

Multivariate Data Analysis

Exercises 3

Released 21 November 2011, work submitted by ~Tuesday 6 December 2009 will be marked and returned. Work submitted after solutions are made available will not be marked

1)

- i) Measurements of cranial length x_{11} and cranial breadth x_{12} on 35 female frogs gave $\bar{x}'_1 = (22.860, 24.397)$ and

$$S_1 = \begin{pmatrix} 17.683 & 20.290 \\ * & 24.407 \end{pmatrix}. \text{ Test the hypothesis that } \mu_{11} = \mu_{12}.$$

- ii) Similar measurements on 14 male frogs gave

$$\bar{x}'_2 = (21.821, 22.843) \text{ and } S_2 = \begin{pmatrix} 18.479 & 19.095 \\ * & 20.756 \end{pmatrix}.$$

Calculate the pooled variance matrix for male & female frogs and test the hypothesis that female & male frogs come from populations with equal mean vectors.

- 2) Using your favourite computer package, access the British Museum Mummy Pots data (see task sheet for week 4) and calculate the two shape variables 'taper' and 'point'.

- i) Do the two batches of pots differ in overall shape as reflected by the calculated shape measures 'taper' and 'point'?
- ii) Do the two batches of pots differ in overall size?
- iii) Without doing any calculations,
- a) would your answer to (ii) be different in any respect if you used the scores on the three PCs calculated from the size variables?
- b) would it make any difference were you to calculate the PCs using the correlation matrix instead of the covariance matrix?

[Suggestion: Read §8.7.4 of the lecture notes]



3) * x_1, \dots, x_n are independent measurements of $N_p(\mu, \sigma^2 I_p)$

- i) Shew that the maximum likelihood estimate of μ , subject to $\mu' \mu = r_0^2$ (a known constant) is the same whether σ is known or unknown.
- ii) Find the maximum likelihood estimate of σ when neither μ nor σ are known.
- iii) Hence, in the case when $\sigma = \sigma_0$ (a known constant) construct the likelihood ratio test of $H_0 : \mu' \mu = r_0^2$ vs $H_A : \mu' \mu \neq r_0^2$ based on n independent observations of $N_p(\mu, \sigma_0^2 I_p)$.
- iv) In an experiment to test the range of a new ground-to-air missile thirty-nine test firings at a tethered balloon were performed and the three dimensional coordinates of the point of ignition of the missile's warhead measured. These gave a mean result of $(0.76, 0.69, 0.66)'$ relative to the site expressed in terms of the target distance. Presuming that individual measurements are independently normally distributed with unit variance, are the data consistent with the theory that the range of the missile was set correctly?







[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[1:10,]; womentrackrecords[46:55,]
      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
belgium  11.41 23.04 52.00  2.00  4.14  8.88  157.85
bermuda  11.46 23.05 53.30  2.16  4.58  9.81  169.98
brazil   11.31 23.17 52.80  2.10  4.49  9.77  168.75
burma    12.14 24.47 55.00  2.18  4.45  9.51  191.02
canada   11.00 22.25 50.06  2.00  4.06  8.81  149.45
chile    12.00 24.52 54.90  2.05  4.23  9.37  171.38
china    11.95 24.41 54.97  2.08  4.33  9.31  168.48
.....
.....
singapor 12.30 25.00 55.08  2.12  4.52  9.94  182.77
spain    11.80 23.98 53.59  2.05  4.14  9.02  162.60
sweden   11.16 22.82 51.79  2.02  4.12  8.84  154.48
switzerl 11.45 23.31 53.11  2.02  4.07  8.77  153.42
taipei   11.22 22.62 52.50  2.10  4.38  9.63  177.87
thailand 11.75 24.46 55.80  2.20  4.72 10.28  168.45
turkey   11.98 24.44 56.45  2.15  4.37  9.38  201.08
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
```

The complete list of countries as held in the data file is

```
rownames(womentrackrecords)
 [1] "argentin" "australi" "austria"  "belgium"  "bermuda"  "brazil"
 [7] "burma"    "canada"   "chile"    "china"    "columbia" "cookis"
[13] "costa"    "czech"    "denmark"  "domrep"   "finland"  "france"
[19] "gdr"      "frg"      "gbni"     "greece"   "guatemal" "hungary"
[25] "india"    "indonesi" "ireland"  "israel"   "italy"    "japan"
[31] "kenya"    "korea"    "dprkorea" "luxembou" "malaysia" "mauritiu"
[37] "mexico"   "netherla" "nz"       "norway"   "png"      "philippi"
[43] "poland"   "portugal" "rumania"  "singapor" "spain"    "sweden"
[49] "switzerl" "taipei"   "thailand" "turkey"   "usa"      "ussr"
[55] "wsamoa"
```

(NB `gdr` the [former] East Germany. `frg` is the [former] West Germany, `gbni` is the UK and `png` is Papua New Guinea.



Below are some preliminary multivariate data analyses:

```
> print(wtrcov.pc$loadings,cutoff=0.001); print(wtrcorr.pc$loadings)
```

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
X100m	-0.010	-0.120	0.326	-0.150	0.925	0.002	-0.017
X200m	-0.025	-0.315	0.880	-0.014	-0.354	-0.025	-0.012
X400m	-0.062	-0.934	-0.328	0.122	0.013	0.022	0.025
X800m	-0.003	-0.026	-0.037	-0.049	-0.015	-0.262	-0.963
X1500m	-0.010	-0.039	-0.055	-0.340	-0.034	-0.899	0.265
X3000m	-0.024	-0.082	-0.088	-0.919	-0.130	0.349	-0.041
marathon	-0.997	0.070	-0.002	0.020	0.002		

Loadings:

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
X100m	-0.368	0.490	0.286	-0.319	-0.231	0.620	
X200m	-0.365	0.537	0.230			-0.711	-0.109
X400m	-0.382	0.247	-0.515	0.347	0.572	0.191	0.208
X800m	-0.385	-0.155	-0.585		-0.620		-0.315
X1500m	-0.389	-0.360		-0.430		-0.231	0.693
X3000m	-0.389	-0.348	0.153	-0.363	0.463		-0.598
marathon	-0.367	-0.369	0.484	0.672	-0.131	0.142	

Various questions will follow later.

