Factor Analysis (Chpater 13)

Factor analysis is a dimension reduction technique where the number of dimensions is specified by the user. The idea is that there are underlying "latent" variables or "factors", and several variables might be measures of the same factor. Here the original variables are considered to be linear combinations of the underlying factors.

The idea is that the number of factors might be lower than the number of variables, and might correspond to theories about the data. For example, you might think there are different cognitive abilities, such as spatial reasoning, analytic reasoning, analogical reasoning, quantitative ability, linguistic ability, etc. A test might have 100 questions that measure these underlying abilities, so different linear combinations of the factors might give the distribution of the answers on the different test questions.

Factor analysis

Variables that are closely related to each other should have relatively high correlation, and variables that are not closely related should have relatively low correlation. Looking at correlations between variables, the ideal correlation matrix might look like this

$$\begin{pmatrix} 1.00 & 0.90 & .05 & .05 & .05 \\ .90 & 1.00 & .05 & .05 & .05 \\ .05 & .05 & 1.00 & .90 & .90 \\ .05 & .05 & .90 & 1.00 & .90 \\ .05 & .05 & .05 & .90 & 1.00 \end{pmatrix}$$

For this matrix, the first two variables are high correlated with each other but to no other variables, and similarly the last three variables are highly correlated to each other but to no other. Thus, the data could plausibly have arisen if there were two factors, with variables 1 and 2 measuring the first factor and variables 3–5 measuring the second factor.

Factor analysis

Factor analysis is similar to principal components in that linear combinations are used for dimension reduction. However

1. In factor analysis, the original variables are linear combinations of the factors. principal components are linear combinations of the original variables.
2. principal components seeks to find linear combinations to explain the total variance $\sum_i s_i^2$, whereas factor analysis tries to account for covariances in the data
3. Factor analysis is somewhat controversial among statisticians partly because solutions are not unique.

Factor analysis

In factor analyis, we treat data as arising from a single factor. We assume that there are$p$ variables and $m < p$ factors, where m is fixed in advance. The factors can be represented by $f_1, \ldots, f_m$. Then for obervation $\mathbf{y}_i$, the model is

$$
\begin{aligned}
y_1 - \mu_1 &= \lambda_{11} f_1 + \lambda_{12} f_2 + \cdots + \lambda_{1m} f_m + \varepsilon_1 \\
y_2 - \mu_2 &= \lambda_{21} f_1 + \lambda_{21} f_2 + \cdots + \lambda_{2m} f_m + \varepsilon_2 \\
&\vdots \\
y_p - \mu_p &= \lambda_{p1} f_1 + \lambda_{p1} f_2 + \cdots + \lambda_{pm} f_m + \varepsilon_p
\end{aligned}
$$

This would look very similar to a regression model if we moved the μs to the right hand sides, except that the factor loadings, λ_{ij} are individualized for each subject, and the factors are unobserved. Also, for each observation, we have p equations rather than one equation per observation that we would normally have in regression.

Factor analysis

The factor loading λ_{ij} indicates the importance of factor j to variable i. For example, if λ_{i2} is large for variables 1–3 and small for variables 4–p, the factor 2 is important for the first three variables but less important for the remaining variables. Hopefully this has some interpretation where the researcher hopes that they can describe something that relates the first three variables but not the remaining ones.

Although the factors are unknown, they are also considered random variables, and in the model we have

$$E(f_i) = 0, \quad Var(f_i) = 1, \quad Cov(f_i, f_j) = 0$$

So the factors are assumed to be independent. The model also assumes

$$E(\varepsilon_i) = 0, \quad Var(\varepsilon_i) = \psi_i$$

In other words, the error terms are allowed to differ for each variable. In addition, it is assumed that $Cov(\varepsilon_i, f_j) = 0$ and $Cov(\varepsilon_i, \varepsilon_j) = 0$.

Factor analysis

Based on the assumptions,

$$Var(y_i) = \lambda_{i1}^2 + \lambda_{i2}^2 + \cdots + \lambda_{im}^2 + \psi_i$$

Factor analysis

The model can also be written in matrix notation as

$$\mathbf{y} - \boldsymbol{\mu} = \mathbf{\Lambda f} + \boldsymbol{\varepsilon}$$

where $\mathbf{y} = (y_1, \ldots, y_p)'$, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)'$, $\mathbf{f} = (f_1, \ldots, f_m)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_p)'$ and $\mathbf{\Lambda} = (\lambda_{ij})$ with $i = 1, \ldots, p$, $j = 1, \ldots, m$.

Because $\boldsymbol{\mu}$ is a constant, the covariance matrix is

$$\begin{aligned} Var(\mathbf{y}) &= Cov(\mathbf{\Lambda f} + \boldsymbol{\varepsilon}) \\ &= Cov(\mathbf{\Lambda f}) + Cov(\boldsymbol{\varepsilon}) \\ &= \mathbf{\Lambda I \Lambda}' + \mathbf{\Psi} \\ &= \mathbf{\Lambda \Lambda}' + \mathbf{\Psi} \end{aligned}$$

Where $\mathbf{\Psi} = \mathrm{diag}(\psi_1, \ldots, \psi_p)$.

Factor analysis: example with 5 variables, 2 factors

$$
\begin{aligned}
y_1 - \mu_1 &= \lambda_{11} f_1 + \lambda_{12} f_2 + \varepsilon_1 \\
y_2 - \mu_2 &= \lambda_{21} f_1 + \lambda_{22} f_2 + \varepsilon_2 \\
y_3 - \mu_3 &= \lambda_{31} f_1 + \lambda_{32} f_2 + \varepsilon_3 \\
y_4 - \mu_4 &= \lambda_{41} f_1 + \lambda_{42} f_2 + \varepsilon_4 \\
y_5 - \mu_5 &= \lambda_{51} f_1 + \lambda_{52} f_2 + \varepsilon_5 .
\end{aligned}
$$

In matrix notation as in (13.3), this becomes

$$
\begin{pmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \\ y_3 - \mu_3 \\ y_4 - \mu_4 \\ y_5 - \mu_5 \end{pmatrix} = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{pmatrix}, \qquad (13.5)
$$

Factor analysis

We can interpret the matrix $\boldsymbol{\Lambda}$ as the covariances between the variables and the factors

$$\boldsymbol{\Lambda} = Cov(\mathbf{y}, \mathbf{f})$$

If the variables are standardized (using z-scores), then $\lambda_{i,j}$ is the correlation between the ith variable and jth factor.

The variance of a variable can be partitioned into a portion due to the factors and the remaining portion,

$$\begin{aligned} Var(y_i) = \sigma_{ii} &= (\lambda_{i1}^2 + \cdots \lambda ip^2) + \psi_i \\ &= h_i^2 + \psi_i \end{aligned}$$

Here h_i^2 is called the **comunality** or **common variance** and ψ is called the **specific variance** or **residual variance**.

Factor analysis

The hope in factor analysis is that

$$\mathbf{\Lambda\Lambda}' + \mathbf{\Psi} \approx \mathbf{\Sigma},$$

the covariance matrix for the original data, but often the approximation is not very good if there are too few factors. If the estimated factor analysis structure doesn' fit the estimate for $\mathbf{\Sigma}$, this indicates the inadequacy of the model and suggests that more factors might be needed.

The book points out that this can be a good thing in that the inadequacy of the model might be easier to see than other statistical settings that require complicated diagnostics.

Factor analysis

An issue that bothers some is that the factor loadings are not unique. If $\mathbf{T}$ is any othogonal matrix, then $\mathbf{T}\mathbf{T}' = \mathbf{I}$. Consequently,

$$\mathbf{y} - \boldsymbol{\mu} = \boldsymbol{\Lambda}\mathbf{T}\mathbf{T}'\mathbf{f} + \boldsymbol{\varepsilon}$$

is equivalent to the model with $\mathbf{T}\mathbf{T}'$ removed, and factor loadings $\boldsymbol{\Lambda}\mathbf{T}$ with factors $\mathbf{T}'\mathbf{f}$ will give equivalent results, and the new model could be written

$$\mathbf{y} - \boldsymbol{\mu} = \boldsymbol{\Lambda}^*\mathbf{f}^* + \boldsymbol{\varepsilon}$$

with $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}\mathbf{T}$ and $\mathbf{f}^* = \mathbf{T}'\mathbf{f}$.

The new factors and factor loadings are different, but the communalities and residual variances are not affected.

Factor analysis

There are different ways to estimate the factor loadings $\boldsymbol{\Lambda}$. The first is called the principal component method, but is unrelated to principal components (!). Four methods listed in the book are

1. "Princpal components" (not PCA)
2. Principal factors
3. Iterated principal factors
4. maximum likelihood

Factor analysis

For the first approach, the idea is to initially factor **S**, the sample covariance matrix of the data, into

$$\mathbf{S} \approx \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Lambda}}'$$

using singular value decomposition, so that

$$S = \mathbf{CDC}' = \mathbf{CD}^{\mathbf{1/2}}\mathbf{D}^{\mathbf{1/2}}\mathbf{C}' = (\mathbf{CD}^{\mathbf{1/2}})(\mathbf{CD}^{\mathbf{1/2}})'$$

where **C** is an orthogonal matrix with normalized eigenvectors and **D** is a diagonal matrix of eigenvalues $\theta_1, \ldots, \theta_p$. θ is used rather than λ because λ is used for the factor loadings (not sure why).

Here $\mathbf{CD}^{\mathbf{1/2}}$ is $p \times p$ but we want $\widehat{\boldsymbol{\Lambda}}$ to be $p \times m$

Factor analysis

To get a matrix with the right dimensions, we use the first m columns of $\mathbf{CD}^{1/2}$ to define $\widehat{\mathbf{\Lambda}}$, assuming that $\theta_1 > \cdots > \theta_m$ and the columns of $\mathbf{C}$ correspond to the eigenvectors with nonincreasing eigenvalues (i.e., rearrange columns of $\mathbf{C}$ and $\mathbf{D}$ if necessary).

To estimate $\mathbf{\Psi}$, we subtract the ith diagonal of $\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}'$ from $\mathbf{S}$:

$$\widehat{\psi}_i = s_{ii} - \sum_{j=1}^{m} \widehat{\lambda}_{ij}^2$$

To confirm that this is right, let $\mathbf{A} = \widehat{\mathbf{\Lambda}}'$. Then the ijth element of $\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}'$ is

$$\sum_{k=1}^{m} \lambda_{ik} a_{kj} = \sum_{k=1}^{m} \lambda_{ik} \lambda jk$$

For the diagonal, $j = i$, and we have λ_{ik}^2 in the summand, and the book uses j instead of k as the index.

Factor analysis

The fact that $m < p$ is what makes $\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}'$ only approximate $\mathbf{S}$. Adding $\widehat{\mathbf{\Psi}}$ means that the original sample variances are recovered exactly in the model but that the covariances are still estimated.

The estimated communality for variable i is the sum of the estimated factor loadings for that variable:

$$\widehat{h}_{ij} = \sum_{j=1}^{m} \widehat{\lambda}_{ij}^2$$

The estimated variance due to the jth factor is

$$\sum_{i=1}^{p} \widehat{\lambda}_{ij}^2 = \theta_j$$

which is the sum of squares of the ith column of $\widehat{\mathbf{\Lambda}}$.

Factor analysis

This equivalence is due to the following

$$\sum_{i=1}^{p} \widehat{\lambda}_{ij}^2 = \sum_{i=1}^{p} (\sqrt{\theta}_j c_{ij})^2 = \theta_j \sum_{i=1}^{p} c_{ij}^2 = \theta_j$$

where $c_{1j}, \ldots, c_{pj}$ is the jth normalized eigenvector.

Factor analysis

The proportion of the total sample variance (adding the variances of all variables separately, regardless of covariances), is therefore

$$\frac{\sum_{i=1}^{p} \widehat{\lambda}_{ij}^2}{\mathrm{tr}(\mathbf{S})} = \frac{\theta_j}{\mathrm{tr}(\mathbf{S})}$$

Note that if variables are standardized, then in place of the covariance matrix **S** we use the correlation matrix **R**, in which case the denominator is p.

The fit of the model can be measured by comparing the covariance matrix with its estimate into an error matrix

$$\mathbf{E} = \mathbf{S} - (\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}' + \widehat{\mathbf{\Psi}})$$

Factor analysis: example with 5 variables, 7 observations

Table 13.1. Perception Data: Ratings on Five Adjectives for Seven People

People	Kind	Intelligent	Happy	Likeable	Just
FSM1[a]	1	5	5	1	1
SISTER	8	9	7	9	8
FSM2	9	8	9	9	8
FATHER	9	9	9	9	9
TEACHER	1	9	1	1	9
MSM[b]	9	7	7	9	9
FSM3	9	7	9	9	7

[a] Female schoolmate 1.
[b] Male schoolmate.

$$\mathbf{R} = \begin{pmatrix} 1.000 & .296 & \mathbf{.881} & \mathbf{.995} & .545 \\ .296 & 1.000 & -.022 & .326 & \mathbf{.837} \\ \mathbf{.881} & -.022 & 1.000 & \mathbf{.867} & .130 \\ \mathbf{.995} & .326 & \mathbf{.867} & 1.000 & .544 \\ .545 & \mathbf{.837} & .130 & .544 & 1.000 \end{pmatrix}.$$

Factor analysis

The eigenvalues of the correlation matrix are

$$3.263 \quad 1.538 \quad 0.168 \quad 0.031 \quad 0$$

indicating that there is collinearity in the columns (they are not linearly independent). The proportion of the variance explained by the first factor is $3.263/5 = 0.6526$ and the proportion explained by the first two together is $(3.263 + 1.538)/5 = 0.9602$, meaning that two factors could account for 96% of the variability in the data. Looking at the correlation matrix, we have strong positive correlations within the sets $\{\mathit{Kind}, \mathit{Happy}, \mathit{Likeable}\}$ and $\{\mathit{Intelligent}, \mathit{Just}\}$, suggesting that these could correspond to separate and somewhat independent factors in the characteristics perceived in people by the subject.

Factor analysis: example with 5 variables, 7 observations

Table 13.2. Factor Loadings by the Principal Component Method for the Perception Data of Table 13.1

Variables	Loadings $\hat{\lambda}_{1j}$	$\hat{\lambda}_{2j}$	Communalities, $\hat{h}_i^2$	Specific Variances, $\hat{\psi}_i$
Kind	.969	$-.231$	.993	.007
Intelligent	.519	.807	.921	.079
Happy	.785	$-.587$	.960	.040
Likeable	.971	$-.210$	.987	.013
Just	.704	.667	.940	.060
Variance accounted for	3.263	1.538	4.802	
Proportion of total variance	.653	.308	.960	
Cumulative proportion	.653	.960	.960	

Factor analysis

$$\hat{\boldsymbol{\Lambda}}\hat{\boldsymbol{\Lambda}}' + \hat{\boldsymbol{\Psi}} = \begin{pmatrix} .969 & -.231 \\ .519 & .807 \\ .785 & -.587 \\ .971 & -.210 \\ .704 & .667 \end{pmatrix} \begin{pmatrix} .969 & .519 & .785 & .971 & .704 \\ -.231 & .807 & -.587 & -.210 & .667 \end{pmatrix}$$

$$+ \begin{pmatrix} .007 & 0 & 0 & 0 & 0 \\ 0 & .079 & 0 & 0 & 0 \\ 0 & 0 & .040 & 0 & 0 \\ 0 & 0 & 0 & .013 & 0 \\ 0 & 0 & 0 & 0 & .060 \end{pmatrix}$$

$$= \begin{pmatrix} 1.000 & .317 & .896 & .990 & .528 \\ .317 & 1.000 & -.066 & .335 & .904 \\ .896 & -.066 & 1.000 & .885 & .161 \\ .990 & .335 & .885 & 1.000 & .543 \\ .528 & .904 & .161 & .543 & 1.000 \end{pmatrix},$$

Factor analysis

In the example, the loadings in the first column give the relative importance of the variables for factor 1, while the loadings in the second column give the relative importance of factor 2. For factor 1, the highest variables are for Kind and Happy, with Likeable being third. The Just variable is somewhat similar to the Happy variable. For factor 2, Intelligent and Just stand out as having much higher correlations than the other variables.

As mentioned previously it is possible to rotate the factor loadings, however. The book points out that you could rotate the loadings (or equivalently, rotate the factors themselves). Plotting the factor loadings on a two-dimensional plot, we can see how to rotate the data to make a clearer (more easily interpretable) contrast between the factors. This is done by choosing $\mathbf{T}$ to be a suitable rotation matrix.

Factor analysis

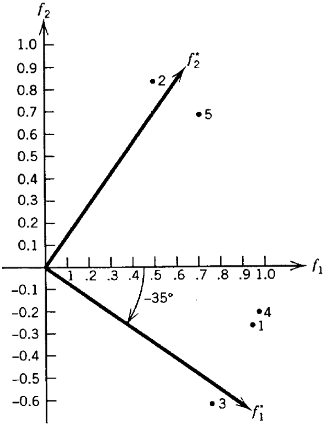


Figure 13.3. Plot of the two loadings for each of the five variables in the perception data of Table 13.1.

Factor analysis

Since there is no unique way to rotate the factors, there is no unique way to interpret the factor loadings. So, for example, variables 3 and 5 are similar on the original factor 1 (but not factor 2), while on the rotated factors, variables 3 and 5 are quite different, with variable 5 being high on factor 2 and variable 3 being high on factor 1. It seems that you could choose whether or not to rotate axes depending on the story you want to tell about variable 3.

The book suggests that when the factors are rotated, the factors could be interpreted as representing humanity (rotated factor 1) and rationality (factor 2). An objection is that this is imposing a theory of preconceived personalities onto the data.

Factor analysis

Another approach called the principle factor method is to estimate factor loadings is to first estimate $\widehat{\boldsymbol{\Psi}}$ and then use

$$\mathbf{S} - \widehat{\boldsymbol{\Psi}} \approx \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Lambda}}'$$

or

$$\mathbf{R} - \widehat{\boldsymbol{\Psi}} \approx \widehat{\boldsymbol{\Lambda}}\widehat{\boldsymbol{\Lambda}}'$$

Here

$$\mathbf{R} - \widehat{\boldsymbol{\Psi}}$$

can be approximated using r_{ij}, the sample correlations, for the off diagonal elements. For the diagonals, these can be estimated using R_i^2, the squared multiple correlation between y_i and the remaining variables. This is computed as

$$\widehat{h}_i = R_i^2 = 1 - \frac{1}{r^{ii}}$$

where r^{ii} denotes the ith diagonal element of $\mathbf{R}^{-1}$.

Using $\mathbf{S}$ instead of $\mathbf{R}$ is similar with

$$widehath_{ii}^2 = s_{ii} - \frac{1}{s^{ii}}$$

These approaches assume that $\mathbf{R}$ or $\mathbf{S}$ are nonsingular. Otherwise, h_i^2 can be estimated using the largest correlation in the ith row of $\mathbf{R}$.

Once $\mathbf{R} - \mathbf{\Psi}$ is estimated, $\widehat{\mathbf{\Lambda}}$ can be estimated using singular value decomposition. The sum of squares of the jth column of $\widehat{\mathbf{\Lambda}}$ is the jth eigenvalue of $\mathbf{R} - \mathbf{\Psi}$, and the sum of squares of the ith row of $\mathbf{R} - \mathbf{\Psi}$ is h_i^2, the communality.

The principal factor method can be iterated to improve estimates. Once $\widehat{\mathbf{\Lambda}}$ is given, this can be used to get improved estimates of the communalities, h_i^2 as

$$\widehat{h}_i^2 = \sum_{i=1}^{m} \widehat{\lambda}_{ij}^2$$

. Then an updated estimate for $\widehat{\mathbf{\Psi}}$ is given by

$$\widehat{\psi}_i = r_{ii} - \widehat{h}_i^2$$

using the updated value for $\widehat{h}_i^2$.

Factor analysis

The three methods — principal components, principal factors, and iterated principal factors — give similar results if the correlations are large and m is small, and if the number of variables p is large. Apparently, the iterative approach can result in an estimated commonality of greater than 1, which corresponds to a ψ value (a variance) being negative. Different software might handle this differently, either reporting an error or truncating estimates of commonalities to 1.

Factor analysis

Maximum likelihood can be done, but assumes a particular distribution for the variables, which is typically multivariate normal, which is often inappropriate for the types of data to which factor analysis is usually applied. This approach is also difficult numerically and requires iterative procedures to approximate the maximum liklelihood solution, and these procedures are not guaranteed to work (they might fail to converge).

Factor analysis

The number of factors might either correspond to a hypothesis about the variables related to the data or some method can be used. Here are some strategies:

1. choose enough factors so that the percentage of variance accounted for,
$$\frac{\sum_{i=1}^{m} \theta_i}{\mathrm{tr}(\mathbf{R})}$$
is sufficiently large, e.g., 90%
2. choose m to be the number of eigenvalues greater than the average eigenvalue
3. make a scree plot of the eigenvalues
4. If using likelihood, do a likelihood ratio test of
$$H_0 : \mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}, H_1 : \mathbf{\Sigma} \neq \mathbf{\Lambda}\mathbf{\Lambda}' + \mathbf{\Psi}$$

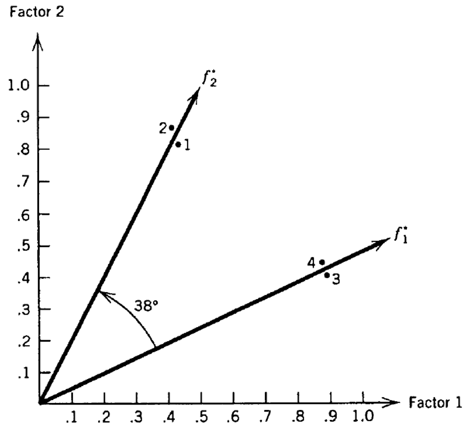
Factor analysis

Because the choice of m might not be obvious, there is a danger that a researcher will choose m to fit a theory about how many factors and which factors are needed to explain the data. This is also a potential danger with principal components or MDS, where a researcher might want to claim that genes correlate with geography for example, and use only two dimensions without testing whether three dimensions would be more appropriate for the data.

We mentioned earlier that factors can be rotated. This can be done visually, using trial and error to get an approximate rotation, or a rotation can be try to optimize some quantity. **Varimax** is an optimization technique that rotates the data so that the squared factor loadings are either minimized (close to 0) or maximized, so that variables are either strongly associated with a factor or fairly unrelated to a factor. This can be applied in two or more dimensions.

Another approach is called **oblique rotation**, in which the factors are made to not be orthogonal. In this case, instead of using an orthogonal transformation matrix $\mathbf{T}$, a nonorthogonal matrix $\mathbf{Q}$ is used instead, so that the new factors are $\mathbf{f}^* = \mathbf{Q}'\mathbf{f}$. If this was applied to the personality data, it would imply that the factors of humanity and rationality (or logicalness) are not independent.

Factor analysis



Factor analysis

It is difficult to say whether a factor loading $\widehat{\lambda}_{ij}$ is statistically signficant. A threshold of 0.3 has been advocated in the past, but the author argues that this is often too low and can result models being difficult to interpret. The book suggests that 0.5 or 0.6 is more useful to consider a factor loading large, although a good threshold also depends on the number of factors, m, with larger m tending to result in smaller factor loadings.

Factor analysis

It is also possible to estimate factors related to a given observation. These are called **factor scores**,

$$\mathbf{f}_i = (\widehat{f}_i, \widehat{f}_2, \ldots, \widehat{f}_m)'$$

where $i = 1, \ldots, n$. These estimate the factor values for each observation. The factor scores can be used to understand the observations themselves or are sometimes as input to MANOVA.

The factor scores are modeled as functions of the original observations

$$\mathbf{f}_i = \mathbf{B}_1'(\mathbf{y}_i - \overline{\mathbf{y}}) + \epsilon$$

where $\mathbf{B}_1 = (\beta_{ij})$ is a matrix of regression coefficients and ϵ is an error term (distinguished from ε used for the factor analysis model).

Factor analysis

The n equations for $\mathbf{f}_i$ can be combined as

these n equations can be combined into a single model,

$$\mathbf{F} = \begin{pmatrix} \mathbf{f}_1' \\ \mathbf{f}_2' \\ \vdots \\ \mathbf{f}_n' \end{pmatrix} = \begin{pmatrix} (\mathbf{y}_1 - \overline{\mathbf{y}})'\mathbf{B}_1 \\ (\mathbf{y}_2 - \overline{\mathbf{y}})'\mathbf{B}_1 \\ \vdots \\ (\mathbf{y}_n - \overline{\mathbf{y}})'\mathbf{B}_1 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_1' \\ \boldsymbol{\epsilon}_2' \\ \vdots \\ \boldsymbol{\epsilon}_n' \end{pmatrix}$$

$$= \begin{pmatrix} (\mathbf{y}_1 - \overline{\mathbf{y}})' \\ (\mathbf{y}_2 - \overline{\mathbf{y}})' \\ \vdots \\ (\mathbf{y}_n - \overline{\mathbf{y}})' \end{pmatrix} \mathbf{B}_1 + \boldsymbol{\Xi}$$

$$= \mathbf{Y}_c\mathbf{B}_1 + \boldsymbol{\Xi} \qquad \text{[by (10.11)]}. \qquad (13.53)$$

Factor analysis

The model essentially looks like a multivariate regression model with $\mathbf{Y}_c$ (the centered version of $\mathbf{Y}$) as the design matrix and $\mathbf{F}$ in place of $\mathbf{Y}$ (the matrix of responses in a usual regression). An important difference here is that $\mathbf{F}$ is unobserved.

If $\mathbf{F}$ had been observed, then the estimate for $\mathbf{B}_1$ would be (using the usual matrix representation of regression)

$$\widehat{\mathbf{B}}_1 = (\mathbf{Y}_c'\mathbf{Y})^{-1}\mathbf{Y}_c\mathbf{F}$$

Factor analysis

In multivariate regression (which we skipped over in chapter 10), you can estimate $\mathbf{B}_1$ as

$$\widehat{\mathbf{B}}_1 = \mathbf{S}_{yy}^{-1}\mathbf{S}_{xy}$$

where $\mathbf{S}_y y$ is the usual covariance matrix $\mathbf{S}$ and $\mathbf{S}_{xy}$ represents covariances between the explanatory variables and response variables, so in our case is $cov(\mathbf{f}, \mathbf{y}) = \widehat{\boldsymbol{\Lambda}}$.

We can estimate $\mathbf{F}$ by

$$\widehat{\mathbf{F}} = \mathbf{Y}_c\mathbf{S}^{-1}\widehat{\boldsymbol{\Lambda}}$$

(or use $\mathbf{R}$ in place of $\mathbf{S}$). Often you would obtain factor scores after doing rotations.

Factor analysis

An example from the book:

Example 13.6. The speaking rate of four voices was artificially manipulated by means of a rate changer without altering the pitch (Brown, Strong, and Rencher 1973). There were five rates for each voice:

$$\begin{aligned} \text{FF} &= 45\% \text{ faster},\\ \text{F} &= 25\% \text{ faster},\\ \text{N} &= \text{normal rate},\\ \text{S} &= 22\% \text{ slower},\\ \text{SS} &= 42\% \text{ slower}. \end{aligned}$$

Factor analysis

There were 20 voices read to 30 judges who judged the voices on a 14 point scale on the following variables: intelligent, ambitious, polite, active, confident, happy, just, likeable, kind, sincere, dependable, religious, good-looking, sociable, and strong. The results from the 30 judges were then averaged, so that there were 20 observations and 15 variables.

Note that in a mixed-model framework, you could use the original 30 judges separate scores, so that you would have 600 observations, but there would be correlation in the data so that scores from the same judge are more likely to be similar. You could also model this using multivariate regression just using the judges as blocks. In mixed models, we think of there being random effects for the judges, who have been selected from some larger population of possible judges.

Factor analysis: from the book

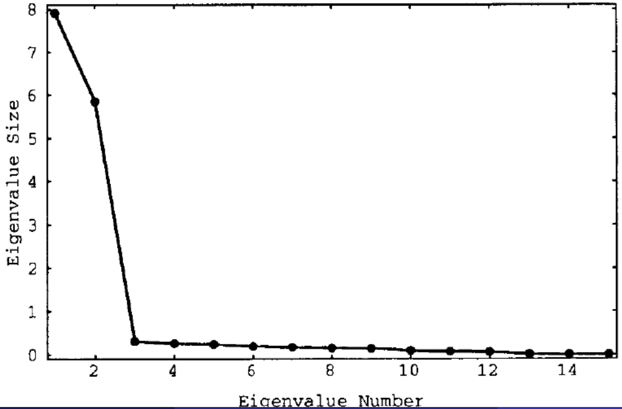
$$
\mathbf{R} = \begin{pmatrix}
1.00 & .90 & -.17 & .88 & .92 & .88 & .15 & .39 & -.02 & -.16 & .52 & -.15 & -.79 & -.78 & .73 \\
.90 & 1.00 & -.46 & .93 & .87 & .79 & -.16 & .10 & -.35 & -.42 & .25 & -.40 & .68 & -.60 & .62 \\
-.17 & -.46 & 1.00 & -.56 & -.13 & .07 & .85 & .75 & .88 & .91 & .68 & .88 & .21 & .31 & .25 \\
.88 & .93 & -.56 & 1.00 & .85 & .73 & -.25 & -.02 & -.45 & -.57 & .10 & -.53 & .58 & .84 & .50 \\
.92 & .87 & -.13 & .85 & 1.00 & .91 & .20 & .39 & -.09 & -.16 & .49 & -.10 & .85 & .80 & .81 \\
.88 & .79 & .07 & .73 & .91 & 1.00 & .27 & .53 & .12 & .06 & .66 & .08 & .90 & .85 & .78 \\
.15 & -.16 & .85 & -.25 & .20 & .27 & 1.00 & .85 & .81 & .79 & .79 & .81 & .43 & .54 & .53 \\
.39 & .10 & .75 & -.02 & .39 & .53 & .85 & 1.00 & .84 & .79 & .93 & .77 & .71 & .69 & .76 \\
-.02 & -.35 & .88 & -.45 & -.09 & .12 & .81 & .84 & 1.00 & .91 & .76 & .85 & .28 & .36 & .35 \\
-.16 & -.42 & .91 & -.57 & -.16 & .06 & .79 & .79 & .91 & 1.00 & .72 & .96 & .26 & .28 & .29 \\
.52 & .25 & .67 & .10 & .49 & .66 & .79 & .93 & .76 & .72 & 1.00 & .72 & .75 & .77 & .78 \\
-.15 & -.40 & .88 & -.53 & -.10 & .08 & .81 & .77 & .85 & .96 & .72 & 1.00 & .33 & .32 & .34 \\
.79 & .68 & .21 & .58 & .85 & .90 & .43 & .71 & .28 & .26 & .75 & .33 & 1.00 & .86 & .92 \\
.78 & .60 & .31 & .54 & .80 & .85 & .54 & .69 & .36 & .28 & .77 & .32 & .86 & 1.00 & .82 \\
.73 & .62 & .25 & .50 & .81 & .78 & .53 & .76 & .35 & .29 & .78 & .34 & .92 & .82 & 1.00
\end{pmatrix}
$$

Factor analysis

The eigenvalues are $7.91, 5.85, .31, .26, \ldots, .002$. Since only the first two eigenvalues are large, this suggests that two factors is reasonable for the data. This can be visualized by a scree plot.

Factor analysis

FACTOR ANALYSIS

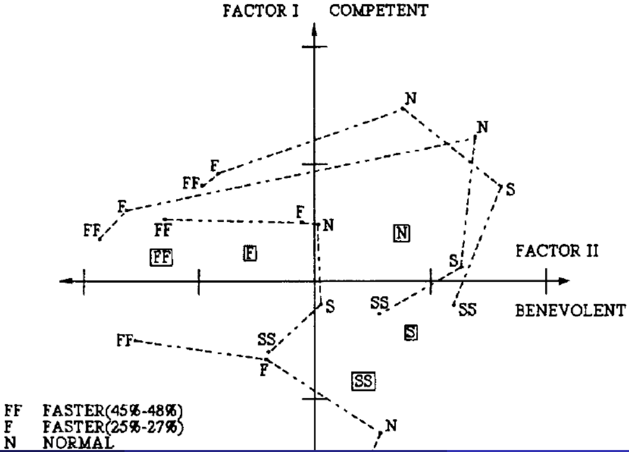


Factor analysis

Table 13.10. Initial and Varimax Rotated Loadings for the Voice Data

Variable	Initial Loadings f_1	Initial Loadings f_2	Rotated Loadings f_1	Rotated Loadings f_2	Communalities
Intelligent	.71	−.65	**.96**	−.06	.93
Ambitious	.48	−.84	**.90**	−.36	.94
Polite	.50	.81	−.12	**.95**	.92
Active	.37	−.91	**.86**	−.48	.97
Confident	.73	−.64	**.97**	−.04	.95
Happy	.83	−.47	**.94**	.15	.91
Just	.71	.58	.20	**.89**	.84
Likeable	.89	.39	.45	**.87**	.95
Kind	.58	.75	−.02	**.95**	.89
Sincere	.52	.82	−.11	**.97**	.95
Dependable	.93	.27	.56	**.79**	.94
Religious	.55	.79	−.07	**.96**	.92
Good looking	.91	−.29	**.89**	.35	.91
Sociable	.91	−.22	**.84**	.40	.87
Strong	.91	−.21	**.84**	.41	.86
Variance accounted for	7.91	5.85	7.11	6.65	13.76

Factor analysis



Factor analysis

As a general interpretation, the researchers categorized the factors as representing benevolence and competence, which you might question. Different researchers might have used different words to describe these groupings of variables. Slower voices were perceived to be more "benevolent" but less "competent", and faster voices were perceived to be more "competent" but less "benevolent".

Factor analysis

The author points out that many statisticians dislike factor analysis partly because of the nonuniqueness of the factor rotations and that this could lead to different interpretations. One question is whether the factors really exist? The previous example seems to suggest that people judge others according to their perceived benevolence or competence. It would have been interesting if these questions had been asked in addition to the attributes such as strength, intelligence, kindness, etc.

Factor analysis

A statistical consultant in a university setting or elsewhere all too often sees the following scenario. A researcher designs a long questionnaire, with answers to be given in, say, a five-point semantic differential scale or Likert scale. The respondents, who vary in attitude from uninterested to resentful, hurriedly mark answers that in many cases are not even good subjective responses to the questions. Then the researcher submits the results to a handy factor analysis program. Being disappointed in the results, he or she appeals to a statistician for help. They attempt to improve the results by trying different methods of extraction, different rotations, different values of m, and so on. But it is all to no avail. The scree plot looks more like the foothills than a steep cliff with gently sloping debris at the bottom. There is no clear value of m. They have to extract 10 or 12 factors to account for, say, 60% of the variance, and interpretation of this large number of factors is hopeless. If a few underlying dimensions exist, they are totally obscured by both systematic and random errors in marking the questionnaire. A factor analysis model simply does not fit such a data set, unless a large value of m is used, which gives useless results.

Factor analysis

A suggestion in the book for determining whether the factors are meaningful is to use replication – either by replicating the study itself or by splitting the data in half and seeing if the same factors seem to emerge for the two halves of the data. A difficult with this is that factor analysis is often applied in cases where there aren't many observations. The book also points out that "there are many data sets for which factor analysis should not be applied".

Factor analysis in R

There are a few ways to do factor analysis in R. Some common ones are the `fa()` and `factanal()` functions. The `fa()` function is in the `psych` package. `factanal()` is built into R but only does maximum likelihood factor analysis.

The `fa()` function is more flexible, for example, it can handle missing data (which would be common in questionnaire data) and doesn't only do maximum likelihood, as well as having more options.

Factor analysis in R

```
fa(r,nfactors=1,n.obs = NA,n.iter=1, rotate="oblimin", scores="regre
residuals=FALSE, SMC=TRUE, covar=FALSE,missing=FALSE,impute="median"
min.err = 0.001,  max.iter = 50,symmetric=TRUE, warnings=TRUE, fm="r
alpha=.1,p=.05,oblique.scores=FALSE,np.obs,use="pairwise",cor="cor"
```

The input is either a correlation matrix, covariance matrix, or the original data matrix. The user specifies the number of factors. The number of observations (number of rows in original data, not the number of variables or number of rows in the correlation matrix) must be specified to get confidence intervals and goodness-of-fit statistics.

There are a surprising number of algorithms to do the rotations, including

```
rotate: "none", "varimax", "quartimax", "bentlerT", "equamax",
    "varimin", "geominT" and "bifactor" are orthogonal rotations.
    "promax", "oblimin", "simplimax", "bentlerQ, "geominQ"
```

Factor analysis in R

`n.iter` specifies the number of iterations to get bootstrapped confidence intervals for factor loadings. If there is only one iteration, then confidence intervals aren't obtained. `fm` specifies the method for doing the factor analysis, with choices including `minres` for Ordinary Least Squares, `wls` for Weighted Least Squares. `ml` for maximum likelihood, `pa` does principal factor rotation (one of the methods in class) and so on.

Example

The following is an example questionnaire data set with 42 questions asking subjects to rate a web site on different variables. A guideline for the variables is the following:

```
TTU Website survey

Legend:

Q1  - Q9  purport to measure "ease of finding information"
Q10 - Q21 purport to measure "web design"
Q22 - Q29 purport to measure "attitude toward TTU"
Q30 - Q35 purport to measure "attitude toward web site"
```

```
Questions:

Q1 Finding information on athletics
Q2 Finding information on on-campus housing
Q3 Finding information on extracurricular activities (clubs)
Q4 Finding information on admissions (fees, general, etc.)
Q5 Finding information on financial aid and scholarship programs
Q6 Finding information on majors/minors
Q7 Finding information on student life (social life)
Q8 Finding information on directions (maps)
Q9 Finding information on admissions criteria
Q10 The use of graphics and pictures was favorable.
Q11 The download speed frustrated me.
Q12 The color coordination was pleasant.
Q13 The virtual tour was valuable. (If applicable).
Q14 I was unable to clearly read the web site's text.
Q15 The search function was worthless.
Q16 I did not think this web site was unique in appearance and conte
Q17 I liked the design of the home page.
```

Q18 Do you believe this web site provided entertainment value?
Q19 Do you believe this web site has a friendly tone?
Q20 I am attracted to web sites that have a friendly tone.
Q21 I prefer for a college/university website to be entertaining.
Q22 Ordinary:Exceptional Product
Q23 Not at all high quality:Extremely high quality
Q24 Poor value:Excellent Value
Q25 Boring:Exciting
Q26 Not a worthwhile university:A worthwhile university
Q27 Unappealing university:Appealing university
Q28 I would not recommend this university
:I would recommend this university
Q29 I would not apply to this university:I would apply
to this university
Q30 This website makes it easy for me to build a
relationship with this university.
Q31 I would like to visit this website again in the future.
Q32 I'm satisfied with the information provided by this web site.

Q33 I feel comfortable in surfing this web site.
Q34 I feel surfing this web site is a good way for me to spend my ti
Q35 Compared with other university web sites, I would rate this one
One of the best:One of the worst
Q36 Would you apply to this university?
Q37 Would you recommend a friend, family member,
or peer to apply to this university?
Q38 Please indicate your gender.
Q39 Please indicate your ethnicity.
Q40 Are you currently enrolled at a university, college,
trade school, etc.?
Q41 If you answered yes, please indicate your classification
Q42 If you are a student, are you considering to transfer?

Example

The data is in a comma delimited file, and looks something like this:

```
Q1,Q2,Q3,Q4,Q5,Q6,Q7,Q8,Q9,Q10,Q11,Q12,Q13,Q14,Q15,Q16,Q17,Q18,Q19,(
28,Q29,Q30,Q31,Q32,Q33,Q34,Q35,Q36,Q37,Q38,Q39,Q40,Q41,Q42
5,4,4,5,5,4,3,3,5,1,4,1,3,4,4,4,2,2,1,2,3,3,3,4,3,4,4,4,4,2,1,1,2,3
5,5,4,4,4,5,5,4,5,2,4,2,2,4,5,4,2,1,1,2,2,4,4,4,3,5,5,5,5,2,2,2,2,2
4,3,3,2,2,4,4,4,3,2,3,2,,4,3,2,1,1,1,,2,2,3,3,3,5,5,5,5,3,2,3,2,3,2
5,4,3,4,3,3,2,4,4,2,4,2,3,5,4,4,2,1,1,2,2,4,4,4,4,5,5,5,5,2,1,1,1,2
3,2,1,2,2,4,3,2,2,2,3,2,2,3,4,3,2,1,1,2,2,4,3,3,4,3,4,3,4,2,3,3,5,4
5,5,5,5,4,5,5,4,5,2,3,2,2,2,4,4,1,2,1,1,2,5,5,5,5,5,5,5,5,1,1,1,2,3
5,4,4,5,5,3,4,4,5,1,3,2,2,4,2,2,2,1,1,2,2,3,3,3,4,5,5,5,5,3,2,2,2,2
5,1,3,5,4,2,2,4,4,2,4,2,4,5,3,1,2,1,1,2,2,4,4,5,4,5,5,5,5,2,1,1,2,3
5,4,4,5,2,4,4,4,4,1,5,1,3,5,5,3,2,2,1,2,4,5,5,5,5,5,5,5,5,2,1,1,1,3
5,4,4,5,4,4,4,5,5,2,4,1,3,5,4,4,2,1,1,2,2,4,4,4,4,5,5,5,5,2,1,1,2,3
```

Note that there is missing data when two commas appear in a row.

Example

The questionnaire was designed with four themes in mind in the first 35 questions: ease of using the website, web design, attitude toward the school, and attitutde toward the web site, so it is plausible that that there are three factors largely influencing the responses.

```
> x <- read.table("http://math.unm.edu/~james/ttu_websurv.csv"
,sep=",",header=T)
> survey <- x[,1-35]
> library(psych)
> a <- fa(survey,nfactors=4,rotate="varimax",fm="pa")
> summary(a)
actor analysis with Call: fa(r = survey, nfactors = 4, rotate = "va

Test of the hypothesis that 4 factors are sufficient.
The degrees of freedom for the model is 662  and the objective func
The number of observations was  328  with Chi Square =  1408.51  wi

The root mean square of the residuals (RMSA) is  0.05
```

Example

To get an idea of the number of factors, one can look at the eigenvalues:

```
> names(a)
 [1] "residual"    "dof"                      "chi"
 [4] "nh"          "rms"                      "EPVAL"
 [7] "crms"        "EBIC"                     "ESABIC"
[10] "fit"         "fit.off"                  "sd"
[13] "factors"     "complexity"               "n.obs"
[16] "objective"   "criteria"                 "STATISTIC"
[19] "PVAL"        "Call"                     "null.model"
[22] "null.dof"    "null.chisq"               "TLI"
[25] "RMSEA"       "BIC"                      "SABIC"
[28] "r.scores"    "R2"                       "valid"
[31] "score.cor"   "weights"                  "rotation"
[34] "communality" "uniquenesses"             "values"
[37] "e.values"    "loadings"                 "fm"
[40] "Structure"   "communality.iterations"   "scores"
[43] "r"           "np.obs"                   "fn"
```

```
> a$e.values
 [1] 8.7740074 2.5751451 2.1574631 1.8840962 1.6223974 1.5288774 1.
 [8] 1.2640660 1.1430268 1.1205339 1.1069691 1.0145828 0.9899176 0.
[15] 0.8568173 0.8539552 0.8063560 0.7693574 0.7291366 0.7087996 0.
[22] 0.6675796 0.6449242 0.6134014 0.5843541 0.5247089 0.4937273 0.
[29] 0.4661804 0.4492236 0.4361704 0.3815501 0.3617987 0.3485546 0.
[36] 0.2869011 0.2764981 0.2549595 0.2340578 0.1778630 0.1316506
> cumsum(a$e.values)/sum(a$e.values)
 [1] 0.2140002 0.2768086 0.3294296 0.3753832 0.4149539 0.4522436 0.
 [8] 0.5156637 0.5435424 0.5708725 0.5978717 0.6226177 0.6467620 0.
[15] 0.6907341 0.7115623 0.7312295 0.7499943 0.7677781 0.7850659 0.
[22] 0.8182225 0.8339523 0.8489133 0.8631659 0.8759637 0.8880058 0.
[29] 0.9110939 0.9220506 0.9326889 0.9419950 0.9508193 0.9593207 0.
[36] 0.9737798 0.9805236 0.9867422 0.9924509 0.9967890 1.0000000
```

Example

A common test is to use the number of factors where the eigenvalues are greater than 1, but this would require 12 factors, and you need 21 factors to explain 80% of the variance, so this is not good. It means that the data will be very hard to interpret in terms of latent variables or factors. On the other hand, it means that the questionnaire is asking different questions and not just asking the same question 10 different ways.

To see some of the other output, the factor analysis gives linear combinations of the factors for each question.

```
> a
Factor Analysis using method =  pa
Call: fa(r = survey, nfactors = 4, rotate = "varimax", fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matri
      PA1   PA2   PA3   PA4    h2   u2 com
Q1   0.17 -0.12  0.31  0.45 0.342 0.66 2.3
Q2   0.09 -0.05  0.40  0.02 0.169 0.83 1.2
Q3   0.13 -0.23  0.51 -0.11 0.342 0.66 1.7
Q4   0.12  0.00  0.50  0.25 0.328 0.67 1.6
```

Example

Even though four factors doesn't fit the data well, we can try to see if we can interpret the factors to some extent. The factor loadings can be made easier to read by only printing those above a certain threshold:

```
> print(a$loadings,cutoff=.5)
```

```
 Loadings:
    PA1    PA2    PA3    PA4
Q1
Q2
Q3
Q4                 0.506
Q5                 0.536
Q6
Q7                 0.515
Q8
Q9
Q10
Q11
Q12                       -0.529
Q13
Q14
Q15
Q16
Q17
Q18
```

Repeating the factor analysis with more factors can show some more groupings of questions, and can change the groupings of the variables.

If your interest is more in redesigning the survey (for example, in order to ask fewer questions), this can still be a useful tool, even if you are not figuring out the number of factors or being able to interpret the factors very easily.

Loadings:

Loadings:

	PA1	PA2	PA6	PA4	PA3	PA10	PA9	PA5	PA8
Q1									
Q2									
Q3						0.648			
Q4									
Q5					0.565				
Q6									
Q7									
Q8									
Q9					0.628				
Q10				0.513					
Q11							0.509		
Q12				0.627					
Q13									0.595
Q14									
Q15									
Q16									
Q17									

Structural Equation Modeling

Structural equation modeling is a topic that we won't go into, but is related to factor analysis. Confirmatory factory analysis (where you test whether your idea of the factors — their number and relationship to the variables) is considered one type of SEM, as are path analysis and latent growth analysis.

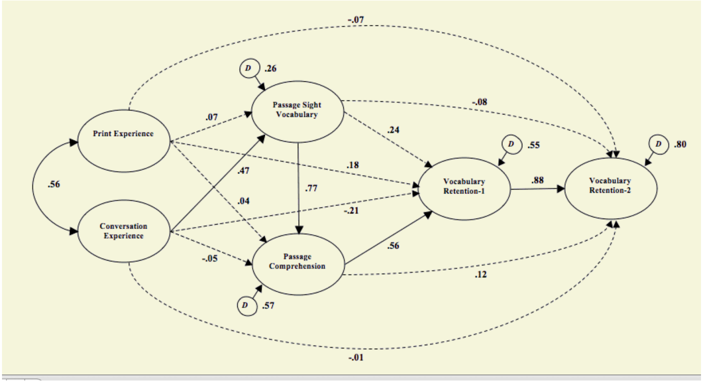
SEM allows modeling relationships between variables, including both measurement variables and latent variables, that can be used to reflect a model. Often this is a type of causal model in which latent variables act as causes affecting observed variables. More generally, latent variables can affect each other in causal ways.

These models can get quite elaborate and often are analyzed using moderately expensive commercial software such as LISREL ($ 495, a lot cheaper than SAS!).

We aren't doing anything with SEM, but I just think you should have

SEM

From http://nflrc.hawaii.edu/rfl/October2008/pulido/pulido.html



Identifiability in SEMs

An interesting issue in SEMs is whether two different graphs could produce the same data. This is a question of model identifiability and is a current area of research in statistics:

```
"Identifiability of parameters in latent structure models with
Elizabeth S. Allman, Catherine Matias, and John A. Rhodes
Annals of Statistics, 37 no.6A (2009) 3099-3132.

"Parameter identifiability of discrete Bayesian networks with
Elizabeth S. Allman, John A. Rhodes, Elena Stanghellini, and M
Journal of Causal Inference, to appear.
```