

各球隊明星球員與球隊勝負之關聯性

國立清華大學經濟系

陳妍廷

廖芷萱

葉千瑀

林品鈞

指導教授：林世昌教授

摘要

本研究旨在探討中華職棒(CPBL)中明星球員表現對球隊勝負之影響，使用 2017 至 2024 年間四支球隊之逐場比賽資料，並採用 Logistic 迴歸模型與隨機森林模型(Random Forest)進行實證分析。

在 Logistic 迴歸模型中，結果顯示明星球員人數與主場優勢對於球隊勝負具統計上顯著的正向影響($p < 0.05$)；然而，進攻與防守表現指標則未達顯著水準，顯示單一績效指標對勝負之影響具侷限性。

為克服傳統迴歸模型之線性限制與變數共線性問題，本研究進一步採用隨機森林模型評估各變數對預測勝負之相對貢獻度。結果指出，明星球員人數仍為影響勝負之最關鍵因子，其次為得分數、外籍球員數、OPS 與打點，突顯核心球員數量與其進攻貢獻在非線性條件下的重要性。

綜合分析結果，明星球員對球隊勝負確實具基礎性影響，但比賽結果仍深受教練調度、臨場表現與戰術安排等複合非線性因素影響，難以單憑個別指標全面評估球隊整體戰力。

Syllabus

This study aims to examine the impact of star player performance on team outcomes in the Chinese Professional Baseball League (CPBL), utilizing game-level data from 2017 to 2024 for four teams. Empirical analyses are conducted using both logistic regression and random forest models.

In the logistic regression model, the results indicate that the number of star players and home-field advantage have a statistically significant positive effect on the likelihood of winning ($p < 0.05$). However, offensive and defensive performance indicators do not reach statistical significance, suggesting that the influence of individual performance metrics on game outcomes may be limited.

To address the linearity and multicollinearity constraints of traditional regression models, this study further employs a random forest model to evaluate the relative importance of each variable in predicting game outcomes. The results highlight that the number of star players remains the most influential factor, followed by runs scored, number of foreign players, OPS, and RBIs. These findings underscore the

importance of core player contributions, particularly in offensive performance, under non-linear conditions.

In conclusion, while star players have a fundamental impact on team outcomes, game results are also shaped by a complex array of non-linear factors, including managerial decisions, in-game performance, and strategic planning. Therefore, it is challenging to comprehensively assess a team's overall strength based solely on individual performance indicators.

壹、緒論

一、研究動機

1996 年至 2009 年，中華職棒歷經了五次的假球事件，而又以 2009 年的黑象事件影響最甚。在黑道以金錢誘惑及暴力脅迫的情況下，使得許多球星牽涉其中，並重創中華職棒形象。在簽賭事件頻傳的幾年間，許多球迷開始對中華職棒喪失信心，進場觀賽人數寥寥可數，單場人數甚至低至 100 人左右。直至中職聯盟與政府建立相關措施，全力防範假球案再度發生，以及增加球員保障後，球迷才漸漸地回到球場。

近年來，在各球團陸續推出多變化的經營模式，及新一代球員的努力下，中華職棒逐漸走出簽賭風波，朝正向發展。而去年的世界十二強棒球賽，中華隊則第一次的打敗日本全明星職業隊，站上世界殿堂。那一晚，全臺灣球迷再一次感受到了棒球所帶來的熱血與感動，更凸顯出臺灣人對棒球的關注度。

目前中華職棒擁有六隊職業隊，包含中信兄弟、統一獅、樂天桃猿、富邦悍將、味全龍和台鋼雄鷹。不論是哪一隊，總有球員是較為核心的存在，其大多擔任王牌投手、中心棒次等角色。普遍認為擁有較多明星球員的球隊，在每隊 120 場例行賽下來，也將會有最佳的勝率，能進入到日後的季後賽，甚至是最後的總冠軍賽。然而，每年的球季結果顯示，並非擁有最強球員的那一隊便會擁有較高的勝場並獲得年度總冠軍。除此之外，在國際賽上也常看見當年度的中華隊評為最強戰力，納入了各隊明星球員，但最後表現不入預期之案例。

綜上所述，本研究將從打擊及投球兩面向，驗證出各隊明星球員的個人表現是否與球隊勝負有重大關聯性。

二、研究目的

基於上述研究動機，為找出影響球隊勝負的因素，本文的自變數¹將從明星球員的兩方面來著手——打擊表現及投球表現。以下為其自變數說明：

- 1. 打擊表現：打擊率（BA）、整體攻擊指數（OPS）、打點（RBI）、得分（R）
- 2. 投球表現：防禦率（ERA）、三振數（K）、每局被上壘率（WHIP）
- 3. 主客場、外籍球員

至於應變數的部分則將使用球隊的單場勝負（以虛擬變數表示）。而近年來雖新增了兩個球隊——味全龍和台鋼雄鷹，但由於本文將探討於 2017 年至 2024 年中的數據，最終決定只以中信兄弟、樂天桃猿（前身為 Lamigo）、富邦悍將、統一獅四支球隊來進行分析。此外，由於每場比賽都有所謂的主客場制度以及外籍球員，其可能影響球隊之表現。因此，本文將主客場、外籍球員及是否換過比賽用球，設為控制變數，以找出明星球員與球隊勝負之最適解釋。由此驗證各隊明星球員在某一年度的整體表現，是否會影響該隊在單場比賽中的勝負結果。

三、名詞解釋

打擊率 (Batting average , BA)	是常見的棒球統計數據之一，即為擊球員上場擔任打擊任務時，擊出安打的或然率。其計算公式為安打/打數。
整體攻擊指數 (On-base percentage Plus Slugging Percentage , OPS)	又稱上壘加長打率，是上壘率 (OBP) 與長打率 (SLG) 的總和。
打點 (Run Batted In , RBI)	是指打者藉由每次打席，使壘包上的跑壘員或打者本身都跑回本壘得分。

¹ 明星球員——打者：出賽數>70 和 OPS>0.8、投手：投球局數>40 和 ERA<3.2

得分 (Run, R)	擊球員將投手投出的球擊出到界內，使跑壘員依序從一壘、二壘、三壘到本壘，每個跑壘員完成此動作時得 1 分。
出賽數 (Games Played, G)	球員上場的次數。
上壘率 (On base percentage, OBP)	常見的棒球統計數據之一，即為擊球員上場擔任打擊任務時，獲得上壘機會的或然率
長打率 (Slugging Percentage, SLG)	衡量擊球員長打能力的數字。
防禦率 (ERA, Earned Run Average)	評估投手表現的最基本也是最重要的指標之一，用來衡量一名投手平均每九局會失掉多少「自責分」(Earned Runs)，這個指標的數值越低代表投手越優秀，因為代表他較少讓對手得分。
三振數 (K, Strikeouts)	衡量投手能力的重要指標，表示投手平均每九局能夠送出多少次三振。
被上壘率 (WHIP)	衡量投手讓打者上壘的頻率的一個指標。該數值越低越好，代表投手控球能力較強，較少讓對方打者上壘

貳、文獻回顧

一、打擊表現

在職業棒球的勝負影響因素中，進攻端的表現向來被視為關鍵。多數實證研究指出，球隊整體進攻效率與明星打者的打擊能力，對戰績具有顯著影響。因此，本研究選擇以明星球員的打擊表現作為主要自變數，探討其與球隊勝負之間的關聯性，所使用的變數包括生涯打擊率(BA)、整體攻擊指數(OPS)、打點

(RBI)與得分(R)。

首先，Lewis（2003）在《Moneyball》²一書中提出傳統棒球評估體系的盲點，主張應以更具代表性的進攻指標來衡量球員價值。其中，OPS（OBP+SLG）被認為能同時反映選球與長打能力，是判斷一位打者對球隊得分貢獻的整體性指標。後續學者如 Bradbury（2007）³與 Albert & Bennett（2001）⁴亦指出，OPS 與球隊勝率之間存在高度正相關，其在迴歸模型中的解釋力甚至超過打擊率（BA）等傳統變數。

儘管如此，傳統指標仍具備一定參考價值。Krautmann（1999）⁵於其分析 MLB 球員薪資與績效關係的研究中，將打點（RBI）與得分（R）視為反映實質進攻產出的指標，並指出這些與最終比賽結果密切相關。特別是 RBI，作為關鍵時刻得分的直接反映，常被用以衡量球員在壓力情境下的貢獻能力。Gius & Johnson（2000）⁶進一步指出得分與球隊勝率具高度一致性，特別是在季後賽等高張力比賽中，進攻核心表現與球隊表現高度同步。

此外，生涯表現變數中的打擊率（BA）亦被視為評估球員穩定性與長期價值的重要指標。Groot 與 Garcia-del-Barrio（2019）⁷針對歐洲棒球聯盟與 MLB 的比較研究發現，長期累積數據能有效剔除偶發性因素，突顯球員真實實力，尤其適用於明星球員評價。

綜合上述文獻可知，本研究選取的四項變數具有紮實的學理依據與實證支持，能從不同角度呈現球員打擊端對球隊戰績的影響程度。透過生涯性指標的納入，更可控制短期表現的隨機波動，提升模型的穩定性與解釋力，進而深入探討明星球員與球隊勝率之間的關聯。

² Lewis, M. (2003). *Moneyball: The Art of Winning an Unfair Game*. W. W. Norton & Company

³ Bradbury, J. C. (2007). *The Baseball Economist: The Real Game Exposed*. Dutton.

⁴ Albert, J., & Bennett, J. (2001). *Curve Ball: Baseball, Statistics, and the Role of Chance in the Game*. Copernicus Books.

⁵ Krautmann, A. C. (1999). What's wrong with Scully-estimates of a player's marginal revenue product. *Economic Inquiry*, 37(2), 369–381.

⁶ Gius, M., & Johnson, D. (2000). An empirical estimation of the determinants of major league baseball attendance. *Journal of Business and Economic Studies*, 6(2), 55–64.

⁷ Groot, L., & Garcia-del-Barrio, P. (2019). Superstar productivity and team performance: Evidence from European baseball. *Applied Economics Letters*, 26(5), 379–384.

二、投手表現

投手表現向來被視為影響棒球比賽勝負的關鍵因素之一。

楊士鋒（2017）⁸以中華職棒 2014 至 2016 年總冠軍賽為研究樣本的實證研究指出，根據防禦率（ERA）、被打擊率等投手指標與球隊比賽勝率有顯著負相關，顯示控球能力與壓制對手得分的能力在季後賽中尤為關鍵。此外，梁功斌（2016）⁹在探討影響中華職棒選手薪資與表現因子的研究中，亦發現投手的勝場數、防禦率皆與其市場價值與球隊表現具有高度相關。該研究也指出，在中職這類中小型聯盟中，少數優秀投手往往對整季戰績具有拉抬效果。陳志成（2016）¹⁰則建構了中華職棒球員績效評估模式，進一步將投手貢獻值量化，結果顯示 WHIP 與勝場率（WPCT）對球隊勝率有高度解釋力，尤其在例行賽中，穩定的投手輪值更有助於拉長勝場優勢。國外 Hakes 與 Sauer（2006）¹¹實證發現，傳統指標如防禦率固然重要，但若加上進階指標如被上壘率(WHIP)，更能有效預測球隊整體防守表現與比賽勝負。

另一方面，外籍球員在中華職棒扮演關鍵角色，尤其在投手戰力的補強方面影響深遠。方進義等人（2013）¹²針對 1990 至 2010 年間的中職資料進行分析，結果顯示外籍球員無論在打擊或投球端，表現均優於本土球員，尤其在勝場率（WPCT）、防禦率（ERA）等項目中，洋將普遍具備更高競爭力。陳志成（2016）也指出，外籍投手通常被安排為先發王牌，且多數在 ERA 與 WHIP 等指標上表現優異，對球隊勝率具有顯著正向貢獻。

綜合上述文獻，我們發現投手的多項表現指標對勝率皆具有顯著影響，而外籍球員的優勢則使其成為各隊爭取佳績的重要戰力來源。因此，我們決定將

⁸ 楊士鋒（2017）。《棒球比賽影響勝負之因素探討：以中華職棒 2014-2016 年總冠軍賽為例》。碩士論文。

⁹ 梁功斌（2016）。影響中華職棒選手薪資與績效表現因子之探討。國立臺灣師範大學運動與休閒學系。

¹⁰ 陳志成（2016）。建構中華職業棒球大聯盟球員績效評估之模式及實證分析。國立臺灣師範大學體育學系碩士論文。

¹¹ Hakes, J. K., & Sauer, R. D. (2006). An introduction to sabermetrics and the economics of baseball. *Journal of Economic Perspectives*, 20(1), 127-142.

¹² 方進義、鄭志富、林欣慧（2013）。中華職棒野手球員的國籍、年資、球隊效果與換隊頻率對其績效之影響：1990 年到 2010 年。《體育學報》，46(3)，291-302。

ERA、WHIP、K 與外籍球員人數納入分析模型中。

三、明星球員

明星球員一直以來都為職棒比賽的重要亮點，中華隊在世界盃十二強中獲得冠軍後，各個球團皆透過高薪盡力延攬或留下明星球員。例如：陳傑憲與其球隊簽訂了總值 2 億元的新合約，味全龍的陳子豪則簽下了 1.3 億元的合約，而中信兄弟的江坤宇則獲得了總值 1.4788 億元的合約，這些合約時長皆長達十年。¹³高額的合約簽訂凸顯了球隊對核心球員的高度依賴，並反映出明星球員在市場上所帶來的經濟價值。

除了明星球員在比賽中發揮自身能力之外，研究也顯示，球員的優異表現對於球迷觀賽意願具有正面影響。在中職球團球賽的收入中，門票及周邊商品大約佔 50%¹⁴（周義峰，2021），由此可知，球迷的觀賽意願及對球隊的喜愛程度對球團收入具有重要影響，並且是球隊能否長期穩定經營的關鍵因素之一。此外，明星選手的卓越表現也會顯著提升球迷的觀賽態度，特別是在當觀眾對比賽產生忠誠度後，這種影響尤為明顯。根據研究，86%的球迷會對職棒觀賽產生忠誠度，並持續關注比賽¹⁵（楊鈞皓，2014）。另外，明星球員數量的多寡，對於觀眾的進場意願也極為重要，劉振家、黃德舜和陳育成指出，職棒球隊明星球員與觀眾人數之關聯是呈現正相關且極具顯著性的¹⁶（劉振家、黃德舜、陳育成，2010）。

在過往的研究當中，我們能得出明星球員對於觀眾是否願意進場觀賽有顯著的影響，且這對於球團持續經營有著重要的貢獻，但較少文獻研究明星球員的個人表現，是否會直接影響球隊勝率，也較少研究聚焦於是否擁有較多明星球員的球隊會有較高的勝率。因此，本研究將從明星球員的打擊表現以及投球

¹³ 路皓惟，〈中職／陳傑憲簽 10 年 2 億巨約！盤點歷史十大合約 大頂薪時代來臨〉，今日新聞 2025 年 3 月 21 日，
<https://www.nownews.com/news/6661619?srsId=AfmBOoqKX8AcKX3JIM4EVTsBv9shgyGd6mbk-sT59SB66JxCURJiyYN>

¹⁴ 周義豐，〈中職玩真的！三六〇〇億元撐腰〉，卓越雜誌 2021 年 6 月 1 日，
<https://www.ecf.com.tw/tw/article/show.aspx?num=6188&page=2&kind=2>

¹⁵ 楊鈞皓，〈明星效應對職棒比賽觀看意願之影響〉，逢甲大學，2014 年 3 月 6 日

¹⁶ 劉振家、黃德舜、陳育成，〈職業運動賽事之觀眾人數及其影響因素分析—美國、日本、韓國、台灣之職業棒球聯盟比較〉，休閒世界研究第八卷第一期，2000 年 3 月，頁 126~142

表現著手，探討各隊核心球員對球隊勝負的影響。

參、模型解釋

一、模型資料

數據蒐集：

數據選取時間從 2017 年至 2024 年，共 8 年。樣本為中信兄弟、樂天桃猿（2020 年以前為 Lamigo）、富邦悍將、統一獅在一軍例行賽中之表現，數據來源為中華職棒大聯盟全球資訊網。

核心球員篩選標準：

1. 打者：出賽數>70 且 OPS>0.8
2. 投手：投球局數>40 且 ERA < 3.2

首先過濾和局資料，並進行標準化處理後，使用 Logistic 迴歸分析，迴歸式如下：

$$\text{logit}(P(\text{Win}_{it}=1)) = \beta_0 + \beta_1 * \text{OPS}_i + \beta_2 * \text{RBI}_i + \beta_3 * R_i + \beta_4 * \text{ERA}_i + \beta_5 * \text{WHIP}_i + \beta_6 * \text{Home}_i + \beta_7 * \text{Star}_i + \beta_8 * \text{Foreign}_i + \beta_9 * \text{ball}_i + \varepsilon_i$$

以下為變數解釋：

Win：為 Dummy Variable，勝為 1，負為 0，和局為 NA 值

OPS：各隊核心球員該年度 OPS 加權平均之結果

RBI：各隊核心球員該年度 RBI 加權平均之結果

ERA：各隊核心球員該年度 ERA 加權平均之結果

WHIP：各隊核心球員該年度 WHIP 加權平均之結果

Home：為 Dummy Variable，主場為 1、客場為 0

Star：各隊該年度符合核心球員標準之數量

Foreign：各隊該年度之洋將數量

ball：是否換過比賽用球，換為 1，沒換為 0

二、迴歸結果

OPS	RBI	R	ERA	WHIP	Home	Star	Foreign	ball
3.614610	5.096886	4.673405	4.757800	4.693540	1.000332	1.405513	1.539819	2.106198

在進行模型解釋之前，本研究亦檢查了模型各變數之間是否存在高度相關性。依結果顯示，各變數 VIF 值已低於常見警告值 5。因此，可認為此模型並未存在嚴重的多重共線性問題，估計結果可信。

	Estimate	Std. Error	z value	Pr(> Z)
(Intercept)	-0.69430	0.13880	-5.002	5.67e-07
OPS	-0.01386	0.06341	-0.219	0.826921
RBI	0.10069	0.07454	1.351	0.176758
R	0.04664	0.07147	0.653	0.514056
ERA	0.01249	0.07143	0.175	0.861146
WHIP	-0.01155	0.07096	-0.163	0.870668
Home	0.24459	0.06551	3.734	0.000189
Star	0.07136	0.01890	3.777	0.000159
Foreign	0.01397	0.02395	0.583	0.559605
ball	0.13747	0.09506	1.446	0.148138

另外，因 Logistic 迴歸結果可能較偏向於勝場的條件，為更平衡及準確地探討勝敗和明星球員表現的關聯性，並提升解釋力，本研究再採用了雙邊設計，以球隊為分析單位，保留單場兩隊的獨立資訊。而為避免同一場比賽資料間的相依性，本研究亦採用叢集式標準誤修正，解決重複根據及相關性問題，迴歸結果如下：

	Estimate	Std. Error	z value	Pr(> Z)
(Intercept)	-0.694302	0.151000	-4.5980	4.265e-06
OPS	-0.013864	0.087925	-0.1577	0.8747080
RBI	0.100688	0.071655	1.4052	0.1599682
R	0.046638	0.049487	0.9424	0.3459793
ERA	0.012493	0.073159	0.1708	0.8644036
WHIP	-0.011553	0.079387	-0.1455	0.8442925
Home	0.244588	0.067052	3.6477	0.0002646
Star	0.071360	0.017253	4.1360	3.534e-05
Foreign	0.013973	0.022645	0.6170	0.5372094
ball	0.137467	0.078637	1.7481	0.0804410

依上述迴歸結果顯示，明星球員的人數對於球隊勝負具有顯著的正向影響，p-value 小於 0.05，表示明星球員越多的球隊獲勝機率越高；主客場在迴歸結果中也呈現顯著性，p-value 小於 0.05，表示主場的確帶來優勢，然而，以各個球隊明星球員的表現切入，對於球隊勝率皆無顯著性，推測原因如下：

1. 單一指標難以反映球隊整體戰力
2. 勝負仍可能受到教練調度、戰術、臨場發揮等複合因素影響

實證結果顯示，核心球員的數量對球隊勝負具有統計上的顯著性。然而，單場比賽的勝敗並不能以單一指標去做定論。

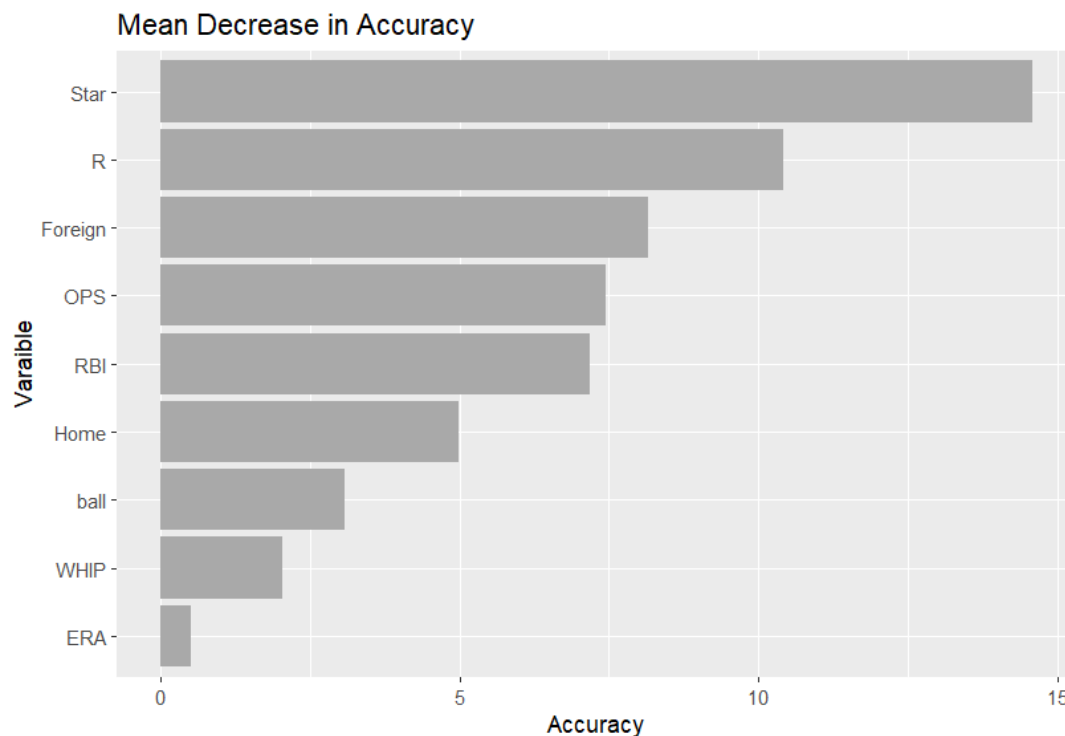
三、隨機森林

為避免迴歸的線性限制及共線性問題，我們再採取了 Random Forest 模型

來去預測每個變數對於模型的貢獻度。結果如下：

	Mean Decrease Accuracy
Star	16.66684056
R	11.12637975
Foreign	7.13062411
RBI	6.53635527
Home	5.88848080
OPS	5.66255271
ball	2.81089668
ERA	-0.05371607
WHIP	-1.29811591

在 Mean Decrease Accuracy 中，其評估了一個變數不存在於模型當中時，會讓此模型減少了多少的預測力，數值愈大代表該變數愈重要。



基於 Logistic 迴歸和 Random Forest 的結果，本研究認為明星球員的表現對於球隊的能否贏球仍具一定的基礎性。然而，各個表現是相輔而成的，一支球隊必須要同時擁有進攻及防守兩者之間的能力，而不能只以單一指標去判斷球隊戰力的強弱。除此之外，每一場比賽中的先發投手、球員發揮、戰術運用、球員調度等非線性因素都可能成為影響球隊勝敗的關鍵因素。

肆、模型改善方向-以美國實例說明

於前方結果中可得知 Logistic 迴歸模型及 random forest 模型兩者的重要變數結果不盡相同，我們推斷在 Logistic 迴歸當中，明星球員單獨各項指標會受到非線性因素，如教練調度、臨場發揮現場支陣容安排等影響，而無法呈現其重要性，因本研究資料集所使用數據為明星球員之年度加權平均表現，難以反映其單場表現的變異性，亦無法掌握賽場上的即時動態與策略改變。

目前中華職棒大聯盟全球資訊網空開數據中，無法取得單場策略及臨場狀況之數據，但在美國職棒中已導入相關技術。自 2025 年春訓起，MLB 開始使用自動電子好球帶挑戰系統（ABS Challenge System），此項系統的加入在投手以及打者的監測表現上有更細緻的紀錄。在投手表現方面，ABS 系統能更精確地分析其投球落點的準確性，特別是對好球帶邊緣的掌控能力；在打者選球部分，則可更清楚掌握其對好球帶的辨識能力，像是 chase rate 與 take rate 等指

標，能更真實反映打者的選球傾向，降低因裁判主觀判決所帶來的誤差。同時，ABS 一致性的判斷標準，於研究模型中增加投手控球表現、選球能力等參數提高模型完整度能讓預測結果能更加準確。

除了近期納入的 ABS 系統外，Statcast 分析系統自 2015 年起於美國職棒啟用¹⁷，亦是在 MLB 中的核心數據資料庫，透過高速攝影機以及雷達技術¹⁸，分析投手的控球能力以及打者的揮棒路徑等。例如以離球初速等進階各項指標評估打者的表現，再搭配上平均的年度基礎指標能更加精準預估球隊整體表現。

目前當場比中的各項進階表現數據在中華職棒當中為各個球隊的內部資料，尚無對外公開也難以取得，本研究也難以取得相關資料。若未來能取得類似於上述的即時及進階數據，將有助於建立更貼近實際情境的預測模型，並進一步提升本研究的實證價值與應用可能性。

伍、結論

本研究以 2017 年至 2024 年間中華職棒四支球隊的逐場比賽資料為基礎，探討明星球員人數、打擊與投球表現，以及外籍球員與主場優勢對球隊勝負之影響。實證方法採用 Logistic 迴歸與隨機森林模型，兼顧統計推論與預測分析，提升模型的整體解釋力與應用價值。

Logistic 迴歸結果顯示，明星球員人數與主場優勢對球隊勝率皆具有顯著正向影響，顯示具備較多核心球員的球隊，以及主場作戰時，明顯較易取得勝利。主場效應的顯著性也反映了球迷支持、熟悉場地與旅程疲勞等因素在賽事中的潛在影響。然而，OPS、ERA、RBI 等個別表現指標未達統計顯著，推測原因為單一年度指標難以反映球隊整體戰力與實際表現波動，例如 OPS 或 ERA 等數值僅呈現年度明星球員加權平均，無法捕捉選手在關鍵比賽或特定情境下的發揮，亦難反映其對勝負的即時貢獻。第二點為比賽結果受無法量化因素影響，如教練調度、戰術安排、對戰組合、臨場狀態等，非線性因素，無法單靠球員個人表現指標完整解釋。

因此，即便明星球員具有較高的個人能力，其貢獻仍需放在整體團隊運作

¹⁷ 邱詠仲，〈機器學習方法運用於棒球教練決策之研究：以投打有利對決與打者價值預測為例〉，國立宜蘭大學，2022 年 7 月

¹⁸ Mike Petriello, "Everything to know about Statcast's new bat-tracking data", May 13th, 2024, <https://www.mlb.com/news/what-you-need-to-know-about-statcast-bat-tracking>

脈絡中解讀，而非以單一數據衡量之。隨機森林模型則進一步強調了非線性關係與變數交互效果的重要性。結果顯示，明星球員人數為預測勝敗最關鍵因素，其次為得分、外籍球員人數、OPS 與打點等，突顯當模型不受限於線性假設時，這些指標的重要性更加顯著。特別是外籍球員人數，雖在 Logistic 模型中未達顯著，但在隨機森林中排名第三，顯示外籍選手對球隊戰力具有潛在貢獻，尤其在先發投手位置上，洋將仍為提升勝率的重要資源。

在 Logistic 模型與隨機森林模型比較中，本研究發現傳統的 Logistic 模型可能因無法處理非線性關係，而限制了明星球員表現變數的解釋力。我們推測明星球員的各項表現可能受到臨場發揮、教練調度與對戰組合等因素干擾，加上本研究採用年度加權平均數據，難以捕捉單場變異性與即時策略變化，致使模型無法呈現其潛在影響力。

目前中華職棒大聯盟雖提供例行數據，但尚未公開如單場策略與即時表現等進階指標。相較之下，美國職棒大聯盟（MLB）已逐步導入如自動電子好球帶挑戰系統與 Statcast 分析系統等科技技術，能夠蒐集打者選球傾向（如 chase rate、take rate）、投手控球精度、擊球初速等資訊，進一步豐富球員評估與賽事預測的資料基礎。這些進階數據可顯著提升模型預測力，並彌補傳統年度統計之不足。

綜合而言，明星球員與主場優勢為中職球隊爭取勝利的核心要素，但比賽勝負仍深受臨場表現、教練調度與策略安排等無法量化的因素影響，明星球員加權平均表現指標雖難以證明對球隊勝負有顯著影響，但透過隨機森林模型分析仍可凸顯其重要性。

陸、參考資料

Lewis, M. (2003). *Moneyball: The Art of Winning an Unfair Game*. W. W. Norton & Company

Bradbury, J. C. (2007). *The Baseball Economist: The Real Game Exposed*. Dutton.

Albert, J., & Bennett, J. (2001). *Curve Ball: Baseball, Statistics, and the Role of Chance in the Game*. Copernicus Books.

Krautmann, A. C. (1999). What's wrong with Scully-estimates of a player's marginal revenue product. *Economic Inquiry*, 37(2), 369–381.

Gius, M., & Johnson, D. (2000). An empirical estimation of the determinants of major league baseball attendance. *Journal of Business and Economic Studies*, 6(2), 55–64.

Groot, L., & Garcia-del-Barrio, P. (2019). Superstar productivity and team performance: Evidence from European baseball. *Applied Economics Letters*, 26(5), 379–384.

楊士鋒（2017）。《棒球比賽影響勝負之因素探討：以中華職棒 2014–2016 年總冠軍賽為例》。碩士論文。

梁功斌（2016）。影響中華職棒選手薪資與績效表現因子之探討。國立臺灣師範大學運動與休閒學系。

陳志成（2016）。建構中華職業棒球大聯盟球員績效評估之模式及實證分析。國立臺灣師範大學體育學系碩士論文。

Hakes, J. K., & Sauer, R. D. (2006). An introduction to sabermetrics and the economics of baseball. *Journal of Economic Perspectives*, 20(1), 127-142.

方進義、鄭志富、林欣慧（2013）。中華職棒野手球員的國籍、年資、球隊效果與換隊頻率對其績效之影響：1990 年到 2010 年。《體育學報》，46(3)，291-302。

路皓惟，〈中職／陳傑憲簽 10 年 2 億巨約！盤點歷史十大合約 大頂薪時代來臨〉，今日新聞 2025 年 3 月 21 日，

<https://www.nownews.com/news/6661619?srsltid=AfmBOoqKX8AcKX3JIM4EVTeSbV9shgyGd6mbk-sT59SB66JxCURJiyYN>

周義豐，〈中職玩真的！三六〇〇億元撐腰〉，卓越雜誌 2021 年 6 月 1 日，

<https://www.ecf.com.tw/tw/article/show.aspx?num=6188&page=2&kind=2>

楊鈞皓，〈明星效應對職棒比賽觀看意願之影響〉，逢甲大學，2014 年 3 月 6 日

劉振家、黃德舜、陳育成，〈職業運動賽事之觀眾人數及其影響因素分析－美國、日本、韓國、台灣之職業棒球聯盟比較〉，休閒世界研究第八卷第一期，2000 年 3 月，頁 126~142

邱詠仲，〈機器學習方法運用於棒球教練決策之研究：以投打有利對決與打者價值預測為例〉，國立宜蘭大學，2022 年 7 月

Mike Petriello, "Everything to know about Statcast's new bat-tracking data", May 13th, 2024, <https://www.mlb.com/news/what-you-need-to-know-about-statcast-bat-tracking>

柒、附錄

程式碼，使用 R 語言：

```
baseball = read.csv("Data set.csv")

## 過濾掉和局 ##

baseball <- baseball %>% filter(Win %in% c(0, 1))

## 表現做標準化處理(因有些是機率有些是加總) ##

baseball$OPS<-scale(baseball$OPS)

baseball$BA<-scale(baseball$BA)

baseball$RBI<-scale(baseball$RBI)

baseball$R<-scale(baseball$R)

baseball$ERA<-scale(baseball$ERA)

baseball$K<-scale(baseball$K)

baseball$WHIP<-scale(baseball$WHIP)

## Logistic 回歸分析 ##

baseball_glm= glm(Win~OPS+RBI+R+ERA+WHIP+Home+Star+Foreign+ball,
family=binomial , data=baseball)

vif(baseball_glm) ## 檢查共線性(BA 跟 OPS 有很大的關連性所以我先去掉 BA)
##

summary(baseball_glm)

## 雙邊設計，解決重複根據 ##

baseball$GameID <- paste0(baseball$Date, "_",
                           pmin(baseball$Team, baseball$Opponent), "_",
                           pmax(baseball$Team, baseball$Opponent))

coeftest(baseball_glm, vcov = vcovCL(baseball_glm, cluster = ~ GameID))
```

```
## 隨機森林，查看變數重要性 ##
```

```
baseball$Win = as.factor(baseball$Win)
```

```
rf_model <- randomForest(Win ~ OPS + RBI + R + ERA + WHIP + Star + Foreign +  
Home + ball,
```

```
data = baseball, ntree = 500, importance = TRUE)
```

```
importance_rf_model <- importance(rf_model)
```

```
imp_df <- as.data.frame(importance_rf_model)
```

```
imp_df$Variable <- rownames(imp_df)
```

```
imp_df <- imp_df %>%
```

```
  arrange(desc(MeanDecreaseAccuracy))
```

```
# 畫圖—變數重要性 ##
```

```
ggplot(imp_df, aes(x = reorder(Variable, MeanDecreaseAccuracy), y =  
MeanDecreaseAccuracy)) +
```

```
  geom_col(fill = "darkgrey") +
```

```
  coord_flip() +
```

```
  labs(
```

```
    title = "Mean Decrease in Accuracy",
```

```
    x = "Variable ",
```

```
    y = "Accuracy "
```

```
  )
```