

# CHAPTER 1

---

## FACTOR ANALYSIS

Factor Analysis is about looking for underlying *relationships* or *associations*. In that way, factor analysis is a correlational study of variables, aiming to group or cluster variables along dimensions. It may also be used to provide an estimate (factor score) of a latent construct which is a linear combination of variables. For example, a standardized test might ask hundreds of questions on a variety of quantitative and verbal subjects. Each of these questions could be viewed as a variable. However, the quantitative questions collectively are meant to measure some *latent* factor, that is the individual's *quantitative reasoning*. A Factor Analysis might be able to reveal these two latent factors (quantitative reasoning and verbal ability) and then also provide an estimate (score) for each individual on each factor.

Any attempt to use factor analysis to summarize or reduce a set of data should be based on a conceptual foundation or hypothesis. It should be remembered that factor analysis will produce factors for most sets of data. Thus, if you simply analyze a large number of variables in the hopes that the technique will "figure it out", your results may look as though they are grasping at straws. The quality or meaning/interpretation of the derived factors is best when related to a conceptual foundation that existed prior to the analysis.

### 1.1 Assumptions of Factor Analysis

1. No outliers in the data set
2. Adequate sample size

- As a rule of thumb, maintain a ratio of variables to factors of at least 3 (some say 5). This depends on the application.
  - You should have at least 10 observations for each variable (some say 20). This often depends on what value of factor loading you want to declare as significant. See Table 1.1 for the details on this.
3. No perfect multicollinearity
  4. Homoskedasticity *not* required between variables
    - All variances not required to be equal
  5. Linearity of variables desired - only models linear correlation between variables
  6. Interval data (as opposed to nominal)
  7. Measurement error on the variables/observations has constant variance and is, on average, 0
  8. Normality is not required

Sample Size Needed for Significance	Factor Loading
350	.30
250	.35
200	.40
150	.45
120	.50
100	.55
85	.60
70	.65
60	.70
50	.75

Table 1.1: Factor loadings are the correlation of each variable and the factor. This table is a guide for the sample sizes necessary to consider a factor loading significant. For example, in a sample of 100, factor loadings of 0.55 are considered significant. In a sample size of 70, however, factor loadings must reach 0.65 to be considered significant. Significance based on 0.05 level, a power level of 80 percent. Source: *Computations made with SOLO Power Analysis, BMDP Statistical Software, Inc., 1993*

## 1.2 Determining Factorability

Before we even begin the process of factor analysis, we have to do some preliminary work to determine whether or not the data even lends itself to this technique. If none of our variables are correlated, then we cannot group them together in any meaningful way! Bartlett's Sphericity Test and the KMO index are two statistical tests for whether or not a set of variables can be factored. These tests *do not* provide information about the appropriate number of factors, only whether or not such factors even exist.

### 1.2.1 Visual Examination of Correlation Matrix

Depending on how many variables you are working with, you may be able to determine whether or not to proceed with factor analysis by simply examining the correlation matrix. With this examination, we are looking for two things:

1. Correlations that are significant at the 0.01 level of significance. At least half of the correlations should be significant in order to proceed to the next step.
2. Correlations are "sufficient" to justify applying factor analysis. As a rule of thumb, at least half of the correlations should be greater than 0.30.

### 1.2.2 Barlett's Sphericity Test

Barlett's sphericity test checks if the observed correlation matrix is significantly different from the identity matrix. Recall that the correlation of two variables is equal to 0 if and only if they are orthogonal (and thus completely uncorrelated). When this is the case, we cannot reduce the number of variables any further, neither PCA nor Factor Analysis will be able to compress the information reliably into fewer dimensions. For Barlett's test,

$$H_0 = \text{The variables are orthogonal}$$

Which implies that there are no underlying factors to be uncovered. Obviously, we must be able to reject this hypothesis for a meaningful result in PCA.

### 1.2.3 Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy

The goal of the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is similar to that of Bartlett's test in that it checks if we can factorize efficiently the original variables. However, the KMO measure is based on the idea of *partial correlation*. The correlation matrix is always the starting point. We know that the variables are more or less correlated, but the correlation between two variables

can be influenced by the others. So, we use the partial correlation in order to measure the relation between two variables by removing the effect of the remaining variables. The KMO index compares the raw values of correlations between variables and those of the partial correlations. If the KMO index is high ( $\approx 1$ ), then PCA can act efficiently; if the KMO index is low ( $\approx 0$ ), then PCA is not relevant. Generally a KMO index greater than 0.5 is considered acceptable to proceed with factor analysis. Table 1.2 contains the information about interpreting KMO results that was provided in the original 1974 paper.

KMO value	Degree of Common Variance
0.90 to 1.00	Marvelous
0.80 to 0.89	Middling
0.60 to 0.69	Mediocre
0.50 to 0.59	Miserable
0.00 to 0.49	Don't Factor

Table 1.2: Interpreting the KMO value. [?]

So, for example, if you have a survey with 100 questions/variables and you obtained a KMO index of 0.61, this tells you that the degree of common variance between your variables is mediocre, on the border of being miserable. While factor analysis may still be appropriate in this case, you will find that such an analysis will not account for a substantial amount of variance in your data. It may still account for enough to draw some meaningful conclusions, however.

### 1.3 Communalities

You can think of **communalities** as multiple  $R^2$  values for regression models predicting the variables of interest from the factors (the reduced number of factors that your model uses). The communality for a given variable can be interpreted as the proportion of variation in that variable explained by the chosen factors.

Take for example the SAS output for factor analysis on the Iris dataset shown in Figure 1.1. The factor model (which settles on only one single factor) explains 98% of the variability in *petal length*. In other words, if you were to use this factor in a simple linear regression model to predict petal length, the associated  $R^2$  value should be 0.98. Indeed you can verify that this is true. The results indicate that this single factor model will do the best job explaining variability in *petal length*, *petal width*, and *sepal length*.

Factor Pattern	
	Factor1
Sepal_Width	-0.46014
Sepal_Length	0.89017
Petal_Width	0.96498
Petal_Length	0.99156

  

Variance Explained by Each Factor
Factor1
2.9184978

  

Final Communality Estimates: Total = 2.918498			
Sepal_Width	Sepal_Length	Petal_Width	Petal_Length
0.21173131	0.79240043	0.93118439	0.98318168

Figure 1.1: SAS output for PROC FACTOR using Iris Dataset

One assessment of how well a factor model is doing can be obtained from the communalities. What you want to see is values that are close to one. This would indicate that the model explains most of the variation for those variables. In this case, the model does better for some variables than it does for others.

If you take all of the communality values,  $c_i$  and add them up you can get a total communality value:

$$\sum_{i=1}^p \hat{c}_i = \sum_{i=1}^k \hat{\lambda}_i$$

Here, the total communality is 2.918. The proportion of the total variation explained by the three factors is

$$\frac{2.918}{4} \approx 0.75.$$

The denominator in that fraction comes from the fact that the correlation matrix is used by default and our dataset has 4 variables. Standardized variables have

variance of 1 so the total variance is 4. This gives us the percentage of variation explained in our model. This might be looked at as an overall assessment of the performance of the model. The individual communalities tell how well the model is working for the individual variables, and the total communality gives an overall assessment of performance.

## 1.4 Number of Factors

A good rule of thumb for determining the number of factors is to only choose factors with associated eigenvalue (or variance) greater than 1. Since the correlation matrix is used for factor analysis, we want our factors to explain more variance than any individual variable from our dataset. If this rule of thumb produces too many factors, it is reasonable to raise that limiting condition only if the number of factors still explains a reasonable amount of the total variance.

## 1.5 Rotation of Factors

The purpose of rotating factors is to make them more interpretable. If factor loadings are relatively constant across variables, they don't help us find latent structure or clusters of variables. This will often happen in PCA when the goal is only to find directions of maximal variance. Thus, once the number of components/factors is fixed and a projection of the data onto a lower-dimensional subspace is done, we are free to rotate the axes of the result without losing any variance. The axes will no longer be principal components! The amount of variance explained by each factor will change, but the total amount of variance in the reduced data will stay the same because all we have done is rotate the basis. The goal is to rotate the factors in such a way that the loading matrix develops a more *sparse* structure. A sparse loading matrix (one with lots of very small entries and few large entries) is far easier to interpret in terms of finding latent variable groups.

The two most common rotations are **varimax** and **quartimax**. The goal of *varimax* rotation is to maximize the squared factor loadings in each factor, i.e. to simplify the columns of the factor matrix. In each factor, the large loadings are increased and the small loadings are decreased so that each factor has only a few variables with large loadings. In contrast, the goal of *quartimax* rotation is to simplify the rows of the factor matrix. In each variable the large loadings are increased and the small loadings are decreased so that each variable will only load on a few factors. Which of these factor rotations is appropriate