

政治大學
統計學系學士班

多變量分析
期中報告

Multivariate Analysis: Midterm Project

指導教授：鄭宗記 教授

學生：陳葳芃 歐西四 107508006

目錄

Part I. Introduction and summarize the data we choose	2
Section 0-1 資料觀察	2
Section 0-2 變數介紹	2
Section 0-3 相關係數矩陣	3
Section 0-4 分析方法及研究目的	3
Part II. The result of data analysis with PCA method	4
Section 2-1 主成分變異解釋表	4
Section 2-2 選取 PC — 根據 Kaiser eigenvalue 選取法則	4
Section 2-3 PCA 結果對照解釋	5
Part III. The result of data analysis with FA method	6
Section 3-1 因素變異解釋表	6
Section 3-2 FA 結果對照解釋	6
Part IV. Conclusion	7
Appendix I _ Programing process (R code)	8

Part I. Introduction and summarize the data we choose

➤ Section0-1 資料觀察

透過先觀察我們欲分析的資料（如表（一）呈現），此筆資料包含了 11 個變數，29 筆的樣本資料；其中第一個變數包含了新北市各行政區的名字，其餘變數則表示各別行政區對於 2018 年 10 大公投案支持程度（各行政區支持度計算方式為有效同意票除以有效投票數）。

行政區	降低火力發電	降低燃煤發電	限制進口核食	婚姻一男一女	禁童性平教育	專法同性婚姻	台灣正名運賽	民法同性婚姻	性平教育實施
板橋	0.7697	0.7498	0.7678	0.6951	0.6549	0.6089	0.4744	0.3580	0.3653
三重	0.7647	0.7402	0.7573	0.6921	0.6427	0.5859	0.5058	0.3631	0.3755
中和	0.7966	0.7829	0.7931	0.7025	0.6693	0.6177	0.4094	0.3391	0.3483
永和	0.8030	0.7851	0.7986	0.6930	0.6585	0.6134	0.3809	0.3413	0.3547
新莊	0.7727	0.7509	0.7735	0.6974	0.6511	0.5911	0.4772	0.3583	0.3659
新店	0.8127	0.7981	0.8021	0.7016	0.6688	0.6183	0.3684	0.3310	0.3397

表(一) 2018 年 10 大公投案新北市各行政區支持度（前六筆資料）

➤ Section0-2 變數介紹

變數名稱	變數解釋	變數型態
行政區	新北市所轄 29 個行政區名稱	類別
降低火力發電	該行政區對提案能源政策「逐年降低火力發電量」支持度百分比	數值
降低燃煤發電	該行政區對提案能源政策「停建燃煤發電廠/機組」支持度百分比	數值
限制進口核食	該行政區對提案食安政策「禁止開放 311 核災地區食品進口」支持度百分比	數值
婚姻一男一女	該行政區對提案法律政策「民法上的定義婚姻應維持一男一女」支持度百分比	數值
禁童性平教育	該行政區對提案教育政策「不應於國民教育階段施行性別平等法中細則——同志教育」支持度百分比	數值
專法同性婚姻	該行政區對提案法律政策「使用民法以外的形式保障同性別兩人永久經營生活的能力」支持度百分比	數值
台灣正名運賽	該行政區對提案運動政策「使用『台灣』之名義參加國際運動賽事」支持度百分比	數值
民法同性婚姻	該行政區對提案法律政策「使用民法形式保障同性別兩人建立婚姻關係」支持度百分比	數值
性平教育實施	該行政區對提案教育政策「應於國民教育階段施行性別平等法中細則——情感教育、性教育、同志教育」支持度百分比	數值
核能運轉延役	該行政區對提案能源政策「應廢除電業法條文『民國 114 年前全部停止運轉/廢除 核能設備』支持度百分比	數值

表(二) 十一筆變數的變數介紹

➤ Section0-3 相關係數矩陣

	降低火力發電	降低燃煤發電	限制進口核食	婚姻一男一女	禁童性平教育	專法同性婚姻	台灣正名運賽	民法同性婚姻	性平教育實施	核能運轉延役
降低火力發電	1.0000	0.9118	0.7622	0.0278	0.3730	0.5821	-0.7617	-0.2237	-0.2606	0.7943
降低燃煤發電	0.9118	1.0000	0.8957	-0.2251	0.1462	0.5801	-0.8833	0.0118	-0.0726	0.8094
限制進口核食	0.7622	0.8957	1.0000	-0.4502	-0.0521	0.4226	-0.9068	0.2458	0.0822	0.7705
婚姻一男一女	0.0278	-0.2251	-0.4502	1.0000	0.8692	0.2448	0.4826	-0.9388	-0.8348	-0.1749
禁童性平教育	0.3730	0.1462	-0.0521	0.8692	1.0000	0.5693	0.1366	-0.9153	-0.8812	0.1416
專法同性婚姻	0.5821	0.5801	0.4226	0.2448	0.5693	1.0000	-0.3439	-0.3662	-0.4090	0.5935
台灣正名運賽	-0.7617	-0.8833	-0.9068	0.4826	0.1366	-0.3439	1.0000	-0.2766	-0.1501	-0.7154
民法同性婚姻	-0.2237	0.0118	0.2458	-0.9388	-0.9153	-0.3662	-0.2766	1.0000	0.8800	0.0863
性平教育實施	-0.2606	-0.0726	0.0822	-0.8348	-0.8812	-0.4090	-0.1501	0.8800	1.0000	-0.0326
核能運轉延役	0.7943	0.8094	0.7705	-0.1749	0.1416	0.5935	-0.7154	0.0863	-0.0326	1.0000

表(三) 變數相關係數矩陣

根據表（三）可得知，在此次公投若探討的兩提案領域同為相近的提案領域（例如：法律、能源、教育……等），則變數之間會出現較強的正相關或負相關(|相關係數| > 0.8)。

➤ Section0-4 分析方法及研究目的

應用到此次分析當中的變數，將採用變數 2~11 項共 11 個數值型變數，透過主成分分析（Principal Components Analysis, PCA）及因素分析（Factory Analysis, FA）來進行維度縮減，試圖從此次眾多提案當中，找出人民關心的實際議題，給出國家建議可以從哪些層面下手了解民意及作為施政方針參考。

Part II. The result of data analysis with PCA method

➤ Section2-1 主成分變異解釋表

第一步，我們試圖嘗試以 PCA 方法將 10 個變數線性組合，組成不同的 components，而從中選取出能最佳解釋原始資料變異的數個 PC，圖（一）為我們初步得到的數據：

```
> summary(pca)
```

Importance of components:

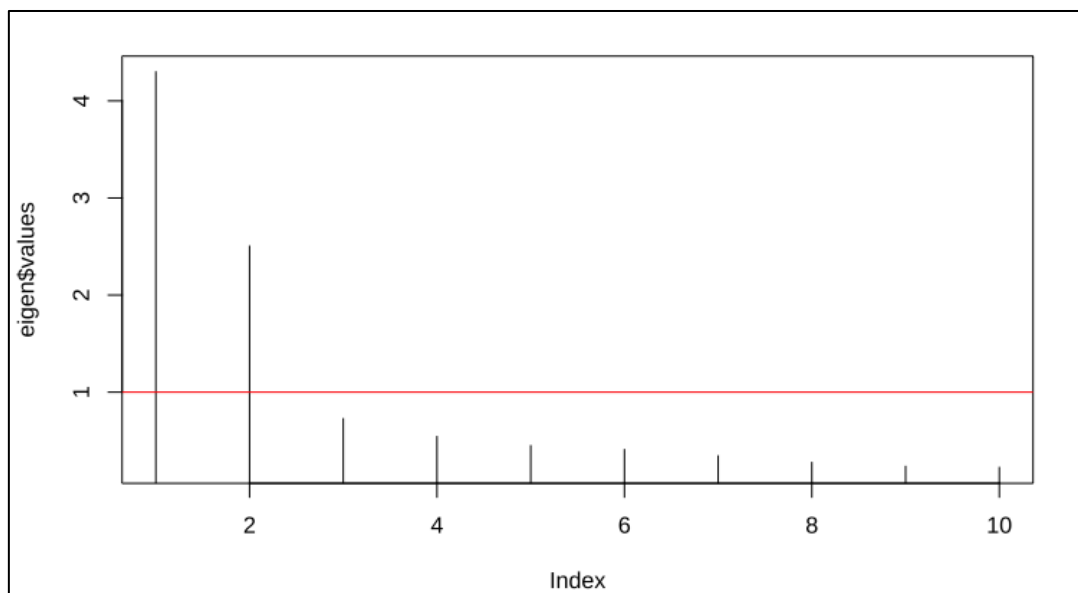
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Standard deviation	2.1744	2.0104	0.71473	0.51993	0.43840	0.29814	0.26151	0.23274	0.17305	0.12442
Proportion of Variance	0.4728	0.4042	0.05108	0.02703	0.01922	0.00889	0.00684	0.00542	0.00299	0.00155
Cumulative Proportion	0.4728	0.8770	0.92806	0.95509	0.97431	0.98320	0.99004	0.99546	0.99845	1.00000

圖（一） 經由線性組合出來的 PCA 變異解釋表

可從變異解釋表中發現，若我們選取 2 個 component 即可獲得不錯的原始資料變異解釋（>85%），若想獲得更佳的解釋程度（>90%）則我們可以選取到第 3 個 component。此外，我們也可以根據 Kaiser 提出的 eigenvalue 選取法則挑選出適當的 PC 個數。

➤ Section2-2 選取 PC— 根據 Kaiser eigenvalue 選取法則

如圖（二）所示，我們可以挑選出 eigenvalue 有大於 1 的 PC，作為我們選取個數的標準，根據圖示結果，我們選擇 **2 個 PC** 即可。



圖（二）依據 Kaiser 的 eigenvalue 選取法則挑選出適合 PC 個數

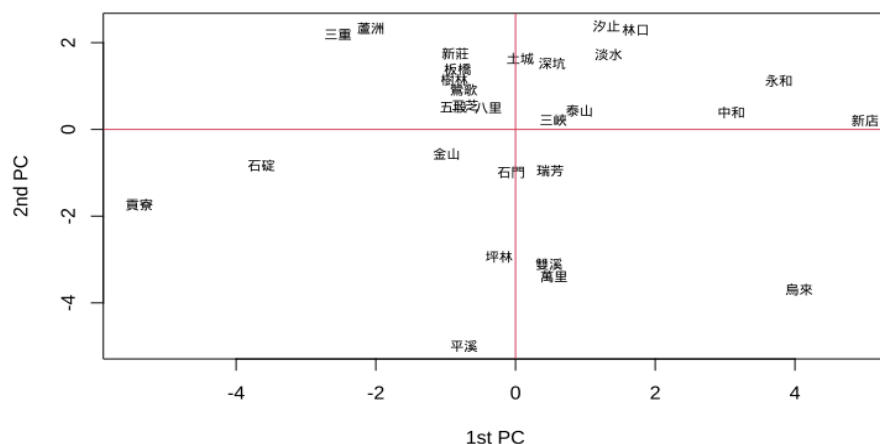
➤ Section 2-3 PCA 結果對照解釋

綜合下圖（三）圖（四），我們可以得出，若在圖（四）靠右邊的行政區，是相對重視環保健康議題，對於對健康環境有疑慮的政策，會有明顯的反彈。而靠近圖片上方的行政區可被歸納在對於多元意識的開放（主要在於性別教育、婚姻價值議題上），反之則可被歸類為較保守的行政區，在新觀念/政令宣導上可能需多加費心力。

```
> pca.factor[,1:2]
```

	PC1	PC2
降低火力發電	0.423264219	-0.10709321
降低燃煤發電	0.447276026	0.01022873
限制進口核食	0.422281465	0.12224230
婚姻一男一女	-0.099964389	-0.47464157
禁童性平教育	0.084934702	-0.47794841
專法同性婚姻	0.296880975	-0.22974540
台灣正名運籌	-0.409693137	-0.15242937
民法同性婚姻	0.008416367	0.48272143
性平教育實施	-0.044337820	0.45892760
核能運轉延役	0.408949485	0.01057101

圖（三）前兩個 PC 的組成成分



圖（四）各行政區在兩個主成分視覺化結果

觀察到上面的分析，我們也更好奇，有沒有更簡單、明確化的因素來解析我們人民關心的議題及相對應表達意向。因此，我們下一部分將會進行因素分析，來檢視此次投票結果。

Part III. The result of data analysis with FA method

➤ Section 3-1 因素變異解釋表

接下來，我們採用因素分析方法嘗試將 10 個變數，結合成一個個不同因素。並透過解讀每個因素代表了什麼意涵，萃取出公投中人民表達出了對哪個議題面向的重視。圖（五）為我們初步得到的數據：

	Factor1	Factor2	Factor3
SS loadings	3.961	3.948	0.635
Proportion Var	0.396	0.395	0.063
Cumulative Var	0.396	0.791	0.854

圖（五） FA 變異解釋表

➤ Section 3-2 FA 結果對照解釋

依據下圖（六）及可視化結果圖（七），可以得知第一因素涵蓋的議題較為全面，也很明顯看出是台灣民眾對於相同性別婚姻及教育的權益保障與否的因素，我們又可簡化為『對於新時代多元性別議題的認同與否』；而第二主成分則明顯可以看出較與能源環境相關議題，比較值得注意的是在第二個因素的組成中，「我國在世界運動賽事正名“台灣”」此項變數高達-0.963 的相關，細究背後原因可能為前三項能源政策變數為當時在最大野黨力推的選舉案，而台灣正名運動更是與當時我國執政黨所認同的理念相近，因而呈現負相關也不奇怪。正因如此，若稱呼為『能源議題的探討』或『執政在野的政策對決』，後者的解釋性可以更較為全面。

Loadings:			
	Factor1	Factor2	Factor3
降低火力發電	0.274	0.863	
降低燃煤發電		0.978	
限制進口核食	-0.188	0.919	
婚姻一男一女	0.928	-0.309	
禁童性平教育	0.937		
專法同性婚姻	0.347	0.398	0.284
台灣正名運賽	0.199	-0.963	
民法同性婚姻	-1.030		0.161
性平教育實施	-0.923		
核能運轉延役	-0.189	0.452	0.709

圖（六） 變數組合出的新因子（Factor）

Part IV. Conclusion

公民投票常常是一國政府施政的方針參考，人民也能透過民意的展現來影響國家前進的方向。從這次 10 項公投案，我們可以清楚知道此次公投主要圍繞兩個主議題：『對於新時代多元性別議題的認同與否』以及『執政在野的政策對決』。分析結果為新北市對於不同多元性別價值與認同感可能較低，從圖（四）PC2 負組成的絕對值普遍大於正組成的絕對值即可以發現。

另外，我們又可以透過因素分析中發現，左右台灣人意見的有時候是政黨所提出的不同主張。但政黨因素這也不是一個絕對值。例如：我們當時選出了擁有獨立黨綱的執政黨，但在面臨國際上國家認同的議題時，卻是希望守護運動員的參賽權利，不希望冒可能喪失參賽權風險而去正名，做出與總統選舉時相反的立場主張。

綜合分析，可以建議新北市政府，若有考慮增設核電廠、燃煤廠，除環境因素外中永和、汐止可能不是一個輕易選定的選擇，而在教育單位方面若有要推廣勢在必行的性平教育觀念的話，平溪、坪林、雙溪、萬里則會是需要投注較多資源心力的地方，來與地方取得雙方能夠接受的共識。

Appendix I _ Programing process (R code)

```
## Multivariate Analysis Midterm Report
```

```
library(readr)                                #input original data set
data <- read.csv("~/Desktop/1112-NCCU/多變量分析/多變量報告期末數據/newtaipei_data.csv")

str(data)                                     #observe the raw data and do some required
transformation
as.character(data[,c(1,2)])
str(data[,2])

data["村里別"][data["村里別"] == ""] <- NA #replace empty blank with NA
new_DF <- data[is.na(data$村里別),]         #to easier analyze, here we just select the
administrative section only
new_DF$行政區別 = gsub("區","",new_DF$行政區別)
data <- na.omit(data)                        #remove NA (the completed administrative
district data) row

df<- unite(data,行政區村里別,c("行政區別","村里別"),sep="_")
df <- data.frame(df, row.names = df[,1])
new_DF <- data.frame(new_DF, row.names = new_DF[,1])

library(tidyverse)                           #rename the column to make the analysis clearly
and easier
df2<-df[, -1]
new_DF<-new_DF[,c(-1,-2)]
lookup <- c(降低火力發電 = "第 7 案", 降低燃煤發電 = "第 8 案",限制進口核食= "第 9 案", 婚姻一男一女
= "第 10 案", 禁童性平教育 = "第 11 案", 專法同性婚姻 = "第 12 案", 台灣正名運賽 = "第 13 案", 民法同性
婚姻 = "第 14 案", 性平教育實施 = "第 15 案", 核能運轉延役 = "第 16 案")
new_DF<-rename(new_DF, all_of(lookup))

corr<-cor(new_DF)                             #observed the
corr2<-as.data.frame(corr)
write_xlsx(corr2,"~/Desktop/corr2.xlsx")

##### PCA
pca<- prcomp(new_DF, center = TRUE, scale = TRUE)
```

```

names(pca)
summary(pca)

install.packages("writexl")
library("writexl")
write_xlsx(new_DF, "~/Desktop/new_DF.xlsx")

pca.factor <- pca$rotation
pca.factor[,1:4]

eigen<-eigen(cor(df2))
plot(eigen$values,type="h")
abline(h=1, col="red") #Kaiser eigenvalue-greater-than-one rule, choose pc1~pc5 by
Kaiser

pcs<-predict(pca)

plot(pcs[,1:2],type="n",xlab='1st PC',ylab='2nd PC')
text(pcs[,1:2], row.names(new_DF), cex=0.7, family = "康熙字典體")
abline(h=0, v=0, col=2)

#####Factor Analysis
sub <- cor(new_DF)
spearman.mle<-factanal(covmat=as.matrix(sub),factors=3)
spearman.mle<-factanal(covmat=as.matrix(sub),factors=3,rotation = 'promax')
#promax is better than varimax
#(as the result of the relationship between variables is related)

spearman.mle
1-spearman.mle$uniq

#Input data
factor.analysis <- factanal(new_DF,factors=2,scores="Bartlett")
plot(factor.analysis$scores[,1],factor.analysis$scores[,2],type="n",xlab= 'Factor 1',
ylab='Factor 2')
text(factor.analysis$scores[,1:2],row.names(new_DF),cex=0.8)
abline(h=0, v=0,col=2)

```

