STATS 4M03 / 6M03

ASSIGNMENT 1

Assignments submitted up to 24 hours late will incur a 30% penalty.
If no assignment is received within 24 hours, the grade will be zero.

Answer all questions, in order, starting each question on a new page.
Questions do not necessarily carry the same amount of marks.
For each question, clearly state any results that you use.

1. Consider the following output from R, where I carry out a principal components anal-ysis based on the correlation matrix associated with ability.cov. For a description of the data, type ?ability.cov in R.

```
> round(cov2cor(ability.cov$cov),3)
          general picture blocks  maze reading vocab
general     1.000   0.466  0.552 0.340   0.576 0.514
picture     0.466   1.000  0.572 0.193   0.263 0.239
blocks      0.552   0.572  1.000 0.445   0.354 0.356
maze        0.340   0.193  0.445 1.000   0.184 0.219
reading     0.576   0.263  0.354 0.184   1.000 0.791
vocab       0.514   0.239  0.356 0.219   0.791 1.000
> #Eigenvalues
> round(eigen(cov2cor(ability.cov$cov))$values,3) [1]
3.077 1.140 0.817 0.411 0.355 0.200
> #Eigenvectors
> round(eigen(cov2cor(ability.cov$cov))$vectors,3)
       [,1]   [,2]   [,3]   [,4]   [,5]   [,6]
[1,] -0.471  0.002  0.072  0.863  0.037 -0.164
[2,] -0.358  0.408  0.593 -0.268  0.531  0.002
[3,] -0.434  0.404  0.064 -0.201 -0.775  0.051
[4,] -0.288  0.404 -0.794 -0.097  0.334  0.052
[5,] -0.440 -0.507 -0.015 -0.100  0.056  0.733
[6,] -0.430 -0.501 -0.090 -0.352  0.021 -0.657
```

(a) What is the sum of the variances of the principal components and what is the sum of the diagonal elements of the correlation matrix?

(b) For each principal component, calculate the proportion of the total variation in the data that it explains.

(c) Interpret this principal components analysis. Do this in two steps:

i. Decide how many principal components to use in your analysis of these data.

ii. Then give an interpretation, in terms of the data, for each of the components you use.

2. Consider the following output from R, where a factor analysis is carried out on the ais data.

```
> data(ais,package="alr3")
> factanal(ais[,2:12],2)

Call:
factanal(x = ais[, 2:12], factors = 2)

Uniquenesses:
    Ht Wt LBM RCC WCC Hc Hg Ferr BMI SSF Bfat 0.353 0.005 0.005
0.591 0.964 0.533 0.509 0.893 0.253 0.044 0.005

Loadings:
      Factor1 Factor2
Ht     0.740   0.316
Wt     0.985   0.164
LBM    0.859   0.509
RCC    0.318   0.555
WCC    0.171
Hc     0.331   0.598
Hg     0.362   0.601
Ferr   0.240   0.223
BMI    0.863
SSF    0.310  -0.928
Bfat   0.163  -0.984
```

	Factor1	Factor2
SS loadings	3.550	3.300
Proportion Var	0.323	0.300
Cumulative Var	0.323	0.623

Test of the hypothesis that 2 factors are sufficient.
The chi square statistic is 1368.85 on 34 degrees of freedom.
The p-value is 6.51e-266

(a) Interpret these factor loadings in terms of the data.

(b) Interpret the Uniqueness values for this analysis.

(c) Repeat this factor analysis with 3 factors.

   i. Interpret the factor loadings.

   ii. Compare your interpretation with that from the 2 factor case.

   iii. How has Uniqueness changed with the addition of an extra factor?

   iv. Which solution do you prefer (two or three factors)? Why?

(d) Try to repeat this factor analysis with 4 factors. What happens? Why?