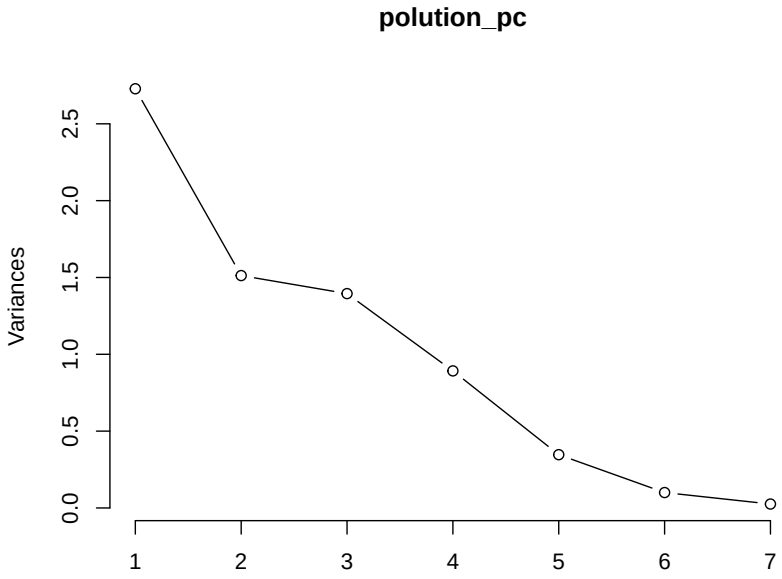
HW3

Han Ambrose

April 17, 2021

Question 4



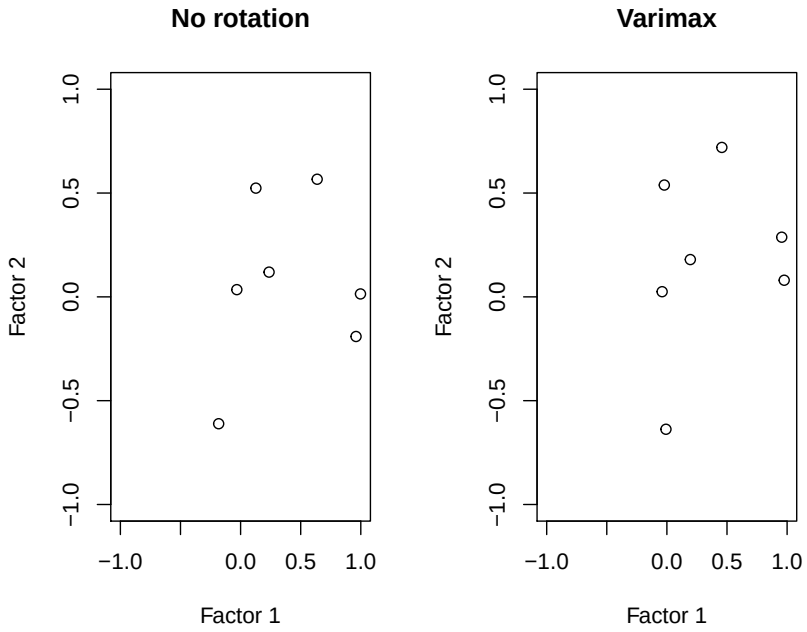
```
Call:
factanal(x = USairpollution, factors = 1, rotation = "none")

Uniquenesses:
    SO2    temp    manu   popul    wind  precip predays
  0.590   0.965   0.005   0.085   0.943   0.999   0.983

Loadings:
        Factor1
SO2      0.640
temp    -0.186
manu     0.998
popul    0.957
wind     0.238
precip
predays  0.130

               Factor1
SS loadings      2.429
Proportion Var   0.347

Test of the hypothesis that 1 factor is sufficient.
The chi square statistic is 83.96 on 14 degrees of freedom.
The p-value is 5.18e-12
```

factor of 1 explains 34% of variation

```
Call:
factanal(x = USairpollution, factors = 2, rotation = "none")

Uniquenesses:
    SO2    temp    manu   popul    wind  precip predays
  0.273   0.593   0.005   0.041   0.930   0.998   0.710

Loadings:
        Factor1 Factor2
SO2      0.637   0.567
temp    -0.182  -0.611
manu     0.997
popul    0.960  -0.191
wind     0.237   0.119
precip
predays  0.127   0.524

               Factor1 Factor2
SS loadings      2.429   1.021
Proportion Var   0.347   0.146
Cumulative Var   0.347   0.493

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 52.99 on 8 degrees of freedom.
The p-value is 1.08e-08
```

fac-

tor of 2 explains 50% of variation



Question 5

```
Importance of components:
                          Comp.1    Comp.2    Comp.3       Comp.4       Comp.5       Comp.6
Standard deviation     2.1596823 1.9779536 1.1930305 9.262084e-03 4.014081e-03 3.413545e-03 3.3
Proportion of Variance 0.4664228 0.3912300 0.1423322 8.578621e-06 1.611285e-06 1.165229e-06 1.1
Cumulative Proportion  0.4664228 0.8576528 0.9999850 9.999936e-01 9.999952e-01 9.999964e-01 9.9
                           Comp.8       Comp.9      Comp.10
Standard deviation     3.077326e-03 2.927989e-03 2.666065e-03
Proportion of Variance 9.469934e-07 8.573119e-07 7.107903e-07
Cumulative Proportion  9.999984e-01 9.999993e-01 1.000000e+00
```

Sparse PCA
The higher the parameters the sparser the components.

```
> PP_spca_penalty <- spca(pitprops,K = 7, type = "Gram", sparse = "penalty", para=c(0.06,0.16,0.1,0.5,0.5,0.
5,.4))
> ### Loadings
> PP_spca_penalty$loadings
          PC1        PC2       PC3 PC4 PC5 PC6 PC7
V1  0.0000000 -0.5079151 0.0000000   0   0   0   0
V2  0.0000000 -0.4978297 0.0000000   0   0   0   0
V3  0.0000000 -0.4937902 0.0000000   0   0   0   0
V4  0.0000000 -0.5003590 0.0000000   0   0   0   0
V5  0.3559174  0.0000000 0.0000000   0   0   0   0
V6  0.3456626  0.0000000 0.0000000   0   0   0   0
V7  0.4439655  0.0000000 0.0000000   0   0   0   0
V8  0.7461466  0.0000000 0.0000000   0   0   0   0
V9  0.0000000  0.0000000 0.6591080   0   0   0   0
V10 0.0000000  0.0000000 0.7520483   0   0   0   0
```

```
> PP_spca_penalty <- spca(pitprops,K = 7, type = "Gram", sparse = "penalty", para=c(0.5,0.5,0.5,0.5,0.5,0.5,.
4))
> ### Loadings
> PP_spca_penalty$loadings
           PC1        PC2      PC3 PC4 PC5 PC6 PC7
V1  0.00000000 -0.1847566 0.000000   0   0   0   0
V2  0.00000000 -0.3360385 0.000000   0   0   0   0
V3  0.00000000 -0.6758950 0.000000   0   0   0   0
V4  0.00000000 -0.6293720 0.000000   0   0   0   0
V5  0.04105809  0.0000000 0.000000   0   0   0   0
V6  0.01570530  0.0000000 0.000000   0   0   0   0
V7  0.25123402  0.0000000 0.000000   0   0   0   0
V8  0.96692763  0.0000000 0.000000   0   0   0   0
V9  0.00000000  0.0000000 0.269623   0   0   0   0
```

```
> PP_spca_varnum <- spca(pitprops,K = 7, type = "Gram", sparse = "varnum", para = c(7,4,4,2,1,1,1))
> ### Loadings
> PP_spca_varnum$loadings
              PC1        PC2         PC3        PC4 PC5 PC6 PC7
V1   0.0000000000 -0.5009531  0.00000000  0.0000000   0   0   0
V2   0.0000000000 -0.4955046  0.00000000  0.0000000   1   0   0
V3  -0.0001205712 -0.5022059  0.00000000  0.0000000   0  -1   0
V4   0.0000000000 -0.5013087  0.00000000  0.0000000   0   0   1
V5   0.4965882465  0.0000000 -0.03875686  0.0000000   0   0   0
V6   0.5012252843  0.0000000 -0.08848437  0.0000000   0   0   0
V7   0.4970217053  0.0000000  0.00000000  0.0000000   0   0   0
V8   0.5008988657  0.0000000  0.00000000  0.0000000   0   0   0
V9   0.0477001959  0.0000000  0.67875136  0.7249099   0   0   0
V10  0.0443593899  0.0000000  0.72798696 -0.6888436   0   0   0
```

Question 6

6ab

This portions are just installation and loading data

6c

```
Call:
lda(r ~ ., data = urine)

Prior probabilities of groups:
        0         1 
0.5714286 0.4285714 

Group means:
   gravity       ph     osmo     cond     urea     calc
0 1.015364 6.125682 561.6591 20.55000 232.4318 2.628864
1 1.021576 5.927273 682.8788 21.37879 302.3636 6.202424

Coefficients of linear discriminants:
                  LD1
gravity 140.610784925
ph       -0.125723350
osmo      0.003408930
cond     -0.117156709
urea     -0.008383726
calc      0.301091251
```

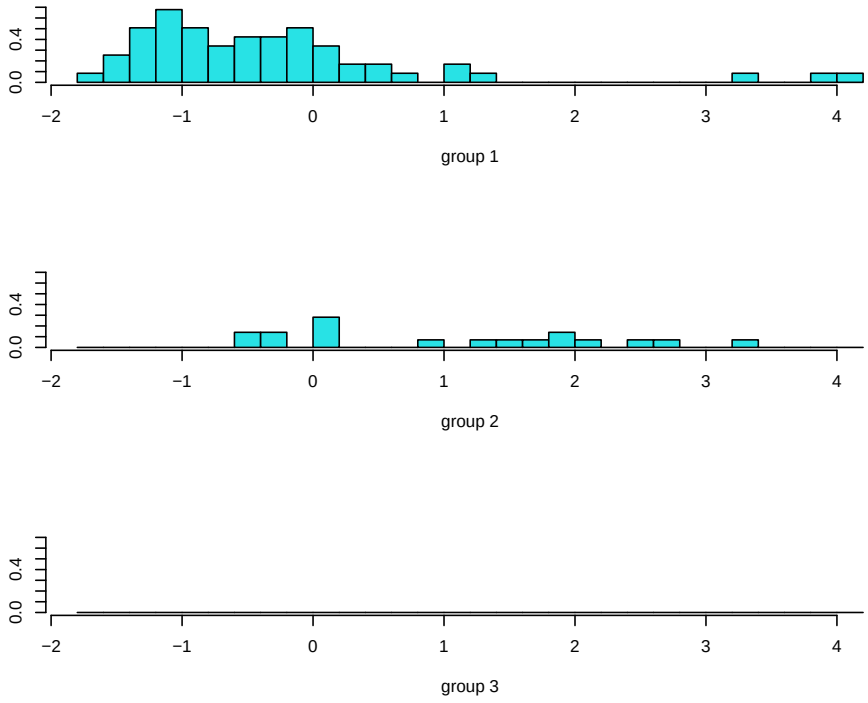
gravity is the dominant feature

6d

79% of data points are predicted correctly

```
> urine$r
 [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
[51] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
> urine.lda.values$class
 [1] 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
[51] 0 0 0 0 1 1 1 0 1 0 1 1 0 0 1 0 0 0 1 1 0 1 1 0 1 1 1
Levels: 0 1
> length(urine$r)
[1] 77
> sum(urine$r==urine.lda.values$class)/length(urine$r)
[1] 0.7922078
```

6e



points are somewhat separated but not quite obvious as there are some points in group 2 are in the region of group 1.