

Multivariate Data Analysis

Quiz



MAS6011/MAS465
2011/2012



[Partially seen] Quiz

Preliminary background information:

Data are available on national track records for various distances held by women from 55 different countries (as they stood at the start of the 1984 Los Angeles Olympics). **The distances are, in metres, 100, 200, 400, 800, 1500, 3000 and marathon.**

A small sample of the data is given below:

```
> womentrackrecords[1:4,]; womentrackrecords[53:55,]
      X100m X200m X400m X800m X1500m X3000m marathon
argentin 11.61 22.94 54.50 2.15 4.43 9.79 178.52
australi 11.20 22.35 51.08 1.98 4.13 9.08 152.37
austria  11.43 23.09 50.62 1.99 4.22 9.34 159.37
be^.....
..
usa      10.79 21.83 50.62 1.96 3.95 8.50 142.72
ussr    11.06 22.19 49.19 1.89 3.87 8.45 151.22
wsamoa  12.74 25.85 58.73 2.33 5.81 13.04 306.00
```

Q1: What units are the data in?

```
> options(digits=3)
> wtr<-womentrackrecords
> wtrcov.pc<-princomp(wtr)
> summary(wtrcov.pc)
Importance of components:
  Comp.1 Comp.2 Comp.3 Comp.4
St dev  30.231 1.994 0.559 0.336
Prop Var 0.995 0.004 0.000 0.000
Cum Prop 0.995 0.9995 0.9998 0.9999
```

Q2: Is the PCA better with covariance or correlations?

Q3: Why?

```
> wtrcorr.pc<-princomp(wtr,cor=T)
> summary(wtrcorr.pc)
Importance of components:
  Comp.1 Comp.2 Comp.3 Cor
St dev  2.409 0.808 0.547 0.354
Prop Var 0.829 0.093 0.042 0.017
Cum Prop 0.829 0.922 0.965 0.983
```

Q4: What other information would be useful?

```
> options(digits=2)
> diag(var(wtr))
      X100m X200m X400m X800m X1500m X3000m marathon
0.204 1.234 7.173 0.012 0.111 0.680 925.957
> sqrt(diag(var(wtr)))
      X100m X200m X400m X800m X1500m X3000m marathon
0.45 1.11 2.68 0.11 0.33 0.82 30.43
```

Q5: covariance or correlation?
Q6: In what units are the variances?

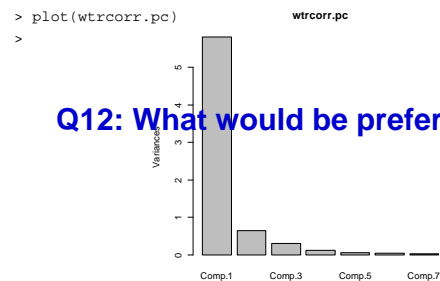
Q7: would it be better to convert all measurements to the same units?

Q8: Why?

```
> summary(wtrcorr.pc)
Importance of components:
  Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7
St dev  2.41 0.808 0.548 0.354 0.2320 0.1976 0.1498
Prop Var 0.83 0.093 0.043 0.018 0.0077 0.0056 0.0032
Cum Prop 0.83 0.923 0.966 0.984 0.9912 0.9968 1.0000
```

Q9: how many components ?

Q10: what would be useful to answer this ?



Q12: What would be preferable?

Q11: useful?



