

國立高雄大學

課程：大數據分析與 AI 創新應用

111學年 第2學期 期末議題專題報告

**題目：運動數據分析**

|  |
| --- |
| 系所：工藝與創意設計學系  姓名： 劉依如  學號：A1111322 |

指導教授：曾智義 教授

中　華　民　國 112 年 05 月27日

目 錄

1. 議題描述及問題設定

1.1議題介紹

1.2分析問題構想

1.3分析目標建立

1.4分析流程圖

1. 檢定假說建立

2.1配速差異假說檢定

2.2步頻差異假說檢定

2.3疲累度差異假說檢定

1. 數據集收集

3.1數據來源

3.2獲取方式

1. 數據清理

4.1整理和清理數據

4.2資料欄位及用途

1. 統計檢定或模型建立

5.1統計檢定

1. 程式編碼及說明
2. 分析結果及意涵
3. 專題心得
4. 議題描述及問題設定

1.1議題介紹

本研究旨在應用課堂所學的跑者數據與健身數據做應用，進行運動數據分析與疲累度、配速、步頻、跑步效率和卡路里消耗的比較，來評斷跑者在一段時間內是否有進步，在哪種距離的跑步下效率最高。

1.2分析問題構想

* 跑者的進步評估：

比較不同時間段的配速、步頻和跑步效率，以評估跑者在一段時間內是否有進步。

比較不同時間段的疲累度，分析是否有明顯降低。

* 最佳距離與效率：

分析不同距離下的跑步效率，觀察哪種距離下跑步效率最高。

* 卡路里消耗：

分析不同時間段的卡路里消耗，觀察跑者在一段時間內的運動量變化。

將卡路里消耗與其他指標進行對比，例如疲累度和配速，分析它們之間的關聯性。

* 數據可視化：

使用散點圖、趨勢圖、直方圖將各項指標進行可視化，呈現數據間的關係和趨勢。

進行統計分析，例如 t 檢定等，以確定結果的統計顯著性。

1.3分析目標建立

* 評估跑者的進步：

比較不同時間段內的配速、步頻、跑步效率等指標。

分析進步的趨勢和幅度，並探討可能影響進步的因素。

比較疲累度在跑步前期與後期的變化，以了解持續運動對疲累度的影響。

* 研究跑步效率和最佳距離：

探討跑步效率在不同距離下的變化，找出最佳跑步距離範圍。

* 分析卡路里消耗和運動量：

統整卡路里的總消耗量，並且與正常人數據做比較查看運動量的大小。

* 數據可視化和統計分析：

使用適當的圖表和圖形來視覺化跑者數據和相關指標之間的關係，如散點圖、趨勢圖、直方圖等。

進行統計分析，例如 t 檢定、相關分析等，以確定數據之間的統計顯著性和相互關聯性。

1.4分析流程圖

一張含有 螢幕擷取畫面, 文字, 字型, 設計 的圖片

自動產生的描述

圖一:分析流程圖

1. 檢定假說建立  
   2.1配速差異假說檢定：

假說 H0：前半部分和後半部分的平均配速沒有顯著差異。

假說 H1：前半部分和後半部分的平均配速存在顯著差異。

2.2步頻差異假說檢定：

假說 H0：前半部分和後半部分的平均步頻沒有顯著差異。

假說 H1：前半部分和後半部分的平均步頻存在顯著差異。

2.3疲累度差異假說檢定：

假說 H0：前半部分和後半部分的疲累度沒有顯著差異。

假說 H1：前半部分和後半部分的疲累度存在顯著差異。

1. 數據集收集

3.1數據來源

創造數據而成，每個數據都有經過嚴謹查證確保數據合理性。

3.2獲取方式

使用隨機生成的函式，限定合理範圍，產生一位跑者的數據庫。

1. 數據清理

4.1整理和清理數據

由於所有數據都是隨機創造，因此沒有空值或異常值等問題出現，故此部分略過。

4.2資料欄位及用途

* Distance(距離)：記錄跑步的總距離，用於分析跑步的遠近和持久度。
* Time(時間)：記錄跑步的總時間，用於計算平均配速和平均速度，以及分析跑步的時長和持久度。
* pace(配速)：計算每公里所需的時間，用於評估跑步的速度和進步情況。
* heart rate(bpm) (心率)：監測跑步期間的心率變化，用於評估運動強度和身體的耐力狀態。
* Steps(步數)：記錄跑步期間的步數，用於評估運動量和活動強度。
* average speed(平均速度)：計算每小時的平均移動速度，用於評估整體運動速度和進步情況。
* running efficiency(跑步效率)：評估跑步的效率和技巧，通常與配速和步頻相關聯。
* MET(代謝當量)：衡量運動的強度，以卡路里消耗和身體代謝為基礎。
* weight(體重)：記錄跑步時的體重，用於計算卡路里消耗和評估運動的效果。
* calorie expenditure(卡路里消耗) ：計算跑步期間消耗的總卡路里，用於評估運動量和能量消耗。
* Fatigue(疲累度)：評估跑步後的身體疲勞程度，用於分析運動後的恢復情況和訓練效果。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| distance | time | pace | heart rate(bpm) | Steps | average speed |
| *9.644279* | *44.22268* | *4.58538* | *144* | *153* | *13.08507* |
| *6.639295* | *30.35073* | *4.571379* | *155* | *187* | *13.12514* |
| *8.790513* | *23.40063* | *2.662032* | *130* | *178* | *22.53917* |
| *6.189303* | *52.62393* | *8.502401* | *132* | *182* | *7.056831* |
| *9.386724* | *30.25359* | *3.223019* | *135* | *173* | *18.61608* |
| *8.855563* | *47.77657* | *5.395091* | *162* | *181* | *11.12122* |
| *7.228557* | *40.9288* | *5.662098* | *171* | *183* | *10.59678* |
| *6.975317* | *21.18643* | *3.037343* | *167* | *189* | *19.75411* |
| *8.499068* | *34.45864* | *4.054402* | *124* | *186* | *14.79873* |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| running efficiency | MET | weight | calorie expenditure | Fatigue |
| *9.086852* | *8* | *62.5* | *368.5223* | *8* |
| *8.467834* | *8* | *62.2* | *251.7087* | *5* |
| *17.33782* | *8* | *61.5* | *191.8852* | *6* |
| *5.346084* | *8* | *63.1* | *442.7427* | *10* |
| *13.78969* | *8* | *62.5* | *252.1133* | *5* |
| *6.864952* | *8* | *62.2* | *396.227* | *9* |
| *6.196947* | *8* | *61.5* | *335.6161* | *8* |
| *11.82881* | *8* | *63.1* | *178.2485* | *5* |
| *11.93446* | *8* | *62.5* | *287.1553* | *5* |

表一:前十項資料欄位

1. 統計檢定或模型建立

5.1統計檢定

用於比較前半部分和後半部分的平均配速、步頻和疲累度之間是否存在顯著差異。使用stats.ttest\_ind函數進行t-檢定，計算t值和p值。。

1. 程式編碼及說明

6.1這段程式碼主要是為了將名為"running\_data.csv"的數據文件讀取到程式中，以便後續進行數據分析和可視化操作。

|  |
| --- |
| 1. import numpy as np 2. import pandas as pd 3. from scipy import stats 4. import matplotlib.pyplot as plt 5. # 讀取數據 6. data = pd.read\_csv("running\_data.csv") |

6.2這段程式碼使用t檢定，對前後部分的配速（pace）數據進行分析和比較，並且繪製出平均配速分布圖。

|  |
| --- |
| 1. # 將數據分為前半部分和後半部分 2. first\_half\_pace = data.loc[0:299, "pace"] 3. second\_half\_pace = data.loc[300:599, "pace"] 4. # 進行 t 檢定 5. t\_statistic, p\_value = stats.ttest\_ind(first\_half\_pace, second\_half\_pace, equal\_var=False) 6. # 繪製前半部分和後半部分平均配速的分布圖 7. plt.figure(figsize=(8, 6)) 8. plt.hist(first\_half\_pace, color="blue", alpha=0.5, label="first\_half\_pace") 9. plt.hist(second\_half\_pace, color="green", alpha=0.5, label="second\_half\_pace") 10. plt.title("Average pace distribution") 11. plt.xlabel("Pace (min/km)") 12. plt.ylabel("Frequency") 13. plt.legend() 14. plt.show() 15. # 印出結果 16. print("t 檢定結果:") 17. print("t 值：{:.2f}".format(t\_statistic)) 18. print("p 值：{:.4f}".format(p\_value)) 19. # 解讀結果 20. alpha = 0.05 21. if p\_value < alpha: 22. print("根據統計結果，在前半部分和後半部分的平均配速之間存在顯著差異。") 23. else: 24. print("根據統計結果，在前半部分和後半部分的平均配速之間沒有顯著差異。") |

6.3這段程式碼使用t檢定，對前後部分的步頻（Steps）數據進行分析和比較，並且繪製出平均步頻分布圖。

|  |
| --- |
| 1. # 將數據分為前半部分和後半部分 2. first\_half\_Steps = data.loc[0:299, "Steps"] 3. second\_half\_Steps = data.loc[300:599, "Steps"] 4. # 進行 t 檢定 5. t\_statistic, p\_value = stats.ttest\_ind(first\_half\_Steps, second\_half\_Steps, equal\_var=False) 6. # 繪製前半部分和後半部分平均步頻的分布圖 7. plt.figure(figsize=(8, 6)) 8. plt.hist(first\_half\_Steps, color="blue", alpha=0.5, label="first\_half\_Steps") 9. plt.hist(second\_half\_Steps, color="green", alpha=0.5, label="second\_half\_Steps") 10. plt.title("Average Steps distribution") 11. plt.xlabel("Steps") 12. plt.ylabel("Frequency") 13. plt.legend() 14. plt.show() 15. # 印出結果 16. print("t 檢定結果:") 17. print("t 值：{:.2f}".format(t\_statistic)) 18. print("p 值：{:.4f}".format(p\_value)) 19. # 解讀結果 20. alpha = 0.05 21. if p\_value < alpha: 22. print("根據統計結果，在前半部分和後半部分的平均步頻之間存在顯著差異。") 23. else: 24. print("根據統計結果，在前半部分和後半部分的平均步頻之間沒有顯著差異。") |

6.4這段程式碼使用t檢定，對前後部分的疲累度（Fatigue）數據進行分析和比較，並且繪製出疲累度直方圖。

|  |
| --- |
| 1. # 將數據分為前半部分和後半部分 2. first\_half\_fatigue = data.loc[0:299, "Fatigue"] 3. second\_half\_fatigue = data.loc[300:599, "Fatigue"] 4. # 進行 t 檢定 5. t\_statistic, p\_value = stats.ttest\_ind(first\_half\_fatigue, second\_half\_fatigue, equal\_var=False) 6. # 繪製前半部分和後半部分疲累度的直方圖 7. plt.figure(figsize=(8, 6)) 8. plt.hist(first\_half\_fatigue, color="blue", alpha=0.5, label="first\_half\_fatigue") 9. plt.hist(second\_half\_fatigue, color="green", alpha=0.5, label="second\_half\_fatigue") 10. plt.title("Fatigue Distribution") 11. plt.xlabel("Fatigue") 12. plt.ylabel("Frequency") 13. plt.legend() 14. plt.show() 15. # 印出結果 16. print("t 檢定結果:") 17. print("t 值：{:.2f}".format(t\_statistic)) 18. print("p 值：{:.4f}".format(p\_value)) 19. # 解讀結果 20. alpha = 0.05 21. if p\_value < alpha: 22. print("根據統計結果，在前半部分和後半部分的疲累度之間存在顯著差異。") 23. else: 24. print("根據統計結果，在前半部分和後半部分的疲累度之間沒有顯著差異。") |

6.5這段程式碼用於繪製散點圖，展示距離（distance）和跑步效率（running efficiency）之間的相關性。

|  |
| --- |
| 1. # 散點圖 2. plt.scatter(data['distance'], data['running efficiency']) 3. plt.xlabel('distance') 4. plt.ylabel('running efficiency') 5. plt.title('Correlation plot between distance and running efficiency') |

6.6這段程式碼用於提取數據中的"calorie expenditure"列，並計算總卡路里消耗。

|  |
| --- |
| 1. # 提取卡路里消耗數據列 2. calorie\_expenditure = data["calorie expenditure"] 3. # 計算總卡路里消耗 4. total\_calories = calorie\_expenditure.sum() 5. print("在599天不懈地跑步下，你的總卡路里消耗：{:.2f}".format(total\_calories)) |

6.7這段程式碼是對跑步數據進行整體分析並給出建議。

|  |
| --- |
| 1. # 整體分析與建議 2. print("=============整體分析與建議=============") 3. if p\_value < alpha: 4. print("根據統計結果，你隨著時間訓練跑步配速和步頻都有顯著的進步。") 5. else: 6. print("根據統計結果，你隨著時間訓練跑步配速和步頻沒有顯著的進步。") 7. if p\_value < alpha: 8. print("根據統計結果，你隨著時間訓練運動後疲累度有明顯降低。") 9. else: 10. print("根據統計結果，你隨著時間訓練運動後疲累度沒有明顯降低。") 11. print("根據距離和跑步效率的相關性分析，你的跑步效率在長距離時是最佳的。") 12. print("在599天不懈地跑步下，你的總卡路里消耗：{:.2f}".format(total\_calories)) 13. if total\_calories > 898500: 14. print("這顯示你的運動量相當大。繼續保持良好的運動習慣！") 15. else: 16. print("這顯示你的運動量不夠多。請加強運動量，加油！") |

1. 分析結果及意涵

跑步配速：根據t檢定結果，前半部分和後半部分的平均配速之間存在顯著差異。這表示隨著時間的訓練，跑者在跑步配速方面有明顯進步。

步頻：根據t檢定結果，前半部分和後半部分的平均步頻之間存在顯著差異。這表示隨著時間的訓練，跑者的步頻有明顯提高。

疲累度：根據t檢定結果，前半部分和後半部分的疲累度之間存在顯著差異。這表示隨著時間的訓練，跑者在運動後的疲累度有明顯降低。

跑步效率：根據距離和跑步效率的相關性分析，跑者的跑步效率在長距離時是最佳的。這意味著跑者在長距離跑步時能夠保持高效率。

總卡路里消耗：在599天的持續跑步下，跑者的總卡路里消耗為253025.04。這顯示跑者的運動量不夠多，需要加強運動量。

綜合分析結果來看，跑者在配速、步頻和疲累度方面都有顯著進步，並且在長距離跑步時能夠保持高效率。然而，跑者的總卡路里消耗相對較低，這意味著跑者的運動量不夠多。建議跑者加強運動量，以進一步提升跑步表現和總卡路里消耗。

t 檢定結果:

t 值：5.66

p 值：0.0000

根據統計結果，在前半部分和後半部分的平均配速之間存在顯著差異。

t 檢定結果:

t 值：39.13

p 值：0.0000

根據統計結果，在前半部分和後半部分的平均步頻之間存在顯著差異。

t 檢定結果:

t 值：34.89

p 值：0.0000

t 檢定結果:

t 值：34.89

p 值：0.0000

根據統計結果，在前半部分和後半部分的疲累度之間存在顯著差異。

在599天不懈地跑步下，你的總卡路里消耗：253025.04

=============整體分析與建議=============

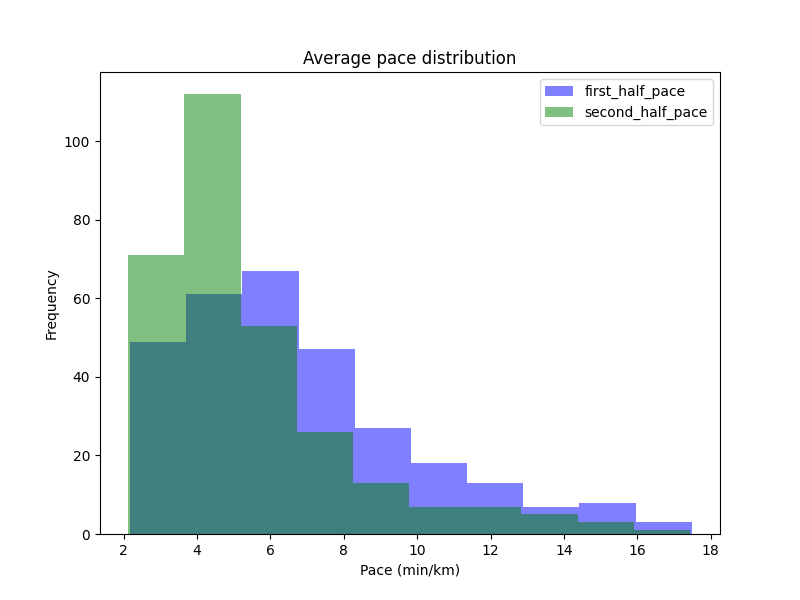
根據統計結果，你隨著時間訓練跑步配速和步頻都有顯著的進步。

根據統計結果，你隨著時間訓練運動後疲累度有明顯降低。

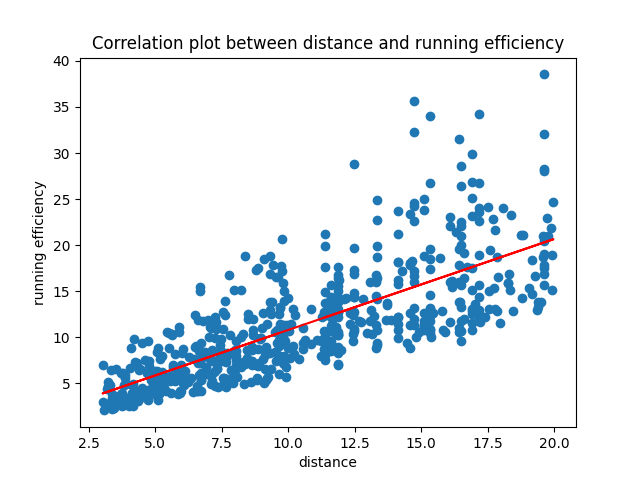
根據距離和跑步效率的相關性分析，你的跑步效率在長距離時是最佳的。

在599天不懈地跑步下，你的總卡路里消耗：253025.04

這顯示你的運動量不夠多。請加強運動量，加油！

一張含有 文字, 圖表, 螢幕擷取畫面, 繪圖 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 圖表, 繪圖 的圖片

自動產生的描述

1. 專題心得

在這個專題中，我對個人的跑步數據進行了大數據分析，並從中獲得了一些有價值的結果和洞察。通過使用統計方法，我能夠對前後部分的跑步數據進行比較，並判斷它們是否存在顯著差異。同時，我還計算了總卡路里消耗，以了解跑者在599天的跑步中消耗的總能量。

這個專題讓我深入了解運動數據中各項數值代表的意涵，並通過統計分析獲得了寶貴的結果和洞察，而我認為這樣的數據，可以應用在各個有執行著跑步計畫的人，無論是業餘還是專業，這個專題的研究都能讓他們進一步了解自己的進步與潛力在哪裡。它不僅讓使用者更加了解自己的訓練成果，也提供了今後訓練的建議和方向。這個專題讓我學習到許多大數據分析的應用以及訓練自我邏輯思維能力，對我而言受益極深。