

# What is factor analysis and how does it simplify research findings?

**9 min read**

There are many forms of data analysis used to report on and study survey data. Factor analysis is best when used to simplify complex data sets with many variables

# What is factor analysis?

Factor analysis is the practice of condensing many variables into just a few, so that your research data is easier to work with.

The theory is that there are deeper factors driving the underlying concepts in your data, and that you can uncover and work with these instead of dealing with the lower-level variables that cascade from them. Factor analysis is also sometimes called “dimension reduction.” You can reduce the “dimensions” of your data into one or more “super-variables,” also known as unobserved variables or latent variables.

These deeper concepts aren't immediately obvious. They might represent traits or tendencies that are hard to measure, such as extraversion or IQ.

As with any kind of process that simplifies complexity, there is a trade-off between the accuracy of the data and how easy it is to work with. With factor analysis, the best solution is the one that yields a simplification that represents the true nature of your data, with minimum loss of precision.

Factor analysis isn't a single technique, but a family of statistical methods that can be used to identify the latent factors driving observable variables. Factor analysis is commonly used in [market research](#), as well as other disciplines like technology, medicine, sociology, field biology, education, psychology and many more.

[Free eBook: 2022 global market research trends report](#)

## Key concepts in factor analysis

One of the most important ideas in factor analysis is **variance** – how much your numerical values differ from the average. When you perform factor analysis, you're looking to understand how the different underlying factors influence the variance among your variables. Every factor will have an influence, but some will explain more variance than others, meaning that the factor more accurately represents the variables it's comprised of.

The amount of variance a factor explains is expressed in an **eigenvalue**. If a factor solution has an eigenvalue of 1 or above, it explains more variance than a single observed variable – which means it can be useful to you in

cutting down your number of variables. Factor solutions with eigenvalues less than 1 account for less variability than a single variable and are not retained in the analysis. In this sense, a solution would contain fewer factors than the original number of variables.

Another important metric is **factor score**. This is a numerical measure that describes how strongly a variable from the original research data is related to a given factor. Another term for this association or weighting towards a certain factor is **factor loading**.

## Types of factor analysis

There are two basic forms of factor analysis, exploratory and confirmatory. Here's how they are used to add value to your research process.

### Confirmatory factor analysis

In this type of analysis, the researcher starts out with a hypothesis about their data that they are looking to prove or disprove. Factor analysis will confirm – or not – where the latent variables are and how much variance they account for.

Principal component analysis is a popular form of confirmatory factor analysis. Using this method, the researcher will run the analysis to obtain multiple possible solutions that split their data among a number of factors. Items that load onto a single factor are more strongly related to one another and can be grouped together by the researcher using their conceptual knowledge.

Using PCA will generate a range of solutions with different numbers of factors, from simplified 1-factor solutions to higher levels of complexity. However, the fewer number of factors employed, the less variance will be accounted for in the solution.

### Exploratory factor analysis

As the name suggests, exploratory factor analysis is undertaken without a hypothesis in mind. It's an investigatory process that helps researchers understand whether associations exist between the initial variables, and if so, where they lie and how they are grouped.

# How to perform factor analysis

Most major statistical software packages, such as SPSS and Stata, include a factor analysis function that you can use to analyze your data. To get started, you will need the variables you are interested in and, if applicable, details of your initial hypothesis about their relationships and underlying variables.

## How factor analysis can help you

As well as giving you fewer variables to navigate, factor analysis can help you understand grouping and clustering in your input variables, since they'll be grouped according to the latent variables.

Say you ask several questions all designed to explore different, but closely related, aspects of [customer satisfaction](#):

- 1 How satisfied are you with our product?
- 2 Would you recommend our product to a friend or family member?
- 3 How likely are you to purchase our product in the future?

But you only want one variable to represent a customer satisfaction score. One option would be to average the three question responses. Another option would be to create a factor dependent variable. This can be done by running PCA and keeping the first Principal Component (also known as a factor). The advantage of PCA over an average is that it automatically weights each of the variables in the calculation.

Say you have a list of questions and you don't know exactly which responses will move together and which will move differently; for example, purchase barriers of potential customers. The following are possible barriers to purchase:

- + Price is prohibitive
- + Overall implementation costs
- + We can't reach a consensus in our organization
- + Product is not consistent with our business strategy
- + I need to develop an ROI, but cannot or have not

- + We are locked into a contract with another product
- + The product benefits don't outweigh the cost
- + We have no reason to switch
- + Our IT department cannot support your product
- + We do not have sufficient technical resources
- + Your product does not have a feature we require
- + Other (please specify)

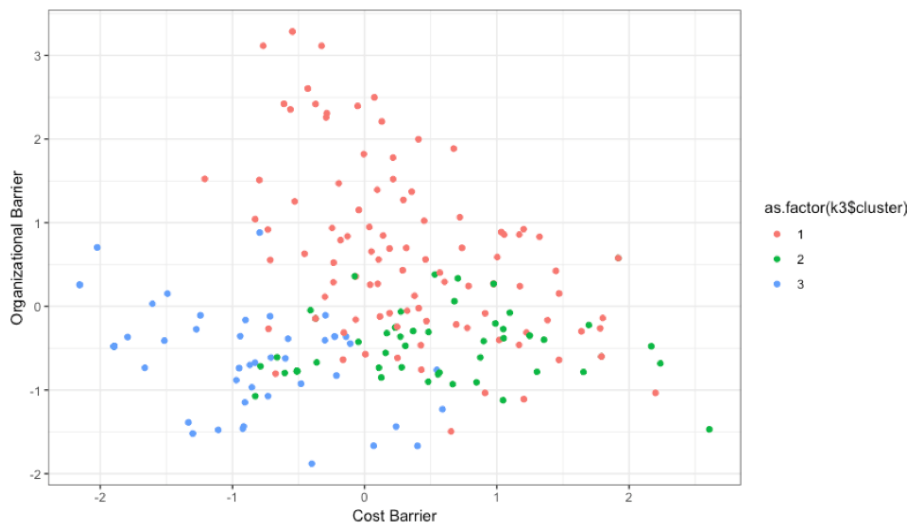
Factor analysis can uncover the trends of how these questions will move together. The following are loadings for 3 factors for each of the variables.

	PC1	PC2	PC3
Price is prohibitive	0.49	-0.16	0.13
Overall implementation costs	0.41	0.38	0.14
We can't reach a consensus in our organization	-0.03	-0.07	0.57
Product is not consistent with our business strategy	0.01	-0.03	0.45
I need to develop an ROI, but cannot or have not	-0.06	0.17	0.62
We are locked into a contact with another product	-0.21	0.08	-0.27
The product benefits don't outweigh the cost	0.68	0.06	-0.27
We have no reason to switch	-0.07	-0.77	-0.25
Our IT department cannot support your product	0.00	0.38	-0.25
We do not have sufficient technical resources	-0.21	0.58	-0.12
Your product does not have a feature we require	-0.40	0.05	0.08
Other (please specify)	-0.50	-0.06	-0.10

Notice how each of the principal components have high weights for a subset of the variables. The first component heavily weights variables related to cost, the second weights variables related to IT, and the third weights variables related to organizational factors. We can give our new super variables clever names.

	Cost	IT	Org
The product benefits don't outweigh the <b>cost</b>	0.68	0.06	-0.27
<b>Price</b> is prohibitive	0.49	-0.16	0.13
Overall implementation <b>costs</b>	0.41	0.38	0.14
We do not have sufficient <b>technical resources</b>	-0.21	0.58	-0.12
Our <b>IT department</b> cannot support your product	0	0.38	-0.25
Product is not consistent with our <b>business strategy</b>	0.01	-0.03	0.45
We can't reach a consensus in our <b>organization</b>	-0.03	-0.07	0.57
I need to develop an <b>ROI</b> , but cannot or have not	-0.06	0.17	0.62
Your product does not have a feature we require	-0.4	0.05	0.08
We are locked into a contract with another product	-0.21	0.08	-0.27
Other (please specify)	-0.5	-0.06	-0.1
We have no reason to switch	-0.07	-0.77	-0.25

If we were to cluster the customers based on these three components, we can see some trends. Customers tend to be high in Cost barriers or Org barriers, but not both.



## Examples of factor analysis studies

Factor analysis, including PCA, is often used in tandem with [segmentation studies](#). It might be an intermediary step to reduce variables before using KMeans to make the segments.

Factor analysis provides simplicity after reducing variables. For long studies with large blocks of Matrix Likert scale questions, the number of variables can become unwieldy. Simplifying the data using factor analysis helps analysts focus and clarify the results, while also reducing the number of dimensions they're clustering on.

## Sample questions

Choosing exactly which questions to perform factor analysis on is both an art and a science. Choosing which variables to reduce takes some experimentation, patience and creativity. Factor analysis works well on Likert scale questions and Sum to 100 question types.

Factor analysis works well on matrix blocks of the following question genres:

Psychographics (Agree/Disagree):

- + I value family
- + I believe brand represents value

Behavioral (Agree/Disagree):

- + I purchase the cheapest option
- + I am a bargain shopper

Attitudinal (Agree/Disagree):

- + The economy is not improving
- + I am pleased with the product

Activity-Based (Agree/Disagree):

- + I love sports
- + I sometimes shop online during work hours

Behavioral and psychographic questions are especially suited for factor analysis.

## Sample output reports

Factor analysis simply produces weights (called loadings) for each respondent. These loadings can be used like other responses in the survey.