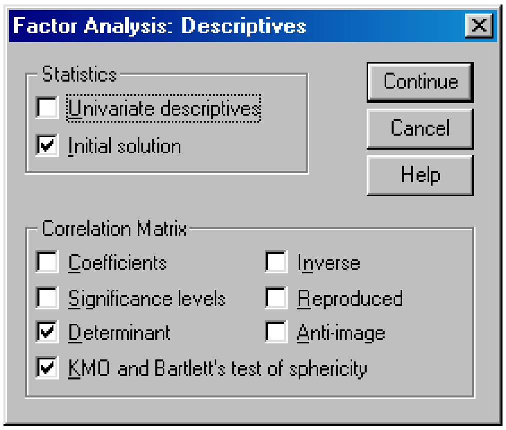
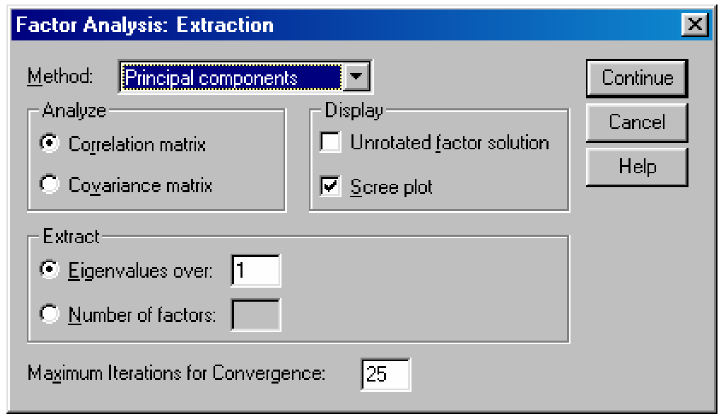
943 -- Determining the Number of Factors

Working again with the data from the adolescents receiving treatment for behavior disorders. But this time the question is about the number and kinds of information available from a series of variables that related to in-home and in-school behavior.

A PC analysis of these variables (without rotation) gives us the following (with some rearrangement)…





Communalities

	Initial
SEI	1.000
SEP	1.000
EI	1.000
EP	1.000
in-school disciplinary action	1.000
suspension from school	1.000
did the court remove the child from treatment	1.000

Extraction Method: Principal Component Analysis.

Sutter-Eiberg Intensity -- frequency of specific behaviors exhibited by child -- teacher rating
Sutter-Eiberg Problem -- how problematic are the same behaviors -- teacher rating
Eiberg Intensity -- parent rating version
Eibher Problem -- parent rating version

Total Variance Explained

	Initial Eigenvalues		
Component	Total	% of Variance	Cumulative %
1	3.456	49.367	49.367
2	1.491	21.306	70.672
3	.924	13.205	83.877
4	.468	6.679	90.556
5	.328	4.690	95.247
6	.226	3.234	98.480
7	.106	1.520	100.000

Extraction Method: Principal Component Analysis.

$\lambda > 1.00$ rule would lead to keeping 2 PCs, but notice the 3rd is "close" (13.2% of variance) and the 4th is "not even close" (6.7%).

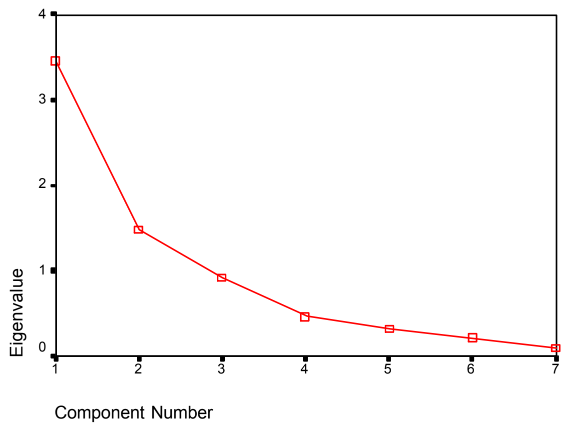
Correlation Matrix[a]

a. Determinant = 1.761E-02

The determinant represents the variances of the variables "minus" the covariances among them.

- So, the more of the variance of the variables that is related to the other variable the smaller the determinant will be.
- If the determinant = 0, it means that at least one of the variables is perfectly correlated with some combination of the other variables -- a singular matrix
- So, the smaller the determinant the more that the variable's variance can be reproduced using the other variables (and PC's formed from them, but …
- If the determinant is "too small" then the mathematics of the factoring formulas become unstable

Scree Plot



The scree plot is sort of ugly! Looks like breaks at 2 and 4 -- suggesting keeping 1 or 3 factors ('67 rule).

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.685
Bartlett's Test of Sphericity	Approx. Chi-Square	164.944
	df	21
	Sig.	.000

The KMO -- indicates the proportion of variance in your variables which is common variance, i.e. which might be caused by underlying factors.

-- values < .50 suggest that the variables won't "factor well"

One of the "Bartlett's Sphericity Tests" is also provided by SPSS -- the one that tests if there's any systematic variance available to be factored.

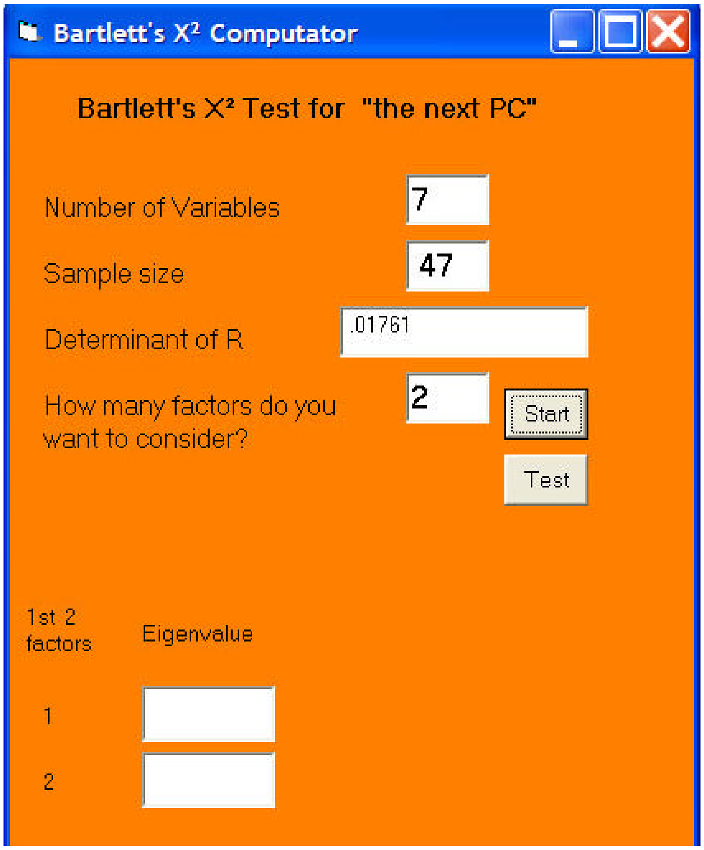
I've provided a program that does the "Keep another factor" X^2 test, let's work with it a bit…

A minimum criterion for keeping a factor is that it is "significant" (probably not a Type I error). This is often referred to as a "minimum" criterion because the test is pretty sensitive, especially when N is large. So, a common result is analogous to having a significant effect, but one with a very small sample size -- the effect is "significant" but not "meaningful".

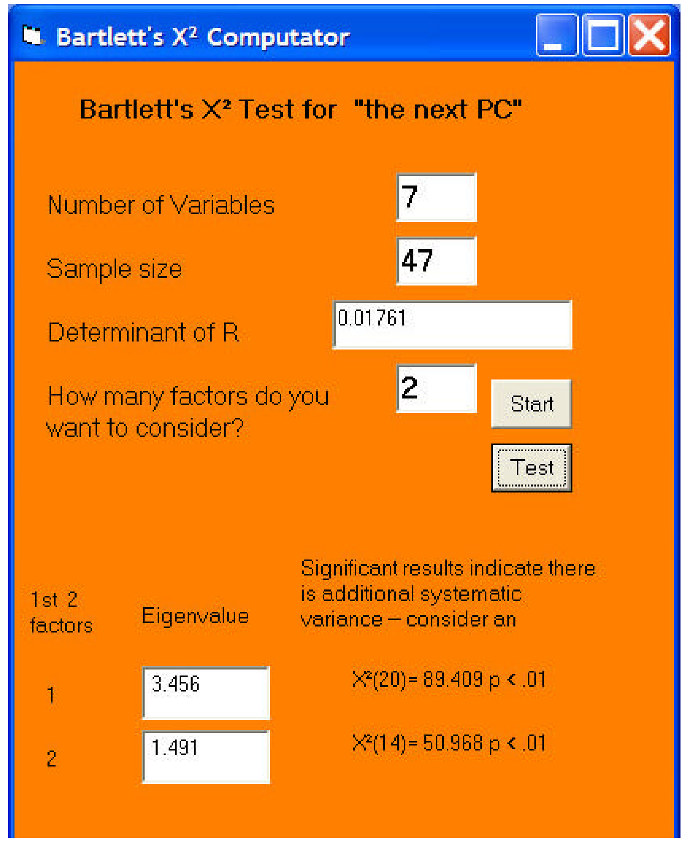
The $\lambda > 1$ rule says keep 2 factors, the 3^{rd} is suggested by the scree, but we might check if the 3^{rd} is "significant". To use the Bartlett's X^2 Computator, we need to know:

- the number of variables → 7
- the sample size → 47
- the determinant of the correlation matrix (R) → .01761
- The number of factors we're sure we want to keep (so we can test the next one) → 2

First we enter these values into the computator and click on the “Start” button



The program then asks for each of the eigenvalues for the factors were are certain we want to keep. Enter all these values and click the “Test” button.



The results shows that there is systematic variance left among the correlations after extracting the second factor – suggesting that the third factor may be a “worthwhile” factor .

Here's Another Example

The researchers wanted to explore the relationship among a selection of dyadic behaviors. The extraction produced…

Total Variance Explained

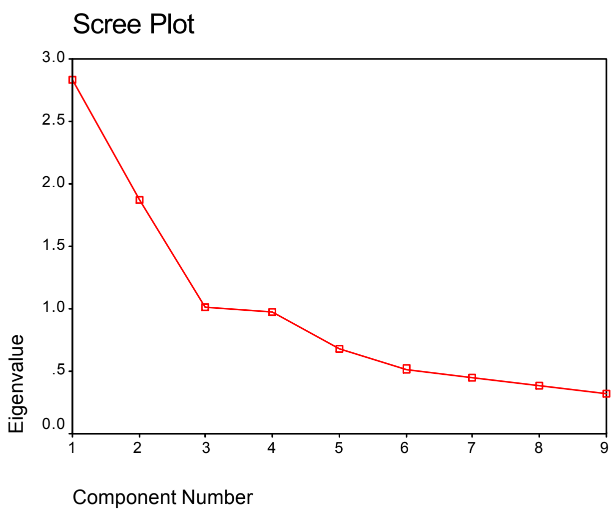
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.829	31.434	31.434	2.829	31.434	31.434
2	1.862	20.690	52.123	1.862	20.690	52.123
3	1.010	11.225	63.348	1.010	11.225	63.348
4	.970	10.783	74.131			
5	.670	7.448	81.579			
6	.514	5.711	87.290			
7	.440	4.893	92.183			
8	.382	4.243	96.426			
9	.322	3.574	100.000			

Extraction Method: Principal Component Analysis.

The $\lambda > 1.00$ rule leads to retaining 3 factors.

But notice that the λ of the 3rd factor is larger than, but very close to 1.00.

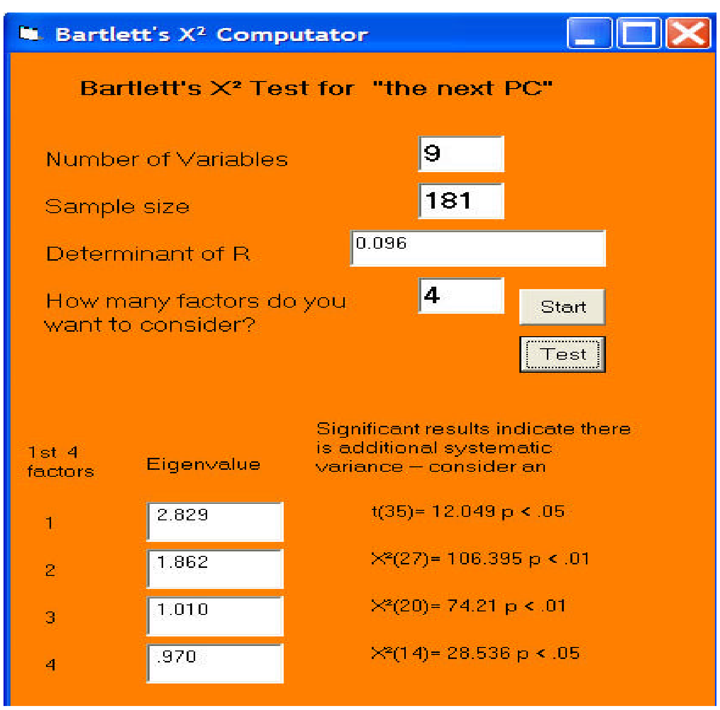
Also notice that the λ of the 4th factor is smaller than, but very close to 1.00



The screen plot looks like …

Major elbow at 3 factors – suggesting 2 or 3 factors, depending upon which version of the scree rule you like.

Maybe another elbow at 5.



Getting the significance tests shows that at least the first 5 factors are based systematic variation.