

應用統計方法期末報告



美國職籃調查與研究



國立陽明交通大學

NATIONAL YANG MING CHIAO TUNG UNIVERSITY

第14組

411350002 李定頤

311554002 林嘉揚

前言：

NBA是世界上最頂級的男子職業籃球聯賽之一，吸引了世界各地最優秀的球員。這個聯盟的比賽風靡全球，吸引了大量的球迷和觀眾。NBA的賽季通常從十月持續到翌年的四月，約莫五月分進入季後賽，最終決出年度的總冠軍隊伍。就在我們撰寫報告的此刻，總冠軍賽已經如火如荼地展開。無數的籃球巨星們在賽場上光芒四射，各顯神通，為了能夠一舉摘下最高榮譽而奮力拚搏。這場熱血沸騰的大賽不僅吸引了全球的球迷注目，球星們的出色表現更是成為了火熱的話題。我們的研究希望能深度探討這些明星球員的獨特特質，增添賞球的樂趣。

研究背景：

NBA(National Basketball Association, 國家籃球協會)是北美的男子職業籃球聯盟。它於1946年6月6日在紐約成立，由30支球隊組成，其中29支位於美國，1支位於加拿大。NBA分為兩個聯盟：東部聯盟和西部聯盟，每個聯盟都有三個分區。NBA的比賽具有高水準的競爭性和精彩的場面，球員們展現出技巧、速度和身體素質的精湛結合。他們在場上進行精彩的進攻和防守，通過投籃、運球、傳球和籃板等動作爭取得分。在NBA中，每支球隊都有自己的特色和球風。比賽中充滿了激烈的競爭，球隊之間進行著激烈的角逐，並爭取進入季後賽以爭奪總冠軍榮譽。NBA的賽事吸引了廣大的球迷和觀眾，無論是在現場觀賽還是通過電視轉播觀看，都能感受到比賽的激情與魅力。



研究目的:

我們研究分成兩個部分。首先透過過去的資料(2021)分析球員本身數據與人氣度是否相關，期望能建立模型。第二部分則透過模型，利用2023年的數據來預測當年全明星的人選。球員的人氣度依據的是全明星投票判斷，而其中球員的數據則來自Kaggle 資料庫，全明星投票數據來自NBA官方網站。

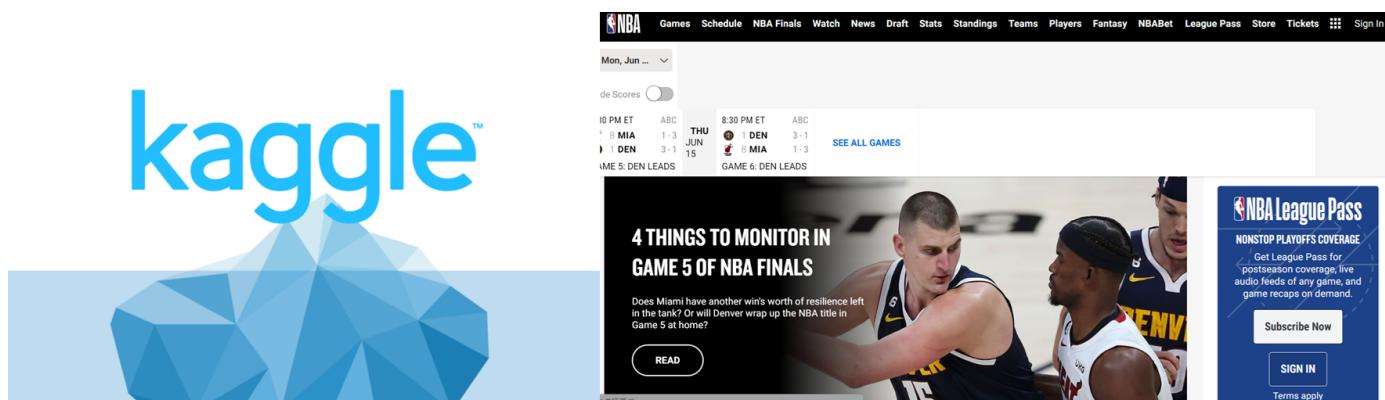
研究方法:

1. 資料來源:

(1) Kaggle是一個全球最大的資料科學社群和機器學習競賽平台。在Kaggle的網站上，用戶可以找到各種開放資料集，並使用這些資料進行數據分析或機器學習等研究。這些資料集涵蓋了多種領域，包括經濟、醫療、科技、社會科學等，甚至有運動相關數據之共享。在Kaggle上，用戶不僅可以找到資料集，還可以找到其他人分享的數據分析方法和機器學習模型，從而學習和交流。

(2) NBA全明星投票:

NBA全明星投票通常在每個賽季的12月25日開始，並持續到次年的1月22日。球迷可以通過NBA App、NBA.com 和Twitter進行投票。每個球迷每天可以投票一次，但在特定的五個日期(12月25日、1月7日、1月13日、1月17日和1月20日)可以進行兩次投票，這是一個新規定。最終的全明星首發陣容將由球迷投票(50%)、媒體投票(25%)和球員投票(25%)共同選出。全明星賽則通常在2月舉行。而相關投票資訊及數據，則可以透過NBA官網(<https://www.nba.com/>)獲得。



2. 研究過程：

首先排除出賽場次以及上場時間少的球員數據，後續針對366名球員數據進行分析：

(1) 球員數據內容包括26項數據：隊伍(Tm)、出賽場次(G)、先發場次(GS)、平均上場時間(MP)、投籃進球數(FG)、投籃出手次數(FGA)、投籃命中率(FG%)、三分球進球數(3P)、三分球出手次數(3PA)、三分球命中率(3P%)、兩分球進球數(2P)、兩分球出手次數(2PA)、兩分球命中率(2P%)、有效命中率(eFG%)、罰球進球數(FT)、罰球次數(FTA)、罰球命中率(FT%)、進攻籃板球(ORB)、防守籃板球(DRB)、籃板球(TRB)、助攻(AST)、抄截(STL)、阻攻(BLK)、失誤(TOV)、個人犯規(PF)、得分(PTS)。

(2) 人氣度根據全明星投票判斷，有以下3項數據：

球迷投票數，球員投票數，媒體投票數。此外根據球員的球場上的位置，區分為控球後衛、得分後衛、小前鋒、大前鋒、以及中鋒，每個位置的職責與打法不同，因此全明星會將投票分為前場(小前鋒、大前鋒、中鋒)以及後場(控球後衛、得分後衛)，所以此研究的模型也會根據位置將數據分開。我們用前場以及後場數據分別對球迷投票數、球員投票數、媒體投票數做multiple regression，來觀察不同類型的投票對於數據的關聯性。

(3) 變數選擇：

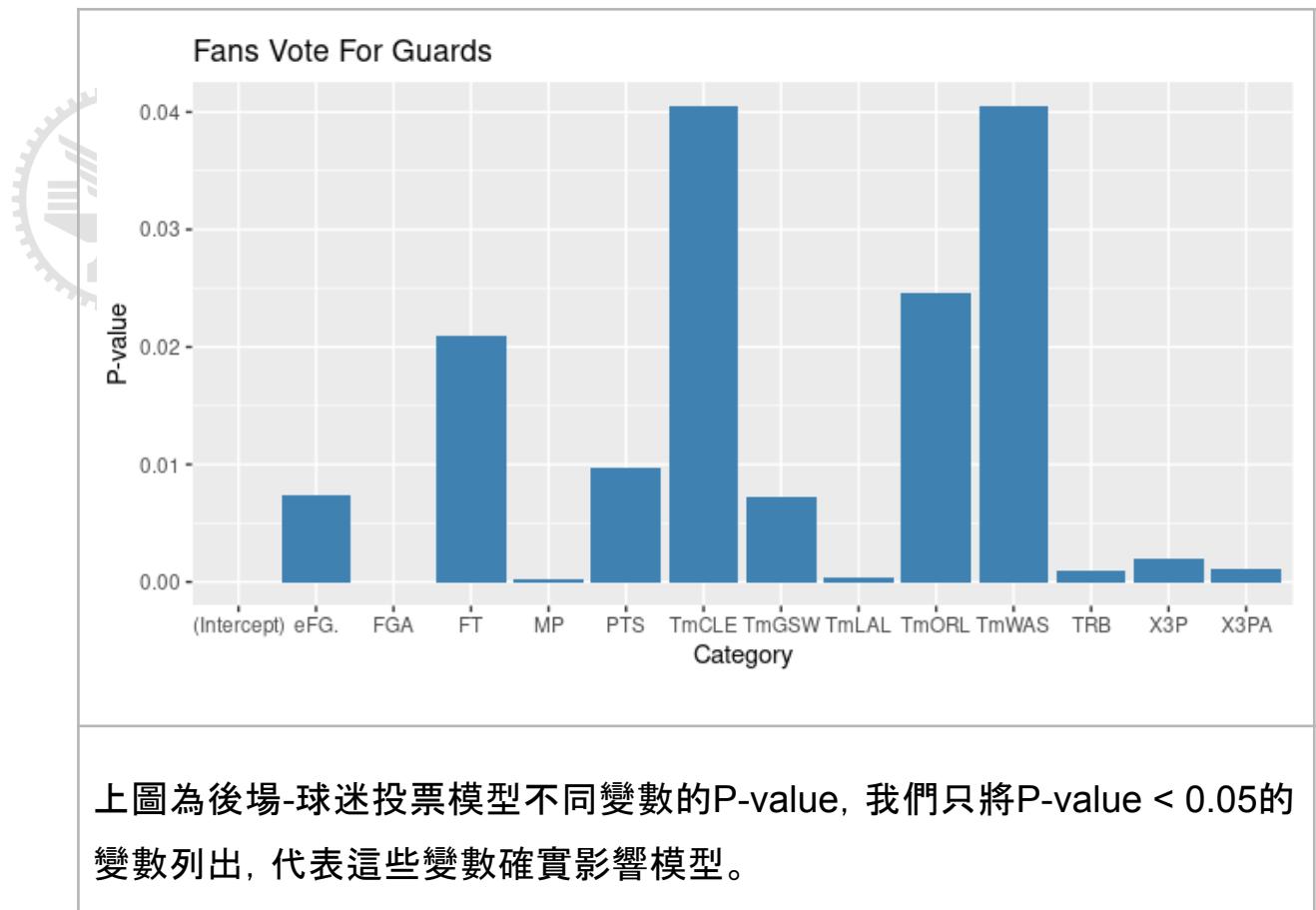
我們使用multiple regression進行分析，為了簡化模型，我們進行變數選擇，採用的方法是Backward regression。

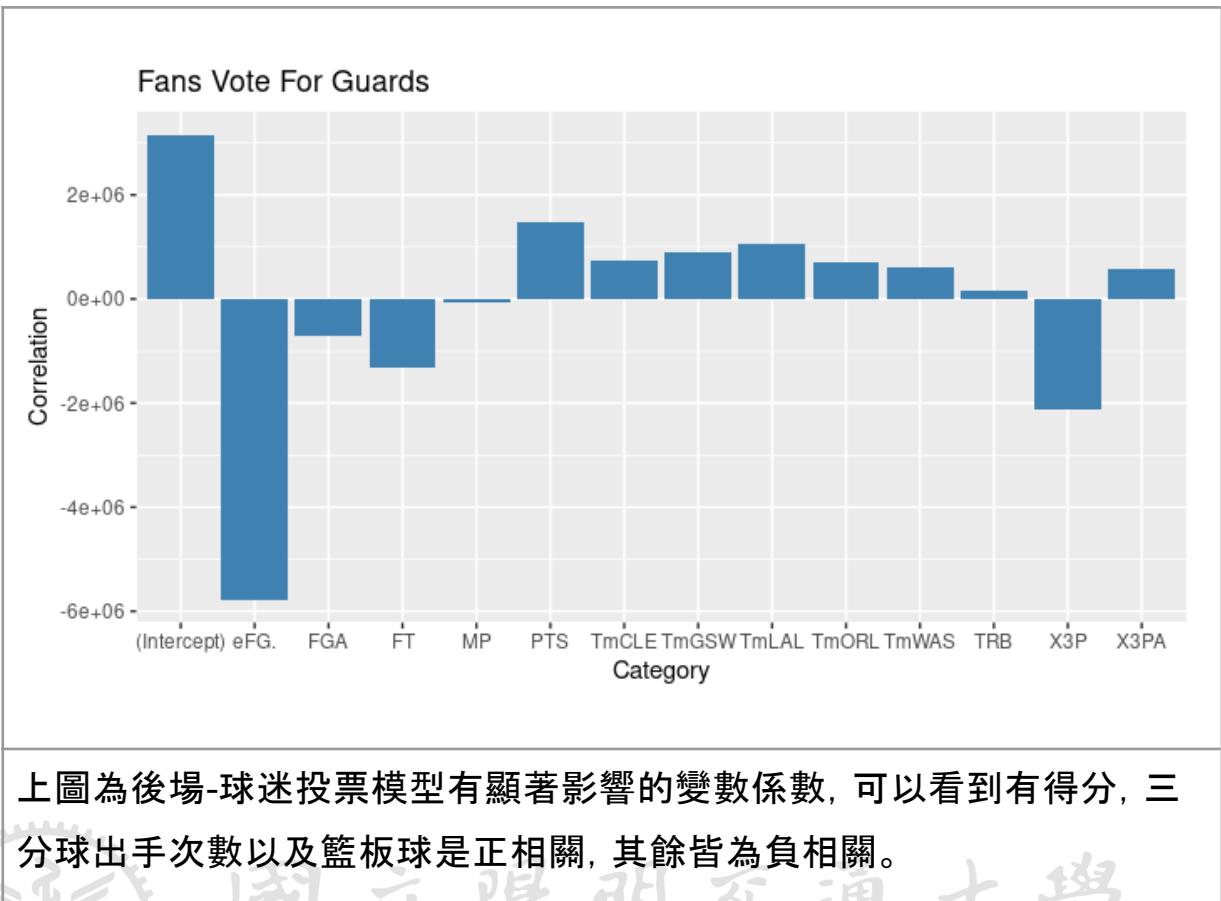
- 在後場 - 球迷投票數的模型中，我們刪除 FG%、DRB、2P、2PA、STL、G、ORB、BLK、FT%、FTA共10個變數，利用其他16個變數與球迷的投票數分析。
- 在前場 - 球迷投票數的模型中，我們刪除 FTA、3P%、G、GS、FG%、FT%、DRB共7個變數。

- 在後場 - 球員投票數的模型中，我們刪除 2PA、2P、AST、G、ORB、DRB、3P%、FG%、BLK、FT%共10個變數。
- 在前場 - 球員投票數的模型中，我們刪除 3P%、DRB、eFG%、FT%、FTA、2P、G、GS共8個變數。
- 在後場 - 媒體投票數的模型，我們刪除 ORB、STL、AST、2PA、2P、BLK、FG%、G、DRB、FT%、3P%共11個變數。
- 在前場 - 媒體投票數的模型中，我們刪除 BLK、eFG%、TOV、3PA、FGA、PF、GS、3P、FTA、DRB、2P%共11個變數。

(4) 模型分析：

- 後場-球迷投票模型：



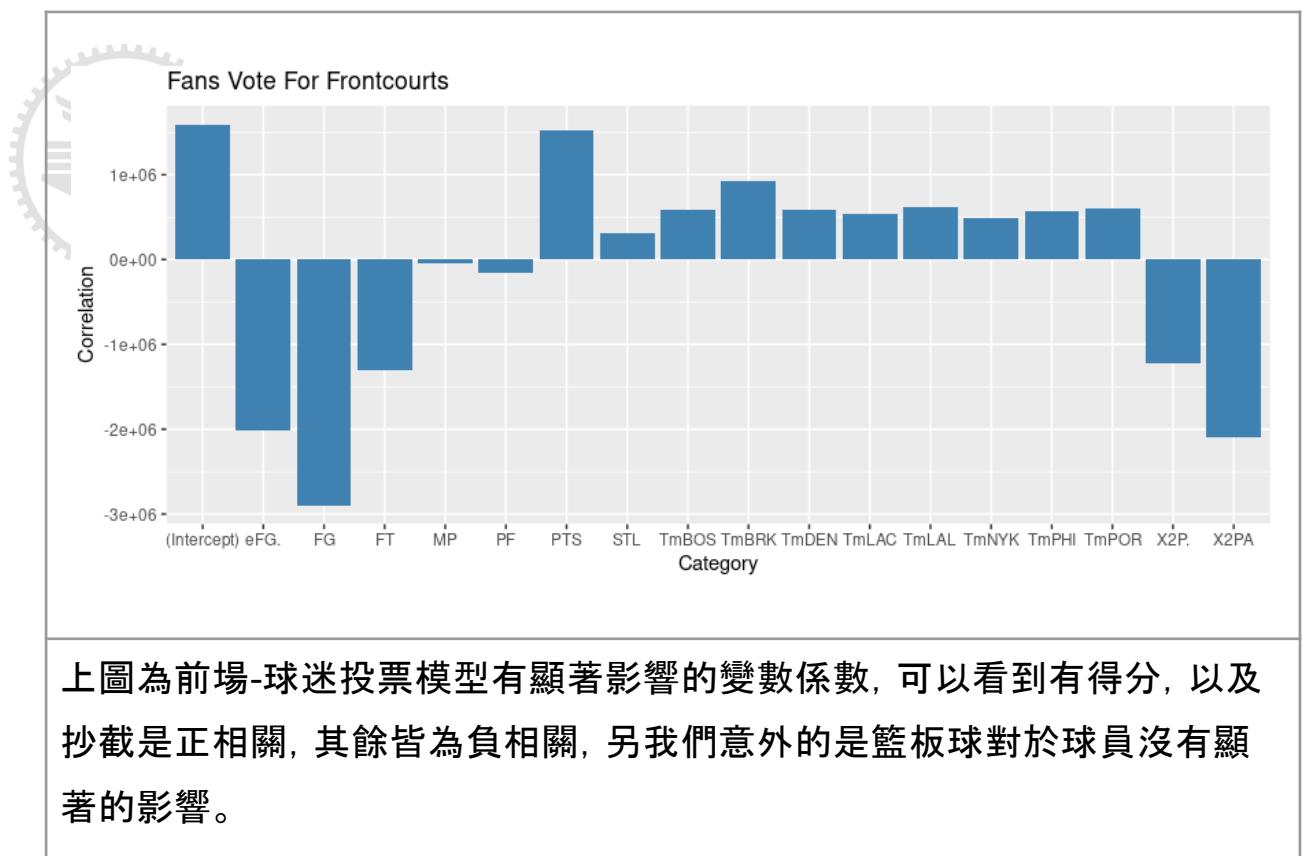
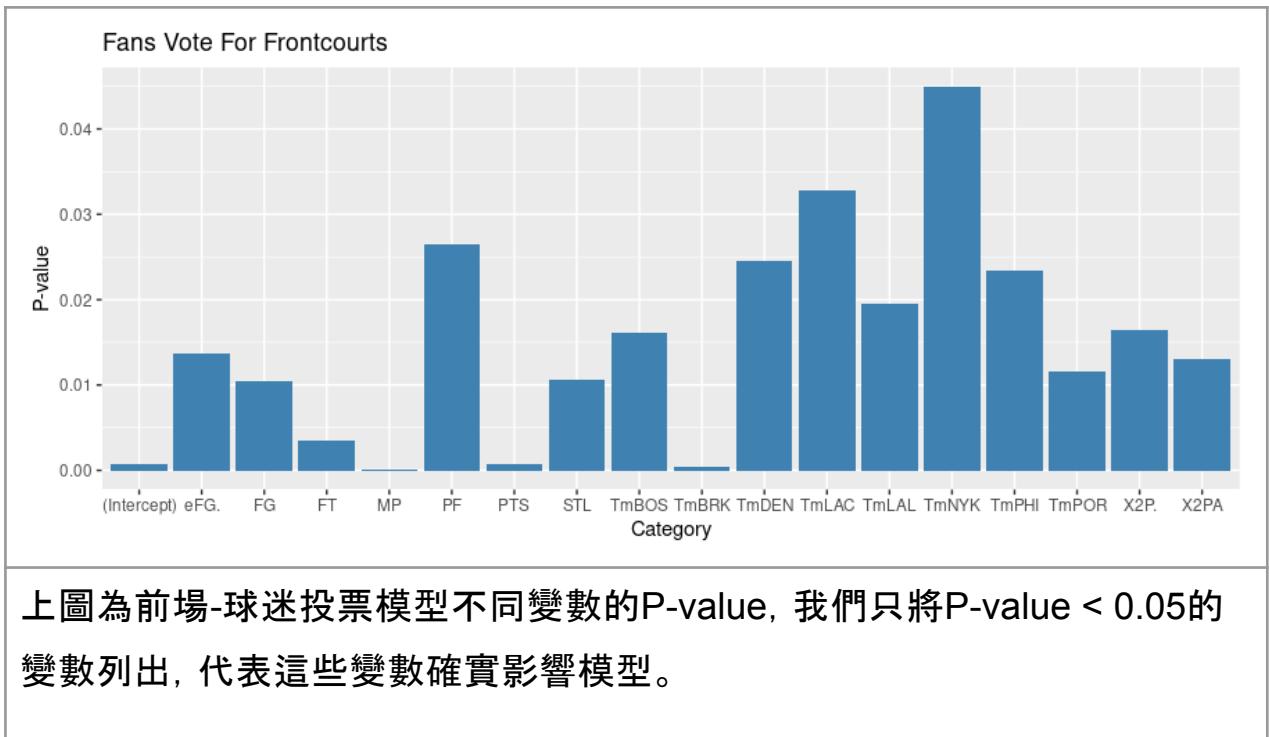


推論原因：我們認為球迷喜愛的後場球員需要具備強大得分力，或是能與中鋒爭搶籃板的能力，因此，得分、籃板數越多越容易得到球迷的青睞，另外我們認為三分球出手次數越多，得分也會越多，不排除有交互作用的產生。

總結來說：擁有得分能力及搶下籃板球的能力，會更易獲得球迷的投票。

除此之外，我們可以看到有一些隊伍也會影響模型，推論原因為剛好在這些隊伍之中有較多球迷以及這種類型的球員，才會有這些結果。

- 前場-球迷投票模型：

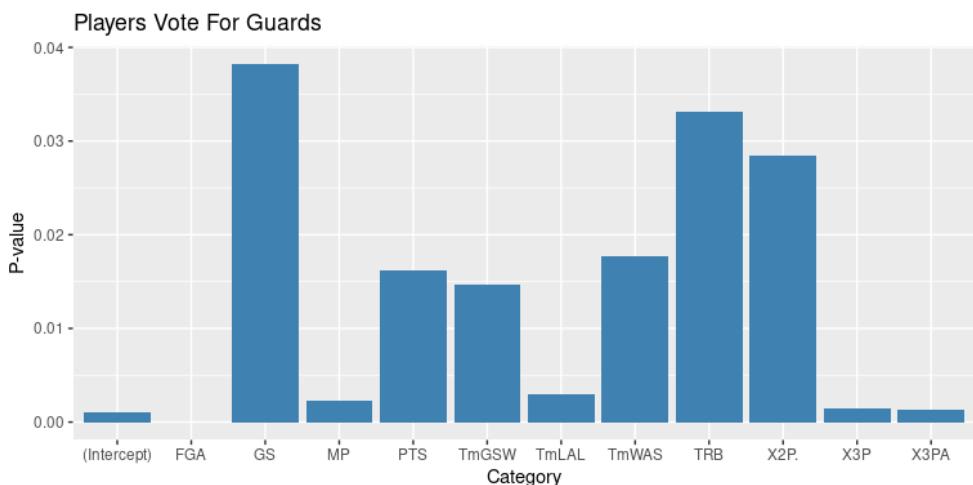


推論原因:我們認為球迷喜愛的前場球員也需要具備強大得分力，因此，得分越多越容易得到球迷的青睞，另外我們推論因為各隊的

中鋒能搶下籃板球的數量是差不多的，因此籃板球不足影響模型，得分還是關鍵。

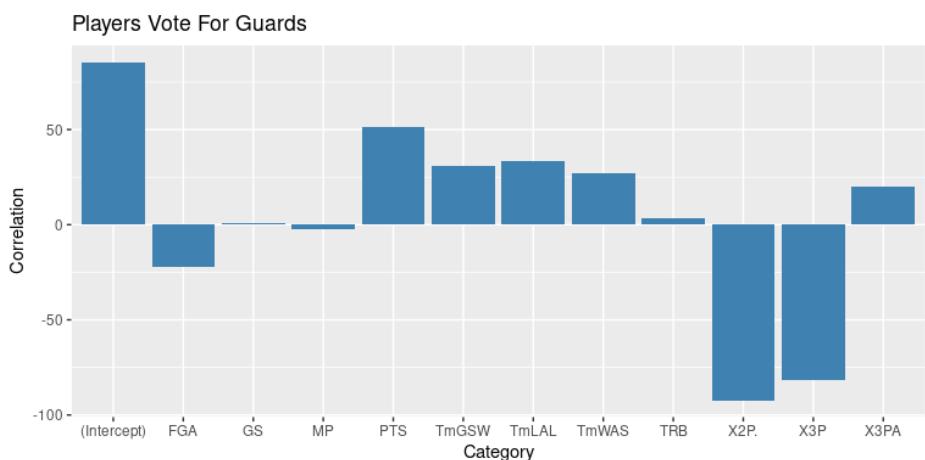
總結來說，擁有得分能力會更容易獲得球迷的投票。

- 後場-球員投票模型：



上圖為後場-球員投票模型不同變數的P-value，我們只將P-value < 0.05的變數列出，代表這些變數確實影響模型。

NATIONAL YANG MING CHIAO TUNG UNIVERSITY



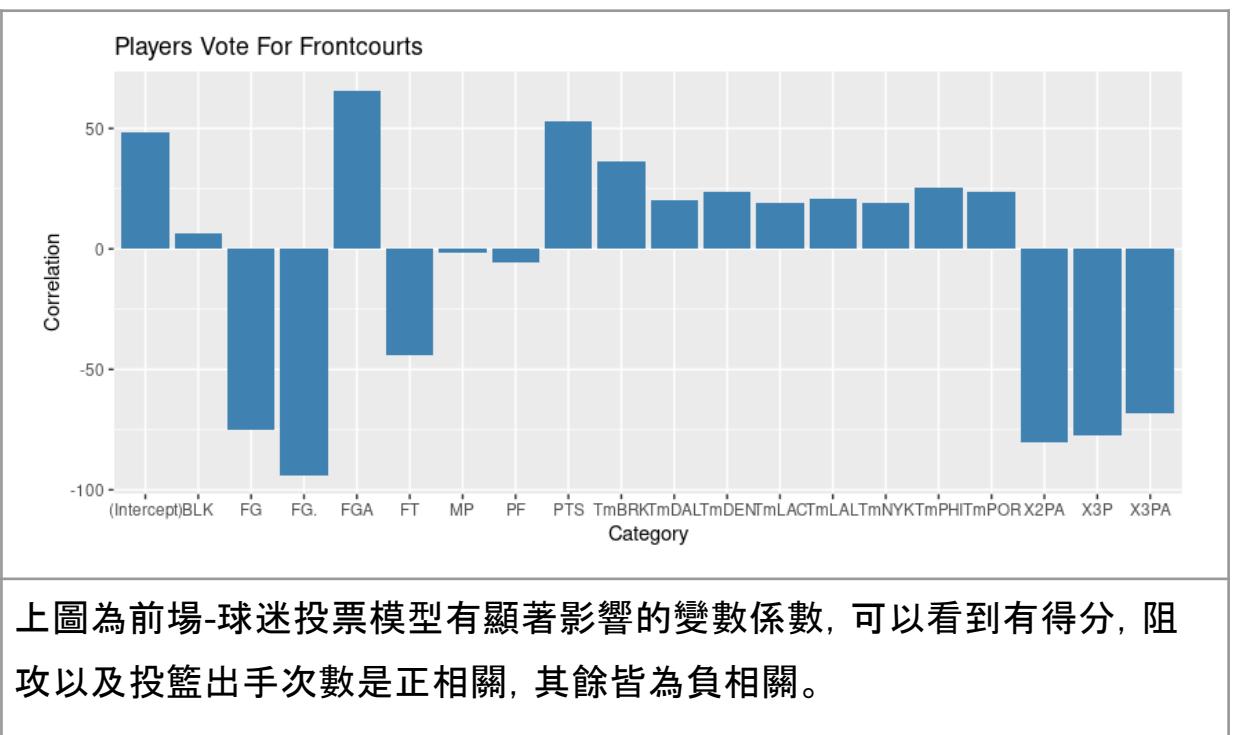
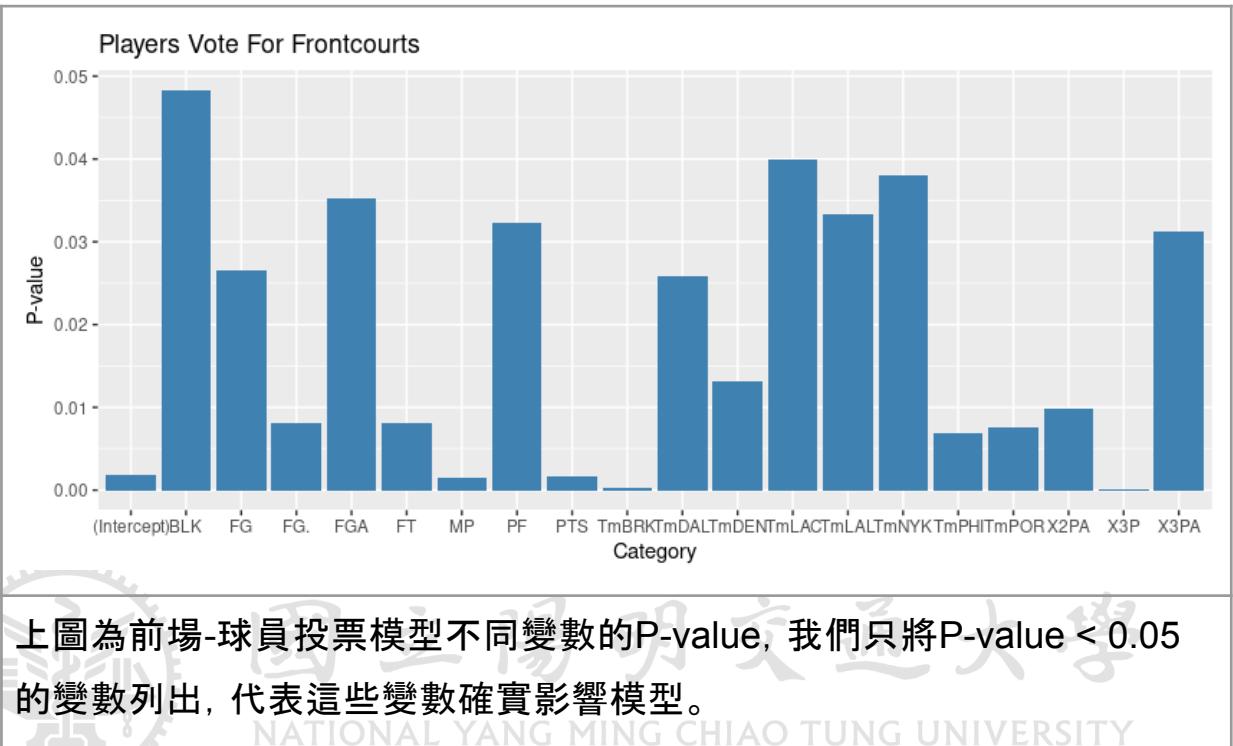
上圖為後場-球員投票模型有顯著影響的變數係數，可以看到有得分，籃板球以及三分球出手次數是正相關，其餘皆為負相關。

推論原因: 我們認為球員會投給具備強大得分力以及能與中鋒爭搶籃板球的後場球員，因此，得分越多越容易得到球員的青

昧，另外我們認為三分球出手次數越多，得分也會越多，不排除有交互作用的產生。

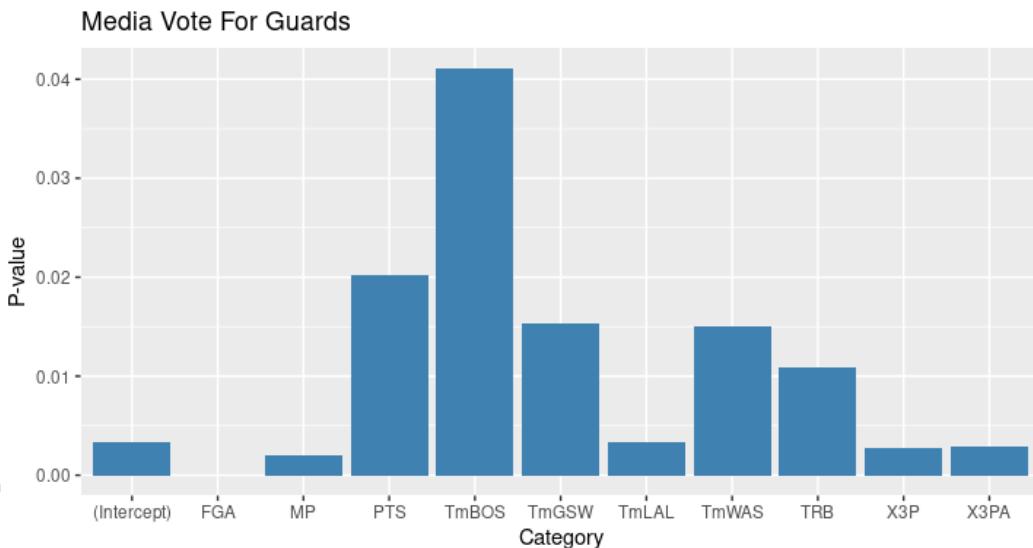
總結來說，擁有得分以及籃板能力會更容易獲得球員的投票。

- 前場-球員投票模型：

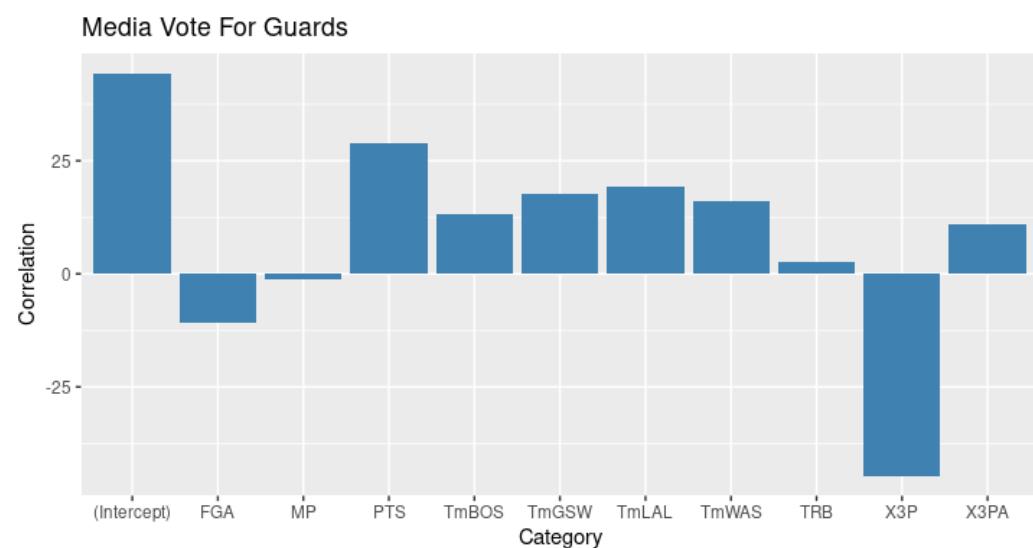


推論原因: 我們認為球員會投給有得分能力以及防守能力的前場球員，因此，能守住內線以及得分會得到球員的青睞，另外我們認為投籃出手次數越多得分就會越多，不排除有交互作用的產生。
總結來說，能夠得分以及阻攻保護內線會更容易獲得球員的投票。

- 後場-媒體投票模型:



上圖為後場-媒體投票模型不同變數的P-value，我們只將P-value < 0.05的變數列出，代表這些變數確實影響模型。

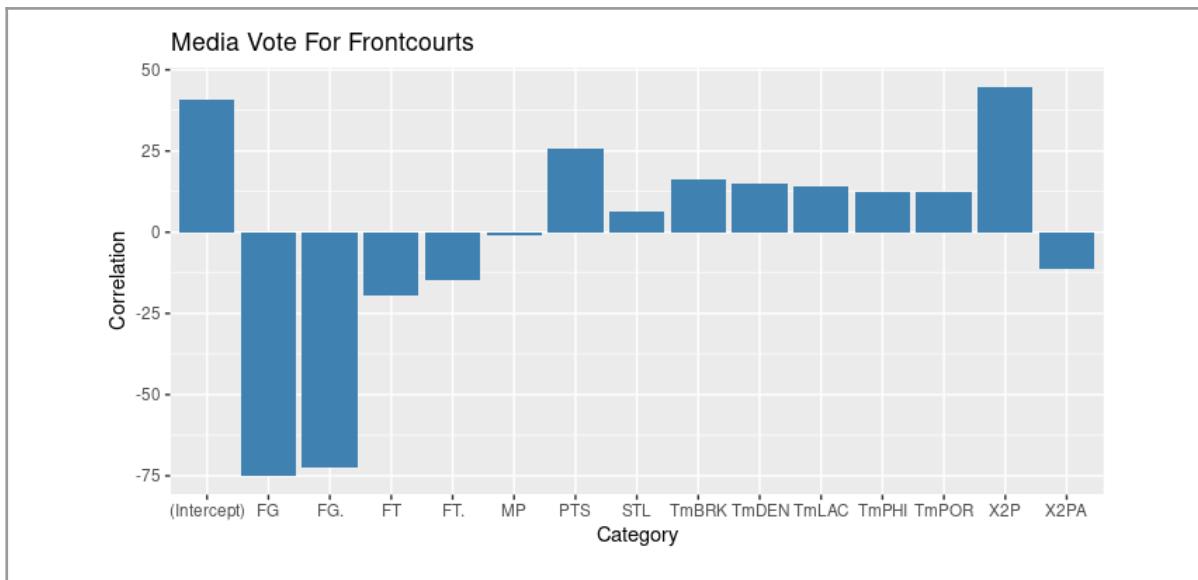
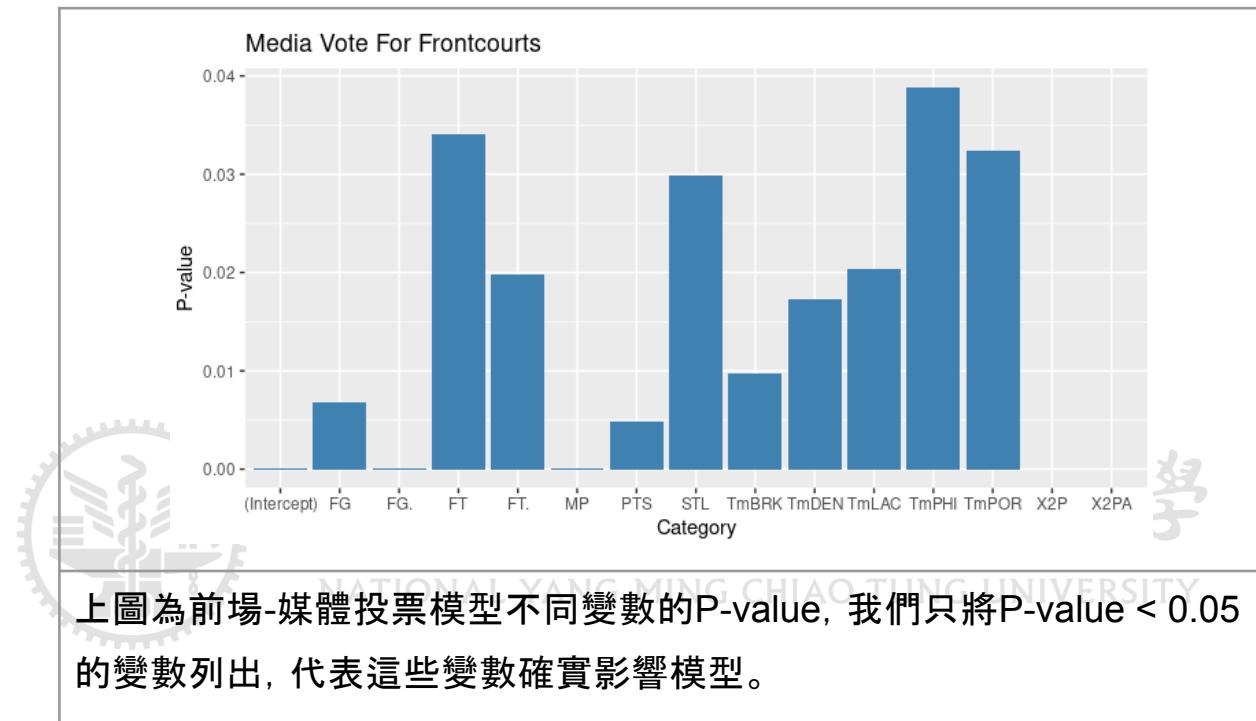


上圖為後場-媒體投票模型有顯著影響的變數係數，可以看到有得分、籃板球以及三分球出手次數是正相關，其餘皆為負相關。

推論原因：上圖的結果和後場-球員投票模型很相似，我們認為媒體喜愛的也是具備強大得分力以及能與中鋒爭搶籃板球的後場球員，因此，得分越多越容易得到媒體的青睞，另外我們認為三分球出手次數越多，得分也會越多，不排除有交互作用的產生。

總結來說，結果與後場-球員投票模型很相似，得分與籃板越多能獲得更多的媒體投票。

- 前場-媒體投票模型：



上圖為前場-媒體投票模型有顯著影響的變數係數，可以看到有得分，抄截以及二分球進球數是正相關，其餘皆為負相關。

推論原因：我們認為媒體會看中有籃下得分能力以及抄截能力的前場球員，因此，能在籃下得分會得到媒體的青睞，另外我們認為二分球進球數越多，總得分也會越多，會有交互作用的產生。
總結來說，能夠得分以及抄截會更容易獲得球員的投票。

(5) 交互作用：

- 後場-球迷投票模型：

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	572983	169907	3.372	0.000941 ***
PTS	-17597	16077	-1.095	0.275400
X3PA	-330129	45956	-7.184	2.67e-11 ***
PTS:X3PA	20505	2480	8.268	5.77e-14 ***

上圖為後場-球迷投票模型的交互作用，可以看到得分和三分球出手次數的確是有交互作用的存在，但是得分這個變數對於模型的關聯性不太顯著了。

- 後場-球員投票模型：

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	17.17308	5.85524	2.933	0.00387 **
PTS	-0.01439	0.55402	-0.026	0.97931
X3PA	-11.42738	1.58370	-7.216	2.24e-11 ***
PTS:X3PA	0.67431	0.08547	7.889	5.11e-13 ***

上圖為後場-球員投票模型的交互作用，可以看到得分和三分球出手次數的確是有交互作用的存在，但是得分的p-value卻很接近1，顯然對於模型的關聯性就完全不顯著了。

- 前場-球員投票模型：

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	22.05723	4.13524	5.334	2.56e-07 ***
PTS	1.85271	0.96715	1.916	0.0568 .
FGA	-9.41897	1.14335	-8.238	2.17e-14 ***
PTS:FGA	0.39953	0.03693	10.818	< 2e-16 ***

上圖為前場-球員投票模型的交互作用，可以看到得分和投籃出手次數的確是有交互作用的存在，但是得分的p-value高於0.05，對於模型就沒有那麼重要了。

- 後場-媒體投票模型：

```
Coefficients:  
Estimate Std. Error t value Pr(>|t|)  
(Intercept) 12.03942 3.30141 3.647 0.000362 ***  
PTS -0.75561 0.31238 -2.419 0.016727 *  
X3PA -5.69322 0.89295 -6.376 1.99e-09 ***  
PTS:X3PA 0.38916 0.04819 8.075 1.76e-13 ***
```

上圖為後場-媒體投票模型的交互作用，可以看到得分和三分球出手次數的確是有交互作用的存在，而且得分的p-value仍然低於0.05，可以說明得分對於模型依然重要。

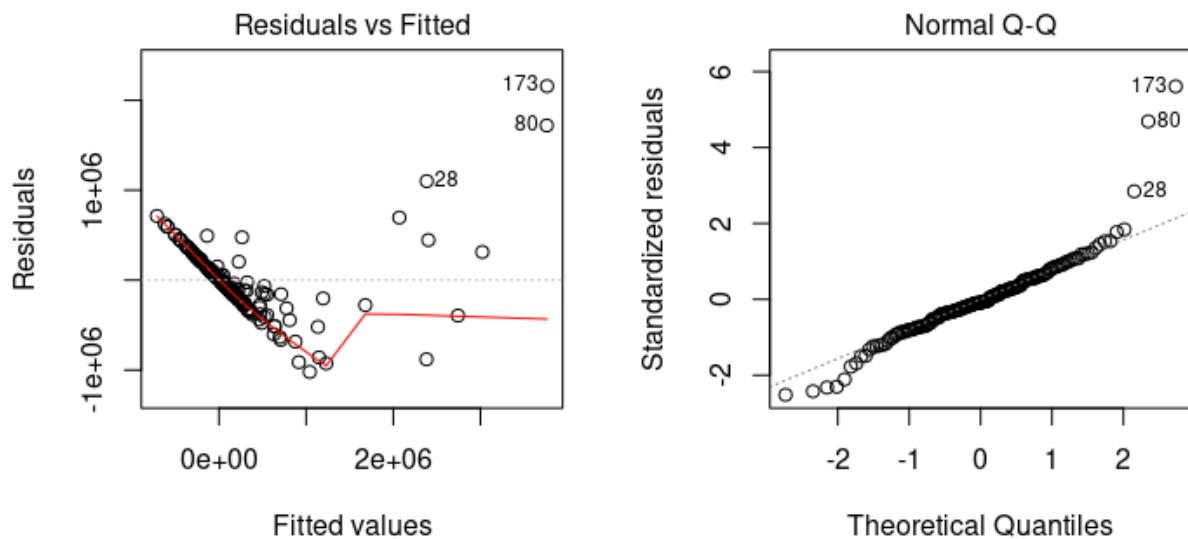
- 前場-媒體投票模型：

上圖為前場-媒體投票模型的交互作用，可以看到得分和二分球出手次數的確是有交互作用的存在，而且得分的p-value非常接近0，可以看出得分對於模型就非常顯著的關聯性。

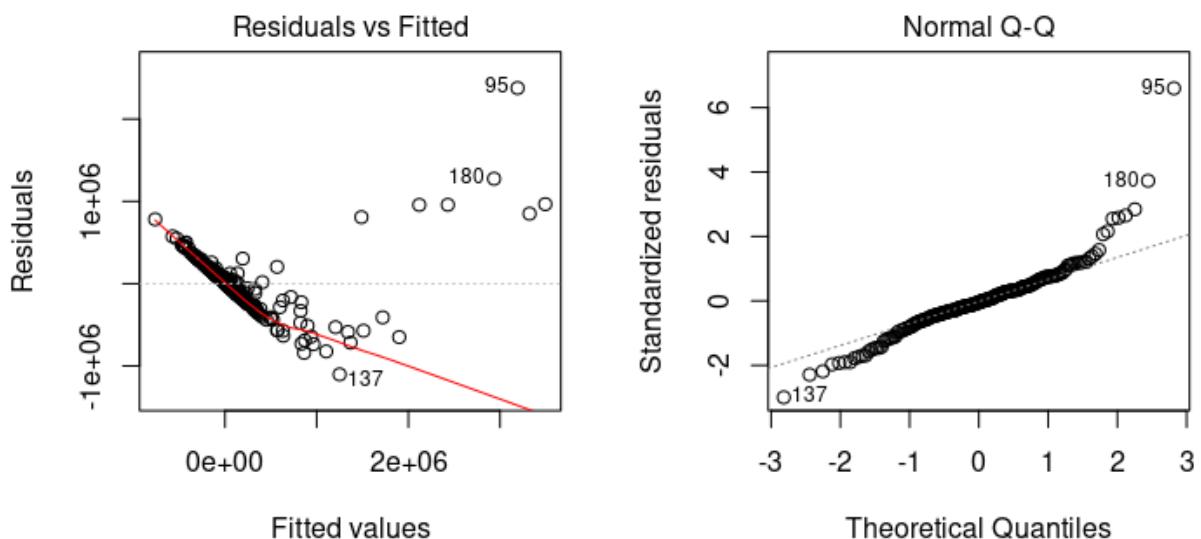
```
Coefficients:  
Estimate Std. Error t value Pr(>|t|)  
(Intercept) 12.0201 2.2786 5.275 3.40e-07 ***  
PTS -1.2299 0.3080 -3.993 9.15e-05 ***  
X2P -7.0297 1.0204 -6.889 6.98e-11 ***  
PTS:X2P 0.5677 0.0491 11.562 < 2e-16 ***
```

(6) 異常值分析：

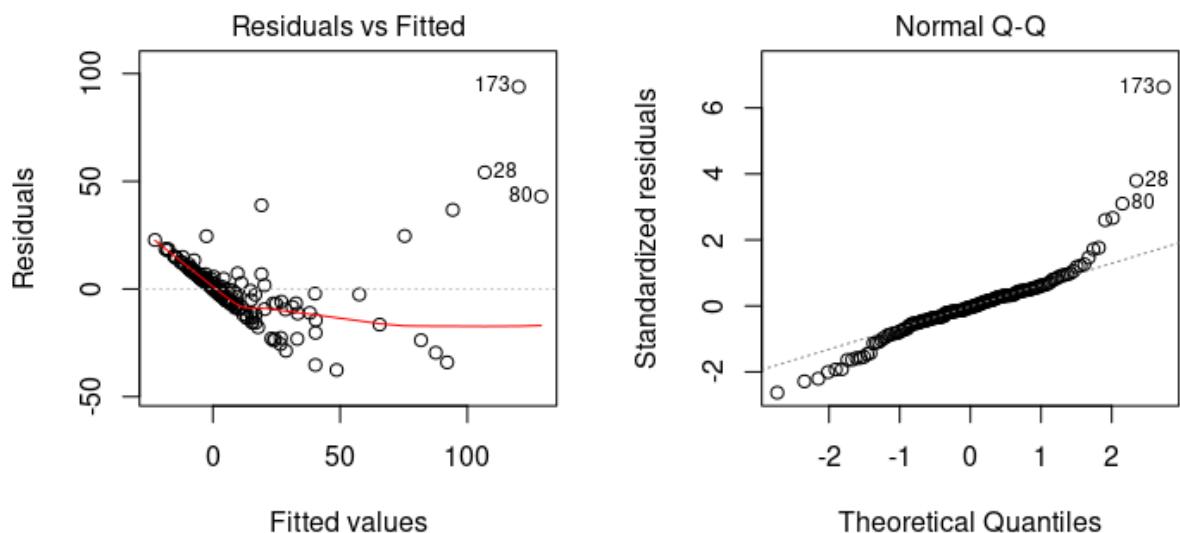
後場-球迷投票模型：



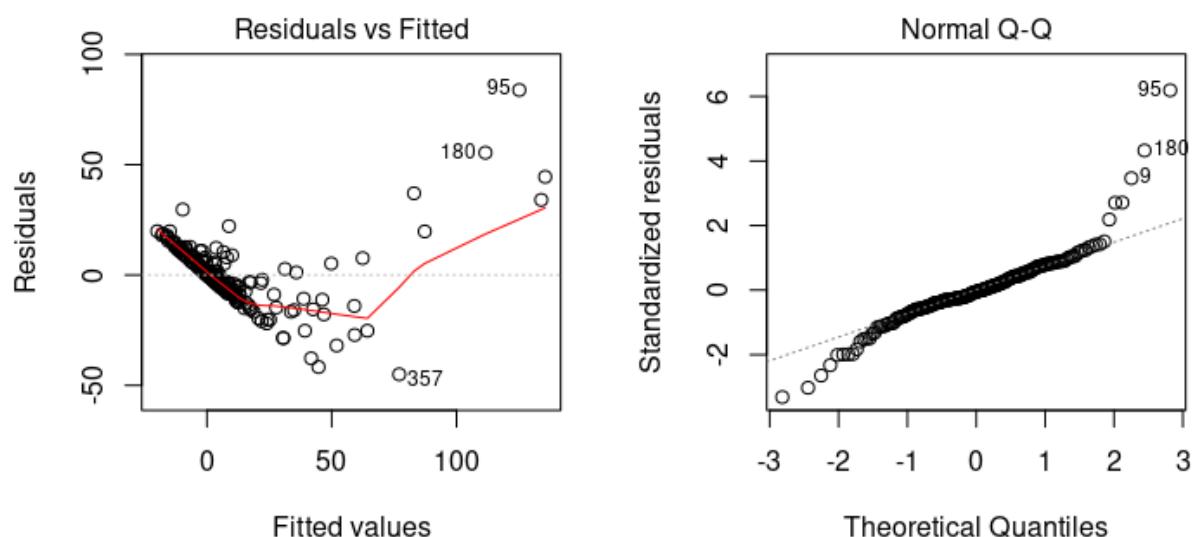
前場-球迷投票模型：



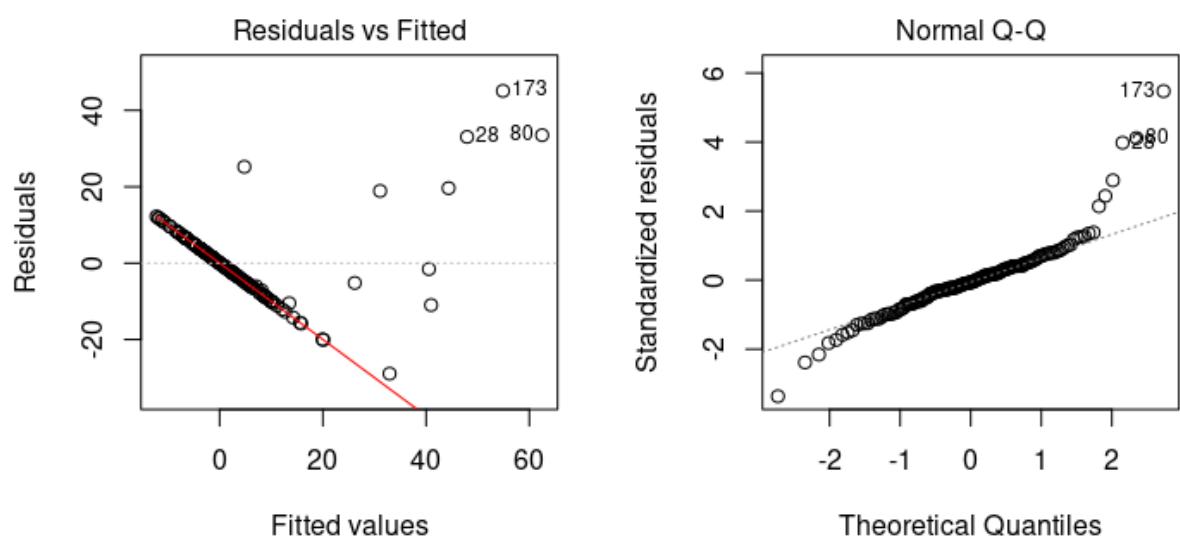
後場-球員投票模型：



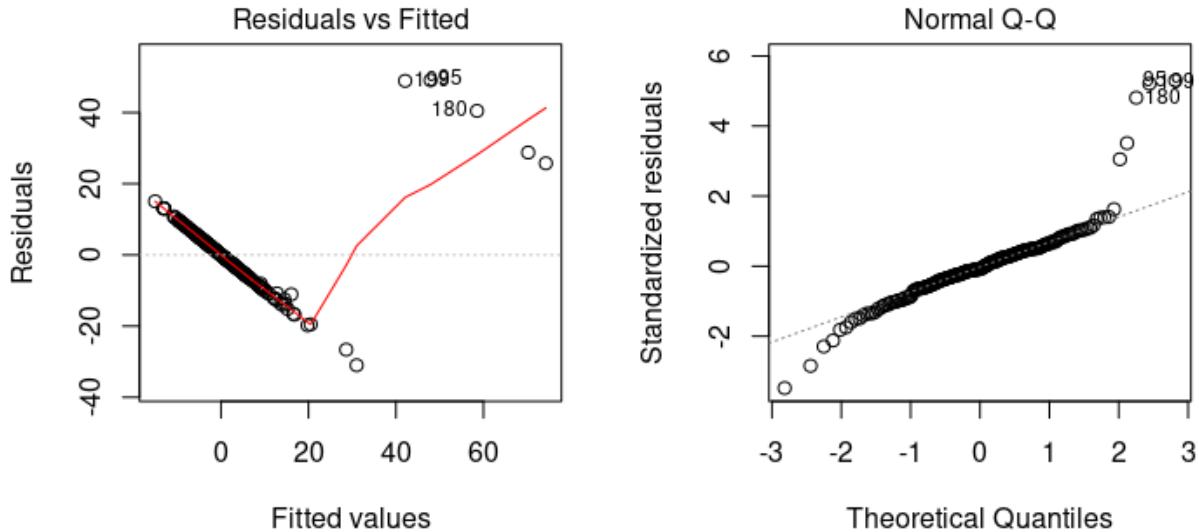
前場-球員投票模型：



後場-媒體投票模型：



前場-媒體投票模型：



以上是模型的殘差圖，我們看到編號28, 80, 173的後場球員以及編號95, 180在各個殘差圖都是異常值，編號28的球員為Bradley Beal、編號80的球員為Stephen Curry、編號173的球員為LeBron James、編號95的球員為Kevin Durant、編號180的球員為Nikola Jokić。

分析這些球員為異常值的原因，我們推論是因為在2021賽季這些球員的表現是所有球員之中最好的，所以他們所獲得的投票比其他球員高出非常多，以平均來看，平均球迷投票為226816票，平均球員投票為10票，平均媒體投票為3票，而這些球員的球迷投票數為500多萬票，球員投票數為160~200票，媒體投票數為90-100票，因此會被認為是異常值。

(7) 全明星結果預測：

NBA分為東區和西區，兩個分區會分別選出12名明星球員，其中包括5名先發球員及7名替補球員，在先發球員部分，由球迷、現役球員及媒體共同票選而出，其中以球迷票選加權分數最高，佔50%，球員和媒體則分別佔25%，而先發球員的組成包括2名後場球員及3名前場球員，至於替補球員部分則包括2名後場球員、3名前場球員及2名不限定位置球員，名單則由30隊教練進行挑選。

我們利用以上6種模型來預測2022-2023年每個球員的各種得票數。預測方法為：

1. 根據球員的角色定位，決定使用後場或是前場模型，控球後衛以及得分後衛使用後場模型，小前鋒、大前鋒以及中鋒使用前場模型。
2. 每位球員都會使用球迷預測模型、球員預測模型、以及媒體預測模型。
3. 根據每個模型的預測結果將得票數由高到低排出名次，得票數最高者名次為1，以此類推。
4. 加權(50%球迷, 25%球員, 25%媒體)之後選出東西區的全明星。

PLAYERS	FANS_VOTES	FANS_RANK	PLAYERS_VO	PLAYERS_RANK	MEDIA_VOTES	MEDIA_RANK	SCORE	DISTRICT	POS	SELECTED	ALL-STAR
Joel Embiid	3783807	4	154	1	85	1	2.5	East	Frontcourt	Starter	Yes
Kevin Durant	3789446	3	144	2	68	4	3	West	Frontcourt	Starter	Yes
Stephen Curry	3838143	2	132	4	63	5	3.25	West	Guard	Starter	Yes
Giannis Antetokounmpo	3397990	5	137	3	80	2	3.75	East	Frontcourt	Starter	Yes
Luka Dončić	3979987	1	127	5	54	8	3.75	West	Guard	Starter	Yes
Nikola Jokić	3086588	7	121	7	73	3	6	West	Frontcourt	Starter	Yes
Damian Lillard	3267539	6	125	6	59	6	6	West	Guard	Reserve	Yes
Donovan Mitchell	2924826	8	111	8	50	9	8.25	East	Guard	Starter	Yes
Jayson Tatum	2885180	9	98	10	59	6	8.5	East	Frontcourt	Starter	Yes
Zion Williamson	2536249	11	101	9	41	11	10.5	West	Frontcourt	Starter	Yes
LeBron James	2599846	10	91	14	36	16	12.5	West	Frontcourt	Reserve	Yes
De'Aaron Fox	2031691	16	94	12	37	14	14.5	West	Guard	Reserve	No
Anthony Davis	2310891	12	96	11	34	23	14.5	West	Frontcourt	Reserve	No
Zach LaVine	2063506	14	88	16	34	23	16.75	East	Guard	Starter	No
Ja Morant	2096536	13	80	18	30	30	18.5	West	Guard	Reserve	Yes
Bradley Beal	1858704	22	86	17	36	16	19.25	East	Guard	Reserve	No
Jordan Poole	1946986	19	75	21	35	20	19.75	West	Guard	Reserve	No
Devin Booker	1960379	18	89	15	32	28	19.75	West	Guard		
Tyrese Haliburton	2042421	15	77	19	30	31	20	East	Guard	Reserve	Yes
Jimmy Butler	1823873	23	75	22	36	16	21	East	Frontcourt	Reserve	No
Shai Gilgeous-Alexander	1760277	26	93	13	35	20	21.25	West	Guard		
Lester Quinones	1759478	27	72	23	37	14	22.75	West	Guard		
Kawhi Leonard	1800152	24	67	26	35	20	23.5	West	Frontcourt	Reserve	No
Jaylen Brown	1878718	21	59	37	36	16	23.75	East	Frontcourt	Reserve	Yes
Julius Randle	1742904	28	64	30	43	10	24	East	Frontcourt	Reserve	Yes
James Harden	1991839	17	65	28	27	36	24.5	East	Guard	Reserve	No
Anthony Edwards	1769222	25	77	20	30	32	25.5	West	Guard		
Kyrie Irving	1889377	20	71	25	26	38	25.75	East	Guard	Reserve	Yes
Paul George	1740715	29	60	36	38	12	26.5	West	Frontcourt		

上圖為我們根據NBA的規則選出全明星的結果。

我們選出的名單如下：

西區 : Kevin Durant, Stephen Curry, Luka Dončić, Nikola Jokić, Damian Lillard, Zion Williamson, LeBron James, De'Aaron Fox, Anthony Davis, Ja Morant, Jordan Poole, Kawhi Leonard

東區 : Joel Embiid, Giannis Antetokounmpo, Donovan Mitchell, Jayson Tatum, Zach LaVine, Bradley Beal, Tyrese Haliburton, Jimmy Butler, Jaylen Brown, Julius Randle, James Harden, Kyrie Irving

2023年NBA全明星的入選名單如下：

西區:Stephen Curry、Luka Dončić、Nikola Jokić、LeBron James、Zion Williamson、Shai Gilgeous-Alexander、Damian Lillard、Ja Morant、Paul George、Jaren Jackson Jr.、Lauri Markkanen、Domantas Sabonis

東區:Kyrie Irving、Donovan Mitchell、Giannis Antetokounmpo、Kevin Durant、Jayson Tatum、Jaylen Brown、DeMar DeRozan、Tyrese Haliburton、Jrue Holiday、Julius Randle、Bam Adebayo、Joel Embiid

我們的模型在預測24名全明星中成功預測了16名全明星，推論沒有成功預測其他8名全明星的原因可能是我們使用了2022-2023賽季整個賽季的數據而不是2022年到全明星賽前的數據，導致我們會選擇某些全明星賽前表現不好的球員但是賽後表現好的球員，準確率因此變差。總結來說，我們的模型預測全明星的準確率達到66.7%，如果能蒐集到賽季初到全明星賽前的球員數據，相信準確率能夠明顯上升。

附件:

在我們提供的附件中有程式碼(final_14.R)以及2個csv檔(nba_stat.csv and nba_2023.csv), nba_stat.csv是用來建立模型的資料, nba_2023.csv是用來預測結果需要的資料，程式會寫入3個csv檔, fans_prediction.csv是球迷模型預測的結果, players_prediction.csv是球員模型預測的結果, media_prediction.csv是媒體模型預測的結果，執行程式可以重現我們做出的結果。

參考資料:

NBA 2020-2021賽季球員數據

<https://www.kaggle.com/datasets/umutalpaydn/nba-20202021-season-player-stats>

NBA 2021全明星投票結果

https://www.basketball-reference.com/allstar/NBA_2021_voting.html

NBA 2022-2023 賽季球員數據

https://www.basketball-reference.com/leagues/NBA_2023_per_game.html

NBA 2023全明星投票結果

https://www.basketball-reference.com/allstar/NBA_2023_voting.html

NBA 全明星投票規則

<https://tw.news.yahoo.com/news/nba%E5%85%A8%E6%98%8E%E6%98%9F%E7%90%83%E5%93%A1%E6%80%8E%E9%BA%BC%E9%81%B8%E5%87%BA%E4%BE%86%E7%9A%84%EF%BC%9F%E5%8F%88%E5%A6%82%E4%BD%95%E9%80%B2%E8%A1%8C%E5%88%86%E9%9A%8A%E5%91%A2%EF%BC%9F-104554991.html>

NBA 2023全明星入選名單

https://en.wikipedia.org/wiki/2023_NBA_All-Star_Game



國立陽明交通大學
NATIONAL YANG MING CHIAO TUNG UNIVERSITY