Q1. 123 people suffering from extreme pain surveyed for on nine statements about pain. Each statement was scored on a scale from 1 to 6 ranging from agreement to disagreement. The nine pain statements were as follows:

1. Whether or not I am in pain in the future depends on the skill of the doctors.
2. Whenever I am in pain, it is usually because of something I have done or not done.
3. Whether or not I am in pain depends on what doctors do for me.
4. I cannot get any help for my pain unless I go to seek medical advice.
5. When I am in pain I know that it is because I have not been taking proper exercise or eating the right food.
6. People's pain results from their own carelessness.
7. I am directly responsible for my pain.
8. Relief from pain is chiefly controlled by the doctors.
9. People who are never in pain are just plain lucky.

The correlation matrix between ratings on nine statements is given by

$$R = \begin{pmatrix} 1.00 & & & & & & & & \\ -0.04 & 1.00 & & & & & & & \\ 0.61 & -0.07 & 1.00 & & & & & & \\ 0.45 & -0.12 & 0.59 & 1.00 & & & & & \\ 0.03 & 0.49 & 0.03 & -0.08 & 1.00 & & & & \\ -0.29 & 0.43 & -0.13 & -0.21 & 0.47 & 1.00 & & & \\ -0.30 & 0.30 & -0.24 & -0.19 & 0.41 & 0.63 & 1.00 & & \\ 0.45 & -0.31 & 0.59 & 0.63 & -0.14 & -0.13 & -0.26 & 1.00 & \\ 0.30 & -0.17 & 0.32 & 0.37 & -0.24 & -0.15 & -0.29 & 0.40 & 1.00 \end{pmatrix}$$

(a) Suggest an appropriate number of factors to estimate the factor loadings using the iterative principal factor method

Using the iterative principal factor method with the chosen number of factors in (a), answer the following questions.

(b) Estimate the communalities.

(c) Estimate the matrix of the specific variances, Ψ .

(d) Estimate the correlation between the 7th statement and the first factor, F_1.

(e) Calculate the residual matrix $R - \hat{\Lambda}\hat{\Lambda}' - \hat{\Psi}$. Is this residual matrix closer to a zero matrix?

Repeat (b) ~ (e) using the rotated estimated loadings using Varimax rotation and compare the results with the ones obtained without rotation.

Q2. A two-factor maximum likelihood analysis solution for a particular set of data yields the estimated loadings shown in the following table.

	F_1	F_2
x_1	0.789	-0.403
x_2	0.834	-0.234
x_3	0.740	-0.134
x_4	0.586	-0.185
x_5	0.676	-0.248
x_6	0.654	0.440
x_7	0.641	0.534
x_8	0.629	0.651
x_9	0.564	0.354
x_{10}	0.808	0.714

(a) By plotting the derived loadings, find the angle corresponding to the minimum rotation of the axes needed to remove those that are negative.
(b) Group variables as two factors based on the rotated loadings in (a).

Q3. Measurements on the six air-pollution variables were recorded at noon in the Los Angeles area on different days. The data (named POLLUTION.DAT) is attached with this file.

X_1	Wind
X_2	Solar radiation
X_3	CO
X_4	NO
X_5	NO_2

X_6	O_3
X_7	HC

(a) Are there some variables which can be recoded? Discuss why.

(b) Obtain the two-factor solution using the maximum likelihood method and interpret factors. Estimate communalities, specific variances, and $\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$.

(c) Obtain the three-factor solution using the maximum likelihood method and interpret factors. Estimate communalities, specific variances, and $\hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$.

(d) Conduct the test for choosing an appropriate number of factors. What number of the factors would you recommend?

(e) Obtain factor scores based on the three-factor solution. (Hint: Factor scores can be obtained by the command, fac.ml$scores, in R, where fac.ml is the output object of the factor analysis.)

(f) Calculate the means and the variances of the factor scores. Compare these values with the theoretical values.

(g) Obtain factor scores based on the simple sum of related manifest variables for each factor of the two-factor solution. Compare these scores with the factor scores obtained in (e).