

Factor Analysis and Related Methods

James H. Steiger

Vanderbilt University

Primary Goals for Factor Analytic Methods

- 1. Structural Exploration**
- 2. Structural Confirmation**
- 3. Data Reduction**
- 4. Attribute Scoring**

Types of Factor Analytic Methods – Some Major Distinctions

Exploratory vs. Confirmatory

Common Factor Analysis vs. Component Analysis

A Typical “Exploratory” Factor Analysis

Preliminary Issue: How many factor analyses are truly completely “exploratory”? (Probably very few.)

Comment. When exploring structure of traits and abilities, we often have ideas, based on our own “commonsense” experience with the subject matter, of how “underlying dimensions” may give rise to observed variables.

Example. How you reacted when I read you the variable names in our “athletic data example.”

Stages in A Typical “Exploratory” Factor Analysis

1. Decide on a Model and Associated Method (Typically Common Factor Analysis or Component Analysis)

2. Decide on a Number of Factors (at least temporarily) ;<)

3. Obtain an Unrotated Solution for the Factor Pattern

4. Evaluate Overall Fit

If fit is inadequate, increase the number of factors and re-extract.

Stages in A Typical “Exploratory” Factor Analysis

5. Rotate to Simple Structure

Typically using orthogonal transformation

6. Name the Factors

Examine the manifest variables that the factors load heavily on. See what they “have in common.”

7. Consider Further Steps

Oblique transformation to improve simple structure.
Dropping manifest variables.

Deciding on a Model

Common Factor Analysis Model

p manifest random variables in the random vector \mathbf{y} .

These represent variables that the data analyst wishes to represent with an *m*-factor common factor model.

The Common Factor Model

The common factor analysis model states that the p observed variables can be expressed as linear functions of m unobserved (“latent”) variables called *common factors*, and that if this is done in the least-squares linear regression sense, i.e., we predict the variables in \mathbf{y} from these common factors with multiple linear regression weights, *the resulting residuals will be uncorrelated*.

The Common Factor Model

Algebraically, we say that

$$\mathbf{y} = \mathbf{F}\mathbf{x} + \mathbf{e} \quad (1)$$

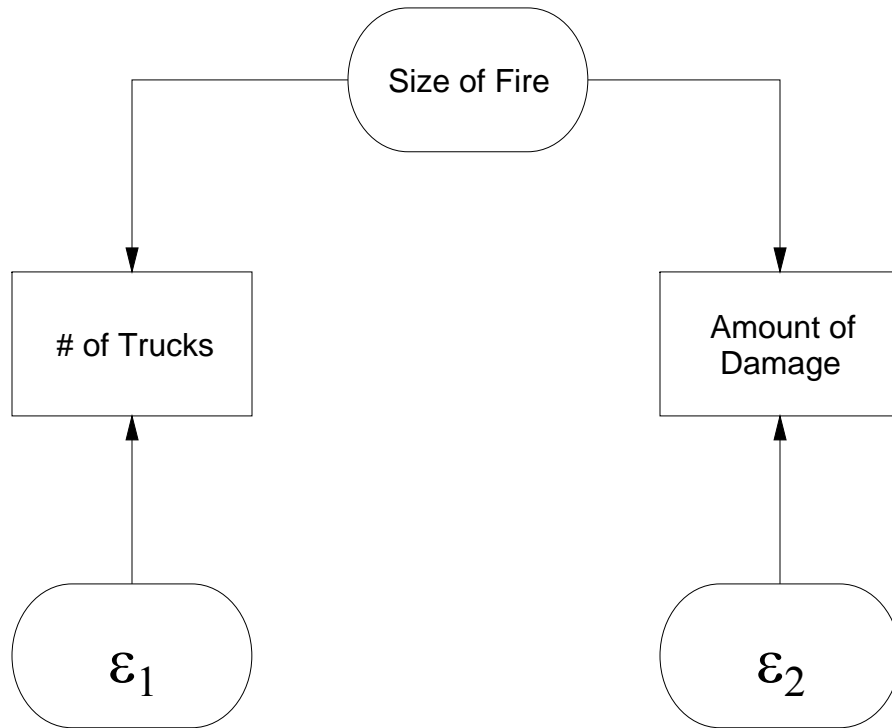
$$E(\mathbf{xx}') = \mathbf{P}, E(\mathbf{xe}') = \mathbf{0}, E(\mathbf{ee}') = \mathbf{U}^2, \quad (2)$$

where \mathbf{U} is a diagonal, positive-definite matrix. \mathbf{F} is the *common factor pattern*, \mathbf{P} the matrix of factor correlations, \mathbf{U}^2 contains the *unique variances* of the variables on its diagonal. If \mathbf{P} is an identity matrix and

the factors are uncorrelated, we say that the common factors are *orthogonal*, otherwise they are *oblique*.

Before exploring the algebra for such a model, we might quickly review some reasons for considering it important. There are many reasons for wanting to fit a common factor model to a set of variables. Here we will consider 4 common, somewhat interrelated ones: (1) The *partial correlation* rationale; (2) The *random noise* rationale; (3) The *true score* rationale, (4) the *data reduction* rationale.

The Partial Correlation-Explanation Rationale



This idea leads to the following notion. If the partial correlations among the variables in set \mathbf{y} with those in set \mathbf{x} partialled out are zero, then in some sense the variables in \mathbf{x} *explain*, or *account for* the correlations among the variables in \mathbf{y} .

With this rationale, we view the “common factors” in \mathbf{x} as the underlying common causes of the variables in \mathbf{y} .

The Random Noise Rationale

In some situations, it is reasonable to hypothesize a physical process that involves several underlying sources of variation that are polluted by random noise. A classic example might be EEG responses to carefully timed standardized auditory signals, recorded at several sensors. It may be that each sensor will pick up output from several unified, consistent sources within the brain, but that these signals will also include random, uncorrelated electrical noise. In this case, the underlying sources are the “common factors” in \mathbf{x} , the observed signals are recorded at \mathbf{y} .

The True Score Rationale

In psychometrics, we commonly measure attributes with devices that are assumed to be degraded by random error. In particular, classical *true score theory* postulates measurements that involve an underlying true score component, and a random error component. If we measure the same ability with several items, this turns out to be a special case of the common factor model. What we are really interested in is the underlying true scores on the variables of interest.

The distinction between the observed scores on measures of a trait, and the underlying trait itself, can be

especially crucial when we seek to establish linear regression relations among variables that have varying amounts of error variance. Observed correlations can be *attenuated* by unreliability, and so the regression relations among the unreliable measures of a set of traits can mislead one about the relations among the traits themselves.

Because of this problem, it is common to try to estimate regression relationships between the common factors underlying a group of measures, rather than the measures themselves.

The Data Reduction Rationale

In many situations, it is computationally inconvenient to operate with a large number of measures. We seek to reduce the number of measures, while simultaneously classifying them into groups, and increasing the reliability of what they measure.

This *data reduction* rationale for factor analysis is a major use for factor analytic technology. We factor analyze a group of items to discover the major sources of variation underlying them, and to find out which items are related to which sources. The resulting information allows us to parcel items into groups, to gain a better understanding of the structure underlying our items, and refine our measures of the sources of variation.

The Fundamental Theorem of Factor Analysis

Recall from a proof we did in class that Equation (1) implies that

$$\Sigma = E(\mathbf{y}\mathbf{y}') = \mathbf{F}\mathbf{P}\mathbf{F}' + \mathbf{U}^2 \quad (3)$$

We are always free to set \mathbf{P} equal to an identity matrix, since the common factors are never observed. So Equation (3) implies that, if the common factor model fits, we can find a diagonal, positive definite matrix \mathbf{U}^2 such that

$$\Sigma - \mathbf{U}^2 = \mathbf{F}\mathbf{F}' \quad (4)$$

A matrix that can be expressed in the form \mathbf{FF}' is said to be Gramian. Since \mathbf{F} has m columns, if it is of full column rank, then \mathbf{FF}' will be *Gramian and of rank m* .

So, in effect, when we fit the common factor model to data, we look for a diagonal matrix that, when subtracted from the covariance matrix Σ of the manifest variables, leaves the matrix Gramian and of rank m . In some fitting algorithms, this involves *iteratively* trying various candidates for \mathbf{U}^2 and testing how close they come to reducing Σ to the desired condition. Later, when we discuss eigenvalues, eigenvectors, and matrix factoring, we shall see that this testing process is relatively routine.

Key Characteristics of the Common Factor Model

Error variables are uncorrelated, leading to the “partial correlation rationale”

Latent variables are “outside the test space,” i.e., cannot be expressed as linear combinations of the manifest variables. This can be seen either as a virtue or a shortcoming.

There are several “indeterminacy problems” to be discussed in detail later. Factor scores cannot be uniquely calculated.

Component Analysis Systems

$$\mathbf{y} = \mathbf{F}\mathbf{x} + \mathbf{e} \quad (5)$$

but

$$\mathbf{x} = \mathbf{B}'\mathbf{y}, \text{ and } \mathbf{e} = (\mathbf{I} - \mathbf{FB}')\mathbf{y} \quad (6)$$

with

$$E(\mathbf{x}\mathbf{e}') = \mathbf{0} \quad (7)$$

As before, this is a linear regression system and exhibits the key properties of such systems. Note that from our basic knowledge of regression algebra, we can say that the covariance matrix of \mathbf{x} is $\mathbf{B}'\Sigma\mathbf{B}$, and, more importantly, \mathbf{F} , the matrix of multiple regression weights, is

$$\mathbf{F} = \Sigma\mathbf{B}(\mathbf{B}'\Sigma\mathbf{B})^{-1} \quad (8)$$

Principal Components Analysis

The *first principal component* of a set of variables is that linear combination which, for a vector of linear weights of fixed length, has maximum variance.

The second principal component is that linear combination which is orthogonal to the first, and otherwise has maximum variance.

The set of linear weights (**B** in Equation (6) above) satisfying this property are given by the *eigenvectors of the covariance matrix Σ for the manifest variables*.

Key Characteristics of Principal Components

The latent variables are “in the test space,” i.e., can be expressed as linear combinations of the manifest variables.

Principal components are maximally efficient at “data reduction,” that is, they account for the maximum amount of variance with the minimum number of variables.

Principal component scores are uniquely defined and easily calculated.

Key Characteristics of Principal Components

Principal components are much easier to compute than common factors when the number of manifest variables is large, and much less subject to numerical problems.

Selecting a Model in SPSS

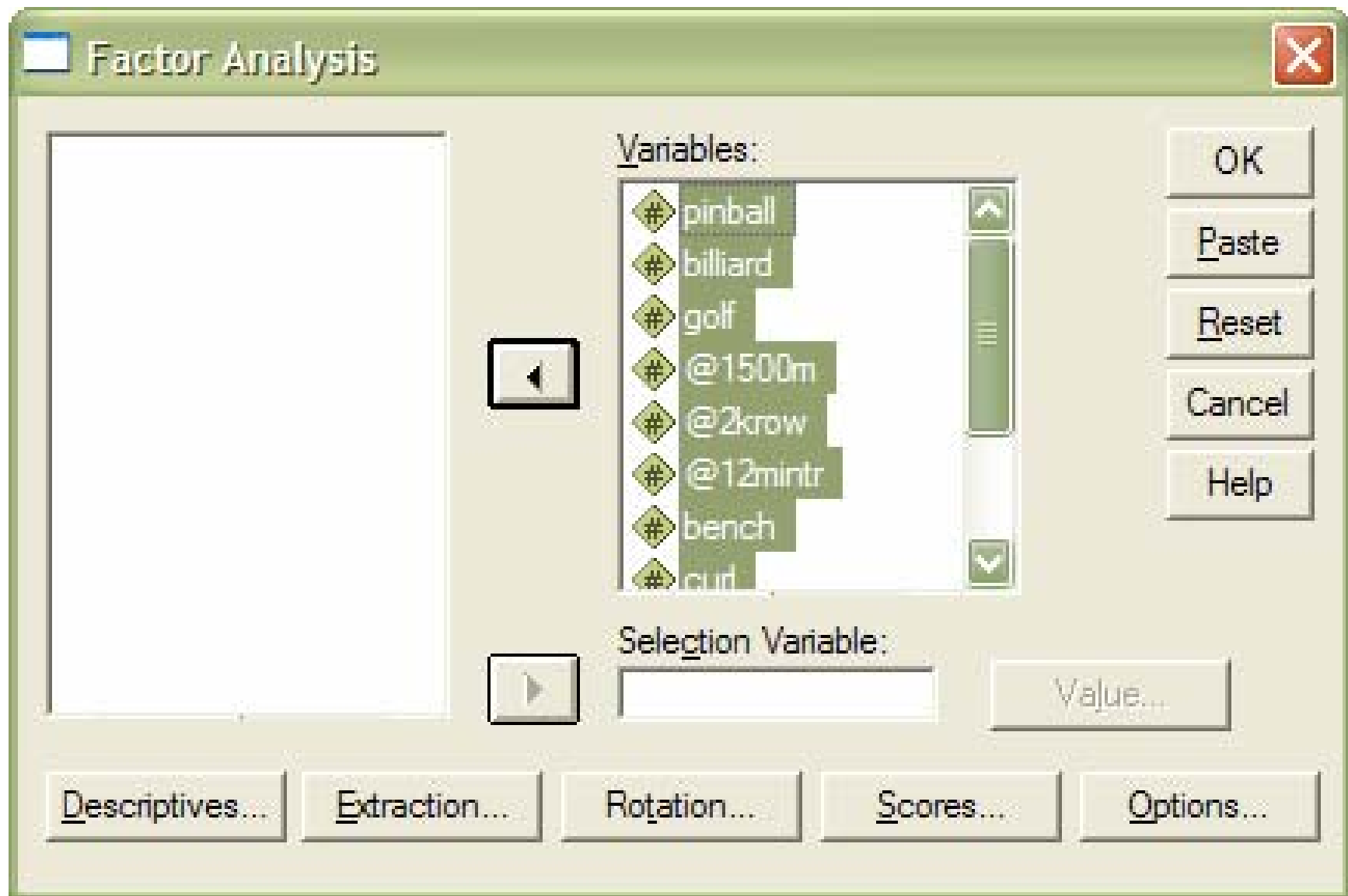
Load the data file into SPSS. Notice there are 1000 observations on 9 variables.

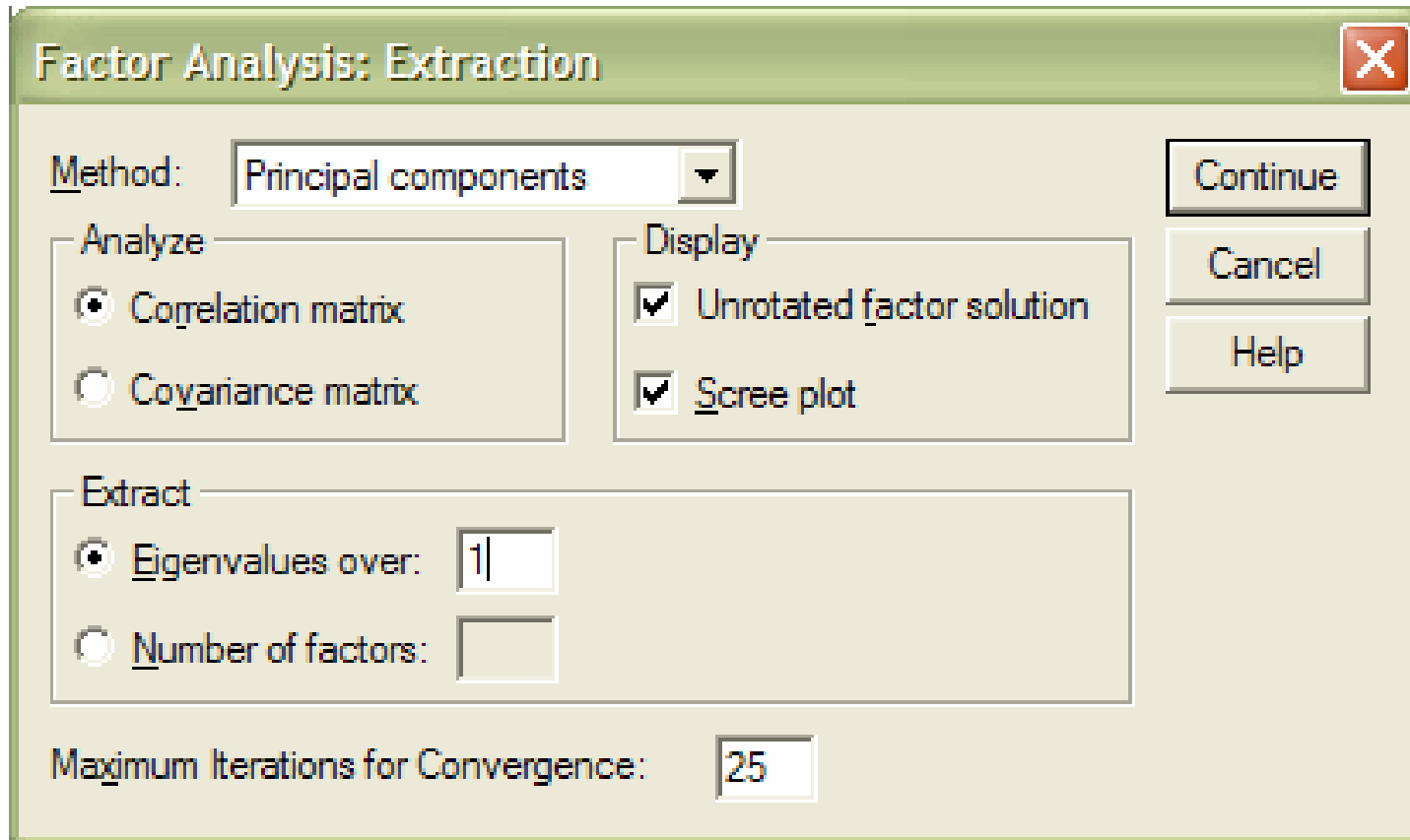
	Name
1	pinball
2	billiard
3	golf
4	@1500m
5	@2krow
6	@12mintr
7	bench
8	curl
9	maxpushu

Analyze Graphs Utilities Window Help

- Reports ▶
- Descriptive Statistics ▶
- Tables ▶
- Compare Means ▶
- General Linear Model ▶
- Mixed Models ▶
- Correlate ▶
- Regression ▶
- Loglinear ▶
- Classify ▶
- Data Reduction ▶**
 - Factor...**
 - Correspondence Analysis...
 - Optimal Scaling...
- Scale ▶
- Nonparametric Tests ▶
- Time Series ▶
- Survival ▶
- Multiple Response ▶
- Missing Value Analysis...

	0m	@2krow	@12mir
	- .95	- .16	.
	.68	.15	.
	- .68	- .52	.
	.93	-1.03	.
	.98	1.22	.
	- .93	- .51	-1
	- .66	- .06	.
	2.04	1.40	.





The above are commonly selected initial options. The principal component solution may be used to approximate the common factor solution quickly, *and* give an indication of the correct number of factors.

Factor Analysis: Rotation



Method

- | | |
|--|---|
| <input type="radio"/> <u>N</u> one | <input type="radio"/> <u>Q</u> uartimax |
| <input checked="" type="radio"/> <u>V</u> arimax | <input type="radio"/> <u>E</u> quamax |
| <input type="radio"/> <u>D</u> irect <u>O</u> blimin | <input type="radio"/> <u>P</u> romax |

Delta:

Kappa

Continue

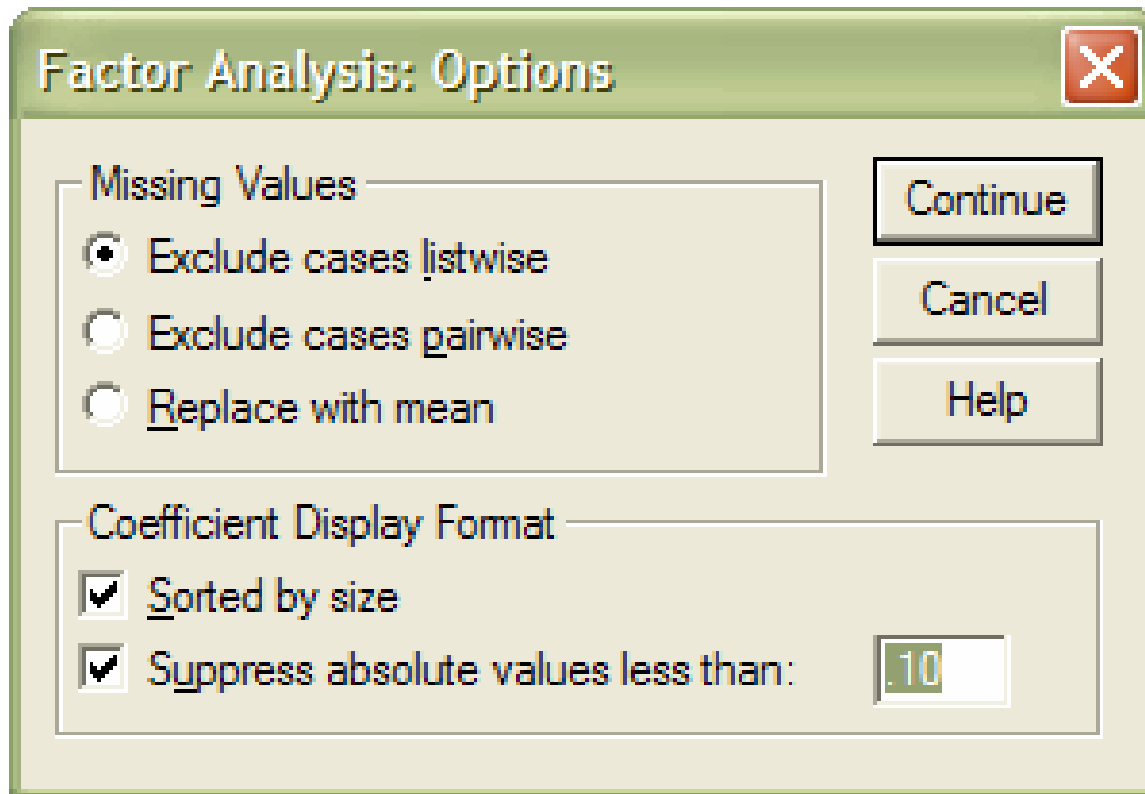
Cancel

Help

Display

- | | |
|--|---|
| <input checked="" type="checkbox"/> <u>R</u> otated solution | <input checked="" type="checkbox"/> <u>L</u> oading plot(s) |
|--|---|

Maximum iterations for Convergence:



Sorting coefficients by size, and suppressing small ones makes the factor pattern *much* easier to read.

Communalities

	Initial	Extraction
PINBALL	1.000	.584
BILLIARD	1.000	.706
GOLF	1.000	.692
@1500M	1.000	.739
@2KROW	1.000	.641
@12MINTR	1.000	.664
BENCH	1.000	.757
CURL	1.000	.679
MAXPUSHU	1.000	.627

Extraction Method: Principal Component Analysis.

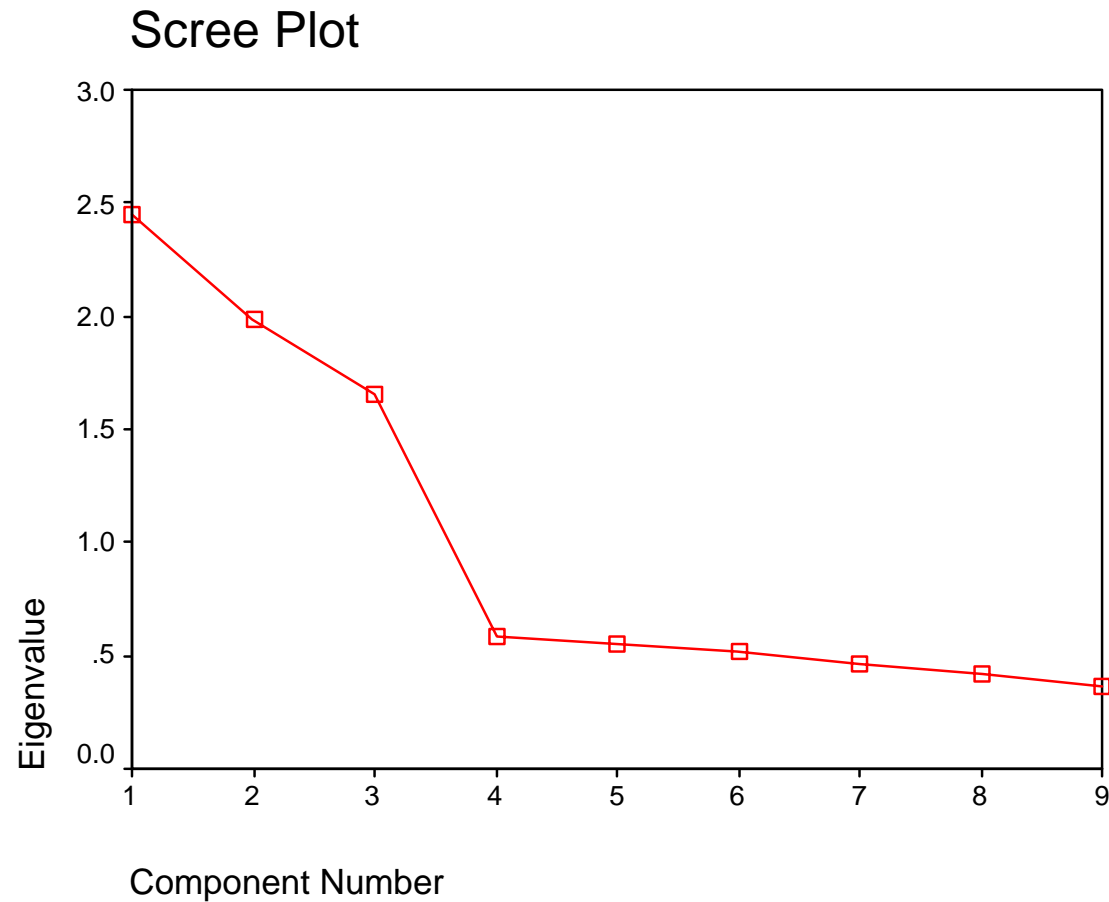
Communalities are the variances accounted for by the factors.

Total Variance Explained

Component	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	2.448	27.199	27.199	2.448	27.199	27.199	2.075	23.057	23.057
2	1.981	22.013	49.212	1.981	22.013	49.212	2.028	22.529	45.587
3	1.660	18.442	67.654	1.660	18.442	67.654	1.986	22.067	67.654
4	.590	6.555	74.209						
5	.552	6.128	80.336						
6	.524	5.821	86.157						
7	.460	5.112	91.269						
8	.417	4.636	95.906						
9	.368	4.094	100.000						

Extraction Method: Principal Component Analysis.

When 3 components are retained, they account for 67.6% of the variance in the 9 variables.



Look for an “elbow” in the Scree plot, and go back one factor from the point of the elbow. 3 components are retained.

Rotational Indeterminacy of F

The existence of orthogonal matrices satisfying $\mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$ implies that even if \mathbf{U} is identified, \mathbf{F} will not be if $m > 1$. Suppose we require m orthogonal factors. If such a model fits, then infinitely many \mathbf{F} matrices will satisfy $\Sigma - \mathbf{U}^2 = \mathbf{F}\mathbf{F}'$, since $\mathbf{F}\mathbf{F}' = \mathbf{F}_1\mathbf{F}_1'$ so long as $\mathbf{F}_1 = \mathbf{F}\mathbf{T}$, for any orthogonal \mathbf{T} .

Simple Structure

Thurstone “solved” this very significant problem with the *simple structure* criterion. Development of “machine rotation” methods and digital computers elevated factor analysis from the status of an esoteric technique understood and practiced by a gifted elite, to a technique accessible (for use and misuse) to virtually anyone. Perhaps lost in the shuffle was the important question of why one would expect to find simple structure in many variable systems.

Rotated Component Matrix^a

	Component		
	1	2	3
@1500M	.836	-.198	
@12MINTR	.810		
@2KROW	.675	.429	
BENCH	-.141	.848	.134
CURL		.817	
MAXPUSHU	.486	.625	
BILLIARD			.839
GOLF			.829
PINBALL		.134	.752

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.

Factor Analysis: Extraction ✖

Method: Maximum likelihood

Analyze

Correlation matrix

Covariance matrix

Display

Unrotated factor solution

Scree plot

Extract

Eigenvalues over: 1

Number of factors: 3

Maximum iterations for Convergence: 50

Continue
Cancel
Help

Goodness-of-fit Test

Chi-Square	df	Sig.
12.940	12	.373

Rotated Factor Matrix^a

	Factor		
	1	2	3
@1500M	.779	-.179	
@12MINTR	.678		
@2KROW	.585	.372	
BENCH	-.119	.816	.137
CURL		.674	
MAXPUSHU	.433	.522	
BILLIARD			.765
GOLF			.734
PINBALL		.131	.590

Extraction Method: Maximum Likelihood.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.

Pattern Matrix^a

	Factor		
	1	2	3
@1500M	.794	-.230	
@12MINTR	.684		
@2KROW	.565	.343	
BENCH	-.171	.829	
CURL		.682	
MAXPUSHU	.403	.505	
BILLIARD			.769
GOLF			.736
PINBALL			.581

Extraction Method: Maximum Likelihood.

Rotation Method: Promax with Kaiser Normalization.

a. Rotation converged in 4 iterations.

Factor Correlation Matrix

Factor	1	2	3
1	1.000	.123	.033
2	.123	1.000	.181
3	.033	.181	1.000

Extraction Method: Maximum Likelihood.

Rotation Method: Promax with Kaiser Normalization.