Confirmatory Factor Analysis
Structural Equation Models

Recall again:

The *Orthogonal* Factor Analysis Model

$$x_1 = \mu_1 + \lambda_{11}F_1 + \lambda_{12}F_2 + \ldots + \lambda_{1m} F_m + \varepsilon_1$$
$$x_2 = \mu_2 + \lambda_{21}F_1 + \lambda_{22}F_2 + \ldots + \lambda_{2m} F_m + \varepsilon_2$$
$$\ldots$$
$$x_p = \mu_p + \lambda_{p1}F_1 + \lambda_{p2}F_2 + \ldots + \lambda_{pm} F_m + \varepsilon_p$$

or

$$\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{LF} + \boldsymbol{\varepsilon}$$

which implies

$$\boldsymbol{\Sigma} = \mathrm{cov}(\boldsymbol{X}) = \boldsymbol{LL}' + \boldsymbol{\Psi}$$

where

$F_1, F_2, \ldots, F_m$ are called the **common factors**

with covariance matrix $\mathrm{cov}(\boldsymbol{F}) = \boldsymbol{I}$

We now change the assumption that the factors are *uncorrelated* with the following:

$$\operatorname{cov}(\boldsymbol{F}) = \boldsymbol{\Phi} = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1m} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m1} & \phi_{m2} & \cdots & \phi_{mm} \end{bmatrix}$$

Then the covariance of the data becomes

$$\boldsymbol{\Sigma} = \operatorname{cov}(\boldsymbol{X}) = \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{L}' + \boldsymbol{\Psi}$$

Path Diagrams

The above equation is just one way to specify the model we want to fit to the data. Alternatively, we may use *path diagrams* to specify the model graphically.

In a path diagram, observed variables are denoted with rectangles and latent (unobservable) variables (factors) are denoted with ovals.

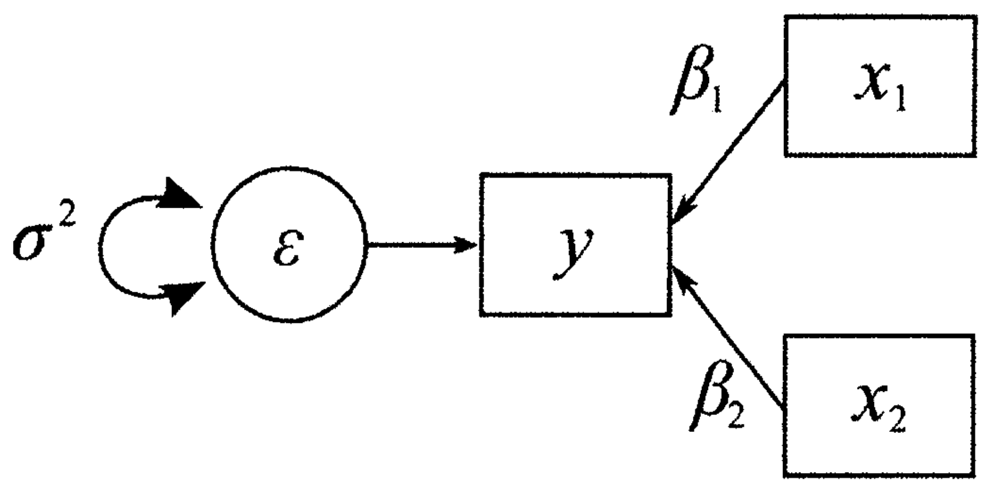
The ovals and rectangles are connected with arrows. Note that a one-headed arrow denotes a structural or unidirectional relationship, while a two-headed arrow denotes a covariance among variables. A two-headed arrow pointing from one variable and back to itself is used to specify the variance of the variable, which is usually an error or "disturbance" variable.

Example:

Consider the two-predictor regression model

$$y = \boldsymbol{\beta}_1 x_1 + \boldsymbol{\beta}_2 x_2 + \boldsymbol{\varepsilon},\ \mathrm{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\sigma}^2$$

It can also be represented with a path diagram as follows:



Example 14.2.2 from Rencher

Table 14.1, which were obtained from a university business statistics class. As shown in Table 14.1, each of the 94 students in the class received scores for laboratory assignments (Lab), homework assignments (HW), pop quizzes (PopQuiz), midterm exam #1 (Exam1), midterm exam #2 (Exam2), and the final exam (FinalExam).

The instructor might hypothesize that the 6-dimensional measure of performance in the class is being driven by an underlying 2-dimensional factor process, with the first factor associated with daily effort and the second factor associated with knowledge mastery.

A possible model is:

$$
\begin{aligned}
\text{Lab} &= \boldsymbol{\mu}_1 + \boldsymbol{\lambda}_{11} f_1 + \boldsymbol{\lambda}_{12} f_2 + \boldsymbol{\varepsilon}_1 \\
\text{HW} &= \boldsymbol{\mu}_2 + f_1 + \boldsymbol{\varepsilon}_2 \\
\text{PopQuiz} &= \boldsymbol{\mu}_3 + \boldsymbol{\lambda}_{31} f_1 + \boldsymbol{\lambda}_{32} f_2 + \boldsymbol{\varepsilon}_3 \\
\text{Exam1} &= \boldsymbol{\mu}_4 + \boldsymbol{\lambda}_{41} f_1 + \boldsymbol{\lambda}_{42} f_2 + \boldsymbol{\varepsilon}_4 \\
\text{Exam2} &= \boldsymbol{\mu}_5 + \boldsymbol{\lambda}_{51} f_1 + \boldsymbol{\lambda}_{52} f_2 + \boldsymbol{\varepsilon}_5 \\
\text{Final Exam} &= \boldsymbol{\mu}_6 + f_2 + \boldsymbol{\varepsilon}_6
\end{aligned}
$$

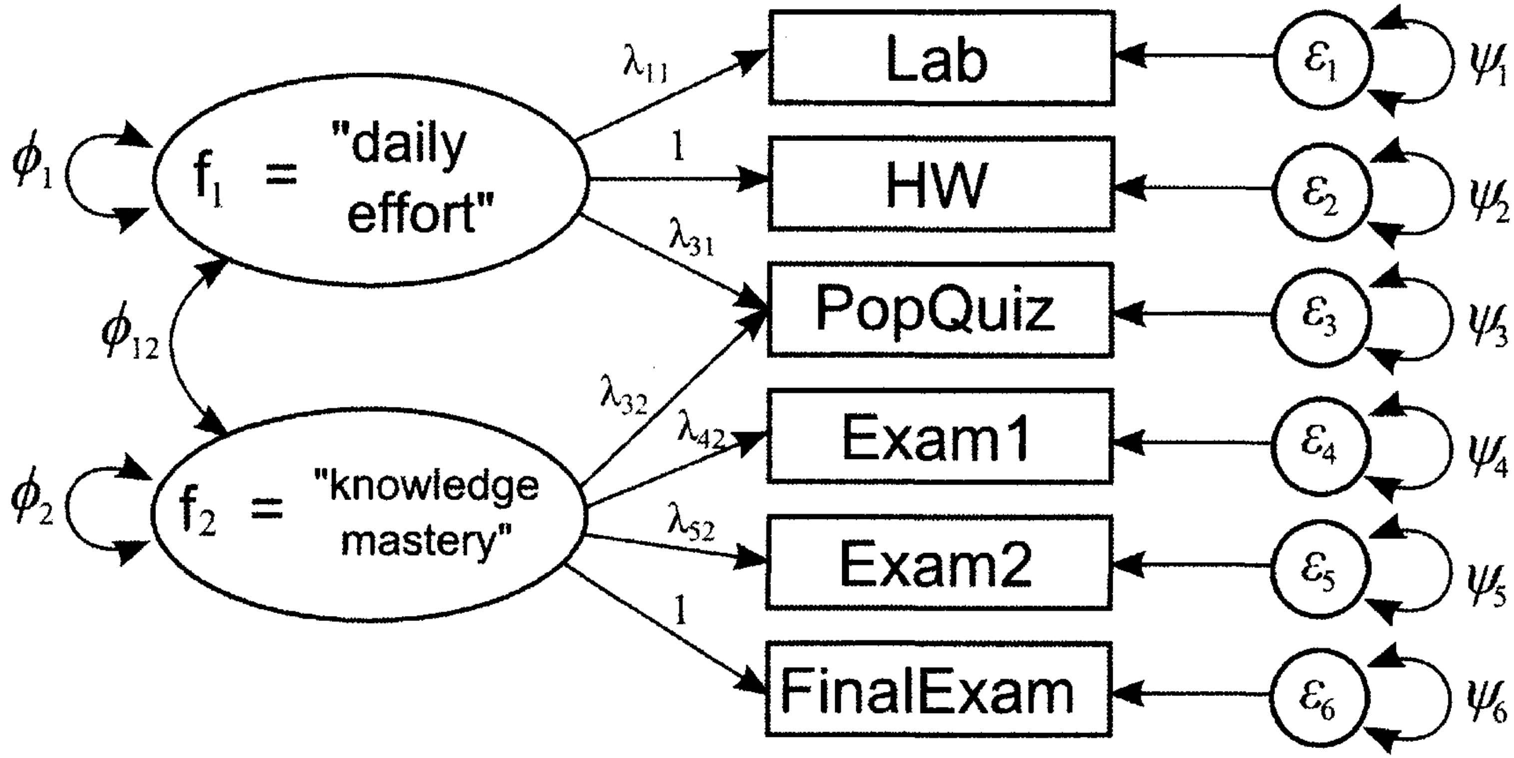
Note: Some of the loadings are assumed to be 1 because of model identifiability issues.

Example (continued):

We will further simplify the model to test the theory that variables mostly load on one of the factors:

$$
\begin{aligned}
\text{Lab} &= \boldsymbol{\mu}_1 + \boldsymbol{\lambda}_{11} f_1 && + \boldsymbol{\varepsilon}_1 \\
\text{HW} &= \boldsymbol{\mu}_2 + \quad f_1 && + \boldsymbol{\varepsilon}_2 \\
\text{PopQuiz} &= \boldsymbol{\mu}_3 + \boldsymbol{\lambda}_{31} f_1 + \boldsymbol{\lambda}_{32} f_2 && + \boldsymbol{\varepsilon}_3 \\
\text{Exam1} &= \boldsymbol{\mu}_4 \qquad\qquad + \boldsymbol{\lambda}_{42} f_2 && + \boldsymbol{\varepsilon}_4 \\
\text{Exam2} &= \boldsymbol{\mu}_5 \qquad\qquad + \boldsymbol{\lambda}_{52} f_2 && + \boldsymbol{\varepsilon}_5 \\
\text{Final Exam} &= \boldsymbol{\mu}_6 \qquad\qquad\quad + f_2 && + \boldsymbol{\varepsilon}_6
\end{aligned}
$$

Example : *Path diagram*



Results from function `sem`

```
             Estimate    Std Error   z value    Pr(>|z|)
lambda11   0.34084083  0.08602003  3.9623428 7.421788e-05 Lab <--- Daily
lambda31   0.12187596  0.09608828  1.2683749 2.046641e-01 PopQuiz <--- Daily
lambda32   0.06069461  0.01736943  3.4943351 4.752440e-04 PopQuiz <--- Knowledge
lambda4    0.88473142  0.12679088  6.9778790 2.996696e-12 Exam1 <--- Knowledge
lambda5    1.41832704  0.19759221  7.1780514 7.071196e-13 Exam2 <--- Knowledge
psi1       0.56426902  0.11775530  4.7918779 1.652275e-06 Lab <--> Lab
psi2      -0.14096627  0.72146127 -0.1953899 8.450877e-01 HW <--> HW
psi3       1.66531161  0.25841865  6.4442393 1.161816e-10 PopQuiz <--> PopQuiz
psi4      66.97340473 13.75491135  4.8690539 1.121338e-06 Exam1 <--> Exam1
psi5     139.20042167 31.61349774  4.4031958 1.066678e-05 Exam2 <--> Exam2
psi6      81.65480010 17.11273586  4.7715807 1.827857e-06 FinalExam <--> FinalExam
phi11      3.30315491  0.85714254  3.8536822 1.163546e-04 Daily <--> Daily
phi12     10.21054128  2.63435835  3.8759120 1.062261e-04 Knowledge <--> Daily
phi22    123.71525855 30.21755333  4.0941521 4.237162e-05 Knowledge <--> Knowledge
```

Example p. 206 of Everitt & Hothorn

Calsyn and Kenny (1977) recorded the values of the following six variables for 556 white eighth-grade students:

SCA: self-concept of ability;

PPE: perceived parental evaluation;

PTE: perceived teacher evaluation;

PFE: perceived friend's evaluation;

EA: educational aspiration;

CP: college plans.

They postulated that two underlying latent variables, *ability* and *aspiration*, generated the relationships between the observed variables.

Proposed model is:

$$\text{SCA} = \boldsymbol{\mu}_1 + \boldsymbol{\lambda}_{11} f_1 + 0 f_2 + \boldsymbol{\varepsilon}_1$$

$$\text{PPE} = \boldsymbol{\mu}_2 + \boldsymbol{\lambda}_{21} f_1 + 0 f_2 + \boldsymbol{\varepsilon}_2$$

$$\text{PTE} = \boldsymbol{\mu}_3 + \boldsymbol{\lambda}_{31} f_1 + 0 f_2 + \boldsymbol{\varepsilon}_3$$

$$\text{PFE} = \boldsymbol{\mu}_4 + \boldsymbol{\lambda}_{41} f_1 + 0 f_2 + \boldsymbol{\varepsilon}_4$$

$$\text{AE} = \boldsymbol{\mu}_5 + 0 f_1 + \boldsymbol{\lambda}_{52} f_2 + \boldsymbol{\varepsilon}_5$$

$$\text{CP} = \boldsymbol{\mu}_6 + 0 f_1 + \boldsymbol{\lambda}_{62} f_2 + \boldsymbol{\varepsilon}_6$$

Note: The model has a total of 13 parameters to estimate, six factor loadings, six specific variances ,and one correlation between ability and aspiration.

.

Example : *Path diagram with estimates*