Lecture notes I: Measurement invariance[1]

Literature.

Mellenbergh, G. J. (1989). Item bias and item response theory. International Journal of Educational Research, 13,127-143.
[a readable discussion of measurement invariance, defined generally]
Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. Psychometrika, 58, 525-543.
[a demanding discussion of measurement invariance in the linear common factor model]
Dolan, C. V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. Multivariate Behavioral Research, 35, 21-50.
[application of MI in the linear factor model to investigate measurement invariance – the example discussed below is based on this paper]

Introduction.

Psychometrics concerns the study of the relationship between latent variables (or traits) and their manifest indicators. Psychometrics has focused largely on the development of statistical models of this relationship. Well known models that relate observed dichotomous indicators to continuous latent variables include the Rasch model and the Birmbaum model. Similarly, there are several well known models for observed polytomous items, such as the graded response model and the partial credit model, and model for continuous items, such as the linear factor model.

Psychometric models (or measurement models) may be viewed as regression models in which we define a single continuous latent trait (e.g., "depression", "perceptual speed", "working memory", "extroversion") as the independent variable, and the observed indicators responses ("do you like to meet new people?" [y/n]; "i find it hard to concentrate" [often / sometimes / seldom / never]) as the dependent variables. If the dependent variable is discrete (dichotomous or polytomous), then the regression model will be (say) a logistic regression model rather than a linear regression model.

As mentioned, psychometric model, which relates a continuous latent variable or trait to continuous indicators, is the linear factor model. Again, the factor model may be viewed as a regression model, but as now both the latent trait and the indicators are continuous, the regression is linear. As such it is very familiar to the standard regression model (see below). There are psychometric models that are suitable to relate discrete latent variables or traits to observed discrete or continuous indicators, but we will not consider these (e.g., latent class model). A taxonomy of psychometric model is provided by the following table 1-1.

Table 1-1 Taxonomy of psychometric models.

	Latent variable / trait / common factor		
		discrete	continuous
observed indicators	discrete	latent class model	IRT: Rasch, Birmbaum, <u>Discrete factor model</u>
	continuous	latent profile model	<u>linear factor model</u>

We will consider the underlined models. However, subject to certain assumptions, the discrete factor model is equivalent to the Birmbaum model and the Rasch model.

[1] Conor V. Dolan c.v.dolan@uva.nl. RM20. MI: continuous & discrete factor models.

Box 1-1: Psychometric modeling: what is an indicator?
The latent variable or trait is the variable that we would like to measure. But because it is latent, we cannot measure it directly. However we can observe the effect of the latent trait on indicators of the trait. What counts as an indicator? From the perspective of the psychometric model, we consider an indicators an observable variable, which is directly and causally dependent on the latent trait. This is simple in theory, but actually difficult in practice. Here psychological theory plays (or should play) an important role. The nature of the latent trait should be theoretically sufficiently developed to inform a choice of indicators. For instance, suppose we want to measure dysthymia. A clinical psychologist should be able to identify for potential indicators ("In the morning, I often feel that I will not be able to cope with the day's events"). The collection of indicators constitutes the items of the psychometric test.

Psychometric modeling serves mainly to demonstrate that the observed item responses are consistent with a single underlying trait. Specifically, this means that the observed item responses covary in a manner that is consistent with the presence of a single latent trait. Equivalently we hypothesize that the item responses covary because they are all influences by the same causal underlying latent trait. If a given model fits the data, we can derive from the fitted model useful information about the quality of the items in the test.

Below we first outline the linear common factor as a measurement model for continuous indicators. With this model place, we shall present the definition of measurement invariance (MI). MI in the linear factor model can be investigated easily in programs like LISREL.

The linear factor model as a measurement model.

We consider the factor model for a single group. First recall the LISREL model without the means. Let $\mathbf{y}_i$ denote the zero mean n_y-vector of observed variable, observed in subject i. Let $\boldsymbol{\eta}_i$ denote the zero mean n_e-vector of latent variables or common factors. The regression of observed $\mathbf{y}$ on latent $\boldsymbol{\eta}$:

$$\mathbf{y}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i, \qquad \text{eq. 1-1}$$

where $\boldsymbol{\Lambda}$ is the n_y x n_e matrix of factor loadings, and $\boldsymbol{\varepsilon}_i$ is a zero mean ny vector or residuals (i.e., in the regression of $\mathbf{y}$ on $\boldsymbol{\eta}$). The regression of components of $\boldsymbol{\eta}$ on components of $\boldsymbol{\eta}$:

$$\boldsymbol{\eta}_i = (\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\zeta}_i, \qquad \text{eq. 1-2}$$

where $\mathbf{B}$ is the n_e x n_e matrix of regression coefficients. The derivation of this is: $\boldsymbol{\eta}_i = \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\zeta}_i$ -> $\boldsymbol{\eta}_i - \mathbf{B}\boldsymbol{\eta}_i = \boldsymbol{\zeta}_i$ -> $(\mathbf{I}-\mathbf{B})\boldsymbol{\eta}_i = \boldsymbol{\zeta}$ -> $(\mathbf{I}-\mathbf{B})^{-1}(\mathbf{I}-\mathbf{B})\boldsymbol{\eta}_i = (\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\zeta}_i$. Of course, if $\mathbf{B}$=0, then we have the identity $\boldsymbol{\eta}_i = \boldsymbol{\zeta}_i$. Here (i.e., $\mathbf{B}=\mathbf{0}$), it is more natural to speak of $\boldsymbol{\eta}_i$ as the latent traits or variables, or common factors. In addition is $n_e = 1$ (single common factor model), the $\mathbf{B}$ will be zero. We assumed $E[\boldsymbol{\zeta}]=0$, $E[\boldsymbol{\varepsilon}]=0$, and $E[\mathbf{y}]=0$[2]. The covariance matrix of **y** equals:

$$\Sigma = \Lambda(I-B)^{-1}\Psi(I-B)^{-1t}\Lambda^{t} + \Theta, \qquad \text{eq. 1-3}$$

where the covariance matrix of $\boldsymbol{\eta}$ equals $(I-B)^{-1}\Psi(I-B)^{-1t}$ (n_e x n_e), the covariance matrix of $\boldsymbol{\zeta}$ equals Ψ (n_e x n_e) and the covariance matrix of $\boldsymbol{\varepsilon}$ equals q (n_y x n_y). If B equals zero, we have (remember that $\Lambda I = \Lambda$):

$$\Sigma = \Lambda\Psi\Lambda^{t} + \Theta, \qquad \text{eq. 1-4}$$

and the covariance matrix of $\boldsymbol{\eta}$ equals Ψ. We now extend the model as follows:

$$\mathbf{y}_i = \boldsymbol{\tau}_y + \Lambda\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i$$
$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + B\boldsymbol{\eta}_i + \boldsymbol{\zeta}_i$$

where $\boldsymbol{\tau}_y$ is the n_y vector of intercepts or indicator means, depending on the details of the model, and $\boldsymbol{\alpha}$ is a vector of intercepts or factor means, depending on the details of the model. To simplify things we shall assume that $B=\mathbf{0}$. The means are:

$$\mathbf{E}[\mathbf{y}] = \tau_y + \Lambda\mathbf{E}[\eta] + \mathbf{E}[\varepsilon]$$
$$\mathbf{E}[\eta] = \alpha + \mathbf{E}[\zeta]$$

$$\mathbf{E}[\varepsilon] = \mathbf{E}[\zeta] = \mathbf{0}$$
$$\mathbf{E}[\eta] = \alpha.$$

Model for means is thus:

$$\mathbf{E}[\mathbf{y}] = \tau + \Lambda\mathbf{E}[\eta]$$
$$\mathbf{E}[\eta] = \alpha,$$

or, given the appropriate substitution:

$$\mathbf{E}[\mathbf{y}] = \tau + \Lambda\alpha. \qquad \text{eq. 1-5}$$

The covariance matrix still equals $\Sigma = \Lambda\Psi\Lambda^{t} + \Theta$ (remember we assumed that $B=\mathbf{0}$), so $\Sigma = \Lambda\Psi\Lambda^{t} + \Theta$. Note that in LISREL the parameter vectors $\boldsymbol{\tau}$ and $\boldsymbol{\alpha}$ are called ty and al. These may appear on the 'mo' line and be specified using 'pa', 'ma', etc. You can refer to specific element in the usual fashion as well (e.g., fi al 1 al 2).

So far we have considered LISREL modeling as a particular instance of covariance structure modeling (particular in the sense that it is limited to the LISREL model). With the model for the means in place, we view LISREL model as a particular instance of mean and covariance structure modeling. Table 1-2 and 1-3 provide an overview of the extended model.

---

[2] Note that $E[\mathbf{y}]$ = mean($\mathbf{y}$), I also employ the notation $\boldsymbol{\mu}_y$ for the mean of **y**.

Table 1-2: LISREL covariance and mean structure in k=1...K populations.

covariance structure	mean structure
$\Sigma_k = \Lambda_k \Psi_k \Lambda_k^t + \Theta_k$	$\mu_{yk} = \tau_k + \Lambda_k \alpha_k$

Table 1-3: LISREL model matrices + dimensions

matrix	LISREL	order	meaning
Λ_k	ly	ny x ne	factor loading matrix ($\mathbf{y}$->η)
Ψ_k	ps	ne x ne	cov/cor matrix of η or ζ
Θ_k	te	ny x ny	cov/cor matrix of residuals(ε)
Σ_k	-	ny x ny	expected model cov. matrix of $\mathbf{y}$
vector	LISREL	dimension	meaning
τ_k	ty	ny x 1	intercept in regression of $\mathbf{y}$ on η
α_k	al	ne x 1	common factor means
μ_{yk}	-	ny x 1	expected means of $\mathbf{y}$

Box 1-2: Scaling in the common factor model

Consider the single common factor model, $\Sigma = \Lambda\Psi\Lambda^t + \Theta$, where Ψ is the variance of the common factor. To fit this model we have to impose some scale on the common factor. Specifically because we cannot observe it, we cannot know its mean or variance. The standard solution to the problem is to either fix the variance to one (Ψ=1), and estimate all factor loadings freely, or to fix a single factor loading to one (say, the j-th loading), and to estimate Ψ freely (which is now a direct function of the j-th observed indicator scale). Now usually we assume the means of all variables in the model to equal zero. But with the introduction of structured means, we have an additional scaling problem: if we cannot observe η, how can be know its mean value? We can solve this problem by fixing the mean of h to zero. So we go from $\mu_{yk} = \tau_k + \Lambda_k\alpha_k$ to simply $\mu_{yk} = \tau_k$. As α_k is zero. Given this constraint the observed means will equal the intercepts τ_k.

Small example: single factor model.

We shall fit a single common factor model to 4 indicators of performal IQ. We shall do this by scaling in Λ (see Box 1-2), and fixing the mean of the common factor to zero. Summary statistics are included in the LISREL input (note: pc=picture completion, pa=picture arrangement, oa=object assembly, ma=matrices).

```
title single factor model including the means
da no=1868 ni=4
cm sy
      8.24
      2.84       8.47
      3.54       3.24       9.06
      2.55       2.40       2.86       9.36
me
     10.41      10.37      10.73      10.41
la
     pc         pa         oa         ma
mo ly=fu,fr ps=sy,fr te=di,fr al=fu,fi ty=fu,fr ne=1 ny=4
pa ly
0 2 3 4
pa te
11 12 13 14
pa ps
21
va 1 ly 1 1 ! scaling - variance of factor
va 0 al 1 ! scaling mean of common factor
ou
```

Note that the means are included in the input, and that the means model of eq 1-5 is specified $\mathbf{E[y]} = \tau + \Lambda\alpha$. However, a is fixed to zero, so that the model is simply $\mathbf{E[y]} = \tau$. That is the estimates in τ will simply equal the observed means. In the output we find:

```
        TAU-Y
                  pc          pa          oa          ma
           --------    --------    --------    --------
              10.41       10.37       10.73       10.41
             (0.07)      (0.07)      (0.07)      (0.07)
             156.70      153.96      154.03      147.02
```

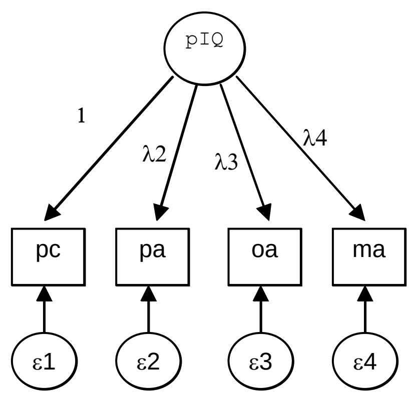


Figure 1-1: Path diagram of single factor model with 4 indicators (scaling of the common factor pIQ achieved by fixing the first factor loading to 1).

Factor modeling in multiple groups

In investigating measurement invariance with respect to group in the linear factor model, it is convenient to fit a given factor model in multiple groups. The multi-group extension is relatively simple as it involve merely stacking LISREL input. To illustrate this, I fit a two group model – but I do so without any constraints over the groups.

```
title single factor model including the means
title whites
da no=1868 ni=4 ng=2
cm sy
      8.24
      2.84      8.47
      3.54      3.24      9.06
      2.55      2.40      2.86      9.36
me
     10.41     10.37     10.73     10.41
la
     pc        pa        oa        ma
mo ly=fu,fr ps=sy,fr te=di,fr al=fu,fi ty=fu,fr ne=1 ny=4
le
PIQ
pa ly
0 2 3 4
pa te
11 12 13 14
pa ps
21
va 1 ly 1 1 ! scaling - variance of factor
va 0 al 1 ! scaling mean of common factor
ou rs
title single factor model including the means
title blacks
da no=306
cm sy
       9.18
       3.40      9.18
       4.39      3.68      8.76
       3.51      3.12      1.81     10.37
me
       8.12      8.10      7.89      8.39
la
     pc        pa        oa        ma
mo ly=fu,fr ps=sy,fr te=di,fr al=fu,fi ty=fu,fr ne=1 ny=4
le
PIQ
pa ly
0 102 103 104
pa te
111 112 113 114
pa ps
121
va 1 ly 1 1 ! scaling - variance of factor
va 0 al 1 ! scaling mean of common factor
ou rs
```

Without going into the details of the results, I merely note that the two group analysis without any constraints over the groups will produce results that are exactly the same as those obtained in two single group analyses.

Measurement invariance in the linear factor model.

The data shown above were IQ tests collected in 1868 white youths and 305 black youths (WISC US norm data). Suppose a researcher carries out a MANOVA to investigate the hypothesis that white youth on average score higher on performal IQ than black youths. Suppose the chosen alpha is 0.01, and the results of the MANOVA are:

```
            Df Pillai approx F num Df den Df    Pr(>F)
gr           1  0.134   84.108      4   2168 < 2.2e-16 ***
```

The test statistic $F(4,2168)=84.108$, and the p-value is $< .01$. The univariate test are also all significant given alpha $=.01/4$. So the researcher concludes that his hypothesis is correct and concludes: "White youth score higher on average than black youths with respect to performal IQ". The researcher tries to publish these results, and receives a review report, including the following comment:

> "The author concludes that the groups differ with respect to performal IQ. This conclusion is based on the supposition that the same construct was measured in both groups. How can the author be so sure of this? How does the researcher know that the difference observed at the level of the observed variables are a function of differences at the level of the latent variable of interest, namely performal IQ?"

This reviewer has an excellent point. How do we know we are measuring the same construct? To answer this question, we have to identify the conditions, which are <u>necessary</u> to establish that we are indeed measuring the same construct in both groups. These conditions are given by the definition of measurement invariance (MI). In considering MI in the linear factor model, we introduce distributional assumptions in the model, and we require the idea of a conditional distribution (see Box 1-3).

We introduce the following distributional assumption in group k:

$$\mathbf{y}_{ki} \sim N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k),\ k=1....K. \qquad \text{eq 1-6}$$

This means that the observed random vector $\mathbf{y}_{ki}$ in group k follows a multivariate normal distribution. The mean vector and covariance matrix are subjected to the linear factor model, so we can write:

$$\mathbf{y}_{ki} \sim N(\boldsymbol{\tau}_k + \boldsymbol{\Lambda}_k\boldsymbol{\alpha}_k,\ \boldsymbol{\Lambda}_k\boldsymbol{\Psi}_k\boldsymbol{\Lambda}_k^{\mathbf{t}} + \boldsymbol{\Theta}_k),\ k=1....K,$$

Note that this distributional assumption is considered within each group. That is, we may consider this distribution a <u>distribution conditional on group</u>. As explained in Box 1-3 we can also condition on the common factor.

Box 1-3. Conditioning

Consider the multi-group linear factor model: $\mathbf{y}_{ki} = \boldsymbol{\Lambda}\boldsymbol{\eta}_{ki} + \boldsymbol{\varepsilon}_{ki}$ (k=1...K groups, and i=1...N_k cases in group k). First consider the model in a given group k. In this model, I condition on a given value of $\boldsymbol{\eta}_k$, $\boldsymbol{\eta}^*$, by considering the model in subject for who $\boldsymbol{\eta}_{ki} = \boldsymbol{\eta}^*$. The mean and covariance matrix of $\mathbf{y}$ in this group of subjects equals:

$E[\mathbf{y}_k|\boldsymbol{\eta}_{ki} = \boldsymbol{\eta}^*] = \boldsymbol{\tau}_k + \boldsymbol{\Lambda}_k \boldsymbol{\eta}^*$, and
$\boldsymbol{\Sigma}_{k|\boldsymbol{\eta}^*} = \boldsymbol{\Theta}_k$

Note that this result is analogous to the results obtain by conditioning on the predictor in the linear regression model: yi = b0 + b1* xi + ei.

$E[y_i|x_i=x^*] = b0 + b1\ x^*$
$var(y_i|x_i=x^*) = \sigma^2_e$

Note that σ^2_e does not depend on the value of x*. This is the assumption of homoskedasticity in the linear regression model. Similarly, note in the factor model that $\boldsymbol{\Theta}_k$ does not depend on $\boldsymbol{\eta}^*$. Again this is the assumption of homoskedasticity, but not defined in the linear factor model. So conditioning on a variable ($\boldsymbol{\eta}$ or x) means considering the model in subject who all have a given identical fixed value on $\boldsymbol{\eta}$ (factor model) or x (regression model). Here we consider the mean and variance of the dependent variable ($\mathbf{y}$ in the factor model or y in the regression model). But we can take a more general approach by considering the condition distribution of y or $\mathbf{y}$. This allows us to adopt a slightly more general approach (the conditional mean and variance are aspects or characteristics of the conditional distribution).

The distribution of the observed data conditional on group is given (i.e., multivariate normality). Within a given group k, we consider the conditional distribution of $\mathbf{y}_{ki}$ given $\boldsymbol{\eta}_k=\boldsymbol{\eta}^*$, $f(\mathbf{y}_{ki}|\ \boldsymbol{\eta}^*)$:

$$\mathbf{y}_{ki}|\ \boldsymbol{\eta}^* \sim N(\boldsymbol{\tau}_k + \boldsymbol{\Lambda}_k \boldsymbol{\eta}^*,\ \boldsymbol{\Theta}_k), \qquad \text{eq 1-7}$$

So $f(\mathbf{y}_{ki}|\ \boldsymbol{\eta}^*)$ is again a multivariate normal distribution, with the specific covariance matrix and mean vector. Specifically, the conditional means and covariance matrix within group k are:

$E[\mathbf{y}_k|\boldsymbol{\eta}_{ki} = \boldsymbol{\eta}^*] = \boldsymbol{\tau}_k + \boldsymbol{\Lambda}_k\boldsymbol{\eta}^*$, and $\boldsymbol{\Sigma}_{k|\boldsymbol{\eta}^*} = \boldsymbol{\Theta}_k$.

The definition of MI in the linear factor model requires the explicit conditioning on group:

Definition of MI: $f(\mathbf{y}_i|\ \boldsymbol{\eta}^*) = f(\mathbf{y}_i|\ \boldsymbol{\eta}^*\ \&\ \text{group}=k)$ eq 1-8.

for all values of $\boldsymbol{\eta}^*$ and all values of k. Now given eq 1-7, this means that $f(\mathbf{y}_{ki}|\ \boldsymbol{\eta}^*)$ should be equal over all groups (k=1...K). Consider just two groups, k=1 and k=2. Conditional distributions in groups 1 and 2 are

$\mathbf{y}_{1i} | \boldsymbol{\eta}^* \sim N(\boldsymbol{\tau}_1 + \boldsymbol{\Lambda}_1 \boldsymbol{\eta}^*, \boldsymbol{\Theta}_1)$
$\mathbf{y}_{2i} | \boldsymbol{\eta}^* \sim N(\boldsymbol{\tau}_2 + \boldsymbol{\Lambda}_2 \boldsymbol{\eta}^*, \boldsymbol{\Theta}_2)$

MI require that these conditional distributions to be equal. So this implies that the distribution $N(\boldsymbol{\tau}_1 + \boldsymbol{\Lambda}_1 \boldsymbol{\eta}^*, \boldsymbol{\Theta}_1)$ should equal the distribution $N(\boldsymbol{\tau}_2 + \boldsymbol{\Lambda}_2 \boldsymbol{\eta}^*, \boldsymbol{\Theta}_2)$. Clearly this is so if and only if $\boldsymbol{\tau}_k$, $\boldsymbol{\Lambda}_k$, and $\boldsymbol{\Theta}_k$ are equal over the groups: $\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \boldsymbol{\tau}$, $\boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_2 = \boldsymbol{\Lambda}$, and $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2 = \boldsymbol{\Theta}$. Only then will we have:

$\mathbf{y}_{ki} | \boldsymbol{\eta}^* \sim N(\boldsymbol{\tau} + \boldsymbol{\Lambda} \boldsymbol{\eta}^*, \boldsymbol{\Theta})$, regardless of group (i.e., in all groups).

So if we take these requires ($\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2$, $\boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_2$, $\boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2$), and consider them in the standard multi-group model, we find that MI in the linear factor model prescribes:

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda\Psi}_k\boldsymbol{\Lambda}^t + \boldsymbol{\Theta} \qquad \text{eq 1.9a}$$
$$\boldsymbol{\mu}_k = \boldsymbol{\tau} + \boldsymbol{\Lambda\alpha}_k \qquad \text{eq 1.9b}$$

If this model is tenable to reasonable approximation, the indictors are measurement invariant with respect to group. Given the context of the factor model (including the distributional assumptions), the factor model is called "strict factorially invariant". So strict factorial invariance with respect to group means measurement invariance with respect to group in the common factor model.

The derivation of MI based on conditional distributions is somewhat abstract, but its consequences are quite concrete. Specifically MI prescribes strict factorial invariance, i.e., specific equality constraints over the groups. Regardless of the derivation, we can may note that the test is MI with respect to group if the observed group differences in summary statistics (means and covariance matrix) are attributable to differences in the means and variance of the latent trait or common factor ($\boldsymbol{\Psi}_k$ and $\boldsymbol{\alpha}_k$). This is logical: if the test measures the same latent variable in the two groups, then that latent variable should be the only source of differences between the groups. The MI model (eq 1.9) may also be view as a model in which the functional relationship between the observed ($\mathbf{y}$) and latent variable ($\boldsymbol{\eta}$) is identical over the groups. To see what this means, consider this in the linear regression model (y on x; rather than $\mathbf{y}$ on $\boldsymbol{\eta}$). Specifically, consider the regressions of y on x in two groups as depicted in Figure 1-2.

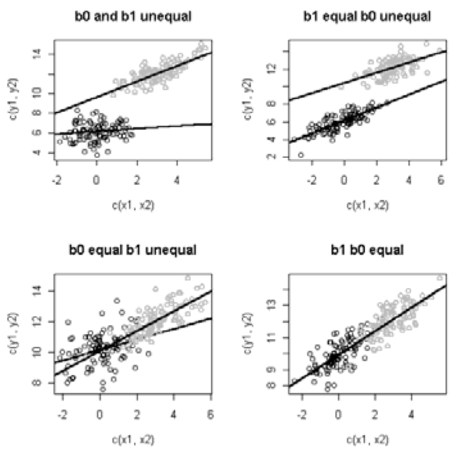


Figure 1-2: regression of y on x in two groups

In Figure 1-2, we display the scatter plot of data in two groups (black dots, gray dots) and the fitted regression line in the two groups. In the factor model, this would be the regression of a given indicator **y** on the factor(s) $\boldsymbol{\eta}$. The model is:

$$y_{ki} = b_{0k} + b_{1k}*x_{ki} + \varepsilon_{ki}$$

where k denotes group (k=1,2) and i denotes case (i=1...N_k). It is only in the bottom right figure that the parameters are equal, i.e., $b_{01} = b_{02} = b_0$, and $b_{11} = b_{12} = b_1$. So in analogy, together $\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \boldsymbol{\tau}$, and $\boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_2 = \boldsymbol{\Lambda}$ imply that the parameters of the regression of indicators on the common factor are identical.

Another consequence of measurement invariance is this. Suppose we are considering a single common factor model. If I choose a given subject from group 1, with latent variable value $\boldsymbol{\eta}^*$ and a given subject from group 2, with latent variable value $\boldsymbol{\eta}^{\bullet}$, where the values are not equal $\boldsymbol{\eta}^{\bullet} \neq \boldsymbol{\eta}^*$. Consider first the situation in which measurement invariance does not hold. For the difference in expected conditional mean (conditional on the latent variable value), we have

$$E[\mathbf{y}_{1i}|\boldsymbol{\eta}_1 = \boldsymbol{\eta}^*] = \boldsymbol{\tau}_1 + \boldsymbol{\Lambda}_1\boldsymbol{\eta}^*$$
$$E[\mathbf{y}_{2i}|\boldsymbol{\eta}_2 = \boldsymbol{\eta}^{\bullet}] = \boldsymbol{\tau}_2 + \boldsymbol{\Lambda}_2\boldsymbol{\eta}^{\bullet}$$

and the difference is complicated:

$$E[\mathbf{y}_{1i}|\boldsymbol{\eta}_k = \boldsymbol{\eta}^*] - E[\mathbf{y}_{1i}|\boldsymbol{\eta}_k = \boldsymbol{\eta}^{\bullet}] = (\boldsymbol{\tau}_1 + \boldsymbol{\Lambda}_1\boldsymbol{\eta}^*) - (\boldsymbol{\tau}_2 + \boldsymbol{\Lambda}_2\boldsymbol{\eta}^{\bullet}) =$$
$$(\boldsymbol{\tau}_1 - \boldsymbol{\tau}_2 + \boldsymbol{\Lambda}_1\boldsymbol{\eta}^* - \boldsymbol{\Lambda}_2\boldsymbol{\eta}^{\bullet}).$$

This is complicated, because the difference between $E[\mathbf{y}_{2i}|\boldsymbol{\eta}_2 = \boldsymbol{\eta}^*]$ and $E[\mathbf{y}_{2i}|\boldsymbol{\eta}_2 = \boldsymbol{\eta}^{\bullet}]$ depends on the parameters $\boldsymbol{\tau}$, the factor loading $\boldsymbol{\Lambda}$, and the latent trait difference ($\boldsymbol{\eta}^{\bullet}$ vs $\boldsymbol{\eta}^*$). Now consider the same comparison, but subject to strict factorial invariance:

$$E[\mathbf{y}_{1i}|\boldsymbol{\eta}_1 = \boldsymbol{\eta}^*] = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}^*$$
$$E[\mathbf{y}_{2i}|\boldsymbol{\eta}_2 = \boldsymbol{\eta}^{\bullet}] = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}^{\bullet}$$

and the difference:

$E[\mathbf{y}_{1i}|\eta_k = \eta^*] - E[\mathbf{y}_{1i}|\eta_k = \eta^\bullet] = (\tau + \Lambda\eta^*) - (\tau + \Lambda\eta^\bullet) =$
$(\tau - \tau + \Lambda\eta^* - \Lambda\eta^\bullet) = \Lambda(\eta^* - \eta^\bullet)$.

The difference now only depends on the latent trait. This is consistent with the idea of measuring the same latent variable in both groups. If you compare individuals with the same latent trait value, then the differences in their expected values should depend only on the latent trait, and on nothing else: $\Lambda(\eta^* - \eta^\bullet)$.

Box 1-4. Conditional means & systematic differences
Given the distributional assumption of normality in the linear factor model we have the unconditional distribution in group k:
$\mathbf{y}_{ki} \sim N(\tau_k + \Lambda_k\alpha_k, \Lambda_k\Psi_k\Lambda_k^t + \Theta_k)$,
and the conditional distribution in group k:
$\mathbf{y}_{ki} | \eta^* \sim N(\tau_k + \Lambda_k\eta^*, \Theta_k)$,
Note that $\mathbf{y}_{ki}|\eta_{ki}=\eta^* = \tau_k + \Lambda_k \eta^* + \varepsilon_{ki}$.
In comparing two subject (i=1,2), with the same latent trait value (η^*) we have:
$\mathbf{y}_{k1}|\eta_{k1}=\eta^* = \tau_k + \Lambda_k \eta^* + \varepsilon_{1i}$
$\mathbf{y}_{k2}|\eta_{k2}=\eta^* = \tau_k + \Lambda_k \eta^* + \varepsilon_{2i}$
These subjects will differ as follows
$\mathbf{y}_{k1}|\eta_{k1}=\eta^* - \mathbf{y}_{k2}|\eta_{k2}=\eta^* = (\tau_k + \Lambda_k \eta^* + \varepsilon_{1i}) - (\tau_k + \Lambda_k \eta^* + \varepsilon_{2i}) = (\varepsilon_{1i} - \varepsilon_{2i})$
This difference is solely a function of error. How will the subjects differ systematically? To answer this question I consider the conditional mean:
$E[\mathbf{y}_{k1}|\eta_{k1}=\eta^*] - E[\mathbf{y}_{k2}|\eta_{k2}=\eta^*] = (\tau_k + \Lambda_k \eta^*) - (\tau_k + \Lambda_k \eta^*) = 0$.
But what does the conditional mean actually represent. You can consider it the means of all subject with $\eta_{ki}=\eta^*$. Or, in a thought experiment, the mean of the scores of a given subject who is tested repeatedly and (ahem...) brainwashed between testing. In theory the expected means of the conditional values alow me to consider the systematic part of the scores (the error is averaged out): $\mathbf{y}_{k2}|\eta_{k2}=\eta^* = \tau_k + \Lambda_k \eta^* + \varepsilon_{2i}$ vs. $E[\mathbf{y}_{k2}|\eta_{k2}=\eta^*] = (\tau_k + \Lambda_k \eta^*)$. And it allows me to consider systematic differences between subjects.

So far have consider measurement invariance of a psychometric test with respect to group. The definition is more general than that however. We can define MI with respect to any variable X:

Definition of MI: $f(\mathbf{y}_i | \eta^*) = f(\mathbf{y}_i | \eta^* \,\&\, X^*)$,

for all values of η and X (about X represented group). That is, if and only if the conditional distribution of $\mathbf{y}$ given (conditional on) η^* (a fixed value of η), equals the conditional distribution of $\mathbf{y}$ given X* and η^* (fixed values of η and X), are the indicators y measurement invariant with respect to X. MI can also be viewed from the perspective of a causal model. That is, the definition implies that that the relationship between X (external variable) and **y** (the indicators) is mediated by **η** (the common factor). Specifically conditioning on **η** will be equivalent to conditioning on X and **η** if and only if the relationship between the indicators y and X is mediated by **η**. We can represent this in a path diagram.

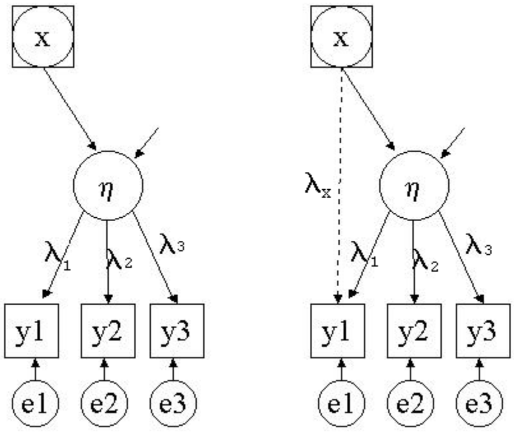


Figure 1-3: Left, the relationship between X and **y** is mediated by η; the test consisting of the indicators y1, y2, and y3 is measurement invariant with respect to X. Right, the direct relationship between X and y1 constitutes a violation of measurement invariance. Specifically, the relationship between X and y1, y2, y3 is not complete mediated by η.

Measurement invariance in the linear factor model: fitting strategy.

We will now consider the practicalities of actually fitting this model. We shall assume that we have obtained a data set in several groups, and that we want to establish measurement invariance with respect to group. We consider a number of increasingly restrictive models. Note that these models are nested, i.e., that each model can be derived from the next model by the imposition of parameter constraints (i.e., equality constraints). This implies that the constraints associated with each model can be tested by means of a likelihood ratio (or log likelihood difference) test. We start off with configural invariance. To ease presentation, we consider just two model.

Model #1: Configural invariance

We fit a two group model, but do not introduce any equality constraints over the groups (example given above).

$\mathbf{\Sigma}_k = \mathbf{\Lambda}_k \ \mathbf{\Psi}_k \ \mathbf{\Lambda}_k^t + \mathbf{\Theta}_k$

$\boldsymbol{\mu}_k = \boldsymbol{\tau}_k + \mathbf{\Lambda}_k \ \boldsymbol{\alpha}_k$

We fit a two group model, but do not introduce any equality constraints over the groups. We do assume that the pattern or configuration of **Λ** and **Θ** are the same, i.e., in the two groups the same indicators load on the same factors. E.g., letting superscripts to denote group,

$$\Lambda_1 = \begin{matrix} \lambda^1_{11} & 0 \\ \lambda^1_{21} & 0 \\ 0 & \lambda^1_{32} \\ 0 & \lambda^1_{42} \end{matrix} \qquad \Lambda_2 = \begin{matrix} \lambda^2_{11} & 0 \\ \lambda^2_{21} & 0 \\ 0 & \lambda^2_{32} \\ 0 & \lambda^2_{42}. \end{matrix}$$

This model actually comprises two independent factor models and thus requires the usual identifying constraints which pertain to a single group factor model. We scale in $\boldsymbol{\Lambda}$ so that we can estimate the factor variances, and fix the factor means to zero.

$$\Lambda_k = \begin{matrix} 1 & 0 \\ \lambda^k_{21} & 0 \\ 0 & 1 \\ 0 & \lambda^k_{42} \end{matrix}$$

$\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda}_k \boldsymbol{\Psi}_k \boldsymbol{\Lambda}_k^t + \boldsymbol{\Theta}_k$

$\boldsymbol{\mu}_k = \boldsymbol{\tau}_k$

Consider the conditional statistics:

$E[\mathbf{y}_{ki} | \boldsymbol{\eta}_{ki}=\boldsymbol{\eta}^*] = \boldsymbol{\tau}_k + \boldsymbol{\Lambda}_k \boldsymbol{\eta}^*$

$cov[\mathbf{y}_{ki} | \boldsymbol{\eta}_{ki}=\boldsymbol{\eta}^*] = cov[\mathbf{y}_{ki} | \boldsymbol{\eta}_{ki}=\boldsymbol{\eta}^*] = \boldsymbol{\Theta}_k$

and note that $E[\mathbf{y}_{1i} | \boldsymbol{\eta}_{1i}=\boldsymbol{\eta}^*] \neq E[\mathbf{y}_{2i} | \boldsymbol{\eta}_{2i}=\boldsymbol{\eta}^*]$ (because of $\boldsymbol{\tau}_1 \neq \boldsymbol{\tau}_2$ & $\boldsymbol{\Lambda}_1 \neq \boldsymbol{\Lambda}_2$).

Model #2: Equal factor loadings (metric invariance).
In the second step towards establishing measurement invariance, we constrain the factor loadings to be equal:

$\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda} \boldsymbol{\Psi}_k \boldsymbol{\Lambda}^t + \boldsymbol{\Theta}_k$

$\boldsymbol{\mu}_k = \boldsymbol{\tau}_k$

$E[\mathbf{y}_{ki} | \boldsymbol{\eta}_{ki}=\boldsymbol{\eta}^*] = \boldsymbol{\tau}_k + \boldsymbol{\Lambda} \boldsymbol{\eta}^*$

$cov[\mathbf{y}_{ki} | \boldsymbol{\eta}_{ki}=\boldsymbol{\eta}^*] = cov[\mathbf{y}_{ki} | \boldsymbol{\eta}^*] = \boldsymbol{\Theta}_k$

$E[\mathbf{y}_{1i} | \boldsymbol{\eta}_{1i}=\boldsymbol{\eta}^*] \neq E[\mathbf{y}_{2i} | \boldsymbol{\eta}_{2i}=\boldsymbol{\eta}^*]$ (because of $\boldsymbol{\tau}_1 \neq \boldsymbol{\tau}_2$).

Model #2 is nested under model #3; the difference in DFs equals....?

Model #3: Equal factor loadings & structured means (strong factorial invariance).
With equal factor loadings we can introduce a model for the means by setting the intercepts $\boldsymbol{\tau}_k$ equal:

$\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda} \boldsymbol{\Psi}_k \boldsymbol{\Lambda}^t + \boldsymbol{\Theta}_k$

$\boldsymbol{\mu}_1 = \boldsymbol{\tau} + \boldsymbol{\Lambda} \boldsymbol{\alpha}_1$

$\boldsymbol{\mu}_2 = \boldsymbol{\tau} + \boldsymbol{\Lambda} \boldsymbol{\alpha}_2$

This model seems to be identified. Suppose we have 1 factor and 4 indicators. We would then have 8 observed means (4 in two groups) and only 4 ($\boldsymbol{\tau}_y$) +2 ($\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$) parameters. However, the model is not identified. We are estimating two factor means, but as ever we cannot estimate means of latent variables (they are latent: this is a scaling problem). There is a simple solution to the problem: fixed the factor mean to equal zero in one group (the reference group, say group 1):

$$\boldsymbol{\mu}_1 = \boldsymbol{\tau}_y + \boldsymbol{\Lambda}(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_1) = \boldsymbol{\tau}_y$$

and estimate the difference in factor mean in groups 2:

$$\boldsymbol{\mu}_2 = \boldsymbol{\tau}_y + \boldsymbol{\Lambda} \ (\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1) = \boldsymbol{\tau}_y + \boldsymbol{\Lambda}(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1) = \boldsymbol{\tau}_y + \boldsymbol{\Lambda} \ \boldsymbol{\delta}_2$$

($\boldsymbol{\delta}_2 = \boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1$ is the latent mean difference in factor means of group 1 and group 2). So we have:

$$\boldsymbol{\Sigma}_k = \boldsymbol{\Lambda} \ \boldsymbol{\Psi}_k \ \boldsymbol{\Lambda}^t + \boldsymbol{\Theta}_k$$

$$\boldsymbol{\mu}_1 = \boldsymbol{\tau} + \boldsymbol{\Lambda}(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_1) = \boldsymbol{\tau}_y$$
$$\boldsymbol{\mu}_2 = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\delta}_2$$

$$E[\mathbf{y}_{ki}|\boldsymbol{\eta}_{ki}=\boldsymbol{\eta}^*] = \boldsymbol{\tau} + \boldsymbol{\Lambda}\,\boldsymbol{\eta}^*$$
$$\text{cov}[\mathbf{y}_{ki}|\boldsymbol{\eta}_{ki}=\boldsymbol{\eta}^*] = \boldsymbol{\Theta}_k$$

note now: $E[\mathbf{y}_{1i}|\boldsymbol{\eta}_{1i}=\boldsymbol{\eta}^*] = E[\mathbf{y}_{2i}|\ \boldsymbol{\eta}_{1i}=\boldsymbol{\eta}^*]$, but
$\text{cov}[\mathbf{y}_{1i}|\boldsymbol{\eta}_{1i}=\boldsymbol{\eta}^*] \neq \text{cov}[\mathbf{y}_{1i}|\boldsymbol{\eta}_{1i}=\boldsymbol{\eta}^*]$ as $\boldsymbol{\Theta}_1 \neq \boldsymbol{\Theta}_2$

Model #3 is nested under model #2; establish the difference in DFs for yourself....

Model #4: Equal factor loadings, equal residuals & structured means (strict factorial invariance).

Finally we add the constraint that the residual variances are equal:

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda} \ \boldsymbol{\Psi}_i \ \boldsymbol{\Lambda}^t + \boldsymbol{\Theta}$$

$$\boldsymbol{\mu}_1 = \boldsymbol{\tau} + \boldsymbol{\Lambda}(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_1) = \boldsymbol{\tau}_y$$
$$\boldsymbol{\mu}_2 = \boldsymbol{\tau} + \boldsymbol{\Lambda}(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1) = \boldsymbol{\tau}_y + \boldsymbol{\Lambda}\boldsymbol{\delta}_2$$

This model represents the strongest form of factorial invariance. If implies:

$$E[\mathbf{y}_{ki}|\boldsymbol{\eta}_{ki}=\boldsymbol{\eta}^*] = \boldsymbol{\tau} + \boldsymbol{\Lambda}\,\boldsymbol{\eta}^*$$
$$\text{cov}[\mathbf{y}_{ki}|\boldsymbol{\eta}_{ki}=\boldsymbol{\eta}^*] = \boldsymbol{\Theta}$$

note that now:

$E[\mathbf{y}_{1i}|\eta_{1i}=\eta^*] = E[\mathbf{y}_{2i}|\eta_{1i}=\eta^*]$, and
$cov[\mathbf{y}_{1i}|\eta_{1i}=\eta^*] = cov[\mathbf{y}_{1i}|\eta_{1i}=\eta^*]$,

More generally if we assume multivariate normality, strict factorial invariance satisfies the requirement:

$f(\mathbf{y}_{ki}|\eta_{ki}=\eta^*) = f(\mathbf{y}_{ki}|\eta_{ki}=\eta^* \ \& \ k)$, or

$f(\mathbf{y}_{1i}|\eta_{1i}=\eta^*) = f(\mathbf{y}_{2i}|\eta_{2i}=\eta^*)$

The present discussion concerned the common factor model, and thus linear regression. However the requirement $f(\mathbf{y}_i|\eta_{1i}=\eta^* \ \& \ k)$ as a condition for unbiasedness (measurement invariance) is general. For instance, if **y** is a dichotomous variable, we could use the normal ogive model to link **y** to η, and arrive at the same requirements. We will return to this later in the course. Model #4 is nested under model #3.

Example
We demonstrate all models using the real data set presented above.

```
N=1868 Group 1 (white youths)
                 pc         pa         oa         ma
            --------   --------   --------   --------
      pc        8.24
      pa        2.84       8.47
      oa        3.54       3.24       9.06
      ma        2.55       2.40       2.86       9.36
        Means
                 pc         pa         oa         ma
            --------   --------   --------   --------
               10.41      10.37      10.73      10.41

N=305 Group 2 (black youths)
                 pc         pa         oa         ma
            --------   --------   --------   --------
      pc        9.18
      pa        3.40       9.18
      oa        4.39       3.68       8.76
      ma        3.51       3.12       1.81      10.37
        Means
                 pc         pa         oa         ma
            --------   --------   --------   --------
                8.12       8.10       7.89       8.39
```

Model #1: configural invariance: no constraints

```
title groups
da No=1868 ni=4 ng=2
cm sy
            8.24
            2.84       8.47
            3.54       3.24       9.06
            2.55       2.40       2.86       9.36
Me
               10.41      10.37      10.73      10.41
la
                pc         pa         oa         ma
mo ly=fu,fr ps=sy,fr te=sy,fr ne=1 ny=4 ty=fu,fr al=fu,fi
le
perfIQ
```

```
pa ly
0 1 1 1
ma ly
1 0 0 0     ! scaling in lambda
ma al
0           ! zero mean factor
pa ps
1
pa te
1
0 1
0 0 1
0 0 0 1
ou
title N=305  Group 2
da no=305 ni=4
cm sy
            9.18
            3.40       9.18
            4.39       3.68       8.76
            3.51       3.12       1.81      10.37
Me
                8.12       8.10       7.89       8.39
la
                  pc         pa         oa         ma
mo ly=fu,fr ps=sy,fr te=sy,fr ne=1 ny=4 ty=fu,fr al=fu,fi
le
  perfIQ
pa ly
0 1 1 1
ma ly
1 0 0 0
ma al
0
pa te
1
0 1
0 0 1
0 0 0 1
ou mi
```

Degrees of Freedom = 4
Minimum Fit Function Chi-Square = 14.96 (P = 0.0048)
Root Mean Square Error of Approximation (RMSEA) = 0.047
Non-Normed Fit Index (NNFI) = 0.97

Does not fit very well judging by the chi2, but N is large. The NNFI and RMSEA both suggest that the model fits well enough. Modification indices in group 2 are:

Modification Indices for THETA-EPS

	pc	pa	oa	ma
pc	- -			
pa	**13.31**	- -		
oa	3.56	3.37	- -	
ma	3.37	3.56	13.31	- -

Expected Change for THETA-EPS

	pc	pa	oa	ma
pc	- -			
pa	-2.68	- -		
oa	1.64	1.16	- -	
ma	1.07	0.97	-1.93	- -

The correlation between pc and pa is overestimated in this group. But we will accept the model as it stands.

MODEL #2 Metric invariance.

```
 title groups
da No=1868 ni=4 ng=2
cm sy
            8.24
            2.84       8.47
            3.54       3.24       9.06
            2.55       2.40       2.86       9.36
Me
               10.41      10.37      10.73      10.41
la
                pc         pa         oa         ma
mo ly=fu,fr ps=sy,fr te=sy,fr ne=1 ny=4 ty=fu,fr al=fu,fi
le
perfIQ
pa ly
0 2 3 4        ! parameter number for equality constraints
ma ly
1 0 0 0
ma al
0
pa ps
1
pa te
1
0 1
0 0 1
0 0 0 1
ou
title N=305  Group 2
da no=305 ni=4
cm sy
             9.18
            3.40       9.18
            4.39       3.68       8.76
            3.51       3.12       1.81      10.37
Me
                8.12       8.10       7.89       8.39
la
                  pc         pa         oa         ma
mo ly=fu,fr ps=sy,fr te=sy,fr ne=1 ny=4 ty=fu,fr al=fu,fi
le
  perfIQ
pa ly
0 2 3 4        ! equality constraints
ma ly
1 0 0 0
ma al
0
pa te
1
0 1
0 0 1
0 0 0 1
ou mi
```

Degrees of Freedom = 7.
Minimum Fit Function Chi-Square = 18.73 (P = 0.0091).
Root Mean Square Error of Approximation (RMSEA) = 0.036
Non-Normed Fit Index (NNFI) = 0.98.

The deterioration in fit not significant: 18.73-14.96 = 3.69, df=3, ns.

MODEL #3, strong factorial invariance

```
title groups
da No=1868 ni=4 ng=2
cm sy
            8.24
```

```
            2.84       8.47
            3.54       3.24       9.06
            2.55       2.40       2.86       9.36
Me
               10.41      10.37      10.73      10.41
la
                pc         pa         oa         ma
mo ly=fu,fr ps=sy,fr te=sy,fr ne=1 ny=4 ty=fu,fr al=fu,fi
le
perfIQ
pa ly
0 2 3 4
ma ly
1 0 0 0
ma al
0
pa ps
1
pa te
1
0 1
0 0 1
0 0 0 1
! pa ty
! 21 22 23 24
ou
title N=305  Group 2
da no=305 ni=4
cm sy
            9.18
            3.40       9.18
            4.39       3.68       8.76
            3.51       3.12       1.81      10.37
Me
                8.12       8.10       7.89       8.39
la
                  pc         pa         oa         ma
mo ly=fu,fr ps=sy,fr te=sy,fr ne=1 ny=4 ty=in al=fu,fr
! note ty=in
le
  perfIQ
pa ly
0 2 3 4
ma ly
1 0 0 0
ma al
-2
pa al  ! estimate the mean in group 2. this parameter is the mean difference!
1
pa te
1
0 1
0 0 1
0 0 0 1

! pa ty               ! i used ty=in on the mo line. parameter number is an alternative
! 21 22 23 24
ou mi
```

Degrees of Freedom = 10
Minimum Fit Function Chi-Square = 21.00 (P = 0.021)
Root Mean Square Error of Approximation (RMSEA) = 0.029
Non-Normed Fit Index (NNFI) = 0.99

MODEL #3, strict factorial invariance

```
title groups
da No=1868 ni=4 ng=2
cm sy
            8.24
            2.84       8.47
            3.54       3.24       9.06
            2.55       2.40       2.86       9.36
Me
               10.41      10.37      10.73      10.41
```

```
la
                pc         pa         oa         ma
mo ly=fu,fr ps=sy,fr te=sy,fr ne=1 ny=4 ty=fu,fr al=fu,fi
le
perfIQ
pa ly
0 2 3 4
ma ly
1 0 0 0
ma al
0
pa ps
1
pa te
1
0 1
0 0 1
0 0 0 1
ou
title N=305  Group 2
da no=305 ni=4
cm sy
             9.18
            3.40       9.18
            4.39       3.68       8.76
            3.51       3.12       1.81      10.37
Me
                8.12       8.10       7.89       8.39
la
                  pc         pa         oa         ma
mo ly=fu,fr ps=sy,fr te=in ne=1 ny=4 ty=in al=fu,fr        ! te=in
le
  perfIQ
pa ly
0 2 3 4
ma ly
1 0 0 0
ma al
-2
pa al
1
ou rs
```

Degrees of Freedom = 14
Minimum Fit Function Chi-Square = 23.52 (P = 0.052)
Root Mean Square Error of Approximation (RMSEA) = 0.023
Non-Normed Fit Index (NNFI) = 0.99

We summarize these results in Table 1-4.

Table 1-4:
Summary of model fits

Model	df	chi2	rmsea	nnfi
#1 (conf)	4	14.9	.047	.97
#2 (metric)	7	18.7	.036	.98
#3 (strong fi)	10	21.0	.029	.99
#4 (strict fi)	14	23.5	.023	.99

Comparisons loglikelihood differences

Models	df	chi2	
#1 vs #2	3	3.8	ns
#2 vs #3	3	2.3	ns
#3 vs #4	4	2.5	ns

Note that I have limited the goodness of fit measure to the chi2, rmsea, and nnfi. Of course, other indices can be used such as the information criteria CAIC or ECVI, or other incremental fit indices, such as CFI.

Based on these results, we accept model #4, i.e., the strict factorial invariance model. This implies that the 4 indicators of Performal IQ are unbiased with respect to group, i.e., we are measuring the same construct in the two groups. Note that we have thus explained the observed group differences in means by positing a latent group difference. Note that this does not tells us anything about the cause of the latent group difference!

Latent means and variances in model #4.

whites	mean	0	variance	3.11 (se. .23)
black	mean	-2.43 (se. .15)	variance	3.54 (se. .45)

Figure 1-4 depicts the latent normal distributions, based on these parameter estimates.

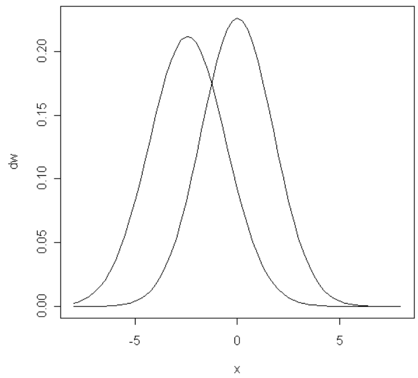


Figure 1-4: distribution of performal IQ in black and whites youths.

Multiple factor model.

Above we fitted a single factor model. But the test of measurement invariance can be carried out equally well in the multiple factor model. Specifically the implications of MI for the multi-group factor model are

$$\Sigma_k = \Lambda \ \Psi_k \ \Lambda^t + \Theta$$

$$\mu_k = \tau + \Lambda \ \alpha_k$$

i.e., strict factorial invariance. This model does not limit the dimensionality of η in anyway, i.e., the model is an n_e common factor model, where $n_e = 1$ or possible $n_e > 1$.

In fitting a multiple factor model in multiple groups, we can again carry out the analysis in the steps outlined above. That is, we can fit the configural invariance model, the metric invariance model, and the strong and strict factorial models. If the strong of strict factorial models fit, we will be modeling the ny observed means differences as a function of the differences in ne common factor models. To illustrate this consider the following data obtained from the WISC US norm data.

Table 1-5: summary stats WISC US norm data (see also appendix)

```
summary stats N=1868
white correlations
1.00
.58 1.00
.51 .43 1.00
.66 .63 .48 1.00
.51 .55 .40 .61 1.00
.34 .33 .42 .36 .23 1.00
.25 .19 .32 .24 .19 .37 1.00
.35 .40 .30 .38 .35 .16 .16 1.00
.37 .37 .26 .39 .34 .18 .19 .34 1.00
.44 .45 .41 .43 .38 .29 .27 .47 .41 1.00
.34 .35 .23 .33 .29 .17 .15 .41 .37 .56  1.00
.26 .25 .29 .29 .23 .28 .25 .15 .22 .30 .20 1.00
.22 .24 .24 .21 .23 .18 .19 .29 .27 .39 .31 .18 1.00
white means
10.41 10.29 10.37 10.42 10.44 10.08 10.09 10.41 10.37 10.39 10.73 10.22 10.41
white stdevs
2.91 3.01 2.84 2.94 2.81 3.00 2.87 2.87 2.91 2.92 3.01 3.30 3.06

summary stats N=305
black correlations
1.00
.55 1.00
.53 .46 1.00
.63 .65 .52 1.00
.49 .48 .39 .63  1.00
.43 .34 .50 .41 .35 1.00
.32 .21 .30 .25 .24 .43 1.00
.42 .43 .32 .43 .44 .28 .29 1.00
.29 .36 .23 .36 .38 .30 .26 .37 1.00
.37 .41 .40 .41 .38 .35 .26 .48 .37 1.00
.31 .36 .28 .34 .35 .25 .17 .49 .41 .57 1.00
.21 .26 .28 .28 .26 .25 .25 .16 .21 .43 .39 1.00
.26 .24 .22 .25 .30 .28 .26 .36 .32 .29 .19 .18 1.00
black means
8.09 7.91 8.63 7.86 7.83 9.18 9.12 8.12 8.10 7.70 7.89 8.86 8.39
black stdevs
2.65 2.92 2.75 2.76 2.53 3.19 2.95 3.03 3.03 2.70 2.96 2.93 3.22
```

Here is the lisrel input file for the configural invariance model.

```
title jensen and reynolds 1982
title MODEL A1.
da no=1868 ng=2 ni=13
km fi=reyn.wh
me fi=reyn.wh
sd fi=reyn.wh
la
 i s a v c ds ts pc pa bd oa co ma
se
 i s a v c ds ts pc pa bd oa co ma /
mo ny=13 ne=3 ly=fu,fr ps=sy,fr te=di,fr ty=fu,fr al=fu,fi
ma al
0 0 0
le
v p m
pa ps
1
1 1
1 1 1
ma ps
0
0 0
0 0 0
ma ly
```

```
0 0 0
0 0 0
0 0 0
1 0 0
0 0 0
0 0 0
0 0 1
0 0 0
0 0 0
0 0 0
0 1 0
0 0 0
0 0 0
pa ly
1 0 1
1 1 0
1 0 1
      0 0 0
1 1 0
0 0 1
      0 0 0
1 1 0
1 1 0
0 1 1
      0 0 0
0 1 1
0 1 1
st 1 all
st .4 ps(2,1) ps(3,1) ps(3,2)
st 10 ty(1)-ty(13)
st 3 te(1)-te(13)
ou rs ad=off it=9999 nd=3 XM MI
title jensen and reynolds 1982
title MODEL A1.
da no=305
km fi=reyn.bl
me fi=reyn.bl
sd fi=reyn.bl
la
 i s a v c ds ts pc pa bd oa co ma
se
 i s a v c ds ts pc pa bd oa co ma /
mo ly=fu,fr ps=sy,fr te=di,fr ty=fu,fr al=fu,fi
le
 v p m
ma al
0 0 0
ma ly
0 0 0
0 0 0
0 0 0
1 0 0
0 0 0
0 0 0
0 0 1
0 0 0
0 0 0
0 0 0
0 1 0
0 0 0
0 0 0
pa ly
1 0 1
1 1 0
1 0 1
      0 0 0
1 1 0
0 0 1
      0 0 0
1 1 0
1 1 0
0 1 1
      0 0 0
0 1 1
0 1 1
st .4 ps(2,1) ps(3,1) ps(3,2)
st 10 ty(1)-ty(13)
```

```
st 5 te(1)-te(13)
ou rs
```

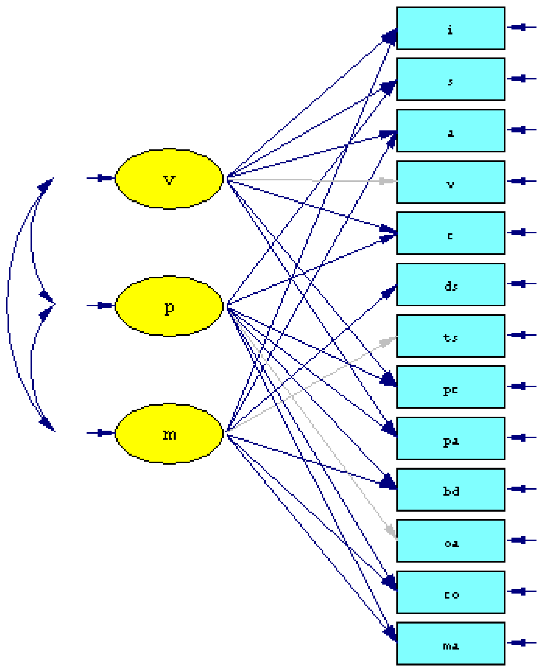


Figure 1-5: Path diagram of model for WISC (scaling in Lambda).

Assignment #1: complete the following table.

Summary of model fits

Model	df	chi2	rmsea	nnfi	CAIC
#1 (conf)	106	239.9	.034	0.991	1124.1
#2 (metric)					
#3 (strong fi)					
#4 (strict fi)					

Comparisons loglikelihood differences

Models	df	chi2
#1 vs #2		
#2 vs #3		
#3 vs #4		

Report the latent mean differences. Are they significant?

Second order factor model

General intelligence, or “g”, is an important construct in IQ research. “g” is viewed as a latent variable that permeates all IQ tests, and thus gives rise to the so-called positive manifold. The positive manifold is a complicated name for a simple phenomenon: correlation among tests of cognitive abilities are almost always positive.

In confirmatory factor analysis of IQ test scores, “g” may be fitted in two ways, 1) as a first order general factor (the “bifactor model”); and 2) as a second order general factor (the “second order factor model”). We consider the latter here in a single group without means. The data are WAIS-III data obtained in a sample of 164 young adult males. We start with an oblique 4 factor model, where we specify the expected factor structure. The common factors are Vocabulary (VO), Perceptual Organization (PO), Working Memory (WM), and Processing Speed (PS). We shall not introduce the means.

```
title young men
! WAIS III
da ng=1 ni=14 no=164 ma=cm
cm sy
    76.213
    32.223     29.376
    14.231      8.457     11.022
    13.135      8.542      5.720     15.761
    28.908     15.727      9.265      7.061     23.232
    31.307     17.711      8.249      5.657     16.128     27.248
     9.793      4.680      4.180      6.621      5.381      5.027
     7.236
    11.654      7.403      4.227      3.656      6.166      6.516
     2.269      9.548
    61.185     36.239     19.663     17.247     24.130     31.713
    14.927     10.278    287.981
    41.354     22.380     15.483     11.234     25.913     21.390
    10.121     10.751     51.575    109.830
    12.261      9.364      6.756      4.736      7.885      8.352
     3.059      5.080     16.589     18.231     15.920
    17.479     10.829      4.566      4.962      9.812     10.277
     4.273      4.663     25.253     21.684      6.488     17.472
    28.582     14.398      9.269      9.169     13.253     14.308
     7.260      8.391     61.877     34.134     13.941      7.462
    71.572
    31.361     19.759     14.341     11.163     20.864     18.058
     9.162     10.114     49.372     59.121     13.555     17.648
    25.364     78.677
la
    VOCAB SIM ARIT DIGIT INFORM COMPRE LN PC COD BD MATRIC PA SS OA
mo ny=14 ne=4 ly=fu,fi ps=sy,fi te=sy,fi
le
VO PO WM PS
fr te 1 1 te 2 2 te 3 3 te 4 4 te 5 5 te 6 6 te 7 7
fr te 8 8 te 9 9 te 10 10 te 11 11 te 12 12 te 13 13 te 14 14
fr te 14 10
pa ly
1 0 0 0 ! voc
1 0 0 0 ! sim
0 1 1 0 ! arit
0 0 1 0 ! digit
1 0 0 0 ! inform
1 0 0 0 ! compre
0 0 1 0 ! ln
0 1 0 0 ! pc
0 0 0 1 ! cod
0 1 0 0 ! bd
0 1 0 0 ! ma
0 1 0 0 ! pa
0 0 0 1 ! ss
0 1 0 0 ! oa
! scaling in Ly
va 1 ly 1 1 ly 8 2 ly 4 3 ly 9 4
fi ly 1 1 ly 8 2 ly 4 3 ly 9 4
pa ps
1
1 1
1 1 1
1 1 1 1
st 3 all
st 1 ps 1 1 -  ps 4 4
st 40 ps 1 1 ps 2 2 ps 3 3 ps 4 4
st 10 te 1 1 te 2 2 te 3 3 te 4 4 te 5 5 te 6 6 te 7 7
st 10 te 8 8 te 9 9 te 10 10 te 11 11 te 12 12 te 13 13 te 14 14
ou rs mi nd=3 ad=off ss
```

The model fits quite well both in terms of the following fit indices:

```
                     Degrees of Freedom = 69
      Minimum Fit Function Chi-Square = 91.209 (P = 0.0380)
     Root Mean Square Error of Approximation (RMSEA) = 0.0383
                Non-Normed Fit Index (NNFI) = 0.970
                Comparative Fit Index (CFI) = 0.978
```

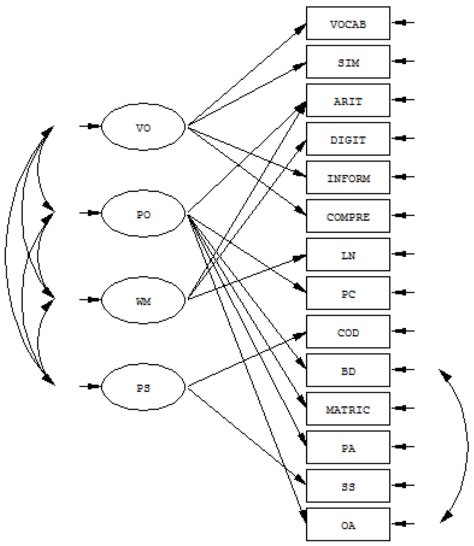
and in terms of the standardized residuals:

```
 - 3|4
 - 2|8
 - 2|20
 - 1|766
 - 1|44433110
 - 0|9988776666665555
 - 0|4444311100000000000000000
   0|1111122233333333344444
   0|56778899
   1|0000122334444
   1|68
   2|02
   2|6
```

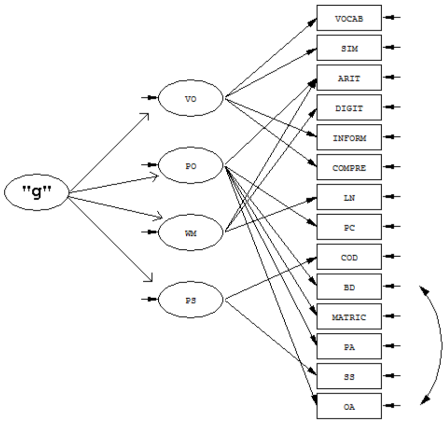
The correlations among the 1st order factors are

	VO	PO	WM	PS
VO	1.000			
PO	0.844	1.000		
WM	0.578	0.630	1.000	
PS	0.663	0.731	0.571	1.000

The second order model includes a common factor upon which the common factor VO, PO, WM and PS load. We specify this in LISREL as follows. Here is the oblique common factor model:



We introduce a 5th common factor, denoted "g".



The first order common factors load on "g", but no observed variable loads directly on "g". Thus in this model "g" does influence all observed variables, but this influence runs via the 1st order common factors. We have

$\eta_{PS} = \beta_{g,\ PS} + \zeta_{PS}$

$\eta_{WM} = \beta_{g,\ WM} + \zeta_{WM}$

$\eta_{PO} = \beta_{g,\ PO} + \zeta_{PO}$

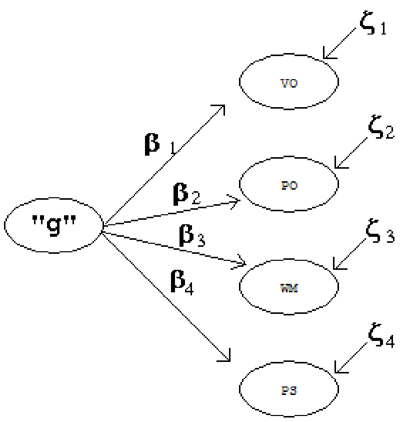
$\eta_{VO} = \beta_{g,\ VO} + \zeta_{VO}$

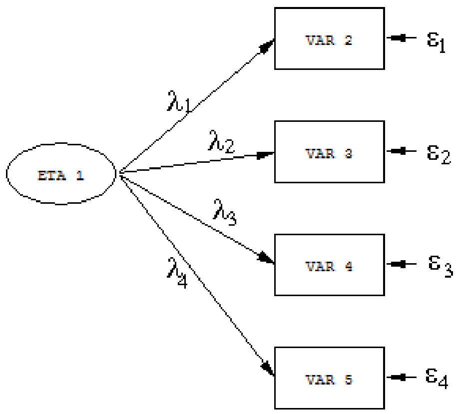
We specify the g factor as follows:

```
title young men
da ng=1 ni=14 no=164 ma=cm
cm sy
    76.213
    32.223    29.376
    14.231     8.457    11.022
    13.135     8.542     5.720    15.761
    28.908    15.727     9.265     7.061    23.232
    31.307    17.711     8.249     5.657    16.128    27.248
     9.793     4.680     4.180     6.621     5.381     5.027
     7.236
    11.654     7.403     4.227     3.656     6.166     6.516
     2.269     9.548
    61.185    36.239    19.663    17.247    24.130    31.713
    14.927    10.278   287.981
    41.354    22.380    15.483    11.234    25.913    21.390
    10.121    10.751    51.575   109.830
    12.261     9.364     6.756     4.736     7.885     8.352
     3.059     5.080    16.589    18.231    15.920
    17.479    10.829     4.566     4.962     9.812    10.277
     4.273     4.663    25.253    21.684     6.488    17.472
    28.582    14.398     9.269     9.169    13.253    14.308
     7.260     8.391    61.877    34.134    13.941     7.462
    71.572
    31.361    19.759    14.341    11.163    20.864    18.058
     9.162    10.114    49.372    59.121    13.555    17.648
    25.364    78.677
la
    VOCAB SIM ARIT DIGIT INFORM COMPRE LN PC COD BD MATRIC PA SS OA
mo ny=14 ne=5 ly=fu,fi ps=sy,fi te=sy,fi be=fu,fr
le
VO PO WM PS g
fr te 1 1 te 2 2 te 3 3 te 4 4 te 5 5 te 6 6 te 7 7
fr te 8 8 te 9 9 te 10 10 te 11 11 te 12 12 te 13 13 te 14 14
fr te 14 10
! note that no indicator loads on the 5th factor.
pa ly
1 0 0 0 0! voc
```

```
1 0 0 0 0! sim
0 1 1 0 0! arit
0 0 1 0 0! digit
1 0 0 0 0! inform
1 0 0 0 0! compre
0 0 1 0 0! ln
0 1 0 0 0! pc
0 0 0 1 0! cod
0 1 0 0 0! bd
0 1 0 0 0! ma
0 1 0 0 0! pa
0 0 0 1 0! ss
0 1 0 0 0! oa
! scaling in Ly
va 1 ly 1 1 ly 8 2 ly 4 3 ly 9 4
fi ly 1 1 ly 8 2 ly 4 3 ly 9 4
pa ps
1
0 1
0 0 1
0 0 0 1
0 0 0 0 1
!
pa be
0 0 0 0 0
0 0 0 0 1
0 0 0 0 1
0 0 0 0 1
0 0 0 0 0
va 1 be 1 5
!
st .5 all
st 1 be 2 5 be 3 5 be 4 5
st 1 ps 1 1 -  ps 4 4
st 40 ps 1 1 ps 2 2 ps 3 3 ps 4 4
st 1 ps 5 5
st 10 te 1 1 te 2 2 te 3 3 te 4 4 te 5 5 te 6 6 te 7 7
st 10 te 8 8 te 9 9 te 10 10 te 11 11 te 12 12 te 13 13 te 14 14
pd
ou rs mi nd=3 ad=off ss
```

The second order model is actually a simple single common factor model:





There is no apparent differences between these two models. Indeed both are characterized by the same problem of identification relating to scaling. In the LISREL job above, we have chose to fix the scale of the first order factors by fixing element in Λ:

```
 ! scaling in Ly
va 1 ly 1 1 ly 8 2 ly 4 3 ly 9 4
fi ly 1 1 ly 8 2 ly 4 3 ly 9 4
```

We have the same scaling to identify the variance of “g”.

```
pa be
0 0 0 0 0
0 0 0 0 1
0 0 0 0 1
0 0 0 0 1
0 0 0 0 0
va 1 be 1 5
```

$\eta_{PS} = \beta_{g,PS}\, g + \zeta_{PS}$

$\eta_{WM} = \beta_{g,WM}\, g + \zeta_{WM}$

$\eta_{PO} = \beta_{g,PO}\, g + \zeta_{PO}$

$\eta_{VO} = g + \zeta_{VO}$

Given this scaling the decomposition of variance is:

$\text{var}(\eta_{PS}) = \beta^2_{g,PS}\,\text{var}(g) + \text{var}(\zeta_{PS})$, etc.

Because we have fixed BE 1 5 to equal 1, we can estimate the variance of “g”. Of course we could also have fixed BE 2 5, BE 3 5, or BE 4 5. As in the standard single common factor model, this is arbitrary. Here are some results. First the model fits well:

```
                    Degrees of Freedom = 71
     Minimum Fit Function Chi-Square = 91.958 (P = 0.0479)
 Root Mean Square Error of Approximation (RMSEA) = 0.0360
           Non-Normed Fit Index (NNFI) = 0.973
           Comparative Fit Index (CFI) = 0.979
```

Standardized residuals are OK, as these range from about -3.4 to about 2.7 and are concentrated about zero:

```
 - 3|4
 - 2|8
 - 2|2
 - 1|9765
 - 1|444322100
 - 0|99987666655555
 - 0|4444421110000000000000000
   0|1111112222233333334444
   0|5677788889
   1|0000012233334
   1|689
   2|1
   2|7
```

We limit our discussion to the second order factor “g”. First, here are the factor loading, which are estimated in BE:

```
       BETA

                  g
            --------
      VO       1.000

      PO       0.266
             (0.041)
               6.411
```

```
WM       0.306
        (0.051)
         6.006

PS       1.303
        (0.220)
         5.920
```

Here are the residuals:

PSI

VO	PO	WM	PS	g
12.959	0.264	4.913	48.990	42.381
(4.152)	(0.282)	(1.178)	(20.115)	(8.184)
3.122	0.935	4.169	2.435	5.179

And here are the reliabilities:

Squared Multiple Correlations for Structural Equations

VO	PO	WM	PS	g
0.766	0.919	0.447	0.595	- -

These are calculated in the standard way, given that we are regressing VO, etc. on "g". For instance the reliability of VO is:

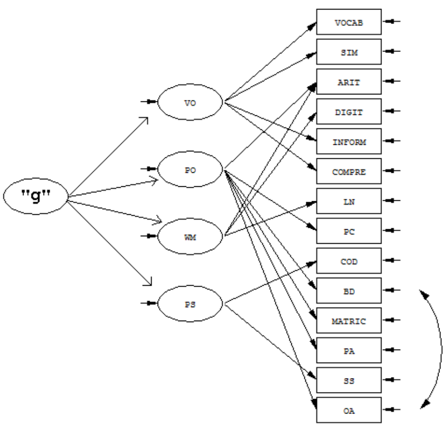
$.766 = (42.381)/(42.381+12.959)$

The reliability of PO is:

$.919=.266^2*42.381 / (.266^2*42.381+0.264)$.

It is interesting to note that "g" explains about 92% of the variance in PO. From a correlational point of view, therefore, "g" and PO are quite hard to distinguish in this model (whether this will generalize to other samples, is an open question).

In addition note that this model:



implies that ***all*** relations between the first-order factors VO, PO, WM and PS are explained by 'g'. It would of course be possible for the first-order factors to show additional relations that are not explained by 'g'. For example, VO and PO could be related beyond their common relation to 'g'. Such additional relations can be accommodated in he model by freeing elements in psi (note that these elements should be interpreted as ***covariances between the residual variances*** of the first-order factors, i.e., relations between those parts of the factors that were not explained by 'g').

An extension of the LISREL model.
Above we considered this LISREL model for the covariance matrix (k stands for group k=1...K, but above K=1):

$$\Sigma_k = \Lambda_k (I-B_k)^{-1} \Psi_k (I-B_k)^{-1t} \Lambda_k^t + \Theta_k$$

The model for the means is (not considered in the example above):

$$\mu_k = \tau_k + \Lambda_k (I-B_k)^{-1} \alpha_k$$

We used this model to fit the second order factor model. This model is quite easy to understand[3], partly because it involves only 4 parameter matrices $\mathbf{\Lambda}$, $\mathbf{B}$, $\mathbf{\Psi}$, and $\mathbf{\Theta}$. However, this is actually just a sub-model of the full LISREL model. We shall not consider the full LISREL model, but we shall consider an extension, which is useful in the light of the previous hierarchical (second-order) analysis. The LISREL model we consider is this:

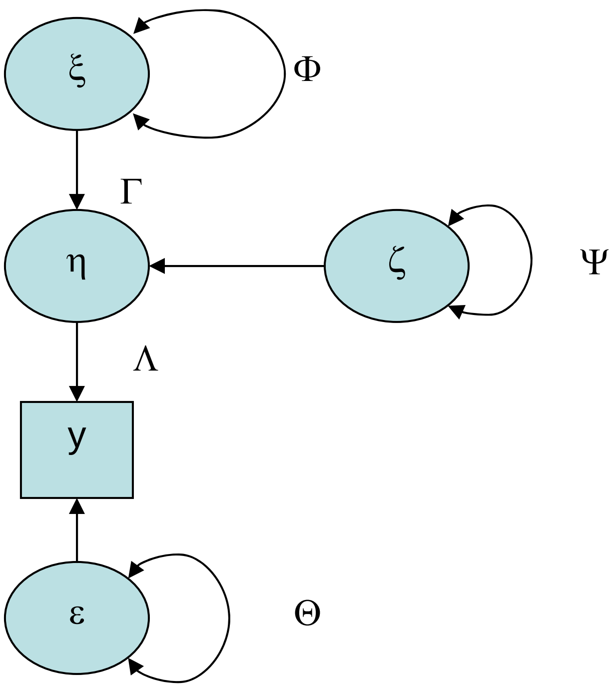
$$\Sigma_k = \Lambda_k (I-B_k)^{-1} (\Gamma_k \Phi_k \Gamma_k^t + \Psi_k) (I-B_k)^{-1t} \Lambda_k^t + \Theta_k$$
$$\mu_k = \tau_k + \Lambda_k (I-B_k)^{-1} (\alpha_k + \Gamma_k \kappa_k)$$

This is complicated! But we can simplify a little if we set $\mathbf{B}_k = \mathbf{0}$:

$$\Sigma_k = \Lambda_k (\Gamma_k \Phi_k \Gamma_k^t + \Psi_k) \Lambda_k^t + \Theta_k$$
$$\mu_k = \tau_k + \Lambda_k (\alpha_k + \Gamma_k \kappa_k) = \tau_k + \Lambda_k \alpha_k + \Lambda_k \Gamma_k \kappa_k)$$

Note the similarity between the familiar common factor model $\Sigma_k = \Lambda_k \Psi_k \Lambda_k^t + \Theta_k$ and the present extension $(\Gamma_k \Phi_k \Gamma_k^t + \Psi_k)$. This matrix $(\Gamma_k \Phi_k \Gamma_k^t + \Psi_k)$ is the covariance matrix of the common factors. Schematically, the extension can be conveyed as follows (with $\mathbf{B}_k = \mathbf{0}$):

[3] Once you have studied it many times and fitted even more times!



We have thus added a variable ξ, which is a latent predictor of η. So working forward from the common factor model we have:

equation for observations y_k =	equations for covariance matrix
$\Lambda_k\ \eta_k + \varepsilon_k$	$\Sigma_k = \Lambda_k \Psi_k \Lambda_k^t + \Theta_k$
$\Lambda_k\ (I-B_k)^{-1}\zeta_k + \varepsilon_k$	$\Sigma_k = \Lambda_k (I-B_k)^{-1} \Psi_k (I-B_k)^{-1t} \Lambda_k^t + \Theta_k$
$\Lambda_k\ (I-B_k)^{-1}(\Gamma_k\xi_k+\zeta_k) + \varepsilon_k$	$\Sigma_k = \Lambda_k (I-B_k)^{-1} (\Gamma_k\Phi_k\Gamma_k^t+\Psi_k)\ (I-B_k)^{-1t}\Lambda_k^t+\Theta_k$

Or in terms of the models for η:

equation for observations	equation for covariance matrix
$\eta_k = \eta_k$	$\Sigma_{\eta k} = \Psi_k$
$\eta_k = (I-B_k)^{-1}\zeta_k$	$\Sigma_{\eta k} = (I-B_k)^{-1}\ \Psi_k\ (I-B_k)^{-1t}$
$\eta_k = (I-B_k)^{-1}(\Gamma_k\xi_k+\zeta_k)$	$\Sigma_{\eta k} = (I-B_k)^{-1}(\Gamma_k\Phi_k\Gamma_k^t+\Psi_k)\ (I-B_k)^{-1t}$

equation for observations y_k =	equations for covariance matrix
$\Lambda_k\ \eta_k + \varepsilon_k$	$\mu_k = \tau_k + \Lambda_k\alpha_k$
$\Lambda_k\ (I-B_k)^{-1}\zeta_k + \varepsilon_k$	$\mu_k = \tau_k + \Lambda_k (I-B_k)^{-1}\alpha_k$
$\Lambda_k\ (I-B_k)^{-1}(\Gamma_k\xi_k+\zeta_k) + \varepsilon_k$	$\mu_k = \tau_k + \Lambda_k (I-B_k)^{-1}(\alpha_k + \Gamma_k\kappa_k)$

Or in terms of the models for η:

equation for observations	equation for covariance matrix
$\eta_k = \eta_k$	$\mu_{\eta k} = \alpha_k$
$\eta_k = (I-B_k)^{-1}\zeta_k$	$\mu_{\eta k} = (I-B_k)^{-1}\alpha_k$
$\eta_k = (I-B_k)^{-1}(\Gamma_k\xi_k+\zeta_k)$	$\mu_{\eta k} = (I-B_k)^{-1}(\alpha_k + \Gamma_k\kappa_k)$

Table 1-6: LISREL covariance and mean structure in k=1...K populations.

covariance structure	mean structure
$\Sigma_k = \Lambda_k (\Gamma_k \Phi_k \Gamma_k^t + \Psi_k) \Lambda_k^t + \Theta_k$	$\mu_k = \tau_k + \Lambda_k(\alpha_k + \Gamma_k\kappa_k)$

matrix	LISREL	order	meaning
Λ_k	ly	ny x ne	factor loading matrix (**y**->η)
Ψ_k	ps	ne x ne	cov/cor matrix of η or ζ
Θ_k	te	ny x ny	cov/cor matrix of residuals(ε)
Φ_k	ph	nk x nk	cov/cor matrix of ξ (2nd order factors)
Γ_k	ga	nk x nk	regression matrix (η->ξ)
Σ_k	-	ny x ny	expected model cov. matrix of **y**

vector	LISREL	dimension	meaning
τ_k	ty	ny x 1	intercept in regression of **y** on η
α_k	al	ne x 1	common factor means of η or ζ
κ_k	al	nk x 1	common factor means of ξ
$\mu_{\mathbf{y}k}$	-	ny x 1	expected means of **y**

Second order factor model, same model, different specification

Above we employed the model $\Sigma_k = \Lambda_k (I-B_k)^{-1} \Psi_k (I-B_k)^{-1t} \Lambda_k^t+\Theta_k$ to fit the second order factor model. We shall now use the following model:

$$\Sigma_k = \Lambda_k (I-B_k)^{-1} (\Gamma_k \Phi_k \Gamma_k^t + \Psi_k) (I-B_k)^{-1t} \Lambda_k^t + \Theta_k$$

to fit the second order factor model. However, now we shall not require the B_k matrix, and because we only have one group, we drop the group index:

$$\Sigma_k = \Lambda (\Gamma \Phi \Gamma^t + \Psi) \Lambda^t + \Theta$$

Here is the LISREL input for this formulation of the model:

```
title young men
!
da ng=1 ni=14 no=164 ma=cm
!
cm sy
    76.213
    32.223     29.376
    14.231      8.457     11.022
    13.135      8.542      5.720     15.761
    28.908     15.727      9.265      7.061     23.232
    31.307     17.711      8.249      5.657     16.128     27.248
     9.793      4.680      4.180      6.621      5.381      5.027
     7.236
    11.654      7.403      4.227      3.656      6.166      6.516
     2.269      9.548
    61.185     36.239     19.663     17.247     24.130     31.713
    14.927     10.278    287.981
    41.354     22.380     15.483     11.234     25.913     21.390
    10.121     10.751     51.575    109.830
    12.261      9.364      6.756      4.736      7.885      8.352
     3.059      5.080     16.589     18.231     15.920
    17.479     10.829      4.566      4.962      9.812     10.277
     4.273      4.663     25.253     21.684      6.488     17.472
    28.582     14.398      9.269      9.169     13.253     14.308
     7.260      8.391     61.877     34.134     13.941      7.462
    71.572
    31.361     19.759     14.341     11.163     20.864     18.058
     9.162     10.114     49.372     59.121     13.555     17.648
    25.364     78.677
la
    VOCAB SIM ARIT DIGIT INFORM COMPRE LN PC COD BD MATRIC PA SS OA
mo ny=14 ne=4 ly=fu,fi ps=sy,fi te=sy,fi nk=1 ga=fu,fr ph=di,fr
le
VO PO WM PS
lk
g
fr te 1 1 te 2 2 te 3 3 te 4 4 te 5 5 te 6 6 te 7 7
fr te 8 8 te 9 9 te 10 10 te 11 11 te 12 12 te 13 13 te 14 14
fr te 14 10
pa ly
1 0 0 0 ! voc
1 0 0 0 ! sim
0 1 1 0 ! arit
0 0 1 0 ! digit
1 0 0 0 ! inform
1 0 0 0 ! compre
0 0 1 0 ! ln
0 1 0 0 ! pc
0 0 0 1 ! cod
0 1 0 0 ! bd
0 1 0 0 ! ma
0 1 0 0 ! pa
0 0 0 1 ! ss
0 1 0 0 ! oa
! scaling in Ly
va 1 ly 1 1 ly 8 2 ly 4 3 ly 9 4
fi ly 1 1 ly 8 2 ly 4 3 ly 9 4
```

```
pa ps
1
0 1
0 0 1
0 0 0 1
pa ga
0 1 1 1
ma ga
1 0 0 0
pa ph
1
st .5 all
st 10 ph 1 1
st 1 ga 2 1 ga 3 1 ga 4 1
st 1 ps 1 1 -  ps 4 4
st 40 ps 1 1 ps 2 2 ps 3 3 ps 4 4
st 10 te 1 1 te 2 2 te 3 3 te 4 4 te 5 5 te 6 6 te 7 7
st 10 te 8 8 te 9 9 te 10 10 te 11 11 te 12 12 te 13 13 te 14 14
pd
ou rs mi nd=3 ad=off ss
```

Here are some results. These are identical to those shown above.

```
         GAMMA

                    g
            --------
       VO      1.000

       PO      0.266
             (0.041)
               6.411

       WM      0.306
             (0.051)
               6.006

       PS      1.303
             (0.220)
               5.920
```

> Note that these gamma-loadings are identical to the beta loadings reported above

```
         Covariance Matrix of ETA and KSI

                   VO         PO         WM         PS          g
            --------   --------   --------   --------   --------
       VO     55.340
       PO     11.262      3.257
       WM     12.969      3.446      8.881
       PS     55.208     14.671     16.895    120.906
        g     42.381     11.262     12.969     55.208     42.381

         PHI

                    g
            --------
              42.381
             (8.184)
               5.179

         PSI
         Note: This matrix is diagonal.

                   VO         PO         WM         PS
            --------   --------   --------   --------
              12.959      0.264      4.913     48.990
             (4.152)    (0.282)    (1.178)   (20.115)
               3.122      0.935      4.169      2.436
```

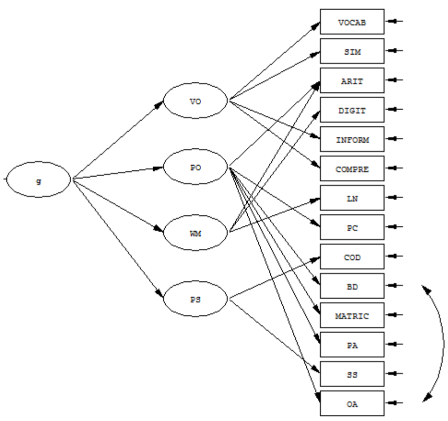
> Ps-parameters are also identical to those reported above

```
         Squared Multiple Correlations for Structural Equations

                   VO         PO         WM         PS
            --------   --------   --------   --------
```

```
      0.766       0.919       0.447       0.595

                         Degrees of Freedom = 71
            Minimum Fit Function Chi-Square = 91.958 (P = 0.0479)
```

The squared multiple correlation .919, corresponds to a correlation between PO and g of sqrt(.919)=.958, or in terms of the parameter estimates: cov(PO,g)/sqrt[var(PO)*var(g)]=

$.958=(42.381*0.266)/[\text{sqrt}(.264+42.381*0.266^2)*\text{sqrt}(42.381)]$. Here finally is the path diagram produced by LISREL.



So we have fitted this model both using $\Sigma = \Lambda(I-B)^{-1}\Psi(I-B)^{-1t}\Lambda^t+\Theta$ and using $\Sigma = \Lambda(\Gamma\Phi\Gamma^t+\Psi)\Lambda^t + \Theta$. The latter model is more suitable for a 2nd order factor model, because the specification using this model is more economical and more elegant. However, these two ways of specifying the model are equivalent.

Multi-group first order factor model: Strict Factorial Invariance.

Here is the input file for a strict factorial invariant model. The data are the same as those shown above (WISC US norm data in 1868 white and 305 black youths).

```
title jensen and reynolds 1982
title MODEL A4.
da no=1868 ng=2 ni=13
km fi=reyn.wh
me fi=reyn.wh
sd fi=reyn.wh
la
 i s a v c ds ts pc pa bd oa co ma
se
 i s a v c ds ts pc pa bd oa co ma /
mo ny=13 ne=3 ly=fu,fr ps=sy,fr te=di,fr ty=fu,fr al=fu,fi
ma al
0 0 0
le
v p m
pa ps
1
1 1
1 1 1
ma ps
0
0 0
0 0 0
ma ly
```

```
0 0 0
0 0 0
0 0 0
1 0 0
0 0 0
0 0 0
0 0 1
0 0 0
0 0 0
0 0 0
0 1 0
0 0 0
0 0 0
pa ly
1 0 1
1 1 0
1 0 1
      0 0 0
1 1 0
0 0 1
      0 0 0
1 1 0
1 1 0
0 1 1
      0 0 0
0 1 1
0 1 1
st 1 all
st .4 ps(2,1) ps(3,1) ps(3,2)
st 10 ty(1)-ty(13)
st 3 te(1)-te(13)
ou rs ad=off it=9999 nd=3 XM MI
title jensen and reynolds 1982
title MODEL A4.
da no=305
km fi=reyn.bl
me fi=reyn.bl
sd fi=reyn.bl
la
 i s a v c ds ts pc pa bd oa co ma
se
 i s a v c ds ts pc pa bd oa co ma /
mo ly=in ps=sy,fr te=in ty=in al=fu,fr
le
 v p m
ma al
0 0 0
st 1 all
st .4 ps(2,1) ps(3,1) ps(3,2)
st 10 ty(1)-ty(13)
st 5 te(1)-te(13)
st -1 al(1)-al(3)
ou rs
```

If you run this model, you will find that the model fits reasonably.

```
                              Degrees of Freedom = 148
             Minimum Fit Function Chi-Square = 327.775 (P = 0.00)
           Root Mean Square Error of Approximation (RMSEA) = 0.0327
         90 Percent Confidence Interval for RMSEA = (0.0278 ; 0.0376)
                     Non-Normed Fit Index (NNFI) = 0.991
```

The latent covariance matrix and means in the white and black samples are:

Ψ_w

Covariance Matrix of ETA

	v	p	m
v	6.290		
p	2.976	4.787	
m	2.221	1.253	2.408

α_w

```
          Mean Vector of Eta-Variables

                 v           p           m
          --------    --------    --------
             0.000*      0.000*       0.000*
```

Ψ_b

```
                 v           p           m
          --------    --------    --------
    v        5.298
    p        2.698       4.425
    m        2.323       1.544       2.532
```

α_b

```
       Mean Vector of Eta-Variables

                 v           p           m
          --------    --------    --------
            -2.634      -2.751      -0.824
```

We will now consider the 2nd order model, using the strict factorial invariance model as the baseline model. We have fitted the first order model:

$$\Sigma_w = \Lambda\Psi_w\Lambda^t + \Theta \qquad \mu_w = \tau \qquad (\alpha_w = 0)$$
$$\Sigma_b = \Lambda\Psi_b\Lambda^t + \Theta \qquad \mu_b = \tau + \Lambda\delta \qquad (\delta = \alpha_b - \alpha_w)$$

We will now fit the second order model, starting with this model:

$$\Sigma_w = \Lambda(\Gamma_w\Phi_w\Gamma_w^t + \Psi_w)\Lambda^t + \Theta \qquad \mu_w = \tau_w$$
$$\Sigma_b = \Lambda(\Gamma_b\Phi_b\Gamma_b^t + \Psi_b)\Lambda^t + \Theta \qquad \mu_w = \tau_b$$

Note that the first elements of Γ_w and Γ_b are fixed to 1. This is a scaling constraint that serves to identify the variance of the second order factor (that is the 1's serve the same purpose as the fixed 1's in the matrix Λ).

Note that the factor loading and residual covariance matrices are equal. The factor covariance matrix and the means are now unconstrained. Here is the input:

```
title jensen and reynolds 1982
title MODEL B1.
da no=1868 ng=2 ni=13
km fi=reyn.wh
me fi=reyn.wh
sd fi=reyn.wh
la
 i s a v c ds ts pc pa bd oa co ma
mo nk=1 ny=13 ne=3 ly=fu,fr ps=di,fr te=di,fr ty=fu,fr al=fu,fi  c
                          ka=fu,fi ga=fu,fr ph=sy,fr
le
 v p m
lk
 g
ma ka
0
pa ps
2 3 4
pa ph
1
ma ly
0 0 0
0 0 0
0 0 0
```

```
1 0 0
0 0 0
0 0 0
0 0 1
0 0 0
0 0 0
0 0 0
0 1 0
0 0 0
0 0 0
pa ly
1 0 1
1 1 0
1 0 1
      0 0 0
1 1 0
0 0 1
      0 0 0
1 1 0
1 1 0
0 1 1
      0 0 0
0 1 1
0 1 1
ma ga
1 0 0
pa ga
0 6 7
ma al
0 0 0
st 1 all
st .2 ga 2 1 ga 3 1
st 1 ph 1 1
st 5 te 1 - te 13
st 10 ty 1 - ty 13
ou ad=off it=500 ns rs
title jensen and reynolds 1982
title MODEL B1.
da no=305
km fi=reyn.bl
me fi=reyn.bl
sd fi=reyn.bl
la
 i s a v c ds ts pc pa bd oa co ma
mo ly=in ps=di,fr te=in ty=fu,fr ga=fu,fr al=fu,fi ka=fu,fi ph=sy,fr
le
 v p m
lk
 g
ma ga
1 0 0
pa ga
0 16 17
pa ka
0
pa ps
12 13 14
ma al
0 0 0
pa al
0 0 0
ou nd=4
```

```
Degrees of Freedom = 138
Minimum Fit Function Chi-Square = 305.7152 (P = 0.00)
```

Assignment #2:

Fit the following models:

model 2) equal Γ

$\Sigma_w = \Lambda(\Gamma\Phi_w\Gamma^t + \Psi_w)\Lambda^t + \Theta \qquad \mu_w = \tau_w$

$\Sigma_b = \Lambda(\Gamma\Phi_b\Gamma^t + \Psi_b)\Lambda^t + \Theta \qquad \mu_b = \tau_b$

```
Degrees of Freedom = 140
Minimum Fit Function Chi-Square = 308.0422 (P = 0.00)
```

model 3) equal Γ and structured means:
equal τ, α_w=0 and α_b=free, κ_w=0 and κ_b=0.

$\Sigma_w = \Lambda(\Gamma\Phi_w\Gamma^t + \Psi_w)\Lambda^t + \Theta \qquad \mu_w = \tau$

$\Sigma_b = \Lambda(\Gamma\Phi_b\Gamma^t + \Psi_b)\Lambda^t + \Theta \qquad \mu_w = \tau + \Lambda(\delta_b) \qquad (\delta_b = \alpha_b)$

```
Degrees of Freedom = 150
Minimum Fit Function Chi-Square = 330.1134 (P = 0.00)
```

model 4) equal Γ and structured means:
equal τ, α_w=0 and α_b=0, κ_w=0 and κ_b=free.

$\Sigma_w = \Lambda(\Gamma\Phi_w\Gamma^t + \Psi_w)\Lambda^t + \Theta \qquad \mu_w = \tau$

$\Sigma_b = \Lambda(\Gamma\Phi_b\Gamma^t + \Psi_b)\Lambda^t + \Theta \qquad \mu_w = \tau + \Lambda\Gamma(\delta_b) \qquad (\delta_b = \kappa_b)$

```
Degrees of Freedom = 152
Minimum Fit Function Chi-Square = 389.7424 (P = 0.0)
```

Report the results in terms of goodness of fit, and in terms of the parameters pertaining to the latent common factors (1st and 2nd order common factors). Consider the Modification indices of the parameters in al in model 4.

Evaluation of fit (see lisrel lecture notes 6).

Fit Measure	Good Fit	Acceptable Fit
χ^2	$0 \leq \chi^2 \leq 2df$	$2df < \chi^2 \leq 3df$
p value	$.05 < p \leq 1.00$	$.01 \leq p \leq .05$
χ^2/df	$0 \leq \chi^2/df \leq 2$	$2 < \chi^2/df \leq 3$
$RMSEA$	$0 \leq RMSEA \leq .05$	$.05 < RMSEA \leq .08$
p value for test of close fit ($RMSEA < .05$)	$.10 < p \leq 1.00$	$.05 \leq p \leq .10$
Confidence interval (CI)	close to $RMSEA$, left boundary of CI $= .00$	close to $RMSEA$
$SRMR$	$0 \leq SRMR \leq .05$	$.05 < SRMR \leq .10$
NFI	$.95 \leq NFI \leq 1.00$[a]	$.90 \leq NFI < .95$
$NNFI$	$.97 \leq NNFI \leq 1.00$[b]	$.95 \leq NNFI < .97$[c]
CFI	$.97 \leq CFI \leq 1.00$	$.95 \leq CFI < .97$[c]
GFI	$.95 \leq GFI \leq 1.00$	$.90 \leq GFI < .95$
$AGFI$	$.90 \leq AGFI \leq 1.00$, close to GFI	$.85 \leq AGFI < .90$, close to GFI
AIC	smaller than AIC for comparison model	
$CAIC$	smaller than $CAIC$ for comparison model	
$ECVI$	smaller than $ECVI$ for comparison model	

Modification indices.

Apart from a summary of fit indices and fitted residuals, LISREL also provides so-called modification indices (MI, obtainable by putting 'mi' on the ou-line) as information related to misspecification.
For every fixed (to zero, or constrained to another values) parameter, LISREL calculates the difference in χ^2 (the expected drop in χ^2) that is to be expected if that parameter was to be estimated freely. The MI can thus be considered a χ^2 statistic with 1 degree of freedom. (LISREL also provides table of Expected Change, which represents the predicted estimated change, in either positive or negative direction, for every fixed parameter. The expected change is however dependent on the scales of the variables, and the scaling choices, so the absolute values are difficult to interpret). At the end of the MI output LISREL prints the largest MI, i.e., the parameter that, if freely estimated, would have the largest beneficial effect on the overall χ^2 fit of the model.

So, if a model does not fit the data neatly, the MI's can be inspected to find out where the largest misfit is located. So the fit of a model can be improved by freeing the parameter(s) with the largest MI. However practical and convenient this procedure seems, there are a few concerns:

1. LISREL simply calculates the expected drop in χ^2 for every constrained parameter, and then advertises the parameter with the largest MI. One should however realize that freeing the parameter with the largest MI (or any other) might not be theoretically

sensible, wise, logical, or justified. It is therefore important to keep in mind what freeing the parameter means for the interpretation of your model (as it might undermine the main goal of your study, or create an improbable, illogical model).

2. One should not upgrade one's model endlessly by freeing one parameter after the other based on the MI's. The more parameters are freed, the more the new model deviates from the original, intended, hypothesized model. Also, the more one capitalizes on chance. Also, if the original model requires many changes, one should consider revising hypotheses and models, rather than desperately attempting to make the model fit the data.

It is possible that model fit improves if one frees the covariance between error terms. Suppose I fit a one-factor model on 4 tests, and the MI's tell me that the fit can be improved greatly by freeing the covariance between the residuals (parts of variance unexplained by the 1 factor) of test 1 and test 3. Of course this is possible: it is conceivable that test 1 and test 3 have something in common over and above their communality with test 2 and 4. One should however realize that such additional relations might be interpreted as an indication that the 1-factor model is too parsimonious, i.e., that a model with more factors actually underlies the data.

Appendix reyn.wh and reyn.bl files

reyn.wh

```
1.00
.58 1.00
.51 .43 1.00
.66 .63 .48 1.00
.51 .55 .40 .61 1.00
.34 .33 .42 .36 .23 1.00
.25 .19 .32 .24 .19 .37 1.00
.35 .40 .30 .38 .35 .16 .16 1.00
.37 .37 .26 .39 .34 .18 .19 .34 1.00
.44 .45 .41 .43 .38 .29 .27 .47 .41 1.00
.34 .35 .23 .33 .29 .17 .15 .41 .37 .56  1.00
.26 .25 .29 .29 .23 .28 .25 .15 .22 .30 .20 1.00
.22 .24 .24 .21 .23 .18 .19 .29 .27 .39 .31 .18 1.00
10.41 10.29 10.37 10.42 10.44 10.08 10.09 10.41 10.37 10.39 10.73 10.22 10.41
2.91 3.01 2.84 2.94 2.81 3.00 2.87 2.87 2.91 2.92 3.01 3.30 3.06
```

reyn.bl

```
1.00
.55 1.00
.53 .46 1.00
.63 .65 .52 1.00
.49 .48 .39 .63  1.00
.43 .34 .50 .41 .35 1.00
.32 .21 .30 .25 .24 .43 1.00
.42 .43 .32 .43 .44 .28 .29 1.00
.29 .36 .23 .36 .38 .30 .26 .37 1.00
.37 .41 .40 .41 .38 .35 .26 .48 .37 1.00
.31 .36 .28 .34 .35 .25 .17 .49 .41 .57 1.00
.21 .26 .28 .28 .26 .25 .25 .16 .21 .43 .39 1.00
.26 .24 .22 .25 .30 .28 .26 .36 .32 .29 .19 .18 1.00
8.09 7.91 8.63 7.86 7.83 9.18 9.12 8.12 8.10 7.70 7.89 8.86 8.39
2.65 2.92 2.75 2.76 2.53 3.19 2.95 3.03 3.03 2.70 2.96 2.93 3.22
```

Lecture notes II: Discete factor model [1]

references

Wirth, R. J. & Edwards, M. C. Item Factor Analysis: Current Approaches and Future Directions, Psychological Methods, Vol. 12, No. 1, 58–79.
[recent review, including a clear explanation of the relation between discrete factor analysis and model from item response theory]

Summary

The aim of the present lecture notes is to introduce the discrete factor model. You are familiar with the standard linear factor model, in which continuously distributed observed variables (indicators) are related to common factors (latent traits) by means of a linear regression model. In this case the common factors are the independent variables and the indicators are the dependent variables. We retain the assumption of continuous common factors, but now switch from continuous indicators to discrete, ordinal indicators. We consider indicators discrete if the response format comprises less than 7 ordered response categories. For instance a three point scale has three ordered response categories, and thus definitely counts as a discrete indicator. Often indicators are dichotomous ("yes / no", "agree / disagree", "correct / incorrect"). Discrete dependent variables cannot be analyzed using the linear model. In standard regression, one uses probit or logit regression in the case of dichotomous dependent variables. Logit & probit regression can be carried out in SPSS. The discrete factor model is based on the probit method. Once the discrete factor model has been introduced, we will return to the theme of measurement invariance (in the next lecture notes).

Software.

We are going to make a switch from LISREL to Mplus. Please go to the Mplus site, download and install the student version of Mplus. http://www.statmodel.com/demo.shtml. Mplus employs a completely different syntax. However, as we shall limit all subsequent models to the 1 or 2 common factor models, I will explain the syntax using examples.

Unweighted Least Squares.

In all analyses caried out so far, the parameters were estimated by the method of maximum likelihood (ML) estimation. This is the most important (certainly most fequently used) estimation technique in LISREL/SEM modeling. However, ML is limited to multivariate normal data. We will now consider another general method of estimation that is based on the principle of least squares minimization, which is important in the analysis of discrete data. We first explain the principle of unweighted least squares (ULS). Let $\mathbf{S}$ denote the p x p observed covariance matrix, and $\mathbf{\Sigma}(\boldsymbol{\theta})$ the pxp model matrix, where the model is the following LISREL submodel (we shall discard $\mathbf{B}$ for now):

$$\mathbf{\Sigma}(\boldsymbol{\theta}) = \mathbf{\Lambda\Psi\Lambda}^t + \mathbf{\Theta},$$

[1] Conor V. Dolan c.v.dolan@uva.nl. RM20. MI: continuous & discrete factor models.

i.e., the common factor model. In ULS estimation we minimize the squared difference between the observed matrix **S** and the model matrix $\Sigma(\theta)$, where θ contains the unknown parameters in the model. The ULS function is:

$$F_{uls}(\theta) = ½\ \mathrm{trace}[\{\mathbf{S}-\Sigma(\theta)\}^2] \qquad \text{eq 2-1}$$ [2]

The minimization problem is to find values of the unknown parameters (collected in q), that minimize the function. As it stands, this function is not very clear. So let's consider a small example: let **S** equal

$$\mathbf{S} = \begin{matrix} a & b \\ b & c, \end{matrix}$$

and $\Sigma(\theta)$ equal

$$\Sigma(\theta) = \begin{matrix} \alpha & \beta \\ \beta & \gamma. \end{matrix}$$

So $\{\mathbf{S}-\Sigma(\theta)\}$ equals

$$\begin{matrix} a-\alpha & b-\beta \\ b-\beta & c-\gamma, \end{matrix}$$

and $\{\mathbf{S}-\Sigma(\theta)\}^2$ equals

$$\begin{matrix} (a-\alpha)^2 + (b-\beta)^2 & (a-\alpha)*(b-\beta)+(c-\gamma)*(b-\beta) \\ (a-\alpha)*(b-\beta)+(c-\gamma)*(b-\beta) & (b-\beta)^2+(c-\gamma)^2 \end{matrix}$$

$½\mathrm{trace}(\{\mathbf{S}-\Sigma(\theta)\}^2)$ equals $½([(a-\alpha)^2+(b-\beta)^2]+[(b-\beta)^2+(c-\gamma)^2])$,

i.e., the sum of the squared differences between the elements of **S** and $\Sigma(\theta)$. Or, in the case of a single factor model with three indicators, $\Sigma(\theta) = \Lambda\Psi\Lambda^t + \Theta$ (Ψ=1, for scaling):

$$\Sigma(\theta) = \begin{matrix} \lambda_1^2+\sigma_{\varepsilon 1}^2 & \lambda_2\lambda_1 & \lambda_3\lambda_1 \\ \lambda_2\lambda_1 & \lambda_2^2+\sigma_{\varepsilon 2}^2 & \lambda_3\lambda_2 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2+\sigma_{\varepsilon 3}^2 \end{matrix}$$

$$\mathbf{S} = \begin{matrix} \mathbf{S}_{11} & \mathbf{S}_{21} & \mathbf{S}_{31} \\ \mathbf{S}_{21} & \mathbf{S}_{22} & \mathbf{S}_{32} \\ \mathbf{S}_{31} & \mathbf{S}_{32} & \mathbf{S}_{33} \end{matrix}$$

$F_{uls}(\theta) = ½\mathrm{trace}[\{\mathbf{S}-\Sigma(\theta)\}^2] =$

$½(\mathbf{S}_{11} -\lambda_1^2-\sigma_{\varepsilon 1}^2)^2 + 2*(\mathbf{S}_{21}- \lambda_2\lambda_1)^2 + (\mathbf{S}_{22} -\lambda_2^2-\sigma_{\varepsilon 2}^2)^2 + \ldots$
$\ldots + 2*(\mathbf{S}_{32}- \lambda_3\lambda_2)^2 + (\mathbf{S}_{33} -\lambda_3^2-\sigma_{\varepsilon 3}^2)^2.$

---

[2] ½ trace$[\{\mathbf{S}-\Sigma(\theta)\}^2]$: $\{\mathbf{S}-\Sigma(\theta)\}$ is a symmetric matrix as $\{\mathbf{S}-\Sigma(\theta)\}^2 = \{\mathbf{S}-\Sigma(\theta)\}\{\mathbf{S}-\Sigma(\theta)\}$; trace(**A**), where **A** is a square matrix, is the operation of summing the diagonal elements of **A**. Trace$[\{\mathbf{S}-\Sigma(\theta)\}^2]$ is therefore a scalar (single number).

LISREL seeks values of the parameters $\boldsymbol{\theta}=[\lambda_1\ \lambda_2\ \lambda_3\ \sigma_{\varepsilon1}^2\ \sigma_{\varepsilon2}^2\ \sigma_{\varepsilon3}^2]$ that minimize this least squares function. In LISREL, ULS estimates are obtained by stating to "uls" on the "ou" line ("ou ml" is the default). Here is the LISREL input:

```
title
da no=100 ni=4 ma=cm
cm sy
  2.00
  1.20 2.44
  0.95 1.14 1.9025
  1.15 1.02 0.8075 1.7225
mo ne=1 ny=4 ly=fu,fr te=di,fr ps=di,fi
ma ps
1
ou nd=3 rs uls
```

Here is the output:

```
 LISREL Estimates (Unweighted Least Squares)

         LAMBDA-Y

               ETA 1
            --------
    VAR 1      1.100
             (0.062)

    VAR 2      1.130
             (0.062)

    VAR 3      0.915
             (0.053)

    VAR 4      0.950
             (0.059)

               THETA-EPS

               VAR 1      VAR 2      VAR 3      VAR 4
            --------   --------   --------   --------
               0.790      1.164      1.066      0.821
             (0.199)    (0.210)    (0.177)    (0.184)

                           Goodness of Fit Statistics

 W_A_R_N_I_N_G: Chi-square, standard errors, t-values and standardized
                residuals are calculated under the assumption of multi-
                variate normality.

                              Degrees of Freedom = 2
       Normal Theory Weighted Least Squares Chi-Square = 3.624 (P = 0.163)
```

ULS is simple to understand, but has the drawback that the standard errors of estimates and the chi2 are not necessarily correct. There are versions of least sqaures estimators which do given correct results. We will first reformulate the ULS function in order to introduce these.

Another more flexible way to represent the ULS function is this: Let $\mathbf{s}$ denote a vector containing the elements in $\mathbf{S}$ and a vector $\boldsymbol{\sigma}(\boldsymbol{\theta})$ containing the element in $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. For example, given

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{matrix} \lambda_1^2+\sigma_{\varepsilon1}^2 & \lambda_2\lambda_1 & \lambda_3\lambda_1 \\ \lambda_2\lambda_1 & \lambda_2^2+\sigma_{\varepsilon2}^2 & \lambda_3\lambda_2 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2+\sigma_{\varepsilon3}^2 \end{matrix}$$

$$\mathbf{S} = \begin{matrix} \mathbf{S}_{11} & \mathbf{S}_{21} & \mathbf{S}_{31} \\ \mathbf{S}_{21} & \mathbf{S}_{22} & \mathbf{S}_{32} \\ \mathbf{S}_{31} & \mathbf{S}_{32} & \mathbf{S}_{33} \end{matrix}$$

$$\mathbf{s} = [\mathbf{S}_{11}\ \mathbf{S}_{21}\ \mathbf{S}_{22}\ \mathbf{S}_{31}\ \mathbf{S}_{32}\ \mathbf{S}_{33}]^t$$

$$\boldsymbol{\sigma}(\boldsymbol{\theta}) = [\lambda_1^2+\sigma_{\varepsilon 1}^2 \quad \lambda_2\lambda_1 \quad \lambda_2^2+\sigma_{\varepsilon 2}^2 \quad \lambda_3\lambda_1 \quad \lambda_3\lambda_2 \quad \lambda_3^2+\sigma_{\varepsilon 3}^2]^t$$

So given p tests, $\mathbf{S}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ are pxp matrices, and $\mathbf{s}$ and $\boldsymbol{\sigma}(\boldsymbol{\theta})$ are q=p*(p+1)/2 vectors. As $\mathbf{S}$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ are symmetrix, $\mathbf{s}$ and $\boldsymbol{\sigma}(\boldsymbol{\theta})$ contain the same information. The ULS function can be formulated as follows:

$$F_{ULS}(\boldsymbol{\theta}) = \{\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\}^t \mathbf{W}_{ULS}^{-1} \{\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\}$$

with $\boldsymbol{\theta}=[\lambda_1\ \lambda_2\ \lambda_3\ \sigma_{\varepsilon 1}^2\ \sigma_{\varepsilon 2}^2\ \sigma_{\varepsilon 3}^2]$, and $\mathbf{W}_{ULS}$ a qxq diagonal matrix, with 1 or .5 on the diagonal. Consider again the simple example:

$$\mathbf{S} = \begin{matrix} a & b, \\ b & c \end{matrix} \qquad \mathbf{s}=[a\ b\ c]$$

and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ equals

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{matrix} \alpha & \beta, \\ \beta & \gamma \end{matrix} \qquad \boldsymbol{\sigma}(\boldsymbol{\theta})=[\alpha\ \beta\ \gamma]$$

$$\mathbf{W}_{ULS} = \begin{matrix} 1 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & 1 \end{matrix}$$

$$F_{ULS}(\boldsymbol{\theta}) = \{\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\}^t \mathbf{W}_{ULS}^{-1} \{\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\} = ((a-\alpha)^2+2*(b-\beta)^2+(c-\gamma)^2),$$

You can choose other matrices $\mathbf{W}$. The function is then generally called the weighted least squares function (WLS):

$$F_{WLS}(\boldsymbol{\theta}) = \{\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\}^t \mathbf{W}^{-1} \{\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\}.$$

There are various choices for $\mathbf{W}$. $\mathbf{W}$ is said to be <u>correct</u> if $\mathbf{W}$ expresses the sampling functuation of the elements in $\mathbf{s}$, i.e., if $\mathbf{W}$ is the covariance matrix of the estimates of $\mathbf{s}$. This may seem like a strange concept (covariance matrix of a covariance matrix?), but actually we are already familiar with it. For instance, the standard error of an estimate may be viewed as the standard deviation of the estimate. Let est(μ) denote the ML estimate of the mean of data vector $\mathbf{x}$, then the standard error equal σ/sqrt(N), and the variance equals σ^2/N, where σ is the standard deviation of $\mathbf{x}$. The variance of an estimated variance (s^2) equals $(2*s^4)/N$, so using WLS to estimate a single variance we would have:

$$F_{WLS}(\boldsymbol{\theta}) = \{\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\}\ \{(2*s^4)/N\}^{-1}\ \{\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\}, \text{ or}$$

$$F_{WLS}(\boldsymbol{\theta}) = \{\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\}\ \{N/(2*s^4)\}\ \{\mathbf{s} - \boldsymbol{\sigma}(\boldsymbol{\theta})\},$$

where the matrix **W** (actually a scalar: $2*s^4)/N\}$ is correctly specified, because it reflects the sampling fluctuation of **s**=[s^2]. We shall use the WLS function in the LISREL modeling of discrete data, where the matrix **W** is chosen to be correct (at least in theory).

LISREL modeling of ordinal data

So far we have assumed that data were multivariate normally distributed, and we used ML estimation to fit LISREL models. Unfortunately there are many situations in which the data is not normally distributed. If the data are continuously distributed, one may consider various data transformations to render the data more normally distributed. There are situations in which transformations work quite well. For instance these data are not normally distributed (skewness=1.029, kurtosis=1.115).

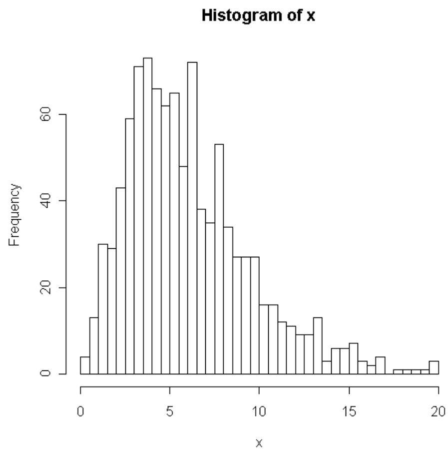


Figure 2-1

However a simple square root transformation helps a lot (skewness=0.263, kurtosis=-0.104):

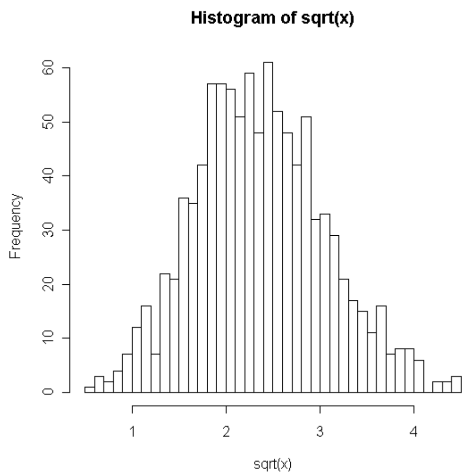


Figure 2-2

In the PRELIS program (part of the LISREL program), you can transform data to so-called normal scores. This transformation renders the skewness and kurtosis as close to the expected values under normality as possible. However, there are situations, in which transformations do not work well, or are simply inappropriate. One such situation is when the data are discrete. For example, in the most extreme case the data may be dichotomous, e.g., scores 0 and 1. For instance, the question "do you drink three or more alcoholic beverages a day" will given rise to the response "yes" or "no", i.e., a dichotomous variable. Less extreme examples are data collected with 3 point scales, or 5 point scales. Of course as the number of response categories increase, the data may start to look normal. Here are some examples (Fig 6-3). Given 7 response categories, the data may start to look normal, as shown in figure 2-3.

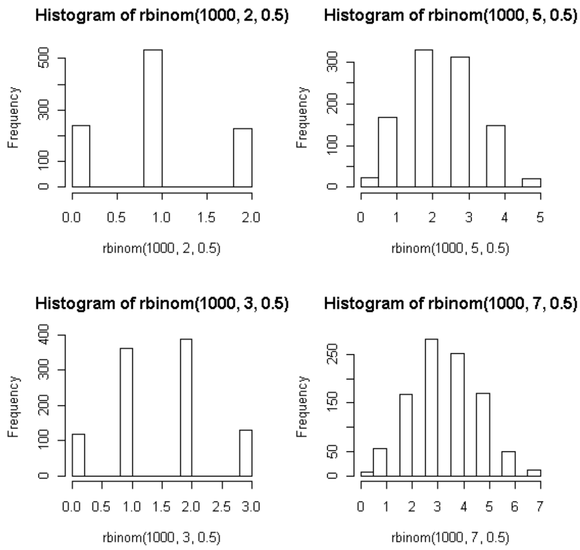


Figure 2-3:

As a rule of thumb we shall call data with 7 or more ordered response categories continuous. If such data appear to be normal (more or less symmetrically distributed), ML estimation will work well enough to be useful. However, this is not the case if the number of categories is 5 or less.

Treating discrete or ordinal as continuous is generally not a good idea. In the first place, the correlations are underestimated. We illustrate this as follows. Consider the following: X is bivariate normal, with zero means, and covariance (correlation) matrix

1	.5
.5	1

Figure 6-4 displays the histogram of 500 estimated correlation between two variables x1 and x2, with each correlation based on N=200 (a simulation study!). Figure 2-4 top: continuous standard normal, Figure 2-4 bottom: discrete, three point scale. The ordinal data was obtained from the continuous data as follows:

```
if x1<-1        y1=0  (if x1 is less than -1, assign 0 to y1)
if x1>-1<1      y1=1  (etc)
if x1>1         y1=2
```

The variable x1 was continuous, standard and normal, but y1 is ordinal, specifically a 3 point scale.

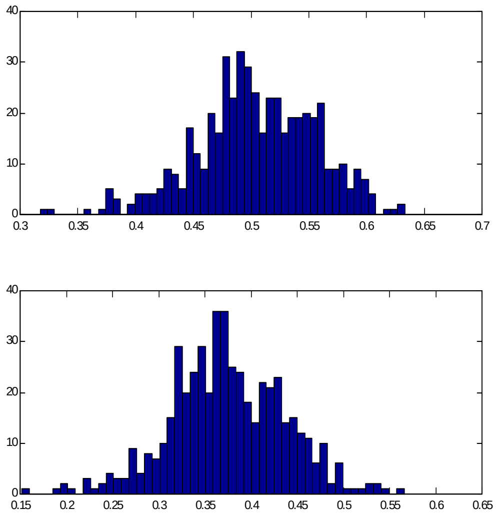


Figure 2-4 top: 500 correlations based on 500 samples of N=200, continuous data; **Figure 2-4** bottom: 500 correlations based on 500 samples of N=200, discretized data (3 point scale).

The mean values and standard deviations of the correlations shown in fig 2-4 are .504 (sd=.049) in the case if the normal data, and .374 (sd=.061) in the case of the three point scale. Given that the true correlation is .5, the observed correlation in the discretized data is clearly underestimated (.374). The degree of underestimation depends in part on distribution of the data. Because the covariance are biased, all parameters in a LISREL model are biased too. In addition to this bias, standard errors are usually overestimated, and the chi2 goodness of fit index does not follow the expected chi2 distribution (under the null hypothesis, i.e., assuming the model fitted is the correct, or true model).

Discrete factor analysis: rationale

Discrete and continuous factor analysis are closely related. Both involve the following model (i for subject, we will assume just one group):

$$\mathbf{y}_i^* = \tau + \Lambda\eta_i + \varepsilon_i.$$

The only difference is that in continuous factor analysis, the indicators $\mathbf{y}_i^*$ are observed, whereas as in discrete factor analysis they are not. What do we observe in discrete factor analysis are discrete (ordinal) responses to the items: $\mathbf{y}_i$, which assume values (say): 0,1,.... Consider a three point scale (0,1,2). The observed discrete responses are related to the latent responses as follows:

y=0 if y*<t1
y=1 if t1<y*<t2
y=2 if y*>t2,

where t1 and t2 are called thresholds (Note: t1 is not τ_1, i.e., a threshold is not an intercept). As demonstrated below, if we observe the discrete **y**, the standard covariance matrix or the Pearson product moment correlation coefficient between y1 and y2 can be calculated. However, if the data are ordinal the correlation between the observed discrete variables underestimates the correlation between y1* and y2* (see Figure 2-4). Thus we require a method to calculate the correlations among the observed variables which takes into account the fact that they are discrete (ordinal) and not continuous. To this end we use the tetrachoric or polychoric correlation matrix (tetrachorics for dichotomous data).

Tetrachoric & polychoric correlation coefficients

Structural equation modeling of ordinal data can be carried out in LISREL by analysing the so-called tetrachoric or polychoric correlation coefficients with WLS estimation. To explain the tetrachoric correlation, let us consider dichotomous variables, i.e., 2 point scales. The tetrachoric correlation is based on the assumption that there is a standard normal distribution underlying the observed dichotomy. Consider the item "do you drink three or more alcoholic beverages a day", with responses coded 0 (yes) 1 (no). Suppose in a sample of N=100 psychology student, you observe an endorsement rate (i.e., response "yes") in 15 cases.

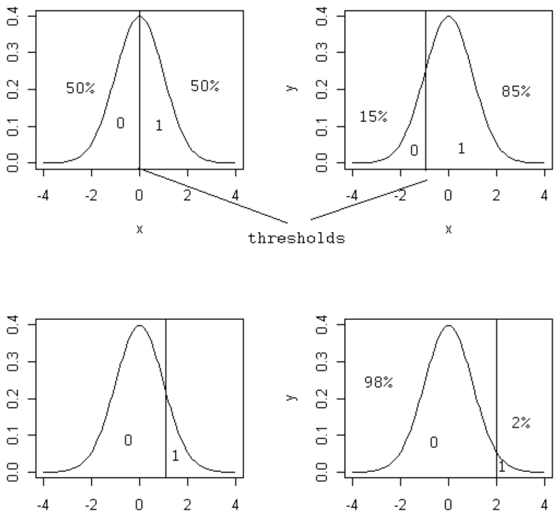
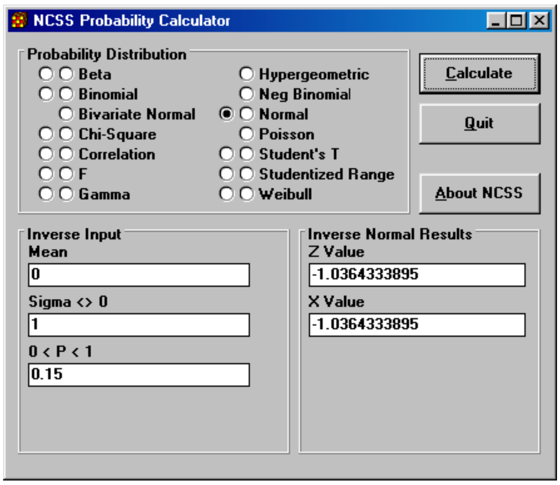


Figure 2-5: thresholds on standard normal distribution

In Figure 2-5, the top right figure shows the underlying standard normal distribution and a threshold (cut-off point) at about −1. According to this model, the tendency to display alcoholic behavior is a continuous variable y*, and the reponse to the item is determined by the subject's position on this underlying variable. There is a point beyond which the reponse is 1. This point is called the <u>threshold</u>, and may be estimated easily: if 15% respond yes (0), then the probability of response yes is point .15. Let Φ(z) denote the cumulative normal distribution, then Φ(z)=.15, and

$\Phi^{-1}(.15)=z$.



Using the NCSS calculator[3], we find that z equals about −1. In ***R*** you can type:

```
> pnorm(1)              this is Φ(1)
[1] 0.8413447
> qnorm(.8413)          this is Φ⁻¹(.8413)
[1] 0.999815
> pnorm(-1)             this is Φ(-1)
 [1] 0.1586553
> qnorm(.15865)         this is Φ⁻¹(.15865)
[1] -1.000022
```

The position of the threshold is an unknown (to be estimated) parameter that depends on the item. For instance, this the item was "do you drink three or more alcoholic beverages a week", and you observed an endorsement rate of about 50%, the top left figure (Figure 5-6) may be appropriate. By defining response probabilities as a function of a continuous but unobserved variable, we can fit the factor model to the continuous unobserved variable (or variables). To this end we need to estimate the correlation between two continuous distributions based on the observed ordinal data. We already know from above (Figure 2-4) that the standard correlation coefficient based on the observed ordinal data underestimates the true correlation between the underlying continuous variables.

A **tetrachoric correlation** is the correlation between the underlying normal distributions, which is calculated on the basis of the observed reponses to two dichotomous items. We can present the observed data in a 2x2 table. E.g., for N=1000:

[3] http://www.ncss.com/download_probcalc.html

		item 1: 0	item 1: 1	marginal
item 2	0	118	372	490
	1	24	486	510
marginal		142	858	1000

or

118	0	0
372	0	1
24	1	0
486	1	1

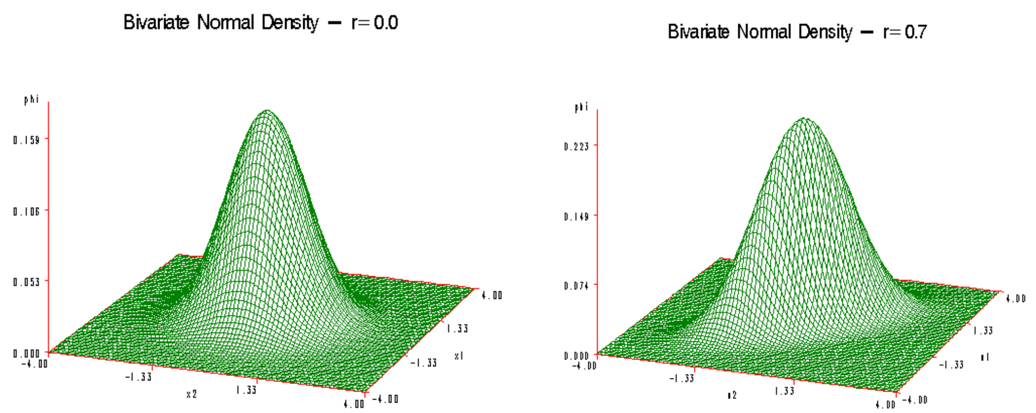
The thresholds are

$\tau 1 = \Phi^{-1}(142/1000) = -1.07$ (item 1).

$\tau 2 = \Phi^{-1}(490/1000) = -0.025$ (item 2).

We now assume that underlying the dichomotomies there is a bivariate standard normal distribution. Let $\phi(z1, z2, \rho)$ denote the standard bivariate normal distribution, where ρ is the correlation between the underlying normal distributions (see **Figure 2-6**).

Figure 2-6: two bivariate normal distributions (ρ=0 and ρ=.7).



The probability of scoring [0,0] equals

$$\int_{-\infty}^{\tau 1} \int_{-\infty}^{\tau 2} \phi(z1, z2, \rho)\, d(z1)\, d(z2) = \Phi(-\infty \ldots \tau 1, -\infty \ldots \tau 2, \rho)$$

The probability of scoring [1,0] equals

$$\int_{\tau 1}^{\infty} \int_{-\infty}^{\tau 2} \phi(z1, z2, \rho)\, d(z1)\, d(z2) = \Phi(\tau 1 \ldots \infty, -\infty \ldots \tau 2, \rho)$$

The probability of scoring [0,1] equals

$$\int_{-\infty}^{\tau 1} \int_{\tau 2}^{\infty} \phi(z1,z2,\rho)\,d(z1)\,d(z2) = \Phi(-\infty\ldots\tau 1,\ \tau 2\ldots-\infty,\ \rho)$$

The probability of scoring [0,1] equals

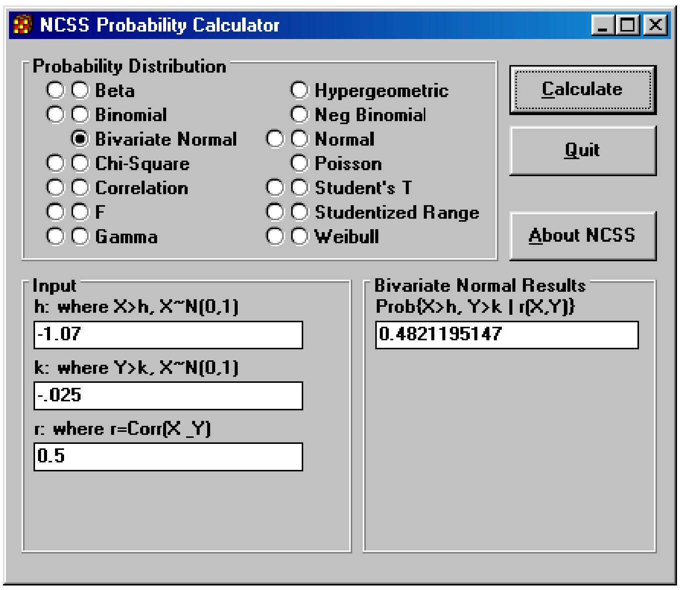
$$\int_{\tau 1}^{\infty} \int_{\tau 2}^{\infty} \phi(z1,z2,\rho)\,d(z1)\,d(z2) = \Phi(\tau 1\ldots\infty,\ \tau 2\ldots\infty,\ \rho)$$

Note that these expressions depend on three unknown quantities: $\tau 1$, $\tau 1$, and ρ. We have already estimated $\tau 1$ (-1.07) and $\tau 2$ (-.025). To estimate ρ, we seek the value of ρ such that the likelihood of the observed count is maximal (i.e., we use ML estimation)

expected count	observed count			
$N*\Phi(-\infty\ldots\tau 1,\ -\infty\ldots\tau 2,\ \rho)$	118	(score	0	0)
$N*\Phi(\tau 1\ldots\infty, -\infty\ldots\tau 2,\ \rho)$	372	(score	0	1)
$N*\Phi(-\infty\ldots\tau 1,\ \tau 2\ldots-\infty,\ \rho)$	24	(score	1	0)
$N*\Phi(\tau 1\ldots\infty,\ \tau 2\ldots\infty,\ \rho)$	486	(score	1	1)

Suppose ρ=.5, then we have

expected count	observed count			
$N*\Phi(-\infty\ldots\tau 1,\ -\infty\ldots\tau 2,.5)=1000*.115$	118	(score	0	0)
$N*\Phi(\tau 1\ldots\infty,-\infty\ldots\tau 2,.5)=1000*.376$	372	(score	0	1)
$N*\Phi(-\infty\ldots\tau 1,\ \tau 2\ldots-\infty,.5)=1000*.027$	24	(score	1	0)
$N*\Phi(\tau 1\ldots\infty,\ \tau 2\ldots\infty,.5)=1000*.482$	486	(score	1	1)



expected: $N*\Phi(\tau 1...\infty, \tau 2...\infty, .5)=1000*.4821=482.1$
observed: 486 (score 1 1)

Given the marginal probabilities and the probability of score 1 1, we can calculate the other probabilities (e.g., prob(1,0)=.858-.482=.376, etc.).

Table: observed counts and expected probabilities based on ρ=.5.

		item 1			
		0	1		
item 2	0	118 (.115)	372 (.376)	490	(.490)
	1	24 (.027)	486 (.482)	510	(.510)
		142 (.142)	858 (.858)	1000	

and thus:

observed count	expected counts	responses	
118	115	0	0
372	376	0	1
24	27	1	0
486	482	1	1

The correlation ρ is estimated by minimizing some function of the difference between the observed counts and the expected counts (based on the current value of ρ). Given the similarity in values of observed and expected, the estimate of .5 is probably close to the ML estimate. The actual maximum likelihood estimate can be obtained from PRELIS, which is part of the LISREL program. You can read the data into LISREL in a number of ways. Given the raw data file ddat1 (1000 x 2), we can use this script (cut and paste this in a lisrel syntax window):

```
title prelis input file
da ni=2 no=1000
ra fi=ddat1
or all
ou ma=pm
```

raw data file ddat1 1000 x 2

all variables are ordinal

But the raw data file actually only contains this information:

freq.	item1	item2
118	0	0
372	0	1
24	1	0
486	1	1

So if you read this table and specify the first column as the weight variable, you will get the same results.

```
title prelis input file
da ni=3 no=0
ra fi=ddats
la
freq itm1 itm2
we 1
or all
ou ma=pm sa=wmat1
```

datafile 4 x 3, freq., item1, item2

118	0	0
372	0	1
24	1	0
486	1	1

weigh data by frequency

The results are (click on the PRELIS icon):

```
 The following lines were read from file E:\lisb\prel2.LS8:

 title prelis input file
 da ni=3 no=0
 ra fi=ddats
 la
 freq itm1 itm2
 we 1
 or all
 ou ma=pm

 Total Sample Size =   1000

 Univariate Marginal Parameters

 Variable     Mean St. Dev.   Thresholds
 --------     ---- --------   ----------
     itm1    0.000    1.000  -0.025
     itm2    0.000    1.000  -1.071

 Univariate Distributions for Ordinal Variables

     itm1 Frequency Percentage Bar Chart
      0      490         49.0
      1      510         51.0

     itm2 Frequency Percentage Bar Chart
      0      142         14.2
      1      858         85.8

 There are 4 distinct response patterns, see FREQ-file.
 The 4 most common patterns are :
     486         1   1
     372         0   1
     118         0   0
      24         1   0

                        Correlations and Test Statistics

           (PE=Pearson Product Moment, PC=Polychoric, PS=Polyserial)
                                        Test of Model         Test of Close Fit
 Variable vs. Variable Correlation  Chi-Squ.  D.F. P-Value       RMSEA  P-Value
 -------- --- -------- -----------  --------  ---- -------       -----  -------
     itm2 vs.     itm1   0.544 (PC)    0.000     0    1.000        0.000   1.000

 Correlation Matrix

                  itm1       itm2
              --------   --------
     itm1      1.000
     itm2      0.544      1.000

 Means

                  itm1       itm2
              --------   --------
                 0.000      0.000

 Standard Deviations

                  itm1       itm2
```

thresholds

tetrachoric correlation

summary statistics of the underlying bivariate standard normal distribution

```
        --------  --------
           1.000     1.000
```

So we find that .544 is the maximum likelihood estimate of the tetrachoric correlation coefficient. This is close to the true value of .5. Simply calculating the Pearson Product Moment correlation coefficient results in a correlation of .277, i.e., as expected, the correlation is underestimated.

WLS estimation

We have seen that we can obtain an estimate of the tetrachoric correlation coefficient. In the case of several variables, we can obtain from PRELIS the tetrachoric correlation matrix. In the case of polytomous data (e.g., 3 or 5 point scales), we can obtain the so-called polychoric correlation matrices. These are based on the same assumption of an underlying (bivariate) standard normal distribution, but involve more thresholds. For instance given a 3 point scale, we have three response categories, and two thresholds. Suppose we observe 158 responses 0, 818 responses 1, and 22 reponses 2 then the tresholds would be about -1 and 2, as shown in Figure 2-6.

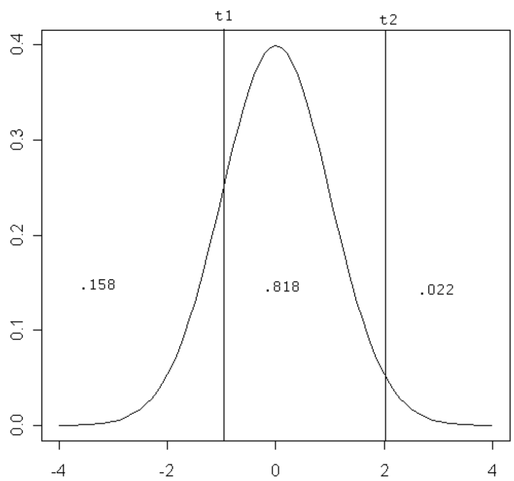


Figure 2-6: three point scale, response requencies determined by the tresholds. Probabilities shown. Suppose the item is scores 0,1,2, then prob(0)=~ .158, prob(1)=~.818, prob(2)=~.022. In a sample of 1000 cases, we could expect 158 scores 0, 818 scores 1 and 22 scores 2.

In addition to the correlation matrix, you can obtain the correct weight matrix for the elements in the correlation matrix. In the following script, the correlation matrix is written to the file ***rmat*** and the weight matrix **W** to the file ***wmat1***:

```
title prelis input file
 da ni=3 no=500
 ra fi=ddat3
la
itm1 itm2 itm3
 or all
 OU MA=PM SM=rmat AC=wmat1 XM XB XT
```

The data file ***ddat3*** contains 3 dichotomous variables observed in 500 cases. This is the (edited) output:

```
Univariate Marginal Parameters

 Variable     Mean St. Dev.   Thresholds
 --------     ---- --------   ----------
     itm1    0.000    1.000   0.035
     itm2    0.000    1.000   0.954
     itm3    0.000    1.000  -1.003

 Univariate Distributions for Ordinal Variables

     itm1 Frequency Percentage Bar Chart
      0      257         51.4
      1      243         48.6

     itm2 Frequency Percentage Bar Chart
      0      415         83.0
      1       85         17.0

     itm3 Frequency Percentage Bar Chart
      0       79         15.8
      1      421         84.2

 There are 7 distinct response patterns, see FREQ-file.
 The 7 most common patterns are :
     171         0   0   1
     166         1   0   1
      74         0   0   0
      73         1   1   1
      11         0   1   1
       4         1   0   0
       1         0   1   0


                         Correlations and Test Statistics

            (PE=Pearson Product Moment, PC=Polychoric, PS=Polyserial)
                                          Test of Model          Test of Close Fit
 Variable vs. Variable Correlation  Chi-Squ.  D.F. P-Value        RMSEA   P-Value
 -------- --- -------- -----------  --------  ---- -------        -----   -------
     itm2 vs.     itm1  0.623 (PC)   0.000      0    1.000         0.000    1.000
     itm3 vs.     itm1  0.753 (PC)   0.000      0    1.000         0.000    1.000
     itm3 vs.     itm2  0.598 (PC)   0.000      0    1.000         0.000    1.000

 Correlation Matrix

                itm1       itm2       itm3
            --------   --------   --------
     itm1      1.000
     itm2      0.623      1.000
     itm3      0.753      0.598      1.000

 Means

                itm1       itm2       itm3
            --------   --------   --------
               0.000      0.000      0.000

 Standard Deviations

                itm1       itm2       itm3
            --------   --------   --------
               1.000      1.000      1.000
```

You may wonder why the means and the standard deviations are zero and one. This is because these pertain to the scale of the unobserved, underlying continuous variables **y*** (where **y** denotes the ordinal variable). These values are due to arbitrary, but convenient, scaling constraints. Because the y* is not observed we have to impose a scale (just as we have to scale the latent variables in a common factor analysis). So y (ordinal) is a function of y* (continuous, not observed), but we have to impose a scale on y*, such that y* standardized with zero mean. Given the assumption of **y*** ~

MVN(μ,Σ), we note that the scaling assumption now implies $\mathbf{y}$* ~ MVN($\mathbf{0}$,$\mathbf{P}$), where $\mathbf{P}$ (greek capital rho) is a correlation matrix.

We can now analyse the data in LISREL using WLS estimation. Fortunately you will have little trouble writing the input file, as most of it is business as usual. The new aspects are shown in green italics: *pm fi=* identifies the location of the polychoric or tetrachoric correlation matrix, and *ac=* identifies the location of the correct weight matrix.

```
title lisrel input file WLS
 da no=500 ni=3 ma=pm
 pm fi=rmat
 ac=wmat3
 mo ly=di,fi ps=sy,fr te=ze ne=3 ny=3
 ma ly
 1 1 1
 ma ps
 1
 0 1
 0 0 1
 pa ps
 0
 1 0
 1 1 0
 ou nd=4
```

Here are the results (edited):

Correlation Matrix

	VAR 1	VAR 2	VAR 3
VAR 1	1.0000		
VAR 2	0.6225	1.0000	
VAR 3	0.7528	0.5984	1.0000

PSI

	ETA 1	ETA 2	ETA 3
ETA 1	1.0000		
ETA 2	0.6225	1.0000	
	(0.0619)		
ETA 3	0.7528	0.5984	1.0000
	(0.0548)	(0.1086)	

Goodness of Fit Statistics
Degrees of Freedom = 3
Minimum Fit Function Chi-Square = 0.0 (P = 1.0000)

The chi2 is zero because this is a saturated model. The estimated correlations in PSI equal the input correlations (only now we have standard errors of the estimates). If we treat the ordinal data as continuous, we would obtain biased correlation coefficients.

ETA 1	1.0000		
ETA 2	0.3376	1.0000	
ETA 3	0.3773	0.1814	1.0000

Generally, if you have discrete or ordinal data, and the assumption of underlying normality is reasonable, you may use LISREL to analyze the tetrachoric or polychoric correlation matrix using WLS. The correlation matrix and the weight matrix can be obtained from PRELIS. Note that WLS usually larger sample sizes than does normal theory ML (analysis of multivariate normal data), especially when the thresholds are extreme. However:

1) When is the assumption of underlying normality reasonable ?
2) What to do when the sample is small ?
3) What to do when the thresholds are extreme ?

If is not reasonable in the case of a nominal variable (sex, political preference). However, it can be difficult to determine whether a variable is nominal. Consider normal (unaffected) vs. personality disordered (affected), In a dimensional model of psychopathology, affected (dysthymic depression, personality disordered) is often viewed as a manifestation of the extreme of a continuous distribution. Here is an extreme example:

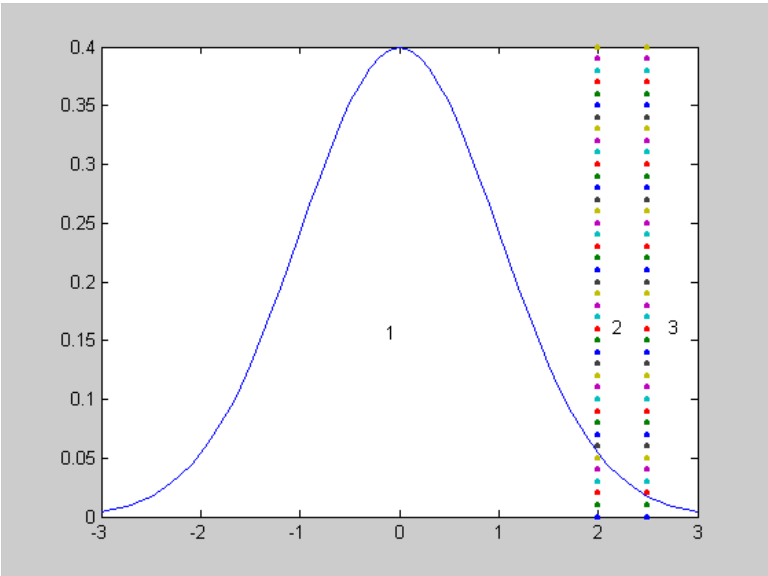


Figure 2-7: extreme responses 2 (mildly affected) and 3 (extremely affected)

1) Normal, p=.9772
2) 2&3 Affected, 1-.9772 = .0228
3) 2 mild p=.0218
4) 3 severe p=.0062

Here the underlying variable is the liability to display psychopathology. If you accept this model, then you consider score 3 to be associated with the extreme of the distribution.

The analysis of ordinal data generally requires large sample sizes. So if the sample is small, you will have to collect more data. But note that seemingly large datasets may be too small to obtain stable estimates of polychoric correlations. Given the example above (prob(severe)=.006), you will require N=10000, if you want to ascertain about 60 (expected value) severely affected cases. Pooling the affecteds can help: given 10000, you will ascertain about 228 cases.

Illustration using PRELIS / LISREL.

It is important to realize that beyond the complications of calculating tetrachoric or polychoric correlations and the weight matrix (all do-able in PRELIS), the actual modeling of the data proceeds along the usual lines: in terms of model specification in LISREL all you now know still applies. PRELIS can also calculate correlations between ordinal variables with varying numbers of categories and continuous variables. The correlation between a continuous variable and a ordinal variable is called a point bi-serial correlation coefficient. PRELIS can also calculate the correct weight matrix in these cases. To illustrate dichotomous factor analysis will fit a single factor model to the following data:

```
0 0 0 0 0   56
0 0 0 0 1   39
0 0 0 1 0    4
0 0 0 1 1    2
0 0 1 0 0   15
0 0 1 0 1   39
0 0 1 1 0    4
0 0 1 1 1   15
0 1 0 0 0   14
0 1 0 0 1   13
0 1 0 1 0    1
0 1 0 1 1    2
0 1 1 0 0   12
0 1 1 0 1   19
0 1 1 1 0    3
0 1 1 1 1   12
1 0 0 0 0   14
1 0 0 0 1   21
1 0 0 1 0    2
1 0 0 1 1    3
1 0 1 0 0    7
1 0 1 0 1   39
1 0 1 1 0    4
1 0 1 1 1   31
1 1 0 0 0    5
1 1 0 0 1    7
1 1 0 1 0    2
1 1 0 1 1    7
```

```
 1 1 1 0 0     7
 1 1 1 0 1    32
 1 1 1 1 0     5
 1 1 1 1 1    64
```

Cut and paste the data to an external file (ddat5). Cut and paste to the SYNTAX window in LISREL (see also the appendix).

```
title prelis input file
da ni=6 no=0
! ddat5 contains the dat as shown above
ra fi=ddat5
la
itm1 itm2 itm3 itm4 itm5 freq
or itm1 itm2 itm3 itm4 itm5
we 6
OU MA=PM SM=rmat5 AC=wmat5 ! XM XB XT
```

Run the suntax. Here are the results:

```
 Total Sample Size =    500

 Univariate Marginal Parameters

 Variable     Mean St. Dev.   Thresholds
 --------     ---- --------   ----------
     itm1    0.000    1.000   0.000
     itm2    0.000    1.000   0.228
     itm3    0.000    1.000  -0.295
     itm4    0.000    1.000   0.462
     itm5    0.000    1.000  -0.496

 Univariate Distributions for Ordinal Variables

     itm1 Frequency Percentage Bar Chart
      0      250         50.0
      1      250         50.0

     itm2 Frequency Percentage Bar Chart
      0      295         59.0
      1      205         41.0

     itm3 Frequency Percentage Bar Chart
      0      192         38.4
      1      308         61.6

     itm4 Frequency Percentage Bar Chart
      0      339         67.8
      1      161         32.2

     itm5 Frequency Percentage Bar Chart
      0      155         31.0
      1      345         69.0


 Correlation Matrix

                itm1       itm2       itm3       itm4       itm5
            --------   --------   --------   --------   --------
     itm1      1.000
     itm2      0.335      1.000
     itm3      0.445      0.371      1.000
     itm4      0.506      0.411      0.571      1.000
     itm5      0.436      0.214      0.528      0.408      1.000
```

Run LISREL

```
title lisrel input file WLS
 da no=500 ni=5 ma=pm
 pm fi=rmat5
 ac=wmat5
 mo ly=fu,fr ps=sy,fi te=di,fr ne=1 ny=5 al=ze ty=ze
pa ly
2 3 4 5 6
pa te
12 13 14 15 16
ma ps
1
st .5 all
ou
```

OUTPUT:

```
 LISREL Estimates (Weighted Least Squares)

         LAMBDA-Y

               ETA 1
            --------
    VAR 1     0.6542
            (0.0537)
             12.1703

    VAR 2     0.4918
            (0.0602)
              8.1670

    VAR 3     0.7789
            (0.0522)
             14.9340

    VAR 4     0.7604
            (0.0542)
             14.0301

    VAR 5     0.6337
            (0.0564)
             11.2400

         THETA-EPS

               VAR 1      VAR 2      VAR 3      VAR 4      VAR 5
            --------   --------   --------   --------   --------
              0.5721     0.7581     0.3933     0.4217     0.5985
            (0.0834)   (0.0742)   (0.0928)   (0.0938)   (0.0843)
              6.8627    10.2111     4.2396     4.4961     7.0983


         Squared Multiple Correlations for Y - Variables

               VAR 1      VAR 2      VAR 3      VAR 4      VAR 5
            --------   --------   --------   --------   --------
              0.4279     0.2419     0.6067     0.5783     0.4015


                          Goodness of Fit Statistics

                             Degrees of Freedom = 5
             Minimum Fit Function Chi-Square = 7.5170 (P = 0.1849)
           Root Mean Square Error of Approximation (RMSEA) = 0.03176

         Standardized Residuals

               VAR 1      VAR 2      VAR 3      VAR 4      VAR 5
            --------   --------   --------   --------   --------
    VAR 1       - -
    VAR 2     0.3120       - -
```

```
    VAR 3    -2.0449    -0.3213      - -
    VAR 4     0.2856     0.9981    -0.8481      - -
    VAR 5     0.5645    -1.7997     1.2912    -1.7096      - -
```

Mplus

To investigate measurement invariance in the discrete factor model, we shall use Mplus (the student version). You can obtain the student version from http://www.statmodel.com/demo.shtml. I assume you have saved the data of the previous illustration in an external file called "ddats5". To fit exaclty the same factor model in Mplus, I specify the following syntax (single factor, one group):

```
Title:
    1-factor CFA 5 dich. items
Data:
       file is ddats5;
Variable:
       names are v1 v2 v3 v4 v5 freq;
       freq is freq;
       usev are v1 v2 v3 v4 v5;
    categorical are v1 v2 v3 v4 v5;
Analysis:
    estimator is wls;
Model:
    f by v1*.5 v2*.5 v3*.5 v4*.5 v5*.5;
    f@1;
    [f@0];
    [v1$1];
    [v2$1];
    [v3$1];
    [v4$1];
 Output:
    standardized tech1 tech2;
```

A major distinction between Mplus and LISREL is that PRELIS is used before LISREL to calculate the correlation matrix and the weight matrix (W). These are then read into the LISREL syntax (pm fi=...., ac=.....). In Mplus, the complete analysis is carried out in one step. This is more convenient.

Mplus	LISREL
f by v1*.5 v2*.5 v3*.5 v4*.5 v5*.5;	pa ly 1 1 1 1 1
f@1; [f@0];	ma ps 1 ma al 0
[v1$1]; [v2$1]; [v3$1]; [v4$1];	thresholds. these are estimated in PRELIS, and are not part of the LISREL input

Here is the output (edited), which is largely simple to follow:

```
UNIVARIATE PROPORTIONS AND COUNTS FOR CATEGORICAL VARIABLES

    V1
      Category 1    0.500      250.000
      Category 2    0.500      250.000
    V2
      Category 1    0.590      295.000
      Category 2    0.410      205.000
    V3
      Category 1    0.384      192.000
      Category 2    0.616      308.000
    V4
      Category 1    0.678      339.000
      Category 2    0.322      161.000
    V5
      Category 1    0.310      155.000
      Category 2    0.690      345.000

THE MODEL ESTIMATION TERMINATED NORMALLY

TESTS OF MODEL FIT

Chi-Square Test of Model Fit

          Value                              7.533
          Degrees of Freedom                     5
          P-Value                           0.1839

Number of Free Parameters                       10

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.032

MODEL RESULTS
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 F        BY
    V1                 0.654      0.054     12.183      0.000
    V2                 0.492      0.060      8.176      0.000
    V3                 0.779      0.052     14.950      0.000
    V4                 0.760      0.054     14.045      0.000
    V5                 0.634      0.056     11.252      0.000

 Means
    F                  0.000      0.000    999.000    999.000

 Thresholds
    V1$1              -0.001      0.056     -0.010      0.992
    V2$1               0.221      0.056      3.920      0.000
    V3$1              -0.292      0.057     -5.133      0.000
    V4$1               0.460      0.058      7.915      0.000
    V5$1              -0.493      0.058     -8.427      0.000

 Variances
    F                  1.000      0.000    999.000    999.000
```

```
R-SQUARE

    Observed                                        Two-Tailed   Residual
    Variable        Estimate       S.E.  Est./S.E.    P-Value    Variance

    V1                 0.428      0.070      6.092      0.000      0.572
    V2                 0.242      0.059      4.088      0.000      0.758
    V3                 0.607      0.081      7.475      0.000      0.393
    V4                 0.578      0.082      7.023      0.000      0.422
    V5                 0.402      0.071      5.626      0.000      0.598

IRT PARAMETERIZATION IN TWO-PARAMETER PROBIT METRIC
WHERE THE PROBIT IS DISCRIMINATION*(THETA - DIFFICULTY)

 Item Discriminations

 F        BY
    V1                 0.865      0.124      6.970      0.000
    V2                 0.565      0.091      6.198      0.000
    V3                 1.242      0.211      5.880      0.000
    V4                 1.171      0.198      5.923      0.000
    V5                 0.819      0.122      6.734      0.000

 Means
    F                  0.000      0.000      0.000      1.000

 Item Difficulties
    V1$1              -0.001      0.086     -0.010      0.992
    V2$1               0.450      0.128      3.509      0.000
    V3$1              -0.375      0.078     -4.820      0.000
    V4$1               0.604      0.089      6.760      0.000
    V5$1              -0.778      0.119     -6.548      0.000

 Variances
    F                  1.000      0.000    999.000    999.000
```

Mplus provides the results in the IRT parameterization. This indicates that the discrete factor model and the two parameter Birmbaum model are actually equivalent. However, we shall limit our presentation to the discrete factor model.

Mplus example: Single group two common factors

```
Title:
    1-factor CFA 6 dich. items
Data:
       file is ddat1;
Variable:
       names are v1 v2 v3 v4 v5 v6;
       usev are v1 v2 v3 v4 v5 v6;
    categorical are v1 v2 v3 v4 v5 v6;
Analysis:
    estimator is WLS;
Model:
    f1 by v1*.5 v2*.5 v3*.5;
    f2 by v4*.5 v5*.5 v6*.5;
    f1@1 f2@1;
    f1 with f2*.4;
    [f1@0 f2@0];
    [v1$1 v1$2];
    [v2$1 v2$2];
    [v3$1 v3$2];
    [v4$1 v4$2];
    [v5$1 v5$2];
    [v6$1 v6$2];
 Output:
    standardized tech1 tech2;
```

Mplus	LISREL
f1 by v1*.5 v2*.5 v3*.5; f2 by v4*.5 v5*.5 v6*.5;	pa ly 1 0 1 0 1 0 0 1 0 1 0 1 st .5 ly 1 1 ly 2 1 ly 3 1 st .5 ly 4 2 ly 5 2 ly 6 2
f1@1 f2@1; f1 with f2*.4; [f1@0 f2@0];	ma ps 1 .4 1 pa al 0 1 0 ma al 0 0
[f1@0 f2@0]; [v1$1 v1$2]; [v2$1 v2$2]; [v3$1 v3$2]; [v4$1 v4$2]; [v5$1 v5$2]; [v6$1 v6$2];	thresholds. these are estimated in PRELIS, and are not part of the LISREL input

Mplus example: Single group two common factors with equality constraints

```
Title:
    1-factor CFA 6 dich. items
Data:
        file is ddat1;
Variable:
        names are v1 v2 v3 v4 v5 v6;
        usev are v1 v2 v3 v4 v5 v6;
    categorical are v1 v2 v3 v4 v5 v6;
Analysis:
    estimator is WLS;
Model:
    f1 by v1*.5 (p1);
    f1 by v2*.5 (p1);
    f1 by v3*.5 (p1);
    f2 by v4*.5 (p2);
    f2 by v5*.5 (p2);
    f2 by v6*.5 (p2);
    f1@1 f2@1;
    f1 with f2*.4;
    [f1@0 f2@0];
    [v1$1];
    [v1$2];
    [v2$1 v2$2];
    [v3$1 v3$2];
    [v4$1 v4$2];
    [v5$1 v5$2];
    [v6$1 v6$2];
 Output:
    standardized tech1 tech2;
```

Mplus	LISREL
f1 by v1*.5 (p1); f1 by v2*.5 (p1); f1 by v3*.5 (p1); f2 by v4*.5 (p2); f2 by v5*.5 (p2); f2 by v6*.5 (p2);	pa ly 3 0 3 0 3 0 0 4 0 4 0 4 st .5 ly 1 1 ly 2 1 ly 3 1 st .5 ly 4 2 ly 5 2 ly 6 2 or: eq ly 1 1 ly 2 1 ly 3 1 eq ly 4 2 ly 5 2 ly 6 2

Equal tresholds:

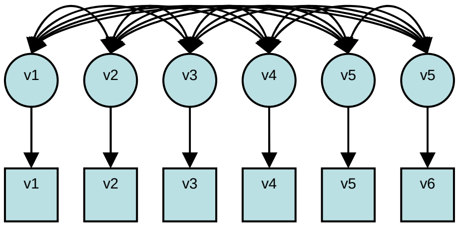
```
    [v1$1] (t1);
    [v1$2] (t2);
    [v2$1] (t1);
    [v2$2] (t2);
```

etc.

Mplus example: Two groups, two common factors

```
Title:
       model 3b
       Multiple-group discrete factor analysis
       1-factor CFA on 5 items
Data:
       file is ddat2;
Variable:
       names are v1 v2 v3 v4 v5 v6 sex;
       usev are v1 v2 v3 v4 v5 v6;
    categorical are v1 v2 v3 v4 v5 v6;
    grouping = sex (1 = female 2 = male)
Analysis:
    parameterization = delta;
Model:
    f1 by v1*.5 v2*.5 v3*.5;
    f2 by v4*.5 v5*.5 v6*.5;
    f1@1 f2@1;
    f1 with f2*.3;
    [f1@0 f2@0];
    [v1$1 v1$2];
    [v2$1 v2$2];
    [v3$1 v3$2];
    [v4$1 v4$2];
    [v5$1 v5$2];
    [v6$1 v6$2];
    {v1@1 v2@1 v3@1 v4@1 v5@1 v6@1};
Model male:
    f1 by v1*.5 v2*.5 v3*.5;
    f2 by v4*.5 v5*.5 v6*.5;
    f1@1 f2@1;
    f1 with f2*.3;
    [f1@0 f2@0];
    [v1$1 v1$2];
    [v2$1 v2$2];
    [v3$1 v3$2];
    [v4$1 v4$2];
    [v5$1 v5$2];
    [v6$1 v6$2];
    {v1@1 v2@1 v3@1 v4@1 v5@1 v6@1};
Output:
    standardized tech1 tech2;
```

Mplus example: Two groups, estimates thresholds and polycorrelations.




```
Title:
       step 1
       multiple-group discrete fa
Data:
       file is ddat2;
Variable:
    names are v1 v2 v3 v4 v5 v6 sex;
    usev are v1 v2 v3 v4 v5 v6;
    categorical are v1 v2 v3 v4 v5 v6;
    grouping = sex (1=female 2=male)
Analysis:
       parameterization = delta;
Model:
    f1 BY v1@1;
    f2 BY v2@1;
    f3 BY v3@1;
    f4 BY v4@1;
    f5 BY v5@1;
    f6 by v6@1
    f1 with f2 f3 f4 f5 f6;
    f2 with f3 f4 f5 f6;
    f3 with f4 f5 f6;
    f4 with f5 f6;
    f5 with f6;
    f1@1 f2@1 f3@1 f4@1 f5@1 f6@1;
    [f1@0 f2@0 f3@0 f4@0 f5@0 f6@1];

    MODEL MALE:
    {v1@1 v2@1 v3@1 v4@1 v5@1 v6@1};
    f1 BY v1@1;
    f2 BY v2@1;
    f3 BY v3@1;
    f4 BY v4@1;
    f5 BY v5@1;
    f6 BY v6@1;
! correlations
    f1 with f2 f3 f4 f5 f6;
    f2 with f3 f4 f5 f6;
    f3 with f4 f5 f6;
    f4 with f5 f6;
    f5 with f6;
!
    f1@1 f2@1 f3@1 f4@1 f5@1 f6@1;
    [f1@0 f2@0 f3@0 f4@0 f5@0 f6@0];
! thresholds
    [v1$1 v1$2 v2$1 v2$2];
    [v3$1 v3$2 v4$1 v4$2 v5$1 v5$2 v6$1 v6$2];
Output:
    standardized tech1 tech2;
```

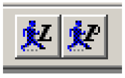
Assignment 1: Here are the Law School Admission Test (LSAT), Section VI data. A famous data set that is often used to illustrate IRT model. The data consist of N=1000, the responses are dichotomous responses to 5 cognitive ability items. Column 1 to 5 are the observed response configurations. The 6th column contains the frequencies. Read the data into PRELIS, calculate the WLS weight matrix and the tetrachoric correlation, write these to external files. In LISREL use WLS estimation to fit the single factor model. Also fit the model with equal factor loadings and residual variance. Compare the model fit and test the restrictions using a likelihood ratio test. Repeat the analysis in Mplus.

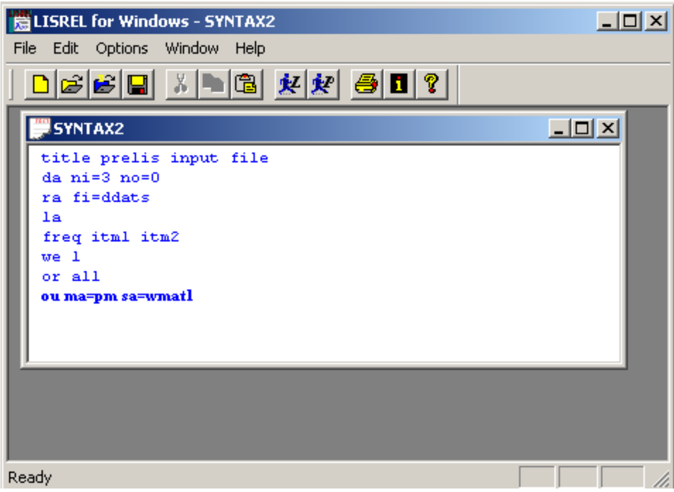
i1	i2	i3	i4	i5	freq
0	0	0	0	0	3
0	0	0	0	1	6
0	0	0	1	0	2
0	0	0	1	1	11
0	0	1	0	0	1
0	0	1	0	1	1
0	0	1	1	0	3
0	0	1	1	1	4
0	1	0	0	0	1
0	1	0	0	1	8
0	1	0	1	1	16
0	1	1	0	1	3
0	1	1	1	0	2
0	1	1	1	1	15
1	0	0	0	0	10
1	0	0	0	1	29
1	0	0	1	0	14
1	0	0	1	1	81
1	0	1	0	0	3
1	0	1	0	1	28
1	0	1	1	0	15
1	0	1	1	1	80
1	1	0	0	0	16
1	1	0	0	1	56
1	1	0	1	0	21
1	1	0	1	1	173
1	1	1	0	0	11
1	1	1	0	1	61
1	1	1	1	0	28
1	1	1	1	1	298

Assignment 2: Use the R script in Appendix B (script 1) to simulate ordinal data (three points scales). Change the script so that the factor loadings are equal over the items, and the thresholds are equal over the items. Fit the true models in Mplus.

Appendix 1: Using PRELIS/LISREL

1) Open LISREL student version, click on FILE, click on NEW, choose SYNTAX Only, click OK. Enter the PRELIS input (cut-and-paste), and save as (FILE, SAVE AS) *yourname.pr2*. Make sure you save the input in the same directory in which you have saved the data (*ddats*). To run the PRELIS input click on the PRELIS icon.





The data are:

118	0	0
372	0	1
24	1	0
486	1	1

Note that PRELIS write the polychoric correlation matrix and the weight matrix to external files (*pm* and *wmat1*).

```
title lisrel input file WLS
 da no=500 ni=5 ma=pm
 pm fi=rmat
 ac=wmat1
etc.
```

Appendix B: data simulation program (R)

```
#
# Script 1 Single group two factor model. 6 three points scales.
#
library(MASS)
np1=500       # sample size
np=np1
ne=2          # number of factors
ny=nv=6       # number of variables
ncat=3        # three points scales (3 categories)
# probabilities response 0,1,2
probs = matrix(c(
.2,.3,.5,
.2,.2,.6,
.3,.3,.4,
.2,.2,.6,
.1,.4,.5,
.2,.4,.4),ny,ncat,byrow=T)
ncat1=ncat+1
cprobs=matrix(0,ny,ncat1)
cprobs[,1]=0
for (i in 1:nv) {
for (j in 1:ncat) {
tmp=0
for (k in 1:j) {
tmp=tmp+probs[i,k]
}
cprobs[i,(j+1)]=tmp   # cumulatie probs cprobs[1,]=0,.2,.5,1
}}
thresholds=qnorm(cprobs)      # thresholds
# define sigma                # create Sigma ly*ps*ly' + te
ly=matrix(c(
.7,0,
.6,0,
.8,0,
0,.7,
0,.6,
0,.8),nv,ne,byrow=T)
ty<-as.matrix(c(0,0,0,0,0,0))        # ty
# group 1
al1=matrix(0,ne,1) # factor mean     # mean of factor zero! al
ps1=matrix(c(1,.5,.5,1),ne,ne,byrow=T)      # factor cov/variance ps
te1=diag(nv)-diag(diag(ly%*%ps1%*%t(ly)))   # te residual
#
mu1=ty+ly%*%al1
sigma1=ly%*%ps1%*%t(ly)+te1
#
rdat1<-mvrnorm(np1,mu=mu1,Sigma=sigma1)     # simulate continuous data
ddat1=matrix(-1,np1,nv)                     # create discrete data
for (k in 1:nv) {
ddat1[,k]=as.numeric(cut(rdat1[,k],thresholds[k,]))-1
}
#
write(t(rdat1),file="rdat1",ncolumn=nv)
write(t(ddat1),file="ddat1",ncolumn=nv)
```

```
#
# Script 2: Two group two factor model. 6 three points scales.
#
library(MASS)
np1=500
np2=500
np=np1+np2
ne=2
ny=nv=6
ncat=3
#
ncat1=ncat+1
ny1=ny+1
# probabilities response 0,1,2
probs = matrix(c(
.2,.3,.5,
.2,.2,.6,
.3,.3,.4,
.2,.2,.6,
.1,.4,.5,
.2,.4,.4),ny,ncat,byrow=T)
cprobs=matrix(0,ny,ncat1)
cprobs[,1]=0
for (i in 1:nv) {
for (j in 1:ncat) {
tmp=0
for (k in 1:j) {
tmp=tmp+probs[i,k]
}
cprobs[i,(j+1)]=tmp
}}
thresholds=qnorm(cprobs)

# define sigma
ly=matrix(c(
.7,0,
.6,0,
.8,0,
0,.7,
0,.6,
0,.8),nv,ne,byrow=T)
ty<-as.matrix(c(0,0,0,0,0,0))
# group 1
al1=matrix(0,ne,1) # factor mean
ps1=matrix(c(1,.5,.5,1),ne,ne,byrow=T)      # factor variance
te1=diag(nv)-diag(diag(ly%*%ps1%*%t(ly)))
#
mu1=ty+ly%*%al1
sigma1=ly%*%ps1%*%t(ly)+te1
#
al2=matrix(c(.5,-.5),ne,1)
mu2=ty+ly%*%al2
ps2=matrix(c(1,.5,.5,1),ne,ne,byrow=T)
te2=te1
sigma2=ly%*%ps2%*%t(ly)+te2
#
#
rdat1=matrix(0,np,ny1)
ddat1=matrix(0,np,ny1)
rdat1[1:np1,1:ny]<-mvrnorm(np1,mu=mu1,Sigma=sigma1)
rdat1[(np1+1):np,1:ny]<-mvrnorm(np2,mu=mu2,Sigma=sigma2)
rdat1[1:np1,ny1]=1
rdat1[(np1+1):np,ny1]=2

for (k in 1:nv) {
ddat1[1:np1,k]=as.numeric(cut(rdat1[1:np1,k],thresholds[k,]))-1
ddat1[(1+np1):np,k]=as.numeric(cut(rdat1[(1+np1):np,k],thresholds[k,]))-1
}
ddat1[1:np1,ny1]=1
ddat1[(np1+1):np,ny1]=2
#
write(t(rdat1),file="rdat2",ncolumn=ny1)
write(t(ddat1),file="ddat2",ncolumn=ny1)
```

Lecture notes III: Measurement invariance with respect to group in the discete factor model[1]

references

Wirth, R. J. & Edwards, M. C. Item Factor Analysis: Current Approaches and Future Directions, Psychological Methods, Vol. 12, No. 1, 58–79.
[recent review, including a clear explanation of the relation between discrete factor analysis and model from item response theory]

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in orderedcategorical measures. Multivariate Behavioral Research, 39(3), 479–515.

Discrete factor analysis, again.

In the previous lecture notes we presented discrete factor analysis. As in standard continuous factor analysis we assumed the following model (i for subject, we will assume just one group):

$$\mathbf{y}_i^* = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i.$$

In continuous factor analysis, the indicators $\mathbf{y}_i^*$ are observed, continuous and multivariate normally distributed. In discrete factor analysis, we observe discrete (ordinal) responses to the items. These are related to the the now unobserved indicators $\mathbf{y}_i^*$, as follows (for a three point scale):

$y=0$ if $y^*<t_1$
$y=1$ if $t_1<y^*<t_2$
$y=2$ if $y^*>t_2$,

where y^* is a given component of $\mathbf{y}^*$. Or, for a dichotomous variable:

$y=0$ if $y^*<t_1$
$y=1$ if $t_1>y^*$.

The parameters t_1 and t_2 are thresholds, i.e., points on the normal distribution, i.e., the distribution of y^*. As depicted in the previous lecture notes:

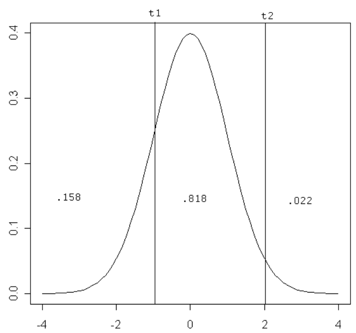


Figure 3-1: latent indicator distributed with thresholds

[1] Conor V. Dolan c.v.dolan@uva.nl. RM20. MI: continuous & discrete factor models.

There are two thing to note. The thresholds are part of a statistical model designed to relate specific discrete outcome (e.g., response to item i is 0) to a probability. In the model this probability of modeled using the probit function, which is just the cumulative normal distribution:

$$\Phi(t_1) = \int_{-\infty}^{t_1} \phi(z_1)\,dz_1 = \Phi(-\infty \ldots t_1)$$

where $z_1 = (y_1^* - \mu_1)/\sigma_1$, and μ_1 and σ_1 are the mean and standard deviation of y^*. Secondly note that y* is now effectively a latent variable, this means that we have to impose some scale on it. Specifically if we cannot observe y*, then now can be known μ_1 and σ_1. This problem is solved by imposing scaling, i.e., $\mu_1=0$ and $\sigma_1=1$ (and so $y_1=z_1$). So the probit may be viewed as a device to assign probabilities to discrete outcome. The choice of the probit is convenient, because it generalizes easily to two items. That is, in the case of two discrete items we can model the joint probabilities of outcomes (e.g., $y_1=0$ and $y_2=0$) using cumulative bi-variate normal:

$$\Phi(t_1,t_2) = \int_{-\infty}^{t_1} \int_{-\infty}^{t_2} \phi(z_1,z_2,\rho)\,d(z_1)\,d(z_2) = \Phi(-\infty \ldots t_1, -\infty \ldots t_2, \rho)$$

which is a function of the thresholds t_1 an t_2, and the correlation ρ between z_1 and z_2. Here again $z_1 = (y_1^* - \mu_1)/\sigma_1$, and $z_2 = (y_2^* - \mu_2)/\sigma_2$, and given imposed scaling $\mu_1=\mu_2=0$ & $\sigma_1=\sigma_2=1$. For instance, suppose r=.35, and t_1=-.7, t_2=-.3. In R we can calculate the marginal and the joint probabilities as follows (using the library mvtnorm[2]):

```
library(mvtnorm)
t1=-.7
t2=-.3
r=.35
ts=c(t1,t2)
sigma=matrix(c(1,r,r,1),2,2,byrow=T)
mean=rep(0,2)
p1=pnorm(t1)
p2=pnorm(t2)
p12=pmvnorm(lower=-Inf, upper=ts, mean=mean,
        corr=sigma)
print(c(p1,p2,p12[1]))
```

0.2419637 (p1) 0.3820886 (p2) 0.1361873 (p12)

Or suppose we observed in a sample of 300 cases the following reponses to a dichotomous item: 73 reponse 0 and 227 response 1. The probability of response 0 is 73/300 = .243. The threshold can be calculated as follows:

```
p=73/300
t1=qnorm(p)
print(t1)
```

[2] You can easily install this and other libraries in R.

i.e., t_1=-.6956. The correlation r can be calculated in PRELIS, Mplus, or in R (library polycor):

```
library(polycor)
polychor(y1,y2, ML = TRUE)
[1] 0.3820016
```

where y_1 contains the responses to the first and y_2 contains the responses to the second item.

Now if we have 5 items, we have the association between the items are a function of the thresholds, which we can estimate readily the thresholds and the correlation matrix of **y*** (using polycor or PRELIS or Mplus).

$$\mathbf{P} = \begin{matrix} 1 & & & & \\ \rho_{21} & 1 & & & \\ \rho_{31} & \rho_{32} & 1 & & \\ \rho_{41} & \rho_{42} & \rho_{43} & 1 & \\ \rho_{51} & \rho_{52} & \rho_{53} & \rho_{54} & 1 \end{matrix}$$

Given responses to 5 ordinal items, and assuming underlying multivariate normality, we can estimate the correlation matrix and subsequently subject this matrix to some model, e.g., a factor model:

$$\mathbf{P} = \Lambda\Psi\Lambda^{t} + \Theta,$$

To fit the model we used a least square estimator, usually called Weighted Least Squares (WLS[3]):

$$F_{WLS}(\boldsymbol{\theta}) = \{\mathbf{r} - \rho(\boldsymbol{\theta})\}^{t}\, \mathbf{W}_{wls}^{-1}\, \{\mathbf{r} - \rho(\boldsymbol{\theta})\},$$

where **r** contains the observed correlations and $\rho(\boldsymbol{\theta})$ contains the expected correlations based on the parameters $\boldsymbol{\theta}$ in the model $\mathbf{P} = \Lambda\Psi\Lambda + \Theta$, i.e., $\boldsymbol{\theta}$ contains the factor loadings and factor correlations, and $\rho=[\rho_{21}, \rho_{31}, \rho_{32},...,\rho_{53}, \rho_{54}]$. The matrix **W** is the covariance matrix of the estimates in **r**. This choice of **W** represents a "correct" weighing of the values in $\{\mathbf{r} - \rho(\boldsymbol{\theta})\}$. This means that the standard errors and the chi2 are - at least - in theory correct. WLS estimation is implemented in LISREL and in Mplus. The main disadvantage of WLS is that is requires large sample sizes to work well, especially when the number of item is large. The influence of the number of items can be appreciated by realizing that **W** is the covariance matrix of the elements in **r**. Given M items, the vector r contains L=M*(M+1)/2-M elements. So the matrix **W** will contain L*(L+1)/2 elements. Consider a numerical example (using R):

```
> getM=function(M) { M*(M+1)/2-M }
```

[3] Both in LISREL and in Mplus, there are robust versions of WLS. These are robust in the sense that they perform well (accurate standard errors, correct chi2 statistics at least in theory). Mplus uses the robust version by default.

```
> getL=function(L) {L*(L+1)/2}
> getL(getM(5:15))

 [1]   55  120  231  406  666 1035 1540 2211 3081 4186 5565
```

So given 5 item L contains 55 elements, but given 15 items W contains 5565 elements! The dependence of WLS on sample size has been solved to a degree by the development of robust WLS and robust ML estimation methods. This issue os somewhat removed from the business at hand (measurement invariance), but we'll consider various estimation procedures below.

Measurement invariance in the discrete factor analysis.

The aim of the present lecture notes is to discuss in terms of Mplus model the steps towards measurement invariance in the discrete (ordinal) factor model using WLS estimation. We limit this presentation to the simple case of 4 3-point items, a single common factor model, and two groups: we want to establish measurement invariance of the items with respect to group. As you may remember from lecture notes I, there is one clear definition of measurement invariance. The definition is general as it applies to any psychometric measurement model. It is far reaching in its consequences. For instance, let us suppose that measurement invariance with respect to group of a set of items holds. This implies that any difference between the groups in the observed summary statistics of the items should be due to differences with respect to the latent traits. For instance, in the linear factor model, we have

$$\Sigma_k = \Lambda\Psi_k\Lambda^t + \Theta$$
$$\mu_k = \tau + \Lambda\alpha_k,$$

that is, group difference in the covariance matrix and groups difference in the means are - given strict factorial invariance, attributable to group differences in the common factor covariance matrices and the common factor means (Ψ_k and α_k, respectively). Taking this perspective on measurement invariance, we shall, in the remainder of the present lecture notes, consider the steps towards measurement invariance in the discrete factor model. We now focus mainly on Mplus program.

The script used to simulate the data (note the parameter values chosen, as you will need them to do the assignments below).

```
#
library(MASS)
np1=500
np2=500
np=np1+np2
ne=1
ny=nv=4
ncat=3
#
ncat1=ncat+1
ny1=ny+1
# probabilities response 0,1,2
probs = matrix(c(
.2,.3,.5,
.2,.2,.6,
.3,.3,.4,
.2,.2,.6),ny,ncat,byrow=T)
```

```
cprobs=matrix(0,ny,ncat1)
cprobs[,1]=0
for (i in 1:nv) {
for (j in 1:ncat) {
tmp=0
for (k in 1:j) {
tmp=tmp+probs[i,k]
}
cprobs[i,(j+1)]=tmp
}}
thresholds=qnorm(cprobs)

# define sigma
# stated as reliablities
# take sqrt to obtaing loadings
rel=c(.5,.6,.55,.45)
ly=matrix(sqrt(rel),nv,ne,byrow=T)
ty<-as.matrix(c(0,0,0,0))
# group 1
al1=matrix(0,ne,1) # factor mean
ps1=matrix(c(1),ne,ne,byrow=T)      # factor variance
te1=diag(nv)-diag(diag(ly%*%ps1%*%t(ly)))
#
mu1=ty+ly%*%al1
sigma1=ly%*%ps1%*%t(ly)+te1
#
al2=matrix(c(-.5),ne,1)
mu2=ty+ly%*%al2
ps2=matrix(c(1),ne,ne,byrow=T)
te2=te1
sigma2=ly%*%ps2%*%t(ly)+te2
#
#
rdat1=matrix(0,np,ny1)
ddat1=matrix(0,np,ny1)
rdat1[1:np1,1:ny]<-mvrnorm(np1,mu=mu1,Sigma=sigma1)
rdat1[(np1+1):np,1:ny]<-mvrnorm(np2,mu=mu2,Sigma=sigma2)
rdat1[1:np1,ny1]=1
rdat1[(np1+1):np,ny1]=2

for (k in 1:nv) {
ddat1[1:np1,k]=as.numeric(cut(rdat1[1:np1,k],thresholds[k,]))-1
ddat1[(1+np1):np,k]=as.numeric(cut(rdat1[(1+np1):np,k],thresholds[k,]))-1
}
ddat1[1:np1,ny1]=1
ddat1[(np1+1):np,ny1]=2
#
write(t(rdat1),file="rdat2",ncolumn=ny1)
write(t(ddat1),file="ddat2",ncolumn=ny1)
```

Steps towards MI

As mentioned we assume that we have measured 4 items in two sample. The items are unidimensional, i.e., within each sample the single common factor model fits well (is correctly specified). In analyzing these data we would want to start with calculating the summary statistics. These are response requencies and the 3x3 cross table of pairs of item responses (remember were considering three point scales). As you can obtain this information easily from SPSS, R, PRELIS, and Mplus, I will not dwell on these statistics. I call the data file ddat2, it contains the item responses to items 1,2,3,4 and a group indicator (1 or 2). Here is the PRELIS input.

```
title prelis input file
da ni=5 no=1000
ra fi=ddat2
la
v1 v2 v3 v4 sex
or v1 v2 v3 v4
sc gr=1                     ! select group 1
```

```
OU MA=PM !SM=rmat AC=wmat1
```

Rather, I will move on to the polychoric correlations and the tresholds.

Mplus parameterization: Delta

The correlation matrix of of the underlying indicators y* is modeled as follow in two groups:

$$\mathbf{P}_1 = \Delta_1(\Lambda_1\Psi_1\Lambda_1^{\mathbf{t}} + \Theta_1)\,\Delta_1^{\mathbf{t}}$$
$$\mathbf{P}_2 = \Delta_2(\Lambda_2\Psi_2\Lambda_2^{\mathbf{t}} + \Theta_2)\,\Delta_2^{\mathbf{t}}$$

Note that the presence of the diagonal matrix Δ is new. In the socalled delta-parameterization, the matrix Θ is constrained as follows: $\text{diag}(\Theta) = \text{diag}(\mathrm{I}) - \text{diag}(\Lambda\Psi\Lambda^{\mathbf{t}})$, where $\text{diag}(\mathrm{I})$ is the identity matrix. So the diagonal elements of Θ are chosen to ensure that the latent indicators have unit variance. In a single group analysis, the default model is:

$$\mathbf{P} = \Delta(\Lambda\Psi\Lambda^{\mathbf{t}} + \Theta)\Delta \text{ with}$$

$\text{diag}(\Theta) = \text{diag}(\mathrm{I}) - \text{diag}(\Lambda\Psi\Lambda^{\mathbf{t}})$, and $\Delta = \mathbf{I}$. This implies that $\mathbf{P}$ is a correlation matrix.

Step 1 towards MI

1) Estimate the polychoric correlation matrices $\mathbf{R}_1$ and $\mathbf{R}_2$, and thresholds $\mathbf{t}_1$ and $\mathbf{t}_2$ in groups 1 and 2, without any constraints.

I do this in Mplus, in a single analysis (I call the group variable sex and the items v1 to v4):

Mplus INPUT step 1

```
Title:
       step 1
       multiple-group discrete fa
Data:
       file is ddat2;
Variable:
    names are v1 v2 v3 v4 sex;
    usev are v1 v2 v3 v4;
    categorical are v1 v2 v3 v4;
    grouping = sex (1 = female 2= male)
Analysis:
!      type = meanstructure;
       parameterization = delta;
Model:
    f1 BY v1@1;                     ! factor loadings
    f2 BY v2@1;
    f3 BY v3@1;
    f4 BY v4@1;
    f1 with f2 f3 f4;               ! factor correlation
    f2 with f3 f4;
    f3 with f4;
    f1@1 f2@1 f3@1 f4@1;            ! factor variances
    [f1@0 f2@0 f3@0 f4@0];          ! factor means
    [v1$1 v1$2 v2$1 v2$2];          ! thresholds
    [v3$1 v3$2 v4$1 v4$2];
    MODEL MALE:
    {v1@1 v2@1 v3@1 v4@1};          ! scale factors diagonal of Delta matrix
    f1 BY v1@1;
    f2 BY v2@1;
```

```
    f3 BY v3@1;
    f4 BY v4@1;
    f1 with f2 f3 f4;
    f2 with f3 f4;
    f3 with f4;
    f1@1 f2@1 f3@1 f4@1;
    [f1@0 f2@0 f3@0 f4@0];
    [v1$1 v1$2 v2$1 v2$2];
    [v3$1 v3$2 v4$1 v4$2];
Output:
    standardized tech1 tech2;
```

Here are the results (edited) in the female sample. First the correlations:

```
Group FEMALE

F1        WITH
    F2                 0.494      0.050      9.861      0.000
    F3                 0.546      0.042     12.987      0.000
    F4                 0.474      0.051      9.215      0.000

 F2        WITH
    F3                 0.605      0.044     13.837      0.000
    F4                 0.439      0.055      7.936      0.000

 F3        WITH
    F4                 0.552      0.047     11.786      0.000
```

And the thresholds:

```
 Thresholds
    V1$1              -0.885      0.065    -13.660      0.000     item 1 theshold 1
    V1$2              -0.055      0.056     -0.984      0.325     item 1 threshold 2
    V2$1              -0.885      0.065    -13.660      0.000     item 2 etc.
    V2$2              -0.316      0.057     -5.536      0.000
    V3$1              -0.601      0.060    -10.032      0.000
    V3$2               0.238      0.057      4.200      0.000
    V4$1              -0.915      0.065    -13.980      0.000
    V4$2              -0.337      0.057     -5.892      0.000
```

Exercise: the polychroic correlation between items 1 and 2 is estimated at .494. What is its true value? The threshold V1$1 is estimated at -.885. What is its true value?

We have estimated the polychoric correlation matrix of the latent indicators: $\mathbf{P}_k = \mathbf{I}(\mathbf{I}\,\mathbf{\Psi}_k\,\mathbf{I})\mathbf{I}^t = \mathbf{\Psi}_k$ and thresholds $\mathbf{t}_k$ (k=1,2). The means of the latent indicartors are zero in both groups mean($\mathbf{y}^*_k$) = $\mathbf{0}$.

Step 2 towards MI

Note that we estimate polychoric correlations subject to the assumption that the underlying indicators $\mathbf{y}^*$, are standardized (mean=0, variance=1):

$y=0$ if $y^*<t_1$
$y=1$ if $t_1<y^*<t_2$
$y=2$ if $y^*>t_2$,

This is an scaling constraint which stems from the fact that we cannot know the mean and variance of a variable which is latent or unobserved (this is just like scaling common factors in a confirmatory factor model). Now given three point scales, we can standardized the **y*** in one group and subject to equal thresholds over the groups estimate the covariance and means of the **y*** in the second group. We illustrate this in using the following R script and figure.

```
x=seq(-4,4,len=100)
d1=dnorm(x)                  # group 1
d2=dnorm(x*1.5-.5)     # group 2
plot(x,d1,type='l',col=2,lwd=3,xlab='latent item y*')
lines(x,d2,type='l',col=4,lwd=4)
lines(rep(qnorm(.2),10),seq(0,.4,len=10),type='b',lwd=2)
lines(rep(qnorm(.2+.3),10),seq(0,.4,len=10),type='b',lwd=2)
```

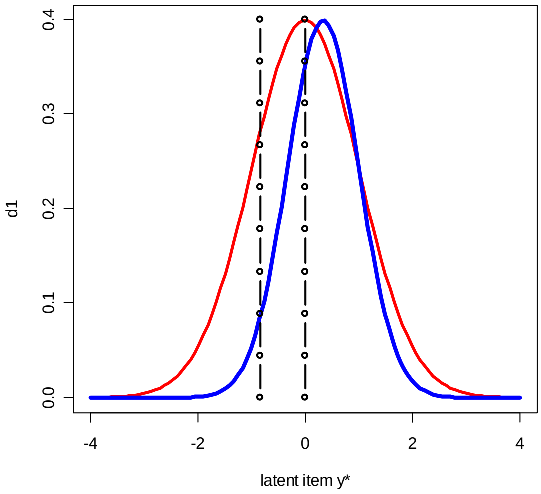


Figure 3-2: latent indicator distributions in two groups with two thresholds.

The thresholds are equal over the groups. Differences in the response frequencies between group 1 (red) and group 2 (blue) are now due to the differnces in the distribution of the latent indicator y*. In the second step, we estimate the polychoric correlation matrix subject to y*~N(0,1) in the first group, and estimate the polychoric covariance matrix and underlying indicator means in the second group.

Mplus input step 2

2) Estimate the polychoric correlation matrix P_1 and thesholds τ in group 1 and estimate the covariance matrix Σ_2 and means μ_2 in group 2. Note that the thresholds are equal over the groups.

```
Title:
       step 2
       multiple-group discrete fa
Data:
       file is ddat2;
Variable:
    names are v1 v2 v3 v4 sex;
    usev are v1 v2 v3 v4;
    categorical are v1 v2 v3 v4;
    grouping = sex (1 = female 2= male)
Analysis:
!      type = meanstructure;
```

```
        parameterization = delta;
Model:
    f1 BY v1@1;
    f2 BY v2@1;
    f3 BY v3@1;
    f4 BY v4@1;
    f1 with f2 f3 f4;
    f2 with f3 f4;
    f3 with f4;
    f1@1 f2@1 f3@1 f4@1;
    [f1@0 f2@0 f3@0 f4@0];
    [v1$1 v1$2 v2$1 v2$2];
    [v3$1 v3$2 v4$1 v4$2];

    MODEL MALE:
    {v1*1 v2*1 v3*1 v4*1};    ! delta elements
    f1 BY v1@1;
    f2 BY v2@1;
    f3 BY v3@1;
    f4 BY v4@1;
    f1 with f2 f3 f4;
    f2 with f3 f4;
    f3 with f4;
    f1@1 f2@1 f3@1 f4@1;
    [f1*0 f2*0 f3*0 f4*0];
!    [v1$1 v1$2 v2$1 v2$2];   ! thresholds not estimated
!    [v3$1 v3$2 v4$1 v4$2];   ! equal to those in group 1
Output:
    standardized tech1 tech2;
```

Output in group 2 (edited):

```
Means           (means of y* in group 2)
    F1               -0.400       0.073      -5.467       0.000
    F2               -0.504       0.076      -6.601       0.000
    F3               -0.548       0.086      -6.388       0.000
    F4               -0.431       0.078      -5.508       0.000

 Thresholds (equal to those in group 1)
    V1$1             -0.885       0.065     -13.659       0.000
    V1$2             -0.055       0.056      -0.984       0.325
    V2$1             -0.885       0.065     -13.658       0.000
    V2$2             -0.316       0.057      -5.539       0.000
    V3$1             -0.601       0.060     -10.029       0.000
    V3$2              0.238       0.057       4.201       0.000
    V4$1             -0.915       0.065     -13.981       0.000
    V4$2             -0.337       0.057      -5.893       0.000

 Variances
    F1                1.000       0.000     999.000     999.000
    F2                1.000       0.000     999.000     999.000
    F3                1.000       0.000     999.000     999.000
    F4                1.000       0.000     999.000     999.000
 Scales         (Matrix Delta in group 2)
    V1                0.976       0.099       9.887       0.000
    V2                0.968       0.124       7.787       0.000
    V3                0.866       0.089       9.713       0.000
    V4                0.965       0.124       7.790       0.000
```

We have estimated the polychoric correlation matrix of the latent indicators and the thresholds in group 1: $\mathbf{P}_1 = \mathbf{I}(\mathbf{I}\,\Psi_1\,\mathbf{I})\,\mathbf{I}^{\mathbf{t}} = \Psi_1$ and thesholds $\mathbf{t}_1$ (k=1,2). The means of the latent indicators are zero in group 1 mean($\mathbf{y}^*_1$) = **0**. In group 2 we have estimated the polychoric covariance matrix of the latent indicators $\mathbf{P}_2 = \Delta(\mathbf{I}\,\Psi_2\,\mathbf{I})\,\Delta^{\mathbf{t}} = \Delta\Psi_2\,\Delta^{\mathbf{t}}$ and the means of the latent indicators mean($\mathbf{y}^*_1$) = [-.400, -.505, -.548, -.421]. Note that Ψ_2 is still standarized. The tresholds in group 2 equal the thresholds in group 1. By constraining the thresholds to be equal, we have sufficient information to estimate the polychoric covariance matrix[4] and means of the latent indicators in group 2. We now assume that the differences in the observed response frequencies are due to the difference in the distribution of the latent indicators.

Step 3(a) towards MI: factor models.

3a) Fit a common factor model the matrices P_1 and Σ_2, without any constraints over groups, i.e., $P_1 = \Lambda_1\,\Psi_1\,\Lambda_1^t + \Theta_1$, and $\Sigma_2 = \Lambda_2\,\Psi_2\,\Lambda_2^t + \Theta_2$. In addition τ are estimated (equal over groups) and μ_2 is estimated (latent indicator means in group 2).

Henceforth we will limit our treatment to the single common factor model, which is the true model. Note a) that standard scaling requirements require some action (I means scaling of the common factor). One possibility is to standardize: $\Psi_1 = \Psi_2 = 1$; b) As before in the delta parameterization, the parameters Θ_1 are not free parameters, as diag($\Lambda_1\,\Psi_1\,\Lambda_1^t + \Theta_1$) = diag($\Lambda_1\,\Lambda_1^t + \Theta_1$) = diag(**I**), where **I** is the identity matrix. In group 2, however no further constraints (beyond scaling) are necessary.

Mplus input step 3a (note the formulation is not standard - see path diagrams below)

```
Title:
    Model 3a
       Multiple-group discrete factor analysis
    1-factor CFA on 4 items
Data:
       file is ddat2;
Variable:
       names are v1 v2 v3 v4 sex;
       usev are v1 v2 v3 v4;
    categorical are v1 v2 v3 v4;
    grouping = sex (1 = female 2= male);
Analysis:
!      type = meanstructure;
    parameterization = delta;
Model:
   f1 by v1;
   f2 by v2;
   f3 by v3;
   f4 by v4;
     f5 by f1* f2 f3 f4;
```

[4] This is so only in the case of three or more point scales. In the case of dichotomous indicators, the Delta matrix cannot be estimated, i.e., it has to be fixed to the identifiy matrix, as in group 1.

```
     f1@0 f2@0 f3@0 f4@0 f5@1;
     [f1@0 f2@0 f3@0 f4@0 f5@0];
     [v1$1 v1$2];
     [v2$1 v2$2];
     [v3$1 v3$2];
     [v4$1 v4$2];
Model male:
     f5 by f1* f2 f3 f4;
     f1@0 f2@0 f3@0 f4@0 f5@1;
     [f1*0 f2*0 f3*0 f4*0 f5@0];
!    {v1@1 v2@1 v3@1 v4@1};
Output:
    standardized tech1 tech2;
```

```
Group FEMALE

 F1       BY
    V1                 1.000      0.000    999.000    999.000

 F2       BY
    V2                 1.000      0.000    999.000    999.000

 F3       BY
    V3                 1.000      0.000    999.000    999.000

 F4       BY
    V4                 1.000      0.000    999.000    999.000

 F5       BY
    F1                 0.681      0.043     15.861      0.000
    F2                 0.716      0.045     15.829      0.000
    F3                 0.827      0.038     22.022      0.000
    F4                 0.661      0.046     14.303      0.000

 Means
    F5                 0.000      0.000    999.000    999.000

 Intercepts
    F1                 0.000      0.000    999.000    999.000
    F2                 0.000      0.000    999.000    999.000
    F3                 0.000      0.000    999.000    999.000
    F4                 0.000      0.000    999.000    999.000

 Thresholds
    V1$1              -0.885      0.065    -13.660      0.000
    V1$2              -0.055      0.056     -0.984      0.325
    V2$1              -0.885      0.065    -13.660      0.000
    V2$2              -0.316      0.057     -5.536      0.000
    V3$1              -0.601      0.060    -10.030      0.000
    V3$2               0.238      0.057      4.198      0.000
    V4$1              -0.915      0.065    -13.979      0.000
    V4$2              -0.337      0.057     -5.893      0.000

 Variances
    F5                 1.000      0.000    999.000    999.000

 Residual Variances
    F1                 0.000      0.000    999.000    999.000
    F2                 0.000      0.000    999.000    999.000
    F3                 0.000      0.000    999.000    999.000
    F4                 0.000      0.000    999.000    999.000

 Scales
```

```
    V1                1.000      0.000    999.000    999.000
    V2                1.000      0.000    999.000    999.000
    V3                1.000      0.000    999.000    999.000
    V4                1.000      0.000    999.000    999.000

Group MALE

 F1       BY
    V1                1.000      0.000    999.000    999.000

 F2       BY
    V2                1.000      0.000    999.000    999.000

 F3       BY
    V3                1.000      0.000    999.000    999.000

 F4       BY
    V4                1.000      0.000    999.000    999.000

 F5       BY
    F1                0.763      0.090      8.461      0.000
    F2                0.854      0.120      7.112      0.000
    F3                0.839      0.103      8.153      0.000
    F4                0.725      0.107      6.802      0.000

 Means
    F5                0.000      0.000    999.000    999.000

 Intercepts
    F1               -0.401      0.073     -5.468      0.000
    F2               -0.504      0.076     -6.596      0.000
    F3               -0.548      0.086     -6.390      0.000
    F4               -0.431      0.078     -5.510      0.000

 Thresholds
    V1$1             -0.885      0.065    -13.660      0.000
    V1$2             -0.055      0.056     -0.984      0.325
    V2$1             -0.885      0.065    -13.660      0.000
    V2$2             -0.316      0.057     -5.536      0.000
    V3$1             -0.601      0.060    -10.030      0.000
    V3$2              0.238      0.057      4.198      0.000
    V4$1             -0.915      0.065    -13.979      0.000
    V4$2             -0.337      0.057     -5.893      0.000

 Variances
    F5                1.000      0.000    999.000    999.000

 Residual Variances
    F1                0.000      0.000    999.000    999.000
    F2                0.000      0.000    999.000    999.000
    F3                0.000      0.000    999.000    999.000
    F4                0.000      0.000    999.000    999.000

 Scales
    V1                0.976      0.099      9.887      0.000
    V2                0.967      0.124      7.787      0.000
    V3                0.866      0.089      9.713      0.000
    V4                0.965      0.124      7.790      0.000
```

We have now estimated in group 1 the polychoric correlation matrix and the thresholds: $\mathbf{P}_1 = \mathbf{I}_1(\mathbf{\Lambda}_1\mathbf{\Psi}_1\mathbf{\Lambda}_1^{\mathbf{t}} + \mathbf{\Theta}_1)\mathbf{I}_1^{\mathbf{t}}$ and $\mathbf{t}_1$. And we have estimated in group 2 the polychoric covariance matrix and the latent indicator means: $\mathbf{P}_2$ =

$\Delta_2(\Lambda_2\Psi_2\Lambda_2^{\mathbf{t}} + \Theta_2)\Delta_2^{\mathbf{t}}$ and mean($\mathbf{y}$*)= [-.401, -.504, -.548,-.431]. Note that $\Lambda_2\Psi_2\Lambda_2^{\mathbf{t}} + \Theta_2$ is still a correlation matrix. The tresholds in group 2 equal those in group 1.

You will have noted that we defined the factor model in the following way:

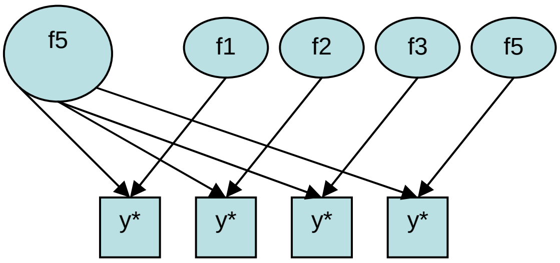


Figure 3-3: residuals as factors.

Rather than in the more nature way:

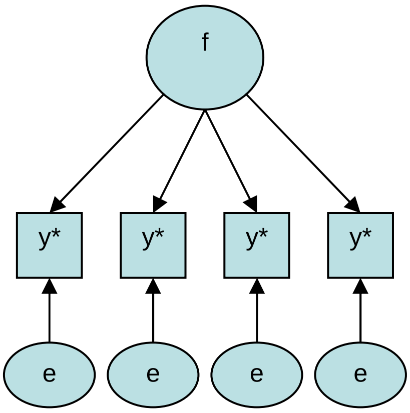


Figure 3-4: more natural: residuals as residuals.

We shall now formulate the model in the more natural way (step 3b).

Step 3b towards MI (alternative to step 3a syntax)

3b) One could also introduce the factor models following step 1. That is estimate the polychoric correlation matrices subject to the factor model: $P_1 = \Lambda_1 \Psi_1 \Lambda_1^t + \Theta_1$, and $P_2 = \Lambda_2 \Psi_2 \Lambda_2^t + \Theta_2$. Given scaling requirements, we require $diag(\Lambda_1 \Lambda_1^t + \Theta_1)=diag(\Lambda_2 \Lambda_2^t + \Theta_2)=diag(I)$, i.e., in both groups the residual covariance matrices Θ_1 and Θ_2 are constrained, and thresholds τ_1 and τ_2 are estimated (NOT equal over the groups).

Mplus input step 3b

```
Title:
    model 3b
       Multiple-group discrete factor analysis
    1-factor CFA on 4 items
Data:
       file is ddat2;
```

```
Variable:
       names are v1 v2 v3 v4 sex;
       usev are v1 v2 v3 v4;
    categorical are v1 v2 v3 v4;
    grouping = sex (1 = female 2= male)
Analysis:
!      type =  meanstructure;
    parameterization = delta;
Model:
    f by v1*.5 v2*.5 v3*.5 v4*.5;
    f@1;
    [f@0];
    [v1$1 v1$2];
    [v2$1 v2$2];
    [v3$1 v3$2];
    [v4$1 v4$2];
Model male:
    f by v1*.5 v2*.5 v3*.5 v4*.5;
    f@1;
    [f@0];
    [v1$1 v1$2];
    [v2$1 v2$2];
    [v3$1 v3$2];
    [v4$1 v4$2];
    {v1@1 v2@1 v3@1 v4@1};

Output:
    standardized tech1 tech2;
```

```
Group FEMALE

 F        BY
    V1                 0.681      0.043     15.860      0.000
    V2                 0.716      0.045     15.828      0.000
    V3                 0.827      0.038     22.020      0.000
    V4                 0.661      0.046     14.302      0.000

 Means
    F                  0.000      0.000    999.000    999.000

 Thresholds
    V1$1              -0.885      0.065    -13.660      0.000
    V1$2              -0.055      0.056     -0.984      0.325
    V2$1              -0.885      0.065    -13.660      0.000
    V2$2              -0.316      0.057     -5.536      0.000
    V3$1              -0.601      0.060    -10.032      0.000
    V3$2               0.238      0.057      4.200      0.000
    V4$1              -0.915      0.065    -13.980      0.000
    V4$2              -0.337      0.057     -5.892      0.000

 Variances
    F                  1.000      0.000    999.000    999.000

 Scales
    V1                 1.000      0.000    999.000    999.000
    V2                 1.000      0.000    999.000    999.000
    V3                 1.000      0.000    999.000    999.000
    V4                 1.000      0.000    999.000    999.000

Group MALE

 F        BY
    V1                 0.745      0.036     20.628      0.000
```

```
   V2                 0.826       0.032      25.849       0.000
   V3                 0.727       0.038      18.886       0.000
   V4                 0.700       0.040      17.394       0.000

 Means
   F                  0.000       0.000     999.000     999.000

 Thresholds
   V1$1              -0.473       0.058      -8.106       0.000
   V1$2               0.337       0.057       5.892       0.000
   V2$1              -0.369       0.057      -6.425       0.000
   V2$2               0.181       0.056       3.218       0.001
   V3$1              -0.045       0.056      -0.805       0.421
   V3$2               0.681       0.061      11.154       0.000
   V4$1              -0.468       0.058      -8.017       0.000
   V4$2               0.090       0.056       1.610       0.107

 Variances
   F                  1.000       0.000     999.000     999.000

 Scales
   V1                 1.000       0.000     999.000     999.000
   V2                 1.000       0.000     999.000     999.000
   V3                 1.000       0.000     999.000     999.000
   V4                 1.000       0.000     999.000     999.000
```

We have now estimated in group 1 the polychoric correlation matrix and the thresholds: $\mathbf{P}_1 = \mathbf{I}_1(\mathbf{\Lambda}_1\mathbf{\Psi}_1\mathbf{\Lambda}_1^{\mathbf{t}} + \mathbf{\Theta}_1)\mathbf{I}_1^{\mathbf{t}}$ and $\mathbf{t}_1$. And we have estimated in group 2 the polychoric covariance matrix and the latent indicator means: $\mathbf{P}_2 = \mathbf{I}_2(\mathbf{\Lambda}_2\mathbf{\Psi}_2\mathbf{\Lambda}_2^{\mathbf{t}} + \mathbf{\Theta}_2)\mathbf{I}_2^{\mathbf{t}}$ and thresholds $\mathbf{t}_2$ (not equal to $\mathbf{t}_1$!). The latent indicator means are zero in both groups mean($\mathbf{y}$*)= [0,0,0,0], and as always $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ are not free parameter matrices.

Step 3a and 3b serve to establish the factor model without further constraints. That is to say the object is only to establish the dimensionality of the set of items. We assume that this dimenionality is identical in the groups (a single factor model), but this is not a requirement that is associated with MI. That is, MI does not require the number of factors to be equal over the groups (2 common factors could be correlated .7 in group 1 and correlated 1 in group 2, i.e., in group 2 the model would be effectively a single common factor model). The model in step 3 fits well, we conclude that the factor models are correctly spectify, and we can continue with step 4.

Step 4 towards MI - equal factor loadings or metric invariance.

4) In step 4, we proceed by constraining the parameters of the factor model. The first step towards MI is to constrains the factor loadings to be equal over the groups: $P_1 = \Lambda\,\Psi_1\,\Lambda^t + \Theta_1$ *and* $\Sigma_2 = \Lambda\,\Psi_2\,\Lambda^t + \Theta_2$*, subject to the equality constraints on the thresholds,* τ*, and, remembering that* μ_2 *is freely estimated in group 2.*

Given the scaling constraint Ψ_1=1, we have $P_1 = \Lambda\,\Lambda^t + \Theta_1$ and $\Sigma_2 = \Lambda\,\Psi_2\,\Lambda^t + \Theta_2$. Note that Ψ_2 is now a free parameter, i.e., the equality constraint on the factor loadings (Λ) enables us to estimate the factor variance in group 2. The hypothesis Ψ_2=1 may be of interest, but is irrelvant to the issue of MI (i.e., MI does not prescribe Ψ_2=1). Note that we still require diag($\Lambda_1\,\Lambda_1^t + \Theta_1$)=diag($\mathbf{I}$).

Mplus step 4 input

```
Title:
       model 4
       Multiple-group discrete factor analysis
    1-factor CFA on 4 items
Data:
       file is ddat2;
Variable:
       names are v1 v2 v3 v4 sex;
       usev are v1 v2 v3 v4;
    categorical are v1 v2 v3 v4;
    grouping = sex (1 = female 2= male);
Analysis:
!      type =  meanstructure;
    parameterization = delta;
Model:
   f1 by v1@1;
   f2 by v2@1;
   f3 by v3@1;
   f4 by v4@1;
   f5 by f1* (1);
   f5 by f2  (2);
   f5 by f3  (3) ;
   f5 by f4  (4);
   f1@0 f2@0 f3@0 f4@0 f5@1;
   [f1@0 f2@0 f3@0 f4@0 f5@0];
   [v1$1 v1$2];
   [v2$1 v2$2];
   [v3$1 v3$2];
   [v4$1 v4$2];
Model male:
   f5 by f1* (1);
   f5 by f2  (2);
   f5 by f3  (3) ;
   f5 by f4  (4);
     f1@0 f2@0 f3@0 f4@0 f5*1;
     [f1*0 f2*0 f3*0 f4*0 f5@0];
!    {v1@1 v2@1 v3@1 v4@1};
Output:
    standardized tech1 tech2;
```

Output

```
Group FEMALE

 F1       BY
    V1                 1.000      0.000    999.000    999.000

 F2       BY
    V2                 1.000      0.000    999.000    999.000

 F3       BY
    V3                 1.000      0.000    999.000    999.000

 F4       BY
    V4                 1.000      0.000    999.000    999.000

 F5       BY
    F1                 0.684      0.039     17.580      0.000
    F2                 0.721      0.042     17.040      0.000
    F3                 0.818      0.034     23.847      0.000
    F4                 0.662      0.043     15.412      0.000

 Means
    F5                 0.000      0.000    999.000    999.000

 Intercepts
    F1                 0.000      0.000    999.000    999.000
    F2                 0.000      0.000    999.000    999.000
    F3                 0.000      0.000    999.000    999.000
    F4                 0.000      0.000    999.000    999.000

 Thresholds
    V1$1              -0.881      0.062    -14.268      0.000
    V1$2              -0.058      0.055     -1.059      0.290
    V2$1              -0.873      0.061    -14.202      0.000
    V2$2              -0.326      0.057     -5.745      0.000
    V3$1              -0.611      0.058    -10.514      0.000
    V3$2               0.247      0.056      4.423      0.000
    V4$1              -0.915      0.062    -14.792      0.000
    V4$2              -0.337      0.057     -5.913      0.000

 Variances
    F5                 1.000      0.000    999.000    999.000

 Residual Variances
    F1                 0.000      0.000    999.000    999.000
    F2                 0.000      0.000    999.000    999.000
    F3                 0.000      0.000    999.000    999.000
    F4                 0.000      0.000    999.000    999.000

 Scales
    V1                 1.000      0.000    999.000    999.000
    V2                 1.000      0.000    999.000    999.000
    V3                 1.000      0.000    999.000    999.000
    V4                 1.000      0.000    999.000    999.000

Group MALE

 F1       BY
    V1                 1.000      0.000    999.000    999.000

 F2       BY
    V2                 1.000      0.000    999.000    999.000
```

```
 F3       BY
    V3                 1.000      0.000    999.000    999.000

 F4       BY
    V4                 1.000      0.000    999.000    999.000

 F5       BY
    F1                 0.684      0.039     17.580      0.000
    F2                 0.721      0.042     17.040      0.000
    F3                 0.818      0.034     23.847      0.000
    F4                 0.662      0.043     15.412      0.000

 Means
    F5                 0.000      0.000    999.000    999.000

 Intercepts
    F1                -0.401      0.073     -5.519      0.000
    F2                -0.509      0.074     -6.891      0.000
    F3                -0.569      0.085     -6.665      0.000
    F4                -0.431      0.078     -5.508      0.000

 Thresholds
    V1$1              -0.881      0.062    -14.268      0.000
    V1$2              -0.058      0.055     -1.059      0.290
    V2$1              -0.873      0.061    -14.202      0.000
    V2$2              -0.326      0.057     -5.745      0.000
    V3$1              -0.611      0.058    -10.514      0.000
    V3$2               0.247      0.056      4.423      0.000
    V4$1              -0.915      0.062    -14.792      0.000
    V4$2              -0.337      0.057     -5.913      0.000

 Variances
    F5                 1.198      0.205      5.859      0.000

 Residual Variances
    F1                 0.000      0.000    999.000    999.000
    F2                 0.000      0.000    999.000    999.000
    F3                 0.000      0.000    999.000    999.000
    F4                 0.000      0.000    999.000    999.000

 Scales
    V1                 0.993      0.076     12.990      0.000
    V2                 1.041      0.095     10.974      0.000
    V3                 0.818      0.065     12.658      0.000
    V4                 0.966      0.089     10.911      0.000
```

We have now estimated in group 1 the polychoric correlation matrix and the thresholds: $\mathbf{P}_1 = \mathbf{I}_1(\mathbf{\Lambda}\mathbf{\Psi}_1\mathbf{\Lambda}^{\mathbf{t}} + \mathbf{\Theta}_1)\mathbf{I}_1^{\mathbf{t}}$ and $\mathbf{t}_1$. Note that due to scaling $\mathbf{\Psi}_1$=1. We have estimated in group 2 the polychoric covariance matrix and the latent indicator means: $\mathbf{P}_2 = \mathbf{\Delta}_2(\mathbf{\Lambda}\mathbf{\Psi}_2\mathbf{\Lambda}^{\mathbf{t}} + \mathbf{\Theta}_2)\mathbf{\Delta}_2^{\mathbf{t}}$ and thresholds $\mathbf{t}_2$ equal to $\mathbf{t}_1$. The latent indicator means are zero in group 1, and estimated in group 2 mean($\mathbf{y}$*) = [-.401,-.509,-.569, -.431]. As always $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ are not free parameter matrices.

Step 5 towards MI: strong factorial invariance

5) In step 5, we proceed by constraining the mean vector in group 2, i.e., $\mu_2 = \Lambda\alpha$, where α is the difference in factor mean between group 1 and group 2. Thus we estimate the thresholds τ, the factor models $P_1 = \Lambda \Lambda^t + \Theta_1$ and $\Sigma_2 = \Lambda \Psi_2 \Lambda^t + \Theta_2$, and the means model $\mu_2 = \Lambda\alpha$.

In terms of the usual taxonomy, this model is the strong factorial invariance model. Here again note that we still require $\mathrm{diag}(\Lambda_1 \Lambda_1^t + \Theta_1) = \mathrm{diag}(\mathbf{I})$.

Mplus input Step 5

```
Title:
 model 5
        Multiple-group discrete factor analysis
    1-factor CFA on 4 items
Data:
        file is ddat2;
Variable:
        names are v1 v2 v3 v4 sex;
        usev are v1 v2 v3 v4;
    categorical are v1 v2 v3 v4;
    grouping = sex (1 = female 2= male)
Analysis:
!       type =  meanstructure;
    parameterization = delta;
Model:
   f1 by v1@1;
   f2 by v2@1;
   f3 by v3@1;
   f4 by v4@1;
   f5 by f1* (1);
   f5 by f2  (2);
   f5 by f3  (3) ;
   f5 by f4  (4);
   f1@0 f2@0 f3@0 f4@0 f5@1;
   [f1@0 f2@0 f3@0 f4@0 f5@0];
   [v1$1 v1$2];
   [v2$1 v2$2];
   [v3$1 v3$2];
   [v4$1 v4$2];
Model male:
   f5 by f1* (1);
   f5 by f2  (2);
   f5 by f3  (3) ;
   f5 by f4  (4);
   f1@0 f2@0 f3@0 f4@0 f5*1;
   [f1@0 f2@0 f3@0 f4@0 f5*0];
!    {v1@1 v2@1 v3@1 v4@1};
Output:
    standardized tech1 tech2;
```

Output

```
Group FEMALE

 F1       BY
    V1                 1.000      0.000    999.000    999.000

 F2       BY
    V2                 1.000      0.000    999.000    999.000

 F3       BY
    V3                 1.000      0.000    999.000    999.000

 F4       BY
    V4                 1.000      0.000    999.000    999.000
```

```
 F5       BY
    F1                0.671      0.036     18.691      0.000
    F2                0.731      0.038     19.146      0.000
    F3                0.823      0.033     24.754      0.000
    F4                0.660      0.038     17.236      0.000

 Means
    F5                0.000      0.000    999.000    999.000

 Intercepts
    F1                0.000      0.000    999.000    999.000
    F2                0.000      0.000    999.000    999.000
    F3                0.000      0.000    999.000    999.000
    F4                0.000      0.000    999.000    999.000

 Thresholds
    V1$1             -0.899      0.056    -15.932      0.000
    V1$2             -0.084      0.049     -1.733      0.083
    V2$1             -0.860      0.058    -14.739      0.000
    V2$2             -0.313      0.049     -6.350      0.000
    V3$1             -0.600      0.055    -10.976      0.000
    V3$2              0.254      0.055      4.626      0.000
    V4$1             -0.916      0.058    -15.703      0.000
    V4$2             -0.341      0.048     -7.037      0.000

 Variances
    F5                1.000      0.000    999.000    999.000

 Residual Variances
    F1                0.000      0.000    999.000    999.000
    F2                0.000      0.000    999.000    999.000
    F3                0.000      0.000    999.000    999.000
    F4                0.000      0.000    999.000    999.000

 Scales
    V1                1.000      0.000    999.000    999.000
    V2                1.000      0.000    999.000    999.000
    V3                1.000      0.000    999.000    999.000
    V4                1.000      0.000    999.000    999.000

Group MALE

 F1       BY
    V1                1.000      0.000    999.000    999.000

 F2       BY
    V2                1.000      0.000    999.000    999.000

 F3       BY
    V3                1.000      0.000    999.000    999.000

 F4       BY
    V4                1.000      0.000    999.000    999.000

 F5       BY
    F1                0.671      0.036     18.691      0.000
    F2                0.731      0.038     19.146      0.000
    F3                0.823      0.033     24.754      0.000
    F4                0.660      0.038     17.236      0.000

 Means
    F5               -0.663      0.083     -8.018      0.000

 Intercepts
    F1                0.000      0.000    999.000    999.000
    F2                0.000      0.000    999.000    999.000
    F3                0.000      0.000    999.000    999.000
    F4                0.000      0.000    999.000    999.000

 Thresholds
    V1$1             -0.899      0.056    -15.932      0.000
    V1$2             -0.084      0.049     -1.733      0.083
    V2$1             -0.860      0.058    -14.739      0.000
    V2$2             -0.313      0.049     -6.350      0.000
    V3$1             -0.600      0.055    -10.976      0.000
    V3$2              0.254      0.055      4.626      0.000
    V4$1             -0.916      0.058    -15.703      0.000
```

```
   V4$2              -0.341      0.048     -7.037      0.000

 Variances
   F5                 1.179      0.200      5.892      0.000

 Residual Variances
   F1                 0.000      0.000    999.000    999.000
   F2                 0.000      0.000    999.000    999.000
   F3                 0.000      0.000    999.000    999.000
   F4                 0.000      0.000    999.000    999.000

 Scales
   V1                 1.017      0.078     13.008      0.000
   V2                 1.034      0.088     11.810      0.000
   V3                 0.824      0.065     12.629      0.000
   V4                 0.977      0.082     11.977      0.000
```

We have now estimated in group 1 the polychoric correlation matrix and the thresholds: $\mathbf{P}_1 = \mathbf{I}_1(\boldsymbol{\Lambda}\boldsymbol{\Psi}_1\boldsymbol{\Lambda}^{\mathbf{t}} + \boldsymbol{\Theta}_1)\mathbf{I}_1^{\mathbf{t}}$ and $\mathbf{t}_1$. We have estimated in group 2 the polychoric covariance matrix and the latent indicator means: $\mathbf{P}_2 = \boldsymbol{\Delta}_2(\boldsymbol{\Lambda}\boldsymbol{\Psi}_2\boldsymbol{\Lambda}^{\mathbf{t}} + \boldsymbol{\Theta}_2)\boldsymbol{\Delta}_2^{\mathbf{t}}$ and thresholds $\mathbf{t}_2$ equal to $\mathbf{t}_1$. The latent indicator means are zero in group 1, and estimated in group 2 as follows: $\boldsymbol{\Lambda}\boldsymbol{\alpha}_2$ ($\boldsymbol{\alpha}_2$ is the common factor mean in group 2, the value of $\boldsymbol{\alpha}_2$ = -.663, its true value is -.5). So mean($\mathbf{y}$*) in group 2 = [(.671*-.663), (.731*-.663),(.823*-.663),(.660*-.663)] = [-0.444873 -0.484653 -0.545649 -0.437580]. As always $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$ are not free parameter matrices. The chi2 of this model is 5.684 with 10 degrees of Freedom.

You will note that we have reverted to the model specification of Figure 3-3. We consider the more natural specification of the same model.

Mplus step 5 input (simpler formulation)

```
Title:
       model 5b alternative simpler
       Multiple-group discrete factor analysis
    1-factor CFA on 4 items
Data:
       file is ddat2;
Variable:
       names are v1 v2 v3 v4 sex;
       usev are v1 v2 v3 v4;
    categorical are v1 v2 v3 v4;
    grouping = sex (1 = female 2= male)
Analysis:
!      type =  meanstructure;
    parameterization = delta;
Model:
    f by v1*.5 v2*.5 v3*.5 v4*.5;
    f@1;
    [f@0];
    [v1$1 v1$2];
    [v2$1 v2$2];
    [v3$1 v3$2];
    [v4$1 v4$2];
Model male:
    f*1;
    [f*0];
!    {v1@1 v2@1 v3@1 v4@1};
Output:
    standardized tech1 tech2;
```

Output

```
Group FEMALE

 F        BY
    V1                 0.671      0.036     18.691      0.000
    V2                 0.731      0.038     19.146      0.000
    V3                 0.823      0.033     24.755      0.000
    V4                 0.660      0.038     17.236      0.000

 Means
    F                  0.000      0.000    999.000    999.000

 Thresholds
    V1$1              -0.899      0.056    -15.932      0.000
    V1$2              -0.084      0.049     -1.733      0.083
    V2$1              -0.860      0.058    -14.739      0.000
    V2$2              -0.313      0.049     -6.350      0.000
    V3$1              -0.600      0.055    -10.976      0.000
    V3$2               0.254      0.055      4.625      0.000
    V4$1              -0.916      0.058    -15.703      0.000
    V4$2              -0.341      0.048     -7.037      0.000

 Variances
    F                  1.000      0.000    999.000    999.000

 Scales
    V1                 1.000      0.000    999.000    999.000
    V2                 1.000      0.000    999.000    999.000
    V3                 1.000      0.000    999.000    999.000
    V4                 1.000      0.000    999.000    999.000

Group MALE

 F        BY
    V1                 0.671      0.036     18.691      0.000
    V2                 0.731      0.038     19.146      0.000
    V3                 0.823      0.033     24.755      0.000
    V4                 0.660      0.038     17.236      0.000

 Means
    F                 -0.663      0.083     -8.018      0.000

 Thresholds
    V1$1              -0.899      0.056    -15.932      0.000
    V1$2              -0.084      0.049     -1.733      0.083
    V2$1              -0.860      0.058    -14.739      0.000
    V2$2              -0.313      0.049     -6.350      0.000
    V3$1              -0.600      0.055    -10.976      0.000
    V3$2               0.254      0.055      4.625      0.000
    V4$1              -0.916      0.058    -15.703      0.000
    V4$2              -0.341      0.048     -7.037      0.000

 Variances
    F                  1.179      0.200      5.892      0.000

 Scales
    V1                 1.017      0.078     13.008      0.000
    V2                 1.034      0.088     11.810      0.000
    V3                 0.824      0.065     12.629      0.000
    V4                 0.978      0.082     11.977      0.000
```

The results, we hope, are identical. But the model is simpler. The chi2 is again 5.684 with 10 degrees of Freedom.

Step 5 towards MI: Switch to theta parameterization.

The delta parameterization does not seem to be suitable to test strict factorial invariance. We therefore switch to the theta parameterization. This involves fixing the residual (error) variances of the latent indicators to arbitrary values (i.e, the parameters in **Θ**). We shall first repeat the step 5 using this parameterization. Note that we shall fix the residual variances to their true values (i.e., .5 .4 .45 .55). This is arbitrary, but facilitates the evaluation of the parameter recovery. The results would not change substantively, if you changed the fixed values to, say, .7,.7,.7,.7 (as you can establish for yourself).

Mplus Step 5 Theta parameterization

```
Title:
       model 5c (model 5 theta)
       Multiple-group discrete analysis
    1-factor CFA on 4 items
Data:
       file is ddat2;
Variable:
       names are v1 v2 v3 v4 sex;
       usev are v1 v2 v3 v4;
    categorical are v1 v2 v3 v4;
    grouping = sex (1 = female 2= male)
Analysis:
!      type =  meanstructure;
    parameterization = theta;
Model:
    f by v1*.5 v2*.5 v3*.5 v4*.5;
    f@1;
    [f@0];
    [v1$1 v1$2];
    [v2$1 v2$2];
    [v3$1 v3$2];
    [v4$1 v4$2];
Model female:
! true value  FIXED
    v1@0.5 v2@0.4 v3@0.45 v4@0.55;
Model male:
    f*1;
    [f*-.5];
!  estimated strong factorial inv.
    v1*0.5 v2*0.4 v3*0.45 v4*0.55;
Output:
    standardized tech1 tech2;
```

OUTPUT

```
Chi-Square Test of Model Fit

          Value                              5.684*
          Degrees of Freedom                    10
          P-Value                           0.8411

Group FEMALE

 F        BY
    V1                 0.640       0.062      10.271       0.000
    V2                 0.679       0.076       8.901       0.000
```

```
    V3                 0.970      0.121      8.004      0.000
    V4                 0.651      0.067      9.729      0.000

 Means
    F                  0.000      0.000    999.000    999.000

 Thresholds
    V1$1              -0.858      0.070    -12.335      0.000
    V1$2              -0.080      0.047     -1.716      0.086
    V2$1              -0.798      0.078    -10.200      0.000
    V2$2              -0.290      0.051     -5.702      0.000
    V3$1              -0.708      0.090     -7.888      0.000
    V3$2               0.299      0.068      4.390      0.000
    V4$1              -0.904      0.075    -12.004      0.000
    V4$2              -0.337      0.052     -6.457      0.000

 Variances
    F                  1.000      0.000    999.000    999.000

 Residual Variances
    V1                 0.500      0.000    999.000    999.000
    V2                 0.400      0.000    999.000    999.000
    V3                 0.450      0.000    999.000    999.000
    V4                 0.550      0.000    999.000    999.000

Group MALE

 F        BY
    V1                 0.640      0.062     10.271      0.000
    V2                 0.679      0.076      8.901      0.000
    V3                 0.970      0.121      8.004      0.000
    V4                 0.651      0.067      9.729      0.000

 Means
    F                 -0.663      0.083     -8.019      0.000

 Thresholds
    V1$1              -0.858      0.070    -12.335      0.000
    V1$2              -0.080      0.047     -1.716      0.086
    V2$1              -0.798      0.078    -10.200      0.000
    V2$2              -0.290      0.051     -5.702      0.000
    V3$1              -0.708      0.090     -7.888      0.000
    V3$2               0.299      0.068      4.390      0.000
    V4$1              -0.904      0.075    -12.004      0.000
    V4$2              -0.337      0.052     -6.457      0.000

 Variances
    F                  1.179      0.200      5.892      0.000

 Residual Variances
    V1                 0.396      0.091      4.368      0.000
    V2                 0.262      0.076      3.465      0.001
    V3                 0.941      0.266      3.535      0.000
    V4                 0.519      0.130      4.005      0.000
```

Step 6 towards MI: strict factorial invariance.

6) In step 6, finally, we constrain the residual variances to be equal over the groups. Thus we estimate the thresholds τ, *the factor models* $P_1 = \Lambda\Lambda^t + \Theta$ *and* $\Sigma_2 = \Lambda\Psi_2\Lambda^t + \Theta$, *and the means model* $\mu_2 = \Lambda\alpha$. *Note that above we required* $diag(\Lambda_1\Lambda_1^t + \Theta) = diag(I)$. *However, in the present model,* Θ *is fixed and constrained to be equal over the groups. This is the theta parameterisation.*

In terms of the usual taxonomy, this model is the strict factorial invariance model. This model represents full measurement invariance.

Mplus in put Step 6 Theta parameterization

```
Title:
       model6
       Multiple-group discrete factor analysis
    1-factor CFA on 4 items
Data:
       file is ddat2;
Variable:
       names are v1 v2 v3 v4 sex;
       usev are v1 v2 v3 v4;
    categorical are v1 v2 v3 v4;
    grouping = sex (1 = female 2= male)
Analysis:
!      type =  meanstructure;
    parameterization = theta;
Model:
    f by v1*.5 v2*.5 v3*.5 v4*.5;
    f@1;
    [f@0];
    [v1$1 v1$2];
    [v2$1 v2$2];
    [v3$1 v3$2];
    [v4$1 v4$2];
Model female:
    v1@0.51 v2@0.64 v3@0.75 v4@0.84;
Model male:
    f*1;
    [f*0];
    v1@0.51 v2@0.64 v3@0.75 v4@0.84;
Output:
    standardized tech1 tech2;
```

Output

```
TESTS OF MODEL FIT

Chi-Square Test of Model Fit

          Value                              14.114*
          Degrees of Freedom                     14
          P-Value                            0.4413


Group FEMALE

 F        BY
    V1                  0.693      0.060     11.534      0.000
    V2                  0.952      0.093     10.234      0.000
    V3                  1.019      0.089     11.403      0.000
    V4                  0.826      0.075     10.979      0.000

 Means
    F                   0.000      0.000    999.000    999.000

 Thresholds
    V1$1               -0.909      0.062    -14.700      0.000
    V1$2               -0.073      0.049     -1.484      0.138
```

```
    V2$1              -1.096      0.088    -12.432      0.000
    V2$2              -0.383      0.069     -5.547      0.000
    V3$1              -0.759      0.078     -9.685      0.000
    V3$2               0.312      0.068      4.573      0.000
    V4$1              -1.129      0.077    -14.609      0.000
    V4$2              -0.414      0.064     -6.445      0.000

 Variances
    F                  1.000      0.000    999.000    999.000

 Residual Variances
    V1                 0.510      0.000    999.000    999.000
    V2                 0.640      0.000    999.000    999.000
    V3                 0.750      0.000    999.000    999.000
    V4                 0.840      0.000    999.000    999.000

Group MALE

 F        BY
    V1                 0.693      0.060     11.534      0.000
    V2                 0.952      0.093     10.234      0.000
    V3                 1.019      0.089     11.403      0.000
    V4                 0.826      0.075     10.979      0.000

 Means
    F                 -0.640      0.082     -7.855      0.000

 Thresholds
    V1$1              -0.909      0.062    -14.700      0.000
    V1$2              -0.073      0.049     -1.484      0.138
    V2$1              -1.096      0.088    -12.432      0.000
    V2$2              -0.383      0.069     -5.547      0.000
    V3$1              -0.759      0.078     -9.685      0.000
    V3$2               0.312      0.068      4.573      0.000
    V4$1              -1.129      0.077    -14.609      0.000
    V4$2              -0.414      0.064     -6.445      0.000

 Variances
    F                  1.179      0.187      6.309      0.000

 Residual Variances
    V1                 0.510      0.000    999.000    999.000
    V2                 0.640      0.000    999.000    999.000
    V3                 0.750      0.000    999.000    999.000
    V4                 0.840      0.000    999.000    999.000
```

We have now estimated in group 1 the polychoric correlation matrix and the thresholds, the latent indicator means are fixed to zero:

$\mathbf{P}_1 = (\Lambda\Psi_1\Lambda^t + \Theta)$ and $\mathbf{t}_1$ (Ψ_1=1) and mean($\mathbf{y}$*) = 0.

We have estimated in group 2 the polychoric covariance matrix and the latent indicator means:

$\mathbf{P}_2 = (\Lambda\Psi_2\Lambda^t + \Theta)$ and thresholds $\mathbf{t}_2$ equal to $\mathbf{t}_1$ and mean($\mathbf{y}$*) = $\Lambda\alpha_2$

The differences between the groups in the correlation matrices and the means of the latent indicators are due solely to differences in the common factor distribution (N(0,1) in group 1 and $N(\alpha_2,\Psi_2)$= N(-0.640,1.179) in group 2. The tresholds which connect the latent indicators to the observed ordinal indicators are equal over the groups.

Just a check.

Here finally is the input for the analyses with all parameters fixed to their true values, using the delta parametrization. This is just a check.

true values from the R script:

```
> al2
     [,1]
[1,] -0.5
> ps1
     [,1]
[1,]    1
> ps2
     [,1]
[1,]    1
> al1
     [,1]
[1,]    0
> ly
          [,1]
[1,] 0.7071068
[2,] 0.7745967
[3,] 0.7416198
[4,] 0.6708204
>
>
> thresholds
     [,1]       [,2]       [,3] [,4]
[1,] -Inf -0.8416212  0.0000000  Inf
[2,] -Inf -0.8416212 -0.2533471  Inf
[3,] -Inf -0.5244005  0.2533471  Inf
[4,] -Inf -0.8416212 -0.2533471  Inf
```

Mplus input: true values all fixed.

```
Title:
       model check fixed to true values
       Multiple-group discrete analysis
    1-factor CFA on 4 items
Data:
       file is ddat2;
Variable:
       names are v1 v2 v3 v4 sex;
       usev are v1 v2 v3 v4;
    categorical are v1 v2 v3 v4;
    grouping = sex (1 = female 2= male)
Analysis:
!      type =  meanstructure;
    parameterization = theta;
Model:
    f by v1@.7071068 v2@.7745969 v3@.7416198 v4@.6708204;
    f@1;
    [f@0];
    [v1$1@-0.8416212   v1$2@0.0000000  ];
    [v2$1@-0.8416212   v2$2@-0.2533471];
    [v3$1@-0.5244005   v3$2@0.2533471];
    [v4$1@-0.8416212   v4$2@-0.2533471];
Model female:
    v1@0.5 v2@0.4 v3@0.45 v4@0.55;
Model male:
    f@1;
    [f@-.5];
    v1@0.5 v2@0.4 v3@0.45 v4@0.55;
Output:
    standardized tech1 tech2;
```

The model should fit the data well!

Chi-Square Test of Model Fit

```
          Value                              28.952*
          Degrees of Freedom                     28
          P-Value                            0.4150

CFI/TLI

          CFI                                 0.999
          TLI                                 1.000

Number of Free Parameters                         0

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                            0.008
```

Assignment:

Using the R code given above (page 4-5), simulate a dataset with parameter values of your own choice, and fit the models in Mplus as described above. In each analysis, state the meaning of the model, list the parameter estimates, and state what they mean.

Lecture note VI.

Reference.

Muthen, B. and Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: multi-group and growthmodeling in Mplus. Mplus Web Notes, no. 4.
[an clear account of ordinal factor analysis in Mplus].

The aims of these final lecture notes are the following:

1) To return to the ordinal factor model with the specific aim of explaining, in more detail, the delta and theta parameterizations in Mplus.

2) To return to the original definition of measurement invariance to outline how the definition gives to a highly constrained multigroup ordinal factor model (note that you have already fitted this model).

3) To briefly discuss other measurement invariance in other measurement models, and finally some remaining details.

1.0: Delta parameterization, theta parameterization.

We return to the *delta* and *theta* parameterizations of the ordinal factor model in Mplus. As before we consider a two group model. Again we assume that the observed indicators in the factor model are ordinal (**y**), and that underlying each ordinal indicator there is a latent continuous indicator (**y***). The latent continuous variance satisfy:

$$\mathbf{y}_i^* = \boldsymbol{\tau} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i.$$

As discussed previously, the observed indicators are related to the latent indicators as follows. In the case of C categories, we have y=c, if $t_c<y^*\leq t_{c+1}$, where c=0,1,...,C-1, and t_0=-inf; t_C=+inf. The parameters t_1 and t_2 are thresholds, i.e., points on the normal distribution, i.e., the distribution of y*. For instance in the case of C=3, we have

$y=0$ if $t_0<y^*\leq t_1$, or $y=0$ if $-\infty<y^*\leq t_1$
$y=1$ if $t_1<y^*\leq t_2$,
$y=2$ if $t_2<y^*<t_3$, or $y=2$ if $t_2<y^*\leq+\infty$

The means and covariance matrix of the underlying indicators **y*** is modeled as follows (in a given group k):

$$\boldsymbol{\mu}_k = \boldsymbol{\Lambda}_k\boldsymbol{\alpha}_k + \boldsymbol{\tau}_k$$
$$\mathbf{P}_k = \boldsymbol{\Delta}_k(\boldsymbol{\Lambda}_k\boldsymbol{\Psi}_k\boldsymbol{\Lambda}_k^{\mathbf{t}} + \boldsymbol{\Theta}_k)\,\boldsymbol{\Delta}_k^{\mathbf{t}}$$

The latent underlying indicators are assumed to be multivariate normally distributed: **y*** ~ N($\boldsymbol{\mu}_k$, $\mathbf{P}_k$).

As they are latent, we have to impose some scale on the **y*** to arrive at identified models. To this end we shall assume that $\boldsymbol{\mu}_k$ is zero and $\mathbf{P}_k$ is a correlation matrix. So that we impose **y*** ~ N($\mathbf{0}_k$, $\mathbf{P}_k$), with diag($\mathbf{P}_k$) = diag(**I**) (unit variances). In fitting this model, the matrix $\boldsymbol{\Theta}_k$ cannot be considered to be free. In the case of a single factor model, we have $var(y_j^*) = \lambda_{j1}\,\psi\,\lambda_{j1} + \sigma^2_{ej} = 1$, where σ^2_{ej} is the j-th diagonal element of $\boldsymbol{\Theta}_k$.

Given appropriate scaling (ψ=1) we can estimate the factor loading. But given the factor loading we already know the value of σ^2_{ej}: σ^2_{ej} = 1 - $\lambda_{j1}\,\psi\,\lambda_{j1}$. More generally we have diag($\mathbf{\Theta}$) = diag($\mathbf{I}$) - diag($\mathbf{\Lambda\Psi\Lambda^t}$). This forms the basis of the <u>delta parameterization</u>. In a single group analysis, μ_k=**0**, Δ_k=**I** and $\mathbf{P}_k$ = $\mathbf{I}_k(\mathbf{\Lambda}_k\mathbf{\Psi}_k\mathbf{\Lambda}_k^{\mathbf{t}} + \mathbf{\Theta}_k)\,\mathbf{I}_k^{\mathbf{t}}$ is a correlation matrix.

Example output (4 three points scales)

```
                      Estimates     S.E.
 F1         BY
    V1                    0.681    0.043        Λ factor loadings
    V2                    0.716    0.045
    V3                    0.827    0.038
    V4                    0.661    0.046

 Means
    F1                    0.000    0.000        α factor mean

 Thresholds
    V1$1                 -0.885    0.065        t11 threshold 1,1
    V1$2                 -0.055    0.056        t12 threshold 1,2
    V2$1                 -0.885    0.065        t21  etc.
    V2$2                 -0.316    0.057        t22
    V3$1                 -0.601    0.060        t31
    V3$2                  0.238    0.057        t32
    V4$1                 -0.915    0.065        t41
    V4$2                 -0.337    0.057        t42

 Variances
    F1                    1.000    0.000        Ψ factor variance

 Scales
    V1                    1.000    0.000        diag(Δ) delta matrix
    V2                    1.000    0.000
    V3                    1.000    0.000
    V4                    1.000    0.000

R-SQUARE

    Observed   Residual
    Variable   Variance   R-Square
                                                residuals are not free parameters:
    V1            0.537      0.463              .537 = 1-.681*1*.681
    V2            0.488      0.512              .488 = 1-.716*1*.716
    V3            0.317      0.683              .317 = 1-.827*1*.827
    V4            0.563      0.437              .563 = 1-.563*1*.653
```

Note that the residual variances are given, but they are a function of the factor loadings. The chi2(2) for this model is 1.603.

An alternative method of treating the elements of $\mathbf{\Theta}$ is by fixing them to arbitrary values, say diag($\mathbf{\Theta}_k$) = [.5,.5,.5,.5]. This parameterization is called the <u>theta parameterization</u>. In a single group analysis again, μ_k = **0**, and $\mathbf{P}_k$ = $\mathbf{\Delta}_k(\mathbf{\Lambda}_k\mathbf{\Psi}_k\mathbf{\Lambda}_k^{\mathbf{t}} + \mathbf{\Theta}_k)\,\mathbf{\Delta}_k^{\mathbf{t}}$ is a correlation matrix. But now $\mathbf{\Delta}_k$ are included and are chosen to ensure that $\mathbf{P}_k$ is indeed a correlation matrix, i.e., diag($\mathbf{\Delta}_k$) = diag$(\mathbf{\Lambda}_k\mathbf{\Psi}_k\mathbf{\Lambda}_k^{\mathbf{t}} + \mathbf{\Theta}_k)^{-1/2}$.

example output

```
                   Estimates     S.E.
 F1        BY
    V1                 0.657    0.077
    V2                 0.725    0.094
    V3                 1.039    0.149
    V4                 0.623    0.077

 Means
    F1                 0.000    0.000

 Thresholds
    V1$1              -0.854    0.071
    V1$2              -0.053    0.054
    V2$1              -0.896    0.082
    V2$2              -0.320    0.061
    V3$1              -0.755    0.098
    V3$2               0.299    0.075
    V4$1              -0.863    0.071
    V4$2              -0.318    0.056

 Variances
    F1                 1.000    0.000

 Residual Variances
    V1                 0.500    0.000               fixed diagonal elements of Θ
    V2                 0.500    0.000
    V3                 0.500    0.000
    V4                 0.500    0.000


R-SQUARE

    Observed   Scale
    Variable   Factors  R-Square

    V1          1.036     0.463                     Δ (scale factors)
    V2          0.988     0.512
    V3          0.796     0.683
    V4          1.061     0.437
```

We can derive 1.036 as 1/sqrt(.657*1*.657+.5), or in matrix terms using R:

```
> ly=matrix(c(.657,.725,1.039,.623),4,1)
> te=diag(.5,4)
> s=ly%*%t(ly)+te
> delta=diag(1/(sqrt(diag(s))))
> diag(delta)
[1] 1.036 0.988 0.796 1.061
```

The chi2(2)=1.603, as expected given that these parameterizations produce equivalent results. Note that $\mathbf{P}_k = \mathbf{\Delta}_k\mathbf{\Lambda}_k\mathbf{\Psi}_k\mathbf{\Lambda}_k^{\mathbf{t}}\mathbf{\Delta}_k^{\mathbf{t}} + \mathbf{\Delta}_k\mathbf{\Theta}_k\mathbf{\Delta}_k^{\mathbf{t}}$ in the theta parameterization must equal $\mathbf{P}_k = \mathbf{\Lambda}_k\mathbf{\Psi}_k\mathbf{\Lambda}_k^{\mathbf{t}} + \mathbf{\Theta}_k$ in the delta parameterization, so (in R):

residual variances (delta%*%te%*%delta): $\mathbf{\Delta}_k\mathbf{\Theta}_k\mathbf{\Delta}_k^{\mathbf{t}}$

```
[1,] 0.5366828 0.0000000 0.0000000 0.0000000
[2,] 0.0000000 0.4875076 0.0000000 0.0000000
[3,] 0.0000000 0.0000000 0.3165517 0.0000000
```

```
[4,] 0.0000000 0.0000000 0.0000000 0.5629813
```

delta parameterization factor loadings (delta%*%ly): $\Delta_k\Lambda_k$

```
[1,] 0.6806741
[2,] 0.7158857
[3,] 0.8267093
[4,] 0.6610739
```

These results equal those obtained in the delta parameterization.

1.1: The two-group models: the six steps towards measurement invariance.
I briefly revisit the models that were discussed in lecture notes III. We again assume that the factor model holds for the latent continuous indicators **y***, and we specifically consider the single factor model[1].

$$\mu_k = \Lambda_k\alpha_k + \tau_k$$
$$\mathbf{P}_k = \Delta_k(\Lambda_k\Psi_k\Lambda_k^{\mathbf{t}} + \Theta_k)\Delta_k^{\mathbf{t}}$$

Step 1: In step one we merely estimated the polychoric correlations and the thresholds of the 4 items simultaneously in the two groups. We used the delta parameterization, so diag(Θ) =diag($\mathbf{I}$) - diag($\Lambda\Psi\Lambda^{\mathbf{t}}$). We obtained the thresholds ($\mathbf{t}_k$) and the polychoric correlations (Ψ_k) in each group. We consider these as simple summary statistics in each group.

Step 2: In step two we imposed the constraint that the thresholds are equal over the groups ($\mathbf{t}_1 = \mathbf{t}_2$). With this constraint, we can estimate the polychoric correlation matrix in one group, and the polychoric covariance matrix in the other group. In addition, we fixed the means of **y*** to zero in the first group, and estimated them freely in the second group. In this analysis we again used the delta parameterization. But while Δ_1 is fixed to an identity matrix in group 1, Δ_1 is freely estimated in group 2 (see Figure 1-1).

$$\mu_1 = \mathbf{0}$$
$$\mathbf{P}_1 = \Psi_2 \text{ (standardized)}$$

$$\mu_2 = \tau_2$$
$$\mathbf{P}_2 = \Delta_2\Psi_2\Delta_2^{\mathbf{t}}$$

This model fitted exactly as well as the step 1 model. This is however is specific to three point scale item. With fewer than 3 response categories this model is not identified, with more that 3 categories, the model represents a testable (df>0) proposition, and therefore may be rejected (in term of poor fit).

[1] Note that the generalization to more than two groups or more than two common factors should not pose any problems.

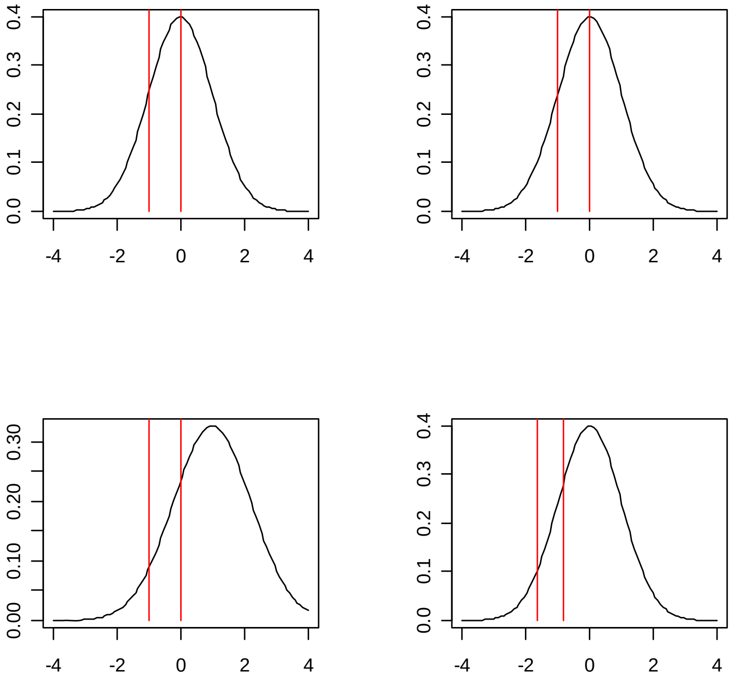


Figure 1.1. Top left and bottom left: same thresholds in two populations, but the populations differ with respect to the distribution of y*, the continuous indicators (m=0,s=1 in group 1; m=-1, s=1.22 in group 2). Top right and bottom right: y* standardized in both group (m=0, s=1). The thresholds in bottom right have changed to ensure that the response frequencies remain the same. That is, the probabilities of responses 0,1, and 2 are the same in bottom right and bottom left.

Step 3: In step three, we retain the equality of the thresholds, and fitted the otherwise unconstrained factor model within the groups. Note that we can convey the model as: $\mathbf{P}_1 = \mathbf{\Lambda}_1 \mathbf{\Psi}_1 \mathbf{\Lambda}_1^t + \mathbf{\Theta}_1$, and $\mathbf{\Sigma}_2 = \mathbf{\Lambda}_2 \mathbf{\Psi}_2 \mathbf{\Lambda}_2^t + \mathbf{\Theta}_2$, but using the delta parameterization we actually fit the model as follows:

$\mathbf{P}_1 = \mathbf{I}[\mathbf{\Lambda}_1 \mathbf{\Psi}_1 \mathbf{\Lambda}_1^t + \mathbf{\Theta}_1]\mathbf{I}$, $\quad \mathbf{\Psi}_1 = 1$ (standard scaling)

$\mathbf{P}_2 = \mathbf{\Delta}_2[\mathbf{\Lambda}_2 \mathbf{\Psi}_2 \mathbf{\Lambda}_2^t + \mathbf{\Theta}_2]\,\mathbf{\Delta}_2$, $\quad \mathbf{\Psi}_2 = 1$ (standard scaling)

where, in both groups $\mathbf{\Theta}_k = \text{diag}(\mathbf{I}) - \text{diag}(\mathbf{\Lambda}_k \mathbf{\Psi}_k \mathbf{\Lambda}_k^{\mathbf{t}})$. In this analysis the means in group 1 are $\boldsymbol{\mu}_1 = \mathbf{0}$ and in group 2, $\boldsymbol{\mu}_2 = \boldsymbol{\tau}_2$, as before.

Step 4: In step four, we retain the equality of the thresholds, and fitted the otherwise factor model subject to equal factor loadings. Using the delta parameterization we actually fit the model as follows:

$\mathbf{P}_1 = \mathbf{I}[\mathbf{\Lambda} \mathbf{\Psi}_1 \mathbf{\Lambda}^t + \mathbf{\Theta}_1]\mathbf{I}$, $\quad \mathbf{\Psi}_1 = 1$ (standard scaling)

$\mathbf{P}_2 = \mathbf{\Delta}_2[\mathbf{\Lambda} \mathbf{\Psi}_2 \mathbf{\Lambda}^t + \mathbf{\Theta}_2]\,\mathbf{\Delta}_2$, $\quad \mathbf{\Psi}_2$ free parameter where, in both groups $\mathbf{\Theta}_k$ = diag($\mathbf{I}$) - diag($\mathbf{\Lambda}_k\mathbf{\Psi}_k\mathbf{\Lambda}_k^{\mathbf{t}}$). In this analysis the means in group 1 are $\boldsymbol{\mu}_1 = \mathbf{0}$ and in group 2, $\boldsymbol{\mu}_2 = \boldsymbol{\tau}_2$, as before.

Step 5: In step five, we retain the equality of the thresholds, and fitted the otherwise factor model subject to equal factor loadings, and structured mean. Using the delta parameterization we actually fit the model as follows:

$\mathbf{P}_1 = \mathbf{I}[\mathbf{\Lambda}\,\mathbf{\Psi}_1\,\mathbf{\Lambda}^t + \mathbf{\Theta}_1]\mathbf{I}$, $\quad \mathbf{\Psi}_1 = 1$ (standard scaling)
$\mathbf{P}_2 = \mathbf{\Delta}_2[\mathbf{\Lambda}\mathbf{\Psi}_2\,\mathbf{\Lambda}^t + \mathbf{\Theta}_2]\,\mathbf{\Delta}_2$, $\quad \mathbf{\Psi}_2$ free parameter

$\boldsymbol{\mu}_1 = \mathbf{0}$
$\boldsymbol{\mu}_2 = \mathbf{\Lambda}\boldsymbol{\alpha}_2$,

where, in both groups $\mathbf{\Theta}_k$ = diag($\mathbf{I}$) - diag($\mathbf{\Lambda}_k\mathbf{\Psi}_k\mathbf{\Lambda}_k^{\mathbf{t}}$). In this analysis the means in group 1 are $\boldsymbol{\mu}_1 = \mathbf{0}$ and in group 2, $\boldsymbol{\mu}_2 = \mathbf{\Lambda}\boldsymbol{\alpha}_2$, as before. You may wonder why we do not fit $\boldsymbol{\mu}_2 = \boldsymbol{\tau} + \mathbf{\Lambda}\boldsymbol{\alpha}_2$, as in the continuous indicator case. This is not possible because of the factor that the continuous indicators are unobserved. That is, the parameters $\boldsymbol{\tau}$ are not identified.

Step 6: In step six we switched to the theta parameterization. We did this because the delta parameterization $\mathbf{\Theta}_k$ = diag($\mathbf{I}$) - diag($\mathbf{\Lambda}\mathbf{\Psi}_k\mathbf{\Lambda}^{\mathbf{t}}$) does not let itself to the imposition of equality constraints on $\mathbf{\Theta}_k$. This is because the matrix $\mathbf{\Lambda}\mathbf{\Psi}_k\mathbf{\Lambda}^{\mathbf{t}}$ is not necessarily equal over the groups ($\mathbf{\Psi}_1$=1, $\mathbf{\Psi}_2$ is freely estimated, and it is not likely that $\mathbf{\Psi}_1$=1). So we fix the parameters in $\mathbf{\Theta}_k$ to equal sensible value. We now fit the model:

$\mathbf{P}_1 = \mathbf{\Delta}_1[\mathbf{\Lambda}\,\mathbf{\Psi}_1\,\mathbf{\Lambda}^t + \mathbf{\Theta}]\mathbf{\Delta}_1$, $\quad \mathbf{\Psi}_1 = 1$ (standard scaling)
$\mathbf{P}_2 = \mathbf{\Delta}_2[\mathbf{\Lambda}\mathbf{\Psi}_2\,\mathbf{\Lambda}^t + \mathbf{\Theta}]\,\mathbf{\Delta}_2$, $\quad \mathbf{\Psi}_2$ free parameter

$\boldsymbol{\mu}_1 = \mathbf{0}$
$\boldsymbol{\mu}_2 = \mathbf{\Delta}_2\mathbf{\Lambda}\boldsymbol{\alpha}_2$,

In this final model the correlation matrices of **y*** in the two group, i.e., $\mathbf{P}_2$ and $\mathbf{P}_2$, differ only because of a difference in factor variance, $\mathbf{\Psi}_1$ vs. $\mathbf{\Psi}_2$. Similarly, $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ differ only as a function of the factor mean ($\boldsymbol{\alpha}_2$). The values in $\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$ differ, but the difference is a function of $\mathbf{\Psi}_1$ vs. $\mathbf{\Psi}_2$: diag($\mathbf{\Delta}_k$) = diag$(\mathbf{\Lambda}\mathbf{\Psi}_k\mathbf{\Lambda}^{\mathbf{t}} + \mathbf{\Theta})^{-1/2}$.

2.0 Measurement invariance in the multigroup ordinal factor model

In lecture notes I, we know that:

> *The distribution of the observed data conditional on group is given (i.e., multivariate normality). Within a given group k, we consider the conditional distribution of $\boldsymbol{y}_{ki}$ given $\eta_k=\eta^*$, $f(\boldsymbol{y}_{ki}|\,\eta^*)$:*
>
> $\boldsymbol{y}_{ki}|\,\eta^* \sim N(\tau_k + \Lambda_k\,\eta^*,\ \Theta_k)$,

So $f(\boldsymbol{y}_{ki}|\ \eta^*)$ is again a multivariate normal distribution, with the specific covariance matrix and mean vector. Specifically, the conditional means and covariance matrix within group k are:

$$E[\boldsymbol{y}_k|\boldsymbol{\eta}_{ki} = \boldsymbol{\eta}^*] = \tau_k + \Lambda_k\boldsymbol{\eta}^*, \text{ and } \Sigma_{k|\eta^*} = \Theta_k.$$

The definition of MI in the linear factor model requires the explicit conditioning on group:

Definition of MI: $f(\boldsymbol{y}_i|\ \eta^*) = f(\boldsymbol{y}_i|\ \eta^* \ \&\ group=k)$

In the case of the ordinal factor model, we can consider the same definition of measurement invariance. Consider a single continuous underlying item y^* and denote the fixed value of $\boldsymbol{\eta}^*$ as $\boldsymbol{\eta}^\bullet$ (to avoid notational mix-up):

$$[y_k^*|\ \eta^\bullet] \sim N(\tau_k + \lambda_k \eta^\bullet,\ \sigma_{ek}),$$

The probability that $y_k^*|\ \eta^\bullet$ is greater than or equal to some point p equals:

$$prob([y_k^* \geq p|\eta^\bullet]) = 1-\Phi(p-(\tau_k + \lambda_k \eta^\bullet)/\sigma_{ek})$$

Note that $b-(\tau_k + \lambda_k \eta^\bullet)/\sigma_{ek})$ is just a standardization to express the value b on the standard normal scale[2]. This allows us to evaluate the probability as $\Phi(z)$, i.e., using the cumulative standard normal distributions from $-\infty$ to z. So $1-\Phi(z)$ is the cumulative standard normal distribution from z to $+\infty$, i.e., $prob(z \geq b) = 1-\Phi(z)$.

Now consider the ordinal item y. For the single item in group k, the condition probability of choosing response category c or great, conditional on a fixed value of η, $\eta^\bullet$, equals:

$$\Phi(y_k \geq c|\eta^*) = 1-\Phi((t_{kc}-(\tau_k+\lambda_k\eta^*))/\sigma_{ek}^2) = 1-\Phi((t_{kc}-\lambda_k\eta^*)/\sigma_{ke}^2)$$

Note that $(t_{kc}-\lambda_k\eta^*)/\sigma_{ek}^2$ again expresses the threshold t_{kc} on the standard normal scale. The intercept τ_k is fixed to zero for reasons of identification. The intercept are not identified. Now clearly

$$F(y \geq c|\eta^*) = F(y \geq c|\eta^* \ \&\ group=k)$$

if and only if $t_{kc} = t_c$, $\sigma_{ek}^2 = \sigma_e^2$, and $\lambda_k = \lambda$. Hence returning to the multigroup model, we define the multigroup ordinal factor model subject to measurement invariance as one in which

$$\boldsymbol{\mu}_k = \boldsymbol{\Lambda\alpha}_k$$
$$\mathbf{P}_k = \boldsymbol{\Delta}_k(\boldsymbol{\Lambda\Psi}_k \boldsymbol{\Lambda}^{\mathbf{t}} + \boldsymbol{\Theta})\ \boldsymbol{\Delta}_k^{\mathbf{t}},$$

[2] That is if $y \sim N(m,s)$, and y is standardized $z=(m-y)/s$, then $z \sim N(0,1)$.

must hold.

3.0 Measurement invariance in other measurement models, say the latent profile model.

The definition of MI with respect to group (k=1...K) is:

Definition of MI: $f(\mathbf{y}_i | \boldsymbol{\eta}^*) = f(\mathbf{y}_i | \boldsymbol{\eta}^* \text{ \& group=k})$ eq 1-8.

for all values of $\boldsymbol{\eta}^*$ and all values of k. We have seen above that this definition can be applied readily in the linear factor model and in the ordinal factor model. It is important to realize that it applies equally well to any measurement model, i.e., any model in which observed indicators of a latent variable are related to the latent variable by means of a explicit function. In the linear factor model, this function is the linear regression function. For instance, consider the following model simple model. We assume that the latent variable is a nominal two class variable (depressed vs. not-depressed; addicted vs. not-addicted; liberal vs. conservative, etc.). The distribution of the latent variable is:

$\eta \sim \text{Bernoulli}(\theta)$, i.e., $\text{prob}(\eta=j) = \theta^{(1-j)} * (1-\theta)^j$,

where j=0,1. So $\text{prob}(\eta=0) = \theta^{(1-0)} * (1-\theta)^0 = \theta$, and $\text{prob}(\eta=1) = \theta^{(1-1)} * (1-\theta)^1 = (1-\theta)$. The latent variable is a dichotomy, i.e. a discrete (nominal) latent variable that can assume just two values (two latent classes). Now we assume that we have continuous indicators of the latent classes, **y**, that are distributed as follows:

$\mathbf{y}|\eta=j \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$,

where the conditional covariance matrix $\boldsymbol{\Sigma}_i$ is diagonal. This assumption can be viewed as the psychometric assumption of local independence: if you condition on the common underlying latent trait (common factor), then the observed item responses are uncorrelated (as we have already seen in the factor model $\mathbf{y}_{ki} | \boldsymbol{\eta}^* \sim N(\boldsymbol{\tau}_k + \boldsymbol{\Lambda}_k \boldsymbol{\eta}^*, \boldsymbol{\Theta}_k)$, where $\boldsymbol{\Theta}_k$ is diagonal).

This model is called a latent profile model. In its more general form the number of latent classes in not restricted to two. Here we consider two classes just to ease presentation. Now suppose that we want to establish measurement invariance of the indicators **y** with respect to, say, sex. We already have $f(\mathbf{y}| \boldsymbol{\eta}^*)$, namely defined as $\mathbf{y}|\eta=j \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$. We require that $f(\mathbf{y}| \boldsymbol{\eta}^*) = f(\mathbf{y}_i | \boldsymbol{\eta}^* \text{ \& group=k})$ so this implies that

$f(\mathbf{y}|\eta=j \text{ \& group=male}) = f(\mathbf{y}|\eta=j \text{ \& group=female})$,

or simply that the conditional distributions be equal over sex. $\mathbf{y}|\eta=j \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ must hold in the male and female sample. Note that this does not mean that the sizes of the latent classes should be equal. That is, the latent distribution may differ over the groups (i.e., sex):

$\eta_k \sim \text{Bernoulli}(\theta_k)$, $\text{prob}(\eta_k=j) = \theta_k^{(1-j)} * (1-\theta_k)^j$, where k denotes group (sex). Of course this is no different from the observation that subject to measurement invariance with respect to groups, the common factor distribution may differ over group.

The latent profile model is a model in which the latent variable of interest is discrete (nominal), and the observed indicators are continuous. As such it fits in our taxonomy of measurement models.

Taxonomy of psychometric models.

	Latent variable / trait / common factor		
observed indicators		discrete	continuous
	discrete	latent class model	IRT: Rasch, Birmbaum, Discrete factor model
	continuous	latent profile model	linear factor model

Generally speaking, in each of the models in this taxonomy, the define 1) a distribution function of the latent variable; 2) a function relating the observed indicators to the latent variable; 3) the distribution of the observed indicators given a fixed value on the latent variable. The constraints associated with measurement invariance with respect to a given variable x, pertain only to the conditional distribution of the indicators given the latent variables. Measurement invariance implies that the parameters of this conditional distribution be invariance for all values of the latent variable and for all values of the variable x (with respect to which measurement invariance is defined).