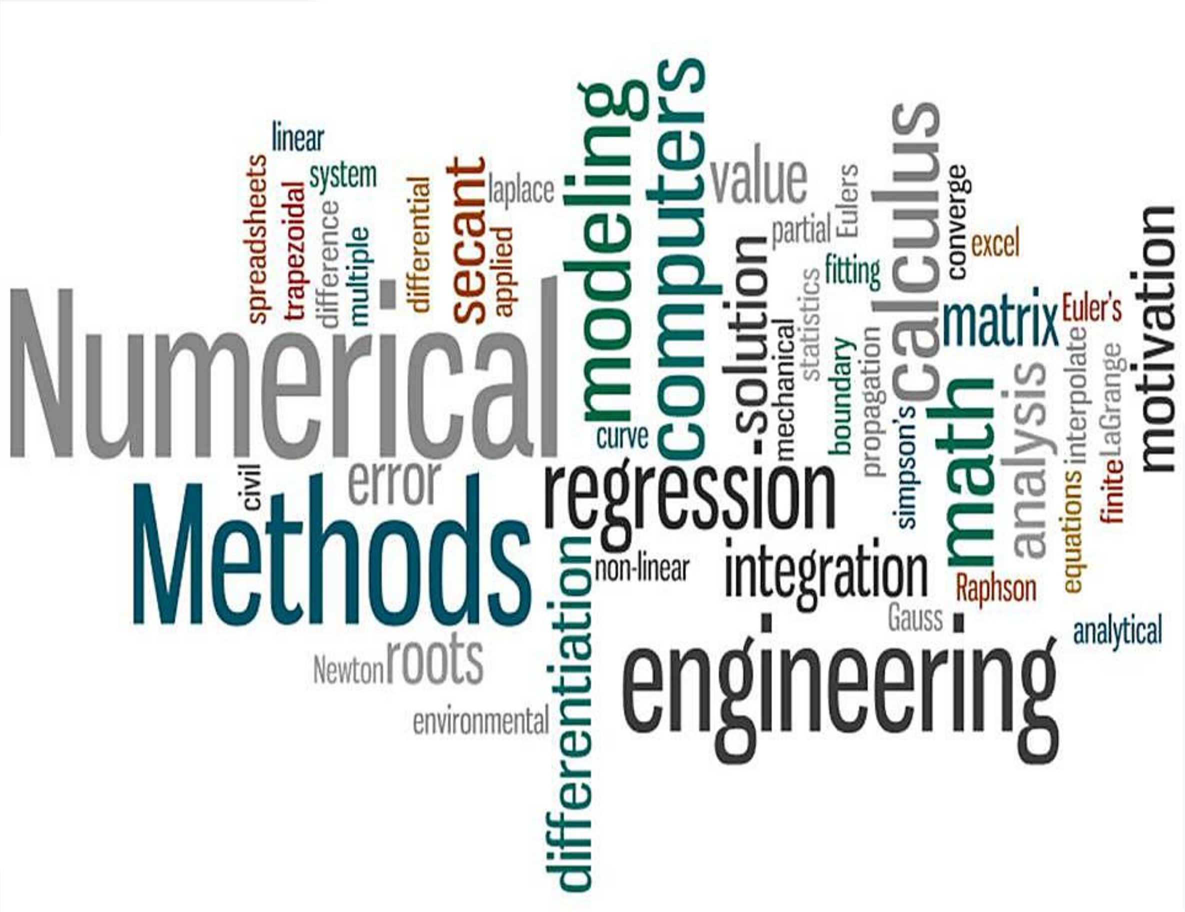
Programming and Numerical Analysis



Curve Fitting

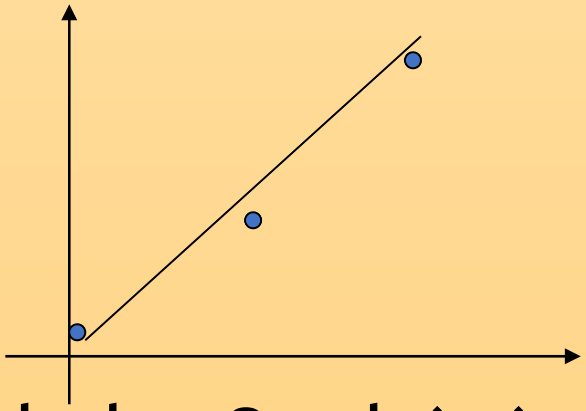
Curve Fitting

- **Curve Fitting:** to fit curves to data points. Two types: **regression** and **interpolation**.

Regression:

- Given a set of Data:

x	0	1	2
y	0.5	10.3	21.3



- Regression: Select a curve that best fits the data. One choice is to find the curve so that the sum of the square of the error is minimized. Experimental results are often of the first type.

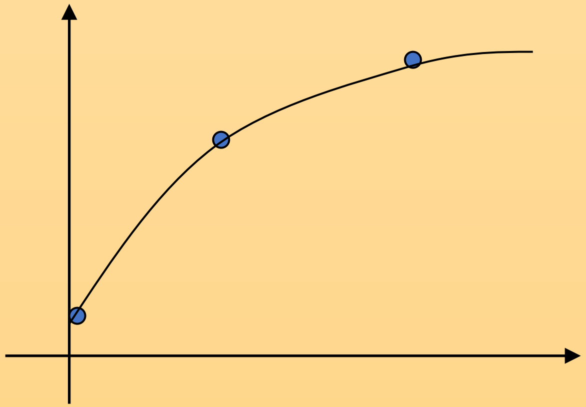
Curve Fitting

- **Curve Fitting:** to fit curves to data points. Two types: **regression** and **interpolation**.

Interpolation:

- **Given a set of Data:**

x_i	0	1	2
y_i	0.5	10.3	15.3



- **Interpolation:** Find a polynomial *P(x)* whose graph passes through all tabulated points.

$$y_i = P(x_i) \quad if \quad x_i \text{ is in the table}$$

Curve Fitting

There are two general approaches for curve fitting:

- **Least Squares regression:**

  Data exhibit a significant degree of scatter. The strategy is to derive a single curve that represents the general trend of the data.

- **Interpolation:**

  Data is very precise. The strategy is to pass a curve or a series of curves through each of the points.

Mathematical Background

Arithmetic mean: The sum of the individual data points (yi) divided by the number of points (n).

$$\bar{y} = \frac{\sum y_i}{n}, i = 1, \ldots, n$$

Standard deviation: The most common measure of a spread for a sample.

$$S_y = \sqrt{\frac{S_t}{n-1}}, \quad S_t = \sum (y_i - \bar{y})^2$$

Mathematical Background (cont'd)

Variance: Representation of spread by the square of the standard deviation.

$$S_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} \qquad S_y^2 = \frac{\sum y_i^2 - \left(\sum y_i\right)^2 / n}{n-1}$$

Coefficient of variation: Has the utility to quantify the spread of data.

$$c.v. = \frac{S_y}{\bar{y}} 100\%$$

Least Square Regression - Linear Regression
الانحدار الخطي البسيط

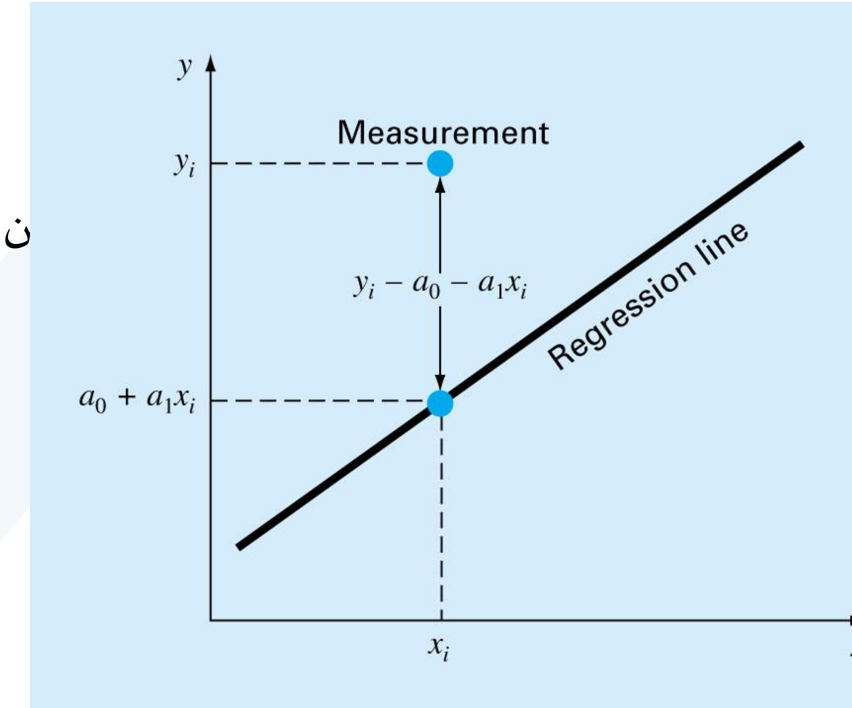
- Linear Regression: Fitting a straight line to a set of paired observations: $(x_1, y_1), (x_2, y_2),\ldots,(x_n, y_n)$.

$y = a_0 + a_1 x + e$

a_1 – slope (ثابت الانحدار أو الجزء المقطوع من المحور Y)

a_0 – intercept (ميل الخط المستقيم)

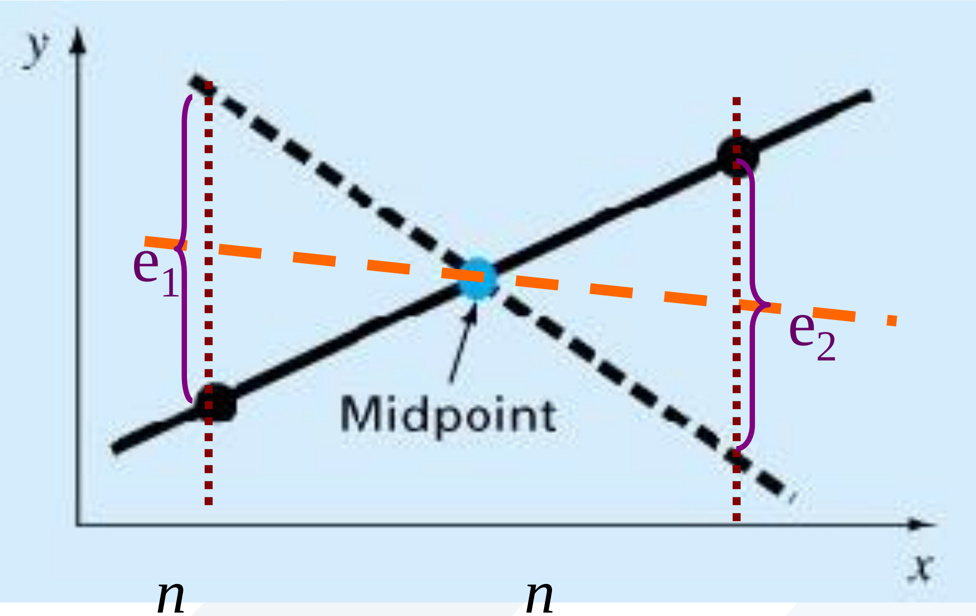
e - error, or residual, between the model and the observations (ن الموديل والقيمة المقيسة



How to find a_0 and a_1 so that the error would be minimum?

Linear Regression – Criteria for a “Best” fit

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



$e_1 = -e_2$

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

Linear Regression – Least Squares Fit

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

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

Yields a unique line for a given set of data.

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 Constants

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

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

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

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

$$\sum a_0 = n a_0$$

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

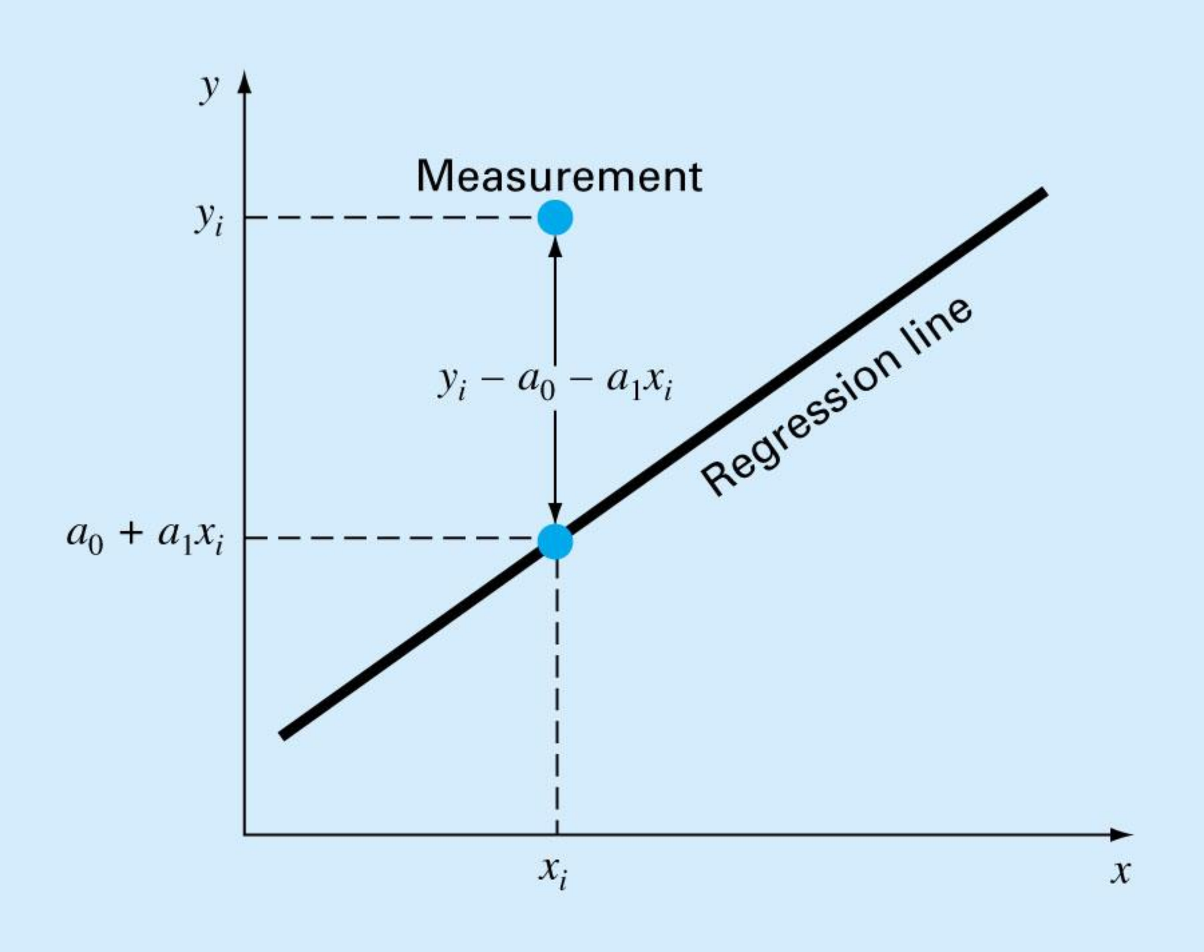
$$\sum y_i x_i = \sum a_0 x_i + \sum a_1 x_i^2$$

2 equations with 2 unknowns, can be solved simultaneously

Linear Regression – Determination of Constants

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

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



Error Quantification of Linear Regression

Total sum of the squares around the mean for the dependent variable, y, is $\mathbf{S_t}$

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

Sum of the squares of residuals around the regression line is **Sr**

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

Error Quantification of Linear Regression

S_t - S_r quantifies the improvement or error reduction due to describing data in terms of a straight line rather than as an average value.

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

r: correlation coefficient

معامل الارتباط عبارة عن مقياس رقمي يقيس قوة الارتباط بين متغيرين

***r*²**: coefficient of determination

r: correlation coefficient

For a perfect fit:

- $S_r = 0$ and $r = r^2 = 1$, signifying that the line explains 100 percent of the variability of the data.
- For $r = r^2 = 0$, $S_r = S_t$, the fit represents no improvement.

Example - Least Squares Fit of a Straight Line

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

- أوجد معادلة الانحدار الخطي البسيط (معادلة خط الانحدار) للبيانات في الجدول التالي، ومن ثم احسب قيمة معامل الارتباط.

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**

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

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

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

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

Example - Least Squares Fit of a Straight Line

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

ثابت الانحدار

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

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

الميل

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

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

Y = 0.07142857 + 0.8392857 x

Example - Least Squares Fit of a Straight Line

- Error Analysis:

x_i	y_i	$(y_i - \bar{y})^2$	$e_i^2 = (y_i - \hat{y})^2$
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
28	**24.0**	**22.7143**	**2.9911**

$$S_t = \sum (y_i - \bar{y})^2 = 22.7143$$

$$S_r = \sum e_i^2 = 2.9911$$

$$r^2 = \frac{S_t - S_r}{S_t} = 0.868$$

$$r = \sqrt{r^2} = \sqrt{0.868} = 0.932$$

Example - Least Squares Fit of a Straight Line

- Error Analysis:

The standard deviation (quantifies the spread around the mean):

$$s_y = \sqrt{\frac{S_t}{n-1}} = \sqrt{\frac{22.7143}{7-1}} = 1.9457$$

The standard error of estimate (quantifies the spread around the regression line)

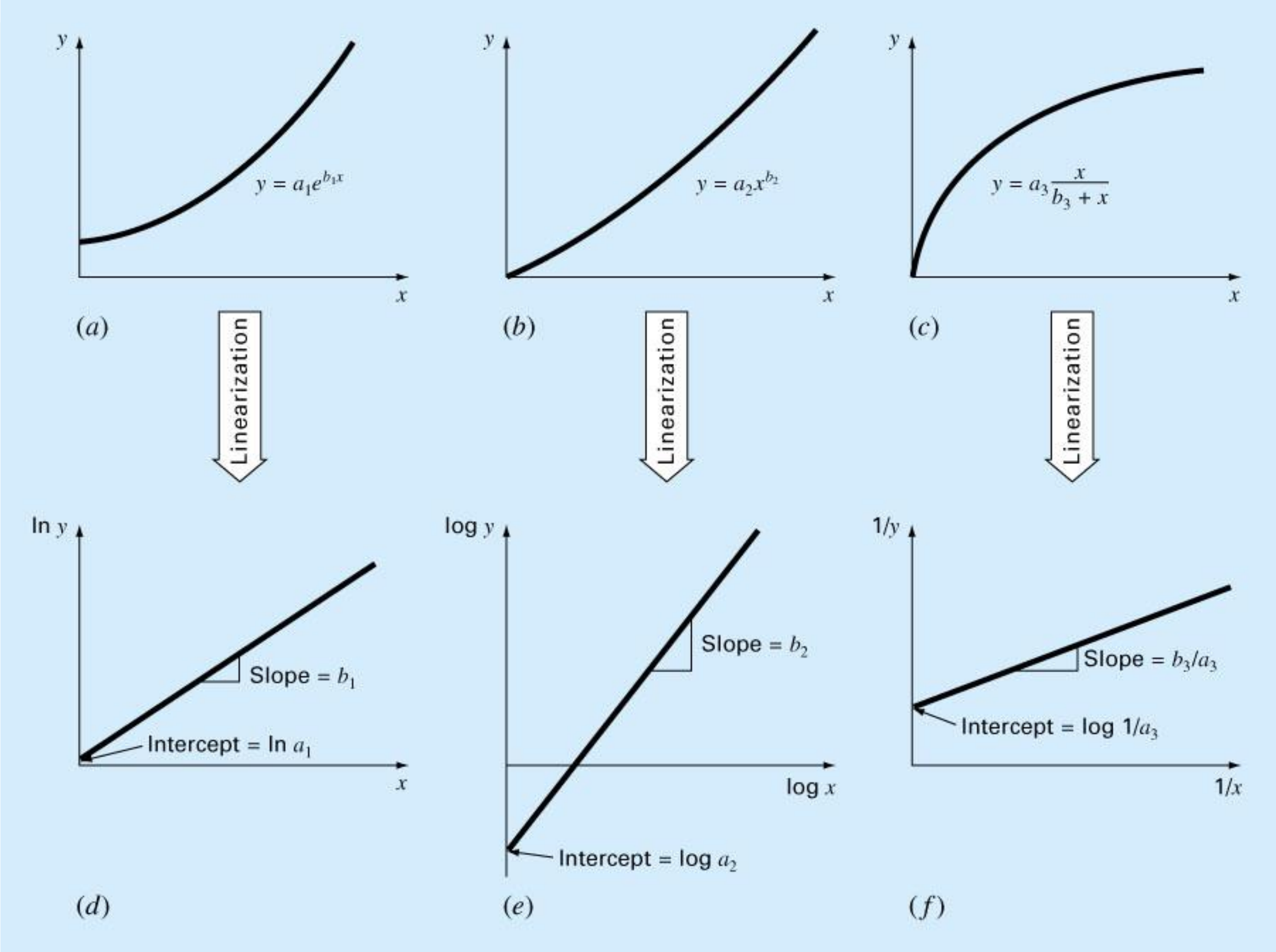
$$s_{y/x} = \sqrt{\frac{S_r}{n-2}} = \sqrt{\frac{2.9911}{7-2}} = 0.7735$$

Because $S_{y/x} < S_y$, the linear regression model has good fitness

Linearization of Nonlinear Relationships

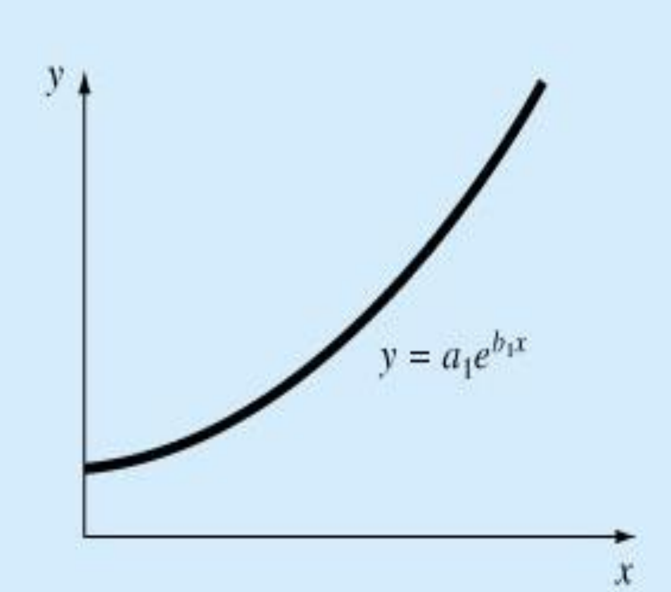
- The relationship between the dependent and independent variables is linear.
- However, a few types of nonlinear functions can be transformed into linear regression problems.
  - The exponential equation.
  - The power equation.
  - The saturation-growth-rate equation.

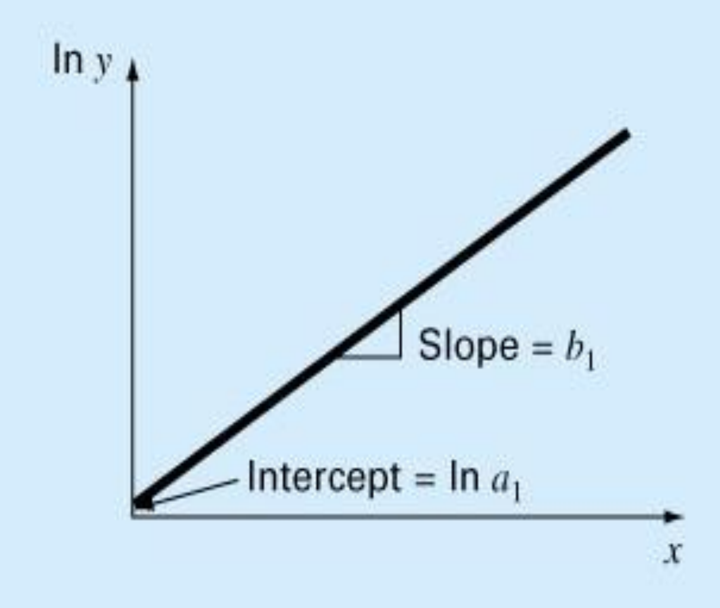
Linearization of Nonlinear Relationships



Linearization of Nonlinear Relationships

1. The Exponential Equation معادلة أسية



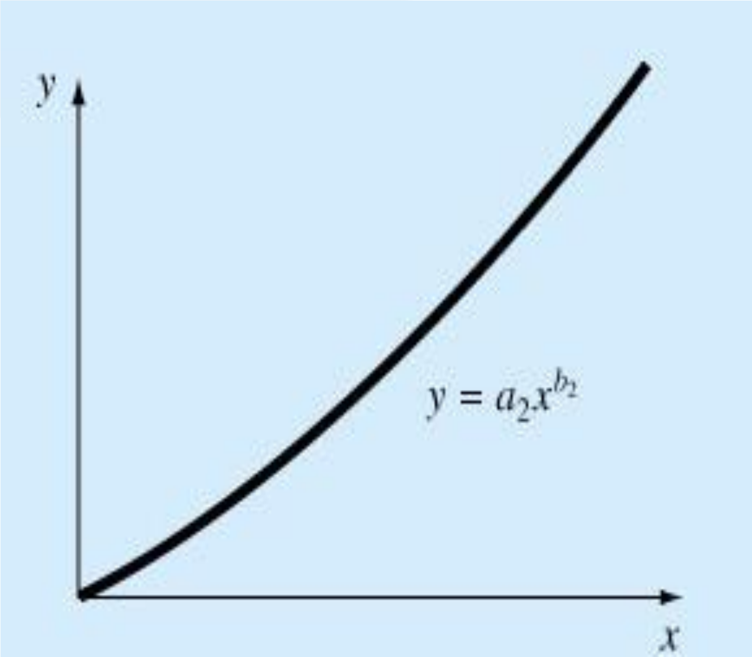


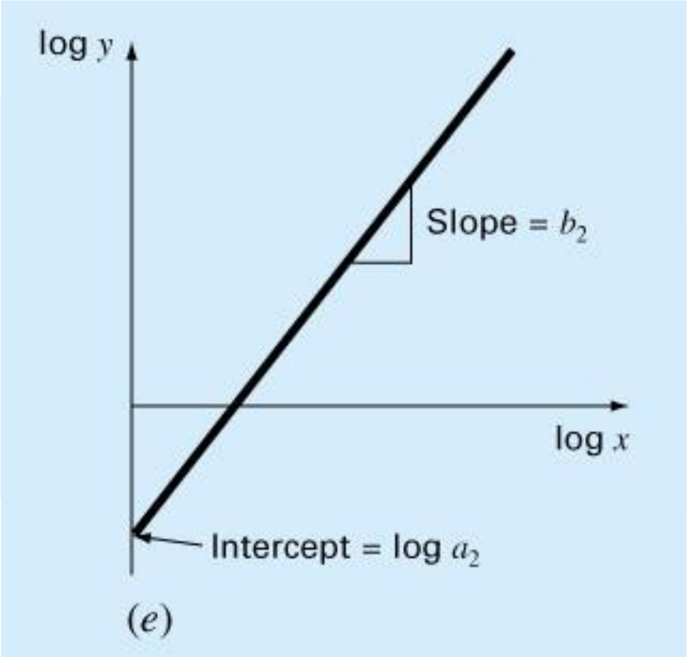
$$\ln y = \ln a_1 + b_1 x$$

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

Linearization of Nonlinear Relationships

2. The Power Equation معادلة قوة جبرية



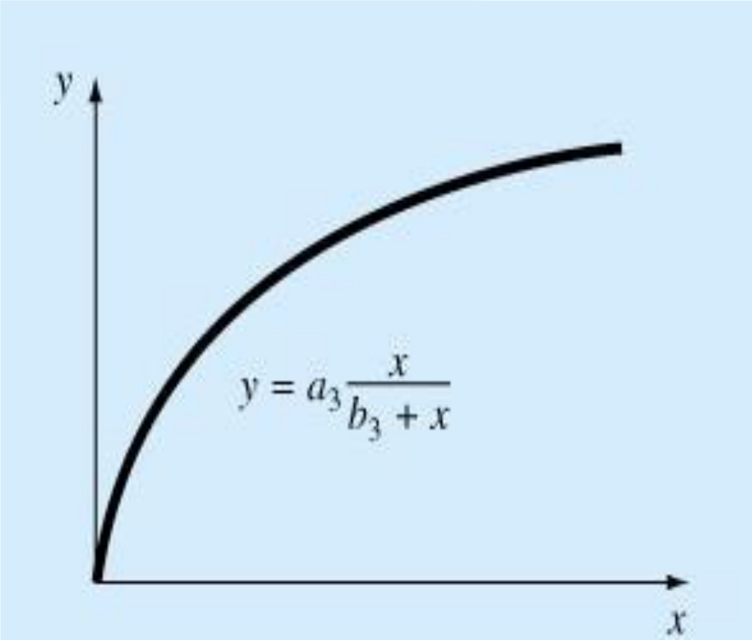


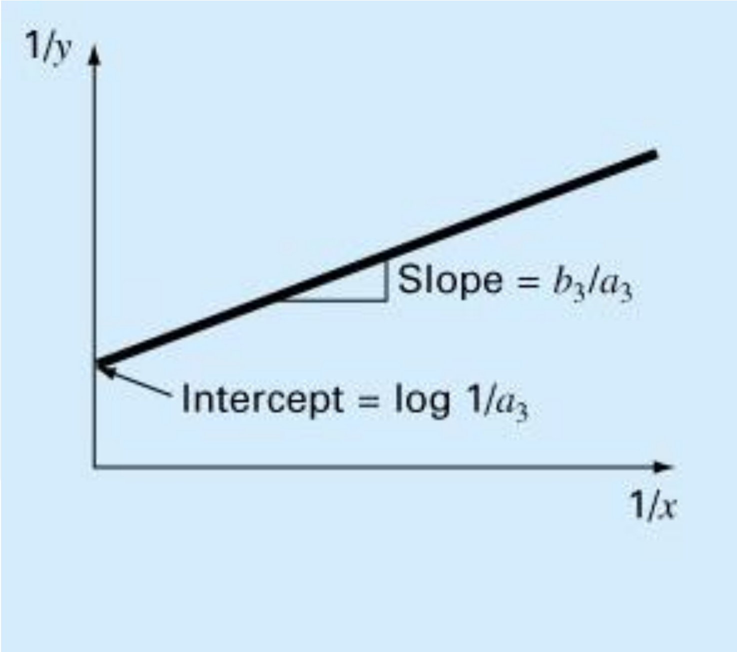
$$\log y = \log a_2 + b_2 \log x$$

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

Linearization of Nonlinear Relationships

3. The Saturation-Growth-Rate Equation معادلة معدل النمو





$$\frac{1}{y} = \frac{1}{a_3} + \frac{b_3}{a_3}\left(\frac{1}{x}\right)$$

$y^* = 1/y$
$a_o = 1/a_3$
$a_1 = b_3/a_3$
$x^* = 1/x$

Example

- Fit the following Equation: $y = a_2 x^{b_2}$ to the data in the following table:

$y = a_2 x^{b_2}$ المطلوب إلباس المعطيات المبينة في الجدول التالي بتابع قوة جبرية من الشكل

x_i	y_i	$X^*=\log x_i$	$Y^*=\log y_i$
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
15	19.7	2.079	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

- Fit the following Equation: $y = a_2 x^{b_2}$ to the data in the following table:

	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**

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

Homework - P_01

- لدراسة علاقة الاستهلاك المحلي y بالإنتاج x لمادة الاسفلت (بالمليون برميل) خلال عدة سنوات، أخذنا عشر قراءات تقريبية كما يلي:

y	6	8	9	8	7	6	5	6	5	5
x	10	13	15	14	9	7	6	6	5	5

- أوجد معادلة الانحدار الخطي البسيط ومعامل الارتباط وتوقع قيمة الاستهلاك عندما يصل الإنتاج إلى 16000000 برميل .

Homework - P_02

المطلوب إلباس المعطيات المبينة في الجدول التالي بـ:

1. تابع معدل النمو.
2. تابع قوة جبرية.

ومن ثم ارسم البيانات والتابع لكل من الحالتين السابقتين.

x	0.75	2	3	4	6	8	8.5
y	1.2	1.95	2	2.4	2.5	2.7	2.6