

110-2 迴歸分析 期末專題報告

透過甲殼類生物身體特徵進行年齡預測：
以波士頓地區螃蟹為例

第四組

應數三乙 林鈺庭 D0856699

應數三乙 賴泓瑜 D0856894

應數三乙 林敬棠 D0856966

應數三乙 黃棣陽 D0857022

應數三乙 鄭煜綸 D0857169

報告日期：2022/06/13

摘要

以各種動物而言，人們常以動物的大小來粗略估計其年齡；以人類為例，我們可以簡易地通過肉眼判斷一個人大概是青年期或是壯年期。反觀甲殼類生物因為其堅硬的結構限制，所以在生長過程是不連續的，也又是說不容易透過外骨骼來對甲殼類生物進行年齡測定。因此要判別一隻甲殼類生物的年齡並非是一件易事。本研究波士頓地區螃蟹資料希望利用螃蟹的身體特徵建構一迴歸模型來預測其年齡。此處將假設該資料的螃蟹為同一時期和同一種類的螃蟹進行下列章節的分析。本研究將在第二章介紹資料以及變數定義；第三章進行探索式資料分析以及視覺化分析；第四章說明研究方法、建立迴歸模型、選擇以及確認最終模型、模型評估，以及模型解釋；最後在第五章作總結與討論。而本研究的結論為模型存在低度擬合的問題，因此還需要多加考慮其他因子來提升預測的效果。

關鍵詞：甲殼類生物、年齡測定、迴歸分析

目錄

1. 研究動機與目的	1
2. 資料說明與變數定義	2
2.1 資料來源與說明	2
2.2 變數定義	2
3. 探索性資料分析	3
3.1 敘述性統計	3
3.2 視覺化呈現與分析	3
4. 研究方法與模型	10
4.1 流程圖	10
4.2 資料前處理	11
4.2 模型選擇	11
4.3 模型評估	16
4.4 模型解釋	16
5. 結果與討論	18
6. 參考資料	18
附錄	20

圖目錄

圖 1 螃蟹性別之圓餅圖.....	3
圖 2 螃蟹長度之盒鬚圖.....	4
圖 3 螃蟹長度之直方圖.....	4
圖 4 螃蟹直徑之盒鬚圖.....	4
圖 5 螃蟹直徑之直方圖.....	5
圖 6 螃蟹高度之盒鬚圖.....	5
圖 7 螃蟹高度之直方圖.....	5
圖 8 螃蟹重量之盒鬚圖.....	6
圖 9 螃蟹重量之直方圖.....	6
圖 10 不含外殼重量之盒鬚圖	6
圖 11 不含外殼重量之直方圖	6
圖 12 內臟重量之盒鬚圖.....	7
圖 13 內臟重量之直方圖.....	7
圖 14 外殼重量之盒鬚圖.....	7
圖 15 外殼重量之直方圖.....	7
圖 16 螃蟹年齡之盒鬚圖.....	8
圖 17 螃蟹年齡之直方圖.....	8
圖 18 各項數值型資料對螃蟹年齡之散佈圖	8
圖 19 相關係數矩陣圖.....	9
圖 20 研究流程圖	10
圖 21 移除離群值後資料盒鬚圖	11

表目錄

表 1 資料變數說明.....	2
表 2 各項變數之統計量.....	3
表 3 檢驗一般線性模型迴歸係數表格.....	12
表 4 檢驗交互作用模型迴歸係數表格.....	12
表 5 步驟 8 模型迴歸係數表格.....	13
表 6 步驟 9 模型迴歸係數表格.....	14
表 7 步驟 10 模型迴歸係數表格.....	15
表 8 三項模型之指標分數與複判定係數.....	15
表 9 刪除資料點.....	20

1. 研究動機與目的

以各種動物而言，人們常以動物的大小來粗略估計其年齡；以人類為例，我們可以簡易地通過肉眼判斷一個人大概是青年期或是壯年期；又或者是魚類生物，研究學者發現可以透過其身上的鱗片或是耳石進行年齡測定(Beamish and McFar Lane, 1983)。反觀甲殼類生物因為其堅硬的結構限制，所以在生長過程是不連續的(Hartnoll, 1982)，因此不容易透過外骨骼來對甲殼類生物進行年齡測定。另外，甲殼類也不存在類似魚類生物身上的鱗片或是耳石可以進行年齡測定。為了達到生長目的，甲殼類動物會在換殼期間迅速生長，等到新的外殼重新生長完畢便又回到之前的成長水準。因此要判別一隻甲殼類生物的年齡並非是一件易事。

事實上，有許多專家學者針對甲殼類生物的年齡測定進行研究。例如，尺寸頻率分析(Size-frequency analysis, Hartnoll, 1982)，透過甲殼類生物的長度或是寬度，以及離散化年齡區間的方法進行年齡測定，該方法也同樣適用於生命週期短的生物年齡測定上。其次還有透過脂褐質(Lipofuscin, Sheehy 1990; Ju et al. 1999; Puckett et al. 2008)，一種大量存在於甲殼類生物大腦的螢光腦色素，的組織切片或螢光分光光度計分析(spectrofluorometric analysis)來實現年齡測定。另外，Butler(1961)則是透過新外殼的寬度以及舊外殼的寬度建構迴歸模型來進行年齡測定。

進行年齡測定的好處在於，生物的年齡信息構成了研究該物種生活史的基礎以及健全管理、生產力，以及估計存活率等重要性。此一方面也可以應用到漁業和水產養殖上，例如在捕撈時以成熟期的甲殼類生物作為目標，而將尚不至成熟期的放回海中生長，避免過度捕撈導致族群滅絕。對於養殖業者而言可以更為妥善管理其養殖的甲殼類生物、決定是否或是何時該捕撈，和達到降低養殖成本等目的。

綜上所述，本研究將以 Kaggle 網站上的波士頓地區螃蟹資料希望利用螃蟹的身體特徵，如螃蟹的高度、長度，以及重量等因子，建構一迴歸模型來預測其年齡。此處將假設該資料的螃蟹為同一時期和同一種類的螃蟹進行下列章節的分析。本研究將在第二章介紹資料以及變數定義；第三章進行探索式資料分析以及視覺化分析；第四章說明研究方法、建立迴歸模型，以及做模型評估；最後在第五章作總結與討論。

2. 資料說明與變數定義

2.1 資料來源與說明

此份報告使用的波士頓螃蟹資料[7]來自 Kaggle 網站¹，其作者為 Gursewak Singh Sidhu。資料內容主要為螃蟹的性別以及外表特徵，如螃蟹的長度、高度，以及重量等，資料總筆數為 2,660 筆。資料目的係透過螃蟹的外表特徵預測其年齡，以幫助養殖業者決定是否以及何時收穫螃蟹進而達到降低成本以及管理螃蟹。

2.2 變數定義

表 1 資料變數說明

特徵名稱	資料型態	特徵說明	變數類型	備註
Sex	類別型	螃蟹性別	自變數	M(雄性)；F(雌性)
Length	數值型	螃蟹長度(單位:英尺)	自變數	1 英尺=30.48 公分 1 盎司=28.35 公克
Diameter	數值型	螃蟹直徑(單位:英尺)	自變數	
Height	數值型	螃蟹高度(單位:英尺)	自變數	
Weight	數值型	螃蟹重量(單位:盎司)	自變數	
Shucked Weight	數值型	不含外殼重量(單位:盎司)	自變數	
Viscera Weight	數值型	內臟重量(單位:盎司)	自變數	
Shell Weight	數值型	外殼重量(單位:盎司)	自變數	
Age	數值型	螃蟹年齡(單位:月)	應變數	

¹ Kaggle 網站創立於 2010 年，是一個提供數據建模以及數據分析的競賽平台，企業或是研究學者亦可在該平台提供數據集或是相關競賽。

3. 探索性資料分析

3.1 敘述性統計

從下表 2 中，首先可以從第一欄確定該份資料不存在遺漏值，因此無須作遺漏值處理。接著比較各項變數的標準差，其中發現到有關螃蟹身上重量資料的標準差皆偏大，也就是說螃蟹身上重量資料的分散程度較大，其中又以螃蟹重量的標準差為最大。此處從最大值和最小值判斷應有離群值存於重量資料之中而導致標準差有較大的現象。

表 2 各項變數之統計量

	Length	Diameter	Height	Weight	Shucked Weight	Viscera Weight	Shell Weight	Age
Count	2660	2660	2660	2660	2660	2660	2660	2660
Mean	1.423	1.116	0.386	28.817	12.421	6.303	8.260	10.915
Std	0.239	0.196	0.094	12.821	5.989	2.880	3.642	3.058
Min	0.387	0.275	0.037	0.439	0.184	0.085	0.141	3.0
25%	1.287	1.0	0.337	19.855	8.093	4.323	5.808	9.0
50%	1.462	1.150	0.387	28.462	12.232	6.166	8.079	10.0
75%	1.587	1.250	0.437	36.560	16.063	8.136	10.347	12.0
Max	2.037	1.625	2.825	80.101	42.184	21.545	28.491	29.0

3.2 視覺化呈現與分析

圖 1 為呈現雄雌螃蟹各自的比例。比例上雄性大於雌性 7.8%，其中資料中的雌性螃蟹筆數為 1,225 筆(46.1%)和雄性螃蟹筆數為 1,435 筆(53.9%)。

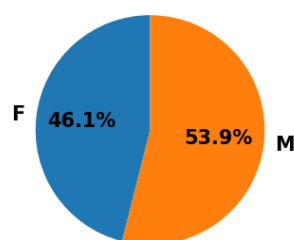


圖 1 螃蟹性別之圓餅圖

圖 2 為螃蟹長度盒鬚圖，雄雌性資料皆存在可能的離群值，其中以雄性資料可能的離群值較多。另外，也從圖 3 可觀察到資料的分佈偏左尾。圖 4 為螃蟹直徑盒鬚圖，資料狀況同長度的狀況，其分佈也偏左尾(如圖 5)。圖 6 的螃蟹高度盒鬚圖中可以明顯看到雄性與雌性各出現一筆明顯的離群值，也需要對此兩筆資料做相關處理。另外，圖 7 可觀察到資料的數值較小，分佈的峰度較大

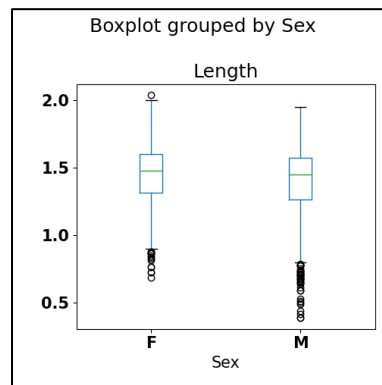


圖 2 螃蟹長度之盒鬚圖

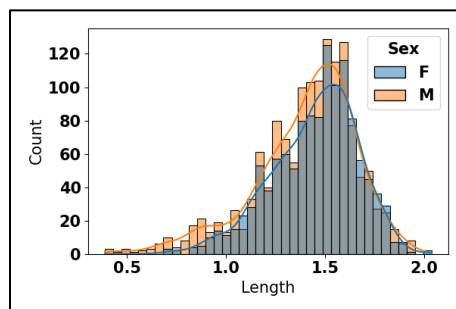


圖 3 螃蟹長度之直方圖

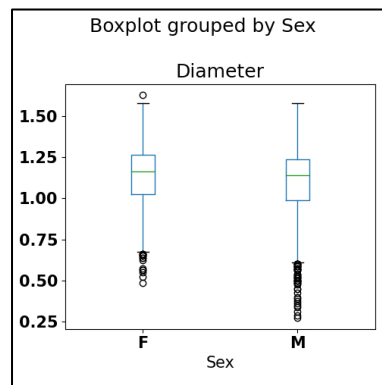


圖 4 螃蟹直徑之盒鬚圖

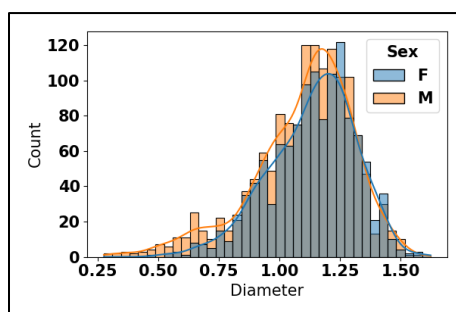


圖 5 螃蟹直徑之直方圖

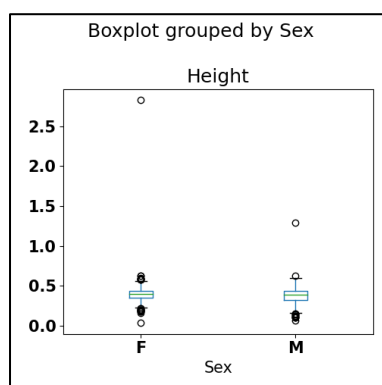


圖 6 螃蟹高度之盒鬚圖

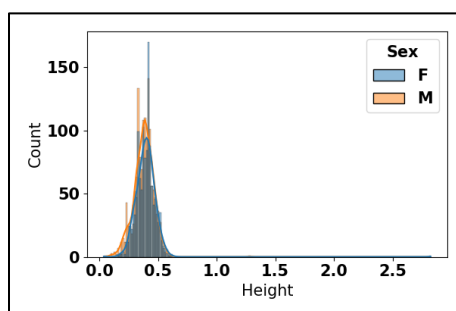


圖 7 螃蟹高度之直方圖

圖 8 為螃蟹重量盒鬚圖，雄雌性資料皆存在可能的離群值，其中以雄性資料可能的離群值較多。另外，圖 9 可觀察到螃蟹重量資料的分佈偏右尾。圖 10 為不含外殼重量盒鬚圖，可以觀察到雌性資料有一筆較為明顯的離群值，因此需做相關處理。另外，圖 11 則可看出可觀察到不含外殼重量資料的分佈偏右尾，且資料較為集中。圖 12 為內臟重量盒鬚圖，可以觀察到雌性資料有一筆較為明顯的離群值以及雄性資料也有兩筆，因此需做相關處理。另外，從圖 13 可看出內臟資料的分佈較近似常態。圖 14 為外殼重量盒鬚圖，也是有些許的離群值，因此需做相關處理。另外，從圖 15 可看出外殼重量資料的分佈近似常態。

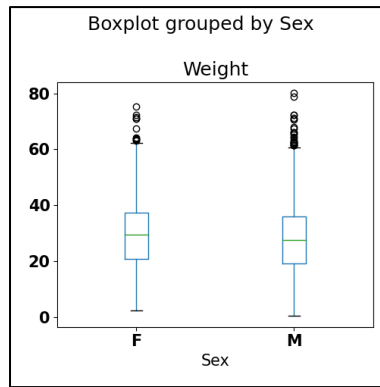


圖 8 螃蟹重量之盒鬚圖

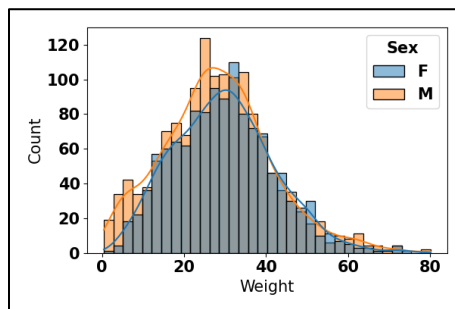


圖 9 螃蟹重量之直方圖

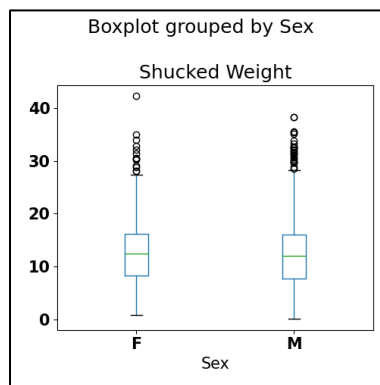


圖 10 不含外殼重量之盒鬚圖

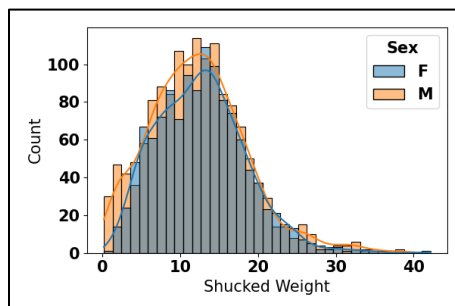


圖 11 不含外殼重量之直方圖

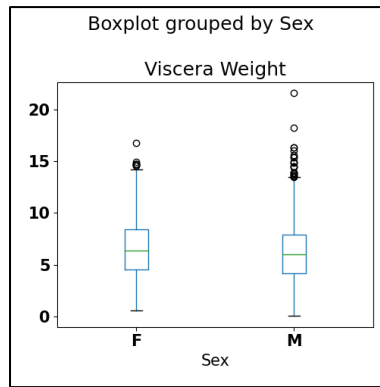


圖 12 內臟重量之盒鬚圖

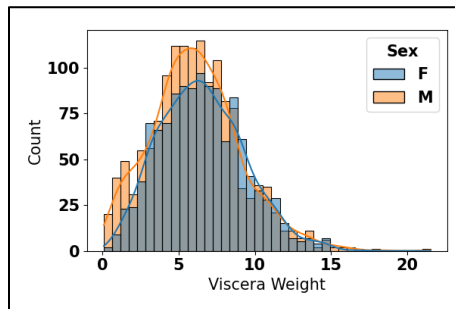


圖 13 內臟重量之直方圖

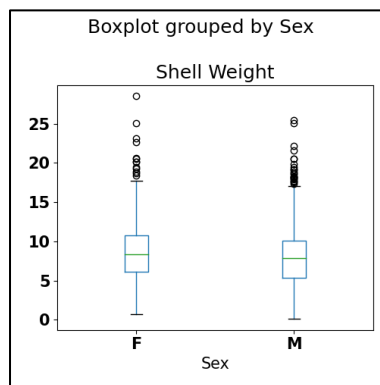


圖 14 外殼重量之盒鬚圖

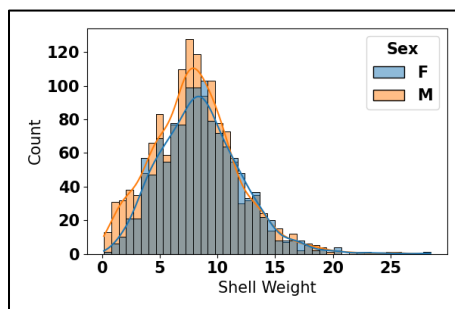


圖 15 外殼重量之直方圖

圖 16 為螃蟹年齡盒鬚圖，可發現到雄性與雌性資料狀況相近。圖 17 為螃蟹年齡直方圖，可觀察到其分佈有明顯的右尾狀況，因此可以考慮對螃蟹年齡做資料轉換。

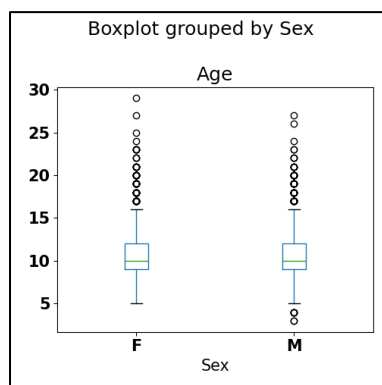


圖 16 螃蟹年齡之盒鬚圖

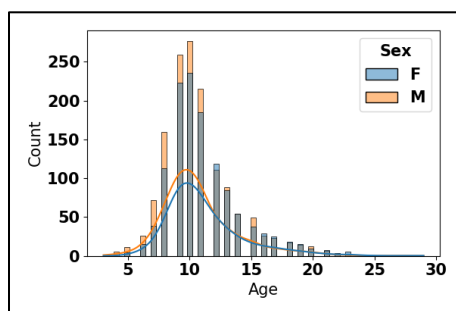


圖 17 螃蟹年齡之直方圖

圖 18 為各項數值型資料對螃蟹年齡散佈圖並以螃蟹性別做區分，可從圖中觀察到各項資料對於螃蟹年齡皆為正相關。

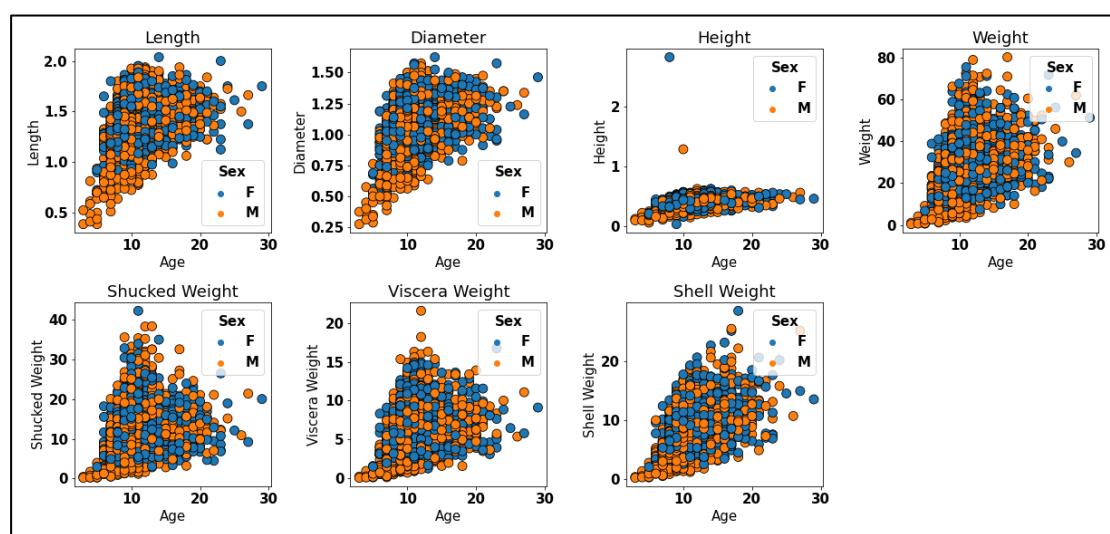


圖 18 各項數值型資料對螃蟹年齡之散佈圖

接著進一步透過圖 19 來觀察各項變數之間的相關係數，此處為計算皮爾森相關係數。與螃蟹年齡相關性最高的是螃蟹外殼重量，其餘預測變數與螃蟹年齡的相關性都滿低的；另外，也可以看到自變數彼此之間的相關性都很高。

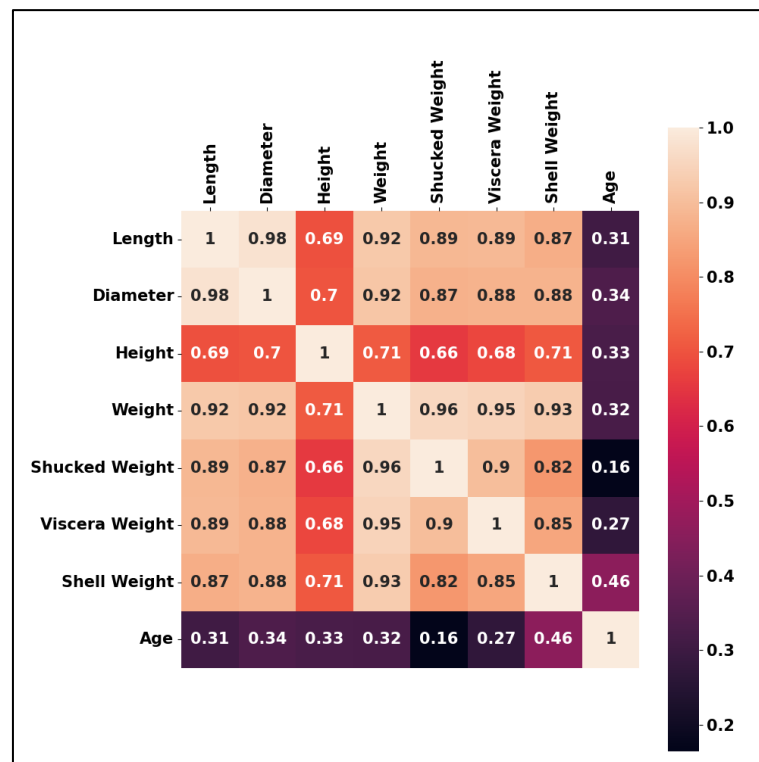


圖 19 相關係數矩陣圖

4. 研究方法與模型

4.1 流程圖

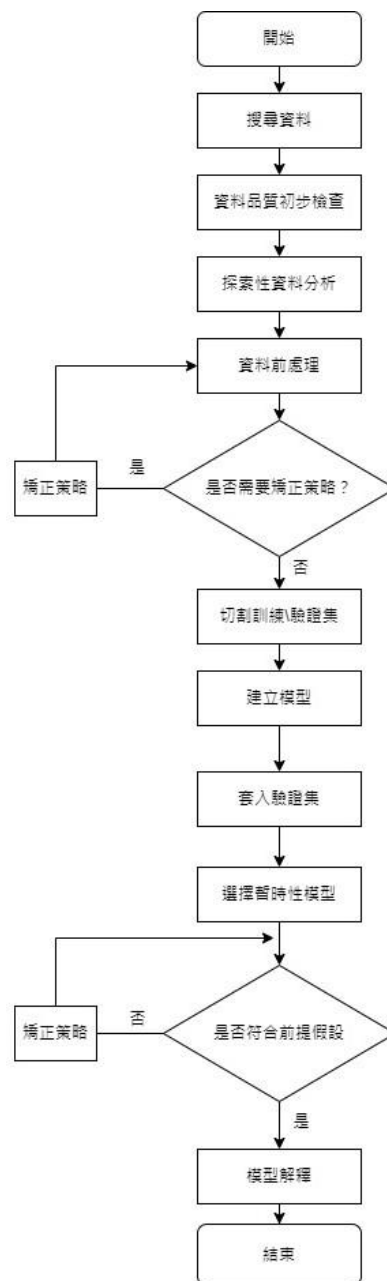


圖 20 研究流程圖

4.2 資料前處理

1. 類別型資料處理：對螃蟹性別做標籤編碼(label encoding)，其中雌性(F)被標為 0；雄性(M)被標為 1。
2. 離群值處理：除了螃蟹年齡之外，將超過該特徵欄位 3 倍四分位數間距(interquartile range, IQR)的資料點移除，移除結果如圖 21。總共移除 11 筆資料點。(參考附錄)

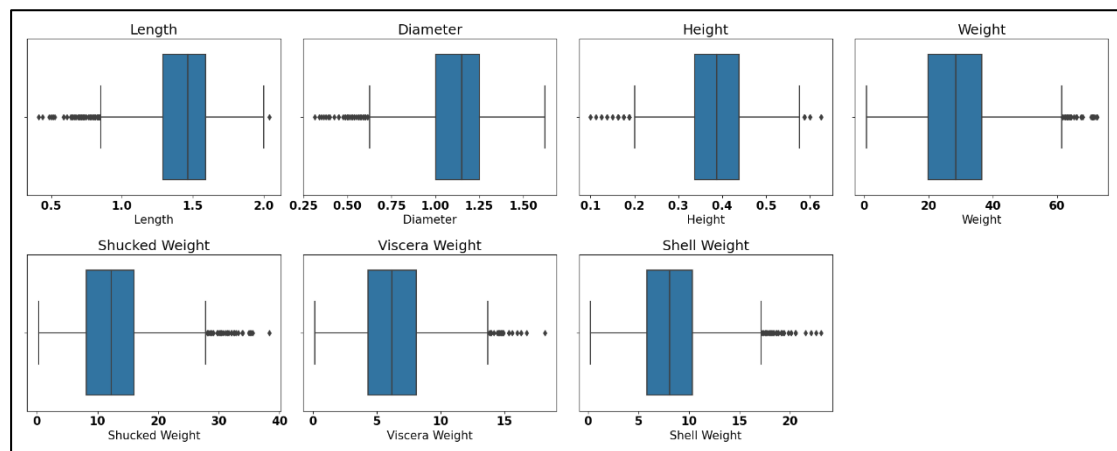


圖 21 移除離群值後資料盒鬚圖

3. 標準化：除了螃蟹年齡之外，對其餘數值型資料做標準化。
4. 切割資料集：本研究將原 2,649 筆資料以 8:2 的比例，分別切割為訓練集(共 2,119 筆)以及驗證集(共 530 筆)。

4.2 模型選擇

1. 先針對自變數建構一般線性模型選擇對模型有效的自變數，其結果如表 3。接著對迴歸係數做推論，此處設定顯著水準 $\alpha = 0.05$ ，則 $t(0.975; 2110) = 1.9610$ 。在此條件下，僅剔除螃蟹性別此一預測變數，其餘皆保留。
2. 接著考慮交互作用，建構交互作用迴歸模型選擇對模型有效的交互作用項，其結果如表 4。接著對迴歸係數做推論，此處設定顯著水準 $\alpha = 0.01$ ，則 $t(0.995; 2091) = 2.5781$ 。在此條件下，保留螃蟹長度×螃蟹直徑、螃蟹重量×不含外殼重量，以及外殼重量×不含外殼重量。

3. 基於上述兩個步驟可得到模型需考慮的所有預測變數，接著使用向前選取法 (forward selection procedure) 來挑選暫時性模型。透過向前選取法，在設定 $\alpha = 0.15$ 的條件下，選擇後三者做為暫時性模型，分別為步驟 8、步驟 9，以及步驟 10 模型。(參照表 5、表 6、表 7)
4. 接著將驗證集分別套入步驟 8、步驟 9，以及步驟 10 模型，並透過模型的複判定係數、調整後的複判定係數、*MSE*，以及 *MAE* 作為挑選最佳模型的指標分數。透過表 8 可以觀察到步驟 8 模型的 *MSE* 以及 *MAE* 最低，因此本研究選擇步驟 8 作為最後模型。

表 3 檢驗一般線性模型迴歸係數表格

Term	Coef.	SE Coef.	T-Value	P-Value
Constant	10.922	0.0766	142.51	0
Length	-0.532	0.26	-2.05	0.041
Diameter	0.724	0.256	2.83	0.005
Height	0.576	0.108	5.35	0
Weight	4.516	0.44	10.26	0
Shucked Weight	-4.244	0.232	-18.26	0
Viscera Weight	-1.144	0.175	-6.53	0
Shell Weight	1.046	0.207	5.06	0
Sex	0.022	0.105	0.21	0.837

表 4 檢驗交互作用模型迴歸係數表格

Term	Coef.	SE Coef.	T-Value	P-Value
Constant	10.9884	0.0824	133.31	0
Length	-0.8	0.273	-2.93	0.003
Diameter	0.443	0.264	1.68	0.094
Height	0.365	0.112	3.25	0.001
Weight	6.388	0.615	10.39	0
Shucked Weight	-5.531	0.32	-17.29	0
Viscera Weight	-1.282	0.228	-5.63	0
Shell Weight	1.006	0.285	3.53	0
Length*Diameter	-0.786	0.259	-3.03	0.002
Length*Height	0.333	0.551	0.6	0.546
Length*Weight	-4.73	1.99	-2.38	0.018
Length*Shucked	2.88	1.07	2.69	0.007

Weight				
Length*Viscera	0.244	0.8	0.31	0.76
Weight				
Length*Shell	1.686	0.93	1.81	0.07
Weight				
Diameter*Height	-0.589	0.525	-1.12	0.262
Diameter*Weight	1.31	2.03	0.65	0.518
Diameter*Shucked				
Weight	0.5	1.1	0.45	0.65
Diameter*Viscera				
Weight	0.422	0.834	0.51	0.613
Diameter*Shell				
Weight	-0.542	0.89	-0.61	0.543
Height*Weight	-0.36	0.748	-0.48	0.631
Height*Shucked				
Weight	0.351	0.395	0.89	0.374
Height*Viscera				
Weight	0.009	0.306	0.03	0.976
Height*Shell				
Weight	0.127	0.294	0.43	0.667
Weight*Shucked				
Weight	0.212	0.424	0.5	0.618
Weight*Viscera				
Weight	0.496	0.655	0.76	0.449
Weight*Shell				
Weight	0.78	0.436	1.79	0.074
Shucked				
Weight*Viscera	-0.069	0.482	-0.14	0.886
Weight				
Shucked				
Weight*Shell	-1.665	0.393	-4.24	0
Weight				
Viscera				
Weight*Shell	-0.626	0.374	-1.68	0.094
Weight				

表 5 步驟 8 模型迴歸係數表格

Term	Coef.	SE Coef.	Type II SS	F-Value	P-Value
------	-------	----------	------------	---------	---------

Intercept	10.97064	0.06271	165413	30608.2	<.0001
Height	0.38348	0.10964	66.10797	12.23	0.0005
Weight	4.63832	0.43025	628.06786	116.22	<.0001
Shucked Weight	-4.74160	0.23059	2285.08558	422.83	<.0001
Viscera Weight	-1.12075	0.16915	237.25435	43.90	<.0001
Shell Weight	1.57343	0.20653	313.67483	58.04	<.0001
Length*Dia meter	-0.92082	0.08992	566.68551	104.86	<.0001
Length*Shu cked Weight	1.60883	0.17751	443.90016	82.14	<.0001
Shucked Weight*Shel l Weight	-0.69086	0.11428	197.51601	36.55	<.0001

表 6 步驟 9 模型迴歸係數表格

Term	Coef.	SE Coef.	Type II SS	F-Value	P-Value
Intercept	11.09625	0.07190	127755	23818.8	<.0001
Length	-0.51075	0.19132	38.22514	7.13	0.0077
Height	0.35118	0.10952	55.14661	10.28	0.0014
Weight	4.85892	0.43166	679.59229	126.70	<.0001
Shucked Weight	-4.82914	0.24075	2158.13051	402.36	<.0001
Viscera Weight	-1.05211	0.17175	201.27330	37.53	<.0001
Shell Weight	1.87547	0.21704	400.49158	74.67	<.0001
Length*Dia meter	-0.32129	0.10011	55.24031	10.30	0.0014
Length*Shu cked Weight	1.15668	0.12495	459.61262	85.69	<.0001

Length*Shell Weight	-1.00398	0.14204	267.96537	49.96	<.0001
---------------------	----------	---------	-----------	-------	--------

表 7 步驟 10 模型迴歸係數表格

Term	Coef.	SE Coef.	Type II SS	F-Value	P-Value
Intercept	11.07451	0.07289	123688	23084.3	<.0001
Length	-0.84252	0.26687	53.40294	9.97	0.0016
Diameter	0.45620	0.25598	17.01841	3.18	0.0749
Height	0.34344	0.10955	52.66146	9.83	0.0017
Weight	4.83813	0.43160	673.29720	125.66	<.0001
Shucked Weight	-4.85444	0.24104	2173.23911	405.60	<.0001
Viscera Weight	-1.04642	0.17169	199.03249	37.15	<.0001
Shell Weight	1.80473	0.22053	358.83550	66.97	<.0001
Length*Diameter	-0.32447	0.10008	56.32260	10.51	0.0012
Length*Shucked Weight	1.16501	0.12498	465.60312	86.90	<.0001
Length*Shell Weight	-0.98456	0.14239	256.19130	47.81	<.0001

表 8 三項模型之指標分數與複判定係數

指標分數	步驟 8	步驟 9	步驟 10
R-sq(Train)	0.4321	0.4307	0.428
R-sq(Test)	0.3812	0.3794	0.3833
R^2-Adj(Train)	0.4294	0.4283	0.4258
R^2-Adj(Test)	0.3693	0.3687	0.3738
MSE(Train)	5.3431	5.3561	5.3813
MSE(Test)	5.1885	5.2032	5.171
MAE(Train)	1.6905	1.692	1.7011
MAE(Test)	1.6839	1.6865	1.6837

4.3 模型評估

在模型評估中，本研究先確認步驟 8 模型的前提假設是否滿足，也就是誤差是否滿足常態性、變異同質性，以及獨立性。

1. 首先利用 Shapiro 檢定驗證誤差是否滿足常態性。此處給定顯著水準 $\alpha = 0.01$ ，透過計算 p 值為接近 0，也就是小於顯著水準，因此拒絕誤差滿足常態性。
2. 為了滿足常態性的前提假設，此處對螃蟹年齡做 Box-Cox 轉換，得到 $\lambda = -0.2635$ ，因此對螃蟹年齡取 -0.2635 次方。
3. 再重新利用 Shapiro 檢定來驗證誤差是否滿足常態性。此處給定顯著水準 $\alpha = 0.01$ ，透過計算 p 值為 0.0154 大於顯著水準，因此不拒絕誤差滿足常態性。
4. 接著利用卡方檢定驗證誤差是否滿足變異同質性。此處給定顯著水準 $\alpha = 0.01$ ，透過計算 p 值為接近 1，也就是遠大於顯著水準，因此不拒絕誤差滿足變異同質性。
5. 最後是利用 Durbin-Watson 檢定驗證誤差是否滿足獨立性。透過計算 Durbin-Watson 統計量可以發現接近 2，也就是說不拒絕誤差滿足獨立性。

另外是共線性檢驗，透過計算 VIF 可以發現螃蟹重量、不含外殼重量，以及螃蟹長度×不含外殼重量的 VIF 值明顯大於 10，因此模型存在共線性。

最後是模型複判定係數為 0.4176，也就是說透過上述預測變數可解釋螃蟹年齡達 0.4176。

4.4 模型解釋

步驟 8 的迴歸方程式如下，

$$\begin{aligned} Age^{-0.2635} = & 10.9706 + 0.383Height + 4.638Weight - 4.742ShuckedWeight \\ & - 1.121VisceraWeight + 1.573ShellWeight - 0.9208Length \\ & * Diameter + 1.609Length * ShuckedWeight \\ & - 0.691ShuckedWeight * ShellWeight, \end{aligned}$$

其中，螃蟹長度*螃蟹直徑可解釋為螃蟹最大面積、螃蟹長度*不含外殼重量可解釋為螃蟹所需的產能，也就是螃蟹需要攝取的能量，以及不含外殼重量*外殼重量為螃蟹所需的耗能。因此模型解釋如下，

1. 當其他數值固定，螃蟹高度變動 1 英尺，年齡會增加 0.383 個月。
2. 當其他數值固定，螃蟹重量變動 1 盎司，年齡會增加 4.638 個月。
3. 當其他數值固定，不含外殼重量變動 1 盎司，年齡會減少 4.742 個月。
4. 當其他數值固定，內臟重量變動 1 盎司，年齡會減少 1.121 個月。
5. 當其他數值固定，外殼重量變動 1 盎司，年齡會增加 1.573 個月。
6. 當其他數值固定，螃蟹最大面積變動 1 平方英尺，年齡會減少 0.9208 個月。
7. 當其他數值固定，螃蟹所需的產能變動 1 單位，年齡會增加 1.609 個月。
8. 當其他數值固定，螃蟹所需的耗能變動 1 單位，年齡會減少 0.691 個月。

5. 結果與討論

事實上，以複判定係數而言，無論是哪個模型其實都有低度擬合(underfitting)的狀況，即便是最終模型也是有相同問題。所以本研究結論為，透過上述的解釋變數無法有效地解釋螃蟹年齡；另外，在部分對於模型的解釋有些不是那麼直觀，甚至有些是違背直覺的，例如螃蟹長度*螃蟹直徑的迴歸係數是負值就不太直觀。當然不排除使用其他機器學習的方法可以做到更好的結果，例如在未進行資料前處理的情況下，直接套入隨機森林模型的複判定係數可以達到 0.61。由此可知，若資料進行處理後預期結果會更好。另外，本研究有考慮二次項迴歸模型，但關係到模型是否可以被清楚地解釋，因此並沒有納入書面報告之中。但考慮二次項迴歸模型複判定係數大約落在 0.48，也是有低度擬合狀況。因此若希望可以透過迴歸分析的方法做出更好的結果可以多納入其他的因子，例如季節、螃蟹種類，或是其他有利的因子。

6. 參考資料

- [1] Beamish, R. J. & G. A. McFarlane, (1983). The forgotten requirement for age validation in fisheries biology. *Transactions of the American Fisheries Society* 112: 735–743.
- [2] Butler, T. H. (1961). Growth and age determination of the Pacific edible crab *Cancer magister* Dana. *Journal of the Fisheries Board of Canada*, 18(5), 873-891.
- [3] Hartnoll, R. G., (1982). Growth. In Bliss, D. E. & L. G. Abele (eds.), *The Biology of Crustacea*, 2, Embryology, Morphology and Genetics. Academic Press, New York: 111–196.
- [4] Sheehy, M. R. J., (1990a). Potential of morphological lipofuscin age-pigmented as an index of crustacean age. *Marine Biology* 107: 439–442.
- [5] Ju, S.-J., D. H. Secor & H. R. Harvey, (1999). Use of extractable lipofuscin for age determination of blue crab *Callinectes sapidus*. *Marine Ecology Progress Series* 185: 171–179.
- [6] Kilada, R., & Driscoll, J. G. (2017). Age determination in crustaceans: a review.

Hydrobiologia, 799(1), 21-36.

[7] Crab Age Prediction | Kaggle

<https://www.kaggle.com/datasets/sidhus/crab-age-prediction?resource=download>

[8] How Old is That Crab?, Alaska Department of Fish and Game

https://www.adfg.alaska.gov/index.cfm?adfg=wildlifenews.view_article&articles_id=845

附錄

表 9 刪除資料點

	Sex	Length	Diameter	Height	Weight	Shucked Weight	Viscera Weight	Shell Weight	Age
693	M	0.3875	0.2875	0.0625	0.6803 88	0.2551 46	0.1417 48	0.2126 21	5
2765	M	0.3875	0.275	0.1	0.4394 17	0.1842 72	0.0850 49	0.1417 48	3
748	M	1.7625	1.4125	1.2875	62.652 4	31.397 07	13.792 03	14.514 94	10
2256	F	1.1375	0.8875	2.825	16.839 6	9.4120 34	3.2885 42	3.7846 58	8
3542	F	1.5875	1.2375	0.0375	32.786 2	14.500 77	8.7316 46	8.1788 31	9
2571	F	1.95	1.575	0.5375	75.324 62	42.184 06	14.132 23	16.612 81	11
772	M	1.9375	1.575	0.625	78.797 44	38.229 3	21.545 62	16.386 01	12
2725	F	1.7875	1.4125	0.6	62.354 73	20.539 21	13.182 52	25.089 31	17
2896	M	1.825	1.4875	0.575	80.101 51	32.502 7	11.878 44	25.429 5	17
3069	M	1.6625	1.3375	0.5625	61.901 13	21.361 35	11.084 65	25.089 31	27
3295	F	1.8125	1.4	0.525	60.696 28	18.427 18	11.283 1	28.491 25	18