

中華民國第 58 屆中小學科學展覽會 作品說明書

高級中等學校組 行為與社會科學科

第二名

052707

BBQ – BaseBall Questions

學校名稱：臺北市立和平高級中學

作者： 高二 朱紫瑄 高二 吳冠逸	指導老師： 黃俊瑋
---------------------------------	------------------

關鍵詞：棒球、統計、進階數據

摘要

本研究中，首先探討美國棒球統計學相關文獻，接著建立中職所有投手的數據庫，以國外的研究理論為基礎，並考量台灣棒球環境與國外的差異，排除守備影響，篩選出能真實反應投手表現的變數。我們先根據職棒 28 年來的數據庫，利用多變量迴歸分析，創造出中職版的投手數據 FIP_A ，並加入新變數「飛球/滾地」，創造出進階版的 FIP'_A 。接著考慮環境的變革，以 2009 年至今穩定維持四隊的數據庫，創造出 FIP_N 與 FIP'_N 。另外，也迴歸創造出不受環境影響的進階公式 $SIERA_A$ 及 $SIERA_N$ 。研究發現，在中職環境裡投手的三振對失分的解釋力低，而全壘打則是影響失分最重要的變數，投手選才應首重被全壘打率低與保送率低的投手。我們也應用這些公式，分析比較中職的投手，提供台灣棒球界選才參考。

壹、研究動機

棒球素有「國球」之美譽，過去在各層級的國際賽場上屢創佳績，並在國際棒壇上佔有一席之地，於各項世界級大賽皆有不錯的成績。棒球，是讓世界看見台灣的重要舞台。不僅於此，棒球不只是一項運動，更是產值驚人的運動產業！美國職棒大聯盟的年營收已突破 3000 億台幣，球隊的總市值更超過 1 兆 4000 億台幣！在如此龐大的經濟產業下，簽一位球員是關乎數億台幣的生意，因此如何正確的評價球員，成為了管理階層的重要課題，而棒球統計學，正是達成此目的的重要工具。

通常在棒球比賽中，評論一個球員時，大家較常看的都是打擊率、自責分率等傳統數據，然而，這些數據的高低往往包含了許多運氣成分。例如研究者 Voros McCracken 在 2001 年提出了新的論述，他認為「球打進場中是否形成安打」有很大的部分和運氣有關！因此，如何改善數據，衡量球員的真正能力與價值，成為了我們所感興趣的課題。

美國的「賽伯社群」（Sabermetrics）是全世界最大的棒球數據研究團體，其中有許多與此相關的研究。例如研究者 Tango 在 2007 年提出了加權上壘率（ $wOBA$ ），可以精準地描述打擊者的表現，他也提出與守備無關的投手數據（ FIP ，*Fielding Independent Pitching*），可以不受守備員的影響，真實地反應出投手的表現，這也是目前美國職棒大聯盟的新主流數據。但是台灣棒球界卻相對缺乏此類進階數據的開發與研究，同時 Tango 所提出的公式，其係數是根據美國職棒大聯盟的歷年數據進行統計迴歸分析得來，若直接套用到中華職棒的球員數據時，往往會因為環境的不同而產生誤差或偏差。

有鑑於近年台灣棒球代表隊在國際賽成績退步，我們好奇是否在選才上出現問題？又像過去王建民、陳偉殷等在大聯盟發光發熱的投手，一夫當關左右比賽勝負。因此我們想要了解中職環境中評價投手最好的數據(變數、公式)是什麼？誰才是中職的最好投手？於是，本

研究中我們利用中華職棒的數據，並以多變量迴歸建造出屬於中華職棒的 *FIP*、*FIP'* 及 *SIERA* 等投手進階數據公式，用以更精確地評估中華職棒球員的表現與能力！

貳、研究目的

本研究主要目的如下：

- 一、聚焦在三振、保送、全壘打、飛球滾地球比率等與守備無關的數據，利用多變量迴歸分析，創造出適用於中華職棒的 *FIP*、*FIP'* 與 *SIERA* 公式，作為分析、比較職棒投手表現的工具。
- 二、考慮中華職棒環境的變革，以近年來穩定維持四隊後的數據，創造出新版的 *FIP*、*FIP'* 與 *SIERA* 公式。
- 三、分析三振、保送、全壘打等變數與投手責任失分的關係，了解中職環境中影響投手失分最重要的變數。
- 四、利用上述投手進階數據，實際分析比較中職投手的表現，提供台灣棒球界選才參考。

參、研究設備及器材

- 一、電腦（利用 EXCEL 軟體進行迴歸分析、WORD 軟體進行文書編輯）
- 二、我們所整理創建的中華職棒數據庫：1. 歷年投手數據庫、2. 投手的生涯數據庫、3. 各年度投手數據庫。

肆、研究過程或方法

一、研究流程與研究方法

首先，由於中華職棒官方網站並未建立整合歷年數據的資料庫，只有 28 年之間每一年的原始數據與基本數據。因此，我們先透過官網的全紀錄查詢，逐年整理所有投手（職棒元年~職棒 28 年）的各項數據做成 EXCEL 檔，建立資料庫以利後續的分析與研究。

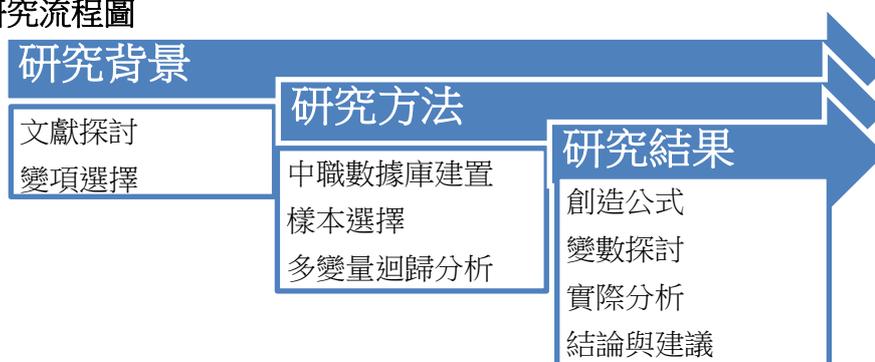
本研究的主要流程如下：

1. 依據文獻探討的結果，探討重要的投手進階數據，以及與這些進階數據相關的變數。
2. 建立中華職棒投手歷史數據庫與各年度數據庫，並從中選出適合本研究的樣本。
3. 考慮投手中與守備、與運氣無關的投手數據「三振」、「保送」與「被全壘打」，並以「自責失分」作為依變項，利用多變量迴歸分析，創造出中職版本的 *FIP_A* 公式。
4. 根據文獻探討，發現「飛球/滾地」比也是反應投手特色且屬於投手可控制的數據。故加入此變數，並利用多變量迴歸創造出進階版的中職 *FIP'_A* 公式。

5. 考慮中職成立至今 28 年間球隊數量與環境有過諸多變革，因此，本研究中再建立 2009 年至今穩定保持四隊的數據庫，利用多變量迴歸分析得到新版本的 FIP_N 與 FIP_N' 公式。
6. 利用中職資料庫現有的基本數據，先造出進階數據 $SIERA$ 所需要的變項，再利用多變量迴歸分析創造出中職版的 $SIERA_A$ 及 $SIERA_N$ 公式。
7. 討論這些公式的意義，並探討、比較公式中各變數對投手責任失分的影響與解釋程度，找出最重要的變數，提供選手參考。
8. 利用本研究所創造的投手進階數據公式，實際進行中華職棒投手的分析與比較。包含個別球員的分析，以及利用投手生涯數據進行中職所有投手與現役投手的排名。

圖 1

本研究的研究流程圖



資料來源：研究者整理繪製

二、文獻探討與前置研究

(一) 本研究中相關棒球術語介紹

1. 自責分 (ER)：投手失分中本身應負責的部分。也就是投手失分 (R) 扣除因隊友守備失誤、妨礙跑壘、捕手捕逸等所造成的失分。
2. 打擊率 (AVG)：評量打者成績的傳統數據，計算方式為 $\frac{\text{安打數}}{\text{打數}}$ 。
3. 加權上壘率 ($wOBA$ ，*Weighted On Base Average*)：利用打擊數據迴歸分析計算出相關打擊結果的係數，給予各種安打不同的權重，有如進階版的上壘率。
4. 妨礙跑壘：防守隊的球員在持球或未處理球之情形下，妨礙跑壘者跑壘的行為。
5. 捕逸：捕手未能確實接捕投手所投出的球。
6. 球打進場內的安打率 ($BABIP$ ，*Batting Average on Balls In Play*)：計算公式為 $BABIP = \frac{\text{安打數}-\text{全壘打數}}{\text{打數}-\text{三振數}-\text{全壘打數}+\text{高飛犧牲打數}}$ ，對於打者與投手兩者皆適用。此數據年與年之

間的相關性非常低，顯現出它不是選手能完全控制。長期下來，多數選手的 *BABIP* 都在 0.29~0.3 左右（也就是球被打進場中後，約有 30% 的球會形成安打）。

7. HR：投手的被全壘打數。

8. HBP：投手的觸身球次數，投出觸身球將會保送打者上一壘。

9. BB：投手四壞球數，投出四壞球將會保送打者上一壘。

10. IBB：投手的故意四壞球數，投手因教練指示或戰術執行投出四壞球數。

11. IP：投手的投球局數。

12. BF：投手所面對的打者次數。

13. 三振率（SO%）：投手面對打者的平均三振率，計算方式為 $\frac{\text{三振數}}{\text{投球人次}} = \frac{\text{SO}}{\text{BF}}$ 。

14. 四壞球率（BB%）：投手面對打者的平均四壞球率，計算方式為 $\frac{\text{四壞球數}}{\text{投球人次}} = \frac{\text{BB}}{\text{BF}}$ 。

（二）傳統數據的介紹

1. 自責分率（*ERA*）

一般大眾聽到「投手數據」時，很直覺地就會想到自責分率（又稱防禦率），它的公式是 $\frac{\text{自責分} \times 9}{\text{投球局數}} = \frac{\text{ER} \times 9}{\text{IP}}$ 。由公式所選取的相關變數可以看出，它的意義是一個投手平均投 9 局的責任失分數。

然而，自責分率太過粗糙，無法實際顯現出一個投手的真實能力。依據這個數據，只能知道一個投手每 9 局的自責失分，但失分這件事有時並不是投手自身所能控制的。舉一個極端的例子，假設有一個投手在自己原本的球隊中投的很「好」，自責分率只有 1.00，但如果將同一位投手換到一個守備很差，守備員都是沒有棒球背景的路人，那他的自責失分率還會這麼「好」嗎？答案一定是否定的。

再舉另一個極端例子，一樣剛剛那個「好」投手，如果在他每次壘上有人的時候就把他換下場，換一個沒有棒球背景的路人當後援投手，這時失分便會記在這個投手的身上，他的自責失分率一定也會跟著變差，所以自責分率也會受到教練換人時機與隊友的影響，由於自責分率看不出這些因素造成的影響，我們才需要更多延伸的投手進階數據。

2. 每局被上壘率（*WHIP*）

每局被上壘率（*WHIP*）的公式是 $\frac{\text{保送數} + \text{安打數}}{\text{投球局數}}$ ，它是用來衡量一個投手每局被打者

上壘的次數，每局被上壘率（*WHIP*）也和自責分率（*ERA*）相似，無法實際了解投手的能力。因為公式中的安打變項也屬於投手自己較無法控制的數據（參見「（三）傳統數據的缺點」中的說明）。因此，大家常用的每局被上壘率（*WHIP*）也非完美的數據，用它來評價投手時，仍需要搭配更多的數據才能完整呈現出投手的表現。

（三）傳統數據的缺點

在（二）傳統數據的介紹中，我們介紹了兩種常見的傳統數據，以及各自的不足之處。這小節裡我們將會論述傳統數據的普遍缺點以及進階數據的精神。

一般而言投手面對打者時，會有「三振」、「保送」、「被全壘打」與「球被打進場中」四種結果（其中前三種情況與隊友的守備無關）。而 Voros McCracken 的研究發現，在職棒這種高層級的比賽環境中，投手的「球被打進場中後是否形成安打」並不是投手本身能完全控制的。無論投手實力如何，大部分都和「運氣」，也就是與球的落點有關，也受守備員能力影響。棒球的防守者通常固定站在某區域，只接得到落到鄰近區域的球，一旦球落到沒人或較偏的區域就會變成安打，換句話說，被安打是「投手+守備+運氣」的產物，所以投手的失分與被上壘也不完全是投手可控制的因素！

表 1

投手數據的年與年相關分析（2002~2012）

變數	年與年間相關係數	變數	年與年間相關係數
GB/FB	0.871	FIP	0.584
GB%	0.839	WHIP	0.430
FB%	0.817	HBP/TBF	0.404
SO/TBF	0.803	HR/TBF	0.390
SO/9	0.803	HR/9	0.390
uBB/TBF	0.711	ERA	0.373
xFIP	0.699	IBB/TBF	0.358
BB/9	0.692	HR/(TBF-BB-HBP-SO)	0.349
K/BB	0.630	BABIP	0.235
HR/FB	-0.029		

Note. From “Basic Pitching Metric Correlation 1955-2012, 2002-2012,” by Matt Klaassen, 2013

Matt Klaassen (2013) 曾利用數十年的大數據進行各項數據年與年之間的相關分析，計算出各項投手數據的年與年之間的相關係數，如表 1 所示是 2002~2012 年的結果，¹其中左欄為各項數據，右欄代表其年與年間的相關係數。從表中可發現：投手的「球被打進場內後形成安打的比率」(BABIP) 的年與年相關係數僅有 0.235，相對偏低，表示同一投手不同年間的「球打進場內形成安打的比率」(BABIP) 變異大。

然而理論上，一位選手的能力與表現，固然會隨年紀而有影響，但除了受傷等極端因素外，同一位選手年與年之間的表現通常不該有巨幅的改變，因此若整個聯盟的投手群的某一數據，在年與年之間的變異過大，我們可以推測可能是這個數據本身存在缺陷，而 BABIP (球打進場內形成安打的比率) 便是如此。由此也可說明，「球被打進場中是否形成安打」並非投手自身能完全控制！

另外，從表中我們也可看出，凡涉及三振、保送與被全壘打的棒球數據，其年與年相關性都比傳統的自責分率 (ERA) 與每局被上壘率 (WHIP) 來得高，特別是三振數 (SO) 與保送數 (BB) 相關數據最高，也就是說，三振數 (SO)、保送數 (BB) 與被全壘打數 (HR) 等與守備無關的數據，較能反應出投手的真正實力，而並非運氣與隊友因素，這些才是一個投手真正能控制的數據。所以，我們在評價投手的表現時，必須把「球打進場中 (即與隊友守備有關)」的情況排除，這就是投手進階數據的重要精神，而我們在本研究中所感興趣的 FIP 與 SIERA 便是與守備無關的投手數據！

(四) 美國大聯盟重要投手進階數據

1. FIP (Fielding Independent Pitching)

FIP 是研究者 Tom Tango 發明的投手進階數據，該公式的變數有四壞球數 (BB)、三振數 (SO)、被全壘打數 (HR)。這項數據的發明是為了排除「投手球被打進場中後形成安打」這項運氣為導向的變數，而且它只考慮被全壘打數、四壞球數與三振數這三個數據，扣除了隊友守備的影響，就如本研究 (三) 傳統數據的缺點中的論述，這三個數據都是投手自己相對能掌控的，不會受到隊友守備影響。同時它也排除了「差勁的後援投手讓遺留在壘包上的跑者得分，導致失分計算在原本的先發投手身上」這件並非投手本身責任的事，也不會有自責分該計算在誰身上的問題，這些都是投手自責失分率所無法解釋的。

2. xFIP (Expected Fielding Independent Pitching)

上述的 FIP 雖然比傳統數據有更好的解釋性，但卻無法反應出各個球場差異造成的影

¹ 部份數據在 2002 年前官網並未統計，因此本研究摘錄 2002~2012 年的表格較完整。

響，特別是有些球場較小或位在山上，較容易形成全壘打，所以研究者 Dave Studeman 發明了這項新的數據。而 *xFIP* 的計算公式基本上與 *FIP* 一樣，差別是 *xFIP* 在全壘打的部分做了校正，把原本 *FIP* 的全壘打改為投手的飛球乘以聯盟平均的 $\frac{\text{全壘打數}}{\text{飛球數}}$ ，如此就能解決球場差異。

3. *SIERA* (*Skill-Interactive ERA*)

SIERA (*Skill-Interactive ERA*) 是研究者 Eric Seidman 和 Matt Swartz 的發明。此公式的變數與其係數如表 2，最左欄是各列標題，代表 *SIERA* 的變數種類，各列中的數字即為該變數的係數（第二和三列分別是美國知名棒球統計社群 Baseball Prospectus 與 Fangraphs 所計算出的版本，兩者的變數一樣，但權重不同）。

表 2

SIERA 的變數與其係數列表

變數	Baseball Prospectus	Fangraphs
(SO/PA)	-16.986	-15.518
(SO/PA)^2	7.653	9.146
(BB/PA)	11.434	8.648
(BB/PA)^2	–	27.252
(netGB/PA)	-1.858	-2.298
+/- (netGB/PA)^2	-6.664	-4.920
(SO/PA)*(BB/PA)	–	-4.036
(SO/PA)*(netGB/PA)	10.130	5.155
(BB/PA)*(netGB/PA)	-5.195	4.546
Constant	6.145	5.534
Year coefficients (versus 2010)	–	From -0.020 to +0.289
% innings as SP	–	0.367

Note. From “New SIERA, Part Two (of Five): Unlocking Underrated Pitching Skills,” by Matt Swartz, 2011

為了更詳細的解釋投手的能力，Eric Seidman 和 Matt Swartz 將原本的 *FIP* 稍作改良加入與飛球 (FB) 和滾地球 (GB) 比率有關變數，進行更複雜的多變量迴歸分析，從年與年相關分析可知，投手被打出的滾地球與高飛球的比率是最穩定的數據之一（可參照（三）傳統

數據的缺點中的表 1)，再加上滾地球不容易造成重傷害，而飛球則容易變成全壘打或長打，所以傾向被打出較多飛球的投手他的成績容易較差，這都會影響投手的數據，所以才將這類變數加到公式中。

（五）數據與環境之關聯

1. 美國 MLB 數據對於我國職棒環境之適用性

由於 *FIP*、*SIERA* 等進階數據都是由美國的研究者所提出，公式中採用的係數都是利用美國職棒大聯盟歷年來的大數據迴歸所得，直接套用到中華職棒投手的分析難免會有些許誤差或偏差。而這是因為兩國的職棒環境與選手素質大不相同所致。有鑑於此，我們便著手建立中華職棒 28 年的數據庫，並利用中職歷年的數據作迴歸分析，創造出適用於台灣環境的公式係數，用以更準確的衡量中華職棒投手的表現。

2. 中職環境變遷對於數據的影響

中華職棒成立於 1989 年，然而至今 28 年間球隊數量有過很大的改變。現在的中華職棒是於 2003 年由「中華職棒大聯盟」與「台灣大聯盟」兩職業聯盟合併而改制而來。在 2003 兩聯盟合併成六隊，最終在 2009 年形成現在中職隊伍長期維持四隊的穩定環境。由於經過了如此劇烈的環境改變，我們決定除了利用職棒元年到今的數據進行迴歸分析、創建公式外，研究中也特別將 2009 年後的數據獨立出來，同樣利用迴歸分析得到另一組公式，適用於近 10 年來穩定環境中的投手分析與比較。

三、建立數據資料庫與樣本的選取

（一）建立歷年投手資料庫

由於中華職棒並未建立整合歷年數據的資料庫，因此我們先整理中華職棒 28 年來的投手數據，創建自己的資料庫，進行數據的管理與使用。創建資料庫的過程如下：

首先，在資料庫及迴歸的取樣方面，我們選擇了投球局數 40 局以上的投手作為本研究中的樣本。這是由於我們迴歸分析時，採用的是每位選手單年度平均每九局的數據，然而當投球局數太少時，極端值對於數據的算術平均數影響會增大。一位投球局數較少的選手，一旦在某次比賽中受到運氣影響，投出了極佳或極差的成績，便會大幅影響他每九局的平均成績，較不具代表性。因此，我們設定出賽局數的門檻，將極端值的影響降低，同時，以 40 局為分界點能讓我們有 1100 個以上的充足數據量進行迴歸分析（總數有 1172 筆），又能減少因極端值產生的影響。

接著我們採用中職官方網站所公布的數據，過濾出該年度出賽 40 局以上的投手，複製貼上到 EXCEL，並將之編號，編號方式如表 3。前 2 碼訂為選手數據的年份代碼，並將中職元年（1990 年）定為 01，後兩碼則是依照該年度投球局數作排序編碼（表 3 中的 IP 即為投球局數）。例如：中職元年出賽局數最多的投手編號定為 0101，出賽局數第 2 多的投手編號定為 0102，以此類推。²

表 3

編碼方式示意表

	NAME	IP	H	HR	BB	SO	ER	BB/IP	HR/IP	SO/IP	ER/IP	G/F	BB9	HR9	SO9	BRA
0101	*黃平洋	246	198	14	48	177	61	0.195	0.057	0.720	0.248	1.461	1.756	0.512	6.476	2.232
0102	*史東J.S	216	173	9	87	110	46	0.403	0.042	0.509	0.213	1.908	3.625	0.375	4.583	1.917
0103	T.M	195.1	174	7	38	105	45	0.195	0.036	0.538	0.231	1.263	1.753	0.323	4.844	2.076
0104	*瑞奇E.B	174	153	4	92	177	50	0.529	0.023	1.017	0.287	1.407	4.759	0.207	9.155	2.586
0105	*賈西L.G	160.2	142	5	80	83	47	0.499	0.031	0.518	0.293	1.528	4.494	0.281	4.663	2.640
0106	*派瑞V.P	155.2	152	8	23	52	53	0.148	0.052	0.335	0.341	1.532	1.334	0.464	3.015	3.073
0107	*涂鴻欽	149	124	3	51	110	41	0.342	0.020	0.738	0.275	1.126	3.081	0.181	6.644	2.477
0108	*康明杉	143	112	6	65	87	36	0.455	0.042	0.608	0.252	1.156	4.091	0.378	5.476	2.266
0109	*杜福明	143	154	7	24	48	50	0.168	0.049	0.336	0.350	1.160	1.510	0.441	3.021	3.147
0110	*陳逸松	129.1	123	9	35	71	51	0.271	0.070	0.550	0.395	1.007	2.440	0.627	4.950	3.555
0111	*黃廣琪	124	131	9	48	57	59	0.387	0.073	0.460	0.476	0.748	3.484	0.653	4.137	4.282
0112	*黃武雄	117	101	5	41	56	26	0.350	0.043	0.479	0.222	2.012	3.154	0.385	4.308	2.000
0113	*張永昌	105.2	123	6	19	57	55	0.181	0.057	0.542	0.523	1.142	1.625	0.513	4.876	4.705
0114	*翁豐境	105.1	133	8	49	50	62	0.466	0.076	0.476	0.590	1.144	4.196	0.685	4.282	5.309
0115	*劉義傳	103	106	5	35	45	44	0.340	0.049	0.437	0.427	1.366	3.058	0.437	3.932	3.845
0116	*李文傳	88.1	94	9	13	12	31	0.148	0.102	0.136	0.352	1.313	1.328	0.919	1.226	3.167
0117	*剛烈J.G	84.1	87	4	50	30	32	0.595	0.048	0.357	0.380	1.724	5.351	0.428	3.210	3.424
0118	*將生D.J	81.2	90	4	45	48	38	0.554	0.049	0.591	0.468	1.621	4.988	0.443	5.320	4.212
0119	*林光宏	80	97	5	27	21	41	0.338	0.063	0.263	0.513	1.000	3.038	0.563	2.363	4.613
0120	*李偉E.R	72.1	88	5	28	26	29	0.388	0.069	0.361	0.402	2.000	3.495	0.624	3.245	3.620

資料來源：由研究者整理繪製

將所有樣本編號之後，由於中職資料庫的數據較陽春，我們需要利用 EXCL 設定程式自己算出中職資料庫所沒有的數據，例如 SO%（SO/BF）、BB%（BB/BF）等，以利將來分析選手時使用。

(二)建立投手生涯數據資料庫

在建立完中職 28 年來的基本數據庫後，我們再利用中職官方網站中的年度累計數據，建立投手生涯數據的資料庫，以利後續分析比較選手的生涯表現。倘若某一選手生涯出賽局數過少，那數據也極容易受運氣及其他因素影響，不足以解釋此選手的的能力與特質，因此，我們以生涯出賽總局數 120 局為標準。接著，我們將生涯出賽總局數 120 局以上的選手數據整理成資料庫，並同樣依投球局數作排序編號，如表 4。

² 資料庫中*指的是退役球員。

表 4

投手生涯資料庫示意表

	NAME	G	GS	GR	IP	WHIP	ERA	BF	H	HR	BB	IBB	HBP	SO	ER
001	潘威倫	307	288	19	1812.1	1.220	3.170	7579	1938	104	265	9	69	1052	638
002	* 勇壯	339	211	128	1570.0	1.250	3.200	6642	1458	110	501	18	63	1286	559
003	* 謝長亨	248	216	32	1491.2	1.200	3.020	6205	1427	77	365	19	27	855	501
004	* 高建三	636	138	498	1383.2	1.350	3.460	5977	1453	89	414	16	62	861	532
005	陽建福	312	199	113	1297.1	1.410	3.990	5671	1389	64	434	13	82	834	575
006	* 陳義信	187	156	31	1275.2	1.120	2.820	5189	1197	109	228	25	36	748	400
007	* 林朝煌	418	103	315	1111.1	1.390	4.110	4901	1164	92	379	22	81	721	508
008	* 翁豐堉	326	151	175	1108.1	1.400	4.120	4843	1180	94	368	12	40	693	507
009	* 風神	181	147	34	1085.0	1.090	2.550	4388	934	47	246	11	53	779	308
010	* 劉義傳	468	33	435	1057.1	1.200	3.120	4447	1061	67	210	45	30	651	367
011	* 林英傑	245	162	83	1055.2	1.230	3.310	4430	1063	54	235	10	36	792	388
012	* 康明杉	200	135	65	1037.1	1.320	3.800	4449	1013	90	354	18	40	522	438
013	* 郭進興	212	117	95	1020.0	1.180	2.930	4202	998	86	205	13	21	580	332
014	* 黃平洋	150	103	47	927.1	1.120	2.540	3786	821	59	214	13	20	593	262
015	* 曾孟承	335	105	230	914.0	1.320	3.760	3906	994	84	211	12	25	440	382
016	羅力	137	133	4	867.2	1.200	3.220	3599	864	64	175	3	21	711	310
017	* 沈鈺傑	526	39	487	817.2	1.390	3.910	3584	923	50	212	26	36	606	355
018	* 吳俊億	304	69	235	814.0	1.410	3.940	3590	862	57	284	13	83	365	356
019	* 陳憲章	173	113	60	800.2	1.330	3.710	3389	915	50	150	15	21	321	330
020	* 柏榕	126	109	17	786.0	1.170	2.840	3263	764	49	154	7	30	541	248

資料來源：由研究者整理繪製

(三)建立各年度投手數據資料庫

除了上述兩個以選手個人成績為基礎的資料庫外，我們又依年度整理建立了中華職棒每一年的投手數據資料庫，目的是為了計算出各年度的平均數據，了解中華職棒投手成績的變化逐年趨勢。

表 5

選手各年資料庫示意表（以 2009 年為例）

	NAME	G	GS	GR	IP	WHIP	ERA	BF	H	HR	BB	IBB	HBP	SO	ER
001	雷克斯A.	26	26	0	159	1.31	3.45	691	168	11	40	0	5	110	61
002	正田樹	27	25	2	158	1.44	4.44	687	191	10	36	0	2	115	78
003	林克謙	28	24	4	145	1.39	4.02	624	160	4	42	3	7	67	65
004	張誌豪	26	19	7	135	1.44	4.31	587	145	10	51	0	3	94	65
005	潘威倫	22	22	0	133	1.30	3.30	576	159	6	15	3	5	81	49
006	廖于誠	22	22	0	130	1.50	3.31	579	124	6	72	1	10	64	48
007	小林亮寬	23	23	0	127	1.61	6.06	589	160	15	45	0	5	78	86
008	許文雄	25	24	1	126	1.64	5.20	571	158	7	49	0	3	77	73
009	林英傑	27	20	7	115	1.34	4.43	495	132	11	23	1	8	93	57
010	麥克M.S	19	19	0	114	1.50	5.45	512	133	15	38	2	4	72	69
011	余文彬	40	11	29	104	1.36	4.33	446	115	5	26	6	3	50	50
012	潘俊榮	28	16	12	100	1.63	5.02	471	127	8	37	0	8	44	56
013	沈鈺傑	65	1	64	100	1.22	3.42	417	101	7	21	4	2	77	38
014	曹錦輝	19	19	0	93	1.37	3.94	406	104	6	24	0	3	64	41
015	陽建福	24	17	7	89	1.84	5.76	422	116	5	48	2	2	40	57
016	朱尉銘	19	14	5	87	1.59	6.18	397	116	12	23	0	4	58	60
017	柳裕展	24	15	9	80	1.87	6.47	386	110	10	41	0	3	61	58
018	林正豐	50	3	47	75	1.52	6.07	347	91	10	24	4	3	44	51
019	克里斯C.	14	14	0	74	1.41	4.48	328	85	4	20	1	1	61	37
020	李濠任	38	5	33	74	1.59	5.35	338	81	3	37	0	5	40	44

資料來源：由研究者整理繪製

四、研究中變數的選擇與調整

文獻探討後，我們參考美國版進階數據公式所採用的變數，並以此為基礎再深入思考如何改良，使之更完善。

(一) 四壞球數 (BB) 中是否扣除故意四壞球數 (IBB)

FIP 與 *SIERA* 中的變數皆有四壞球數 (BB)，但投手的四壞球數裡包含了故意四壞球數，而故意四壞的戰術是投手接收經由教練團所發出的指示，相同的情況下有的教練會選擇用戰術、有的則不會，**完全依照教練的決定**，並非投手能控制！考慮到進階數據公式的本意是衡量投手的真正價值，應採用投手自身所能控制的數據來做計算。因此我們選擇將四壞球數 (BB) 扣除故意四壞球數 (IBB)。

(二) 迴歸時的 Y 變數採用自責分 (ER) 或失分 (R)

美國研究者將迴歸時的應變數 Y 定為失分。然而，相較於失分，自責分的計算會扣除守備失誤、捕手漏接等非投手能控制的運氣成分，讓我們知道**在運氣相等的狀況下各投手的實際失分**。因此我們認為，與採用失分相比，採用自責分較能顯現投手的實力，與本研究的目的較相符，在思量後選擇採用自責分作為應變數 Y 來進行迴歸分析。

(三) 進階版的 *FIP'* 納入「滾飛比」

Tom Tango 所發明的 *FIP* 公式包含的變數有保送 (BB)、三振 (SO)、被全壘打 (HR)，我們的研究中除了參考國外的研究，迴歸創造出中職版的 *FIP* 公式外，另外也加入滾飛球比 (G/F，投手被擊出滾地球數與飛球數的比率)，算出進階版的 *FIP'*。選取理由如下：

依滾飛球比劃分投手，大致能區分出滾地球型投手與飛球型投手兩類。滾地球型投手的定義為滾飛比大於 1 (即投手被擊出的滾地球數大於飛球數) 的投手；飛球型投手的定義則為滾飛比小於 1 (即投手被擊出的飛球數大於滾地球數) 的投手。

然而，滾地球型投手與飛球型投手何者較為優秀一直存在爭議，兩方的支持者都有各自支持的理由。表 6 為 Neil Weinberg 的研究結果，表中列出投手被打出滾地球與高飛球的打擊率以及純長打率，其中滾地球的被安打率為 0.239，高於高飛球的 0.207，乍看之下或許飛球型投手略勝一籌，不過考量表 6 的純長打率數據，不難想像飛球較滾地球容易形成二壘、三壘或全壘打等長打，導致投手大量失分。

表 6

滾地球、飛球的被安打率、純長打率

球種 (Type)	被安打 (AVG)	純長打率 (ISO)
滾地球 (GB)	0.239	0.020
飛球 (FB)	0.207	0.378

Note. From “Which is Better? A Ground Ball Pitcher or a Fly Ball Pitcher,” by Neil Weinberg, 2014

一如知棒球論壇 Fangraphs 的分析師 Neil Weinberg 所言 “One type isn’t better than the other, they’re just different.” (Weinberg, 2014) 即兩種型態的投手皆有各自的優勢與劣勢，其優勢與劣勢是影響失分的因素之一，也代表投手各自的能力，且 G/F 的年與年相關係數高達 0.871 (參照 (三) 傳統數據的缺點中表 1) 屬於投手穩定的特性。因此我們在研究中將 G/F (滾飛比) 這項變數加入 FIP’ 公式。然而，因為中職數據庫缺乏滾地球與飛球總數，無法計算聯盟平均滾飛比，我們在計算聯盟平均時以「滾地球出局數/飛球出局數」(GO/FO) 代替投手被擊出滾地球數與飛球數的比率 (G/F)。

(四) SIERA 公式相關變數的取代

美國版本的投手數據 SIERA 包含變數 netGB，該變數指的是「投手被擊出滾地球數 (GB) 減投手被擊出飛球數 (FB)」(即 GB-FB)，但是中華職棒官方網站的資料庫缺少這兩項數據，於是我們用「滾地球出局數 (GO) 減飛球出局數 (AO)」來代替，並以現有的數據為基礎，計算出 SIERA 公式所需要的各項變數。

伍、研究結果

一、創建中職版 FIP 與進階版 FIP’ 公式

我們創建資料庫之後，先算出每位選手單年度的平均每 9 局數據，包含「每 9 局三振數 (SO9)」、³「每 9 局保送數 (BB9)」與「每 9 局被全壘打數 (HR9)」等。⁴接著利用 EXCEL 進行多變量迴歸分析，將 FIP 的自變數 X 的範圍設定為「BB9」、「SO9」、「HR9」；而進階版 FIP’ 的變數 X 範圍則為上述 FIP 的自變數再加上「G/F」(即包含 BB9、SO9、HR9、G/F)，應變數 Y 皆為每 9 局責任失分 (ERA)。

我們將多變量迴歸所得結果，整理化為表 7 的形式，其中我們將「用全部年份的數據

³ 棒球統計學中 SO9 常被用來表示每 9 局三振數，即 $9 \times (\text{總三振} / \text{總局數})$ 。

⁴ 以下迴歸所用到的三振、保送、被全壘打與責失分等變數，皆為每 9 局的平均數據。

庫」迴歸所得的公式，表示成 FIP_A 及 FIP'_A ；將「用合併成四隊後（2009 年後）的新數據庫」迴歸所得的公式，表示成 FIP_N 及 FIP'_N 。並在表格的最下列，將大聯盟版本的公式列出比較：

表 7

FIP 及 FIP' 公式迴歸結果列表

適用年度	變數有無滾飛比	公式	R^2
全部年度的 中職環境	變數無滾飛比	FIP_A $= \frac{3.31 * BB + 16.37 * HR - 1.25 * SO}{IP} + 2.31$	0.515
	變數加入滾飛比	FIP'_A $= \frac{3.31 * BB + 16.15 * HR - 1.27 * SO}{IP} - 0.07(G/F) + 2.43$	0.516
合併四隊後的 (2009 年後) 中職環境	變數無滾飛比	FIP_N $= \frac{3.49 * BB + 15.74 * HR - 1.61 * SO}{IP} + 2.90$	0.546
	變數加入滾飛比	FIP'_N $= \frac{3.47 * BB + 16.23 * HR - 1.51 * SO}{IP} - 0.28(G/F) + 2.47$	0.553
美國大聯盟	變數無滾飛比	$FIP = \frac{13 * HR + 3 * BB - 2 * SO}{IP} + \text{常數}$	

資料來源：研究者自行繪製

此外，我們在研究中也提出一個簡化版的公式，方便快速計算、進行投手比較：

$$FIP_s = \frac{3 * BB + 17 * HR}{IP} + 1.5$$

二、創鍵中職版 $SIERA$ 公式

依據文獻探討的結果，我們利用 EXCEL 算出必要的變數後，同樣用分析工具箱裡的迴歸分析，將 $SIERA$ 的變數 X 的範圍設定為表 8 中最左欄的變數，應變數 Y 的則設定為自責分 (ERA)。

進行多變量迴歸後，所得公式結果各表 8 所示。最左欄是 $SIERA$ 所包含的變數，列中的數字即為該變數的係數。我們將「用全部年份的數據」迴歸的公式，以及「用合併成四隊後（2009 年後）的數據」所得的公式分別表示成 $SIERA_A$ 及 $SIERA_N$ ，並在表格的最右二列將美國知名棒球統計社群 Baseball Prospectus 與 Fangraphs 的公式一併列出比較：

表 8

SIERA 迴歸結果列表及美國版公式對照表

變數	$SIERA_A$ (全部年度)	$SIERA_N$ (合併成四隊後)	Baseball Prospectus	Fangraphs
(SO/PA)	-11.066	-0.141	-16.986	-15.518
(SO/PA) ²	2.411	-30.265	7.653	9.146
(BB/PA)	21.837	26.471	11.434	8.648
(BB/PA) ²	-39.067	-51.411	–	27.252
(netGB/PA)	-1.015	-4.283	-1.858	-2.298
+/(netGB/PA) ²	-4.801	-7.195	-6.664	-4.920
(SO/PA)*(BB/PA)	1.724	6.952	–	-4.036
(SO/PA)*(netGB/PA)	0.625	18.944	10.130	5.155
(BB/PA)*(netGB/PA)	-21.054	10.881	-5.195	4.546
Constant	4.131	3.313	6.145	5.534

資料來源：研究者自行繪製

上述表格是為了方便讀者比較各家公式，以下我們把這些公式用數學式表示：

1. 用全部年份的數據迴歸所得的公式為：

$$SIERA_A =$$

$$SIERA_A =$$

$$-11.066 * \frac{SO}{PA} + 2.411 * \left(\frac{SO}{PA}\right)^2 + 21.837 * \frac{BB}{PA} - 39.067 * \left(\frac{BB}{PA}\right)^2 - 1.015 * \frac{netGB}{PA} - 4.801 * \left(\frac{netGB}{PA}\right)^2$$

$$+ 1.724 * \left(\frac{SO}{PA} * \frac{BB}{PA}\right) + 0.625 * \left(\frac{SO}{PA} * \frac{netGB}{PA}\right) - 21.054 * \left(\frac{BB}{PA} * \frac{netGB}{PA}\right) + 4.131$$

2. 用合併成四隊後的數據迴歸所得的公式為：

$$SIERA_N =$$

$$-0.141 * \frac{SO}{PA} - 30.265 * \left(\frac{SO}{PA}\right)^2 + 26.471 * \frac{BB}{PA} - 51.411 * \left(\frac{BB}{PA}\right)^2 - 4.283 * \frac{netGB}{PA} - 7.195 * \left(\frac{netGB}{PA}\right)^2$$

$$+ 6.952 * \left(\frac{SO}{PA} * \frac{BB}{PA}\right) + 18.944 * \left(\frac{SO}{PA} * \frac{netGB}{PA}\right) + 10.881 * \left(\frac{BB}{PA} * \frac{netGB}{PA}\right) + 3.313$$

陸、討論

一、公式的分析與比較

(一) 環境變革對公式係數之影響

我們的 *FIP* 公式排除掉守備因素與運氣成份，利用投手三振、四壞與保送來評價投手的表現。因此，利用 *FIP* 公式計算所得數據其意義為：長期下來，若投手的運氣與隊友的守備能力是聯盟的平均水準，則預期此投手平均每 9 局的責任失分數。因此，假設某個投手在比賽中平均每投 9 局，投出 5 次保送，10 次三振，被打 1 支全壘打，則利用合併成四隊後的公式可算出他的

$$FIP_N = \frac{3.49*5+15.74*1-1.61*10}{9} + 2.90 \approx 4.80$$

也就是在中華職棒的環境中，若運氣與隊友的守備能力是聯盟的平均水準，則一位平均每 9 局「投出 5 次保送，10 次三振，被打 1 支全壘打」的投手，可以預期他平均每 9 局的失分為 4.80。

此外，從上面的數據與中職合併成四隊版的公式來看，中華職棒的投手若平均每 9 局多被打一支全壘打，將會多失去 $\frac{15.74}{9} \approx 1.75$ 分；若平均每 9 局多投 1 次保送，將會多失去 $\frac{3.49}{9} \approx 0.39$ 分；若平均每 9 局多投 1 次三振，將會少失去 $\frac{1.61}{9} \approx 0.18$ 分。

若再拿本研究中「用中華職棒所有年度的數據庫來迴歸」所得的公式與「用合併成四隊後的數據來迴歸」所得的新公式做比較，可以看出三振數 (SO) 在合併成四隊後的公式中係數變大了。此變化的意義是，在新環境下投手投三振減少失分的效果變好了。而保送與被全壘打對於失分的影響則略增加。

另一方面，由我們迴歸所得的公式可發現，中職版的 *FIP* 公式與美國職棒大聯盟版的公式 $FIP = \frac{13*HR+3*BB-2*SO}{IP} + \text{常數}$ ，明顯不同（其中大聯盟的常數通常為 3.2，小聯盟為 3.6）。因此，過去台灣棒球界或媒體利用美國版的 *FIP* 公式分析台灣選手時，並未考慮環境差異，而產生諸多的偏差與錯誤的比較。

以前面的例子來看，若運氣與隊友的守備能力是聯盟平均水準，則長期下來此投手在中華職棒的平均每 9 局失分為 4.80。

然而，若利用美國大聯盟版的公式可算出他的 $FIP = \frac{3*5-2*10+13*1}{9} + 3.2 \approx 4.09$ 。亦

即同樣表現的投手，在大聯盟的環境中平均 9 局的失分為 4.09，比中華職棒的环境少了許

多。而小聯盟版的公式算出來的 $FIP = \frac{3*5-2*10+13*1}{9} + 3.6 \approx 4.49$ 。即同樣表現的投手，

在小聯盟的環境中平均 9 局的失分為 4.49，也比中華職棒少。

再以中華職棒的投手陳鴻文當實際的例子來看，如果用美國大聯盟版的公式代入他在中華職棒 2017 年的各項數據，會計算出他的 FIP 約等於 4.85。而若以本研究做出的中華職棒版公式來計算，他的 FIP 約會等於 5.5！不難想像，大聯盟、小聯盟與中華職棒的平均水準的守備能力差異很大，若隊友的守備能力是大聯盟的平均水準，則可預期投手長期下來的平均失分會少許多。因此，若直接拿美國版的公式來分析、比較中華職棒的投手時，將會產生較大的偏差，這也是本研究的重要意義與價值所在。

（二）各項公式與變數的探討

1. Fip 變數不包含安打的原因

從表 9 中可以看出我們所創造出的 FIP 公式（公式 1）其判定係數（*Coefficient of Determination*） R^2 為 0.492，也就是公式中三項變數對於投手失分有 51.5% 的解釋力。我們也試著將這三項變數再加上傳統的安打這項變數，迴歸得到表 9 中的公式 3，其 R^2 則高達 0.810。也就是三振、四壞、全壘打、安打這四項變數的迴歸式對於投手的失分有極高的解釋力。

表 9

四個迴歸模型的比較

	公式	R^2
公式 1	$FIP_A = \frac{3.31*BB + 16.37*HR - 1.25*SO}{IP} + 2.31$	0.515
公式 2	$ER_1 = \frac{3.16*BB + 16.9*HR}{IP} + 1.51$	0.480
公式 3	$ER_3 = \frac{2.96*BB + 13.84*HR + 0.09*SO + 4.71*(H - HR)}{IP} - 2.88$	0.810
公式 4	$ER_4 = \frac{2.97*BB + 13.83*HR + 4.67*(H - HR)}{IP} - 2.79$	0.810

資料來源：研究者自行繪製

無論由常識推理或由這兩個模型的比較，都可以看出安打是影響投手失分的一項重要變數，既然安打這麼重要，為何我們的 FIP 公式中並沒有安打這項變數呢？雖然包含安打這項

變數的模型對於失分有極高的解釋力，但是正如本研究中一再強調的「球打進場中是否形成安打並不是投手能控制」，它是「投手+守備+運氣」的產物，投手每一季的被安打數都會因為運氣與守備等其他外在因素而有所起伏，因此利用包含安打這項變數的公式模型預測球員成績不符合我們研究的精神，且無法預測投手的「運氣、隊友守備為聯盟平均水準」時的實際表現，也就是說此公式對於我們評價投手的表現時，並沒有實值的意義，我們只能由此推論「安打」確實是一項影響投手失分的重要變數。

2. 三振 (SO) 的重要性

從表 9 四個公式比較我們也發現，三振 (SO) 這項變數的有無對於公式的 R^2 並沒有顯著的差異。其中，以「三振、保送、被全壘打」迴歸所得的公式 1 (即 *FIP* 公式)，其 R^2 為 0.515，但以「保送、被全壘打」所迴歸所得的公式 2，其 R^2 則為 0.480，兩個模型雖然差了三振這項變數，但是 R^2 卻只有小幅的變動，也就是加入三振這項變數並沒有提高太多模型對於失分的解釋力。而公式 3 與公式 4 的比較結果也相同，刪掉三振後的模式其 R^2 幾乎沒有改變。另外，我們也計算出投手每 9 局三振與投手每 9 局責失分的相關係數 $r = -0.209$ ，綜合來看，三振在中職環境中對於失分的解釋度偏低，投手的三振能力在中職並非影響失分最重要的因素，也因此投手的三振能力在中職環境中的重要性也許不如傳統認知，或不如在美國大聯盟環境中重要。而表 10 也可看出，三振對保送與被全壘打的相關係數也極低，因此，三振在整個模式的解釋力低，應不是受到保送或全壘打的影響。

表 10

每 9 局失分、保送、全壘打、三振相關係數表

	每 9 局責失分	每 9 局保送	每 9 局被全壘打	每 9 局三振
每 9 局責失分	1	0.388	0.607	-0.209
每 9 局保送	0.388	1	0.149	0.066
每 9 局被全壘打	0.607	0.149	1	-0.084
每 9 局三振	-0.209	0.066	-0.084	1

資料來源：研究者自行繪製

過去棒球界或許多棒球研究都很重視三振這項數據，然而我們這次的研究中反而發現，在中職的環境中投手的三振對於失分並沒有太大的解釋力，這個現象很可能是因為過去中職打者的打擊觀念，常常為了避免被三振而追求「碰」球，而不是把球打結實也比較不追求長

打，加上過去大量使用犧牲觸擊戰術，造成打者自動出局，也使得三振能力較差的投手其三振能力不足的事實被掩蓋，也不會出現因為三振能力不足而讓失分大量上升。雖然三振對於失分的解釋力不大，但是本著 *FIP* 的精神，我們創造的公式還是將三振這項變數納入其中。

3.全壘打與保送的重要性

由表 10 可知，中職投手全壘打與責失的相關係數為 0.607、保送與責失的相關係數為 0.388，都比「三振」來得高。同時，為了清楚比較保送、安打、三振、全壘打對於失分的解釋力，我們也嘗試將表 9 公式 3 的各變數先作標準化後再進行迴歸，利用各項係數來分析各變數對於失分的解釋力，我們發現安打的係數為最高的 0.589，而全壘打的係數為 0.461、保送的係數則為 0.314，但是球打進場中是否成為安打不是投手能控制的，主要受運氣與守備員影響，因此安打對於失分的高解釋度在預測投手未來表現並沒有實質意義。而全壘打與保送是安打之外最能解釋失分又是投手相對能控制的數據。因此，從數據分析我們也認為，投手選才時應首要選擇被全壘打率低與保送率低的投手，才是符合中職環境的好投手。

事實上，若基於統計學或數學的簡約精神“Simple is beautiful”，在中職環境中「三振」這項變數甚至可以被納入公式中。因此，我們也基於表 9 中的公式 2，提出簡化版的公式方便計算、進行投手比較：

$$FIP_s = \frac{3 * BB + 17 * HR}{IP} + 1.5$$

公式中僅包含「保送」與「全壘打」這兩項對失分有足夠解釋力，且是投手相對能控制的變數。只要將投手的保送與被全壘打數代入，就可簡單計算結果並比較投手。例如中職 29 年至 6 月 3 日止，投手倪福德共出賽 24 局，投出 3 次保送，被打 2 支全壘打，代入此公式可以很快地計算出

$\frac{3 * 3 + 17 * 2}{24} + 1.5 \approx 3.3$ ；而陳禹勳出賽 $24 \frac{2}{3}$ 局，投出 11 次保送，被打 1 支全

壘打，代入可以計算出 $\frac{3 * 11 + 17 * 1}{24} + 1.5 \approx 3.6$ 。

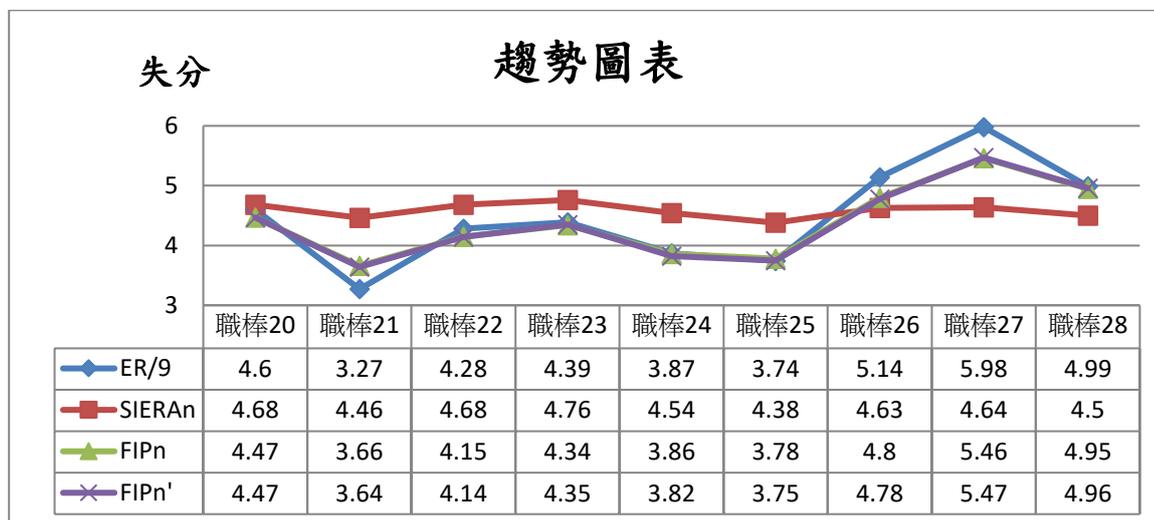
(三) 各數據聯盟平均的趨勢比較

從表 11 趨勢中我們可以看出整個聯盟的平均 $SIERA_N$ 是相對穩定的數據，而其他三項數據較一致但起伏大，特別在中職 27 年時都有暴增的情形。這是因為在這一年中職的比賽用球彈性係數較往年高出許多，造成該年的總全壘打產量高達 429 支，較前一年的 209 支全壘打高出許多。而 ERA 、 FIP 、 FIP_n 都深受全壘打的影響，因此全壘打的暴增使得三項數據跟著增加。然而，球的彈性係數增高造成投手被全壘打數增加，其增加的失分並不能完全怪罪

於投手，但 *SIERA* 避免了這樣的情形，它以三振、保送、滾地球與飛球的相關比率作為變數，取代全壘打，讓公式能更真實的表現出球投手的表現。

表 11

中職 20 年~29 年各項投手數據聯盟平均值



資料來源：研究者自行繪製

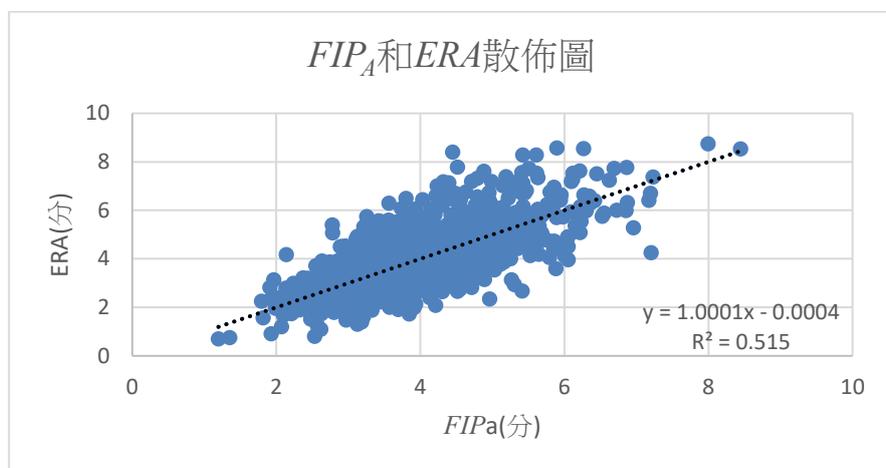


圖 2 FIP 與 ERA 的散佈圖和線性迴歸式

資料來源：研究者自行繪製

另一方面，從表 11 也可看出每一年整個聯盟的平均投手自責失分率 (ERA) 與兩個版本的 FIP 的趨勢走向一致。同時，我們算出 FIP 公式與 ERA 的相關係數為 0.712，均顯示這兩個公式所得結果具有一致性，而圖 2 中 FIP 與 ERA 的線性迴歸式的 R^2 達 0.515。加上 FIP 只需要用到三振、保送與被全壘打三項變數，對於初學者反而簡單易懂。因此，我們也主張利用本研究所得的 FIP 公式取代傳統用來評價投手的自責失分率。

二、實際應用與分析球員

(一) 個別案例

傳統數據常常會因為守備失誤、運氣等因數，造成投手帳面數據與投手實際表現不相符的情形，從進階數據的觀點切入，可以推論出不同的結論，我們在此舉出兩個例子說明：

1. 陳禹勳

中職 26 年（2015 年）國際棒球總會舉辦第一屆世界 12 強競賽，陳禹勳在進入中職的第二年即獲得中華隊總教練的青睞進入國家隊。對於陳禹勳入選國家隊，外界議論紛紛，許多人認為陳禹勳 2015 年的成績不盡理想，自責分率甚至高達 4.941，比聯盟平均的 4.74 高，不該入選國手。然而，面對負面評論，陳禹勳在 2015 年世界 12 強賽上共計出賽 5.1 局，繳出 5 次三振 0 失分的好表現，而後 2016 年與 2017 年兩年穩定的表現也讓他獲得 2017 年度救援王的獎項，徹底抹滅外界以往的疑慮。如果我們利用進階分析可發現：陳禹勳的 FIP'_N 從入聯盟四年來皆低於聯盟平均，也就是說，若他隊友的守備能力與他的運氣是聯盟平均水準的狀況下，他的失分應低於聯盟平均，由此可見他的真實表現被傳統數據「低估」了。以下是我們的分析：

陳禹勳每年面對打者的三振率（SO%）一直以來皆高於聯盟平均（如表 12），其中 2015 年他的 SO% 為 0.180（聯盟平均為 0.159），也就是每面對 100 位打者，陳禹勳能比聯盟平均實力多投出 2.1 次三振，可見他的三振能力是在聯盟平均水準之上。另外，陳禹勳的被全壘打率（HR%）從進入聯盟到現在均較聯盟平均低，2015 年他的 HR% 更是只有 0.012（聯盟平均 0.021），也就是每面對 100 位打者，陳禹勳比聯盟平均少被打將近一次的全壘打。陳禹勳於 2015 年的 FIP'_N 為 3.906（聯盟平均為 4.080），可見在扣除守備及隊友等運氣成分的情況下，也預期他的失分應比聯盟平均略少。

表 12

球員陳禹勳在各年間的數據列表（中職元年為 1990 年）

中職年份	NAME	FIP'n	ERA	SO%	BB%	SO/BB	HR%	G/F
25年	陳禹勳	2.987	3.654	0.197	0.060	2.362	0.007	1.397
26年	陳禹勳	3.906	4.941	0.180	0.079	2.000	0.012	2.036
27年	陳禹勳	3.864	3.411	0.239	0.093	2.380	0.019	1.000
28年	陳禹勳	3.693	2.646	0.269	0.104	2.414	0.017	1.157

資料來源：研究者自行繪製

根據上述進階數據的分析，我們認為陳禹勳中職 26 年(2015 年)的表現還是優於聯盟平均，只因為這年陳禹勳的自責分率高達 4.941，比聯盟平均高，大家就憑此認為他的表現不優秀。再從表 13 可看出職棒 20 年~職棒 25 年這 6 年之間的聯盟平均自責分率為 3.825，而中職 26 年（2015 年）聯盟平均為 4.741，比前面 6 年的平均高了將近 1 分，而大眾批評陳禹勳時未能察覺聯盟整體環境落差這項盲點，還停留在過去聯盟平均自責分率僅有 3.825 的標準來評斷球員，才會認為陳禹勳在這年的表現不佳。

更重要的，投手的自責分率很容易因為運氣或守備受影響，因此我們以 FIP'_N 等進階數據分析出陳禹勳 2015 年很可能只是運氣差或受到隊友守備影響，導致實際自責分率偏高，其實他真正的實力並不差，不難預期他往後的成績會漸漸回升，而陳禹勳在 2016 年與 2017 年的卓越表現便是最好的印證。

表 13

各年份的全聯盟投手平均數據列表

中職年份	ERA	SO%	BB%	SO/BB	HR%	G/F
24年	3.66	0.139	0.062	2.258	0.01	1.09
25年	3.422	0.157	0.061	2.601	0.011	1.082
26年	4.741	0.159	0.072	2.224	0.021	1.081
27年	5.647	0.174	0.081	2.157	0.03	1.111
28年	4.503	0.191	0.078	2.44	0.024	1.103
20~25年	3.825	0.146	0.067	2.185	0.012	1.185
20~26年	3.946	0.148	0.068	2.191	0.013	1.171
27~28年	5.085	0.182	0.08	2.291	0.027	1.107
全部年度	3.672	0.152	0.069	2.189	0.016	1.269

資料來源：研究者自行繪製

2.陳鴻文

中華職棒的另一個選手陳鴻文也是個好的例子，中職 27 年他的自責分率（ ERA ）高達 4.869，比生涯平均 3.861 高出許多，隔年他遭中信兄弟隊以戰力不符為由釋出球隊，但也隨即獲得富邦悍將隊的賞識，爭取到一份新合約，富邦隊怎麼會選擇別人捨棄的球員呢？富邦與中信隊究竟何隊的操控較合理呢？又近幾年來，許多球迷們心中開始納悶：「這幾年的陳鴻文到底怎麼了？真的退步了嗎？」就讓我們用進階數據一探究竟吧！

中職 27 年陳鴻文的自責分率從中職 26 年的 2.943 衝高至 4.869，這意味他每九局的自責失分比去年多 1.926 分，從傳統數據自責分率觀察，陳鴻文中職 27 年大大退步，這或許就是中信釋出他的原因。不過我們分析陳鴻文中職進階數據後發現，陳鴻文職棒 27 年的表現比

職棒 26 年的表現還佳。

由表 14 可看出，職棒 26 年陳鴻文的平均投出四壞球率（BB%）為 0.093（聯盟平均為 0.072），但職棒 27 年他的 BB%大幅降低至 0.057（聯盟平均為 0.081），也就是中職 27 年他平均每面對 100 位打者能比職棒 26 年時少投出 3.6 次四壞球保送，如此可以降低打者上壘的機會，而且他大幅降低的四壞球率（BB%）是在整個聯盟的平均四壞球率（BB%）上升的環境下，可見中職 27 年陳鴻文控球能力進步非常多。此外，SO/BB 為投出三振數與四壞球數的比率，數值愈大的話，可以顯現投手的控球表現與三振能力愈佳，從表 15 可看出陳鴻文的 SO/BB 除了職棒 26 年微幅差於聯盟平均，其他四年皆大幅優於聯盟平均，可見他的控球與三振能力在聯盟中算是強者，這或許就是富邦隊看上陳鴻文的原因。

表 14

球員陳鴻文在各年間的數據列表

中職年份	NAME	FIP'n	ERA	SO%	BB%	SO/BB	HR%	G/F
24年	陳鴻文	3.179	2.970	0.212	0.048	4.413	0.016	1.027
25年	陳鴻文	2.735	2.673	0.154	0.045	3.262	0.004	0.950
26年	陳鴻文	5.300	2.943	0.223	0.093	2.294	0.040	0.687
27年	陳鴻文	4.793	4.869	0.221	0.057	3.630	0.041	0.735
28年	陳鴻文	5.525	5.850	0.208	0.055	3.364	0.049	1.015

資料來源：研究者自行繪製

但是，問題出在他的被全壘打率（HR%）從中職 26 年開始飆高至 0.040（聯盟平均為 0.021），中職 27 年高達 0.041（聯盟平均 0.030），職棒 28 年也持續上升至 0.049（聯盟平均為 0.024），連續三年的被全壘打率皆高出聯盟平均許多，可見他這幾年失分大幅增加的原因可能是被全壘打數增加的結果。但是依據 Matt Klaassen 的研究指出，各個投手的全壘打與飛球比（HR/FB）的年與年相關性極低，相關係數僅有 -0.029，因此可推論飛球是否形成全壘打，並不是投手能完全控制，他有時取決於球場大小、風向、運氣、球的彈性係數等諸多環境因素。因此，我們決定使用排除全壘打這項變數的進階數據——*SIERA* 來分析陳鴻文的表現。

從表 15 的數據我們可看出，陳鴻文每一季的 *SIERA* 皆優於聯盟平均，我們舉陳鴻文遭到中信隊釋出的職棒 27 年為例，他的 *SIERA_N* 為 3.415（聯盟平均為 4.534），為聯盟第六低（職棒 27 年投球局數 40 局以上之球員共有 47 位），從陳鴻文 *SIERA* 明顯低於聯盟平均可以得知，他主要是因為被全壘打率（HR%）攀升，失分進而增加。若不考慮全壘打，則陳鴻文在場上實際的表現其實在聯盟中名列前茅，只是受到球的彈性或運氣等因素影響，造成他的帳面數據不盡理想。我們合理推論如果聯盟球的彈性係數改回正常後，他的表現有可能

回到正常水準，但若用球未改變，則他的投手生涯會因而大受影響。

表 15

球員陳鴻文在各年間的 *SIERA* 列表

	姓名	聯盟平均	<i>SIERAn</i>
職棒 24 年	陳鴻文	4.444	3.349
職棒 25 年	陳鴻文	4.265	3.987
職棒 26 年	陳鴻文	4.477	4.097
職棒 27 年	陳鴻文	4.534	3.415
職棒 28 年	陳鴻文	4.323	3.573

資料來源：研究者自行繪製

(二) 以投手生涯數據作選手排名

我們創造的公式，除了可用於個別選手的分析外，還可利用來比較球員的生涯數據或是比較現役投手，進行投手的排名。在生涯數據排名的部份，由於選手出賽的年份橫跨中職元年至今，因此我們用 *SIERAA* 與 *FIPA* 作為依據為各選手排名；至於現役投手的比較部份，則以 *SIERAN* 與 *FIPN* 作為選手排名的依據。本土投手的生涯排名如表 16 所示，另外，我們也列出依據傳統數據投手自責失分率 (*ERA*) 所得的排名，方便比較。

表 16

本土投手

SIERA 生涯數據排名

FIP 生涯數據排名

ERA 生涯數據排名

排名	選手	<i>SIERAA</i>	排名	選手	<i>FIPA</i>	排名	選手	<i>ERA</i>
1	林恩宇	2.99	1	* 李明進	2.69	1	* 林恩宇	2.37
2	李明進	3.05	2	* 林恩宇	2.73	2	* 黃平洋	2.54
3	曹峻揚	3.21	3	* 郭李建夫	2.89	3	* 郭源治	2.67
4	林柏佑	3.24	4	* 買嘉瑞	3.04	4	* 郭李建夫	2.68
5	郭源治	3.28	5	羅錦龍	3.05	5	* 許銘偉	2.78
6	林英傑	3.29	6	* 林英傑	3.06	6	* 陳義信	2.82
7	劉義傳	3.39	7	潘威倫	3.13	7	* 李明進	2.92
8	潘威倫	3.42	8	* 曹峻揚	3.21	8	* 郭進興	2.93
9	倪福德	3.42	9	鄭凱文	3.22	9	傅于剛	2.97
10	陳鴻文	3.45	10	* 李風華	3.24	10	* 謝長亨	3.02

資料來源：研究者自行繪製

我們發現林恩宇、李明進、曹竣揚、林英傑、潘威倫等 5 位投手同時名列 *SIERA* 與 *FIP* 的排行榜，而最頂尖的林恩宇、李明進同時名列三份榜單。至於出現在自責失分率 (*ERA*) 的名單，則主要包含了早期大家熟悉的四大天王黃平洋、陳義信、謝長亨，以及曾旅日的名投郭源治與郭李建夫，可見中職早期選才或評價投手時著重的是傳統數據。

至於本土現役投手的排名如表 17 所示，我們同樣列出依據傳統數據投手自責失分率 (*ERA*) 所得的排名，方便作比較。

表 17

現役投手排名

SIERA 現役投手排名

FIP 現役投手排名

ERA 現役投手排名

排名	選手	SIERAN	排名	選手	FIPN	排名	選手	ERA
1	林柏佑	3.65	1	潘威倫	3.51	1	傅于剛	2.97
2	潘威倫	3.73	2	羅錦龍	3.51	2	潘威倫	3.17
3	傅于剛	3.76	3	鄭凱文	3.55	3	陳禹勳	3.62
4	倪福德	3.76	4	陳禹勳	3.81	4	羅錦龍	3.65
5	陳鴻文	3.8	5	陽建福	4.03	5	陳鴻文	3.75
6	鄭凱文	3.89	6	林克謙	4.04	6	鄭凱文	3.78
7	羅錦龍	4.11	7	林其緯	4.12	7	陽建福	3.99
8	陳禹勳	4.19	8	林柏佑	4.14	8	林克謙	4.11
9	蔡明晉	4.22	9	傅于剛	4.15	9	倪福德	4.13
10	陳韻文	4.27	10	王鏡銘	4.23	10	林羿豪	4.19

資料來源：研究者自行繪製

我們發現 *SIERA* 與 *FIP* 的的排行榜上的羅錦龍、陳禹勳、潘威倫、鄭凱文、陽建福、傅于剛、林克謙與倪福德等投手，也都列名 *ERA* 的排行榜上。也就是說，本研究利用 2009 年後穩定四隊的數據庫所發展創造的 FIP_N 與 $SIERA_N$ ，不僅排除了守備與運氣成份，且近年來依此標準所選出的最佳投手，多數也符合傳統數據的標準。因此，我們也認為本研究發展的工具，可用來取代傳統投手自責失分率。

三、未來展望與研究限制

本研究中創造出中華職棒版的 *FIP* 與 *SIERA* 等公式，未來還可再嘗試創造出中華職棒版本的 $xFIP$ 公式。同時，我們也將依據本研究創造出中華職棒版的投手進階數據公式，並參考國外研究建立的標準，重新比較、評價中華職棒史上的所有投手，提供棒球界作為選才的參考依據。

另外，本研究主要著重在投手數據，尚未討論到打者與守備相關的進階數據，這也是值得研究的主題。因此，未來可以再建立中華職棒 28 年來的打擊數據資料庫與守備資料庫，並著重在以下幾個研究方向：

1. 創造出中華職棒版的打者相關的進階數據公式，例如目前公認最重要的加權平均上壘率（*wOBA*）公式，再實際利用這些新公式，重新逐年分析比較中華職棒的打者。
2. 研究並發展中華職棒的守備進階數據。
3. 利用我們所創造的投手與打者的進階數據，參考國外研究的轉換公式與相關理論，計算出每位選手對球隊的勝場貢獻，更精準地計算出每位選手的價值，並且將投手與打者放在同一天秤上比較。

在本研究中我們建立數據庫的過程，發現缺了許多中職官網沒有公布的數據，導致本研究發展受限。我們期許中職能統計並公布更多的數據，像是用來計算 *SIERA* 的滾地球總數和飛球總數，或是投手與打者的內野飛球率、平飛球率、高飛球率、滾地球率等，以及投手被二壘打安、三壘安打總數等，能讓相關研究更完整且有更多發展的空間。

柒、結論

本研究先進行文獻探討，以前人的研究與理論作為基礎，並考量台灣棒球環境與國外的差異，創造出適用於中華職棒環境的各項投手進階數據公式。

首先，我們整理中職元年～中職 28 年所有投手的數據建立數據資料庫。接著，利用 EXCEL 軟體中的多變量迴歸分析，創造出中華職棒版的與守備無關的投手數據公式：

$$FIP_A = \frac{3.31 * BB + 16.37 * HR - 1.25 * SO}{IP} + 2.31$$

與進階版的

$$FIP_A' = \frac{3.31 * BB + 16.15 * HR - 1.27 * SO}{IP} - 0.07(G/F) + 2.43$$

另外，我們也迴歸創造出更加穩定，且不受球場環境影響的中職版 *SIERA*_A 公式：

$$\begin{aligned} SIERA_A = & -11.066 * \frac{SO}{PA} + 2.411 * \left(\frac{SO}{PA}\right)^2 + 21.837 * \frac{BB}{PA} - 39.067 * \left(\frac{BB}{PA}\right)^2 - 1.015 * \frac{netGB}{PA} - 4.801 * \left(\frac{netGB}{PA}\right)^2 \\ & + 1.724 * \left(\frac{SO}{PA} * \frac{BB}{PA}\right) + 0.625 * \left(\frac{SO}{PA} * \frac{netGB}{PA}\right) - 21.054 * \left(\frac{BB}{PA} * \frac{netGB}{PA}\right) + 4.131 \end{aligned}$$

上述三個公式適用於中職歷史所有投手的分析與比較。

此外，我們也考慮中華職棒這 28 年來的環境變革，利用 2009 年至今穩定維持四隊的數據庫，創造第 2 個版本的 FIP 公式：

$$FIP_N = \frac{3.49 * BB + 15.74 * HR - 1.61 * SO}{IP} + 2.90$$

與進階版的

$$FIP'_N = \frac{3.47 * BB + 16.23 * HR - 1.51 * SO}{IP} + 0.28(G/F) + 2.47$$

並創造出中職版的 $SIERA_N$ 公式：

$$\begin{aligned} SIERA_N = & -0.141 * \frac{SO}{PA} - 30.265 * \left(\frac{SO}{PA}\right)^2 + 26.471 * \frac{BB}{PA} - 51.411 * \left(\frac{BB}{PA}\right)^2 - 4.283 * \frac{netGB}{PA} - 7.195 * \left(\frac{netGB}{PA}\right)^2 \\ & + 6.952 * \left(\frac{SO}{PA} * \frac{BB}{PA}\right) + 18.944 * \left(\frac{SO}{PA} * \frac{netGB}{PA}\right) + 10.881 * \left(\frac{BB}{PA} * \frac{netGB}{PA}\right) + 3.313 \end{aligned}$$

上述三個公式，適用於中職近年來穩定維持四隊環境的投手分析與比較。

我們分析各變數對投手失分的解釋程度時發現：在中華職棒環境中，**被全壘打是影響投手失分最重要的因素**，而投手的三振數對於自責失分的解釋力很低，這與美國大聯盟的結果，以及傳統棒球界的認知不同。從上述研究結果，我們也建議：投手選才時首要選擇**被全壘打率低與保送率低的投手**，才是符合中職環境的好投手。

基於此，我們提出簡化版的公式方便計算、投手的比較：

$$FIP_S = \frac{3 * BB + 17 * HR}{IP} + 1.5$$

最後，還有一點必需注意，若只看單一種數據評判球員或是下結論作決策，容易造成偏頗，因此評價棒球選手時需要綜合考量多種數據，才能完整地看出選手真正的能力與價值！我們也期望台灣棒球界能實際利用這些重新改造過的進階數據公式，分析比較中華職棒的投手，作為選才的參考依據。

捌、參考資料及其他

1. Cameron,D. (2012, June. 18). Re:Tim Lincecum: Bad Luck or Bad Pitches [FANGRAPHS]. Retrieved from <https://www.fangraphs.com/blogs/tim-lincecum-bad-luck-or-bad-pitches/>
2. Klaassen,M. (2013, January 23). Re:Basic Pitching Metric Correlation 1955-2012,2002-2012 [FANGRAPHS].Retrieved from <https://www.fangraphs.com/blogs/basic-pitching-metric-correlation-1955-2012-2002-2012/>
3. Pawlikowski,J. (2010, February 26). Re:Hit the Ball in the Air Against the Twins [FANGRAPHS].Retrieved from <https://www.fangraphs.com/blogs/hit-the-ball-in-the-air-against-the-twins/>
4. Stats All Forks. (2014, January 20). Re:2013 in baseball :Without the Luck[FANGRAPHS]. Retrieved from <https://www.fangraphs.com/community/2013-in-baseball-without-the-luck/>
5. Tango,T.,Lichtman,M.,Dophin,A.(2014, April 28).*The Book: Playing The Percentages In Baseball*.Seattle, WA, and Charleston, SC,United States:CreateSpace Independent Publishing Platform.
6. Voros McCracken. (2001, January 23).Pitching and Defense: How Much Control Do Hurlers Have?[Baseball Prospectus].Retrieved from <https://www.baseballprospectus.com/news/article/878/pitching-and-defense-how-much-control-do-hurlers-have/>
7. Weinberg,N. (2014, December 12). Re:Which is Better?A Ground Ball Pitcher or a fly ball pitcher [FANGRAPHS].Retrieved from <https://www.fangraphs.com/library/which-is-better-a-ground-ball-pitcher-or-a-fly-ball-pitcher/>
8. *BABIP* [FANGRAPHS]. Retrieved from <https://www.fangraphs.com/library/pitching/babip/>
9. *xFIP* [FANGRAPHS]. Retrieved from <https://www.fangraphs.com/library/pitching/xfip/>
10. *FIP* [FANGRAPHS]. Retrieved from <https://www.fangraphs.com/library/pitching/fip/>
- 11.吳明隆（2011）。論文寫作與量化研究。臺北市：五南。
- 12.邱皓政（2006）。量化研究法（二）統計原理與分析技術。臺北市：雙葉書廊。

【評語】 052707

針對職棒資料做長期的分析，進而提供國內職棒投手表現的量化數據，值得鼓勵。

此研究加入滾飛比後的解釋量並沒有太大差異而三振的係數與其他變項係數相比也不高，加入傳統變項後的三振係數甚至可忽略不計，因而剩下兩個主要預測因子，但是這樣的結果只能讓一般球迷使用，對於選手的專業評估是否具有效度仍待進一步研究。

摘要

本研究首先進行文獻探討，接著建立中職所有投手的數據庫，以國外的研究理論為基礎，並考量台灣棒球環境與國外的差異，排除守備影響，篩選出能真實反應投手表現的變數。我們先以職棒 28 年來的數據庫，利用多變量迴歸分析創造出中職版的 FIP_A ，並加入新變數「滾飛比」創造出進階版的 FIP'_A 。再考慮環境變革，以穩定維持四隊的數據庫創造出 FIP_N 與 FIP'_N 。另外，也迴歸創造出不受環境影響的進階公式 $SIERA_A$ 及 $SIERA_N$ 。研究發現，在中職環境裡投手的三振對失分的解釋力低，投手選才應首重被全壘打率低與保送率低的投手。我們也應用這些公式分析比較中職的投手，提供台灣棒球界選才參考。

壹、研究動機與文獻探討

棒球是產值驚人的運動產業，如何正確地評價球員與選才，成為管理階層的重要課題，而**棒球統計學**，正是達成此目的的重要工具。評論棒球投手時，大家常看自責分率等傳統數據，然而當投手面對打者時，會有「三振」、「保送」、「被全壘打」及「球被打進場中」四種情況。Voros McCracken (2001) 指出「**球被打進場中是否形成安打**」包含了許多**運氣**成分且與**隊友守備**有關。因此，如何改善傳統數據以衡量投手真正的能力與表現是我們感興趣的問題。

Tango (2007) 提出**與守備無關的投手數據** (FIP , *Fielding Independent Pitching*)，只考慮投手的四壞球數、三振數與被全壘打數，**排除守備員與運氣的影響，真實地反應投手的表現**，是目前美國職棒大聯盟的主流數據。而 $SIERA$ (*Skill-Interactive ERA*) 是更複雜的公式，加入與飛球和滾地球比率有關的變數。

台灣棒球界相對缺乏此類進階數據的開發與研究，若直接套用美國大聯盟的公式到中華職棒球員的數據，會因為**環境的不同**而產生誤差或偏差。因此，我們考量台灣棒球環境與國外的差異，建造中華職棒的歷史數據庫，為中華職棒量身打造各項投手進階數據公式，更精確地評估中職投手的表現與能力。

貳、研究目的

- 一、聚焦在三振、保送、全壘打、飛球滾地球比率等與守備無關的數據，利用多變量迴歸分析，創造出適用於中華職棒的 FIP 、 FIP' 與 $SIERA$ 公式，作為分析、比較職棒投手表現的工具。
- 二、考慮中華職棒環境的變革，以近年來穩定維持四隊後的數據，創造出新版的 FIP 、 FIP' 與 $SIERA$ 公式。
- 三、分析三振、保送、全壘打等變數與投手責任失分的關係，了解中職環境中影響投手失分最重要的變數。
- 四、利用上述投手進階數據，實際分析比較中職投手的表現，提供台灣棒球界選才參考。

參、研究設備與器材

- 一、電腦、利用 EXCEL 軟體進行迴歸分析、WORD 軟體進行文書編輯。
- 二、我們所整理創建的中華職棒數據庫：1. 歷年投手數據庫、2. 投手的生涯數據庫、3. 各年度投手數據庫。

肆、研究流程與方法

- 一、進行文獻探討，討論重要的投手進階數據，以及相關的重要變數。
- 二、建立中華職棒投手 28 年的**歷史投手數據庫**、**各年度數據庫**與**生涯數據庫**，並篩選出適合的樣本。
- 三、以投手數據中**與守備無關的「三振」、「保送」與「被全壘打」**作為解釋變項，並以「自責失分」作為依變項，利用多變量迴歸分析創造出中職版本的 FIP_A 公式。
- 四、「**滾地與飛球比**」與守備無關且是投手可控制的數據，故加入此變數創造出中職版 FIP'_A 公式。
- 五、以近年穩定**維持四隊的數據庫**，創造出新版本的 FIP_N 與 FIP'_N 公式，適用近年或現役選手的分析。
- 六、探討 $SIERA$ 的意義，並利用多變量迴歸分析創造出中職版的 $SIERA_A$ 及 $SIERA_N$ 公式。
- 七、討論這些公式的意義，比較公式中各變數對投手失分的解釋程度，找出影響投手失分最重要的變數，提供選才參考。並造出簡化版的 FIP_S 公式，供一般球迷使用。
- 八、實際進行中職投手的分析與比較：個別球員的分析、中華職棒史上所有投手與現役投手的排名。

本研究的研究流程圖

研究背景

文獻探討
變項選擇

研究方法

中職數據庫建置
樣本選擇
多變量迴歸分析

研究結果

創造公式
公式分析與比較
變數探討
實際球員分析
結論與建議

伍、研究結果與討論

一、建立中華職棒歷年投手數據資料庫

我們從中職官方網站，逐年整理中職 28 年每一投手的各項數據，建立 (一) 中華職棒投手 28 年完整數據庫供研究使用。並以投球局數 40 局為門檻，篩選出 1172 筆資料作為本研究的樣本。接著，再建立中華職棒投手 (二) 各年度投手數據庫與 (三) 投手生涯數據庫，同樣篩選出合適的樣本。

NAME	IP	H	HR	BB	SO	ER	BB/P	HR/P	SO/P	ER/P	G/F	BB9	HR9	SO9	ERA	NAME	G	GS	GR	P	WHIP	ERA	BF	H	HR	BB	ER	SO	ER	NAME	G	GS	GR	P	WHIP	ERA	BF	H	HR	BB	ER	SO	ER		
0101 * 潘韋廷	246	198	14	48	177	61	0.195	0.057	0.720	0.248	1.461	1.756	0.512	6.476	2.232	001 潘韋廷	307	208	19	1812	1.220	3.170	7759	8388	104	265	9	69	3902	436	001 潘韋廷	26	26	0	159	1.31	3.45	691	168	11	46	0	5	110	61
0102 * 史東志	216	173	9	87	110	46	0.403	0.042	0.509	0.213	1.908	3.625	0.375	4.583	1.917	002 史東志	339	211	128	1570	1.250	3.200	6540	1458	110	501	18	63	1286	359	002 史東志	27	25	2	158	1.44	4.44	687	191	10	36	0	2	115	78
0103 * T.M	195	174	7	38	105	45	0.195	0.036	0.538	0.231	1.263	1.753	0.323	4.844	2.076	003 林克謙	28	24	4	145	1.39	4.02	624	160	4	42	3	7	87	65	003 林克謙	28	24	4	145	1.39	4.02	624	160	4	42	3	7	87	65
0104 * 楊瑞麟	174	153	4	92	177	50	0.259	0.023	1.017	0.287	1.407	4.759	0.207	9.155	2.586	004 楊瑞麟	28	19	7	135	1.44	4.31	587	145	10	51	0	3	94	65	004 楊瑞麟	28	19	7	135	1.44	4.31	587	145	10	51	0	3	94	65
0105 * 黃國川	160	142	5	80	83	47	0.499	0.031	0.518	0.293	1.528	4.494	0.281	4.663	2.640	005 黃國川	312	199	113	1267	1.410	3.990	5971	1369	64	434	13	82	634	275	005 黃國川	22	22	0	133	1.30	3.30	576	159	6	15	3	5	91	49
0106 * 周煥平	155	126	9	25	52	59	0.148	0.052	0.335	0.341	1.538	1.334	0.444	3.015	3.073	006 周煥平	187	136	51	1275	1.120	2.920	5189	1197	109	228	25	36	746	400	006 周煥平	22	22	0	133	1.30	3.30	576	159	6	15	3	5	91	49
0107 * 徐維謙	149	124	3	51	116	41	0.242	0.020	0.738	0.275	1.126	3.081	0.181	6.844	2.477	007 徐維謙	183	103	80	1111	1.300	4.110	4801	1184	32	379	22	31	721	330	007 徐維謙	23	23	0	127	1.61	6.06	589	160	15	45	0	5	78	86
0108 * 唐明松	143	112	6	65	87	36	0.455	0.042	0.608	0.252	1.156	4.091	0.378	5.476	2.266	008 唐明松	136	151	175	1101	1.400	4.120	4843	1180	94	368	12	40	669	307	008 唐明松	25	24	1	126	1.64	5.20	571	158	7	49	0	3	77	73
0109 * 杜福明	143	154	7	24	46	50	0.168	0.049	0.336	0.350	1.160	1.510	0.441	3.021	3.147	009 杜福明	181	147	34	1065	1.090	2.550	4388	934	47	246	11	33	779	338	009 杜福明	27	20	7	115	1.34	4.43	495	132	11	23	1	8	93	37
0110 * 陳逸松	129	123	9	35	71	51	0.271	0.070	0.550	0.395	1.007	2.440	0.627	4.950	3.555	010 陳逸松	169	31	435	1057	1.200	3.120	4447	1061	67	210	45	30	651	307	010 陳逸松	19	19	0	114	1.50	3.45	512	133	15	36	2	4	72	69
0111 * 謝廣興	124	131	9	48	57	59	0.387	0.073	0.460	0.476	0.748	3.484	0.653	4.137	4.282	011 謝廣興	245	162	83	1053	1.230	3.310	4430	1053	54	325	19	36	792	388	011 謝廣興	40	11	29	104	1.36	4.33	446	115	5	26	6	3	50	50
0112 * 黃武雄	117	101	5	41	56	26	0.350	0.043	0.479	0.222	2.012	3.154	0.385	4.308	2.000	012 黃武雄	230	135	95	1071	1.130	3.800	4449	1013	30	334	18	40	732	438	012 黃武雄	19	19	0	93	1.37	3.94	406	104	6	24	0	3	64	41
0113 * 張金龍	105	123	6	19	37	35	0.181	0.057	0.542	0.523	1.142	1.825	0.513	4.676	4.705	013 張金龍	117	95	160	1.180	2.970	4262	986	36	305	13	21	569	332	013 張金龍	19	14	5	87	1.59	6.18	397	116	12	23	0	4	58	60	
0114 * 黃聖璋	105	133	8	49	50	62	0.466	0.076	0.476	0.590	1.144	4.196	0.685	4.282	5.309	014 黃聖璋	139	103	47	927	1.120	2.940	3786	921	39	244	13	20	581	362	014 黃聖璋	19	19	0	93	1.37	3.94	406	104	6	24	0	3	64	41
0115 * 劉亮輝	103	106	5	35	45	44	0.340	0.049	0.437	0.427	1.366	3.058	0.437	3.932	3.845	015 劉亮輝	135	105	30	814	1.150	3.750	3065	994	84	211	12	25	440	382	015 劉亮輝	24	17	7	89	1.84	5.76	422	116	5	48	2	2	40	57
0116 * 李文偉	88	94	9	13	12	31	0.148	0.102	0.136	0.352	1.313	1.328	0.919	1.226	3.167	016 李文偉	137	133	4	897	1.200	3.220	2699	964	64	175	3	21	711	310	016 李文偉	14	5	9	80	1.07	6.47	396	110	10	41	0	3	61	58
0117 * 謝冠忠	84	89	4	50	30	32	0.595	0.048	0.357	0.380	1.724	5.351	0.428	3.210	3.424	017 謝冠忠	52	39	407	817	1.300	3.910	3584	923	50	212	26	36	606	355	017 謝冠忠	30	3	47	75	1.52	6.07	347	91	10	24	4	3	44	51
0118 * 許生口	81	90	4	45	48	38	0.354	0.049	0.591	0.468	1.621	4.988	0.443	5.320	4.212	018 許生口	104	69	235	814	1.410	3.940	2560	862	37	284	13	13	365	256	018 許生口	14	14	0	74	1.41	4.48	328	85	4	20	1	1	61	37
0119 * 林光宏	80	97	5	27	21	41	0.338	0.063	0.263	0.513	1.000	3.038	0.563	2.363	4.613	019 林光宏	173	113	60	800	1.180	3.710	3389	915	30	130	15	21	511	330	019 林光宏	14	14	0	74	1.41	4.48	328	85	4	20	1	1	61	37
0120 * 李傑志	72	1	88	5	28	29	0.388	0.069	0.361	0.402	2.989	3.495	0.434	3.245	3.620	020 李傑志	128	109	17	786	1.170	2.910	2863	768	49	154	7	30	541	248	020 李傑志	38	5	33	74	1.59	5.35	338	81	3	37	0	5	40	44

中華職棒投手 28 年完整數據庫

各年度投手數據庫

投手生涯數據庫

二、中職版的 FIP 與 FIP' 公式

FIP 公式以及 FIP' 公式是**利用與守備無關、且是投手相對能控制的三振、保送、被全壘打與滾飛比等變數，預測當隊友的守備能力是聯盟平均水準時，一位投手平均每 9 局的責任失分**，且能真實地**量測投手個人的表現**。我們利用研究中建立的中職數據庫，並使用 EXCEL 進行多變量迴歸分析，創造出中華職棒版的 FIP 公式及獨創的中職 FIP' 公式：

適用範圍	FIP (變數無滾飛比)	FIP' (變數加入滾飛比)
全部年度	$FIP_A = \frac{3.31 * BB + 16.37 * HR - 1.25 * SO}{IP} + 2.31$	$FIP'_A = \frac{3.31 * BB + 16.15 * HR - 1.27 * SO}{IP} - 0.07(G/F) + 2.43$
合併四隊後 (2009 年後)	$FIP_N = \frac{3.49 * BB + 15.74 * HR - 1.61 * SO}{IP} + 2.90$	$FIP'_N = \frac{3.47 * BB + 16.23 * HR - 1.51 * SO}{IP} - 0.28(G/F) + 2.47$

三、變數的比較與 FIP_S 公式

我們對各變數做了一系列的探討後發現：**保送與被全壘打**是影響投手失分的重要因素，而去掉「三振」這項變數並沒有讓各公式的解釋力 (R²) 明顯下降，表示中職環境中**投手的三振對失分的解釋力低**。基於此，我們提出簡化版的 FIP_S 公式方便計算、進行投手比較：

$$FIP_S = \frac{3 * BB + 17 * HR}{IP} + 1.5$$

此公式中僅包含「保送」與「全壘打」這兩項對失分具有足夠解釋力，並且是投手相對能控制的變數。

四、中職版的 SIERA 公式

SIERA (Skill-Interactive ERA) 這項數據是**利用投手的技能 (三振、保送、滾飛比) 與它們的交互作用來描述投手表現，預測投手的失分**。我們利用多變量迴歸分析，創造出適用於中職歷史投手分析的公式 SIERA_A：

$$SIERA_A = -11.066 * \frac{SO}{PA} + 2.411 * \left(\frac{SO}{PA}\right)^2 + 21.837 * \frac{BB}{PA} - 39.067 * \left(\frac{BB}{PA}\right)^2 - 1.015 * \frac{netGB}{PA} - 4.801 * \left(\frac{netGB}{PA}\right)^2 + 1.724 * \left(\frac{SO}{PA} * \frac{BB}{PA}\right) + 0.625 * \left(\frac{SO}{PA} * \frac{netGB}{PA}\right) - 21.054 * \left(\frac{BB}{PA} * \frac{netGB}{PA}\right) + 4.131$$

並用「合併成四隊後 (2009 年後) 的數據庫」創造出適用於近年與現役選手分析的公式 SIERA_N：

$$SIERA_N = -0.141 * \frac{SO}{PA} - 30.265 * \left(\frac{SO}{PA}\right)^2 + 26.471 * \frac{BB}{PA} - 51.411 * \left(\frac{BB}{PA}\right)^2 - 4.283 * \frac{netGB}{PA} - 7.195 * \left(\frac{netGB}{PA}\right)^2 + 6.952 * \left(\frac{SO}{PA} * \frac{BB}{PA}\right) + 18.944 * \left(\frac{SO}{PA} * \frac{netGB}{PA}\right) + 10.881 * \left(\frac{BB}{PA} * \frac{netGB}{PA}\right) + 3.313$$

以中華職棒歷年的數據來看，此公式計算所得的投手數據更加穩定，且能更精準地衡量一個投手的真實表現。

五、討論

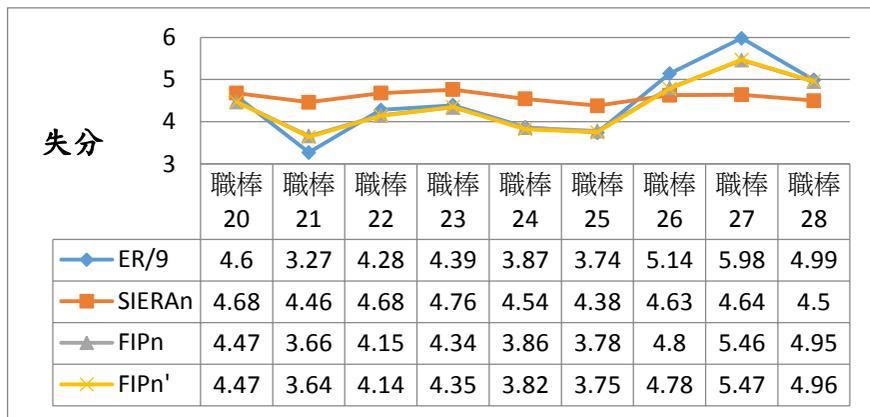
(一) 公式的分析與比較

1. **傳統投手數據在量測選手的表現時缺乏信度與效度**，本研究所發展的公式可克服這些缺點。而中職版 FIP 公式中的係數與美國職棒大聯盟版的公式 $FIP = \frac{13 * HR + 3 * BB - 2 * SO}{IP} + 常數$ ，明顯不同。若直接套用美國版的公式來分析比較台灣投手，會因未考慮環境的差異，產生較大的偏差與錯誤的比較，這是本研究的重要意義與價值所在。

2. 分析各變數後發現：在中職環境中，保送與被全壘打是影響投手失分的重要因素，三振對失分的解釋力偏低。基於模型簡化的精神，我們去除「三振」這項變數，提出簡化版的公式方便計算進行投手比較：

$$FIP_S = \frac{3 * BB + 17 * HR}{IP} + 1.5$$

3. 中職 20 年~29 年各項投手數據聯盟平均值趨勢



- *SIERA* 這項數據利用投手的技能（三振、保送、滾飛比）與它們的交互作用來描述投手表現，其變數共有 9 項。
- 整個聯盟的平均 *SIERA_N* 是相對穩定的數據，其他三項數據隨環境的改變較起伏大。

（二）實際應用與分析球員

1. 投手個別案例分析：陳禹勳、陳鴻文、希克等。
2. 利用 *SIERA_A* 與 *FIP_A* 進行中職歷史上所有投手生涯數據的排名與比較。
3. 利用 *SIERA_N* 與 *FIP_N* 進行中職現役投手生涯數據的排名與比較。

SIERA_A 生涯數據排名

排名	選手	SIERAA
1	林恩宇	2.99
2	李明進	3.05
3	曹峻場	3.21
4	林柏佑	3.24
5	郭源治	3.28
6	林英傑	3.29
7	劉義傳	3.39
8	潘威倫	3.42
9	倪福德	3.42
10	陳鴻文	3.45

FIP_A 生涯數據排名

排名	選手	FIPA
1	* 李明進	2.69
2	* 林恩宇	2.73
3	* 郭李建夫	2.89
4	* 賈嘉瑞	3.04
5	羅錦龍	3.05
6	* 林英傑	3.06
7	潘威倫	3.13
8	* 曹峻場	3.21
9	鄭凱文	3.22
10	* 李風華	3.24

SIERA_N 現役投手排名

排名	選手	SIERAN
1	林柏佑	3.65
2	潘威倫	3.73
3	傅于剛	3.76
4	倪福德	3.76
5	陳鴻文	3.8
6	鄭凱文	3.89
7	羅錦龍	4.11
8	陳禹勳	4.19
9	蔡明晉	4.22
10	陳韻文	4.27

FIP_N 現役投手排名

排名	選手	FIPN
1	潘威倫	3.51
2	羅錦龍	3.51
3	鄭凱文	3.55
4	陳禹勳	3.81
5	陽建福	4.03
6	林克謙	4.04
7	林其緯	4.12
8	林柏佑	4.14
9	傅于剛	4.15
10	王鏡銘	4.23

陸、結論與未來展望

一、結論

- （一）本研究進行文獻探討後，以前人的研究與理論作為基礎，並考量台灣棒球環境與國外的差異，先建立中華職棒 28 年的投手數據庫，再利用多變量迴歸分析創造出 *FIP_A*、*FIP_A'* 與 *SIERA_A* 三類中華職棒版的投手進階數據公式，適用於中職歷年投手的分析與比較。
- （二）考慮中華職棒的環境變革，利用 2009 年至今穩定四隊的數據庫，創造出穩定四隊版的三類公式 *FIP_N*、*FIP_N'* 與 *SIERA_N*，適用於近年合併四隊之後的球員分析與比較。
- （三）我們分析各變數對投手失分的解釋程度時發現：在中華職棒環境中，投手的三振數對於自責失分的解釋力低，這與美國大聯盟的結果，以及傳統棒球界的認知不同。基於此，我們提出簡化版的 *FIP_S* 公式方便計算與投手的比較。並提出建議：中職環境的投手選才，應首要選擇被全壘打率低與保送率低的投手。
- （四）所有的棒球數據都有其限制，若只看單一種數據評判球員或是下結論、作決策，容易造成偏頗，因此評價棒球選手時需要綜合考量多種數據，才能完整地看出選手真正的能力與價值。

二、未來研究展望

- （一）建立中華職棒 28 年來的打者數據資料庫，創造出中華職棒版的加權平均上壘率 (*wOBA*) 等進階數據公式，再實際利用這些新公式，重新逐年分析比較中華職棒的打者。
- （二）利用我們創造的投手與打者的進階數據，參考國外研究的轉換公式與相關理論，計算出每位選手對球隊的勝場貢獻，更精準地計算出每位選手的價值，並且將投手與打者放在同一天秤上比較。

重要參考文獻

1. Klaassen, M. (2013). *Basic Pitching Metric Correlation 1955-2012, 2002-2012*.
2. Tango, T., Lichtman, M., Dopin, A. (2014). *The Book: Playing The Percentages In Baseball*. Seattle, WA, and Charleston, SC, United States: CreateSpace Independent Publishing Platform.
3. Voros McCracken (2001). *Pitching and Defense: How Much Control Do Hurlers Have?*