KEY

Assignment 2 (Due at 1 p.m. on Tuesday, April 18, 2006)

This is an open-book, open-note assignment. Collaboration is not allowed. Late assignments will not be accepted. Please show all your work, as no credit will be given for unsupported answers. Please try to fit your answers into the spaces provided.

Good luck!

Consider the data given in Table 1.7, p. 43, in our text. (See also Exercise 8.13, p. 471.) It was decided to perform a factor analysis to determine if the radiotherapy symptoms could be combined into a smaller number of reaction indices.

1. Is it more appropriate to factor analyze the correlation matrix R or the covariance matrix S for these data? Give a reason for your answer.

6/ The correlation matrix R, since the dependent variables are measured in different units.

2. (a) The eigenvalues of the matrix you specified in Question 1 above, along with the proportion variance explained are given below:

	1	2	3	4	5	6
Eigenvalue	2.83	1.06	0.80	0.66	0.40	0.25
Proportion	.47	.18	.13	.11	.07	.04
Cum Prop	.47	.65	.78	STOP		

Using these results, determine the appropriate number of factors to extract using each of the following methods:

(i) KG Criterion

8/ $f = 2$ since the 1st 2 eigenvalues are ≥ 1

(ii) % Variance Explained

8/ $f = 3$ since 3 factors are required to account for at least 75% of the total variance

(b) The average squared partial correlation was obtained after f principal components were extracted for $f = 1, 2, 3, 4, 5$. The results are summarized in the following table:

f	MAP Criterion
1	.256
2	.384
3	.544
4	.643
5	1.000

Based on this criterion, what is the appropriate number of PC's to extract? Give a reason for your answer.

$f = 1$ since this value minimizes the MAP criterion

6/

(c) Maximum likelihood (ML) factors were also considered for this problem. The Lawley-Maxwell test was performed after f ML factors were extracted for $f = 0, 1, 2$. The results are summarized in the following table:

f	p-value
0	< .0001
1	.031
2	.146

According to this method, what is the appropriate number of ML factors to extract? Give a reason for your answer.

$f = 2$, since we continue to extract ML factors until the Lawley-Maxwell test becomes non-significant.

8/

2. (a) The decision was made to extract f = 3 principal components. The varimax rotated loadings and communality estimates are given below:

	Factor1	Factor2	Factor3
X1	0.73789	0.19566	0.16115
X2	0.87800	-0.09157	-0.01524
X3	0.15915	0.93091	0.09556
X4	0.71742	0.33250	-0.07015
X5	0.73721	0.49192	-0.12358
X6	0.00678	0.05964	0.98308

Final Communality Estimates:

X1	X2	X3	X4	X5	X6
0.609	0.780	0.901	0.630	0.801	0.970

Do these results indicate an adequate factor solution? Why or why not?

No. Factor 2 and Factor 3 are both specific factors. This indicates an inadequate factor solution.

12/

(b) The decision was made to remove X_3 ("SLEEP") and X_6 ("SKIN REACTION") and repeat the factor analysis. The eigenvalues of the new correlation matrix, along with the proportion variance explained are given below:

	1	2	3	4
Eigenvalue	2.59	.68	.46	.27
Proportion	.65	.17	.12	.07
Cum Prop	.65	.82		

The following rotated loadings were obtained:

	Factor1	Factor2
X4	0.92504	0.19496
X5	0.81965	0.39820
X1	0.19117	0.88997
X2	0.35842	0.76477

Do these results indicate an adequate factor solution? Why or why not?

Yes, the 2 factors together explain over 75% of the variance and there are no specific or trivial factors.

12/

(c) Use the rotated loadings in Part (b) above to calculate the estimated communalities for this factor solution. After examining these values, comment on the apparent quality of the factor solution.

12

	Variable	Communality
X_1	Symptoms	.829
X_2	Activity	.713
X_4	Eat	.894
X_5	Appetite	.830

Provides an adequate factor solution since all $\hat{h}_i^2 \geq .50$

For symptoms: $\hat{h}_1^2 = (.19117)^2 + (.88997)^2$
$= .829$

For activity: $\hat{h}_2^2 = (.35842)^2 + (.76477)^2$
$= .713$

For eat: $\hat{h}_4^2 = (.92504)^2 + (.19496)^2$
$= .894$

For appetite: $\hat{h}_5^2 = (.81965)^2 + (.39820)^2$
$= .830$

8

(d) After examining the rotated loadings in Part (b) above, name the two factors.

Factor 1: Diet (or "food consumption" or "nutrition")

Factor 2: General Condition (or perhaps "quality of life" or "health impact")

(e) Give the formulas for the factor scores for each of the two rotated factors using the "rough approximation" method covered in class.

8/

Factor 1 : $X_4 + X_5$

Factor 2 : $X_1 + X_2$

(f) It was decided to use four indices to describe a patient's reaction to radiotherapy: the two original variables, SLEEP and SKIN REACTION, and the two factors that you named in part (d) above.

Suppose that an individual undergoing radiotherapy has the following vector of standardized values for the 6 variables.

X_1	X_2	X_3	X_4	X_5	X_6
[-1.0	1.5	-.5	1.5	1.75	-2.0]

12/

Calculate the factor scores for this individual using the formulas you derived in part (e) above. Would you characterize this patients' overall reaction to radiotherapy as "below average," "average," or "above average")? Give a reason for your answer.

Factor 1 : $1.5 + 1.75 = 3.25$

Factor 2 : $-1.0 + 1.5 = 0.50$

Sleep : $-.5$

Skin Reaction : -2.0

I would characterize their overall reaction as above average (that is, as "good"). They are eating well, have few skin problems, are slightly above the mean in terms of general condition and slightly below the mean in terms of sleep.