

統計系112學年度專題研討暨專題競賽

開發以瀏覽器為介面的迴歸分析系統與說明網站

參賽人員：

410979068 施尚丞

410984014 鍾艾珊

410878022 徐子旺

410978074 張博鈞

410973041 傅紀茹

410973022 李嫚庭

113年5月20日

摘要

迄今，各界需依賴統計分析驅動管理與決策，迴歸分析更受到廣泛應用。然而，過高的專業操作和語言門檻、檔案格式的限制及昂貴的軟體費用，使初學者在執行迴歸分析時面臨重重困難。因此，本產品利用 Python 的 Streamlit 套件，在瀏覽器上開發一個具有輔助教學和實際分析能力的迴歸分析系統，能針對所需的資料進行預測及分析，並為使用者提供清晰的使用介面及適當的提示、引導和反饋。同時，使用 HTML / CSS 架設迴歸分析說明網站，幫助用戶輕鬆掌握迴歸分析的基本理論和專有名詞。此產品將提供更全面且友善的迴歸分析平台，幫助非統計專業人士學習並執行迴歸分析，降低用戶的學習和使用門檻外，促進統計知識的普及和應用，提高教學與學習的效率，從而提升社會的科技素養。

關鍵詞：迴歸分析、使用者介面、Streamlit、網站部署

目錄

- 壹、研究動機與研究目的.....1
- 貳、文獻回顧與探討.....2
 - 一、迴歸分析.....2
 - 二、使用者介面.....2
 - 三、小結.....3
- 參、研究方法及步驟.....3
 - 一、迴歸分析流程實現.....3
 - 二、網站介面開發.....4
 - 三、教育資源補充.....4
- 肆、產品介紹.....4
 - 一、產品概述.....4
 - 二、迴歸分析實作網站功能介紹.....4
 - 三、統計理論網站介紹.....11
 - 四、兩產品之間的互動性.....12
 - 五、使用場景.....12
- 伍、討論.....12
- 陸、結論.....12
 - 一、教育面.....12
 - 二、學習面.....13
 - 三、社會面.....13

壹、研究動機與研究目的

當今各式統計方法的運用和迴歸分析已經成為各領域中解決問題和做出預測的重要工具；然而，對於非統計專業人士而言，嘗試進行迴歸分析可能會面臨諸多困難，比如複雜的數學原理、繁瑣的程式碼以及晦澀難懂的數據結果等。

同時現有的統計分析軟體存在成本昂貴或使用複雜的問題，包括需編寫程式碼、缺乏友善的使用者介面，以及提供的資訊過於簡單等，如下表所示：

表 1 現有迴歸分析網站或軟體

網站/軟體名稱	優點	缺點
3x86.com	- 功能單一，適合快速計算	- 無法導入檔案，執行較複雜的 data
clspaces.com	- 提供線性回歸計算功能	- 僅能操作範例，UI 老舊，無法實際操作大量數據
datatab.net	- 搭配說明，教學資源豐富	- 操作手續繁雜，無法直觀執行迴歸分析
statskingdom.com	- 提供線性回歸計算	- 功能單一，不支援其他檔案類型，初學者不易使用
SPSS	- 強大的統計分析功能	- 軟體較昂貴，學習曲線陡峭
SAS	- 強大的數據分析能力	- 軟體昂貴，語法操作複雜

本研究期望應用 Python 的 Streamlit 套件在瀏覽器上建立 UI 介面，開發一個同時具備具有輔助教學和實際分析能力的迴歸分析系統；期望此系統將能解決現有軟體缺陷，滿足各界對於使用友好、價格合理的迴歸分析工具的需求。同時，期望為非統計專長之用戶提供一個更加方便及全面的操作方法，更易於取得所需的數據，從而在制定決策時作出更精準和高效的判斷；而在教育領域，期望此系統能提供有價值的教學工具，提高教學效果，促進學習成效；結合上述研究動機，因此有以下幾點研究目的：

- 1、協助企業深入了解各種變數之間的關係，進而預測可能的趨勢和結果；利用深度的數據洞察幫助企業優化營運、提高產品效能，並更有針對性的滿足客戶所需，使其在競爭激烈的市場中保持敏捷和競爭力。
- 2、透過在系統中提供解說和引導，使學生能更好的進行學習。
- 3、將研究結果實際發佈於公開網站，使一般使用者能夠在其最熟習的瀏覽器上使用此研究，促進迴歸分析對於各界的幫助。

貳、文獻回顧與探討

一、迴歸分析

根據 Chatterjee & Hadi (2013) 統整，用戶執行迴歸分析，需進行以下步驟：

- (一) 問題陳述：決定需要透過迴歸分析解決的問題為何，起初的問題選擇會影響到後續的模型制定和決策，若選擇到不適合的主題將會耗費大量精力與時間。
- (二) 選擇潛在相關變數：在問題陳述後需要決定問題的應變數與自變數。
- (三) 資料選擇：蒐集所需的資料，這些資料通常會有 n 個觀測對象，每一個觀測對象的數據由各自變數的測量值組成。
- (四) 模型設定：根據上述的條件與限制，選擇使用線性迴歸分析模型、複迴歸分析模型或是其他。
- (五) 估計模型參數：使用最小平方法 (OLS) 或其他適當方法來估計模型參數。
- (六) 模型擬合：評估數據對模型的擬合程度與有效性。
- (七) 模型診斷與改進：進行模型診斷來檢查模型假設是否得到滿足，例如殘差的獨立性、常態性和等方差性，並識別模型中可能存在的問題，如異常值或多重共線性。若存在問題，則針對變量或模型調整。
- (八) 執行迴歸分析，產出結論與預測。

二、使用者介面

使用者介面 (User interface, UI) 設計的目標在於簡化並高效地促使使用者與系統之間的互動，讓使用者能夠在完成任務時較少遇到阻礙。良好的使用者介面設計不僅使操作更加直觀，還能減少使用者在處理介面時的認知負擔，否則設計不良的使用者介面容易造成用戶的挫折感與情緒上的不滿。

Python 是一種非常強大的編程語言，Streamlit 是一個 Python 函式庫的開源軟體，它讓機器學習和數據科學的專業人士輕鬆創建客製化的 Web 應用程式。Pramod Singh (2021) 認為 Streamlit 是 Flask 的替代方案，用於將機器學習模型部署為 Web 服務。使用 Streamlit 的最大優勢在於它允許在應用程式的 Python 文件中使用 HTML 代碼。基本上，它無需單獨製作前端用戶界面的模板和 CSS 格式。Shukla, S., et al. (2021) 認為它能夠利用極少的代碼行來建構各種應用程式。且在 Streamlit 中編碼相當容易部署、管理並與其他應用程式協作。聲明一個變數與添加一個小部件是相同的，無需編寫後端代碼、定義路徑、處理 HTTP 請求、連接到前端，也無需編寫 HTML、CSS 或 JavaScript。透過使用 Streamlit，避免了大量模板、HTML、CSS，使深度學習模型的部署變得更加簡單。

根據 Shukla, S., et al. (2021) 做的一項研究，這項研究使用機器學習 (ML) 演算法和 Python，嘗試透過 Streamlit 程式庫以非常簡單有效的方式建立一個有助於預測個人貸款批准的 Web 應用程式。研究結果顯示，透過 Streamlit 函式庫，建立和運行任何資料科學和 ML 模型的 Web 應用程式相較於傳統方法更容易。研究者認為 Streamlit 易於學習，並能輕鬆將 ML 和資料科學專案轉換為 Web 應用程式。這使得人們可以輕鬆地將 Python 程式碼轉換為 Web 應用程式。因此最後我們選擇採用 Python 和 Streamlit 來實現使用者介面的設計，這不僅使得介面的開發更加高效，同時也提供了更簡單且易於操作的方式。

三、小結

迴歸分析在研究和實務的應用上一直扮演著至關重要的角色，其應用領域從商學、工程、醫學、政治乃至心理學都有著舉足輕重的地位。然而對於普通民眾或未受過統計知識的人而言，完整執行迴歸分析之步驟卻十分困難且花費昂貴。本研究將會以Python開發一迴歸分析之使用者介面，來幫助這些使用者更容易地進行迴歸分析。本研究旨在降低進行迴歸分析的技術門檻。通過提供步驟指導、自動化的數據處理以及圖形化的結果展示，用戶可以不需要深厚的統計背景知識也能夠完成從數據輸入到結果解釋的全過程。

參、研究方法及步驟

本研究以 Python 為基礎,開發一個名為 RegressifyXpert 的網站介面系統,提供使用者進行多元迴歸分析。系統架構分為三大部分：

一、迴歸分析流程實現

首先，本研究將提出迴歸分析之流程及方法，同時標註各步驟、相關參數的估計方法、模型假設的檢定方式和資料轉換步驟。接著，透過系統介面提供包括參數估計、各式檢定以及相關的圖表的生成步驟。針對本產品所採用的迴歸分析方法，我們將依序以下列步驟完成：

- (一) 資料預處理：辨識遺失值，自訂無效值設定，根據使用者需求刪減資料。
- (二) 資料視覺化：繪製多種視覺化圖表（圓餅圖、直方圖、散佈矩陣圖等），探索變數關係。
- (三) 模型配適：提供使用者多項式模型進行選用，可自行加入交互作用項，或是用常用的 log, square transformation 對變數進行轉換後加入模型。預設以最小平方方法估計係數，計算殘差，也提供加權最小平方方法供使用者選用。
- (四) 殘差診斷：提供圖表與假設檢定兩種方法進行診斷。針對小樣本資料，我們需要透過一些非正式的殘差診斷圖討論資料是否偏離假設；而在大樣本下，建議同時透過殘差診斷圖和正式的統計檢定來診斷誤差項之各項假設。
 - (1) 獨立性診斷：提供 Sequence Plot Of Residual 及 Durbin-Watson Test。
 - (2) 線性關係診斷：提供 Residual Plot 及 F-test lack of fit。
 - (3) 同質變異性診斷：提供 Residual Plot 及 Brown Forsythe Test, Breusch Pagan Test, White Test 三種常見的檢定方式。
 - (4) 常態性診斷：提供 QQ Plot, Histogram 及四種常見的常態檢定 (Shapiro-Wilk Test, Kolmogorov-Smirnov Test, Cramer-von Mises Test, Anderson-Darling Test)。
- (五) 模型修正：在滿足不同假設的情況下，我們將針對模型及資料採用不同的處理方式，並希望透過修正能適當的分析資料並進行推論與預測。
 - (1) 不滿足獨立性的假設下，將引導使用者調整模型以納入可能遺漏變數、使用其他分類模型或利用時間序列模型等。
 - (2) 若不滿足線性關係診斷，則引導使用者對自變數進行轉換。
 - (3) 若不滿足線性關係診斷，則引導使用者對自變數進行轉換。

(4) 若不滿足同質變異性診斷，提供加權最小平方法 (Weighted least squares) 使得不同觀測值的變異性被納入考慮，從而獲得更準確的估計

(5) 若不滿足常態性診斷，系統提供 Bootstrap method 進行區間估計，並且在大樣本下，可透過中央極限定理，推論模型係數近似常態分配，可依據此推論結果進一步計算估計誤差及信賴區間。在小樣本的情況之下，本產品也提供 Box-Cox 轉換將數據轉換成不同的形式，包括對數轉換、平方根轉換方式等。

(六) 模型結果：透過模型配適的估計結果進行統計推論，提供之資訊包含 整體模型的 F 檢定、參數的 T 檢定、R 平方、標準誤、信賴區間等，幫助使用者進一步根據模型估計結果進行預測。

二、網站介面開發

將應用 Python 的 Streamlit 套件建立多頁式的 UI 介面，包含資料預處理、視覺化探索、篩選、模型配適、殘差分析、模型評估等功能頁面；預設支援 CSV 格式數據輸入，可處理數值及類別型變數，並提供一階、二階多項式迴歸模型選項，還可依照使用者需求加入交互作用項。期望透過此介面設計，即使是不熟悉程式語言的使用者也能輕鬆進行迴歸分析，無需直接接觸程式碼，只需透過友善的介面操作即可達成目的；相較於市面上非程式類的資料分析工具，此產品除了常規分析外，我們還包含了針對殘差的各種統計檢定和資料轉換與模型使用的建議，幫助使用者更適切地進行資料分析。

三、教育資源補充

將應用 HTML / CSS 建置 RegressifyXpert 的說明網站，網站包含模型介紹、專有名詞介紹、問題解決導引等頁面，並在介面中加入〔小常識補給站〕、〔常見問題〕與〔例〕以解釋統計理論概念，期望幫助使用者更了解迴歸分析。

綜合上述，本研究系統期望實現迴歸分析完整流程，提供資料預處理到模型評估的各項功能，並針對違反模型假設情況給予修正建議。同時搭配 HTML / CSS 說明網站及介面提示，達到實用分析及統計理論教育的雙重目的，有助於不同程度使用者的使用。

肆、產品介紹

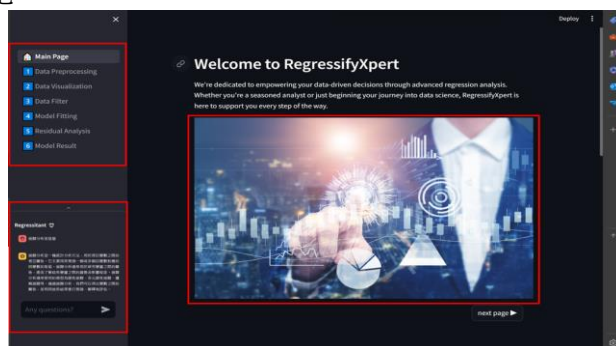
一、產品概述

為使本產品兼具實用性及教育意義，我們最終開發了一套迴歸分析工具，包含以 Streamlit 架構的迴歸分析實作網站和迴歸分析相關之統計理論網站，透過兩個網站的串聯，使用者在實作資料分析的同時，也能透過自製的統計知識網站提供理論的輔助，讓大眾了解分析的各個理論方法及概念，能更輕鬆地進行數據分析。

二、迴歸分析實作網站功能介紹

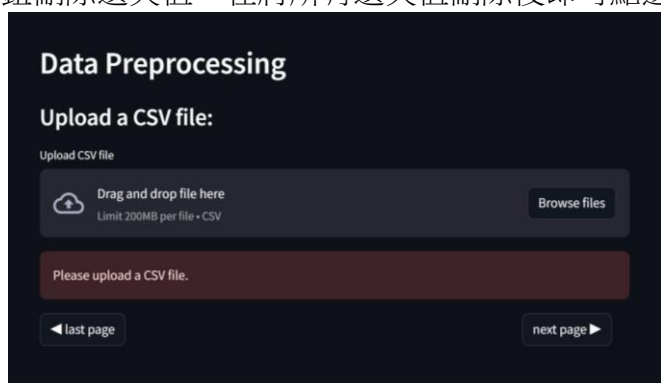
(一) 首頁：如(附圖一)，左上角為目錄，包含整個系統所有的分頁，使用者可點選目錄或最下方之換頁鍵，進行頁面切換；左下角為 GPT 的聊天視窗，使用者可透過該聊天視窗初步的針對任何資料分析及模型建構的問題，答疑解惑。此外，首

頁包含本產品的文字介紹，以及網站使用說明的示範動畫，透過動圖可了解本網站使用流程及相關功能。

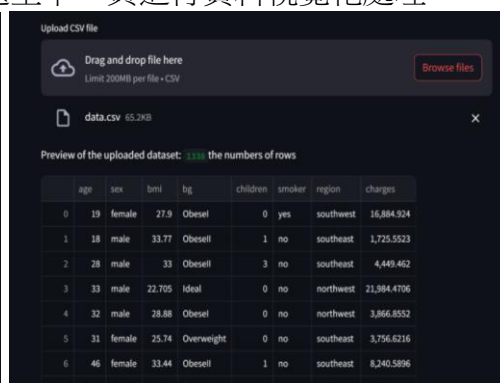


圖一、系統首頁

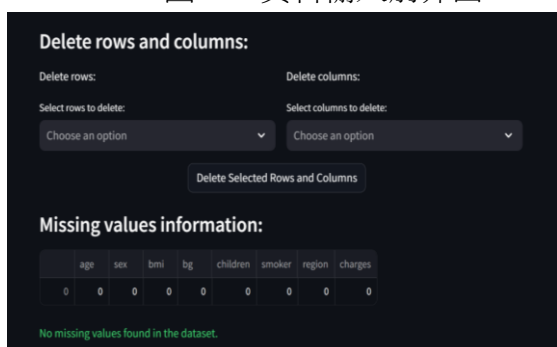
(二) 資料輸入及預處理頁面：圖二為資料尚未輸入時的介面，在點選資料檔後，可直接透過該頁面查看資料概況及資料筆數（附圖三），同一頁面往下滑為資料刪除區域（附圖四），透過點選可刪除特定變數及觀察值，接著下方為遺失值區塊，可直接查看各變數的遺失值數量，若資料存在遺失值，可透過按鈕點選刪除包含遺失值的整筆資料，而本頁面最下方區塊可處理潛在遺失值（附圖五），若資料本身包含非 NA 之遺失值，如問卷資料經常以 99999 設定遺失值，則可透過本區塊手動設定特定變數之遺失值代表的文字或數字，進階查詢所有遺失值的數量，並可點選按鈕刪除遺失值，在將所有遺失值刪除後即可點選至下一頁進行資料視覺化處理。



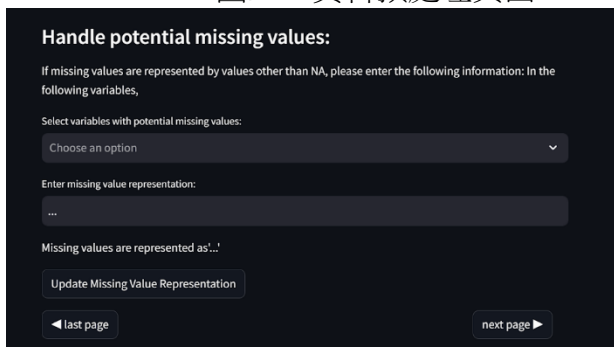
圖二、資料輸入前介面



圖三、資料預處理頁面



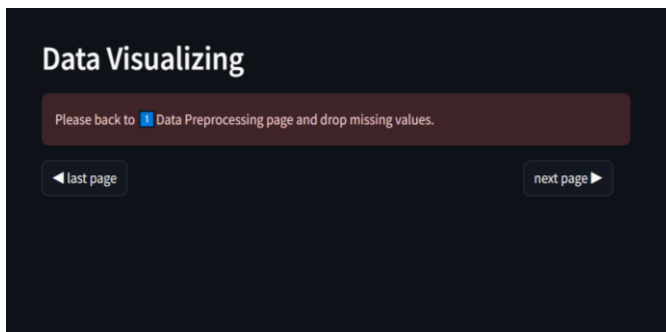
圖四、變數及資料刪除區域



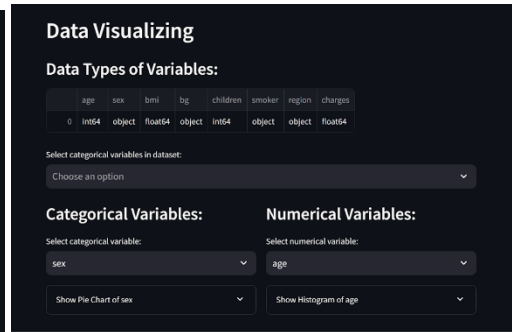
圖五、遺失值處理區域

(三) 資料視覺化頁面：若資料存在遺失值且使用者尚未點選刪除遺失值，並直接進入視覺化頁面，則本頁面會顯示提示訊息，引導使用者回到前一頁刪除遺失值（附圖六），執行完成後才能接續進行資料視覺化分析（附圖七），在資料視覺化分頁中，首先第一部分為顯示各變數型態及調整類別型變數的區域，透過點選變數名稱，可手動將非文字的類別型變數正確辨別其變數型態，接著往下滑為單一變數資料視覺化的區域（附圖八）、兩連續變數之散佈圖及其相關係數（附圖九）、所

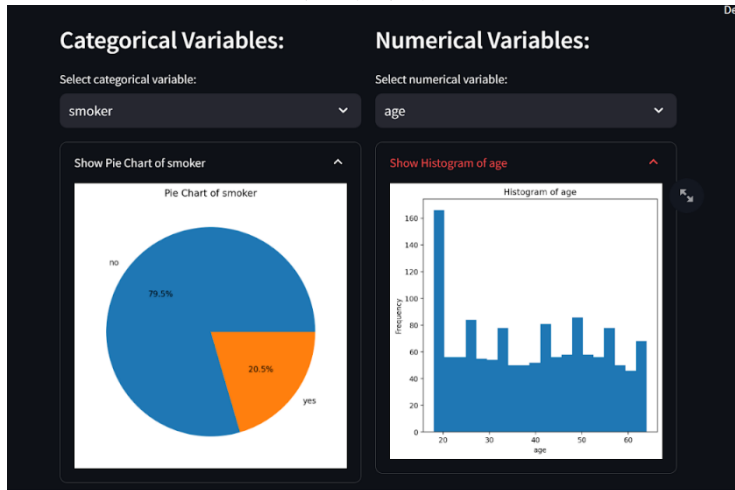
有連續型變數繪製之散佈矩陣圖（附圖十），及其對應的相關係數矩陣（附圖十一）。



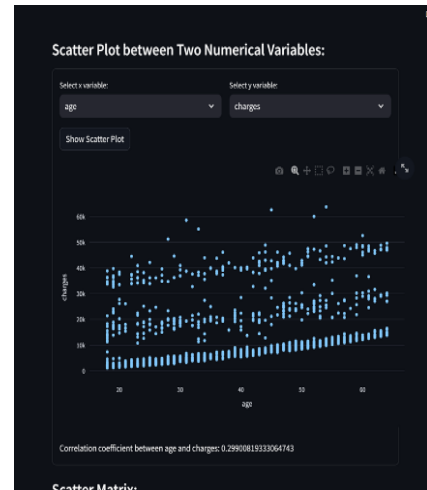
圖六、系統阻擋提示



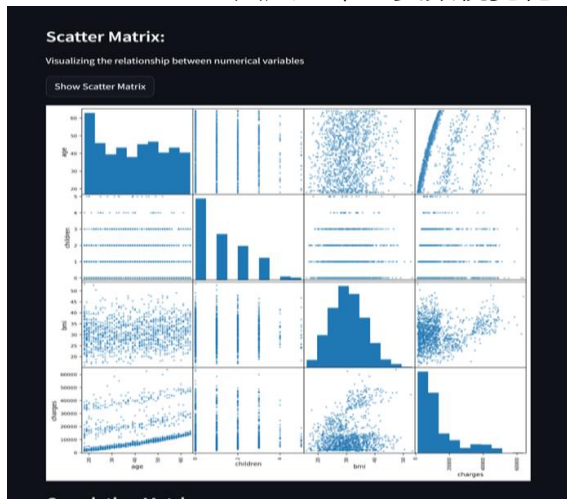
圖七、變數類別顯示區域



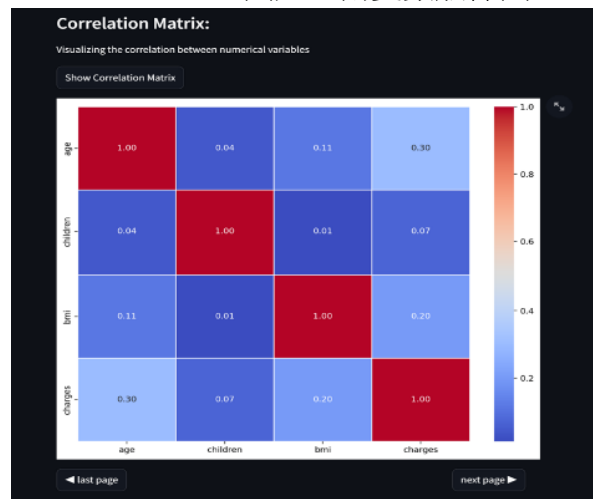
圖八、單一變數視覺化



圖九、兩變數散佈圖

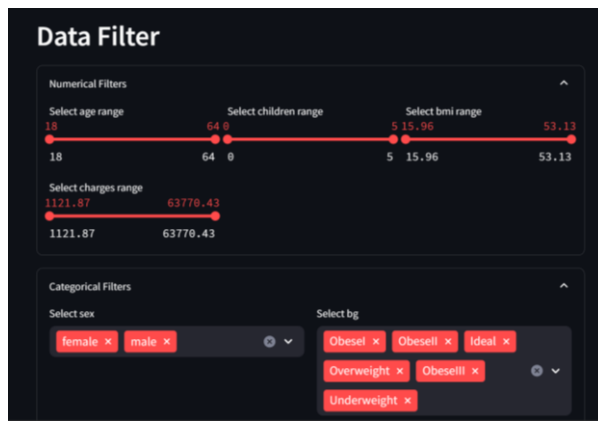


圖十、散佈矩陣圖



圖十一、相關係數矩陣

（四）資料篩選頁面：在資料視覺化後，使用者可透過本頁面進行資料篩選（附圖十二），透過拖移和點選的方式，可輕鬆進行特定欄位之資料篩選，並且頁面下方將顯示篩選後的資料（附圖十三），提供使用者查看與確認。



圖十二、資料篩選設定區域

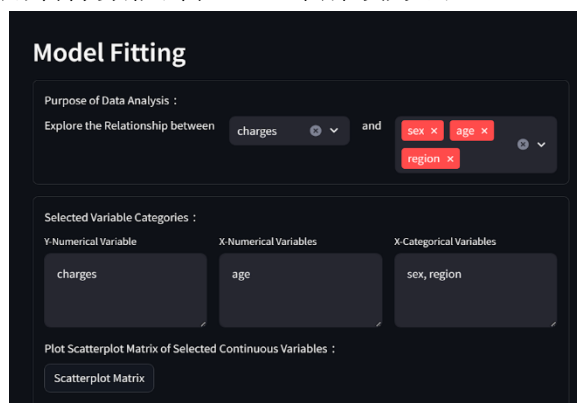
	age	sex	bmi	bg	children	smoker	region	charges
0	19	female	27.9	Obese	0	yes	southwest	16,884.924
1	18	male	33.77	Obese	1	no	southeast	1,725.5523
2	28	male	33	Obese	3	no	southeast	4,449.462
3	33	male	22.705	Ideal	0	no	northwest	21,984.4706
4	32	male	28.88	Obese	0	no	northwest	3,866.8552
5	31	female	25.74	Overweight	0	no	southeast	3,756.6216
6	46	female	33.44	Obese	1	no	southeast	8,240.5896
7	37	female	27.74	Obese	3	no	northwest	7,281.5056
8	37	male	29.83	Obese	2	no	northeast	6,406.4107
9	60	female	25.84	Overweight	0	no	northwest	28,923.1369

圖十三、篩選後資料呈現

(五) 迴歸模型建構頁面：根據篩選後的資料挑選模型自變數及應變數(附圖十四)，並接續根據所挑選的變數之變數型態進行劃分，顯示出各變數為連續型或類別型，提供使用者再次確認納入模型之自變數的變數型態是否正確，若使用者此時發現變數型態有誤，可回到資料視覺化的頁面調整，再接續進行模型配適。

此外，頁面下方也能查看使用者所挑選之連續型自變數的散佈矩陣圖(附圖十五)，方便使用者分析各連續型自變數和應變數的關係，及自變數彼此之間的關係，而在圖形下方(附圖十六)，也以文字顯示散佈矩陣圖的查看重點及分析要點，幫助初學者進一步理解圖形之用意，並且若模型之自變數包含類別型變數，將進一步顯示類別型變數轉換為虛擬變數的過程(附圖十七)。

最後，本頁最下方為迴歸函數建構區域(附圖十八)，提供使用者多樣且常見的自變數及應變數型態做挑選，包含變數之次方項、兩兩交互作用項、指對數型態及自訂之 k 次方變數型態，透過點選各型態之變數，建構迴歸估計函數，並在下方顯示出使用者所建構的函數型態及模型假設(附圖十九)，在使用者確認完所要配適的模型後，即可點選下方 Run Model 按鈕，以 OLS 方法估計係數，得到估計的迴歸方程式與模型配適圖和殘差圖(附圖二十)，同時其中的 interpretation 按鈕提供使用者係數估計值之正確解讀方法。



圖十四、變數選取及分類區域



圖十五、模型之散佈矩陣圖

Interpreting Scatterplot Matrix:

- Please note the linear relationship between the independent and dependent variables. Observe if there is a linear relationship between `children` and `charges`. Observe if there is a linear relationship between `age` and `charges`. These relationships may suggest that `children`, `age` has a significant impact on predicting `charges`.
- Please note the correlation between independent variables: observe if there is any relationship between the variables in `children`, `age` pair by pair.

Dealing with Categorical Variables:

- The number of categories for categorical variable `region` is 4, with values ['southwest' 'southeast' 'northwest' 'northeast']. Converting them into dummy variables:
`region_southeast` `region_northwest` `region_northeast`
 which `region_southeast` is:

圖十六、解讀散佈矩陣圖之提示

Dealing with Categorical Variables:

- The number of categories for categorical variable `region` is 4, with values ['southwest' 'southeast' 'northwest' 'northeast']. Converting them into dummy variables:
`region_southeast` `region_northwest` `region_northeast`
 which `region_southeast` is:

$$region_{southeast} = \begin{cases} 1, & \text{if } region = southeast \\ 0, & \text{otherwise} \end{cases}$$

which `region_northwest` is:

$$region_{northwest} = \begin{cases} 1, & \text{if } region = northwest \\ 0, & \text{otherwise} \end{cases}$$

which `region_northeast` is:

$$region_{northeast} = \begin{cases} 1, & \text{if } region = northeast \\ 0, & \text{otherwise} \end{cases}$$

圖十七、虛擬變數設定區域

Model Selection

Select the dependent variable form

- ☒ charges
- ☐ log(charges)
- ☐ charges^2

Select independent variables for the first-order form

Select All

Select independent variables for the second-order form

Select independent variables for the interaction form

Select independent variables for the log form

Select independent variables for the exp form

圖十八、變數型態挑選區域

log(children) x exp(age)

Select independent variables for the custom form

Choose an option

children^[-1.0]

Please input the order of the custom form

-1.00

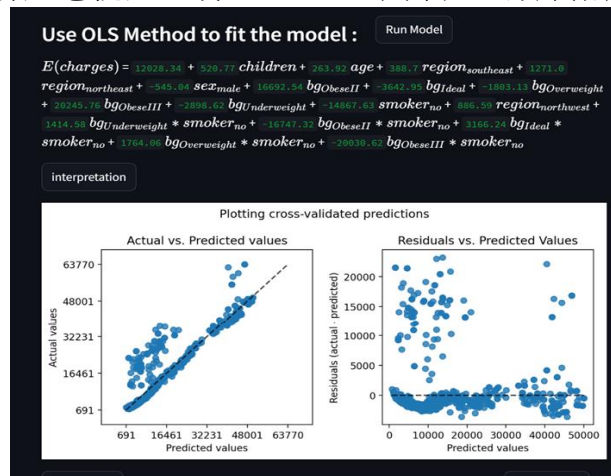
The multiple regression equation with an intercept term can be written as:

$$charges = \beta_0 + \beta_1 children + \beta_2 age + \beta_3 region_{southeast} + \beta_4 region_{northwest} + \beta_5 region_{northeast} + \beta_6 sex_{male} + \beta_7 bg_{ObeseII} + \beta_8 bg_{Ideal} + \beta_9 bg_{Overweight} + \beta_{10} bg_{Underweight} + \beta_{11} smoker_{no} + \beta_{12} region_{northwest} + \beta_{13} children^2 + \beta_{14} bg_{Underweight} * smoker_{no} + \beta_{15} bg_{ObeseII} * smoker_{no} + \beta_{16} log(children) + \beta_{17} exp(age) + \epsilon$$

Assumptions of the error term ϵ :

- The error term ϵ has a mean of zero, i.e., $E(\epsilon) = 0$.
- The error term ϵ has constant variance, i.e., $Var(\epsilon) = \sigma^2$.
- The error term ϵ is normally distributed.

圖十九、顯示配飾之迴歸函數

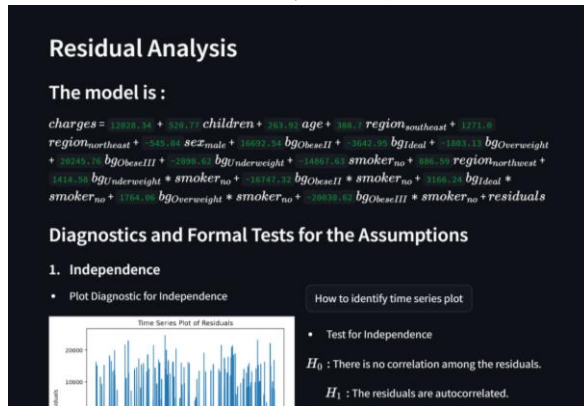


圖二十、模型配飾圖及殘差圖

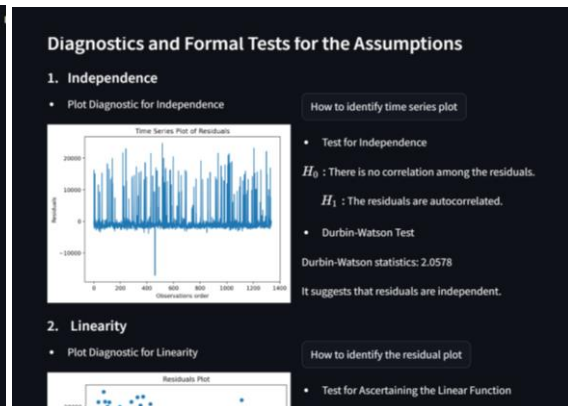
(六) 殘差分析頁面：

在估計完係數後，使用者將透過本頁進行殘差診斷及統計檢定，判斷是否違反模型假設以及辨別該模型之適當性，首先頁面最上方再次顯示估計之迴歸方程式，方便使用者查看（附圖二十一），接著進行四項假設之診斷，各項診斷包含圖形診斷及多種正式統計檢定（附圖二十二 - 圖二十五），最後在下方顯示診斷結果（附圖二十六），包含模型可能違反的假設，而又考量到統計檢定之檢定力容易受樣本大小影響，因此雖然診斷結果之區塊，預設為根據統計檢定結果顯示，但在設計上仍然保留在小樣本下，使用者能根據圖形判別，進一步變更殘差診斷結果的彈性，同時在小樣本下，系統也會自動顯示文字，提醒使用者應同時考量圖形診斷之必要（附圖二十七），並且針對初學者，也可透過各圖形之超連結，進一步的在統計知識輔助網站，查閱圖形診斷之指引。

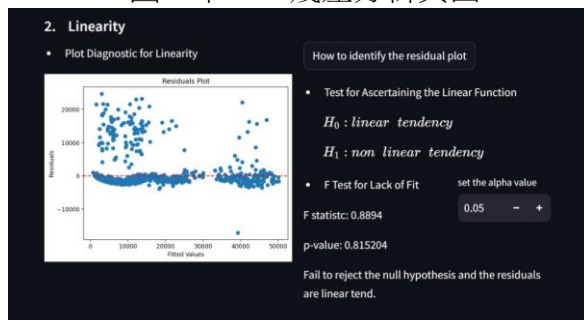
最後，根據診斷結果，將在下方顯示因應的補救方法，包含若殘差不滿足彼此獨立的假設，將引導使用者至前一頁添加相關的模型變數，或以 Bootstrap 方法估計係數之信賴區間（附圖二十八）；若模型違反線性關係之假設，將引導使用者至前一頁嘗試適合的變數型態，或仍需添加相關的模型自變數（附圖二十九）；若模型違反同質變異之假設，則將提供 Box-Cox 轉換，或 WLS 估計方式進行係數估計（附圖三十）；若殘差不滿足常態分配的假設，則將提供 Box-Cox 轉換，或以 Bootstrap 方法估計係數之 95% 信賴區間（附圖三十一）。



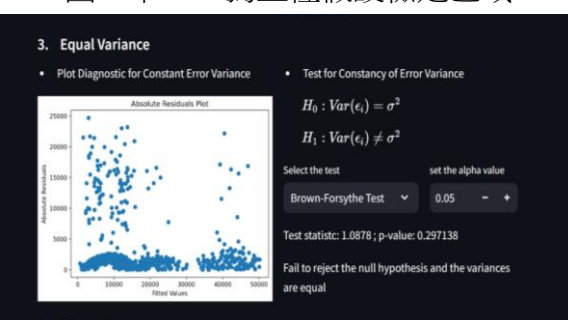
圖二十一、殘差分析頁面



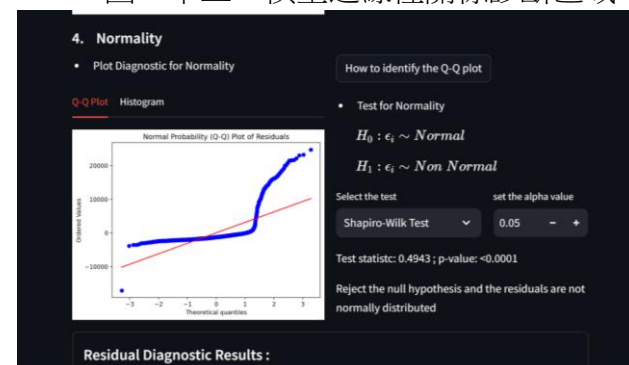
圖二十二、獨立性假設檢定區域



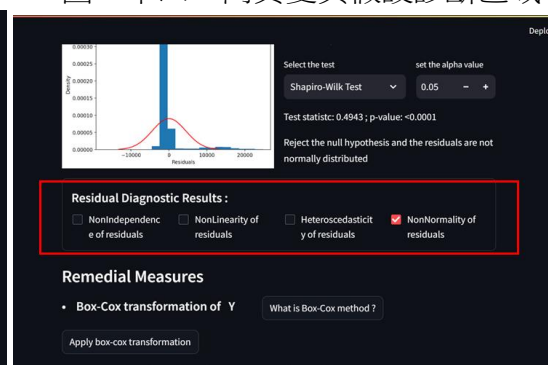
圖二十三、模型之線性關係診斷區域



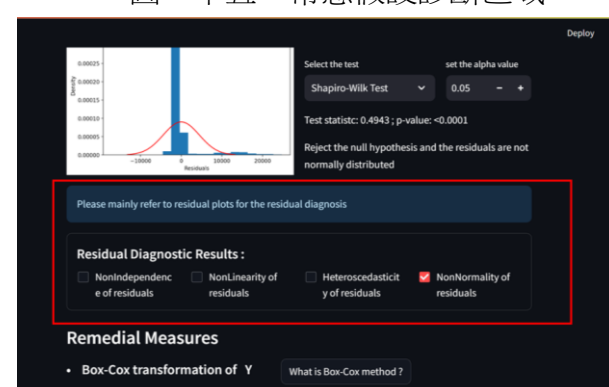
圖二十四、同質變異假設診斷區域



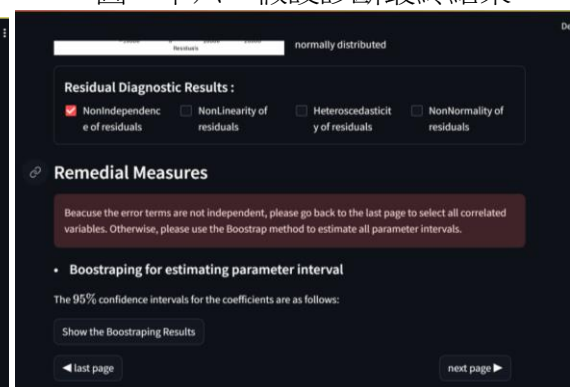
圖二十五、常態假設診斷區域



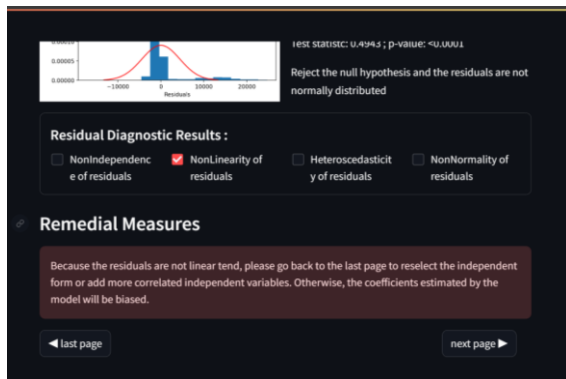
圖二十六、假設診斷最終結果



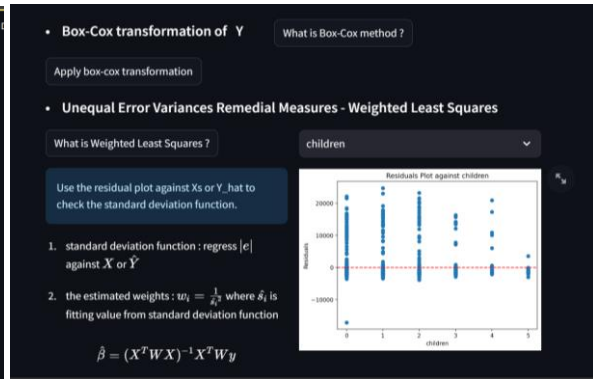
圖二十七、小樣本之診斷提示



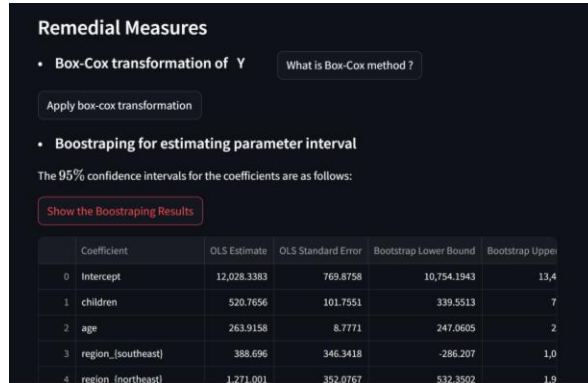
圖二十八、違反獨立性之系統提示



圖二十九、違反線性關係之系統提示



圖三十、異質變異補救區域



圖三十一、不滿足常態分配之補救區域

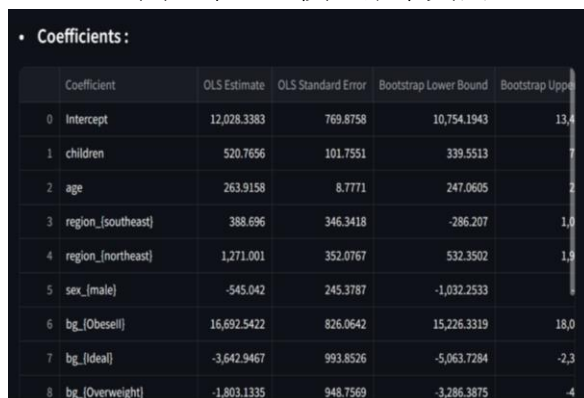
(七) 模型結果頁面：本頁將根據前述調整後之模型、不同之係數估計方法及區間估計方法產出相對應之統計報表、迴歸估計式及提供相關參數之解釋（附圖三十二至圖三十五）。



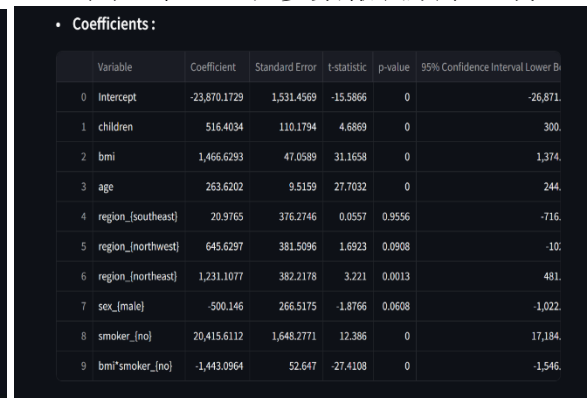
圖三十二、模型結果頁面



圖三十三、各參數報表顯示區域



圖三十四、Bootstrap區間估計報表



圖三十五、個別係數估計及檢定區域

三、統計理論網站介紹

(一) 首頁：本產品的核心目的為「一站式提供所有您所需的迴歸分析知識」，因此，我們提供在實做網站中使用到的模型介紹、專業名詞解析、問題解決、統計推論與一個釋例的詳細說明供使用者參考，並且以實作網站中的頁面分類專有名詞，讓使用者能更輕易地尋找到所需資訊。

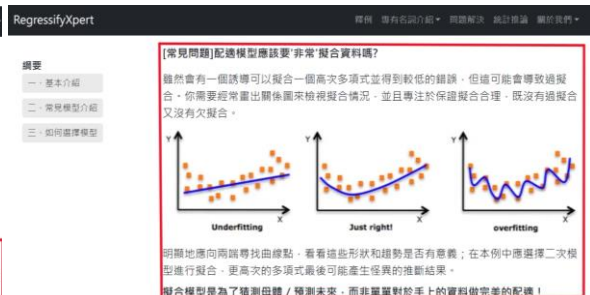


圖三十六、統計輔助網站首頁介紹

(二) 專有名詞介紹：本研究不僅會將使用之統計相關名詞進行介紹，必要時也會提供圖示供使用者進行參考，並增添「常見問題」的內容，幫助使用者能夠更加深入了解迴歸分析。

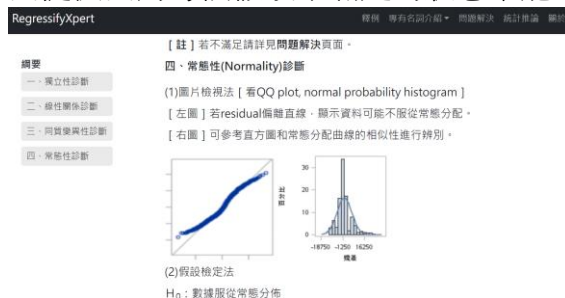


圖三十七、名詞解釋頁面

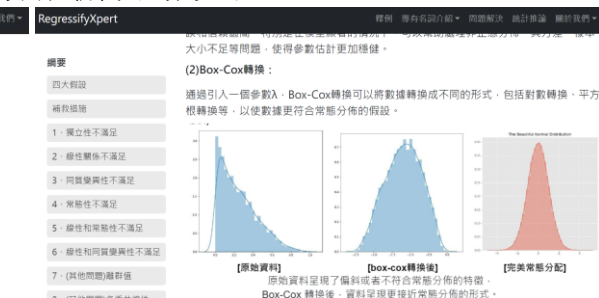


圖三十八、常見問題指引頁面

(三) 問題解決：在殘差診斷的專有名詞解釋頁面提供四個假設的圖片檢視法和假設檢定法，但在實作時必定會出現不滿足假設的情況發生，因此我們在問題解決頁面提供面對每個假設不滿足的狀態下能夠嘗試解決的方法。

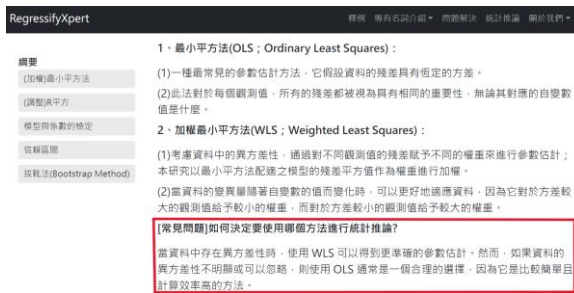


圖三十九、殘差診斷指引頁面(一)



圖四十、殘差診斷指引頁面(二)

(四) 統計推論：針對配適模型時使用到的方法與實作網站中的「模型結果」頁面所提供之檢定和數據進行解釋，並提供「小常識補給站」、「常見問題」與「例」幫助使用者更了解迴歸分析。



圖四十一、估計方法介紹頁面



圖四十二、模型結果解讀指引

四、兩產品之間的互動性

用戶在使用實作網站進行資料分析的過程中，可以隨時點擊界面上的超連結，直接訪問相關的輔助網頁，獲取相應的理論知識和方法解釋。透過這樣的連結方式，不僅提高了產品使用的便利性，也能有效地提升用戶的學習效率和理解深度。

五、使用場景

(一) 學術研究：研究人員可以使用該工具進行數據分析及迴歸模型的建構，並獲取相關的圖表及數據報表。

(二) 商業分析：有數據分析需求的民眾，可以應用該工具進行市場趨勢分析、銷售預測等商業分析任務。

(三) 教育培訓：可作為教學工具，幫助學生直觀地學習迴歸分析理論與技術。

伍、討論

本研究開發一款基於 Python 的 Streamlit 套件的迴歸分析系統，旨在為用戶提供直觀且功能齊全的分析平台。除了支持視覺化、模型擬合、殘差診斷和統計推論等核心功能外，該系統還具有教育意義，通過使用 HTML / CSS 架設迴歸分析說明網站幫助用戶理解迴歸分析的基本概念和理論背景。儘管如此，仍有以下幾個方面值得進一步優化和擴展：

1. **支持多種數據格式導入**：目前僅支持 CSV 格式的數據導入。未來，可以考慮對 EXCEL、JSON 等常見格式的支持，提高數據導入的靈活性。
2. **豐富模型種類**：目前主攻多項式迴歸分析，對類別變數的處理有限。未來可嘗試納入分類模型（如羅吉斯迴歸、決策樹）和時間序列模型（如自迴歸模型、指數平滑模型等），擴大系統的應用範疇。

陸、結論

透過完整的迴歸分析輔助系統，我們的目標是讓非統計專業人士和學習者能夠輕鬆進行資料分析，促進數據分析知識的普及和應用。使用者只需上傳資料，系統即可進行預處理，包括資料篩選和類別變數轉換，並提供所有相關視覺化資訊。接下來，使用者可以進入模型配適頁面，享有各種客製化選擇，從自變數到交互作用皆可自由調整。最後，系統還提供殘差診斷結果，讓使用者能夠判斷模型配適的結果是否符合各種假設，並提供相應的補救措施和統計推論的判定教學。在分析過程中，若遇到艱澀難懂的專有名詞，使用者只需點擊即可前往我們專門建立的外部網站，其中提供專有名詞的解釋、釋例以及對迴歸分析的相關知識整理。

本研究期望透過這套系統達到以下三方面的預期結果：

一、教育面

(一) 增加教學效率：在教課期間，校方可搭配此整合迴歸分析平台，幫助學生掌握基本技術操作，配合指引學習如何選擇合適的模型與評估模型的有效性，從而減少教師須不斷親自指導之情形。

(二) 促進跨領域學習：迴歸分析的應用跨越多個學科，本研究開發之整合工具能夠激發學生探索迴歸分析在不同領域中的應用，促進跨學科學習和研究。

二、學習面

(一) 自主學習：為非專業用戶提供的步驟指導和圖形化反饋，使他們在無需外部幫助的情況下自主學習和實踐迴歸分析，提升自學能力。

(二) 統計知識普及化：透過完善的學習動線規劃，降低迴歸分析的學習門檻，使更多人能夠接觸和學習這一重要的數據分析工具，促進統計學知識的普及。

(三) 增強互動學習：通過開發實用、易懂的迴歸分析使用者介面，用戶可以直觀地看到不同變量之間關係的視覺化，從而更好地理解迴歸分析的原理和應用，提高學生的學習動機和效率。

三、社會面

(一) 提高決策質量：在商業、政策制定和科研學術等領域，迴歸分析是一個重要的決策工具，本研究開發易於使用的迴歸分析工具，可以幫助更多的管理者利用客觀數據進行決策，提高其準確性和效率。

(二) 降低成本支出：企業端可以使用該平台進行迴歸分析，除了減少軟體的訂閱費用外，也可以減少學習統計相關知識的時間成本，更有效的進行業務相關的分析，強化產品、顧客管理、市場研究的數據基礎。

柒、參考文獻

Chatterjee, S., & Hadi, A. S. (2012). *Regression analysis by example* (5th ed.). Wiley.

Shukla, S., Maheshwari, A., & Johri, P. (2021). **Comparative Analysis of ML Algorithms & Streamlit Web Application**. Greater Noida: IEEE.

Singh, P. (2021). **Machine Learning Deployment as a Web Service**. In: *Deploy Machine Learning Models to Production*. Apress, Berkeley, CA.

學號	姓名	分工內容
410979068	施尚丞	實作網站(OPENAI機器人)、 統計理論網站(部署框架)、書面、PPT
410984014	鍾艾珊	實作網站(前半部版面設計與製作)、 統計理論網站(網站內容與排版)、 釋例(數據分析)、書面、PPT
410973041	傅紀茹	實作網站(主要製作與統整)、書面、PPT
410973022	李嫚庭	實作網站(前半部版面設計與製作)、 統計理論 網站(輔助與排版)、書面、PPT
410978074	張博鈞	釋例(篩選資料、初步分析)、 提供部分模型結果及殘差頁面之程式碼
410878022	徐子旺	-