Multivariate Analysis for the Behavioral Sciences, Second Edition (Chapman and Hall/CRC, 2019) **Solutions to Exercises of Chapter 13: Principal Components Analysis**

Kimmo Vehkalahti and Brian S. Everitt

6 December 2018

Solutions

Exercise 13.2

```
R <- matrix(c(
  1,      0.6579, 0.0034,
  0.6579, 1,     -0.0738,
  0.0034,-0.0738, 1
  ), ncol=3, byrow=T)

R_pc <- princomp(covmat=R)

# get component variances
varpc <- R_pc$sd^2
R_pc$loadings[, 1]   %*% diag(varpc[1])   %*% t(R_pc$loadings[, 1])
```

```
##             [,1]        [,2]         [,3]
## [1,]  0.49406120  0.49715685 -0.052913081
## [2,]  0.49715685  0.50027190 -0.053244619
## [3,] -0.05291308 -0.05324462  0.005666897
```

```
R_pc$loadings[, 1:2] %*% diag(varpc[1:2]) %*% t(R_pc$loadings[, 1:2])
```

```
##            [,1]        [,2]        [,3]
## [1,] 0.83339546  0.82554604  0.02293278
## [2,] 0.82554604  0.83130595 -0.09345488
## [3,] 0.02293278 -0.09345488  0.99770997
```

```
# just to show the three components reproduce R
R_pc$loadings[, 1:3] %*% diag(varpc[1:3]) %*% t(R_pc$loadings[, 1:3])
```

```
##        [,1]    [,2]    [,3]
## [1,] 1.0000  0.6579  0.0034
## [2,] 0.6579  1.0000 -0.0738
## [3,] 0.0034 -0.0738  1.0000
```

Exercise 13.3

```
R <- matrix(c(
  1    , 0.402, 0.396, 0.301, 0.305, 0.339, 0.340,
  0.402, 1    , 0.618, 0.150, 0.135, 0.206, 0.183,
  0.396, 0.618, 1    , 0.321, 0.289, 0.363, 0.345,
  0.301, 0.150, 0.321, 1    , 0.846, 0.759, 0.661,
  0.305, 0.135, 0.289, 0.846, 1    , 0.797, 0.800,
  0.339, 0.206, 0.363, 0.759, 0.797, 1    , 0.736,
  0.340, 0.183, 0.345, 0.661, 0.800, 0.736, 1
  ), ncol=7,byrow=T)

labels <- c("HL", "HB", "FB", "LFinL", "LForL", "LFootL", "Height")
dimnames(R) <- list(labels, labels)

R
```

```
##           HL    HB    FB LFinL LForL LFootL Height
## HL     1.000 0.402 0.396 0.301 0.305  0.339  0.340
## HB     0.402 1.000 0.618 0.150 0.135  0.206  0.183
## FB     0.396 0.618 1.000 0.321 0.289  0.363  0.345
## LFinL  0.301 0.150 0.321 1.000 0.846  0.759  0.661
## LForL  0.305 0.135 0.289 0.846 1.000  0.797  0.800
## LFootL 0.339 0.206 0.363 0.759 0.797  1.000  0.736
## Height 0.340 0.183 0.345 0.661 0.800  0.736  1.000
```

```
R_pc <- princomp(covmat = R)
summary(R_pc, loadings = TRUE)
```

```
## Importance of components:
##                            Comp.1    Comp.2     Comp.3    Comp.4     Comp.5
## Standard deviation     1.9492241 1.2256950 0.80610632 0.6000474 0.58237656
## Proportion of Variance 0.5427821 0.2146183 0.09282963 0.0514367 0.04845178
## Cumulative Proportion  0.5427821 0.7574004 0.85023003 0.9016667 0.95011851
##                            Comp.6     Comp.7
## Standard deviation     0.48502898 0.33751644
## Proportion of Variance 0.03360759 0.01627391
## Cumulative Proportion  0.98372609 1.00000000
##
## Loadings:
##        Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
## HL     -0.276 -0.365  0.882
## HB     -0.212 -0.639 -0.258  0.687
## FB     -0.295 -0.512 -0.381 -0.699 -0.101
## LFinL  -0.438  0.235         0.102 -0.619  0.318  0.503
## LForL  -0.456  0.277         0.113         0.290 -0.785
## LFootL -0.450  0.178                      -0.870
## Height -0.436  0.180               0.770  0.233  0.353
```

Exercise 13.4

```
prestige <- structure(
  c(82, 90, 76, 90, 87, 93, 90, 88, 89, 97, 59, 73, 81, 45, 39, 34, 41, 16,
    33, 53, 67, 57, 26, 29, 10, 15, 19, 10, 13, 24, 20, 7, 16, 11, 8, 41,

    23.8, 37.5, 37, 20.7, 10.6, 14.2, 45.6, 31.9, 24.3, 31.9, 16, 16.8, 64.8,
    47.3, 21.9, 16.5, 32.4, 24.1, 32.7, 30.8, 34.2, 34.5, 24.4, 29.4, 14.4,
    41.7, 19.2, 24.9, 17.9, 15.7, 36, 24.4, 42.2, 38.2, 20.3, 47.6,

    3977, 5509, 4303, 4091, 2410, 4366, 6448, 4590, 6284, 8302, 3176, 3456,
    4700, 3806, 2828, 3480, 3771, 2543, 2450, 3447, 4648, 3303, 2693, 3353,
    1898, 2410, 3424, 2213, 2590, 2915, 2357, 1942, 2249, 2551, 1866, 2866,

    14.4, 16, 15.6, 16, 16, 16, 16, 16, 16, 16, 15.8, 16, 12.2, 11.6, 12.7,
    12.2, 12.7, 12.1, 8.7, 11.1, 8.8, 9.6, 9.4, 9.3, 10.3, 8.2, 9.2, 8.9, 9.6,
    9.6, 8.8, 9.8, 8.7, 8.5, 8.2, 10.6),

.Dim = c(36L, 4L),
.Dimnames = list(c("Accountants", "Architects", "Authors", "Chemists",
                   "Clergymen", "Academics", "Dentists", "Civengineers",
                   "Lawyers", "Physicians", "Socialwk", "Teachers",
                   "Mangmanuf", "Mangretail", "Bookkeepers", "Mail-carriers",
                   "Insurag", "Salesman", "Carpenters", "Electricians",
                   "Locmeng", "Machinists", "Mechanics", "Plumbers",
                   "Parkingatt", "Miners", "Railwaydr", "Taxidr", "Truckdr",
                   "Machoper", "Barbers", "Waiters", "Cooks", "Watchmen",
                   "Janitors", "Policemen"),
                 c("Prestige", "Suicide", "Medinc", "Mediansch")
                 )
  )
head(prestige); tail(prestige)
```

```
##             Prestige Suicide Medinc Mediansch
## Accountants       82    23.8   3977       14.4
## Architects        90    37.5   5509       16.0
## Authors           76    37.0   4303       15.6
## Chemists          90    20.7   4091       16.0
## Clergymen         87    10.6   2410       16.0
## Academics         93    14.2   4366       16.0
```

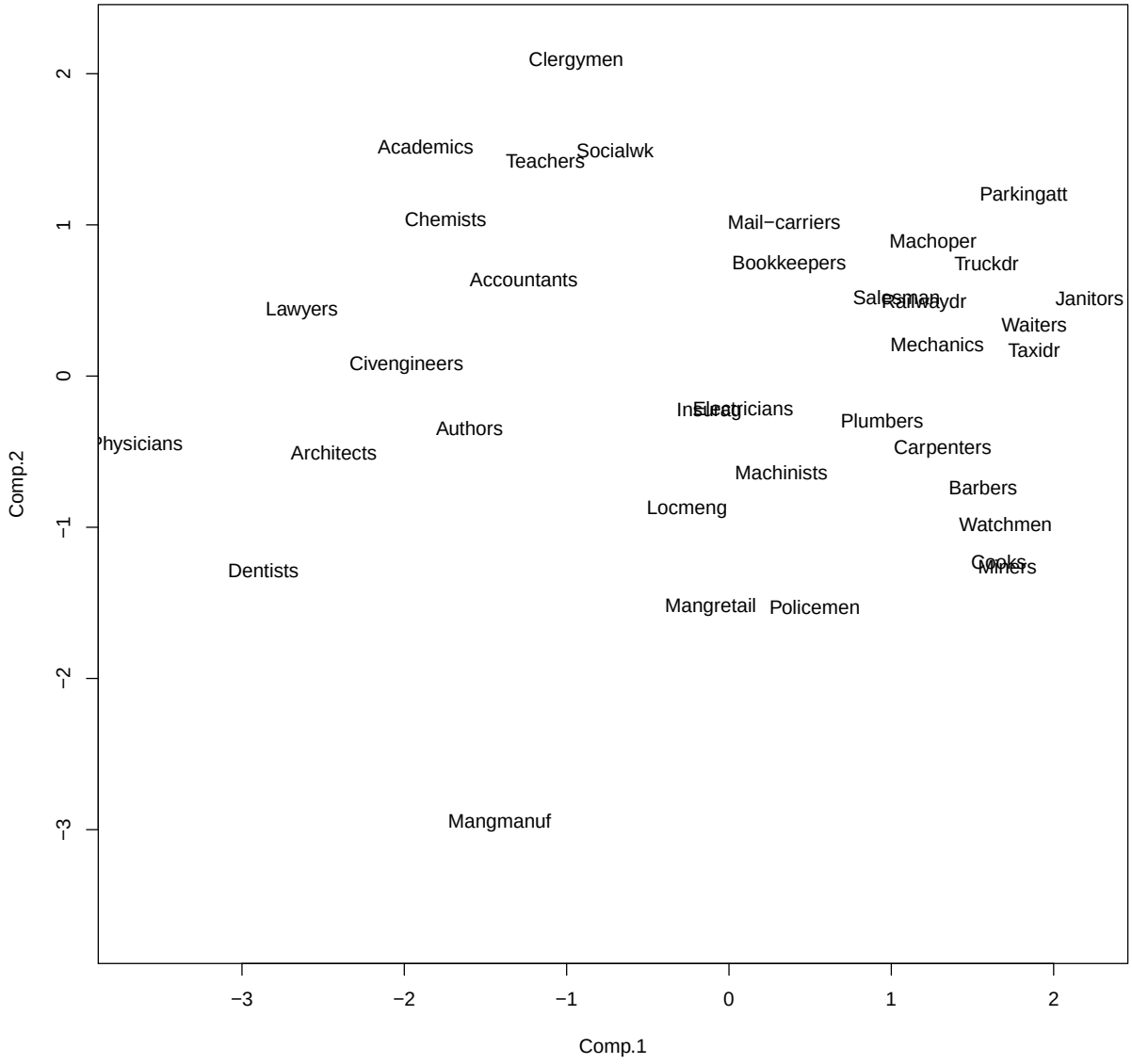
```
##           Prestige Suicide Medinc Mediansch
## Barbers         20    36.0   2357        8.8
## Waiters          7    24.4   1942        9.8
## Cooks           16    42.2   2249        8.7
## Watchmen        11    38.2   2551        8.5
## Janitors         8    20.3   1866        8.2
## Policemen       41    47.6   2866       10.6
```

```
prestige_pc <- princomp(prestige, cor = TRUE)
summary(prestige_pc, loadings = TRUE)
```

```
## Importance of components:
##                           Comp.1    Comp.2     Comp.3     Comp.4
## Standard deviation     1.5956984 1.0505238 0.50377308 0.31041740
```

```
## Proportion of Variance 0.6365634 0.2759000 0.06344683 0.02408974
## Cumulative Proportion  0.6365634 0.9124634 0.97591026 1.00000000
##
## Loadings:
##            Comp.1 Comp.2 Comp.3 Comp.4
## Prestige   -0.604         0.257  0.754
## Suicide           -0.929  0.331 -0.143
## Medinc     -0.559 -0.203 -0.785 -0.172
## Medianschy -0.563  0.308  0.456 -0.617
```

```r
# use first two component scores to plot the data
xlim <- range(prestige_pc$scores[, 1])
plot(prestige_pc$scores[, 1:2], ylim = xlim, type = "n")
text(prestige_pc$scores[, 1:2], row.names(prestige), cex=1) # abbreviate() may also be used
```



```r
options(digits=3)
cbind(prestige_pc$scores[, 1:2]) # helpful in interpreting the plot
```

```
##                Comp.1 Comp.2
## Accountants   -1.2639  0.637
## Architects    -2.4338 -0.503
## Authors       -1.5976 -0.342
## Chemists      -1.7453  1.039
## Clergymen     -0.9406  2.085
## Academics     -1.8678  1.519
## Dentists      -2.8681 -1.285
## Civengineers  -1.9862  0.074
## Lawyers       -2.6307  0.432
## Physicians    -3.6504 -0.459
## Socialwk      -0.7013  1.496
## Teachers      -1.1306  1.426
## Mangmanuf     -1.4123 -2.956
## Mangretail    -0.1122 -1.531
## Bookkeepers    0.3712  0.739
## Mail-carriers  0.3403  1.017
## Insurag       -0.1209 -0.232
## Salesman       1.0325  0.521
## Carpenters     1.3167 -0.483
## Electricians   0.0874 -0.210
## Locmeng       -0.2587 -0.879
## Machinists     0.3231 -0.634
## Mechanics      1.2835  0.207
## Plumbers       0.9434 -0.294
## Parkingatt     1.8168  1.196
## Miners         1.7146 -1.261
## Railwaydr      1.2020  0.487
## Taxidr         1.8793  0.171
## Truckdr        1.5876  0.747
## Machoper       1.2580  0.885
## Barbers        1.5657 -0.734
## Waiters        1.8806  0.340
## Cooks          1.6618 -1.225
## Watchmen       1.7047 -0.977
## Janitors       2.2223  0.513
## Policemen      0.5291 -1.525
```

Division into professional and nonprofessional perhaps?