Enhanced Programme on Promoting Mathematical Modelling for Teachers and Students in Secondary Schools
Student Workshop 2025/26 (Senior)
推廣中學教師及學生數學建模計劃
學生工作坊 2025/26（高中）

Part I: Mathematical Modelling
第一部份：數學建模與數據分析

Prof. Gary Pui-Tung Choi 蔡沛彤教授

Dr. Jeff Chak-Fu Wong 黃澤富博士

Department of Mathematics, The Chinese University of Hong Kong

香港中文大學數學系

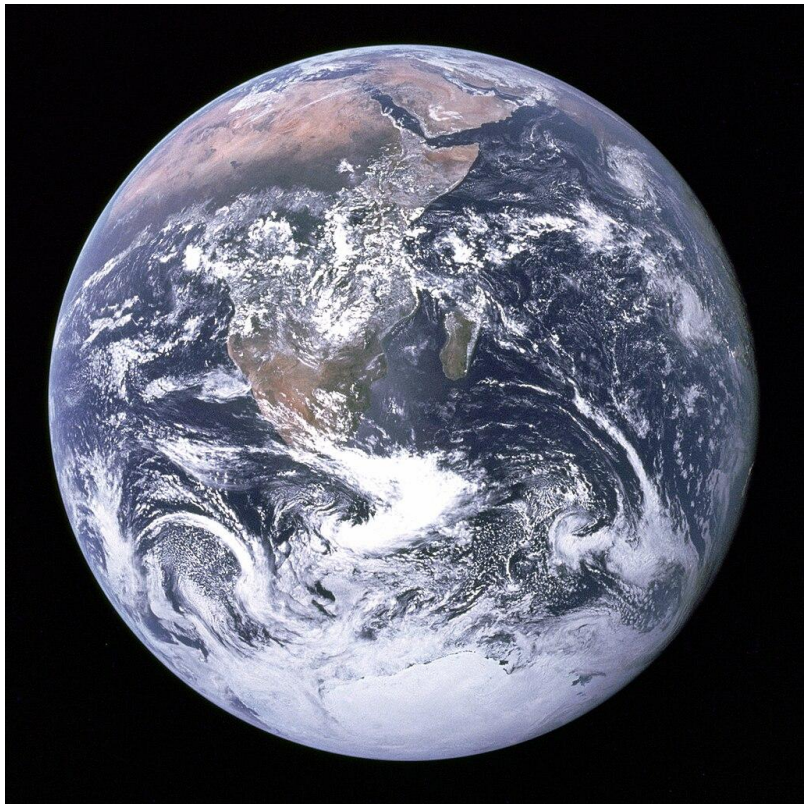
What is Mathematical Modelling?

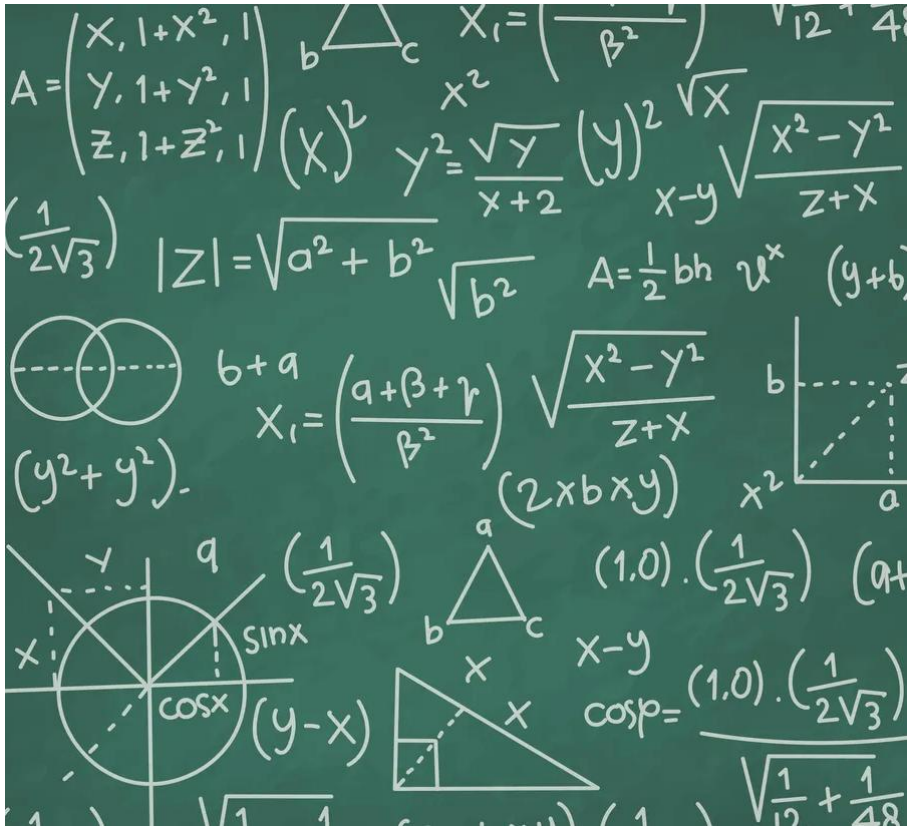
甚麼是數學建模？

Mathematical Modelling is … 數學建模是 ...

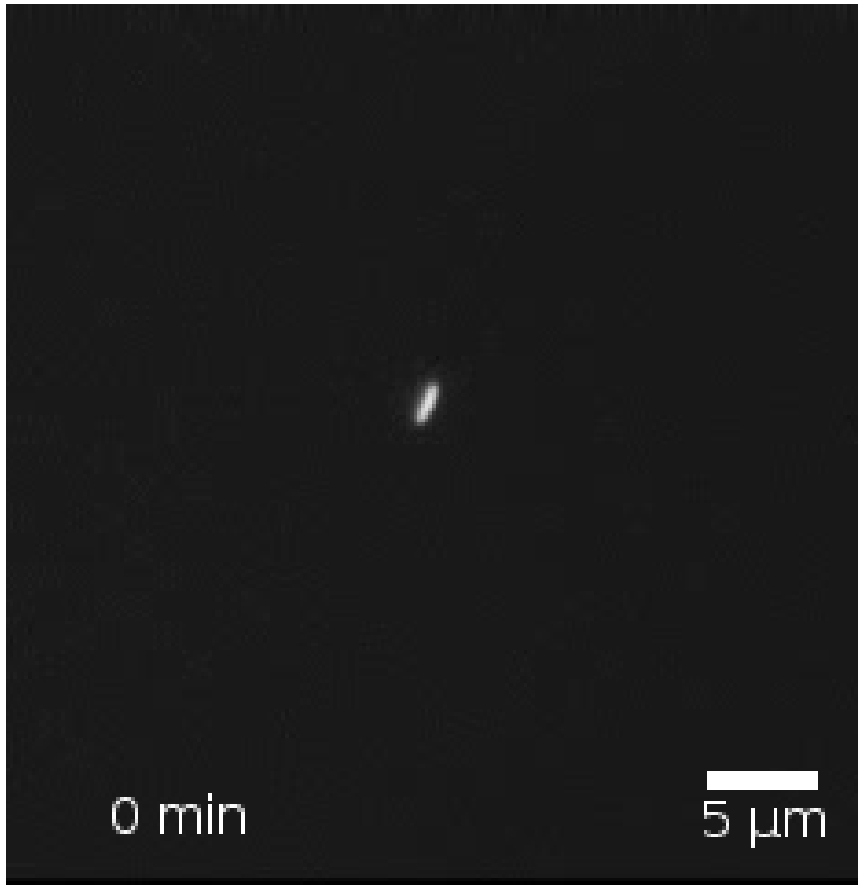
Understanding a real-world problem using mathematics

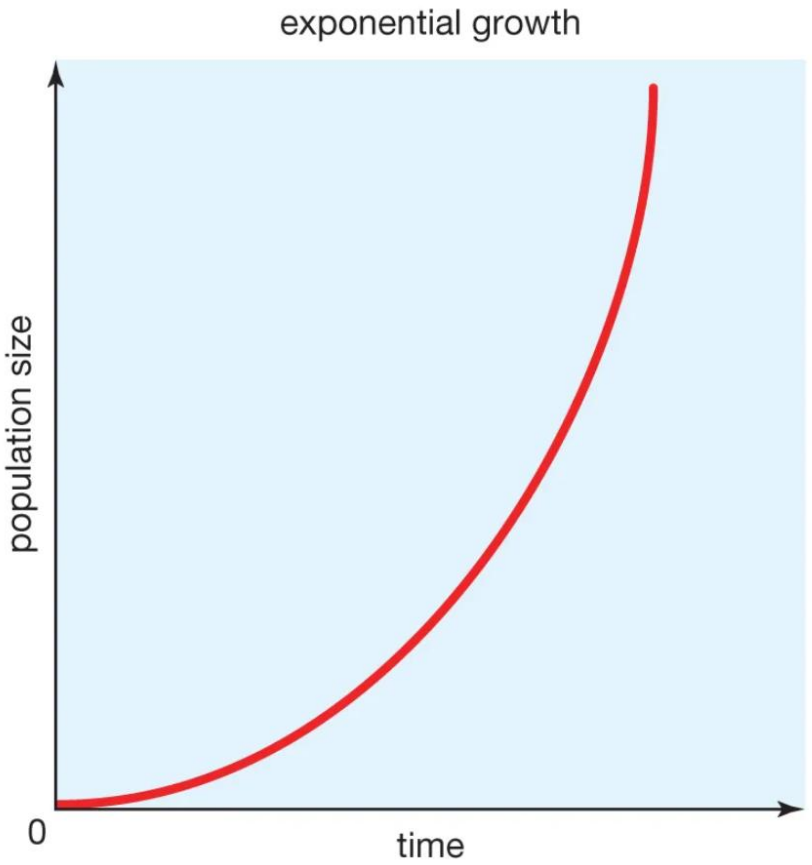
用數學了解現實生活問題





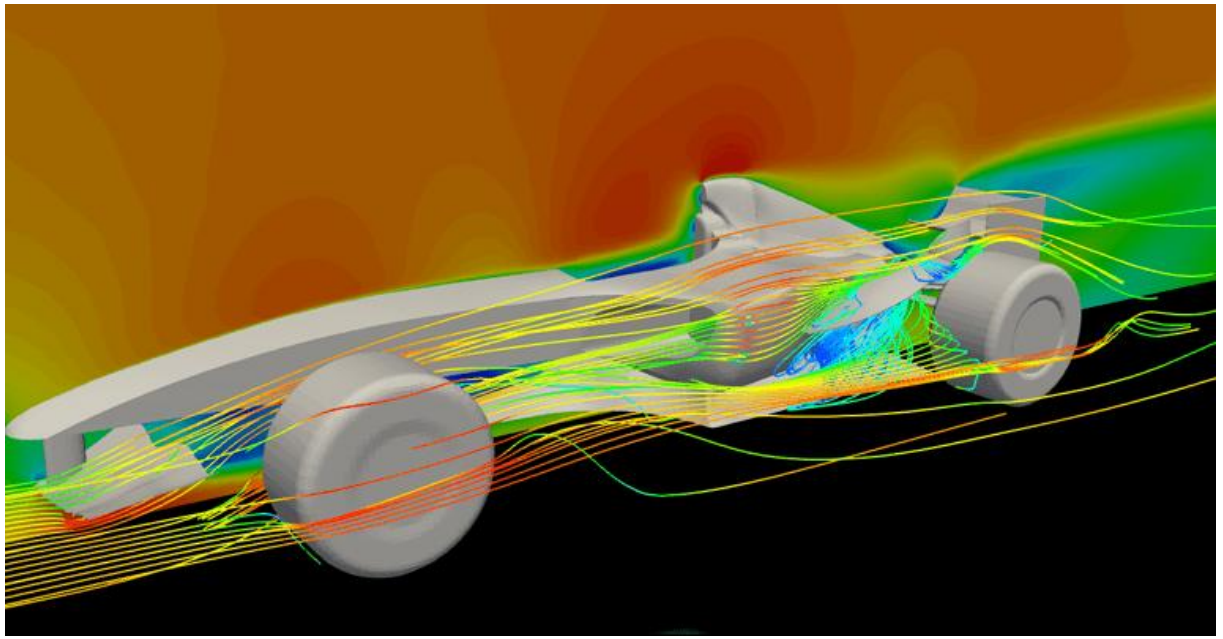
Mathematical Problems in Real Life 現實生活中的數學問題





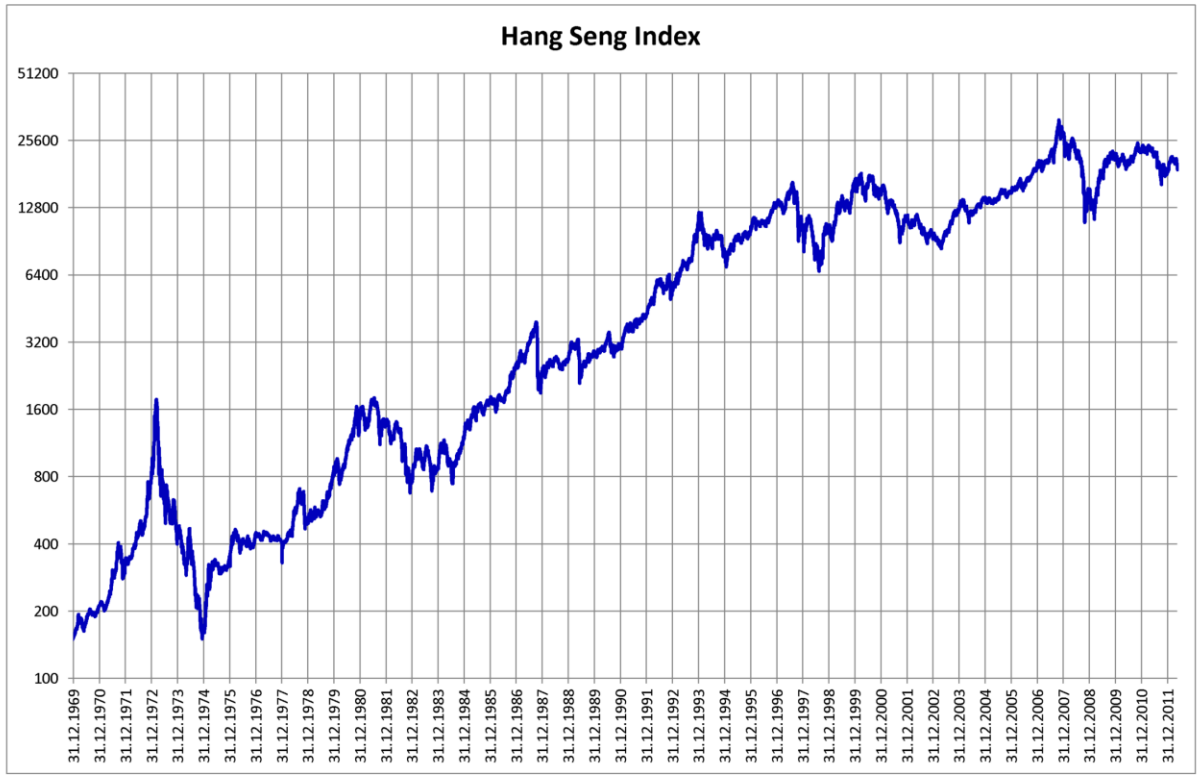
Source:
https://en.wikipedia.org/wiki/Exponential_growth

Mathematical Problems in Real Life 現實生活中的數學問題



Source:
https://www.themanual.com/auto/types-of-car-racing/
https://www.simscale.com/blog/cfd-analysis-for-beginners/

Mathematical Problems in Real Life 現實生活中的數學問題



Source:
https://en.wikipedia.org/wiki/Hang_Seng_Index

Mathematical Problems in Real Life 現實生活中的數學問題

Source:
https://en.wikipedia.org/wiki/Snow
A. Stomakhin et al., ACM Trans. Graph. (2013) https://www.youtube.com/watch?v=O0kyDKu8K-k

Features of Mathematical Modelling 數學建模的特點

- **Interdisciplinary 跨學科**
  - Science 科學
  - Technology 科技
  - Engineering 工程
  - Art 藝術
  - Mathematics 數學

Features of Mathematical Modelling 數學建模的特點

- **Interdisciplinary 跨學科**
  - Science 科學
  - Technology 科技
  - Engineering 工程
  - Art 藝術
  - Mathematics 數學

- **Target-oriented but exploratory** 既具明確目標亦富探究性
  - Tackle a concrete real-world problem
    解決一個具體的現實生活問題
  - Explore different mathematical approaches
    探索不同的數學方法

Features of Mathematical Modelling 數學建模的特點

- **Interdisciplinary 跨學科**
  - Science 科學
  - Technology 科技
  - Engineering 工程
  - Art 藝術
  - Mathematics 數學

- **Target-oriented but exploratory** 既具明確目標亦富探究性
  - Tackle a concrete real-world problem
    解決一個具體的現實生活問題
  - Explore different mathematical approaches
    探索不同的數學方法

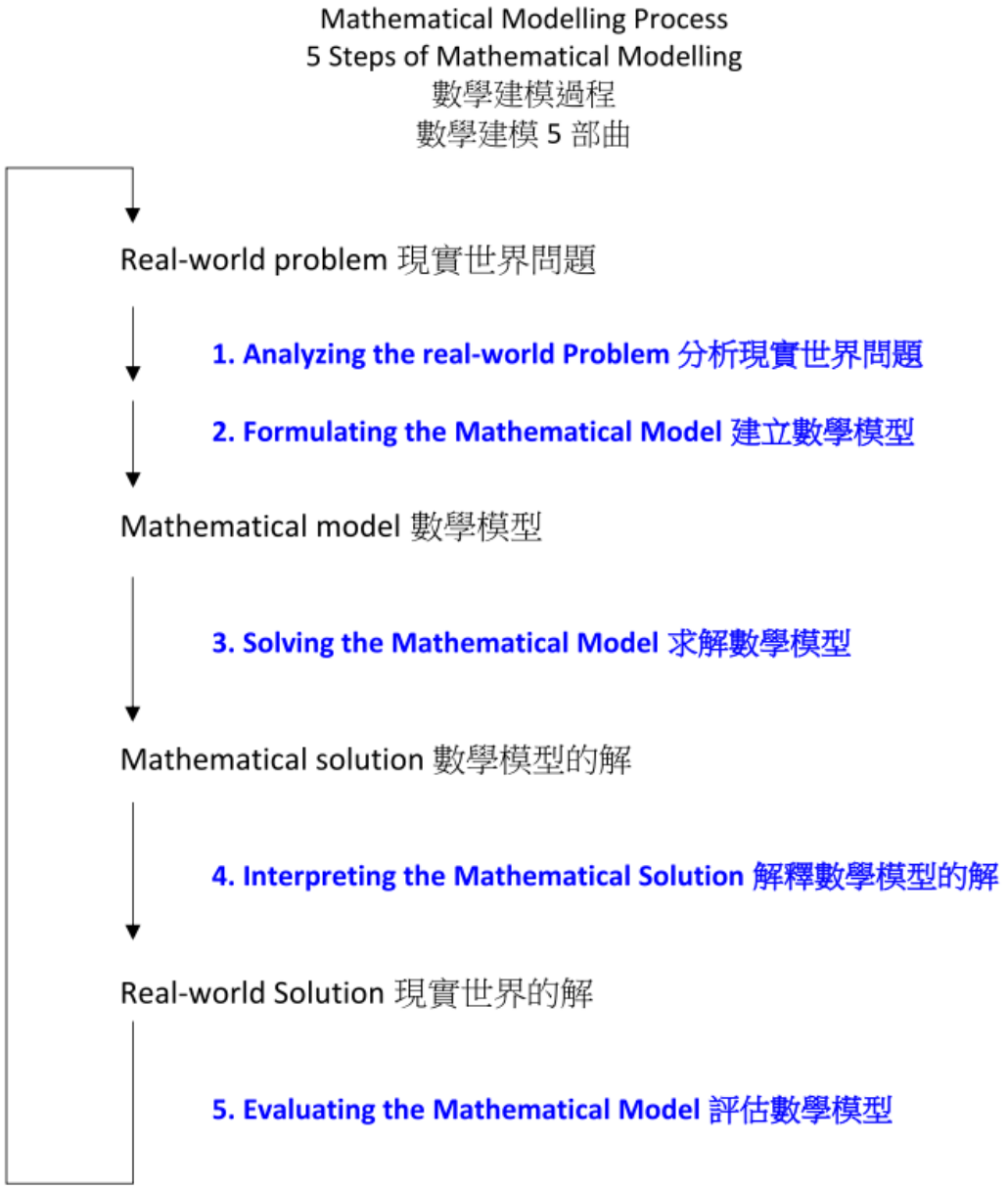
- **Creative but rigorous** 既需要創意亦需要嚴謹性
  - No "correct answer" 沒有「正確答案」
  - Require justification of the models developed需要理據支持建立的模型

Mathematical Modelling Process

數學建模過程

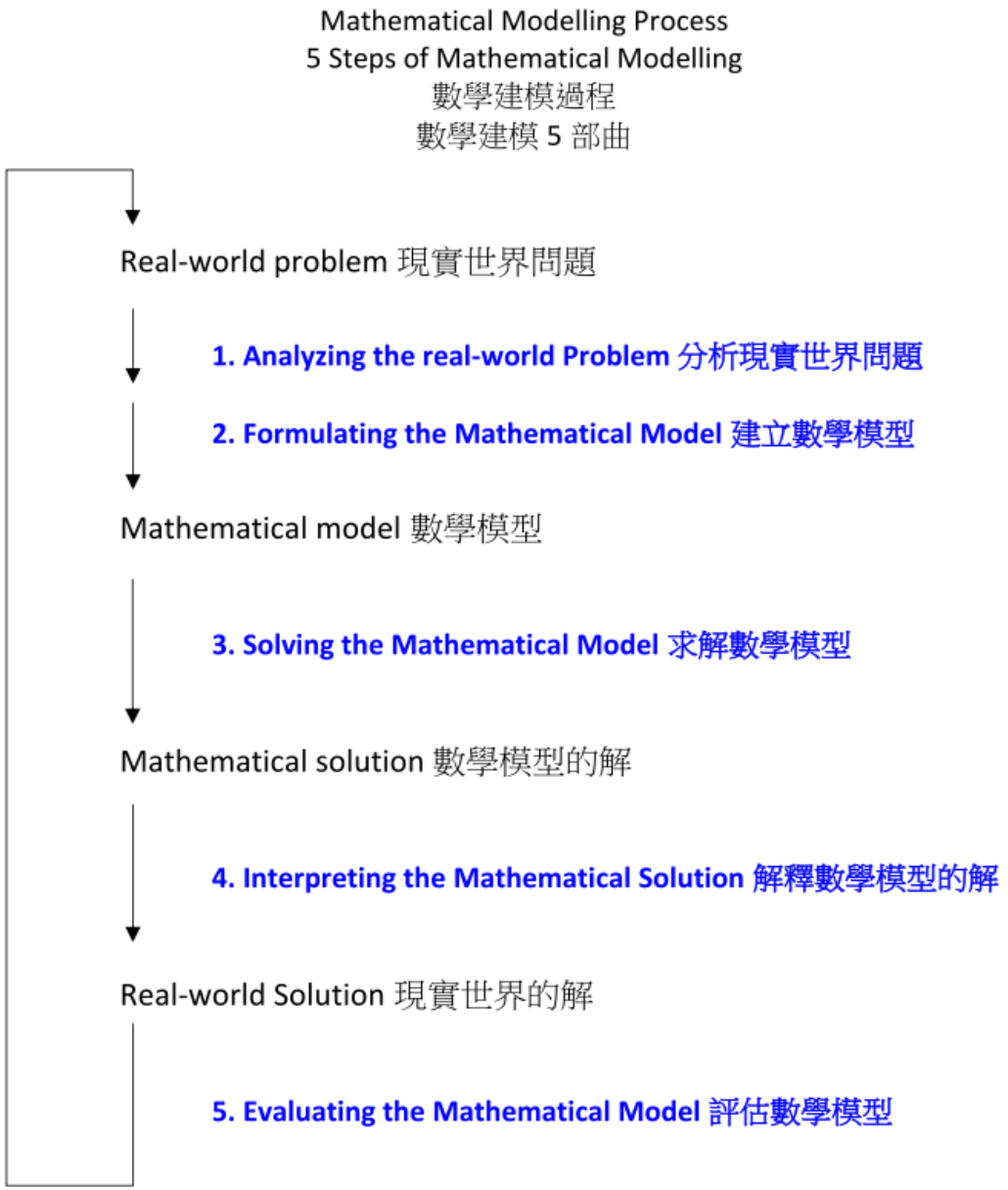
Mathematical Modelling Process 數學建模過程

- **1. Analyzing the real-world problem**
  分析現實世界問題
  - Problems arising in everyday life, society, and the workplace
    日常生活、社會和工作場所中出現的問題
  - Understand the problem background
    了解問題背景
  - Locate relevant information
    尋找相關資訊



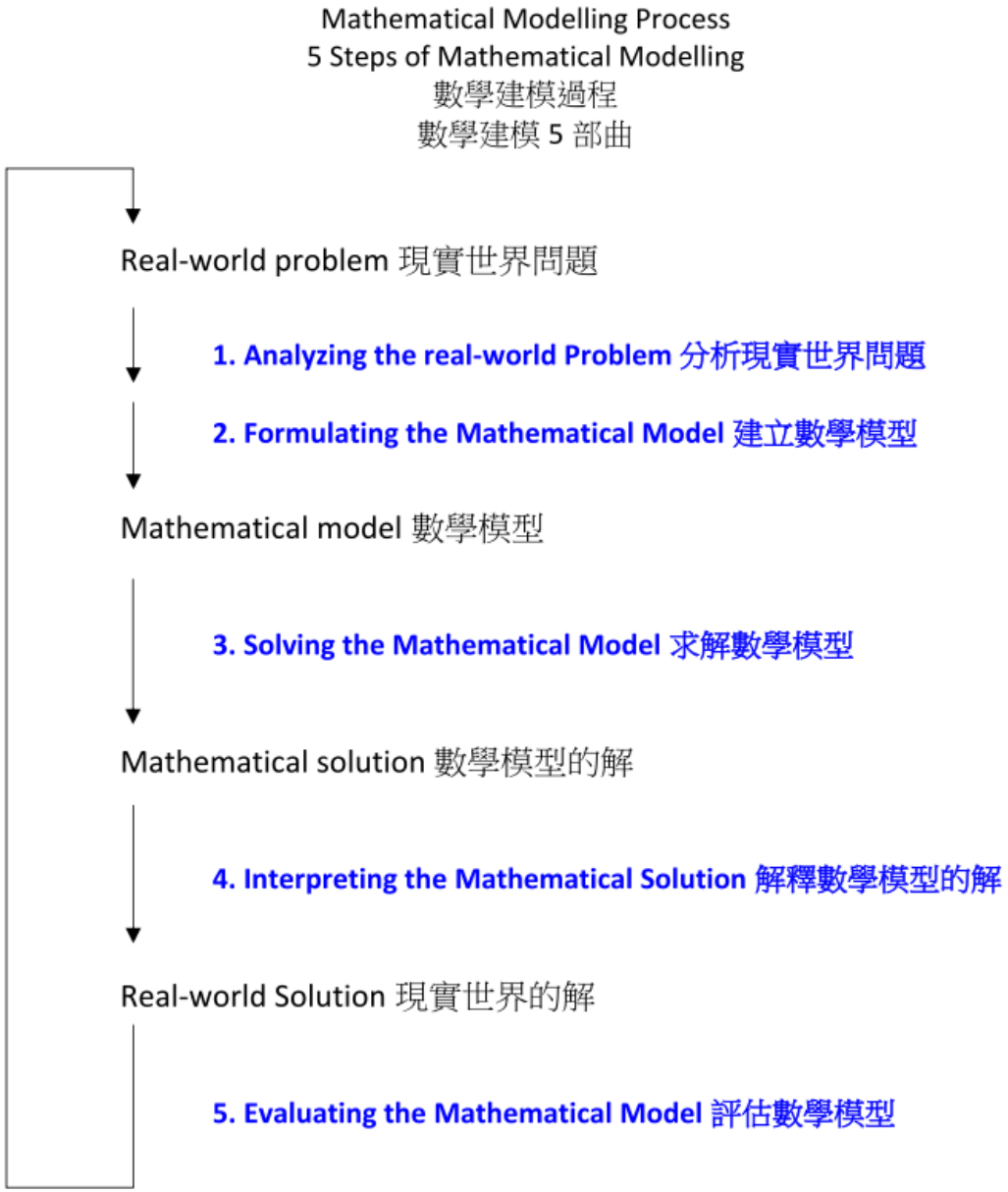
Mathematical Modelling Process 數學建模過程

- **2. Formulating the mathematical model**
  建立數學模型
  - Make suitable assumptions 作出適當假設
  - Identify important factors 找出重要因素
  - Collect the corresponding data 收集相應數據
  - Construct suitable models 建構合適的模型
  - "All models are wrong, but some are useful"
    「所有模型都是錯的，但有些是有用的」



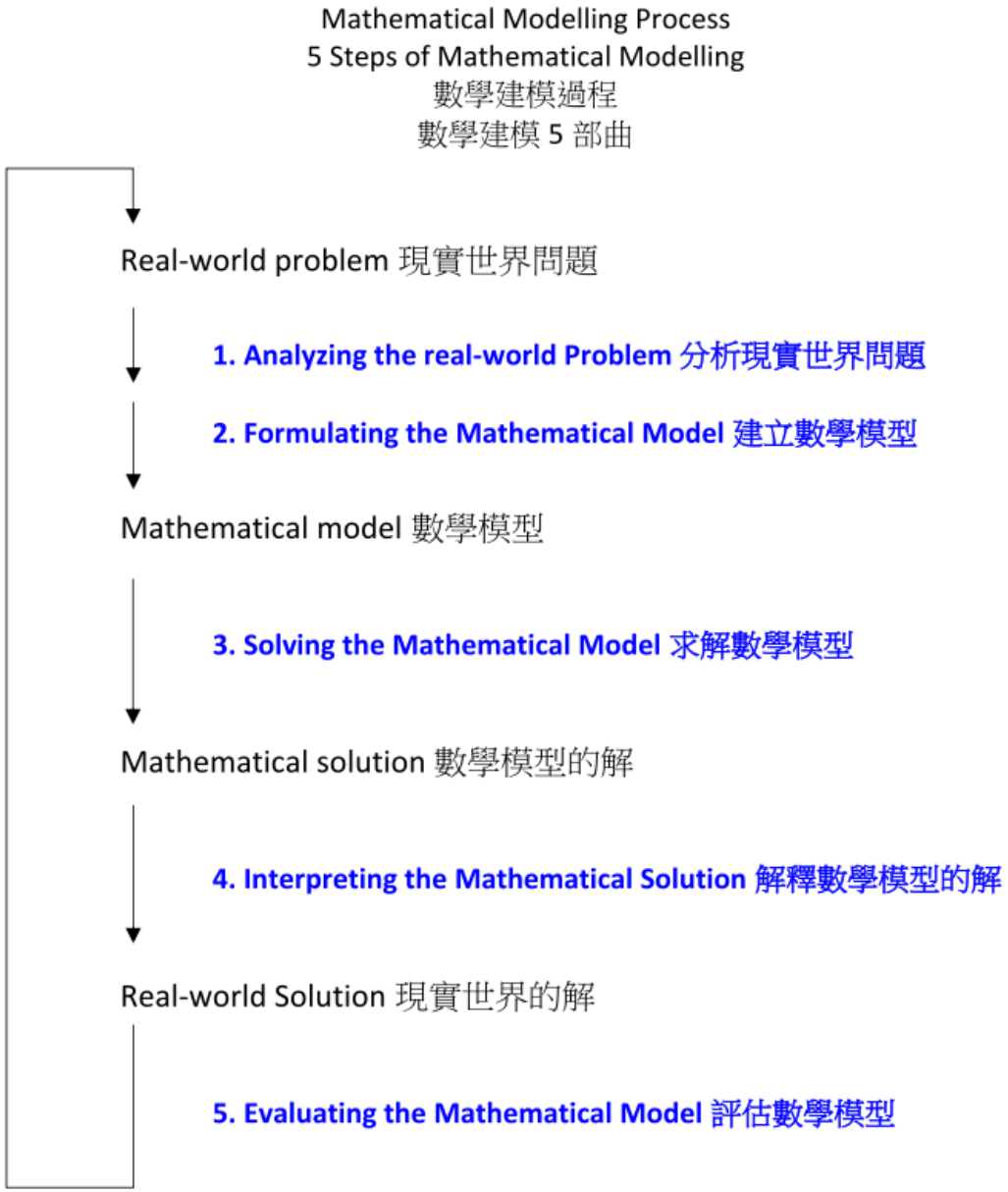
Mathematical Modelling Process 數學建模過程

- **3. Solving the mathematical model**
  求解數學模型
  - Utilize theoretical and/or computational tools to solve the model
    運用理論和/或計算工具求解模型
  - Mathematical derivation 數學推導
  - Using IT tools for exactly getting or approximating the solution
    利用 IT 工具精確計算或估算模型的解



Mathematical Modelling Process 數學建模過程

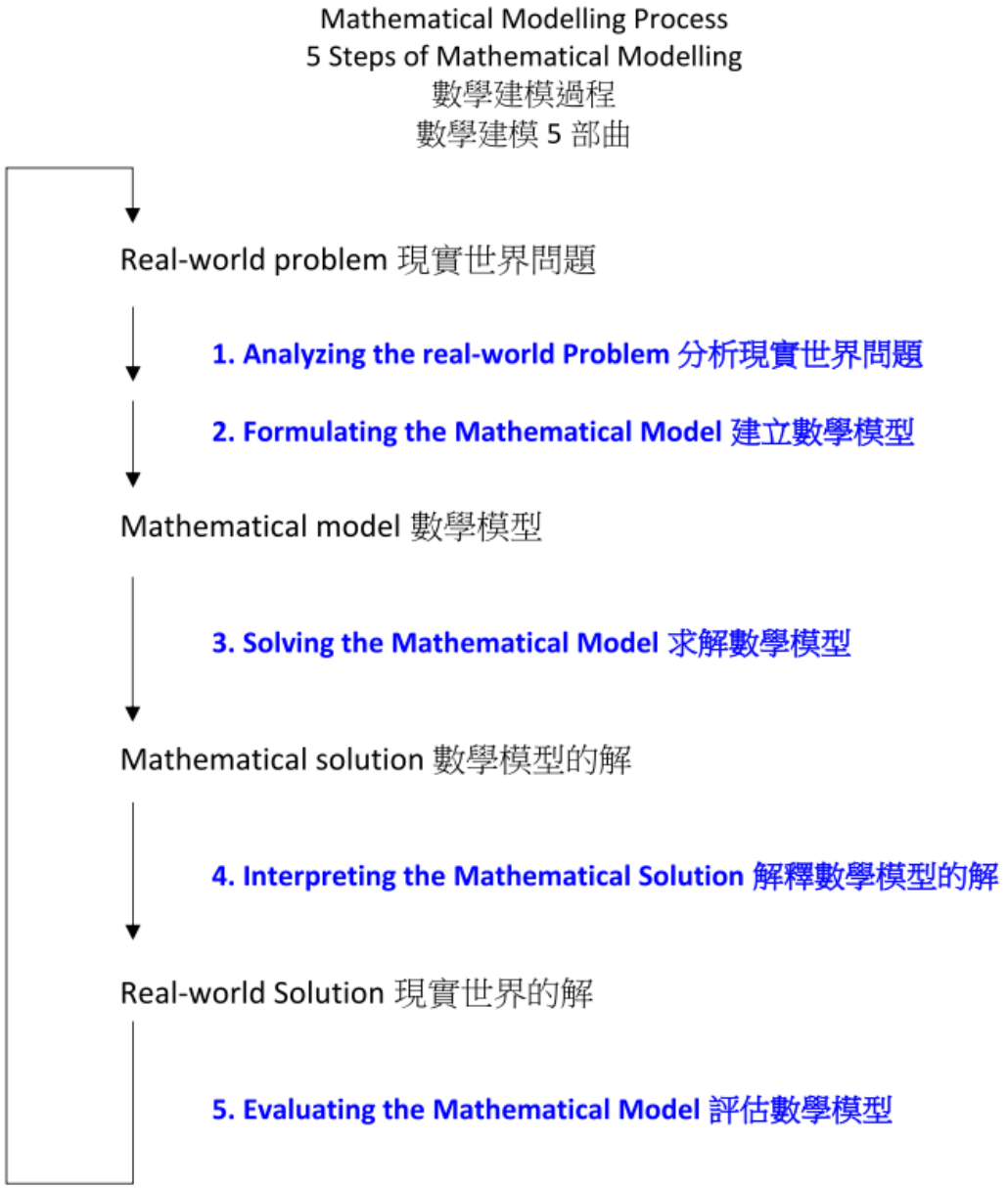
- **4. Interpreting the mathematical solution: 解釋數學模型的解**
  - Converting numerical results into practical terms, e.g., explaining what a calculated value represents in the real-world problem 將數值結果轉換為實際意思，例如解釋計算值在現實問題中的意義
    - Best-fit parameters in the solution 最佳擬合參數
    - Value at a certain time point 特定時間點的值
    - Graphical visualization 圖形視覺化
  - Focusing on the implication of the solution to address the real-world problem 著重闡述模型解對實際問題的意義
  - Emphasizing how the model findings align with or illuminate the aspects of the real-world problem, without yet judging the model's overall quality 強調模型結果如何與現實問題的各個方面相契合或闡明實際問題，而未需評估模型的整體品質



Mathematical Modelling Process 數學建模過程

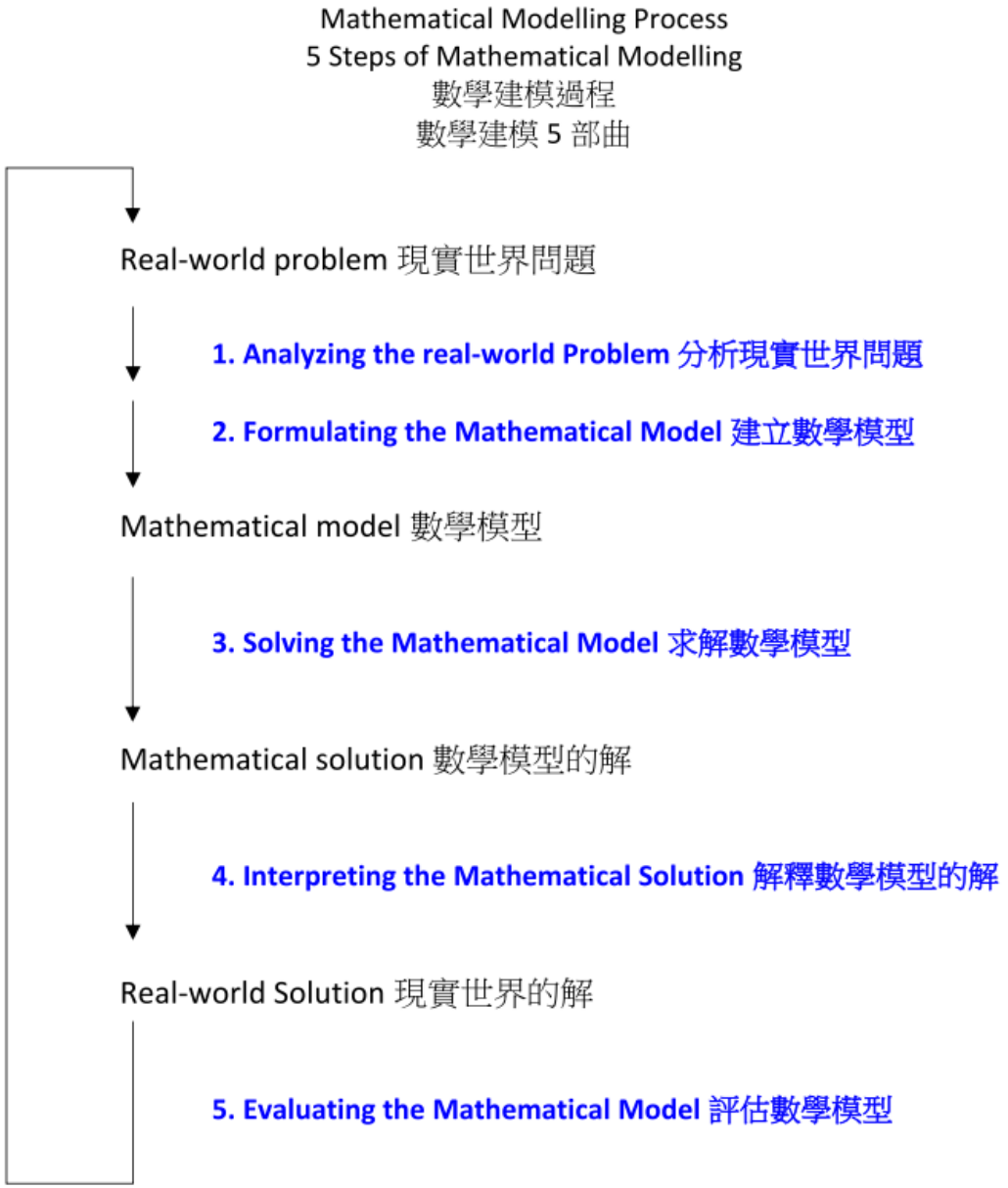
- **5. Evaluating the mathematical model**
  評估數學模型

  - Assess the accuracy to judge whether the model adequately represents the real-world problem 評估模型的準確性，以判斷模型是否能充分代表實際問題
  - Check whether overfitting/underfitting occurs 檢查是否有過擬合/欠擬合現象
  - Test the model against other data
    使用其他數據測試模型



Mathematical Modelling Process 數學建模過程

- Mathematical modelling is an iterative process!
  數學建模是一個循環過程！
  - Identify room for improvement
    找出改進空間
  - Refine the model as needed
    根據需要完善模型
  - Repeat (1) – (5)
    重複步驟 (1) – (5)

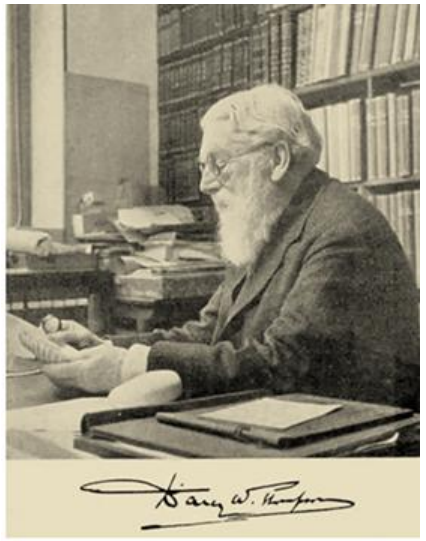


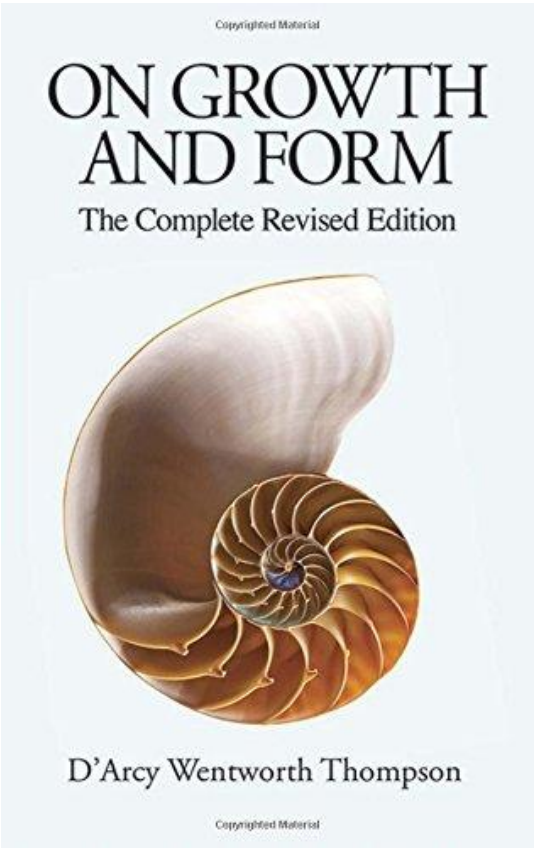
Example: Modelling the Growth and Form in Biology
生物學中形狀生長的建模

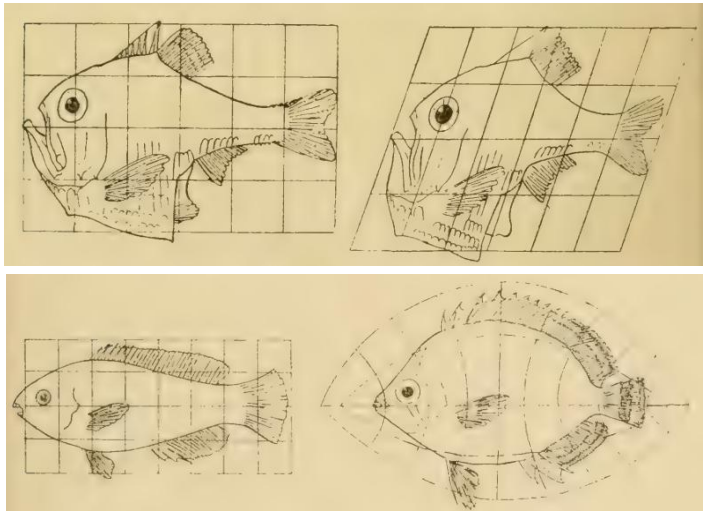
- **1. Analyzing the real-world problem:**
  分析現實世界問題

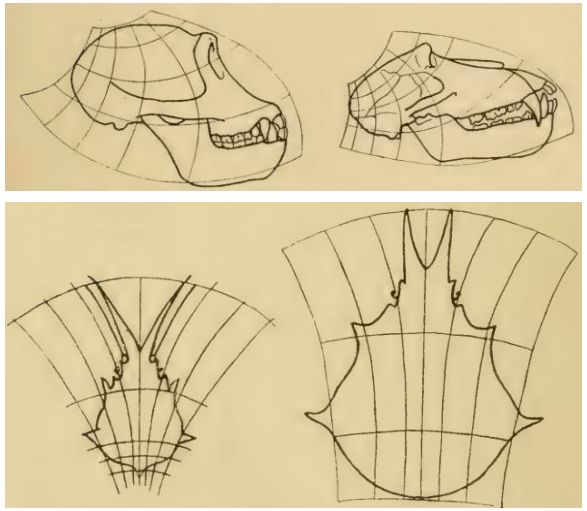
  - Studying the growth of biological shapes in the world
    研究生物形態在自然界中的生長

  - How to represent the change in shapes?
    如何表示形態的變化？







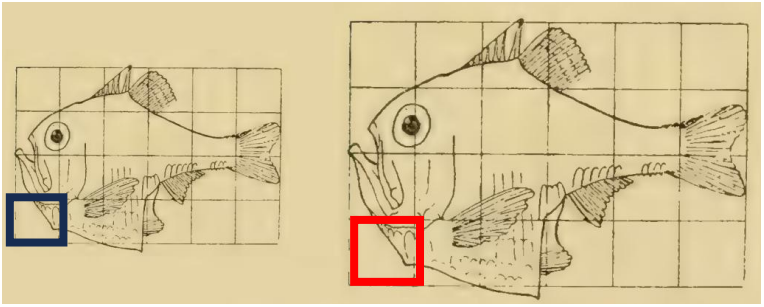


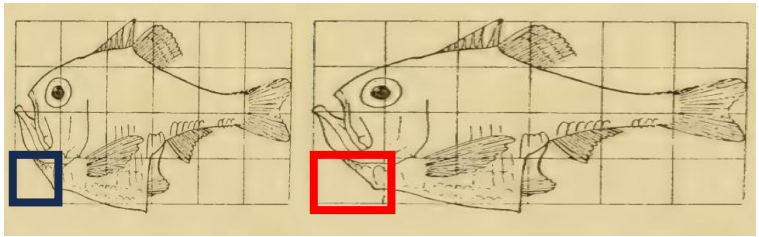
D'Arcy Thompson, *On Growth and Form* (1917)

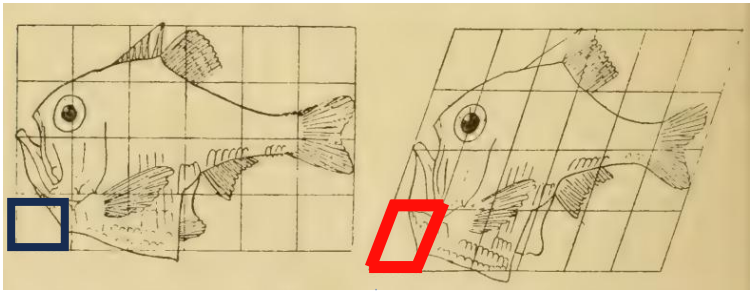
Example: Modelling the Growth and Form in Biology
生物學中形狀生長的建模

- **2. Formulating the mathematical model:**
  建立數學模型

  - Assume 2D shapes 假設為二維圖形
  - Collect data from images or scans
    從影像或掃描收集數據
  - Constructing a model 建立模型 :
    - Uniform scaling 均勻縮放:
      $$(x, y) \to (mx, my)$$
    - Non-uniform scaling 非均勻縮放:
      $$(x, y) \to (mx, ny)$$
    - Scaling and tilting 縮放及傾斜:
      $$(x, y) \to (ax + by, cx + dy)$$

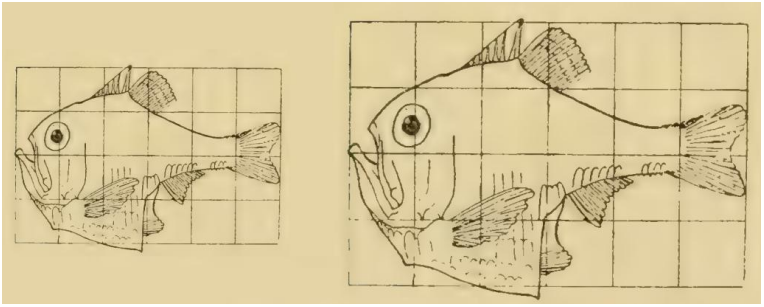




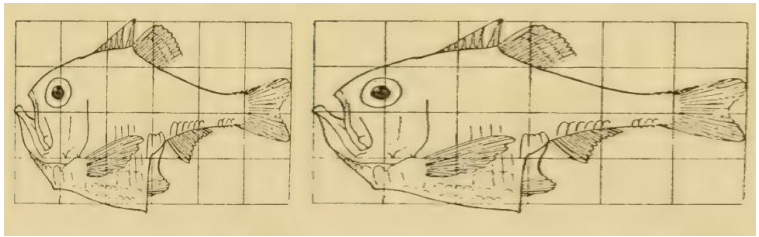


Example: Modelling the Growth and Form in Biology
生物學中形狀生長的建模

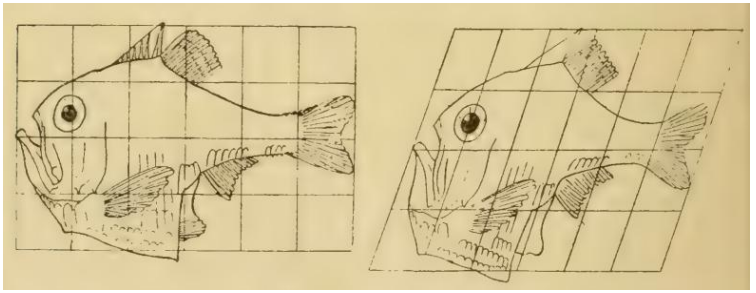
- **3. Solving the mathematical model:**
  求解數學模型



- Utilize IT tools to extract coordinates from the image data
  利用 IT 工具從影像資料中擷取座標

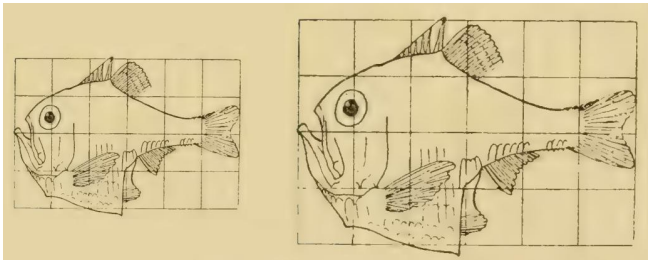


- Solve for the best fit parameters for the chosen model
  求解所選模型的最佳擬合參數

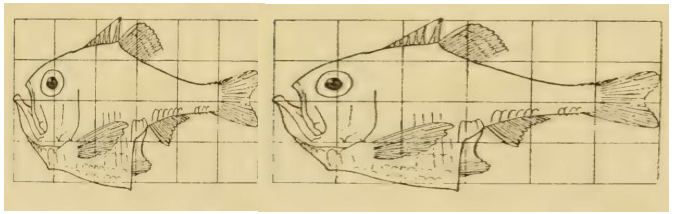


Example: Modelling the Growth and Form in Biology
生物學中形狀生長的建模

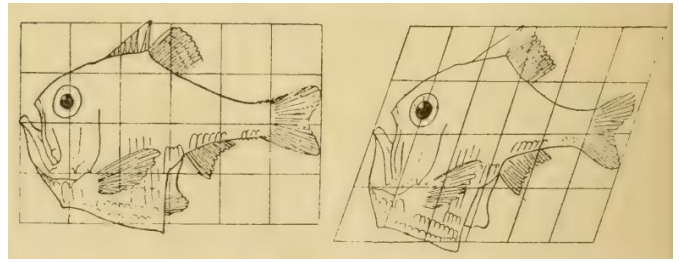
- **4. Interpreting the mathematical solution:**
  解釋數學模型的解



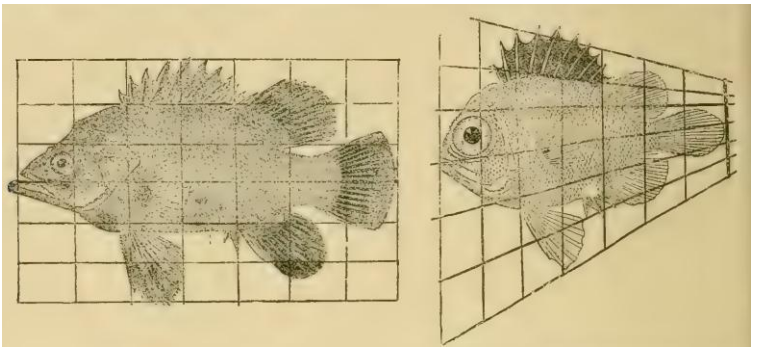
- Are some of the fitted values much larger than the others?
  某些擬合值是否遠大於其他值？
  - If so, it implies that the species grow more rapidly in one direction than some others
    如果是，則表示物種在某些方向上的生長速度比其他方向更快。



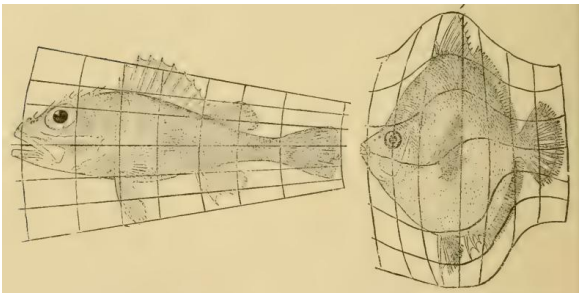
- Do we have a generally good fit for all species and all growth periods?
  我們是否能對所有物種和所有生長階段都得到良好的擬合？
  - If so, it means that they all grow constantly over time
    如果是，則表示它們隨時間推移持續生長。
  - If not, it means that the growth rate can be different for different species and/or different developmental stages
    如果不是，則表示不同物種和/或不同階段的生長速度可能不同。



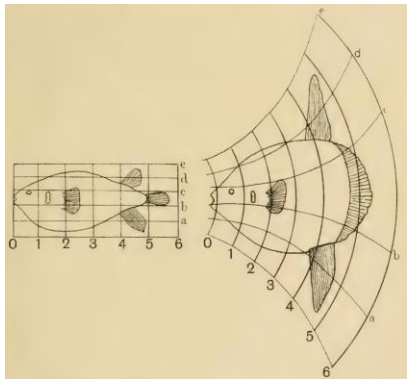
Example: Modelling the Growth and Form in Biology
生物學中形狀生長的建模



- **5. Evaluating the mathematical model 評估數學模型:**
  - Calculate the average/maximum/minimum error to assess the model accuracy
    計算平均誤差/最大誤差/最小誤差以評估模型精度



  - Test the model with more other species 使用更多物種測試模型
  - Refine the model 改進模型:
    - Use a nonlinear (curve-like) growth model instead of a linear model?
      使用非線性（曲線式）生長模型代替線性模型？
    - For different classes of biological shapes, we may need different models?
      對於不同類型的生物形態，我們可能需要不同的模型？
    - For different regions or growth periods, we may need different models?
      對於不同的部份或生長階段，我們可能需要不同的模型？



How to apply mathematical modelling in real life?

如何在現實生活中應用數學建模？

Mathematical Problems in Real Life 現實生活中的數學問題

- **Consideration in buying products**
  購買貨品時的考慮
  - Price of the products 貨品價格
  - Quality of the products 貨品質素
    - Capacity 容量
    - Durability 耐用性
    - Speed 速度
    - …

Mathematical Problems in Real Life 現實生活中的數學問題

- **Estimating future population growth**
  估算未來人口增長
  - Year 年份
  - Population 人口數字

World Population by Year

Year	World Population	Yearly Change
2024	**8,161,972,572**	0.87 %
2023	**8,091,734,930**	0.88 %
2022	**8,021,407,192**	0.84 %
2021	**7,954,448,391**	0.86 %
2020	**7,887,001,292**	0.97 %
2019	**7,811,293,698**	1.05 %
2018	**7,729,902,781**	1.10 %
2017	**7,645,617,954**	1.15 %
2016	**7,558,554,526**	1.18 %
2015	**7,470,491,872**	1.20 %
2014	**7,381,616,244**	1.23 %
2013	**7,291,793,585**	1.26 %
2012	**7,201,202,485**	1.27 %

Source: https://www.worldometers.info/world-population/

Mathematical Problems in Real Life 現實生活中的數學問題

- In many problems, we will obtain **data points** and analyze them
  在許多問題中，我們會取得**數據點**並作出分析

(Price 價錢, Capacity 容量)

(390, 512)
(800, 1024)
(1450, 2048)
⋮
(2000, 4096)

(Year, population in billion)
(年份, 人口(十億))

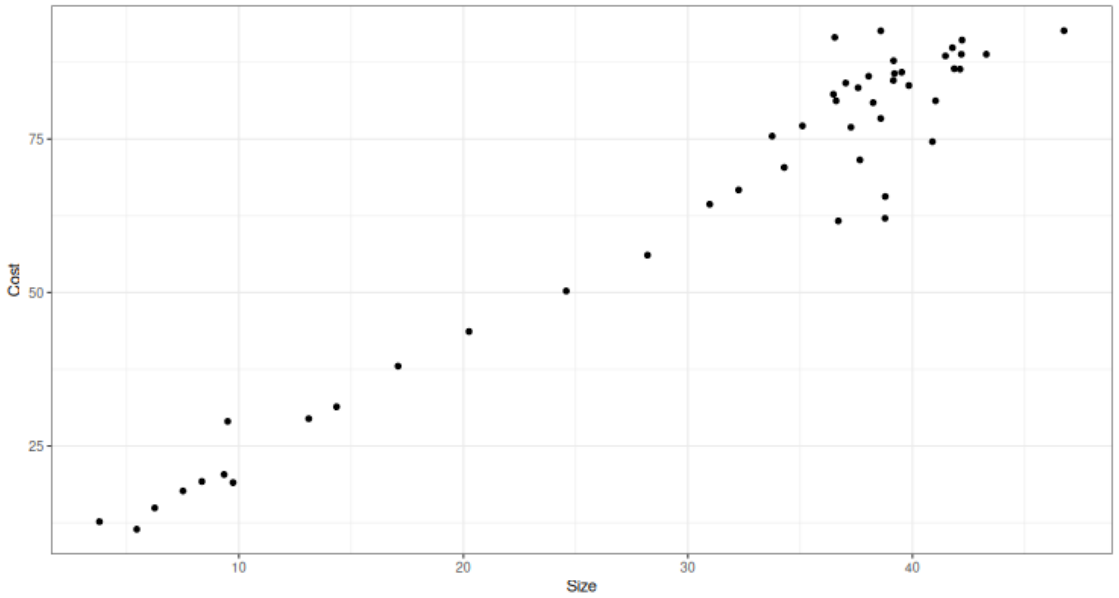
(2012, 7.20)
(2013, 7.29)
(2014, 7.38)
⋮
(2023, 8.09)
(2024, 8.16)

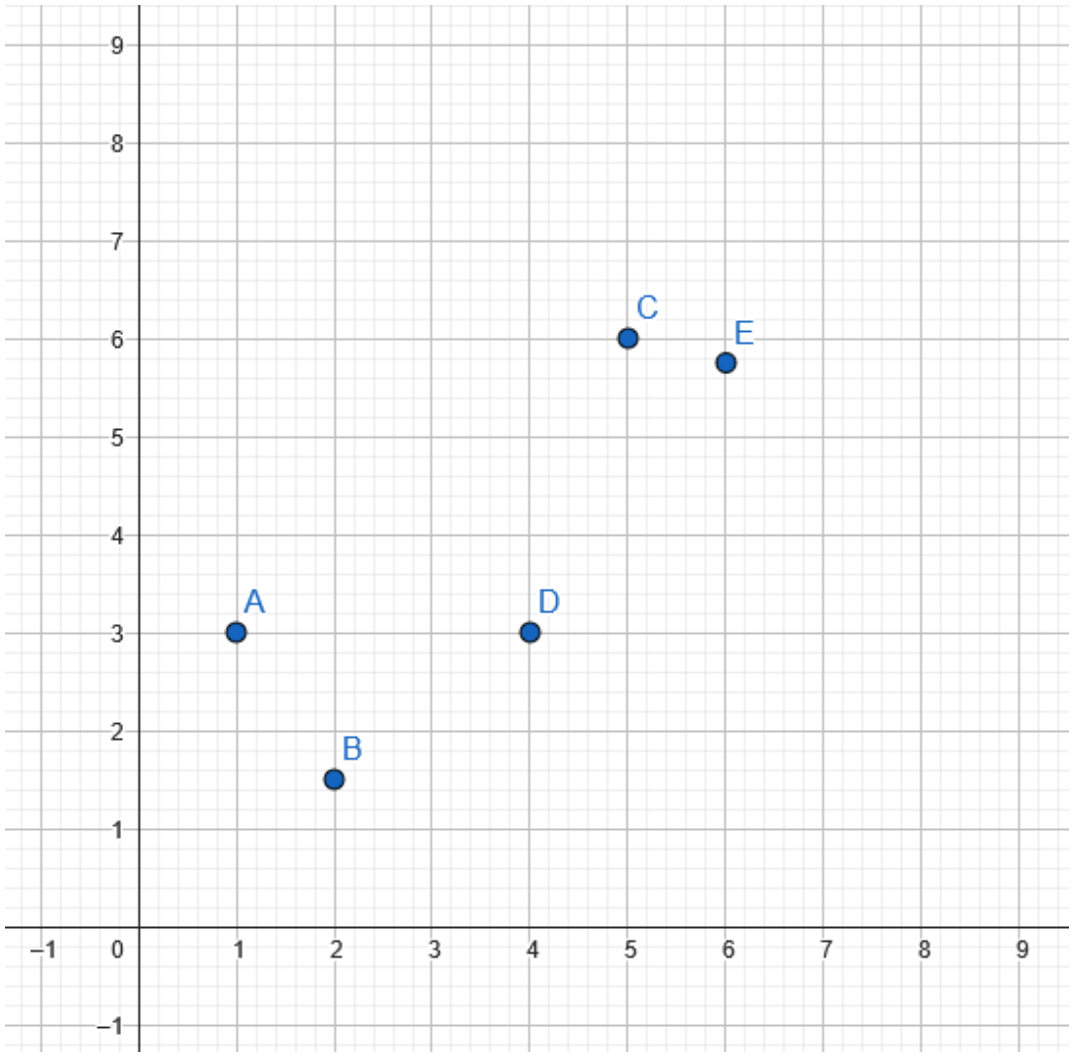
Mathematical Problems in Real Life 現實生活中的數學問題

- **Data points 數據點**

$A = (1, 3),\ \ B = (2, 1.5),\ C = (5, 6),$
$D = (4, 3),\ E = (6, 5.75),\ \ldots$

- What if we have more data points?
  如果我們有更多數據點，怎麼辦？

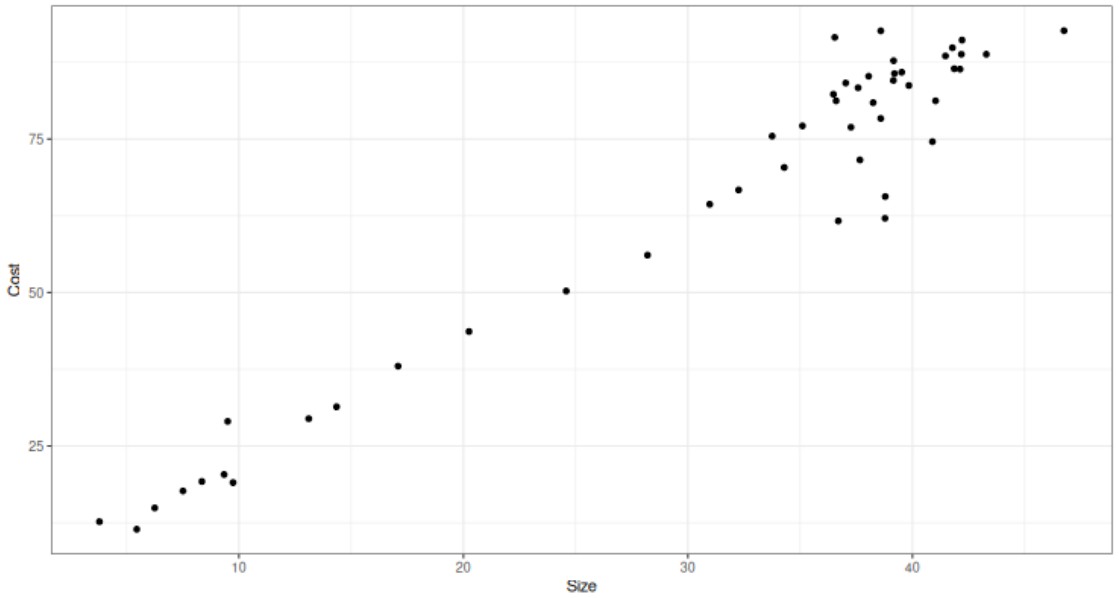


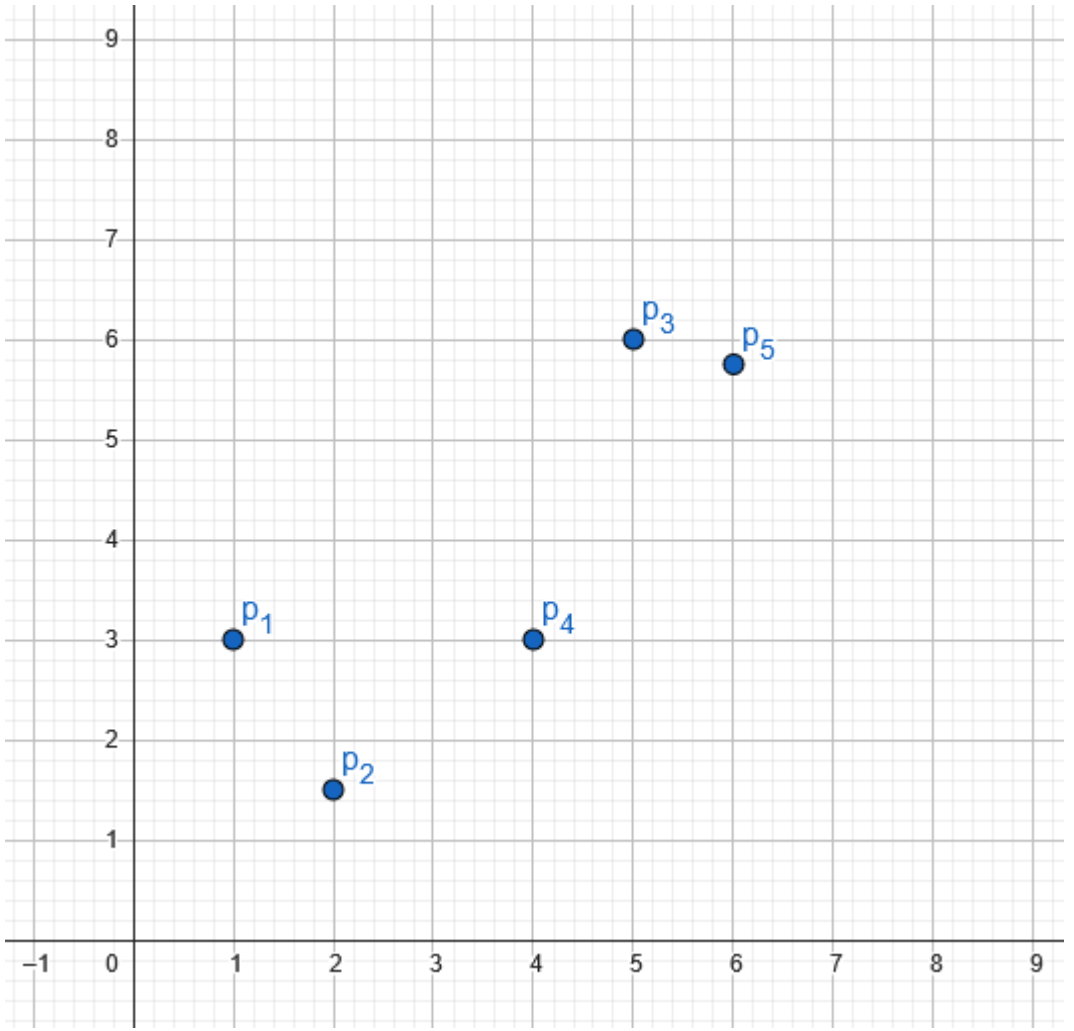


Mathematical Problems in Real Life 現實生活中的數學問題

- In this case, it is common to use subscripts to represent different data points:
  在這情況下，我們通常會用下標表示不同數據點：

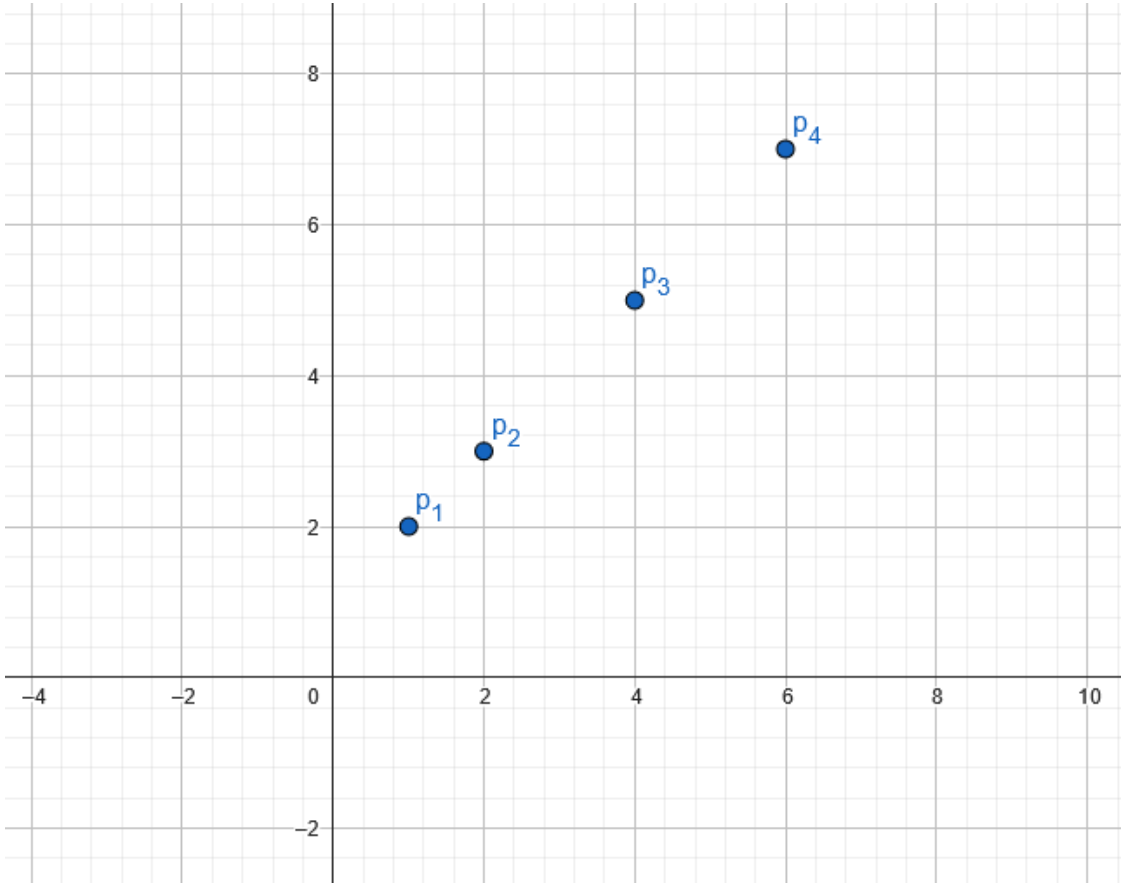
$$p_1 = (x_1, y_1)$$
$$p_2 = (x_2, y_2)$$
$$p_3 = (x_3, y_3)$$
$$\vdots$$
$$p_n = (x_n, y_n)$$





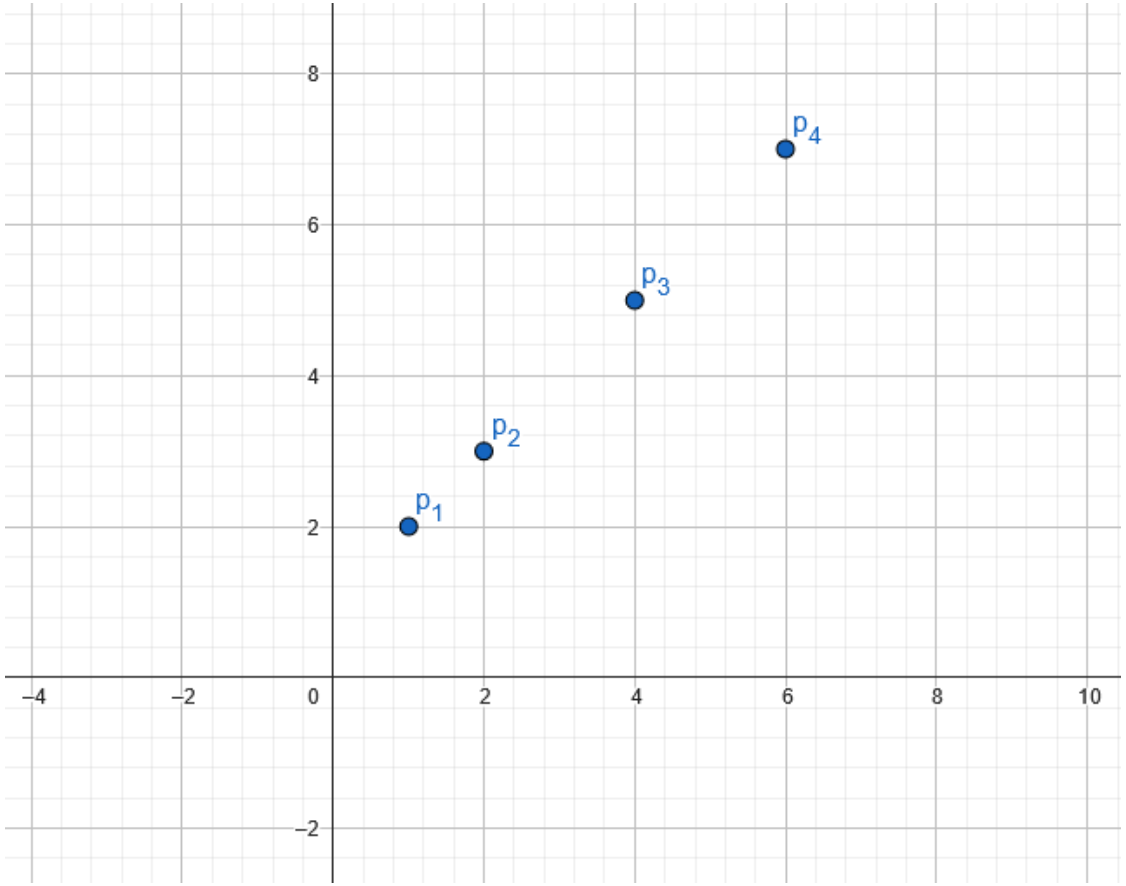
How to Find Trends from Data? 如何從數據中找出規律？

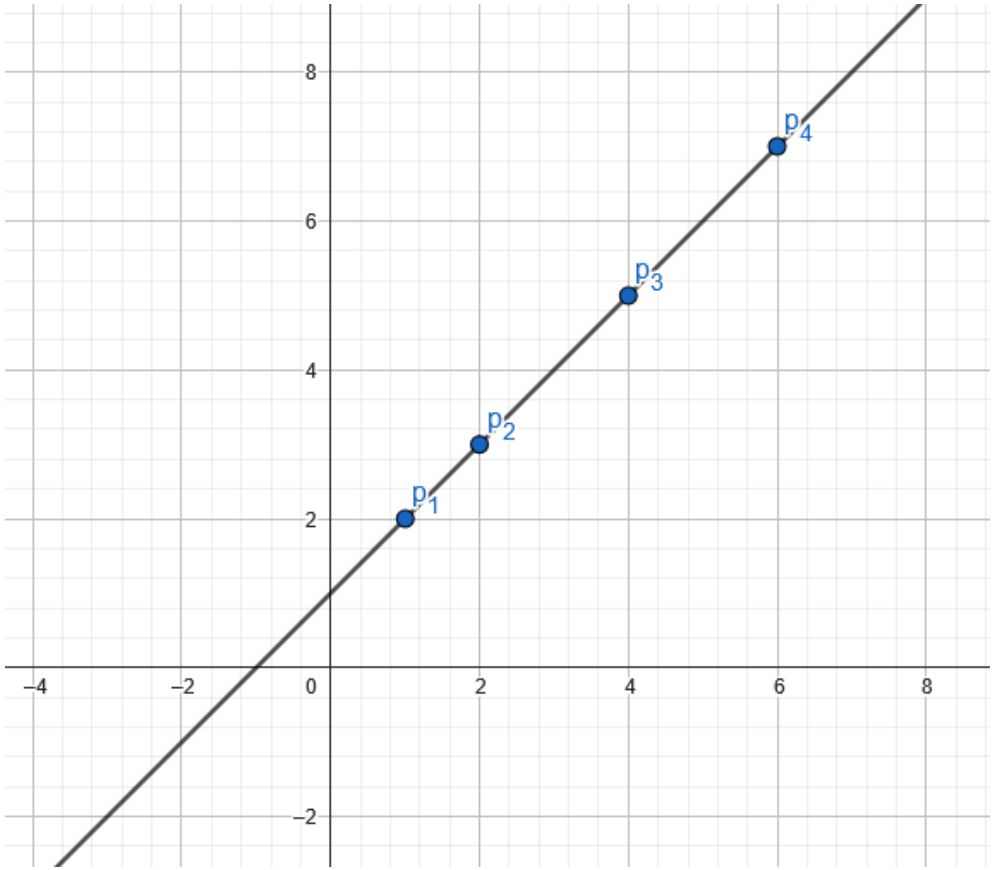
- In some cases, we can easily see the trend in a dataset and even draw a straight line passing through all points
  在某些情況下，我們可以容易從數據點中找出規律，甚至用一條直線穿過所有點



How to Find Trends from Data? 如何從數據中找出規律？

- In some cases, we can easily see the trend in a dataset and even draw a straight line passing through all points
  在某些情況下，我們可以容易從數據點中找出規律，甚至用一條直線穿過所有點



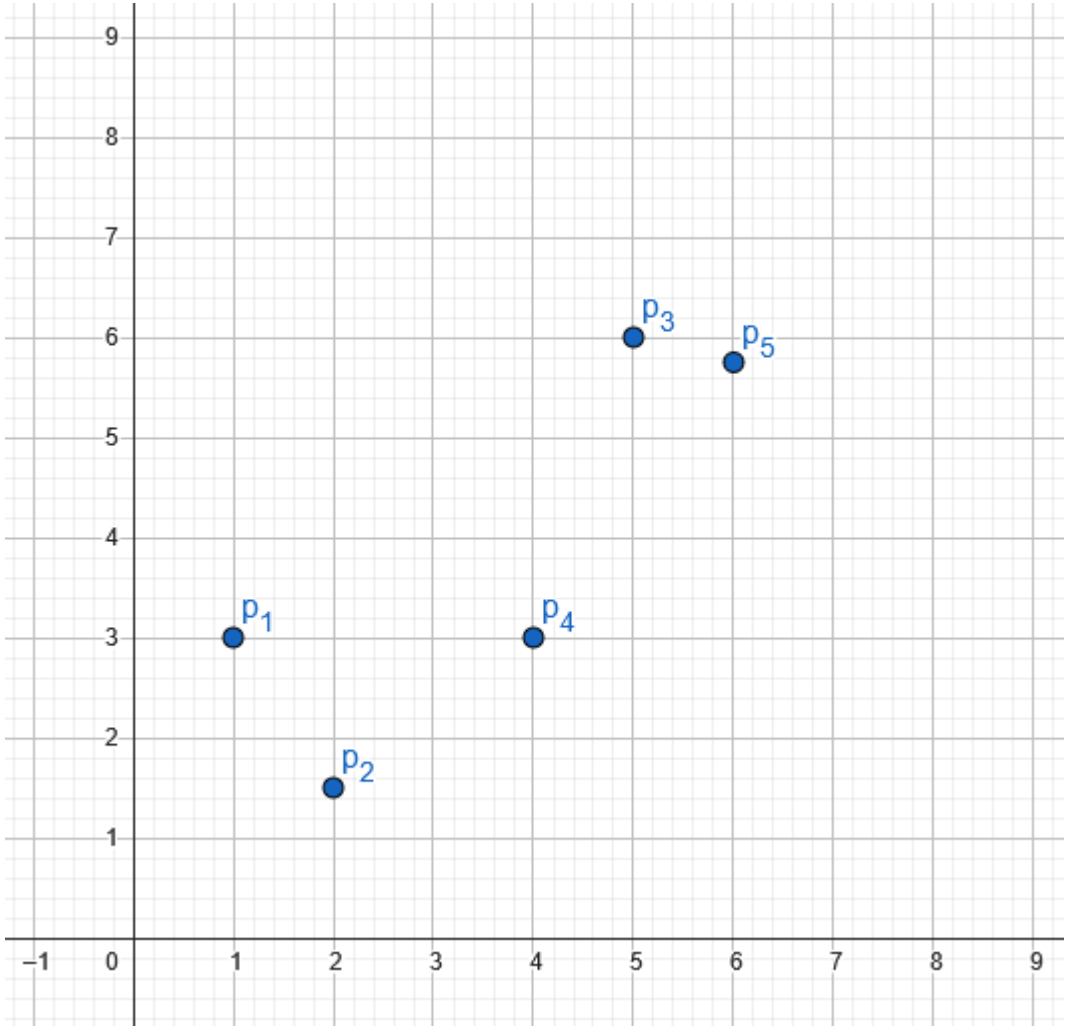


How to Find Trends from Data? 如何從數據中找出規律？

- **What if the data points do not form a straight line?**
  但是，如果數據點並不形成一條直線，怎麼辦？

- In this case, can we find a “best-fit” straight line?
  在某些情況下，我們能找到一條「最佳擬合」直線嗎？

- How to define “best-fit”?
  如何定義「最佳擬合」？



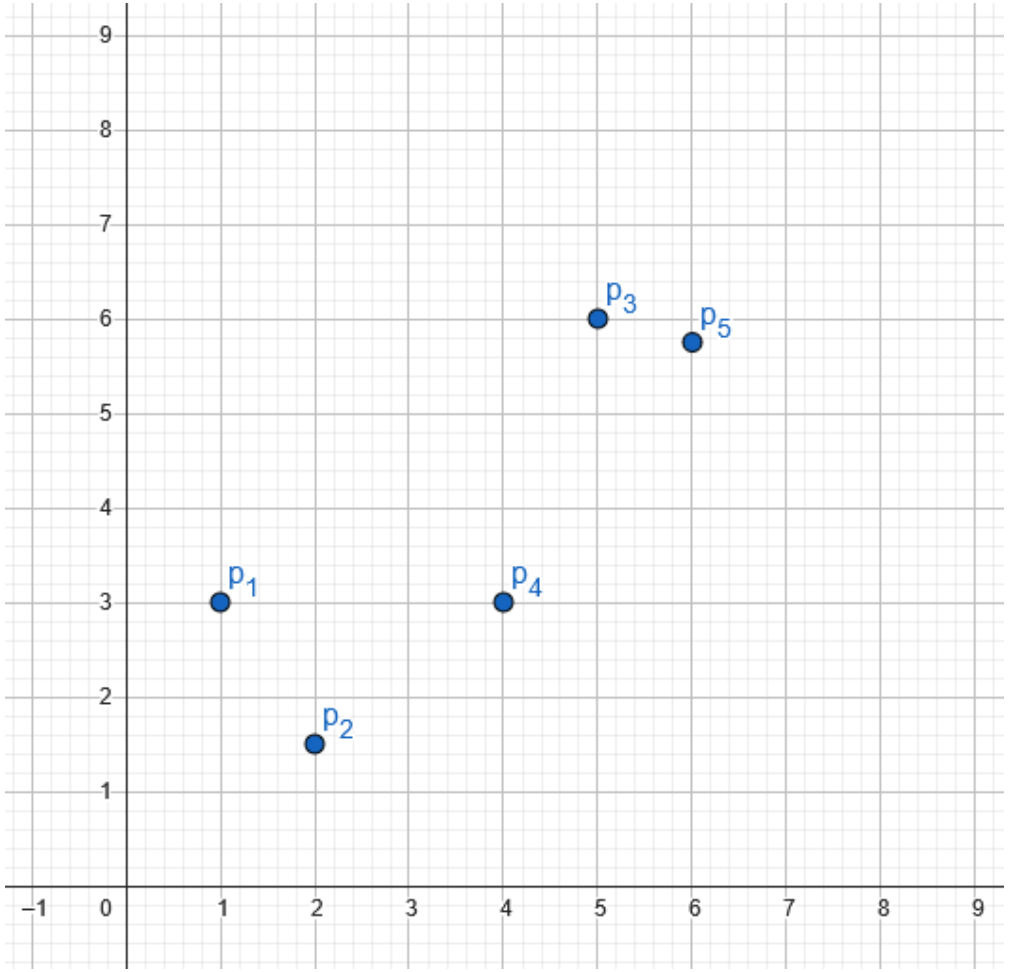
How to Find Trends from Data? 如何從數據中找出規律？

- **Exercise 練習:**
  - Use our Find What Fits R Shiny tool
    利用我們的找出擬合線 R Shiny工具
    https://www.math.cuhk.edu.hk/app/mathmodel/tool.html

  - Create a data file using the data points shown on RHS
    使用右邊的數據點建立文檔

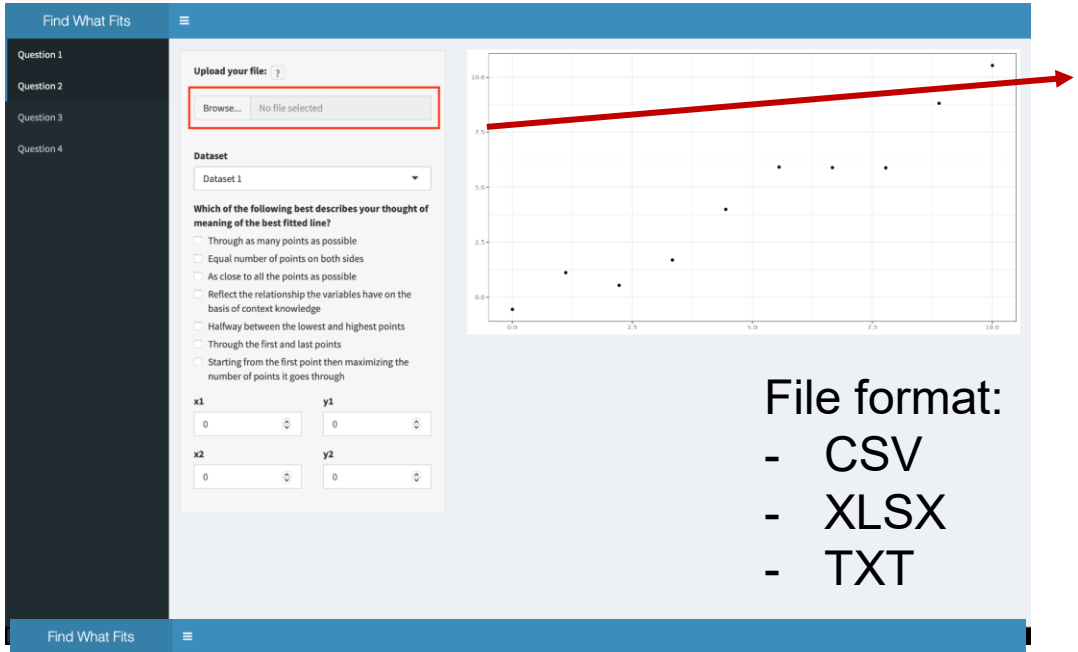
  - Consider different criteria to define a “best-fit” straight line
    考慮不同的標準來定義「最佳擬合」直線

x	y
1	3
2	1.5
5	6
4	3
6	5.75



Finding What Fits with R Shiny 利用R Shiny工具找出最佳直線

- Step 1: Upload the dataset
  步驟 1：上傳資料集

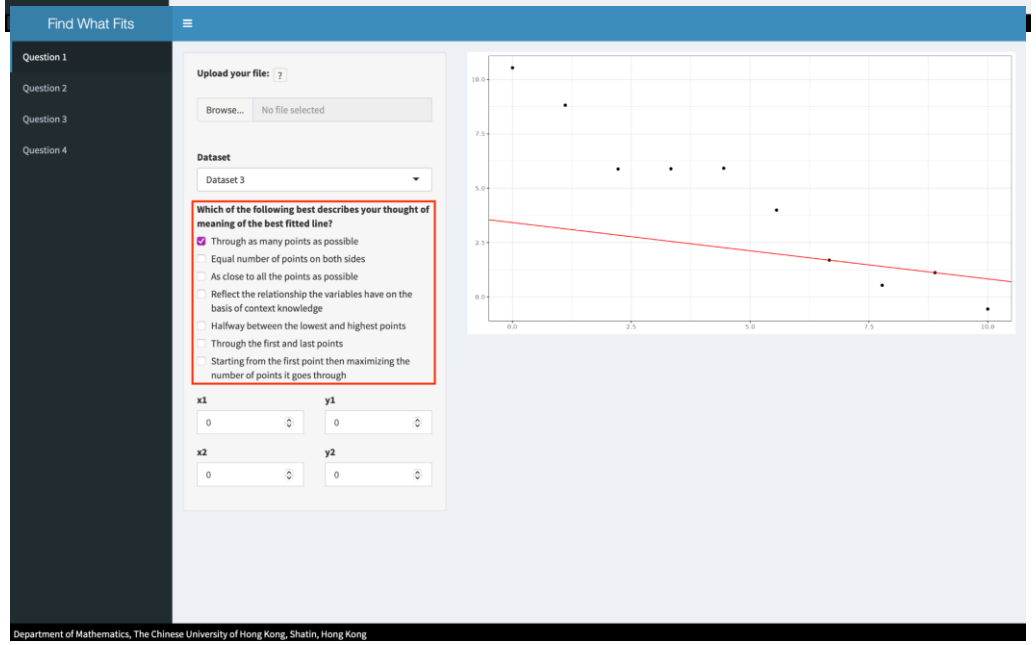


File format:
- CSV
- XLSX
- TXT

For the CSV/XLSX file, there should be 2 columns of data.

x	y
1.50	40.60
3.80	65.00
8.90	70.80
10.00	78.00
12.00	84.00
20.00	108.50

- Step 2: Explore different definitions of “best-fit” line
  步驟 2：探索「最佳擬合」線的不同定義



For the TXT file, the entries should be comma-separated.

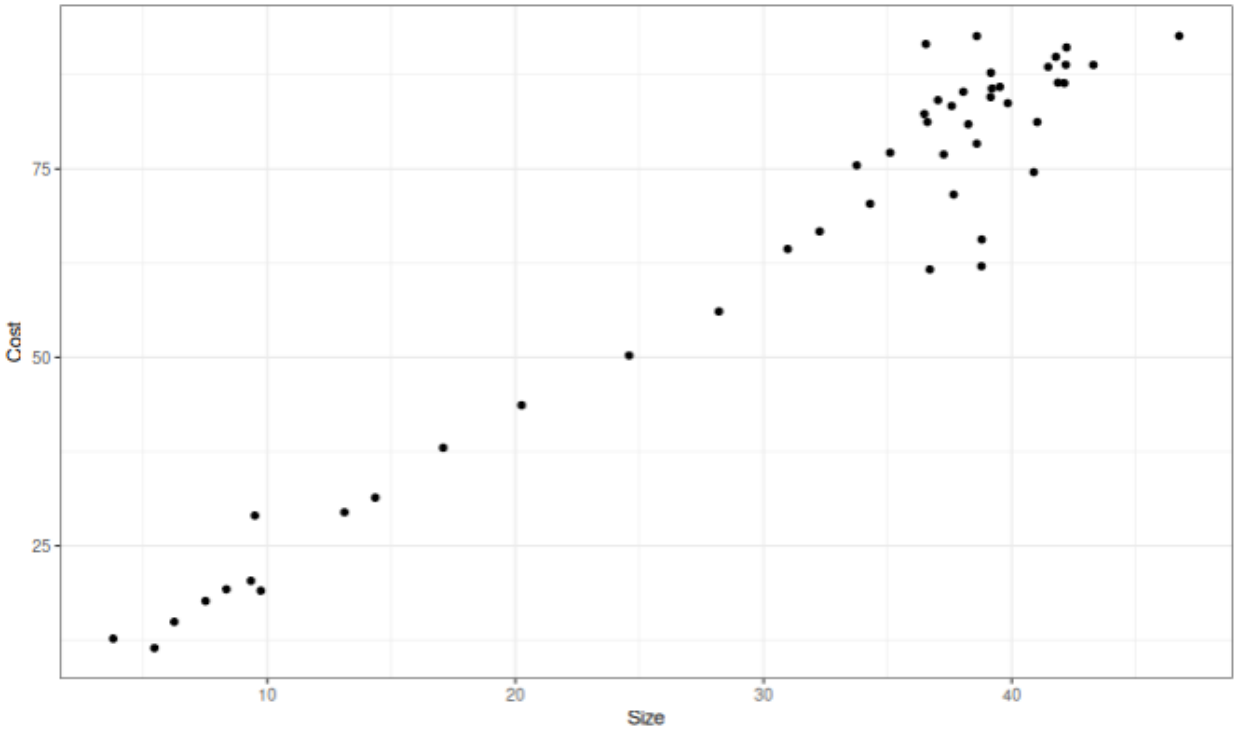
```
x,y
1.5,40.6
3.8,65
8.9,70.8
10,78
12,84
20,108.5
```

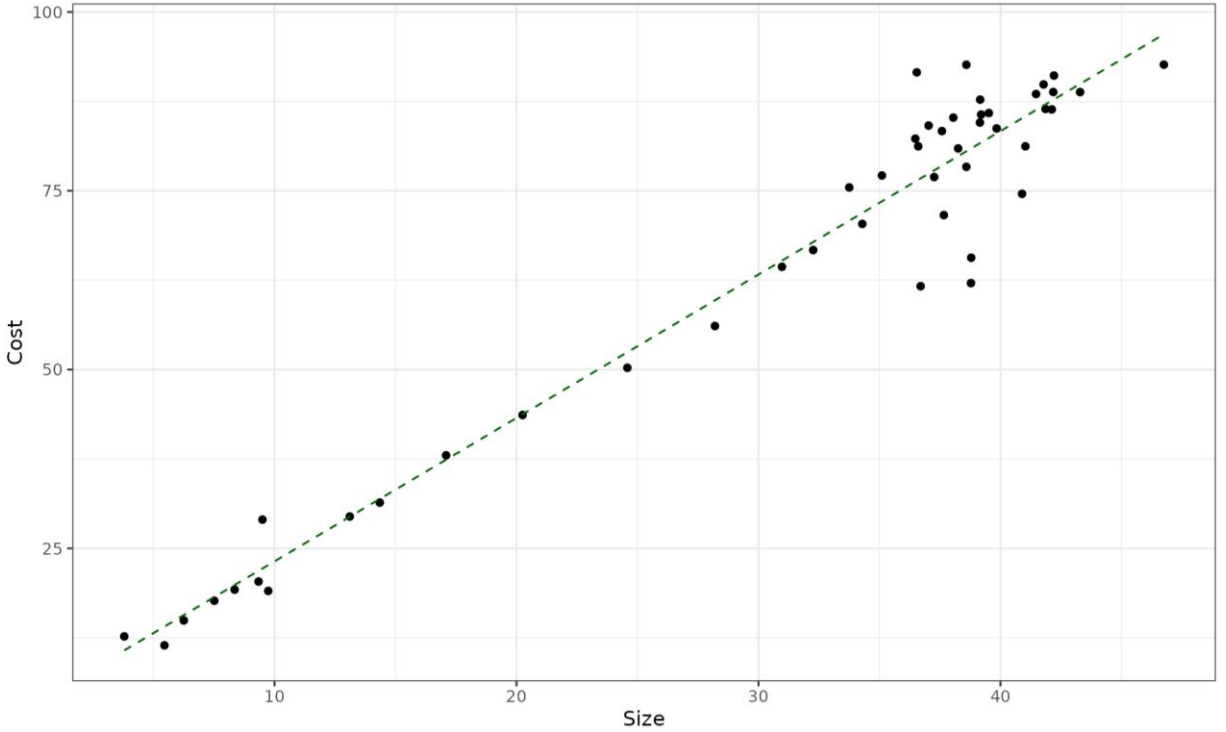
Mathematical Modelling Methods and Tools

數學建模方法和工具

Linear Regression 線性迴歸

- As we can see, there can be many definitions of “best-fit”!
  我們可以看到，「最佳擬合」可以有很多種定義！
- For consistency, mathematicians define the “best-fit” straight line as the line that minimizes the residual sum of squares
  為了保持一致性，數學家將「最佳擬合」直線定義為最小化殘差平方和的直線

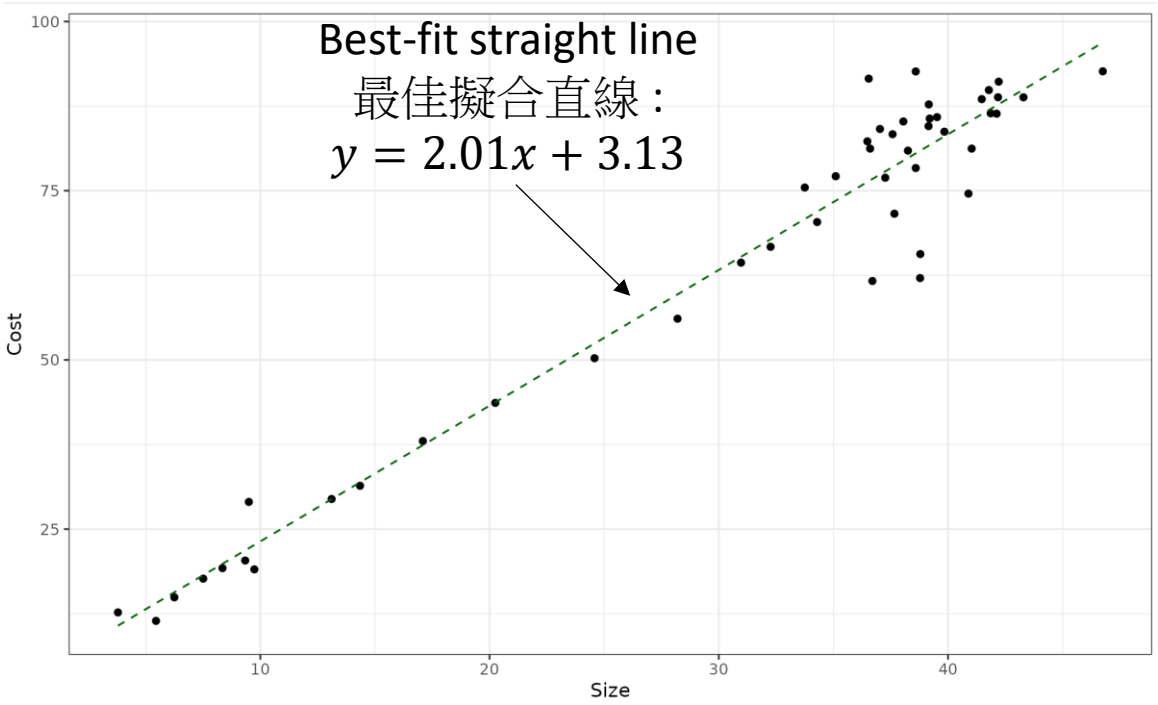




Linear Regression 線性迴歸

- **The "best-fit" straight line is defined as the line $y = ax + b$ that minimizes the residual sum of squares *RSS*, where**
  最佳擬合直線定義為最小化殘差平方和 *RSS* 的直線 $y = ax + b$ ，其中

$$RSS = \big(y_1 - (ax_1 + b)\big)^2 + \big(y_2 - (ax_2 + b)\big)^2 + \big(y_3 - (ax_3 + b)\big)^2 + \cdots + \big(y_n - (ax_n + b)\big)^2$$



- $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ are the given data points 是給定的數據點

Background Knowledge: Summation Notation 求和符號

- **Summation 求和符號**

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

Example:

If $x_1 = 2, x_2 = 5, x_3 = 3, x_4 = 12$, then

$$\sum_{i=1}^{4} x_i = 2 + 5 + 3 + 12 = 22, \qquad \sum_{i=1}^{4} x_i^2 = 2^2 + 5^2 + 3^2 + 12^2 = 182$$

More examples:

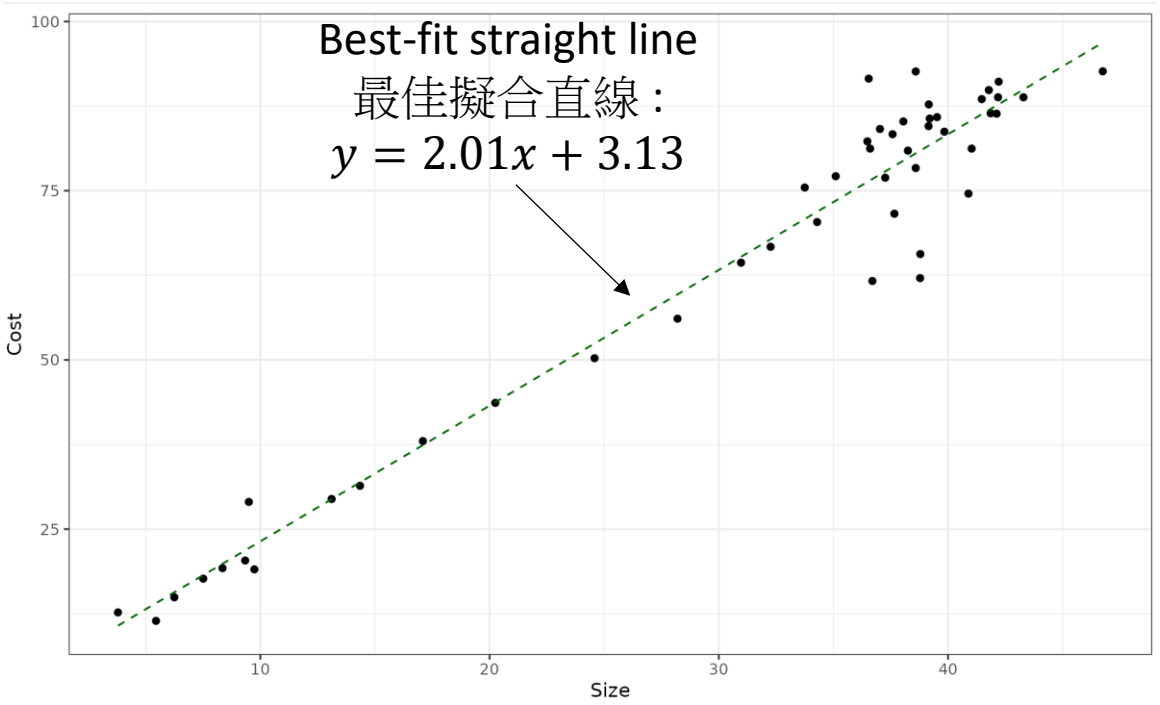
$$\sum_{k=2}^{5} 2k = 2(2) + 2(3) + 2(4) + 2(5) = 28$$

$$\sum_{m=-1}^{2} \frac{2}{m+3} = \frac{2}{-1+3} + \frac{2}{0+3} + \frac{2}{1+3} + \frac{2}{2+3} = \frac{77}{30}$$

Linear Regression 線性迴歸

- **The "best-fit" straight line is defined as the line $y = ax + b$ that minimizes the residual sum of squares *RSS*, where**
  最佳擬合直線定義為最小化殘差平方和 *RSS* 的直線 $y = ax + b$ ，其中

$$RSS = \big(y_1 - (ax_1 + b)\big)^2 + \big(y_2 - (ax_2 + b)\big)^2 + \cdots + \big(y_n - (ax_n + b)\big)^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$



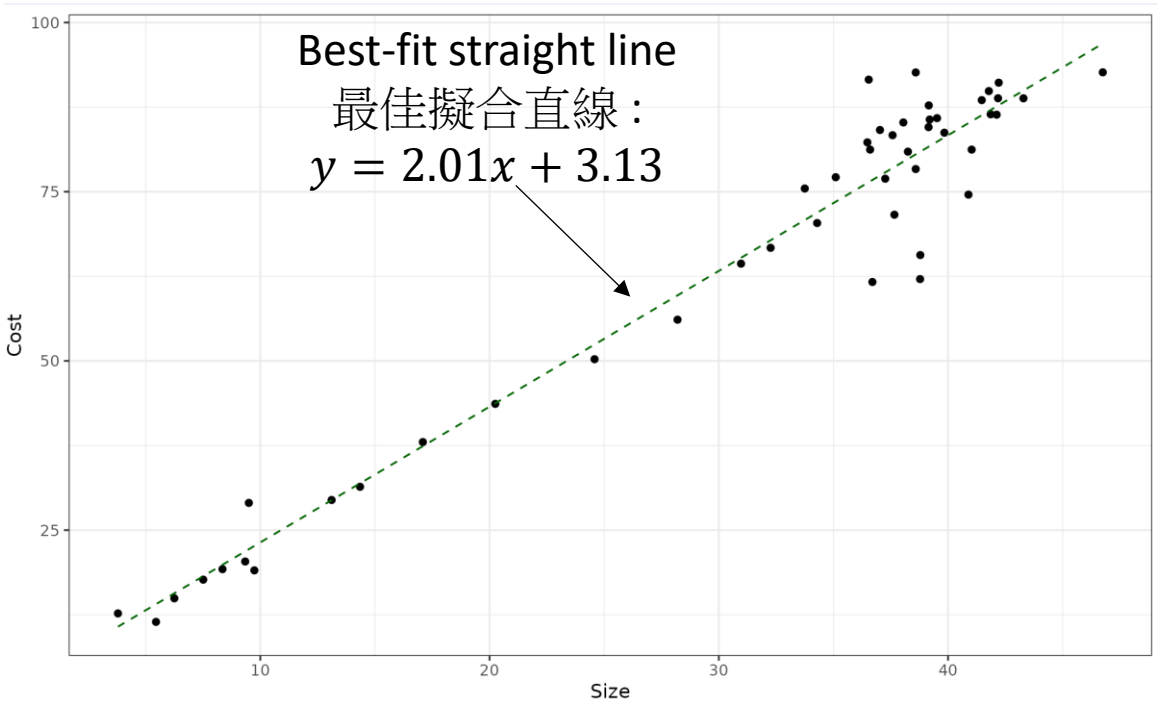
- $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ are the given data points 是給定的數據點
- $\hat{y}_i = ax_i + b$ is the fitted value 是擬合值

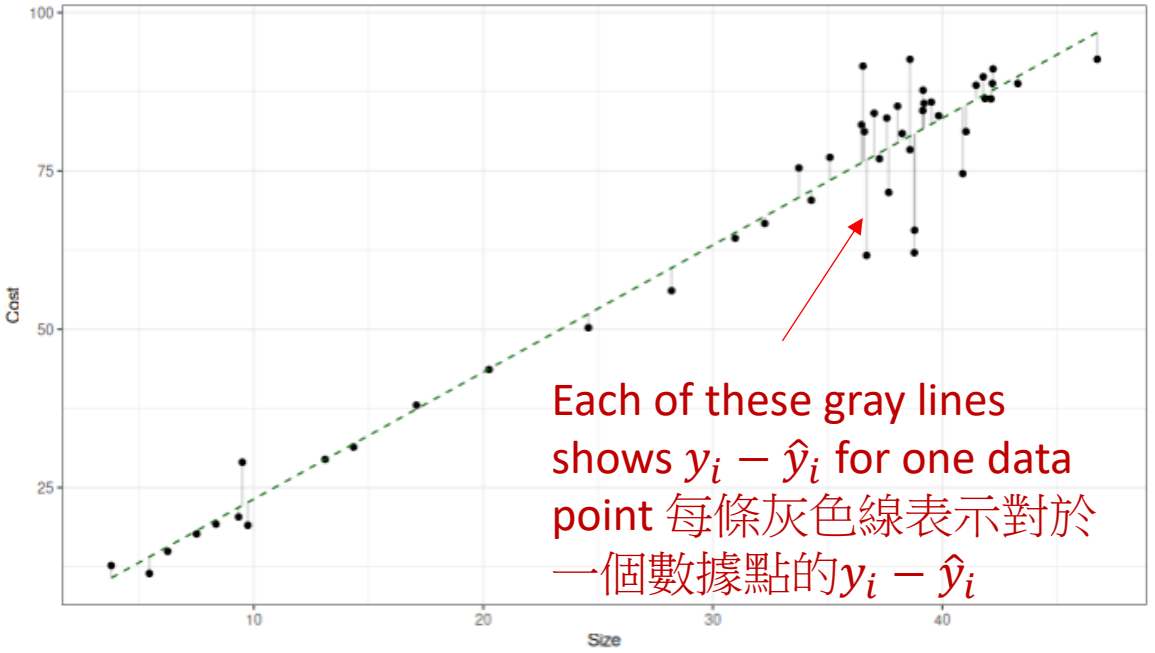
Linear Regression 線性迴歸

- **The "best-fit" straight line is defined as the line $y = ax + b$ that minimizes the residual sum of squares *RSS*, where**
  最佳擬合直線定義為最小化殘差平方和 *RSS* 的直線 $y = ax + b$ ，其中

$$RSS = \big(y_1 - (ax_1 + b)\big)^2 + \big(y_2 - (ax_2 + b)\big)^2 + \cdots + \big(y_n - (ax_n + b)\big)^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

  - $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ are the given data points 是給定的數據點
  - $\hat{y}_i = ax_i + b$ is the fitted value 是擬合值





Linear Regression 線性迴歸

- A very natural question: How can we find the values of a and b in the best-fit model?
  一個非常自然的問題：我們如何在最佳擬合模型中找到 a 和 b 的值？

- Mathematically, we can derive the solution using Calculus and Algebra to get the optimal a and b for $y = ax + b$:
  數學上，我們可以使用微積分和代數推導出在 $y = ax + b$ 中最優的 a 和 b:

Summation 求和符號

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

Example:
If $x_1 = 2, x_2 = 5, x_3 = 3, x_4 = 12,$
then

$$\sum_{i=1}^{4} x_i = 2 + 5 + 3 + 12 = 22,$$

$$\sum_{i=1}^{4} x_i^2 = 2^2 + 5^2 + 3^2 + 12^2 = 182$$

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

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

- Detailed derivation can be found in our e-book 詳細推導可參考我們的電子書

Background Knowledge: Vectors and Matrices 向量和矩陣

- **Vector** 向量 :

$$\begin{pmatrix}0\\3\end{pmatrix}, \begin{pmatrix}1\\2\\4\end{pmatrix}, (2 \quad 5 \quad 8), \dots$$

- **Matrix** 矩陣 :

$$\begin{pmatrix}0\\3\end{pmatrix}, \begin{pmatrix}1 & 2\\1 & -3\end{pmatrix}, \begin{pmatrix}a & b\\c & d\end{pmatrix}, \begin{pmatrix}1 & 0\\3 & 4\\6 & 2\end{pmatrix}, \dots$$

- **Basic operations** 基本操作 :

$$\begin{pmatrix}a & b\\c & d\end{pmatrix} + 2\begin{pmatrix}e & f\\g & h\end{pmatrix} = \begin{pmatrix}a+2e & b+2f\\c+2g & d+2h\end{pmatrix}, \qquad \begin{pmatrix}a & b\\c & d\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}ax+by\\cx+dy\end{pmatrix}$$

- **Determinant** of square matrices 方陣的行列式:

$$\begin{vmatrix}a & b\\c & d\end{vmatrix} = ad - bc$$

Linear Regression 線性迴歸

- Deriving the solution using Calculus and Algebra 使用微積分和代數推導:

$$U = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

Differentiating with respect to a:
對 a 求導：

$$\frac{\partial U}{\partial a} = 0$$
$$\Rightarrow \frac{\partial}{\partial a} \sum_{i=1}^{n} (y_i - a - bx_i)^2 = 0$$
$$\Rightarrow 2 \sum_{i=1}^{n} (y_i - a - bx_i)(-1) = 0$$
$$\Rightarrow \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$
$$\Rightarrow \sum_{i=1}^{n} y_i - na - b \sum_{i=1}^{n} x_i = 0$$
$$\Rightarrow \sum_{i=1}^{n} y_i = na + b \sum_{i=1}^{n} x_i$$

Differentiating with respect to b:
對 b 求導：

$$\frac{\partial U}{\partial b} = 0$$
$$\Rightarrow \frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - a - bx_i)^2 = 0$$
$$\Rightarrow 2 \sum_{i=1}^{n} (y_i - a - bx_i)(-x_i) = 0$$
$$\Rightarrow \sum_{i=1}^{n} (y_i x_i - a x_i - b x_i^2) = 0$$
$$\Rightarrow \sum_{i=1}^{n} y_i x_i - a \sum_{i=1}^{n} x_i - b \sum_{i=1}^{n} x_i^2 = 0$$
$$\Rightarrow \sum_{i=1}^{n} y_i x_i = a \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} x_i^2$$

Then we have a system of linear equations:
我們得出以下方程組：

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

Linear Regression 線性迴歸

- We can solve the system of equations
  我們可以求解方程組

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

using the following formula (called the Cramer's rule):
利用以下公式（稱為克萊瑪法則）：

$$a = \frac{\begin{vmatrix} \sum y & \sum x \\ \sum yx & \sum x^2 \end{vmatrix}}{\begin{vmatrix} n & \sum x \\ \sum x & \sum x^2 \end{vmatrix}} = \frac{\sum y \sum x^2 - \sum x \sum yx}{n\sum x^2 - (\sum x)^2} \qquad b = \frac{\begin{vmatrix} n & \sum y \\ \sum x & \sum yx \end{vmatrix}}{\begin{vmatrix} n & \sum x \\ \sum x & \sum x^2 \end{vmatrix}} = \frac{n\sum yx - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$$

(covered in M2!)

Linear Regression 線性迴歸

- Too complicated? Don't worry!
  太複雜了？不用擔心！

- Computationally, we can use our Linear Regression R Shiny tool
  to find the best-fit straight line $y = ax + b$
  在計算方面，我們可以使用 R Shiny 線性迴歸工具來找到最佳擬合直線 $y = ax + b$

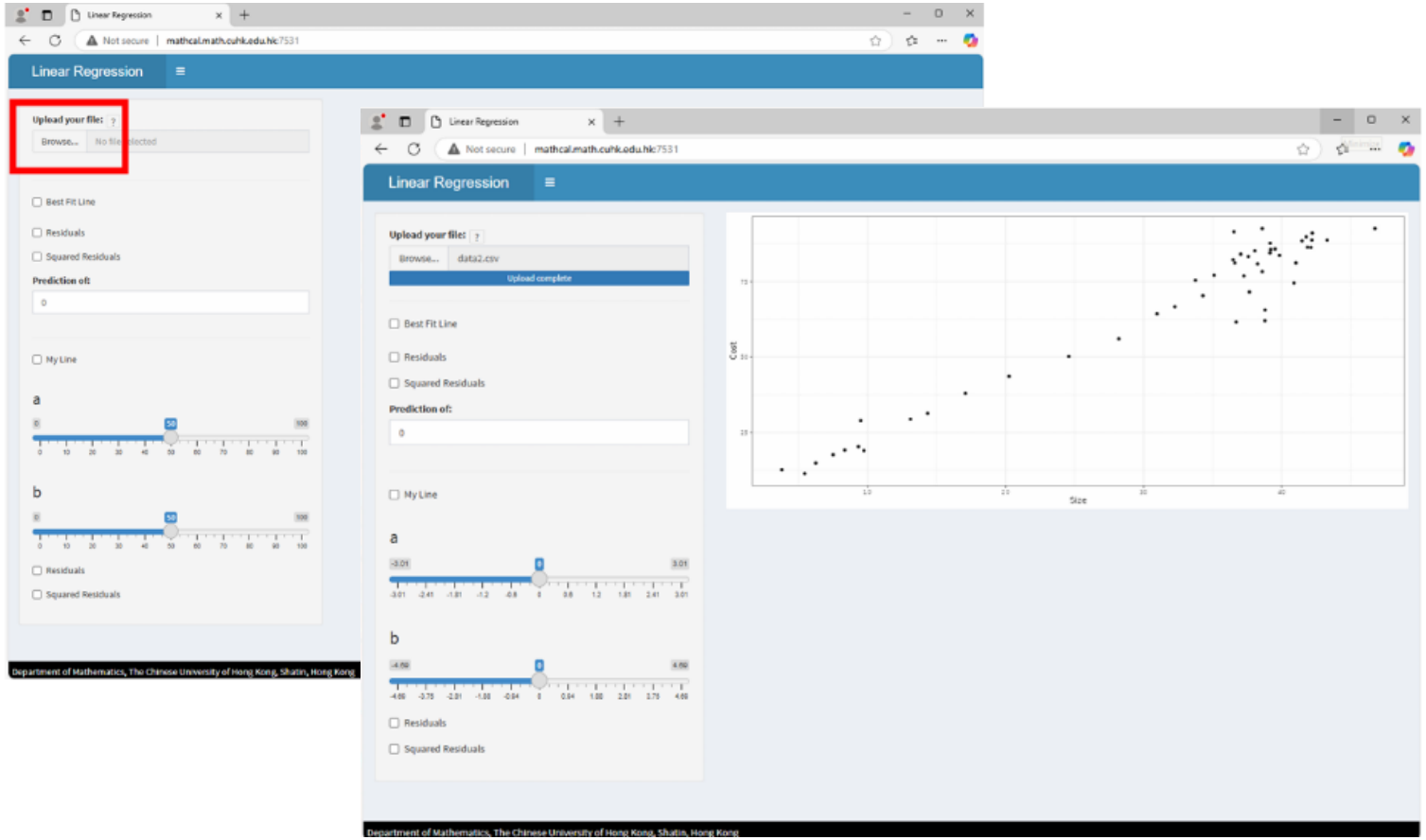
  https://www.math.cuhk.edu.hk/app/mathmodel/tool.html

Linear Regression with R Shiny 利用 R Shiny 進行線性迴歸

- https://www.math.cuhk.edu.hk/app/mathmodel/tool.html

File format:
- CSV
- XLSX
- TXT

Step 1: Upload your own dataset
步驟 1：上傳資料集



For the CSV/XLSX file, there should be 2 columns of data.

	A	B
1	Size	Cost
2	46.75	92.64
3	42.18	88.81
4	41.86	86.44
5	43.29	88.8
6	42.12	86.38
7	41.78	89.87
8	41.47	88.53
9	42.21	91.11
10	41.03	81.22
11	39.84	83.72
12	39.15	84.54
13	39.2	85.66

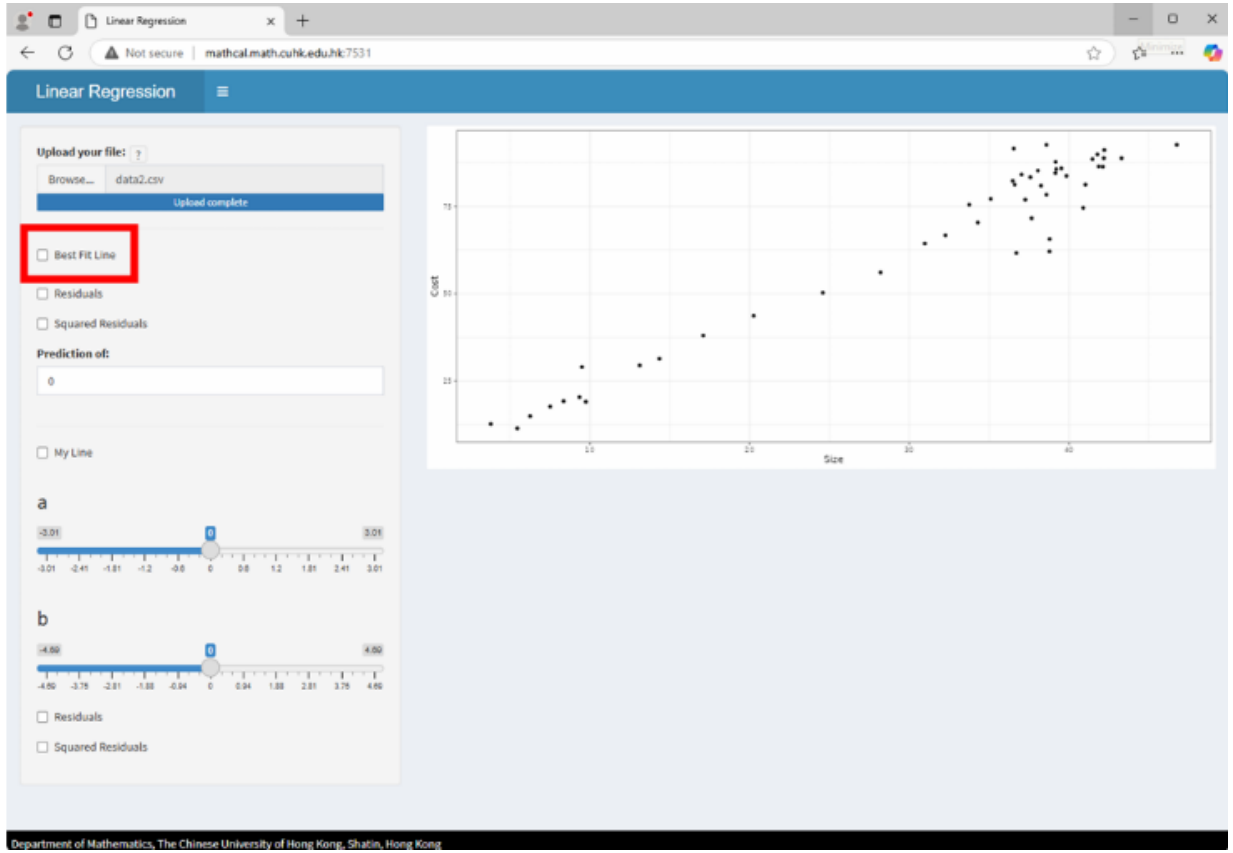
For the TXT file, the entries should be comma-separated.

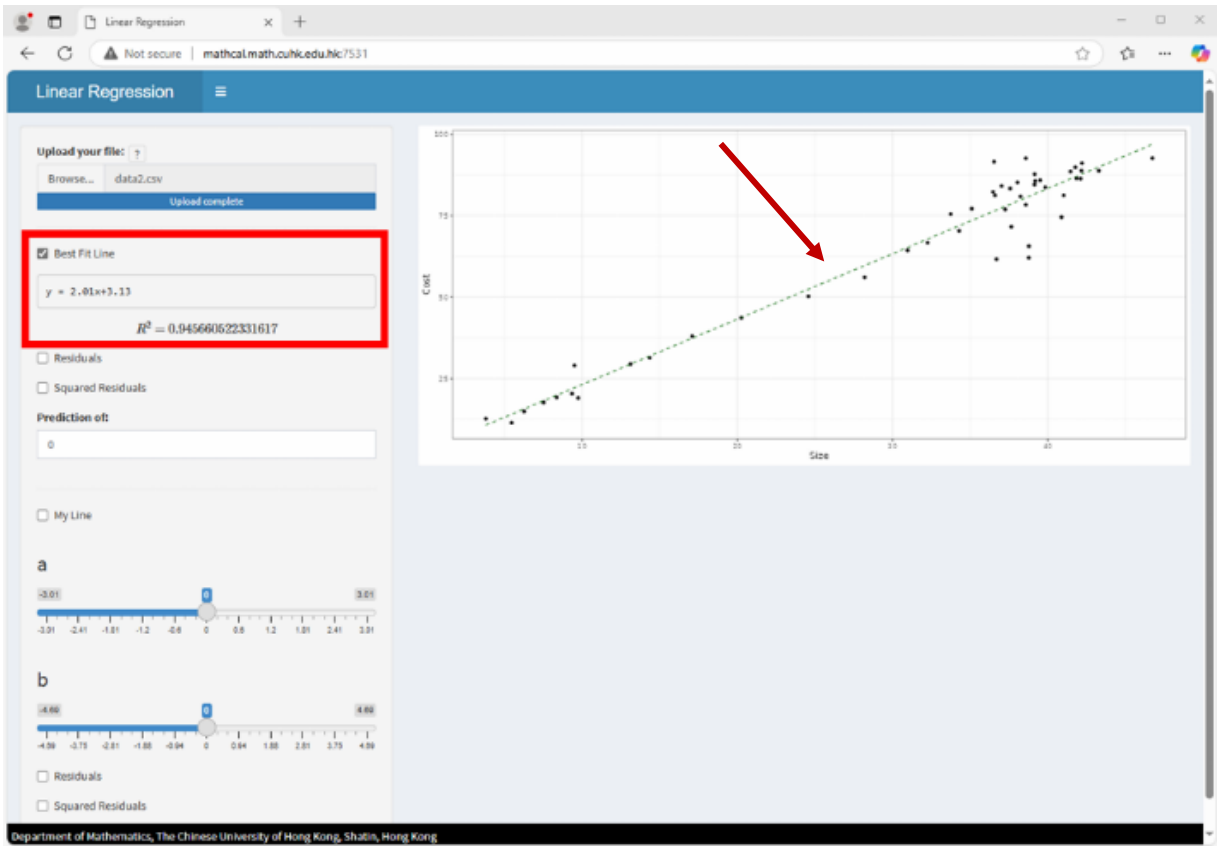
```
x,y
1.5,40.6
3.8,65
8.9,70.8
10,78
12,84
20,108.5
```

Linear Regression with R Shiny 利用 R Shiny 進行線性迴歸

- https://www.math.cuhk.edu.hk/app/mathmodel/tool.html

Step 2: Find the best-fit line (with the equation and the R^2 value)
步驟 2：找到最佳擬合線（包含方程式和 R^2值）

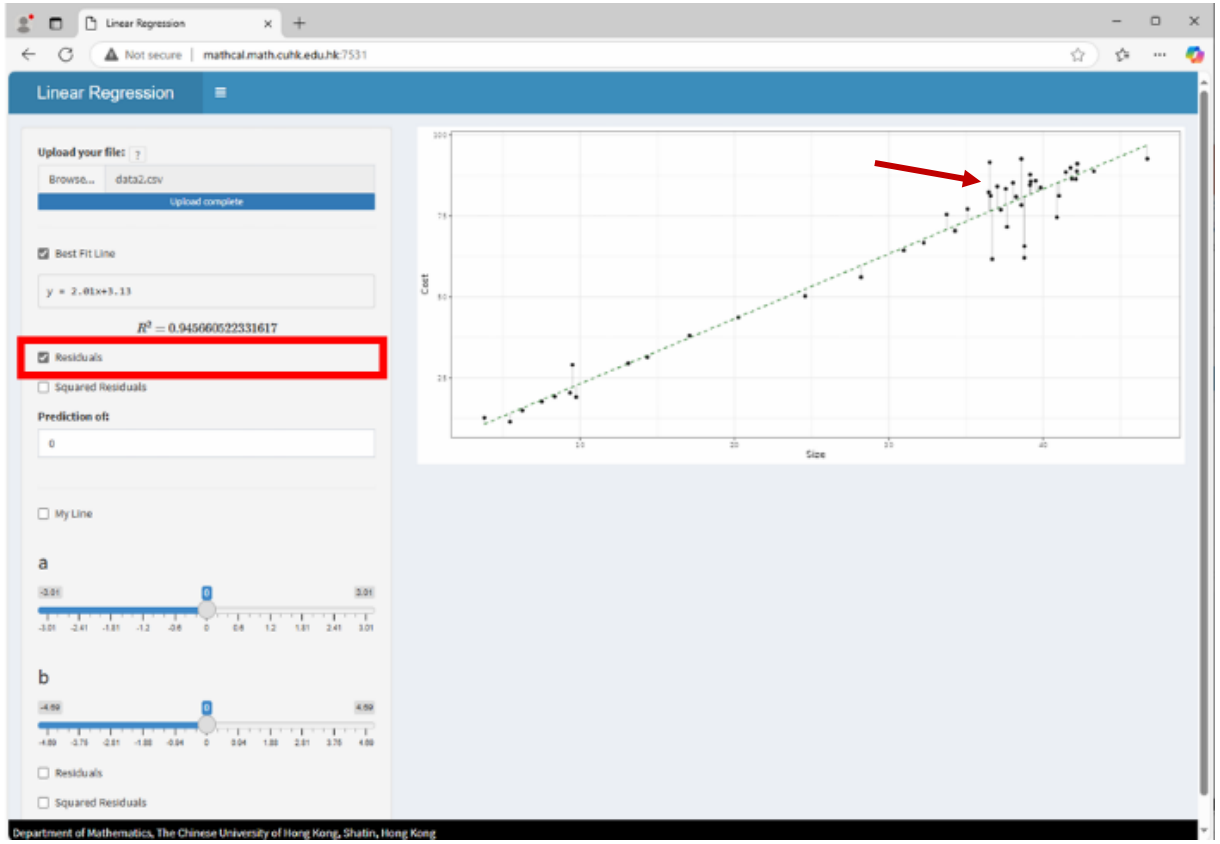




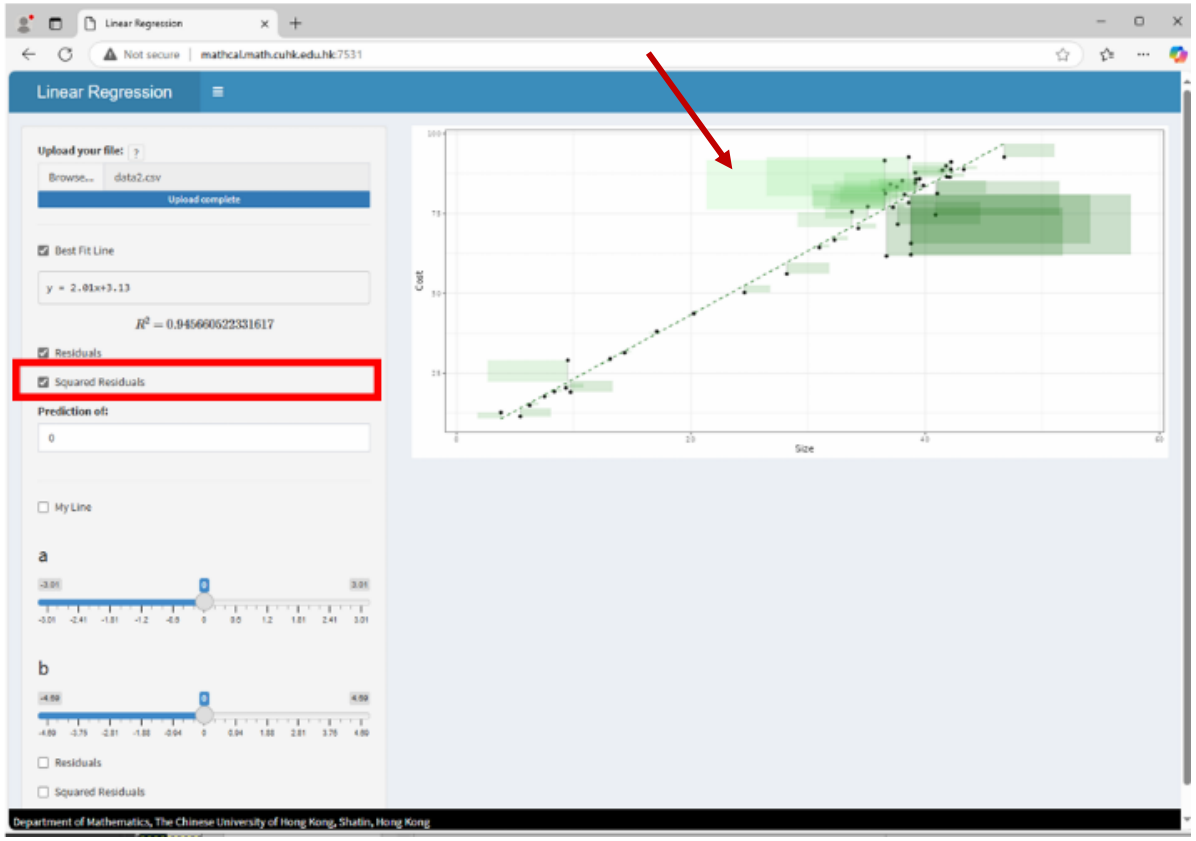
Linear Regression with R Shiny 利用 R Shiny 進行線性迴歸

- https://www.math.cuhk.edu.hk/app/mathmodel/tool.html

Display the residuals 顯示殘差



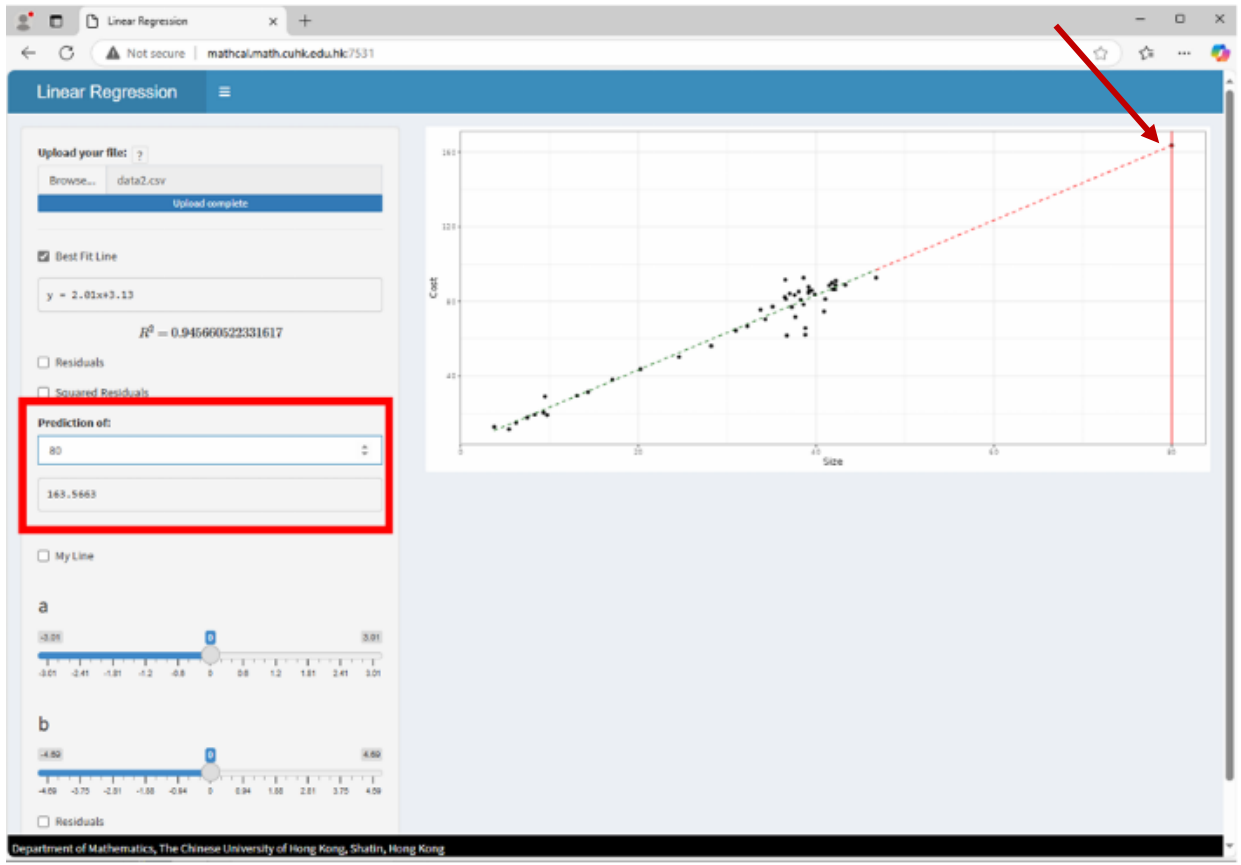
Display the squared residuals 顯示殘差平方



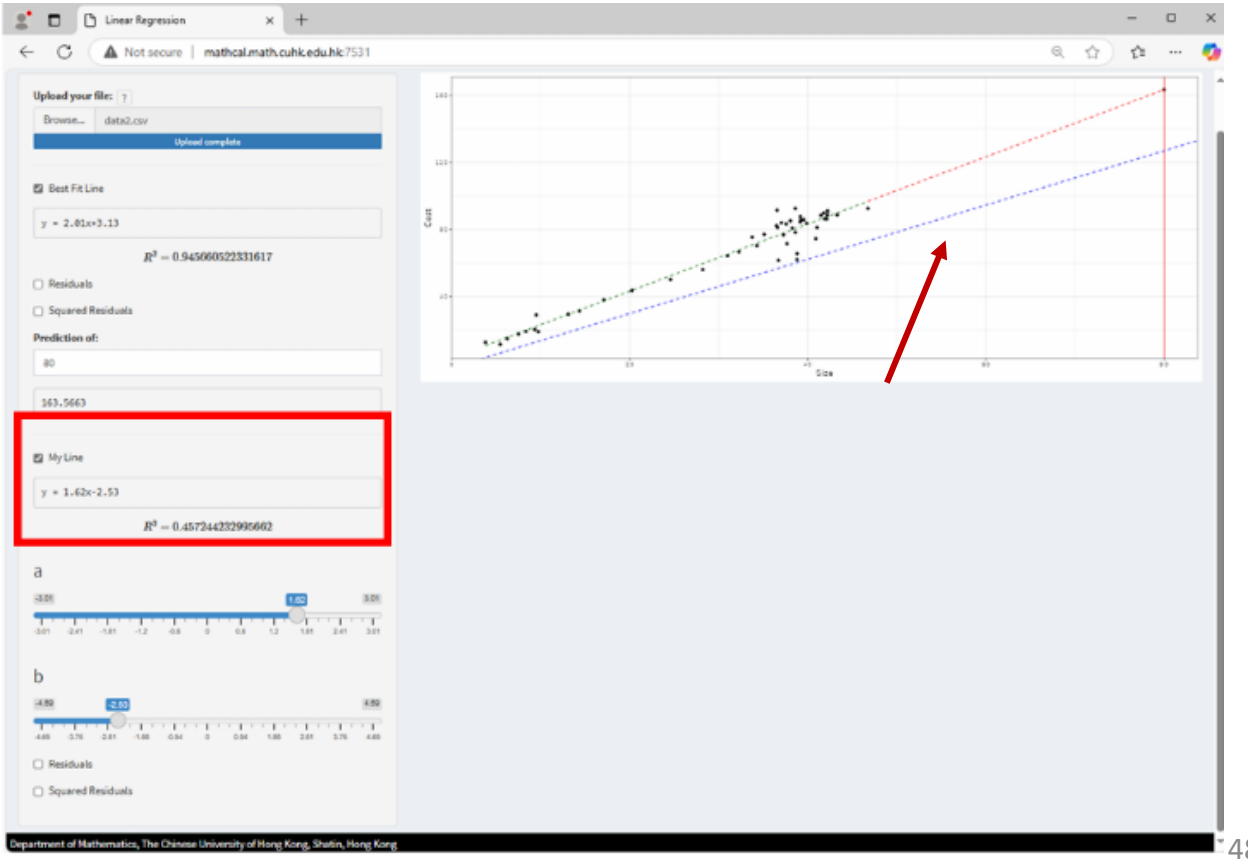
Linear Regression with R Shiny 利用 R Shiny 進行線性迴歸

- https://www.math.cuhk.edu.hk/app/mathmodel/tool.html

Find the predicted value at a specific point
找出指定點的預測值



Define your own “best-fit” line
自行定義「最佳擬合」線



Linear Regression 線性迴歸

- **Exercise 練習**

  - Predicting the world population growth is important in social science.
    估計世界人口增長是社會科學的重要問題。
  - Consider the dataset on the RHS.
    考慮右邊的數據集。
  - Utilize the Linear Regression RShiny tool
    利用線性迴歸 RShiny 工具:
    https://www.math.cuhk.edu.hk/app/mathmodel/tool.html
    1. Find the best-fit linear model.
       找出最佳線性模型。
    2. Predict the world population in Year 2030, 2040 and 2050.
       估計在2030, 2040 和2050年的世界人口。

Year 年份	World Population (in billion) 世界人口 (十億)
2010	7.0
2012	7.2
2015	7.5
2019	7.8
2022	8.0
2023	8.1

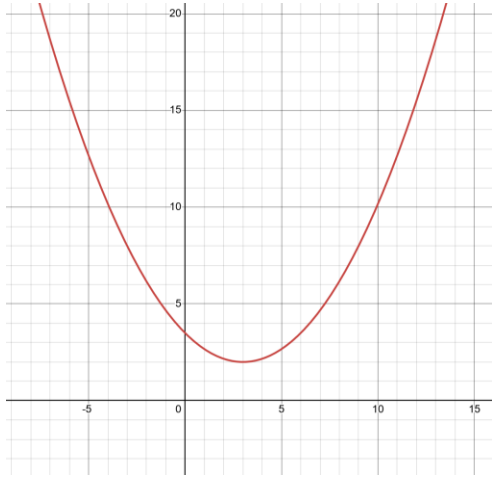
Source: Worldometer (www.Worldometers.info)

Non-Linear Regression 非線性迴歸

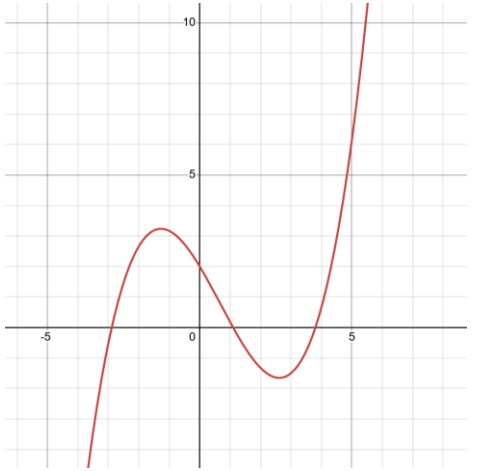
- **What if the data points do not form a linear trend? 如果數據點不形成線性趨勢，怎麼辦？**

- Extend the method to other polynomials!
  將方法擴展到其他多項式！
  - Quadratic model 二次多項式模型:
    $$y = ax^2 + bx + c$$
  - Cubic model 三次多項式模型:
    $$y = ax^3 + bx^2 + cx + d$$
  - Quadric model 四次多項式模型:
    $$y = ax^4 + bx^3 + cx^2 + dx + e$$
  - Quintic model 五次多項式模型:
    $$y = ax^5 + bx^4 + cx^3 + dx^2 + ex + f$$
  - General polynomial model 一般多項式模型:
    $$y = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

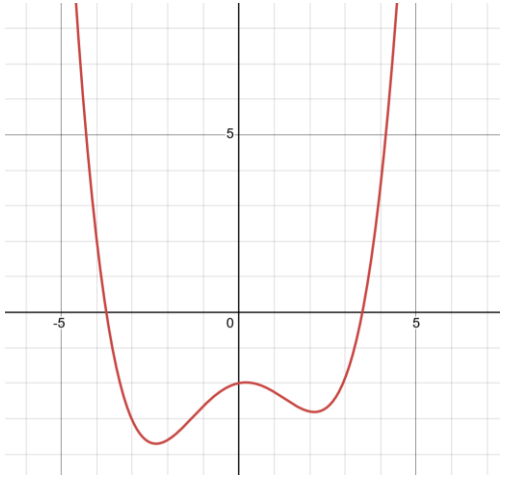
Quadratic model
二次多項式模型



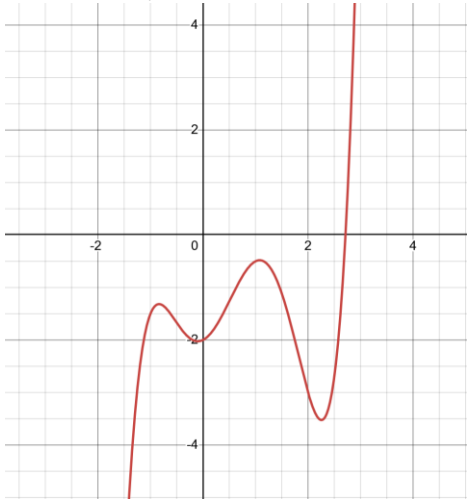
Cubic model
三次多項式模型



Quadric model
四次多項式模型



Quintic model
五次多項式模型



Non-Linear Regression 非線性迴歸

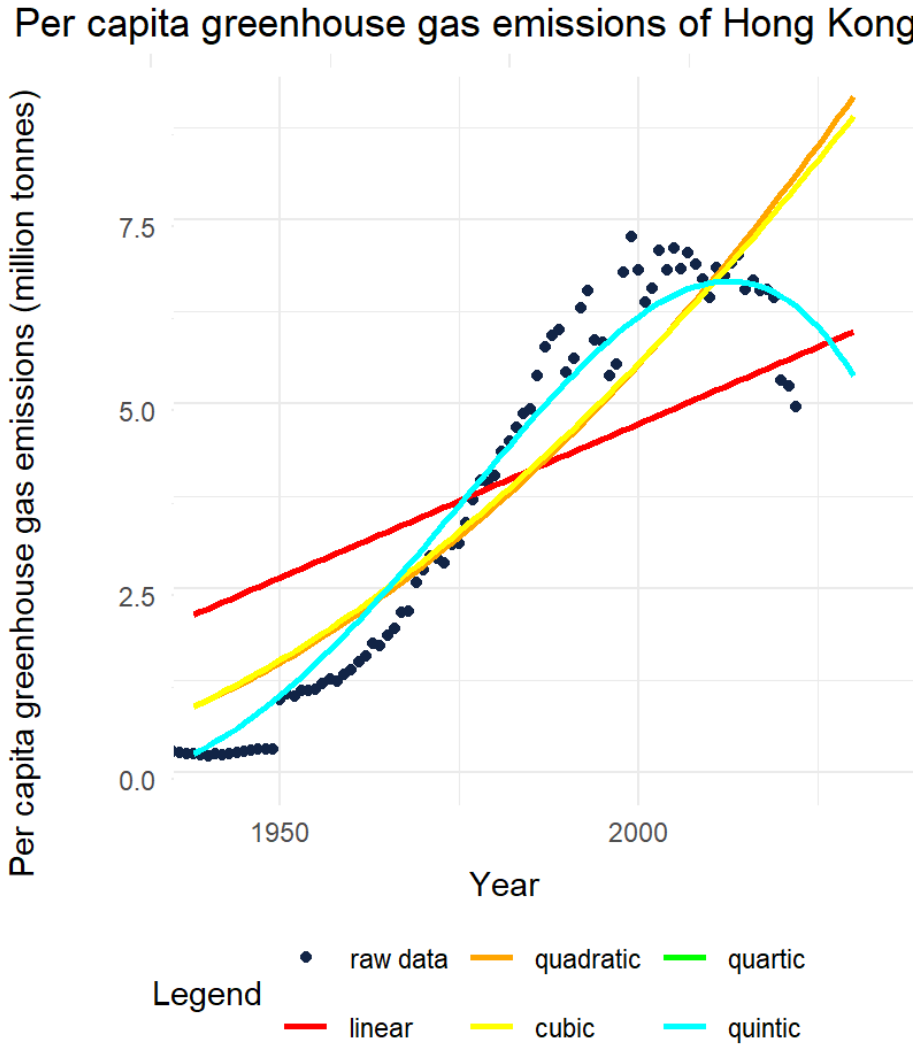
- Find the coefficients of $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ that minimizes the residual sum of squares
  找出最小化殘差平方和的函數 $f(x)$系數

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

- Similar to the linear model, we can mathematically derive a formula to find the values of all coefficients for the best-fit model.
  與線性模型類似，我們可以從數學上推導出一條公式來找到最佳擬合模型的所有系數的值。

- The formula is much more complicated!
  但是，公式要複雜得多！

- We may also consider models other than polynomials
  我們亦可考慮多項式以外的模型



Background knowledge: Logarithm 對數公式

- **Logarithm 對數公式**
  - $y = x^a \Leftrightarrow a = \log_x y$
    (for $x, y > 0$ with $x \neq 1$)
  - $\log a + \log b = \log ab$ (for $a, b > 0$)
  - $\log a - \log b = \log \frac{a}{b}$ (for $a, b > 0$)
  - $\log_b a = \frac{\log_c a}{\log_c b}$
    (for $a, b, c > 0$ with $b \neq 1$ and $c \neq 1$)
  - $\log_a 1 = 0$ (for $a > 0$ with $a \neq 1$)
  - $\log_a a = 1$ (for $a > 0$ with $a \neq 1$)

Example:

- $\log_{10} 100 = 2$ (since $100 = 10^2$)
- $\log_3 81 = 4$ (since $81 = 3^4$)
- $\log 6 = \log 2 + \log 3$
- $\log x^2 y^3$
  $= \log x^2 + \log y^3$
  $= 2 \log x + 3 \log y$

Non-Linear Regression 非線性迴歸

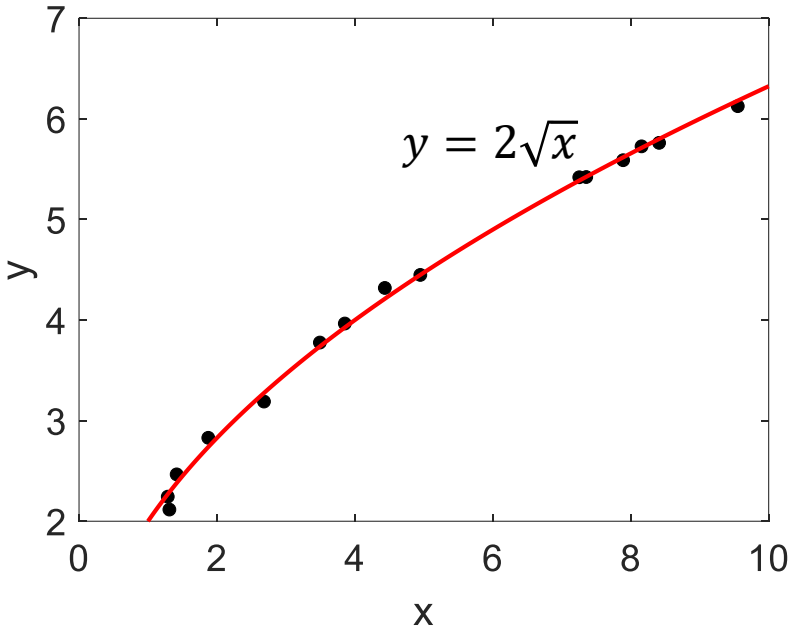
- **Power model 冪模型**

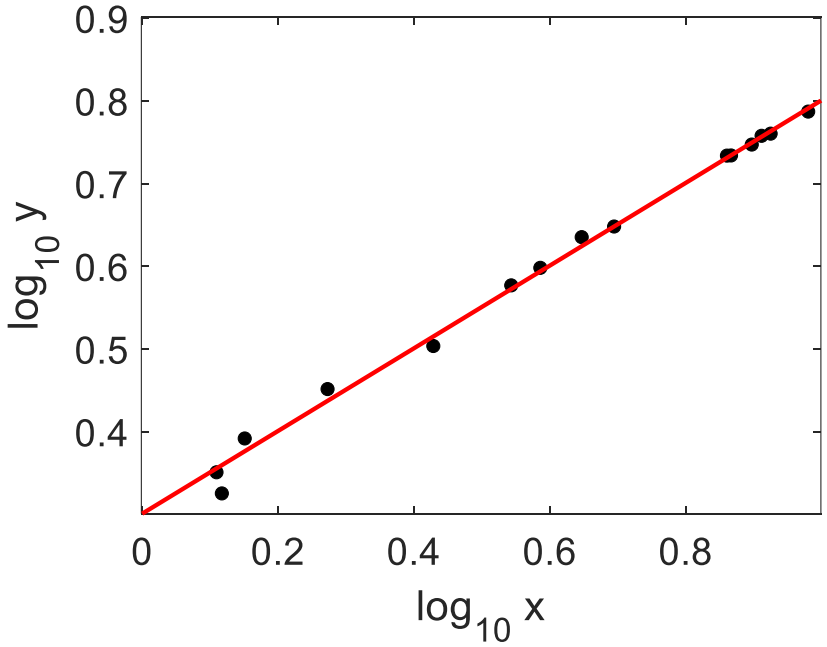
$$y = ax^b$$

- Linearization 線性化:

$$\log y = \log\left(ax^b\right) = \log a + b \log x$$

- Can then treat it as a linear regression problem with data points $(\log x, \log y)$
  然後可以將其視為數據點 $(\log x, \log y)$ 的線性迴歸問題





Non-Linear Regression 非線性迴歸

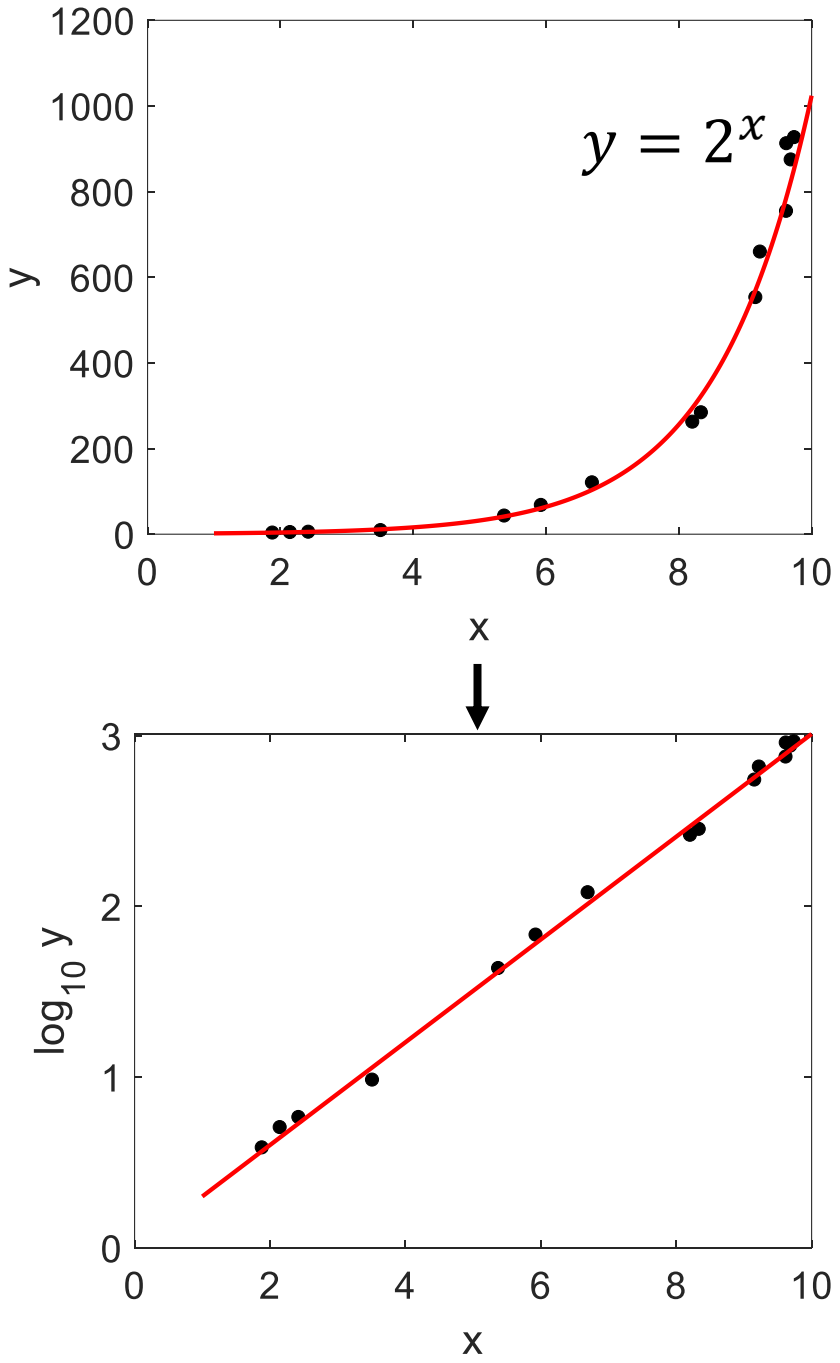
- **Exponential model 指數模型**

$$y = ab^x$$

- Linearization 線性化：

$$\log y = \log(ab^x) = \log a + x \log b$$

- Can then treat it as a linear regression problem with data points $(x, \log y)$
然後可以將其視為數據點 $(x, \log y)$ 的線性迴歸問題

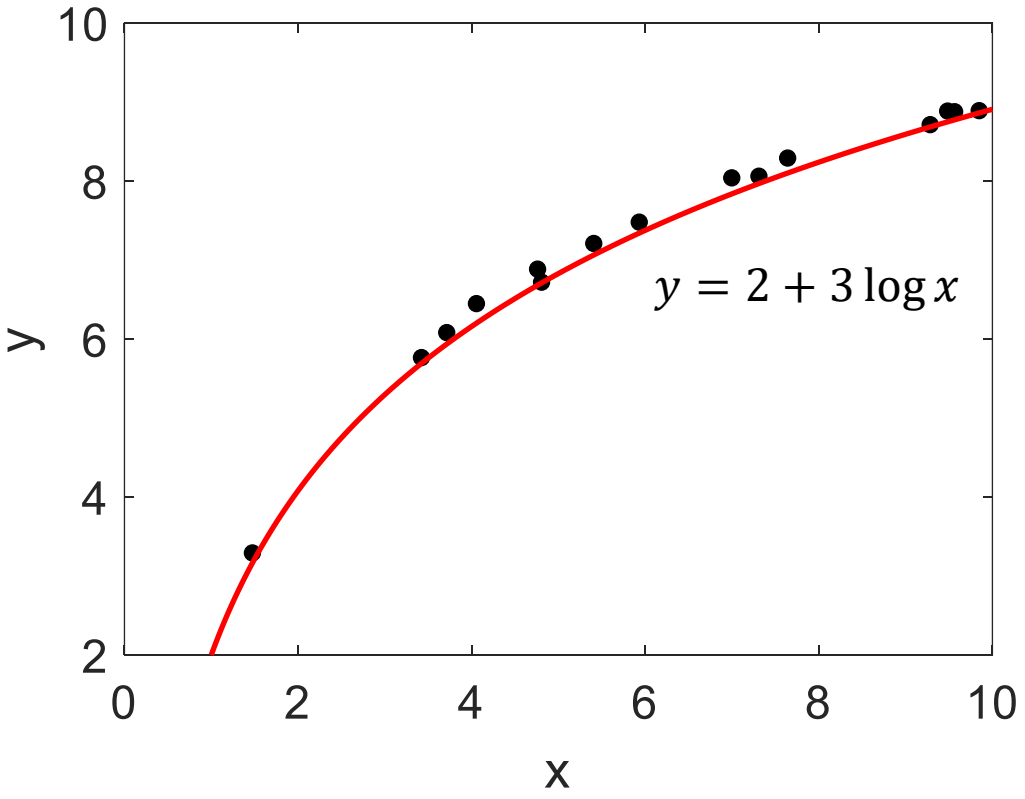


Non-Linear Regression 非線性迴歸

- **Logarithmic model 對數模型**

$$y = a + b \log x$$

  - Can treat it as a linear regression problem with data points $(\log x, y)$
    可以將其視為數據點 $(\log x, y)$ 的線性迴歸問題



Non-Linear Regression 非線性迴歸

- Computationally, we can use our Non-linear Regression R Shiny tool
  在計算方面，我們可以使用 R Shiny 非線性迴歸工具
  https://www.math.cuhk.edu.hk/app/mathmodel/tool.html

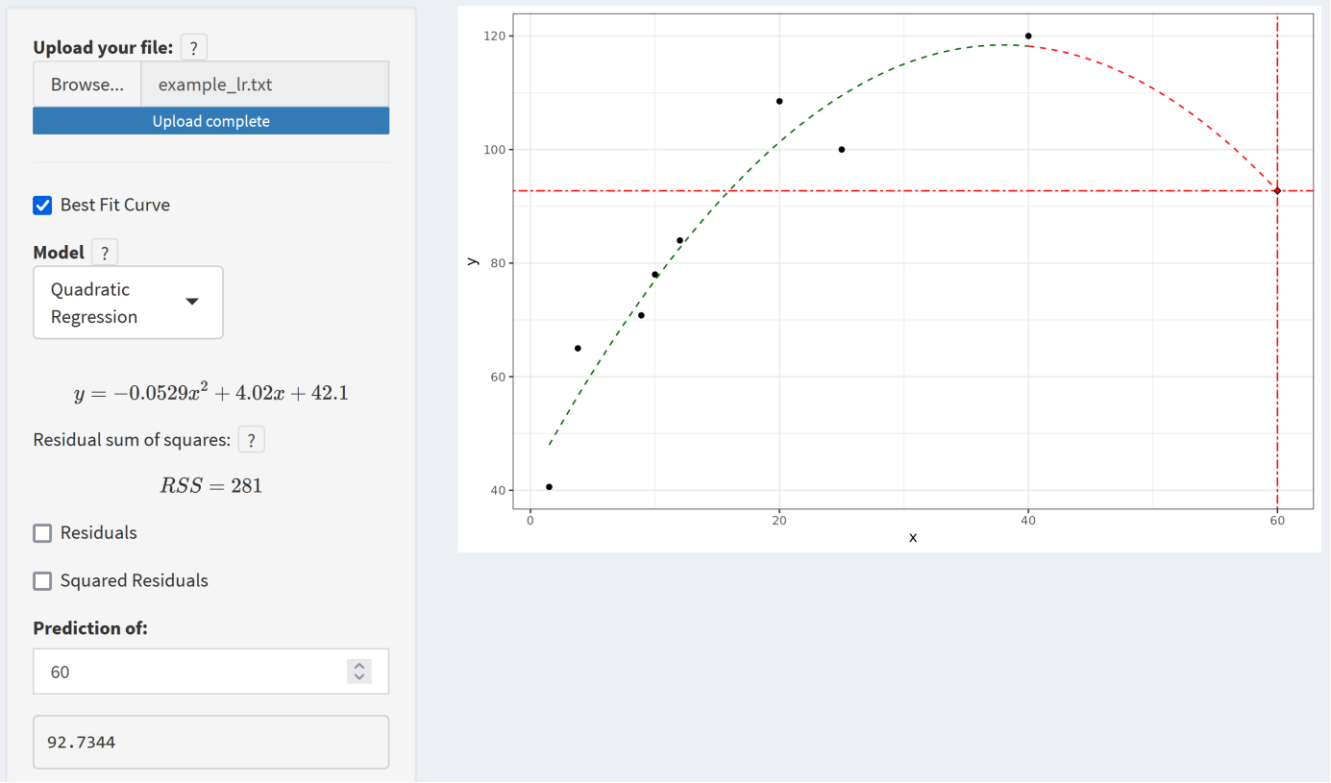
  - Linear model 線性模型
  - Quadratic model 二次模型
  - Cubic model 三次模型
  - Polynomial model 多項式模型
  - Power model 冪模型
  - Exponential model 指數模型
  - Logarithmic model 對數模型

Non-Linear Regression with R Shiny (XY data)
利用 R Shiny 進行非線性迴歸 (XY數據)

- https://www.math.cuhk.edu.hk/app/mathmodel/tool.html

- Same file format as in the Linear Regression R Shiny tool
  與線性迴歸 R Shiny 工具中的檔案格式相同

- Many models available
  多種模型可供選擇:
  - Linear model 線性模型
  - Quadratic model 二次模型
  - Cubic model 三次模型
  - Polynomial model 多項式模型
  - Power model 冪模型
  - Exponential model 指數模型
  - Logarithmic model 對數模型



Non-Linear Regression with R Shiny (XY data)
利用 R Shiny 進行非線性迴歸 (XY數據)

- **Exercise 練習**
  - Consider the average global life expectancy data on the RHS.考慮右邊的全球平均預期壽命數據。

  - Utilize the Non-Linear Regression RShiny tool:
    利用非線性迴歸 RShiny 工具：
    https://www.math.cuhk.edu.hk/app/mathmodel/tool.html
    - Try the Linear, Quadratic and Cubic models. Compare their RSS values.
      嘗試線性、二次和三次模型。比較它們的RSS值。

    - Predict the average life expectancy in Year 2030, 2040 and 2050 using each approach.
      使用每種方法預測 2030 年、2040 年和 2050 年的平均壽命。

Year 年份	Life expectancy 預期壽命
1950	46.4
1960	47.8
1970	56.3
1980	60.5
1990	64
2000	66.4
2010	70.1
2020	71.9

See more:
https://ourworldindata.org/life-expectancy

Overfitting and model validation

過度擬合及模型驗證

Overfitting and model validation 過度擬合及模型驗證

- As we can see earlier 正如我們早前看到：

  Increasing the degree of the polynomial → Reducing the RSS!

  增加多項式的次數 → 減少 RSS！

- In fact, for any given set of n data points, we can always find a degree $(n-1)$ polynomial that exactly passes through all points. In this case, we have RSS = 0 (Try it using the R Shiny tool!)

  事實上，對於任何給定的 n 個數據點，我們總可以找到一個$(n-1)$次多項式，能完全通過所有數據點。在這種情況下，RSS = 0（試試使用 R Shiny 工具！）

- However, does that reflect the actual trend of the data?

  但這真的反映出數據的實際趨勢嗎？

- We have to be careful about the issue of overfitting

  我們必須小心過度擬合的問題

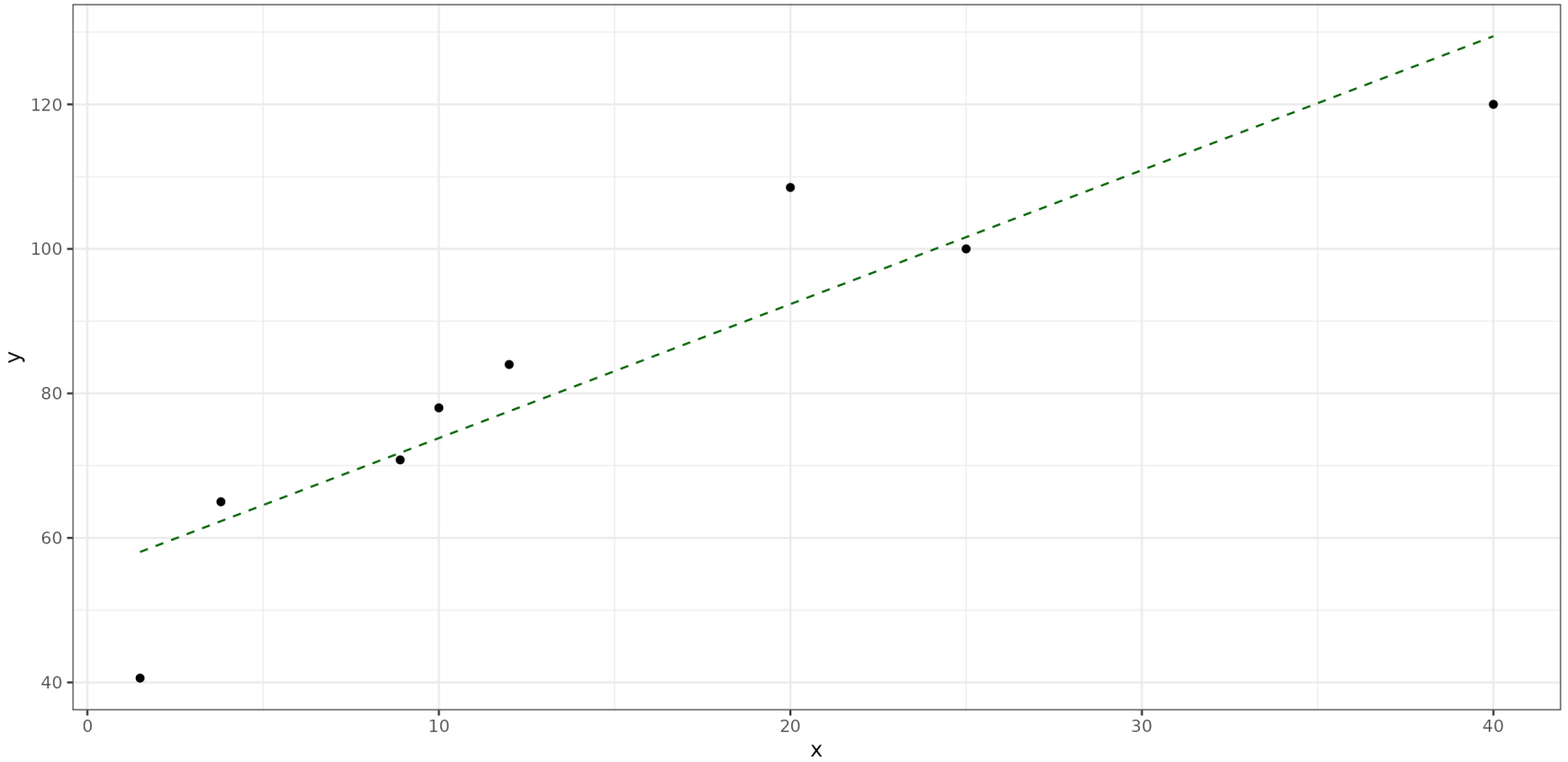
The Issue of Overfitting 過度擬合

- We have to be careful about the issue of overfitting
  我們必須小心過度擬合的問題

$$y = 1.85x + 55.3$$

Residual sum of squares: ?

$$RSS = 725$$



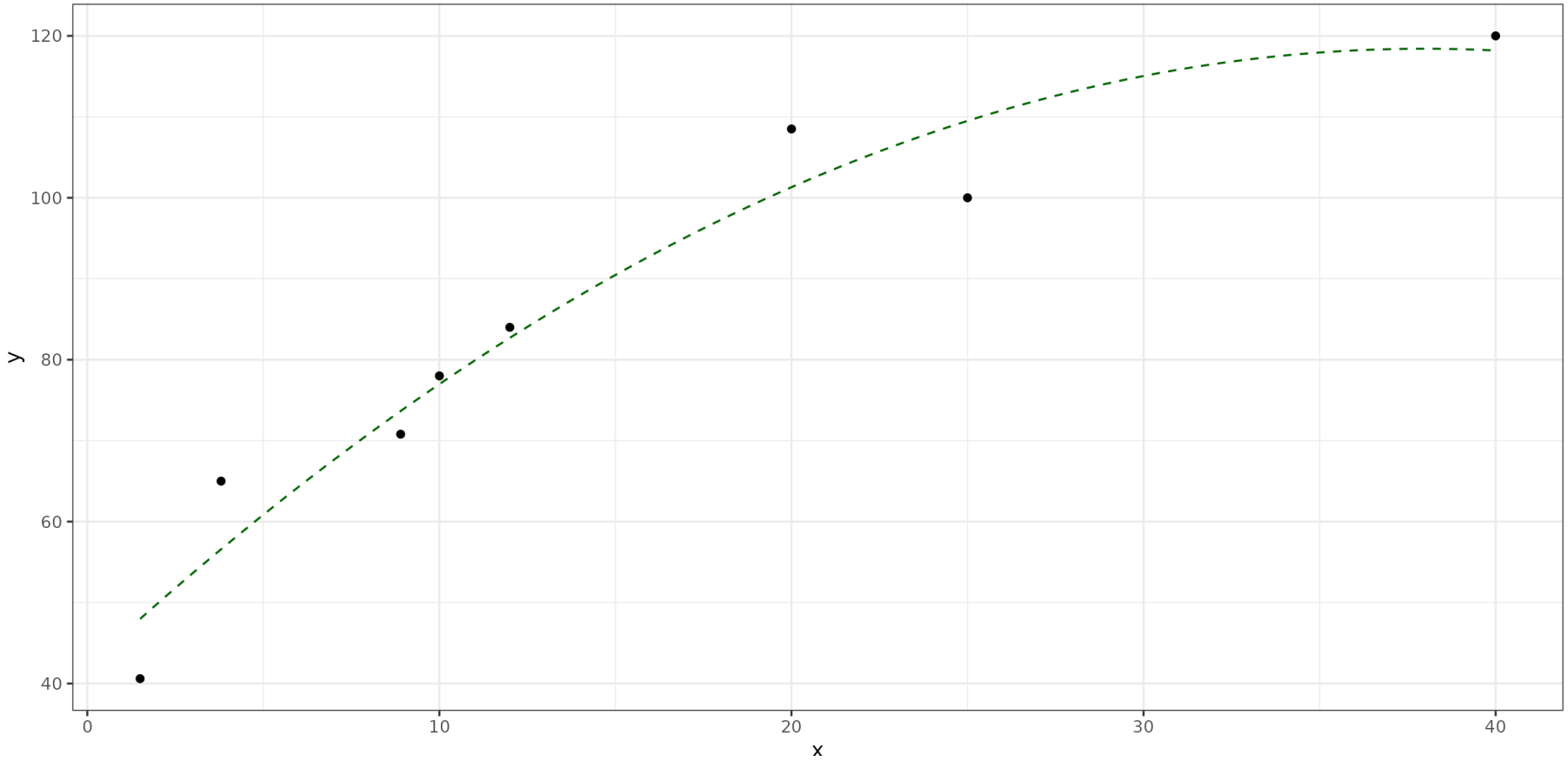
The Issue of Overfitting 過度擬合

- We have to be careful about the issue of overfitting
  我們必須小心過度擬合的問題

$$y = -0.0529x^2 + 4.02x + 42.1$$

Residual sum of squares: ?

$$RSS = 281$$



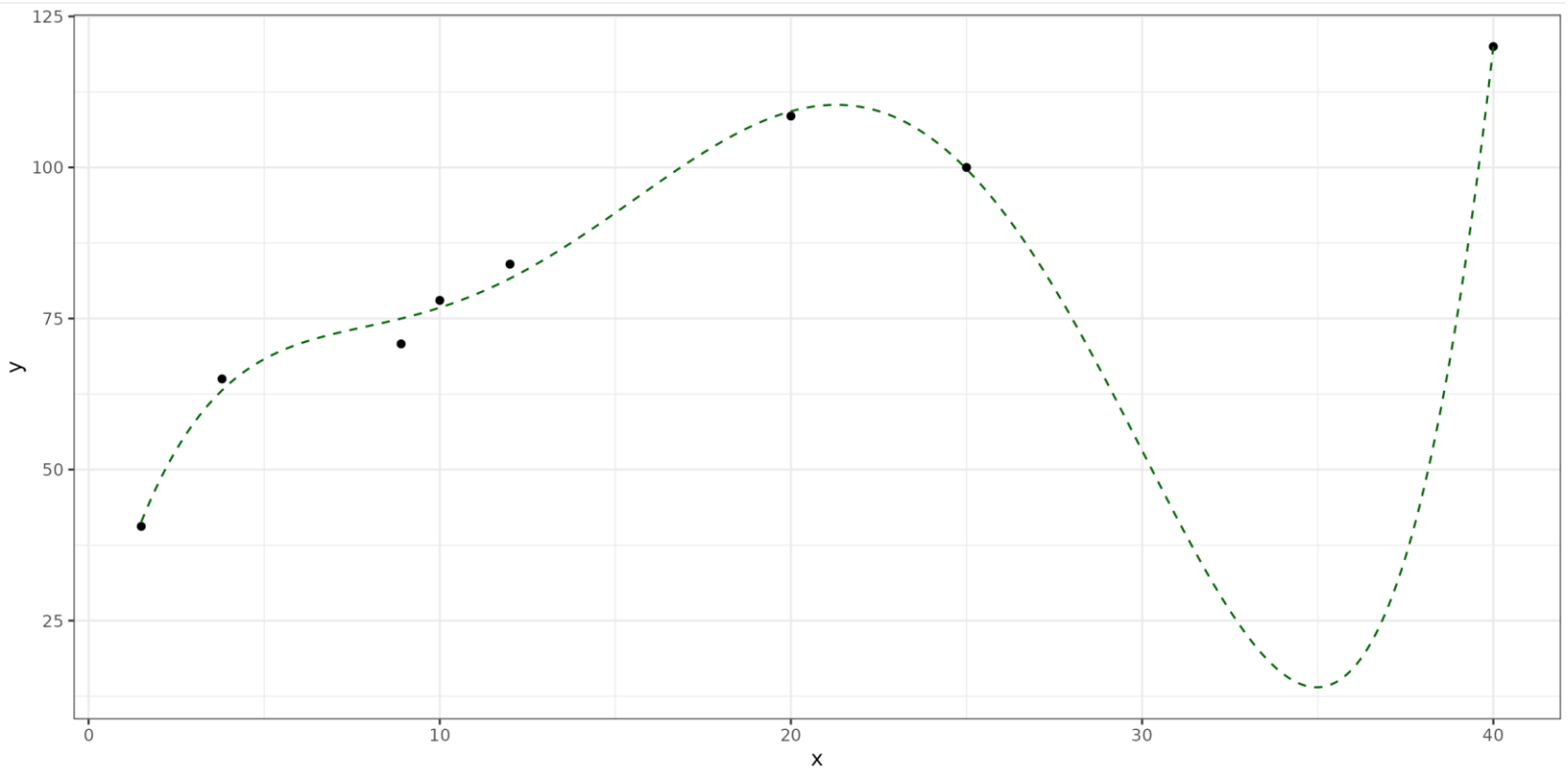
The Issue of Overfitting 過度擬合

- We have to be careful about the issue of overfitting
  我們必須小心過度擬合的問題

$$y = 0.000102x^5 - 0.00909x^4 + 0.28x^3 - 3.74x^2 + 23.8x + 13.2$$

Residual sum of squares: ?

$$RSS = 29.5$$



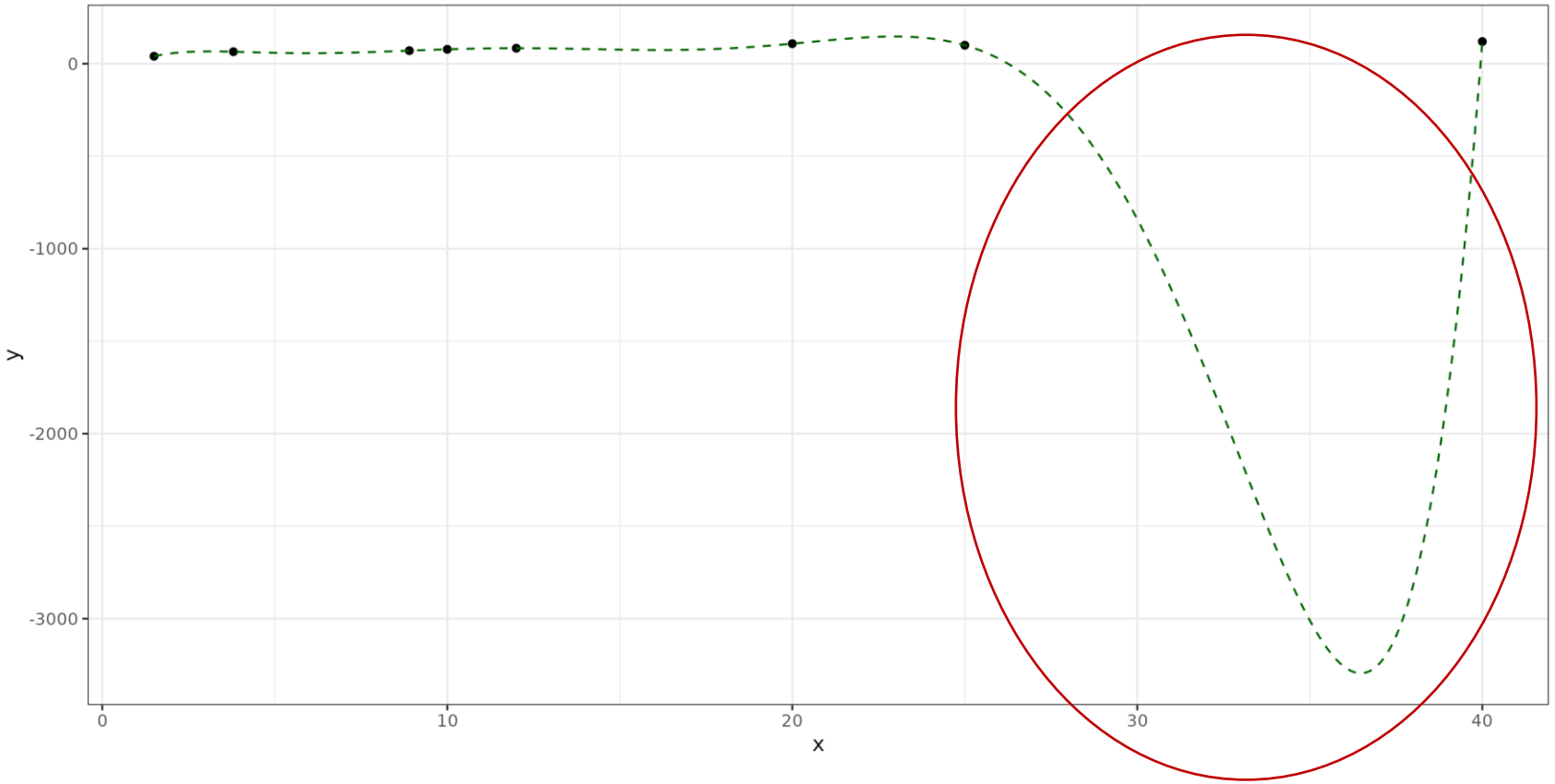
The Issue of Overfitting 過度擬合

- We have to be careful about the issue of overfitting
  我們必須小心過度擬合的問題

$$y = 6.89 \cdot 10^{-6}x^7 - 0.000777x^6 + 0.034x^5 - 0.742x^4 + 8.6x^3 - 51.8x^2 + 148x - 89.8$$

Residual sum of squares: ?

$$RSS = 0$$



What's happening??

Overfitting and model validation 過度擬合及模型驗證

- How to obtain a reasonable model that can explain the data trend?
  如何獲得一個合理的、能夠解釋資料趨勢的模型？

- **Method of divided difference 差分法**
  - For finding a suitable polynomial model 找出合適的多項式模型

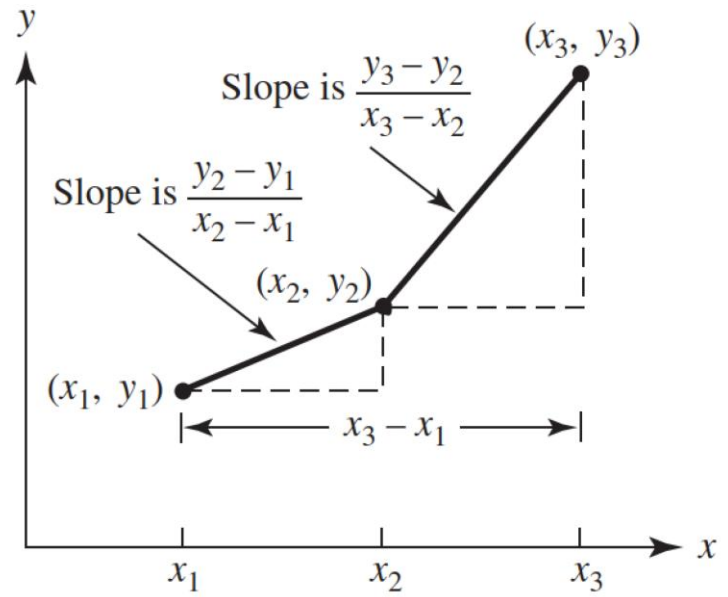
- **Model validation 模型驗證**
  - Applicable to more general models 可用於更一般的模型
  - Training 訓練
  - Testing 測試

Method of divided difference 差分法

- For finding a suitable polynomial model 找出合適的多項式模型

- Given three data points $(x_1, y_1), (x_2, y_2), (x_3, y_3)$, we can consider
  對於三個數據點 $(x_1, y_1), (x_2, y_2), (x_3, y_3)$, 我們可以考慮
  - The first divided difference 一階差分 $\frac{y_2 - y_1}{x_2 - x_1}$, $\frac{y_3 - y_2}{x_3 - x_2}$
  - The second divided difference 二階差分 $\dfrac{\frac{y_3 - y_2}{x_3 - x_2} - \frac{y_2 - y_1}{x_2 - x_1}}{x_3 - x_1}$

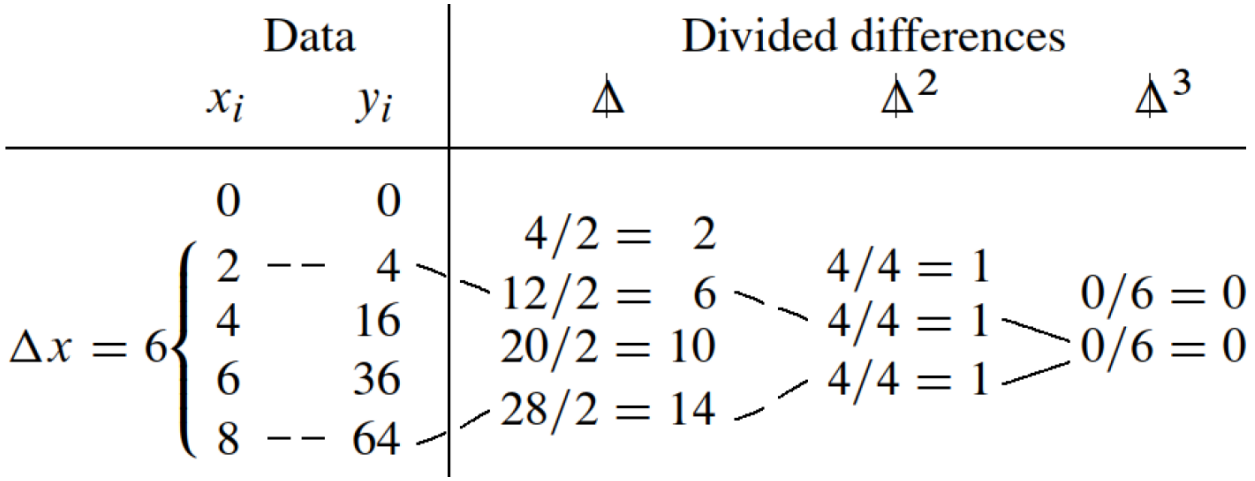
Data		First divided difference	Second divided difference
x_1	y_1		
		$\frac{y_2 - y_1}{x_2 - x_1}$	
			$\dfrac{\frac{y_3 - y_2}{x_3 - x_2} - \frac{y_2 - y_1}{x_2 - x_1}}{x_3 - x_1}$
x_2	y_2		
		$\frac{y_3 - y_2}{x_3 - x_2}$	
x_3	y_3		



Method of divided difference 差分法

- Given multiple data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, we can consider
  對於多個數據點 $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, 我們可以考慮
  - The first divided difference 一階差分
  - The second divided difference 二階差分
  - The third divided difference 三階差分
  - …
  - The $(n-1)$-th divided difference $(n-1)$階差分

x_i	0	2	4	6	8
y_i	0	4	16	36	64



Method of divided difference 差分法

- If the k-th divided differences are all 0, it means that the data points form a $(k-1)$-th degree polynomial
  如果所有 k 次差分值都為 0，代表數據點構成一條 $(k-1)$ 次多項式

- In practice, if the k-th divided differences are close to 0, we can consider fitting the dataset using a $(k-1)$-th degree polynomial
  實際上，如果所有 k 次差分值都已接近 0，我們可以考慮使用 $(k-1)$ 次多項式

x_i	0	2	4	6	8
y_i	0	4	16	36	64

$y = x^2$

	Data		Divided differences		
	x_i	y_i	Δ	Δ^2	Δ^3
	0	0			
			$4/2 = 2$		
$\Delta x = 6$	2	4		$4/4 = 1$	
			$12/2 = 6$		$0/6 = 0$
	4	16		$4/4 = 1$	
			$20/2 = 10$		$0/6 = 0$
	6	36		$4/4 = 1$	
			$28/2 = 14$		
	8	64			

Method of divided difference 差分法

- Example: Find a suitable polynomial model to represent the stopping distance as a function of the speed of the car
  例子： 找出一個合適的多項式模型來表示停車距離與汽車速度的關係

Speed v (mph)	20	25	30	35	40	45	50	55	60	65	70	75	80
Distance d (ft)	42	56	73.5	91.5	116	142.5	173	209.5	248	292.5	343	401	464

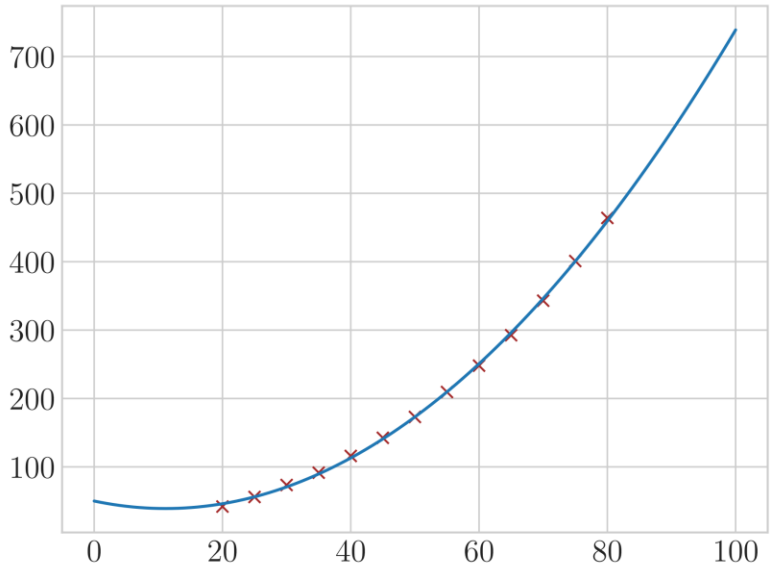
Method of divided difference 差分法

- Example: Find a suitable polynomial model to represent the stopping distance as a function of the speed of the car
  例子： 找出一個合適的多項式模型來表示停車距離與汽車速度的關係

Speed v (mph)	20	25	30	35	40	45	50	55	60	65	70	75	80
Distance d (ft)	42	56	73.5	91.5	116	142.5	173	209.5	248	292.5	343	401	464

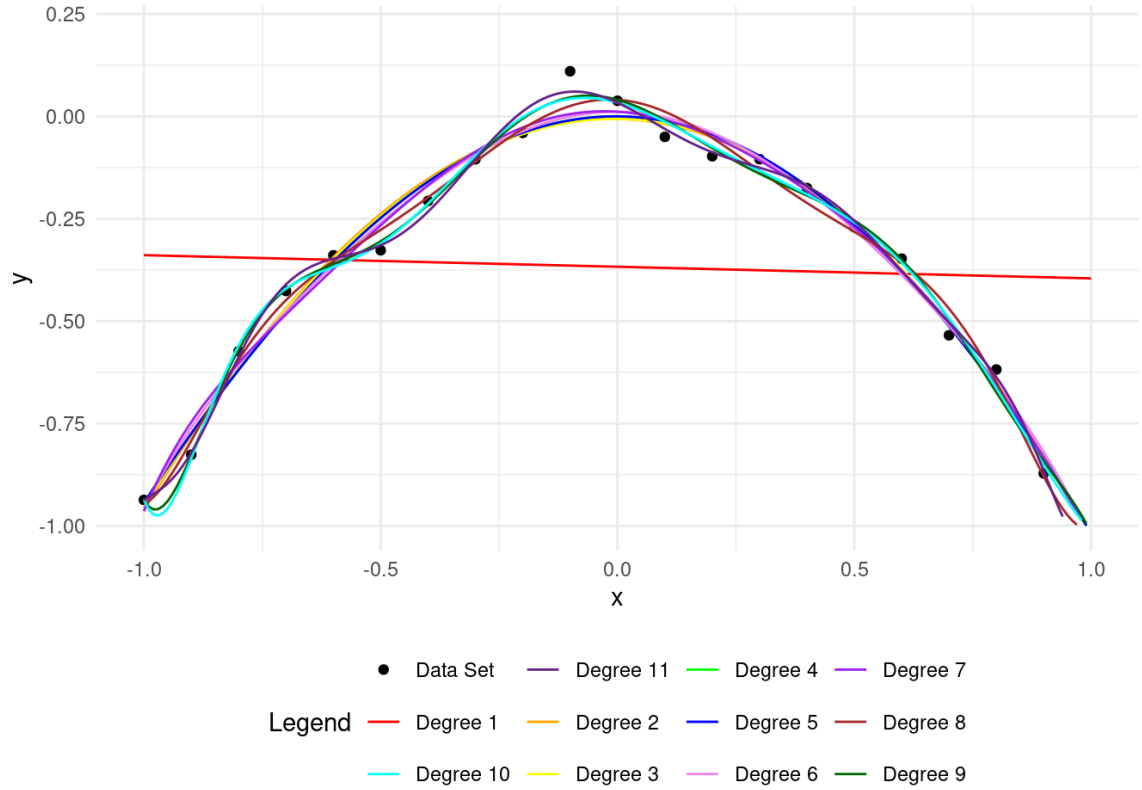
Data		Divided differences			
v_i	d_i	Δ	Δ^2	Δ^3	Δ^4
20	42				
		2.2800			
25	56		0.0700		
		3.5000		−0.0040	
30	73.5		0.0100		0.0006
		3.6000		0.0080	
35	91.5		0.1300	−0.0060	−0.0007
		4.9000			
40	116		0.0400	0.0027	0.0004
		5.3000			
45	142.5		0.0800	0.0027	0.0000
		6.1000			
50	173		0.1200	−0.0053	−0.0004
		7.3000			
55	209.5		0.0400	0.0053	0.0005
		7.7000			
60	248		0.1200	0.0000	−0.0003
		8.9000			
65	292.5		0.1200	0.0020	0.0001
		10.1000			
70	343		0.1500	−0.0033	−0.0003
		11.6000			
75	401		0.1000		
		12.6000			
80	464				

$$d = 0.0886\,v^2 - 1.9701v + 50.0594$$



Model validation 模型驗證

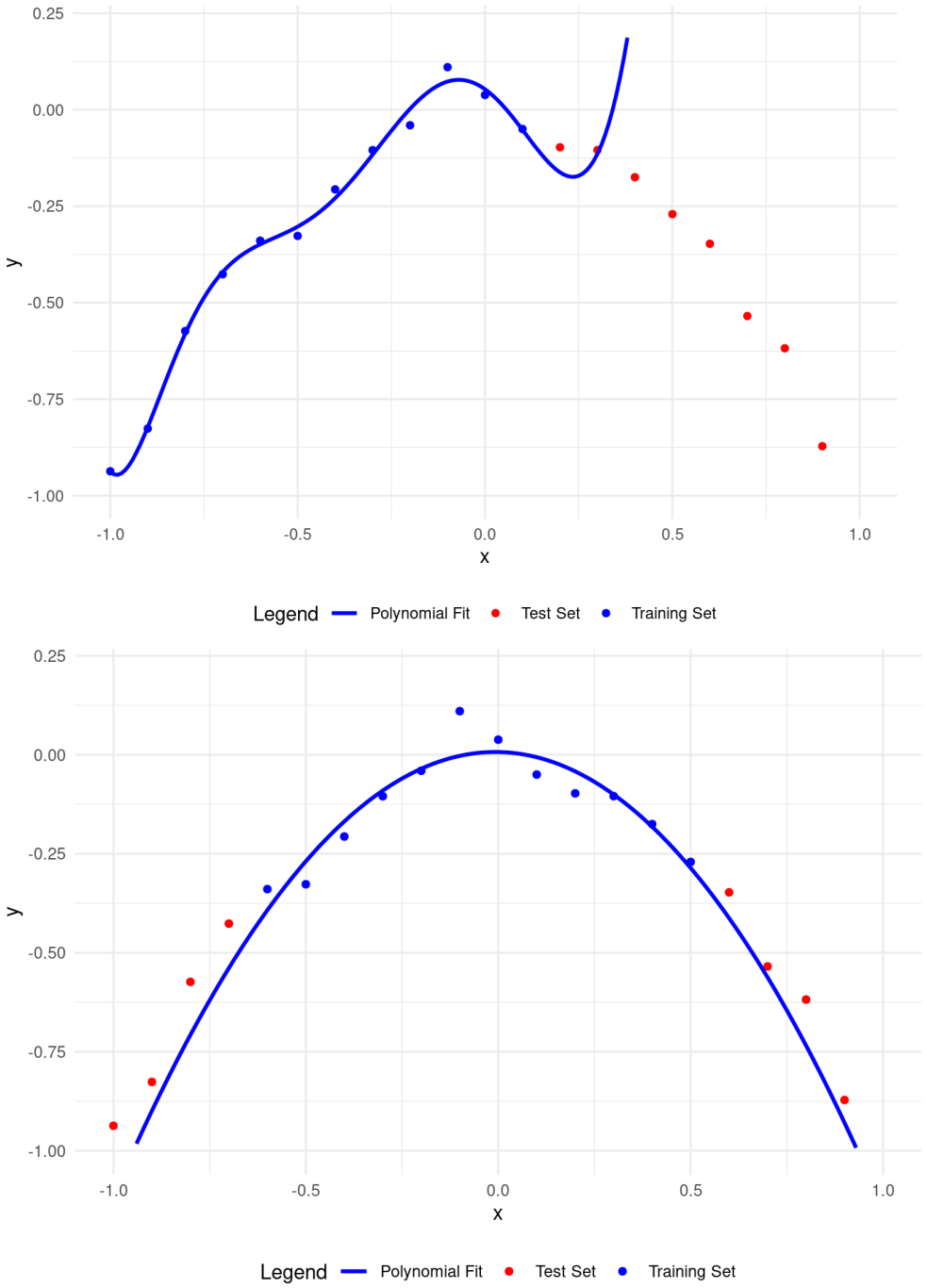
- If we construct a best-fit model based on all data points, overfitting may occur:
  如果我們根據所有數據點建立最佳擬合模型則可能會發生過度擬合：
  - very accurate with the training data
    在現有數據上非常準確
  - but loses accuracy with new data
    但在新數據上會失去準確性



Model validation 模型驗證

- Solution: Divide the dataset into
  解決方法：將數據集劃分為
  - Training data 訓練數據
    - For constructing a best-fit model
      建構最佳擬合模型
  - Testing data 測試數據
    - For testing whether the model gives good accuracy when handling new data
      用於測試模型在處理新資料時是否具有良好的準確性

- Overfitting and Underfitting with Polynomial Regression 多項式迴歸的過度擬合與欠擬合
  http://mathcal.math.cuhk.edu.hk:7542/



Evaluating the accuracy 準確性的評估方法

- **Coefficient of determination 決定係數 (R^2):**

$$R^2 = 1 - \frac{RSS}{TSS}$$

  - RSS is the residual sum of squares RSS 是殘差平方和:

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \cdots + (y_n - \hat{y}_n)^2$$

    with $\hat{y}_i$ being the predicted value based on the chosen model
    其中 $\hat{y}_i$ 是基於所選模型的預測值

  - TSS is the total sum of squares TSS 是總平方和:

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y})^2 = (y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \cdots + (y_n - \bar{y})^2$$

    with $\bar{y} = \frac{1}{n}(\sum_{i=1}^{n} y_i) = \frac{y_1 + y_2 + \cdots + y_n}{n}$ being the mean of the given data
    其中$\bar{y} = \frac{1}{n}(\sum_{i=1}^{n} y_i) = \frac{y_1 + y_2 + \cdots + y_n}{n}$ 是已有數據的平均值

Evaluating the accuracy 準確性的評估方法

- **Coefficient of determination 決定係數 (R^2):**

$$R^2 = 1 - \frac{RSS}{TSS}$$

- A model exactly matching all observed values will give $R^2 = 1$
  如果一個模型完全吻合所有觀測值， 則 $R^2 = 1$

- A baseline model $y = \bar{y}$ (which always predicts $\bar{y}$ regardless of the value of x) will give $R^2 = 0$
  一個基準模型 $y = \bar{y}$（無論 x 值為何，永遠預測平均值 $\bar{y}$ ）會得到 $R^2 = 0$

- R^2 can be negative if the chosen model "fits worse than a horizontal line"!
  如果所選模型的擬合「比一條水平線還差」， R^2會出現負值！

Evaluating the accuracy 準確性的評估方法

- **Mean squared error 均方誤差 (MSE) :**

$$MSE = \frac{RSS}{n} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

  - Value can be from 0 to ∞ 數值可以為 0 到 ∞
  - Easy to compute 易於計算

- **Root mean squared error 均方根誤差 (RMSE):**

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

  - Value can be from 0 to ∞ 數值可以為 0 到 ∞
  - With the square root, the error is in the same units as the original data (hence more intuitive, but relatively hard to compute)
    加上平方根後，誤差的單位與原始數據相同（因此更直觀，但相對難以計算）

Evaluating the accuracy 準確性的評估方法

- **Mean absolute error 平均絕對誤差 (MAE):**

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

  - Here, $|x|$ is the absolute value (絕對值) of x: $|x| = \begin{cases} x \text{ if } x \geq 0 \\ -x \text{ if } x < 0 \end{cases}$
  - Value can be from 0 to ∞ 數值可以為 0 到 ∞
  - The error is in the same units as the original data 誤差的單位與原始數據相同

- **Mean absolute percentage error 平均絕對百分比誤差 (MAPE)**

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%$$

  - Value can be from 0% to 100%, hence giving an intuitive interpretation in terms of relative error 數值可以從 0% 到 100%，因此在相對誤差方面提供直觀的解釋
  - But small or close-to-zero values of some y_i may disproportionately affect the MAPE score 但某些 y_i 的值很小或接近零時，可能會不成比例地影響 MAPE 值

Exhaustive cross-validation strategies 窮舉交叉驗證

- **Leave-one-out cross validation 留一交叉驗證 (LOOCV)**

- Select one candidate form of model 選擇一種候選模型

- For n data points 對於 n 個數據點：
  - Split the dataset into training (all but the i-th data point) and testing set (the i-th data point)
    將數據集分割為訓練集（除了第 i 個數據點的所有點）和測試集（僅第 i 個數據點）
  - Train the model using the training set 使用訓練集來訓練模型
  - Test it using the testing set 使用測試集來測試模型
  - Evaluate the accuracy 評估準確度
  - Repeat the above processes for $i = 1,2,\ldots,n$ 重複上述過程，對 $i = 1,2,\ldots,n$ 逐一進行
  - Summarize the accuracy 總結準確度

Exhaustive cross-validation strategies 窮舉交叉驗證

- **Leave-p-out cross validation 留 p 交叉驗證 (LpOCV)**
  - Similar to LOOCV 類似於LOOCV
  - but consider all C_p^n combinations of selecting p data points from n data points
    但考慮從 n 個資料點中選擇 p 個資料點的所有組合 C_p^n

- Advantage: Less affected by outliers 優點：受異常值的影響較小

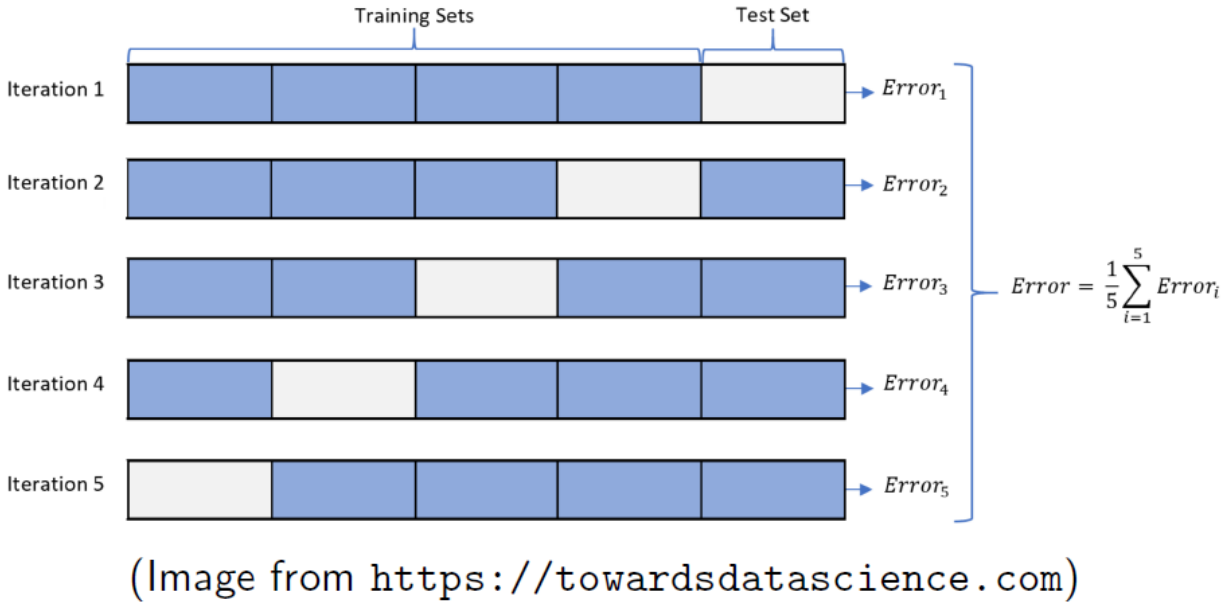
- Disadvantage: More computationally expensive 缺點：計算成本較高

- Example: 例如：
  - For a dataset with100 points, every time we select 2 data points as testing set
    對於一個包含 100 個數據點的數據集，每次選擇 2 個數據點作為測試集
  - Then we have $C_2^{100} = \frac{100\times99}{2} = 4950$ combinations!
    那我們就有 $C_2^{100} = \frac{100\times99}{2} = 4950$ 種組合！

Non-exhaustive cross-validation strategies非窮舉交叉驗證

- **k-fold cross validation k 折交叉驗證**
  - The original dataset is randomly reordered and partitioned into k equally-sized (or as-equal-as-possible) blocks
    將數據集隨機排列及分成 k 個大小相等（或盡可能相等）的組別

  - For $i = 1, \ldots, k$, 對於 $i = 1, \ldots, k$,
    - Train the model on all the data except block i. 用除了第 i 組以外的所有數據訓練模型。
    - Evaluate the model (compute the model error) using block i.
      使用第 i 組評估模型（計算誤差）

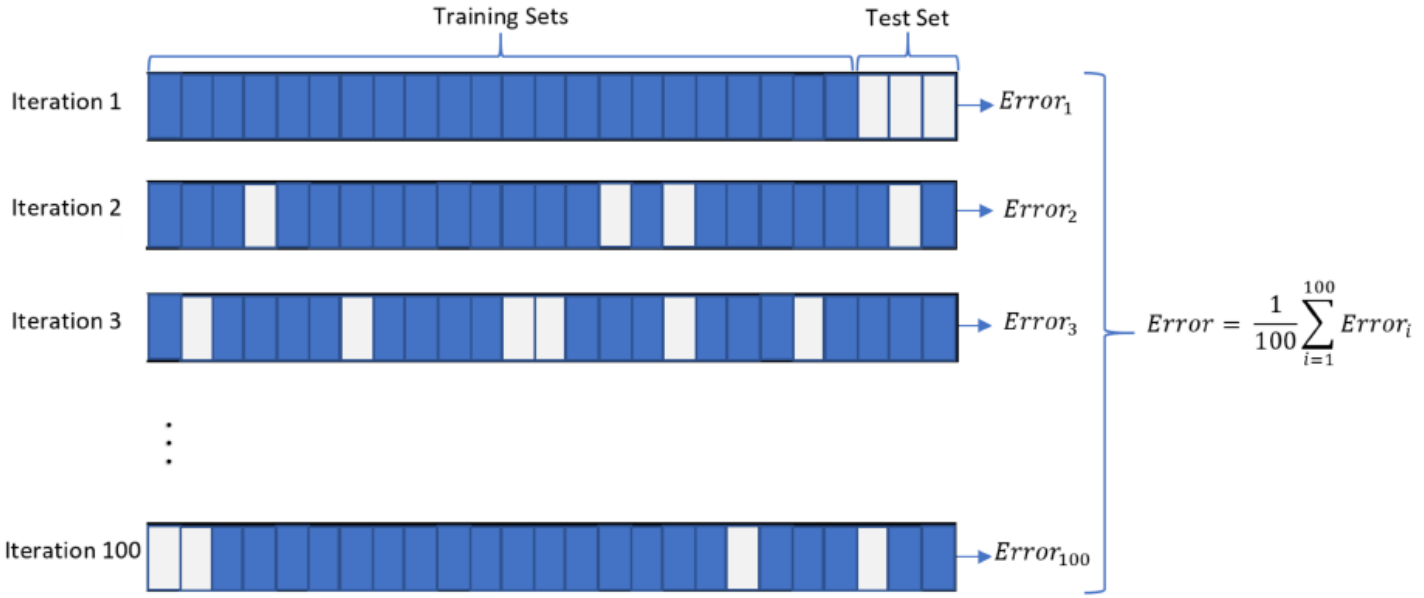


(Image from https://towardsdatascience.com)

  - Take the average of all k model errors.
    計算所有 k 個模型誤差的平均值。

Non-exhaustive cross-validation strategies非窮舉交叉驗證

- **Monte Carlo cross-validation 蒙特卡羅交叉驗證：**
  - Randomly select (without replacement) a certain percentage of the data to form the training set and then assign the rest of the points to the test set.
    隨機抽取（不放回）一定比例的數據構成訓練集，其餘數據構成測試集。
  - Train the model based on the training set and evaluate the model error using the test set. 基於訓練集訓練模型，並使用測試集評估模型誤差。
  - Repeat the above multiple times
    重複上述步驟多次
  - Take the average of all model errors
    計算所有模型誤差的平均值



(Image from https://towardsdatascience.com)

Another consideration in model evaluation: Sensitivity analysis
模型評估的另一個考慮因素：敏感度分析

- In reality, there may be noise in the input data, fluctuations in different factors, …
  實際上，輸入數據中可能存在噪聲，不同因素也可能出現波動

- Besides considering the accuracy of a model using validation methods, we can also evaluate the model's **stability** by performing **Sensitivity Analysis**:
  除了使用驗證方法評估模型的準確性之外，我們還可以透過進行**敏感度分析**來評估模型的**穩定性**：
  - See how much the output will change when the inputs or parameters change
    觀察當輸入或參數改變時，輸出會發生多大的變化
  - i.e., running "what-if" scenarios
    運行"假設分析"場景

Another consideration in model evaluation: Sensitivity analysis
模型評估的另一個考慮因素：敏感度分析

- Common sensitivity analysis methods: 常用的敏感度分析方法：
  - **One-at-a-Time (OAT):** Change one input variable/factor while holding others constant to see the effect on the output.
    一次一變(OAT)：改變一個輸入變數/因素，同時保持其他不變，以觀察對輸出的影響。

  - **Scenario Analysis**: Testing predefined sets of input changes (e.g. best-case, worst-case)
    情境分析：測試給定的輸入變化集合（例如最佳情況、最壞情況）

  - **Monte Carlo Simulation**: Using random sampling to test numerous input combinations to understand a range of possible outcomes and their likelihood.
    蒙地卡羅模擬：使用隨機抽樣測試眾多輸入組合，以了解可能結果的範圍及其可能性。

Modelling Using Multiple Regression 多元迴歸

- What if there are multiple factors to be considered? 如果有多個因素需要考慮怎麼辦？
- How to determine whether a factor is important? 如何判斷一個因素是否重要？
- Example:
  How does advertisement in different media channels (TV, radio, newspaper) affect the product sales?
  不同媒體（電視、廣播、報紙）上的廣告如何影響產品銷售？

TV	radio	newspaper	sales
230.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	9.3
151.5	41.3	58.5	18.5
180.8	10.8	58.4	12.9
8.7	48.9	75.0	7.2

- **Multiple regression**: we can consider
  多元迴歸：我們可以考慮

$$y = f(x_1, x_2, \ldots, x_d) = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_d x_d$$

and minimize 並最小化

$$\text{RSS} = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_{i,1} - \cdots - a_d x_{i,d})^2$$

Modelling Using Multiple Regression 多元迴歸

- Approaches 方法:
  - Consider each variable separately
    分別考慮每個變數

$$\text{sales} = \beta_0 + \beta_1 \text{TV}$$

  - Including the variables 包括所有變數

$$\text{sales} = \beta_0 + \beta_1 \text{TV} + \beta_2 \text{radio} + \beta_3 \text{newpaper}$$

TV	radio	newspaper	sales
230.1	37.8	69.2	22.1
44.5	39.3	45.1	10.4
17.2	45.9	69.3	9.3
151.5	41.3	58.5	18.5
180.8	10.8	58.4	12.9
8.7	48.9	75.0	7.2

  - Including all variables and their interaction terms 包括所有變數及它們互相的影響

$$\begin{aligned}\text{sales} = \beta_0 &+ \beta_1 \text{TV} + \beta_2 \text{radio} + \beta_3 \text{newpaper} \\ &+ \beta_4 (\text{TV} \cdot \text{radio}) + \beta_5 (\text{TV} \cdot \text{newpaper}) + \beta_6 (\text{radio} \cdot \text{newpaper})\end{aligned}$$

  - And many more possibilities 還有更多可能性

- Detailed description 詳細說明: http://mathcal.math.cuhk.edu.hk:7541/

Modelling Using Multiple Regression 多元迴歸

- **Interpretation of importance of different factors via multiple regression**:
  透過多元迴歸分析不同因素的重要性：

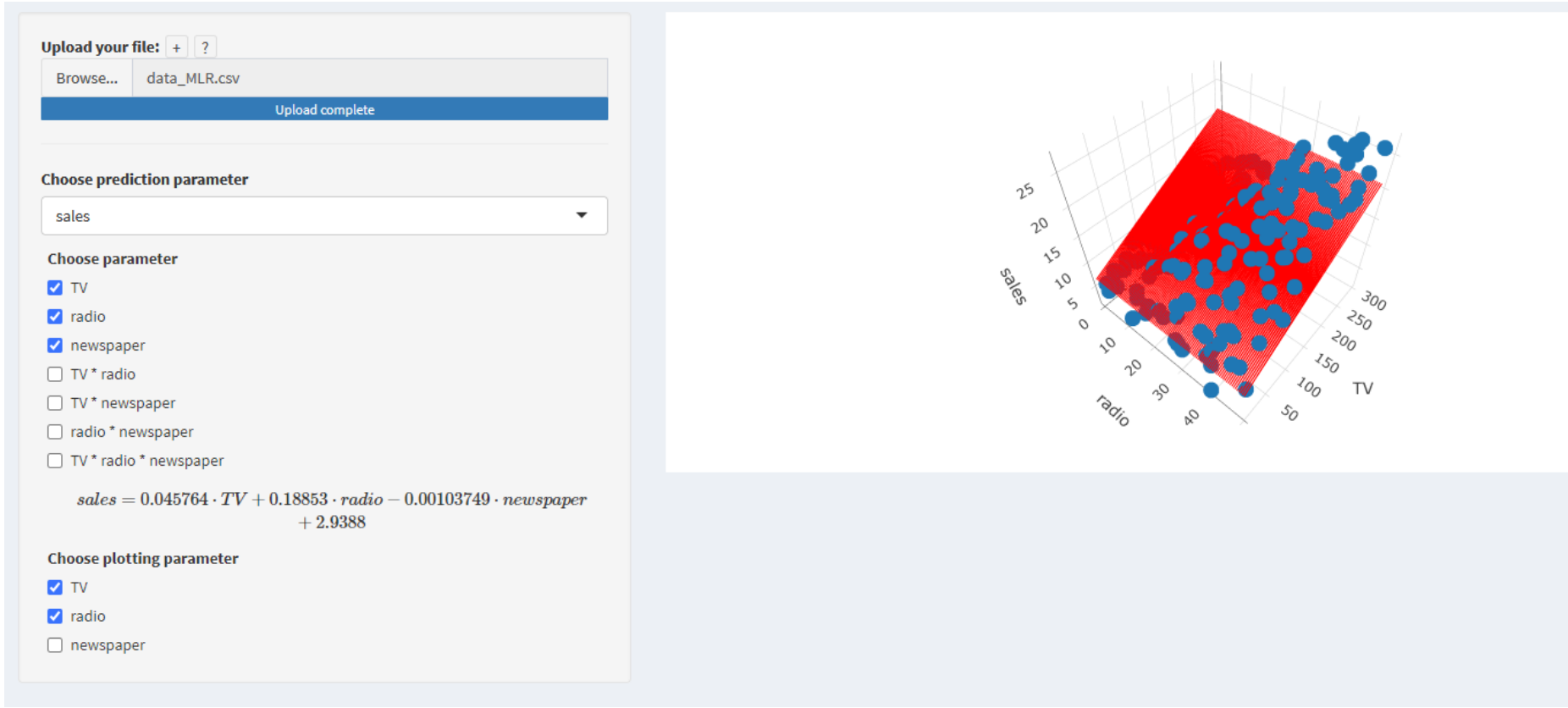
  - We first standardize all variables (i.e., transform the data values for each variable to have a mean of 0 and a standard deviation of 1) and then solve for the best-fit multiple regression model
    首先，我們將所有變數進行標準化處理（即將每個變數的資料值轉換為平均值為 0、標準差為 1 的平均值），然後求解最佳擬合的多元迴歸模型
  - The magnitude of the resulting coefficients can indicate the relative importance of the variables 所得系數的大小可以表示變數的相對重要性
  - Can be used for feature selection or model simplification
    可用於特徵選擇或模型簡化

- **Multi-Regression R Shiny Tool 多元迴歸 R Shiny 工具**
  https://mathmodelcuhk.shinyapps.io/multi-regression/

Multi-Regression R Shiny Tool 多元線性迴歸 R Shiny工具

- https://mathmodelcuhk.shinyapps.io/multi-regression/
  - Performing multiple linear regression 進行多元線性迴歸
  - Input data format: CSV, XLSX, or TXT file containing the data points (at least 2 columns of data) 輸入數據格式：包含數據點的 CSV、XLSX 或 TXT 檔案（最少 2 列資料）
  - Sample data file 參考數據檔案: [data_MLR.csv]

	A	B	C	D
1	TV	radio	newspape	sales
2	230.1	37.8	69.2	22.1
3	44.5	39.3	45.1	10.4
4	17.2	45.9	69.3	9.3
5	151.5	41.3	58.5	18.5
6	180.8	10.8	58.4	12.9
7	8.7	48.9	75	7.2
8	57.5	32.8	23.5	11.8
9	120.2	19.6	11.6	13.2
10	8.6	2.1	1	4.8
11	199.8	2.6	21.2	10.6
12	66.1	5.8	24.2	8.6
13	214.7	24	4	17.4
14	23.8	35.1	65.9	9.2
15	97.5	7.6	7.2	9.7
16	204.1	32.9	46	19
17	195.4	47.7	52.9	22.4
18	67.8	36.6	114	12.5



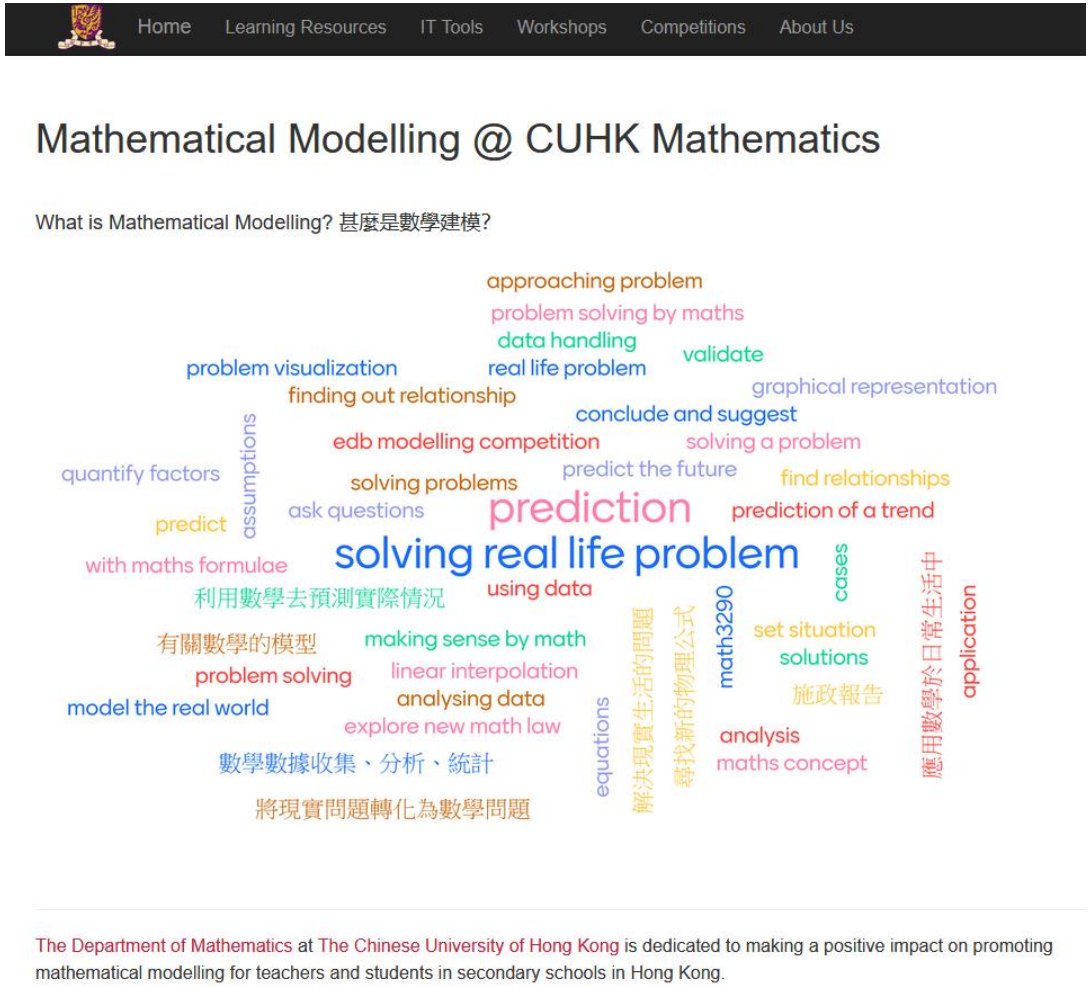
Mathematical Modelling Resources 數學建模資源

- **Mathematical Modelling @ CUHK Mathematics** website 網頁: https://www.math.cuhk.edu.hk/app/mathmodel/

- Learning Resources 學習資源
  - E-book 電子書
  - Exercises 練習
  - Workshop materials 工作坊資源

- IT tools for math modelling 數學建模的IT工具
  - Computing and visualization 計算與可視化
  - AI 人工智能

- A large variety of real-life examples 不同現實生活例子



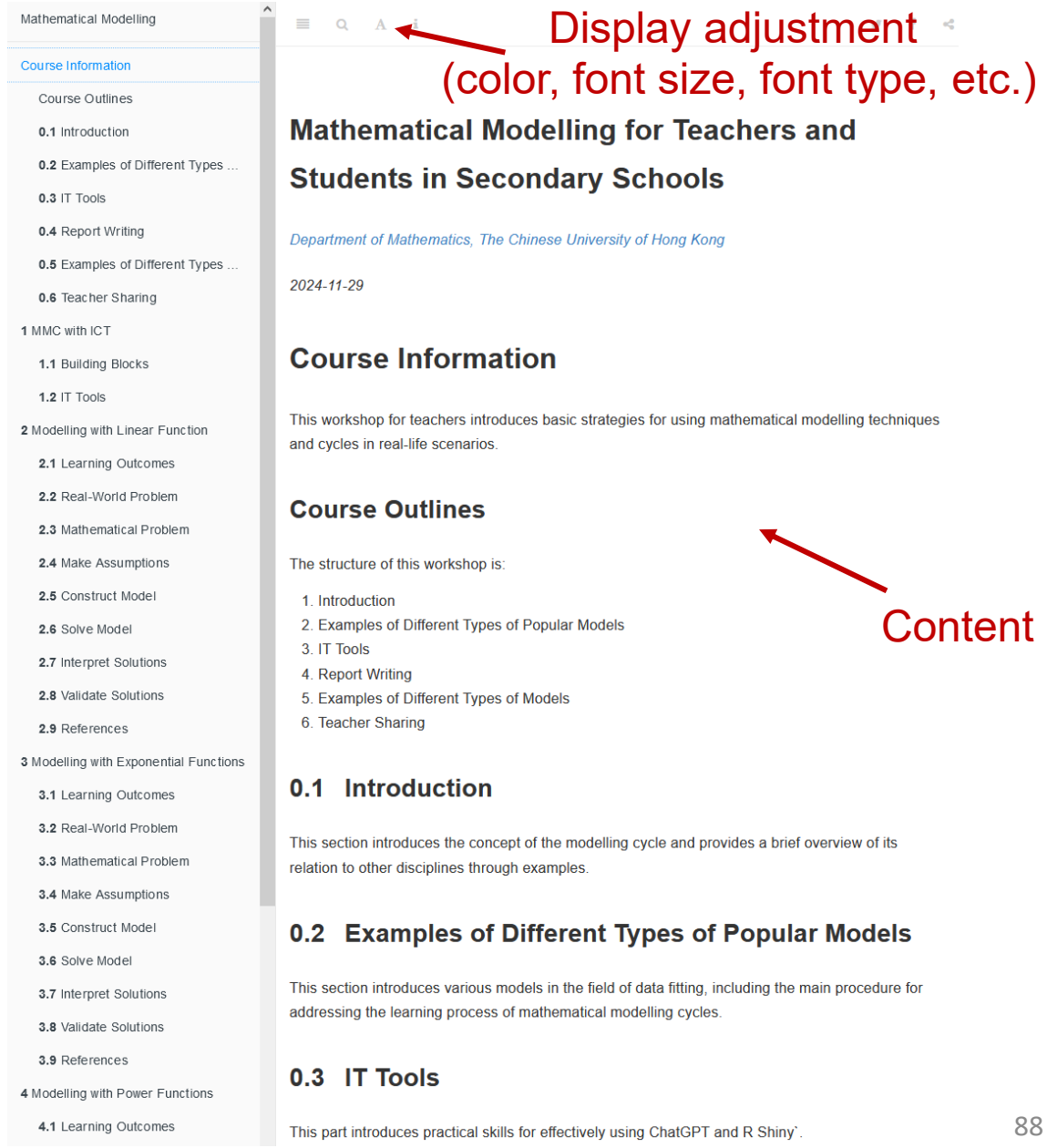
Mathematical Modelling Resources 數學建模資源

- **Mathematical Modelling e-book:**
  https://www.math.cuhk.edu.hk/~mathcal/MM/
  Username: mathmodel
  Password: mm@2024

- Mathematical concepts and derivations
  數學概念及推導

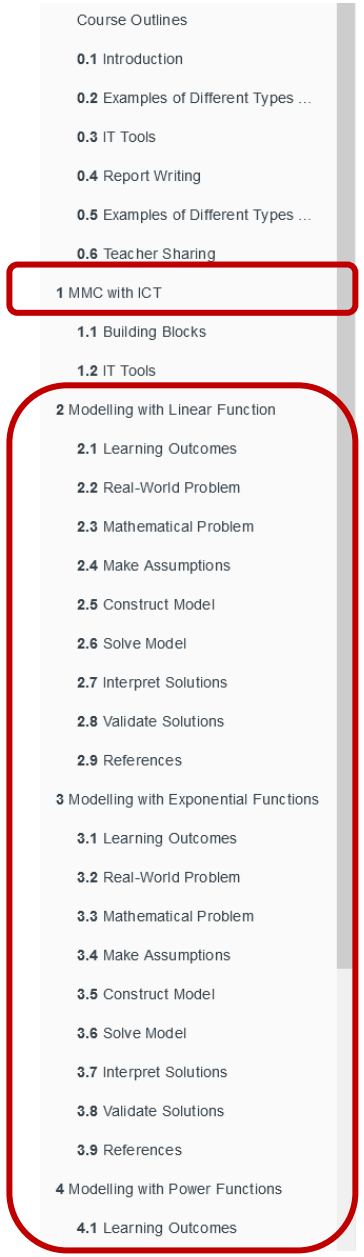
- IT tool usage IT 工具使用

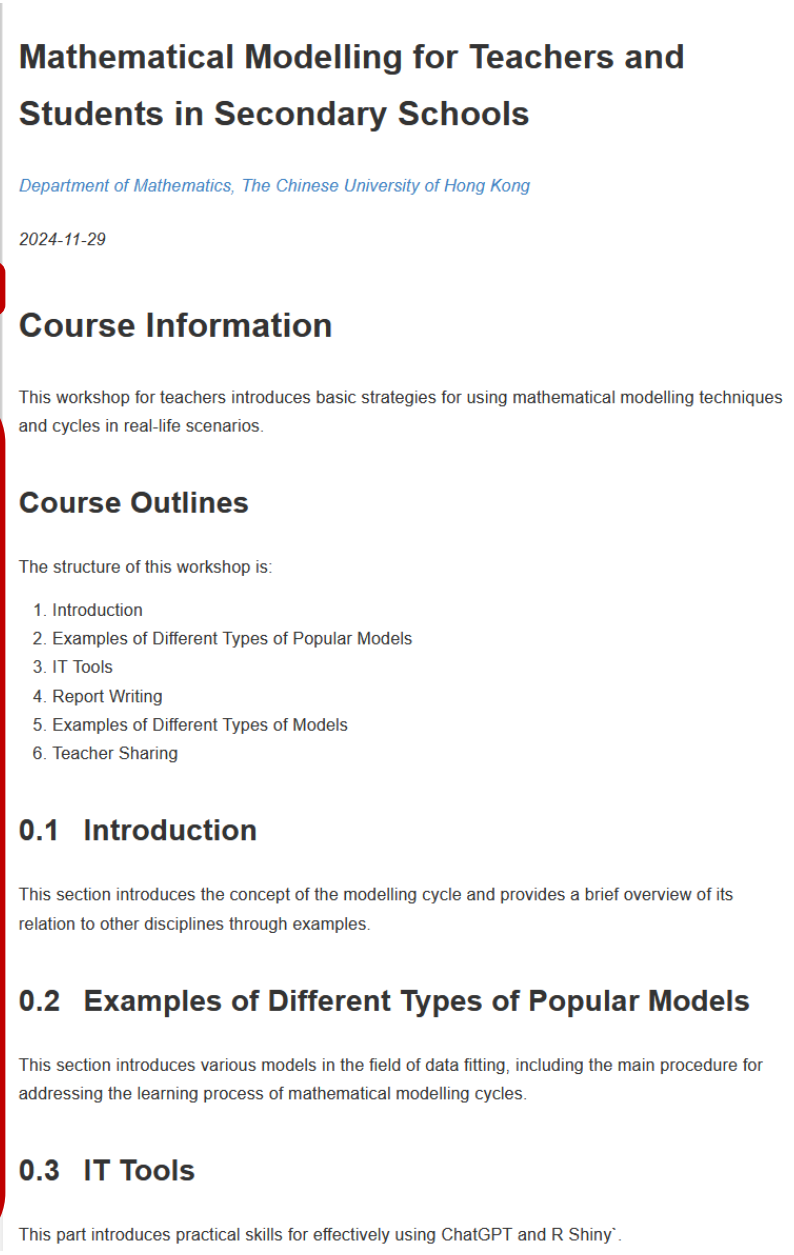
- Different examples 不同範例



Mathematical Modelling Resources 數學建模資源

- **Mathematical modelling concepts 數學建模概念**
  - Math modelling process 數學建模過程
  - Key steps關鍵步驟
- **Common models 常用模型：**
  - Linear functions線性函數
  - Exponential functions指數函數
  - Power functions冪函數
  - Trigonometric functions三角函數
  - Sigmoidal functions S形函數
  - All with detailed formulations, derivations, and examples 均提供詳細的公式、推導過程和範例
- **More advanced math tools 進階數學工具:**
  - Probability 概率, network model 網絡模型, …
  - Introduced via specific real-life problems (price prediction, social network etc.) 透過具體問題（價格預測、社交網路等）介紹





Mathematical Modelling for Teachers and Students in Secondary Schools

Department of Mathematics, The Chinese University of Hong Kong

2024-11-29

Course Information

This workshop for teachers introduces basic strategies for using mathematical modelling techniques and cycles in real-life scenarios.

Course Outlines

The structure of this workshop is:

1. Introduction
2. Examples of Different Types of Popular Models
3. IT Tools
4. Report Writing
5. Examples of Different Types of Models
6. Teacher Sharing

0.1 Introduction

This section introduces the concept of the modelling cycle and provides a brief overview of its relation to other disciplines through examples.

0.2 Examples of Different Types of Popular Models

This section introduces various models in the field of data fitting, including the main procedure for addressing the learning process of mathematical modelling cycles.

0.3 IT Tools

This part introduces practical skills for effectively using ChatGPT and R Shiny`.

Mathematical Modelling @ CUHK Mathematics:
https://www.math.cuhk.edu.hk/app/mathmodel

Contact:
mathmodel@math.cuhk.edu.hk

Thank you!