Chapter 17: Least squares method

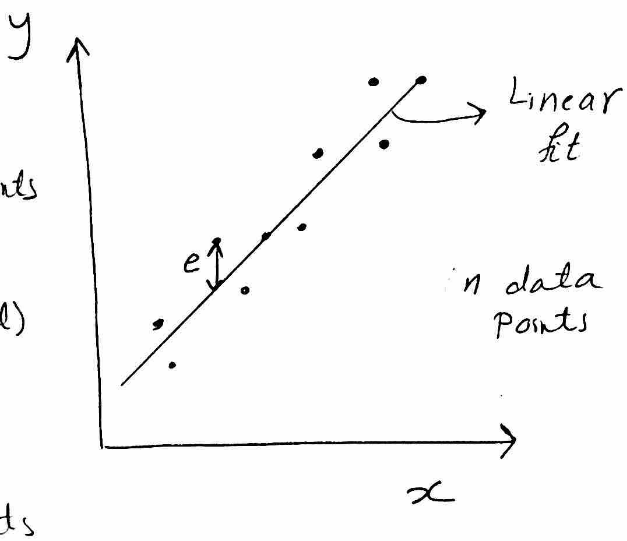
* Here we will learn how to obtain a curve fit (equation) of some data points

1- Linear curve fit

a_0, a_1 constants

$$y(x) = a_0 + a_1 x + \underbrace{e}_{\text{error (residual)}}$$

$e = y - a_0 - a_1 x$ ← For one data points



For all data points

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)$$, n # of data points

For least squares method

$$\sum_{i=1}^{n} e_i^2 = \underbrace{S_r}_{\text{Sum of the squares of the residuals}} = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

* Our goal ~~here~~ is to minimize S_r (or $\sum_{i=1}^{n} e_i^2$).
(To find best fit)

* In other words, we need to obtain a_0 and a_1 that minimize S_r $\left(\sum_{i=1}^{n} e_i^2\right)$.

* For this reason $\Rightarrow$ $\frac{\partial S_r}{\partial a_0} = 0$ and $\frac{\partial S_r}{\partial a_1} = 0$

$$\frac{\partial S_r}{\partial a_0} = -\cancel{2} \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} a_0 + \sum_{i=1}^{n} a_1 x_i = \sum_{i=1}^{n} y_i$$

$$n a_0 + \sum_{i=1}^{n} a_1 x_i = \sum_{i=1}^{n} y_i \quad \text{— Eq (1)}$$

$$\frac{\partial S_r}{\partial a_1} = -\cancel{2} x_i \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i) = 0$$

$$\Rightarrow \sum_{i=1}^{n} a_0 x_i + \sum_{i=1}^{n} a_1 x_i^2 = \sum_{i=1}^{n} x_i y_i \quad \text{— Eq (2)}$$

Matrix Form

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{Bmatrix}$$

↗ Normal Equations

* Solve normal equations, to get:

$$a_0 = \bar{y} - a_1\bar{x}$$, $\bar{x}$ and $\bar{y}$ are the mean values

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

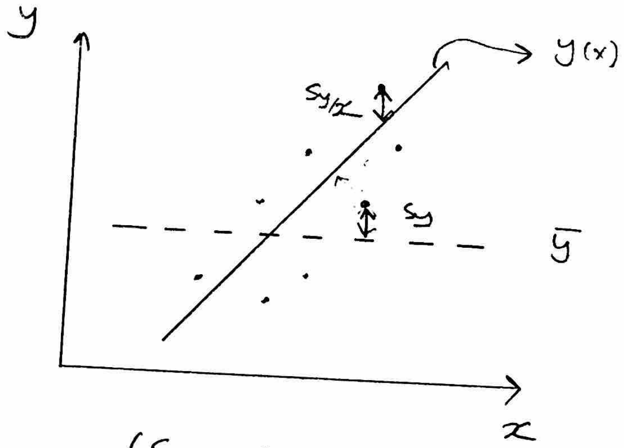
$$a_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

* Standard deviation (S_y)

measures the spread around the mean

$$S_y = \sqrt{\frac{S_t}{n-1}}$$, $S_t: \sum_{i=1}^{n}(y_i - \bar{y})$, n # of data points

The total sum of the squares around the mean.



* Variance (S_y^2)

$$S_y^2 = \frac{S_t}{n-1}$$

* Standard error of the estimate ($S_{y/x}$)

measures the spread around the fit line

$$S_{y/x} = \sqrt{\frac{S_r}{n-(m+1)}}$$, m: fit order (linear m=1)

$$\Rightarrow S_{y/x} = \sqrt{\frac{S_r}{n-2}}$$

n: # of data points

* **Coefficient of determination (r^2)**

→ measures the goodness of the fit.

$$r^2 = \frac{S_t - S_r}{S_t}$$

For linear curve fit

$$r^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2 - \sum_{i=1}^{n}(y_i - a_0 - a_1 x_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})}$$

* **Correlation Coefficient ($r = \sqrt{r^2}$)**

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)}}$$

$$0 < r^2 < 1 \rightarrow \text{good fit}$$

$$0 < r < 1$$

bad fit

Higher r and r^2 ⇒ good

Lower r and r^2 ⇒ Bad

Example

$n = 7$

x_i	y_i	$x_i y_i$	x_i^2	$(y_i - \bar{y})^2$	$(y_i - a_0 - a_1 x_i)^2$
1	0.5	0.5	1	8.58	0.17
2	2.5	5.0	4	0.86	0.56
3	2.0	4.0	9	2.04	0.35
4	4.0	16.0	16	0.326	0.33
5	3.5	17.5	25	0.005	0.59
6	6.0	36.0	36	6.61	0.80
7	5.5	38.5	49	4.29	0.20
Σx_i 28	Σy_i 24	$\Sigma x_i y_i$ 119.5	$\sum_{i=1}^{n} x_i^2$ 140	$S_t = \Sigma (y_i - \bar{y})^2$ 22.714	$S_r = \Sigma (y_i - a_0 - a_1 x_i)^2$ 2.991

Find ① Linear Curve fit ($y = a_0 + a_1 x$)

② S_y and S_y^2

③ $S_{y/x}$

④ r^2 and r

(1) Linear Curve fit

$$\begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{Bmatrix}$$

$$\begin{bmatrix} 7 & 28 \\ 28 & 140 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix} = \begin{Bmatrix} 24 \\ 119.5 \end{Bmatrix} \xrightarrow{\text{Solve}} \begin{matrix} a_0 = 0.0714 \\ a_1 = 0.8393 \end{matrix}$$

$$\Rightarrow \boxed{y(x) = 0.0714 + 0.8393\,x}$$ Linear Curve fit

(2) S_y, S_y^2

$$S_y = \sqrt{\frac{S_t}{n-1}}, \quad S_t = \sum_{i=1}^{n} (y_i - \bar{y})^2, \quad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{24}{7} = 3.43$$

$$S_y = \sqrt{\frac{22.714}{7-1}} = 1.946$$

$$S_y^2 = (1.946)^2 = 3.787$$

, For point (1)

$y_i = 0.5$, $\bar{y} = 3.43$

$$(y_i - \bar{y})^2 = (0.5 - 3.43)^2 = 8.58$$

(3) $S_{y/x}$

$$S_{y/x} = \sqrt{\frac{S_r}{n-2}}, \quad S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

$$= \sqrt{\frac{2.991}{7-2}} = 0.77,$$

For point (1)

$y_i = 0.5$, $x_i = 1$ $\qquad a_0 = 0.0714$, $a_1 = 0.8393$

$$(y_i - a_0 - a_1 x_i)^2 = (0.5 - 0.0714 - 0.8393(1))^2 = 0.17$$

(4) r^2, r

$$r^2 = \frac{S_t - S_r}{S_t} = \frac{22.714 - 2.991}{22.714} = 0.87$$

$$r = \sqrt{r^2} = \sqrt{0.87} = 0.93$$

* **Application for linear curve fit**

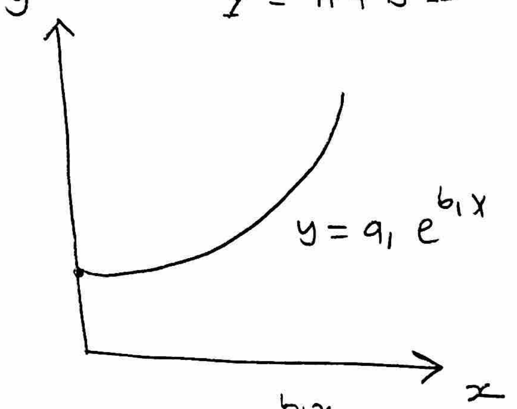
① Exponential function

$$y(x) = a_1 e^{b_1 x}$$ ⇒ linearization take ln

$$\ln y(x) = \ln a_1 + \ln e^{b_1 x}$$

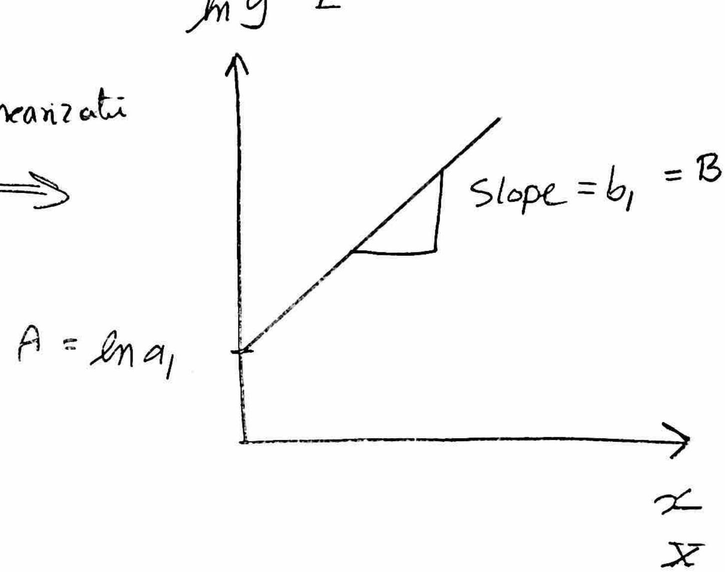
$$\ln y = \ln a_1 + b_1 x$$ → Linear Equation

$$Y = A + B X$$



$y(x) = a_1 e^{b_1 x}$

Linearization ⇒

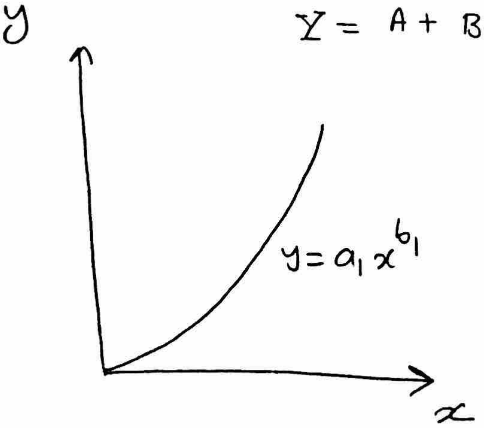


② Power Equation

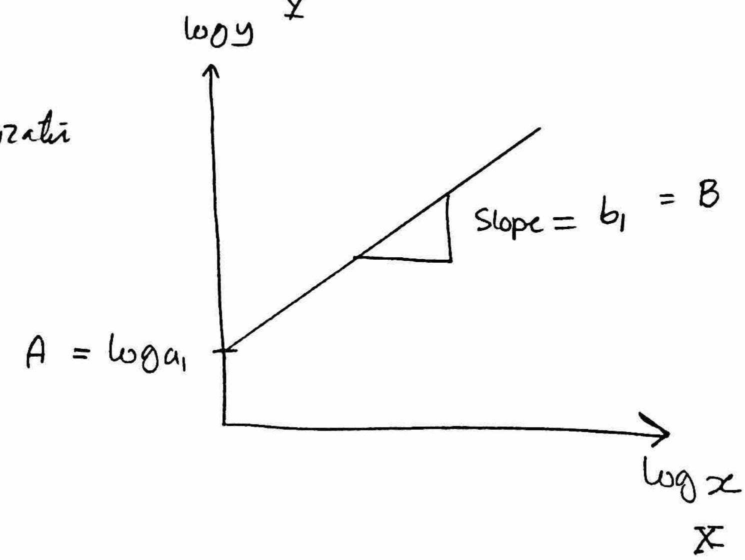
$$y(x) = a_1 x^{b_1}$$ take log (or ln)

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

$$Y = A + B X$$



linearization ⇒



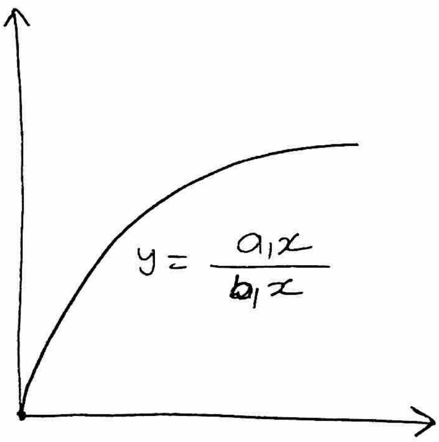
3 Growth - function

$$y(x) = \frac{a_1 x}{b_1 + x} \quad , \quad \text{take } (\ )^{-1} \quad \text{inverse}$$

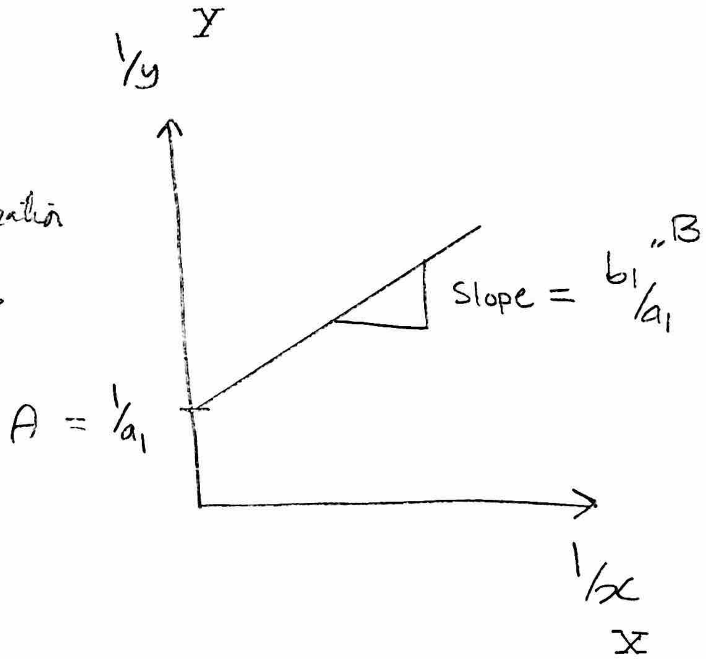
$$\frac{1}{y} = \frac{b_1 + x}{a_1 x} = \frac{b_1}{a_1 x} + \frac{x}{a_1 x}$$

$$\Rightarrow \quad \frac{1}{y} = \frac{b_1}{a_1} \cdot \frac{1}{x} + \frac{1}{a_1}$$

$$Y = A + BX$$



linearization $\Longrightarrow$



See Problems 17.10 — 17.15 For more examples

Chapter 17 : Least Squares Method

* Linear Curve fit $\quad y(x) = a_0 + a_1 x$

* Second order fit $\quad y(x) = a_0 + a_1 x + a_2 x^2$

  ↳ How to obtain ?

$$y = a_0 + a_1 x + a_2 x^2 + \underbrace{e}_{\text{residual}}$$

$$e = y - a_0 - a_1 x - a_2 x^2$$

Apply LSM

$$\sum_{i=1}^{n} e_i^2 = S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2$$

* Now, we need to find a_0, a_1 and a_2 $\xrightarrow{\text{minimize}}$ $S_r \left(\sum_{i=1}^{n} e_i^2 \right)$

$$① \leftarrow \quad \frac{\partial S_r}{\partial a_0} = -2 \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2) = 0$$

$$② \leftarrow \quad \frac{\partial S_r}{\partial a_1} = -2 \sum_{i=1}^{n} x_i (y_i - a_0 - a_1 x_i - a_2 x_i^2) = 0$$

$$③ \leftarrow \quad \frac{\partial S_r}{\partial a_2} = -2 \sum_{i=1}^{n} x_i^2 (y_i - a_0 - a_1 x_i - a_2 x_i^2) = 0$$

Thus,

$$n a_0 + \left(\sum_{i=1}^{n} x_i\right) a_1 + \left(\sum_{i=1}^{n} x_i^2\right) a_2 = \sum_{i=1}^{n} y_i$$

$$\left(\sum_{i=1}^{n} x_i\right) a_0 + \left(\sum_{i=1}^{n} x_i^2\right) a_1 + \left(\sum_{i=1}^{n} x_i^3\right) a_2 = \sum_{i=1}^{n} x_i y_i$$

$$\left(\sum_{i=1}^{n} x_i^2\right) a_0 + \left(\sum_{i=1}^{n} x_i^3\right) a_1 + \left(\sum_{i=1}^{n} x_i^4\right) a_2 = \sum_{i=1}^{n} x_i^2 y_i$$

Matrix Form

Normal Equations

$$\begin{bmatrix} n & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \Sigma y_i \\ \Sigma x_i y_i \\ \Sigma x_i^2 y_i \end{Bmatrix}$$

3 Equations and 3 unknowns, solve $\longrightarrow$ a_0, a_1 and a_2

$$S_{y/x} = \sqrt{\frac{S_r}{n-(m+1)}}, \quad m=2 \Rightarrow S_{y/x} = \sqrt{\frac{S_r}{n-3}}$$

$$r^2 = \frac{S_t - S_r}{S_t}, \quad S_t = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$S_r = \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)^2$$

This procedure can be followed for any other polynomial curve fit

Example Find $y = a_0 + a_1 x + a_2 x^2$

x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$
0	2.1	0	0	0	0	0
1	7.7	1	1	1	7.7	7
2	13.6	4	8	16	27.2	53.6
3	27.2	9	27	81	81.6	244.8
4	40.9	16	64	256	163.6	654.4
5	61.1	25	125	625	305.5	1527.5
$\Sigma x_i = 15$	Σy_i 152.6	Σx_i^2 55	Σx_i^3 225	Σx_i^4 979	$\Sigma x_i y_i$ 585.6	$\Sigma x_i^2 y_i$ 2488.8

$$\begin{bmatrix} n & \Sigma x_i & \Sigma x_i^2 \\ \Sigma x_i & \Sigma x_i^2 & \Sigma x_i^3 \\ \Sigma x_i^2 & \Sigma x_i^3 & \Sigma x_i^4 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \Sigma y_i \\ \Sigma x_i y_i \\ \Sigma x_i^2 y_i \end{Bmatrix}$$

$$\begin{bmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 152.6 \\ 585.6 \\ 2488.8 \end{Bmatrix}$$

$a_0 = 2.48$, $a_1 = 2.36$, $a_2 = 1.86$

$$y(x) = 2.48 + 2.36x + 1.86x^2$$

* Find ① $S_{y/x}$, ② S_y , ③ r^2 and ④ r

x_i	y_i	$(y_i - \bar{y})^2$	$(y_i - a_0 - a_1x_i - a_2x_i^2)^2$
0	2.1	544.44	0.14332
1	7.7	314.47	1.00286
2	13.6	140.03	1.08158
3	27.2	3.12	0.80491
4	40.9	239.22	0.61951
5	61.1	1272.11	0.09439
Σ	152.6	2513.39	3.74657

S_t $\quad$ S_r

$\bar{y} = \dfrac{\Sigma y_i}{n} = \dfrac{152.6}{6} = 25.43$

① $S_{y/x} = \sqrt{\dfrac{S_r}{n-3}} = \sqrt{\dfrac{3.75}{6-3}} = 1.12$

② $S_y = \sqrt{\dfrac{S_t}{n-1}} = \sqrt{\dfrac{2513.4}{6-1}} = 22.42$

③ $r^2 = \dfrac{S_t - S_r}{S_t} = \dfrac{2513.4 - 3.75}{2513.4} = 0.99851$

④ $r = \sqrt{r^2} = 0.999$

End of Chapter 17