2024 年【科學探究競賽-這樣教我就懂】

普高組 成果報告表單

題目名稱：利用 Desmos 數學軟體進行函數擬合

一、摘要

本題目發想自高中數學課程：二維數據分析的單元內，有提及如何計算迴歸直線，也就是線性擬合，本團隊思考該怎樣的手段才能進行曲線擬合。後續本團隊先在數學軟體 Desmos 上學習如何給定散佈點進行函數擬合。後續在已知的五個監測點歷經 15 週的懸浮汙染物濃度的測量值，本研究團隊透過曲線擬合的技術，透過根式/分式函數與指數函數的伸縮與平移，成功擬合出兩種 $R^2=1$ 的函數，並用以估計第 20 週與第 30 週時湖中汙染物濃度。透過兩種不同的曲線擬合的估計結果相當接近，依第一週時五個監測點的數據得知湖中汙染物濃度平均值為 10.83592 mg/L。經過本研究團隊的估計，第 20 週的汙染濃度應介於 2.941 mg/L ~ 2.959 mg/L，第 30 週的汙染濃度應介於 2.777 mg/L ~ 2.812 mg/L。

二、探究題目與動機

在高中的數學課程中，曾在二維數據分析的單元內，學習如何給出數個散佈點進而計算其迴歸直線以及其相關係數，也就是線性迴歸擬合，但是本團隊思考是否有方法可以找出一個函數將所有的散佈點都通過，也就是曲線擬合。本團隊使用 2024 第十屆國際數學建模挑戰賽（IMMC）大中華區秋季賽 A 題的數據進行研究，題目中提出江蘇省地區之太湖，由於太湖受上游造紙工廠排放廢水導致湖中水資源有嚴重的汙染問題。本研究期待參考題目中所給的 15 週湖泊內汙染物濃度的數據，進行曲線擬合推測第 20 週及第 30 週的湖泊汙染物濃度，並學習如何進行曲線擬合。

三、探究目的與假設

(一) 學習如何利用數學軟體 Desmos 進行曲線擬合。

(二) 若上游造紙廠每週排汙流速和濃度不變，利用表 (一) 的數值建立合適的數學模型，預測第 20 週和第 30 週的汙染物濃度。

表 (一) 造紙懸浮顆粒污染物濃度測量值 (濃度單位 mg/L)

監測時間	監測點 1	監測點 2	監測點 3	監測點 4	監測點 5
第 1 週	10.8366	10.8286	10.8418	10.8313	10.8413
第 2 週	9.35012	9.33735	9.36265	9.36667	9.36858
第 3 週	8.16433	8.16756	8.1424	8.13454	8.15016
第 4 週	7.1481	7.16879	7.17691	7.16098	7.15679
第 5 週	6.3267	6.35153	6.33866	6.38672	6.33468
第 6 週	5.64729	5.69483	5.68283	5.72216	5.72069
第 7 週	5.15103	5.18964	5.19767	5.1787	5.18579
第 8 週	4.69592	4.72477	4.69179	4.7502	4.76333
第 9 週	4.42042	4.32338	4.3661	4.3312	4.35653
第 10 週	4.0741	4.02744	4.14201	4.02247	4.09824
第 11 週	3.82257	3.8737	3.88571	3.85653	3.86492
第 12 週	3.66742	3.59325	3.64282	3.66924	3.69893
第 13 週	3.50337	3.47648	3.5374	3.48706	3.54225
第 14 週	3.31069	3.31237	3.34224	3.34739	3.30775
第 15 週	3.26862	3.24237	3.26983	3.27488	3.28471

四、探究方法與驗證步驟

(一) 首先將欲擬合的數據輸入到 desmos 數學軟體上，並且打上擬合的指令，以下舉出四種不同的函數圖形的指令寫法。如果想透過線性函數圖形進行擬合，則指令需打上 $y_1 \sim x_1$；若想透過對數函數進行擬合，那指令需打上 $y_1 \sim \log_a x_1$；若想透過根式/分式函數進行擬合，則指令需打上 $y_1 \sim x_1^{\,a}$；若想透過指數函數函數進行擬合，則指令需打上 $y_1 \sim a^{x_1}$，若想使得 R^2 提高，將函數進行伸縮與平移，進而優化擬合結果。

(二) 學習完如何曲線擬合後，由於賽題給了 5 個觀測點測得的數據，且賽題中並未提到觀測點的相對位置。因此本研究先計算每一週在 5 個觀測點中測得汙染物濃度的平均值，並以此數據建立數學模型，預測第 20 週和；第 30 週的汙染物濃度。

表 (二) 造紙懸浮污染物濃度平均值 (濃度單位 mg/L)

監測時間	平均值	監測時間	平均值	監測時間	平均值
第 1 週	10.83592	第 6 週	5.69356	第 11 週	3.86069
第 2 週	9.35707	第 7 週	5.18057	第 12 週	3.65433
第 3 週	8.15180	第 8 週	4.72520	第 13 週	3.50931
第 4 週	7.16231	第 9 週	4.35953	第 14 週	3.32409
第 5 週	6.34766	第 10 週	4.07285	第 15 週	3.26808

透過表 (二) 之數值，將表中的數值改為二維數據放置在直角坐標平面上。其中橫軸為時間軸 (單位：週)；縱軸代表濃度 (單位：mg/L)。透過 Geogebra 數學軟體造出 15 個散佈點如圖 (一)。由右圖可以明顯觀察出 15 個散佈點的濃度，隨著時間的推移有著遞減的趨勢。此外，由於時間為連續變量，若假想將此 15 個散佈點透過一個曲線連接，顯然會是一個非線性且凹向上的遞減函數圖形，視覺感受上甚至有著當時間遞增時，濃度會收斂的狀態。為此，本研究希望可以找到適當的函數進行擬合，期待能同時達到上述的幾個觀察到的條件與特色：1. 凹口向上且遞減。2. 時間遞增時濃度收斂。3. 盡可能使得已有的 15 個散佈點都在擬合函數圖形上，即曲線擬合用於統計分析時常見的 $r^2 \approx 1$ (廖培凱，2023)。本研究將曲線擬合函數分為以下四種討論：

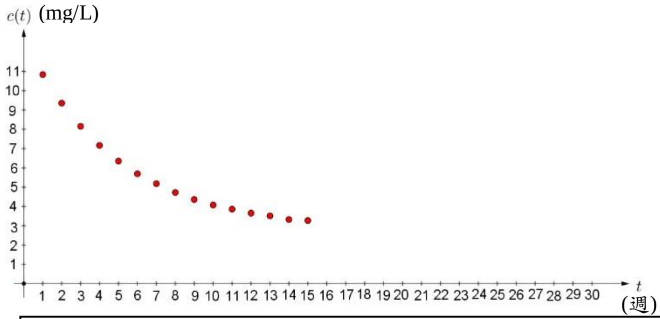

圖 (一) 造紙懸浮污染物濃度平均值散佈圖

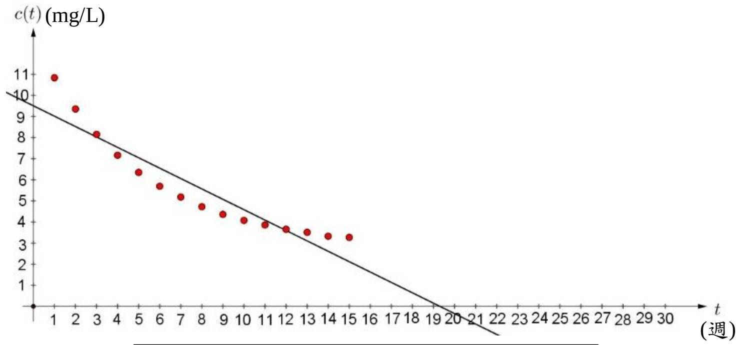

圖 (二) 1 次多項式函數擬合

(一) 多項式函數

本研究先利用 $y = at + b$ 函數進行擬合，擬合出的函數圖形如圖 (二)，其方程式為 $y = -0.492645\,t + 9.50802$ 。顯然此方程式即為迴歸直線，其中 $r^2 = 0.8752$。由於 r^2 的值與 1 相差甚遠，因此本篇研究不考慮以 1 次多項式函數進行擬合用以估計。由代數基本定理可知，15 個點必可唯一決定一個 14 次多項式函數，因此本研究在利用多項式函數進行擬合時，從 2

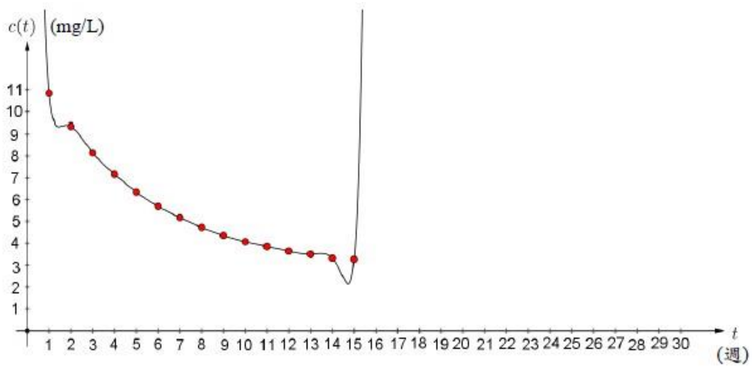

圖 (三) 14 次多項式函數擬合

次、3 次多項式函數、...、最高討論到 14 次多項式函數。在上述的擬合過程中，本研究將各個不同次數的多項式函數擬合結果繪製出來都有發現其中不合理之處。以圖 (三) 為例，由於圖 (三) 是 14 次多項式函數圖形，顯然其擬合結果會通過所有已知的 15 個點，即 $r^2=1$。但觀察 $t>15$ 時發現，函數圖形呈現相當陡峭的遞增，顯然這樣的函數雖然 $r^2=1$，但並不適合用以估計後續各個時間點的觀測濃度。因此本研究不考慮多項式函數用以估計。

(二) 對數函數

對數函數 $y=\log_a t$，其中 $a,t>0$，$a\neq 1$，若底數 $0<a<1$，則會是遞減且凹向上的函數圖形，為了讓函數擬合結果的 r^2 足夠接近於 1，因此本研究將對數函數做伸縮及平移，進行擬合。

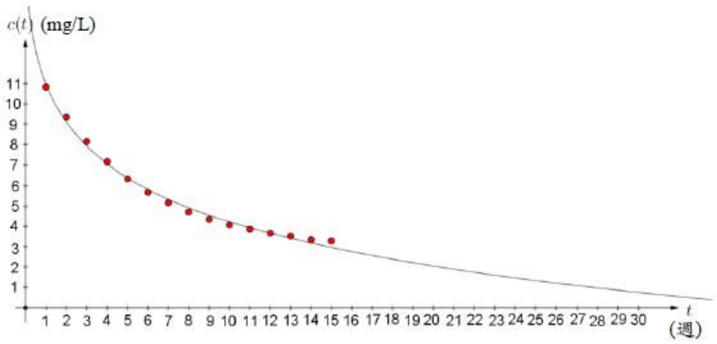

圖 (四) 對數函數擬合

本研究以 $y=a\log_b(ct+d)+f$ 函數進行擬合，擬合出的函數圖形如圖 (四)，其方程式為 $y=-9.4618\log_{18.9}(29.532t+7.62744)+22.6269$，且 $r^2=0.9943$。觀察圖 (四) 會發現此函數雖為凹向上遞減，但濃度值在遞減過程中並不會有收斂的效果。因此本篇研究不考慮以對數函數進行擬合用以估計。

(三) 根式/分式函數

根式/分式函數 $y=t^a$，若指數 $a<0$ 且 a 為整數時，稱此函數為分式函數。若指數 $a<0$ 且 a 為非整數之有理數時，稱此函數為根式函數。此類函數在第一象限內會是遞減且凹向上的圖形。為了讓擬合函數的 r^2 足夠接近於 1，本研究透過數學軟體 Desmos，將已知的 15 個點坐標輸入形成點坐標表格 (x_1, y_1)，接著透過指令 $y_1\sim a(bx_1+c)^d+f$ 進行擬合，擬合出的函數圖形如圖 (五)，其方程式為

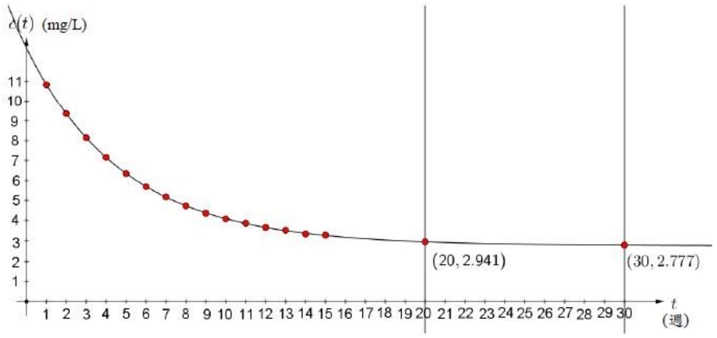

圖 (五) 根式/分式函數擬合

$y=6.8308\times 10^{14}(0.00723072\,t+1.8432)^{-52.1049}+2.74659$，且 $r^2=1$。透過此一函數與直線 $t=20$ 和 $t=30$ 分別交於 $(20, 2.941)$、$(30, 2.777)$。因此本研究以根式函

數預測第 20 週和第 30 週的汙染物濃度，分別為 2.941 mg/L 和 2.777 mg/L。

(四) 指數函數

指數函數 $y = a^x$，若底數 $0 < a < 1$，則會是遞減且凹向上的函數圖形，為了讓擬合函數的 r^2 足夠接近於 1，本研究透過數學軟體 Desmos，將已知的 15 個點坐標輸入形成點坐標表格 (x_1, y_1)，接著透過指令 $y_1 \sim ab^{x_1} + c$ 進行擬合，擬合出的函數圖形如圖 (六)，其方程式為 $y = 9.86222 \times 0.815918^t + 2.79045$，且 $r^2 = 1$。透過此一函數與直線 $t = 20$ 和 $t = 30$ 分別交於 (20 , 2.959)、(30 , 2.812)。因此本研究以指數函數預測第 20 週和第 30 週的汙染物濃度，分別為 2.959 mg/L 和 2.812 mg/L。

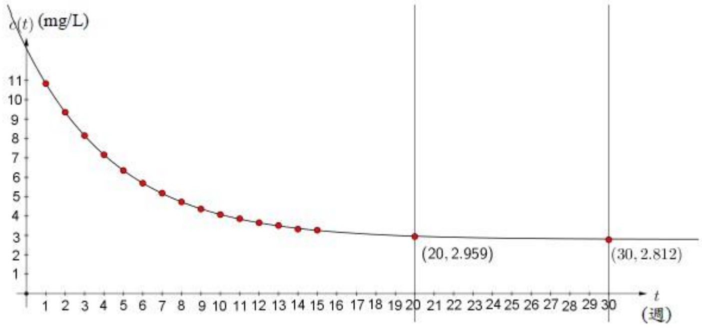

圖 (六) 指數函數擬合

至此本研究透過 $y = 6.8308 \times 10^{14}(0.00723072\,t + 1.8432)^{-52.1049} + 2.74659$ 和 $y = 9.86222 \times 0.815918^t + 2.79045$，兩個 $r^2 = 1$ 的函數進行預測，預測結果如下表 (四) 所示：

表 (四) 擬合函數用以估計第 20、30 週懸浮汙染物之濃度

函數	第 20 週	第 30 週
$y = 6.8308 \times 10^{14}(0.00723072\,t + 1.8432)^{-52.1049} + 2.74659$	2.941 mg/L	2.777 mg/L
$y = 9.86222 \times 0.815918^t + 2.79045$	2.959 mg/L	2.812 mg/L

因此本篇研究預測第 20 週的汙染濃度應介於 2.941 mg/L ~ 2.959 mg/L，第 30 週的汙染濃度應介於 2.777 mg/L ~ 2.812 mg/L。

五、結論與生活應用

本研究透過曲線擬合的手段成功預測第 20 週的汙染物濃度應介於 2.941 mg/L ~ 2.959 mg/L；第 30 週的污染物濃度應為 2.777 mg/L ~ 2.812 mg/L。

參考資料

[1] 廖培凱、王玟綺、江雅芬、施琇敏、賴政泓、王瑞陽、蘇麗敏、單維彰（2023 年 10 月 27 日）。探究單元教案設計 - 厚此薄彼的擬合探究。取自 https://reurl.cc/GKyN7G

[2] Carrot Cheng（2021 年 3 月 20 日）。回歸分析 (Regression Analysis) 的 R 平方 (R Squared) 與調整後 R 平方 (Adjusted R Squared)。取自 https://reurl.cc/q0AGOg

[3] IMMC 2024 大中華區秋季賽 A 題，Lake pollution control for sustainable development，https://reurl.cc/M4v1dk