

東海大學資訊工程學系研究所

碩士論文

指導教授：許玟斌 博士

一個棒球賽局勝負比之預測模式

A Prediction Model of Baseball Game Odds

研究生：陳英俊

中華民國 102 年 6 月

摘要

攻守數據收集及分析對於職業棒球比賽是非常重要的，藉由量化的資料，可以了解球員及球隊在臨場各種可能性的發生率。然而職業棒球比賽數據繁多，在長年累月收集的資料數量及規模，需要花費十分龐大的心力來進行分析，所以已不是人類計算分析能力所能處理的範圍，勢必得藉由電腦強大的運算分析能力來代為處理。

本研究依據球隊的攻守數據，建立球隊贏球機率的計算模式。因棒球守備和攻擊的數據分析指標太複雜，故先利用主成份分析(Principal Component Analysis)之方法，先找出影響比賽勝負的綜合性指標。用主成份分析，找出原有守備和攻擊的數據數之線性組合，利用共變異數矩陣 S ，求出特徵值及單位特徵向量，則特徵值最大的為第一主成份，並取變異數最大的主成份來當作總合指標。迴歸分析可用來找出兩個或兩個以上計量變數間的關係，並進而從一群變數中可以預測資料趨勢，所以可用來分析勝率與攻守數據的關係。設定勝率為自變數，攻擊能力與守備能力因素為依變數，可由函數關係中計算出勝率估計值。羅序模式是以機率的概念來解釋受試者能力和測驗項目間之關係，就可推估出受試者正確反應的機率，所以將之應用在預測職業棒球未來比賽之結果。將兩兩球隊之勝率估計值帶入羅序模式(Rasch Model)，計算甲球隊贏乙球隊的機率。我們相信本研究建立的模式，可提供球隊在賽前針對不同對戰球隊，安排最佳的攻守陣容，以增加贏球的機率。

關鍵字：棒球賽局、主成份分析、迴歸分析、羅序模式

Abstract

Abstract

Records from a few metrics in measuring the capabilities of offensive and defensive performance in a baseball game are quite important. With those quantitative data, researchers may be able to figure out the probabilities of various events in a match. However, the mass amount of statistics collected over years and/or large scale, it is not possible to analyze without much effort. Thus, using computer and suitable models is the only choice left to discover the characteristics from game data.

In our study, we aim to build a model for calculating the odds of the two teams in a game based on their individual historical offensive and defensive records. Due to the many indexes or metrics, we use principal component analysis to form some composite measures as the factors of match results. Then a regression model was developed to relate the winning rate and the composite measures. In item response theory, the Rasch model is a popular method in estimating the probability of correct response of examinee on a test item. We apply Rasch model to estimate the probability of team A beats team B, or vice versa. We believe our model is able to provide information for arranging offensive and defensive strategies to order to increase the winning chance.

Keyword : Baseball Game, Principal Component Analysis, Regression Analysis, Rasch Model

致 謝

工作多年後，繼續攻讀在職進修碩士專班，多年的光陰稍縱即逝。本論文能順利的完成，承蒙師長諄諄教誨，首先要感謝恩師指導教授 許玟斌教授，在論文的撰寫過程中細心的指導與教誨，處處地關懷，學者之風範讓我受教良多，使本論文得以順利完成。浩蕩師恩，銘感於心，感激之情，難以言表！

在論文口試期間，承蒙口試委員 胡學誠博士、徐麗蘋博士、陳怡靜博士在百忙之中對於本論文句斟字酌、立論之周延，不吝指教，提供許多寶貴的建議，使本論文能更臻完備，在此，謹致上最高的敬意與謝意。

陳英俊 謹誌於

私立東海大學資訊工程與科學研究所

中華民國102年6月

目 錄

摘 要.....	II
ABSTRACT.....	III
致 謝.....	IV
目 錄.....	V
圖目錄.....	VII
表目錄.....	VIII
第 一 章 緒論.....	1
1.1 研究動機.....	1
1.2 研究目的.....	2
1.3 研究流程.....	2
第 二 章 文獻探討.....	3
2.1 棒球比賽.....	3
2.2 主成份分析.....	4
2.3 迴歸分析法.....	7
2.4 羅序模式.....	9
第 三 章 研究方法.....	12
3.1 研究方法及架構.....	12
3.2 彙整攻擊能力與守備能力因素.....	12
3.3 發展勝率迴歸方程式.....	14
3.4 建立勝負比機率模式.....	14

第四章	實例分析.....	15
4.1	資料數據.....	15
4.2	主成份分析.....	18
4.3	求出勝率估計值.....	20
4.4	結果與分析.....	22
第五章	結論與建議	24
	參考文獻.....	25

圖目錄

圖 1、研究流程圖.....	2
圖 2、四條典型的二個參數試題特徵曲線.....	10

表目錄

表 1、甲球隊攻擊數據資料.....	15
表 2、乙球隊攻擊數據資料.....	16
表 3、丙球隊攻擊數據資料.....	16
表 4、丁球隊攻擊數據資料.....	17
表 5、甲球隊守備數據資料.....	17
表 6、乙球隊守備數據資料.....	18
表 7、丙球隊守備數據資料.....	18
表 8、丁球隊守備數據資料.....	19
表 9、攻擊主成份分析結果-成份矩陣.....	19
表 10、守備主成份分析結果-成份矩陣.....	20
表 11、標準化係數.....	21
表 12、球隊勝率估計值.....	22
表 13、甲球隊與其他球隊贏球機率預測結果.....	22
表 14、乙球隊與其他球隊贏球機率預測結果.....	23
表 15、丙球隊與其他球隊贏球機率預測結果.....	23
表 16、丁球隊與其他球隊贏球機率預測結果.....	23

第一章 緒論

棒球運動是非常注重數據收集及分析的運動，數據可分為記錄性數據及統計數據，投手的記錄性數據有勝敗、救援成功、中繼成功、三振、四死保送、投球數、好壞球數、被安打、被全壘打、投球局數、失分、責失…等，打、跑者的記錄性數據有打席、打數、安打、全壘打、被三振、獲四死球、盜壘成敗、打點、得分…等；防守則有守備機會、守備成功、失誤、助殺…等；而根據這些記錄性數據計算出各種統計數據，如與投手有關的自責分率、K/9（平均九局三振）值、BB/9（平均每九局四死球數）、勝率…等；與打者和跑者有關的打擊率、長打率、上壘率、盜壘成功率、得點圈打擊率…等；還有與守備有關的守備率、狙殺率…等。職業棒球比賽數據繁多，在長年累月收集的資料數量及規模，已不是人類計算分析能力所能處理的範圍，勢必得藉由電腦強大的運算分析能力來代為處理。

1.1 研究動機

發展職業運動是一個國家運動水準高、運動人口多、運動迷市場穩固的象徵，民國七十九年臺灣由四個企業投入職棒市場，至此職棒運動邁向一個新的紀元。而隨著職業運動的成立與休閒觀念的興起，運動幾乎已深入社會大眾生活之中。運動產業不單是帶動了民眾的運動意願，更由此延伸了許多相關產業的發展，如廣告、運動彩券、授權產品、運動服裝等等，所以每場比賽的勝負，可影響到整個產業鍊的經濟活動[3]。

1.2 研究目的

職業棒球比賽數據繁多，如何從已知的比賽資料中，來預測雙方的比數及勝負，是一項非常必要的研究。本研究依據球隊的攻守數據，依據球隊的攻守數據，建立球隊贏球機率的計算模式。我們相信本研究建立的模式，可提供球隊在賽前針對不同對戰球隊，安排最佳的攻守陣容，以增加贏球的機率。

1.3 研究流程

本論文首先必須先確定所要研究問題及範圍，並對於問題的原因進行探討；然後開始收集各項資料進行分析與探討，並決定用什麼方法解決該問題。下圖 1 研究流程圖

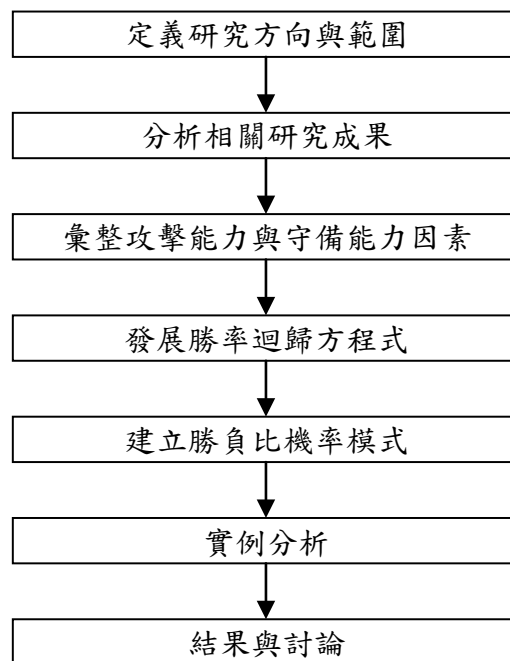


圖 1、研究流程圖

第二章 文獻探討

2.1 棒球比賽

職業棒球運動棒球是一高度團隊比賽的運動，它是由攻擊與防守兩要素所構成的運動[1]。棒球運動的紀錄模式是最完備、最詳細，因為它可以反映出比賽之全貌，只要一看比賽記錄，一場活生生的比賽，彷彿在眼前重現[2]。

Barker 對二十個職業聯盟球隊的比賽記錄做了因素分析研究，其結果為打擊力，對比賽勝率的影響力占 52%，而球隊攻擊力的影響依序為連續安打、打數、長打率，投手與守備則佔剩下之 48%。研究同時顯示強打率對於球隊勝率來說，遠不如連續安打重要[5]。

Scully 指出影響球隊勝率之因素，以球隊長打率、上壘率、分區冠軍(虛擬變數)與勝差場數為解釋變數，發現球隊長打率、球隊上壘率與分區冠軍對球隊獲勝率有顯著正向影響；球隊勝率的解釋變數中(球隊長打率與球隊上壘率)，是由個人表現所成的。故可推論，球隊勝率會直接受到球員表現所影響，球隊的表現是由球員表現累積而成，球員的表現直接地影響了球隊的勝率[6]。

2.2 主成份分析

主成份分析(Principal Component Analysis)[10]是一個歷史悠久的多變數分析方法，它的出現可以追溯到1901年，由Karl Pearson所提出，但是今天所較通用的方法則是出現在1933年，由Harold Hotelling在他論文裡提出。主成份分析就是尋找幾個解釋變數的線性組合，他們一方面要能保持原來變數的資訊(即具有代表性)，而且主成份間不能重疊(即有獨立性)，更重要的是能以少數幾個主成份代替原來多個解釋變數(即有精簡性，或是維度縮減)。主成份分析所著重的在於如何「轉換」原始變項使之成為一些互相獨立的線性組合變數，而且經由線性組合而得的主成份仍保有原變數最多的資訊，其關鍵在「變異數」問題，利用求特徵值(eigenvalue)及特徵向量(eigenvector)之方法，過濾出佔最大變異數的型態，最大變異數者則為第一主成份。由於主成份分析有簡化問題複雜度的優點，而且並不需要事先對於問題中所遇到大量變數間的關係加以瞭解就可以進行，所以非常適用於變數數量龐大或是變數間關係不明的問題上。

2.2.1 主成份分析說明

主成份分析所著重的在於如何「轉換」原始變項使之成為一些互相獨立的線性組合變數，而且經由線性組合而得的主成份仍保有原變數最多的資訊，其關鍵在「變異數」問題，利用求特徵值及特徵向量之方法，過濾出佔最大變異數的型態，此即為最主要之型態，且主成份會滿足下列性質：

1. 每一主成份為原來變數的線性組合。
2. 第一主成份必需能解釋最大的資料變異量。
3. 第二主成份必須能解釋扣除第一主成份後所剩餘的最大資料變異量。

4. 第 k 個主成份則解釋扣除 $k-1$ 個主成份後所剩餘的最大資料變異量。
5. 此 K 個主成份間是不相關的。

由以上性質可以看出，主成份座標中各座標軸互相正交，並且此主成份座標軸為唯一，也就是說在相同維度的空間中，再也找不到比此主成份座標軸可以解釋更多的資料變異量的另一組正交座標軸了。

2.2.2 主成份分析的目的

1. 概述變數間的關係。
2. 可將原來變數轉換成新的沒有相關的變數。
3. 可用來簡化多變量資料的維度，即降低變數個數，但亦會喪失部分資訊。
4. 可解決迴歸分析裡共線性問題。
5. 可用來作一組變數的綜合指標，如物價指數...

2.2.3 主成份分析的方法

主成份分析方法即為找出原有變數之線性組合使其變異數最大

假設有一個 $N \times M$ 階原始資料矩陣

$$\begin{aligned}
 \mathbf{X} &= [x_1 \ x_2 \ \cdots \ x_m \ \cdots \ x_M] \\
 &= \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nM} \\ \vdots & \vdots & \cdots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{bmatrix} \\
 &= \{x_{nm}\}_{n=1 \sim N, m=1 \sim M}
 \end{aligned}$$

其中 m 代表變數個數， N 代表資料點個數， \mathbf{x}_m 、 μ_m 、 σ_m^2 分別代表第 m 個變數的行向量、平均數及變異量。

$$\mathbf{x}_m = [\chi_{1m} \chi_{2m} \cdots \chi_{nm} \cdots \chi_{Nm}]^T$$

$$\mu_m = \frac{1}{N} \sum_{n=1}^N \chi_{nm}$$

$$\sigma_m^2 = \frac{1}{N} \sum_{n=1}^N (\chi_{nm} - \mu_m)^2$$

在進行主成份分析計算之前，首先將資料矩陣平均置中，亦即將矩陣中各元素減去各變數的平均數，使得矩陣中各行向量平均數為零，則可得到

$$\begin{aligned} \check{\mathbf{X}} &= \begin{bmatrix} \check{\mathbf{x}}_1 & \check{\mathbf{x}}_2 & \cdots & \check{\mathbf{x}}_m & \cdots & \check{\mathbf{x}}_M \end{bmatrix} \\ &= \begin{bmatrix} \check{\chi}_{11} & \check{\chi}_{12} & \cdots & \check{\chi}_{1M} \\ \vdots & \vdots & \vdots & \vdots \\ \check{\chi}_{n1} & \check{\chi}_{n2} & \cdots & \check{\chi}_{nM} \\ \vdots & \vdots & \cdots & \vdots \\ \check{\chi}_{N1} & \check{\chi}_{N2} & \cdots & \check{\chi}_{NM} \end{bmatrix} \\ &= \left\{ \check{\chi}_{nm} \right\}_{n=1 \sim N, m=1 \sim M} \end{aligned}$$

其中 $\check{\chi}_{nm} = \chi_{nm} - \mu_m$

其互變異矩陣(covariance matrix)可以計算如下

$$S = \text{cov}\left(\overset{\vee}{\mathbf{X}}\right) = \frac{\overset{\vee}{\mathbf{X}}^T \overset{\vee}{\mathbf{X}}}{N-1}$$

$$= \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1j} & \cdots & S_{1M} \\ S_{21} & S_{22} & \cdots & S_{2j} & \cdots & S_{2M} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ S_{i1} & S_{i2} & \cdots & S_{ij} & \cdots & S_{iM} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ S_{M1} & S_{M2} & \cdots & S_{Mj} & \cdots & S_{MM} \end{bmatrix}$$

其中
$$S_{ij} = \frac{\overset{\vee}{\mathbf{X}}_i^T \overset{\vee}{\mathbf{X}}_j}{N-1} \quad i=1\sim M, j=1\sim M$$

此互變異矩陣為一對稱矩陣，其中對角線上的各元素($i = j$)代表了各原始變數的變異數值，而非對角線上的元素則為其所在行、列代表的變數間的共變異數(covariance)。各變數的變異數平均顯示出資料點在各變數軸上散佈的寬度，而共變異數則顯示了變數間分布相關性的多寡。所以藉由計算此互變異矩陣之特徵值及特徵向量，便可以了解資料點的分布情況。

2.3 迴歸分析法

迴歸分析法，是在分析市場現象自變數和依變數之間相關關係的基礎上，建立變數之間的迴歸方程，並將迴歸方程作為預測模型，根據自變數在預測期的數量變化來預測依變數關係大多表現為相關關係，因此，迴歸分析法是一種重要的市場預測方法，當我們在對市場現象未來發展狀況和水平進行預測時，如果能將影響市場預測對象的主要因素找到，並且能夠取得其數量資料，就可以採用迴歸

分析法進行預測。它是一種具體的、行之有效的、實用價值很高的常用市場預測方法。

2.3.1 迴歸分析法的分類

迴歸分析法有多種類型。依據相關關係中自變數的個數不同分類，可分為一元迴歸分析法和多元迴歸分析法。在一元迴歸分析法中，自變數只有一個，而在多元迴歸分析法中，自變數有兩個以上。依據自變數和依變數之間的相關關係不同，可分為線性迴歸預測和非線性迴歸預測[4]。

2.3.2 迴歸分析法的步驟

1.根據預測目標，確定自變數和依變數

明確預測的具體目標，也就確定了依變數。如預測具體目標是下一年度的銷售量，那麼銷售量 Y 就是依變數。通過市場調查和查閱資料，尋找與預測目標的相關影響因素，即自變數，並從中選出主要的影響因素。

2.建立迴歸預測模型

依據自變數和依變數的歷史統計資料進行計算，在此基礎上建立迴歸分析方程，即迴歸分析預測模型。

3.進行相關分析

迴歸分析是對具有因果關係的影響因素（自變數）和預測對象（依變數）所進行的數理統計分析處理。只有當變數與依變數確實存在某種關係時，建立的迴歸方程才有意義。因此，作為自變數的因素與作為依變數的預測對象是否有關，相關程度如何，以及判斷這種相關程度的把握性多大，就成為進行迴歸分析必須要解決的問題。進行相關分析，一般要求出相關關係，以相關係數的大小來判斷

自變數和依變數的相關的程度。

4. 檢驗迴歸預測模型，計算預測誤差

迴歸預測模型是否可用於實際預測，取決於對迴歸預測模型的檢驗和對預測誤差的計算。迴歸方程只有通過各種檢驗，且預測誤差較小，才能將迴歸方程作為預測模型進行預測。

5. 計算並確定預測值

利用迴歸預測模型計算預測值，並對預測值進行綜合分析，確定最後的預測值。

2.4 羅序模式

羅序模式通行於歐洲地區的心理計量學界，以及美國芝加哥大學等大學，兩個參數對數形模式(Two-Parameter Logistic Model)[5]：這個模式的數學公式如下所示：

$$P_i(\theta) = \frac{e^{\alpha_i(\theta-b_i)}}{1 + e^{\alpha_i(\theta-b_i)}} \quad i=1,2,\dots,n \quad \dots\dots \text{公式(1)}$$

其中， $P_i(\theta)$ 表示任何一位能力為 θ 的考生答對試題 i 或在試題 i 上正確反應的機率； b_i 表示試題難度(difficulty)參數； α_i 表示試題鑑別度(item discrimination)，用來描述試題所具有鑑別力大小的特性； n 是該測驗的試題總數； e 代表以底為2.718的指數；且 $P_i(\theta)$ 是一種S形曲線，其值介於0與1之間。典型的二個參數的試題特徵曲線，可參見圖2所示。

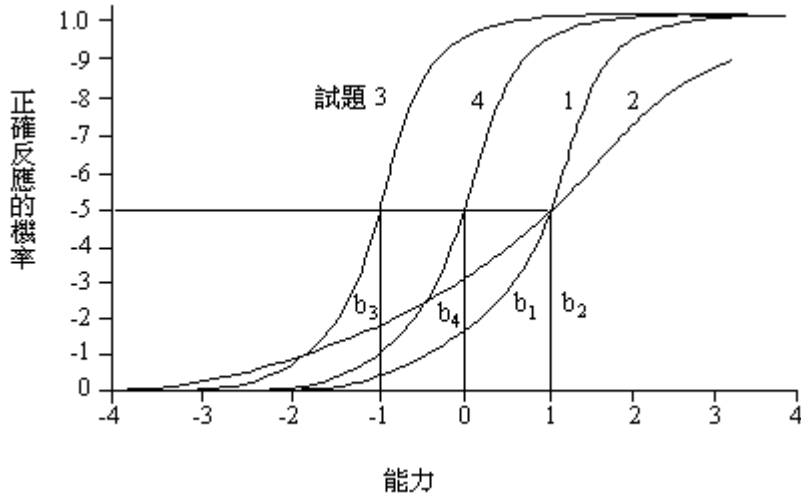


圖 2、四條典型的二個參數試題特徵曲線

試題鑑別度參數的值，剛好與在點的試題特徵曲線的斜率(slope)成某種比例。試題特徵曲線愈陡(steeper)的試題比稍平滑的試題，具有較大的鑑別度參數值；換句話說，鑑別度愈大的試題，其區別出不同能力水準考生的功能愈好，亦即分辨的效果愈好。事實上，該試題能否區別出以能力水準為，上下兩組（即高於和小於等於）不同能力考生的有效性，是與對應於量尺的試題特徵曲線的斜率成某種比例。

理論上，值的範圍在之間，但學者們通常捨棄負的值不用，因為該試題反向區別不同能力水準的考生，此外，帶有負值的試題特徵曲線代表著：能力愈高的考生答對某試題的機率愈低，這似乎與學理相違背，所以負的值不用。通常，值也不可能太大，常用的值範圍介於 0 與 2 之間；值愈大，代表試題特徵曲線愈陡，試題愈有良好的分辨能力；值愈小，代表試題特徵曲線愈平坦，正確反應的機率與能力間成一種緩慢增加的函數關係，亦即試題愈無法明顯有效地分辨出考生的能力水準。

很明顯的，二個參數對數形模式是由一個參數對數形模式延伸演變而來，亦即把試題鑑別度參數考慮進一個參數對數形模式裡，便成為二個參數對數

形模式。圖 2 所示，四條試題特徵曲線的試題參數分別為，這些參數決定試題特徵曲線的形狀不會是平行的，因為有不同大小的試題鑑別度值存在的關係。當這四條試題特徵曲線的值都相等時，這些曲線便成平行的 S 形曲線，如圖一所示；因此，我們可以這麼說：一個參數對數形模式是二個參數對數形模式的一種特例，亦即把試題鑑別度參數都設定成一致時（通常設定），公式二的數學式子便簡化成公式一的數學式子，這種說法於是成立。

由圖 2 亦可知，這些曲線的下限值都是零，亦即二個參數對數形模式並不把考生的猜題因素考慮在內，這點假設與一個參數對數形模式雷同。猜題因素不存在的假設，往往使二個參數對數形模式適用於自由反應(free-response)的試題分析，或試題不太困難的單選題測驗分析，對於有良好施測指導語的能力測驗資料亦可適用。

二個參數對數形模式是由 Birnbaum (1968)修改自 Lord (1952)的原始二個參數常態肩形模式(Normal Ogive Model)而來，由於它比常態肩形模式易於計算和解釋，目前已取代常態肩形模式，而成為主要的試題反應模式[7]。

第三章 研究方法

3.1 研究方法及架構

本研究建模式可應用於預測職業棒球未來比賽之結果，羅序模式是以機率的觀念來解釋受試者能力和測驗反應間之關係，就可推估出受試者正確反應的機率，所以將之應用在預測職業棒球未來比賽之結果上。本研究使用職業棒球比賽的攻守數據，並利用主成份分析方法定義影響比賽勝負的綜合性指標，接著應用線性迴歸分析，求出各球隊勝率估計值，針對兩兩球隊之勝率估計值，帶入羅序模式找出甲球隊贏乙球隊的機率。

3.2 彙整攻擊能力與守備能力因素

棒球是由守備力和攻擊力所組成的運動，所以一個球隊的戰力可從攻擊和守備兩方的數據來評估。攻擊能力是由打擊和跑壘的數據進行分析，打擊和跑壘的數據有打席、打數、安打、全壘打、被三振、獲四死球、盜壘成敗、打點、得分…等。守備能力是由投球和守備的數據進行分析，投球和守備的數據有失分、被安打、被全壘打、三振、四壞球、雙殺、失誤…等。

因棒球守備和攻擊的數據分析指標太複雜，故利用主成份分析之方法，先找出影響比賽勝負的綜合性指標。用主成份分析，找出原有守備和攻擊的數據數之線性組合，利用共變異數矩陣 S ，求出特徵值及單位特徵向量，則特徵值最大的為第一主成份，並取變異數最大的主成份來當作總合指標。

假設將攻擊能力原始數據變數 x_1, x_2, \dots, x_p 做線性組合，轉換為一組新的變數 z_1, z_2, \dots, z_p ,

$$\begin{aligned} z_1 &= a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p \\ z_2 &= a_{21}x_1 + a_{22}x_2 + \dots + a_{2p}x_p \\ &\vdots \\ z_p &= a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pp}x_p \end{aligned}$$

或表示為 $z = Ax$

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_p \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{bmatrix}$$

選擇一組組合係數，讓新變數 z 的變異數最大，即

$$\max_{\mathbf{u}} E(z^2) \equiv \max_{\mathbf{u}} \mathbf{u}^T \Sigma_X \mathbf{u}$$

新變數的變異數為

$$\text{var}(z) = E(z^2) = \mathbf{u}^T \Sigma_X \mathbf{u} = \lambda \mathbf{u}^T \mathbf{u} = \lambda$$

當 λ 等於 Σ_X 最大的特徵值時，其相對的特徵向量 \mathbf{v}_1 便是最佳的組合係數。此時的新變數稱為第一個主成分

$$z_1 = \mathbf{v}_1(1)x_1 + \mathbf{v}_1(2)x_2 + \dots + \mathbf{v}_1(p)x_p$$

3.3 發展勝率迴歸方程式

利用迴歸分析可用來找出兩個或兩個以上計量變數間的關係，並進而從一群變數中可以預測資料趨勢，所以可用來分析勝率與攻守數據的關係。設定勝率為依變數，攻擊能力與守備能力因素為自變數，可由函數關係中計算出勝率估計值。

3.4 建立勝負比機率模式

由於一般化線性模式是以機率的觀念來解釋受試者能力和測驗反應間之關係，可以將兩兩球隊之勝率估計值，帶入計算得出正確的反應，經過多次的實驗後，就可以計算出甲球隊贏乙球隊的機率。

第四章 實例分析

4.1 資料數據

攻守資料數據來源主要以中華職棒大聯盟為主，先將資料建立資料庫後，再使用 SPSS 統計軟體進行資料分析。

表 1、甲球隊攻擊數據資料

年度	打擊率	得分	安打	全壘打	保送	被三振	盜壘	上壘率	長打率
1	0.261	379	807	36	266	342	60	0.318	0.353
2	0.256	326	753	44	224	323	63	0.306	0.36
3	0.266	428	778	66	353	339	137	0.341	0.392
4	0.271	397	811	56	285	302	110	0.331	0.391
5	0.302	447	936	63	201	328	78	0.342	0.438
6	0.279	416	950	57	251	375	67	0.327	0.381
7	0.278	457	944	61	277	448	58	0.33	0.394
8	0.289	502	964	73	342	521	54	0.352	0.408
9	0.265	427	951	52	359	606	91	0.33	0.365
10	0.252	374	796	43	325	606	108	0.319	0.346

表 2、乙球隊攻擊數據資料

年度	打擊率	得分	安打	全壘打	保送	被三振	盜壘	上壘率	長打率
1	0.247	328	741	36	313	439	58	0.316	0.338
2	0.244	337	706	34	328	495	54	0.319	0.331
3	0.261	422	775	73	398	454	39	0.347	0.39
4	0.266	400	793	40	366	439	68	0.344	0.361
5	0.295	461	927	52	305	405	56	0.354	0.411
6	0.282	420	951	50	336	418	76	0.345	0.385
7	0.299	591	1048	72	328	518	118	0.356	0.429
8	0.28	524	929	95	437	618	87	0.361	0.427
9	0.262	513	923	46	424	625	131	0.339	0.36
10	0.267	414	813	47	332	549	171	0.335	0.369

表 3、丙球隊攻擊數據資料

年度	打擊率	得分	安打	全壘打	保送	被三振	盜壘	上壘率	長打率
1	0.253	340	739	37	338	442	97	0.328	0.355
2	0.259	386	755	42	351	457	116	0.336	0.35
3	0.248	300	729	40	328	493	89	0.322	0.334
4	0.28	495	952	59	392	523	118	0.351	0.394
5	0.242	406	791	53	401	734	103	0.324	0.339
6	0.261	409	862	47	363	593	76	0.332	0.363
7	0.261	410	880	51	325	584	51	0.325	0.356
8	0.301	625	1092	105	388	631	47	0.366	0.446
9	0.285	595	998	84	442	578	46	0.36	0.421
10	0.287	655	1201	80	481	736	104	0.356	0.402

表 4、丁球隊攻擊數據資料

年度	打擊率	得分	安打	全壘打	四死球	三振	盜壘	上壘率	長打率
1	0.269	423	910	64	349	645	82	0.335	0.379
2	0.275	495	922	57	365	566	67	0.344	0.387
3	0.272	501	968	77	440	579	61	0.35	0.397
4	0.232	304	723	36	253	562	86	0.289	0.312
5	0.263	396	797	60	265	424	114	0.321	0.375
6	0.246	364	760	47	277	476	62	0.306	0.338
7	0.276	417	831	54	317	511	69	0.343	0.39
8	0.275	473	945	62	343	551	102	0.339	0.39
9	0.259	450	864	55	342	649	96	0.326	0.356
10	0.259	431	886	45	258	563	74	0.309	0.353

表 5、甲球隊守備數據資料

年度	失分率	被安打	被全壘打	四死球	奪三振	雙殺	失誤	守備率	WHIP
1	3.907	823	57	240	355	205	145	0.96	1.371
2	3.462	785	59	275	439	189	116	0.967	1.308
3	3.277	771	74	297	415	209	125	0.964	1.361
4	3.474	827	54	216	469	196	110	0.97	1.409
5	3.11	796	53	169	465	236	120	0.967	1.325
6	3.223	877	63	266	482	216	117	0.971	1.273
7	4.276	1003	69	330	630	285	120	0.97	1.442
8	3.821	847	67	372	736	275	119	0.97	1.393
9	4.408	1041	72	457	625	266	114	0.973	1.293
10	3.881	813	64	396	523	224	122	0.967	1.359

表 6、乙球隊守備數據資料

年度	失分率	被安打	被全壘打	四死球	奪三振	雙殺	失誤	守備率	WHIP
1	5.604	1118	110	376	404	234	144	0.965	1.297
2	4.301	913	76	390	523	229	146	0.959	1.381
3	3.441	904	60	353	722	259	129	0.968	1.490
4	3.641	827	50	358	592	235	135	0.958	1.489
5	3.327	780	52	323	427	207	138	0.969	1.474
6	3.973	860	72	361	478	254	148	0.979	1.344
7	3.833	731	78	427	558	247	145	0.980	1.453
8	2.818	755	52	364	733	247	111	0.973	1.271
9	3.331	784	61	396	748	233	134	0.973	1.284
10	3.679	896	62	362	628	267	119	0.975	1.415

表 7、丙球隊守備數據資料

年度	失分率	被安打	被全壘打	四死球	奪三振	雙殺	失誤	守備率	WHIP
1	3.907	823	57	240	355	231	122	0.978	1.383
2	3.462	785	59	275	439	236	123	0.951	1.486
3	3.277	771	74	297	415	230	144	0.971	1.393
4	3.474	827	54	216	469	228	128	0.951	1.343
5	3.11	796	53	169	465	220	146	0.952	1.414
6	3.223	877	63	266	482	235	115	0.959	1.302
7	4.276	1003	69	330	630	266	148	0.971	1.408
8	3.821	847	67	372	736	235	143	0.969	1.434
9	4.408	1041	72	457	625	266	110	0.972	1.477
10	3.881	813	64	396	523	282	115	0.959	1.321

表 8、丁球隊守備數據資料

年度	失分率	被安打	被全壘打	四死球	奪三振	雙殺	失誤	守備率	WHIP
1	3.204	795	58	302	450	135	118	0.957	1.358
2	3.216	806	45	323	529	132	126	0.972	1.333
3	3.481	799	50	322	462	120	136	0.970	1.444
4	3.021	893	46	279	657	125	124	0.954	1.456
5	3.453	883	57	303	634	110	111	0.976	1.429
6	3.792	883	62	295	585	127	117	0.975	1.434
7	4.214	875	58	395	523	124	134	0.978	1.453
8	4.691	878	65	485	671	145	149	0.958	1.310
9	3.692	882	53	413	572	122	130	0.954	1.382
10	4.989	1060	95	510	697	137	137	0.955	1.442

4.2 主成份分析

因影響比賽勝負的攻守數據指標太複雜，故利用主成份分析之方法，分別找出影響比賽勝負之攻擊與守備綜合數據指標，萃取第一和第二主成份。

表 9、攻擊主成份分析結果-成份矩陣

	第一主成份	第二主成份
得分	0.9	0.299
安打	0.848	0.067
全壘打	0.827	-0.515
盜壘	-0.425	0.394
保送	0.744	-0.039
被三振	0.525	-0.743
打擊率	0.783	0.542
上壘率	0.856	0.421
長打率	0.95	-0.103

表 10、守備主成份分析結果-成份矩陣

	第一主成份	第二主成份
失分率	0.639	0.598
被安打	0.699	0.438
被全壘打	0.818	-0.418
四死球	0.771	0.247
奪三振	0.471	-0.741
WHIP	0.736	0.632
失誤	-0.782	0.512
雙殺	0.787	0.206
守備率	0.794	-0.522

4.3 求出勝率估計值

將由主成份分析結果的攻擊與守備成份矩陣，分別與攻擊與守備數據矩陣相乘。以相乘結果設為自變數，另設定勝率為依變數，帶入線性迴歸方程式，可求得常數 β_0 與標準化係數 β_1 、 β_2 、 β_3 、 β_4 。假設 Y_i 為 i 年勝率估計值， X_{i1} 為 i 年攻擊數據與攻擊第一主成份的積， X_{i2} 為 i 年攻擊數據與攻擊第二主成份的積， X_{i3} 為 i 年守備數據與守備第一主成份的積， X_{i4} 為 i 年守備數據與守備第二主成份的積。

表 12、球隊勝率估計值

	勝率估計值
甲球隊	0.4028
乙球隊	0.3903
丙球隊	0.5376
丁球隊	0.4256

4.4 結果與分析

將上述求出之勝率估計值帶入羅序模式(公式 2)運算，即可以計算出球隊贏球機率 $P(X)$ 。設 $c1$ =甲球隊勝率估計值， $c2$ =乙球隊勝率估計值，經由公式計算結果 $P(X)=0.5031$ ，由此可知甲球隊贏乙球隊的機率為 0.5031。

$$P(x) = \frac{1}{1 + e^{-(c1-c2)}} \quad \dots\dots \text{公式(2)}$$

表 13、甲球隊與其他球隊贏球機率預測結果

	贏球機率預測
甲球隊 VS 乙球隊	0.5031
甲球隊 VS 丙球隊	0.4663
甲球隊 VS 丁球隊	0.49431

表 14、乙球隊與其他球隊贏球機率預測結果

	贏球機率預測
乙球隊 VS 甲球隊	0.4968
乙球隊 VS 丙球隊	0.4632
乙球隊 VS 丁球隊	0.4911

表 15、丙球隊與其他球隊贏球機率預測結果

	贏球機率預測
丙球隊 VS 甲球隊	0.5336
丙球隊 VS 乙球隊	0.5367
丙球隊 VS 丁球隊	0.5279

表 16、丁球隊與其他球隊贏球機率預測結果

	贏球機率預測
丁球隊 VS 甲球隊	0.5056
丁球隊 VS 乙球隊	0.5088
丁球隊 VS 丙球隊	0.4720

第五章 結論與建議

本研究嘗試建立一個球隊贏球機率的計算模式，利用職業棒球比賽的攻守數據資料，使用主成份分析來將眾多的攻守數據，轉換成簡單的綜合性指標，讓評估更加容易，經由羅序模式，可以正確的反應出受測者的反應也就是球隊贏球機率，最後建立一個參考之準則，可提供球隊在賽前針對不同對戰球隊，安排最佳的攻守陣容，以增加贏球的機率。由於本研究主要以中華職棒大聯盟攻守資料數據，來進行資料萃取與分析，然而有些影響比賽勝負的因素，無法從數據上分析，如選手士氣、場地狀況、裁判判定標準、球隊管理狀況..等因素，若能將這些因素加以資料化，並加入本研究建立的模式，可更加客觀地預測比賽的勝負。

參考文獻

1. 葉志仙，”棒球比賽成績記錄分析研究”， 1996， 1-38。
2. 葉麗琴，”慢式壘球記錄法之研究分析”， 1996， 50-53。
3. 程紹同、方信淵、洪嘉文、廖俊儒、謝一睿，”運動行銷商戰剖析”， 2001， 35-46。
4. 陳耀茂，”多變量分析導論”， 2000， 12-15。
5. Barker, D. G., “The factor structure of major league baseball records” in Research ,1994.
6. Scully, G. W., “Performance in major league baseball” in The American Economic ,1997.
7. Baker, F. B. Item response theory: parameter estimation techniques. N. Y. Marcel, 1992.
8. Hambleton, R. K. and Swaminathan, H. Item response theory: 1985.
9. Hambleton, R. K. and Swaminathan, H. and Roger, H. J. Fundamentals of item response theory . New burry Park, CA:SAGE , 1991.
10. Kouzani, A. Z. ; He, F. ; Sammut, K. ” Quadtree principal component analysis and its application to facial expression classification” ,1999