STAT 530 - Fall 2003 - Homework 4

Due: Friday, October 17th

The following problems use the data set:
`http://www.stat.sc.edu/~habing/courses/data/a1291space.txt`
(Note that the set has no column titles so `head=FALSE`, `dim(a1291)` should be 6000 by 24.)

The data contains the results of 6,000 examinees who took the December 1991 analytical reasoning portion of the Law School Admissions test. There were 24 test questions based around for paragraphs; questions 1-7, questions 8-13, questions 14-19, and questions 20-24. Each column is a separate test question (variable) and each row is a separate examinee (observation). A one in position (i, j) means examinee i got question j correct. A zero means that it was incorrect.

1) Use the description of the exam and the three common methods we discussed to determine how many factors seem to underly this exam. Which of the three methods doesn't seem helpful in this case?

2) Conduct a factor analysis of this dataset using the number of factors you chose in part 1, using a varimax rotation.

a) Report the factor loadings, indicating which seem to have practical significance, and those that seem to have borderline practical significance. Does the pattern in the factor loadings seem expected?

b) Check that the covariances of the errors seem to generally have the correct pattern. Does the fit seem particularly good?

c) Would you describe the size of the communalities as being large or small? Can you relate this to why one of the methods in question 1 didn't seem to work very well?

3) Look at the form of the individual observations in this data set. Is there anything about data of this sort that makes you question the validity of using either principal components or factor analysis on it? (The answer to this question has nothing to do with what you found in 1 or 2... you can find the answer simply by comparing this data set to the data we looked at in the other examples and by looking at the formulas we use for PCA and FA.)