國立中山大學應用數學系研究所

碩士論文

台灣馬拉松選手之成績預測

Race time prediction for Taiwan marathoners

研究生：江承鴻 撰

指導教授：羅夢娜 博士

中華民國　九十七　年 七 月

致謝詞

本論文得以順利完成，並通過學位考試，首先要感謝羅夢娜教授這幾年來盡心盡力的指導，尤其在資料分析與處理上給予莫大的幫助，每當遇到瓶頸時，都能不厭其煩的一一解說，不時的指點正確的方向，使我在這些年中獲益匪淺，在此致上由衷的感激。

論文口試期間承蒙郭教授美惠、張教授福春、陳教授瑞彬、黃教授錦輝等諸位口試委員提供精闢的剖析與寶貴的建議，使論文得以去蕪存菁，更為嚴謹，也在此致上由衷的敬意。

還要感謝路跑界的許多前輩們大力幫忙，讓問卷得以順利收集到足夠樣本。尤其感謝長青慢跑協會提供機會，讓晚輩於西子灣馬拉松比賽當天發放問卷，也感謝女友及孫大哥於比賽當天到處奔波請人填寫問卷。更感激參與這份問卷的諸位熱情的跑友，若沒有你們寶貴的資料，一切分析將無法如期進行，實在是萬分感謝各位熱心又熱血的跑友。

在求學過程中，感謝家人的支持，才能心無旁騖的完成碩士學位，謝謝父母的辛苦付出，以及精神上的鼓勵。除此也感謝女友的陪伴與體諒，讓愛跑步的我，可以兼顧課業與練習，並於學生生涯最後一年拿到全國大專校院運動會一般組五千與一萬公尺兩面金牌。

回首這三年的求學日子，感謝很多人直接或間接的默默幫忙，方能一路順利度過，在此對曾經幫助及關心過我的人獻上心中無限的感激。最後，謹以此文獻給家人、師長與朋友，與你們共同分享這份成果。

江承鴻 謹識

2008.07.14

摘 要

知名運動專家 **Pete Riegel** 在 1977 年提出預測跑步成績的公式，本文探討是否也適用於台灣馬拉松選手的成績預測。藉由問卷方式收集到 204 筆有效數據，問卷中加入一些可能影響成績的變數。例如：性別、年齡、跑齡、身高、體重、跑馬次數、每週練習量與次數。接下來利用複迴歸(Multiple Regression)及切片反向迴歸(Sliced Inverse Regression)來增加預測準確性。最佳模型估計結果有八成選手成績的誤差在 15 分鐘內，而 **Riegel** 的模型只有六成二。

關 鍵 字：馬拉松、複迴歸、切片反向迴歸

Abstract

Pete Riegel, a well-known sport expert, proposed the formula of race time prediction in 1977. This article discusses whether it is also suitable for Taiwan marathoners. We compiled two hundred and four effective datum by questionary. Some variables possible to affect the running result are added in this work, namely: sex, age, the year of run , height, weight, the race number of marathon, the quantity and the frequency of practices each week. Next, we use multiple regression and sliced inverse regression to increase the accuracy of the running time prediction. The best model, found here has eighty percentage's player with predictive error within fifteen minutes, which is better than the original model by Riegel(1977) with only having sixty-two percentages.

Key words: Marathon, multiple regression, sliced inverse regression.

Contents

表格 **1**

圖形 **2**

1. 前言 **3**
- **1.1** 研究背景與動機 3
- **1.2** 研究問題與目的 3
- **1.3** 名詞解釋 6

2. 研究方法與步驟 **8**
- **2.1** 問卷設計及樣本收集 8
- **2.2** 分析方法與工具 9
- **2.3** 切片反向迴歸分析法 9
- **2.4** 分析步驟 10

3. 統計分析結果 **11**
- **3.1** 變數基本分析 11
- **3.2** 模型比較 17
- **3.3** 刪掉離群值 21
- **3.4** 殘差分析 23

4. 結論與建議 **24**

參考文獻 **24**

附錄 **26**

表格

表一：1964年東京奧運會馬拉松分段時間(阿貝貝選手) 3
表二：優秀運動員的肌肉纖維組成類型 4
表三：馬拉松成績的敘述統計量 11
表四：5、10及21公里成績的敘述統計量 12
表五：年齡與跑齡的敘述統計量 13
表六：身高與體重的敘述統計量 14
表七：跑馬次數的敘述統計量 15
表八：每週練習量及次數的敘述統計量 16
表九：各模型之誤差的敘述統計量 20
表十：各模型對照表(刪掉3倍標準差) 21
表十一：各模型之誤差的敘述統計量(刪掉3倍標準差) 21
表十二：各模型對照表(刪掉2倍標準差) 22
表十三：各模型之誤差的敘述統計量(刪掉2倍標準差) 22

圖形

圖一：馬拉松成績的直方圖 11

圖二：5公里成績的直方圖 12

圖三：10公里成績的直方圖 12

圖四：21公里成績的直方圖 12

圖五：年齡的直方圖 13

圖六：跑齡的直方圖 13

圖七：身高的直方圖 14

圖八：體重的直方圖 14

圖九：跑馬次數的直方圖 15

圖十：每週練習量的直方圖 16

圖十一：每週練習次數的直方圖 16

圖十二：M1模型之散佈圖 17

圖十三：$ln(\frac{t_2}{t_1})$與前兩個SIR變量的平面圖 18

圖十四：$ln(\frac{t_2}{t_1})$與β_2'X的平面圖 19

圖十五：$ln(\frac{t_2}{t_1})$與$ln(\frac{d_2}{d_1})$的關係圖 19

圖十六：各模型之誤差的盒鬚圖 20

圖十七：常態機率圖(刪掉3倍標準差) 23

圖十八：殘差的趨勢圖(刪掉3倍標準差) 23

圖十九：殘差的時間圖(刪掉3倍標準差) 23

1. 前言

1.1 研究背景與動機

韓國影星曹承佑主演的勵志電影**馬拉松小子**，片中楚原第一次參加馬拉松並沒有順利完成，是被送上救護車回來的；不過在第二次參賽，就以低於三小時的佳績完成了。為什麼第一次無法完成比賽，原因是出在哪裡呢？如果賽前知道自己大概的成績，於比賽途中將體力作合理的分配，也就是俗稱的配速，這樣一來對於馬拉松成績勢必有所提升。配速不好的運動員，不是成績不佳，就是中途停頓棄權。就以蟬聯兩屆奧運馬拉松冠軍阿貝貝來說，就是最佳的例子。許玉芳(2006)提到阿貝貝於1964年東京奧運時，每五公里區間時間(如表一)，其中最快與最慢時間只差76秒。在長達2小時以上的激烈比賽中，能夠如此冷靜的控制速度，實屬不易。

表一：1964年東京奧運會馬拉松分段時間(阿貝貝選手)

5k	10k	15k	20k	25k	30k	35k	40k	42.195k
15'19"	14'55"	15'11"	15'23"	15'42"	16'10"	16'11"	16'09"	7'01"

許績勝(2000)建議想在馬拉松創造優異成績，應該在途中以平均且穩定的速度來跑，才可以減少無謂的體力耗損，達到體力的最高經濟效益。許樹淵(1984)討論到從力學的角度來看，能量消耗是和速度成平方的關係，即速度增加二倍，能量消耗增加四倍，因此長距離最忌忽然加速或變速，馬拉松更需依據此原理來跑。所以賽前了解自身實力，是非常重要的，然而目前國內顯少有預測馬拉松成績的相關研究，因而觸發本文的研究動機。

1.2 研究問題與目的

Riegel (1977)於跑者雜誌提出下列預測跑步成績的公式：

$$\frac{t_2}{t_1} = \left(\frac{d_2}{d_1}\right)^{1.06} \tag{1.1}$$

其中d_1代表舊距離，t_1代表跑舊距離所花時間，d_2代表新距離(文中提到的d_2皆為42.195公里)，t_2代表跑新距離所花時間(文中t_2皆為馬拉松完成時間)。d_1和d_2單位一樣，t_1和t_2單位一樣。本文要探討的問題是該模型適用於國內馬拉松選手嗎？在已知一組t_1及d_1之下，再加入一些可能會影響成績的變數，進一步提出新模型，期望能夠增加預測的準確性。這樣一來，讓國內馬拉松選手能夠初步了解大致成績，以穩定配速，跑出個人最佳成績。影響馬拉松成績的原因，本文將其分成生理、心理、環境和經驗這四方面。

一、生理方面：

(一)最大攝氧量(maximum volume of oxygen)。最大攝氧量是指人體於運動時每分鐘能被身體利用的最大氧氣數值。由於最大攝氧量是受人體體型影響，故可以用每公斤體重的攝氧量來表達，常用單位為毫升/公斤/分鐘。最大攝氧量數值愈高，顯示心肺功能愈好，數值愈低則心肺功能愈差。可藉由運動強度、時間長度與練習頻率來改善最大攝氧量，所以在問卷中收集了**每週練習量**與**每週練習次數**這兩項變數。

(二)跑步經濟性(running economy)。過去很多研究的結果證實，最大攝氧量並無法完全解釋選手在運動成績上的差異。兩位擁有不同最大攝氧量的選手，其運動成績確有顯著差異；而另外兩位擁有相同成績的選手，其最大攝氧量卻不一定相同。這樣的研究結果顯示，仍有其他因素影響耐力運動的表現，後來很多研究一致地指出跑步經濟性也是影響耐力運動的重要因素。林信甫等人(2003)建議改善跑步經濟性可藉由高強度與長間歇跑的訓練方式，而短間歇跑則較無顯著效果。

(三)肝臟與肌肉肝醣儲存(liver and muscle glycogen stores)。1967年由同一所瑞典的實驗室發表的三項研究報告，讓世人了解**肌肝醣**對運動能力的重要性。Hermansen et al.(1967)證明了肌肝醣在激烈運動會有系統地耗盡，當運動員覺得疲勞時，其體內的肌肝醣量已降至零；Ahlborg et al.(1967)證實選手疲勞的時間點與肌肉中的原肌肝醣儲存量有關；Bergstrom et al.(1967)在實驗室中證明控制飲食中的碳水化合物含量的確可以改變肌肝醣的濃度及疲勞點的時間。林貴福(2002)說到基於肌肝醣含量是耐力賽的關鍵，有人提出肌肝醣超補法。首先，從事時間較長且費力的運動去耗盡肌肝醣。再來，一邊持續訓練，一邊連續吃三天的含蛋白質與脂肪食物。最後，接下來三天不做任何運動，吃高碳水化合物食物。

(四)乳酸閾值(lactate threshold)。乳酸閾值的意義表示當運動達到一定激烈情況時，血液中乳酸濃度開始有系統地增加。Londeree(1986)證實身體的疲勞與血液中所含乳酸濃度息息相關，因此以乳酸濃度來判定耐力運動的表現是非常合乎邏輯的。Priest et al.(1987) 建議練習時的強度等於或略高於乳酸閾值，能提供最大有氧能力的改善。

(五)肌肉的組成。林貴福等人(2002)提到骨骼肌依其生化和收縮性可分為三大類，兩種快縮肌纖維**Type IIb** 纖維和**Type IIa** 纖維，和一種慢縮肌**Type I** 纖維。從下表可以了解慢縮肌纖維比例高的選手，有利於長距離比賽。

表二：優秀運動員的肌肉纖維組成類型

運動	%慢縮肌纖維	%快縮肌纖維
長距離跑者	70−80	20−30
短距離跑者	25−30	70−75

(六)性別。在各項運動裡，肌力強弱與否，往往是運動員取勝的重要因素，此情況不僅在爆發性的項目上相當明顯，連長時間的耐力運動中，最後的衝刺決勝，也相當受肌力的優劣所左右。黃永任(1994)提到針對肌力而言，一般女性為男性的三分之二，若以上半身、下半身做比較，女性上半身肌力約為男性的50%至60%，而下半身的肌力為男性的70%至80%。因此，性別的確會影響運動的表現。

(七)年齡。林貴福等人(2002)討論到年齡對運動成績有很大影響，老化會伴隨著肌肉的流失，因年齡因素造成肌肉量流失有兩階段，首先是肌肉流失**慢**的階段，從25歲到50歲肌肉量約減少10%；之後就是**快**的階段，從50到80歲又減少了40%的肌肉量。雖然肌肉量會隨老化而減少，不過規律的運動卻能改善肌力和肌耐力。

(八)體重。陳孟欣(2004)提到體重是跑步時最直接且主要的負擔，跑相同的距離則體重較重者其能量消耗也較多，因此在一定速度下跑步來測量耗氧量時，體重越重的人其氧氣消耗也相對的越多；可是以絕對攝氧量除以體重所得的相對攝氧量卻是相差不遠的，由此可見體重的影響力。飲食中把握八分飽並配合適當練習量的原則，可以鍛鍊出更適於長跑的體態。

二、心理方面：許樹淵(2000)建議賽前保持良好的心理狀態，去除多餘心理活動的能量消耗，避免造成不必要的能量損失而影響比賽。轉移緊張、害怕或興奮感，以輕鬆愉快的心情來面對比賽，對成績有所幫助。

三、環境方面：當天溫度與濕度、風向與風力、高度與坡度都會直接影響選手成績。甚至空氣污染太嚴重，也會減低心肺功能，導致心肺所受的壓力變大，直接影響選手成績。值得一提的是**溫度**對選手的影響是不公平的，例如：比賽6點起跑，優秀選手在9點前就能抵達終點，而普通選手大約10點完成，顯而易見的這一個多小時的差距，溫度可能就上升許多，對選手成績勢必造成相當影響。

四、經驗方面：跑馬拉松是門大學問，途中可能發生的狀況很多。例如：想上廁所、身體某些地方起水泡、喝水過多導致腹痛、抽筋、脫水等等。途中的補給也很重要，每個水站必須喝水或運動飲料，如汗流得多則補充鹽片。最常遇到的就是配速過快，馬拉松真正起點在32公里處，之前的距離只能算是熱身，必須穩紮穩打，保留體力，不可輕舉妄動，否則未來痛苦不堪。所以問卷中也收集了**跑齡**與**跑馬拉松次數**此兩變數，當成經驗指標。

綜合以上影響馬拉松成績因素，本文除了收集跑友的一組t_1及d_1，還挑選了**每週練習量、每週練習次數、性別、年齡、身高、體重、跑齡、跑馬次數**這些選項進入問卷當中。

1.3 名詞解釋

一、馬拉松：

馬拉松距離為42.195公里，即標準四百公尺操場跑105圈又195公尺，是一項長距離的賽跑。張保羅(2001)文中提到馬拉松從古自今，都是一個令人矚目的比賽項目，它經歷了漫長的歷史沿革，流傳著一個悲壯動人的故事。

在公元前490年，波斯帝國入侵希臘，十萬大軍壓境雅典城下。六十多歲的米爾蒂亞德統率全城軍民奮守孤城，準備與波斯人決一死戰，禦敵軍於雅典城郊約40公里的**馬拉松鎮**。希臘軍民出奇制勝，將波斯人擊潰於海灘之上，此戰使波斯入侵者丟盔卸甲，潰不成軍，這就是歷史聞名的馬拉松戰役。希臘人獲勝後，為將此捷報傳到雅典，主帥米爾蒂亞德傳令士兵**菲利浦斯**(Pheidipides)前往雅典，當菲利浦斯從馬拉松鎮的戰場上一口氣跑到雅典時，已經筋疲力盡，他用最後一口氣高喊：「我們勝利了！」之後，應聲倒下，光榮犧牲。

兩千多年後，顧拜旦(Coubertin)建議奧運會設立一個以馬拉松為名的長跑項目，紀念馬拉松戰役，緬懷菲利浦斯英靈。此建議被當時希臘委員會接受，此後，馬拉松賽跑即登上國際體育競技場。1896年第一屆雅典奧運會就設有此項，但是馬拉松跑的距離也經幾番波折，直到1924年第八屆巴黎奧運會上，才把42.195公里的距離固定下來，延續至今。通常馬拉松賽都是從田徑場出發，最後又回到田徑場抵達終點，中間段落都是漫長的公路。為此，參加選手還要掌握場地跑與公路跑不同的技術。場地跑的跑道比較鬆軟，富有彈性；公路則有瀝青路和水泥路面，堅硬且較滑，這些都需要選手進行適應性的練習。

二、複迴歸(Multiple Regression)：

迴歸分析的主題就是在討論變數間的關係，並根據一些相關的理論來建立模型。在迴歸分析中，必有一因變數(dependent variable)，一般以y表示；且可能有數個自變數(independent variable)，一般以x_i表示。如自變數只有一個，稱為簡單迴歸(Simple Regression)；若自變數有兩個以上，則稱為複迴歸(Multiple Regression)。若複迴歸模型為線性，稱為複線性迴歸，Montgomery et al.(2001)提到模型如下(假設有k個自變數，n筆數據)：

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

三、切片反向迴歸(Sliced Inverse Regression，簡稱SIR)：

Li (1991)提到SIR模型如下：

$$Y = f(X_1, X_2, \ldots, X_P, \epsilon) = g(\beta_1'\mathbf{X}, \beta_2'\mathbf{X}, \ldots, \beta_K'\mathbf{X}, \epsilon)$$

，其中ϵ與$\mathbf{X}$互相獨立，g是R^{k+1}空間上任意的函數。模型中描述一種理想的情況，即在沒有假設ϵ服從$N(0, \sigma^2)$，而且還可在沒有損失任何資訊的前提下，將多個自變數$\mathbf{X} = (X_1, X_2, \ldots, X_P)'$，例如是$P$度空間，降為較小的$K$度空間，$Y$和$\mathbf{X}$之關係僅透過$K$個$\beta_1'\mathbf{X}$、$\beta_2'\mathbf{X}$、...、$\beta_K'\mathbf{X}$之線性組合。SIR目的不在於對迴歸函數的估計，而是有效的維度縮減(effective dimension reduction，簡稱e.d.r.)。即藉由原有的多個自變數$\mathbf{X}$，投影在e.d.r.方向而減少$\mathbf{X}$的維度，並且能夠保存全部$\mathbf{X}$與Y的所有訊息。e.d.r.方向的定義為，在SIR模型之下，由β_1、β_2、...、β_K所生成的向量空間，被稱為e.d.r.空間，而在e.d.r.空間的非零向量皆可被稱為e.d.r.方向。如果反向切片迴歸的**有效的維度減少方向之個數大於一**，則最小平方法一定會漏掉價值攸關的資訊，由此可見SIR的一大優勢。即使**有效的維度減少方向等於一**，如果資料數據透過SIR來看是明顯地非線性，若用線性模型來逼近，所得結果將令人不滿意。

2. 研究方法與步驟

綜合第一章探討影響馬拉松成績的因素，本文挑選了**每週練習量**、**每週練習次數**、**性別**、**年齡**、**身高**、**體重**、**跑齡**、**跑馬次數**這些選項進入問卷當中。

2.1 問卷設計及樣本收集

問卷分成三大部分：

第一部份：基本資料。這部份是要初步了解選手間的個別差異。

(一)曾經是學校校隊、國家代表隊或以上皆非
(二)性別與年齡
(三)身高與體重
(四)跑齡
(五)初馬成績、比賽年度及名稱

第二部份：創下個人最佳馬拉松成績當時狀況。這部份是要取得選手在創下個人最佳成績當時的生理與練習狀況。會選擇取得**馬拉松最佳成績**的原因在於，最佳成績只有一場，選手比較容易回想起當時練習的狀況。

(一)最佳成績在第幾次馬拉松
(二)成績、比賽年度及名稱
(三)當時年齡
(四)體重
(五)當時一週練習量與練習次數

第三部份：創下個人最佳馬拉松成績「之前」、「最近」且「短於42.195公里」的跑步成績。這部份是為了取得(1.1)式中的d_1與t_1。

(一)當時5公里成績、比賽年度與名稱
(二)當時10公里成績、比賽年度與名稱
(三)當時21公里成績、比賽年度與名稱
(四)其它距離成績、比賽年度與名稱

樣本收集來源有網路問卷以及手寫問卷。網路部分，張貼於跑者廣場、孵蛋箱討論區、鳳山慢跑協會以及長青慢跑協會討論區等等網站，供跑友上網填寫。手寫部分，感謝長青慢跑協會於今年一月十三日，在西子灣舉辦一場優質的馬拉松比賽，提供場地發送問卷供跑友填寫。扣除一些填寫不完全的問卷，總共有204筆有效問卷，詳細馬拉松問卷內容及範例，請見附錄(A)與(B)。

2.2 分析方法與工具

利用到的統計方法，主要有迴歸分析與切片反向迴歸法。本文使用到的數學及統計軟體為$Mathematica$ 5.0版與$Splus$ 7.0版。

2.3 切片反向迴歸分析法

假設y代表反應變數，$\mathbf{x}=(x_1,x_2,...,x_p)'$為$p$個自變數。切片反向迴歸法主要目的是將$p$個自變數$\mathbf{x}$，降為較小的$k$度空間，$y$和$\mathbf{x}$之關係僅透過$k$個$\beta_1'\mathbf{x}$、$\beta_2'\mathbf{x}$、$\cdots$、$\beta_k'\mathbf{x}$之線性組合。切片反向迴歸模型如下：

$$y=f(x_1,x_2,...,x_p,\epsilon)=g(\beta_1'\mathbf{x},\beta_2'\mathbf{x},...,\beta_k'\mathbf{x},\epsilon)$$

反向切片迴歸分析法步驟如下：

一、將y排序後，儘可能地等分成H片。

二、令n_h為第h片的個數，計算每一片內$\mathbf{x}$的樣本平均數，$\overline{\mathbf{x}}_h=n_h^{-1}\Sigma_{(i)\in h}\mathbf{x}_{(i)}$。

三、樣本平均數為$\overline{\mathbf{x}}=n^{-1}\Sigma_{i=1}^n x_i$，計算組間加權共變異數矩陣(covariance matrix)。

$$\widehat{\Sigma}_\eta=n^{-1}\Sigma_{h=1}^H n_h(\overline{\mathbf{x}}_h-\overline{\mathbf{x}})(\overline{\mathbf{x}}_h-\overline{\mathbf{x}})'$$

四、計算全部$\mathbf{x}$的共變異數矩陣，$\widehat{\Sigma}_x=n^{-1}\Sigma_{i=1}^n(\mathbf{x}_i-\overline{\mathbf{x}})(\mathbf{x}_i-\overline{\mathbf{x}})'$。

五、藉由執行$\widehat{\Sigma}_\eta$對$\widehat{\Sigma}_x$的特徵值分解，前p個較大的特徵值可求得SIR方向：

$$\widehat{\Sigma}_\eta\widehat{\beta}_i=\widehat{\lambda}_i\widehat{\Sigma}_x\widehat{\beta}_i$$

$$\widehat{\lambda}_1\geqslant\widehat{\lambda}_2\geqslant\cdots\geqslant\widehat{\lambda}_p$$

，其中第i個特徵向量$\widehat{\beta}_i$被稱為第i個SIR方向。

六、將$\mathbf{x}$沿著SIR方向投影，即使用每一個SIR方向來構成$\mathbf{x}$之線性組合，我們稱$\widehat{\beta}_1'\mathbf{x}$為第一個SIR變量(variate)，$\widehat{\beta}_2'\mathbf{x}$為第二個SIR變量，之後依此類推。

七、以y為縱軸、SIR變量為橫軸，畫平面圖或立體圖。

2.4 分析步驟

問卷收集了影響馬拉松成績的變數，包括d_1、t_1、性別、年齡、身高、體重、跑齡、跑馬次數、每週練習量與次數。模型不考量變數之間的關係，直接進行複線性迴歸。

變數選取方式為**反向淘汰法**(backward)，即先將所有變數放入模型中，然後在每次淘汰中，剔出對模型的貢獻最小者，直到剩下的變數皆有一定的貢獻為止(p -value皆小於0.05)。而切片反向迴歸法的模型，首先將**全部變數放進去**，但效果和只放**尚未切片前的顯著變數**一樣，所以最終選取的變數為尚未切片前的顯著變數。接下來將204筆有效數據帶入估計好的模型當中，經過簡單的轉換，可以得到$\widehat{t_2}$，再拿實際的t_2與之相減，稱為誤差(即$t_2 - \widehat{t_2}$)。然後利用敘述統計量，如：最小值(minimum)、中位數(median)、最大值(maximum)、差距(range)、四分位距(interquartile range)、平均數(mean)與標準差(standard deviation)；以及一些統計圖，如：直方圖(histogram)、散佈圖(scatter plot)與盒形圖(boxplot)，來比較模型間的預測能力。接下來刪掉離群值(outliers)，最後找出最佳模型，並對該模型進行殘差分析，看看是否滿足殘差的基本假設。

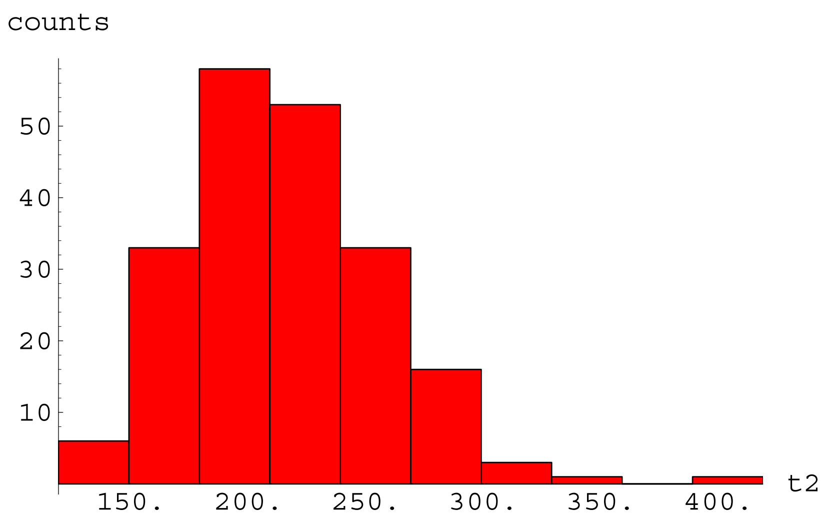
3. 統計分析結果

3.1 變數基本分析

針對收集到的變數，做基本的了解。

一、馬拉松最佳成績(即t_2)

由圖一可看出，樣本中選手的成績大多介於180到240分鐘，也就是說國內選手的成績大約在3至4小時。由表三得知，最好成績約為2小時17分，最差成績為6小時37分，差距高達4小時以上。由此可見若用一般模型來估計，誤差會很大。樣本中馬拉松平均成績約3小時36分，有75%選手成績在4時04分內。



x軸為馬拉松成績(分鐘)　　y軸為個數

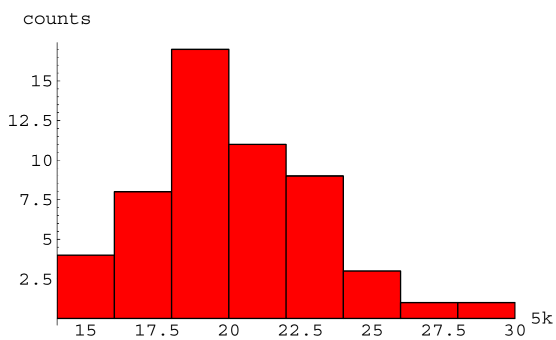
圖一：馬拉松成績的直方圖

表三：馬拉松成績的敘述統計量

	馬拉松成績(分鐘)
最小值	137.32
1/4位數	185.77
中位數	214.58
3/4位數	243.55
最大值	397.00
平均值	216.10
標準差	41.15

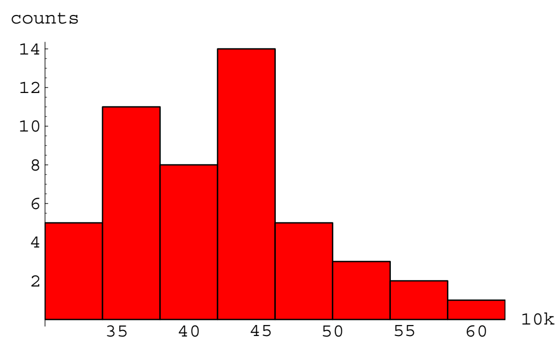
二、在不同距離d_1之下的成績t_1

樣本中d_1只有在5、9、9.6、10、11、13、14.6、15、19、21、25與30，這12處有數值。其中填寫5公里成績的選手佔了26%、10公里有24%、21公里有42%，這三類總共92%，所以特別挑選出來觀察其變化。由圖二可看出，5公里成績大多介於18到24分鐘；由圖三可看出，10公里成績大多介於34到46分鐘；由圖四可看出，21公里成績大多介於80到110分鐘。由表四得知，5公里平均成績約20分，10公里平均成績約42分，21公里平均成績約91分。換算可得三者速度分別為每小時15、14.29、13.85公里，速度的差異並沒有想像中大。



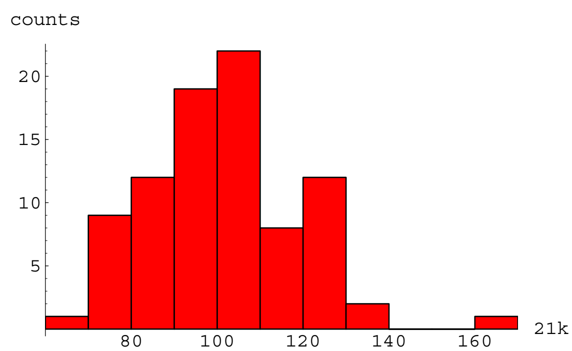
x軸為5公里成績(分鐘)　y軸為個數

圖二：5公里成績的直方圖



x軸為10公里成績(分鐘)　y軸為個數

圖三：10公里成績的直方圖



x軸為21公里成績(分鐘)　y軸為個數

圖四：21公里成績的直方圖

表四：5、10及21公里成績的敘述統計量

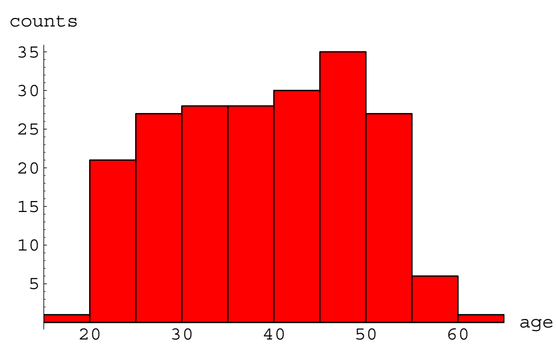
	5公里(分)	10公里(分)	21公里(分)
最小值	14.62	30.43	69.55
最大值	28.50	59.18	160.50
平均值	20.06	42.14	90.95
標準差	3.03	6.97	16.97

三、性別

樣本中，男生有191人，女生13人，分別佔94%與6%，男生參與馬拉松的比例明顯高於女生。

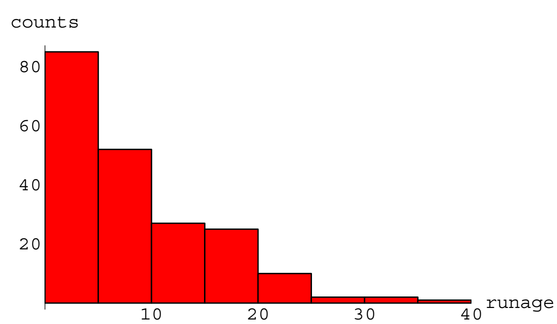
四、年齡與跑齡

由圖五看出，參與馬拉松的年齡層以20到55歲居多，可見跑步這項運動老少皆宜。圖六可觀察到，跑齡大部分在10年內。由表五知，最年輕者是17歲，最年長60歲，平均約38.5歲，有75%在47歲內。而跑齡最大值35年，平均約8年，有75%跑齡在12年內。



x軸為年齡　y軸為個數

圖五：年齡的直方圖



x軸為跑齡(年)　y軸為個數

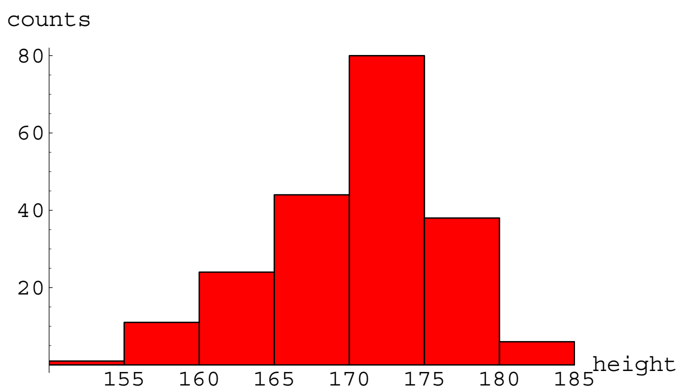
圖六：跑齡的直方圖

表五：年齡與跑齡的敘述統計量

	年齡	跑齡
最小值	17	1
1/4位數	30	3
中位數	39	5
3/4位數	47	12
最大值	60	35
眾數	29	3
平均值	38.50	8.02
標準差	10.30	6.54

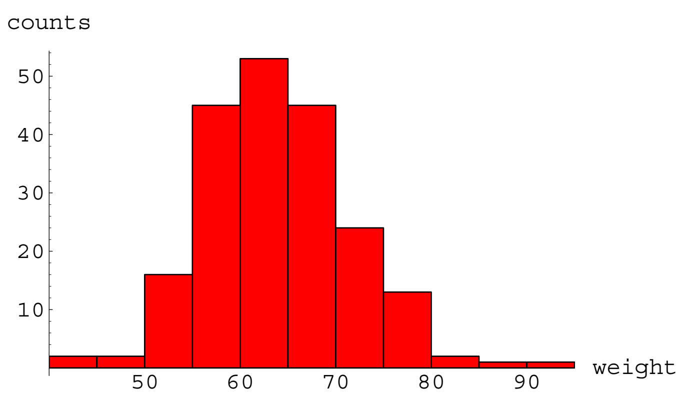
五、身高與體重

由圖七看出，身高大多介於165至180公分之間。從圖八知，體重大約介於55到70公斤。由表六知，平均身高約170公分，有75%在174公分內。而平均體重約63公斤，有75%在68公斤以內。



x軸為身高(公分)　y軸為個數

圖七：身高的直方圖



x軸為體重(公斤)　y軸為個數

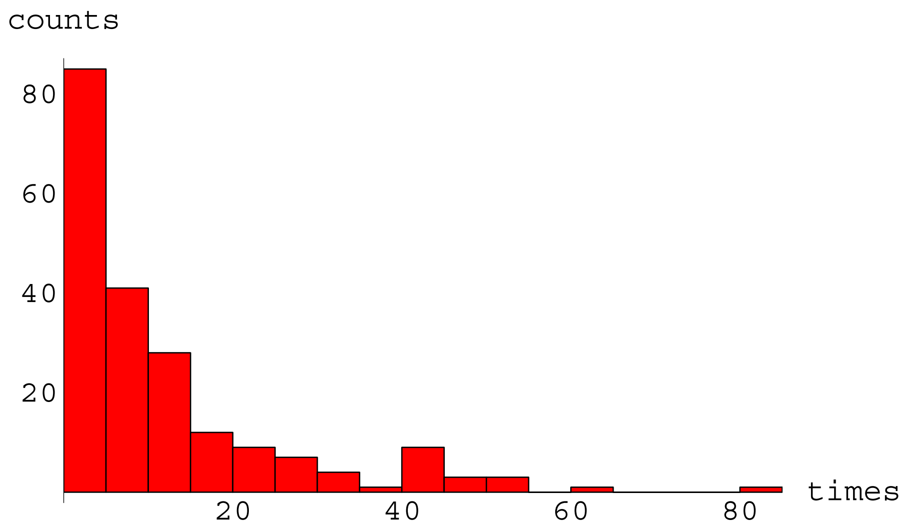
圖八：體重的直方圖

表六：身高與體重的敘述統計量

	身高(公分)	體重(公斤)
最小值	152	42
1/4位數	166	58
中位數	171	63
3/4位數	174	68
最大值	183	93
平均值	169.74	63.19
標準差	5.91	7.56

六、跑馬次數

由圖九可看出，大部分選手參加次數為15次以內。由表七知，平均約為11次，不過仍有選手高達80次之多。值得一提的是眾數為1次，也就是說第一次參加馬拉松的人數最多，共佔了32人之多，大約是全部的16%。再由3/4位數了解到75%選手參加次數在14次以內。



x軸為跑馬次數　　y軸為個數

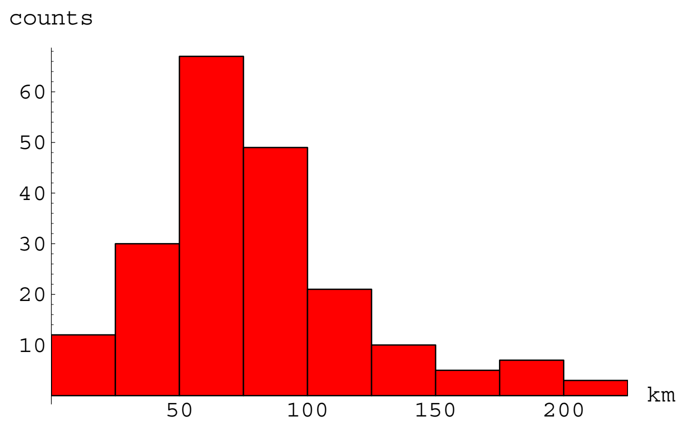
圖九：跑馬次數的直方圖

表七：跑馬次數的敘述統計量

	跑馬次數
最小值	1
1/4位數	2
中位數	6
3/4位數	14
最大值	80
眾數	1
平均值	11.19
標準差	13.46

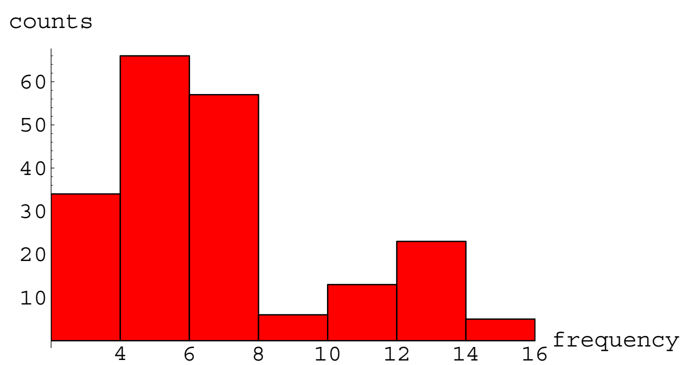
七、每週練習量與練習次數

由圖十看出，每週練習量大多介於50至100公里之間。由圖十一，每週練習次數大多在8次以內。由表八得知，75%選手每週練習量在90公里內，每週練習量平均約78公里；75%選手每週練習次數在7次以內，每週練習次數平均約6次。



x軸為練習量(公里)　y軸為個數

圖十：每週練習量的直方圖



x軸為練習量(公里)　y軸為個數

圖十一：每週練習次數的直方圖

表八：每週練習量及次數的敘述統計量

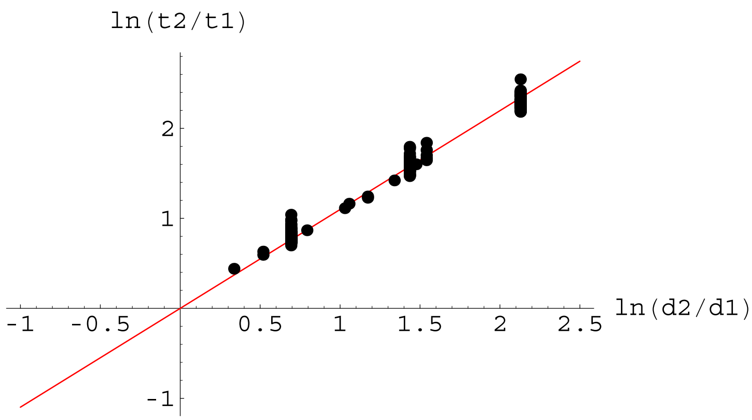
	練習量(公里)	練習次數
最小值	15	2
1/4位數	50	4
中位數	70	6
3/4位數	90	7
最大值	210	14
眾數	60	4
平均值	77.56	6.28
標準差	39.73	3.21

3.2 模型比較

推廣(1.1)模型，依序提出五個新模型，與之比較。

一、M1模型：

(1.1)式中指數部分為固定值1.06，M1模型為利用最小平方法，重新估計指數部分。估計過程為先將(1.1)式取自然對數(natural logarithm，數學表示式為ln)，然後改寫成$ln(\frac{t_2}{t_1}) = \beta ln(\frac{d_2}{d_1})$的形式，接著利用最小平方法估得$\widehat{\beta} = \frac{\Sigma ln(\frac{t_{2i}}{t_{1i}}) ln(\frac{d_{2i}}{d_{1i}})}{\Sigma ln(\frac{d_{2i}}{d_{1i}})^2} = 1.09835$。M1模型即為$\widehat{ln(\frac{t_2}{t_1})} = 1.09835 ln(\frac{d_2}{d_1})$。下圖直線為配適結果，但在四處的誤差明顯過大，有很大的改善空間。



x軸為$ln(\frac{d_2}{d_1})$ y軸為$ln(\frac{t_2}{t_1})$

圖十二：M1模型之散佈圖

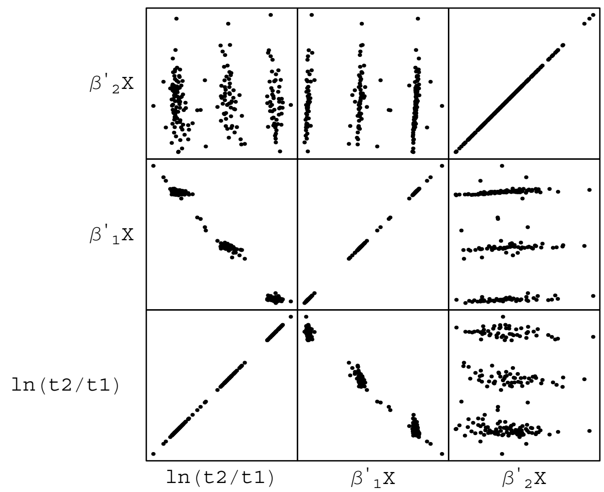
二、M2模型：

針對M1模型，再加入其他八項可能會影響馬拉松成績的變數，包括性別、年齡、身高、體重、跑齡、跑馬次數、每週練習量與次數。接下來利用迴歸分析配適模型，結果為$\widehat{ln(\frac{t_2}{t_1})}$=.172+1.058$ln(\frac{d_2}{d_1})$−.001×年齡−.001×跑馬次數−.001×每週練習量。

三、M3模型：

利用2.3節提到的SIR分析法，應用到M2模型中。Hsing et al.(1992)證實理論上切片個數H可以取到$n/2$，實際上使用較大的H並無顯著的幫助。本文切片個數取5的倍數，最多取到40。接著套用SIR分析法，得到SIR變量，然後觀察SIR變量

與$ln(\frac{t_2}{t_1})$的圖形。由圖十三，可看出$\beta_1'X$與$ln(\frac{t_2}{t_1})$呈現高度線性關係，接著利用迴歸分析來建立模型。切片個數H取25時，得到模型為$\widehat{ln(\frac{t_2}{t_1})} = .172 - 1.054 \times \beta_1'X$，其中$\beta_1' = (-1.0000, .0015, .0009, .0003)$，$X' = (ln(\frac{d_2}{d_1})$,年齡,跑馬次數,每週練習量)。



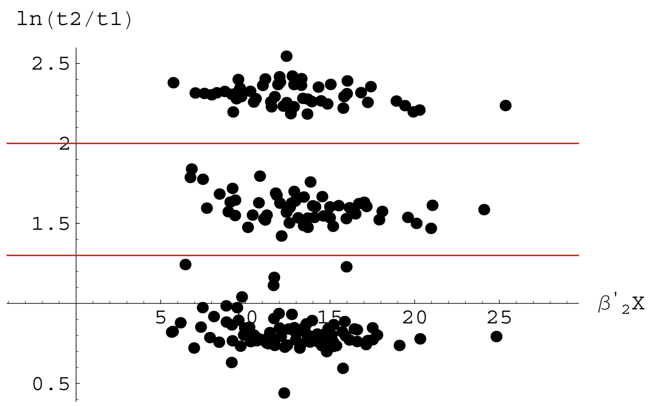
圖十三：$ln(\frac{t_2}{t_1})$與前兩個SIR變量的平面圖

四、M4模型：

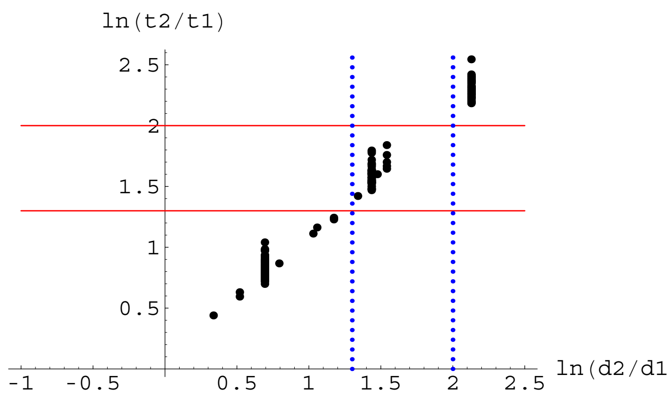
由圖十四，可看出$\beta_2'X$與$ln(\frac{t_2}{t_1})$似乎呈現三條不同截距與斜率的直線，所以將指標函數(indicator function)加入M3模型中，看看是否對預測有所幫助。在此，指標函數(I_1, I_2)的表示式為(1)的形式。配適模型首先以完整模型$ln(\frac{t_2}{t_1}) = a + b \times \beta_1'X + c_1 \times I_1 + c_2 \times I_2 + d_1 \times \beta_1'X \times I_1 + d_2 \times \beta_1'X \times I_2$下手，結果得到顯著變數為$\beta_1'X$、$I_2$、$\beta_1'X \times I_1$、$\beta_1'X \times I_2$這四項，所以模型為$\widehat{ln(\frac{t_2}{t_1})} = -.194 - 1.232 \times \beta_1'X + .420I_2 - .099 \times \beta_1'X \times I_1 + .272 \times \beta_1'X \times I_2$ ，其中$\beta_1' = (-1.0000, .0015, .0009, .0003)$，$X' = (ln(\frac{d_2}{d_1})$,年齡,跑馬次數,每週練習量)。由圖十五看出，若$ln(\frac{t_2}{t_1})$的值大於2，則$ln(\frac{d_2}{d_1})$的值也大於2；若$ln(\frac{t_2}{t_1})$的值介於1.3到2之間，則$ln(\frac{d_2}{d_1})$的值也介於此區間；若$ln(\frac{t_2}{t_1})$的值小於1.3，則$ln(\frac{d_2}{d_1})$的值也小於1.3。因此，指標函數$(I_1, I_2)$由(1)改寫成(2)。

$$(I_1, I_2) = \begin{cases} (0,0) & ,若\quad 2 \leqslant ln(\frac{t_2}{t_1}) \\ (1,0) & ,若\quad 1.3 \leqslant ln(\frac{t_2}{t_1}) < 2 \\ (0,1) & ,若\quad ln(\frac{t_2}{t_1}) < 1.3 \end{cases} \tag{1}$$

$$(I_1, I_2) = \begin{cases} (0,0) & ,若\quad 2 \leqslant ln(\frac{d_2}{d_1}) \\ (1,0) & ,若\quad 1.3 \leqslant ln(\frac{d_2}{d_1}) < 2 \\ (0,1) & ,若\quad ln(\frac{d_2}{d_1}) < 1.3 \end{cases} \tag{2}$$



圖十四：$ln(\frac{t_2}{t_1})$與$\beta_2'X$的平面圖



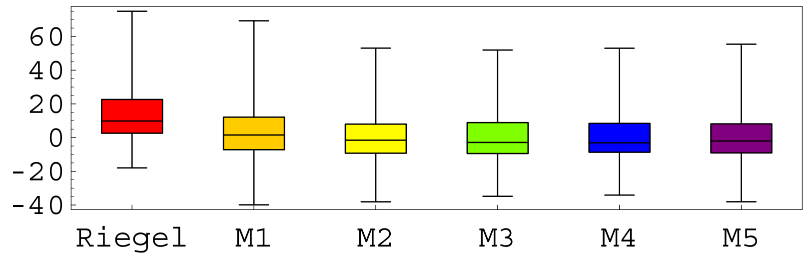
圖十五：$ln(\frac{t_2}{t_1})$與$ln(\frac{d_2}{d_1})$的關係圖

五、M5模型：

針對M4模型的**跑馬次數**(簡稱T)作適當的調整，避免因為T無上限的增加而造成「跑越多次成績越好的假象」。調整方法為利用**分位數**(quantile)將T等分成p部分，再用階梯函數表示之。文中試過p=3,...,10，其中以p=3表現最佳(絕對誤差小於15分鐘比例最高)。當p=3時，則T以1/3分位數(等於3)及2/3分位數(等於10)區分成3群。第一群為3次以內，取其平均為**2次**；第二群為4到10次，取其平均為**7次**；第三群為超過10次，取第二群的上界**10次**。調整過後的T，在此簡稱T^*，表示式如(3)。接下來配適過程和M4模型相同，配適結果為$\widehat{ln(\frac{t_2}{t_1})} = -.120 + 1.209 \times \beta_1^{*'}X^* + .361I_2 + .088 \times \beta_1^{*'}X^* \times I_1 - .238 \times \beta_1^{*'}X^* \times I_2$，其中$\beta_1^{*'} = (1.0000, -.0015, -.0055, -.0003)$，$X^{*'} = (ln(\frac{d_2}{d_1})$,年齡,$T^*$,每週練習量$)$，$(I_1, I_2)$表示式如(2)，$T^*$表示式如(3)。

$$T^* = \begin{cases} 2 & ,若 \quad T \leqslant 3 \\ 7 & ,若 \quad 3 < T \leqslant 10 \\ 10 & ,若 \quad 10 < T \end{cases} \tag{3}$$

將M1、M2、M3、M4及M5模型，取指數後再乘以t_1得到$\widehat{t_2}$，然後與真實值t_2相減得到誤差。從各模型之誤差的盒形圖(圖十六)，可看出M2、M3、M4與M5模型估計的誤差範圍遠小於其它模型，接下來進一步比較其數值的差異。



x軸為模型　y軸為誤差(分鐘)

圖十六：各模型之誤差的盒鬚圖

由表九可得，M5模型估計誤差所得到的**標準差**最小，且絕對誤差小於15分鐘比例和M4並列最高，比(1.1)模型高出16.18%。本文選擇最佳模型優先以**誤差小於15分鐘比例最高**為第一考量，其次是**標準差最小**，故M5為目前最佳模型。

表九：各模型之誤差的敘述統計量

模型	Riegel	M1	M2	M3	M4	M5
最小值	−17.97	−39.89	−38.09	−34.84	−34.12	−38.06
中位數	9.83	1.55	−1.55	−2.88	−3.10	−2.02
最大值	74.96	69.32	53.10	51.95	53.03	55.42
四分位距	20.02	19.30	17.26	18.35	17.12	17.25
差距	92.94	109.21	91.19	86.80	87.15	93.48
平均數	14.15	4.11	0.39	0.66	0.63	0.59
標準差	16.91	17.96	15.00	15.14	14.92	14.35
% $\lvert$誤差$\rvert \leqslant 15$	61.76	70.59	74.02	75.49	77.94	77.94

單位：分鐘

3.3 刪掉離群值

針對上一小節提出的M5模型，分別刪掉誤差超過3倍標準差(即43.06)的第17、64、104這三點，以及超過2倍標準差(即28.71)的第3、17、31、35、40、64、83、104、111、141、149、159、174、203這十四點，然後重複上一小節處理的所有流程，表十為刪掉3倍標準差的模型。

表十：模型對照表(刪掉3倍標準差)

	模型
Riegel	$\widehat{ln(\frac{t_2}{t_1})} = 1.06ln(\frac{d_2}{d_1})$
M1	$\widehat{ln(\frac{t_2}{t_1})} = 1.09641ln(\frac{d_2}{d_1})$
M2	$\widehat{ln(\frac{t_2}{t_1})}=.155+1.057ln(\frac{d_2}{d_1})-.001\times$年齡$-.001\times T-.001\times$每週練習量
M3	$\widehat{ln(\frac{t_2}{t_1})} = .183 - 1.054 \times \beta_1^{'}X$
M4	$\widehat{ln(\frac{t_2}{t_1})} = -.179 - 1.231 \times \beta_1^{'}X + .419I_2 - .103 \times \beta_1^{'}X \times I_1 + .277 \times \beta_1^{'}X \times I_2$
M5	$\widehat{ln(\frac{t_2}{t_1})} = -.074 + 1.191 \times \beta_1^{*'}X^* + .321I_2 + .083 \times \beta_1^{*'}X^* \times I_1 - .220 \times \beta_1^{*'}X^* \times I_2$

註：$\beta_1^{'} = (-1.0000, .0017, .0006, .0005)$，$\beta_1^{*'} = (1.0000, -.0016, -.0054, -.0004)$，$X^{'} = (ln(\frac{d_2}{d_1})$,年齡,$T$,每週練習量)，$X^{*'} = (ln(\frac{d_2}{d_1})$,年齡,$T^*$,每週練習量)，其中$T$為跑馬次數，$T^*$為調整過後的跑馬次數。

$$T^* = \begin{cases} 2 & ,\text{若} \quad T \leqslant 3 \\ 7 & ,\text{若} \quad 3 < T \leqslant 10 \\ 10 & ,\text{若} \quad 10 < T \end{cases}$$

由表十一可知，由M5模型估計誤差所得到的**四分位距**、**差距**與**標準差**為所有模型中最小，且絕對誤差小於15分鐘比例更高達80.10%，比(1.1)模型足足多了17.41%，接下來觀察刪掉2倍標準差的結果如何。

表十一：各模型之誤差的敘述統計量(刪掉3倍標準差)

模型	Riegel	M1	M2	M3	M4	M5
最小值	−17.97	−38.73	−35.77	−35.64	−34.86	−38.59
1/4位數	2.61	−6.27	−8.76	−8.76	−8.36	−8.37
中位數	9.36	1.50	−1.71	−2.52	−2.93	−1.51
3/4位數	22.50	11.63	8.48	8.92	8.71	8.11
最大值	62.37	53.82	42.26	40.60	40.06	34.41
四分位距	19.89	17.90	17.24	17.68	17.07	16.48
差距	80.34	92.55	78.04	76.24	74.92	73.01
平均數	13.34	3.80	0.29	0.48	0.45	0.39
標準差	15.61	16.60	13.94	14.10	13.80	13.21
$\%$ |誤差| $\leqslant 15$	62.69	72.64	76.12	77.61	80.10	80.10

表十二：模型對照表(刪掉2倍標準差)

	模型
Riegel	$\widehat{ln(\frac{t_2}{t_1})} = 1.06ln(\frac{d_2}{d_1})$
M1	$\widehat{ln(\frac{t_2}{t_1})} = 1.09324ln(\frac{d_2}{d_1})$
M2	$\widehat{ln(\frac{t_2}{t_1})}$=.128+1.058$ln(\frac{d_2}{d_1})$−.001×年齡−.001× T−.0004×每週練習量
M3	$\widehat{ln(\frac{t_2}{t_1})} = .128 + 1.051 \times \beta_1^{'}X$
M4	$\widehat{ln(\frac{t_2}{t_1})} = -.318 + 1.266 \times \beta_1^{'}X + .506I_2 + .111 \times \beta_1^{'}X \times I_1 - .310 \times \beta_1^{'}X \times I_2$
M5	$\widehat{ln(\frac{t_2}{t_1})} = -.320 - 1.276 \times \beta_1^{*'}X^* + .525I_2 - .116 \times \beta_1^{*'}X^* \times I_1 + .328 \times \beta_1^{*'}X^* \times I_2$

註：$\beta_1' = (1.0000, -.0013, -.0002, -.0001)$，$\beta_1^{*'} = (-1.0000, .0010, .0036, .0001)$，$X' = (ln(\frac{d_2}{d_1})$,年齡,$T$,每週練習量)，$X^{*'} = (ln(\frac{d_2}{d_1})$,年齡,$T^*$,每週練習量)，其中$T$為跑馬次數，$T^*$為調整過後的跑馬次數。

$$T^* = \begin{cases} 2.5 & ,\text{若} \quad T \leqslant 4 \\ 8 & ,\text{若} \quad 4 < T \leqslant 11 \\ 11 & ,\text{若} \quad 11 < T \end{cases}$$

由表十一可知，由M5模型估計誤差所得到的**四分位距**、**差距**與**標準差**為所有模型中最小，且絕對誤差小於15分鐘比例更高達81.58%，比(1.1)模型足足多了16.32%。

表十三：各模型之誤差的敘述統計量(刪掉2倍標準差)

模型	Riegel	M1	M2	M3	M4	M5
最小值	−14.21	−32.30	−30.19	−25.59	−26.93	−21.30
1/4位數	2.51	−5.12	−8.15	−8.53	−7.67	−7.40
中位數	9.07	1.56	−1.67	−2.17	−2.71	−1.93
3/4位數	19.45	11.26	8.26	9.60	9.13	8.43
最大值	59.48	46.71	34.67	38.66	34.18	29.85
四分位距	16.94	16.38	16.41	18.13	16.80	15.83
差距	73.69	79.01	64.86	64.25	61.11	51.15
平均數	11.88	3.26	0.28	0.65	0.66	0.61
標準差	13.12	13.94	11.72	12.19	11.84	11.26
% |誤差| $\leqslant$ 15	65.26	75.79	81.58	80.00	80.53	81.58

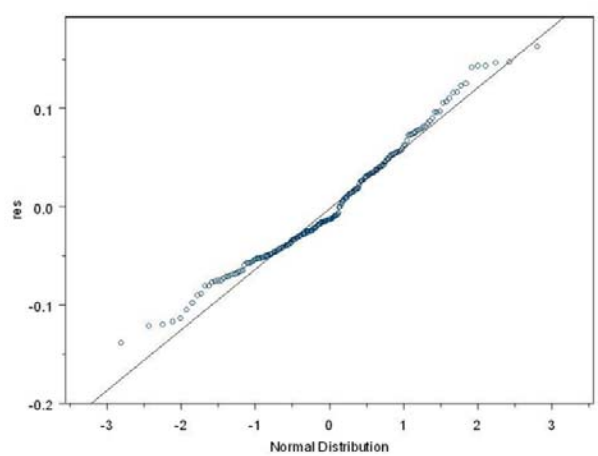
3.4 殘差分析

針對刪掉超過誤差之3倍及2倍標準差後的M5模型，進行殘差分析。在迴歸分析中，殘差主要假設有三點，依序如下：

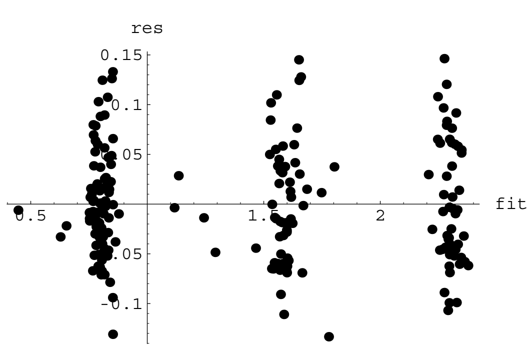
(一)殘差為常態分佈。由單一樣本常態性檢定(*One sample Kolmogorov–Smirnov test of Normality*)，得到在**虛無假設是殘差符合常態分佈**之下刪掉誤差超過3倍標準差的p -value=0.068(>0.05)，故刪掉3倍標準差結果為接受常態性假設；但刪掉2倍標準差的p -value=0.0067(<0.05)，所以刪掉2倍標準差結果並不符合常態性假設，接下來就以刪掉3倍標準差為分析的重點。刪掉誤差的3倍標準差，雖然勉強通過常態性假設，不過由圖十七得知有偏斜(skew)情況，需要經過一些處理。

(二)殘差的變異數一致。由圖十八之殘差趨勢圖(patterns for residual plot)，看出殘差的變異程度差不多，所以符合殘差的變異數一致的假設。

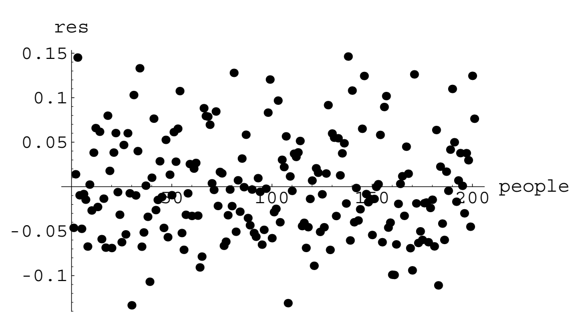
(三)殘差之間為獨立。由圖十九之殘差時間圖，看出殘差之間並沒有明顯的相關性，所以符合殘差獨立性的假設。



圖十七：常態機率圖(刪掉3倍標準差)



圖十八：殘差的趨勢圖(刪掉3倍標準差)



圖十九：殘差的時間圖(刪掉3倍標準差)

4. 結論與建議

本研究中最佳預測模型建立過程為先將(1.1)模型轉換成$ln(\frac{t_2}{t_1}) = \beta ln(\frac{d_2}{d_1})$的形式，然後利用最小平方法估計$\widehat{\beta}$(即M1模型)，接著加入收集到的自變數(即M2)，再來利用SIR分析法(即M3)，之後加入指標函數(即M4)，最後將**跑馬次數** T改寫成**階梯函數** T^*(即M5)。需要加入指標函數原因為樣本中d_1主要是5、10與21公里這三類，所以模型自然呈現三種不同的斜率與截距。而將跑馬次數改寫原因是避免造成「跑越多次成績越好的假象」。最後再刪掉誤差超過3倍標準差的離群值，得到最佳模型M5，該模型所需資料為最近比賽或測驗成績、年齡、跑馬次數與每週練習量這四項，適用範圍為**男性**且**年齡介於17至60歲**，**跑馬次數80次內**，**每週練習量210公里內**。模型如下：

$$\widehat{t_2} = t_1 \times e^{-.074+1.191\times\beta_1^{*'}X^*+.321I_2+.083\times\beta_1^{*'}X^*\times I_1-.220\times\beta_1^{*'}X^*\times I_2}$$

，其中$\beta_1^{*'} = (1.0000, -.0016, -.0054, -.0004)$，$X^{*'} = (ln(\frac{d_2}{d_1})$,年齡,$T^*$,每週練習量)，$(I_1, I_2)$及$T^*$表示式如下。

$$(I_1, I_2) = \begin{cases} (0,0) & ,\text{若} \quad 2 \leqslant ln(\frac{d_2}{d_1}) \\ (1,0) & ,\text{若} \quad 1.3 \leqslant ln(\frac{d_2}{d_1}) < 2 \\ (0,1) & ,\text{若} \quad ln(\frac{d_2}{d_1}) < 1.3 \end{cases}$$

$$T^* = \begin{cases} 2 & ,\text{若} \quad T \leqslant 3 \\ 7 & ,\text{若} \quad 3 < T \leqslant 10 \\ 10 & ,\text{若} \quad 10 < T \end{cases}$$

該模型估計結果為有八成選手的馬拉松成績會在誤差15分鐘以內，而Riegel提出的(1.1)模型只有六成二，預測公式的實際應用，請參考附錄(C)。若能收集到每位跑友的兩筆d_1與t_1，或是比賽時的地形(高度與坡度)與氣候(溫度、濕度、風力與風向)，這些環境因素納入模型中，相信對成績的預測更有幫助。

參考文獻

中文部分：

1. 林信甫、莊泰源 (2003)。跑步經濟性及其相關影響因素探討。中華體育季刊，**17**(3)，53 - 60。

2. 林貴福、徐台閣、吳慧君 (2002)。運動生理學。台北市：藝軒圖書出版社，439-145。

3. 許玉芳 (2006)。2004雅典奧運馬拉松分段時間之分析。國立體育學院運動技術研究所碩士論文。

4. 陳孟欣 (2004)。長跑選手的體重控制與運動表現之探討。大專體育，70，170 - 174。

5. 張保羅 (2001)。田徑。台北市：國家出版社，69 - 71。

6. 許樹淵 (1984)。運動員體格成績分析：第一屆世界田徑錦標賽。台北市：中華民國田徑協會出版，137 - 153。

7. 許樹淵 (2000)。運動生理心理學。台北市：師大書苑，969-970。

8. 許績勝 (2000)。馬拉松。國立體育學院教練研究所技術報報書，1 - 96。

9. 黃永任 (1994)。運動科學講座。台北市：八熊星出版社，33-36。

英文部分：

1. Ahlborg, B. , Bergstrom, J. , Ekelund, L.G. , and Hultman, E. (1967). Muscle glycogen and muscle electrolytes during prolonged physical exercise. *Acta Physiologica Scandinavica*, 70 : 129 - 42.

2. Bergstrom, J. , Hermansen, L. , Hultman, E. , and Saltin, B. (1967). Diet, muscle glycogen and physical performance. *Acta Physiologica Scandinavica*, 71 : 140 - 50.

3. Hermansen, L. , Hultman, E. , and Saltin, B. (1967). Muscle glycogen during prolonged severe exercise. *Acta Physiologica Scandinavica* 71 : 129 - 39.

4. Hsing, T. and Carroll, R. J. (1992). An asymptotic theory for sliced inverse regression. *Ann. Statistic.* 20, 1040-1061.

5. Li, K. C. (1991). Sliced inverse regression for dimension reduction. *Joumal of the American Statistical Association*, **86**(414).

6. Montgomery, D. C. , Peck , E.A. ,and Vining , G.G. (2001). Introduction to linear regression analysis. New York : Wiley, 67-75.

7. Priest, J.W. , and Hagan, R.D. (1987). The effects of maximum steady-state pace on running performance. *British Journal of Sports Medicine* 21 : 18 - 21.

8. Riegel, P. (1977). Time Predicting. *Runner's World*, August.

9. Londeree, B.R. (1986). The use of laboratory test results with long distance runners. *Sports Medicine* 3 : 201 - 13.

附錄

(A) 馬拉松問卷內容

研究題目：「台灣馬拉松選手之成績預測」

親愛的跑友您好：

感謝您抽空填寫這份問卷。本研究是想藉由「馬拉松選手」的下列各項資料，來探討影響國內選手表現的因素，並加以分析，進一步建立預測模型，請依照您個人的寶貴經驗回答下列問題。問卷內容僅作為學術研究，個人資料將不會對外公開。感謝您參與本問卷的填寫。敬祝您　平安健康。

國立中山大學應用數學系-統計組
指導教授：羅夢娜　教授
研究生：江承鴻　敬上

第一部份：基本資料

1.曾經是 □ 學校校隊 □ 國家代表隊 □ 皆非。專長：_________公尺
2.性別：_____，年齡：_____歲。身高：_______，體重：_____公斤，跑齡：_____。
3.初馬成績：__時___分___秒，比賽年度：民國____年，比賽名稱：____________。
4.E-mail(以便提供分析結果)：________________________________。

第二部份：創下個人最佳馬拉松成績當時狀況

1.最佳成績在第幾馬：_____，成績：__時___分___秒，比賽年度：民國____年，
比賽名稱：____________。
2.當時年齡：_____歲，當時體重：_____公斤。
3.當時一週練習次數(如：早晚都練算2次)：_____次。
當時一週練習公里數(含熱身與收操)：_______公里。

第三部份：創下個人最佳馬拉松成績之前跑步成績(盡量在同一年，以增加預測準確性)

當時5、10、21公里或其它距離之最佳成績
_____公里，成績：__時___分___秒，比賽年度：民國____年，比賽名稱：____________。
_____公里，成績：__時___分___秒，比賽年度：民國____年，比賽名稱：____________。
_____公里，成績：__時___分___秒，比賽年度：民國____年，比賽名稱：____________。

(B) 馬拉松問卷範例

研究題目：「台灣馬拉松選手之成績預測」

親愛的跑友您好：

感謝您抽空填寫這份問卷。本研究是想藉由「馬拉松選手」的下列各項資料，來探討影響國內選手表現的因素，並加以分析，進一步建立預測模型，請依照您個人的寶貴經驗回答下列問題。問卷內容僅作為學術研究，個人資料將不會對外公開。感謝您參與本問卷的填寫。敬祝您　平安健康。

國立中山大學應用數學系-統計組
指導教授：羅夢娜　教授
研究生：江承鴻　敬上

第一部份：基本資料

1.曾經是 ■ 學校校隊　□ 國家代表隊　□ 皆非。專長：五千、一萬公尺
2.性別：男，年齡：24歲。身高：167公分，體重：56公斤，跑齡：6年。
3.初馬成績：2時52分29秒，比賽年度：民國96年，比賽名稱：2007台北國際馬拉松。
4.E-mail(以便提供分析結果)：略。

第二部份：創下個人最佳馬拉松成績當時狀況

1.最佳成績在第幾馬：1，成績：2時52分29秒，比賽年度：民國96年，比賽名稱：2007台北國際馬拉松。
2.當時年齡：24歲，當時體重：55公斤。
3.當時一週練習次數(如：早晚都練算2次)：10次。
當時一週練習公里數(含熱身與收操)：125公里。

第三部份：創下個人最佳馬拉松成績之前跑步成績(盡量在同一年，以增加預測準確性)

當時5、10、21公里或其它距離之最佳成績
5公里，成績：0時16分25秒，比賽年度：民國96年，比賽名稱：2007全國大專院校運動會。
10公里，成績：0時34分42秒，比賽年度：民國96年，比賽名稱：2007台南古都國際馬拉松。
21公里，成績：1時15分46秒，比賽年度：民國96年，比賽名稱：2007 親六堆跑客庄路跑賽。

(C)預測公式的實際應用

一、請至下面網址下載預測公式(若以複製網址的方式，請重新輸入∼ 此符號)：

http://ms90.nttu.edu.tw/∼u9003087/Marathon.xls

二、填寫方式：

1. **A3**請填寫最近比賽或測驗成績的距離(單位為公里)。
2. **B3**、**C3**、**D3**請分別填寫跑該距離所花的時、分、秒。
3. **E3**請填寫年齡。
4. **F3**填寫**目前**要跑的馬拉松是第幾場。
5. **G3**請填寫**一個禮拜**所跑的公里數。
6. 輸入以上資料後，結果請見**M3**(單位為分鐘)。

三、舉例說明：

已知阿哲選手賽前5公里成績15分05秒，年齡23歲，目前要跑第12場馬拉松，每週練習量是190公里。所以**A3為5**，**B3為0**，**C3為15**，**D3為5**，**E3為23**，**F3為12**，**G3為190**，預測結果阿哲有143分鐘完成馬拉松的實力。