Mahidol University
Faculty of Information
and Communication Technology

General Linear Least-Squares

Rawesak Tanawongsuwan, Ph.D.
rawesak.tan@mahidol.ac.th

This slide is part of teaching materials for ITCS122 Numerical Methods
Semester 2/2023, Calendar year 2024

Recall least-squares fit of a straight line

- A curve-fitting strategy by approximating the shape of the data without necessarily matching or passing through the individual points.
- We have seen a simple way is to fit a straight line to a set of data : $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ by minimizing the discrepancy between data points and the line.

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

  - a_0 and a_1 are coefficient representing the intercept and the slope.
  - e is the error or residual between the model and the data

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

  - the residual e is the discrepancy between true value of y and the approximate value $a_0 + a_1 x$, predicted by the linear equation.

Minimizing the sum of the squares of the residuals (*least squares* criterion)

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

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

$$\frac{\partial S_r}{\partial a_1} = -2\sum[(y_i - a_0 - a_1 x_i)x_i] \quad \Rightarrow \quad 0 = \sum x_i y_i - \sum a_0 x_i - \sum a_1 x_i^2$$

Note that $\sum a_0 = n a_0$

$$\left.\begin{aligned} n a_0 + \left(\sum x_i\right) a_1 &= \sum y_i \\ \left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 &= \sum x_i y_i \end{aligned}\right\}$$

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

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

Polynomial regression

- Now we will fit polynomials to a set of data : $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ by minimizing the discrepancy between data points and polynomials.
- For fitting a second-order polynomial or quadratic

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

  - a_0, a_1 and a_2 are coefficients of a second-order polynomial.
  - e is the error or residual between the model and the data

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

  - the residual e is the discrepancy between true value of y and the approximate value $a_0 + a_1 x + a_2 x^2$, predicted by the quadratic equation.

Minimizing the sum of the squares of the residuals (*least squares* criterion) for a second-order polynomial

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

$$\frac{\partial S_r}{\partial a_0} = -2\sum\left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right) \quad \Rightarrow \quad 0 = \sum y_i - \sum a_0 - \sum a_1 x_i - \sum a_2 x_i^2$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum\left[\left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right) x_i\right] \quad \Rightarrow \quad 0 = \sum x_i y_i - \sum a_0 x_i - \sum a_1 x_i^2 - \sum a_2 x_i^3$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum\left[\left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right) x_i^2\right] \quad \Rightarrow \quad 0 = \sum x_i^2 y_i - \sum a_0 x_i^2 - \sum a_1 x_i^3 - \sum a_2 x_i^4$$

Note that $\sum a_0 = n a_0$

Determining a least-squares second-order polynomial is equivalent to solving a system of three simultaneous linear equations

$$na_0 + \left(\sum x_i\right)a_1 + \left(\sum x_i^2\right)a_2 = \sum y_i$$

$$\left(\sum x_i\right)a_0 + \left(\sum x_i^2\right)a_1 + \left(\sum x_i^3\right)a_2 = \sum x_i y_i$$

$$\left(\sum x_i^2\right)a_0 + \left(\sum x_i^3\right)a_1 + \left(\sum x_i^4\right)a_2 = \sum x_i^2 y_i$$

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

Quantification of error of second-order polynomial regression

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

- The square of the discrepancy between the data and a single estimate of the measure of central tendency – the mean

$$S_r = \sum\left(y_i - a_0 - a_1x_i - a_2x_i^2\right)^2$$

- The square of the vertical distance between the data and another measure of central tendency – the second-order polynomial curve

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

- Standard deviation quantifies the spread of the data around the mean

$$s_{y/x} = \sqrt{\frac{S_r}{n-3}}$$

- Standard error of the estimate quantifies the spread of the data around the second-order polynomial regression curve

The standard error of the estimate $s_{y/x}$

- The subscript notation "y/x" represents the error for a predicted value of y corresponding to a particular value of x.
- Dividing by $n-3$ comes from three data-derived estimates a_0, a_1 and a_2 were used to compute S_r, therefore we have lost three degrees of freedom.
- Another justification for dividing by $n-3$ is that there is no such thing as the "spread of data" around a second-order polynomial curve connecting three points. Therefore, for the case where $n=3$, $s_{y/x}$ yields a meaningless result of infinity.

Goodness of fit (still the same as before)

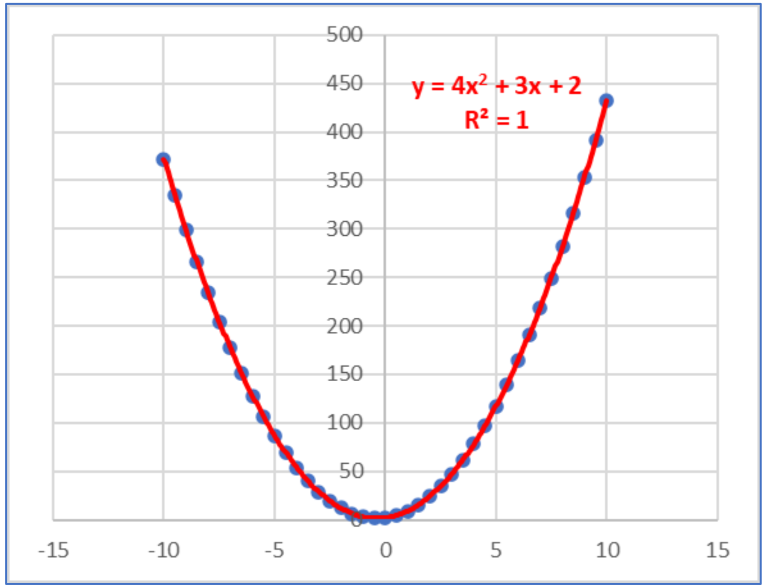
- The difference $S_t - S_r$ quantifies the improvement or error reduction due to describing the data in terms of a second-order polynomial rather than an average value (mean).
- Since the magnitude of this quantify is scale-dependent, we normalize to S_t to be

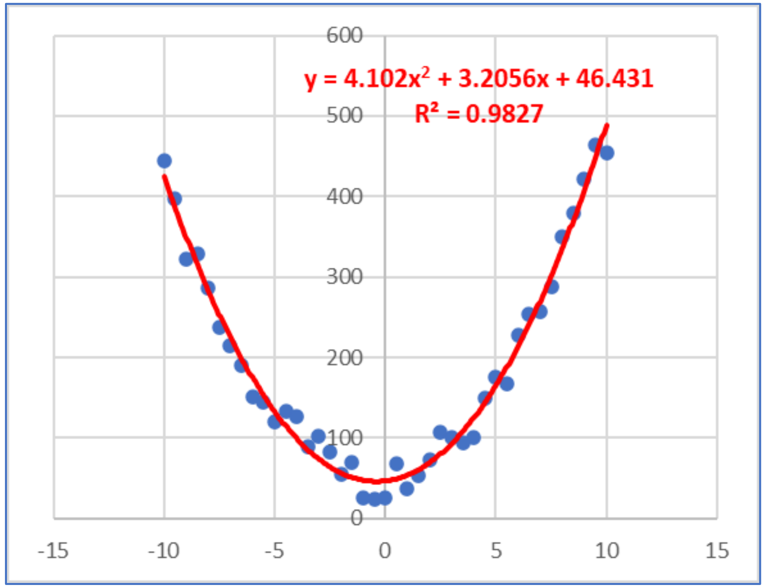
$$r^2 = \frac{S_t - S_r}{S_t}$$

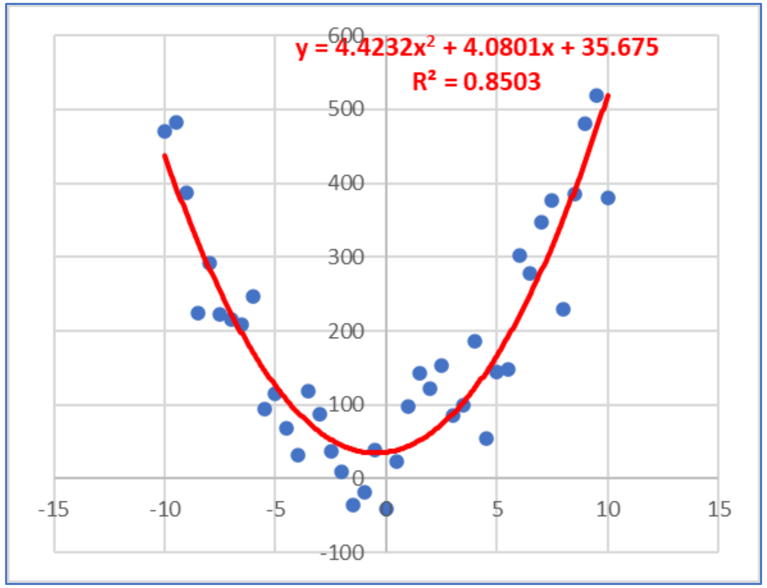
- r^2 is called *the coefficient of determination*
- r is called *the correlation coefficient* → $r = \sqrt{r^2}$

Goodness of fit

- For a perfect fit, $S_r = 0$ and $r^2 = 1$ → the second-order polynomial explains 100% of the variability of the data.
- For $r^2 = 0$ and $S_r = S_t$ → the second-order polynomial represents no improvement.







Example : Polynomial regression

	x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$	$(y_i - \bar{y})^2$	$\left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)^2$
	0	**2.1**	0	0	0	0	0	544.44	0.14334
	1	**7.7**	1	1	1	7.7	7.7	314.47	1.00280
	2	**13.6**	4	8	16	27.2	54.4	140.03	1.08160
	3	**27.2**	9	27	81	81.6	244.8	3.12	0.80497
	4	**40.9**	16	64	256	163.6	654.4	239.22	0.61937
	5	**61.1**	25	125	625	305.5	1527.5	1272.11	0.09449
$\sum$	**15**	**152.6**	**55**	**225**	**979**	**585.6**	**2488.8**	**2513.39**	**3.74657**

Example : Polynomial regression

- Simultaneous linear equations

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

$$\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \\ \\ \end{Bmatrix}$$

- Solved coefficients

$$\begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \\ \\ \end{Bmatrix}$$

- The least-squares quadratic equation

$y =$

Example : Polynomial regression

- The standard error of the estimate

$$s_{y/x} =$$

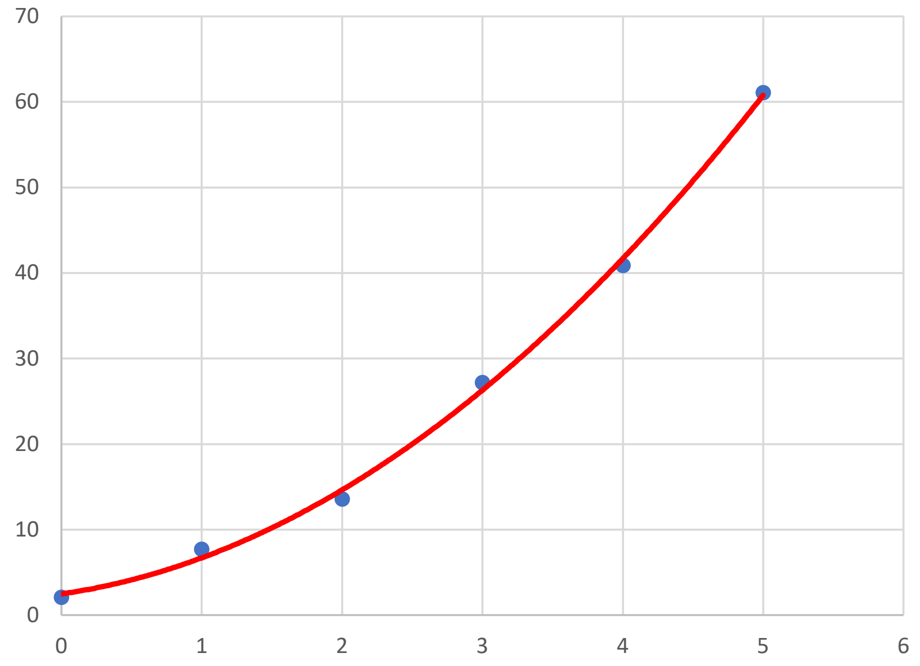
- The coefficient of determination

$$r^2 =$$

- The correlation coefficient

$$r =$$

- Goodness of fit of a second-order polynomial



For fitting an $m^{\mathbf{th}}$-order polynomial

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m + e$$

- Determining the coefficients of an m^{th}-order polynomial is equivalent to solving a system of $m + 1$ simultaneous linear equations.

- The standard error of the estimate

$$s_{y/x} = \sqrt{\frac{S_r}{n - (m + 1)}}$$

- The coefficient of determination

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

Multiple linear regression

- **Multiple linear regression** is used to estimate the relationship between two or more independent variables and one dependent variable.
- You can use multiple linear regression when you want to know:
  - How strong the relationship is between two or more independent variables and one dependent variable (e.g. how rainfall (x_1), temperature (x_2), and amount of fertilizer added (x_3) affect crop growth (y)).
  - The value of the dependent variable at a certain value of the independent variables (e.g. the expected yield of a crop at certain levels of rainfall, temperature, and fertilizer addition).

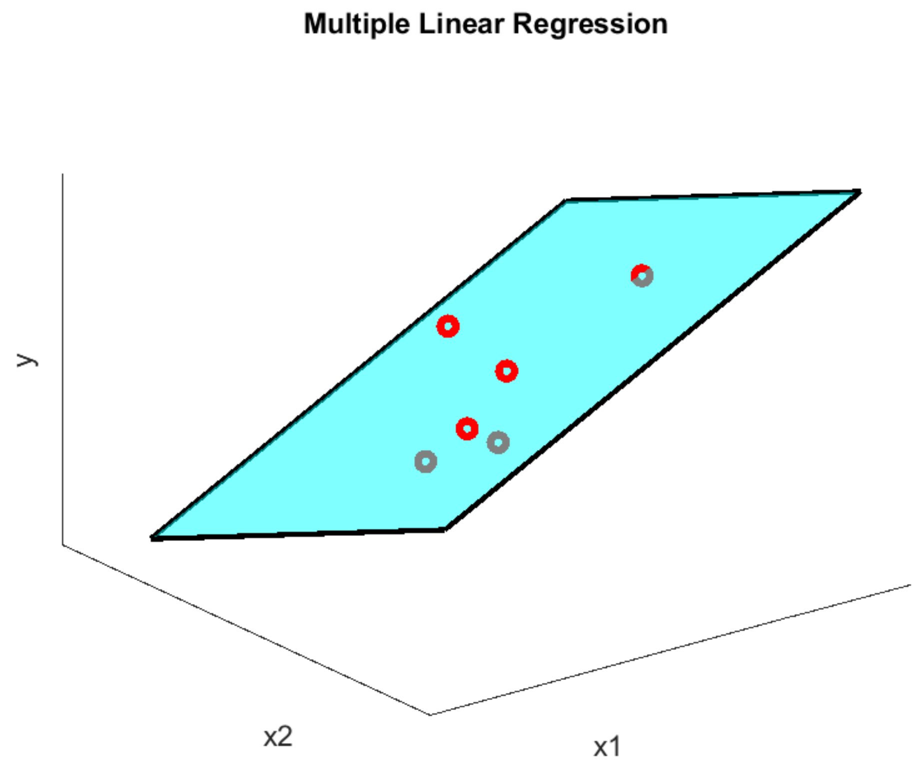
https://www.scribbr.com/statistics/multiple-linear-regression/

Multiple linear regression

- A linear function of x_1 and x_2

$$y = a_0 + a_1 x_1 + a_2 x_2 + e$$

- For the two-dimensional case, the regression "line" becomes a "plane".



Multiple linear regression

- Now we will fit a linear function to a set of data :

$$\left(x_{1,1}, x_{2,1}, y_1\right), \left(x_{1,2}, x_{2,2}, y_2\right), \dots, \left(x_{1,n}, x_{2,n}, y_n\right)$$

  by minimizing the discrepancy between data points and a linear function.

- For fitting a linear function

$$y = a_0 + a_1 x_1 + a_2 x_2 + e$$

  - a_0, a_1 and a_2 are coefficients of a linear function.
  - e is the error or residual between the model and the data

$$e = y - a_0 - a_1 x_1 - a_2 x_2$$

  - the residual e is the discrepancy between true value of y and the approximate value $a_0 + a_1 x + a_2 x_2$, predicted by the linear function.

Minimizing the sum of the squares of the residuals (*least squares* criterion) for a linear function

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

$$\frac{\partial S_r}{\partial a_0} = -2\sum\left(y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i}\right) \quad \Rightarrow \quad 0 = \sum y_i - \sum a_0 - \sum a_1 x_{1,i} - \sum a_2 x_{2,i}$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum\left[\left(y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i}\right) x_{1,i}\right] \quad \Rightarrow \quad 0 = \sum x_{1,i} y_i - \sum a_0 x_{1,i} - \sum a_1 x_{1,i}^2 - \sum a_2 x_{1,i} x_{2,i}$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum\left[\left(y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i}\right) x_{2,i}\right] \quad \Rightarrow \quad 0 = \sum x_{2,i} y_i - \sum a_0 x_{2,i} - \sum a_{1,i} x_{2,i} - \sum a_2 x_{2,i}^2$$

Note that $\sum a_0 = n a_0$

Determining a least-squares multiple linear function is equivalent to solving a system of three simultaneous linear equations

$$na_0 + \left(\sum x_{1,i}\right)a_1 + \left(\sum x_{2,i}\right)a_2 = \sum y_i$$

$$\left(\sum x_{1,i}\right)a_0 + \left(\sum x_{1,i}^2\right)a_1 + \left(\sum x_{1,i}x_{2,i}\right)a_2 = \sum x_{1,i}y_i$$

$$\left(\sum x_{2,i}\right)a_0 + \left(\sum x_{1,i}x_{2,i}\right)a_1 + \left(\sum x_{2,i}^2\right)a_2 = \sum x_{2,i}y_i$$

$$\begin{bmatrix} n & \sum x_{1,i} & \sum x_{2,i} \\ \sum x_{1,i} & \sum x_{1,i}^2 & \sum x_{1,i}x_{2,i} \\ \sum x_{2,i} & \sum x_{1,i}x_{2,i} & \sum x_{2,i}^2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_{1,i}y_i \\ \sum x_{2,i}y_i \end{Bmatrix}$$

Quantification of error of a multiple linear regression

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

- The square of the discrepancy between the data and a single estimate of the measure of central tendency – the mean

$$S_r = \sum \left(y_i - a_0 - a_1 x_{1,i} - a_2 x_{2,i}\right)^2$$

- The square of the vertical distance between the data and another measure of central tendency – the plane

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

- Standard deviation quantifies the spread of the data around the mean

$$s_{y/x} = \sqrt{\frac{S_r}{n-3}}$$

- Standard error of the estimate quantifies the spread of the data around the multiple linear regression curve

Goodness of fit (still the same as before)

- The difference $S_t - S_r$ quantifies the improvement or error reduction due to describing the data in terms of a multiple linear function rather than an average value (mean).
- Since the magnitude of this quantify is scale-dependent, we normalize to S_t to be

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

- r^2 is called *the coefficient of determination*
- r is called *the correlation coefficient* → $r = \sqrt{r^2}$

For fitting an m-dimension linear function

$$y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m + e$$

- Determining the coefficients of an m-dimension linear function is equivalent to solving a system of $m + 1$ simultaneous linear equations.

- The standard error of the estimate

$$s_{y/x} = \sqrt{\frac{S_r}{n - (m + 1)}}$$

- The coefficient of determination

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

Example : Multiple linear regression

	x_{1i}	x_{2i}	y_i	x_{1i}^2	x_{2i}^3	$x_{1i}x_{2i}$	$x_{1i}y_i$	$x_{2i}y_i$	$(y_i - \bar{y})^2$	$(y_i - a_0 - a_1x_{1i} - a_2x_{2i})^2$
	0	**0**	**3**	**0**	**0**	**0**	**0**	**0**	**25**	**2.3339**
	2	**1**	**10.5**	**4**	**1**	**2**	**21**	**10.5**	**6.25**	**2.7616**
	2.5	**2**	**8.5**	**6.25**	**4**	**5**	**21.25**	**17**	**0.25**	**0.2552**
	1	**3**	**0.5**	**1**	**9**	**3**	**0.5**	**1.5**	**56.25**	**0.0485**
	4	**6**	**2.5**	**16**	**36**	**24**	**10**	**15**	**30.25**	**0.1628**
	7	**2**	**23**	**49**	**4**	**14**	**161**	**46**	**225**	**0.2083**
$\sum$	**16.5**	**14**	**48**	**76.25**	**54**	**48**	**213.75**	**90**	**343**	**5.7703**

Example : Multiple linear regression

- Simultaneous linear equations

$$\begin{bmatrix} n & \sum x_{1,i} & \sum x_{2,i} \\ \sum x_{1,i} & \sum x_{1,i}^2 & \sum x_{1,i}x_{2,i} \\ \sum x_{2,i} & \sum x_{1,i}x_{2,i} & \sum x_{2,i}^2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_{1,i}y_i \\ \sum x_{2,i}y_i \end{Bmatrix}$$

$$\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \\ \\ \end{Bmatrix}$$

- Solved coefficients

$$\begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \\ \\ \end{Bmatrix}$$

- The least-squares quadratic equation

$$y =$$

Example : Multiple linear regression

- The standard error of the estimate

$s_{y/x} =$

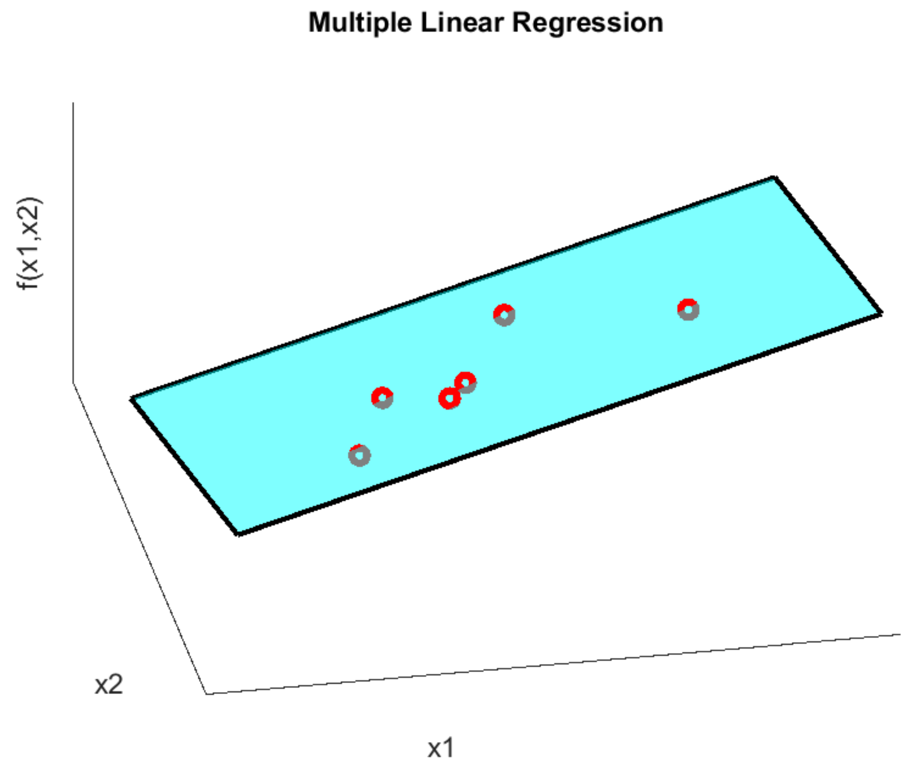
- The coefficient of determination

$r^2 =$

- The correlation coefficient

$r =$

- Goodness of fit of a multiple linear function



General linear least squares

- The general linear least-squares model :

$$y = a_0 z_0 + a_1 z_1 + a_2 z_2 + \cdots + a_m z_m + e$$

where z_0, z_1 and z_2 are $m + 1$ basis functions.

- All 3 types of regression that we have studied belong to the general model :

Simple linear $\quad y = a_0 + a_1 x + e$
- $z_0 = 1, z_1 = x$

Polynomial $\quad y = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m + e$
- $z_0 = 1, z_1 = x, z_2 = x^2, \dots, z_m = x^m$

Multiple linear $\quad y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_m x_m + e$
- $z_0 = 1, z_1 = x_1, z_2 = x_2, \dots, z_m = x_m$

General linear least squares

- "linear" refers to the model's dependence on its parameters (a's)
  - The functions themselves can be nonlinear.
  - For example,
    - Polynomial $y = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m + e$
    - Sinusoids $y = a_0 + a_1 \cos(\omega x) + a_2 \sin(\omega x)$

- "nonlinear" refers to the model that can not be manipulated into the linear format
  - For example
    - $y = a_0(1 - e^{-a_1 x})$

General linear least squares

$$y = a_0 z_0 + a_1 z_1 + a_2 z_2 + \cdots + a_m z_m + e$$

- General linear least-squares model can be expressed in matrix notation as

$$\{y\} = [Z]\{a\} + \{e\}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} z_{01} & z_{11} & \cdots & z_{m1} \\ z_{02} & z_{12} & \cdots & z_{m2} \\ \vdots & \vdots & & \vdots \\ z_{0n} & z_{1n} & \cdots & z_{mn} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

Minimizing the sum of the squares of the residuals (*least squares* criterion)

$$S_r = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{n} a_j z_{ji} \right)^2$$

This quantity can be minimized by taking its partial derivative with respect to each of the coefficients and setting the resulting equations equal to zero.

The outcome is the normal equations in matrix form as

$$\big[[Z]^T[Z]\big]\{a\} = \{[Z]^T\{y\}\}$$

It can be shown that this normal equation is equivalent to the normal equations developed for simple linear, polynomial, and multiple linear regression.

Goodness of fit

$$r^2 = \frac{S_t - S_r}{S_t} = 1 - \frac{S_r}{S_t}$$

$$r^2 = 1 - \frac{\sum (y_i - \widehat{y_i})^2}{\sum (y_i - \overline{y_i})^2}$$

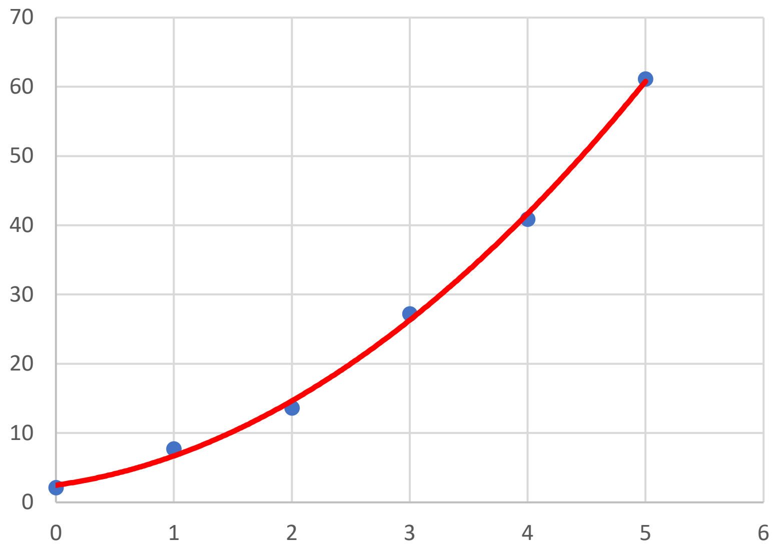
$\hat{y}$ = the prediction of the least-squares fit.

$y_i - \hat{y}$ = the residuals between the best-fit curve and the data

$= \{y\} - [Z]\{a\}$

Example : Polynomial regression

	x_i	y_i	x_i^2	x_i^3	x_i^4	$x_i y_i$	$x_i^2 y_i$	$(y_i - \bar{y})^2$	$\left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)^2$
	0	**2.1**	0	0	0	0	0	544.44	0.14334
	1	**7.7**	1	1	1	7.7	7.7	314.47	1.00280
	2	**13.6**	4	8	16	27.2	54.4	140.03	1.08160
	3	**27.2**	9	27	81	81.6	244.8	3.12	0.80497
	4	**40.9**	16	64	256	163.6	654.4	239.22	0.61937
	5	**61.1**	25	125	625	305.5	1527.5	1272.11	0.09449
Σ	**15**	**152.6**	**55**	**225**	**979**	**585.6**	**2488.8**	**2513.39**	**3.74657**



Example : Using the general linear least squares

```matlab
x = [0 1 2 3 4 5]';
y = [2.1 7.7 13.6 27.2 40.9 61.1]';
dataset = [x y]
```

```
dataset = 6×2
         0    2.1000
    1.0000    7.7000
    2.0000   13.6000
    3.0000   27.2000
    4.0000   40.9000
    5.0000   61.1000
```

```matlab
Z = [ones(size(x)) x x.^2]

% the unknown coefficients
a = (Z'*Z) \ (Z'*y)
```

```
Z = 6×3
     1     0     0
     1     1     1
     1     2     4
     1     3     9
     1     4    16
     1     5    25

a = 3×1
    2.4786
    2.3593
    1.8607
```

```matlab
% the sum of the squares of the residuals
% between the best-fit curve and the data
Sr = sum((y-Z*a).^2)

% the sum of the squares of the residuals
% between the data points and the mean
St = sum((y-mean(y)).^2)

% the coefficient of determination
r2 = 1 - Sr/St

% the standard error of the estimate
syx = sqrt(Sr/(length(x)-length(a)))
```

```
Sr = 3.7466

St = 2.5134e+03

r2 = 0.9985

syx = 1.1175
```