

# thuchanh\_w6

Bùi Khánh Duy

2023-04-12

```
dl = read.csv("table2.csv")
head(dl[, -1])
```

```
##      S02 Nhietdo Soluong DanSo  Gio LuongMua SoNgay
## 1  46    47.6     44   116  8.8    33.36   135
## 2  11    56.8     46   244  8.9     7.77    58
## 3  24    61.5    368   497  9.1    48.34   115
## 4  47    55.0    625   905  9.6    41.31   111
## 5  11    47.1    391   463 12.4    36.11   166
## 6  31    55.2     35    71  6.5    40.75   148
```

```
x <- dl[, -1]
s <- cov(x)
```

s

```
##              S02          Nhietdo          Soluong          DanSo          Gio
## S02          614.555263 -87.5942105 13991.7079 11906.028947 1.6578947
## Nhietdo      -87.594211  49.7832632 -1136.8953 -513.852105 0.4431579
## Soluong     13991.707895 -1136.8952632 522007.0816 509426.497368 327.5315789
## DanSo       11906.028947 -513.8521053 509426.4974 538344.871053 361.1789474
## Gio         1.657895    0.4431579    327.5316    361.178947 1.9810526
## LuongMua     4.381658    27.4665421 -273.0543    -5.212026 -1.4161053
## SoNgay      240.300000   -97.8400000  2133.5421    566.142105 0.7473684
##              LuongMua          SoNgay
## S02          4.381658 240.3000000
## Nhietdo      27.466542 -97.8400000
## Soluong     -273.054289 2133.5421053
## DanSo       -5.212026 566.1421053
## Gio         -1.416105 0.7473684
## LuongMua    121.223773 111.8605789
## SoNgay     111.860579 680.9368421
```

1) Tính các giá trị riêng và vectơ riêng tương ứng của ma trận tương quan mẫu.

```
tquan_mau = cor(x)
eicor = eigen(tquan_mau)
eicor
```

```
## eigen() decomposition
## $values
## [1] 2.91084504 1.56072435 1.30735066 0.88023318 0.19687611 0.12416937 0.01980129
##
## $vectors
##          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## [1,]  0.52096782 -0.1529578 -0.01199717  0.25631027  0.74634780  0.19935977
## [2,] -0.28763736  0.4912635  0.48916810  0.14049617  0.08690719  0.63969175
## [3,]  0.54849916  0.2261182  0.07151683  0.14926148 -0.18130420  0.02441424
## [4,]  0.50836681  0.3157964  0.12859145  0.13855793 -0.47966584 -0.08883438
## [5,]  0.19519472  0.3642556  0.05993107 -0.86840514  0.25366269 -0.08192149
## [6,] -0.02360551 -0.2112159  0.82632278  0.04392958  0.11659393 -0.50508574
## [7,]  0.21888566 -0.6380541  0.22917056 -0.34209837 -0.30744539  0.52984886
##
##          [,7]
## [1,] -0.20634486
## [2,] -0.01191515
## [3,]  0.76626256
## [4,] -0.60673788
## [5,] -0.02101819
## [6,]  0.03722565
## [7,] -0.01277291
```

2) Đưa ra tỉ lệ biến sai tổng cộng của từng thành phần chính.

```
pcacov = princomp(covmat=s)
summary(pcacov, loadings=T)
```

```
## Importance of components:
##
##              Comp.1      Comp.2      Comp.3      Comp.4
## Standard deviation 1019.8015639 144.52594715 2.603745e+01 1.187979e+01
## Proportion of Variance 0.9794455 0.01967164 6.384778e-04 1.329126e-04
## Cumulative Proportion 0.9794455 0.99911713 9.997556e-01 9.998885e-01
##
##              Comp.5      Comp.6      Comp.7
## Standard deviation 1.028333e+01 3.343802e+00 1.203265e+00
## Proportion of Variance 9.959015e-05 1.053004e-05 1.363551e-06
## Cumulative Proportion 9.999881e-01 9.999986e-01 1.000000e+00
##
## Loadings:
##              Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
## S02              0.227  0.907  0.324  0.121
## Nhietdo          -0.107 -0.222  0.374  0.892
## Soluong          0.701  0.708
## DanSo            0.713 -0.699
## Gio                                     0.997
## LuongMua          0.212 -0.300  0.839 -0.398
## SoNgay            0.942 -0.181 -0.225  0.162
```

```
pcacor = princomp(covmat=tquan_mau)
summary(pcacor, loadings=T)
```

```
## Importance of components:
```

```
##               Comp.1   Comp.2   Comp.3   Comp.4   Comp.5
## Standard deviation 1.706120 1.2492895 1.1433944 0.9382074 0.44370724
## Proportion of Variance 0.415835 0.2229606 0.1867644 0.1257476 0.02812516
## Cumulative Proportion 0.415835 0.6387956 0.8255600 0.9513076 0.97943276
##               Comp.6   Comp.7
## Standard deviation 0.35237675 0.140717070
## Proportion of Variance 0.01773848 0.002828756
## Cumulative Proportion 0.99717124 1.000000000
##
## Loadings:
##      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
## S02      0.521 0.153      0.256 0.746 0.199 0.206
## Nhietdo -0.288 -0.491 -0.489 0.140      0.640
## Soluong 0.548 -0.226      0.149 -0.181      -0.766
## DanSo   0.508 -0.316 -0.129 0.139 -0.480      0.607
## Gio     0.195 -0.364      -0.868 0.254
## LuongMua      0.211 -0.826      0.117 -0.505
## SoNgay 0.219 0.638 -0.229 -0.342 -0.307 0.530
```

Tỉ lệ biến sai tổng cộng của từng thành phần chính lần lượt là: [0.415835, 0.2229606, 0.1867644, 0.1257476, 0.02812516, 0.01773848, 0.002828756]

### 3) Biểu diễn thành phần chính thứ nhất theo các biến ban đầu.

$$Comp.1 = 0.521 \cdot S02 - 0.288 \cdot Nhietdo + 0.548 \cdot Soluong + 0.508 \cdot DanSo + 0.195 \cdot Gio + 0.219 \cdot SoNgay$$

### 4) Cần bao nhiêu thành phần chính để thu được 95% thông tin về tập dữ liệu

ban đầu?

Từ dữ liệu ở trên, ta kết luận được cần 4 thành phần chính

### 5) Vẽ biểu đồ biplot cho bộ dữ liệu.

```
library(stats)
pca<-princomp(x)
biplot(pca,col=c("gray","black"))
```

```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

