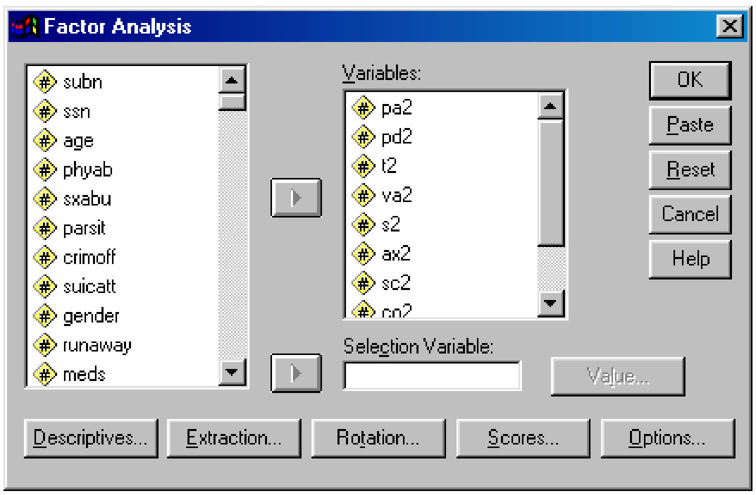
PC Example

A community-based treatment center for adolescents with behavior disorders collects daily data for all residents, including the variables listed below. The question asked was, "How many different 'kinds' of information are we getting from these eight measures?" We decided to use data from the first full month in residence -- by then the "kids" have settled in and are "being themselves" but treatment effects are negligible.

Whether or not this occurred each day	Daily rating using 1-5 scale
physical aggression (pa2) property damage (pd2) theft (t2) extreme verbal abuse patients/staff (va2)	Sad (s2) anxious (ax2) self-confidence (sc2) compliance (co2)

Analyze→ Data Reduction → Factor



Highlight and move the variables into the window…

- Variables should be quantitative or binary

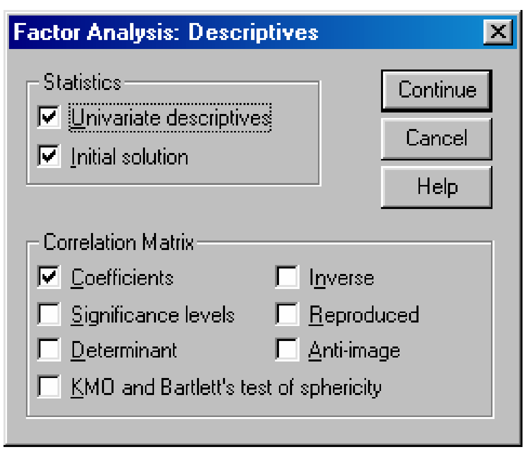
Then work your way through the different specifications to get the analysis and output you want.

We'll keep this first one simple…

Remember that you should do a careful "data screening" before factoring, checking for discontinuities, skewness, outliers, etc.

Be sure **not** to include…

- constants (i.e., need variance to have covariance)
- "elements & aggregates" (e.g., total, V, Q & A GRE)



Descriptives window

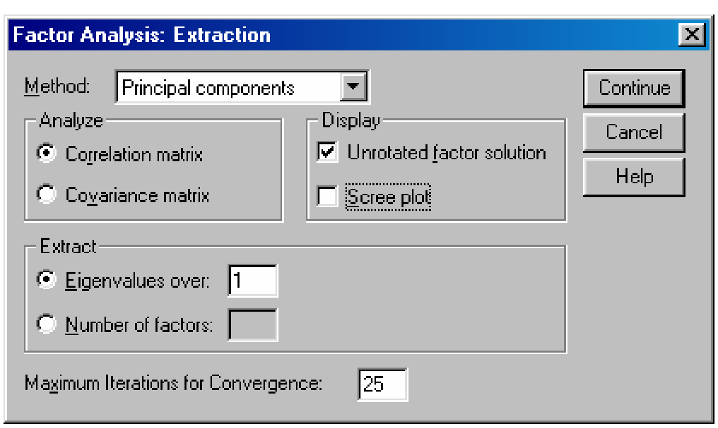
- Basics now -- more later

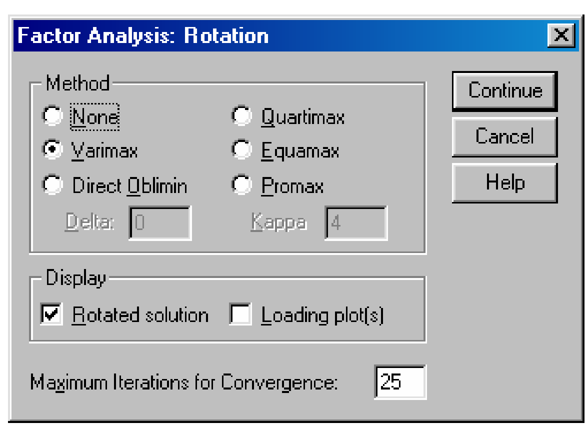
Extraction window

- Lots of decisions to make here
- We'll start with "PC" factoring
- Notice you can determine the number of factors to be extracted

Rotation window

- The "bridge" between mathematical and interpretive aspects of factoring
- Varimax is probably the most commonly used





Output (skipping univariates and correlation matrix to safe a bit space -- also the listing order is changed just a bit)

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	3.048	38.097	38.097	3.048	38.097	38.097	2.304	28.798	28.798
2	1.709	21.363	59.459	1.709	21.363	59.459	2.093	26.163	54.961
3	1.340	16.746	76.205	1.340	16.746	76.205	1.699	21.244	76.205
4	.636	7.953	84.158						
5	.483	6.036	90.194						
6	.340	4.244	94.438						
7	.240	3.000	97.438						
8	.205	2.562	100.000						

Extraction Method: Principal Component Analysis.

Initial Solution — Extracted Solution — Rotated Solution

The initial solution has as many components (factors) as there are variables -- complete components solution --

- The eigenvalues (λ) and % variance are examined to determine the number of factors (% = λ/k)
- The default (and probably most common rule) is the λ > 1.00 rule -- applied here we'd keep 3 factors which would account for 76.205% of the variance in the original 8 variables

The extracted solution has the chosen number of factors -- called a truncated components solution

- Notice that it is a repeat of the information from the first however many factors are kept

The rotated solution is the extracted solution after "changing the viewing angle" to improve interpretability

- Notice that the total variance explained is the same
- Notice that there is some shifting in which factors explain how much variance

Component Matrix[a]

	Component		
	1	2	3
physical aggression	.758	.413	1.164E-03
property damage	.693	.489	-.199
theft	.362	.656	-.204
extreme verbal abuse	.826	6.589E-02	.235
sad	.540	-.510	.441
anxious	.654	-.335	.507
self-confidence	-.349	.539	.669
compliance	-.580	.450	.551

Extraction Method: Principal Component Analysis.

a. 3 components extracted.

The unrotated Component Matrix usually isn't interpreted.

The communalities are the % variance of each variable that is accounted for by the solution. Sometimes low communalities suggest the utility of additional factors.

Communalities

	Initial	Extraction
physical aggression	1.000	.746
property damage	1.000	.759
theft	1.000	.603
extreme verbal abuse	1.000	.742
sad	1.000	.747
anxious	1.000	.797
self-confidence	1.000	.861
compliance	1.000	.843

Extraction Method: Principal Component Analysis

Rotated Component Matrix[a]

	Component		
	1	2	3
physical aggression	.807	.301	-6.51E-02
property damage	.853	9.875E-02	-.148
theft	.751	-.186	6.583E-02
extreme verbal abuse	.562	.645	-.105
sad	-8.06E-02	.846	-.155
anxious	.110	.884	-5.44E-02
self-confidence	3.339E-02	-2.60E-02	.927
compliance	-.165	-.209	.879

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 4 iterations.

Factor "interpretation" focuses on the rotated solution.

The component matrix (often called the Structure Matrix) shows the correlation between each factor and each variable.

Factors are named or interpreted based upon the variables that "load on" or correlate with each. "Cutoffs" for identifying what variables load on which factors vary amazingly across disciplines, authors, and content areas -- .3 - .4 is "modal".

Factorings of this sort often reveal "internalizing" vs. "externalizing" factors. This solution was interestingly different.

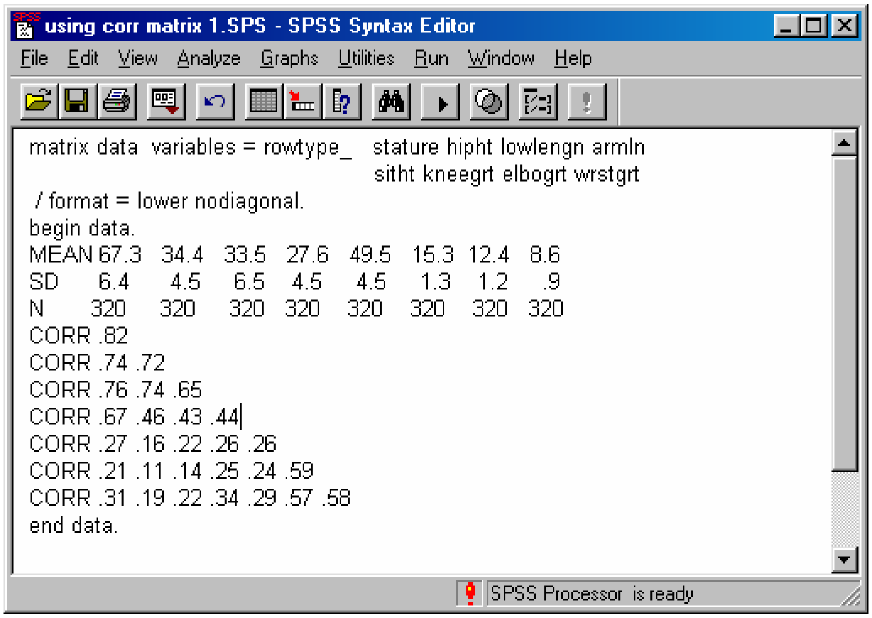
- Factor 1 is probably an "externalizing" factor
- Factor 2 would be a "internalizing" factor, except for the loading by extreme verbal abuse -- which also loads on F1
  - That extreme verbal abuse loads on both factors suggests that this behavior is both tied to other forms of "acting out" and might also be a response to "negative internal states" (i.e., being sad and/or anxious)
- Factor 3 is compliance & self-confidence
  - Notice that these behaviors are separated from the others & include an "internalizing" & "externalizing"
  - Notice that these are both positively weighted -- increases in one "go with" increases in the other

943 Another Factoring Example -- with matrix input

One nice thing about SPSS is that you can start a factor analysis with only the correlation matrix (you do not need the original data file). Articles often include the correlation matrix from a factoring (or other analyses), giving you a chance to "check-up on" or "extend" the published treatment of the data.

Inputting the Data

File→ New→ Syntax to open a "Syntax file"




```
matrix data variables = rowtype_ stature hipht lowlengn armln
                         sitht kneegrt elbogrt wrstgrt
 / format = lower nodiagonal.
begin data.
MEAN 67.3  34.4  33.5  27.6  49.5  15.3  12.4  8.6
SD    6.4   4.5   6.5   4.5   4.5   1.3   1.2   .9
N     320   320   320   320   320   320   320   320
CORR .82
CORR .74 .72
CORR .76 .74 .65
CORR .67 .46 .43 .44
CORR .27 .16 .22 .26 .26
CORR .21 .11 .14 .25 .24 .59
CORR .31 .19 .22 .34 .29 .57 .58
end data.
```

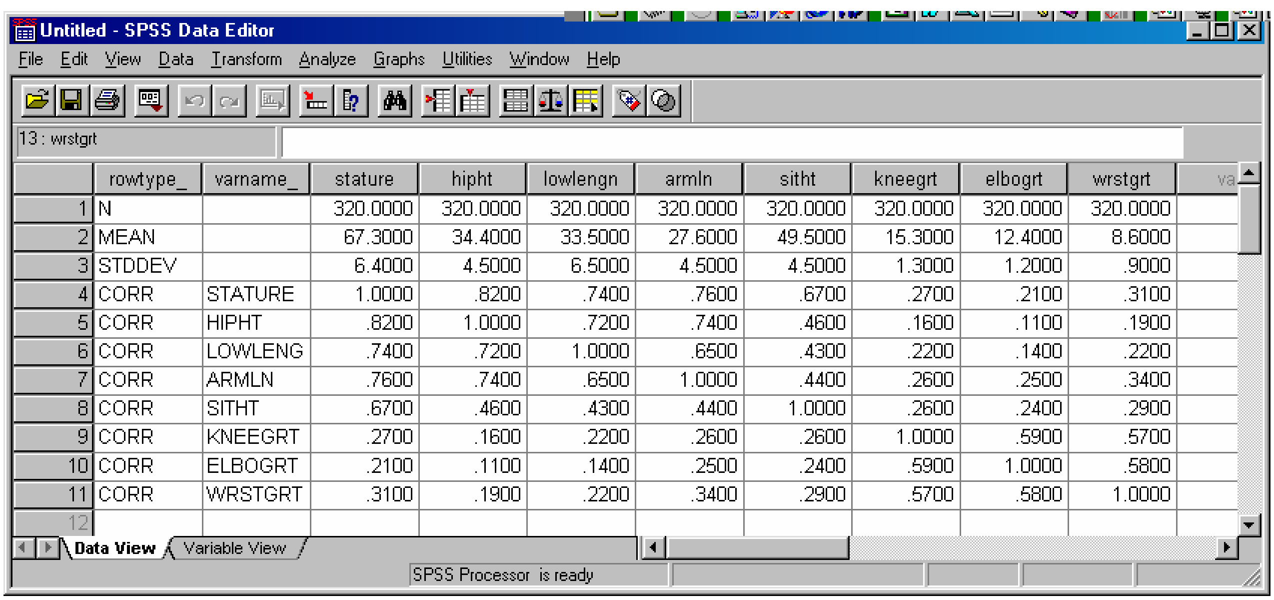
You will use these syntax commands verbatim -- only the variable names and the data values will change.

You need not have the mean and/or sd data to do this.

You will need to include the "N" of the sample in order to conduct significance tests (more later).

Notice that the correlation matrix includes only the lower off-diagonal values (without the 1.00's), so there are only 7 rows of correlations for the 8x8 correlation matrix.

Run → All to put these data into the SPSS Data Editor



	rowtype_	varname_	stature	hipht	lowlengn	armln	sitht	kneegrt	elbogrt	wrstgrt
1	N		320.0000	320.0000	320.0000	320.0000	320.0000	320.0000	320.0000	320.0000
2	MEAN		67.3000	34.4000	33.5000	27.6000	49.5000	15.3000	12.4000	8.6000
3	STDDEV		6.4000	4.5000	6.5000	4.5000	4.5000	1.3000	1.2000	.9000
4	CORR	STATURE	1.0000	.8200	.7400	.7600	.6700	.2700	.2100	.3100
5	CORR	HIPHT	.8200	1.0000	.7200	.7400	.4600	.1600	.1100	.1900
6	CORR	LOWLENG	.7400	.7200	1.0000	.6500	.4300	.2200	.1400	.2200
7	CORR	ARMLN	.7600	.7400	.6500	1.0000	.4400	.2600	.2500	.3400
8	CORR	SITHT	.6700	.4600	.4300	.4400	1.0000	.2600	.2400	.2900
9	CORR	KNEEGRT	.2700	.1600	.2200	.2600	.2600	1.0000	.5900	.5700
10	CORR	ELBOGRT	.2100	.1100	.1400	.2500	.2400	.5900	1.0000	.5800
11	CORR	WRSTGRT	.3100	.1900	.2200	.3400	.2900	.5700	.5800	1.0000

If you use Analyze → Data Reduction → Factor… you'll get a factor analysis, **but not the right one!!!**

Instead you must add the following line to the syntax file and run it. factor matrix in(cor=*).

This command will get you a factoring with all the "defaults" (PC, $\lambda > 1$, Varimax, etc.). To get a full listing of the available syntax commands and their meanings go to…

Help → Topics → Statistical Analysis → Data Reduction → Factor Analysis procedure
→ Syntax → Factor Command Syntax (and click the "Display" button)

About the variables (all measured in inches)…

stature -- standing height measured in inches
hipht -- hip height measures in inches
lowleng -- lower leg length (inseam)
armln -- arm length shoulder to fingertips
sitht -- height when seated on a standard 18" chair

kneegrt -- distance around the upper knee
elbogrt -- distance around the lower elbow
wrstgrt -- distance around the wrist

Considering the variables and the patterns in the correlation matrix, it is easy to expect that we will find 2 factors -- one including the 5 "length" measures and one including the 3 "girth" measures. Let's see…

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.033	50.413	50.413	4.033	50.413	50.413	3.572	44.653	44.653
2	1.776	22.195	72.608	1.776	22.195	72.608	2.236	27.955	72.608
3	.652	8.149	80.757						
4	.445	5.567	86.324						
5	.408	5.101	91.425						
6	.329	4.110	95.535						
7	.232	2.899	98.434						
8	.125	1.566	100.000						

Extraction Method: Principal Component Analysis.

Component Matrix[a]

	Component	
	1	2
STATURE	.906	-.270
HIPHT	.817	-.396
LOWLENGN	.788	-.314
ARMLN	.836	-.195
SITHT	.688	-5.89E-02
KNEEGRT	.502	.678
ELBOGRT	.453	.730
WRSTGRT	.541	.643

Extraction Method: Principal Component Analysis.
a. 2 components extracted.

Rotated Component Matrix[a]

	Component	
	1	2
STATURE	.930	.169
HIPHT	.908	1.575E-02
LOWLENGN	.845	7.589E-02
ARMLN	.834	.204
SITHT	.640	.258
KNEEGRT	.141	.832
ELBOGRT	7.456E-02	.856
WRSTGRT	.193	.818

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

Communalities

	Initial	Extraction
STATURE	1.000	.894
HIPHT	1.000	.825
LOWLENGN	1.000	.720
ARMLN	1.000	.737
SITHT	1.000	.477
KNEEGRT	1.000	.711
ELBOGRT	1.000	.738
WRSTGRT	1.000	.707

Extraction Method: Principal Component Analysis.

The 2 factors with $\lambda > 1.00$ accounted for 72.6% of the variance in the 8 variables.

The rotated solution shows that expected structure -- a "length" group factor and a ”girth" group factor.

The communalities show that the variables are "well accounted for" except for sitting height. Possible explanation? The sample includes both men and women, which differ more on "back length" than any of the other variables. So, that variability might be harder to "account for" using the resulting factors.

The language of factor interpretation:

Describing Each Factor

- One set of descriptive words has to do with how many variables comprise the factor
  - general factor - - All or nearly all variables load
  - group factor - - Subset of variables load
  - unique factor - - Single item loads
  - undefined factor - - "Not sure" what the factor "is" because of few or "odd" loadings

- Another set of descriptive words has to do with the "polarity" of the factor
  - Unipolar factor - - All variables that load on the factor have positive structure weights (loadings). Usually all the loading variables are positively correlated with the factor because they are positively correlated with each other
  - Bipolar factor - - Some variables load with positive weights (e.g., < .40) while other variables load with negative weights (e.g., < -.40). Usually the loading variables have a mixture of positive and negative correlations with the factor because they have a mixture of positive and negative correlations with each other.

Describing Variables

- univocal item - - variable that loads on only one factor
- multivocal item - - variable that loads on more than one factor

Practice with Factor Interpretation:

In a "second study" a few variables were added and the factoring re-done.

Factor Matrix:

		FACTOR 1	FACTOR 2	FACTOR 3
STATURE	(standing height)	.92991	-.24829	.32467
BODYFAT	(% body fat)	.52453	-.31432	.64561
LOWLEGLN	(from knee to floor)	.74915	-.24610	.32413
WRSTGRT	(girth around waist)	.63449	.57476	-.21345
ARMLN	(arm length)	.80174	-.14265	-.01232
KNEEGRT	(girth around knee)	.75599	.59979	.21112
HIPHT	(from hip to floor)	.81381	-.35745	.21222
ELBOGRT	(girth around elbow)	.75446	.66015	.22312
WEIGHT	(body weight)	.53148	.32498	.51322
TONE	(index of fitness)	.16547	.12267	-.63671
SITHT	(sitting height)	.60619	-.01780	-.21368

Interpret this factor solution with a cutoff of .4 -- using the appropriate language form above.

Consider a cutoff .3 -- what factors "change" -- which cutoff do you like better? Why ?