$\Sigma_{i+1} = L_i Q_i (= Q_i' \Sigma_i Q_i)$, $i = 1, 2, \ldots$ (The Gram-Schmidt orthogonalization is a way of finding Q_i and L_i; the QR method replaces a lower triangular matrix

Factor Analysis:-

The essential purpose of factor analysis is to decribe, if possible the covariance relationships among many variables ~~in~~ terms of a few underlying but unobservable, random quantities called factors.

Basically the factor model is motivat-ed the following arguments.

1. Suppose variables can be grouped by their correlations.

2. suppose all variables with in a particular group are highly correlated among themselves.

3. But have relatively small-correlations with variables in a different group.

Then it is conceivable that each

... single

For Example:

Correlations from the group of test Scores in classics, French, English, Mathematics, and Music collected by Spearman suggested an underlying "intelligence" factor.

A 2^{nd} group of variables, representing Physical-fitness scores, if a variable, might correspond to another factor. It is this type of structure that factor analysis seeks to confirm.

Factor analysis can be considered an extension of principal component analysis. Both can be viewed as attempts to approximate the covariance matrix Σ. However, the approximation based on the factor analysis model is more elaborate. The primary question in factor analysis is whether the data are consistent with a prescribed structure.

Difference between PCA and FA:

1. Principal component analysis involves extracting linear composites of observed variables.

2. Factor analysis is based on a

formal model predicting observed variables from theoretical latent factors.

In psycalogy these two techniques are often applied in the construction of multi scale tests to determine which items load on which scales. They typically yield similar substantive conclusions.

This helps to explain why some statistics packages seem to bundle them together.

I have also seen situations where "Principal component analysis" is incorrectly labelled "factor analysis".

In terms of a simple rule of thumb, suggest that you:

1. Run factor analysis if you assume or wish to test a theoretical model of latent factors causing observed variables.

2. Run PCA if you want to simply reduce your correlated observed to a smaller set of important independent composite variables.

The PC's of a data set are unique. The output of a factor analysis is only unique upto orbitary rotations. As such if you see a factor analysis that makes a lot of sense there's theorm to ask how much the results have been massaged.

Factor analysis has some other issues that make it a little undesirable as a dimension reduction technique.

1. Find the using SPSS correlation matrix
2. What are the PC using PCA
3. Apply factor analysis : 5 which variables coming to whether.

variables / subjects	1	2	3	4	5	6	7	8	9	10
1	59	70	81	74	71	71	73	69	87	100
2	60	71	82	85	61	58	68	59	98	100
3	58	61	61	72	60	32	57	56	95	85
4	43	54	58	79	66	24	61	60	79	85
5	59	65	74	91	66	73	78	66	87	95
6	55	68	59	70	63	50	69	66	92	90
7	70	64	78	80	74	73	75	70	99	100
8	63	65	68	66	71	63	71	69	87	100
9	69	64	76	76	61	71	74	68	96	100
10	59	62	71	87	71	68	82	70	89	100
11	36	60	69	75	63	54	71	58	68	100
12	63	24	53	55	42	27	57	30	65	90
13	58	65	69	76	60	52	68	57	84	100
14	54	55	57	56	62	53	66	65	71	100
15	54	23	61	72	54	21	61	22	94	100
16	60	59	66	57	57	32	66	59	76	85

3/10/11 Factor Analysis Model:

Orthogonal Factor model with m common factors:

$$X_{(p\times1)} = \mu_{(p\times1)} + L_{(p\times m)}F_{(m\times1)} + \varepsilon_{(p\times1)}$$

μ_i = mean of variable

ε_i = i^{th} Specific factor

F_j = j^{th} Common factor

l_{ij} = loading of the i^{th} variable on the j^{th} factor

The unobservable random vectors F & ε satisfy the following conditions:

F and ε are independent.

$E(F) = 0$, $\operatorname{cov}(F) = I$

$E(\varepsilon) = 0$, $\operatorname{cov}(\varepsilon) = \psi$, where, ψ is a diagonal matrix.

The orthogonal factor model implies a covariance structure for X. From the model,

$$(X-\mu)(X-\mu)' = (LF+\varepsilon)(LF+\varepsilon)'$$

$$= (LF+\varepsilon)[(LF)'+\varepsilon']$$

$$= LF(LF)' + \varepsilon(LF)' + LF\varepsilon' + \varepsilon\varepsilon'$$

so that, $\Sigma = \operatorname{cov}(X) = E(X-\mu)(X-\mu)'$

$$= LE(FF')L' + E(\varepsilon F')L' + LE(F\varepsilon') + E(\varepsilon\varepsilon')$$

$$= LL' + \psi$$

also by independence, $\operatorname{cov}(\varepsilon, F) = E(\varepsilon, F') = 0$

$$(X-\mu)F' = (LF+\varepsilon)F' = LFF' + \varepsilon F'$$

$$\operatorname{cov}(X, F) = E(X-\mu)F' = LE(FF') + E(\varepsilon F') = L$$

Covariance structure for the orthogonal Factor Model:

1. $\operatorname{cov}(X) = LL' + \psi$

or

$var(X_i) = l_{i1}^2 + \dots + l_{im}^2 + \psi_i$

$cov(X_i, X_k) = l_{i1} \cdot l_{k1} + \dots + l_{im} l_{km}$

2. $cov(X, F) = L$

or

$cov(X_i, F_j) = l_{ij}$.

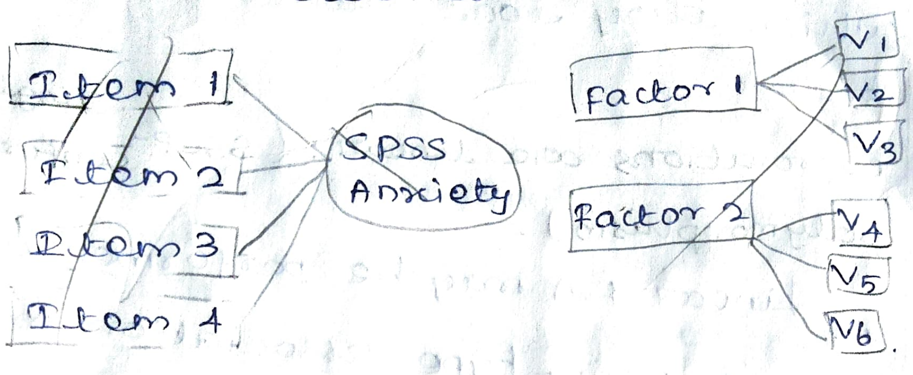
11/10/19 **Factor Loading :**

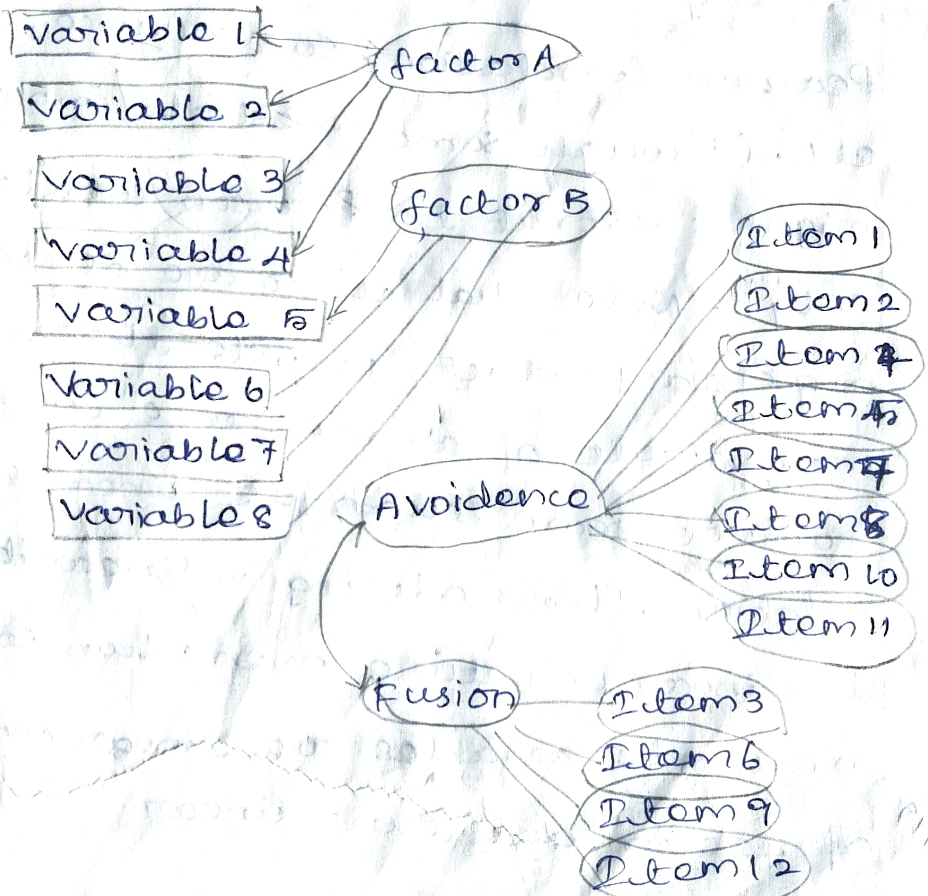
Factor Loading is basically the correlation co-efficient for the variable and factor. Factor Loading shows the variance explained by the variable on that particular factor. In the structural equation modeling approach, as a rule of thumb 0.7 or higher factor loading represents that the factor extracts sufficient variance from that variable.

Factor loading are part of the outcome from factor analysis. which servies as a data reduction method designed to explain the correlation between observed variables using a smaller number of factor. Because F.A is a widly used method in social and behavior research, an indepth examination of factor loading and the related factor loading matrix will facilitate a better understanding and use of the technique.

F.L are co-efficients found in either a factor pattern matrix or a factor structure matrix. The former

...s of regression co-efficients that multiply common factor to predict observed variables, where as the latter matrix is made up of product-moment correlation co-efficients between common factor and observed variables.





Assumptions of factor Analysis:

1. Scale (Interval or ratio) input variables:

That means the items are either continuous measures or are conceptualized as continuous while measured on data in linear F A. Binary data should be avoided. Linear F.A assumes that latent common and unique factors are continuous. Therefore, observed variables which they load should be continuous, too.

2. correlations are linear : (SSCP- single strand conformation polymorphism).

Linear F.A may be performed based on any SSCP-type association matrix; Pearson Correlation, co-variance, cosine, etc. (though some methods/implementations might restrict to Pearson correlations only). Note that these are all linear-algebra products despite that the magnitude of a co-variance co-efficient reflects more than just linearity in relation, the modeling in linearity FA is linear in nature even when co-variance are used; variables are linear combinations of factor and thus linear

$y = ab^x$

$a + bx$

28/10/19

3. No outliers :

That's as with any non ~~Subust~~ roubust methods Pearson correlation and ~~smaller~~ similar SSCP type association are sensitive of outliers so watch out.

Most of the rationable for rotating come from Thurstone (1947) and cattell (1978) who defended it use because this procedure simplifies the factor structure and therefore makes it interpretation easier and more reliable (ie., easier to replicate with different data sample).

Thurstone suggested five criteria to indentify a simple structure. According to there criteria still often reported in the literature a matrix if lodings simple if.

1. Each row contains at least one zero's.

2. For each column. There are at least as many zero's.

3. For any pairs of factor. there are some variable with zero loading on one factor and large loading on the other factor.

4. For any pair of factors there is a sizable proportion of zero loading.

5. For any pair of factors, there is only a small number of large loadings.

Rotations of factor can be done graphically. but are mostly obtained analytically and necessitate to specify mathematically the nation of simple structure in order to implement it with a computer program.

Orthogonal Rotation:

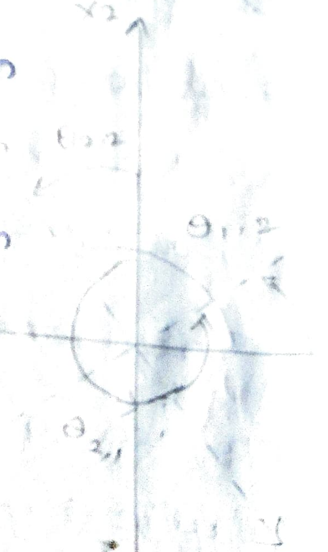
An orthogonal rotation is specified by a rotation matrix donated R, where the rows stands for the original factors and the columns for the new factors. At the intersection of row 'm' and columns 'n' we have the cosine of the angle between the original axis and the new one

$$\vartheta_{m,n} = \cos \theta_{m,n}.$$

$$R = \begin{bmatrix} \cos \theta_{1,1} & \cos \theta_{1,2} \\ \cos \theta_{2,1} & \cos \theta_{2,2} \end{bmatrix} = \begin{bmatrix} \cos \theta_{1,1} & -\sin \theta_{1,1} \\ \sin \theta_{1,1} & \cos \theta_{1,1} \end{bmatrix}$$

With the values of $\theta_{1,1} = 15$ degrees. The rotation matrix has the important property of being orthogonal because it corresponds to a matrix of direction cosines and therefore $R^T R = I$

An orthogonal rotation in 2 dimensions. The angles of rotation between and old axis 'm' and new axis 'n' is donated by $\theta_{m,n}$.



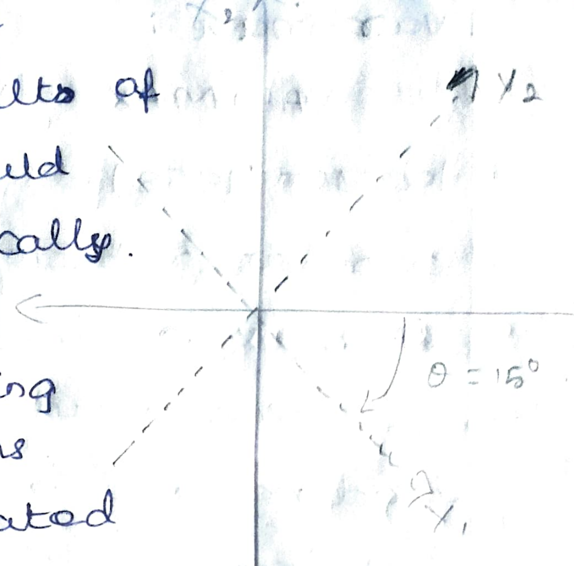
VARIMAX which was developed by kaiser (1958) is indubitably by most popular rotation method by far. For VARIMAX a simple solution means that each factor has a small number of large loading and a large number of zero loading.

Formally varimax searches for a rotation of the orthogonal factor such that the variance of the loading is maximized, which amounts to maximizing $V = \Sigma(q^4_{j,l} - \bar{q}^2_{j,l})^2$ with $q^2_{j,l}$ being the squared loading of the j^{th} variable on the 'l' factor, and $\bar{q}^2_{j,l}$ being the mean of the squared loading. $\bar{q}^2_{j,l}$ being the

Oblique Rotations:

In oblique rotations the new axes are free to take any position in the factor space, but the degree of correlation allowed among factors is in general, small because two highly correlated factor are better interpreted as only one factor. oblique rotations therefore relax the orthogonality constraint in order to gain simplitity in the interpretation.

Even though, in the principle, the results of oblique rotation could be presented graphically. They are almost interpreted by looking at the correlations between the rotated axis and the original variables.



These correlations are interpreted as loading.

4. Reasonably high correlations are Presented :

FA is the analysis of correlatedness - what its use when all or almost all correlations are weak? - no use. However, what is ["reasonably high correlation" depend on the field of study. There is also an interesting and varied question whether very high correlations should be accepted. To test statistically if the data are not uncorrelated]. Bartlett's test of sphericity can be used.

5. Partial Correlations are weak, and factor can be enough depend :

FA assumes that factors are

paors of correlated items. Infact, there even an advice not to extract factors loading decently less than 3 times in explatatory FA; and in confirmatory FA only 3+ is guaranteed identified structure. A technical problem of extraction called Heywood case has, as one of the reason's behind the two-few-items-on factor situation.

Kaiser-Mayer-Olkin (KMO) "sampling adequacy measure" estimates for you how work are partial correlations in the data relative the full correlations, and it can be computed for every item, for the whole correlation Matrix.

6) NO Multi collinearity:-

FA Model assumes that all items each posses unique factor and these factors are orthogonal therefore 2 items must define a plane 3 times - a 3d space etc.. p correlated vectors must span p-dim space to accomodata their p mutually perpendicular unique components. So no singlarity for their critical reasons. Not that complate multi collinearity is allowed through; yet it may cause computational problems in most of FA algorithms.

7) Distribution:-

In general, linear FA does not require normality of the input data. Moderately skewed distributions are acceptable. Bimodality is not a contra indication. Normality is indeed assumed for unique factors in the model (they serve as regressional errors) - but not the input data for common factors and the input data. Still multivariate normality of the data can be required as additional assumption by some methods of extraction (namely maximum likelihood) and by performing some asymptotic testing.

Maximum likelihood estimation:-

It has probably not escaped your notice that the estimation procedure above requires a starting guess as to 4. This makes its consistency somewhat shaky. On the other hand, we know from our elementary statistics course that MLE's are generally consistent unless we choose a spectacularly bad model.

We have so far got away with

making assumptions about the means and covariances of the factor scores F. to get an actual likelihood, we need to assume something about their distribut as well.

The usual assumption is that $F_{ir} \sim N(0,1)$ and that the factor scores are independent across factors $r = 1, \dots q$ and individuals $i = 1, 2, \dots n$. with the assumption, the factor have a multivariate normal distribution $X \sim N(0, \psi + W^TW)$: This means that the log likelihood is

$$L = \frac{-np}{2} \log 2\pi - \frac{n}{2} \log |\psi + W^TW| - \frac{n}{2} tr((\psi + W^TW)^{-1} V).$$

one can either try direct numerical maximization or use a two-stage procedure. starting, once again, with a guess as to ψ, one finds that the optimal choice of $\psi^{1/2} W^T$ is given by the matrix whose columns are the q leading eigen vectors of $\psi^{1/2} V \psi^{1/2}$.

The differences between the MLE and the "Principal factors". approach can be substantial.

If the data appear to be normally distributed then the additional efficiency of MLE is highly worthwhile

Also, as we'll see next time, it is a lot easier to test model assumptions is one uses the MLE.

Application for Principal component Analysis:

1 2m

PCA is predominantly used as dimensionality reduction computer vision and image compression. It is also used for finding patters in data high dimension in the field of finance data mining, bioinformatics, psychology, etc.

Assumptions of FA:

1 2m

1. Scale input variable (Interval, ra
2. correlations are linear
3. No outliers.
4. Reasonably high correlation are present
5. Partial correlations are week, and factor can be enough defined
6. No multicollinearity
7. Distribution.