

# 108學年度專題報告競賽

題目：探討棒球場上薪資與表現：

以美國職棒與中華職棒為例

系所班別：國立臺北大學統計系

姓名學號：林偲晴(410578028)

鄧揚耀(410578045)

張苙庭(410578050)

陳威傑(410578068)

甘珮儒(410474212)

報告日期：2020/06/05

## 摘要

### 壹、前言

- 研究背景：棒球從日本殖民時代開始便是台灣的國球，然而就在去年 2019 年時，代表台灣的中華隊在各層級的棒球國際賽取得佳績，更在爭取奧運資格的世界 12 強棒球賽取得第五名的優秀成績，這也促使了我們想深入研究棒球這項運動。藉此研究我們主要想了解球團薪資與球團和球員表現之間的關係，以及比較美國職棒和中華職棒可能存在的差異。

- 目的：透過比較美國職棒和中華職棒的差異，以及分析薪資與球團表現間的影響，找出台灣棒球的缺陷與迷思並且將其改善。更希望透過這項研究使政府對台灣棒球政策有所改善，讓更多企業能夠支持及投資於這項運動，也能透過數據分析，讓球團簽訂薪資時更有依據，也更有保障。

- 重要性：球員薪資與球員表現確實存在一定關係，球團訂定球員薪資合約時，顯然已將球員表現納入考量。然而，簽定薪資合約的時間點也是必須探討的重要因素之一。球團與球員洽談未來一年薪資的時間，通常為前一年球季結束後，球團並不能完全預測未來一年球員的表現，只能依照球員前一年的表現給予評估後的薪資；換句話說，球團簽定薪資合約是有風險的，故我們想透過研究試圖降低此風險。

### 貳、文獻回顧

#### （一）薪資

於 Team Payroll Versus Performance in Professional Sports: Is increased Spending Associated with Greater Success? (Grant Shorin, 2017)中提及，球團越高的薪資支出與球隊越高的勝率有關。四種運動（MLB、NBA、NHL、NFL）在例行賽中，由計量經濟分析顯示薪資花費越高的球隊在比賽上有著更好的表現。而在季後賽中，僅有 NBA 和 NHL 達到統計上的顯著，然而若移除了團隊固定效果後，MLB 亦於球隊薪資和冠軍勝率具有顯著正向關係。

#### （二）薪資不平等

於 Team payroll and team performance in major league baseball: 1985-2002 (Frederick Wiseman, Sangit Chatterjee, 2003)中提及，球隊內巨大的不平等

對於運動的競爭平衡有所影響，且當球隊中個人的薪資越不平等時，球隊的勝場數越高。

於 Testing Causality Between Team Performance and Payroll The Cases of Major League Baseball and English Soccer(Stephen G. Hall, Andrew Zimbalist, 2013)中的數據顯示，在 1990 年代，球隊工資與績效之間的橫截面相關性顯著增加。

又於 The Effect of Salary Distribution on Production: An Analysis of Major League Baseball (R.Todd Jewell, David J. Molina, 2004)中顯示，MLB 中的薪資隨著球隊內與球隊間的薪水不平等而升高。其中升高的不平等也可能影響單一球隊或整個聯盟的成功。並且發現，以一支球隊的獲勝百分比衡量，MLB 球隊內部的薪資分配確實對球隊的成功產生了顯著的負面影響。

而又於 The relationship between payroll and performance disparity in major league baseball: an alternative measure (Daniel Mizak, Anthony Stair, 2004)中顯示於 MLB 內球隊間越大的薪資不平等，會導致越大的勝利不平等和越大的競爭不平等。亦發現自 1990 年代中期以來，尤其是在美國職棒的美國聯盟中，薪資差距越來越大，表明奢侈稅一直無效，並且在不久的將來可以預見更大的績效差距。

### （三）研究方法

於 Team Payroll Versus Performance in Professional Sports: Is increased Spending Associated with Greater Success? (Grant Shorin, 2017)中提及不同的研究方法。經驗分析可平衡固定影響的迴歸模型以控制一些特權因素（例如：教練、管理、地點等等）。而後續分析可以延伸不同的成功標準以了解薪資如何與其他結果相關聯，例如：排名、最終決戰勝率、團隊價值等等。然而，這個研究可能因專注於團隊層級的薪資支出，卻無法看到任何球員層級的淺在關係而產生限制。

### （四）有效支出

於 Team Payroll Versus Performance in Professional Sports: Is increased Spending Associated with Greater Success? (Grant Shorin, 2017)中指出，平均而言，增加美國職業棒球大聯盟，NBA 和 NHL 的球隊薪資支出是比較明智的，因為如果球隊取得更好的現場表現，可以抵消大部分支出，還可以為團隊如何最有效地花錢實現組織目標提供有價值的見解。

## 參、資料介紹及前置處理

### • 資料來源

與棒球相關的資料來自 Lahman's Baseball Database 與 cpbl stats。

勝率 (Win\_percentage) 由每個團隊的獲勝除以一年中的總比賽次數除以 100 得出。勝率取值介於 0 到 100 之間。團隊薪水，投手薪水或擊球手薪水總計為所有球員，投手或擊球手的薪水，以百萬美元為單位。工資數據是營業日工資。在一個賽季中球員可能受傷，被交易或被召喚而無法踢球的情況下，一個賽季中不參加任何一場比賽的球員的薪水不包括在球隊總工資中。此外，為消除通貨膨脹的影響，所有薪資變量均除以 CPI，然後轉換為它們的實際價值。

美國職棒大聯盟 (Major League Baseball，簡稱：MLB，或大聯盟)，是世界水準最高的職業棒球比賽，由國家聯盟和美國聯盟在 1902 年成立，與職業籃球 (NBA)、美式足球 (NFL)、冰上曲棍球 (NHL) 並稱北美四大職業體育競賽。目前美國職棒大聯盟共有三十支球隊，分屬兩聯盟。其中國家聯盟十五隊，美國聯盟十五隊。兩聯盟各分為三區(東區，中區，西區)，各分區的冠軍球隊及兩聯盟的外卡球隊均可參加季後賽，爭奪世界大賽冠軍。

中華職業棒球大聯盟 (CPBL，英文全稱為 Chinese Professional Baseball League，簡稱中華職棒大聯盟或中華職棒) 是於 2003 年由「中華職棒聯盟」與「台灣大聯盟」兩職業聯盟合併而改制。1989 年中華職棒聯盟成立，1990 年 03 月 17 日中華職棒開幕戰開打，正式宣告台灣進入職棒元年。1996 年台灣職業棒球大聯盟成立，1997 年 02 月 28 日台灣大聯盟開幕戰揭幕，開啟我國職棒進入兩聯盟並存競爭的新時代。2003 年 01 月 13 日因兩聯盟長期惡鬥，包含爭奪球員、相互挖角、黑球或票房等問題，導致雙邊聯盟呈現「黑暗期」。在時任總統陳水扁指示體委會協調之下，兩聯盟合併 (實質為台灣大聯盟合成兩隊併入中華職棒)，並定名為「中華職業棒球大聯盟」；同年 03 月 01 日，合併後的中華職棒大聯盟開幕戰登場。自此，台灣走向單一聯盟化，而兩聯盟並立的時代正式走入歷史。目前有富邦悍將隊、樂天桃猿隊、中信兄弟隊、統一 7-ELEVEn 獅隊、味全龍隊。

#### • 資料處理

首先，將資料分成打者，投手，團隊討論，並取出其 1985-2016 年的資料，成立新的資料集。再將新資料集中的打者資料與投手資料，將同一年度 stint>1 的數據合併，清理資料；並且增加傳統以及進階棒球數據變數進入打者與投手資料，完成資料集的前置處理；另外將團隊資料新增 2 個變數。最後一步：我們將打者資料篩選出完成規定打席之打者(打席須  $\geq 3.1 * \text{Game}$ )；然後將投手資料篩選出投球局數  $\geq 120$  為先發投手， $< 120$  為後援投手。結合以下傳統棒球數據&進階棒球數據，後完成資料清理，進

行統計分析，以下為打者與投手的數據計算方式：

### 打者數據

#### AVG(或 BA) 打擊率(Batting Average)

計算公式：

$$AVG = H / AB$$

#### OBP 上壘率(On Base Percentage)

計算公式：

$$OBP = (H + BB + HBP) / (AB + BB + HBP + SF)$$

備註: IBB 早已包含在 BB 中，切記不要重複計算

#### TB 壘打數(Total Bases On Safe Hits)

計算公式：

$$TB = 1 \times 1B + 2 \times 2B + 3 \times 3B + 4 \times HR$$

#### IsoP 純長打率(Isolated Power)

計算公式：

$$IsoP = SLG - AVG$$

純長打率=(長打率-打擊率)

#### IsoD (Isolated Discipline) — 純選球率

計算公式：

$$IsoD = (OBP - AVG)$$

純選球率=(上壘率-打擊率)

說明：

純選球率是判斷打者選球力的指標，比單純的上壘率好，因為可以有效降低上壘率被打擊率烘托的效應。

一般來說，一位打者的純選球率要是大於0.100，即為絕佳的一棒人選；及格的一棒，最起碼的要求也是要大於0.060。

#### SecA (Secondary AVG) — 第二打擊率

計算公式：

$$SecA = (BB + (TB - H) + (SB - CS)) / AB$$

第二打擊率=(四壞球+(總壘打數 - 安打)+(盜壘-盜壘刺))/打數

#### EqA (Equivalent Average)

計算公式:

$$EqA = (H + TB + 1.5 \times (BB + HBP) + SB) / (AB + BB + HBP + CS + SB/3)$$

EqA = (安打+總壘打數+1.5x(四壞球+觸身球)+盜壘)/(打數+四壞球+觸身球+盜壘刺+盜壘/3)

#### RC (Run Create) — 創造得分

計算公式：

$$A = H + BB + HBP - CS - GIDP$$

$$B = TB + 0.24 \times (BB - IBB + HBP) + 0.62 \times SB + 0.5 \times (SH + SF) - 0.3 \times SO$$

$$C = AB + BB + HBP + SH + SF$$

$$RC(1) = A \times B / C$$

$$RC(2) = [(2.4xC + A) \times (3xC + B) / 9xC] - 0.9Xc$$

A = 安打+四壞球+觸身球-盜壘刺-雙殺打

B = 總壘打數+0.24x(四壞數-故意四壞+觸身球)+0.62x盜壘+0.5x(犧牲短打+犧牲高飛)-0.3x三振

C = 打數+四壞球+觸身球+犧牲短打+犧牲飛球

$$\text{創造得分}(1) = A \times B / C$$

$$\text{創造得分}(2) = [(2.4xC + A) \times (3xC + B) / 9xC] - 0.9Xc$$

說明：

RC，創造得分，是一個比打點及得分更好的數據，可以看出一位打者整季替球隊貢獻幾分。

RC的公式共有非常多種版本，

RC(1)的為2002年前的版本，是假定打線9位都是同一選手時，該位選手創造出來的得分。

RC(2)所介紹的是Bill James於2002年提出的最新版本，

此版本是假設該位球員與其他8位都是OBP.300及SLG.400的隊友（約是聯盟平均成績）時，替球隊創造出來的得分效益。

提出此一版本的目地是在修正當某位選手有過於極端突出的表現時,RC(1)的版本會出現高估的效應.

至於最原始版本「RC=(安打數+四壞球)x(總壘打數)/打席」則計算上比較簡單,目前也仍有人在使用!!

### OPS+ (Adjusted Production) — 修正能力值

計算公式：

$$OPS+ = (OBP/Lg\ OBP) + (SLG/Lg\ SLG) - 1$$

$$OPS+ = (\text{上壘率}/\text{聯盟平均上壘率}) + (\text{長打率}/\text{聯盟平均長打率}) - 1$$

### ABR (Adjusted Batting Runs) — 修正每打數所創造的分數

計算公式：

$$ABR = (0.47) \times 1B + (0.78) \times 2B + (1.09) \times 3B + (1.40) \times HR + (0.33) \times (BB + HBP) - (0.25) \times (AB - H) - (0.50) \times (H + BB + HBP - LOB - R - CS)$$

$$ABR = (0.47 \times \text{一安}) + (0.78 \times \text{二安}) + (1.09 \times \text{三安}) + (1.40 \times \text{全壘打}) + 0.33 \times (\text{四壞} + \text{觸身}) - 0.25 \times (\text{打數} - \text{安打}) - 0.50 \times (\text{安打} + \text{四壞} + \text{觸身} - \text{殘壘數} - \text{得分} - \text{盜壘刺})$$

### 投手數據

#### ERA自責分率(Earned Run Average)

計算公式：

$$ERA = (ER \times 9) / IP$$

備註：因為合併資料會造成自責分率有誤，所以要重新計算。

### WHIP每局被上壘率(Walks Plus Hits per Inning Pitched)

計算公式:

$$WHIP = (H + B B) / IP$$

### ERA+ (Earned Runs Average Plus) — 修正ERA

計算公式：

$$ERA+ = LgERA / ERA$$

$$\text{修正ERA} = \text{聯盟ERA平均} / ERA$$

說明：

ERA+是將原本的ERA去除掉「球場因素」，這個指標可以視為投手的PRO+(或OPS+)，可以測量出這個投手相較於聯盟

所有其他投手的平均責失分比率。一個投手本身的ERA越低，當然ERA+就越高，也就代表此投手越好。

如同OPS+一樣，當一個投手的ERA+是100%時，那就是聯盟平均水準；若是120%，就是高於聯盟平均水準20%。

### APR (Adjusted Preventing Runs) — 修正失分

計算公式：

$$APR = IP / 9 * (LgERA - ERA)$$

$$\text{修正失分} = \text{投球局數} / 9 * (\text{聯盟平均ERA} - ERA)$$

說明：

當ERA已經經過球場修正後，這個指標可以測量出一個投手與一個聯盟平均等級的投手在一個「中立」的球場投球時，在同樣的投球局數下，可以少失多少(自責)分。

#### 【註】

此處的ERA可以全部換成為RA

$$\text{亦即，} APR = IP / 9 * (LgRA - RA)$$

## 研究方法

### (一) 探索性資料分析 Exploratory Data Analysis (EDA)：

分析數據，概括呈現數據集中的主要特點，探討各變數跟薪資的關係。

### (二) 主成分迴歸 Principal Component Regression (PCR)：

找出重要的變數以利後續的預測模型。

### (三) 偏最小平方迴歸 Partial Least Squares Regression (PLSR)：

預測線性迴歸模型。

### (四) 多元線性迴歸 Multiple Linear Regression：

建立出迴歸模型，藉此預測薪資。

## 統計方法

### (一)多元線性迴歸 (Multiple Linear Regression)

多元線性迴歸可以用來解釋多個獨立的變數間的關係，或預測變數間的關係。其中依變項於模型中由多個自變項與相應係數與常數項的函數計算而成。

模型假設： $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$

其中， $\beta_0$  為截距

$\beta_1, \beta_2 \cdots \beta_k$  為迴歸係數

$\varepsilon$  為誤差項

估計式： $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_k x_k$

且，誤差項需同時滿足三項條件：

(1)常態性(Normality)假設：誤差須為常態分佈，則資料也會呈現常態分佈。

(2)獨立性(Independency)假設：誤差之間需彼此獨立。

(3)變異數同質性(Constant Variance)假設：預測出來的變數需具有相同的變異數。

並且， $R$  值為觀察值與預測值間關係的測量方法， $R^2$  為  $R$  的平方，可用於呈現模型解釋力的程度。

### (二)主成份迴歸 Principal Component Regression

主成份迴歸為多元線性迴歸的另一種方法，以主成份分析法為基礎，從原始變數中找出少數幾個主成份來預測變量間的內部結構，可被用於估計未知的迴歸係數，並且可用於解決變數間共線性問題。

一般的迴歸模型等式為 $Y = XB + e$ 。其中， $Y$  為依變數， $X$  為自變數， $B$  為預估計的迴歸模型係數， $e$  則為殘差。

在最小平方法中，迴歸係數是利用 $\hat{B} = (X'X)^{-1}X'Y$  估計。

而在主成份迴歸中，我們將自變項轉置為主成份。寫為 $X'X = PDP'$  =  $Z'Z$ 。其中， $D$  為 $X'X$ 的特徵值之對角矩陣， $P$  則為 $X'X$ 的特徵向量，並且創造新變數  $Z$  為原變數  $X$  的加權平均。

為使均方誤差少於最小平方法，估計方程式為 $A = (Z'Z)^{-1}Z'Y = D^{-1}Z'Y$ 。

### (三)決策樹 Decision Tree

決策樹是一種預測模型，他是一種自變項與依變項之間的映射關係。

## 肆、研究結果



## 伍、結論

## 參考文獻

1. The Effect of Salary Distribution on Production: An Analysis of Major League Baseball (R.Todd Jewell, David J. Molina, 2004)
2. The relationship between payroll and performance disparity in major league baseball: an alternative measure (Daniel Mizak, Anthony Stair, 2004)
3. Team payroll and team performance in major league baseball: 1985-2002 (Frederick Wiseman, Sangit Chatterjee, 2003)
4. Team Payroll Versus Performance in Professional Sports: Is increased Spending Associated with Greater Success? (Grant Shorin, 2017)
5. Testing Causality Between Team Performance and Payroll The Cases of Major League Baseball and English Soccer(Stephen G. Hall, Andrew Zimbalist, 2013)
6. Predicting Salaries of Major League Baseball Players(Rhonda Magel,Michael Hoffman, 2015)
7. Determinants of Major League Baseball Player Salaries(Tyler Wasserman, 2013)
8. Salary Inequality, Team Success and the Superstar Effect(Philippe Cyrenne, 2014)
9. An Exploration into the Relationship of MLB Player Salary and Performance(Nicholas Dorsey, 2015)
10. ANALYSIS OF SALARY FOR MAJOR LEAGUE BASEBALL PLAYERS(Michael Glenn Hoffman, 2014)

資料來源

<http://article.sapub.org/10.5923.j.sports.20150502.02.html>

[https://surface.syr.edu/cgi/viewcontent.cgi?article=1098&context=honors\\_capstone](https://surface.syr.edu/cgi/viewcontent.cgi?article=1098&context=honors_capstone)

<http://economics.uwinnipeg.ca/RePEc/winwop/2014-02.pdf>

[https://pdfs.semanticscholar.org/e11f/e9cfd092b52de3ded392be807f02d06dc37a.p  
df](https://pdfs.semanticscholar.org/e11f/e9cfd092b52de3ded392be807f02d06dc37a.pdf)

[https://library.ndsu.edu/ir/bitstream/handle/10365/27563/Analysis%20of%20Salary  
%20for%20Major%20League%20Baseball%20Players.pdf?sequence=1&isAllowed=y](https://library.ndsu.edu/ir/bitstream/handle/10365/27563/Analysis%20of%20Salary%20for%20Major%20League%20Baseball%20Players.pdf?sequence=1&isAllowed=y)