THE CHINESE UNIVERSITY OF HONG KONG
Department of Mathematics
Exercises on Modelling Data using Functions

Learning Outcomes 學習目標

- Apply the principle of least squares.
  應用最小二乘原理
- Use the procedure for fitting any curve or functional relationship.
  使用擬合任何曲線或函數關係的程序
- Calculate the residuals.
  計算殘差
  - Fit a straight line to the given data.
    將直線擬合到給定數據
  - Fit a second-degree parabola to the given data.
    將二次拋物線擬合到給定數據
  - Fit a power curve $Y = aX^b$ to the given data.
    將冪曲線$Y = aX^b$ 擬合到給定數據
  - Fit exponential curves of the forms $Y = ab^X$ and $Y = ae^X$, where $e = 2.71828182846$.
    將形式為$Y = ab^X$ 和$Y = ae^X$ 的指數曲線擬合到給定數據，其中$e = 2.71828182846$
- Visualize the data using a scatter plot and the best-fitted line/curve in the same figure using the Linear and Nonlinear RShiny tools.
  使用Linear and Nonlinear RShiny 工具在同一圖中使用散點圖和最佳擬合線/曲線可視化數據。

Objectives 目標

Our aim is to estimate the parameters of a straight line, a second-degree parabola, a power curve $Y = aX^b$, and exponential curves using the method of least squares. We will use the Linear and Nonlinear RShiny apps as validation tools to check our hand-calculated findings. Additionally, calculating the residual sum of squares and the coefficients of determination will help us understand how well the mathematical models perform.

我們的目標是使用最小二乘法估計直線、二次拋物線、冪曲線$Y = aX^b$ 和指數曲線的參數。我們將使用Linear and Nonlinear RShiny 應用程式作為驗證工具來檢查我們計算的結果。此外，計算殘差平方和和決定係數將幫助我們了解數學模型的表現。

Problem 1

Do the following:
執行以下步驟：

Enter the Y values and X values into the Linear and Nonlinear Regression RShiny Tool apps to examine different regression models. Describe the trends of the linear and nonlinear regression lines and curves.

在Linear and Nonlinear Regression RShiny Tool 應用程式中，輸入Y 值和X 值，以檢查不同的回歸模型。描述線性和非線性回歸線和曲線的趨勢。

1. Find a straight line that best fits the following data:
   找到一條最適合以下數據的直線：

x	1	6	11	16	20	26
y	13	16	17	23	24	31

2. Find a straight line that best fits the following data:
   找到一條最適合以下數據的直線：

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

3. Find a straight line that best fits the following data:
   找到一條最適合以下數據的直線：

Year	Twitter users
11.00	68
11.25	85
11.50	101
11.75	117
12.75	185

4. Fit a second-degree parabola to the following data:
   將一條二次拋物線擬合到以下數據：

x	y
0	1
1	3
2	4
3	5
4	6

5. Fit the power curve $Y = aX^b$ to the following data:
   將冪曲線$Y = aX^b$ 擬合到以下數據：

x	y
6	9
2	11
10	12
5	8
8	7

6. Fit the power curve $Y = aX^b$ to the following data:
將冪曲線$Y = aX^b$ 擬合到以下數據：

x	y
0.135	22000
0.67	13000
2.45	8000
75	2550
1775	900

7. Fit the exponential curve $Y = ab^X$ to the following data:
將指數曲線$Y = ab^X$ 擬合到以下數據：

x	y
2	1
4	3
6	6
8	12
10	24

8. Fit an exponential curve $Y = ae^{bX}$ to the following data:
將指數曲線$Y = ae^{bX}$ 擬合到以下數據：

x	y
0	1481.66
2	2333.32
4	3501.39
6	5938.71
8	8838.60

Problem 2

Fill in all the missing places in **Solution of Question 1**.
填寫問題1的解答中所有缺失的地方。

Given the dataset:
給定數據集：

x	1	6	11	16	20	26
y	13	16	17	23	24	31

Let the straight line be given by the equation $Y = a + bX$. To determine the values of a and b for this line, we use the normal equations:
設直線方程為$Y = a + bX$，為了確定這條線的a和b值，我們使用正則方程式：

$$\begin{aligned}\sum_{i=1}^{n} y_i &= na + b\sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} y_i x_i &= a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2\end{aligned}$$

can be written as
可以寫作

$$\begin{aligned}\sum y &= na + b\sum x \\ \sum yx &= a\sum x + b\sum x^2\end{aligned}$$

We need to calculate $\sum y$, $\sum x$, $\sum xy$, and $\sum x^2$, which can be obtained from the following table:
我們需要計算$\sum y$, $\sum x$, $\sum xy$, and $\sum x^2$，這些可以從下表中獲得：

i	x	y	x^2	xy
1	1	13	1	13
2	6	16	36	96
3	11	17	121	187
4	16	23	256	368
5	20	24	400	480
6	26	31	676	806
$n=6$	$\sum x = 80$	$\sum y = 124$	$\sum x^2 = 1490$	$\sum xy = 1950$

For example,
例如，

$$\begin{aligned}\sum x &= \sum_{i=1}^{6} x_i \\ &= x_1 + x_2 + x_3 + x_4 + x_5 + x_6 \\ &= 1 + 6 + 11 + 16 + 20 + 26 \\ &= \underline{\qquad},\end{aligned}$$

$$\begin{aligned}\sum y &= \sum_{i=1}^{6} y_i \\ &= y_1 + y_2 + y_3 + y_4 + y_5 + y_6 \\ &= 13 + 16 + 17 + 23 + 24 + 31 \\ &= \underline{\qquad},\end{aligned}$$

$$\begin{aligned}\sum y &= \sum_{i=1}^{6} x_i^2 \\ &= x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 \\ &= 1^2 + 6^2 + 11^2 + 16^2 + 20^2 + 26^2\end{aligned}$$

$$= 1 + 36 + 121 + 256 + 400 + 676$$
$$= \_\_\_\_\_\_,$$

and
及

$$\begin{aligned} \sum xy &= \sum_{i=1}^{6} x_i y_i \\ &= x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 + x_5y_5 + x_6y_6 \\ &= 13 + 96 + 187 + 368 + 480 + 806 \\ &= \_\_\_\_\_\_ \end{aligned}$$

By substituting the values of $\sum y$, $\sum x$, $\sum xy$, and $\sum x^2$ into the normal equations, we obtain:
將$\sum y$、$\sum x$、$\sum xy$ 和$\sum x^2$ 的值代入正則方程式，我們得到：

$$\begin{aligned} \_\_\_\_\_\_ &= \_\_\_\_\_\_ a + \_\_\_\_\_\_ b \quad (1) \\ \_\_\_\_\_\_ &= \_\_\_\_\_\_ a + \_\_\_\_\_\_ b \quad (2) \end{aligned}$$

Solve equations (1) and (2) with these values for a and b using the method of elimination, and verify the equation for the line of best fit is $\widehat{Y} = 11.3227 + 0.7008X$ using a pen-and-paper approach.
使用消元法求解方程(1)和(2)中的a 和b，並透過手算驗證最佳擬合直線的方程為$\widehat{Y} = 11.3227 + 0.7008X$。

Compute the predicated value $\widehat{y_i}$ at x_i, e.g.,
計算在x_i 處的預測值$\widehat{y_i}$ ，例如：

$$\begin{aligned} \widehat{y_1} &= 11.3227 + 0.7008x_i = 11.3227 + 0.7008(1) = \_\_\_\_\_\_ \\ \widehat{y_2} &= 11.3227 + 0.7008x_2 = 11.3227 + 0.7008(6) = \_\_\_\_\_\_ \\ \widehat{y_3} &= 11.3227 + 0.7008x_3 = 11.3227 + 0.7008(11) = \_\_\_\_\_\_ \\ \widehat{y_4} &= 11.3227 + 0.7008x_4 = 11.3227 + 0.7008(16) = \_\_\_\_\_\_ \\ \widehat{y_5} &= 11.3227 + 0.7008x_5 = 11.3227 + 0.7008(20) = \_\_\_\_\_\_ \\ \widehat{y_6} &= 11.3227 + 0.7008x_6 = 11.3227 + 0.7008(26) = \_\_\_\_\_\_ \end{aligned}$$

Compute the residuals $y_i - \widehat{y_i}$ at x_i, e.g.,
計算在x_i 處的殘差$y_i - \widehat{y_i}$ ，例如：

$$\begin{aligned} y_1 - \widehat{y_1} &= \_\_\_\_\_\_ \\ y_2 - \widehat{y_2} &= \_\_\_\_\_\_ \\ y_3 - \widehat{y_3} &= \_\_\_\_\_\_ \\ y_4 - \widehat{y_4} &= \_\_\_\_\_\_ \\ y_5 - \widehat{y_5} &= \_\_\_\_\_\_ \\ y_6 - \widehat{y_6} &= \_\_\_\_\_\_ \end{aligned}$$

Compute the residual sum of squares, e.g.,
計算殘差平方和，例如：

$$\sum_{i=1}^{n=6} (y_i - \widehat{y_i})^2$$

$$= (y_1 - \widehat{y_1})^2 + (y_2 - \widehat{y_2})^2 + (y_3 - \widehat{y_3})^2 + (y_4 - \widehat{y_4})^2 + (y_5 - \widehat{y_5})^2 + (y_6 - \widehat{y_6})^2$$
$$= \_\_\_\_\_\_$$

Compute the coefficient of determination, e.g.,
計算決定係數，例如：

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}$$
$$= \_\_\_\_\_\_$$

where
其中

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{1}{n}\left(Y_1 + Y_2 + Y_3 + Y_4 + Y_5 + Y_6\right) = \_\_\_\_\_\_$$

Discuss the applications of R-squared in linear regression.
討論R 平方在線性回歸中的應用。

Problem 3

Fill in all the missing places in **Solution of Question 4**.
填寫問題4 的解答中所有缺失的地方。

x	y
0	1
1	3
2	4
3	5
4	6

Let $\widehat{Y} = a + bX + cX^2$ be the second-degree parabola, and we have to determine a, b and c. Normal equations for the second-degree parabola are
設$\widehat{Y} = a + bX + cX^2$ 為二次拋物線，我們需確定a、b 和c。二次拋物線的正則方程式為

$$\begin{aligned}
\sum y &= na + b\sum x + c\sum x^2 \\
\sum xy &= a\sum x + b\sum x^2 + c\sum x^3 \\
\sum x^2 y &= a\sum x^2 + b\sum x^3 + c\sum x^4
\end{aligned}$$

To solve the above normal equations, we need $\sum y$, $\sum x$, $\sum xy$, $\sum x^2y$, $\sum x^2$, $\sum x^3$, and $\sum x^4$, which are obtained from the following table:
為了解上述正則方程式，我們需要$\sum y$、$\sum x$、$\sum xy$、$\sum x^2y$、$\sum x^2$、$\sum x^3$ 和$\sum x^4$，這些可以從下表中獲得：

i	x	y	xy	x^2	x^2y	x^3	x^4
1	0	1	0	0	0	0	0
2	1	3	3	1	3	1	1
3	2	4	8	4	16	8	16
4	3	5	15	9	45	27	81
5	4	6	24	16	96	64	256
$n=5$	$\sum x = 10$	$\sum y = 19$	$\sum xy = 50$	$\sum x^2 = 30$	$\sum x^2y = 160$	$\sum x^3 = 100$	$\sum x^4 = 354$

Substituting the values of $\sum y$, $\sum x$, $\sum xy$, $\sum x^2y$, $\sum x^2$, $\sum x^3$, and $\sum x^4$ into the above normal equations, we have:
將$\sum y$、$\sum x$、$\sum xy$、$\sum x^2y$、$\sum x^2$、$\sum x^3$ 和$\sum x^4$ 的值代入上述正則方程式，我們得到：

$$\begin{array}{rcrcrcr} \sum y &=& na &+& b\sum x &+& c\sum x^2 \\ \sum xy &=& a\sum x &+& b\sum x^2 &+& c\sum x^3 \\ \sum x^2y &=& a\sum x^2 &+& b\sum x^3 &+& c\sum x^4 \end{array}$$

$$\begin{array}{rcrcrcrl} \_\_\_\_\_\_ &=& \_\_\_\_\_\_a &+& \_\_\_\_\_\_b &+& \_\_\_\_\_\_c & (1) \\ \_\_\_\_\_\_ &=& \_\_\_\_\_\_a &+& \_\_\_\_\_\_b &+& \_\_\_\_\_\_c & (2) \\ \_\_\_\_\_\_ &=& \_\_\_\_\_\_a &+& \_\_\_\_\_\_b &+& \_\_\_\_\_\_c & (3) \end{array}$$

Solve equations (1), (2) and (3) with these values for a, b and c using the method of elimination, and verify the equation for the second degree of parabola of best fit is $\widehat{Y} = 1.114 + 1.7717X - 0.1429X^2$ using a pen-and-paper approach.
使用消元法求解方程(1)、(2)和(3)中的a、b 和c，並透過手算驗證最佳擬合二次拋物線的方程為$\widehat{Y} = 1.114 + 1.7717X - 0.1429X^2$。

Compute the predicated value $\widehat{y_i}$ at x_i, e.g.,
計算在x_i 處的預測值$\widehat{y_i}$ ，例如：

$$\begin{aligned} \widehat{y_1} &= 1.114 + 1.7717x_1 - 0.1429x_1^2 = \_\_\_\_\_\_ \\ \widehat{y_2} &= 1.114 + 1.7717x_2 - 0.1429x_2^2 = \_\_\_\_\_\_ \\ \widehat{y_3} &= 1.114 + 1.7717x_3 - 0.1429x_3^2 = \_\_\_\_\_\_ \\ \widehat{y_4} &= 1.114 + 1.7717x_4 - 0.1429x_4^2 = \_\_\_\_\_\_ \\ \widehat{y_5} &= 1.114 + 1.7717x_5 - 0.1429x_5^2 = \_\_\_\_\_\_ \end{aligned}$$

Compute the residuals $y_i - \widehat{y_i}$ at x_i, e.g.,
計算在x_i 處的殘差$y_i - \widehat{y_i}$ ，例如：

$$\begin{aligned} y_1 - \widehat{y_1} &= \_\_\_\_\_\_ \\ y_2 - \widehat{y_2} &= \_\_\_\_\_\_ \\ y_3 - \widehat{y_3} &= \_\_\_\_\_\_ \\ y_4 - \widehat{y_4} &= \_\_\_\_\_\_ \\ y_5 - \widehat{y_5} &= \_\_\_\_\_\_ \end{aligned}$$

Compute the residual sum of squares, e.g.,
計算殘差平方和，例如：

$$\sum_{i=1}^{n=5}(y_i-\widehat{y_i})^2=(y_1-\widehat{y_1})^2+(y_2-\widehat{y_2})^2+(y_3-\widehat{y_3})^2+(y_4-\widehat{y_4})^2+(y_5-\widehat{y_5})^2$$
$$= \underline{\qquad}$$

Compute the coefficient of determination, e.g.,
計算決定係數，例如：

$$R^2 = \underline{\qquad}$$

Discuss the applications of R-squared in nonlinear regression.
討論R 平方在非線性回歸中的應用。

Problem 4

Fill in all the missing places in **Solution of Question 7**.
填寫問題7 的解答中所有缺失的地方。

x	y
2	1
4	3
6	6
8	12
10	24

Let the exponential curve be $\widehat{Y}=ab^X$, and the normal equations for estimating a and b are
設指數曲線為$\widehat{Y}=ab^X$，用於估計a 和b 的正則方程式為

$$\begin{aligned}\sum u &= nA + B\sum x\\ \sum ux &= A\sum x + B\sum x^2\end{aligned}$$

where $a=\log^{-1}(A)=10^A$ and $b=\log^{-1}(B)=10^B$.
其中$a=\log^{-1}(A)=10^A$ 且$b=\log^{-1}(B)=10^B$。

Note: Here we are using log with base 10.
注意：這裡我們使用以10 為底的log。

To find the values of a and b from the above normal equations, we require $\sum u$, $\sum x$, $\sum ux$ and $\sum x^2$, which are being calculated in the following table:
為了從上述的正則方程式中找到a 和b 的值，我們需要$\sum u$、$\sum x$、$\sum ux$ 和$\sum x^2$，這些在以下表格中計算：

i	x	y	$u = \log y$	ux	x^2
1	2	1	0	0	4
2	4	3	0.477121255	1.908485019	16
3	6	6	0.77815125	4.668907502	36
4	8	12	1.079181246	8.633449968	64
5	10	24	1.380211242	13.80211242	100
$n = 5$	$\sum x = 30$	$\sum y = 46$	$\sum u = 3.714664993$	$\sum ux = 29.01295495$	$\sum x^2 = 220$

Substituting the values of $\sum u, \sum x, \sum ux$ and $\sum x^2$ in the above normal equations, we get
將$\sum u$、$\sum x$、$\sum ux$ 和$\sum x^2$ 的值代入上述正則方程式，我們得到

$$\begin{aligned} \_\_\_\_\_\_ &= \_\_\_\_\_\_A + \_\_\_\_\_\_B \quad (1) \\ \_\_\_\_\_\_ &= \_\_\_\_\_\_A + \_\_\_\_\_\_B \quad (2) \end{aligned}$$

Solve equations (1) and (2) with these values for A and B using the method of elimination. Then, use the properties of logarithmic functions to obtain a and b, and verify whether the equation for the exponential curve of best fit is $\widehat{Y} = 0.5424(1.4727)^X$ using a pen-and-paper approach.
使用消元法求解方程(1)和(2)中的A 和B。然後，利用對數函數的性質求得a 和b，並透過手算驗證最佳擬合指數曲線的方程是否為$\widehat{Y} = 0.5424(1.4727)^X$。

Compute the predicated value $\widehat{y_i}$ at x_i, e.g.,
計算在x_i 處的預測值$\widehat{y_i}$，例如：

$$\begin{aligned} \widehat{y_1} &= 0.5424(1.4727)^{x_1} = \_\_\_\_\_\_ \\ \widehat{y_2} &= 0.5424(1.4727)^{x_2} = \_\_\_\_\_\_ \\ \widehat{y_3} &= 0.5424(1.4727)^{x_3} = \_\_\_\_\_\_ \\ \widehat{y_4} &= 0.5424(1.4727)^{x_4} = \_\_\_\_\_\_ \\ \widehat{y_5} &= 0.5424(1.4727)^{x_5} = \_\_\_\_\_\_ \end{aligned}$$

Compute the residuals $y_i - \widehat{y_i}$ at x_i, e.g.,
計算在x_i 處的殘差$y_i - \widehat{y_i}$，例如：

$$\begin{aligned} y_1 - \widehat{y_1} &= \_\_\_\_\_\_ \\ y_2 - \widehat{y_2} &= \_\_\_\_\_\_ \\ y_3 - \widehat{y_3} &= \_\_\_\_\_\_ \\ y_4 - \widehat{y_4} &= \_\_\_\_\_\_ \\ y_5 - \widehat{y_5} &= \_\_\_\_\_\_ \end{aligned}$$

Compute the residual sum of squares, e.g.,
計算殘差平方和，例如：

$$\begin{aligned} \sum_{i=1}^{n=5} (y_i - \widehat{y_i})^2 &= (y_1 - \widehat{y_1})^2 + (y_2 - \widehat{y_2})^2 + (y_3 - \widehat{y_3})^2 + (y_4 - \widehat{y_4})^2 + (y_5 - \widehat{y_5})^2 \\ &= \_\_\_\_\_\_ \end{aligned}$$

Compute the coefficient of determination, e.g.,
計算決定係數，例如：

$$R^2 = \_\_\_\_\_\_$$

Discuss the applications of R-squared in nonlinear regression.
討論R 平方在非線性迴歸中的應用。