P: ISSN No. 0976-8602 RNI No.UPENG/2012/426228 VOL.-III, ISSUE-IV, October-2014
E: ISSN No. 2349 - 9443

Asian Resonance

Factor Analysis: A Tool For Research

Abstract

'Factor analysis' is a useful tool for investigating variable relationships for complex concepts such as socioeconomic status, dietary patterns, or psychological scales.

It allows researchers to investigate concepts that are not easily measured directly by collapsing a large number of variables into a few interpretable underlying factors.

Praveen Sahu
Assistant Professor,
Deptt.of Commerce and Management
V.S. Govt. P.G. College,
Gwalior,

Kaushlendra Awasthi
Assistant Professor,
Madhav College,
Gwalior

Keywords: Research, Variables, Factor Analysis.

Introduction

Factor Analysis – An Introduction

Factor analysis is the name given to a group of statistical techniques that can be used to analyze interrelationships among a large number of variables and to explain these variables in terms of their common underlying dimensions (factors). The approach involves condensing the information contained in a number of original variables into a smaller set of dimensions (factors) with a minimum loss of information In more technical terms Factor analysis addresses the problem of analyzing the structure of the interrelationships (correlations) among a large number of variables (e.g., test scores, test items, questionnaire responses) by defining a set of common underlying dimensions, known as factors.

Objectives

Factor analysis is used mostly for data reduction purposes :

1. To get a small set of variables (preferably uncorrelated) from a large set of variables (most of which are correlated to each other)
2. To create indexes with variables that measure similar things (conceptually).

Review of Literature

Paatero, P. and Tapper, U., 1993. Analysis of different modes of factor analysis as least squares fit problems. Chemometrics and intelligent Laboratory Systems, 18: 183–194.

It is shown that each mode of principal component analysis or 'factor analysis' is equivalent to solving a certain least squares problem where certain error estimators σ_{ij} are assumed for the measured data matrixX_{ij}. Selecting the mode (e.g. Q-mode) implicitly selects a scaling transformation as a preparatory step. Each scaling corresponds optimally to a certain σ. It is shown that the customary modes (Q-mode and R-mode) corresponds to such error estimates which do not normally occur in chemistry or physics. The best posssible scaling ('optimal scaling') and a near-optimal scaling are introduced. The Quail Roost II air pollution simulation data sets are studied as examples:

Evolving factor analysis, a new multivariate technique in chromatographyMarcel Maeder, Arne Zilian Overlapping peaks are a general problem in chromatography. Modern multichannel detectors such as the diode-array detector allow multivariate techniques for a computational resolution. Evolving factor analysis (Efa) is a recently developed method for a completely model-free resolution of overlapping peaks into concentration profiles and absorption spectra. Efa is successfully tested with real chromatograms. The requirements concerning the quality of the measured data are discussed and related to the scope and fields of application of Efa.

Warne, R. T., & Larsen, R. (2014). Evaluating a proposed modification of the Guttman rule for determining the number of factors in an exploratory factor analysis.Psychological Test and Assessment Modeling, 56, 104-123. Ritter, N. (2012). A comparison of distribution-free and non-distribution free methods in factor analysis. Paper presented at Southwestern Educational Research Association (Sera) Conference 2012, New Orleans, LA (ED529153).

Subbarao, C.; Subbarao, N.V.; Chandu, S.N. (December 1996). "Characterisation of groundwater contamination using factor analysis". Environmental Geology 28 (4):175 180. doi:10.1007/s002540050091.

What is a Factor ?

The key concept of factor analysis is that multiple observed variables have similar patterns of responses because of

For example, people may respond similarly to questions about income, education, and occupation, which are all associated with the latent variable socioeconomic status.

In every factor analysis, there is the same number of factors as there are variables. Each factor captures a certain amount of the overall variance in the observed variables, and the factors are always listed in order of how much variation they explain.

The eigen value is a measure of how much of the variance of the observed variables a factor explains.

Any factor with an eigenvalue ≥1 explains more variance than a single observed variable.

So if the factor for socioeconomic status had an eigenvalue of 2.3 it would explain as much variance as 2.3 of the three variables. This factor, which captures most of the variance in those three variables, could then be used in other analyses.

The factors that explain the least amount of variance are generally discarded. Deciding how many factors are useful to retain will be the subject of another post.

Two Types of Factor Analysis

Exploratory

It is exploratory when you do not have a pre-defined idea of the structure or how many dimensions are in a set of variables.

Confirmatory

It is confirmatory when you want to test specific hypothesis about the structure or the number of dimensions underlying a set of variables (i.e. in your data you may think there are two dimensions and you want to verify that).

Basic Idea of Factor Analysis as a Data Reduction Method

Suppose we conducted a study in which we measure 100 people's height in inches and centimetres. Thus, we would have two variables that measure height. If in future studies, we want to research, for example, the effect of different nutritional food supplements on height, would we continue to use both measures? Probably not; height is one characteristic of a person, regardless of how it is measured.

Let's now extrapolate from this study to something that you might actually do as a researcher. Suppose we want to measure people's satisfaction with their lives. We design a satisfaction questionnaire with various items; among other things we ask our subjects how satisfied they are with their hobbies (item 1) and how intensely they are pursuing a hobby (item 2). Most likely, the responses to the two items are highly correlated with each other. Given a high correlation between the two items, we can conclude that they are quite redundant.

Combining Two Variables into a Single Factor

Correlation between two variables can be summarized using a scatter plot. A regression line can then be fitted that represents the ‘best’ summary of the linear relationship between the variables. If we could define a variable that would approximate the regression line in such a plot, then that variable would capture most of the ‘essence’ of the two items. Subjects' single scores on that new factor, represented by the regression line, could then be used in future data analyses to represent that essence of the two items. In a sense we have reduced the two variables to one factor. Note that the new factor is actually a linear combination of the two variables.

Principal Components Analysis

The example described above, combining two correlated variables into one factor, illustrates the basic idea of factor analysis, or of principal components analysis to be precise (we will return to this later). If we extend the two-variable example to multiple variables, then the computations become more involved, but the basic principle of expressing two or more variables by a single factor remains the same.

Extracting Principal Components

We do not want to go into the details about thecomputational aspects of principal components analysis here, which can be found elsewhere (references were provided at the beginning of this section). However, basically, the extraction of principal components amounts to a variance maximizing (varimax) rotation of the original variable space. For example, in a scatter plot we can think of the regression line as the original X axis, rotated so that it approximates the regression line. This type of rotation is called variance maximizing because the criterion for (goal of) the rotation is to maximize the variance (variability) of the "new" variable (factor), while minimizing the variance around the new variable (see Rotational Strategies).

Generalizing to the Case of Multiple Variables

When there are more than two variables, we can think of them as defining a ‘space,’ just as two variables defined a plane. Thus, when we have three variables, we could plot a three- dimensional scatter plot, and, again we could fit a plane through the data.

With more than three variables it becomes impossible to illustrate the points in a scatter plot, however, the logic of rotating the axes so as to maximize the variance of the new factor remains the same.

Multiple Orthogonal Factors

After we have found the line on which the variance is maximal, there remains some variability around this line. In principal components analysis, after the first factor has been extracted, that is, after the first line has been drawn through the data, we continue and define another line that maximizes the remaining variability, and so on. In this manner, consecutive factors are extracted. Because each consecutive factor is defined to maximize the variability that is not captured by the preceding factor, consecutive factors are independent of each other. Put another way, consecutive factors are uncorrelated or orthogonal to each other.

How many Factors to Extract ?

Remember that, so far, we are considering principal components analysis as a data reduction method, that is, as a method for reducing the number of variables. The question then is, how many factors do we want to extract? Note that as we extract consecutive factors, they account for less and less variability. The decision of when to stop extracting factors basically depends on when there is only very little "random" variability left. The nature of this decision is arbitrary; however, various guidelines have been developed, and they are reviewed in 'Reviewing the Results of a Principal Components Analysis under Eigen values and the Number-of-Factors Problem.'

Reviewing the Results of a Principal Components Analysis

Without further ado, let us now look at some of the standard results from a principal components analysis. To reiterate, we are extracting factors that account for less and less variance. To simplify matters, you usually start with the correlation matrix, where the variances of all variables are equal to 1.0. Therefore, the total variance in that matrix is equal to the number of variables. For example, if we have 10 variables each with a variance of 1 then the total variability that can potentially be extracted is equal to 10 times 1. Suppose that in the satisfaction study introduced earlier we included 10 items to measure different aspects of satisfaction at home and at work. The variance accounted for by successive factors would be summarized as follows

Statistical Factor Analysis	Eigenvalues (factor.sta) Extraction: Principal components			
Value	Eigen value	% Total Variance	Cumulative Eigen value	Cumulative %
1	6.118369	61.18369	6.11837	61.1837
2	1.800682	18.00682	7.91905	79.1905
3	.472888	4.72888	8.39194	83.9194
4	.407996	4.07996	8.79993	87.9993
5	.317222	3.17222	9.11716	91.1716
6	.293300	2.93300	9.41046	94.1046
7	.195808	1.95808	9.60626	96.0626
8	.170431	1.70431	9.77670	97.7670
9	.137970	1.37970	9.91467	99.1467
10	.085334	.85334	10.00000	100.0000

Eigen Values

In the second column (Eigenvalue) above, we find the variance on the new factors that were successively extracted. In the third column, these values are expressed as a percent of the total variance (in this example, 10). As we can see, factor 1 accounts for 61 percent of the variance, factor 2 for 18 percent, and so on. As expected, the sum of the eigenvalues is equal to the number of variables. The third column contains the cumulative variance extracted. The variances extracted by the factors are called the eigenvalues. This name derives from the computational issues involved.

Eigenvalues and the Number-of-Factors Problem

Now that we have a measure of how much variance each successive factor extracts, we can return to the question of how many factors to retain. As mentioned earlier, by its nature this is an arbitrary decision. However, there are some guidelines that are commonly used, and that, in practice, seem to yield the best results.

The Kaiser Criterion

First, we can retain only factors with eigenvalues greater than 1. In essence this is like saying that, unless a factor extracts at least as much as the equivalent of one original variable, we drop it. This criterion was proposed by Kaiser (1960), and is probably the one most widely used. In our example above, using this criterion, we would retain 2 factors (principal components).

The Scree Test

A graphical method is the scree test first proposed by Cattell (1966). We can plot the eigenvalues shown above in a simple line plot.

Which Criterion to Use

Both criteria have been studied in detail (Browne, 1968; Cattell & Jaspers, 1967; Hakstian, Rogers, & Cattell, 1982; Linn, 1968; Tucker, Koopman & Linn, 1969). Theoretically, you can evaluate those criteria by generating random data based on a particular number of factors. You can then see whether the number of factors is accurately detected by those criteria. Using this general technique, the first method (Kaiser Criterion) sometimes retains too many factors, while the second technique (scree test) sometimes retains too few; however, both do quite well under normal conditions, that is, when there are relatively few factors and many cases. In practice, an additional important aspect is the extent to which a solution is interpretable. Therefore, you usually examine several solutions with more or fewer factors, and choose the one that makes the best "sense." We will discuss this issue in the context of factor rotations below.

Principal Factors Analysis

Before we continue to examine the different aspects of the typical output from a principal components analysis, let us now introduce principal factors analysis. Let us return to our satisfaction questionnaire example to conceive of another "mental model" for factor analysis. We can think of subjects' responses as being dependent on two components. First, there are some underlying common factors, such as the "satisfaction-with-hobbies" factor we looked at before. Each item measures some part of this common aspect of satisfaction. Second, each item also captures a unique aspect of satisfaction that is not addressed by any other item.

Communalities

If this model is correct, then we should not expect that the factors will extract all variance from our items; rather, only that proportion that is due to the common factors and shared by several items. In the language of factor analysis, the proportion of variance of a particular item that is due to common factors (shared with other items) is called communality. Therefore, an additional task facing us when applying this model is to estimate the communalities for each variable, that is, the proportion of variance that each item has in common with other items. The proportion of variance that is unique to each item is then the respective item's total variance minus the communality.

A common starting point is to use the squared multiple correlation of an item with all other items as an estimate of the communality. Some authors have suggested various iterative "post-solution improvements" to the initial multiple regression communality estimate; for example, the so-called MINRES method (minimum residual factor method; Harman & Jones, 1966) will try various modifications to the factor loadings with the goal to minimize the residual (unexplained) sums of squares.

Factor Analysis as a Classification Method

Let us now return to the interpretation of the standard results from a factor analysis. We will henceforth use the term factor analysis generically to encompass both principal components and principal factors analysis. Let us assume that we are at the point in our analysis where we basically know how many factors to extract. We may now want to know the meaning of the factors, that is, whether and how we can interpret them in a meaningful manner. To illustrate how this can be accomplished, let us work "backwards," that is, begin with a meaningful structure and then see how it is reflected in the results of a factor analysis. Let us return to our satisfaction example; shown below is the correlation matrix for items pertaining to satisfaction at work and items pertaining to satisfaction at home.

Statistical Factor Analysis	**Correlations (factor.sta) Casewise deletion of MD n=100**					
Variable	**WORK_1**	**WORK_2**	**WORK_3**	**HOME_1**	**HOME_2**	**HOME_3**
WORK_1	1.00	.65	.65	.14	.15	.14
WORK_2	.65	1.00	.73	.14	.18	.24
WORK_3	.65	.73	1.00	.16	.24	.25
HOME_1	.14	.14	.16	1.00	.66	.59
HOME_2	.15	.18	.24	.66	1.00	.73
HOME_3	.14	.24	.25	.59	.73	1.00

The work satisfaction items are highly correlated amongst themselves, and the home satisfaction items are highly intercorrelated amongst themselves. The correlations across these two types of items (work satisfaction items with home satisfaction items) is comparatively small. It thus seems that there are two relatively independent factors reflected in the correlation matrix, one related to satisfaction at work, the other related to satisfaction at home.

Factor Loadings

Let us now perform a principal components analysis and look at the two-factor solution. Specifically, let us look at the correlations between the variables and the two factors (or "new" variables), as they are extracted by default; these correlations are also called factor loadings.

Statistical Factor Analysis	**Factor Loadings (Unrotated) Principal Components**	
Variable	**Factor 1**	**Factor 2**
WORK_1	.654384	.564143
WORK_2	.715256	.541444
WORK_3	.741688	.508212
HOME_1	.634120	-.563123
HOME_2	.706267	-.572658
HOME_3	.707446	-.525602
Expl.Var	2.891313	1.791000
Prp.Totl	.481885	.298500

Apparently, the first factor is generally more highly correlated with the variables than the second factor. This is to be expected because, as previously described, these factors are extracted successively and will account for less and less variance overall.

Rotating the Factor Structure

We could plot the factor loadings shown above in a scatter plot. In that plot, each variable is represented as a point. In this plot we could rotate the axes in any direction without changing the relative locations of the points to each other; however, the actual coordinates of the points, that is, the factor loadings would of course change. In this example, if you produce the plot it will be evident that if we were to rotate the axes by about 45 degrees we might attain a clear pattern of loadings identifying the work satisfaction items and the home satisfaction items.

Rotational Strategies

There are various rotational strategies that have been proposed. The goal of all of these strategies is to obtain a clear pattern of loadings, that is, factors that are somehow clearly marked by high loadings for some variables and low loadings for others. This general pattern is also sometimes referred to as simple structure (a more formalized definition can be found in most standard textbooks). Typical rotational strategies are varimax, quartimax, and equamax.

The idea of the varimax rotation is described before (see Extracting Principal Components), and it can be applied to this problem as well. As before, we want to find a rotation that maximizes the variance on the new axes; put another way, we want to obtain a pattern of loadings on each factor that is as diverse as possible, lending itself to easier interpretation. Below is the table of rotated factor loadings.

Statistical Factor Analysis	Factor Loadings (Varimax normalized) Extraction: Principal components	
Variable	**Factor 1**	**Factor 2**
WORK_1	.862443	.051643
WORK_2	.890267	.110351
WORK_3	.886055	.152603
HOME_1	.062145	.845786
HOME_2	.107230	.902913
HOME_3	.140876	.869995
Expl.Var	2.356684	2.325629
Prp.Totl	.392781	.387605

Interpreting the Factor Structure

Now the pattern is much clearer. As expected, the first factor is marked by high loadings on the work satisfaction items, the second factor is marked by high loadings on the home satisfaction items. We would thus conclude that satisfaction, as measured by our questionnaire, is composed of those two aspects; hence we have arrived at a classification of the variables.

Oblique Factors

Some authors (e.g., Cattell & Khanna; Harman, 1976; Jennrich & Sampson, 1966; Clarkson & Jennrich, 1988) have discussed in some detail the concept of oblique (non-orthogonal) factors, in order to achieve more interpretable simple structure. Specifically, computational strategies have been developed to rotate factors so as to best represent "clusters" of variables, without the constraint of orthogonality of factors. However, the oblique factors produced by such rotations are often not easily interpreted.

To return to the example discussed above, suppose we would have included in the satisfaction questionnaire above four items that measured other, "miscellaneous" types of satisfaction. Let us assume that people's responses to those items were affected about equally by their satisfaction at home (Factor 1) and at work (Factor 2). An oblique rotation will likely produce two correlated factors with less-than- obvious meaning, that is, with many cross-loadings.

1. Hierarchical Factor Analysis

Instead of computing loadings for often difficult to interpret oblique factors, you can use a strategy first proposed by Thompson (1951) and Schmid and Leiman (1957), which has been elaborated and popularized in the detailed discussions by Wherry (1959, 1975, 1984).

In this strategy, you first identify clusters of items and rotate axes through those clusters; next the correlations between those (oblique) factors is computed, and that correlation matrix of oblique factors is further factor-analyzed to yield a set of orthogonal factors that divide the variability in the items into that due to shared or common variance (secondary factors), and unique variance due to the clusters of similar variables (items) in the analysis (primary factors). To return to the example above, such a hierarchical analysis might yield the following factor loadings

Statistical Factor Analysis	Secondary & Primary Factor Loadings		
Factor	**Second. 1**	**Primary 1**	**Primary 2**
WORK_1	.483178	.649499	.187074
WORK_2	.570953	.687056	.140627
WORK_3	.565624	.656790	.115461
HOME_1	.535812	.117278	.630076
HOME_2	.615403	.079910	.668880
HOME_3	.586405	.065512	.626730
MISCEL_1	.780488	.466823	.280141
MISCEL_2	.734854	.464779	.238512
MISCEL_3	.776013	.439010	.303672
MISCEL_4	.714183	.455157	.228351

Careful examination of these loadings would lead to the following conclusions:

1. There is a general (secondary) satisfaction factor that likely affects all types of satisfaction measured by the 10 items;
2. There appear to be two primary unique areas of satisfaction that can best be described as satisfaction with work and satisfaction with home life.

   Wherry (1984) discusses in great detail examples of such hierarchical analyses, and how meaningful and interpretable secondary factors can be derived.

Confirmatory Factor Analysis

Over the past 15 years, so-called confirmatory methods have become increasingly popular (e.g., see Jöreskog and Sörbom, 1979). In general, you can specify a priori, a pattern of factor loadings for a particular number of orthogonal or oblique factors, and then test whether the observed correlation matrix can be reproduced given these specifications. Confirmatory factor analyses can be performed via Structural Equation Modeling (SEPATH)

Miscellaneous Other Issues and Statistics

Factor Scores

We can estimate the actual values of individual cases (observations) for the factors. These factor scores are particularly useful when you want to perform further analyses involving the factors that you have identified in the factor analysis.

Reproduced and Residual Correlations

An additional check for the appropriateness of the respective number of factors that were extracted is to compute the correlation matrix that would result if those were indeed the only factors. That matrix is called the reproduced correlation matrix. To see how this matrix deviates from the observed correlation matrix, you can compute the difference between the two; that matrix is called the matrix of residualcorrelations. The residual matrix may point to "misfits," that is, to particular correlation coefficients that cannot be reproduced appropriately by the current number of factors.

Matrix Ill-Conditioning

If, in the correlation matrix there are variables that are 100% redundant, then the inverse of the matrix cannot be computed. For example, if a variable is the sum of two other variables selected for the analysis, then the correlation matrix of those variables cannot be inverted, and the factor analysis can basically not be performed. In practice this happens when you are attempting to factor analyze a set of highly intercorrelated variables, as it, for example, sometimes occurs in correlational research with questionnaires. Then you can artificially lower all correlations in the correlation matrix by adding a small constant to the diagonal of the matrix, and then restandardizing it. This procedure will usually yield a matrix that now can be inverted and thus factor-analyzed; moreover, the factor patterns should not be affected by this procedure. However, note that the resulting estimates are not exact.

Discussion and Conclusion

Applying factors analysis approach, this study proposes a theoretical model to investigate main factors that contribute to successful information .This study is expected to contribute to both academics and management practices. The key concept of factor analysis is multiple observed variables have similar patterns of responses. In every factor analysis there is the same number of factors as they are variables.

Factor analysis is a statistical technique that is used to determine the extent to which a group of measures share common variance. Factor analysis is sometimes termed a "data reduction" technique because the method is frequently used to extract a few underlying components (or factors) from a large initial set of observed variables. It is extensively used in psychological research concerned with the construction of scales intended to measure attitudes, perceptions, motivations, and so forth. Business-related applications are numerous and examples include the development of scales used to measure customer satisfaction with products and employee work attitudes.

Factor analysis, however, has applicability outside of the realm of psychological research. It may be used, for example, by financial analysts to identify groups of stocks in which prices fluctuate in similar ways. And factor analysis often plays a crucial role in establishing the validity of employment tests and performance appraisal methods, thus helping a firm defend itself against employment discrimination.

References

1. Bartholomew, D.J.; Steele, F.; Galbraith, J.; Moustaki, I. (2008). Analysis of Multivariate Social Science Data. Statistics in the Social and Behavioral Sciences Series (2nd ed.). Taylor & Francis. ISBN 1584889608.
2. Harman, Harry H. (1976). Modern Factor Analysis. University of Chicago Press. pp. 175, 176. ISBN 0-226-31652-1.
3. Polit DF Beck CT (2012). Nursing Research: Generating and Assessing Evidence for Nursing Practice, 9th ed. Philadelphia, USA: Wolters Klower Health, Lippincott Williams & Wilkins.
4. Meng, J. (2011). "Uncover cooperative gene regulations by microRNAs and transcription factors in glioblastoma using a nonnegative hybrid factor model". International Conference on Acoustics, Speech and Signal Processing.
5. Liou, C.-Y.; Musicus, B.R. (2008). "Cross Entropy Approximation of Structured Gaussian Covariance Matrices". Ieee Transactions on Signal Processing 56(7): 3362–3367. doi:10.1109/TSP.2008.917878.
6. Bandalos, D.L.; Boehm-Kaufman, M.R. (2008). "Four common misconceptions in exploratory factor analysis". In Lance, Charles E.; Vandenberg, Robert J. Statistical and Methodological Myths and Urban Legends: Doctrine, Verity and Fable in the Organizational and Social Sciences. Taylor & Francis. pp. 61–87. ISBN 978-0-8058-6237-9.
7. Larsen, R., & Warne, R. T. (2010). Estimating confidence intervals for eigenvalues in exploratory factor analysis.Behavior Research Methods, 42, 871-876. doi:10.3758/BRM.42.3.871
8. [b] Warne, R. T., & Larsen, R. (2014). Evaluating a proposed modification of the Guttman rule for determining the number of factors in an exploratory factor analysis.Psychological Test and Assessment Modeling, 56, 104-123.
9. * Ledesma, R.D.; Valero-Mora, P. (2007). "Determining the Number of Factors to Retain in EFA: An easy-to-use computer program for carrying out Parallel Analysis".Practical Assessment Research & Evaluation 12 (2): 1–11.
10. Velicer, W.F. (1976). "Determining the number of components from the matrix of partial correlations".Psychometrika 41: 321–327. doi:10.1007/bf02293557.
11. Ruscio, John; Roche, B. (2012). "Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure".Psychological Assessment 24: 282–292.doi:10.1037/a0025697.

12. Garrido, L. E., & Abad, F. J., & Ponsoda, V. (2012). A new look at Horn's parallel analysis with ordinal variables. Psychological Methods. Advance online publication. doi:10.1037/a0030005
13. Courtney, M. G. R. (2013). Determining the number of factors to retain in EFA: Using the SPSS R-Menu v2.0 to make more judicious estimations. Practical Assessment, Research and Evaluation, 18(8).Available online:http://pareonline.net/getvn.asp?v=18&n=8
14. Russell, D.W. (December 2002). "In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin". Personality and Social Psychology Bulletin 28 (12): 1629–46.doi:10.1177/014616702237645.
15. Sternberg, R. J. (1977). Metaphors of Mind: Conceptions of the Nature of Intelligence. New York: Cambridge University Press. pp. 85–111.[verification needed]
16. Richard B. Darlington (2004) "Factor Analysis". Retrieved July 22, 2004.[dead link]
17. [b] Fabrigar et al. (1999). "Evaluating the use of exploratory factor analysis in psychological research.". Psychological Methods.
18. ^Suhr, Diane (2009). "Principal component analysis vs. exploratory factor analysis". SUGI 30 Proceedings. Retrieved 5 April 2012.
19. SAS Statistics. "Principal Components Analysis".SAS Support Textbook.
20. Meglen, R.R. (1991). "Examining Large Databases: A Chemometric Approach Using Principal Component Analysis". Journal of Chemometrics 5 (3): 163–179.doi:10.1002/cem.1180050305/.
21. Brown, J. D. (January 2009). "Principal components analysis and exploratory factor analysis – Definitions, differences and choices.". Shiken: Jalt Testing & Evaluation Sig Newsletter. Retrieved 16 April 2012.c gvbn