# Lab 3 - Multivariate Statistical Methods

*Gustav Sternelöv*

*30 november 2015*

## Assignment 1

### a)

The sample correlation matrix R:

```
##                100       200       400       800      1500      3000
## 100      1.0000000 0.9410886 0.8707802 0.8091758 0.7815510 0.7278784
## 200      0.9410886 1.0000000 0.9088096 0.8198258 0.8013282 0.7318546
## 400      0.8707802 0.9088096 1.0000000 0.8057904 0.7197996 0.6737991
## 800      0.8091758 0.8198258 0.8057904 1.0000000 0.9050509 0.8665732
## 1500     0.7815510 0.8013282 0.7197996 0.9050509 1.0000000 0.9733801
## 3000     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
## 100      0.6689597
## 200      0.6799537
## 400      0.6769384
## 800      0.8539900
## 1500     0.7905565
## 3000     0.7987302
## Marathon 1.0000000
```

The eigenvalues:

```
##    values1    values2    values3    values4    values5    values6
## 5.80762446 0.62869342 0.27933457 0.12455472 0.09097174 0.05451882
##    values7
## 0.01430226
```

The eigenvectors:

```
##    vectors.1  vectors.2  vectors.3   vectors.4   vectors.5   vectors.6
## 1 -0.3777657 -0.4071756 -0.1405803  0.58706293 -0.16706891  0.53969730
## 2 -0.3832103 -0.4136291 -0.1007833  0.19407501  0.09350016 -0.74493139
## 3 -0.3680361 -0.4593531  0.2370255 -0.64543118  0.32727328  0.24009405
## 4 -0.3947810  0.1612459  0.1475424 -0.29520804 -0.81905467 -0.01650651
## 5 -0.3892610  0.3090877 -0.4219855 -0.06669044  0.02613100 -0.18898771
## 6 -0.3760945  0.4231899 -0.4060627 -0.08015699  0.35169796  0.24049968
## 7 -0.3552031  0.3892153  0.7410610  0.32107640  0.24700821 -0.04826992
##     vectors.7
## 1  0.08893934
## 2 -0.26565662
## 3  0.12660435
## 4 -0.19521315
```

```
## 5  0.73076817
## 6 -0.57150644
## 7  0.08208401
```

## b)

The first two principal components for the standardized variables are determined. The correlations between the standardized variables are shown in the table below.

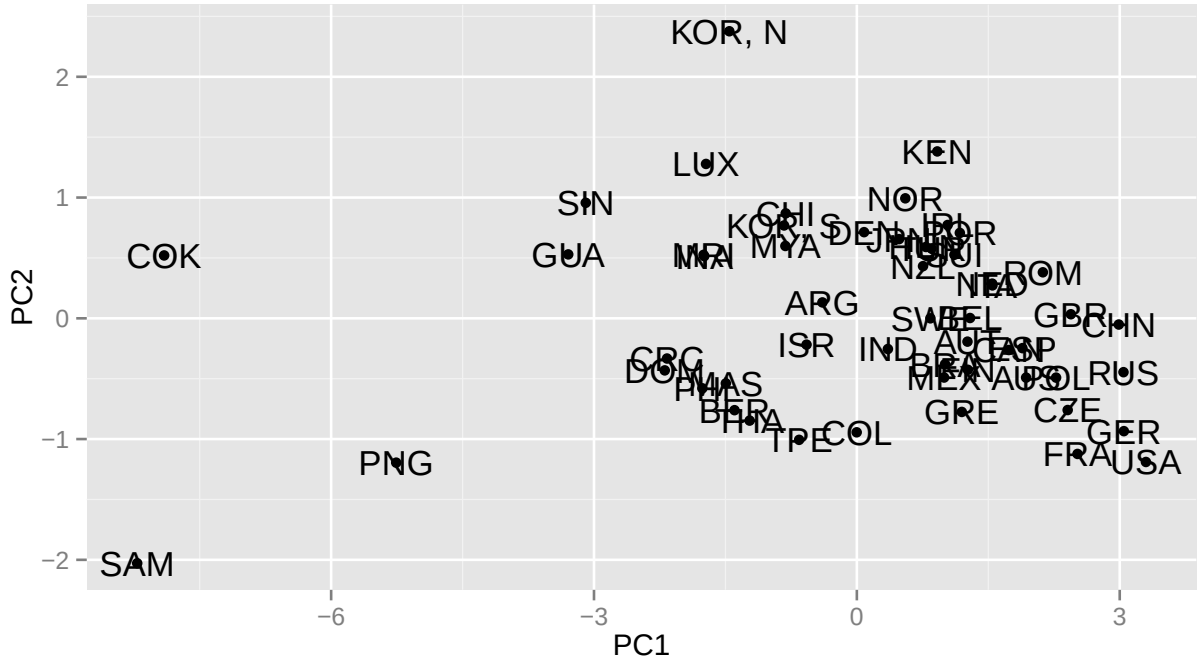
```
##                      X100       X200       X400       X800      X1500
## prComp.x...1. -0.9103780 -0.9234990 -0.8869307 -0.9513832 -0.9380805
## prComp.x...2.  0.3228503  0.3279673  0.3642220 -0.1278522 -0.2450762
##                     X3000   Marathon
## prComp.x...1. -0.9063506 -0.8560043
## prComp.x...2. -0.3355481 -0.3086096
```

The first component explains 82.966 % of the sample variance and the second component explains 8.981 % of the sample variance. The cumulative % of the sample variance explained by the two components then is 91.947 %.

## c)

The principal components are shown in a table below. These components are interpretet by looking at the values in the table, but also by a plot where the scores for each country for the respective component are plotted against each other.

```
##                 PC1        PC2
## X100     -0.3777657  0.4071756
## X200     -0.3832103  0.4136291
## X400     -0.3680361  0.4593531
## X800     -0.3947810 -0.1612459
## X1500    -0.3892610 -0.3090877
## X3000    -0.3760945 -0.4231899
## Marathon -0.3552031 -0.3892153
```




The first principal component seem to measure the excellence of a nation since countries like USA, Germany and Russia, well-known top nations, are among those with the highest scores and less well-known countries in the field of athletics like Samoa and Cook Islands have the lowest scores.
Regarding the second component the values points out the relative strength for a country in shorter versus longer running distances. If a country is better at shorter distances it obtains a lower value and if it is better at longer distances the value becomes higher. If the difference between how well the country performs at shorter and longer distances is small, the value for the second component also will become small.

### d)

The first column in the table below gives the 6 highest scores for PC1 and the second column the countries who have the highest scores. The third column gives the 6 lowest scores in descending order and the fourth the countries who have the lowest scores.

```
##        PC1 country     PC1.1 country.1
## 1 3.299149     USA -2.192410       DOM
## 2 3.047517     GER -3.093920       SIN
## 3 3.042948     RUS -3.294124       GUA
## 4 2.989467     CHN -5.257450       PNG
## 5 2.518346     FRA -7.906227       COK
## 6 2.442706     GBR -8.213415       SAM
```

When athletic excellence is compared between countries, the ones who have the highest scores are all well-known top nations. The reverse can be said about the countries with the lowest scores since they are fairly small countries who cannot compete with USA, Germany and so on when athletic excellence is compared.

## Assignment 2