

美國兩大職業運動聯盟非傳統數據 對於比賽結果影響之分析

0812255 應數 12 曾昱仁

0812227 應數 12 謝翊庭

0812213 應數 12 江皓霖

目 次

一、前言	03
(一) 研究背景與動機	03
(二) 美國職業棒籃球比賽規則簡介及賽制說明	04
二、假說發展	05
三、研究方法	06
(一) 研究資料定義與來源	06
(二) 研究變數	07
四、結果分析	08
五、結論	17
六、參考資料	18

一、前言

(一)研究背景與動機

前陣子棒球經典賽剛落幕，美國職業棒球大聯盟(MLB)也隨即點燃戰火，從中我們發現美國隊、多明尼加隊的成員都是由許多頂薪的球員組成，美國隊成功晉級冠軍戰，反觀多明尼加在小組賽遺憾出局，另外薪資並沒有這麼可觀的日本隊，卻成功捧起冠軍金盃，因此我們想探討在 MLB 的球隊薪資是不是對季賽勝率也有影響。

另外一個點我們發現，今年的光芒隊到目前為止打出了讓所有人都意想不到的好成績，現在的勝率甚至比 MLB 死球年代後單季最多勝的 2001 年水手隊還要高。不過，其實他們今年的團隊薪資可是位在聯盟中的倒數幾名，隊上也沒有任何一名球員的薪資能夠排在 MLB 前 15 名。相較於光芒隊，大都會隊今年到目前的戰績就顯得讓人失望透頂。雖然他們仍然有五成的勝率，不過他們的團隊薪資可是全聯盟最高，薪水前 15 名的球員也佔了三位，卻無法把雄厚的資金實力轉換成場上的好成績。於是，我們對 MLB 隊伍中是否擁有前 15 名高薪的選手跟能不能把隊伍帶進季後賽的關係感到十分好奇。

為了觀察是否其他團隊運動也會有相同趨勢，我們加入美國職業籃球聯賽(NBA)一同分析。跟棒球相比，籃球在個人的表現方面能對球隊有更高的貢獻影響力，所以各個球隊主要都是圍繞著當家球星來設計戰術並給予更多上場時間與發揮空間，但是相對來說歷史上也有不少球隊是透過團隊實力來彌補自己缺乏球星的劣勢，舉例來說，今年的邁阿密熱火隊，先發有三位是低順位甚至落選秀，但他們卻以東區第八的身分一路闖進東區冠軍賽，因此我們想看看究竟組巨頭能有更高勝率，還是其實薪資對於勝率的影響其實並不大。

從 MLB 與 NBA 中，我們發現有不少地區是同時擁有 NBA 及 MLB 球隊的，以洛杉磯而言，NBA 方面就有快艇隊與湖人隊，MLB 球隊是道奇隊及天使隊，那我們想研究看看說一個地區的人口數會不會也對勝率有影響，因為球迷進場越多、周邊賣得越好，相對來說球隊就有更多資金能去訓練球員，進而提高勝率。

(二)美國職業棒籃球比賽規則簡介及賽制說明

NBA 比賽規則和賽制的簡介：

一場 NBA 比賽分為四節，每節 12 分鐘，共 48 分鐘。如果比賽在正規時間內出現平局，則進行加時賽。每支 NBA 球隊由 12 名球員組成，其中 5 名球員可同時上場比賽。在季後賽部分，東區與西區的前 8 名球隊將晉級季後賽，並以淘汰制進行比賽，直到決出東區和西區的冠軍。最後，兩區的冠軍球隊將進行總決賽，爭奪 NBA 總冠軍。額外說明 NBA 季後賽一些改制，自 2020-2021 賽季起，新增附加賽賽制，第 7 種子由第七名及第八名球隊對決中勝方獲得，第 8 種子則由第七名及第八名球隊對決中負方與第九名及第十名球隊對決中勝方之間爭奪，每場均以一場定勝負形式進行，故第 8 種子有可能為各區第九名甚至更差的第十名，此方法增加了在季中季末異軍突起的球隊晉級機會。

以下說明 NBA 和 FIBA 的主要差別，比賽時間方面 NBA 為四節，每節 12 分鐘；而 FIBA 為四節，每節 10 分鐘。三分線位置：在 NBA（約 7.24 米）中，三分線距離籃框的距離比 FIBA（約 6.75 米）更遠，因此有些球員在 NBA 表現稍微遜色，在國際賽中採用的 FIBA 規則反而打出身價。裁判判決和視頻回放方面，在 NBA 中，使用視頻回放系統來審查部分判決，例如是否成功投進三分球或是否在犯規發生時有出界。而 FIBA 的視頻回放系統則相對較少使用，裁判主要依賴於實地裁判判決。

MLB 比賽規則和賽制的簡介：

一場正規的 MLB 比賽總共有 9 局，每局由兩支球隊輪流進行守備和進攻。如果比賽在 9 局結束時出現平局，將進入額外的延長賽。而每支球隊開季可以登錄 26 名球員，包括先發投手、後援投手、捕手、內野手及外野手。至於季後賽的部分，最新的賽制從 2022 賽季開始啟用。一樣三個分區冠軍隊不變外，增加成三隊外卡，外卡系列賽採取三戰兩勝制，並由三個分區冠軍戰績較差的隊伍與外卡三隊分兩組競爭，勝者晉級分區系列賽。分區系列賽是採取五戰三勝制，由戰績第一及第二的分區冠軍對上兩支外卡系列賽的勝隊。並且，分區冠軍擁有主場優勢，獲勝者則晉級到聯盟冠軍賽。而兩聯盟（美國聯盟和國家聯盟）冠軍賽之勝方將晉級最後的「世界大賽」，先拿到四勝的球隊即成為 MLB 當年度的總冠軍。

以下解釋 MLB 和 WBC 之間的主要差異：

MLB 是美國的職業棒球聯盟，由美國和加拿大的球隊組成。而 WBC 是國際棒球比賽，各國代表隊取得資格參賽，球員都來自不同國家；MLB 的賽季通常從四月持續到十月，球隊之間進行多場比賽，共計 162 場的例行賽。WBC 則是每四年舉辦一次，比賽時程相對較緊湊；在 MLB 中，球員沒有國籍的限制。但在 WBC 裡，球員必須擁有適當的國籍資格才能代表所屬國家參賽；MLB 作為一個職業棒球聯盟，擁有廣大的球迷基礎和商業影響力。WBC 則更側重於國際間的棒球競技，也受到參賽國家的極高關注與球迷支持。

二、假說發展

NBA 及 MLB 的球隊都有發展循環週期，會經過重建 > 培育自家球員 > 再補足角色或明星球員來完成球隊拼圖 > 嘗試衝擊總冠軍 > 失利後再次重建，在這個週期中，球隊總薪資在補足球員階段薪資會開始大幅上升，大部分球隊戰績也會相對有所提升。另外，我們猜測大城市的球隊因為球迷消費力較強或是球迷數量眾多，球隊有更多錢培養球員或是簽大咖球星，讓勝率能提高。綜合以上所述，我們提出假說一。

假說一：在兩聯盟中季賽勝率的高低跟所屬城市人口以及球隊總薪資高低皆呈正相關，且薪資之相關性**更高**。

除此之外，在 WBC 世界棒球經典賽時，在滿場的球迷的支持下，台灣打出了可圈可點的成績，相信也出乎了許多人的意料，所以我們認為進場的觀眾人數也有機會提高勝率，根據上述我們的猜想，我們提出假說二。

假說二：在兩聯盟中季賽勝率的高低跟進場的觀眾人數比勝率的高低跟所屬城市人口之相關性**更高**。

另外，球隊的若是總收入較其他球隊來的高，則可以將此收入轉化成對球隊的投資，如增加球員的總量、升級球場訓練設備、提升訓練品質等。根據以上的條件，我們提出假說三。

假說三：在兩聯盟中季賽勝率的高低跟各隊的總收入比例(門票收入、周邊收入等)呈**正相關**。

MLB 一隊球員數為 26 人，NBA 一隊為 12 人，在人數上 MLB 一隊大約是 NBA 的兩倍。籃球在調度上，教練基本上只會運用 7~8 人來面對比賽，大部分球員的實力和薪水呈現正相關，因此較高的總薪資一定會對球隊勝率有所提升。反觀棒球而言，一場比賽就至少需要 9 個人，還不包括比賽中後段上來中繼後援的投手群，以及關鍵時刻一些功能性的代打代守代跑等。也就是說，棒球相比籃球是更注重團隊的，每個球員每場表現的波動對於球隊勝負的結果產生了更高的不確定性。綜合以上所述，我們提出假說四。

假說四：在薪資經過標準化下，NBA 的季賽勝率受球隊總薪資影響程度**高於** MLB。

球星的價值就是在關鍵時刻能跳出來為球隊建功。以調度面來說，在 NBA，關鍵時刻球的流動到出手基本上都會教給球星親自去操刀，在防守或進攻下，球星在大量的球權及上場時間，也相對能創造更高的效率值，讓教練在調度上能更放心，因此在一個 5v5

的籃球賽，是否擁有球星對勝率上會有更直接關係；而在 MLB 的比賽裡，投、打、守三方面缺一不可，球隊內就算擁有頂級的王牌投手，若是無法把足夠多的分數打回來，球隊是無法贏球的；同樣地，如果隊上有眾星雲集的打者群，可是牛棚一直把比賽搞砸的話，一樣沒辦法取得比賽的勝利。綜合以上所述，我們提出假說五。

假說五：在 NBA 中，擁有薪資排名前 15 名球員的球隊晉級季後賽的機率高於其他球隊；但相反地，在 MLB 裡，是否擁有薪資排名前 15 名球員的球隊與晉級季後賽的機率會接近無相關性。

在職業運動領域中，教練在球隊的成功與否扮演著至關重要的角色。他們的領導能力、戰術安排和球員管理能力都可以直接影響球隊的表現。隨著教練在聯賽中的歷練增加，他們累積了豐富的比賽經驗和適應不同情況的能力。這些經驗使得教練能夠更好地分析球場上的局勢，做出明智的決策。他們對球隊成員的特點和能力有更深入的了解，能夠更好地調整陣容、制定戰術，以及在關鍵時刻做出正確的決定。綜合以上所述，我們提出假說六。

假說六：在 MLB 和 NBA 中，教練的帶隊的經驗越多會間接影響到球隊整體勝率。

三、研究方法

（一）研究資料定義與來源

我們選定 2000~2022 的資料，因為 NBA 賽季都會跨年度，MLB 則否，所以我們統一以起始年度作為基準，舉例來說，NBA 的 00~01 賽季資料我們會設定成西元 2000 年的資料，如此一來沒有時間的問題就能跟 MLB 相互討論。

薪水部分，NBA 與 MLB 各 30 支球隊，球隊總薪水以及球員薪資排名，NBA 採用的是球迷屋的資料各有 690 組資料，MLB 採用的是 Steve O's Baseball Umpire Resources 的資料也是各有 690 組資料。

球隊勝率與晉級季後賽部分，NBA 在球隊勝率是採用 TeamRankings 的資料，晉級季後賽分別是維基百科在 2000~2022 年的資料，總共也都是各 690 組資料；MLB 的資料皆是從 MLB 官網取得，都是各 690 組資料。

人口的部分我們是取得 2000、2010、2020、2022 年美國各城市的人口數，資料來源是 Biggest US Cities，有些球隊不在百大或是不在美國的城市，我們就是收集自維基百科，總共 1380 筆資料。觀眾人數是從 ESPN 網站上蒐集 2000~2022 年的資料。

NBA 總收入的部分我們是取得 2001 至 2021 年各球隊的總收入，而 MLB 總收入的部分則是取得 2001 年至 2022 年的各球隊總收入，資料來源為 Statista。

NBA 和 MLB 的教練年資部分，我們是從 Basketball reference 和 Baseball reference 網站蒐集 2000 至 2022 年的資料。

（二）研究變數

1. 球隊所屬城市人口 (POPULATION)

我們將此變數設定為自變數，使用上面提到之四個年度的人口並運用內插法得到其餘年度城市人口。

2. 觀眾人數 (ATTENDANCE)

我們將此變數設定為自變數。

3. 各隊總收入 (REVENUE_N)

我們將此變數設定為自變數，且將各球隊該年度的總收入經過正規化之處理。

4. 球隊總薪資 (WAGE)

我們將此變數設定為自變數，且把各球隊該年度的總薪資經過標準化之處理。

5. 是否擁有薪資排名前五名的球員 (SALARY)

我們將此變數設定為二元的自變數，若球隊擁有任何一名薪資排名前五名的球員，則變數值設為 1。反之，將設為 0。

6. 季賽勝率 (WIN RATE)

我們將此變數設定為應變數，而勝率的取值範圍從 0 到 1。

7. 晉級季後賽與否 (PLAYOFF)

我們將此變數設定為二元的應變數，若球隊有晉級該年度的季後賽，則變數值設為 1。反之，將設為 0。

8. 球隊教練年資 (COACH_EXP)

我們將此變數設定為自變數，以年作為單位去做計算。

四、結果分析

假說一：在兩聯盟中季賽勝率的高低跟所屬城市人口以及球隊總薪資高低皆呈正相關，且薪資之相關性**更高**。

```
Call:
lm(formula = WIN.RATE ~ WAGE + POPULATION, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.35769 -0.09287  0.01026  0.09469  0.32726

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.088e-01  6.478e-03  78.548  < 2e-16 ***
WAGE         5.305e-02  5.317e-03   9.978  < 2e-16 ***
POPULATION  -1.136e-08  2.813e-09  -4.038  5.99e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1327 on 687 degrees of freedom
Multiple R-squared:  0.1291,    Adjusted R-squared:  0.1266
F-statistic: 50.94 on 2 and 687 DF,  p-value: < 2.2e-16
```

表一 NBA 季賽勝率跟總薪資及人口之複迴歸分析結果

```
Call:
lm(formula = WIN.RATE ~ WAGE + POPULATION, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.213930 -0.051277  0.001096  0.051940  0.205413

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.005e-01  3.471e-03 144.211  <2e-16 ***
WAGE         2.886e-02  3.066e-03   9.411  <2e-16 ***
POPULATION  -3.650e-10  1.486e-09  -0.246   0.806
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06969 on 687 degrees of freedom
Multiple R-squared:  0.1414,    Adjusted R-squared:  0.1389
F-statistic: 56.59 on 2 and 687 DF,  p-value: < 2.2e-16
```

表二 MLB 季賽勝率跟總薪資及人口之複迴歸分析結果

1. 兩者迴歸模型之 Adjusted R-squared 的值差不多，不過都偏低。
2. 兩者總薪資的相關係數皆為正，並且都具有統計顯著性，在 95%的信心水準下。
3. NBA 總薪資的相關係數比 MLB 總薪資的相關係數還要高。

4. 兩者人口的相關係數都非常接近 0，代表人口對於勝率在兩聯盟中都幾乎沒有影響。

從上面的分析結果可得知，假說一裡人口對球隊的季賽勝率相關程度與我們的預期結果不一致。

假說二：在兩聯盟中季賽勝率的高低跟進場的觀眾人數比勝率的高低跟所屬城市人口之相關性更高。

```
Call:
lm(formula = WIN.RATE ~ ATTENDANCE, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.38441 -0.10141  0.00702  0.10213  0.32763

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.061e-01  2.145e-02  18.935  < 2e-16 ***
ATTENDANCE   1.298e-07  3.121e-08   4.159  3.6e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1403 on 688 degrees of freedom
Multiple R-squared:  0.02452, Adjusted R-squared:  0.02311
F-statistic: 17.3 on 1 and 688 DF, p-value: 3.601e-05

Call:
lm(formula = WIN.RATE ~ POPULATION, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.38934 -0.09467  0.00866  0.10315  0.34130

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.984e-01  6.836e-03  72.911  <2e-16 ***
POPULATION  -4.146e-09  2.907e-09  -1.427   0.154
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1419 on 688 degrees of freedom
Multiple R-squared:  0.002949, Adjusted R-squared:  0.0015
F-statistic: 2.035 on 1 and 688 DF, p-value: 0.1542
```

表三 NBA 勝率跟觀眾人數與勝率跟人口數之迴歸分析結果

```
Call:
lm(formula = WIN.RATE ~ ATTENDANCE, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.205047 -0.052191 -0.000328  0.046587  0.291913

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.251e-01  7.593e-03   55.98  <2e-16 ***
ATTENDANCE   3.286e-08  3.121e-09  10.53  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06975 on 688 degrees of freedom
Multiple R-squared:  0.1388, Adjusted R-squared:  0.1375
F-statistic: 110.9 on 1 and 688 DF, p-value: < 2.2e-16

Call:
lm(formula = WIN.RATE ~ POPULATION, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.230936 -0.056468  0.001375  0.056607  0.222042

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.903e-01  3.499e-03  140.121  < 2e-16 ***
POPULATION   6.438e-09  1.378e-09   4.672  3.59e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07399 on 688 degrees of freedom
Multiple R-squared:  0.03075, Adjusted R-squared:  0.02934
F-statistic: 21.83 on 1 and 688 DF, p-value: 3.592e-06
```

表四 MLB 勝率跟觀眾人數與勝率跟人口數之迴歸分析結果

季賽勝率跟觀眾人數的迴歸模型分析

1. MLB 跟觀眾人數迴歸模型之 Adjusted R-squared 的值高於其他三者，不過偏低，而其他三者之 Adjusted R-squared 的值都接近 0，也代表著模型幾乎沒有任何解釋能力。

2. 兩者的相關係數都非常接近 0，代表觀眾人數對於勝率在兩聯盟中都幾乎沒有影響。

季賽勝率跟人口的迴歸模型分析

1. 兩者迴歸模型之 Adjusted R-squared 的值都離 0 很近，也就是說模型幾乎沒有任何解釋能力。

2. 兩者人口的相關係數都非常接近 0，代表人口對於勝率在兩聯盟中都幾乎沒有影響。

從上面的分析結果可得知，進場的觀眾人數對於勝率的相關程度在相關係數方面雖高於人口對於勝率的相關程度，但相關係數仍不高，與我們預測不一致。

假說三：在兩聯盟中季賽勝率的高低跟各隊的總收入比例(門票收入、周邊收入等)呈正相關。

```
Call:
lm(formula = WIN.RATE ~ REVENUE_N, data = NBA_DATA)

Residuals:
    Min       1Q   Median       3Q      Max
-0.37210 -0.10499  0.01314  0.10314  0.31216

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.456623   0.008755  52.155  < 2e-16 ***
REVENUE_N    0.122481   0.023900   5.125 4.05e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1396 on 585 degrees of freedom
Multiple R-squared:  0.04296,    Adjusted R-squared:  0.04133
F-statistic: 26.26 on 1 and 585 DF,  p-value: 4.054e-07
```

表五 NBA 勝率跟正規化的各隊收入之迴歸分析結果

```
Call:
lm(formula = WIN.RATE ~ REVENUE_N, data = MLB_DATA)

Residuals:
    Min       1Q   Median       3Q      Max
-0.230729 -0.054738  0.007179  0.050453  0.191351

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.468649   0.004866  96.316  < 2e-16 ***
REVENUE_N    0.118099   0.014145   8.349 7.34e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07127 on 482 degrees of freedom
Multiple R-squared:  0.1264,    Adjusted R-squared:  0.1245
F-statistic: 69.71 on 1 and 482 DF,  p-value: 7.337e-16
```

表六 MLB 勝率跟正規化的各隊收入之迴歸分析結果

1. 兩者總薪資的相關係數皆為正，並且都具有統計顯著性，在 95%的信心水準下。
2. NBA 各隊收入的相關係數與 MLB 各隊收入的相關係數兩者數值差不多。
3. 兩者迴歸模型之 Adjusted R-squared 的值以 MLB 較 NBA 的來的多上一些，不過整體差不多，兩者數值都偏低。

從上面的分析結果可得知，假說三的各隊的總收入比例對球隊的季賽勝率具有些許相關性，然而可以從偏小的 Adjusted R-squared 得出此變數對於勝率的影響依然不彰。

假說四：在薪資經過標準化下，NBA 的季賽勝率受球隊總薪資影響程度高於 MLB。

```
Call:
lm(formula = WIN.RATE ~ WAGE, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.39047 -0.09424  0.01276  0.09713  0.33956

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.492459    0.005107  96.419  <2e-16 ***
WAGE         0.047529    0.005195   9.149  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1342 on 688 degrees of freedom
Multiple R-squared:  0.1085,    Adjusted R-squared:  0.1072
F-statistic: 83.71 on 1 and 688 DF,  p-value: < 2.2e-16
```

表七 NBA 勝率跟總薪資之迴歸分析結果

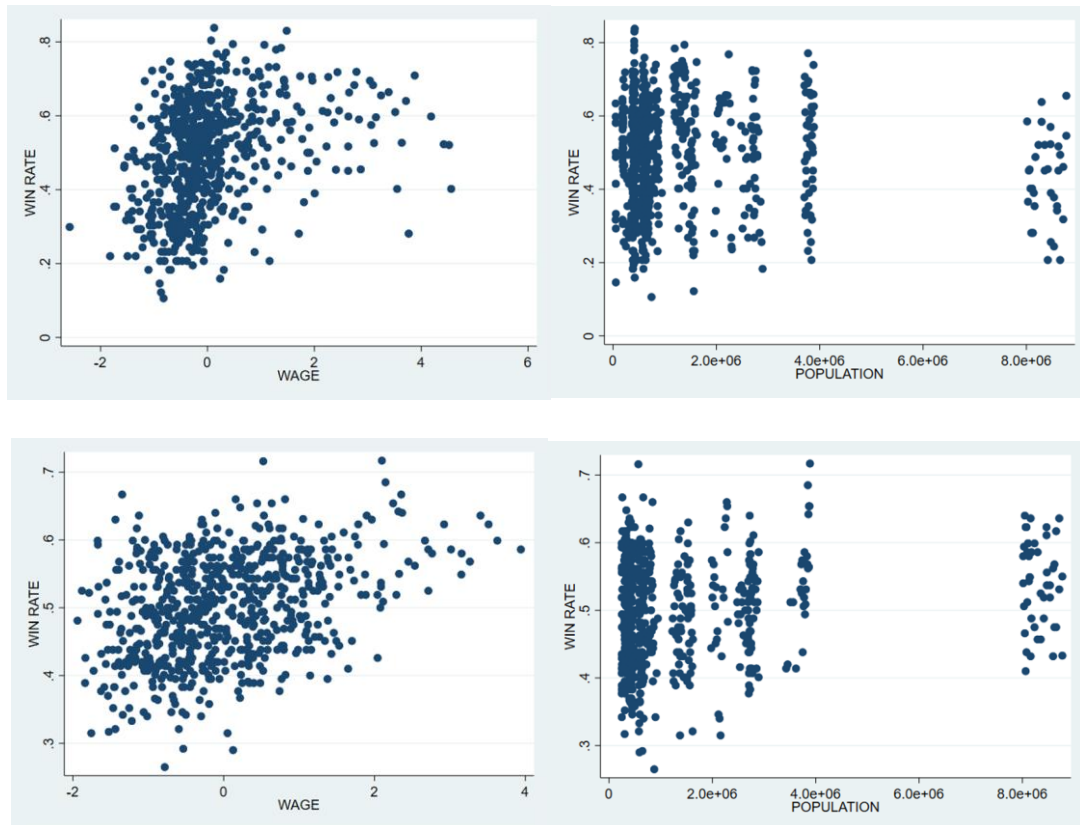
```
Call:
lm(formula = WIN.RATE ~ WAGE, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.213550 -0.051340  0.001007  0.051659  0.205375

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.499998    0.002651 188.59  <2e-16 ***
WAGE         0.028490    0.002677  10.64  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06964 on 688 degrees of freedom
Multiple R-squared:  0.1414,    Adjusted R-squared:  0.1401
F-statistic: 113.3 on 1 and 688 DF,  p-value: < 2.2e-16
```

表八 MLB 勝率跟總薪資之迴歸分析結果



圖一(左上)NBA 勝率與總薪資數據分配點圖 圖二(右上)NBA 勝率與人口數據分配點圖
圖三(左下)MLB 勝率與總薪資數據分配點圖 圖四(右下)MLB 勝率與人口數據分配點圖

季賽勝率跟總薪資的迴歸模型分析

1. 兩者迴歸模型之 Adjusted R-squared 的值差不多，不過都偏低。
2. 兩者總薪資的相關係數皆為正，並且都具有統計顯著性，在 95%的信心水準下。
3. NBA 總薪資的相關係數比 MLB 總薪資的相關係數還要高。

點圖之視覺化分析

1. MLB 總薪資資料的變異程度比 NBA 總薪資資料的變異程度來的高。
2. NBA 季賽勝率資料的變異程度比 MLB 季賽勝率資料的變異程度來的高。
3. 兩聯盟個別球隊所屬的城市人口資料的相似性很高。

從上面的分析結果可得知，假說四的數據結果有符合預期。

假說五：在 NBA 中，擁有薪資排名前 15 名球員的球隊晉級季後賽的機率高於其他球隊；但相反地，在 MLB 裡，是否擁有薪資排名前 15 名球員的球隊與晉級季後賽的機率會接近無相關性。

(一)使用 Logistic 迴歸模型

```
. logistic playoff wage population i.salary
```

Logistic regression		Number of obs = 690	
		LR chi2(3) = 69.76	
		Prob > chi2 = 0.0000	
Log likelihood = -442.97266		Pseudo R2 = 0.0730	

playoff	Odds ratio	Std. err.	z	P> z	[95% conf. interval]	
wage	1.793097	.2044414	5.12	0.000	1.434014	2.242096
population	.9999999	4.82e-08	-2.92	0.004	.9999998	1
i.salary	1.719471	.3203133	2.91	0.004	1.193515	2.477204
_cons	1.09822	.137685	0.75	0.455	.8589612	1.404124

Note: _cons estimates baseline odds.

表九 NBA 是否晉級季後賽跟總薪資、人口及薪資排名前 15 名之 Logistic 迴歸分析結果

```
. logistic playoff wage population i.salary
```

Logistic regression		Number of obs = 690	
		LR chi2(3) = 54.99	
		Prob > chi2 = 0.0000	
Log likelihood = -399.76236		Pseudo R2 = 0.0644	

playoff	Odds ratio	Std. err.	z	P> z	[95% conf. interval]	
wage	2.062201	.2431846	6.14	0.000	1.63664	2.598416
population	1	4.61e-08	-0.24	0.808	.9999999	1
i.salary	.7108482	.1566983	-1.55	0.122	.4614656	1.095001
_cons	.4821648	.0607354	-5.79	0.000	.3766819	.6171862

Note: _cons estimates baseline odds.

表十 MLB 是否晉級季後賽跟總薪資、人口及薪資排名前 15 名之 Logistic 迴歸分析結果

Logistic 迴歸模型分析

1. 兩者迴歸模型之 Pseudo R-squared 的值差不多，不過都偏低。
2. NBA 薪資排名前 15 名的相關係數為正，並且具有統計顯著性，在 95%的信心水準下。
3. MLB 薪資排名前 15 名的相關係數為負，而且沒有統計顯著性，在 95%的信心水準下。

4. NBA 薪資排名前 15 名的相關係數絕對值大於 MLB 薪資排名前 15 名的相關係數絕對值。

勝算比數值分析

MLB 總薪資勝算比 > NBA 總薪資勝算比 > NBA 薪資排名前 15 勝算比 > MLB 人口勝算比
 $= 1 \approx$ NBA 人口勝算比 > MLB 薪資排名前 15 勝算比

(二)使用 Probit 迴歸模型

```
. probit playoff wage population i.salary
```

```
Iteration 0: log likelihood = -477.85408
Iteration 1: log likelihood = -443.29767
Iteration 2: log likelihood = -443.21229
Iteration 3: log likelihood = -443.21229
```

```
Probit regression                                Number of obs =   690
                                                LR chi2(3)      =   69.28
                                                Prob > chi2     =  0.0000
Log likelihood = -443.21229                    Pseudo R2      =  0.0725
```

playoff	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
wage	.3441432	.0650484	5.29	0.000	.2166508	.4716357
population	-8.57e-08	2.80e-08	-3.06	0.002	-1.41e-07	-3.09e-08
i.salary	.3412867	.1149374	2.97	0.003	.1160136	.5665599
_cons	.0491091	.0767859	0.64	0.522	-.1013885	.1996067

```
. logistic playoff wage population i.salary
```

```
Logistic regression                                Number of obs =   690
                                                LR chi2(3)      =   69.76
                                                Prob > chi2     =  0.0000
Log likelihood = -442.97266                    Pseudo R2      =  0.0730
```

playoff	Odds ratio	Std. err.	z	P> z	[95% conf. interval]	
wage	1.793097	.2044414	5.12	0.000	1.434014	2.242096
population	.9999999	4.82e-08	-2.92	0.004	.9999998	1
i.salary	1.719471	.3203133	2.91	0.004	1.193515	2.477204
_cons	1.09822	.137685	0.75	0.455	.8589612	1.404124

Note: _cons estimates baseline odds.

表十一 NBA 是否晉級季後賽跟總薪資、人口及薪資排名前 15 名之 Probit 迴歸分析結果


```
. probit playoff wage population i.salary
```

```
Iteration 0: log likelihood = -427.25944
Iteration 1: log likelihood = -399.87345
Iteration 2: log likelihood = -399.78536
Iteration 3: log likelihood = -399.78536
```

Probit regression

Number of obs = 690
LR chi2(3) = 54.95
Prob > chi2 = 0.0000
Pseudo R2 = 0.0643

Log likelihood = -399.78536

playoff	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
wage	.4379143	.0700056	6.26	0.000	.3007058	.5751228
population	-6.20e-09	2.80e-08	-0.22	0.825	-6.12e-08	4.88e-08
i.salary	-.2084918	.1330583	-1.57	0.117	-.4692813	.0522978
_cons	-.4451036	.0761976	-5.84	0.000	-.5944483	-.295759

```
. logistic playoff wage population i.salary
```

Logistic regression

Number of obs = 690
LR chi2(3) = 54.99
Prob > chi2 = 0.0000
Pseudo R2 = 0.0644

Log likelihood = -399.76236

playoff	Odds ratio	Std. err.	z	P> z	[95% conf. interval]	
wage	2.062201	.2431846	6.14	0.000	1.63664	2.598416
population	1	4.61e-08	-0.24	0.808	.9999999	1
i.salary	.7108482	.1566983	-1.55	0.122	.4614656	1.095001
_cons	.4821648	.0607354	-5.79	0.000	.3766819	.6171862

Note: _cons estimates baseline odds.

表十二 MLB 是否晉級季後賽跟總薪資、人口及薪資排名前 15 名之 Probit 迴歸分析結果

1. 兩者迴歸模型之 Pseudo R-squared 的值差不多，不過都偏低。
2. NBA 薪資排名前 15 名的相關係數為正，並且具有統計顯著性，在 95%的信心水準下。
3. MLB 薪資排名前 15 名的相關係數為負，而且沒有統計顯著性，在 95%的信心水準下。
4. NBA 薪資排名前 15 名的相關係數絕對值大於 MLB 薪資排名前 15 名的相關係數絕對值。
5. Logistic 迴歸模型 與 Probit 迴歸模型 跑出來的分析結果一致性很高。

從上面的分析結果可得知，假說五的數據結果也有符合預期。

假說六：在 MLB 和 NBA 中，教練的帶隊的經驗越多會間接影響到球隊整體勝率。

```
Call:
lm(formula = WIN.RATE ~ COACH_EXP, data = 統計數據...NBA)

Residuals:
    Min       1Q   Median       3Q      Max
-0.38048 -0.09583  0.01204  0.09871  0.35784

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.467511   0.007643  61.170 < 2e-16 ***
COACH_EXP    0.006324   0.001389   4.554 6.21e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.14 on 688 degrees of freedom
Multiple R-squared:  0.02927,    Adjusted R-squared:  0.02786
F-statistic: 20.74 on 1 and 688 DF,  p-value: 6.215e-06
```

表十三 MLB 勝率與教練年資之迴歸分析結果

```
Call:
lm(formula = WIN.RATE ~ COACH_EXP, data = 統計數據...MLB)

Residuals:
    Min       1Q   Median       3Q      Max
-0.225713 -0.054701  0.001863  0.056863  0.214759

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.4854008   0.0043640 111.227 < 2e-16 ***
COACH_EXP    0.0033680   0.0007682   4.384 1.35e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07413 on 688 degrees of freedom
Multiple R-squared:  0.02718,    Adjusted R-squared:  0.02577
F-statistic: 19.22 on 1 and 688 DF,  p-value: 1.345e-05
```

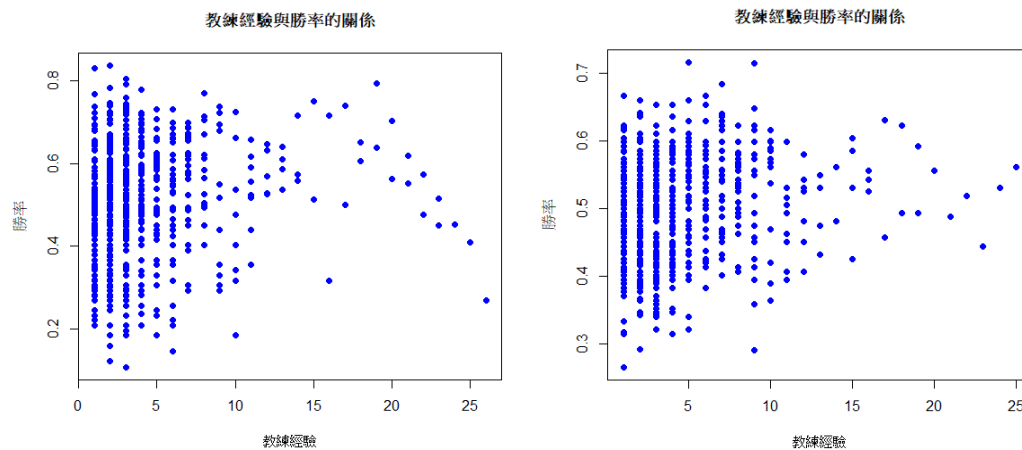
表十四 MLB 勝率與教練年資之迴歸分析結果

Logistic 迴歸模型分析

1. 教練經驗對 MLB 和 NBA 的勝率有顯著的正向影響：根據模型的係數估計，每增加一個單位的教練經驗，預測的勝率將分別增加 0.0033680、0.006324。這意味著具有較豐富的教練經驗的球隊在一定程度上更有可能取得較高的勝率。

2. 模型的解釋力有限：兩者的多重 R 平方值分別為 0.02718 和 0.02927，說明模型能夠解釋目標變數（勝率）的約 3% 的變異性。這意味著還有其他未考慮的因素對於解釋勝率是必要的。

3. 統計顯著性：根據 t 值和 p 值，在 95%的信心水準下，教練經驗的係數在統計上是顯著的。這意味著教練經驗的影響不太可能是由於隨機變異引起的，而是真正與勝率之間存在關聯。



圖五(左)NBA 勝率與教練年資分配點圖 圖六(右)MLB 勝率與教練年資分配點圖

點圖之視覺化分析

1. MLB 教練年資的變異程度比 NBA 教練年資的變異程度較來的高。
2. MLB 教練年資的分佈與 NBA 教練年資的分佈相似。

五、結論

我們發現上述迴歸模型中, Adjusted R-squared 和 Pseudo R-squared 的值都偏低, 代表模型中可被解釋的變異程度較低。我們覺得是因為有其他影響勝負的關鍵因素未被我們納入模型中, 或者是因為蒐集到的數據集裡, 有一些 Outlier(離群值)導致模型的配適度較低。而且, 其實不管是 NBA 還是 MLB, 每支球隊在常規季賽中跟其他球隊的對戰次數不盡相同, 另外進到了季後賽不同分區競爭度通常也不太一樣。這告訴了我們影響球隊勝負方面的表現之不可控因素太多, 使得無法單純用一兩個變因去很好地解釋迴歸後的結果。

六、參考資料

NBA 球員及球隊薪水:

<https://www.giumiwu.com/standings/salaries/1/2022>

MLB 球員及球隊薪水:

<http://www.stevetheump.com/>

NBA&MLB 球隊所屬城市人口:

<https://www.biggestuscities.com/>

NBA&MLB 球隊各隊收入:

<https://www.statista.com/>

NBA 晉級季後賽球隊(以 00~01 賽季為例):

<https://zh.wikipedia.org/zh-tw/2001%E5%B9%B4NBA%E5%AD%A3%E5%90%8E%E8%B5%9B>

NBA 季賽勝率:

<https://www.teamrankings.com/nba/stat/win-pct-all-games>

MLB 晉級季後賽球隊/季賽勝率:

<https://www.mlb.com/>

NBA 教練年資:

[Basketball Statistics & History of Every Team & NBA and WNBA Players | Basketball-Reference.com](#)

MLB 教練年資:

[MLB Stats, Scores, History, & Records | Baseball-Reference.com](#)