10largest

Yuchen Li

2/6/2022

```r
# 10largest.r
#
d0=read.table("10largest.txt")
names(d0)=c("sales","profit","assets")
d0
```

```
##     sales profit  assets
## 1  108.28  17.05 1484.10
## 2  152.36  16.59  750.33
## 3   95.04  10.91  766.42
## 4   65.45  14.14 1110.46
## 5   62.97   9.52 1031.29
## 6  263.99  25.33  195.26
## 7  265.19  18.54  193.83
## 8  285.06  15.73  191.11
## 9   92.01   8.10 1175.16
## 10 165.68  11.13  211.15
```

```r
# only sales and profit
d1=d0[,c(1,2)]
str(d1)
```

```
## 'data.frame':    10 obs. of  2 variables:
##  $ sales : num  108.3 152.4 95 65.5 63 ...
##  $ profit: num  17.05 16.59 10.91 14.14 9.52 ...
```

```r
#
# Principal Components Analysis
# R will center the data by default
#
# prcomp
#
pr1 = prcomp(d1)
names(pr1)
```

```
## [1] "sdev"     "rotation" "center"   "scale"    "x"
```

```r
#
# sdev:     square root of eigenvalues
# rotation: matrix with eigenvectors
```

```r
# center:   col means of original -uncentered- data
# x:        transformed dataset
#
# eigenvalues of covariance matrix of centered dataset
#
pr1$sdev^2
```

```
## [1] 7488.80605   13.83751
```

```r
#
# eigenvectors of covariance matrix of centered dataset
#
pr1$rotation
```

```
##               PC1         PC2
## sales  0.99917338  0.04065165
## profit 0.04065165 -0.99917338
```

```r
# columns are eigenvectors (loadings)
#
pr1$center
```

```
##   sales  profit
## 155.603  14.704
```

```r
pr1$scale
```

```
## [1] FALSE
```

```r
#
# transformed data (score vectors) Z_ij
#
pr1$x
```

```
##              PC1        PC2
##  [1,] -47.188513 -4.2678188
##  [2,]  -3.163650 -2.0162743
##  [3,] -60.667170  1.3288779
##  [4,] -90.101405 -3.1013344
##  [5,] -92.767166  1.4140305
##  [6,] 108.729370 -6.2111060
##  [7,] 109.652353  0.6220633
##  [8,] 129.391697  4.2374888
##  [9,] -63.808896  4.0133806
## [10,]   9.923381  3.9806923
```

```r
#
# save eigenvectors in rot
rot = pr1$rotation
#
```

```r
# data used in prcomp was centered as shown below
#
d1centered = scale(d1,scale=F)
#
# scale function creates a matrix
class(d1centered)
```

```
## [1] "matrix" "array"
```

```r
d1centered
```

```
##          sales profit
##  [1,] -47.323  2.346
##  [2,]  -3.243  1.886
##  [3,] -60.563 -3.794
##  [4,] -90.153 -0.564
##  [5,] -92.633 -5.184
##  [6,] 108.387 10.626
##  [7,] 109.587  3.836
##  [8,] 129.457  1.026
##  [9,] -63.593 -6.604
## [10,]  10.077 -3.574
## attr(,"scaled:center")
##   sales  profit
## 155.603  14.704
```

```r
#
# centered columns have mean zero
#
colMeans(d1centered)
```

```
##         sales        profit
## -4.263256e-15 -7.105427e-16
```

```r
#
# transformed data (score vectors) Z_ij
#
# multiply centered data by matrix of eigenvectors
#
d1centered%*%rot
```

```
##              PC1        PC2
##  [1,] -47.188513 -4.2678188
##  [2,]  -3.163650 -2.0162743
##  [3,] -60.667170  1.3288779
##  [4,] -90.101405 -3.1013344
##  [5,] -92.767166  1.4140305
##  [6,] 108.729370 -6.2111060
##  [7,] 109.652353  0.6220633
##  [8,] 129.391697  4.2374888
##  [9,] -63.808896  4.0133806
## [10,]   9.923381  3.9806923
```

```r
#
# same as
#
pr1$x
```

```
##               PC1        PC2
##  [1,] -47.188513 -4.2678188
##  [2,]  -3.163650 -2.0162743
##  [3,] -60.667170  1.3288779
##  [4,] -90.101405 -3.1013344
##  [5,] -92.767166  1.4140305
##  [6,] 108.729370 -6.2111060
##  [7,] 109.652353  0.6220633
##  [8,] 129.391697  4.2374888
##  [9,] -63.808896  4.0133806
## [10,]   9.923381  3.9806923
```

```r
#
# covariance matrix of PC1,PC2 is diagonal matrix
# with eigenvalues lambda in main diagonal
# showing that PC1, PC2 are uncorrelated
#
Sigma_PC = var(pr1$x)
Sigma_PC
```

```
##              PC1          PC2
## PC1 7.488806e+03 9.094500e-14
## PC2 9.094500e-14 1.383751e+01
```

```r
#
# main diagonal is equal to variances of columns of original data (X1,X2)
#
diag(Sigma_PC)
```

```
##        PC1        PC2
## 7488.80605   13.83751
```

```r
sum(diag(Sigma_PC))
```

```
## [1] 7502.644
```

```r
#
# this sum is equal to
# the sum of column variances
#
apply(d1,2,var)
```

```
##      sales     profit
## 7476.45325   26.19032
```

```
sum(apply(d1,2,var))
```

```
## [1] 7502.644
```

```
#
# Eigenvalue Analysis
#
# eigenvalues and eigenvectors of Covariance Matrix
#
Sigma_X = var(d1centered)
eigen(Sigma_X)
```

```
## eigen() decomposition
## $values
## [1] 7488.80605   13.83751
##
## $vectors
##             [,1]        [,2]
## [1,] -0.99917338  0.04065165
## [2,] -0.04065165 -0.99917338
```

```
#
# eigen() and prcomp() results agree
#
# scaled data
#==================================================================
#
d1scaled = scale(d1,scale=T)
d1scaled
```

```
##             sales     profit
##  [1,] -0.54729875  0.4584138
##  [2,] -0.03750586  0.3685287
##  [3,] -0.70042166 -0.7413563
##  [4,] -1.04263517 -0.1102069
##  [5,] -1.07131681 -1.0129654
##  [6,]  1.25351457  2.0763447
##  [7,]  1.26739278  0.7495632
##  [8,]  1.49719279  0.2004827
##  [9,] -0.73546414 -1.2904367
## [10,]  0.11654226 -0.6983678
## attr(,"scaled:center")
##   sales  profit
## 155.603  14.704
## attr(,"scaled:scale")
##     sales    profit
## 86.466486  5.117647
```

```
#
# # PC1, PC2 on scaled data
#
```

```
pr3 = prcomp(d1,scale=T)
d3x = pr3$x
d3x
```

```
##               PC1         PC2
##  [1,] -0.06285118 -0.71114613
##  [2,]  0.23406850 -0.28710980
##  [3,] -1.01949095  0.02894513
##  [4,] -0.81518244 -0.65932636
##  [5,] -1.47381011 -0.04126065
##  [6,]  2.35456604 -0.58182874
##  [7,]  1.42620322  0.36616083
##  [8,]  1.20043788  0.91691247
##  [9,] -1.43252821  0.39242485
## [10,] -0.41141275  0.57622840
```

```
#
# Covariance Matrix LAMBDA on PC1,PC2 on scaled data
#
Sigma_PC = var(d3x)
Sigma_PC
```

```
##              PC1          PC2
## PC1 1.686136e+00 3.361674e-16
## PC2 3.361674e-16 3.138640e-01
```

```
#
# eigenvalues of PC1, PC2
#
diag(Sigma_PC)
```

```
##      PC1      PC2
## 1.686136 0.313864
```

```
sum(diag(Sigma_PC))
```

```
## [1] 2
```

```
#
```