

STAT660/FES758b
Multivariate Statistics
Homework #6 OPTION B: Factor Analysis
Due : Wednesday, 4/25/18 – Submit on CANVAS by midnight

You should answer the questions below for the Ohio Crime Dataset, OR you should analyze them for your own data (all the better!). You may need to modify the questions I ask as appropriate - use your good statistical judgment. There are examples in the notes of how to use SAS, SPSS, and R for Factor Analysis.

A 1996 mail survey in Ohio evaluated attitudes towards crime, treatment of offenders, as well as other demographic/attitudinal information. We will use factor analysis to evaluate 3 sets of questions asked as part of the survey to identify latent factors. The following files contain information for this analysis :

Ohiocrime.xls : contains data for 43 questions. Each question is measured on a six point scale.

OhioCrimeQuestions.pdf : Word document containing information on questions (i.e. variable labels and descriptions, more information on conduct of survey)

OhioCrime.pdf (optional) : a more complete description of the dataset, with information on other questions, etc. that we will not use in this analysis.

Your goal is to identify latent factors which explain correlations observed between relevant variables.

- 1) Look through indicators (questions). Think about which indicators might be related through latent factors. (nothing to turn in here)
- 2) Compute the correlation matrix between all indicators (you may want to do this in batches). Comment on relationships you do/do not observe.
- 3) Compute KMO or other measure (i.e. just look at matrix produced above) to comment on suitability of data for factor analysis.
- 4) Use Principle Components (or appropriate option in Factor Analysis) to decide on a number of latent factors. You can use Scree Plot, eigenvalue>1, or parallel analysis.

NOTE : I DO NOT recommend eigenvalue>1 for the crime attitude data set. Use scree plot elbow or parallel analysis. The relevant parallel analysis values are

provided below (based on 488 observations without missing data and 43 variables) :

LONGMAN	ALLEN
1.592104	1.715118
1.519501	1.656772
1.473414	1.614956
1.443388	1.57978
1.406284	1.550125
1.373723	1.5253
1.344394	1.50854
1.316281	1.498921
1.292872	1.486652
1.265411	1.479616

- 5) Perform a series of factor analyses using orthogonal models. First, try at least two extraction methods (choose from Principle Components, Principle Axis Factoring, Iterative Principle Components, Maximum Likelihood). Use some method for comparing extraction methods to choose a 'best' method (i.e. RMSR or # residuals greater than .05).
- 6) Once you've chosen an extraction method, try a varimax and/or a quartimax rotation. Pick one of these rotations and discuss the interpretation of the final factors. Make one or more loading plots as appropriate.

Note : a loading plot may aid in deciding which variables load more heavily on which variables.