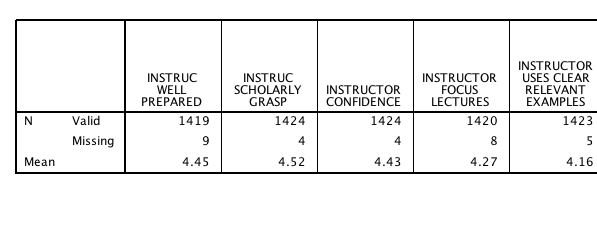
# Complete Example PCA (Factor and SPSS)

**Research Question:** Is there an underlying pattern to the way that students answer questions on their teaching evaluations?

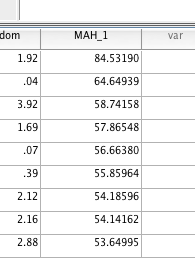
* What components influence their answering?

Assumption Checks:

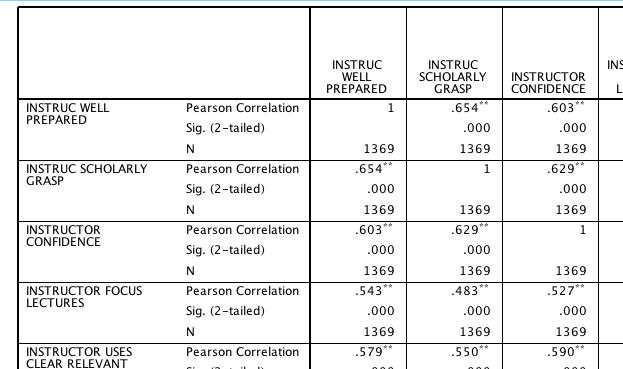
1. Number of variables – There are 12 questions on the end of year evaluations.
2. Sample Size – we have a large sample = N = 1428
3. Missing data – we do have lots of missing data.
   1. You can mean replace or use a different replacement method.



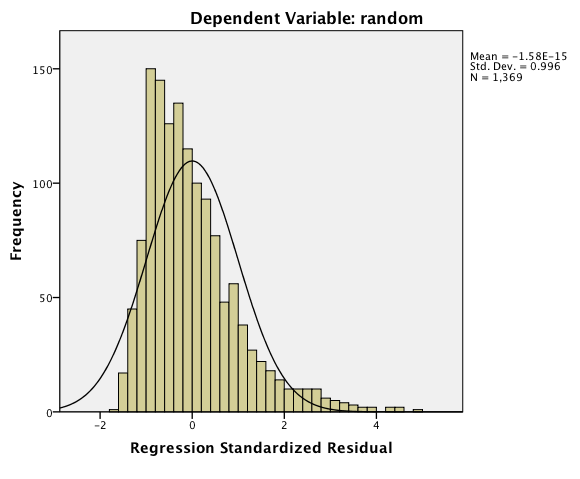
1. Outliers – in this case I will have to create a random variable to figure out the outliers.
   1. Transform > compute.
   2. Label the variable something you will remember (random).
   3. Use a random number generator (like RV ChiSq) – don’t forget you need a number to fill in the question mark.
   4. Hit ok.
   5. Run a fake regression to get your Mahalanobis values.
      1. Analyze > Regression > Linear.
      2. Random variable in the DV.
      3. All your variables in the IV.
      4. Save > Mahalanobis.
   6. We have 59 people over our cut off of 32.91 for 12 variables p<.001.
   7. Sometimes outliers don’t do a thing to the fit of the model, sometimes they do. If you have a small sample size and don’t want to eliminate that many outliers, you can test with and without.
   8. With my very large sample size, I’m going to exclude them.



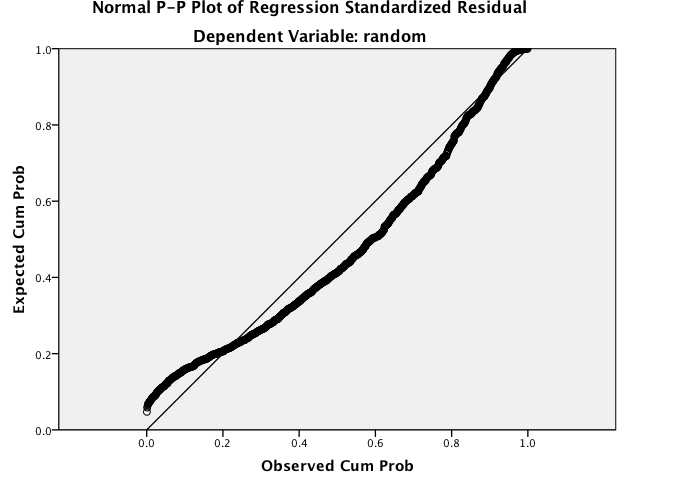
1. Multicollinearity
   1. Analyze > correlate > bivariate.
   2. Here’s the thing – you **want** them to be correlated – but not so much that mathematically it explodes.
   3. Check for r values over .95 or .99.



1. Normality – rerun your fake regression and ask for the residual plots (you especially have to rerun if you excluded outliers).
   1. Analyze > regression > linear.
   2. Plots > ZPRED in Y, ZRESID in X, normality pp, histogram.
   3. Not 100% normal, but we have a lot of participants, so it should be ok.

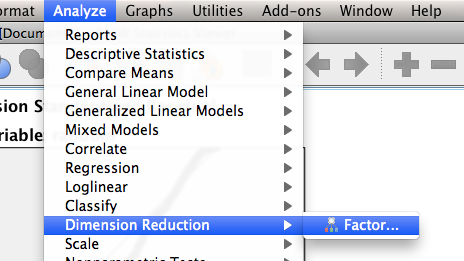


1. Linearity

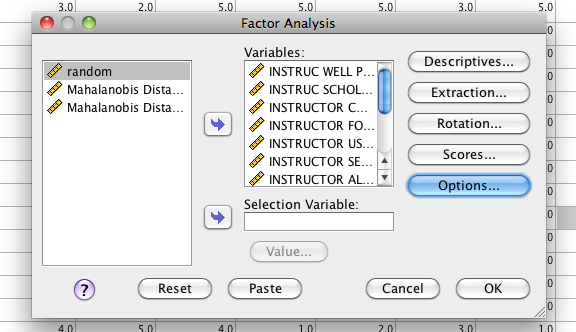


**How to Run Analysis SPSS:**

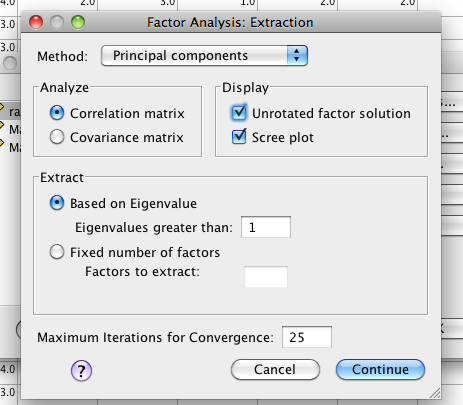
1. Analyze > Dimension Reduction > Factor.



1. Put all your variables in the variables box.



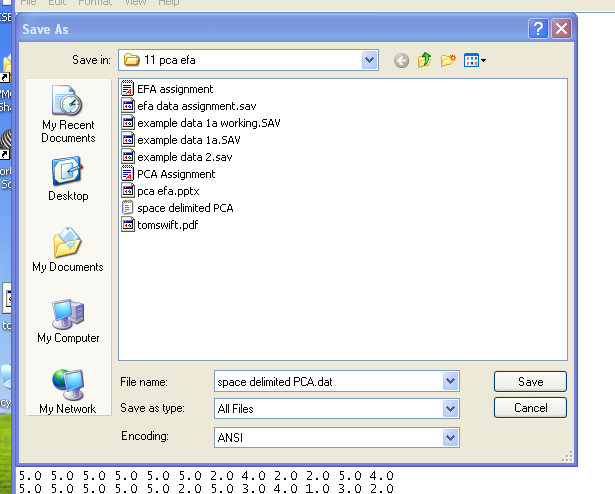
1. Hit extraction.
   1. For PCA – make sure the method says Principal components.
   2. Check for a scree plot – to help you decide on the number of factors.
   3. If you do not know how many factors you want to run, just let it decide based on Eigenvalues. You can change it later.
   4. If you do know the number of factors you want, select the second option.



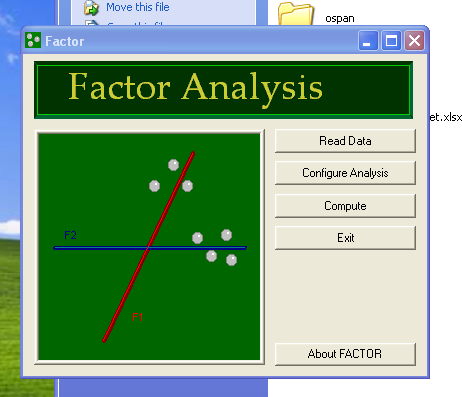
1. PCA does not rotate, so you do not need the rotation options.

**How to run in Factor:**

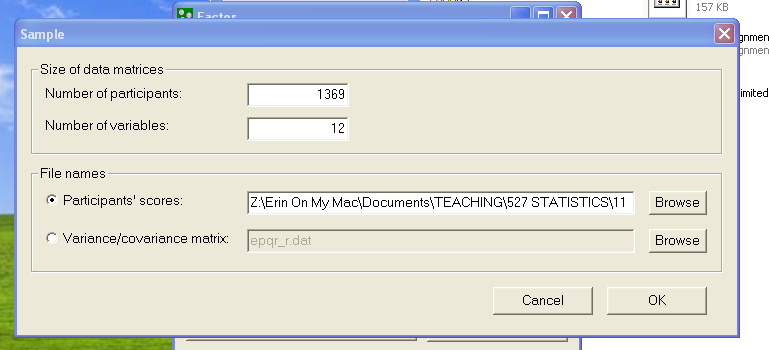
1. First you need to save your data as text.
2. You can do File > Save as > Tab Delimited.



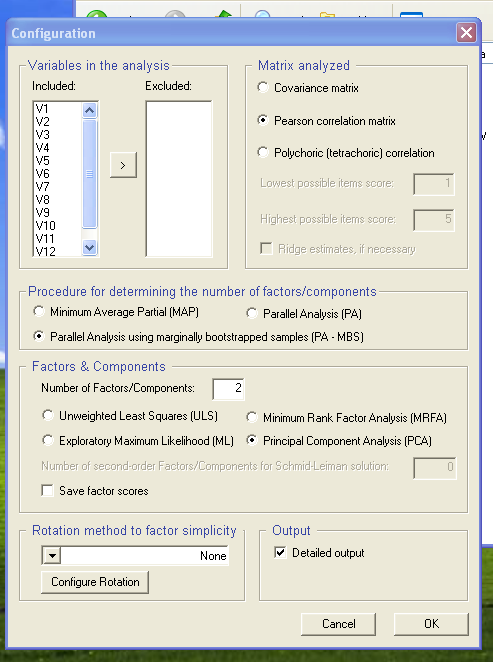
1. Open Factor.



1. First, hit read data.
2. Enter the number of ROWS in the dataset for participants. How many people did you have?
3. Enter the number of COLUMNS for variables. How many questions did you have?

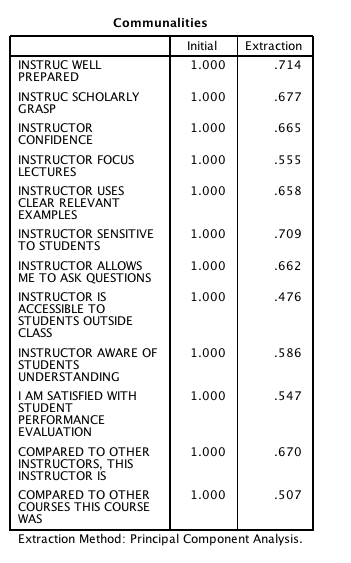


1. Hit ok. If you’ve done this right – the front screen of Factor should say “Ready!”.
2. Hit configure analysis.
3. Variable screen:
   1. Variables – the most confusing part to people is this variable side. They are all included automatically. If you move them to the right, you will leave them out. This set up is backwards from SPSS.
   2. Procedures for factors components: Leave on Parallel Analysis Bootstrapped. This option will tell you how many components are greater than chance.
   3. Under Factors/Components be sure to pick Principal Components Analysis!
   4. Under rotation – pick none so you don’t confuse yourself when looking at the output.
   5. Hit Ok.
4. Hit compute – you’ll get an open text file when done.

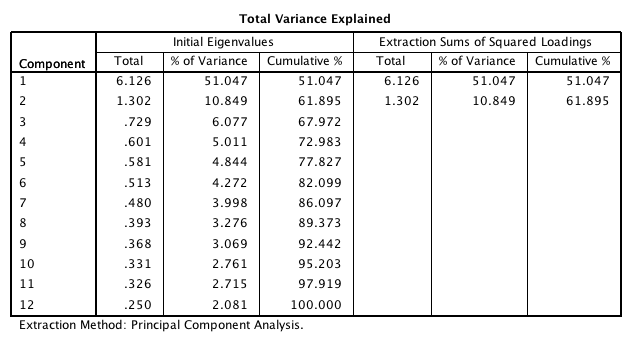
****

**Reading the Output SPSS:**

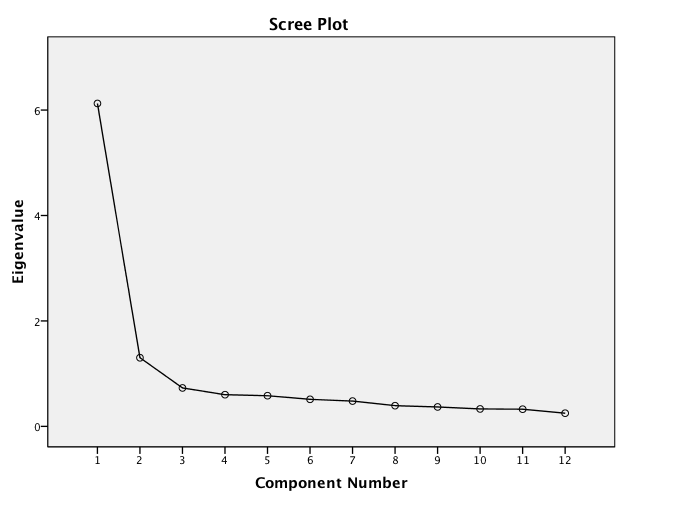
1. Communalities – communality is the amount of variance accounted for in that question alone. We got 71% of the variance covered in our well prepared variable.



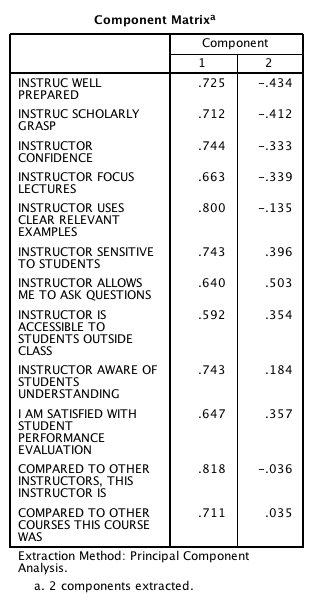
1. Total variance explained – This box gives you the eigenvalues – you well get as many eigenvalues as you have questions.
   1. The “total” column is the eigenvalue – meaning the grouping of the variance that is the highest.
   2. The % variance is the amount of variance taken up by just that component.
   3. Cumulative is total variance of those eigenvalues added up in order.

****

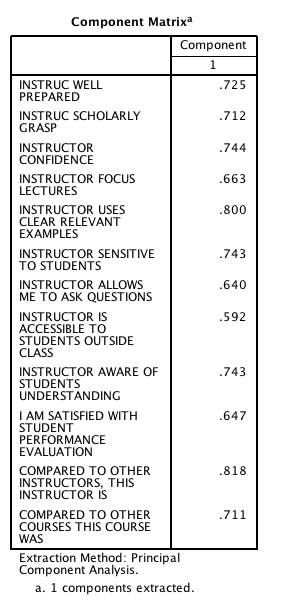
1. The scree plot is a visual depiction of the eigenvalues. You use this plot to figure out how many components to run. For example, there’s a very large drop off after 1 eigenvalue, so I’d probably use one component.

****

1. Component Matrix – is the important box! You will look here for loadings. Here are the rules again:
   1. You want things to load over >.300.
   2. You want things to not split components or have two loadings over >.300 (eek!).
   3. You want things to load on at least one component.
   4. Our problem here is that everything loads on the first component, so when we include the second one, they start to split. This pattern is a strong indication that there’s only one component to answering teaching eval questions.

****

1. Unless everything goes perfectly, you’ll probably have to run at least one or two more models. Here I’m going to run the 1 component model due to the scree plot and the matrix loadings. (in both programs). I’d only report the final model however.



**How to read the Output Factor:**

1. A lot of this output will be redundant / the same. The first thing you want to find is the parallel analysis.

PARALLEL ANALYSIS (PA - MBS)

Lattin, Carroll, & Green (2003)

Variable Real-data Mean of random 95 percentile of random

eigenvalues eigenvalues eigenvalues

1 6.12504\* 1.15439 1.19081

2 1.30208\* 1.11437 1.14266

3 0.72941 1.08521 1.10791

4 0.60141 1.05790 1.07796

5 0.58121 1.03247 1.05028

6 0.51270 1.00921 1.02636

7 0.47983 0.98701 1.00526

8 0.39305 0.96336 0.98055

9 0.36828 0.94004 0.95903

10 0.33142 0.91439 0.93370

11 0.32589 0.88777 0.91106

12 0.24969 0.85388 0.88118

\* Advised number of dimensions: 2

1. You are looking here for the advised number of dimensions. What a parallel analysis tests is a fake sample to see if the eigenvalues are greater than chance. Here it actually suggests two, but as seen in the SPSS output above, that was kind of a mess.
2. Next look for Fit Indices (root mean squared residual – sometimes listed as RMSR mostly listed as RMR in reporting). It’s at the bottom.

Root Mean Square of Residuals (RMSR) = 0.0602

* 1. RMR <.10 is Ok fit.
  2. RMR <.05 is good fit.
  3. We’re close, but the loading matrix was still bad with two factors.

1. Rerun the analysis and change to 1 component.
2. Look at the fit indices then – the Parallel analysis will not change.
   1. Root Mean Square of Residuals (RMSR) = 0.0877
   2. RMR almost always goes DOWN with more components (because you are accounting for more variance), so the decrease didn’t put me above .10, which I would say is good business.

**Example Write Up:**

Results

A Principal Components Analysis was used to analyze end of term teaching evaluations with SPSS and FACTOR. We examined these questions to understand the pattern of answers given to teaching evaluations and to see if those patterns produced distinct components. Data were screened for multivariate assumptions, and 59 participants were excluded as multivariate outliers (Mahalanobis X2(12) > 32.91) leaving 1,369 participants in the analysis.

A parallel analysis indicated that two components were greater than chance, while a scree plot examination indicated that only one component might best it the data. Both one-component and two-component models were run. Table 1 shows the component loadings for the one-component model. All questions loaded highly on the first component, and when a second component was added, questions began to split loadings. The RMR was better for the two-component model (0.06) than the one-component model (0.09), but there was not a significant loss of fit by eliminating the component. The one-component model indicated that participants used the same cognitive processes to make teaching evaluation ratings across all questions and did not have multiple patterns to answering questions.

Table 1. *Component Loadings for One-Component Model*

|  |  |
| --- | --- |
| Question | Loading |
| Instructor Well Prepared | 0.725 |
| Scholarly Grasp | 0.712 |
| Confidence | 0.744 |
| Focus Lectures | 0.663 |
| Clear Relevant Examples | 0.800 |
| Sensitive To Students | 0.743 |
| Allows Me to Ask Questions | 0.640 |
| Accessible Outside Class | 0.592 |
| Aware of Student Understanding | 0.743 |
| Satisfied with Student Performance Evaluations | 0.647 |
| Compared to Other Instructors | 0.818 |
| Compared to Other Courses | 0.711 |
|  |  |