

# 王柏融逐球資料分析

作者:林聖皓

日期:2020.6

## 目錄

一、 動機 .....	1
1 動機說明 .....	1
2 資料來源 .....	2
二、 資料蒐集 .....	2
三、 數據分析 .....	4
1 打擊熱區 .....	4
2 壞球分析與本壘板紀律 .....	6
3 決策樹分析 .....	10
四、 結論 .....	15
1 分析出的可能結果 .....	15
2 分析的缺點 .....	15

## 一、動機

### 1 動機說明

由於我本身就是王柏融的球迷，綜觀他最近幾年的職業成績，可以發現在中職的成績與日職時打出的成績有很大的落差。我想藉由逐球數據來觀察，究竟王柏融的弱點是在哪邊，又或是因為聯盟轉換的落差才導致他在日本職棒打不出理想的成績。

攤開完整年度打擊成績來看，王柏融是近年在中華職棒打擊成績最優秀的選手：

	OPS+	WRC+
2016	172.1	175.6
2017	196.6	186.4
2018	154.6	151.1

圖表 1 王柏融年度 OPS+與 WRC+

2017 年打出 OPS+:196.6 、WRC+:186.4 的成績，是其生涯最高的一年。之後 2018 年的 OPS+:154.6、WRC+:151.1，雖然這樣的成績仍然相當優秀，但與其前一年相比有小幅落差。後來 2019 於日本職棒的打擊 OPS 為 0.647，顯見聯盟轉換對其打擊成績影響。也有可能是在 2018 時，就能找出打擊成績下修的原因。

為了要分析詳細的因素，我將 2017、2018 年的打擊逐球資料做紀錄，本篇文只對 2017 年做資料分析，最後再與現有的日本職棒網站數據做比較，找出可能遇到的打擊問題。

## 2 資料來源

由於目前國內職棒並沒有公開的逐球資料庫，在這個因素下，我決定自行對王柏融選手於 2017 年的桃園主場逐球打擊紀錄做資料蒐集。會選擇桃園球場做紀錄是因為轉播單位通常會提供參考的打擊框，相對較方便我辨別好壞球的位置。透過 cpbltv 的重播，我逐場逐球用 Access 記錄每一顆投球的結果，最後再用 python 做逐球的資料視覺化處理，解讀不同的打擊結果。

## 二、資料蒐集

使用 Access 軟體記錄逐球數據

game	pa	pitch_count	ball_strike	pitch_speed	pitch_type	pitch_position	swing	result	kzone_edge	pitcher
4	1	1	1-0	143	1	92		ball		布魯斯
4	1	2	2-0	145	1	93		ball		布魯斯
4	1	3	2-1	145	1	4		strike	✓	布魯斯
4	1	4	2-2	133	3	6	✓	foul		布魯斯
4	1	5	2-2	147	1	7	✓	foul	✓	布魯斯
4	1	6	2-2	131	3	9	✓	fo		布魯斯
4	2	1	0-0	140	1	1	✓	go		布魯斯
4	3	1	0-0	146	1	61		ball		陳韻文
4	3	2	1-0	141	1	2	✓	fo		陳韻文
4	4	1	0-0	132	2	8	✓	foul	✓	江承峰
4	4	2	0-1	129	2	91		ball		江承峰
4	4	3	1-1	142	1	33	✓	strike		江承峰
4	4	4	1-2	141	1	22		ball		江承峰
4	4	5	2-2	130	2	91	✓	foul		江承峰
4	4	6	2-2	129	2	91	✓	go		江承峰
6	1	1	0-0	138	1	72		ball		林其緯
6	1	2	1-0	118	4	72		ball		林其緯
6	1	3	2-0	137	1	5	✓	go		林其緯
6	2	1	0-0	137	1	13	✓	foul		林其緯
6	2	2	0-1	108	4	33		ball	✓	林其緯
6	2	3	1-1	139	1	31		strike	✓	林其緯
6	2	4	1-2	129	2	93		ball		林其緯
6	2	5	2-2	122	3	4	✓	lb	✓	林其緯
6	3	1	0-0	130	1	5	✓	lb		林威志
6	4	1	0-0	140	1	91		strike	✓	王鐘銘

由左而右的 columns 分別代表

1.game：比賽場次

2.pa：該比賽的第 n 次打席

3.pitch\_count：該打席的第 n 顆球

4.ball\_strike：好壞球比，左邊為壞球右邊為好球

5.pitch\_speed：該投球球速

**6.pitch\_type**：球種分析

(1 代表直球，2 代表變速球或指叉球，3 代表滑球或卡特球，4 代表曲球)

**7.pitch\_position**：投球進壘點

(總共分為 25 個位置，中心九個為好球帶的九宮格，其餘位置代表如下圖所示，  
視角為投手看向打者的方向)

**8.swing**：判斷是否出棒，以布林值儲存

**9.result**：該球的擊球結果

**10.kzone\_edge**：判斷是否在九宮格邊緣一顆球內，以布林值儲存

**11.pitcher**：對戰投手

12	13	23	33	32
11	1	2	3	31
41	4	5	6	61
71	7	8	9	91
72	73	83	93	92

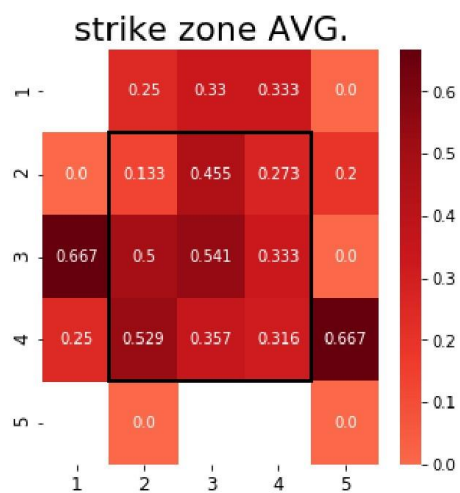
圖表 2 好球帶標號，紅色為好球區

### 三、數據分析

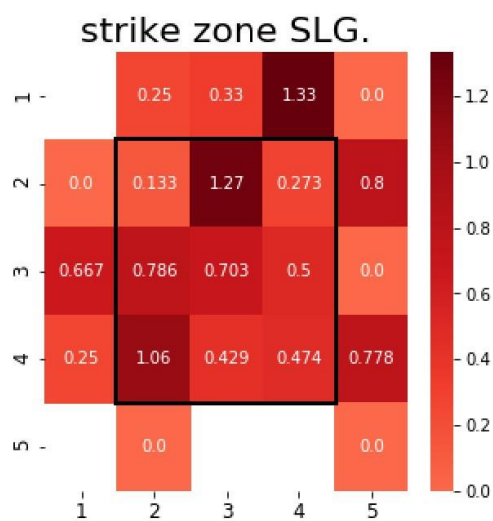
#### 1 打擊熱區

打擊熱區是一個能直觀上看出打擊者擅長與弱點的打擊位置圖。

我利用手邊的資料，使用 python 的 seaborn 套件內的 heatmap 為王柏融在 2017 年的打擊資料做出打擊熱區圖。

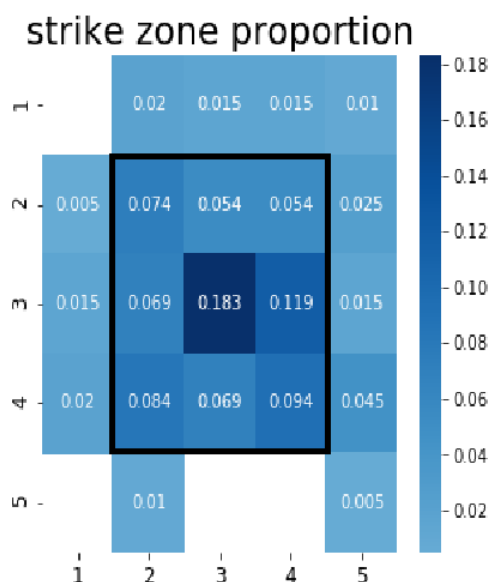


圖表 3 打擊率熱區



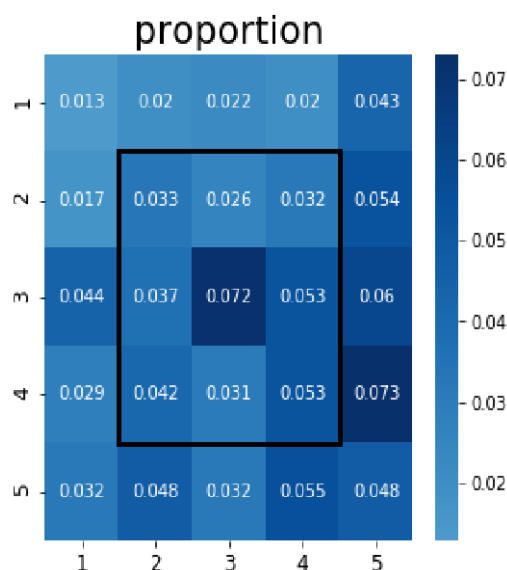
圖表 4 長打率熱區

其實從這兩張打擊率與長打率熱區的圖不難發現，王柏融在好球帶 2、4、5、7 的位置打得特別好。弱點也非常明顯，像是 1 號內角高的位置就打得特別差。除了這兩張圖，我也很好奇王柏融所打得球當中進壘點通常在什麼位置，或是說投手面對他時，較常將球丟在什麼地方。



圖表 5 (每次擊球出去的進壘點分布，包括安打、出局等等)

從圖表 5 可以發現，5 號位置的比例相對來說是非常高的。這告訴了我們，王柏融在所有打進場內的球當中，紅中的比例是所有位置中最多的，甚至有接近兩成的比例都是打紅中的球。接下來我想弄得更明白，究竟是王柏融看球看得非常清楚才能打到這麼多紅中球，還是說其實投手面對他時丟到紅中的比例也是這麼高，於是我整理了下一張圖表。



圖表 6(所有球路的進壘點分布比例，包括好壞球、界外球等等)

根據圖表 6，我們可以發現在好球帶內面對到最多的位置是 5 號紅中球。再看其他位置的比例，外角球明顯也是多於內角球。從這樣的資料我們可以得知，投手所投出的球當中，紅中的比例相對高非常多。也因此王柏融很容易將這樣的球打出不錯的結果。

## 2 壞球分析與本壘板紀律

根據上述的資料以及我在做逐球紀錄時所看的影片畫面，我認為投手對於王柏融所投出的球當中，有很高的比例都是屬於壞球，甚至有不少球和好球帶有很大的偏差，這樣可能造成俗稱的壞球太好選的現象。

於是我對現有的 2017 年逐球資料做壞球的比例計算，在這邊我定義新的名詞「當然壞球」，意指進壘點不在好球帶內且偏離好球帶一顆球以上。



當然壞球數: 448  
 總球數 929  
 當然壞球被判為好球數: 0  
 當然壞球率: 48.22%  
 當然壞球出棒總數: 48  
 當然壞球出棒率: 10.71%  
 揮空率: 43.75%

圖表 7(當然壞球比例)

計算出來結果發現，當然壞球的比例有將近五成這麼高。詳細情況，將計算本壘板紀律後與大聯盟平均的本壘板紀律做比較。

接下來，我計算出他的本壘板紀律，從結果觀察一些客觀的現象。

Stat	Average
O-Swing	15.42%
Z-Swing	75.57%
Swing	38.21%
O-Contact	73.03%
Z-Contact	92.48%
Contact	87.61%
Zone	37.89%
F-Strike	34.40%
<u>SwStr</u>	4.74%

圖表 8(本壘板紀律)

Stat	Average
O-Swing	30%
Z-Swing	65%
Swing	46%
O-Contact	66%
Z-Contact	87%
Contact	80%
Zone	45%
F-Strike	59%
SwStr	9.5%

圖表 9(mlb 平均本壘板紀律表)

O-Swing% :球丟在好球帶外時的出棒比例

Z-Swing% :球丟在好球帶內時的出棒比例

Swing :總出棒比例

O-Contact :好球帶外的擊球比例

Z-Contact :好球帶內的擊球比例

Zone :好球帶內球數/總球數

以 O-Swing%及 Z-Swing% 而言，我們可以知道王柏融的本壘板紀律數據是非常誇張的，特別是壞球的出棒比例，甚至是只有 mlb 統計平均值的一半，我們可以說王柏融對於壞球的辨識度很高，很少去追打好球帶以外的球。

至於 Zone 的比例，37.89%相較於 mlb 的 45%也是低了不少。是不是壞球太明顯，這個還缺乏更多數據佐證，但我們確實可以說王柏融遇到的壞球是比正常值高很多，其中也有很高的比例是屬於當然壞球。

以下是我從 [https://baseballdata.jp/playerB/1800127\\_2.html](https://baseballdata.jp/playerB/1800127_2.html) 日本數據網站找到的資料。我挑出王柏融在日本職棒 2019 年的選球眼數據觀察，前後分別是 O-Swing %與 Z-Swing %。

ボール球計	見極数	見極率
683	517	75.70%

圖表 10 (O-Swing%= 100%-75.70%=24.3%)

見逃	空振	スイング率
261	47	63.75%

圖表 11 (Z-Swing%)

簡單從這兩個數據來看，我們已經可以發現這兩數據已經和他在中華職棒的數據相差非常多，特別是 Z-Swing %已經比 mlb 的平均值 65%還低了，可以知道他在日本職棒對於好球帶內的進攻慾望變非常低。再來看看之前在中職很誇張的 O-Swing%，我們可以發現雖然說這項數據仍然比 mlb 平均值優異，但與原本的 15.42%相比，已經上升不少。當然畢竟日本職棒投手的強度一定比中職投手強很多，不過我們可以發現，日本職棒投手所丟出的球路肯定比中職投手的球路更具引誘性。整體而言，王柏融在日職的 O-Swing%表現還是相當不錯，也許王柏融本身選球能力就相當不錯，自然能選擇掉好球帶外的球路。

### 3 決策樹分析

由於逐球資料本身幾乎可以說是全類別資料，若要做一些機器學習分析勢必得將資料做過 one-hot-encoding 轉換。有些回歸分析等等的方式也不是那麼容易就能直接代入套件使用。在考量過很多目前我所學過的方法，我認為決策樹分析會是很好的方式，來幫助我了解到哪些變數可能是影響王柏融打得好與打不好的因素。以下將解釋我所使用的方式與輸出結果，最後分析得出結論。

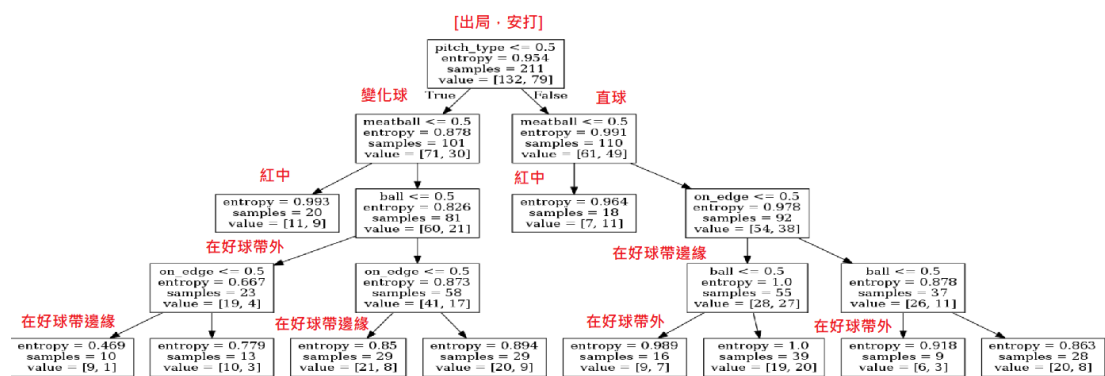
	A	B	C	D	E
1	ball	meatball	on edge	pitch type	result
2	1	0	1	0	0
3	0	1	0	1	0
4	0	1	1	0	0
5	1	1	1	0	1
6	0	1	1	0	0
7	1	1	1	1	0
8	1	1	0	0	0
9	1	0	1	0	1
10	1	1	0	1	0
11	0	1	0	1	1
12	1	1	1	0	0
13	1	1	0	1	1
14	1	1	0	1	1
15	0	1	0	1	1

圖表 12(選出可能影響的變數，並做二元分類)

變數

1. ball：是否在好球帶內
2. meatball：是否在紅中 5 號位置
3. on\_edge：是否在好球帶邊緣
4. pitch\_type：直球或變化球
5. result：打擊結果為出局或安打

我抓出這前四個可能影響 columns 當作自變數，並把其中一些數據做二元分類。例如我從原始數據的 pitch\_position 這項 column 抓出是否在好球帶內與是否為紅中 5 號的位置做出兩種變數，原始資料的 pitch\_type column 也有四種球路，為了方便起見，我只分類出變化球與直球兩種。最後將擊球結果 result 整理出兩種分類，分別是出局與安打。來觀察看看究竟是什麼樣的變數會對 result 的影響最大。

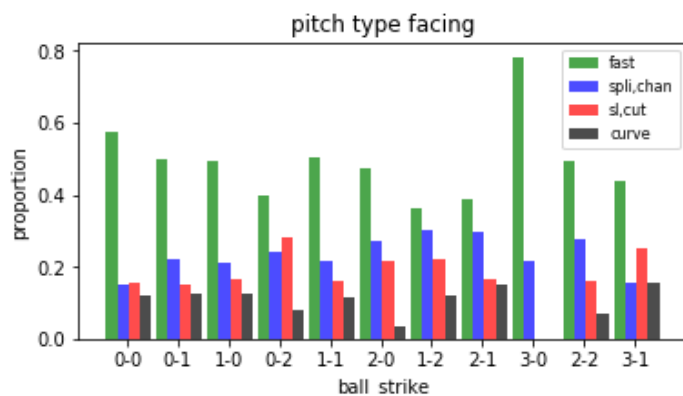


圖表 13(決策樹分析)

決策樹分析的結果相當有趣，我原本以為影響最大的變數會是紅中與否，不過跑出來的結果是直球或變化球才是影響最大的變數。金字塔最上面是原始數據，value 值的左邊代表出局數，右邊代表安打數。經過第一次分析之後我們可以發現，王柏融變化球的打擊結果比較不理想。再看下去可以看到變化球若不是丟到紅中的話，他的打擊率會下降更多，這說明了一個結果，王柏融也許對於不是太甜的變化球會相較之下苦手許多，倘若投手要做攻擊或許可以積極進攻好球帶，只要丟進好球帶但不是紅中的位置，就能得到不錯的效果。

來看直球的部分，當直球丟到紅中的時候，確實打得非常理想，安打數已經比出局數多，不過對於是否在好球帶邊緣或是否在好球帶內，沒有看出明顯的差異。

再來我對各種球數狀況來觀察王柏融面對到的球種，看看直球跟變化球的比例。



圖表 14(在不同情況下各球種比例)

從圖表 14 我們可以觀察到其實普遍直球的比例還是都有超過或接近一半) 0.5，雖說普遍投手所丟的直球比例就會比變化球高，而且更多時候變化球都是拿來搭配使用，不過在面對像王柏融這樣打變化球與直球有明顯落差的選手，增加變化球的使用會是可行的策略，而且結果就如決策樹所分析的結果，若變化球不要丟進紅中有機會得到不錯的結果。

接下來我做簡單的計算，來看看為什麼投手丟直球的比例會這麼高，是不是因為投手的變化球控球較差，不容易將球丟進好球帶。

變化球壞球總球數 206  
變化球壞球比例: 0.450  
直球壞球總球數 217  
直球壞球比例: 0.472

圖表 15

計算後的結果也算是稍微出乎我的意料，可以發現其實投手丟出的壞球比例，兩者之間並沒有太大的差異。

那是不是投手對於變化球的掌控不是那麼好，容易將變化球丟進不理想的位置？

變化球紅中總球數	32
變化球紅中比例	0.070
-----	
直球紅中總球數	35
直球紅中比例	0.076

圖表 16

很明顯經過計算的結果可以發現，兩者之間也沒有特別大的差異。

當然這當中有很多技術層面要考量，並非簡單估算就能得到正確的結論，如果說變化球的比例拉高，王柏融對於變化球的應對是不是也會變得更理想，這也是蠻值得觀察的現象。

變速、指叉球當然壞球數	110
總球數	929
變速、指叉當然壞球率	55.00%
變速、指叉當然壞球出棒總數	20
變速、指叉當然壞球出棒率	18.18%
揮空率	60.00%
-----	
滑球、卡特球當然壞球數	58
總球數	929
滑球、卡特當然壞球率	29.00%
滑球、卡特當然壞球出棒總數	6
滑球、卡特當然壞球出棒率	10.34%
揮空率	50.00%
-----	
曲球當然壞球數	50
總球數	929
曲球當然壞球率	25.00%
曲球當然壞球出棒總數	2
曲球當然壞球出棒率	4.00%
揮空率	100.00%

圖表 17(變化球打擊結果)

我做了簡單的計算來觀察哪一種變化球種對王柏融而言是較具有引誘性。

其中變速球及指叉球這樣的縱向球路，當其進壘點偏離好球帶時，王柏融的出棒率約為 18%，比圖表 8 的 O-Swing% 還高，而且這當中還不包括在好球帶邊緣的壞球出棒比例。

變速、指叉球壞球數: 135  
總球數 929  
變速、指叉壞球率: 67.50%  
變速、指叉壞球出棒總數: 29  
變速、指叉壞球出棒率: 21.48%  
揮空率: 48.28%

圖表 18(變速、指叉打擊結果)

從這樣的結果我們可以看到若變速、指叉丟在好球帶外時，不論是否在邊緣，王柏融的出棒比例大約是 21.5%，而揮空率也有約一半左右，從這樣的結果我們可以推論出王柏融可能是較苦手於變速及指叉球。



## 四、結論

### 1 分析出的可能結果

1. 王柏融不擅長打內角高的 1 號位置，經過我的計算我發現這個位置的打擊結果是 1-11，11 個打數當中只有 1 支一壘安打。但這樣的樣本數還不算太多，經過更多的統計才能得到更正確的結果。
2. 投手的進壘點位置不理想，從圖表 5、圖表 6 很明顯可以發現紅中的比例相當高。換個角度說王柏融對於紅中的球掌握度非常好，投手應該避免將球丟到 5 號的位置。
3. 王柏融比較不擅長打變化球，透過決策樹圖表 13 可以發現變化球相較於直球打的不理想。但以投手的配球來看，投手並沒有因此丟更多的變化球面對王柏融，或許可以試著丟更多的變化球來面對他，會得到不錯的效果。
4. 最後從圖表 18 可以發現王柏融是最不會打變速球或是指叉球的。如果說投手要挑一種變化球路來面對他，那變速球或指叉球會是最好的選擇。
5. 以日本職棒而言，日本職棒的投手有很多是擅長指叉球，這剛好是王柏融最不拿手的球路。又者日本職棒的投手相對於中華職棒投手來說控球是更精準的，如果日本的投手能有效的將變化球丟在非 5 號紅中位置的話，那王柏融的打擊結果可能會受到影響。

### 2 分析的缺點

1. 數據不夠多而且不準確。因為我只有做 60 場總共 900 多球的紀錄，以技術上而言，人工的紀錄沒辦法很有效的判讀出是否在好球帶以內或外面，更難的情況是很難精準辨別每個進壘點的位置或是球種的結果。只能依照我本身的經驗做紀錄，與美國 mlb 的機器記錄相比，這樣會造成很大的人工誤差。

2. 缺少更多的日本職棒數據做為佐證，目前我找到的進階數據也不多，而且2019年王柏融後來的受傷，打席數並沒有非常多，可能要蒐集到更多的資料才可以得知其他結果。