

다변량

6장 과제

과목	다변량데이터분석
담당교수	임태진 교수님
전공	산업정보시스템공학과
학번	20201368
이름	한채원
제출일	2022.11.15

1. 표준화를 수행한 결과

```
> setwd("c:/temp/data")
> euroemp = read.csv('Euroemp.csv',row.names=1)
> ecor = cor(euroemp[,-1])
> round(ecor,3)

      AGR    MIN    MAN    PS    CON
AGR  1.000  0.316 -0.254 -0.382 -0.349
MIN  0.316  1.000 -0.672 -0.387 -0.129
MAN -0.254 -0.672  1.000  0.388 -0.034
PS   -0.382 -0.387  0.388  1.000  0.165
CON -0.349 -0.129 -0.034  0.165  1.000
SER -0.605 -0.407 -0.033  0.155  0.473
FIN -0.176 -0.248 -0.274  0.094 -0.018
SPS -0.811 -0.316  0.050  0.238  0.072
TC   -0.487  0.045  0.243  0.105 -0.055
      SER    FIN    SPS    TC
AGR -0.605 -0.176 -0.811 -0.487
MIN -0.407 -0.248 -0.316  0.045
MAN -0.033 -0.274  0.050  0.243
PS   0.155  0.094  0.238  0.105
CON  0.473 -0.018  0.072 -0.055
SER  1.000  0.379  0.388 -0.085
FIN  0.379  1.000  0.166 -0.391
SPS  0.388  0.166  1.000  0.475
TC   -0.085 -0.391  0.475  1.000

> euroemp_pca = prcomp(euroemp[,-1],scale = TRUE)
> eigen_euroemp = euroemp_pca$sdev^2
> names(eigen_euroemp) = paste("PC",1:9,sep='')
> round(eigen_euroemp,3)
  PC1  PC2  PC3  PC4  PC5  PC6  PC7
3.112 1.809 1.496 1.063 0.710 0.311 0.293
  PC8  PC9
0.204 0.000

> (sumlam = sum(eigen_euroemp))
[1] 9
> propvar = eigen_euroemp/sumlam;cumvar=cumsum(propvar)
> mat = rbind(eigen_euroemp,propvar,cumvar)
> rownames(mat) = c('Eig-val','Prop-var','Cum-prop-var')
```

```
> round(mat,3)
```

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Eig-val	3.112	1.809	1.496	1.063	0.710	0.311	0.293	0.204	0
Prop-var	0.346	0.201	0.166	0.118	0.079	0.035	0.033	0.023	0
Cum-prop-var	0.346	0.547	0.713	0.831	0.910	0.945	0.977	1.000	1

```
> round(euroemp_pca$rotation,3)
```

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
AGR	0.511	0.023	0.279	-0.016	0.024	-0.042	-0.164	-0.540	0.582
MIN	0.375	0.000	-0.515	-0.114	-0.346	0.199	0.213	0.449	0.419
MAN	-0.246	-0.432	0.502	-0.058	0.234	-0.031	0.236	0.432	0.447
PS	-0.316	-0.109	0.294	-0.023	-0.854	0.206	-0.061	-0.155	0.030
CON	-0.222	0.242	-0.072	-0.783	-0.062	-0.503	-0.020	-0.031	0.129
SER	-0.382	0.408	-0.065	-0.169	0.267	0.673	0.175	-0.202	0.245
FIN	-0.131	0.553	0.096	0.489	-0.131	-0.406	0.458	0.027	0.191
SPS	-0.428	-0.055	-0.360	0.317	0.046	-0.158	-0.621	0.041	0.410
TC	-0.205	-0.517	-0.413	0.042	0.023	-0.142	0.492	-0.502	0.061

2. 표준화를 수행하지 않은 결과

```
> euroemp_pca2 = prcomp(euroemp[,-1],scale = F)
```

```
> eigen_euroemp2 = euroemp_pca2$sdev^2
```

```
> names(eigen_euroemp2) = paste("PC",1:9,sep="")
```

```
> round(eigen_euroemp2,3)
```

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
	244.717	118.202	42.015	26.395	10.437	4.831	0.732	0.308	0.000

```
> (sumlam2 = sum(eigen_euroemp2))
```

```
[1] 447.6381
```

```
> propvar2 = eigen_euroemp2/sumlam2;cumvar2=cumsum(propvar2)
```

```
> mat2 = rbind(eigen_euroemp2,propvar2,cumvar2)
```

```
> rownames(mat2) = c('Eig-val','Prop-var','Cum-prop-var')
```

```
> round(mat2,3)
```

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Eig-val	244.717	118.202	42.015	26.395	10.437	4.831	0.732	0.308	0
Prop-var	0.547	0.264	0.094	0.059	0.023	0.011	0.002	0.001	0
Cum-prop-var	0.547	0.811	0.905	0.964	0.987	0.998	0.999	1.000	1

```
> round(euroemp_pca2$rotation,3)
```

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
AGR	0.741	0.298	0.263	-0.345	-0.193	0.087	-0.104	0.095	0.330
MIN	0.340	-0.532	-0.610	0.175	0.134	0.151	-0.196	0.103	0.330
MAN	-0.273	0.727	-0.448	0.069	0.128	0.117	-0.189	0.132	0.330
PS	-0.017	0.012	-0.004	0.008	0.025	-0.052	0.012	-0.938	0.341
CON	-0.049	-0.032	0.054	0.293	-0.346	-0.800	-0.122	0.156	0.329
SER	-0.198	-0.094	0.400	0.500	-0.344	0.544	-0.101	0.085	0.332
FIN	-0.047	-0.076	0.428	0.055	0.809	-0.120	-0.056	0.148	0.334
SPS	-0.464	-0.290	0.024	-0.710	-0.181	0.047	-0.205	0.108	0.328
TC	-0.034	-0.014	-0.110	-0.049	-0.042	0.024	0.919	0.135	0.344

3. 결과 해석

(1) 상관행렬의 고유값

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Eig-val	3.112	1.809	1.496	1.063	0.710	0.311	0.293	0.204	0
Prop-var	0.346	0.201	0.166	0.118	0.079	0.035	0.033	0.023	0
Cum-prop-var	0.346	0.547	0.713	0.831	0.910	0.945	0.977	1.000	1

(2) 분산-공분산 행렬의 고유값

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Eig-val	244.717	118.202	42.015	26.395	10.437	4.831	0.732	0.308	0
Prop-var	0.547	0.264	0.094	0.059	0.023	0.011	0.002	0.001	0
Cum-prop-var	0.547	0.811	0.905	0.964	0.987	0.998	0.999	1.000	1

다음과 같이 살펴보면 상관행렬과 분산-공분산 행렬의 고유값이 다를 수 있다. 상관행렬에서는 제1 주성분이 자료 변동량의 34.6%를 차지하지만 분산-공분산 행렬에서는 제1 주성분이 자료 변동량의 54.7%를 차지하는 것을 알 수 있다. 상관행렬에서는 변동량의 83%를 설명하기 위해 4개의 주성분이 필요하지만 분산-공분산 행렬에서는 변동량의 81%를 설명하기 위해 2개의 주성분이 필요하다는 것을 알 수 있다.

(3) 상관행렬의 고유벡터

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
AGR	0.511	0.023	0.279	-0.016	0.024	-0.042	-0.164	-0.540	0.582
MIN	0.375	0.000	-0.515	-0.114	-0.346	0.199	0.213	0.449	0.419
MAN	-0.246	-0.432	0.502	-0.058	0.234	-0.031	0.236	0.432	0.447
PS	-0.316	-0.109	0.294	-0.023	-0.854	0.206	-0.061	-0.155	0.030
CON	-0.222	0.242	-0.072	-0.783	-0.062	-0.503	-0.020	-0.031	0.129
SER	-0.382	0.408	-0.065	-0.169	0.267	0.673	0.175	-0.202	0.245
FIN	-0.131	0.553	0.096	0.489	-0.131	-0.406	0.458	0.027	0.191
SPS	-0.428	-0.055	-0.360	0.317	0.046	-0.158	-0.621	0.041	0.410
TC	-0.205	-0.517	-0.413	0.042	0.023	-0.142	0.492	-0.502	0.061

(4) 분산-공분산 행렬의 고유벡터

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
AGR	0.741	0.298	0.263	-0.345	-0.193	0.087	-0.104	0.095	0.330
MIN	0.340	-0.532	-0.610	0.175	0.134	0.151	-0.196	0.103	0.330
MAN	-0.273	0.727	-0.448	0.069	0.128	0.117	-0.189	0.132	0.330
PS	-0.017	0.012	-0.004	0.008	0.025	-0.052	0.012	-0.938	0.341
CON	-0.049	-0.032	0.054	0.293	-0.346	-0.800	-0.122	0.156	0.329
SER	-0.198	-0.094	0.400	0.500	-0.344	0.544	-0.101	0.085	0.332
FIN	-0.047	-0.076	0.428	0.055	0.809	-0.120	-0.056	0.148	0.334
SPS	-0.464	-0.290	0.024	-0.710	-0.181	0.047	-0.205	0.108	0.328
TC	-0.034	-0.014	-0.110	-0.049	-0.042	0.024	0.919	0.135	0.344

고유벡터에서도 차이를 보인다는 것을 알 수 있다. 상관행렬에서의 제2 주성분은 MAN, TC 수치와 CON, SER, FIN 수치 대비를 나타낸다. 분산-공분산 행렬에서의 제2 주성분은 AGR, MAN 수치와 MIN, SER, SPS 수치 대비를 나타낸다. 이와 같이 어떤 직종이 영향을 미치는지의 결과가 다르다는 것을 확인할 수 있다.

설문조사처럼 같은 scale 점수화가 된 경우에는 공분산행렬 사용해도 되지만 변수들의 scale 이 위의 데이터처럼 많이 다른 경우 특정 변수가 전체적인 경향을 좌우하기 때문에 값이 다르게 나옴을 확인할 수 있었다. 이러한 편향을 줄이기 위해 상관계수 행렬을 사용하여 분석하는 것이 더 좋다.