

動物收容所收容率與認領養率對自然死亡率的關聯性研究--以公立收容所為例

研究動機：

在美國，動物收容所面臨著資源短缺和動物數量激增的挑戰。根據2023年的報導，疫情後經濟壓力、獸醫費用上升和住房限制等因素，導致美國多處動物收容所犬隻爆滿，收容所動物數量增加了近25萬隻。此外，通貨膨脹和經濟不景氣也使得美國養狗成本上升，進一步加劇了收容所的壓力。在此背景下，許多收容所和動物福利組織開始探索新的方法來應對挑戰。雖然目前缺乏直接使用大數據技術成功降低死亡率的報導，但隨著科技的進步，未來可能會有更多相關的案例出現。

在台灣，隨著流浪動物數量的增加，動物收容所面臨著日益嚴峻的收容壓力。在資源有限的情況下，高收容率可能導致收容所的照護品質下降，進而增加動物的死亡率。了解收容率與死亡率之間的關聯性，對於制定有效的政策和改善收容所管理至關重要。

研究背景：

2015年，農委會修正《動物保護法》，宣布自2017年2月6日起，全國公立動物收容所禁止撲殺「經通知或公告逾12日而無人認領、認養或無適當之處置」的動物，曾經的12夜悲歌就此消失，但是零撲殺會帶來什麼問題呢？據農委會統計，全國遊蕩犬數量在110年已達159,697隻，且相對於107年的146,773不減反增。這意味著收容所的動物只會不斷增加，各地收容所都面臨超收的問題，政府投入相當多的預算依據「源頭減量、改善收容、多元認養」原則，推動各項措施，但成效有限。

研究目的：

動物福利是社會文明進步的象徵，減少收容所中動物的死亡率是提升動物福利的核心目標之一。透過科學數據分析，研究收容率如何影響死亡率，將為改善收容所內部運作提供依據，並幫助相關單位提升動物的生活品質。

政府與非營利組織在資源分配時，通常缺乏基於數據的分析支持。本研究希望透過探索收容率與死亡率的關聯性，提供決策參考，幫助資源更有效率地分配給有迫切需求的收容所。

大數據技術在許多領域已證明其價值，卻很少被應用於動物收容所管理。本研究旨在利用公開數據，建立量化模型，找出關鍵影響因素，為未來的智能化收容所管理奠定基礎。

動物收容所的高死亡率問題往往被社會忽視。本研究不僅希望透過數據揭露現狀，更期望以數據引發公眾對流浪動物議題的關注，促進更多人參與解決流浪動物問題的行動。

研究計劃:

一、初步資料準備

- 載入並清理資料，包括異常值處理與缺失值補全。
- 視覺化數據分佈，識別初步趨勢。

二、描述性與推論性分析

- 生成摘要統計與視覺化圖表。
- 計算變數間的相關性並進行假設檢定。

三、建模與分群

- 執行迴歸模型，量化收容率與死亡率的關係。
- 應用 K-Means 分群法，探索資料中的潛在模式。

四、視覺化

- 建立進階視覺化圖表，呈現多變量關係與地區趨勢。

五、報告與結果

- 總結每個分析步驟的發現與見解。
- 提出對收容所管理與政策制定的啟示與建議，為動物福利與資源分配提供實證依據。

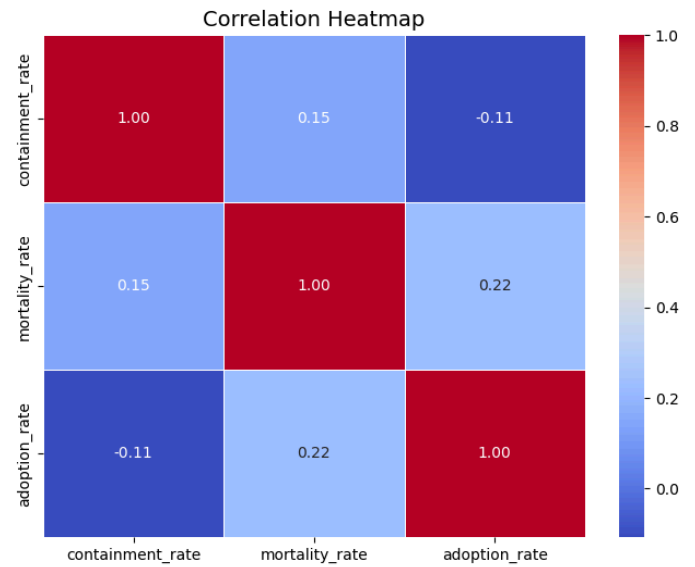
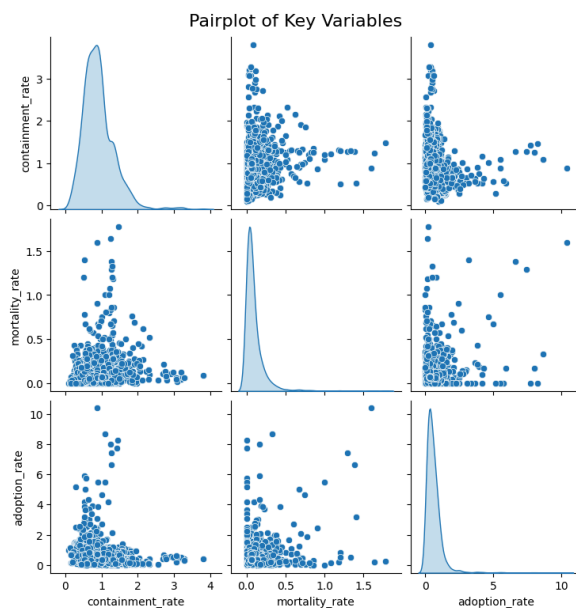
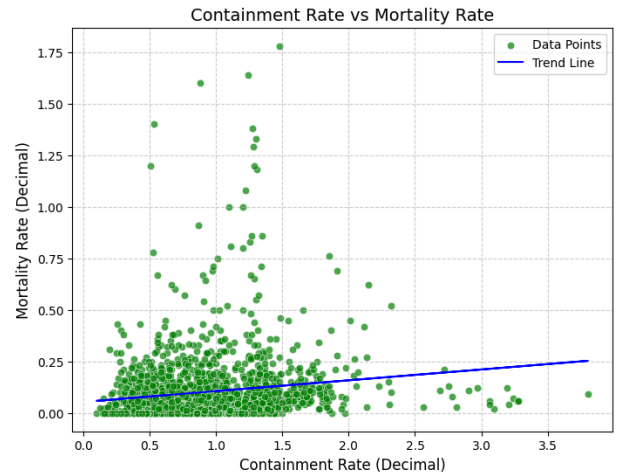
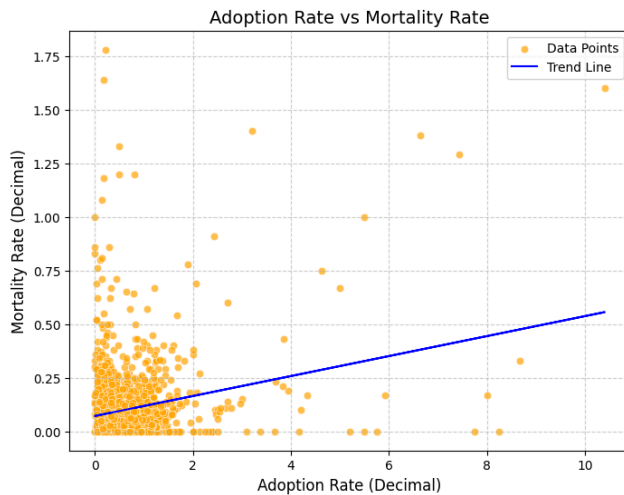
研究方法:

一、資料清理與前處理

- 異常值處理
 - 計算數值變數（收容率、死亡率、認領養率）的四分位距（Interquartile Range, IQR）。
 - 移除超出四分位距 1.5 倍範圍的數據點，以確保分析的穩健性。
- 缺失資料處理
 - 對於缺失值，根據資料特性採取平均值、中位數或 KNN 方法進行插補。
 - 對缺失值過多的資料列，考量移除以避免影響分析結果。

二、探索性資料分析 (EDA)

- 變數間的關聯性分析
 - 使用相關矩陣與散點圖分析收容率、死亡率與認領養率等關鍵變數之間的關係。
 - 針對主要變數計算皮爾森相關係數 (Pearson Correlation) 及 p 值，以驗證變數間的顯著性關聯。



2.描述性統計

總結數值欄位的集中趨勢(平均數、中位數)與離散程度(標準差、變異數)，全面了解資料的分佈特性。

a. 收容率盒鬚圖

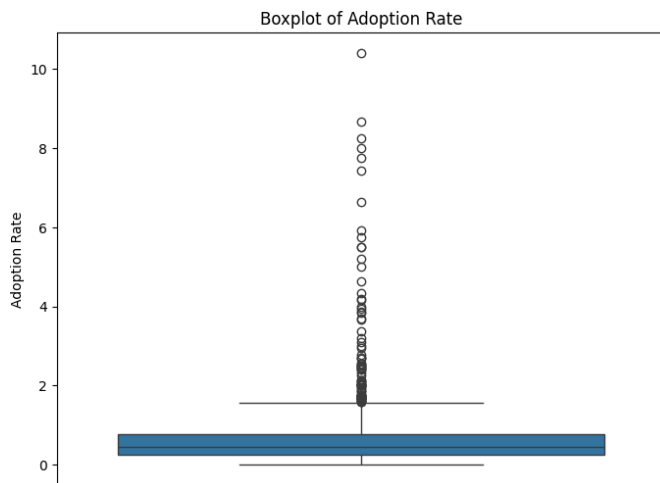
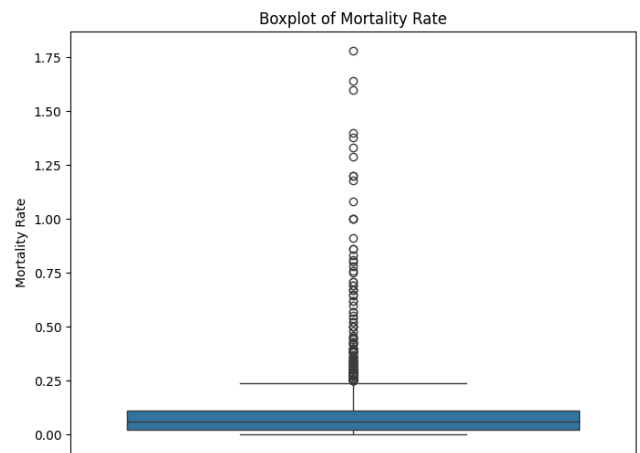
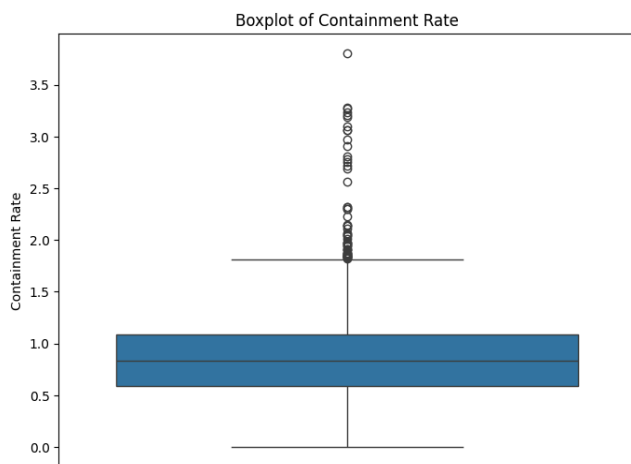
- 中位數: 接近0.5, 表示大多數數據點的收容率都集中在這個值附近。
- 四分位範圍: 主要集中在0到1.0之間, 這表示大部分的收容率都在這個範圍內。
- 異常值: 有相當數量的異常值分布在1.0到3.5之間, 指出有些收容所的收容率遠高於大多數。

b. 死亡率盒鬚圖

- 中位數: 非常接近0.25, 顯示大部分收容所的死亡率低於這個數值。
- 四分位範圍: 數據非常集中, 四分位範圍接近0到0.5, 表示大多數收容所的死亡率都很低。
- 異常值: 有數據點高達1.5以上, 這些可能是管理或條件極差的收容所。

c. 認領養率盒鬚圖

- i. 中位數:約0.5, 和收容率的中位數相似, 表明一半的收容所有較低的認領養率。
 - ii. 四分位範圍:主要集中在0到約2.0, 這表明大多數收容所的認領養率相對較低。
 - iii. 異常值:高達10的異常值顯示有些收容所的認領養率異常高, 可能因特殊事件或促銷活動。
- d. 單變量分佈
- i. 收容率:呈現右偏分佈, 大多數數據集中在較低的數值, 但也有一些較高的異常值。
 - ii. 死亡率:同樣呈現右偏分佈, 絕大多數的數據點都集中在低值區域, 顯示大部分收容所的死亡率較低。
 - iii. 認領養率:分佈非常右偏, 大部分數據集中在極低的範圍內, 但有少數收容所表現出非常高的認領養率。



分析結果解釋

- 收容率對死亡率:散點圖顯示收容率與死亡率之間的正相關, 這可能暗示當收容所收容的動物數量增加時, 可能由於資源分配問題導致死亡率上升。
- 認領養率對死亡率:另一散點圖顯示認領養率與死亡率之間的輕微負相關, 這表示提高認領養率可能有助於減少死亡率, 可能由於認領養出的動物減少了收容所的負擔。

三、特徵工程

1. 虛擬變數 (Dummy Variables)

a. 將類別欄位 (如縣市) 轉換為虛擬變數, 方便後續的迴歸與分群分析。

```
Mean Squared Error: 0.0197672749307188
Model Coefficients: [ 0.00768715  0.10044819  0.02612542  0.03724786  0.03054995 -0.01848178
-0.02395075  0.06880426 -0.01476442  0.03451724  0.02497621 -0.01991036
 0.17258037  0.09121976  0.02741796  0.02019268  0.01569657  0.00369104
-0.01220635  0.00647847  0.25395202  0.00392027]
Coefficient for Containment Rate: 0.00768715169941142
```

分析結果解釋:

MSE: 0.01967

這個值表示模型預測的平均誤差的平方, 較低的MSE指示著較好的模型預測性能。

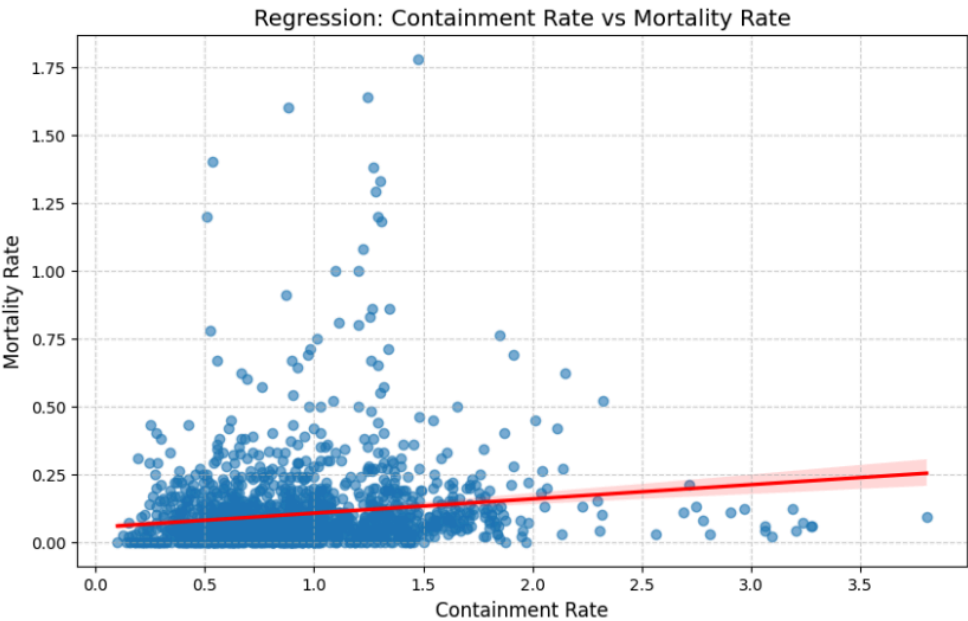
收容率的係數: 0.00768715

這個係數表示當收容率增加一個單位時, 死亡率預計會增加大約 0.0077 個單位, 暗示了收容率和死亡率之間存在輕微的正相關關係。

四、統計與機器學習分析

1. 迴歸建模

a. 執行線性迴歸分析, 評估收容率與認領養率對死亡率的預測能力。

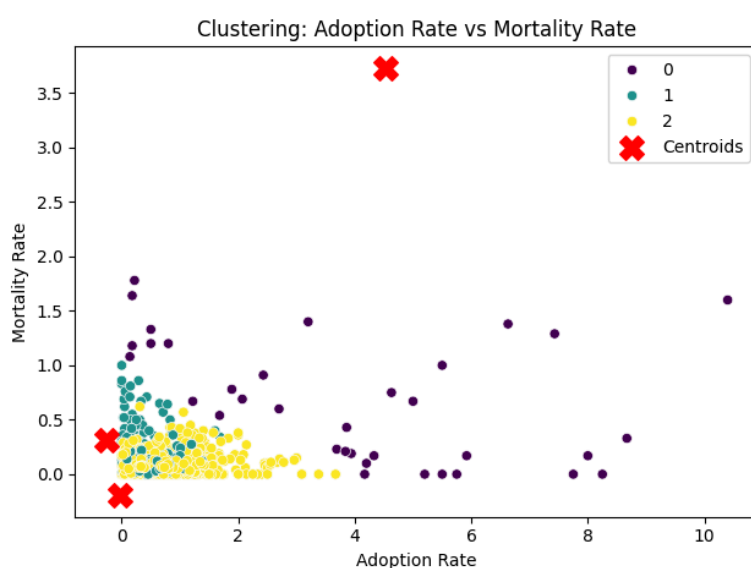


模型解釋：

趨勢線表明，在控制其他因素不變的情況下，收容率的增加與死亡率輕微上升有關。這可能表明當收容所負載增加時，可能因資源、管理或醫療照顧不足等因素導致死亡率上升。

2. 分群分析

- 應用 K-Means 分群法，辨識不同收容所或地區的模式與趨勢。
- 使用標準化的數值變數進行分群，確保變數間權重一致。
- 程式碼整合：已準備好標準化資料以供 K-Means 使用，可能需進一步處理 NaN 值。



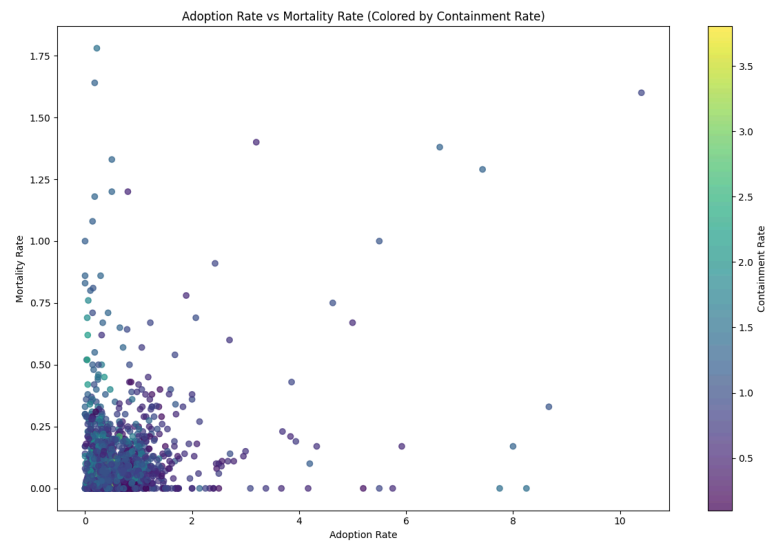
聚類分析結果：

- 低認領養率和低死亡率(黃色點)：這一群組包括死亡率和認領養率都較低的點，可能代表管理良好的收容所，其中動物被認領的比例不高，但死亡率也相對較低。
 - 中等認領養率和中等死亡率(藍色點)：死亡率和認領養率都位於中等範圍的收容所，可能面臨中等程度的運營壓力。
 - 高認領養率和高死亡率(紫色點)：死亡率和認領養率都相對較高，這可能表明一些收容所即使認領養率高，也可能因各種壓力或挑戰導致較高的死亡率。
-

五、進階視覺化

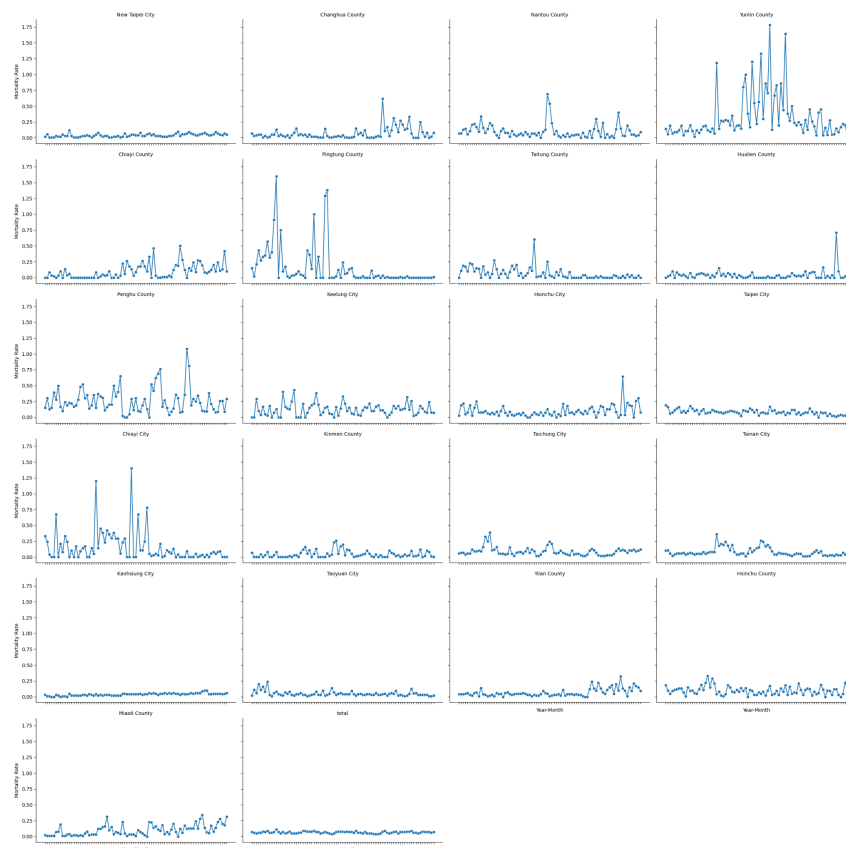
1. 多變量散點圖

- 以顏色區分方式，同時視覺化認領養率、死亡率與收容率，呈現變數間的交互影響。

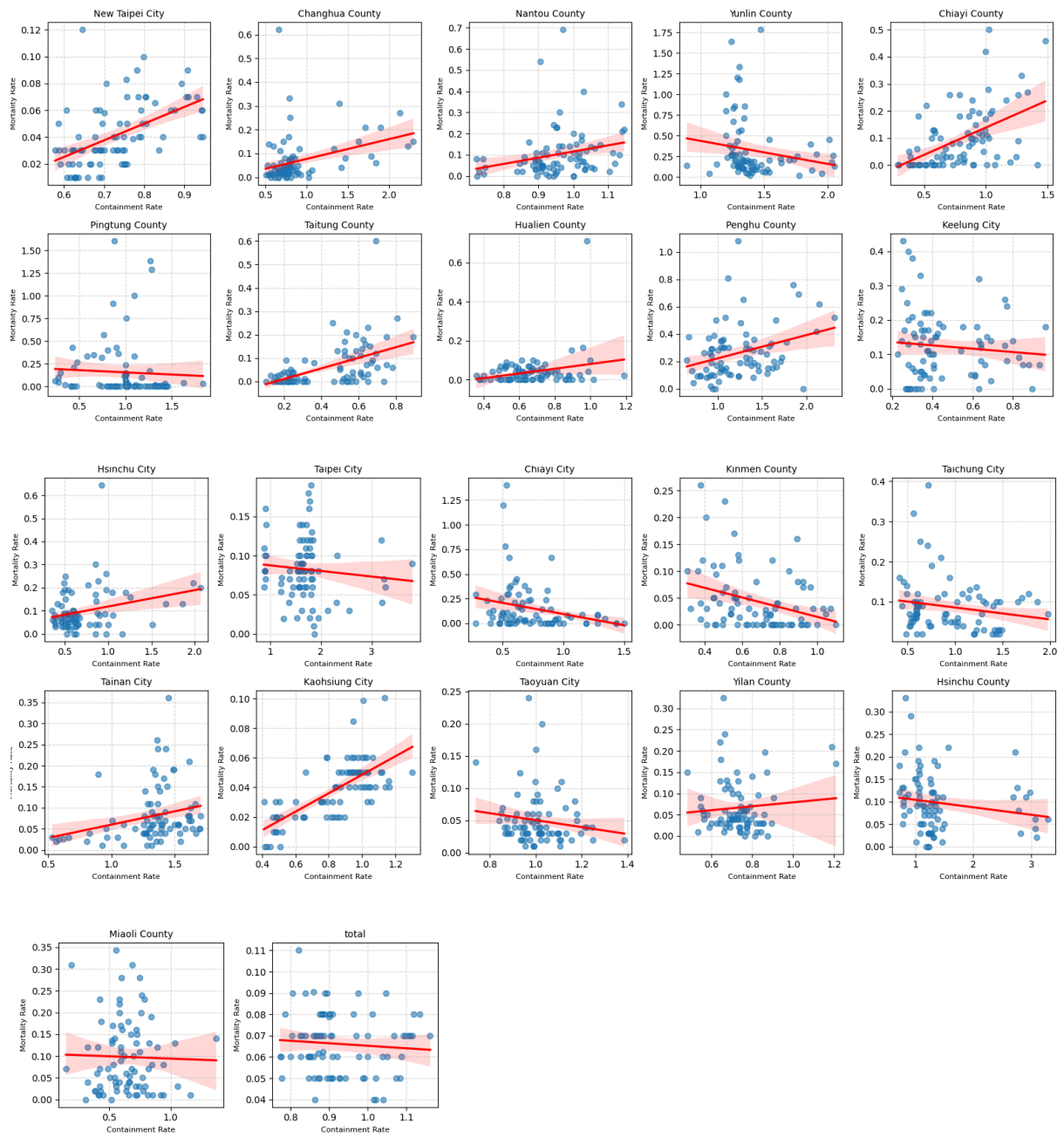


2. 分面趨勢分析

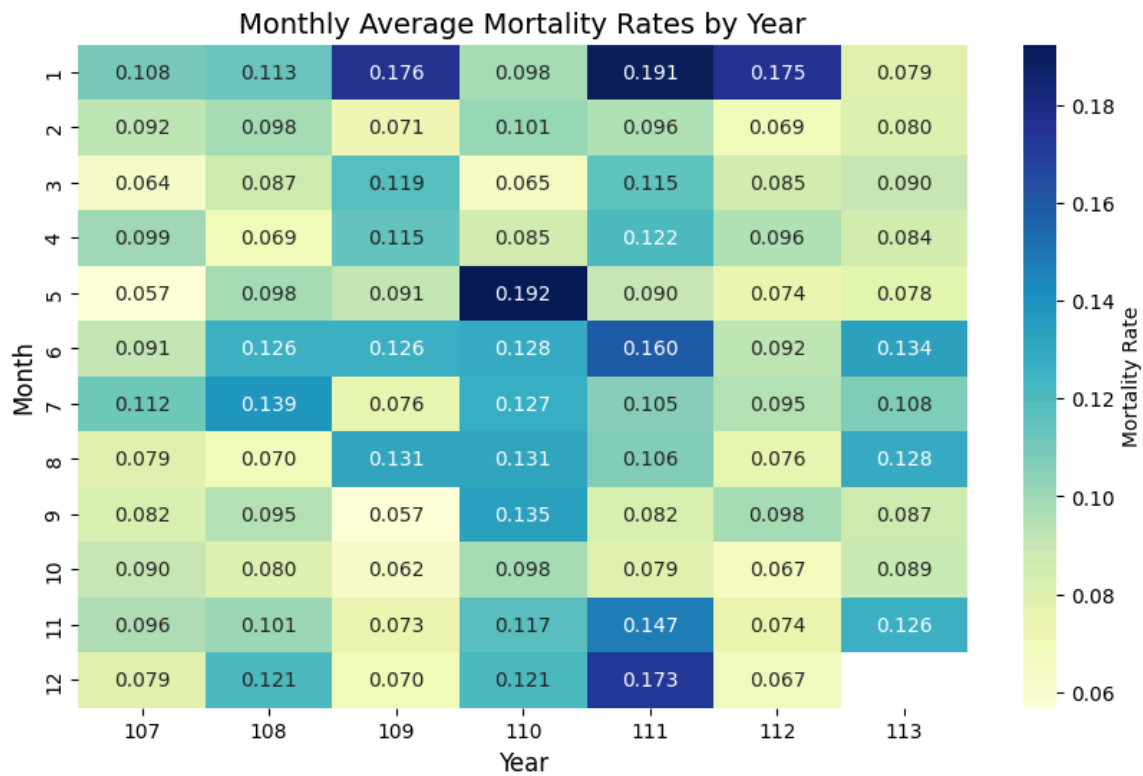
生成分面折線圖，展示各地區的動物收容趨勢與變化。



3. 分區迴歸分析

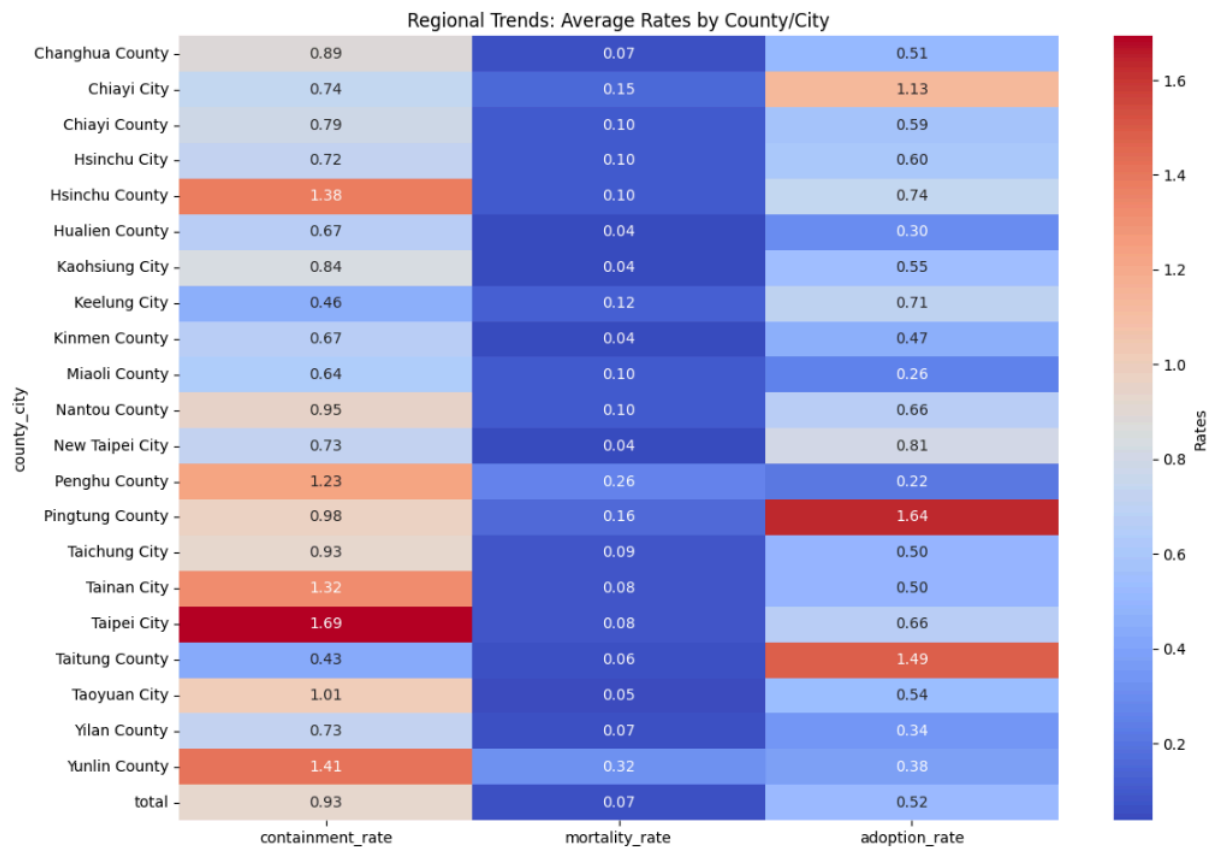


4. 熱圖分析各年份平均死亡率



分析結果解釋：我們可以看到在冬季、夏季，氣溫較為極端的狀態下，死亡率都偏高，所以我們推論氣候對於死亡率有些微影響，我們推薦政府可以考慮在極端氣候月份增強收容所的設施，例如改善冷暖氣系統，以保護動物免受極端氣候的影響。

5. 熱圖分析分區比率平均



分析結果解釋：

高收容率與高死亡率：某些地區如台東縣和雲林縣展示出較高的收容率和死亡率，這可能指出收容所面臨的壓力大，可能因資源分配不足導致較高的死亡率。

低死亡率與高認領養率：如彰化縣和苗栗縣，展示了低死亡率和相對較高的認領養率，這可能表示這些地區的動物福利政策或社區參與度較高。

特別注意點：屏東縣的認領養率特別高，而其收容率和死亡率並不顯著，這可能反映出有效的社區支持和認領養推廣策略。

六、結論

本研究顯示，雖然收容率和認領養率無法完全解釋收容所自然死亡率的變化，但它們之間的弱正相關在統計上具有顯著性，且在地區層面上存在明顯差異。

顯示政府可以從這方面下手，例如擴展收容所、增加認領養率等，在最後的熱區分析圖(屏東縣)，可以看出很高的認領養率也許對降低死亡率有幫助，但是認領養率需要大眾有意願才能提高，在推行上也許有困難，因此嚴格的源頭管理也非常重要，不過我們能夠提出目前能夠達到的方法。

整體而言，比率上也許相關性非常低，但是分區探討後，發現各區的狀況是非常不同的，所以在制定相關政策時，政府應考慮地區差異，針對性地分配資源，優化管理系統，並鼓勵公眾參與。通過這些措施，可逐步改善全台收容所內動物的福利，最終實現流浪動物照護與收容系統的全面提升。

資料來源：

<https://www.sPCA.org.tw/post/dpm>

https://tw.news.yahoo.com/%E5%8B%95%E7%89%A9%E6%94%B6%E5%AE%B9%E6%89%80%E9%81%AD%E7%88%86%E7%85%A7%E9%A1%A7%E7%92%B0%E5%A2%83%E5%B7%AE-%E8%BE%B2%E5%A7%94%E6%9C%83-%E5%8D%94%E5%8A%A9%E5%9C%B0%E6%96%B9%E6%94%BF%E5%BA%9C%E6%94%B9%E5%96%84-061600448.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAACt7TBR_zSgwuBUBOMTjL6_4zRL8rKviBzwB7qLtRhylGsSVa4-j_TCkJb5fvkGMHpbBHxVU57CToL4546iQ_sbGYWmrLUcHWFpFCrcEtAvAKNg1WKaZHceSEzqnCjRimknuKzTaTYVU7nufco-uCJbViM7zFILRLqdmux3zq3Ad

<https://ndltd.ncl.edu.tw/cgi-bin/gs32/gswweb.cgi/login?o=dnclcdr&s=id=%22105NTU05541008%22.&searchmode=basic#XXX>

<https://e-info.org.tw/node/238347>

<https://tw.news.yahoo.com/%E5%85%AC%E7%AB%8B%E5%8B%95%E7%89%A9%E6%94%B6%E5%AE%B9%E6%89%80%E9%81%9B%E7%8B%97%E8%AA%BF%E6%9F%A5-%E5%8B%95%E7%9D%A3%E7%9B%9F-%E6%85%98-071930853.html>

數據來源：[動物保護資訊網](#) x DS_dataset_shelter.xlsx