hw4_prob1

Haoyu_Zhao, 2016012390

April 26, 2017

a

All of the correlation is positive between x_1, x_2, x_3, x_4. The correlation between x_1, x_2 and x_1, x_4 is relatively large. They are about 0.5.

b

```r
#the data input
vec <- c(1,0.4919,0.2635,0.4653,-0.2277,0.0652,
         0.4919,1,0.3127,0.3506,-0.1917,0.2045,
         0.2636,0.3127,1,0.4108,0.0647,0.2493,
         0.4653,0.3506,0.4108,1,-0.2249,0.2293,
         -0.2277,-0.1917,0.0647,-0.2249,1,-0.2144,
         0.0652,0.2045,0.2493,0.2293,-0.2144,1)

R <- matrix(vec, c(6,6))
#print the data
print(R)
```

```
##         [,1]    [,2]   [,3]    [,4]    [,5]    [,6]
## [1,]  1.0000  0.4919 0.2636  0.4653 -0.2277  0.0652
## [2,]  0.4919  1.0000 0.3127  0.3506 -0.1917  0.2045
## [3,]  0.2635  0.3127 1.0000  0.4108  0.0647  0.2493
## [4,]  0.4653  0.3506 0.4108  1.0000 -0.2249  0.2293
## [5,] -0.2277 -0.1917 0.0647 -0.2249  1.0000 -0.2144
## [6,]  0.0652  0.2045 0.2493  0.2293 -0.2144  1.0000
```

```r
R_p <- R[1:4, 1:4]

#applying the PCA to the first 4 component
R_p.pc <- princomp(R_p)

summary(R_p.pc)
```

```
## Importance of components:
##                           Comp.1    Comp.2   Comp.3       Comp.4
## Standard deviation     0.3973105 0.3105456 0.221815 1.747316e-08
## Proportion of Variance 0.5201241 0.3177589 0.162117 1.005981e-15
## Cumulative Proportion  0.5201241 0.8378830 1.000000 1.000000e+00
```

```r
#compute the eigenvector and the eigen value of R_p
e <- eigen(R_p)
print(e)
```

```
## $values
## [1] 2.1539189 0.7875498 0.6156590 0.4428723
```

```
##
## $vectors
##            [,1]       [,2]       [,3]       [,4]
## [1,] -0.5265384 -0.4571582  0.2491110 -0.6719808
## [2,] -0.5033080 -0.4119833 -0.6142538  0.4467865
## [3,] -0.4427711  0.7584337 -0.3680028 -0.3056057
## [4,] -0.5228691  0.2146031  0.6520812  0.5053996
```

From the summary of the PCA and the eigen value and the eigen vector, we can do the following analysis.

The first PC is a weighted sum of the 4 elements(omit all the minus sign), and the second is contrast Bluegill and Crappie with the 2 bass. The third PC contrast Bluegill and Largemouth with others and the fourth contrast Bluegill and Smallmouth with others.

C

Applying the PC analysis to all the 6 elements.

```r
#do the PC analysis
R.pc <- princomp(R)
#print the summary
summary(R.pc)
```

```
## Importance of components:
##                           Comp.1    Comp.2    Comp.3     Comp.4     Comp.5
## Standard deviation     0.6901267 0.4044247 0.3033912 0.24598212 0.17317918
## Proportion of Variance 0.5791430 0.1988857 0.1119268 0.07357584 0.03646864
## Cumulative Proportion  0.5791430 0.7780287 0.8899555 0.96353136 1.00000000
##                              Comp.6
## Standard deviation     3.535557e-09
## Proportion of Variance 1.520001e-17
## Cumulative Proportion  1.000000e+00
```

```r
#print the eigenvalue and the eigenvectors
e <- eigen(R)
print(e)
```

```
## $values
## [1] 2.3549250 1.0718567 0.9842486 0.6643858 0.5003897 0.4241942
##
## $vectors
##            [,1]        [,2]        [,3]        [,4]       [,5]       [,6]
## [1,]  0.4753543  0.02188910  0.47987482  0.04564932  0.3578888  0.6425051
## [2,]  0.4719328 -0.01924243  0.20906455  0.70296351 -0.1771358 -0.4556533
## [3,]  0.3931540 -0.56061775 -0.26440849 -0.17551515 -0.5973982  0.2713733
## [4,]  0.4963538 -0.07723024  0.03226116 -0.60426514  0.3238962 -0.5260914
## [5,] -0.2563177 -0.80502150  0.01294351  0.21817123  0.4823355 -0.0768220
## [6,]  0.2910014  0.17559671 -0.80925398  0.24539324  0.3822272  0.1524794
```

From this analysis, we can see that the first PC contrast Walleye with the other places, the second PC is largely the sum of the Smallmouth and the Walleye, the third is almost the contrast between Bluegill and Northren pike because the coeffient in the eigenvector is much bigger than the others. We can also apply the same analysis to the last 3 PCs but becuase the first 3 is the most important and we do not interpret the last 3 PCs.