### International Business Research Methods

## Assignment\_09 The Principle Component Analysis

Student/ Student ID /Department

蘇弈慎/ R66061069 /國企所碩一

Professor: 江明憲老師

Date: 2017/12/12

Table of content

[壹、 Output and Analysis 3](#_Toc500670061)

[一、 Principle Component Analysis 3](#_Toc500670062)

[a. The Meaning of Principle Component Analysis 3](#_Toc500670063)

[b. Problem Statement 3](#_Toc500670073)

[c. Output (run by SPSS) 4](#_Toc500670077)

[d. Code and Output (run by R) 6](#_Toc500670078)

1. Output and Analysis
   1. Principle Component Analysis
      1. The Meaning of Principle Component Analysis

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to

convert a set of observations of possibly correlated variables into a set of values of linearly

uncorrelated variables called principal components. The number of distinct principal components is equal

to the smaller of the number of original variables or the number of observations minus one. This

transformation is defined in such a way that the first principal component has the largest

possible variance (that is, accounts for as much of the variability in the data as possible), and each

succeeding component in turn has the highest variance possible under the constraint that it

is orthogonal to the preceding components. The resulting vectors are an uncorrelated orthogonal basis set.

PCA is sensitive to the relative scaling of the original variables.

* + 1. Problem Statement

近年來小球戰術在NBA受到重視與接受，無疑是因為邁阿密熱火隊在教練執行小球戰術的帶領下四度進入總決賽，而金州勇士隊則是由Steve Kerr總教練帶領，以小球戰術拿下總冠軍。過去總認為季後賽是節奏緩慢、內線拚搏的比賽，這種想法也在小球戰術的風潮下漸漸式微。小球戰術是以快速多變的打法、外線強大的火力、靈活好用的內線為戰術風格。在這種戰術體系下，球隊每場得分偏高，因為在頻繁快攻之下進攻回合數增加，加上擁有高命中率的射手的緣故。

本命題以 NBA 2016-2017 球季 例行賽 30隊的數據為資料，以PTS(每場得分)、3PM(每場三分球進球數)、REB(每場總籃板球數)、TOV(每場總失誤)為進行主成分分析的四個變數。

資料來源：<http://stats.nba.com/teams/traditional/?sort=W_PCT&dir=-1&Season=2016-17&SeasonType=Regular%20Season>

* + 1. Output (run by SPSS)

**因數分析**

|  |  |  |  |
| --- | --- | --- | --- |
| **敘述統計** | | | |
|  | 平均值 | 標準差 | 分析 N |
| PTS | 105.590 | 4.1289 | 30 |
| 3PM | 9.650 | 1.6197 | 30 |
| REB | 43.520 | 1.7165 | 30 |
| TOV | 13.960 | 1.2705 | 30 |

* + - 1. Descriptive statistics

由上表可得知,30隊在這四項變數的平均值以及標準差，而從標準差可以看出，30隊在三分球的命中數、每場總籃板球和每場總失誤數並沒有非常大的差別，但在得分能力上，強隊和弱隊較有明顯的差別。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **相關性矩陣** | | | | | |
|  | | PTS | 3PM | REB | TOV |
| 相關性 | PTS | 1.000 | .575 | .360 | .314 |
| 3PM | .575 | 1.000 | -.115 | .103 |
| REB | .360 | -.115 | 1.000 | .245 |
| TOV | .314 | .103 | .245 | 1.000 |

* + - 1. Correlation Matrix

由上表相關性矩陣可得知，每場總得分(PTS)與每場三分球命中數(3PM)呈現中度正相關，與每場總籃板球數(REB)和每場總失誤(TOV)呈現低度正相關。每場三分球命中數(3PM)與每場總籃板球數(REB)呈現低度負相關，與每場總失誤(TOV)呈現低度正相關。每場總籃板球數(REB)和每場總失誤(TOV)呈現低度正相關。

|  |  |  |
| --- | --- | --- |
| **KMO 與 Bartlett 檢定** | | |
| Kaiser-Meyer-Olkin 取樣適切性量數。 | | .417 |
| Bartlett 的球形檢定 | 近似卡方分配 | 23.235 |
| 自由度 | 6 |
| 顯著性 | .001 |

|  |  |  |
| --- | --- | --- |
| **共同性** | | |
|  | 初始 | 萃取 |
| PTS | 1.000 | .833 |
| 3PM | 1.000 | .907 |
| REB | 1.000 | .763 |
| TOV | 1.000 | .479 |
| 擷取方法：主成分分析。 | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **解說總變異量** | | | | | | |
| 成分 | 初始固有值 | | | 擷取平方和負荷量 | | |
| 總計 | 變異的 % | 累加 % | 總計 | 變異的 % | 累加 % |
| 1 | 1.808 | 45.212 | 45.212 | 1.808 | 45.212 | 45.212 |
| 2 | 1.173 | 29.317 | 74.528 | 1.173 | 29.317 | 74.528 |
| 3 | .757 | 18.919 | 93.447 |  |  |  |
| 4 | .262 | 6.553 | 100.000 |  |  |  |
| 擷取方法：主成分分析。 | | | | | | |

* + - 1. Total Variance Explained

總計為每個成分的特徵值，變異的%為第i個成分解釋變異的百分比。

|  |  |  |
| --- | --- | --- |
| **成分矩陣a** | | |
|  | 成分 | |
| 1 | 2 |
| PTS | .903 | -.132 |
| 3PM | .647 | -.699 |
| REB | .487 | .725 |
| TOV | .581 | .376 |
| 擷取方法：主成分分析。 | | |
| a. 已擷取 2 個成分。 | | |

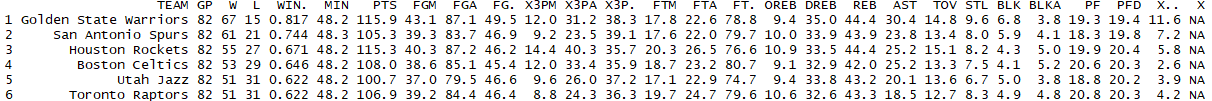
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **重新產生的相關性** | | | | | |
|  | | PTS | 3PM | REB | TOV |
| 重新產生的相關性 | PTS | .833a | .676 | .344 | .475 |
| 3PM | .676 | .907a | -.192 | .113 |
| REB | .344 | -.192 | .763a | .555 |
| TOV | .475 | .113 | .555 | .479a |
| 殘差b | PTS |  | -.101 | .016 | -.161 |
| 3PM | -.101 |  | .077 | -.011 |
| REB | .016 | .077 |  | -.310 |
| TOV | -.161 | -.011 | -.310 |  |
| 擷取方法：主成分分析。 | | | | | |
| a. 重新產生的共同性 | | | | | |
| b. 殘差是在觀察相關性與重新產生的相關性之間計算的。具有 4 (66.0%) 個非多餘殘差，其絕對值大於 0.05。 | | | | | |

* + 1. Code and Output (run by R)

#讀取資料

NBA<-read.csv("C:/Users/Shen/Desktop/碩一上課程/國際企業研究方法/Assignment 09/2016-2017 NBA Team stats.csv", header=T, sep=",")

head(NBA)



#選擇每場得分、每場三分球命中數、每場總籃板及每場總失誤四個變數進行主成份分析

NBA.PCA<-prcomp(~PTS+X3PM+REB+TOV, data=NBA, center=TRUE, scale=TRUE)

NBA.PCA

Standard deviations:

[1] 1.3447920 1.0828947 0.8699116 0.5119838

Rotation:

PC1 PC2 PC3 PC4

PTS -0.6715209 -0.1216007 0.23408324 0.6924435

X3PM -0.4812712 -0.6452178 0.03614256 -0.5922548

REB -0.3618377 0.6696986 0.50680822 -0.4046266

TOV -0.4318694 0.3470030 -0.82888128 -0.0776761

#萃取轉軸後的因素負荷量

load=NBA.PCA$rotation

#將因素負荷量排序

#繪製因素負荷量點圖

sorted.loadings=load[order(load[,1]),1]

dotchart(sorted.loadings, main="Loading Plot for PC1", xlab="Variable Loadings", cex=1.5, col="#FF0000")

sorted.loadings=load[order(load[,2]),2]

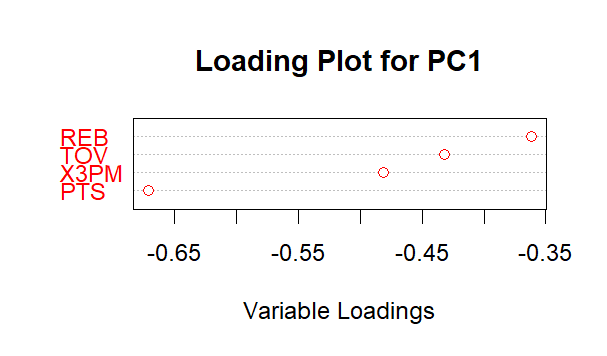
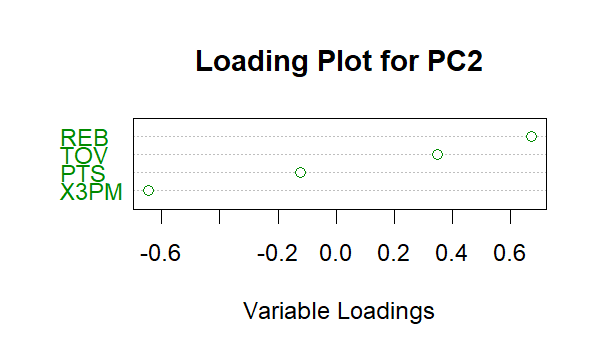
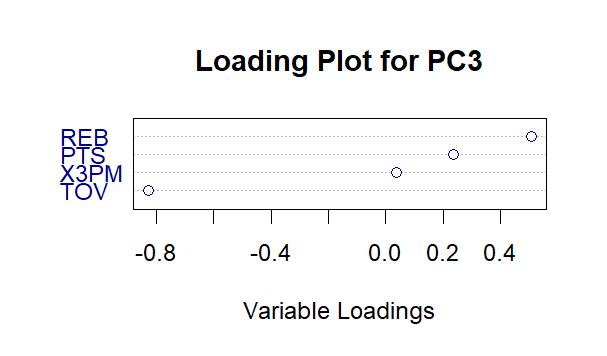
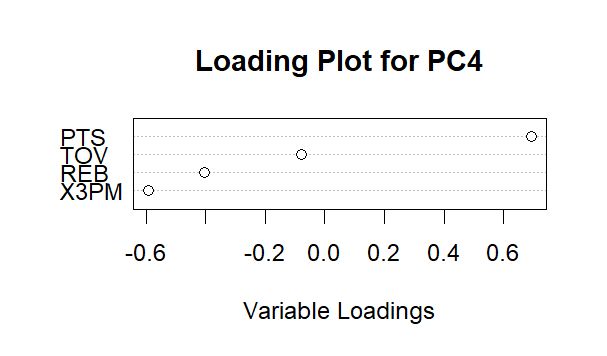
dotchart(sorted.loadings, main="Loading Plot for PC2", xlab="Variable Loadings", cex=1.5, col="#008B00")

sorted.loadings=load[order(load[,3]),3]

dotchart(sorted.loadings, main="Loading Plot for PC3", xlab="Variable Loadings", cex=1.5, col="#00008B")

sorted.loadings=load[order(load[,4]),4]

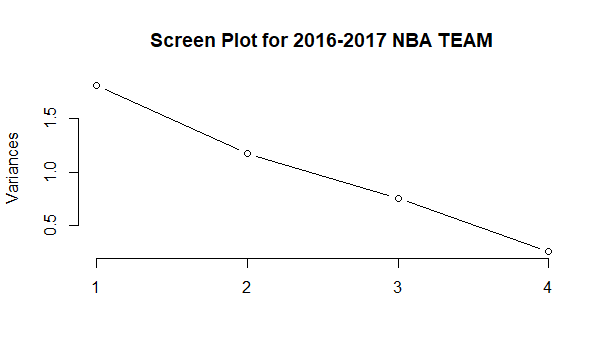
dotchart(sorted.loadings, main="Loading Plot for PC4", xlab="Variable Loadings", cex=1.5, col="#000000")



由上面的因素負荷圖可以更清楚呈現變數與因子間的關係。

　　　#繪製陡坡圖

plot(NBA.PCA, type="line", main="Screen Plot for 2016-2017 NBA TEAM")



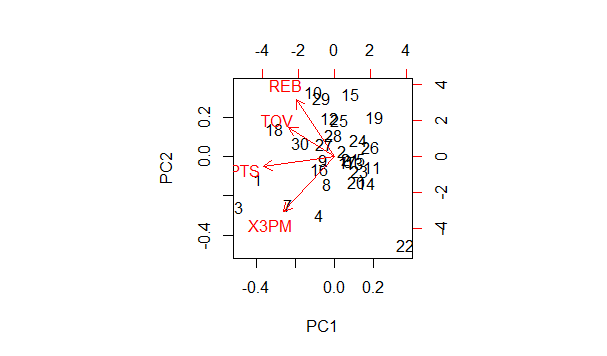
#因子的標準差平方就是特徵值

NBA.PCA$sdev^2

[1] 1.8084656 1.1726609 0.7567461 0.2621274

#因素負荷量圖

biplot(NBA.PCA, choices=1:2)



從因素負荷圖可以看見 NBA 30支球隊的特性，可以見得高排名的球隊幾乎都是擅長得分、投三分球的球隊(1,3,4,7)，擅長搶籃板的球隊(10,29)戰績排名多是中間偏後，說明籃板球的多寡並不是影響戰績的最關鍵因素，而失誤較高的球隊(18,30)戰績排名也都是中間偏後，說明過多的失誤將會讓球隊喪失勝利的機會。

#主成份分析摘要，包含解釋變異量

summary(NBA.PCA)

Importance of components:

PC1 PC2 PC3 PC4

Standard deviation 1.3448 1.0829 0.8699 0.51198

Proportion of Variance 0.4521 0.2932 0.1892 0.06553

Cumulative Proportion 0.4521 0.7453 0.9345 1.00000