CUFE, M. Sc., 2015-2016

Computers & Numerical Analysis (STR 681)

Lecture 10
CURVE FITTING

Dr. Maha Moddather
Structural Engineering Department
Faculty of Engineering – Cairo University
mahamoddather@eng.cu.edu.eg

Spring 2016

Introduction

- Data is often given for discrete values along a continuum.
- However, you may require estimates at points between the discrete values.
- The present part describes techniques to fit curves to such data to obtain intermediate estimates.
- In addition, you may require a simplified version of a complicated function.

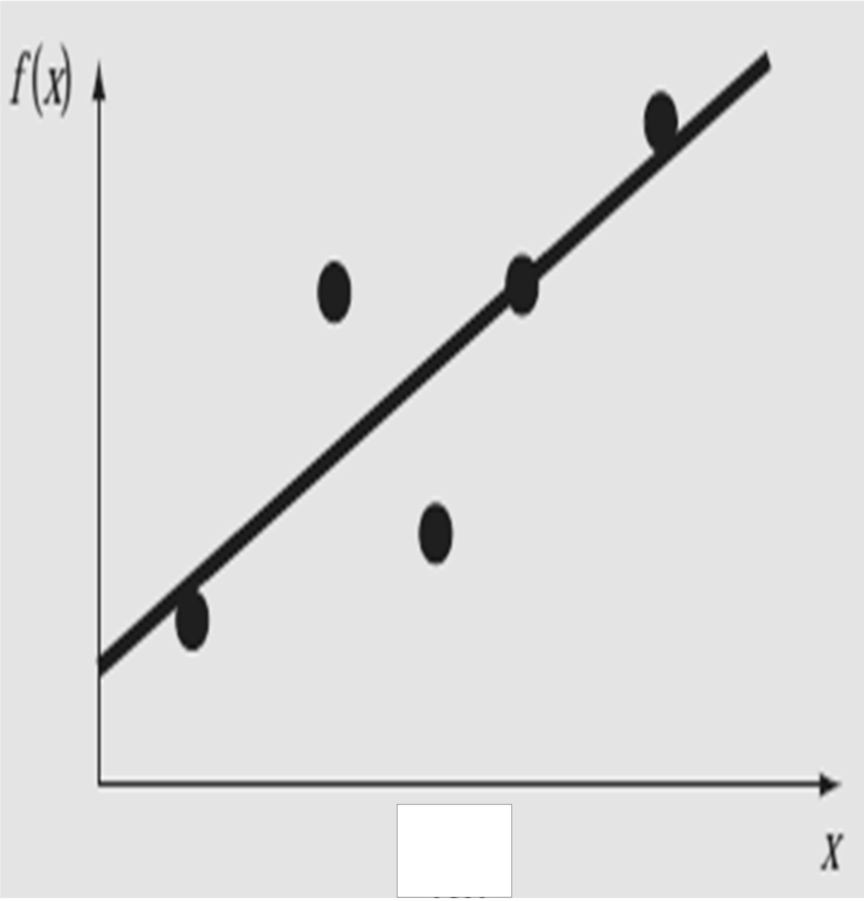
Introduction

- One way to do this is to compute values of the function at a number of discrete 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.
- There are two general approaches for curve fitting that are distinguished from each other on the basis of the amount of error associated with the data.

Introduction

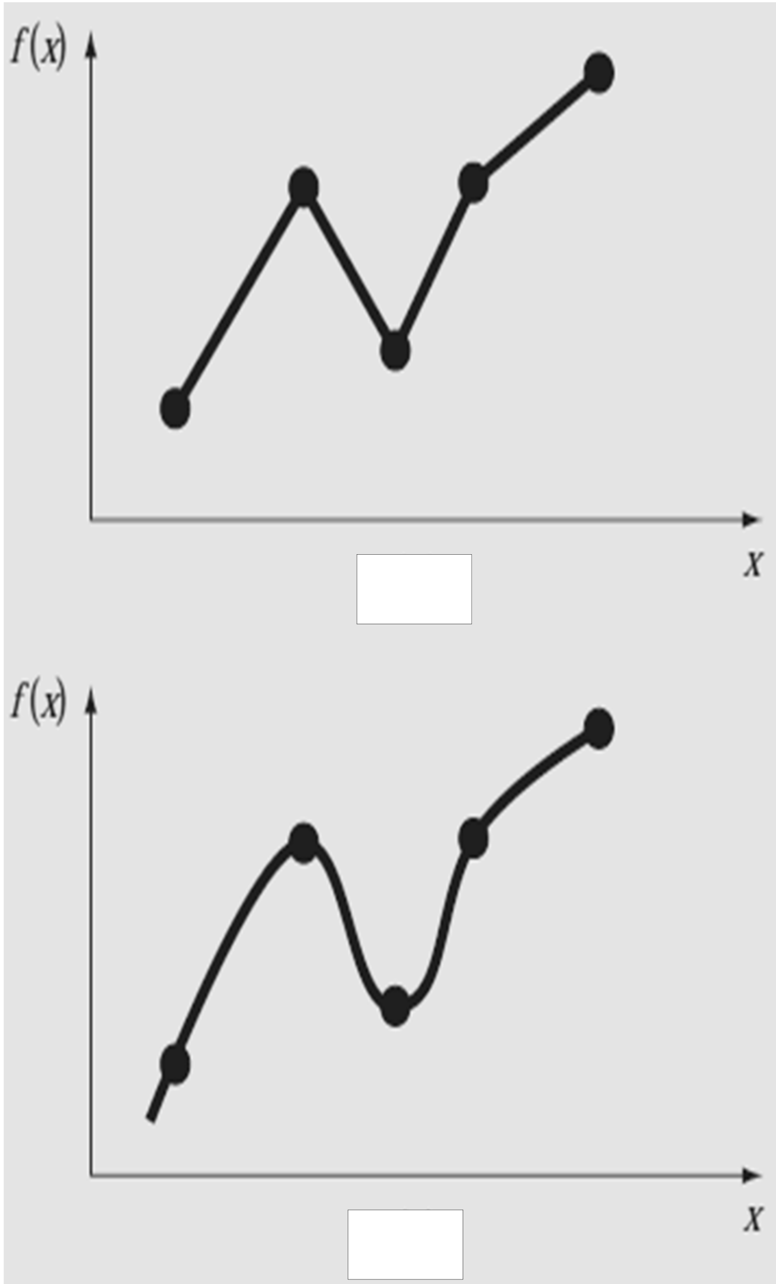
First, where the data exhibits **a significant degree of error or "noise"**, the strategy is to derive a single curve that represents the general trend of the data.

- Because any individual data point may be incorrect, we make no effort to intersect every point. Rather, **the curve is designed to follow the pattern of the points taken as a group**. One approach of this nature is called ***least-squares regression.***



Introduction

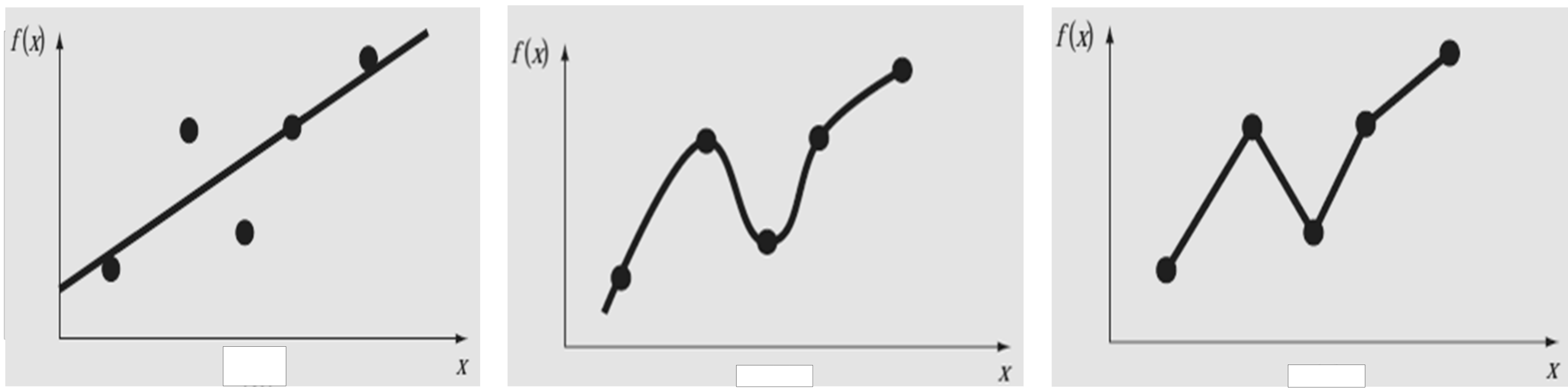
- Second, where the data is known to be **very precise**, the basic approach is to **fit a curve or a series of curves that pass directly through each of the points**. Such data usually originates from tables. Examples are values for the density of water or for the heat capacity of gases as a function of temperature.



- The estimation of values between well-known discrete points is called ***interpolation***.

Noncomputer methods

- The simplest method for fitting a curve to data is to plot the points and then sketch a line that visually conforms to the data.
- Although this is a valid option when quick estimates are required, the results are dependent on the subjective viewpoint of the person sketching the curve.



Engineering Practice

- Throughout our engineering career, we have frequent occasion to estimate intermediate values from tables.

- Although many of the widely used engineering properties have been tabulated, there are a great many more that are not available in this convenient form.

- Special cases and new problem contexts often require that you measure your own data and develop your own predictive relationships.

Engineering Practice

- ***Trend analysis*** may be used to predict or forecast values of the dependent variable. This can involve ***extrapolation*** beyond the limits of the observed data or ***interpolation*** within the range of the data.

- ***hypothesis testing:*** an existing mathematical model is compared with measured data. If the model coefficients are unknown, it may be necessary to determine values that best fit the observed data.

Curve Fitting

- Least Square Regression
- Interpolation
- Fourier Approximation

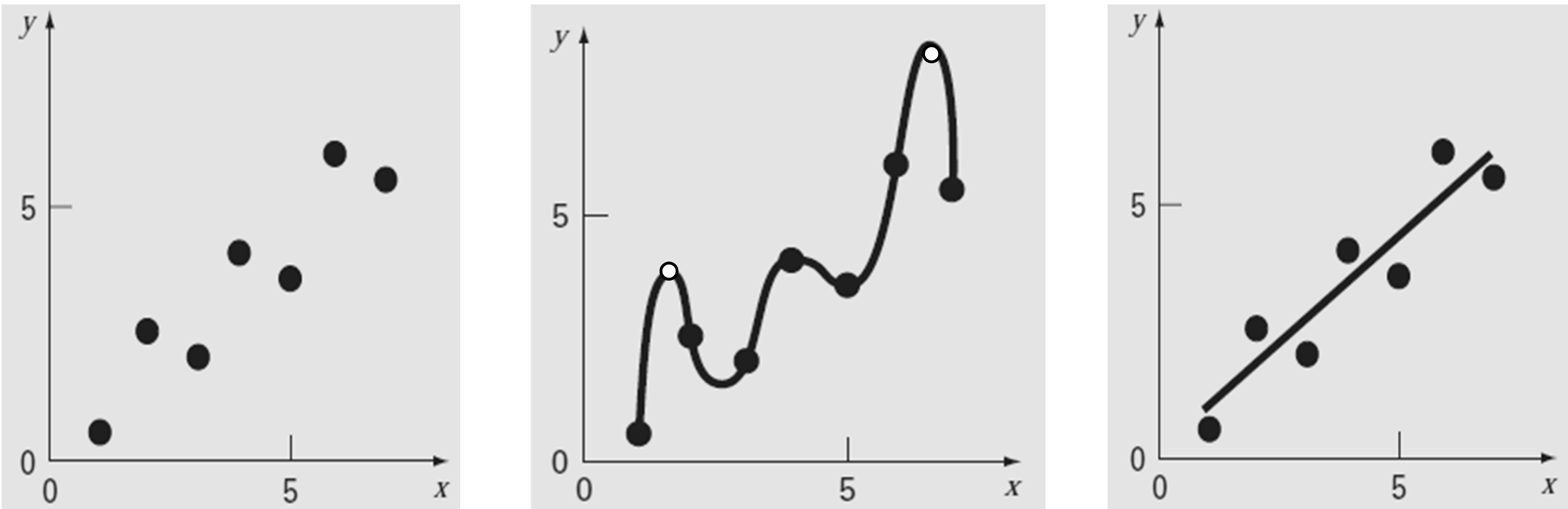
LEAST SQUARE REGRESSION

Least Square Regression

- $\boxed{\text{Linear Regression}}$
- Multiple Regression
- General linear least squares
- Nonlinear Regression

Least Square Regression

- Where substantial error is associated with data, polynomial interpolation is inappropriate and may yield unsatisfactory results when used to predict intermediate values.



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 mathematical expression for the straight line is:

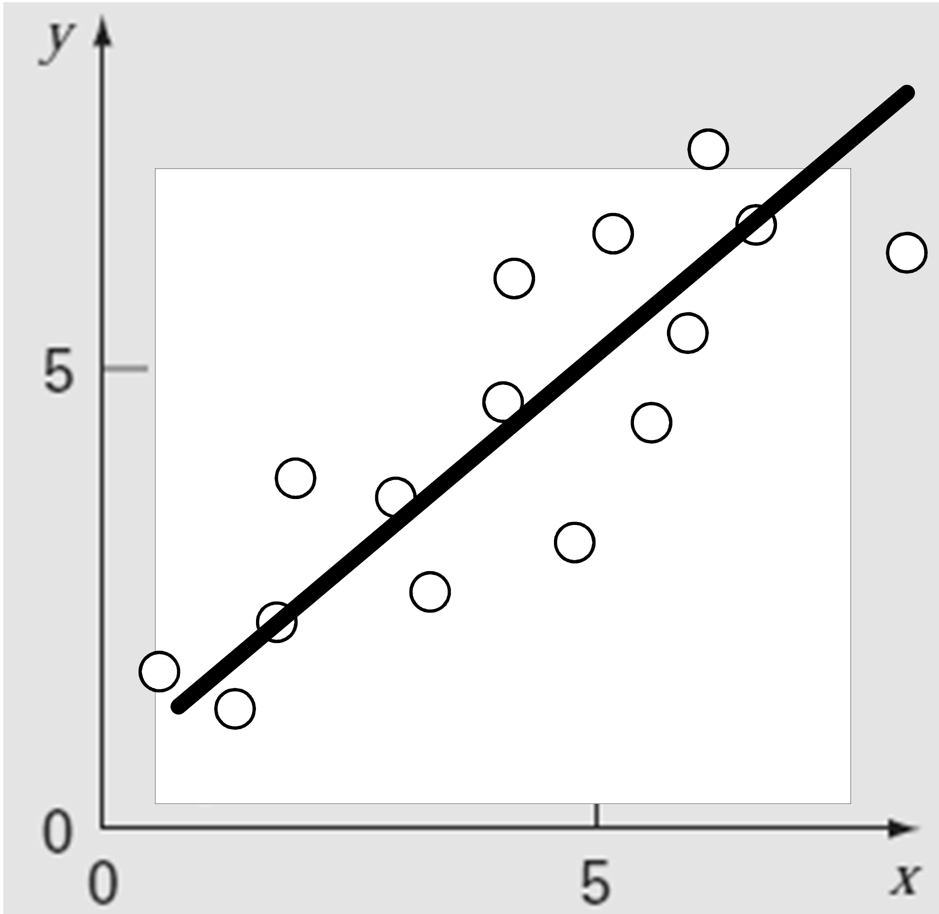
$$y = a_0 + a_1 x + e$$

a_0 and a_1 are coefficients representing the intercept and the slope

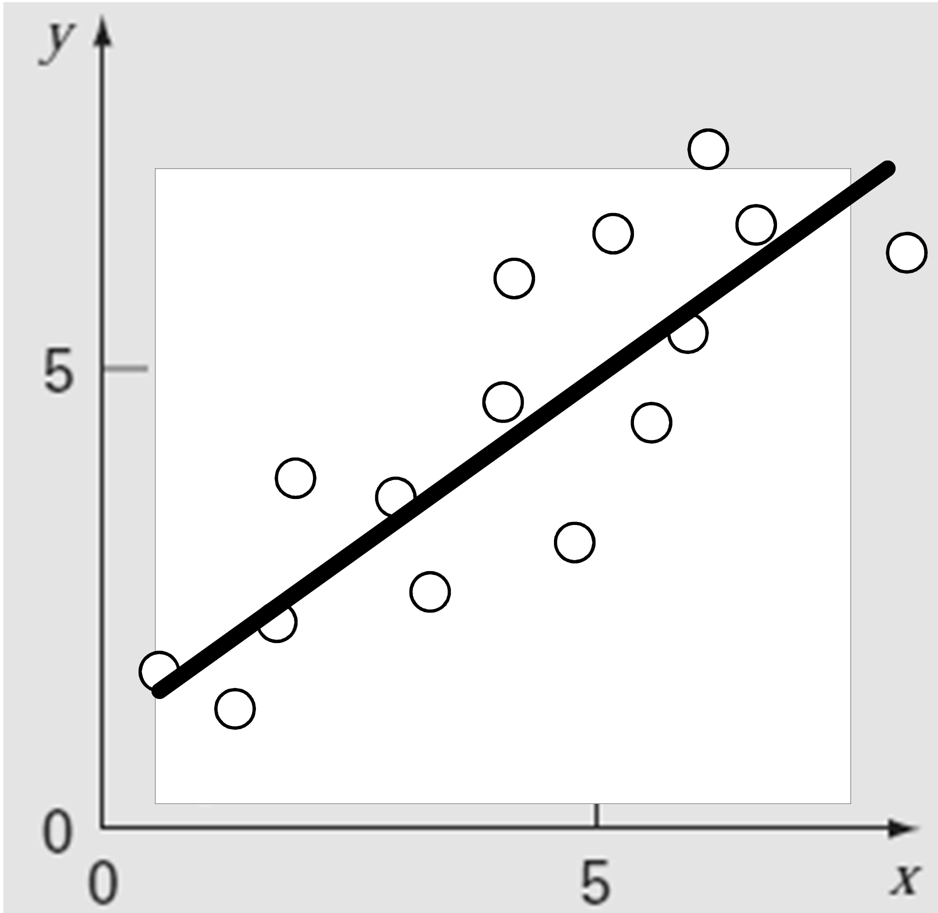
e is the error, or residual, between the model and the observations

$$e = y - a_0 - a_1 x$$

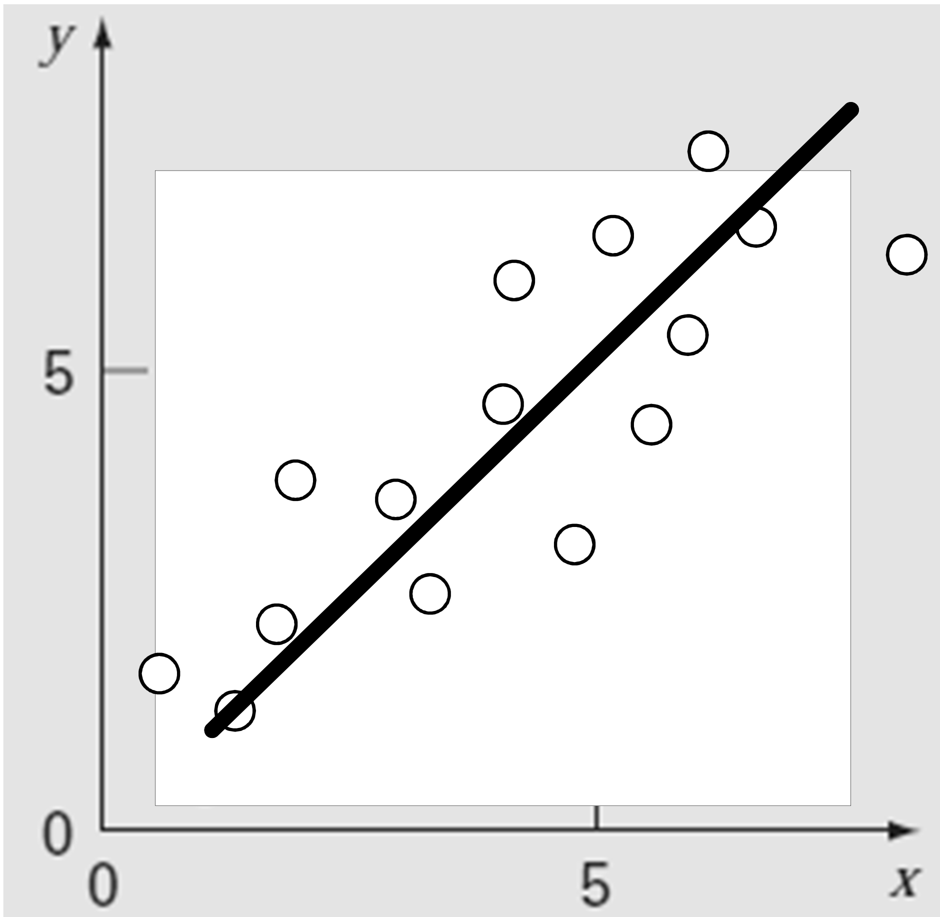
Linear Regression



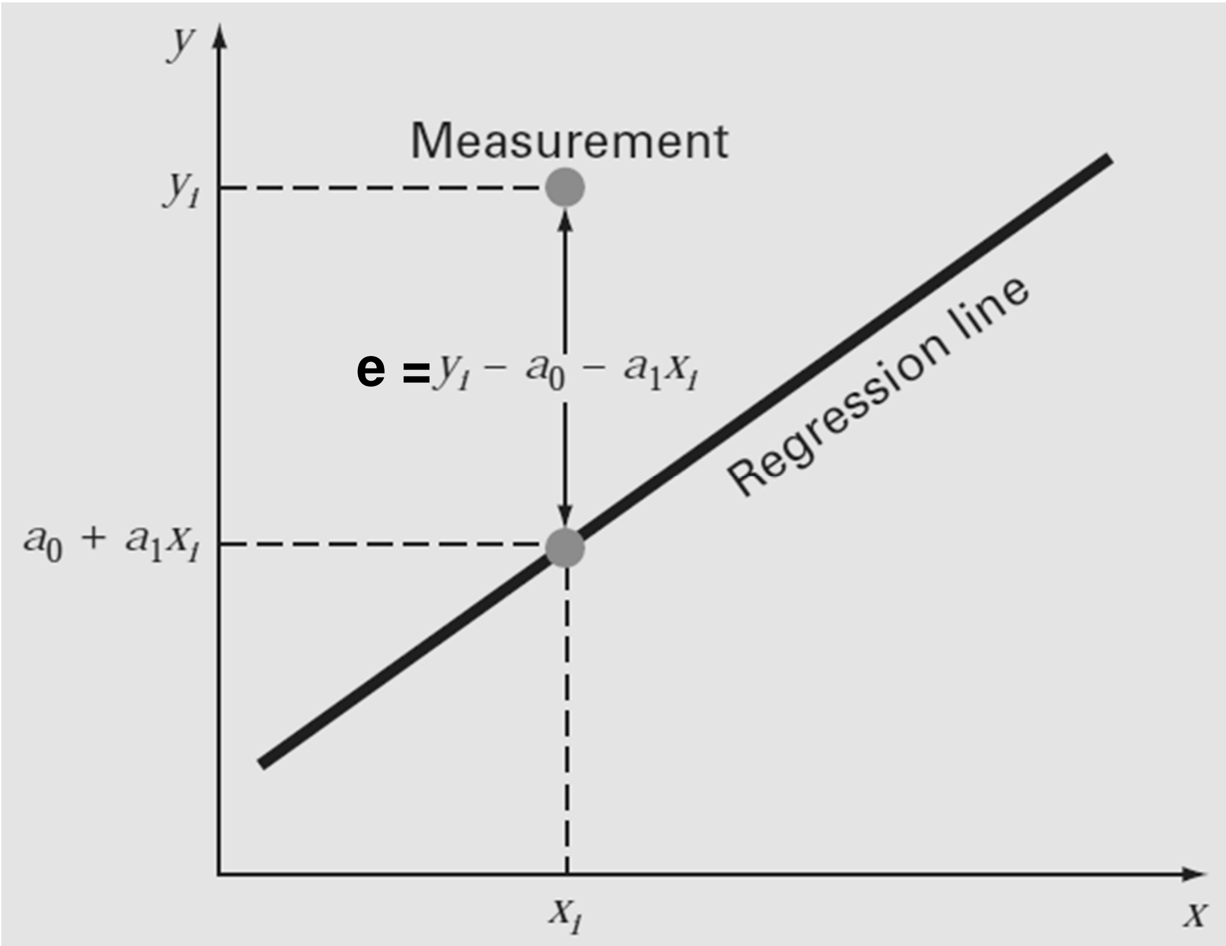
Linear Regression



Linear Regression



Linear Regression



We need a criteria to fit data !!

Minimum error for all points

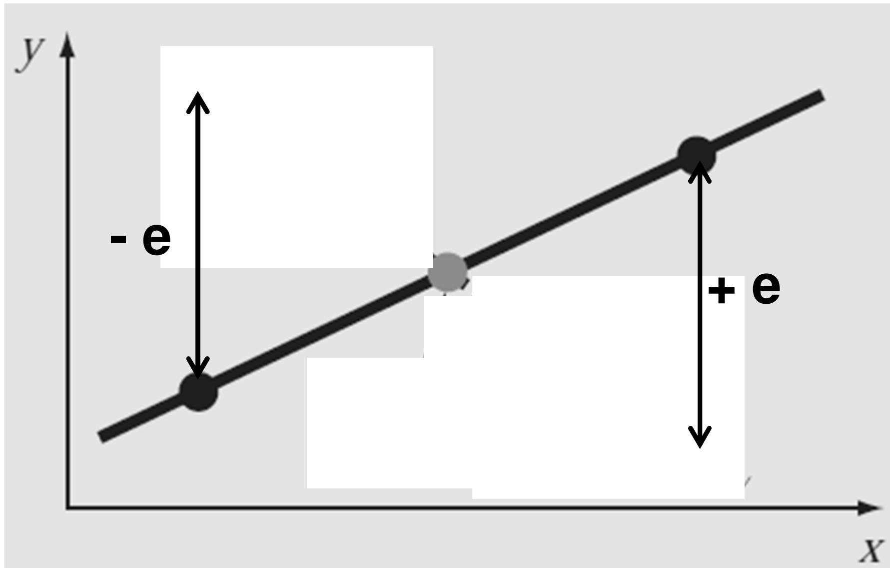
One best solution

Linear Regression

➨ One strategy for fitting a "best" line through the data would be to minimize the sum of the residual errors for all the available data:

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

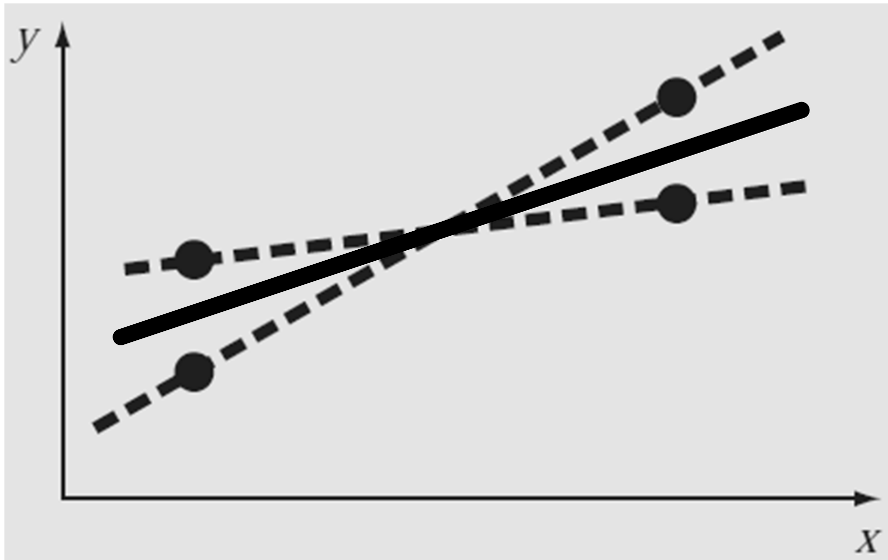
Σe = zero !!



Linear Regression

- One strategy for fitting a “best” line through the data would be to minimize the sum of the residual errors for all the available data:

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



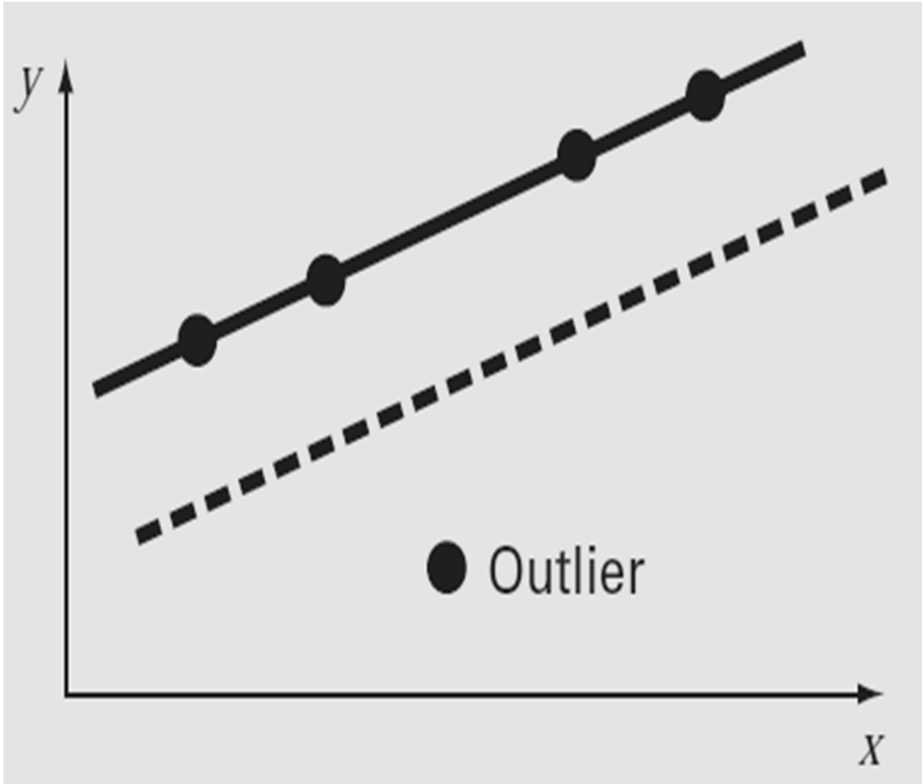
more than one solution !!

Linear Regression

➨ One strategy for fitting a "best" line through the data would be to minimize the sum of the residual errors for all the available data:

The mini-max criterion:

The line is chosen that minimizes the maximum distance that an individual point falls from the line.



Linear Regression

How to calculate a_o and a_1?

- One strategy for fitting a "best" line through the data would be to minimize the sum of the residual errors for all the available data:

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_{i,\text{measured}} - y_{i,\text{model}})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

- This criterion has a number of advantages, overcomes the shortcomings of the aforementioned approaches and the fact that it yields a unique line for a given set of data.

Linear Regression

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

- Differentiate S_r with respect to each coefficient:

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

$$\frac{\partial S_r}{\partial a_1} = -2 \sum [(y_i - a_0 - a_1 x_i) x_i]$$

- If these derivatives = zero ➔ minimum S_r

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

Linear Regression

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

$$n a_0 + \left(\sum x_i\right) a_1 = \sum y_i \qquad \left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 = \sum x_i y_i$$

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

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

Linear Regression

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

x_i	y_i
1	0.5
2	2.5
3	2.0
4	4.0
5	3.5
6	6.0
7	5.5

Linear Regression

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

$$n = 7$$

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

$$\sum x_i = 28$$

$$\sum y_i = 24$$

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

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

$$a_1 = \frac{7(119.5) - 28(24)}{7(140) - (28)^2} = 0.8392857$$

$$a_0 = 3.428571 - 0.8392857(4) = 0.07142857$$

$$\sum x_i y_i = 119.5$$

$$\sum x_i^2 = 140$$

Linear Regression

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

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

$$\sum x_i = 28$$

$$\sum y_i = 24$$

$$\sum x_i y_i = 119.5$$

$$\sum x_i^2 = 140$$

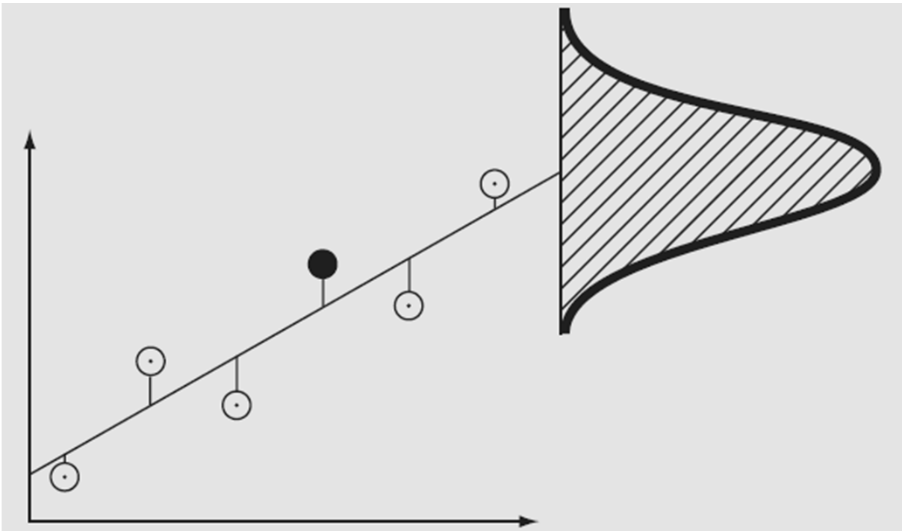
x_i	y_i	x_i^2	$x_i y_i$
1	0.5	1	0.5
2	2.5	4	5
3	2	9	6
4	4	16	16
5	3.5	25	17.5
6	6	36	36
7	5.5	49	38.5
28	24	140	119.5

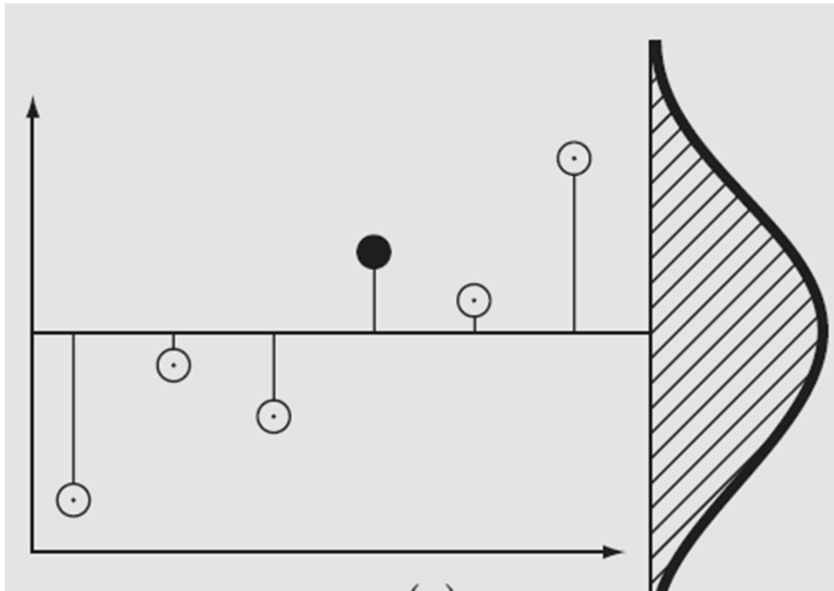
Linear Regression

Quantification of data spread around regression line:

➨ $S_{y/x}$: The standard error of estimates ➔ $$s_{y/x} = \sqrt{\frac{S_r}{n-2}}$$

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





(a)

Linear Regression

Quantification of Error reduction:

- r^2: Coefficient of determination ➔ $$r^2 = \frac{S_t - S_r}{S_t}$$

$$S_t = \Sigma(y_i - \bar{y})^2$$

- r: correlation coefficient $(= \sqrt{r^2})$

$$r = \frac{n\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)}{\sqrt{n\Sigma x_i^2 - (\Sigma x_i)^2}\sqrt{n\Sigma y_i^2 - (\Sigma y_i)^2}}$$

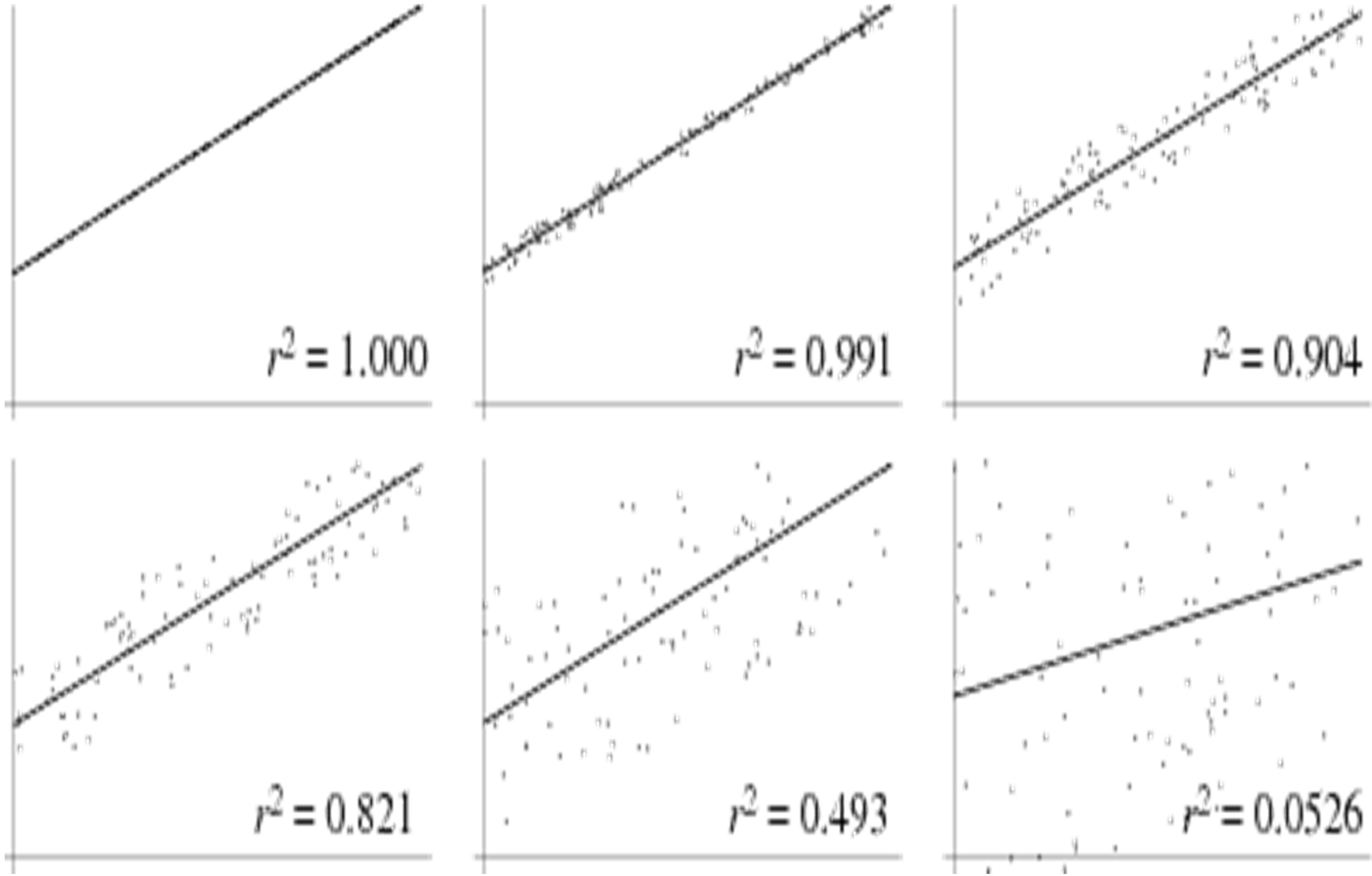
Linear Regression

x_i	y_i	$(y_i - \bar{y})^2$	$(y_i - a_0 - a_1 x_i)^2$
1	0.5	8.5765	0.1687
2	2.5	0.8622	0.5625
3	2.0	2.0408	0.3473
4	4.0	0.3265	0.3265
5	3.5	0.0051	0.5896
6	6.0	6.6122	0.7972
7	5.5	4.2908	0.1993
Σ	24.0	22.7143	2.9911

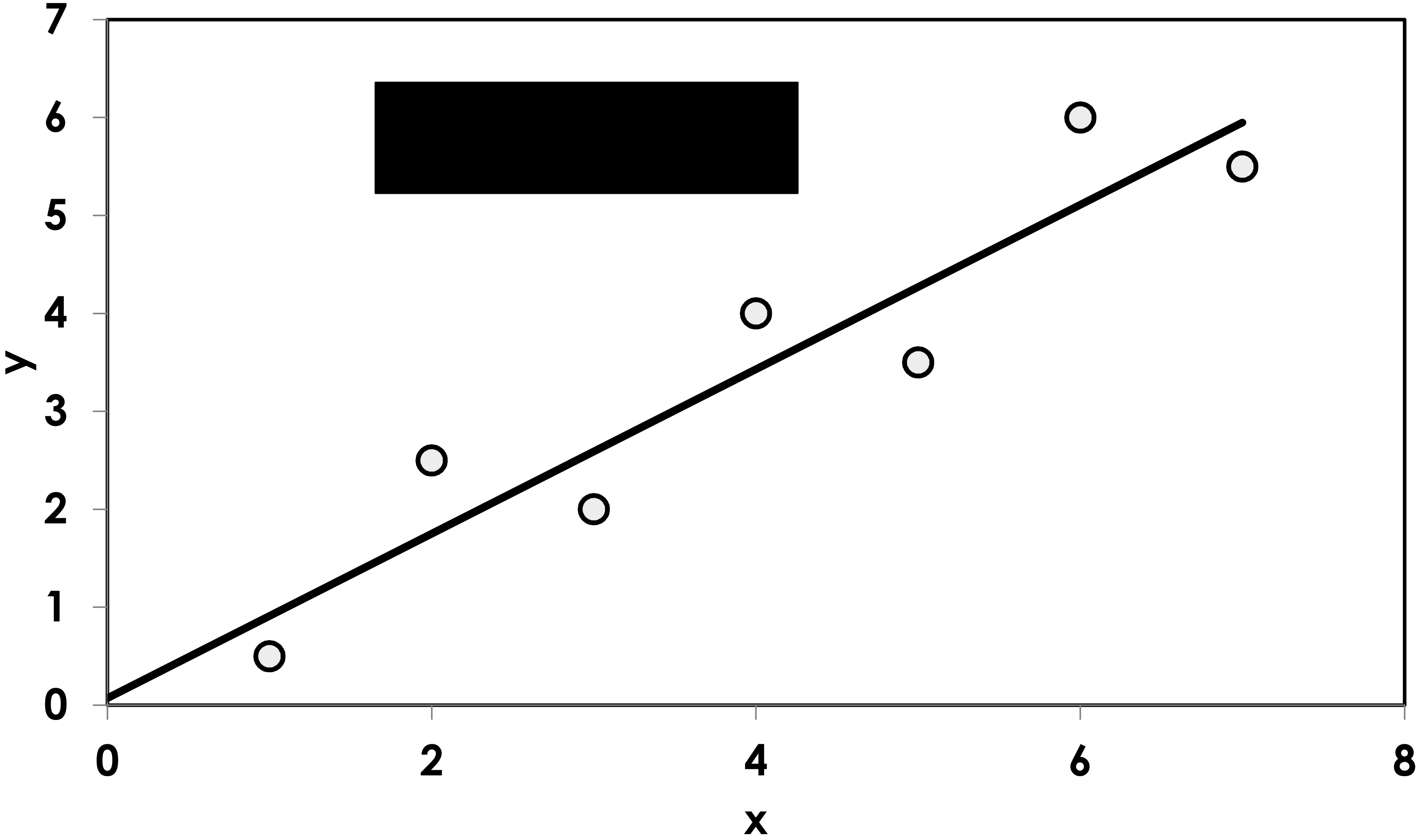
$$s_{y/x} = \sqrt{\frac{2.9911}{7-2}} = 0.7735$$

$$r^2 = \frac{22.7143 - 2.9911}{22.7143} = 0.868 \qquad r = \sqrt{0.868} = 0.932$$

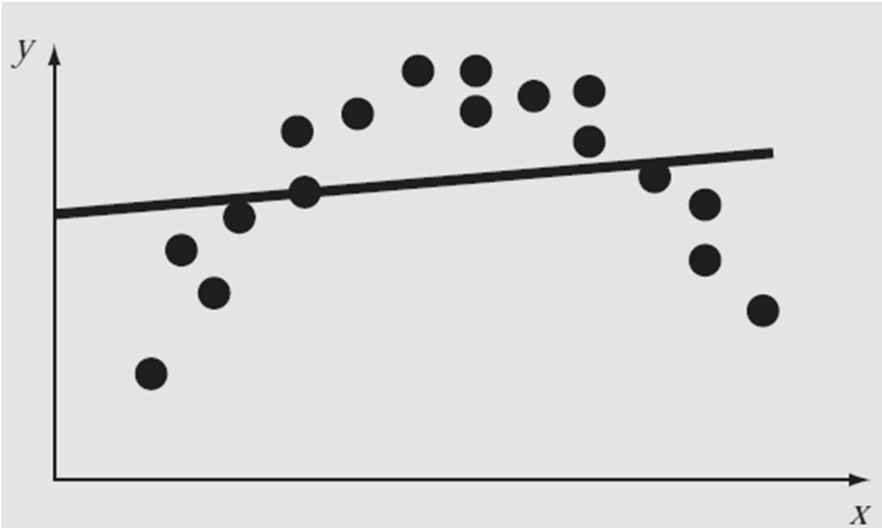
Linear Regression

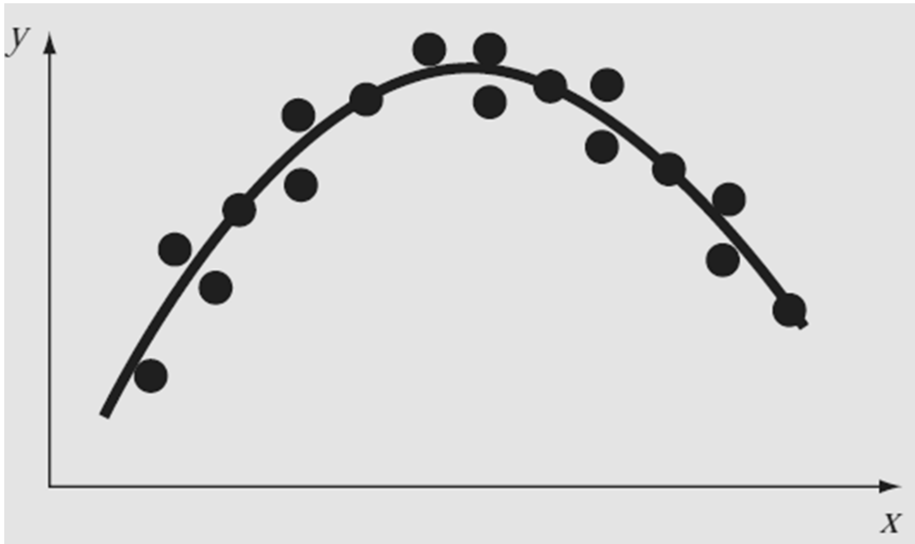


Linear Regression



Linear Regression





Always inspect the plot of points along with your regression curve !!

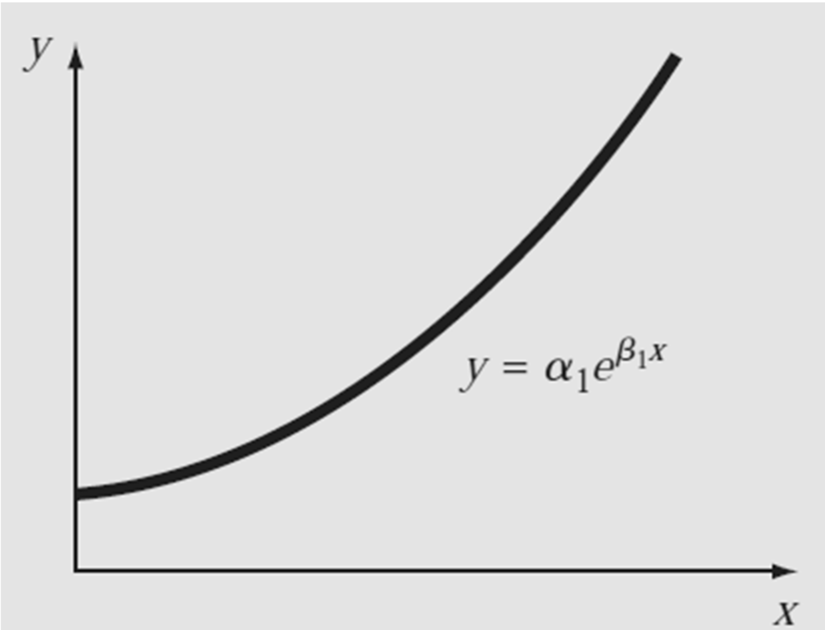
Linearization of Nonlinear Relationships

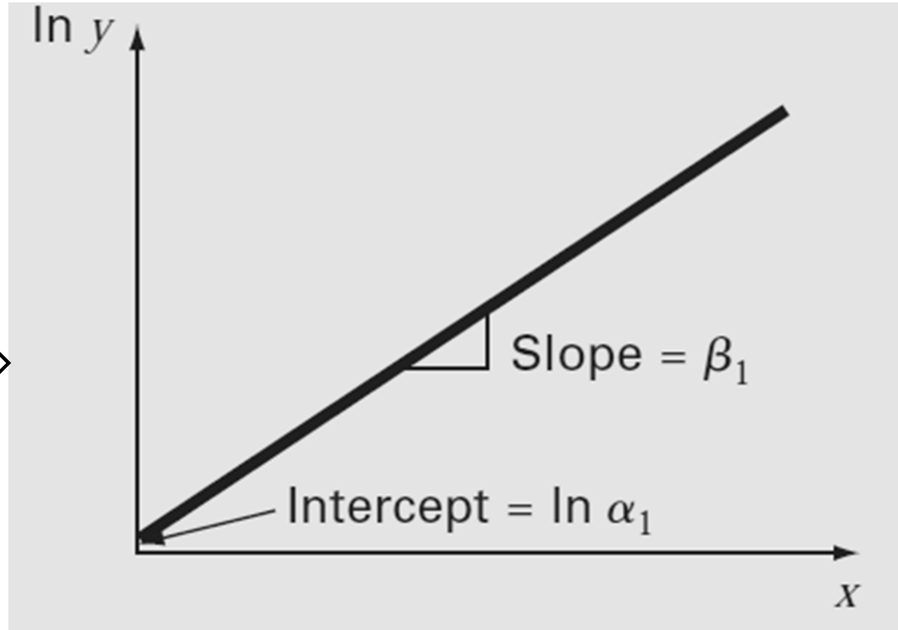
- Linear regression provides a powerful technique for fitting a best line to data.
- This is not always the case, and the first step in any regression analysis should be to plot and visually inspect the data to ascertain whether a linear model applies.
- Some data are curvilinear. In some cases, techniques such as polynomial regression are appropriate.
- For others, transformations can be used to express the data in a form that is compatible with linear regression.

Linearization of Nonlinear Relationships

$$\ln y = \ln \alpha_1 + \beta_1 x \ln e$$

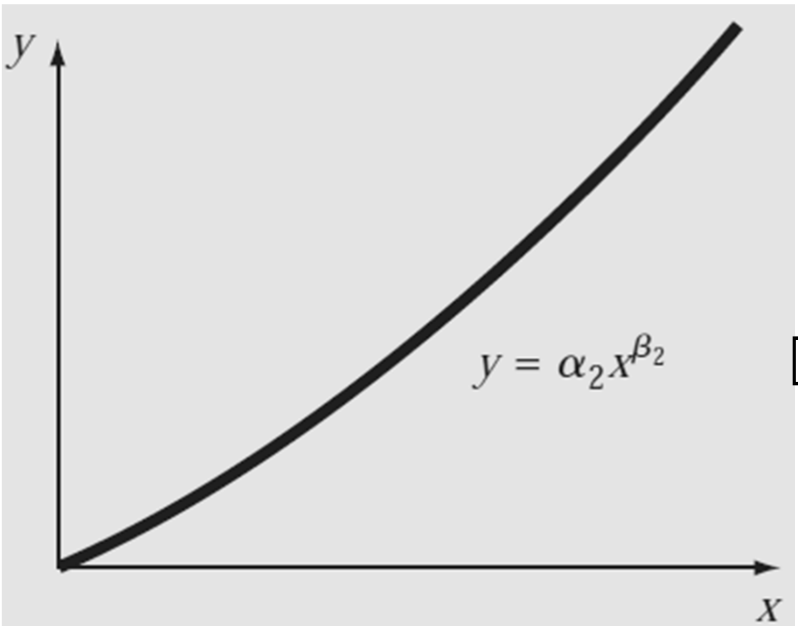
$$\ln y = \ln \alpha_1 + \beta_1 x$$



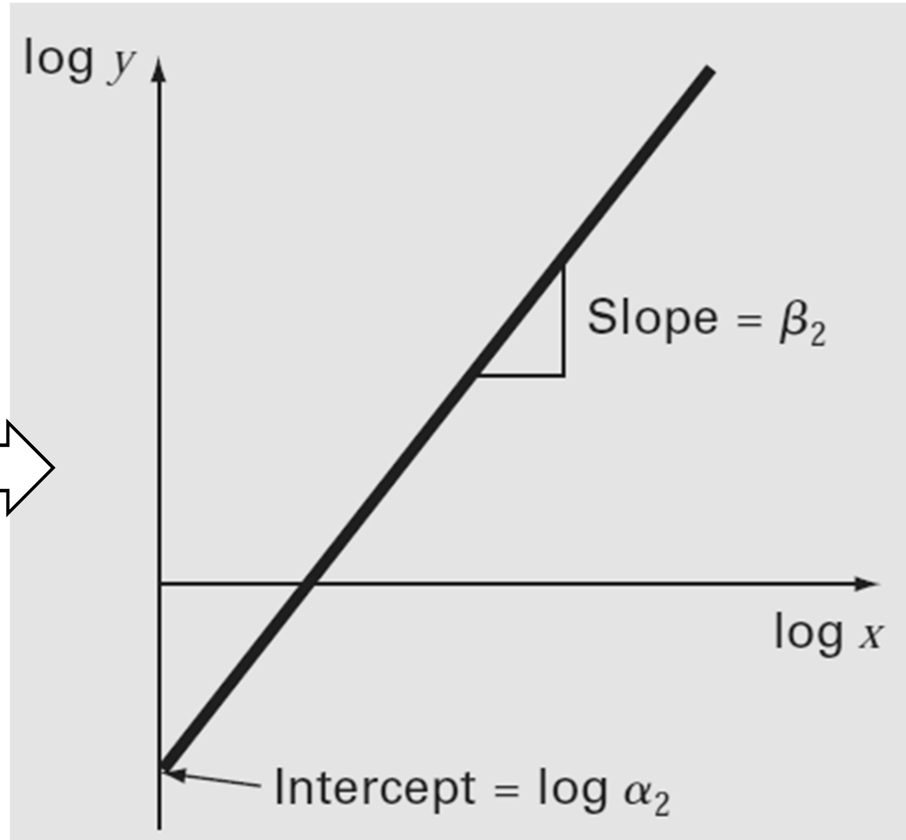


Exponential
equation

Linearization of Nonlinear Relationships

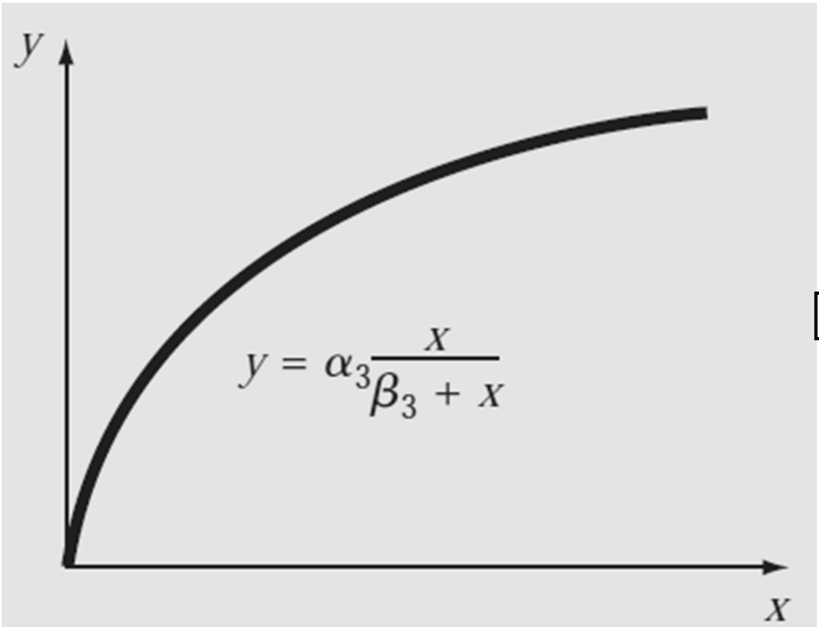


Power equation

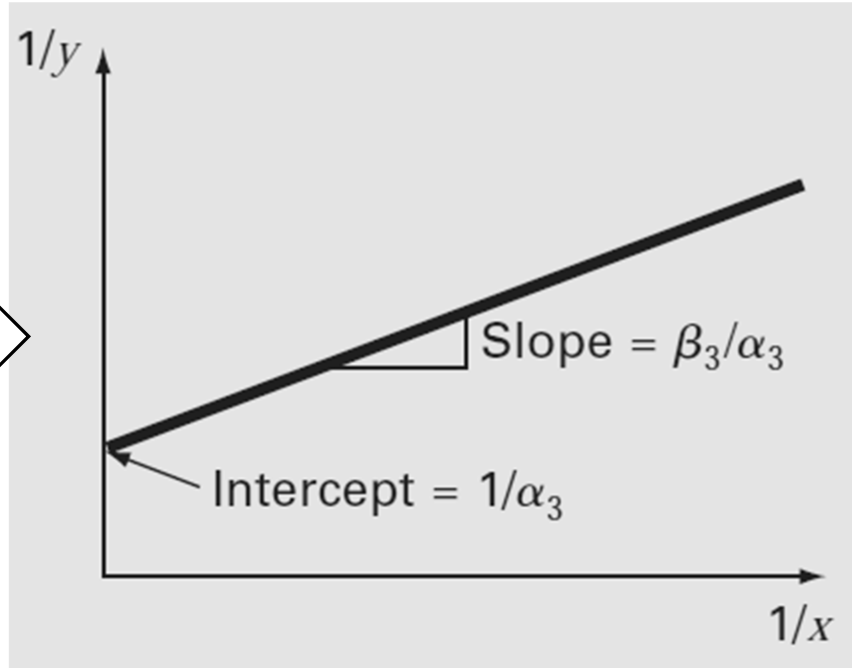


$$\log y = \beta_2 \log x + \log \alpha_2$$

Linearization of Nonlinear Relationships



Saturation growth rate equation



$$\frac{1}{y} = \frac{\beta_3}{\alpha_3}\frac{1}{x} + \frac{1}{\alpha_3}$$

Linear Regression

➨ Fit a **power equation** to the x and y values in the following Table:

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

Linear Regression

$$y = \alpha_2 x^{\beta_2} \Rightarrow \log y = \beta_2 \log x + \log \alpha_2$$

New x and y ➔ use linear regression

$a_1 = \beta_2$

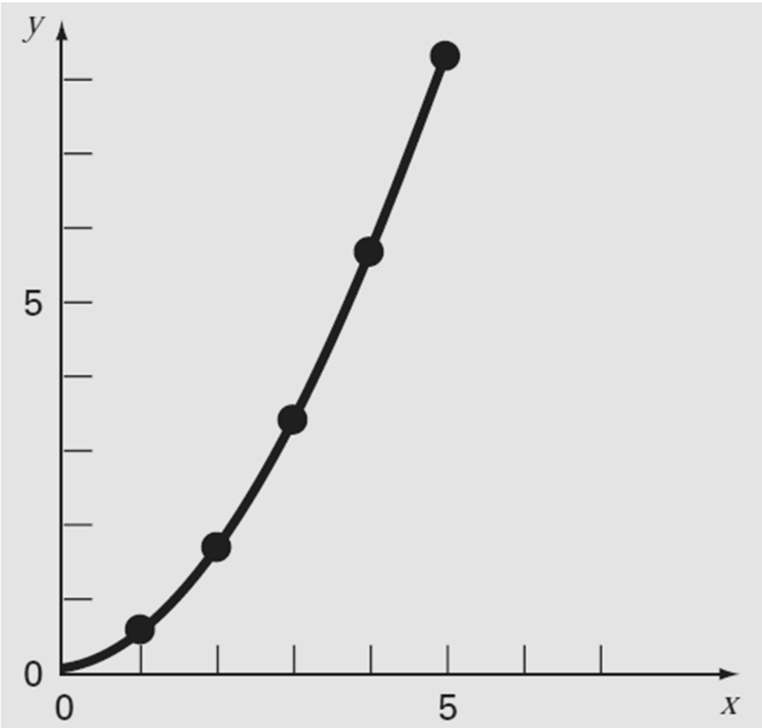
$a_0 = \log \alpha_2$

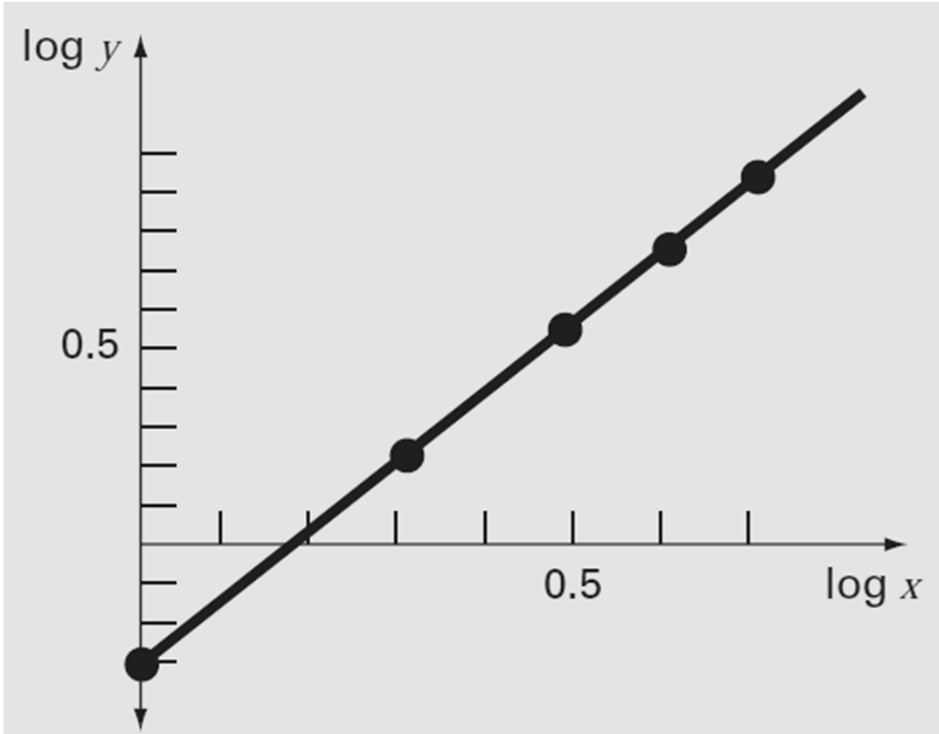
x	y	log x	log y
1	0.5	0	−0.301
2	1.7	0.301	0.226
3	3.4	0.477	0.534
4	5.7	0.602	0.753
5	8.4	0.699	0.922

Linear Regression

$$y = 0.5x^{1.75}$$

$$\log y = 1.75 \log x - 0.300$$





INTERPOLATION

Interpolation

- $\boxed{\text{➨ Newton Polynomial}}$
- ➨ Lagrange Polynomial
- ➨ Polynomial Coefficient
- ➨ Inverse Interpolation
- ➨ Splines
- ➨ Multidimensional Interpolation

Interpolation

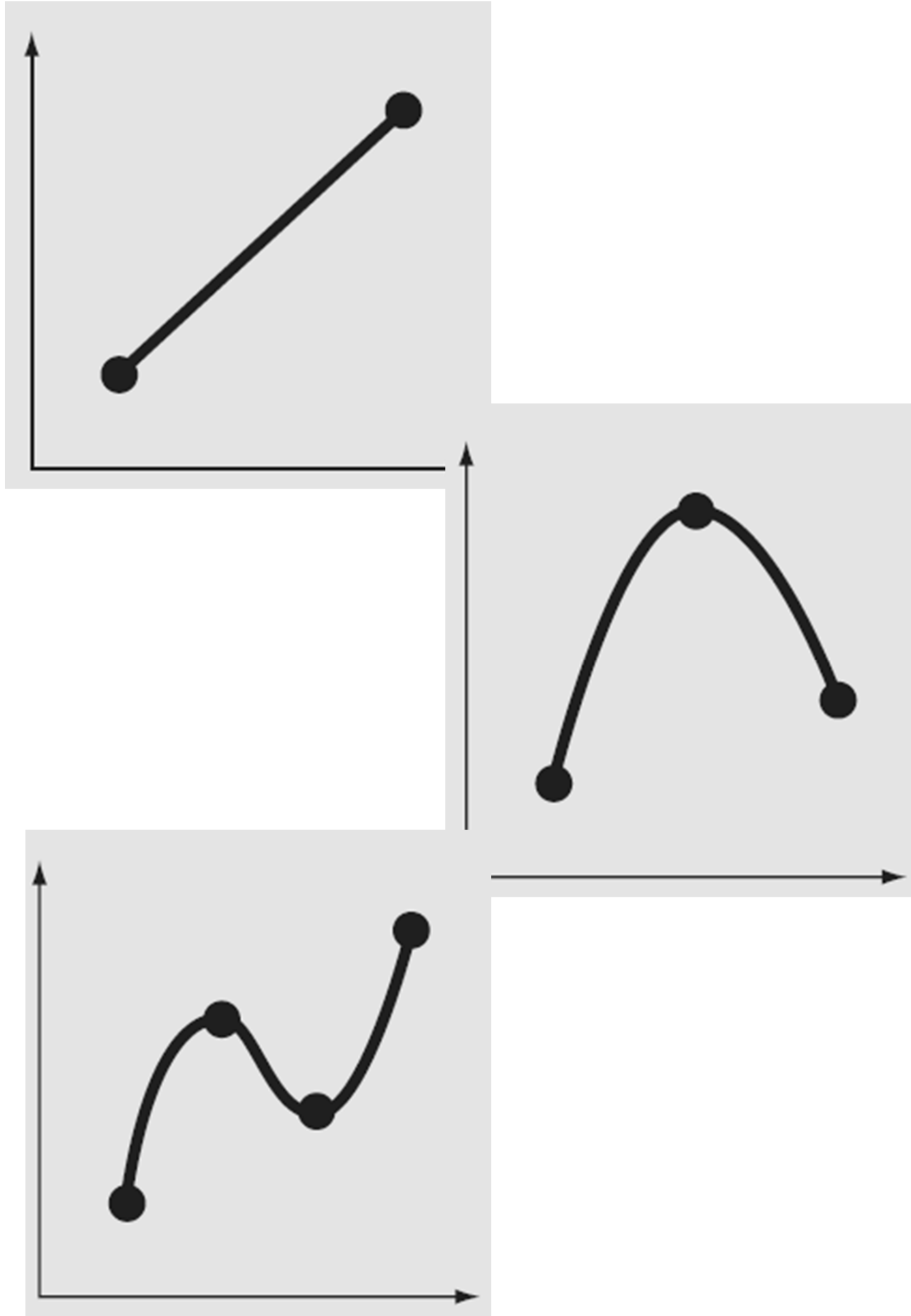
- You will frequently have occasion to estimate intermediate values between precise data points.
- The most common method used for this purpose is polynomial interpolation with the the general formula for an nth-order polynomial is for $n+1$ data points.

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

- There is one and only one polynomial of order n that passes through all the points.

Interpolation

- There is only one straight line (first-order polynomial) that connects two points.
- Only one parabola (second order polynomial) connects a set of three points.
- Only one cubic (third order polynomial) connects a set of four points.

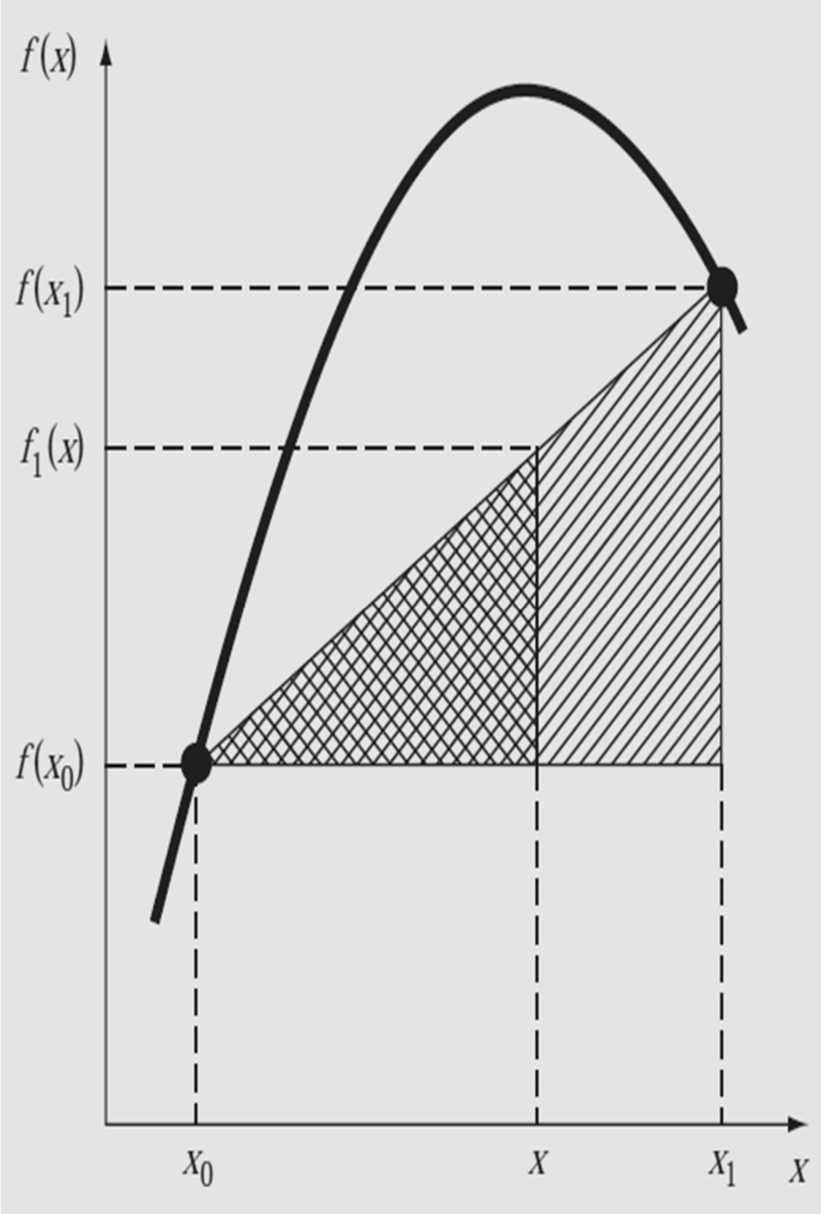


Newton Polynomial

- The simplest form of interpolation is to connect two data points with a straight line. This technique, called ***linear interpolation,*** is depicted graphically, using similar triangles:

$$\frac{f_1(x) - f(x_0)}{x - x_0} = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

$$f_1(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0)$$



Newton Polynomial

- Estimate the natural logarithm of 2 using linear interpolation.
- If ln 1 = 0 and ln 6 = 1.791759.
- If ln 1 = 0 and ln 4 =1.386294.

(ln 2 = 0.6931472)

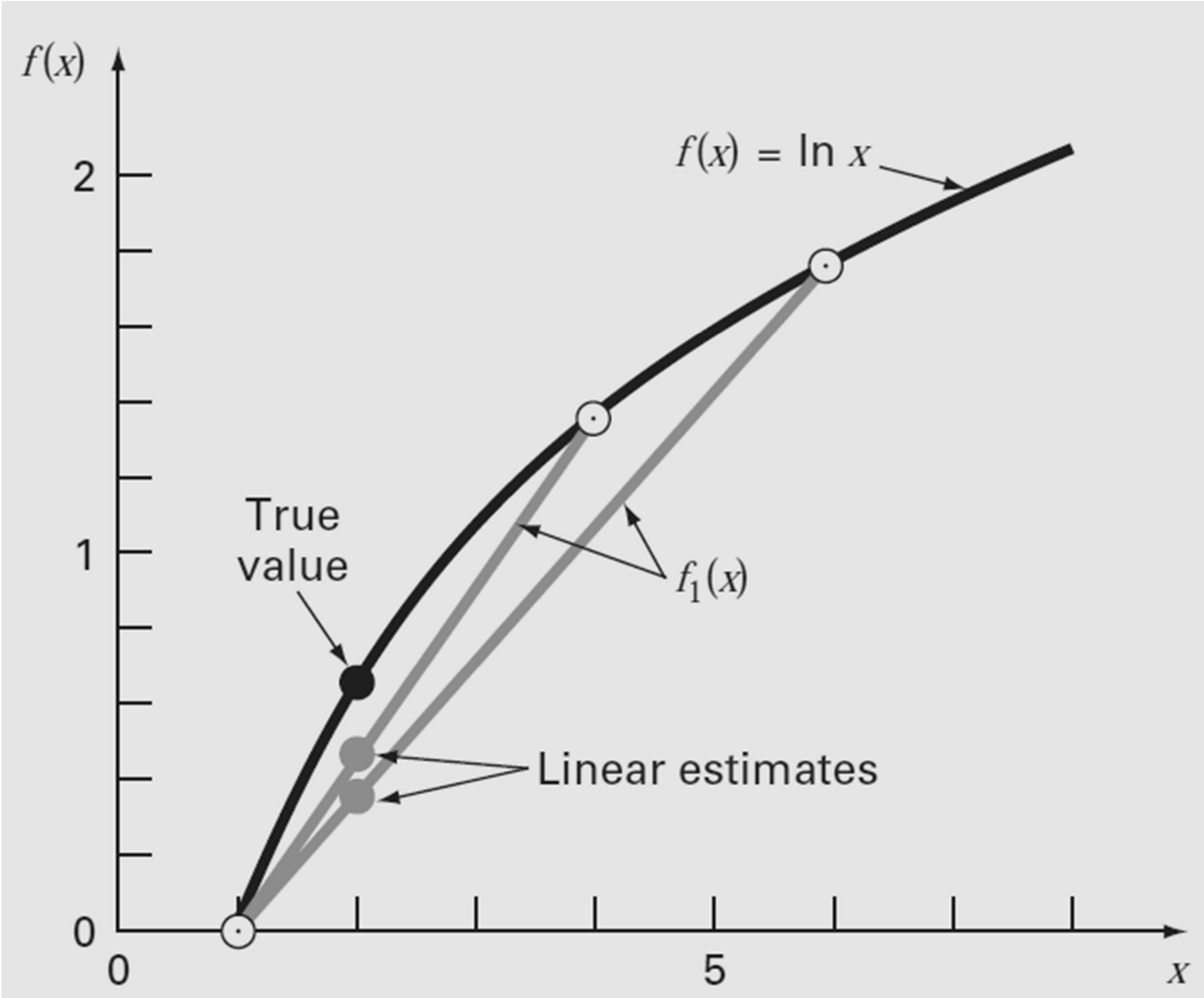
Newton Polynomial

$$f_1(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0)$$

$$f_1(2) = 0 + \frac{1.791759 - 0}{6 - 1}(2 - 1) = 0.3583519 \qquad \varepsilon_t = 48.3\%$$

$$f_1(2) = 0 + \frac{1.386294 - 0}{4 - 1}(2 - 1) = 0.4620981 \qquad \varepsilon_t = 33.3\%$$

Newton Polynomial



Newton Polynomial

- The error in previous Example resulted from our approximating a curve with a straight line.
- A strategy for improving the estimate is to introduce some curvature into the line connecting the points.
- If three data points are available, this can be accomplished with **a second-order polynomial** (also called a quadratic polynomial or a ***parabola***):

$$f_2(x) = a_0 + a_1 x + a_2 x^2$$

Newton Polynomial

$$f_2(x) = a_0 + a_1 x + a_2 x^2$$

$$\Downarrow$$

$$f_2(x) = b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1)$$

$$\boxed{\begin{aligned} a_0 &= b_0 - b_1 x_0 + b_2 x_0 x_1 \\ a_1 &= b_1 - b_2 x_0 - b_2 x_1 \\ a_2 &= b_2 \end{aligned}}$$

Newton Polynomial

$$b_0 = f(x_0)$$

$$b_1 = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

$$b_2 = \frac{\dfrac{f(x_2) - f(x_1)}{x_2 - x_1} - \dfrac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0}$$

Newton Polynomial

- Estimate the natural logarithm of 2 using **quadratic interpolation**.
- If ln 1 = 0, ln 4 = 1.386294, and ln 6 = 1.791759.

(ln 2 = 0.6931472)

Newton Polynomial

$$x_0 = 1 \qquad f(x_0) = 0$$

$$x_1 = 4 \qquad f(x_1) = 1.386294$$

$$x_2 = 6 \qquad f(x_2) = 1.791759$$

$$b_0 = 0$$

$$b_1 = \frac{1.386294 - 0}{4 - 1} = 0.4620981$$

$$b_2 = \frac{\dfrac{1.791759 - 1.386294}{6 - 4} - 0.4620981}{6 - 1} = -0.0518731$$

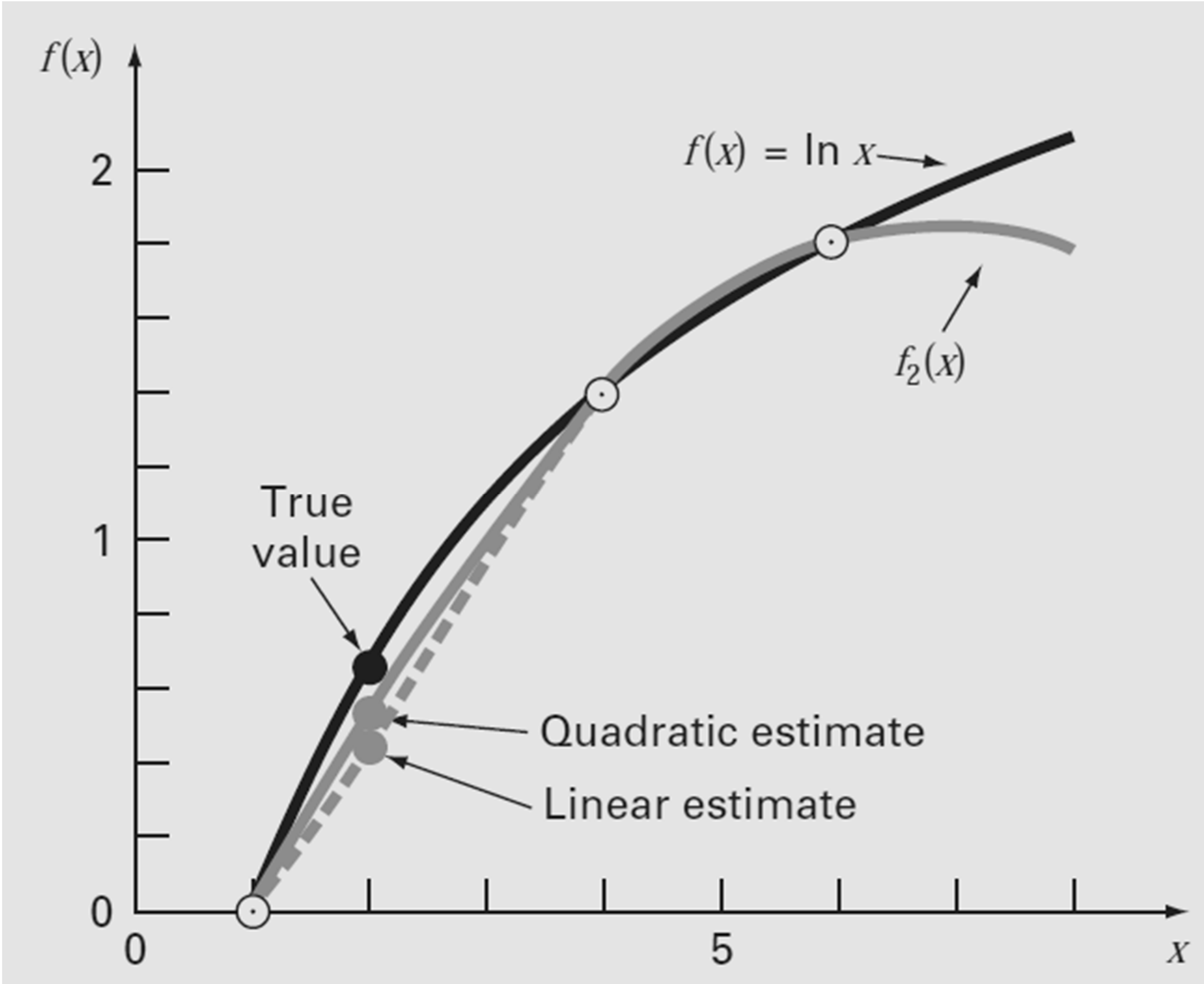
Newton Polynomial

$$f_2(x) = 0 + 0.4620981(x - 1) - 0.0518731(x - 1)(x - 4)$$

$$f_2(2) = 0.5658444$$

$$\varepsilon_t = 18.4\%$$

Newton Polynomial



Grading System