Multivariate hw5

106070020

2021年5月11日

```
dat2 <- as.matrix(read.table("C:/Users/eva/Desktop/作業 上課資料(清大)/大四下/多變量/hw5/T4-6.DAT"), header=T)
dat3 <- as.matrix(read.table("C:/Users/eva/Desktop/作業 上課資料(清大)/大四下/多變量/hw5/T8-4.DAT"), header=T)
```

Q2

(a)

```
dat2<-dat2[,1:5]
var(dat2)
```

```
##            V1         V2          V3         V4         V5
## V1  34.750209 -4.2766846 -18.0717949 -15.972868   5.716458
## V2  -4.276685 17.5134168   0.4197973  -7.868217  -8.723315
## V3 -18.071795  0.4197973  29.8447227   9.348837 -13.942159
## V4 -15.972868 -7.8682171   9.3488372  33.042636  -9.941860
## V5   5.716458 -8.7233154 -13.9421586  -9.941860  26.957961
```

```
pc2 <- prcomp(dat2)
var_pc2<-(pc2$sdev)^2
#eigenvalue
var_pc2
```

```
## [1] 68.752385 31.508994 23.100973 16.354182  2.392411
```

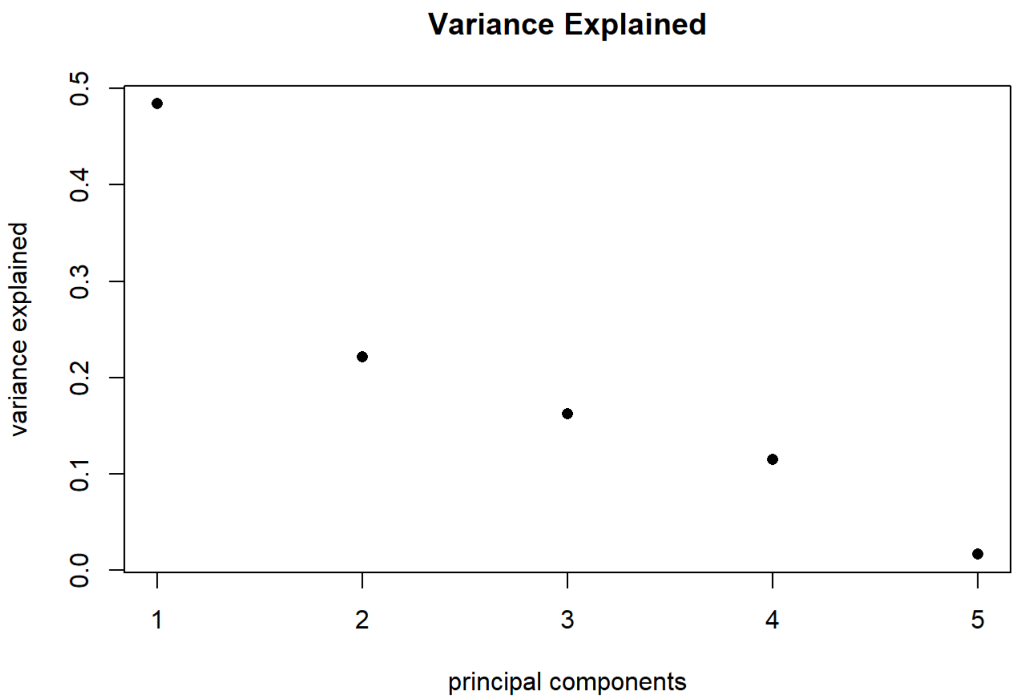
```
cov_mat2<-pc2$rotation
#eigenvector
cov_mat2
```

```
##            PC1         PC2        PC3         PC4        PC5
## V1 -0.57943538  0.07917988  0.6428795 -0.30939267 -0.3859629
## V2  0.04165689  0.61192825 -0.1399143  0.51462195 -0.5825777
## V3  0.52428496  0.21883511 -0.1192554 -0.73403767 -0.3524249
## V4  0.49309245 -0.57215650  0.4221873  0.30427403 -0.3983365
## V5 -0.38013742 -0.49398633 -0.6120997 -0.08970196 -0.4782893
```

```
#variance explained
pc2$sdev^2/sum(pc2$sdev^2)
```

```
## [1] 0.48380054 0.22172422 0.16255819 0.11508200 0.01683505
```

```
#scree plot
plot(pc2$sdev^2/sum(pc2$sdev^2), pch = 16, xlab = "principal components",
     ylab = "variance explained", main="Variance Explained")
```



Using the scree plot and the proportion of variance explained, it seems that 4 components should be retained. These components explain almost all (98%) of the variabilty.

(b)

```
mainpc2<-cov_mat2[,1:2]
rownames(mainpc2)=c("Indep", "Supp", "Benev", "Conform", "Leader")
mainpc2
```

```
##                 PC1         PC2
## Indep   -0.57943538  0.07917988
## Supp     0.04165689  0.61192825
## Benev    0.52428496  0.21883511
## Conform  0.49309245 -0.57215650
## Leader  -0.38013742 -0.49398633
```

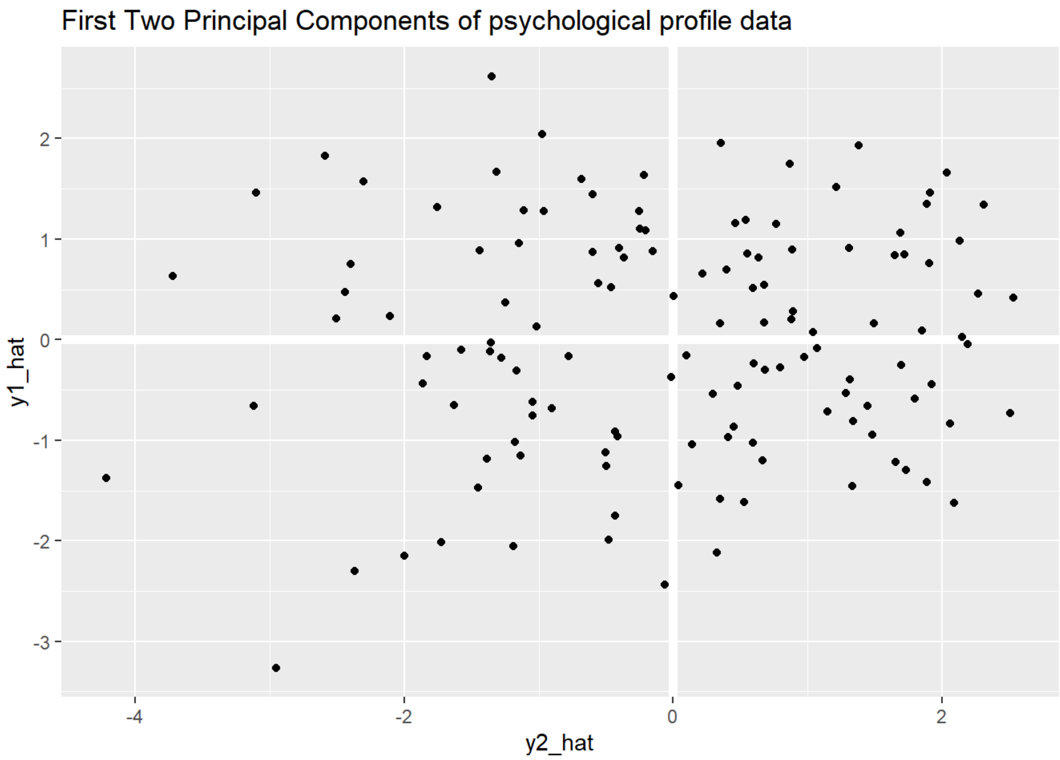
The first component contrasts independence and leadership with benevolence and conformity. The second component contrasts support with conformity and leadership and so on.

(c)

```
library(ggplot2)
dat22<-scale(dat2)
PC1 <- as.matrix(dat22) %*% mainpc2[,1]
PC2 <- as.matrix(dat22) %*% mainpc2[,2]
PC <- data.frame(PC1, PC2)
head(PC)
```

```
##          PC1        PC2
## 1 -1.1739871 -1.0139877
## 2  2.0550938 -0.8368591
## 3  0.2192151  0.6537238
## 4 -0.2519305  1.2808495
## 5  1.9044990  0.7568771
## 6  0.5938083 -1.0296499
```

```
ggplot(PC, aes(PC1, PC2)) +
  modelr::geom_ref_line(h = 0) +
  modelr::geom_ref_line(v = 0) +
  geom_point(mapping=aes(PC1, PC2),data=PC) +
  xlab("y2_hat") +
  ylab("y1_hat") +
  ggtitle("First Two Principal Components of psychological profile data ")
```



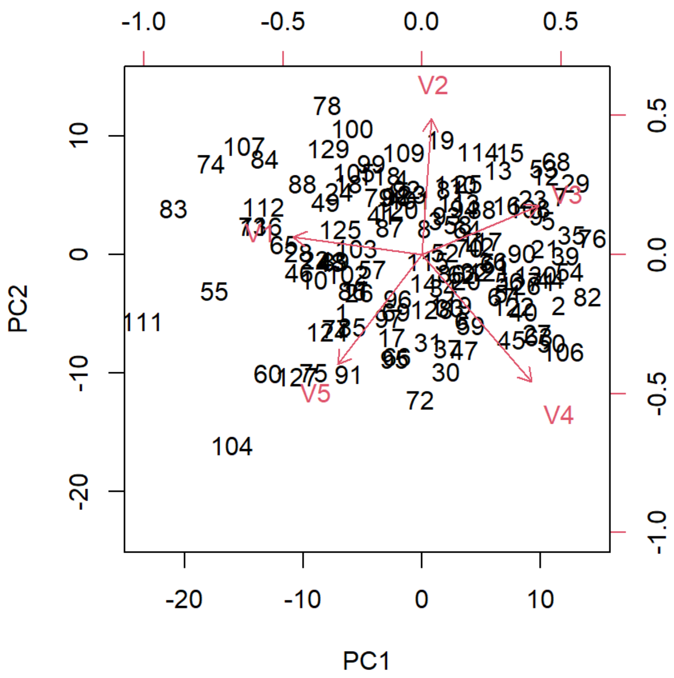
> The two dimensional plot of the scores on the first two components suggests that the two socioeconomic levels cannot be distinguished from one another nor can the two genders be distinguished.

(d)

```
R<-cor(dat22)
R
```

```
##             V1          V2          V3         V4         V5
## V1  1.0000000 -0.17335767 -0.56116271 -0.4713753  0.1867690
## V2 -0.1733577  1.00000000  0.01836202 -0.3270797 -0.4014696
## V3 -0.5611627  0.01836202  1.00000000  0.2977052 -0.4915331
## V4 -0.4713753 -0.32707967  0.29770524  1.0000000 -0.3331093
## V5  0.1867690 -0.40146956 -0.49153305 -0.3331093  1.0000000
```

```
biplot(pc2, scale = 0)
```



> The components are very similar to those obtained from the correlation matrix R. All four of the components represent contrasts of some forms.

Q3

```r
library(aTSA)
pc3 <- prcomp(dat3)
var(dat3)
```

```
##              V1           V2           V3           V4           V5
## V1 4.332695e-04 0.0002756679 1.590265e-04 6.411929e-05 8.896616e-05
## V2 2.756679e-04 0.0004387172 1.799737e-04 1.814512e-04 1.232623e-04
## V3 1.590265e-04 0.0001799737 2.239722e-04 7.341348e-05 6.054612e-05
## V4 6.411929e-05 0.0001814512 7.341348e-05 7.224964e-04 5.082772e-04
## V5 8.896616e-05 0.0001232623 6.054612e-05 5.082772e-04 7.656742e-04
```

```r
var_pc3<-(pc3$sdev)^2
var_pc3
```

```
## [1] 0.0013676780 0.0007011596 0.0002538024 0.0001426026 0.0001188868
```

```r
cov_mat3<-pc3$rotation
cov_mat3
```

```
##           PC1        PC2         PC3        PC4         PC5
## V1 -0.2228228  0.6252260 -0.32611218  0.6627590 -0.11765952
## V2 -0.3072900  0.5703900  0.24959014 -0.4140935  0.58860803
## V3 -0.1548103  0.3445049  0.03763929 -0.4970499 -0.78030428
## V4 -0.6389680 -0.2479475  0.64249741  0.3088689 -0.14845546
## V5 -0.6509044 -0.3218478 -0.64586064 -0.2163758  0.09371777
```

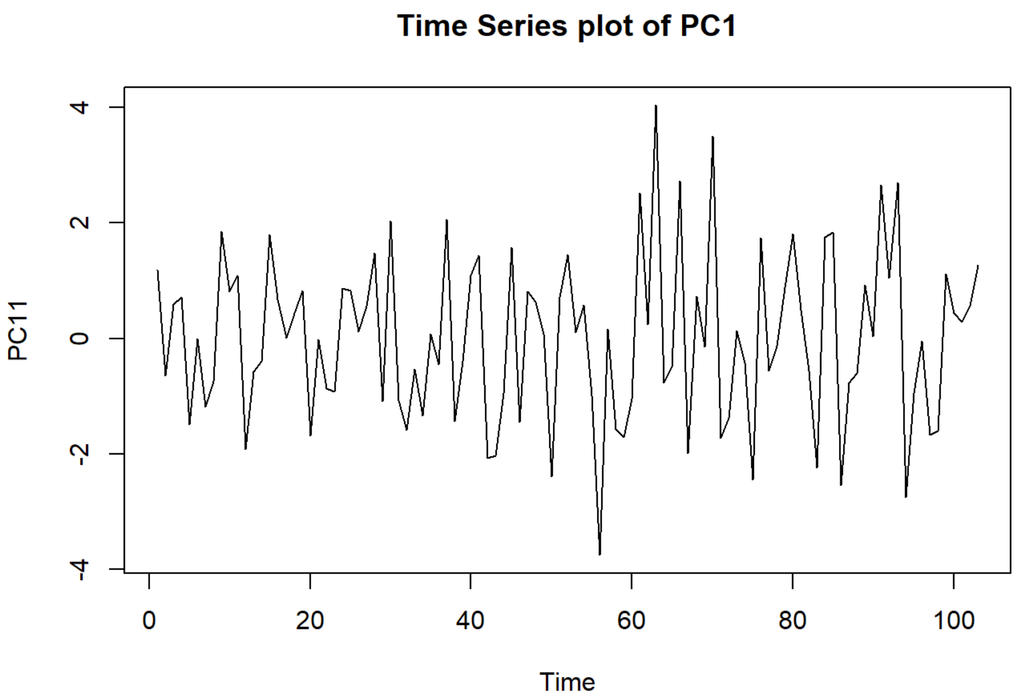
```r
var_ex3<-pc3$sdev^2/sum(pc3$sdev^2)
cumsum(var_ex3)
```

```
## [1] 0.5292607 0.8005936 0.8988095 0.9539935 1.0000000
```

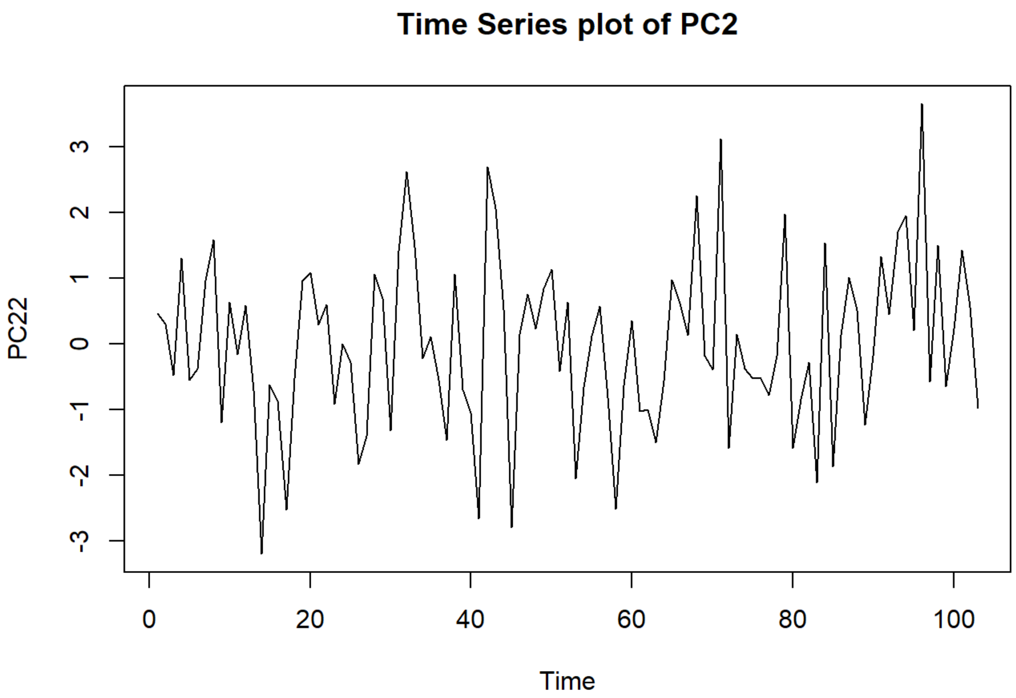
```
mainpc3<-cov_mat3[,1:2]
dat33<-scale(dat3)
PC11 <- as.matrix(dat33) %*% mainpc3[,1]
PC22 <- as.matrix(dat33) %*% mainpc3[,2]
PC3 <- data.frame(PC11, PC22)
head(PC3)
```

```
##          PC11       PC22
## 1  1.17916624  0.4539448
## 2 -0.64389065  0.2963530
## 3  0.58827842 -0.4737926
## 4  0.71230848  1.2957392
## 5 -1.49423658 -0.5557366
## 6 -0.01040056 -0.3835861
```

```
ts.plot(PC11, main='Time Series plot of PC1')
```




```
ts.plot(PC22, main='Time Series plot of PC2')
```




```
adf.test(ts(PC11))
```

```
## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##      lag    ADF p.value
## [1,]   0 -10.87    0.01
## [2,]   1  -7.65    0.01
## [3,]   2  -6.40    0.01
## [4,]   3  -5.45    0.01
## [5,]   4  -6.07    0.01
## Type 2: with drift no trend
##      lag    ADF p.value
## [1,]   0 -10.82    0.01
## [2,]   1  -7.61    0.01
## [3,]   2  -6.37    0.01
## [4,]   3  -5.42    0.01
## [5,]   4  -6.03    0.01
## Type 3: with drift and trend
##      lag    ADF p.value
## [1,]   0 -10.78    0.01
## [2,]   1  -7.59    0.01
## [3,]   2  -6.35    0.01
## [4,]   3  -5.41    0.01
## [5,]   4  -6.01    0.01
## ----
## Note: in fact, p.value = 0.01 means p.value <= 0.01
```

```
adf.test(ts(PC22))
```

```
## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##      lag   ADF p.value
## [1,]   0 -9.57    0.01
## [2,]   1 -7.00    0.01
## [3,]   2 -5.25    0.01
## [4,]   3 -4.30    0.01
## [5,]   4 -3.78    0.01
## Type 2: with drift no trend
##      lag   ADF p.value
## [1,]   0 -9.52    0.01
## [2,]   1 -6.96    0.01
## [3,]   2 -5.22    0.01
## [4,]   3 -4.28    0.01
## [5,]   4 -3.76    0.01
## Type 3: with drift and trend
##      lag   ADF p.value
## [1,]   0 -9.59  0.0100
## [2,]   1 -7.08  0.0100
## [3,]   2 -5.33  0.0100
## [4,]   3 -4.45  0.0100
## [5,]   4 -3.92  0.0163
## ----
## Note: in fact, p.value = 0.01 means p.value <= 0.01
```

> According to the adf.test, the two time series plot are both stationary, as p-value is less than 0.01.