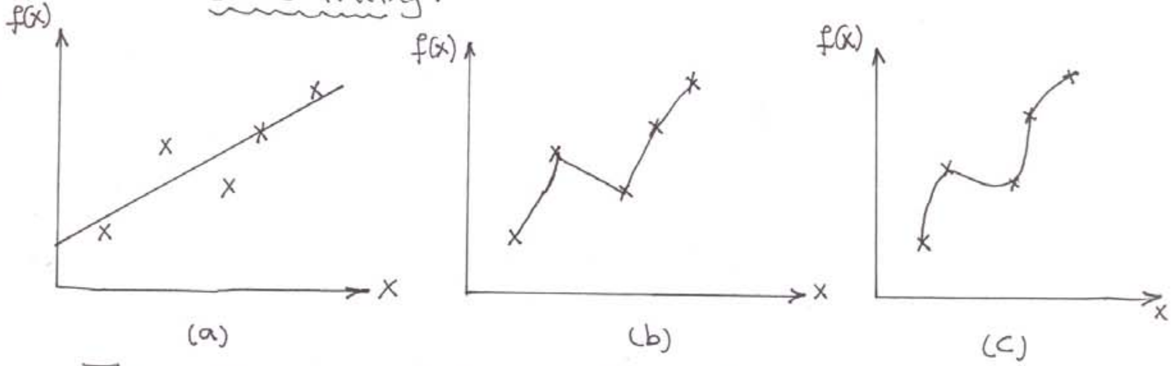
Curves Fitting

One way to do Fit the data is to compute values of the function at a number of descrete values along the range of interest. Then, a simpler function may be derived to fit these values. Both of these applications are known as curve fitting.



Three attempts to fit a best curve through five data points
ⓐ Least-squares regression ⓑ linear interpolation ⓒ curvilinear interpolation

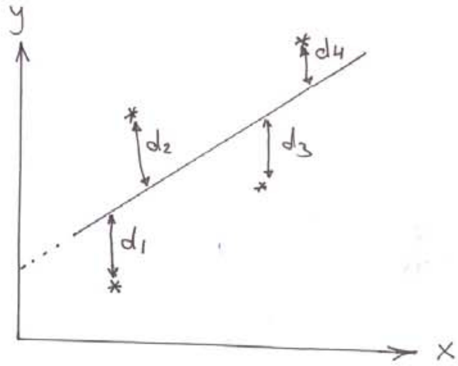
Least-Squares Regression

linear Regression

The simplest example of a least-squares approximation is fitting a straight line to a set of paired observations: $(x_1, y_1), (x_2, y_2) ---- (x_n, y_n)$ the mathmatical expression for the straight line is

$$\boxed{\bar{y} = a_0 + a_1 x}$$

$$\text{Deviation} = d = y - \acute{y}$$



where

$$d_1 = y_1 - \bar{y}_1 = y_1 - f(x_1)$$
$$d_2 = y_2 - \bar{y}_2 = y_2 - f(x_2)$$
$$d_3 = y_3 - \bar{y}_3 = y_3 - f(x_3)$$
$$\vdots \qquad \vdots$$
$$d_m = y_m - \bar{y}_m = y_m - f(x_m)$$

m = No. of points

Now we applied to minimize the sum of the squares of the residuals between the measured y and the y calculated with the linear model

$$S = \sum_{i=1}^{m} d_i^2 = \sum_{i=1}^{m} (y_i - \bar{y}_i)^2 = \sum_{i=1}^{m} (y_i - a_0 - a_1 x_i)^2$$

to find a_0, a_1 must $\frac{\partial S}{\partial a_0} = \frac{\partial S}{\partial a_1} = 0$ minimum values

Then,

$$\frac{\partial S}{\partial a_0} = -2 \sum (y_i - a_0 - a_1 x_i) = 0$$

$$\frac{\partial S}{\partial a_1} = -2 \sum \left[(y_i - a_0 - a_1 x_i) x_i \right] = 0$$

Now, realizing that $\sum_{i=1}^{m} a_0 = m a_0$, we can express the equations as a set of two simultaneous linear equations with two unknowns (a_0 and a_1)

$$n a_0 + a_1 \cdot (\sum x_i) = (\sum y_i)$$
$$(\sum x_i) a_0 + a_1 \cdot (\sum x_i^2) = (\sum x_i y_i)$$

there are called the normal equations they can be solved as

$$a_0 = \frac{\sum y_i \cdot \sum x_i^2 - \sum x_i \cdot \sum x_i \cdot y_i}{m \sum x_i^2 - (\sum x_i)^2} \qquad a_1 = \frac{m \sum x_i y_i - \sum x_i \cdot \sum y_i}{m \sum x_i^2 - (\sum x_i)^2}$$

ex. Use linear regression to fit the following experimental data :

x:	1	3	4	6	8	9	11	14
y:	1	2	4	4	5	7	8	9

sol.

Let $\bar{y} = a_0 + a_1 x$ Then $m = 8$

i	x_i	y_i	x_i^2	$x_i \cdot y_i$
1	1	1	1	1
2	3	2	9	6
3	4	4	16	16
4	6	4	36	24
5	8	5	64	40
6	9	7	81	63
7	11	8	121	88
8	14	9	196	126
Σ	56	40	564	364

$$a_0 = \frac{\Sigma y_i \cdot \Sigma x_i^2 - \Sigma x_i \cdot \Sigma x_i y_i}{m \Sigma x_i^2 - (\Sigma x_i)^2}$$

$$\Rightarrow a_0 = \frac{40 * 524 - 56 * 364}{8 * 524 - 56^2}$$

$$\Rightarrow a_0 = 6/11$$

also

$$a_1 = \frac{m \Sigma x_i y_i - \Sigma x_i \Sigma y_i}{m \Sigma x_i^2 - (\Sigma x_i)^2}$$

$$\Rightarrow a_1 = \frac{8 * 364 - 56 * 40}{8 * 524 - 56^2}$$

$$\Rightarrow a_1 = 7/11 \qquad \Rightarrow \bar{y} = \frac{6}{11} + \frac{7}{11} x \quad \text{or } 11\bar{y} - 7x = 6$$

ex: Fit a straight line to the following data

x :	1	2	3	4	5	6	7
y :	0.5	2.5	2.0	4.0	3.5	6.0	5.5

sol: the following quantities can be computed

$m = 7$, $\Sigma x_i y_i = 119.5$, $\Sigma x_i^2 = 140$, $\Sigma x_i = 28$

$\Sigma y_i = 24$

Then $a_0 = 0.07142857$

$a_1 = 0.8392857$

Linearization of nonlinear relationships

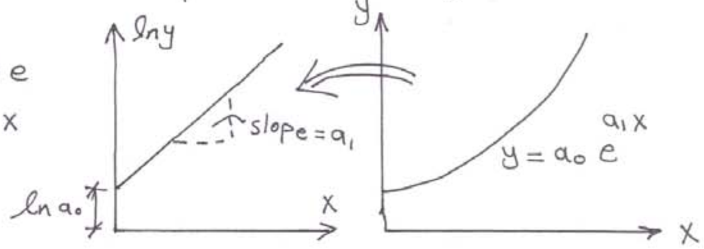
Linear regression provides a powerful technique for fitting a best line to data.

Case ① $\Rightarrow$ Exponential equation $\quad y = a_0 e^{a_1 X}$

then,

$\ln y = \ln a_0 + a_1 X \ln e$

$\Rightarrow \ln y = \ln a_0 + a_1 X$

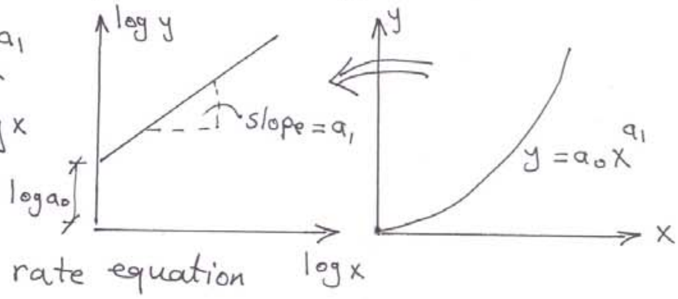


Case ② $\Rightarrow$ power equation $\quad y = a_0 X^{a_1}$

then

$\log y = \log a_0 + \log X^{a_1}$

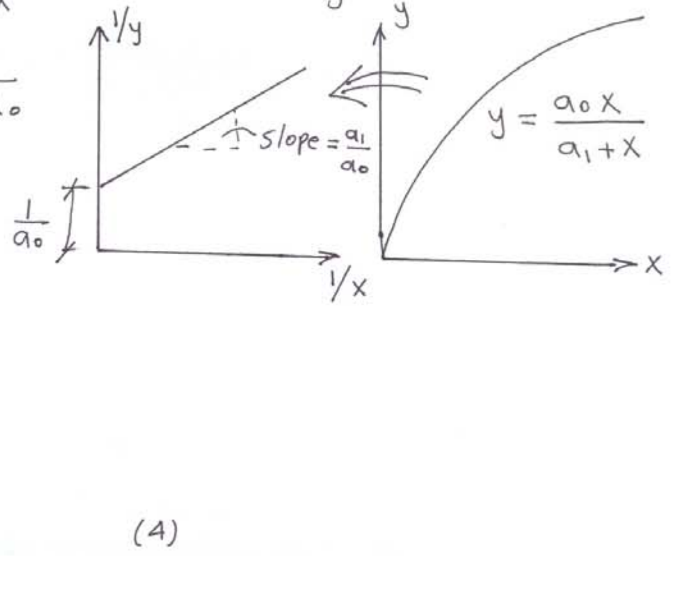
$\log y = \log a_0 + a_1 \log X$



Case ③ $\Rightarrow$ Growth-rate equation

$$y = a_0 \cdot \frac{X}{a_1 + X} \qquad \text{by inverting}$$

$$\Rightarrow \frac{1}{y} = \frac{a_1}{a_0} \qquad + \frac{1}{a_0}$$



ex: Fit the data in the following table using a logarithmic transformation of the data

x :	1	2	3	4	5
y :	0.5	1.7	3.4	5.7	8.4

sol: logarithmic transformation $\Rightarrow$ applied for power eq.

$$y = a_0 x^{a_1} \xrightarrow[\text{to}]{\text{Linearization}} \log y = \log a_0 + a_1 \log x$$

then

$$Y = a_0 + a_1 X$$

	x	y	log x	log y	$(\log x)^2$	$\log x \cdot \log y$
	1	0.5	0.000	-0.301	0.00	0.00
	2	1.7	0.301	0.226	0.090	0.080
	3	3.4	0.477	0.534	0.227	0.254
	4	5.7	0.602	0.753	0.362	0.453
	5	8.4	0.699	0.922	0.488	0.644
Σ			2.079	2.134	1.167	1.431

then
$$a_0 = \frac{\Sigma \log y \cdot \Sigma \log^2 x - \Sigma \log x \cdot \Sigma \log x \cdot \log y}{m \cdot \Sigma \log^2 x - (\Sigma \log x)^2} = \frac{2.134 \times 1.167 - 2.079 \times 1.431}{5 \times 1.167 - 2.079^2}$$

$$\Rightarrow a_0 = -0.320 = \log a_0. \quad \Rightarrow a_0 = 0.478$$

also
$$a_1 = \frac{m \cdot \Sigma \log x \cdot \log y - \Sigma \log x \cdot \Sigma \log y}{m \; \Sigma \log x^2 - (\Sigma \log x)^2} = \frac{5 \times 1.431 - 2.079 \times 2.134}{5 \times 1.167 - 2.079^2}$$

$$\Rightarrow a_1 = 1.796 = a_1$$

then $y = 0.478 \, X^{1.796}$ or $\log y = -0.32 + 1.796 \log x$

Polynomial Regression

Let $\bar{y} = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$

$$\Rightarrow S^t = \sum_{i=1}^{m} [y_i - \bar{y}] = \sum_{i=1}^{m} [y_i - a_0 - a_1 x_i - a_2 x_i^2 - \cdots - a_n x_i^n]$$

$$\Rightarrow \frac{\partial S}{\partial a_0} = -2 \sum [y_i - a_0 - a_1 x_i - a_2 x_i^2 - \cdots a_n x_i^n] = 0$$

$$\Rightarrow \frac{\partial S}{\partial a_1} = -2 \sum [x_i (y_i - a_0 - a_1 x_i - a_2 x_i^2 - \cdots - a_n x_i^n)] = 0$$

$$\Rightarrow \frac{\partial S}{\partial a_2} = -2 \sum [x_i^2 (y_i - a_0 - a_1 x_i - a_2 x_i^2 - \cdots a_n x_i^n)] = 0$$

$$\vdots \qquad \vdots$$

$$\frac{\partial S}{\partial a_n} = -2 \sum [x_i^n (y_i - a_0 - a_1 x_i - a_2 x_i^2 - \cdots - a_n x_i^n)] = 0$$

then

$$a_0 m + a_1 \sum x_i + a_2 \sum x_i^2 + \cdots + a_n \sum x_i^n = \sum y_i$$
$$a_0 \sum x_i + a_1 \sum x_i^2 + a_2 \sum x_i^3 + \cdots + a_n \sum x_i^{n+1} = \sum x_i y_i$$
$$a_0 \sum x_i^2 + a_1 \sum x_i^3 + a_2 \sum x_i^4 + \cdots + a_n \sum x_i^{n+2} = \sum x_i^2 \cdot y_i$$
$$\vdots$$
$$\Rightarrow a_0 \sum x_i^n + a_1 \sum x_i^{n+1} + a_2 \sum x_i^{n+2} + \cdots + a_n \sum x_i^{2n} = \sum x_i^n \cdot y_i$$

<u>ex</u>: fit a second-order polynomial to the data in the following table:

x :	0	1	2	3	4	5
y :	2.1	7.7	13.6	27.2	40.9	61.1

<u>sol</u> $\bar{y} = a_0 + a_1 x + a_2 x^2$

then $n = 2$, $m = 6$

$\sum x_i = 15$, $\sum y_i = 152.6$, $\sum x_i^2 = 55$, $\sum x_i^3 = 225$

$\sum x_i^4 = 979$, $\sum x_i y_i = 585.6$, $\sum x_i^2 \cdot y_i = 2488.8$

$\Rightarrow$

$$\Rightarrow \quad 6a_0 + 15a_1 + 55a_2 = 152.6$$
$$15a_0 + 55a_1 + 225a_2 = 585.6$$
$$55a_0 + 225a_1 + 979a_2 = 2488.8$$

by Guass method

$a_0 = 2.47857$

$a_1 = 2.35929$ $\quad \Rightarrow y = 2.47857 + 2.35929x + 1.8607x^2$

$a_2 = 1.86071$