Lab 3 732A97

Raymond Sseguya

2019-12-07

Data

```
library(dplyr)

trackrcs <- read.table("T1-9.dat",
  col.names = c("countries", "x100m", "x200m",
        "x400m", "x800m", "x1500m", "x3000m", "marathon"))
rownames(trackrcs) <- (trackrcs)[,1]
```

Question 1: Principal components, including interpretation of them

a) Obtain the sample correlation matrix R for these data, and determine its eigenvalues and eigenvectors.

```
S <- cov((trackrcs)[,-1])
R <- cov2cor(S); R
```

```
##              x100m     x200m     x400m     x800m    x1500m    x3000m
## x100m    1.0000000 0.9410886 0.8707802 0.8091758 0.7815510 0.7278784
## x200m    0.9410886 1.0000000 0.9088096 0.8198258 0.8013282 0.7318546
## x400m    0.8707802 0.9088096 1.0000000 0.8057904 0.7197996 0.6737991
## x800m    0.8091758 0.8198258 0.8057904 1.0000000 0.9050509 0.8665732
## x1500m   0.7815510 0.8013282 0.7197996 0.9050509 1.0000000 0.9733801
## x3000m   0.7278784 0.7318546 0.6737991 0.8665732 0.9733801 1.0000000
## marathon 0.6689597 0.6799537 0.6769384 0.8539900 0.7905565 0.7987302
##           marathon
## x100m    0.6689597
## x200m    0.6799537
## x400m    0.6769384
## x800m    0.8539900
## x1500m   0.7905565
## x3000m   0.7987302
## marathon 1.0000000
```

b) Determine the first two principal components for the standardized variables. Prepare a table showing the correlations of the standardized variables with the components, and the cumulative percentage of the total (standardized) sample variance explained by the two components.

the first two principal components for the standardized variables

```
res=prcomp((trackrcs)[,-1], scale. = TRUE)

# Each PC is a linear combination of the original variables
#### res$rotation
res$rotation[,1:2]
```

```
##                PC1        PC2
## x100m    -0.3777657  0.4071756
## x200m    -0.3832103  0.4136291
## x400m    -0.3680361  0.4593531
## x800m    -0.3947810 -0.1612459
## x1500m   -0.3892610 -0.3090877
## x3000m   -0.3760945 -0.4231899
## marathon -0.3552031 -0.3892153
```

correlation of the standardized variables with the components

```
CorWithPC <- cor( t(scale((trackrcs)[,-1])),  res$rotation[,1:2] )
CorWithPC
```

```
##               PC1          PC2
## ARG  -0.438037879  0.202079509
## AUS  -0.291997723 -0.697604768
## AUT   0.847704287 -0.271463401
## BEL   0.109195340  0.002591797
## BER   0.624314679 -0.718694306
## BRA   0.021484856 -0.676133654
## CAN   0.315710477 -0.534890462
## CHI   0.110157560  0.850572320
## CHN   0.552766808 -0.101456102
## COL  -0.387269032 -0.809584489
## COK   0.669238352  0.418809631
## CRC  -0.629065898 -0.439021331
## CZE   0.299939530 -0.598517614
## DEN  -0.047823857  0.958227842
## DOM  -0.438173784 -0.641470441
## FIN  -0.273226905 -0.765893001
## FRA   0.187191789 -0.946533854
## GER   0.116536774 -0.962377701
## GBR  -0.373542566  0.077219176
## GRE   0.153920114 -0.731302098
## GUA  -0.278595537  0.792584770
## HUN   0.171576469  0.844203249
## INA  -0.465608590  0.584080299
## IND   0.636958939 -0.418322902
## IRL  -0.274350156  0.791341688
## ISR  -0.500275551 -0.498372960
## ITA   0.510425707  0.678934090
## JPN  -0.635906243  0.753460359
## KEN  -0.107481518  0.904527496
## KORS -0.640749240  0.682166246
## KORN  0.003796409  0.896190585
## LUX  -0.241381463  0.873332288
## MAS  -0.168333644 -0.593428743
## MRI   0.479189202  0.733624293
## MEX  -0.683892340 -0.441971198
## MYA   0.209421271  0.827091539
## NED   0.369678314  0.432312011
## NZL   0.118190761  0.814989175
## NOR  -0.320502294  0.886700480
## PNG   0.457777161 -0.479234975
## PHI  -0.043993952 -0.495302008
## POL   0.251925206 -0.969982671
## POR   0.154134596  0.939543376
## ROM   0.458751400  0.465524691
## RUS   0.648562276 -0.690875300
## SAM  -0.431149206 -0.683730154
## SIN  -0.582405129  0.673578150
```

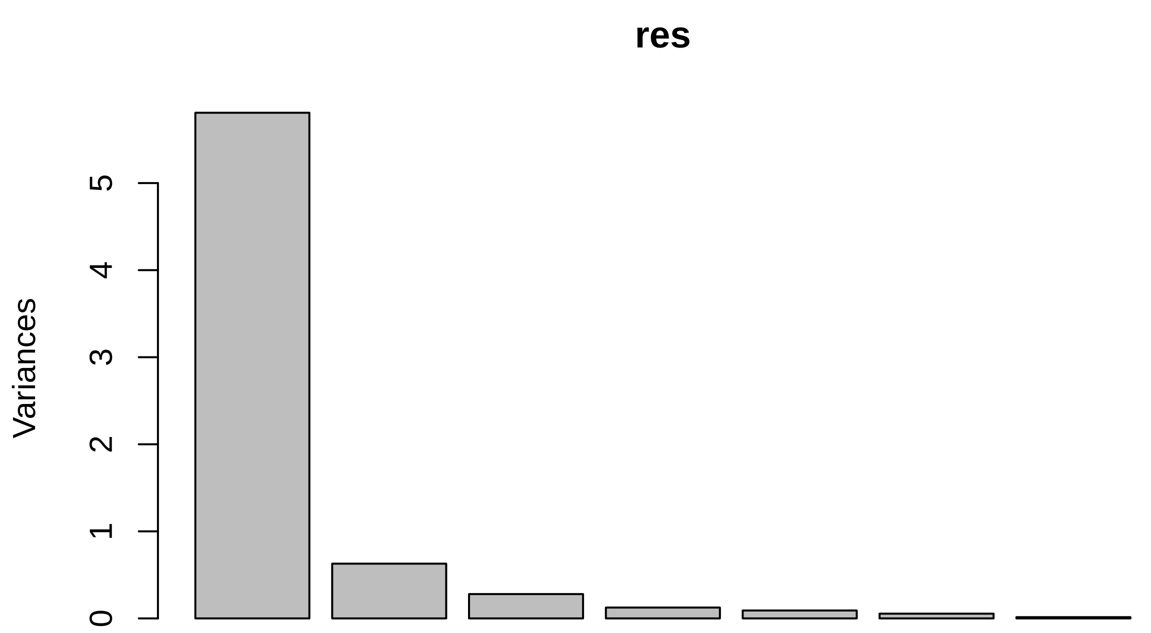
```
## ESP   0.390327787 -0.709215238
## SWE   0.617016907 -0.003157090
## SUI   0.216540220  0.850862874
## TPE  -0.095764950 -0.783031341
## THA  -0.003608943 -0.832174030
## TUR   0.452267074  0.517762412
## USA   0.138598843 -0.838662319
```

cummulative percentage of total standardized sample variance explained by the 2 components

```
eigenvalues=res$sdev^2
# proportion of variation
sprintf("%2.3f",eigenvalues/sum(eigenvalues)*100)[1:2]
```

```
## [1] "82.966" "8.981"
```

```
screeplot(res)
```




```
CorWithPC %>% apply(MARGIN=2,FUN=abs) %>%
  apply(MARGIN=2,FUN=function(a) 100*cumsum(a)/sum(a))
```

```
##              PC1         PC2
```

```
## ARG    2.349178   0.5809434
## AUS    3.915148   2.5864357
## AUT    8.461348   3.3668457
## BEL    9.046958   3.3742967
## BER   12.395130   5.4404177
## BRA   12.510352   7.3841843
## CAN   14.203493   8.9219013
## CHI   14.794263  11.3671487
## CHN   17.758727  11.6588173
## COL   19.835634  13.9862318
## COK   23.424729  15.1902366
## CRC   26.798381  16.4523465
## CZE   28.406944  18.1729805
## DEN   28.663421  20.9277188
## DOM   31.013327  22.7718346
## FIN   32.478631  24.9736437
## FRA   33.482532  27.6947638
## GER   34.107514  30.4614322
## GBR   36.110806  30.6834239
## GRE   36.936272  32.7857901
## GUA   38.430368  35.0643334
## HUN   39.350525  37.4912708
## INA   41.847563  39.1704000
## IND   45.263545  40.3730056
## IRL   46.734873  42.6479752
## ISR   49.417828  44.0807107
## ITA   52.155218  46.0325280
## JPN   55.565555  48.1985954
## KEN   56.141974  50.7989546
## KORS  59.578283  52.7600638
## KORN  59.598643  55.3364557
## LUX   60.893161  57.8471341
## MAS   61.795927  59.5531384
## MRI   64.365797  61.6621806
## MEX   68.033481  62.9327708
## MYA   69.156598  65.3105151
## NED   71.139167  66.5533369
## NZL   71.773018  68.8962889
## NOR   73.491858  71.4453984
## PNG   75.946896  72.8231156
## PHI   76.182834  74.2470226
## POL   77.533897  77.0355540
## POR   78.360514  79.7365777
## ROM   80.820778  81.0748801
## RUS   84.298988  83.0610264
## SAM   86.611222  85.0266316
## SIN   89.734635  86.9630515
## ESP   91.827945  89.0019219
## SWE   95.136979  89.0109980
## SUI   96.298275  91.4570807
## TPE   96.811858  93.7081595
## THA   96.831213  96.1005150
## TUR   99.256701  97.5889918
## USA  100.000000 100.0000000
```

c) Interpret the two principal components obtained in Part b. (Note that the first component is essentially a normalized unit vector and might measure the athletic excellence of a given nation. The second component might measure the relative strength of a nation at various running distances.)

d) Rank the nations based on their score on the first principal component. Does this ranking correspond with your inituitive notion of athletic excellence for the various countries?

```r
countries_CorWithPC = cbind.data.frame(countries = trackrcs[,1],
                                correlation = (CorWithPC[,1]),
                                unsigned_correlation = abs(CorWithPC[,1]))

countries_CorWithPC[,1][order(countries_CorWithPC[,3], decreasing = TRUE)]
```

```
##  [1] AUT  MEX  COK  RUS  KORS IND  JPN  CRC  BER  SWE  SIN  CHN  ITA  ISR
## [15] MRI  INA  ROM  PNG  TUR  DOM  ARG  SAM  ESP  COL  GBR  NED  NOR  CAN
## [29] CZE  AUS  GUA  IRL  FIN  POL  LUX  SUI  MYA  FRA  HUN  MAS  POR  GRE
## [43] USA  NZL  GER  CHI  BEL  KEN  TPE  DEN  PHI  BRA  KORN THA
## 54 Levels: ARG AUS AUT BEL BER BRA CAN CHI CHN COK COL CRC CZE DEN ... USA
```

```r
countries_CorWithPC[,1][order(countries_CorWithPC[,2], decreasing = TRUE)]
```

```
##  [1] AUT  COK  RUS  IND  BER  SWE  CHN  ITA  MRI  ROM  PNG  TUR  ESP  NED
## [15] CAN  CZE  POL  SUI  MYA  FRA  HUN  POR  GRE  USA  NZL  GER  CHI  BEL
## [29] BRA  KORN THA  PHI  DEN  TPE  KEN  MAS  LUX  FIN  IRL  GUA  AUS  NOR
## [43] GBR  COL  SAM  ARG  DOM  INA  ISR  SIN  CRC  JPN  KORS MEX
## 54 Levels: ARG AUS AUT BEL BER BRA CAN CHI CHN COK COL CRC CZE DEN ... USA
```

We get a different ranking when we use correlations directly (with their signs) from when we use unsigned correlations (the magnitude of the correlations). The ranking seems to be very innacurate about the athletic excellence for the various countries.

Question 2: Factor analysis

```r
factanal(trackrcs[,-1], factors = 3, covmat = S) # varimax is the default
```

```
##
## Call:
## factanal(x = trackrcs[, -1], factors = 3, covmat = S)
##
## Uniquenesses:
##    x100m    x200m    x400m    x800m   x1500m   x3000m marathon
##    0.106    0.005    0.133    0.047    0.005    0.041    0.225
##
## Loadings:
##          Factor1 Factor2 Factor3
## x100m    0.815   0.413   0.245
## x200m    0.886   0.410   0.203
## x400m    0.797   0.311   0.367
## x800m    0.512   0.617   0.556
## x1500m   0.449   0.849   0.270
## x3000m   0.361   0.866   0.280
## marathon 0.380   0.553   0.571
##
##                Factor1 Factor2 Factor3
## SS loadings      2.824   2.593   1.022
## Proportion Var   0.403   0.370   0.146
## Cumulative Var   0.403   0.774   0.920
##
## The degrees of freedom for the model is 3 and the fit was 0.2033
```

```r
factanal(trackrcs[,-1], factors = 3, covmat = R)
```

```
##
## Call:
## factanal(x = trackrcs[, -1], factors = 3, covmat = R)
##
## Uniquenesses:
##    x100m    x200m    x400m    x800m   x1500m   x3000m marathon
##    0.106    0.005    0.133    0.047    0.005    0.041    0.225
##
## Loadings:
##          Factor1 Factor2 Factor3
## x100m    0.815   0.413   0.245
## x200m    0.886   0.410   0.203
## x400m    0.797   0.311   0.367
## x800m    0.512   0.617   0.556
## x1500m   0.449   0.849   0.270
## x3000m   0.361   0.866   0.280
## marathon 0.380   0.553   0.571
##
##                Factor1 Factor2 Factor3
## SS loadings      2.824   2.593   1.022
```

```
## Proportion Var   0.403   0.370   0.146
## Cumulative Var   0.403   0.774   0.920
##
## The degrees of freedom for the model is 3 and the fit was 0.2033
```

```r
psych::principal(cov2cor(S), nfactors=3, rotate="varimax")
```

```
## Principal Components Analysis
## Call: psych::principal(r = cov2cor(S), nfactors = 3, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##           RC2  RC1  RC3   h2    u2 com
## x100m    0.85 0.41 0.23 0.94 0.061 1.6
## x200m    0.86 0.40 0.25 0.96 0.037 1.6
## x400m    0.86 0.26 0.36 0.93 0.065 1.5
## x800m    0.54 0.59 0.54 0.93 0.072 3.0
## x1500m   0.44 0.82 0.34 0.99 0.010 1.9
## x3000m   0.35 0.85 0.37 0.98 0.020 1.7
## marathon 0.33 0.44 0.82 0.98 0.019 1.9
##
##                        RC2  RC1  RC3
## SS loadings           2.92 2.33 1.47
## Proportion Var        0.42 0.33 0.21
## Cumulative Var        0.42 0.75 0.96
## Proportion Explained  0.43 0.35 0.22
## Cumulative Proportion 0.43 0.78 1.00
##
## Mean item complexity =  1.9
## Test of the hypothesis that 3 components are sufficient.
##
## The root mean square of the residuals (RMSR) is  0.02
##
## Fit based upon off diagonal values = 1
```

```r
psych::fa(cov2cor(S), nfactors=3, rotate="varimax")
```

```
## Factor Analysis using method =  minres
## Call: psych::fa(r = cov2cor(S), nfactors = 3, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##           MR2  MR3  MR1   h2     u2 com
## x100m    0.83 0.41 0.23 0.90 0.0993 1.6
## x200m    0.88 0.40 0.21 0.98 0.0160 1.5
## x400m    0.80 0.31 0.35 0.87 0.1338 1.7
## x800m    0.53 0.60 0.54 0.94 0.0622 3.0
## x1500m   0.46 0.85 0.26 1.00 0.0018 1.8
## x3000m   0.38 0.85 0.30 0.95 0.0457 1.7
## marathon 0.37 0.56 0.59 0.80 0.2002 2.7
##
##                        MR2  MR3  MR1
## SS loadings           2.88 2.54 1.03
## Proportion Var        0.41 0.36 0.15
## Cumulative Var        0.41 0.77 0.92
## Proportion Explained  0.45 0.39 0.16
## Cumulative Proportion 0.45 0.84 1.00
```

```
##
## Mean item complexity =  2
## Test of the hypothesis that 3 factors are sufficient.
##
## The degrees of freedom for the null model are  21  and the objective function was  11.62
## The degrees of freedom for the model are 3  and the objective function was  0.23
##
## The root mean square of the residuals (RMSR) is  0
## The df corrected root mean square of the residuals is  0.01
##
## Fit based upon off diagonal values = 1
```