Factor Analysis Continued

Psy 524

Ainsworth

Equations – Extraction Principal Components Analysis

	Variables			
Skiers	Cost	Lift	Depth	Powder
S1	32	64	65	67
S2	61	37	62	65
S3	59	40	45	43
S4	36	62	34	35
S5	62	46	43	40

Equations – Extraction

- Correlation matrix w/ 1s in the diag

	Cost	Lift	Depth	Powder
Cost	1	-0.952990	-0.055276	-0.129999
Lift	-0.952990	1	-0.091107	-0.036248
Depth	-0.055276	-0.091107	1	0.990174
Powder	-0.129999	-0.036248	0.990174	1

  - Large correlation between Cost and Lift and another between Depth and Powder
  - Looks like two possible factors

Equations – Extraction

- Reconfigure the variance of the correlation matrix into eigenvalues and eigenvectors

	2.016356722	0.000000000	0.000000000	0.000000000
	0.000000000	1.941495414	0.000000000	0.000000000
L =	0.000000000	0.000000000	0.037784214	0.000000000
	0.000000000	0.000000000	0.000000000	0.004363650
	0.3525740484	0.6142350748	0.6627227864	0.2433214370
	-0.2513116190	-0.6636944396	0.6760812698	0.1981572044
V =	-0.6273115310	0.3224031907	0.2748380988	-0.6534527107
	-0.6473131033	0.2797876809	-0.1678590017	0.6888598954

Equations – Extraction

- $L=V'RV$
  - Where L is the eigenvalue matrix and V is the eigenvector matrix.
  - This diagonalized the R matrix and reorganized the variance into eigenvalues
  - A 4 x 4 matrix can be summarized by 4 numbers instead of 16.

Equations – Extraction

- R=VLV’
  - This exactly reproduces the R matrix if all eigenvalues are used
    - SPSS matrix output ‘factor_extraction.sps’
  - Gets pretty close even when you use only the eigenvalues larger than 1.
    - More SPSS matrix output

Equations – Extraction

- Since $R=VLV'$

$$R = V\sqrt{L}\sqrt{L}V'$$

$$R = (V\sqrt{L})(\sqrt{L}V')$$

$$V\sqrt{L} = A, \sqrt{L}V' = A'$$

$R = AA'$, where A is the loading matrix

and A' is the transpose of the loading matrix.

See SPSS output from matrix syntax.

Equations – Extraction

	Factor 1	Factor 2
Cost	-0.401	0.907
Lift	0.251	-0.954
Depth	0.933	0.351
Powder	0.957	0.288

- Here we see that factor 1 is mostly Depth and Powder (Snow Condition Factor)
- Factor 2 is mostly Cost and Lift, which is a resort factor
- Both factors have complex loadings

Equations – Orthogonal Rotation

- Factor extraction is usually followed by rotation in order to maximize large correlation and minimize small correlations
- Rotation usually increases simple structure and interpretability.
- The most commonly used is the Varimax variance maximizing procedure which maximizes factor loading variance

Equations – Orthogonal Rotation

- The unrotated loading matrix is multiplied by a transformation matrix which is based on angle of rotation

$$A_{unrotated}\Lambda = A_{rotated}$$

$$\Lambda = \begin{bmatrix} \cos\Psi & -\sin\Psi \\ \sin\Psi & \cos\Psi \end{bmatrix}, \text{ where } \Psi \text{ is the angle of rotation}$$

$$\text{if } \Psi = 19 \text{ then } \Lambda = \begin{bmatrix} .946 & -.326 \\ .326 & .946 \end{bmatrix}$$

See SPSS matrix syntax.

Equations – Other Stuff

	Factor 1	Factor 2	Communalities (h^2)
COST	−.086	.981	$\sum a^2 = .970$
LIFT	−.071	−.977	$\sum a^2 = .960$
DEPTH	.994	.026	$\sum a^2 = .989$
POWDER	.997	−.040	$\sum a^2 = .996$
SSLs	$\sum a^2 = 1.994$	$\sum a^2 = 1.919$	3.915
Proportion of variance	.50	.48	.98
Proportion of covariance	.51	.49	

- Communalities are found from the factor solution by the sum of the squared loadings
  - 97% of cost is accounted for by Factors 1 and 2

Equations – Other Stuff

	Factor 1	Factor 2	Communalities (h^2)
COST	−.086	.981	$\sum a^2 = .970$
LIFT	−.071	−.977	$\sum a^2 = .960$
DEPTH	.994	.026	$\sum a^2 = .989$
POWDER	.997	−.040	$\sum a^2 = .996$
SSLs	$\sum a^2 = 1.994$	$\sum a^2 = 1.919$	3.915
Proportion of variance	.50	.48	.98
Proportion of covariance	.51	.49	

- Proportion of variance in a variable set accounted for by a factor is the SSLs for a factor divided by the number of variables
  - For factor 1 1.994/4 is .50

Equations – Other Stuff

	Factor 1	Factor 2	Communalities (h^2)
COST	−.086	.981	$\sum a^2 = .970$
LIFT	−.071	−.977	$\sum a^2 = .960$
DEPTH	.994	.026	$\sum a^2 = .989$
POWDER	.997	−.040	$\sum a^2 = .996$
SSLs	$\sum a^2 = 1.994$	$\sum a^2 = 1.919$	3.915
Proportion of variance	.50	.48	.98
Proportion of covariance	.51	.49	

- The proportion of covariance in a variable set accounted for by a factor is the SSLs divided by the total communality (or total SSLs across factors)
  - 1.994/3.915 = .51

Equations – Other Stuff

- The residual correlation matrix is found by subtracting the reproduced correlation matrix from the original correlation matrix.
  - See matrix syntax output
  - For a “good” factor solution these should be pretty small.
  - The average should be below .05 or so.

Equations – Other Stuff

- Factor weight matrix is found by dividing the loading matrix by the correlation matrix
  - See matrix output

$$B = R^{-1}A$$

Equations – Other Stuff

- Factors scores are found by multiplying the standardized scores for each individual by the factor weight matrix and adding them up.

$$F = ZB$$

Equations – Other Stuff

- You can also estimate what each subject would score on the (standardized) variables

$$Z = FA'$$

Equations – Oblique Rotation

- In oblique rotation the steps for extraction are taken
  - The variables are assessed for the unique relationship between each factor and the variables (removing relationships that are shared by multiple factors)
  - The matrix of unique relationships is called the pattern matrix.
  - The pattern matrix is treated like the loading matrix in orthogonal rotation.

Equations – Oblique Rotation

- The Factor weight matrix and factor scores are found in the same way
- The factor scores are used to find correlations between the variables.

$$pattern = A = \begin{vmatrix} -.079 & .981 \\ -.078 & -.978 \\ .994 & .033 \\ .977 & -.033 \end{vmatrix}$$

Equations – Oblique Rotation

$$R^{-1}A = B = \begin{vmatrix} .104 & .584 \\ .081 & -.421 \\ .159 & -.020 \\ .856 & .034 \end{vmatrix}$$

$$F = ZB = \begin{vmatrix} 1.12 & -1.18 \\ 1.01 & .88 \\ -.46 & .68 \\ -1.07 & -.98 \\ -.59 & .59 \end{vmatrix}$$

Equations – Oblique Rotation

- Once you have the factor scores you can calculate the correlations between the factors (phi matrix; Φ)

$$\Phi = \left(\frac{1}{N-1}\right) F'F$$

Equations – Oblique Rotation

$$\Phi = \frac{1}{4}\begin{bmatrix} 1.12 & 1.01 & -0.46 & -1.07 & -0.59 \\ -1.18 & 0.88 & 0.68 & -0.98 & 0.59 \end{bmatrix} * \begin{bmatrix} 1.12 & -1.18 \\ 1.01 & .88 \\ -.46 & .68 \\ -1.07 & -.98 \\ -.59 & .59 \end{bmatrix} = \begin{bmatrix} 1.00 & -.01 \\ -.01 & 1.00 \end{bmatrix}$$

Equations – Oblique Rotation

- The structure matrix is the (zero-order) correlations between the variables and the factors.

$$C = A\Phi$$

$$C = \begin{bmatrix} -.079 & .981 \\ -.078 & -.978 \\ .994 & .033 \\ .977 & -.033 \end{bmatrix} * \begin{bmatrix} 1.00 & -.01 \\ -.01 & 1.00 \end{bmatrix} = \begin{bmatrix} -.069 & .982 \\ -.088 & -.977 \\ .994 & .023 \\ .997 & -.043 \end{bmatrix}$$

Equations – Oblique Rotation

- With oblique rotation the reproduced factor matrix is found be multiplying the structure matrix by the pattern matrix.

$$R_{rep} = CA'$$

What else?

- How many factors do you extract?
  - One convention is to extract all factors with eigenvalues greater than 1 (e.g. PCA)
  - Another is to extract all factors with non-negative eigenvalues
  - Yet another is to look at the scree plot
  - Number based on theory
  - Try multiple numbers and see what gives best interpretation.

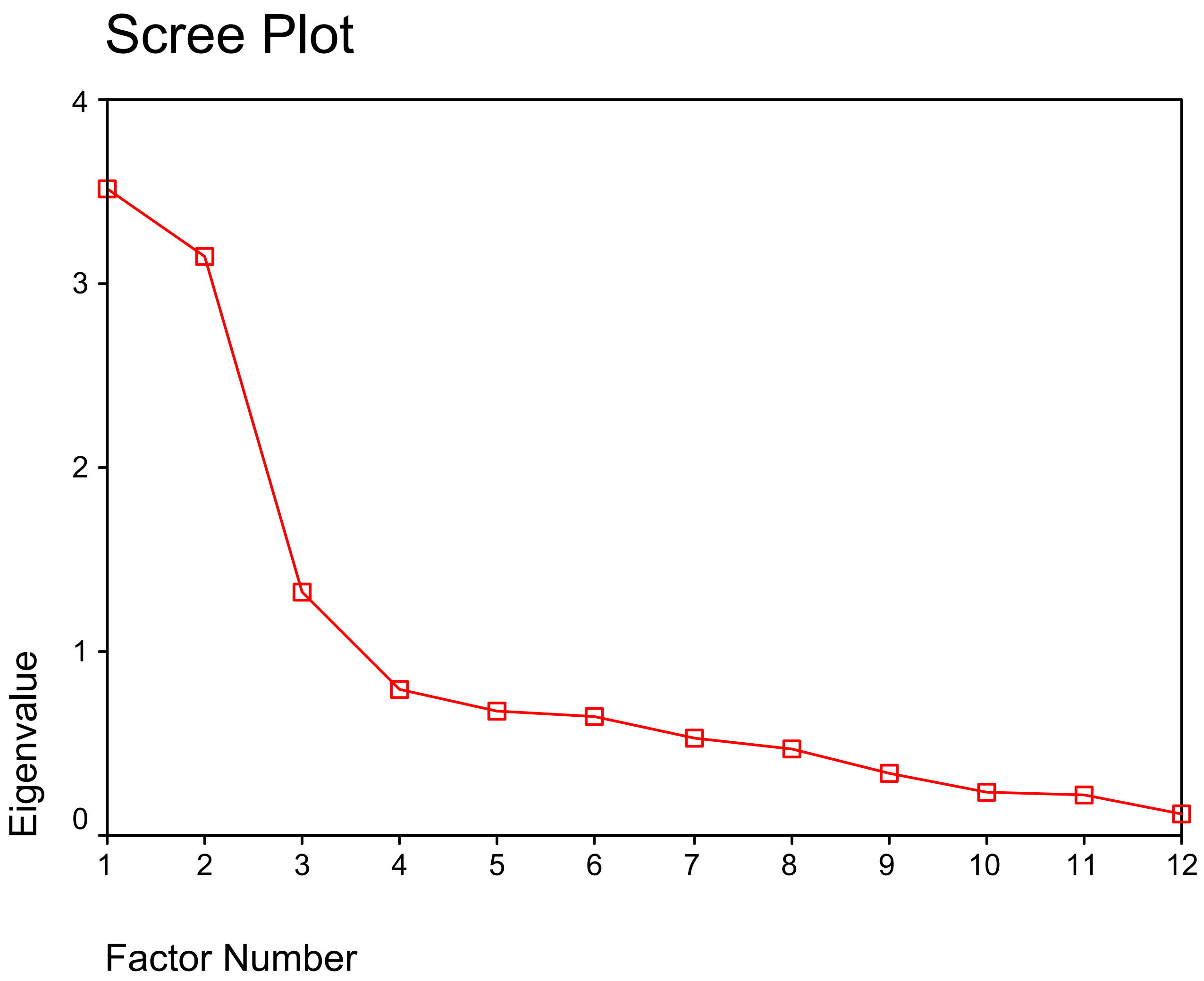
Eigenvalues greater than 1

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.513	29.276	29.276	3.296	27.467	27.467	3.251	27.094	27.094
2	3.141	26.171	55.447	2.681	22.338	49.805	1.509	12.573	39.666
3	1.321	11.008	66.455	.843	7.023	56.828	1.495	12.455	52.121
4	.801	6.676	73.132	.329	2.745	59.573	.894	7.452	59.573
5	.675	5.623	78.755						
6	.645	5.375	84.131						
7	.527	4.391	88.522						
8	.471	3.921	92.443						
9	.342	2.851	95.294						
10	.232	1.936	97.231						
11	.221	1.841	99.072						
12	.111	.928	100.000						

Extraction Method: Principal Axis Factoring.

Scree Plot



Scree Plot
4
3
2
1
0
Eigenvalue
1
2
3
4
5
6
7
8
9
10
11
12
Factor Number

What else?

- How do you know when the factor structure is good?
  - When it makes sense and has a simple (relatively) structure.
- How do you interpret factors?
  - Good question, that is where the true art of this comes in.