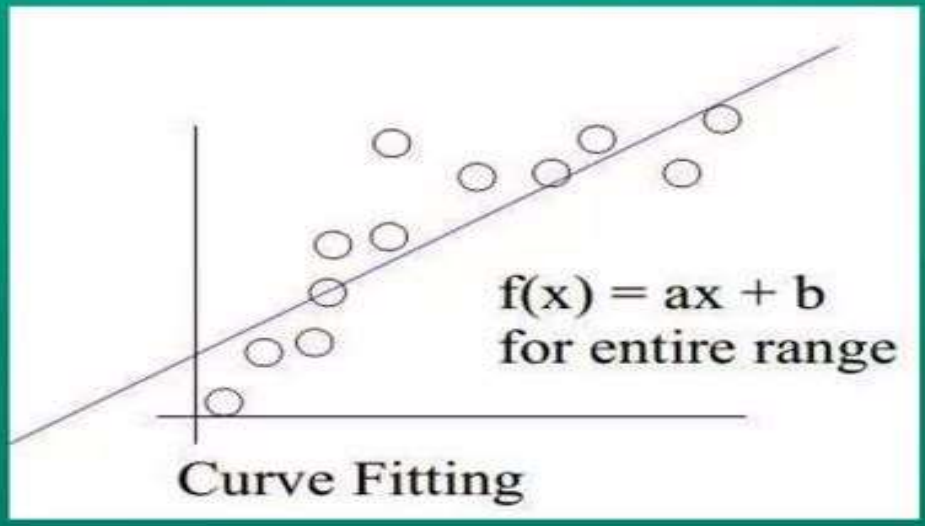
Interpolation & Curve fitting

- In many application areas, one is faced with the test of describing data, often measured, with an analytic function. There are two approaches to this problem:-
- 1. In Interpolation, the data is assumed to be correct and what is desired is some way to descibe what happens between the data points.
- 2. The other approach is called curve fitting or regression, one looks for some smooth curve that ``best fits'' the data, but does not necessarily pass through any data points.

fppt.com

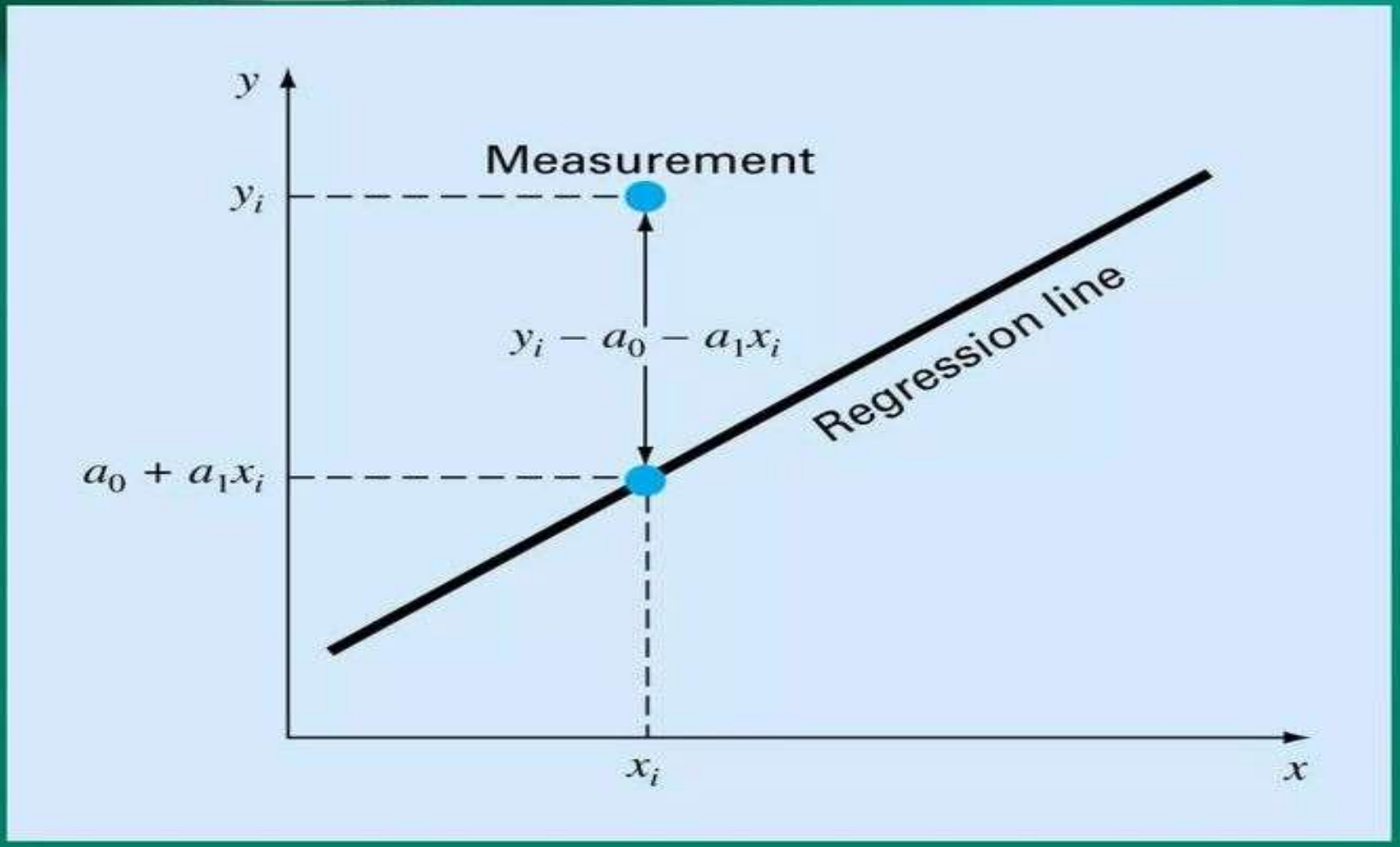
General approach for curve fitting





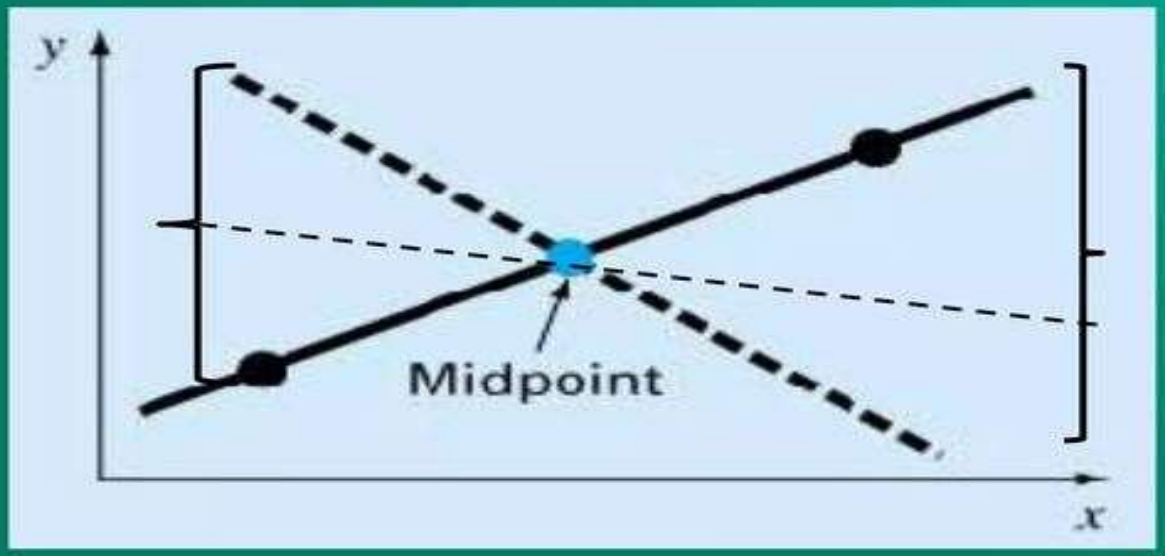
fppt.com

Linear Regression: Residual



Linear Regression: Criteria for a "Best" Fit

$$\min \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)$$



$e_1 = -e_2$

Linear Regression: Least Squares Fit

$$\min S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

The coefficients a_0 and a_1 that minimize S_r must satisfy the following conditions:

$$\begin{cases} \dfrac{\partial S_r}{\partial a_0} = 0 \\ \dfrac{\partial S_r}{\partial a_1} = 0 \end{cases}$$

Linear Regression: Determination of a_o and a_1

$$\frac{\partial S_r}{\partial a_o} = -2\sum (y_i - a_o - a_1 x_i) = 0$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum [(y_i - a_o - a_1 x_i) x_i] = 0$$

$$0 = \sum y_i - \sum a_0 - \sum a_1 x_i$$

$$0 = \sum y_i x_i - \sum a_0 x_i - \sum a_1 x_i^2$$

$$\sum a_0 = n a_0$$

$$n a_0 + \left(\sum x_i\right) a_1 = \sum y_i$$

$$\sum y_i x_i = a_0 \sum x_i + a_1 \sum x_i^2$$

$\}$ 2 equations with 2 unknowns, can be solved simultaneously

$$\Sigma y = a_0 n + a_1 \Sigma x$$

$$\Sigma xy = a_0 \Sigma x + a_1 \Sigma x^2$$

Linear Regression: Determination of ao and a1

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

$$a_0 = \bar{y} - a_1 \bar{x}$$

Example

- The table blew gives the temperatures T in C and Resistance R in Ω of a circuit if $R=a_0 + a_1T$
- Find the values of a_0 and a_1

T	10	20	30	40	50	60
R	20.1	20.2	20.4	20.6	20.8	21

Solution

T=Xi	R=yi	$Xi^2 = T^2$	Xiyi=TR	g(xi)=Y
10	20.1	100	201	20.05
20	20.2	400	404	20.24
30	20.4	900	612	20.42
40	20.6	1600	824	20.61
50	20.8	2500	1040	20.80
60	21	3600	1260	20.98
$\sum xi = 210$	$\sum yi = 123.1$	$\sum xi^2 = 9100$	$\sum xiyi = 4341$	

Solution

- $6a_0+210a_1=123.1$ $a_0=19.867$
- $210a_0+9100a_1=4341$ $a_1=0.01857$
- $g(x)=19.867+0.01857*T$

fppt.com

Least Squares Fit of a Straight Line: Example

- Fit a straight line to the x and y values in the following Table:

x_i	y_i	$x_i y_i$	x_i^2
1	0.5	0.5	1
2	2.5	5	4
3	2	6	9
4	4	16	16
5	3.5	17.5	25
6	6	36	36
7	5.5	38.5	49
28	**24**	**119.5**	**140**

$$\sum x_i = 28 \quad \sum y_i = 24.0$$

$$\sum x_i^2 = 140 \quad \sum x_i y_i = 119.5$$

$$\bar{x} = \frac{28}{7} = 4$$

$$\bar{y} = \frac{24}{7} = 3.428571$$

fppt.com

Least Squares Fit of a Straight Line: Example

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

$$= \frac{7\times 119.5 - 28\times 24}{7\times 140 - 28^2} = 0.8392857$$

$$a_0 = \bar{y} - a_1\bar{x}$$

$$= 3.428571 - 0.8392857\times 4 = 0.07142857$$

$$Y = 0.07142857 + 0.8392857\ x$$

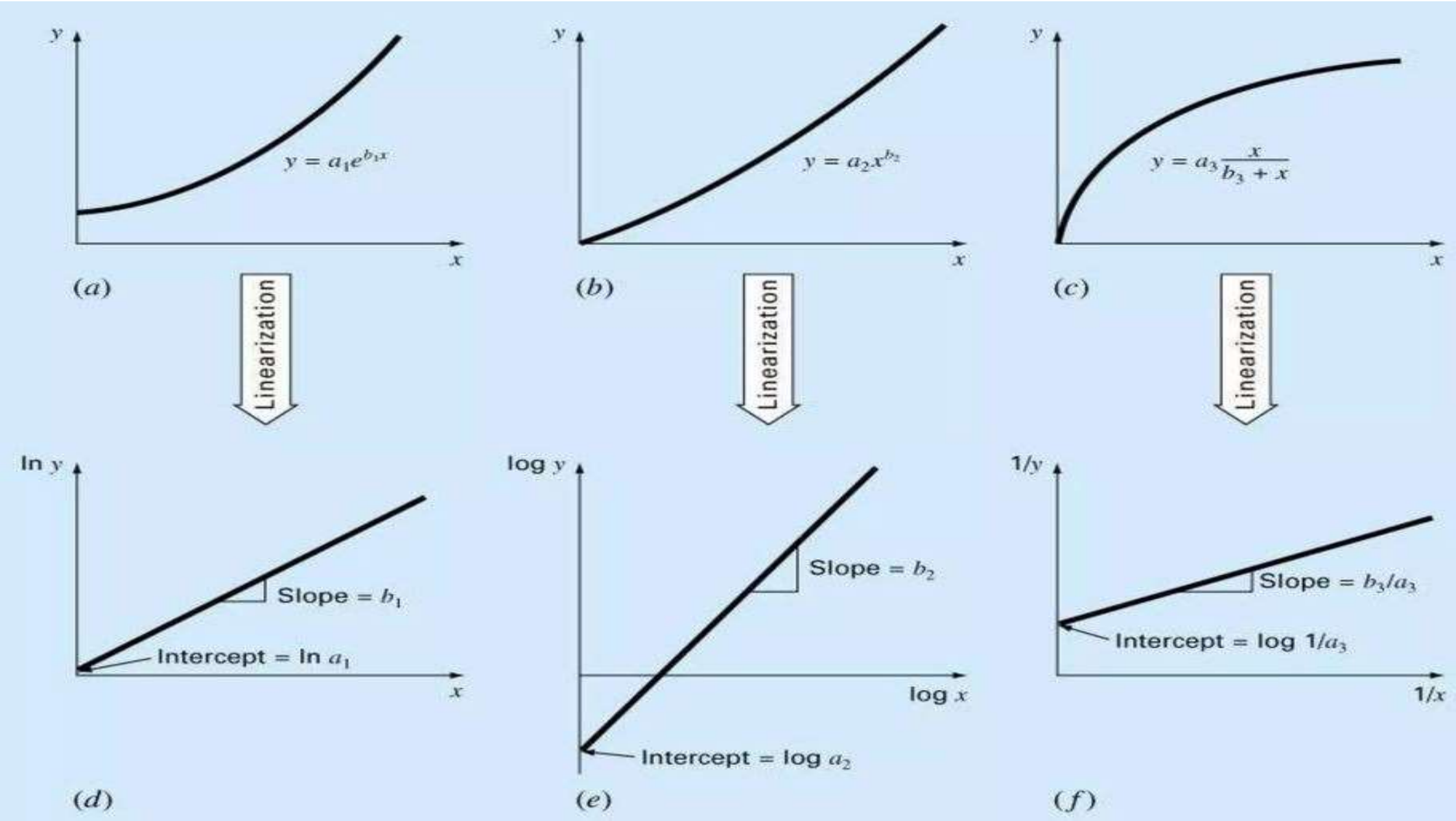
fppt.com

COEFFICIENT OF CORELATION

The coefficient of correlation is given by

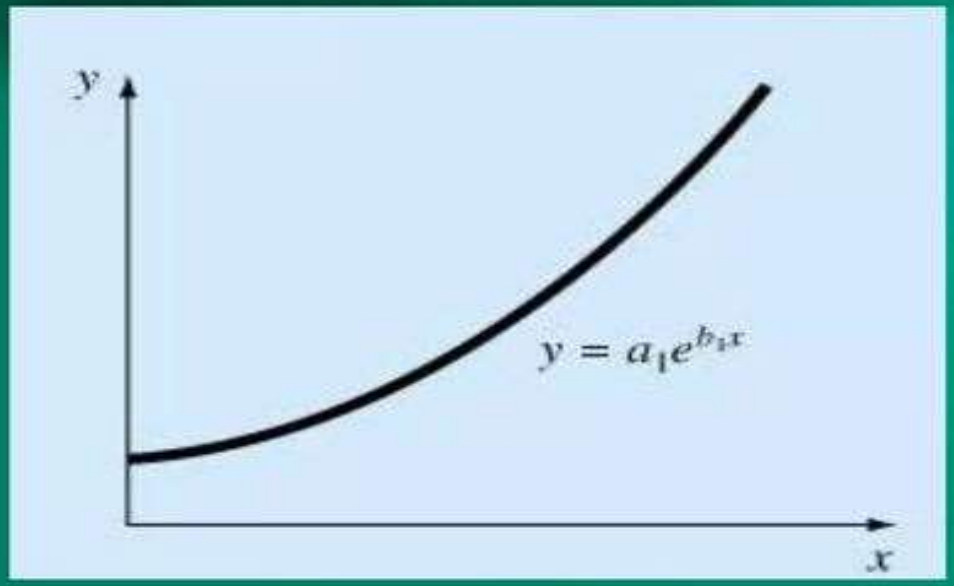
$$r = \frac{N(\sum x_i Y_i) - (\sum x_i)(\sum Y_i)}{\sqrt{N(\sum x_i^2) - (\sum x_i)^2}\sqrt{N\sum Y_i^2 - (\sum Y_i)^2}}$$

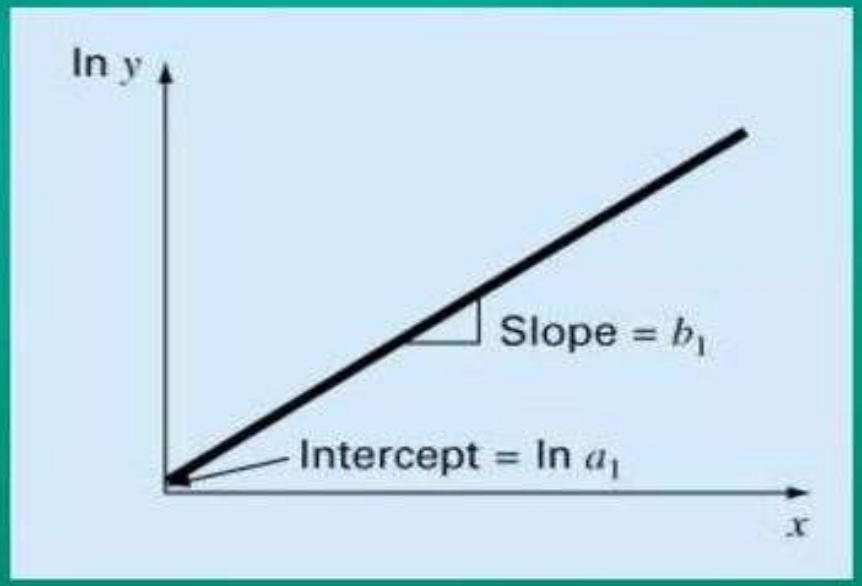
For a perfect correlation r = 1.0. If r is between 0.6 and 1.0, it is generally taken as a good correlation.


y
$y = a_1 e^{b_1 x}$
x
(a)
Linearization
y
$y = a_2 x^{b_2}$
x
(b)
Linearization
y
$y = a_3 \frac{x}{b_3 + x}$
x
(c)
Linearization
ln y
Slope = b_1
Intercept = ln a_1
x
(d)
log y
Slope = b_2
log x
Intercept = log a_2
(e)
1/y
Slope = b_3/a_3
Intercept = log $1/a_3$
1/x
(f)

Linearization of Nonlinear Relationships

1. The exponential equation.



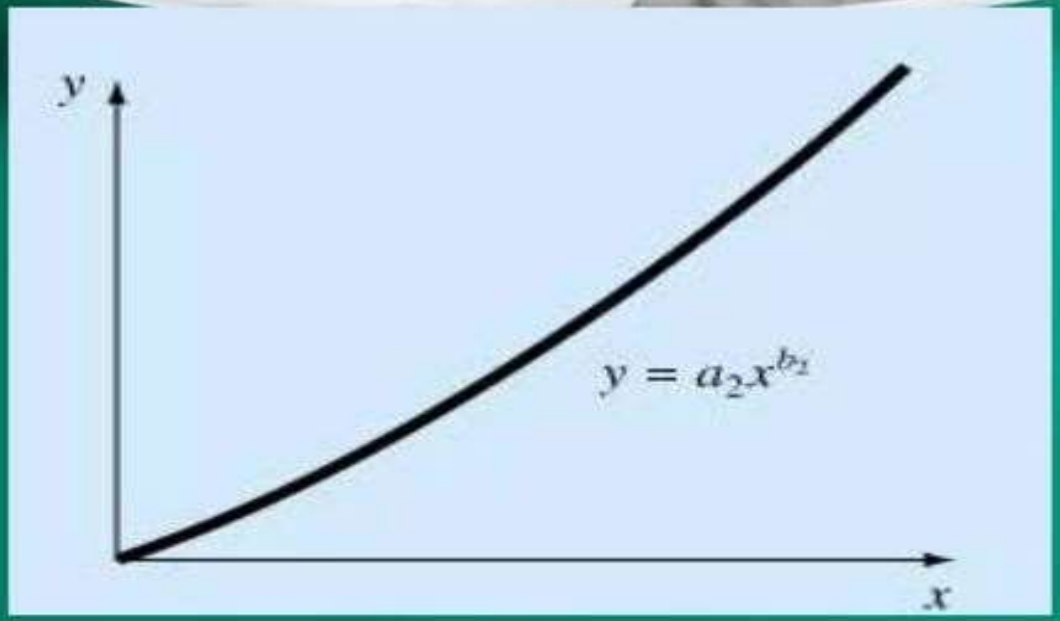


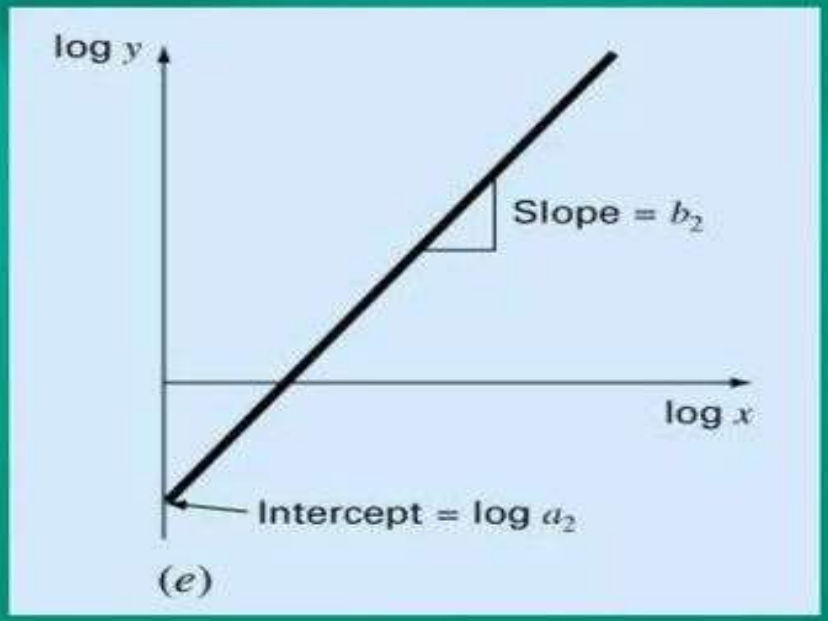
$$\ln y = \ln a_1 + b_1 x$$

$$y^* = a_o + a_1 x$$

Linearization of Nonlinear Relationships

2. The power equation





$$\log y = \log a_2 + b_2 \log x$$

$$y^* = a_o + a_1 x^*$$

Example

Fit the following Equation:

$$y = a_2 x^{b_2}$$

To the data in the following table:

x_i	y_i	$X^*=\log x_i$	$Y^*=\log y_i$
1	0.5	0	0.602
2	1.7	0.301	0.753
3	3.4	0.301	0.699
4	5.7	.226	0.922
5	8.7	.447	2.079
15	19.7	.534	2.141

$$\log y = \log(a_2 x^{b_2})$$

$$\log y = \log a_2 + b_2 \log x$$

$$\text{let } Y^* = \log y,\ X^* = \log x,$$

$$a_0 = \log a_2,\ a_1 = b_2$$

$$Y^* = a_0 + a_1 X^*$$

Example

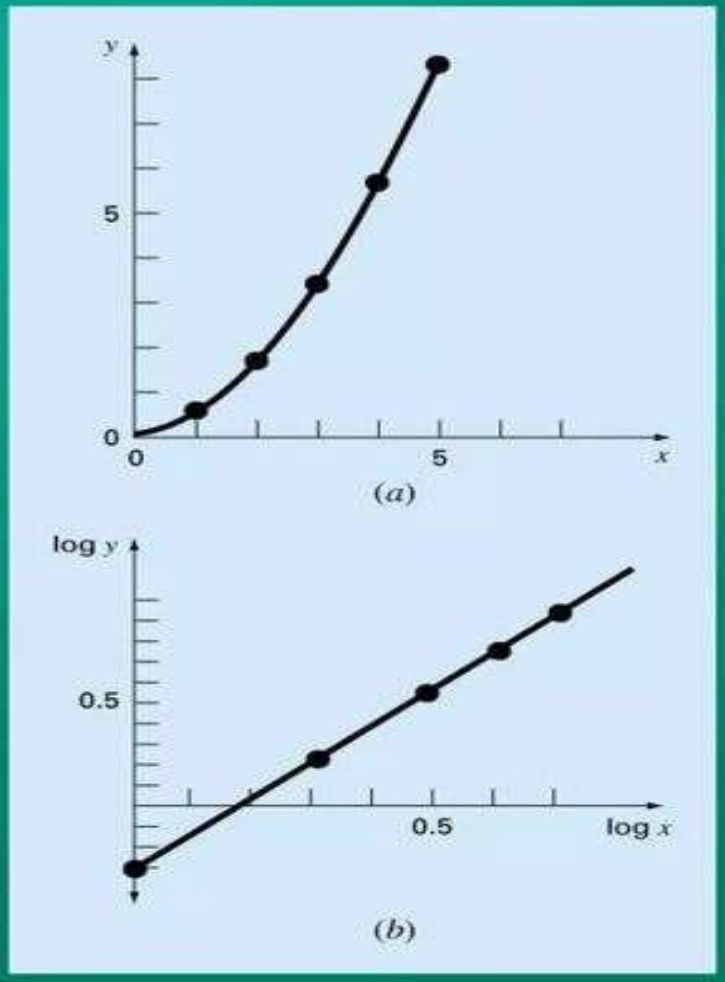
Xi	Yi	X^*_i=Log(X)	Y^*_i=Log(Y)	X*Y*	X*^2
1	0.5	0.0000	-0.3010	0.0000	0.0000
2	1.7	0.3010	0.2304	0.0694	0.0906
3	3.4	0.4771	0.5315	0.2536	0.2276
4	5.7	0.6021	0.7559	0.4551	0.3625
5	8.4	0.6990	0.9243	0.6460	0.4886
Sum	**15**	**19.700**	**2.079**	**2.141**	**1.424**	**1.169**

$$\begin{cases} a_1 = \dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2} = \dfrac{5\times 1.424 - 2.079\times 2.141}{5\times 1.169 - 2.079^2} = 1.75 \\ a_0 = \bar{y} - a_1\bar{x} = 0.4282 - 1.75\times 0.41584 = -0.334 \end{cases}$$

Linearization of Nonlinear Functions: Example

$\log y = -0.334 + 1.75 \log x$

$$y = 0.46x^{1.75}$$



Next, suppose we want to fit a curve

$$y = ae^{bx} \tag{7.13}$$

to a given set of data. Taking logarithm of both sides of Eq.(7.13), we get

$$\ln y = \ln a + bx \tag{7.14}$$

This is similar to Eg. (7.1) and we obtain the following expressions for a and b:

$$b = \frac{N(\sum x_i \ln Y_i) - (\sum x_i)(\sum \ln Y_i)}{N\sum (x_i)^2 - (\sum x_i)^2} \tag{7.15a}$$

$$\ln a = \frac{1}{N}(\sum \ln Y_i - b\sum x_i) \tag{7.15b}$$

The coefficient of

$$r = \frac{N(\sum x_i \ln Y_i) - (\sum x_i)(\sum \ln Y_i)}{\left(\sqrt{N\sum (x_i)^2 - (\sum x_i)^2}\right)\left(\sqrt{N\sum (\ln Y_i)^2 - (\sum \ln Y_i)^2}\right)}$$

ample 7.3

Fit the curve $y = ae^{bx}$ to the following set of data:

x	1	2	3	4	5	6	7	8
Y	15.3	20.5	27.4	36.6	49.1	65.6	87.8	117.6

ution We have N = 8. The various quantities required by Eqs.(7.15) and (7.16) are given **ow.**

x_i	Y_i	x_i^2	$\ln Y_i$	$(\ln Y_i)^2$	$x_i \ln Y_i$
1	15.3	1	2.7279	7.4410	2.7279
2	20.5	4	3.0204	9.1230	6.0408
3	27.4	9	3.3105	10.9838	9.9316
4	36.6	16	3.6000	12.9603	14.4002
5	49.1	25	3.8939	15.1621	19.4693
6	65.6	36	4.1836	17.5023	25.1015
7	87.8	49	4.4751	20.0262	31.3254
8	117.6	64	4.7673	22.7270	38.1383
$\Sigma =$ 36		204	29.9787	115.9260	147.1350

$$\therefore b = \frac{8 \times 147.1350 - 36 \times 29.9787}{8 \times 204 - 36^2} = \frac{97.8468}{336} = 0.2912$$

$$\ln a = \frac{1}{8}(29.9787 - 0.2912 \times 36) = 2.4369$$

$$\therefore a = e^{2.4369} = 11.4374$$

So, the fitted curve is

$$y = 11.4374e^{0.2912x}$$

The correlation coefficient r is given by

$$r = \frac{8 \times 147.1250 - 36 \times 29.9787}{\sqrt{8 \times 204 - 36^2} \times \sqrt{8 \times 115.9260 - 29.9787^2}} = \frac{97.8468}{98.1751} = 0.9967$$