# Lecture 11 - Latent variable models 2

Björn Andersson

*University of Oslo*

November 24 2021

# Factor scores and sum scores

- **The observed score is an unbiased estimator of the true score with single factor models**
  - This applies also to the case when the factor loadings are not all equal or when the unique variances are not the same for all the items
  - With parallel items (equal factor loadings and equal error variances), it is also the least-squares and maximum likelihood estimator (MLE)
- With a general single factor model, a different estimator is the MLE

$$\hat{t}_i^{\text{MLE}} = \frac{\sum_{j=1}^{m}(\lambda_j/\Psi_j^2)(x_{ji}-\mu_j)}{\sum_{j=1}^{m}\lambda_j^2/\Psi_j^2} + \sum_{j=1}^{m}\mu_j.$$

  - This estimator gives different weights to each item score, depending on the factor loadings and the unique variances.

# Table of Contents

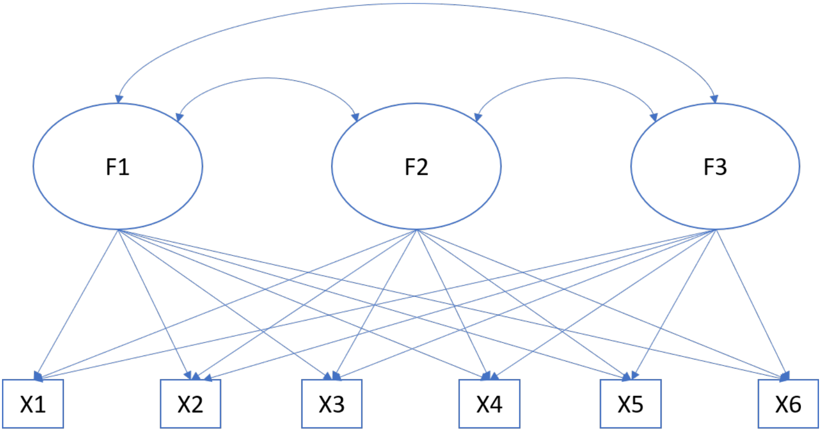
# The multiple factor model

For an item score $X_j$, we have the following model with, e.g., three dimensions

$$X_j = \mu_j + \lambda_{j1} F_1 + \lambda_{j2} F_2 + \lambda_{j3} F_3 + \epsilon_j,$$

where the factors $\mathbf{F} = (F_1, F_2, F_3)'$ have mean 0 and correlation matrix $\Sigma$ and where $\epsilon_j$ is independent of $\mathbf{F}$.

# Graphical illustration



# Interpretation of the model

- $X_j$ is the observed item score variable - the *indicator*
- $F_1$, $F_2$ and $F_3$ are the measures of the attribute - the *factors*
- $\epsilon_j$ is the item-specific error
- $\mu_j$ is the item difficulty level
- $\lambda_{1,j}$ is the *factor loading*, which indicates how "strongly" the item measures the factor $F$

# The factor loadings

$\lambda_{1,j}$ is the factor loading of factor 1 for item $j$

- The factor loading measures the sensitivity of each item with respect to the attribute
- An item with a high factor loading is a better indicator of the attribute $F_1$ than an item with a low factor loading
- The factor loading is a measure of the ability of the item to distinguish between individuals with high and low values of the factor

# The factor loadings

With the SWLS we had the following estimated factor loadings:

| | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 |
|---|---|---|---|---|---|
| $\lambda_j$ | 1.29 | 1.10 | 1.15 | 0.95 | 1.19 |
| $\Psi_j^2$ | 0.90 | 1.27 | 1.14 | 1.86 | 1.95 |

- This means that an individual with a factor score of 2 has an expected score of $\mu_1 + 1.29 \times 2$ for item 1, while an individual with a factor score of 1 has an expected score of $\mu_1 + 1.29 \times 1$ for the same item.
- The error variance for item 1 is the lowest while the factor loading is the largest, among all items. This means that item 1 contributes the most to the reliability of the sum scores among all the items.
- Item 4 has the lowest factor loading, meaning that the average difference between the item scores of individuals with factor scores of 2 and 1 is the lowest among all items.

# Model assumptions

What are the implicit assumptions of the factor model?

- The model assumes conditional independence between item scores (given the factors, the item scores are independent of each other)
- We also assume continuous item scores
- For the typical estimation methods (ML) we assume that factors are normal distributed

# Relaxing the model assumptions

- We can specify models that do not assume normally distributed factors
- We can model residual dependence between item scores
- We can specify a model which explicitly takes into account binary and ordinal data
- We can combine all the three above things

In this course we will not pursue such extensions to the main model. How to handle these issues is touched upon in the Item response theory and Measurement models courses given during the spring.

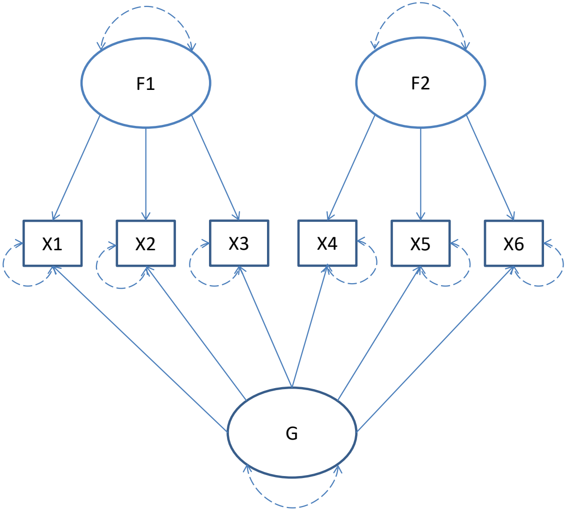
# A special case: the bifactor model

Let's say we have six item scores and three factors. In case there is a general factor $G$ and two subfactors $F_1$ and $F_2$, the equations for each item score are:

$$
\begin{aligned}
X_1 &= \mu_1 + \lambda_{G,1} G + \lambda_{1,1} F_1 + \epsilon_1 \\
X_2 &= \mu_2 + \lambda_{G,2} G + \lambda_{1,2} F_1 + \epsilon_2 \\
X_3 &= \mu_3 + \lambda_{G,3} G + \lambda_{1,3} F_1 + \epsilon_3 \\
X_4 &= \mu_4 + \lambda_{G,4} G + \lambda_{2,4} F_2 + \epsilon_4 \\
X_5 &= \mu_5 + \lambda_{G,5} G + \lambda_{2,5} F_2 + \epsilon_5 \\
X_6 &= \mu_6 + \lambda_{G,6} G + \lambda_{2,6} F_2 + \epsilon_6
\end{aligned}
$$

Here, $G, F_1$ and $F_2$ are **uncorrelated**.

# Graphical illustration



# Interpreting the bifactor model

- Is this a realistic and interpretable measurement model?
- We assume uncorrelated latent variables so the different factors have to be interpreted in this way - kind of like a "residual variance" explained by another independent factor specific to a subset of the items
  - Not easily justified to call subfactors of mathematics "algebra" and "geometry", for example
  - More like: "an algebra/geometry-related factor that is uncorrelated with the common factor"
- We can use the model as a tool to evaluate "how unidimensional" a test is

## Common variance

We define common variance as the variance explained by the model, i.e. the total variance minus the variance of the error score. In a bifactor model, the factors are all uncorrelated so the variance of the item score is just the sum of variances for all the parts:

$$\begin{aligned}\mathrm{Var}(X_1) &= \mathrm{Var}(\mu_1 + \lambda_{G,1} G + \lambda_{1,1} F_1 + \epsilon_1)\\ &= \lambda_{G,1}^2 \times \mathrm{Var}(G) + \lambda_{1,1}^2 \times \mathrm{Var}(F_1) + \mathrm{Var}(\epsilon_1)\\ &= \lambda_{G,1}^2 + \lambda_{1,1}^2 + \mathrm{Var}(\epsilon_1).\end{aligned}$$

Considering all six items in the scale, we then have that the total common variance is equal two the sum of the first two terms in the above equation, i.e.

$$\sum_{j=1}^{6} \lambda_{G,j}^2 + \sum_{j=1}^{3} \lambda_{1,j}^2 + \sum_{j=4}^{6} \lambda_{2,j}^2.$$

# Common variance explained by the general factor

The variance explained by the common factor is equal to $\sum_{j=1}^{6} \lambda_{G,j}^2$. We can then look at how much of the common variance that is explained by the general factor to get an idea of to what degree the test as a whole is unidimensional. We obtain the explained common variance:

$$\text{ECV} = \frac{\sum_{j=1}^{6} \lambda_{G,j}^2}{\sum_{j=1}^{6} \lambda_{G,j}^2 + \sum_{j=1}^{3} \lambda_{1,j}^2 + \sum_{j=4}^{6} \lambda_{2,j}^2}.$$

A generally recommended cut-off for "sufficient" unidimensionality is $\text{ECV} \geq 0.7$.

# Example: cognitive scale MoCA

- The Montreal Cognitive Assessment (MoCA) is a cognitive scale used as a screening tool for mild cognitive impairment
- Data were available from the baseline assessment of a longitudinal study on health and well-being of Cantonese-speaking older persons residing in public rental housing estates in Hong Kong.
- Participants in the longitudinal study were randomly sampled in three age strata (65-74 years, 75-84 years, 85 years and older).
- Those with known dementia diagnosis or a known psychiatric disorder were excluded from this analysis.
- Relevant data collected between July and November 2014 were retrieved from 1,876 participants for this analysis.

# Example: cognitive scale MoCA

- To investigate the strength of unidimensionality, we specified a bifactor model with item bundles defined based on prior information
- We estimated the model and calculated the explained common variance from the general factor
- From the parameter estimates, the explained common variance was 0.76 which supported a unidimensional interpretation of the scale scores
- Based on these results, we proceeded to evaluate the item properties with respect to education level using a unidimensional model

# Table of Contents

# Model evaluation procedures: hypothesis testing

- One way of assessing model fit is to look at the chi-square statistic between a model without restrictions and the restricted model fitted (e.g. a single factor model)
- This is what is reported in the default output in lavaan
- Generally speaking, the chi-square statistic is not very informative since it investigates an absolute model fit - factor models will practically never fit perfectly

# Model evaluation procedures: model-implied covariance matrix

Let $s_{jk}$ be the sample covariance between item scores $j$ and $k$ and let $\sigma_{jk}$ be the model-implied covariance between item scores $j$ and $k$, which is a function of the estimated factor model parameters.

- The McDonald book suggests to always inspect the difference between the model-implied covariance matrix and the observed covariance matrix
- We calculate the **discrepancy matrix** which is the matrix of differences between the $s_{jk}$ and $\sigma_{jk}$ for all item scores $j$ and $k$
- If the absolute values are all less than 0.1 between the correlation matrices, McDonald suggests that the fit is acceptable
- Inspecting the discrepancy matrix can also inform where in the model possible misfit is present

# Model-implied covariance matrix: SWLS

An independent-clusters model with two correlated factors yielded the following discrepancy matrix when using the correlation matrix:

$$\mathbf{S_X} - \hat{\Sigma}_\mathbf{X} = \begin{pmatrix} 0.000 & & & & \\ 0.016 & 0.000 & & & \\ -0.010 & -0.010 & 0.000 & & \\ 0.010 & -0.063 & 0.046 & 0.000 & \\ -0.017 & 0.013 & 0.016 & 0.000 & 0.000 \end{pmatrix}$$

From inspection of the matrix, does the model fit well?

# Model evaluation procedures: GFI

We can define

$$q_u = \frac{1}{m^2}\sum_{j=1}^{m}\sum_{k=1}^{m}(s_{jk} - \sigma_{jk})^2,$$

which is the mean of the squared differences between the sample covariances and the model-implied covariances. Besides these, we also want to account for the magnitude of the sample covariances. If we also compute the statistic

$$c = \frac{1}{m^2}\sum_{j=1}^{m}\sum_{k=1}^{m} s_{jk}^2,$$

we can achieve this.

# Model evaluation procedures: GFI

We then define the goodness-of-fit index as

$$\text{GFI} = 1 - q_u/c.$$

If the fit is good, the GFI will be close to 1. We've already mentioned some guidelines that a GFI $\geq 0.9$ is "acceptable" and a GFI $\geq 0.95$ is "good".

# Model evaluation procedures: SRMR/SRMSR

Another fit measure is the Standardized root mean squared residual (SRMR, also SRMSR) index. Like the GFI, it considers the differences between the sample covariances $s_{jk}$ and the model-implied covariances $\sigma_{jk}$:

$$\mathrm{SRMR} = \sqrt{\frac{2\sum_{j=1}^{m}\sum_{k=1}^{j}\left[(s_{jk}-\sigma_{jk})/(s_{jj}s_{kk})\right]^2}{m(m+1)}},$$

where $m$ is the number of observed variables and $s_{jj}$ is the sample standard deviation for item $j$. Hu and Bentler (1999) recommends SRMR $< 0.08$ as a rule of thumb for assessing model fit.

Note: This definition of SRMR is from Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6(1)*, 1-55.

# Model evaluation procedures: RMSEA

Another goodness-of-fit measure also incorporates parsimony (penalty is added for more parameters in the model). Define $L$ as the likelihood value at the MLE, let $df$ be the degrees of freedom and let $n$ be the sample size. We can then calculate

$$d = \frac{L - df}{n},$$

which is an estimate of the "error of approximation of the model to the population". From this statistic we can compute the Root Mean Square Error of Approximation, defined as

$$\mathrm{RMSEA} = \sqrt{d/df},$$

which is equal to zero if the model fits perfectly. McDonald mentions that RMSEA $< 0.05$ is an acceptable fit while Hu and Bentler (1999) suggests using RMSEA $< 0.06$.

# Model selection procedures

- Hypothesis tests
- Information criteria

# Likelihood ratio tests

- We fit nested models and compare the likelihood between the models
- If there is no significant result, we prefer the simpler model
- Example: We want to choose between a factor model with two factor loadings restricted to be equal and a model where they are not restricted to be equal.

# Restricted factor loadings likelihood ratio tests

Let $L_c$ denote the constrained log-likelihood and let $L_u$ denote the unconstrained log-likelihood, corresponding to the resticted model and the unrestricted model. We proceed as follows:

- We select a significance level $\alpha$
- We estimate the restricted model and the unrestricted model and obtain the values of the likelihood at the MLEs.
- We calculate the test statistic

$$T = -2 \log \frac{L_c}{L_u} = -2(\log L_c - \log L_u).$$

- We reject the null hypothesis of no difference between the restricted and unrestricted models if the test statistic has a large enough value, according to the selected $\alpha$

# Pitfalls of likelihood ratio tests

- Multiple comparison problem
  - If we do several comparisons these need to be taken into account with respect to $\alpha$
  - If we do not know how many comparisons we will do, controlling the overall significance level is a difficult problem to solve
- Violation of conditions
  - The null hypothesis may imply that the parameters are on the boundary of the parameter space
- The test statistic has a $\chi^2$-distribution only with large samples

# Model selection: information criteria

- For likelihood methods, it is possible to select models based on different information criteria.
- The information criteria are based on the value of the likelihood at the MLE and the number of parameters in the model, and possible other things such as the sample size
- The model which has the lowest value of the criterion is selected
- A potential problem is to decide which criterion to use, because they may suggest different models

# AIC and BIC

Let $k$ denote the number of free parameters in the model and let $L$ denote the likelihood. The Akaike Information Criterion (AIC) is defined as

$$\mathrm{AIC} = 2k - 2\log(L)$$

The AIC only penalizes for including more parameters in the model. The Bayesian Information Criterion (BIC) also accounts for the sample size. Let $n$ denote the sample size. The BIC is defined as

$$\mathrm{BIC} = k\log(n) - 2\log(L).$$

The BIC has attractive large sample properties for some (but not all) types of models. This means that if you use the BIC to select a model among a number of different models and the true model is among those models, the BIC will select the correct model with probability 1 as the sample size tends to infinity.

# BIC model selection: Thurstone data

During Lab 3, you were tasked with estimating three models:

1. A single-factor model
2. A three-factor independent clusters model with correlated factors where the dimensions are defined by:
   - Verbal comprehension: Sentences, Vocabulary, Sent.Completion
   - Word fluency: First.Letters, Four.Letter.Words, Suffixes
   - Reasoning: Letter.Series, Pedigrees, Letter.Group
3. A bifactor model with three uncorrelated subfactors defined by the same items as in 2.

Which one did you choose?

# BIC model selection: Thurstone data

The results gave:

1. A single-factor model: 4655.247
2. A three-factor independent clusters model with correlated factors: 4475.063
3. A bifactor model with three uncorrelated subfactors: 4493.185

Thus, we select model 2: the correlated factors model.