

逢甲大學學生報告 ePaper

報告題名：

NBA 球員得分之迴歸分析

—以 Steve Nash 為例

The Analysis of Regression on NBA Players' Scores—

Take Steve Nash for Example

作者：徐子淳、李振維、胡智凱、周興文

系級：統計系三年甲班

學號：D9963395、D0087675、D9963216、D9963161

開課老師：高秀蘭 教授

課程名稱：迴歸分析

開課系所：統計學系

開課學年：101 學年度 第一學期

中文摘要

在一場籃球比賽中，球員的比賽結果往往是大家的焦點所在。但是在很多情況下，出手次數、命中率、籃板、失誤、抄截、助攻，都會影響球員的表現。我們擷取美國職業籃球聯賽中現役湖人隊後衛奈許在 2007-2009 單場表現數據來觀察出手次數、命中率、籃板、失誤、抄截、助攻，對於得分的影響。

在此報告中，主要探討在各項反應變數裡，何者對解釋變數(得分)具有較大的解釋能力。先對各項反應變數做適合度檢定，來判斷是否與解釋變數(得分)有線性迴歸關係。在確定之間具有線性迴歸關係後，再對模型來進行選取，針對模型選取，本篇報告對此使用了三種選取方法，第一為前進選取法(Forward method)，第二為後退選取法(Backward method)，第三是逐步迴歸法(Stepwise method)，而本報告也藉由使用其他選取法尋找變數，並且與前三種選取方法來綜合比較來選取解釋變數，也就是出手次數和命中率，以求得最佳的迴歸模型。

在得到迴歸模型後，我們也針對模型使用殘差檢定來檢測樣本是否服從常態分配、變異數是否為常數及迴歸模型是否為最佳的線性迴歸模型。

而最後的結論我們得到，當不考慮其他變數，出手次數越多、命中率越高，則得分也會越高。

關鍵字： Steve Nash、得分、迴歸、殘差分析、選取法

Abstract

In a basketball game, what people focus on is usually the result. However, what can affect players' performance includes steals, errors, rebounds, assists and field goals. We take Steve Nash, one of the Lakers' guards, as an analytical basis to understand the effects of steals, errors, rebounds, assists and field goals upon scores. The data range from 2007 to 2009.

This report focuses on all response variables and tries to find which one has more expansion in explanatory variables (scores). At first, we conducted fitness tests to recognize if a response variable has any relationship with explanatory variables. And then, after we collected and calculated the data, we were sure that they have relationships of linear regression. Then, we used three methods (forward method, backward method and stepwise method) to test the model. In addition to the three methods, we used other methods to make a comparison. Finally, we found that the greatest explanatory variables are field goal attempts and percentage. The field goal attempts and percentage contribute to the best model of regression.

After getting the model, we used the residual analysis to test if it has the relationship of normal distribution, if the variance is constant, and if the model of regression is the best model linear regression.

We came to the conclusion that when the other conditions are fixed, more field goal attempts and higher field goal percentage lead to more scores.



Keyword : Steve Nash; scores; regression; residual analysis; selection methods

目錄

一、研究動機與目的及流程圖.....	6
二、基本統計資料分析.....	8
2-1. 基本敘述統計量.....	8
2-2. Correlation.....	13
三、原始模式檢定.....	15
3-1. 建立迴歸模式.....	15
3-2. 參數檢定.....	16
3-3. 模型適合度檢定.....	19
3-4. 模型解釋能力.....	20
四、模式的選取方法.....	21
4-1. 前進選擇法(Forward)slentry=0.05.....	21
4-2. 後退選取法(Backward)slentry=0.05.....	22
4-3. 逐步迴歸法 slentry=0.05 slstay=0.05.....	23
4-4. 其他選取法.....	25
4-5. 結論.....	28
五、殘差檢定.....	29
5-1. 檢測常態.....	29
5-2. 檢測獨立.....	33
5-3. 檢測變異數.....	34
六、結論.....	38
七、附錄.....	39
7-1. 原始檔案.....	39
參考文獻與資料來源.....	48

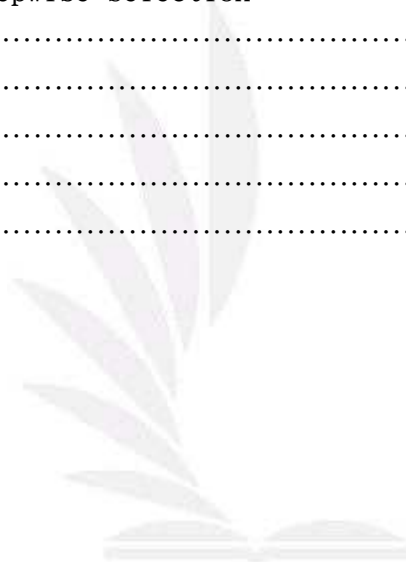
圖目錄

圖 1. 得分之相對次數直方圖	9
圖 2. 出手相對次數直方圖	9
圖 3. 命中率相對次數直方圖	10
圖 4. 籃板數相對次數直方圖	10
圖 5. 失誤相對次數直方圖	11
圖 6. 抄截相對次數直方圖	11
圖 7. 助攻相對次數直方圖	12
圖 8. 莖葉圖和箱形圖	30
圖 9. 常態分配圖	31
圖 10. 原始資料部分(Nomal q-q plot)	31
圖 11. 經各種檢定後留下的變數部分($X1&X2$)(Nomal q-q plot)	32
圖 12. 標準化殘差值次數分配直方圖	33
圖 13. 檢測變異數是否齊一及是否線性相關	34
圖 14. 原始資料殘差圖	35
圖 15. 最後留下變數部分殘差圖	36



表目錄

表 1. 基本統計量表	8
表 2. 變數相關性表	13
表 3. 參數估計表	15
表 4. 適合度檢定表	19
表 5. 模型解釋能力分析表	20
表 6. Summary of Forward Selection	21
表 7. Summary of Backward Elimination	22
表 8. 逐步迴歸 STEP1-1	23
表 9. 逐步迴歸 STEP1-2	23
表 10. 逐步迴歸 STEP2-1	24
表 11. 逐步迴歸 STEP2-2	24
表 12. Summary of Stepwise Selection	25
表 13. 其他選取法	25
表 14. 選取模式表	28
表 15. 常態檢定表	29
表 16. 獨立檢測表	33
表 17. 殘差值估計表	37



一、研究動機與目的

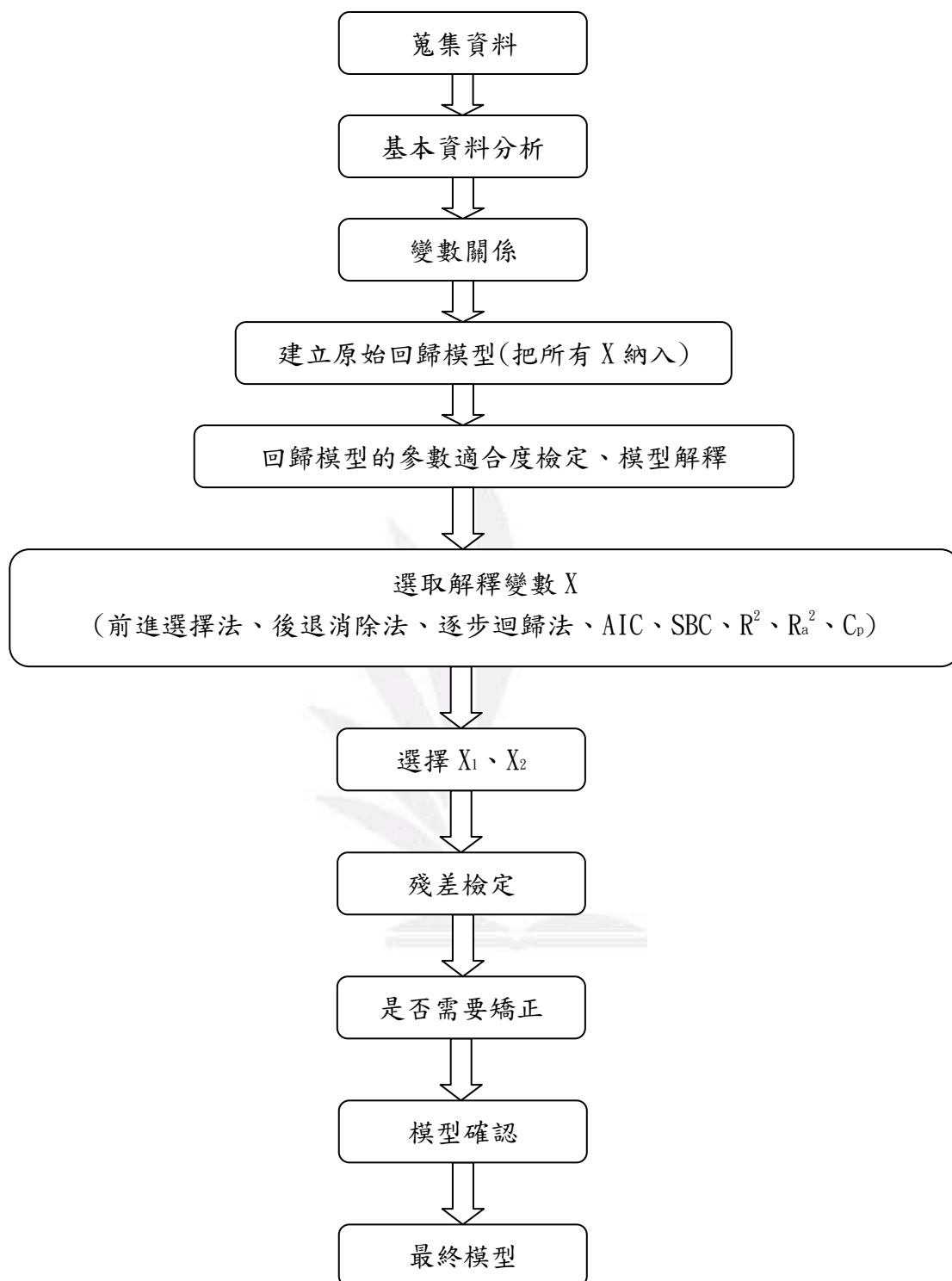
籃球是起源於美國的體育運動，籃球運動經過百餘年的發展後，在各方面有長足的改變，如：球員裝備、器材改良、規則、裁判專業、戰術、球員素質的提升，使籃球比賽更為精采。而近年來網路與媒體的普及，各項精采球賽的轉播，如 NBA 籃球、亞洲錦標賽，還有國內的超級籃球聯賽 SBL 等熱門賽事，深受國人的喜愛，籃球已是國人受歡迎的運動之一。

NBA 是世界上最迷人、運動競技水準最高的籃球比賽，其技巧、爆發力、速度、融合各個球員特殊技藝所展現的驚心動魄球賽。NBA 職籃源自 1949 年至今已有 50 年歷史。其採用獨特籃球規則，不同於國際規則。

其中我們找了洛杉磯湖人隊的控球後衛 Steve Nash 來做分析，Nash 的組織、傳球、突破、投籃，所有的靈感是來自於他的創意，他即興創作，隨意揮灑，藉著單檔切入、變速和他自己的風格。沒有什麼複雜的戰術框框，更不是刻意走位、掩護下的產物。球在 Nash 手中，就像一科會思考力、判斷力的大腦，在 Nash 的指揮下，任何時間都能找到場上隊友，輕鬆得兩分。

因此，我們以 NBA 這項熱門的運動比賽作為我們這次的研究主題，然後我們找了 NASH 的 2007-2009 年兩年的出場資料來了解他個人與得分的情況之關係，探討得分與命中率、出手次數、攻守之相關性。

流程圖



二、基本統計資料分析

2-1. 基本敘述統計量

Y：得分 x1：出手次數 x2：命中率 x3：籃板數
x4：失誤次數 x5：抄截次數 x6：助攻次數 N：場次

表 1. 基本統計量表

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Y (得分)	155	16.32903	7.27664	2531	0	37.00000
x1 (出手數)	155	11.69677	4.54469	1813	0	27.00000
x2 (命中率)	155	50.18774	16.82697	7779	0	100.00000
x3 (籃板)	155	3.25806	1.87200	505.00000	0	9.00000
x4 (失誤)	155	3.50323	1.98492	543.00000	0	10.00000
x5 (抄截)	155	0.69677	0.90722	108.00000	0	4.00000
x6 (助攻)	155	10.41290	3.75169	1614	2.00000	21.00000

Y：得分

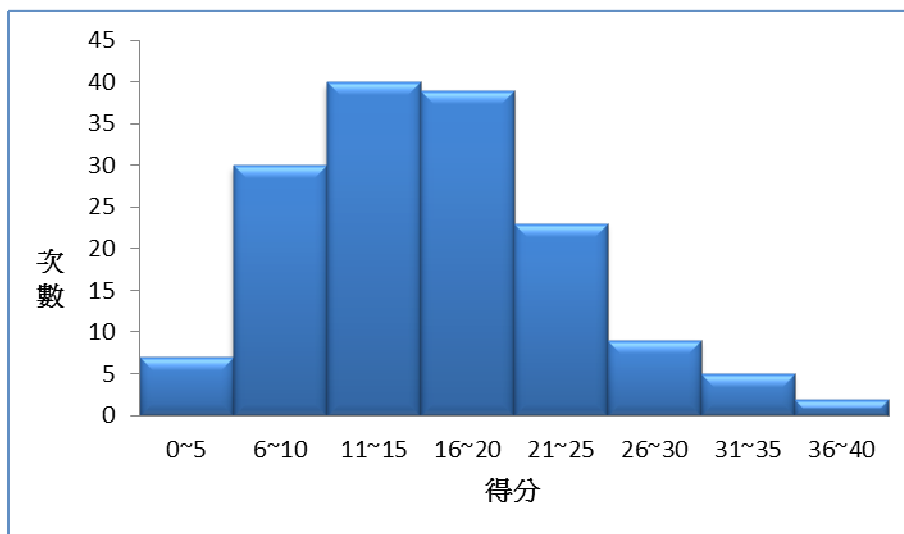


圖 1. 得分之相對次數直方圖

由上圖可知，在 11-15 的個數最多，有 40 個；在 31-35 及 36-40 的個數則最少，分別為 5 個和 2 個；總體來說集中於 6-25；呈現右尾分配。

(1) X_1 ：出手次數

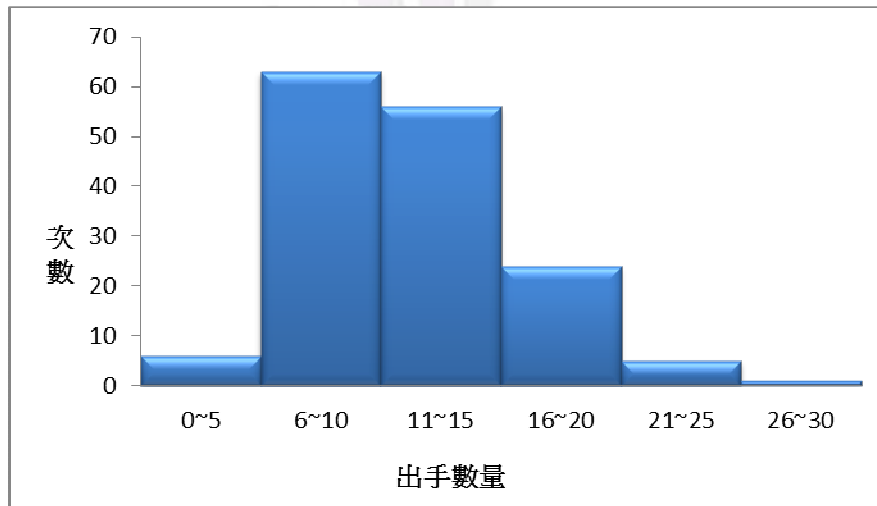


圖 2. 出手相對次數直方圖

由上圖可知，在 6-10 的最多，有 63 次；而在 26-30 的最少，只有 1 次；大多集中在 6-20，呈右尾分配。

(2) X_2 ：命中率

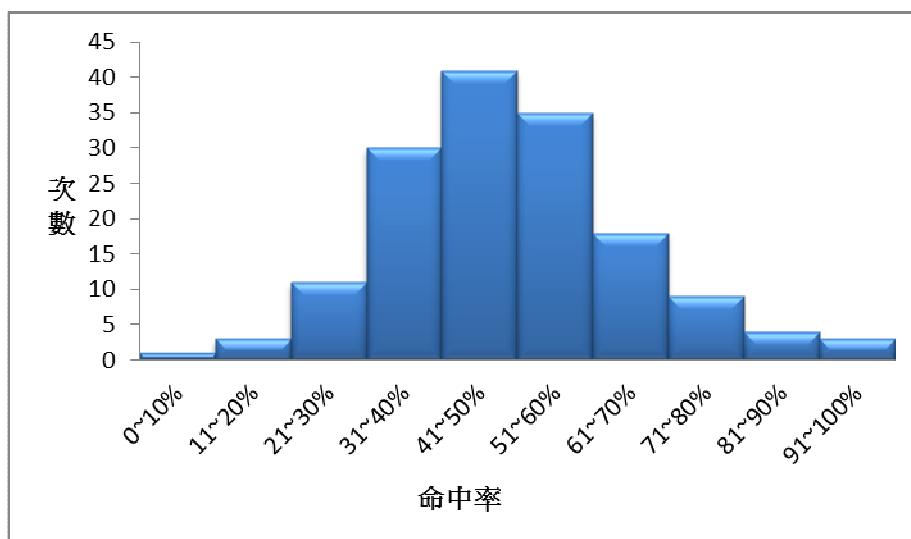


圖 3. 命中率相對次數直方圖

由上圖可知，在 40%~50% 的最多，有 41 個；0~10% 的最少，只有 1 個；在 71% 以上的不多，大致為鐘形分配。

(3) X_3 ：籃板數

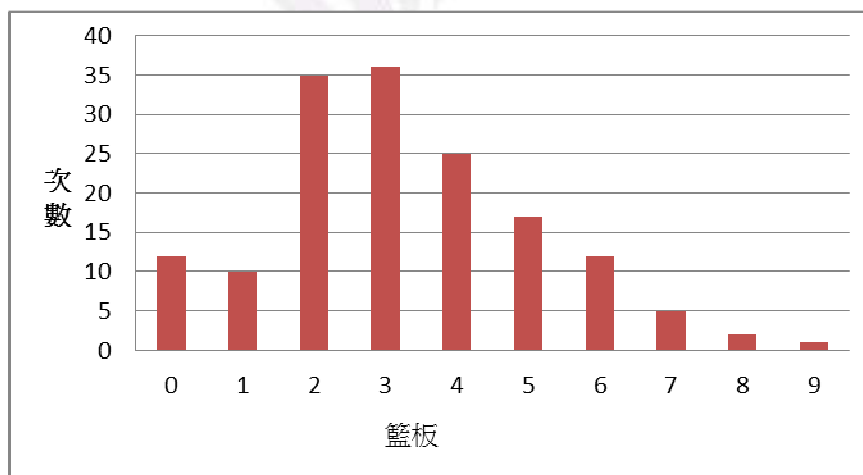


圖 4. 籃板數相對次數直方圖

由上圖可知，在 3 的最多，有 36 個；在 9 的最少，只有 1 個；多集中在 2~5；呈右尾分配。

(4) X_4 ：失誤次數

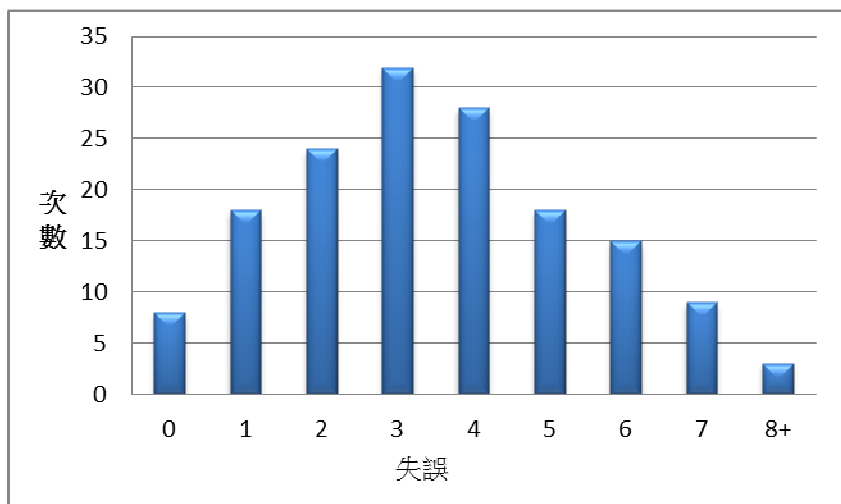


圖 5. 失誤相對次數直方圖

由上圖可知，失誤在 3 的最多，有 32 次；8 以上的最少，有 3 次；，大致為鐘形分配。

(5) X_5 ：抄截次數

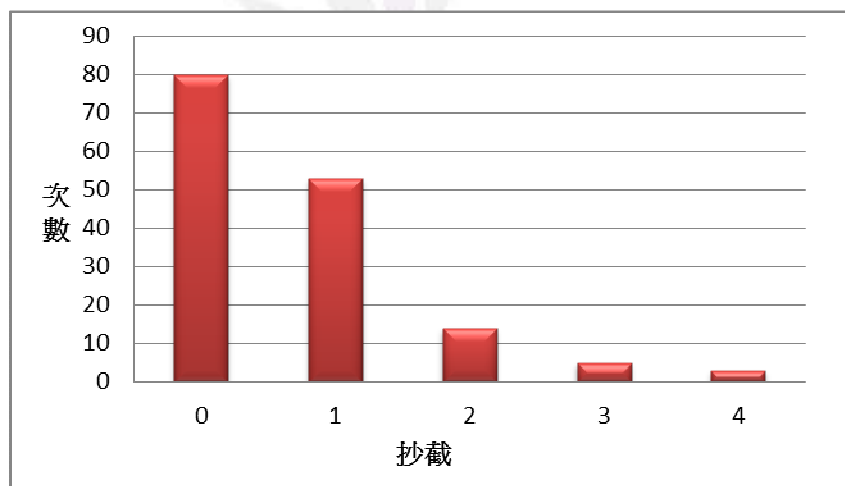


圖 6. 抄截相對次數直方圖

由上圖可知，抄截次數在 0 次最多，有 80 次；4 次最少，有 3 次。圖形成左尾分配。

(6) X_6 ：助攻次數

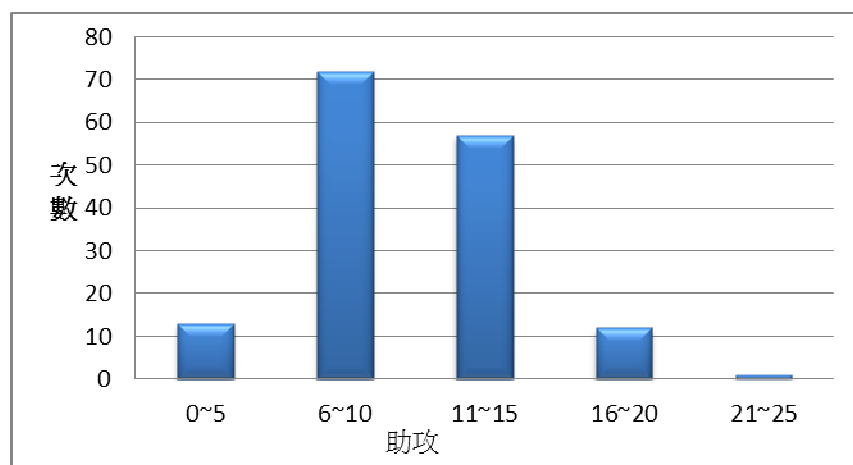


圖 7. 助攻相對次數直方圖

由上圖可知，助攻次數在 6~10 次最多，有 72 次；在 21~25 最少，有 1 次。圖形接近鐘型分配。



2-2. Correlation

表 2. 變數相關性表

Pearson Correlation Coefficients, N = 155							
Prob > r under H0: Rho=0							
	y	x1	x2	x3	x4	x5	x6
y 得分	1.00000	0.77491 <.0001	0.49758 <.0001	0.09193 0.2553	0.15615 0.0523	0.10177 0.2076	0.06064 0.4535
x1 出手數	0.77491 <.0001	1.00000	0.02631 0.7452	0.04055 0.6164	0.13724 0.0886	0.15395 0.0558	0.09232 0.2532
x2 命中率	0.49758 <.0001	0.02631 0.7452	1.00000	0.04172 0.6062	0.08668 0.2835	0.00716 0.9296	-0.06564 0.4171
x3 籃板數	0.09193 0.2553	0.04055 0.6164	0.04172 0.6062	1.00000	0.09938 0.2186	0.05020 0.5351	0.20663 0.0099
x4 失誤	0.15615 0.0523	0.13724 0.0886	0.08668 0.2835	0.09938 0.2186	1.00000	0.08889 0.2714	0.08876 0.2721
x5 抄截	0.10177 0.2076	0.15395 0.0558	0.00716 0.9296	0.05020 0.5351	0.08889 0.2714	1.00000	0.07709 0.3404
x6 助攻	0.06064 0.4535	0.09232 0.2532	-0.06564 0.4171	0.20663 0.0099	0.08876 0.2721	0.07709 0.3404	1.00000

如何判斷：

(1) 相關係數 r 值介於 1 和 -1 之間，可分為高度、中度、低度正（負）相關。

(2) 相關係數 r 值為介判定原則為：

(a) $0.7 < |r| \leq 1$ → 高度相關

(b) $0.3 \leq |r| \leq 0.7$ → 中度相關

(c) $0 \leq |r| < 0.3$ → 低度相關

由上表可知得分(Y)和出手次數(X_1)的相關程度是 0.77491，得分(Y)和命中率(X_2)的相關程度是 0.49758，得分(Y)和籃板數(X_3)的相關程度判定是

0.09193，得分(Y)和失誤次數(X_4)的相關程度是 0.15615，得分(Y)和抄截次數(X_5)的相關程度是 0.10177，得分(Y)和助攻數(X_6)的相關程度是 0.06064。根據 r 的判定原則得知，得分(Y)和出手次數(X_1)呈高度相關，得分(Y)和命中率(X_2)呈中度相關，得分(Y)和籃板數(X_3)失誤次數(X_4)、抄截次數(X_5)、助攻數(X_6)呈現低度相關。



三、原始模式檢定

3-1. 建立迴歸模式

影響得分(Y)的因素有出手次數(X₁)、命中率(X₂)、籃板(X₃)、失誤(X₄)、抄截(X₅)和助攻(X₆)。

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + \varepsilon_i, \quad i=1, 2, 3, \dots, 155。$$

$$E(Y_i) = \bar{\beta}_0 + \bar{\beta}_1 X_{1i} + \bar{\beta}_2 X_{2i} + \bar{\beta}_3 X_{3i} + \bar{\beta}_4 X_{4i} + \bar{\beta}_5 X_{5i} + \bar{\beta}_6 X_{6i} + \varepsilon_i, \quad i=1, 2, 3, \dots, 155。$$

為驗證以上的基本假設，所以我們將所有的可能變數皆列入解釋變數。

令 X₁=出手次數 X₂=命中率 X₃=籃板數

X₄=失誤次數 X₅=抄截次數 X₆=助攻次數

以下我們利用 SAS 建立迴歸模型，建立迴歸模型。

表 3. 參數估計表

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	Intercept	1	-9.03176	1.24125	-7.28	<0.0001
x1	出手數	1	1.22000	0.05539	22.03	<0.0001
x2	命中率	1	0.20603	0.01474	13.98	<0.0001
x3	籃板數	1	0.14978	0.13503	1.11	0.2691
x4	失誤	1	0.02626	0.12664	0.21	0.8360
x5	抄截	1	-0.18163	0.27574	-0.66	0.5111
x6	助攻	1	0.02854	0.06773	0.42	0.6741

經由上列資料可以得到：

$$\begin{aligned}\widehat{\beta}_0 &= -9.03176 & \widehat{\beta}_1 &= 1.22000 & \widehat{\beta}_2 &= 0.20603 \\ \widehat{\beta}_3 &= 0.14978 & \widehat{\beta}_4 &= 0.02626 & \widehat{\beta}_5 &= -0.18163 \\ & & \widehat{\beta}_6 &= 0.02854 & & \end{aligned}$$

Model :

$$\widehat{Y} = -9.03176 + 1.22X_1 + 0.20603X_2 + 0.14978X_3 + 0.02626X_4 - 0.18163X_5 + 0.02854X_6$$

3-2. 參數檢定

我們得到 β 值之後，接下來則是要判斷各個解釋變數出手次數(X_1)、命中率(X_2)、籃板數(X_3)、失誤次數(X_4)、抄截次數(X_5)、助攻次數(X_6)之間是否有線性相關。

(1) 我們想要判斷出手次數(X_1)與得分(Y)之間是否存在線性關係，首先我們先假定其他變數固定的情況下，統計基本假設：

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

因為 $p\text{-value} < 0.0001 < \alpha = 0.05$ 所以拒絕 H_0 的假設，表示我們有充分的資料顯示 $\beta_1 \neq 0$ 。即表示出手次數(X_1)與得分(Y)有存在線性關係。

(2) 我們想要判斷命中率(X_2)與得分(Y)之間是否存在線性關係，首先我們先假定其他變數固定的情況下，統計基本假設：

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

因為 $p\text{-value} < 0.0001 < \alpha = 0.05$ 所以拒絕 H_0 的假設，表示我們有充分的資料顯示 $\beta_2 = 0$ 。即表示命中率(X_2)與得分(Y)之間存在線性關係。

(3) 我們想要判斷籃板數(X_3)與得分(Y)之間是否存在線性關係，首先我們先假定其他變數固定的情況下，統計基本假設：

$$H_0 : \beta_3 = 0$$

$$H_1 : \beta_3 \neq 0$$

因為 $p\text{-value}=0.2691 > \alpha=0.05$ 所以不拒絕 H_0 的假設，表示我們沒有充分的資料顯示 $\beta_3 \neq 0$ 。即表示籃板(X_3)與得分(Y)沒有存在線性關係。

(4) 我們想要判斷失誤次數(X_4)與得分(Y)之間是否存在線性關係，首先我們先假定其他變數固定的情況下，統計基本假設：

$$H_0 : \beta_4 = 0$$

$$H_1 : \beta_4 \neq 0$$

因為 $p\text{-value}=0.8360 > \alpha=0.05$ 所以不拒絕 H_0 的假設，表示我們沒有充分的資料顯示 $\beta_4 \neq 0$ 。即表示失誤次數(X_4)與得分(Y)之間沒有存在線性關係。

(5) 我們想要判斷抄截次數(X_5)與得分(Y)之間是否存在線性關係，首先我們先假定其他變數固定的情況下，統計基本假設：

$$H_0 : \beta_5 = 0$$

$$H_1 : \beta_5 \neq 0$$

因為 $p\text{-value}=0.5111 > \alpha=0.05$ 所以不拒絕 H_0 的假設，表示我們沒有充分的資料顯示 $\beta_5 \neq 0$ 。即表示抄截(X_5)與得分(Y)之間沒有存在線性關係。

(6) 我們想要判斷助攻次數(X_6)與得分(Y)之間是否存在線性關係，首先我們先假定其他變數固定的情況下，統計基本假設：

$$H_0 : \beta_6 = 0$$

$$H_1 : \beta_6 \neq 0$$

因為 $p\text{-value}=0.6741 > \alpha=0.05$ 所以不拒絕 H_0 的假設，表示我們沒有充分

NBA 球員得分之迴歸分析〈以 Steve Nash 為例〉

的資料顯示 $\beta_6 \neq 0$ 。即表示助攻次數(X_6)與得分(Y)之間沒有存在線性關係。



3-3. 模型適合度檢定

以下為我們利用SAS 做適合度檢定，建立回歸模型：

表 4. 適合度檢定表

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	6774.22955	1129.03826	121.09	<0.0001
Error	148	1379.98980	9.32426		
Corrected Total	154	8154.21935			

統計假設如下：

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$$

$$H_1: \beta_i \text{ 不完全為 } 0, i = 1, 2, 3, 4, 5, 6$$

因為 $p\text{-value} < 0.0001 < \alpha = 0.05$ 所以拒絕 H_0 的假設，表示我們有充分的資料顯示 β_i 不完全為 0，即表示解釋變數(出手次數、命中率、籃板、失誤、抄截、助攻)與反應變數(得分)有線性迴歸相關。

3-4. 模型解釋能力

表 5. 模型解釋能力分析表

Root MSE	3.05356	R-Square	0.8308
Dependent Mean	16.32903	Adj R-Sq	0.8239
CoeffVar	18.70022		

$$R\text{-Square} = SSR/SSTO = 0.8308$$

$$\text{Adj R-Square} = 1 - MSE/MSTO = 0.8239$$

在全模型下的 R^2 值為0.8308，表示樣本回歸直線已解釋了總變異的83.08%，而經過校正 R_a^2 只比 R^2 低0.0069，所以代表 X_i ， $i=1, 2$ 對 Y 具有相當程度的解釋能力，因此下一步要做選取變數的分析來判定最終的模型。

四、模式的選取方法

4-1. 前進選擇法(Forward)slentry=0.05

在 SAS 的 PROC REG 程序中選取解釋變數，若採行前進選擇法(FORWARD)，其選擇進入模式內的解釋變數之基本方法是，在給定已進入模式的變數之條件下，在尚未進入模式內的個別變數中，選取進入模式之後其參數估計值的 F 統計量最顯著者(亦即 p-value 最小者)優先進入，直到模式外的每一個變數加進模式內，其參數估計值的 F 統計量皆不顯著時，則選取解釋變數進入模式的工作停止。此外，前進選擇法有一原則，一旦變數進入模式後，則此變數將不會再由模式中剔除，始終保留在模式內。

令 X_1 =出手次數 X_2 =命中率 X_3 =籃板數

X_4 =失誤次數 X_5 =抄截次數 X_6 =助攻次數

表 6. Summary of Forward Selection

Summary of Forward Selection								
Step	Variable Entered	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x1	出手數	1	0.6005	0.6005	198.380	229.97	<.0001
2	x2	命中率	2	0.2279	0.8284	1.1026	201.80	<.0001

Step 1：系統選取的解釋變數是 X_1 ，其 F 統計量為 229.97 (p-value< α =0.05)。

Step 2：系統在其餘 5 個變數(X_2 、 X_3 、 X_4 、 X_5 、 X_6)中選取最顯著的變數 X_2 ，其 F 統計量為 201.80(p-value< α =0.05)進入模式內。之後，系統發現尚未進入模式的 X_3 、 X_4 、 X_5 、 X_6 ，在顯著水準 0.05 下並不顯著，因此不放進模式中。利用前進選則法選取模式解釋變數(X_1 、 X_2)的線性回歸模式完成，其模型如下：

$$\hat{Y} = -8.31205 + 1.22062X_1 + 0.20650X_2$$

4-2. 後退選取法(Backward)slentry=0.05

在 SAS 的 PROC REG 語法中，採用後退刪去法，其選擇解釋變數的基本方式是將所有解釋變數皆放入模式中，然後再逐步地把不顯著的變數從模式中剔除，剔除的原則是，在每個階段在模式內找尋停留在模式中其 F 統計量最小最不顯著（亦即 p-value 最大者）的變數，將其剔除，一直到模式內的變數皆顯著時則停止。此外，後退刪去法有一個原則，就是在每個階段中被剔除在模式外的變數，將不會再進入此模式之內。

令 X_1 =出手次數 X_2 =命中率 X_3 =籃板
 X_4 =失誤 X_5 =抄截 X_6 =助攻

表 7. Summary of Backward Elimination

Summary of Backward Elimination								
Step	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x4	失誤	5	0.0000	0.8307	5.0430	0.04	0.8360
2	x6	助攻	4	0.0002	0.8305	3.2324	0.19	0.6630
3	x5	抄截	3	0.0004	0.8301	1.6213	0.39	0.5314
4	x3	籃板數	2	0.0017	0.8284	1.1026	1.51	0.2218

Step 1：系統將所有解釋變數都放入模式內，而由模式所求得的 F 統計量得知，變數 X_4 的 F 統計量為 0.04 (p-value=0.8360>0.05) 是最不顯著的，所以此變數應首先被剔除。

Step 2：系統再根據其餘的變數(X_1 、 X_2 、 X_3 、 X_5 、 X_6)再重新建立模式，而由模式所求得的 F 統計量得知，變數 X_6 的 F 統計量為 0.19(P-Value=0.6630>0.05) 是最不顯著的，所以剔除此變數。

Step 3：系統再根據其餘的變數(X_1 、 X_2 、 X_3 、 X_5)再重新建立模式，而由模式所求得的 F 統計量得知，變數 X_5 的 F 統計量為 0.39(P-Value=0.5314>0.05) 是最不顯著的，所以剔除此變數。

Step 4：系統再根據其餘的變數(X_1 、 X_2 、 X_3)再重新建立模式，而由模式所求得的 F 統計量得知，變數 X_3 的 F 統計量為 1.51(P-Value=0.2218>0.05) 是最不顯著的，所以剔除此變數。

Step 5：系統再根據其餘的變數(X_1 、 X_2)再重新建立模式而由模式所求得的 F 統計量得知，該階段兩個解釋變數的 F 統計量再給定的顯著水準 0.05 下皆具顯著性。此時，利用後退消除法選取的解釋變數(X_2 、 X_4 、 X_5)的線性迴歸模式完成，其模型如下：

$$\hat{Y} = -8.31205 + 1.22062X_1 + 0.20650X_2$$

4-3. 逐步迴歸法 slentry=0.05 slstay=0.05

逐步迴歸法其實是交叉混合應用了「前進選擇法」及「後退消除法」的觀念。首先，我們先使用前進選擇法將模式外所剩餘的變數選取進入模式後，其 F 值最顯著(亦即 p-value 最小者)進入模式中；而後再利用後退消除法對已經留在模式內的變數重新進行檢定工作。而此時若有某個變數的 F 統計量不顯著時，則將此變數移除到模式之外，直到

- (1)在模式外的變數其欲進入模式內的 F 統計量皆不顯著時或
- (2)欲再進入模式中的變數是之前才被系統由模式中剔除出來的變數時就停止執行的工作。

此時，留在模式內的變數才是最後所選取的解釋變數。

Step 1

Variable x1 Entered: R-Square = 0.6005 and C(p) = 198.3796

表 8. 逐步迴歸 STEP1-1

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4896.51500	4896.51500	229.97	<.0001
Error	153	3257.70436	21.29219		
Corrected Total	154	8154.21935			

表 9. 逐步迴歸 STEP1-2

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	1.81644	1.02626	66.70291	3.13	0.0787
x1	1.24073	0.08182	4896.51500	229.97	<.0001

Step 2

Variable x2 Entered: R-Square = 0.8284 and C(p) = 1.1026

表 10. 逐步迴歸 STEP2-1

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	6754.62439	3377.31219	366.79	<.0001
Error	152	1399.59497	9.20786		
Corrected Total	154	8154.21935			

表 11. 逐步迴歸 STEP2-2

Variable	Parameter Estimate	Standard Error	Type II SS	F Value	Pr > F
Intercept	-8.31205	0.98175	660.04459	71.68	<.0001
x1	1.22062	0.05382	4735.76114	514.32	<.0001
x2	0.20650	0.01454	1858.10939	201.80	<.0001

All variables left in the model are significant at the 0.0500 level.

No other variable met the 0.0500 significance level for entry into the model.

表 12. Summary of Stepwise Selection

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	x1		出手數	1	0.6005	0.6005	198.380	229.97	<.0001
2	x2		命中率	2	0.2279	0.8284	1.1026	201.80	<.0001

Step 1：所選取的變數是 X_1 ，其 $F=229.97$ ($p\text{-value}<.0001<0.05$)是五個解釋變數中最顯著的；接著系統將會檢定模式內的唯一解釋變數 X_1 ，結果發現是顯著的，因此不再將其由模式中剔除。

Step 2：由剩下的五個解釋變數中(X_2 、 X_3 、 X_4 、 X_5 、 X_6)選取 F 統計量較高且仍顯著的變數 X_2 ，其 $F=201.80$ ($p\text{-value}<.0001<0.05$)，是五個解釋變數中最顯著的；系統將會檢定模式內唯一的解釋變數 X_2 ，發現是顯著的，進入模式內。接著，系統會重新檢定新建立的模式中四個解釋變數(X_3 、 X_4 、 X_5 、 X_6)皆不顯著，因此由模式中剔除。最後，系統在逐步迴歸法所選取的(X_1 、 X_2)三個變數所建立迴歸模式如下：

$$\hat{Y} = -8.31205 + 1.22062X_1 + 0.20650X_2$$

4-4. 其他選取法

表 13. 其他選取法

Number in Model	R-Square	Adjusted R-Square	C(p)	AIC	MSE	SBC	Variables in Model
1	0.6005	0.5979	198.3796	476.0297	21.29219	482.11655	x1
1	0.2476	0.2427	506.9996	574.1517	40.10037	580.23858	x2
1	0.0244	0.0180	702.1926	614.4178	51.99599	620.50464	x4
2	0.8284	0.8261	1.1026	347.0795	9.20786	356.20980	x1 x2
2	0.6042	0.5989	197.1731	476.6006	21.23557	485.73089	x1 x3
2	0.6030	0.5978	198.1684	477.0456	21.29663	486.17590	x1 x4
3	0.8301	0.8267	1.6213	347.5423	9.17737	359.71598	x1 x2 x3
3	0.8288	0.8254	2.6890	348.6518	9.24330	360.82553	x1 x2 x6
3	0.8287	0.8253	2.7790	348.7450	9.24886	360.91869	x1 x2 x5
4	0.8305	0.8260	3.2324	349.1362	9.21438	364.35335	x1 x2 x3 x5
4	0.8302	0.8257	3.4617	349.3758	9.22863	364.59292	x1 x2 x3 x6
4	0.8301	0.8256	3.5841	349.5035	9.23624	364.72059	x1 x2 x3 x4
5	0.8307	0.8250	5.0430	350.9380	9.26437	369.19856	x1 x2 x3 x5 x6
5	0.8306	0.8249	5.1776	351.0788	9.27279	369.33940	x1 x2 x3 x4 x5
5	0.8303	0.8246	5.4339	351.3467	9.28883	369.60727	x1 x2 x3 x4 x6
6	0.8308	0.8239	7.0000	352.8930	9.32426	374.19695	x1 x2 x3 x4 x5 x6

(1) 複判定係數法(R^2) Coefficient of Multiple Determination

$R^2 = SSR/SSTO = 1 - SSE/SSTO$ ，它表示 Y 之總變異中和使用諸 X 變數有關的部份。高的 R^2 值並不一定表示配適的模型有用。且即使 R^2 很高，其 MSE 可能太大而實際上卻需要高精確度故不符實用。在 SAS 的選取變數方法中，選擇 R^2 法並不

是在多個解釋變數中找出唯一的最佳迴歸模式出來，而是分別在不同個數的解釋變數組合中列出變數所有可能組合情形的 R^2 出來，而不作其他任何的估計；著手尋找所有自變數的任意組合中，相互比較，以選取最大的 R-square 為最佳、最有效的迴歸模式。

由此種方法的分析可發現投入變數 X_1 、 X_2 時， $R^2=0.8284$ 與投入 X_1 、 X_2 、 X_3 、 X_4 、 X_5 、 X_6 時的 $R^2=0.8308$ ，相差並不大，及表示多投入 X_3 、 X_4 、 X_5 、 X_6 變數的貢獻並沒有增加很多，所以選擇投入 X_1 、 X_2 解釋變數為最佳的迴歸模型組合。

(2) 校正的複判定係數法 (R_a^2)

Adjusted Coefficient of Multiple Determination

$$R_a^2 = 1 - (SSE/n-p) / (SSTO/n-1) = 1 - (n-1/n-p) * (SSE/SSTO)$$

在迴歸模型中增加 X 變數只會提高 R^2 而不會使它降低，因為更多 X 變數只可能降低 SSE，而對一組給定反應值 SSTO 固定。由於 R^2 總是可因採用大量預測變數而增大，故有些學者建議採取校正的複判定係數。是將計算 R^2 公式的平方和以其對應自由度除之。

其結果與 R^2 法相似，但 R_a^2 是將不同個數的解釋組合視為不同分群並再行根據 R^2 的大小來排列，而是將所有解釋變數的組合視為一整群，並各自的根據 R_a^2 的大小來排列之，所以依其法以找 R_a^2 最大者或 MSE 最小者來判斷。此法當模型引進新預測變數時實際上可能降低，因為 SSE 降低的幅度可能不及其自由度 $n-p$ 損失所抵消的效果。

而我們發現當投入 X_1 、 X_2 、 X_3 時，其 $R_a^2=0.8267$ 為最大，且 $MSE=9.17737$ 為最小，所以選擇投入解釋變數 X_1 、 X_2 、 X_3 當作最佳迴歸模式組合。

(3) Cp 準則

此一準則是關心每一個被考慮的子集，其 n 個配適值的總均方誤差 (total mean squared error) 估算所有自變數的任意組合中之 C_p 值，相互比較，當 C_p 值估算與變數個數 (p) 最接近時為最佳組合。

由此方法，我們可發現投入變數 X_1 、 X_2 、 X_4 、 X_5 、 X_6 時的 $C_p=6.2305$ 最接近 $p(=6)$ ，所以選擇投入 X_1 、 X_2 、 X_4 、 X_5 、 X_6 解釋變數當作最佳迴歸模式組合。

(4) AIC、SBC 法

估算所有自變數的任意組合中之 SBC、AIC 值，相互比較，選取 SBC、AIC 值最小者為最佳的迴歸模式。

由此兩種方法我們發現投入變數 X_1 、 X_2 時 AIC=347.0795 為最小及 SBC=356.20980 也是為最小。所以選 X_1 、 X_2 解釋變數當作最佳迴歸模式組合。

4-5. 結論

在信心水準為 95% 的情形下：

表 14. 選取模式表

方法	選擇的變數
前進選擇法(FORWARD)	X_1 、 X_2
後退消除法(BACKWARD)	X_1 、 X_2
逐步迴歸法(STEPWISE)	X_1 、 X_2
複判定係數法(R^2)	X_1 、 X_2
校正的複判定係數法(R_a^2)	X_1 、 X_2 、 X_3
Cp 準則	X_1 、 X_2 、 X_4 、 X_5 、 X_6
SBC、AIC 法	X_1 、 X_2

綜合以上各種檢定方法，因此我們應該選擇的變數為出手次數(X_1)、命中率(X_2)來進行迴歸分析，以求得最佳迴歸模式：

$$\hat{Y} = -8.31205 + 1.22062X_1 + 0.20650X_2$$

五、殘差檢定

當模式為： $\hat{Y} = -8.31205 + 1.22062X_1 + 0.20650X_2$

5-1. 檢測常態

(1) 方法 1

設 H_0 ：誤差項具常態分配性質

H_1 ：誤差項不具常態分配性質

表 15. 常態檢定表

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.987274	Pr < W	0.2015
Kolmogorov-Smirnov	D	0.043675	Pr > D	>0.1500
Cramer-von Mises	W-Sq	0.067844	Pr > W-Sq	>0.2500
Anderson-Darling	A-Sq	0.462275	Pr > A-Sq	>0.2500

根據 Shapiro-Wilk test，P-value=0.2015>0.05 之信賴水準，所以不拒絕 H_0 ：常態分配性質。

(2)方法 2

①由莖葉圖和箱形圖可看出，圖形均呈現左右對稱的形狀，所以我們可以得知其服從常態分配。

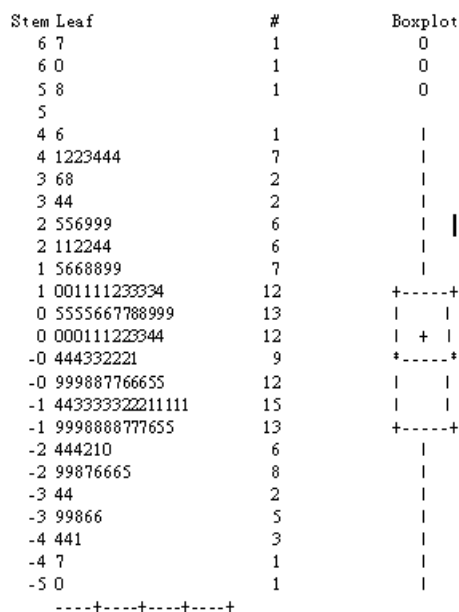


圖 8. 莖葉圖和箱形圖

②由常態分配圖來看，所有的數值在圖上呈現斜率約為 1 的圖形，所以我們可以得知其服從常態分配。

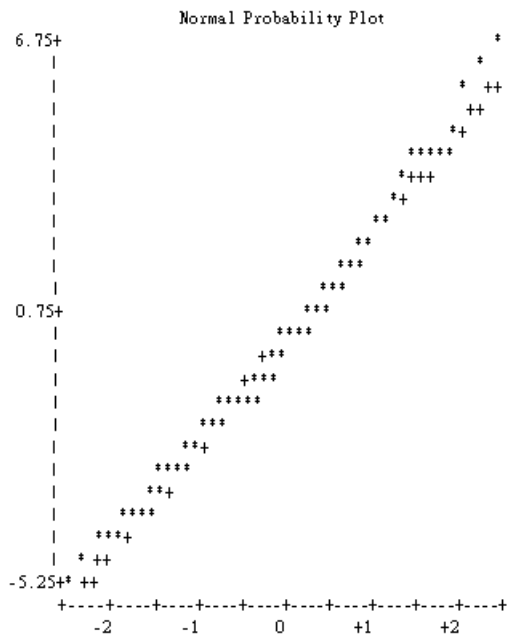


圖 9. 常態分配圖

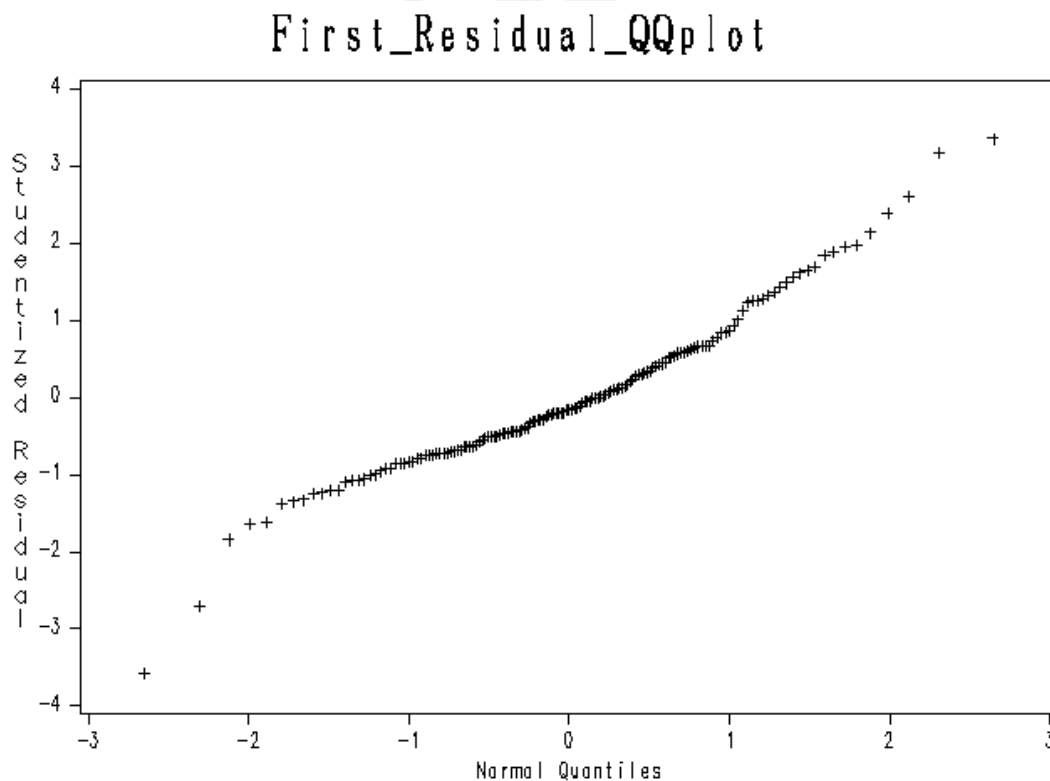


圖 10. 原始資料部分(Nomal q-q plot)

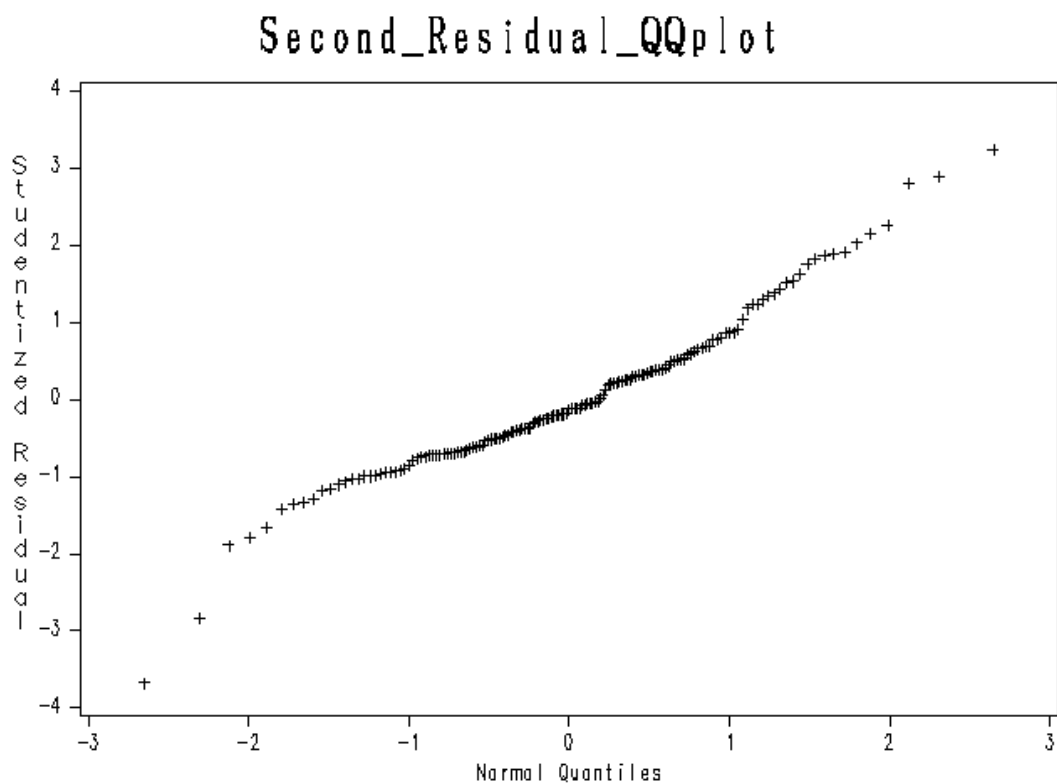


圖 11. 經各種檢定後留下的變數部分(X1&X2)(Normal q-q plot)

③標準化殘差值次數分配直方圖

此直方圖可作為檢定樣本觀察值是否符合常態性的基本假設，當殘差值的次數分配與圖中的鐘型曲線（標準常態分配曲線）完全符合時，即表示樣本的標準化殘差值呈完全常態分配，本研究的殘差值雖並不完全符合標準常態分配圖但也不能否決觀察值不符合常態性假設。由於在實際研究中多少會產生有抽樣誤差（sampling error）的存在，因此以樣本殘差值常態機率分佈圖，來檢定較為適當。

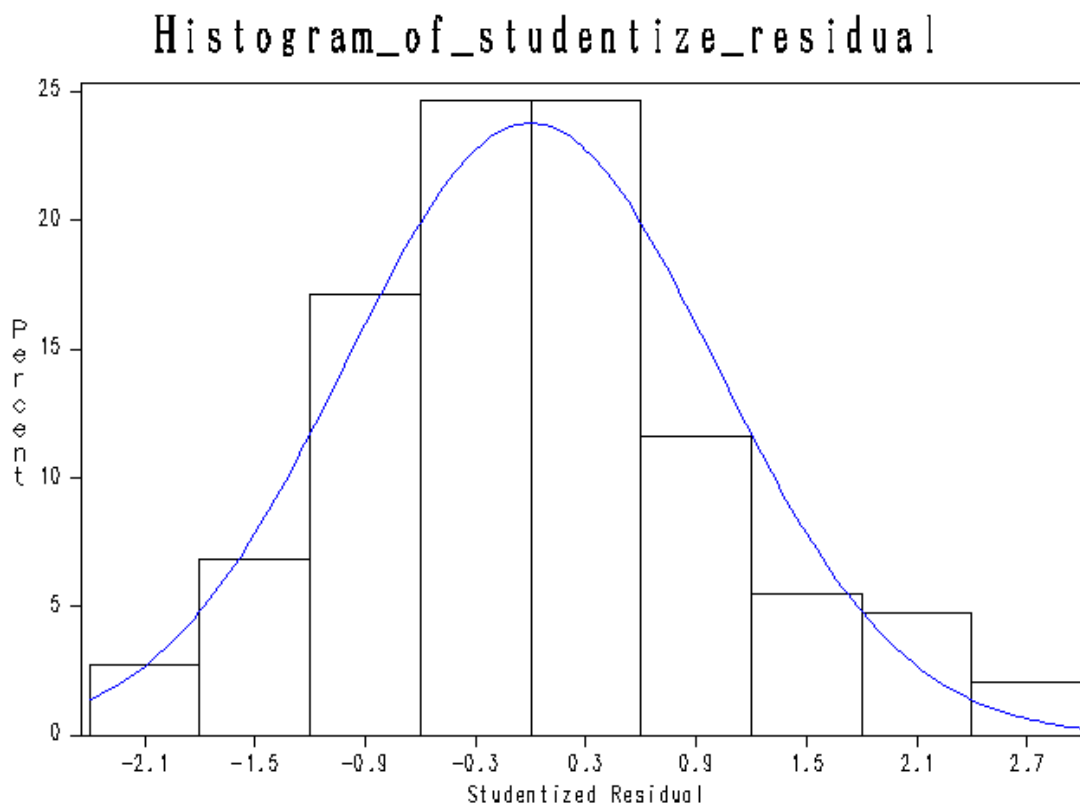


圖 12. 標準化殘差值次數分配直方圖

5-2. 檢測獨立

表 16. 獨立檢測表

Dependent Variable: y

Durbin-Watson D	1.956
Pr < DW	0.3920
Pr > DW	0.6080
Number of Observations	146

Durbin-Watson D	1.956
1st Order Autocorrelation	0.020

若 $D > DU \rightarrow$ 結論不拒絕 H_0 ；若 $D < DL \rightarrow$ 結論拒絕 H_0 ；若 $DU > D > DL \rightarrow$ 無結論經過判斷 $\rightarrow D = 1.956 > DU$ ，所以不拒絕 H_0 ，則可以視為該模式的殘差項皆無相關存在，即誤差項具獨立性。

5-3. 檢測變異數

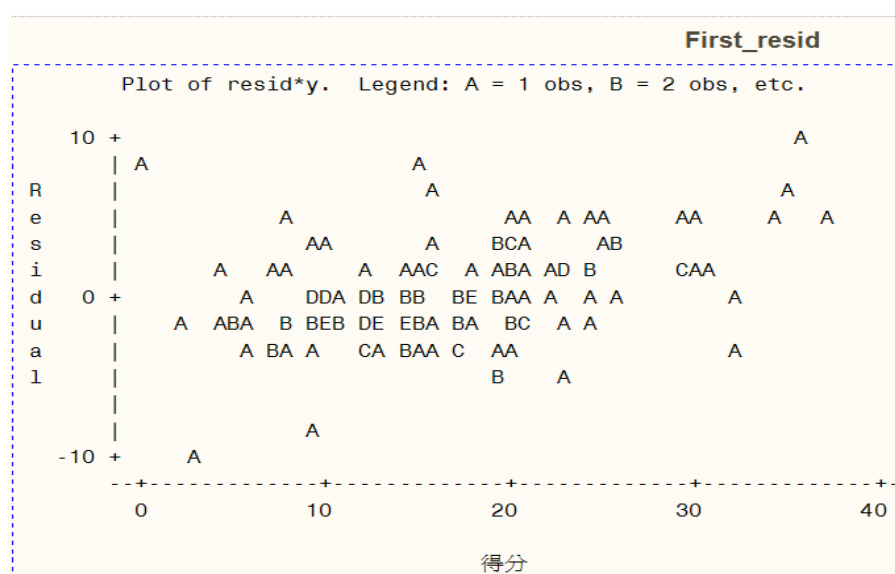


圖 13. 檢測變異數是否齊一及是否線性相關

(1) 第一部分:原始資料

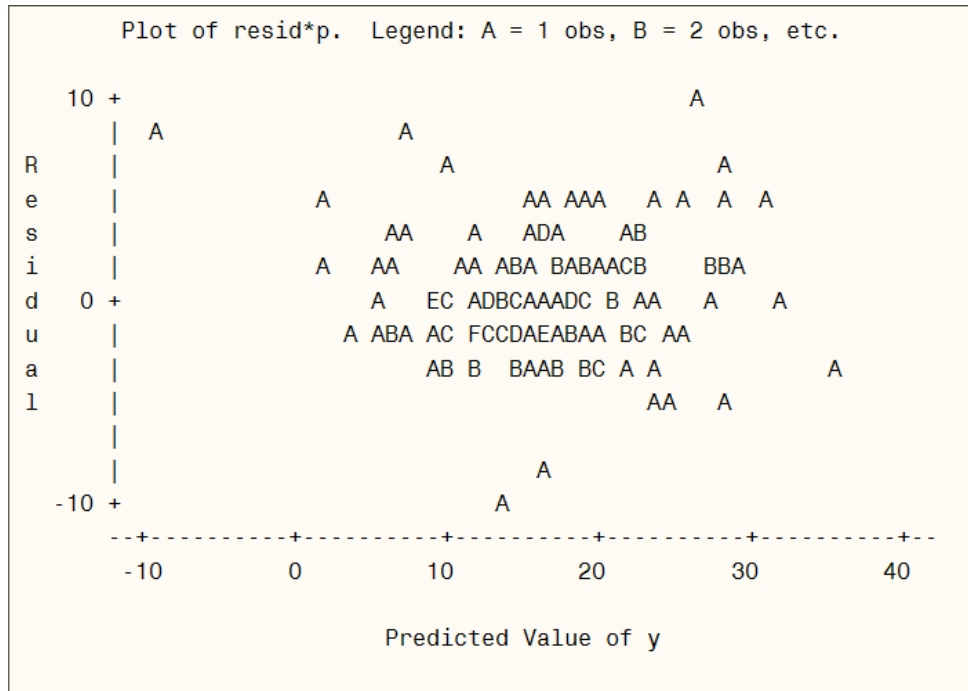


圖 14. 原始資料殘差圖

(2) 第二部分:最後留下變數部分

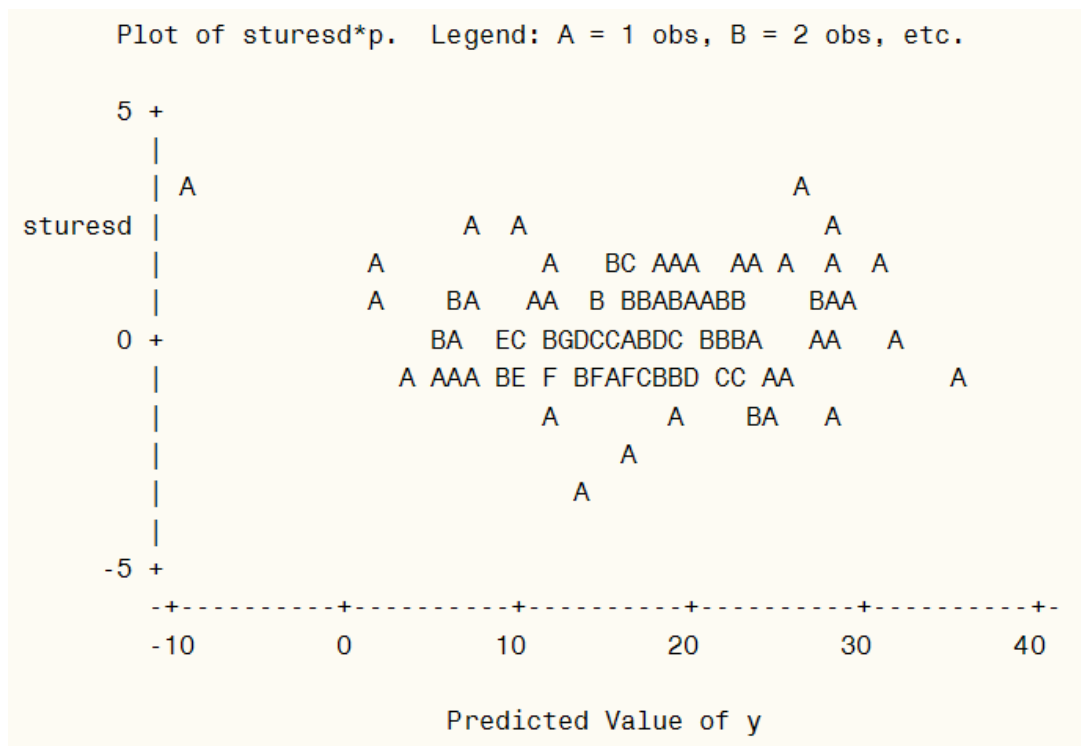


圖 15. 最後留下變數部分殘差圖



可知，迴歸函數具直線性及誤差變異數為常數。另外我們也可以由下表得知，90%殘差值落在1.4526392至-1.1799720之間。

表 17. 殘差值估計表

Quantile	Estimate
100% Max	2.9169027
99%	2.6508793
95%	1.8220605
90%	1.4526392
75% Q3	0.5758795
50% Median	-0.0703397
25% Q1	-0.6839298
10%	-1.1799720
5%	-1.6367895
1%	-2.1100747
0% Min	-2.1758434

六、結論

我們的報告主要是要探討「出手次數」、「命中率」、「籃板」、「失誤」、「抄截」、「助攻」此六個變數，對得分的影響。我們使用前進選取法、後退消除法、逐步迴歸法、其他選取法(R^2 、 R_a^2 、 C_p 法、SBC法和AIC法)來刪去不合適的變數，選擇最佳的變數組合。在檢定過後，我們選擇刪除「籃板」、「失誤」、「抄截」、「助攻」這四項不適合的預測變數。

接著利用剩下的兩個變數預測一個複迴歸模式，得到一個複迴歸模式為

$$\hat{Y} = -8.31205 + 1.22062X_1 + 0.20650X_2$$

之後以此迴歸模式進行殘差分析，符合三大假設：獨立性、常態性、一致性，故此迴歸模式為最佳線性迴歸模式。

在最佳線性迴歸模式 $\hat{Y} = -8.31205 + 1.22062X_1 + 0.20650X_2$ 中，各參數代表的意義為：

β_0 ：當不考慮 X_1 、 X_2 影響的情況下，即 X_1 、 $X_2=0$ 時， Y 平均的值為 -8.31205

β_1 ：當其他變數 X_2 不變時，每增加一單位的 X_1 ， Y 的平均值會增加 1.22062 即不考慮命中率的情況下，當出手次數每增加一單位，得分之平均數則會增加 1.22062 分。

β_2 ：當其他變數 X_1 不變時，每增加一單位的 X_2 ， Y 的平均值會增加 0.20650，即不考慮出手次數情況下，當命中率每增加一單位，得分之平均數會增加 0.20650 分。

在此份報告中，校正後的 R-square=0.8239，對於模型可解釋的總變異來說，解釋能力是很高的。因此我們可以說：當出手次數越多、命中率越高、得分會越高。

當球員出手次數越多、命中率越高，可能代表球員不斷努力或是積極爭取機會，在相對的情況之下，球員得分也會相對增加。

七、附錄

7-1. 原始檔案

Obs	出手數	命中率	籃板數	助攻	抄截	失誤	得分	n	F9
1	1	100.0	2	5	0	1	3	1	
2	11	36.4	9	14	2	5	13	2	
3	9	22.2	6	9	0	1	6	3	
4	10	40.0	5	10	0	0	12	4	
5	11	45.5	2	8	0	2	16	5	
6	17	23.5	5	10	0	5	14	6	
7	6	83.3	3	11	1	7	14	7	
8	14	42.9	4	18	3	4	17	8	
9	18	61.1	3	8	1	0	36	9	
10	9	22.2	3	11	1	3	10	10	
11	8	25.0	2	10	1	4	5	11	
12	10	50.0	2	9	0	1	12	12	
13	14	64.3	3	9	0	1	23	13	
14	10	50.0	4	5	0	6	15	14	
15	13	69.2	2	11	1	6	23	15	
16	16	56.3	0	8	0	1	22	16	
17	9	77.8	2	8	1	2	18	17	
18	13	53.8	4	13	0	3	21	18	
19	7	57.1	3	8	0	2	8	19	
20	12	58.3	5	14	0	4	19	20	
21	14	50.0	6	15	1	3	17	21	

NBA 球員得分之迴歸分析〈以 Steve Nash 為例〉

Obs	出手數	命中率	籃板數	助攻	抄截	失誤	得分	n	F9
22	11	36.4	4	13	0	5	12	22	
23	12	50.0	3	4	0	1	19	23	
24	13	46.2	3	11	0	6	21	24	
25	6	16.7	0	13	0	4	8	25	
26	11	72.7	2	13	0	4	25	26	
27	8	50.0	2	6	0	4	11	27	
28	12	58.3	7	5	2	5	18	28	
29	17	52.9	2	8	1	1	26	29	
30	17	52.9	6	13	0	5	24	30	
31	13	53.8	4	12	2	7	20	31	
32	6	50.0	3	12	0	4	10	32	
33	13	30.8	4	8	2	2	13	33	
34	23	56.5	4	12	1	10	32	34	
35	10	30.0	3	11	0	5	8	35	
36	18	33.3	2	8	1	3	17	36	
37	8	50.0	4	10	3	4	10	37	
38	10	50.0	2	7	0	6	14	38	
39	11	81.8	4	9	0	7	26	39	
40	16	43.8	3	16	0	4	21	40	
41	23	56.5	4	10	2	2	37	41	
42	11	45.5	3	9	0	3	13	42	
43	6	100.0	4	13	0	1	17	43	
44	15	33.3	0	20	2	5	13	44	
45	14	35.7	3	13	0	6	14	45	

NBA 球員得分之迴歸分析〈以 Steve Nash 為例〉

Obs	出手數	命中率	籃板數	助攻	抄截	失誤	得分	n	F9
46	17	76.5	2	7	1	3	35	46	
47	8	37.5	2	4	1	4	9	47	
48	11	45.5	4	10	1	3	13	48	
49	14	64.3	3	11	0	4	23	49	
50	11	63.6	4	10	1	5	17	50	
51	11	36.4	5	15	2	2	12	51	
52	7	42.9	6	10	0	4	10	52	
53	7	42.9	1	12	0	3	9	53	
54	18	44.4	2	14	3	3	24	54	
55	7	42.9	6	16	0	0	7	55	
56	15	60.0	3	18	1	5	21	56	
57	8	37.5	8	10	0	3	10	57	
58	12	33.3	3	7	0	2	12	58	
59	12	83.3	6	11	1	6	29	59	
60	13	30.8	0	14	2	4	11	60	
61	14	35.7	4	15	0	2	14	61	
62	9	44.4	5	19	0	3	12	62	
63	7	57.1	2	18	0	3	10	63	
64	9	66.7	6	17	1	3	18	64	
65	10	30.0	3	15	0	1	9	65	
66	8	25.0	6	14	0	2	15	66	
67	20	55.0	2	6	0	3	29	67	
68	14	57.1	2	13	0	8	23	68	
69	9	66.7	2	10	1	4	20	69	

NBA 球員得分之迴歸分析〈以 Steve Nash 為例〉

Obs	出手數	命中率	籃板數	助攻	抄截	失誤	得分	n	F9
70	10	70.0	4	15	0	1	18	70	
71	14	35.7	3	12	1	4	14	71	
72	11	54.5	3	15	1	5	19	72	
73	11	27.3	2	15	4	5	10	73	
74	6	33.3	3	12	1	6	5	74	
75	13	69.2	5	11	0	2	19	75	
76	16	68.8	2	8	0	4	30	76	
77	19	63.2	5	11	0	6	34	77	
78	8	50.0	6	7	0	0	10	78	
79	19	63.2	4	10	1	5	30	79	
80	15	46.7	0	3	1	5	19	80	
81	13	46.2	7	12	1	7	18	81	
82	7	57.1	2	10	0	3	10	82	
83	10	70.0	0	12	0	4	18	83	
84	3	33.3	3	7	1	2	4	84	
85	4	100.0	3	6	2	2	9	85	
86	9	77.8	2	13	1	1	24	86	
87	12	33.3	2	8	1	2	12	87	
88	19	63.2	3	9	1	3	29	88	
89	15	60.0	1	17	3	1	25	89	
90	21	57.1	4	14	0	5	31	90	
91	21	33.3	2	6	0	2	20	91	
92	13	46.2	2	5	0	3	15	92	
93	10	40.0	4	14	0	3	12	93	

NBA 球員得分之迴歸分析〈以 Steve Nash 為例〉

Obs	出手數	命中率	籃板數	助攻	抄截	失誤	得分	n	F9
94	10	50.0	3	9	1	0	15	94	
95	12	66.7	0	4	0	0	17	95	
96	18	55.6	5	10	0	2	24	96	
97	8	50.0	7	9	1	2	11	97	
98	14	57.1	1	8	3	7	18	98	
99	15	60.0	2	6	0	2	20	99	
100	17	58.8	0	13	0	4	23	100	
101	22	45.5	2	11	0	3	23	101	
102	27	51.9	5	13	1	1	32	102	
103	17	70.6	5	10	1	3	29	103	
104	14	57.1	7	8	4	7	20	104	
105	10	80.0	2	5	0	6	22	105	
106	20	40.0	0	11	1	2	19	106	
107	9	55.6	3	8	1	0	13	107	
108	11	54.5	3	12	0	3	21	108	
109	4	50.0	1	10	1	1	8	109	
110	8	12.5	2	8	0	3	2	110	
111	12	41.7	3	21	1	2	15	111	
112	19	47.4	5	8	4	5	19	112	
113	16	37.5	3	9	0	5	16	113	
114	6	50.0	3	9	1	3	9	114	
115	14	50.0	2	10	0	7	15	115	
116	12	41.7	2	18	0	3	16	116	
117	13	46.2	5	15	0	4	14	117	

NBA 球員得分之迴歸分析〈以 Steve Nash 為例〉

Obs	出手數	命中率	籃板數	助攻	抄截	失誤	得分	n	F9
118	7	57.1	3	13	1	6	14	118	
119	6	33.3	0	5	1	6	4	119	
120	9	33.3	6	19	0	4	9	120	
121	7	71.4	4	8	0	5	12	121	
122	6	33.3	6	18	1	2	9	122	
123	8	37.5	4	6	0	3	6	123	
124	10	50.0	7	14	0	6	20	124	
125	13	53.8	5	6	1	2	23	125	
126	8	62.5	5	12	1	4	14	126	
127	9	44.4	3	12	2	0	9	127	
128	9	44.4	4	12	0	4	16	128	
129	7	42.9	1	11	1	3	9	129	
130	0	0.0	0	2	0	1	0	130	
131	8	62.5	4	8	1	6	13	131	
132	11	45.5	2	11	2	3	16	132	
133	14	35.7	1	11	1	6	15	133	
134	18	44.4	6	6	2	3	21	134	
135	18	44.4	5	10	1	4	21	135	
136	12	16.7	1	9	0	4	7	136	
137	8	75.0	4	10	1	5	19	137	
138	14	28.6	5	9	1	4	12	138	
139	11	63.6	1	10	1	1	20	139	
140	20	55.0	1	9	2	3	26	140	
141	10	60.0	3	6	2	4	20	141	

NBA 球員得分之迴歸分析〈以 Steve Nash 為例〉

Obs	出手數	命中率	籃板數	助攻	抄截	失誤	得分	n	F9
142	14	57.1	8	15	0	7	20	142	
143	7	85.7	3	7	0	8	16	143	
144	9	33.3	1	10	0	2	8	144	
145	10	60.0	3	8	0	2	14	145	
146	10	60.0	3	7	0	4	17	146	
147	10	40.0	4	3	0	2	10	147	
148	11	27.3	2	6	0	6	12	148	
149	9	33.3	5	7	1	1	16	149	
150	7	28.6	0	5	0	3	6	150	
151	5	40.0	3	6	1	3	7	151	
152	7	71.4	2	11	1	4	12	152	
153	10	60.0	3	7	1	1	20	153	
154	16	56.3	4	9	0	7	24	154	
155	11	45.5	2	13	0	3	13	155	



7-2. SAS 程式碼

```
odsrtffile="h:\picture.rtf";
PROCIMPORTOUT= WORK.master
DATAFILE= "H:\第 6 組報告\nash.xls"
DBMS=EXCEL REPLACE;
    RANGE="Sheet1$";
    GETNAMES=YES;
    MIXED=NO;
    SCANTEXT=YES;
    USEDATE=YES;
    SCANTIME=YES;
RUN;

*迴規模式確立;
data a;
setwork.master ;
label y=得分 x1='出手數' x2='命中率' x3='籃板數' x4='失誤' x5='抄截'
x6='助攻' ;
procprintdata=a label;
title'原始資料';
run;

procglmdata=a ;
model y=x1-x6;
title'原始迴歸 model';
run;

proccorr;
var y x1-x6;
title'檢定是否有相關性';
run;

procregdata=a;
model y=x1-x6/selection=forward sle=0.05;
title'forward';
```

NBA 球員得分之迴歸分析〈以 Steve Nash 為例〉

```
proc reg data=a;
model y=x1-x6 / selection=backward sls=0.05;
title 'backward';
proc reg data=a;
model y=x1-x6 / selection=stepwise sls=0.05 sle=0.05;
title 'stepwise';
proc rsquare data=a adjrsqcpaicsbcmse;
model y=x1-x6 best=3;
title '其他選取變數方法';
run;

data b;
set a;

proc corr data=b;
var y x1 x2;
title '檢定是否有相關性';

data c;
set b;
n=_n_;
proc univariate normal plot;
var r;
title '殘差檢定';
run;

data d;
set b;
yhatt = -8.31205 + 1.22062 * X1 + 0.20650 * X2;
residual = (y - yhatt);
title 'yhatt 和殘差值';
proc print data=d;
run;
```


參考文獻與資料來源

1. 王國川(2004)。圖解SAS視窗在迴歸分析上的應用。台北：五南。
2. 林豐政與溫博仕(2009)。SAS應用之統計分析。台北：華泰文化事業公司。
3. 資料來源：NBA.COM

www.nba.com/suns

