

# Multivariate Statistical Methods - Lab 1

*Gustav Sternelöv*

*Wednesday, November 11, 2015*

## Assignment 1

a)

Mean value, standard deviation, variance, min value, max value and median for the seven variables are displayed in the tables below.

### Mean

```
##          100          200          400          800          1500          3000
## 11.357778 23.118519 51.989074 2.022407 4.189444 9.080741
## Marathon
## 153.619259
```

### Standard\_deviation

```
##          100          200          400          800          1500          3000
## 0.39410116 0.92902547 2.59720188 0.08687304 0.27236502 0.81532689
## Marathon
## 16.43989508
```

### Variance

```
##          100          200          400          800          1500
## 1.553157e-01 8.630883e-01 6.745458e+00 7.546925e-03 7.418270e-02
##          3000 Marathon
## 6.647579e-01 2.702702e+02
```

### Max

```
##          100          200          400          800          1500          3000 Marathon
##          12.52          25.91          61.65          2.29          5.42          13.12          221.14
```

### Min

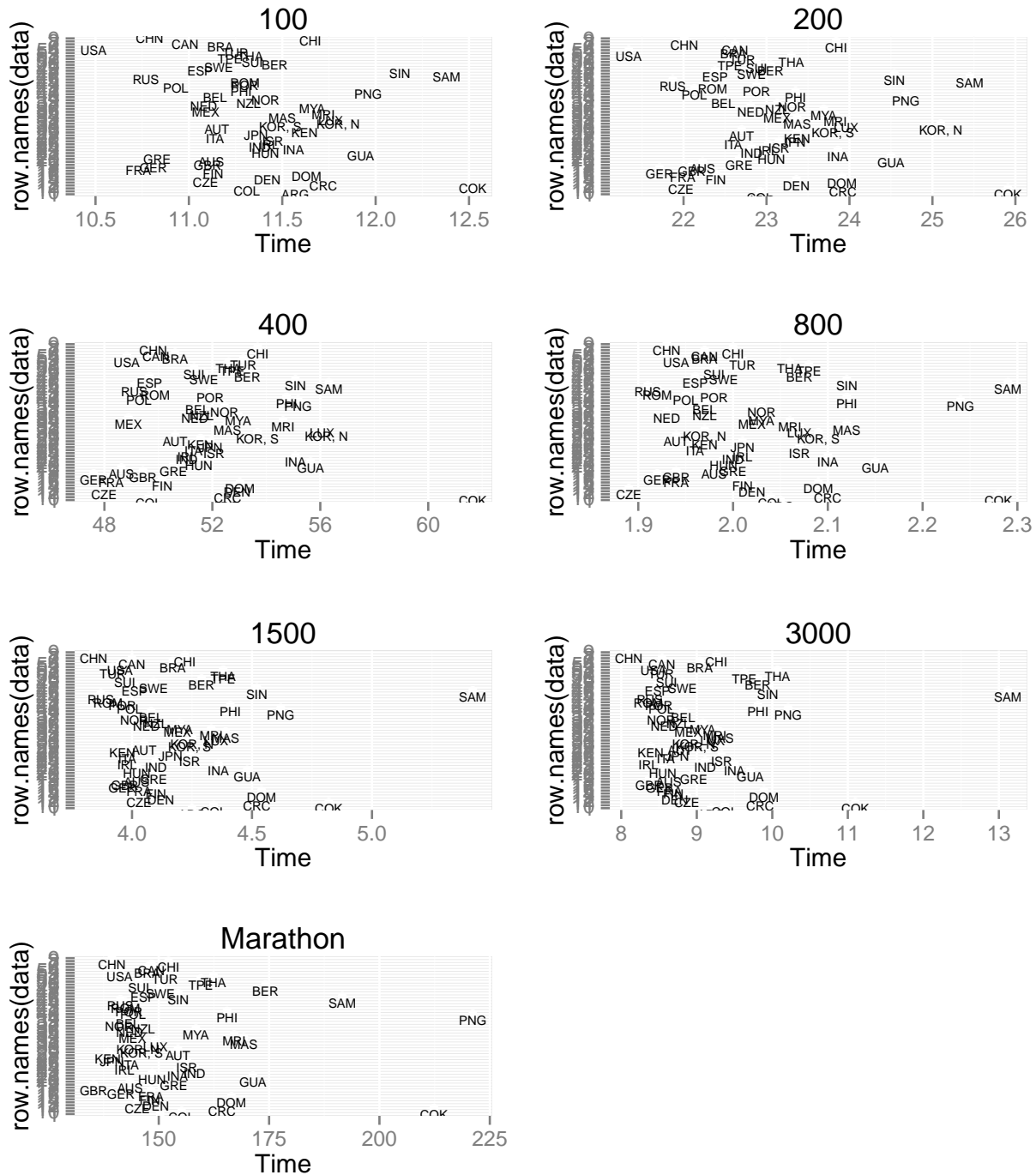
```
##          100          200          400          800          1500          3000 Marathon
##          10.49          21.34          47.60          1.89          3.84          8.10          135.25
```

### Median

```
##          100          200          400          800          1500          3000 Marathon
##          11.325          22.980          51.645          2.005          4.100          8.845          148.430
```

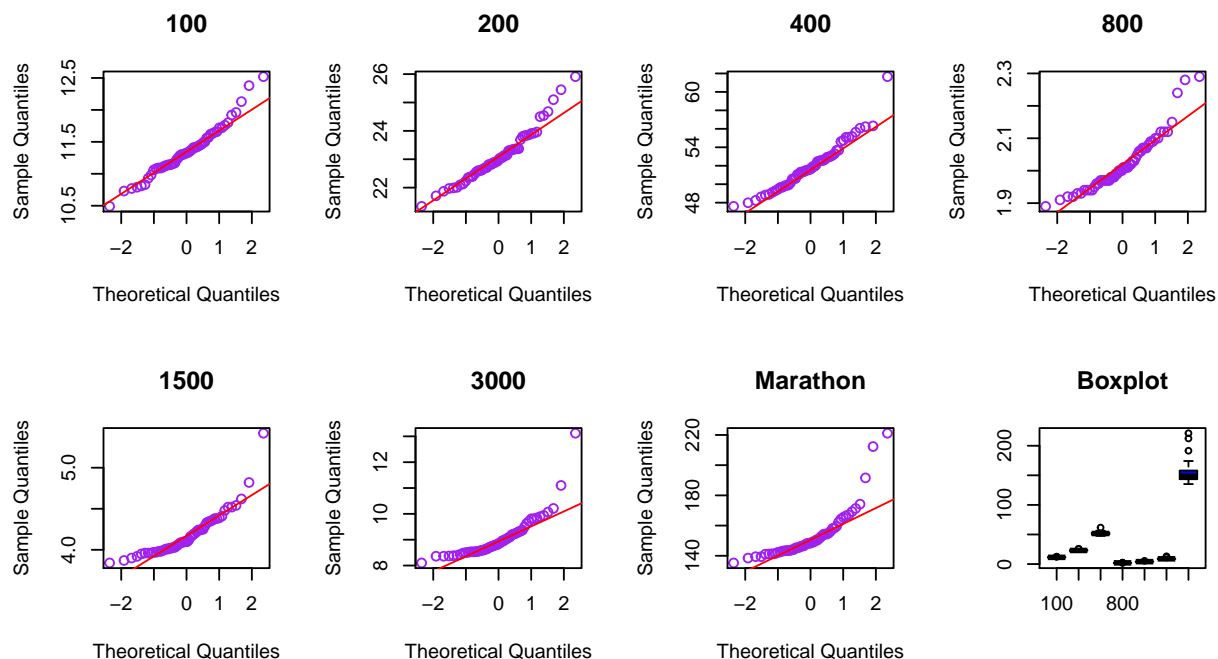
b)

Dot plot for each variable to investigate if any extreme values can be found.



An interpretation of the dot plots is that extreme values can be seen in most of the graphs. The most extreme countries seem to be Samoa and Cook Islands who has the most extreme values for several variables.

Examining if the variables seem to be normally distributed by looking at the following plots. A box plot is also included to illustrate how much the values differ for the respective distance.



It is concluded that the observed values for all the variables are lying quite well along the red line. This indicates that they can be thought of as being approximately normally distributed. Regarding the box plot it can easily be seen that the values differ far more for marathon than for instance 100m.

## Assignment 2

a)

Covariance and correlation matrices:

```
##           100           200           400           800           1500
## 100      0.15531572  0.3445608  0.8912960  0.027703564  0.08389119
## 200      0.34456080  0.8630883  2.1928363  0.066165898  0.20276331
## 400      0.89129602  2.1928363  6.7454576  0.181807932  0.50917683
## 800      0.02770356  0.0661659  0.1818079  0.007546925  0.02141457
## 1500     0.08389119  0.2027633  0.5091768  0.021414570  0.07418270
## 3000     0.23388281  0.5543502  1.4268158  0.061379315  0.21615514
## Marathon 4.33417757 10.3849876 28.9037314 1.219654647 3.53983732
##           3000    Marathon
## 100      0.23388281  4.334178
## 200      0.55435017 10.384988
## 400      1.42681579 28.903731
## 800      0.06137932  1.219655
## 1500     0.21615514  3.539837
## 3000     0.66475793 10.706091
## Marathon 10.70609113 270.270150

##           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
```

```

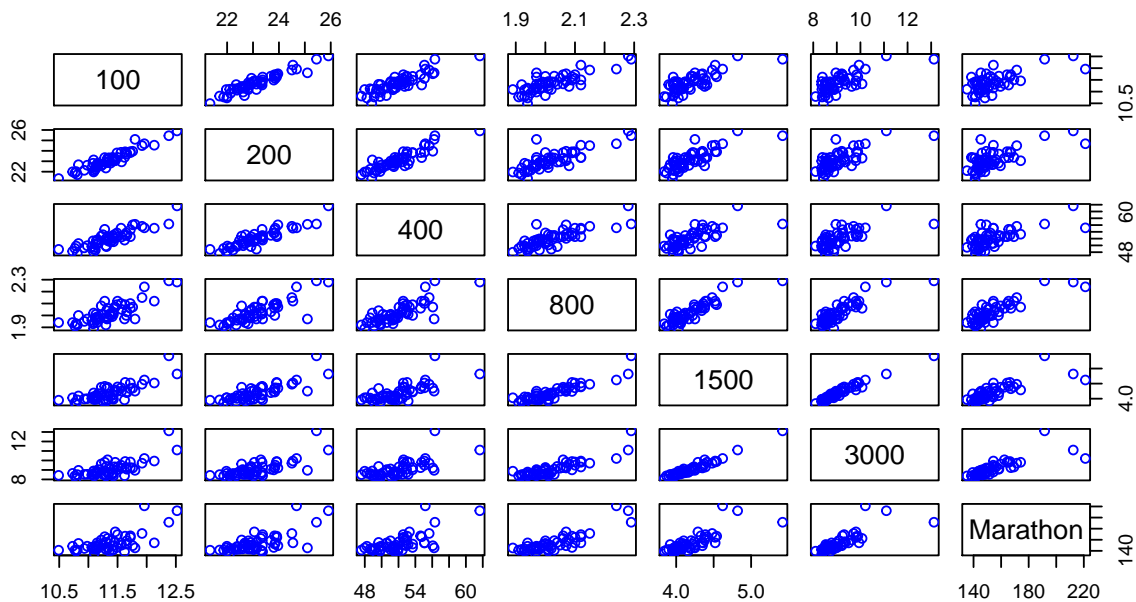
## 400      0.8707802 0.9088096 1.0000000 0.8057904 0.7197996 0.6737991
## 800      0.8091758 0.8198258 0.8057904 1.0000000 0.9050509 0.8665732
## 1500     0.7815510 0.8013282 0.7197996 0.9050509 1.0000000 0.9733801
## 3000     0.7278784 0.7318546 0.6737991 0.8665732 0.9733801 1.0000000
## Marathon 0.6689597 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 correlation is stronger for more similar distances. For example the 100m have the strongest correlation to 200m, second strongest to 400m and so on. Regarding the variance it seem to be rising for longer distances.

b)

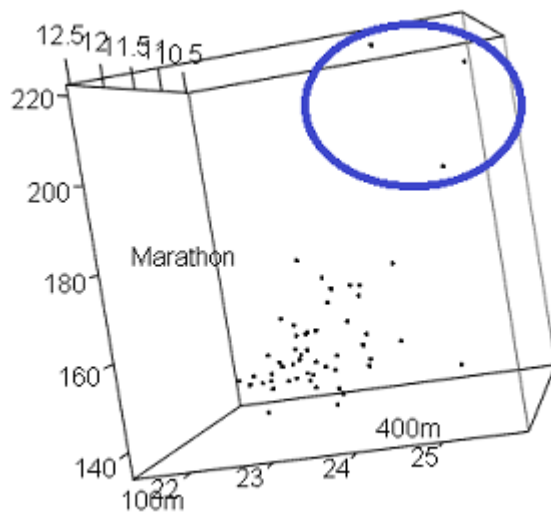
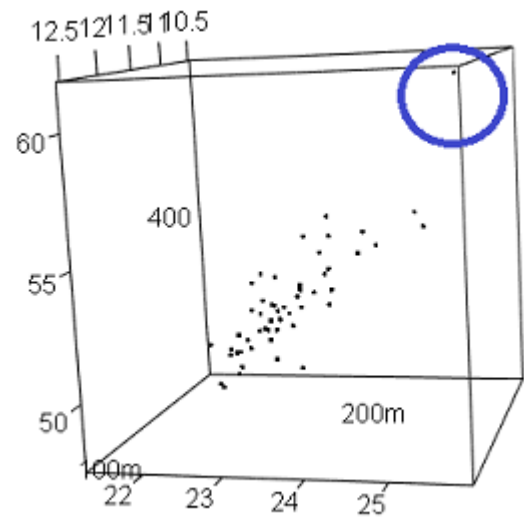
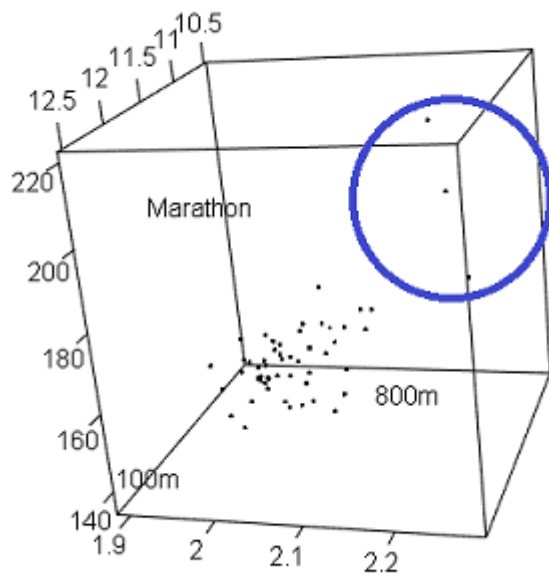
A scatterplot matrix



The longer distance, the more apparent extreme values. When looking at for example Marathon versus the other variables there are three extreme values and for 3000m there are at least one extreme value.

c)

Another way of visually investigating which countries that are the most extreme is by looking at three-dimensional scatterplots.



### Assignment 3

a)

By looking at the matrix scatterplot, 3D plots and dot plots the most extreme countries seem to be Samoa, Cook Islands, Papua New Guinea and perhaps also North Korea.

b)

The five most extreme countries according to the squared Euclidean distance:

```
## data...1. sqDiagdistance
## 40 PNG 4573.5413
```

```
## 11      COK      3553.9691
## 46      SAM      1484.1570
## 5       BER      425.0219
## 19      GBR      345.6425
```

c)

When the Euclidean distances are computed for standardized data the following countries are the most extreme:

```
##      data...1. newSqDistance
## 46      SAM      75.58280
## 11      COK      64.60116
## 40      PNG      34.22891
## 54      USA      12.87689
## 47      SIN      11.44486
```

d)

A third way to measure the distance is by computing Mahalanobis distances between the countries. The five most extreme values according to this measure are given by the following table.

```
##      data...1. MahanabisD
## 46      SAM      35.01406
## 40      PNG      30.50725
## 31      KOR, N    26.16714
## 11      COK      19.83400
## 35      MEX      14.23093
```

e)

The countries Samoa, Cook Islands and Papua New Guinea are in the top five for all of the computed measures, but as mentioned in the exercise the results differ between the respective distance measures. For example Sweden are ranked at place 48 in for the squared Euclidean distance, 50 for the Euclidean distance on standardized data and at place 54 for Mahalanobis distances.