Slide 1

Linear Factor Analysis - The Model

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \boldsymbol{\epsilon} \tag{1}$$

$\mathbf{x}$: p-dimensional vector of continuous observed variables

$\mathbf{f}$: q-dimensional vector of underlying latent factors. Often called "common factors". Assume $\mathbf{f}$ is random such that $E(\mathbf{f}) = \mathbf{0}$ and $Var(\mathbf{f}) = \boldsymbol{\Phi}$

$\boldsymbol{\epsilon}$: p-dimensional vector of random error. Often called "unique factors" or "specific factors". Assume $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $Var\boldsymbol{\epsilon} = \boldsymbol{\Psi}$. Element along the diagonal f $\boldsymbol{\Psi}$ often called "uniquenesses" or "specific variances"

$\boldsymbol{\Lambda}$: $p \times q$ matrix of scalars called "factor loadings". This matrix describes how the observed variables $\mathbf{x}$ are related to the latent factors $\mathbf{f}$.

$\boldsymbol{\mu}$: $p \times 1$ vector of scalars. Often ignored, most software assumes by default that $\boldsymbol{\mu} = \mathbf{0}$ and analyze centered $\mathbf{x}$ variables, i.e. analyze $\mathbf{x} - \bar{\mathbf{x}}$

Slide 2

Linear Factor Analysis - The Model

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \boldsymbol{\epsilon}$$

An assumption is made for model (1) that $Cov(\mathbf{f}, \boldsymbol{\epsilon}) = \mathbf{0}$. This is a very critical assumption because what it means is that the variability in $\mathbf{x}$ can be separated into two additive parts, one coming from the **common factors** and one coming from **specific factors** or the errors. That is there is no covariance between the two. Specifically,

$$\begin{aligned} Var(\mathbf{x}) &= \boldsymbol{\Lambda} Var(\mathbf{f})\boldsymbol{\Lambda}' + Var(\boldsymbol{\epsilon}) + 2Cov(\mathbf{f}, \boldsymbol{\epsilon}) \\ Var(\mathbf{x}) &= \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}. \end{aligned}$$

It is assumed that equation (1) holds for each individual in the population and thus for $i = 1 \ldots n$ independently sampled individuals we have

$$\mathbf{x}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f}_i + \boldsymbol{\epsilon}_i.$$

Slide 3

As an example of how to write model (1) , here are the equations when $p = 5$ and $q = 2$ and it is assumed that $\mathbf{\Psi}$ is a diagonal matrix, for $i = 1 \ldots n$ we have:

$$\begin{aligned} x_{1i} &= \mu_1 + \lambda_{11} f_{1i} + \lambda_{12} f_{2i} + \epsilon_{1i} \\ x_{2i} &= \mu_2 + \lambda_{21} f_{1i} + \lambda_{22} f_{2i} + \epsilon_{2i} \\ x_{3i} &= \mu_3 + \lambda_{31} f_{1i} + \lambda_{32} f_{2i} + \epsilon_{3i} \\ x_{4i} &= \mu_4 + \lambda_{41} f_{1i} + \lambda_{42} f_{2i} + \epsilon_{4i} \\ x_{5i} &= \mu_5 + \lambda_{51} f_{1i} + \lambda_{52} f_{2i} + \epsilon_{5i} \end{aligned}$$

or in matrix/vector notation

$$\begin{pmatrix} x_{1i} \\ x_{2i} \\ x_{3i} \\ x_{4i} \\ x_{5i} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \end{pmatrix} + \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \\ \lambda_{31} & \lambda_{32} \\ \lambda_{41} & \lambda_{42} \\ \lambda_{51} & \lambda_{52} \end{pmatrix} \begin{pmatrix} f_{1i} \\ f_{2i} \end{pmatrix} + \begin{pmatrix} \epsilon 1i \\ \epsilon 2i \\ \epsilon 3i \\ \epsilon 4i \\ \epsilon 5i \end{pmatrix}$$

and we also must specify the variance structure of $\mathbf{f}$ and $\boldsymbol{\epsilon}$

$$Var(\mathbf{f}_i) = Var \begin{pmatrix} f_{1i} \\ f_{2i} \end{pmatrix} = \mathbf{\Phi} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{pmatrix}$$

$$Var(\epsilon_i) = Var \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \epsilon_{3i} \\ \epsilon_{4i} \\ \epsilon_{5i} \end{pmatrix} = \mathbf{\Psi} = \begin{pmatrix} \psi_1 & 0 & 0 & 0 & 0 \\ 0 & \psi_2 & 0 & 0 & 0 \\ 0 & 0 & \psi_3 & 0 & 0 \\ 0 & 0 & 0 & \psi_4 & 0 \\ 0 & 0 & 0 & 0 & \psi_5 \end{pmatrix}$$

Slide 4

Communalities

- Communalities are the diagonal elements of $\mathbf{\Lambda\Lambda}'$ when the factors are assumed to be uncorrelated
- Communalities are the part of the variance of each observed variable which is due to the q underlying factors

$$\begin{aligned} x_1 &= \mu_1 + \lambda_{11} f_1 + \lambda_{12} f_2 + \epsilon_1 \\ x_2 &= \mu_2 + \lambda_{21} f_1 + \lambda_{22} f_2 + \epsilon_2 \\ x_3 &= \mu_3 + \lambda_{31} f_1 + \lambda_{32} f_2 + \epsilon_3 \\ x_4 &= \mu_4 + \lambda_{41} f_1 + \lambda_{42} f_2 + \epsilon_4 \\ x_5 &= \mu_5 + \lambda_{51} f_1 + \lambda_{52} f_2 + \epsilon_5 \end{aligned}$$

$$\begin{aligned} Var(x_1) &= \lambda_{11}^2 + \lambda_{12}^2 + \psi_1 \\ Var(x_2) &= \lambda_{21}^2 + \lambda_{22}^2 + \psi_2 \\ Var(x_3) &= \lambda_{31}^2 + \lambda_{32}^2 + \psi_3 \\ Var(x_4) &= \lambda_{41}^2 + \lambda_{42}^2 + \psi_4 \\ Var(x_5) &= \lambda_{51}^2 + \lambda_{52}^2 + \psi_5 \end{aligned}$$

The communality for x_2 is $\lambda_{21}^2 + \lambda_{22}^2$. On the other hand, ψ_2 is the part of the variability in x_2 that is unique to x_2, i.e. is not shared with other observed variables, "is not common".

Slide 5

Exploratory factor analysis vs. Confirmatory factor analysis (EFA vs. CFA)

EFA general purposes:

- To determine how many underlying factors are necessary to explain most of the correlations and variance in the data.
- To determine the relationship via **rotation** between each of these underlying factors with each of the observed variables in a meaningful way so that the factors can be interpreted and named.
- To weed out observed variables that do not tend to measure well the underlying factors shared by the other variables.
- To propose blocks of variables that may be subsequently be used to create a simple sum scale.
- To propose a CFA model

In EFA every element in $\mathbf{\Lambda}$ is estimated and it is assumed that $\mathbf{\Psi}$ is diagonal. Also, it is common to assume that $Var(\mathbf{f}) = \mathbf{\Phi} = \mathbf{I}$, i.e. the factors are uncorrelated with variance 1 (but this is not a necessary assumption, it is dropped when examining oblique rotations).

Slide 6

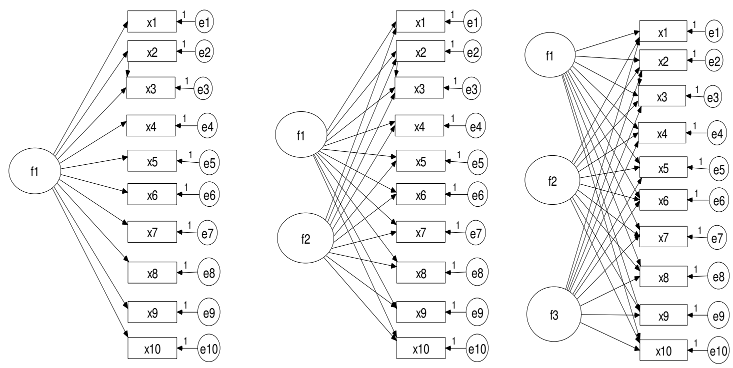
Exploratory factor analysis vs. Confirmatory factor analysis (EFA vs. CFA)

CFA general purposes:

- To define a measurement model for the relationship between multivariate observations and underlying factors
- To test the statistical significance of factor loadings and correlations. Note this testing cannot currently be done in the EFA model. Thus one may be interested in testing whether rotated factor loadings from an EFA that look "close to zero" are, in fact, significantly different from zero or not.
- To test whether the measurement model for one group is the same as the measurement model for some other group
- As a precursor to a Structural equation model

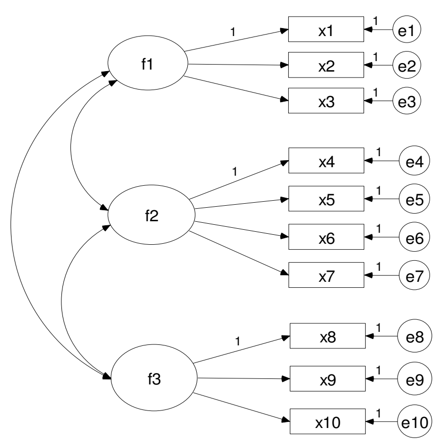
In CFA usually several elements in $\mathbf{\Lambda}$ are fixed to zero and it is possible to consider correlated $\boldsymbol{\epsilon}$ which means that $\mathbf{\Psi}$ is not necessarily diagonal. Furthermore, it is usually assumed that the factors are correlated so that no restriction is placed on $\mathbf{\Phi}$.

Slide 7



In EFA every element in $\mathbf{\Lambda}$ is estimated and it is assumed that $\mathbf{\Psi}$ is diagonal. Also, it is common to assume that $Var(\mathbf{f}) = \mathbf{\Phi} = \mathbf{I}$, i.e. the factors are uncorrelated with variance 1 (but this is not a necessary assumption, it is dropped when examining oblique rotations).

Slide 8



In CFA usually several elements in $\mathbf{\Lambda}$ are fixed to zero and it is possible to consider correlated $\boldsymbol{\epsilon}$ which means that $\mathbf{\Psi}$ is not necessarily diagonal. Furthermore, it is usually assumed that the factors are correlated so that no restriction is placed on $\mathbf{\Phi}$.

Slide 9

Exploratory factor analysis vs. Confirmatory factor analysis (EFA vs. CFA)

What are the differences in the two models?

In the EFA model, there are arrows going from every factor to every observed variable. Also there are no double headed curved arrows going from factor to factor.

In the CFA model, there are clearly a lot less arrows going from factors to observed variable and in particular each observed variable has exactly 1 arrow hitting it - this is called simple structure. Also there are curved double headed arrows going from each factor to each other factor.

Look at papers:

- Hurley, A., Scandura, T, Scriesheim, C. Brannick, M., Seers, A, Vandenberg, R., and Williams, L. (1997) "Exploratory and confirmatory factor analysis: guidelines, issues, and alternatives" *Journal of organizational behavior*, 18, 667-683.
- Gerbing DW and Hamilton JG "Viability of Exploratory factor analysis as a precursor to confirmatory factor analysis", *Structural equation modeling*, 3(1), 62-72.

Slide 10

Outline of Exploratory Factor Analysis

- Factor extraction method
  - principal component analysis
    Starts with correlation matrix and extracts factors (by calculating eigenvectors and eigenvalues) from it directly.
  - common factor analysis (Maximum Likelihood, Least squares, or Principal factor method)
    Starts with correlation matrix but then makes a "reduced correlation matrix" meant to represent the correlation of the "common part", then extracts factors (by calculating eigenvectors and eigenvalues or using a discrepancy function) from the "reduced correlation matrix".
- Choose the number of factors
  - examine eigenvalues: % variance explained, eigenvalue > 1, scree plot
  - Model fit statistics: Chi-square value, RMSR root mean square residuals for correlation matrix, RMSEA (Cudeck) function of chi-square taking into account sample size, $< .05$ good fit, $< .08$ ok fit

Slide 11

Outline of Exploratory Factor Analysis - Continued

- Interpreting the factors - Rotation
  - rotate to simple structure - orthogonal (varimax) or oblique (promax) correlated factors
  - examining size of loadings (stat sig?), cross loadings
- Determining quality of factors and items
  - subject matter interpretation
  - number of items loading on the factor
  - variance explained by the factor
  - eliminate poor items and factors and repeat EFA
- Possible next steps
  - Verify structure using confirmatory factor analysis
  - Calculate factor score estimates $E(f|X)$, Mplus produces a "factor determinacy" value which represents correlation between factor and the estimated factor score
  - Calculate simple sum scores of items associated with each f with simple structure, use Cronbach alpha as estimate of reliability

Slide 12

Maximum number of factors possible to extract

- When the PCA method is used, it is possible to extract as many factors as there are variables.
- For the common factor methods which assume that the factors explain only the systematic part of the data (not the specific errors), the number of factors is constrained. The following must be true $\frac{p(p+1)}{2} - pq - p + \frac{q(q-1)}{2} \geq 0$ So for example when $p = 3$ or $p = 4$, q must be no greater than 1. When $p = 20$, q can be up to 14.

This restriction is related to the degrees of freedom associated with the common factor model. The number of parameters being estimated must be less than or equal to the amount of unconstrained data. For the EFA model the d.f. are $\frac{p(p+1)}{2} - pq - p + \frac{q(q-1)}{2}$ The first term is the number of unique elements in the sample covariance matrix $\mathbf{S}$, then we are going to estimate pq factor loadings in $\mathbf{\Lambda}$ and p variances in $\mathbf{\Psi}$ and since we have to put restrictions in order to fix rotation we get those $\frac{q(q-1)}{2}$ d.f. back. These $\frac{q(q-1)}{2}$ d.f. correspond to fixing the $\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda}$ off diagonal elements being fixed to zero.

Slide 13

Rotation in EFA - this is an issue related to model identification

No matter what estimation procedure, for exploratory factor analysis we get estimates that look like:

$$\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Lambda}}\hat{\mathbf{\Lambda}}' + \hat{\mathbf{\Psi}}$$

We can get the exact same $\hat{\mathbf{\Sigma}}$ by taking

$$\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Lambda}}TT'\hat{\mathbf{\Lambda}}' + \hat{\mathbf{\Psi}}$$

where T is a $q \times q$ matrix such that $TT' =$ Identity matrix.

Thus

$$\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Lambda}}^*\hat{\mathbf{\Lambda}}^{*\prime} + \hat{\mathbf{\Psi}}$$

where $\hat{\mathbf{\Lambda}}^*$ is the **orthogonally rotated** factor loading matrix, i.e. $\hat{\mathbf{\Lambda}}^* = \hat{\mathbf{\Lambda}}T$. BUT, There are an infinite number of matrices T that satisfy $TT' =$ I.

Slide 14

Rotation

- Why do we bother to rotate? – We want to find out which variables "stick together"
- How to choose a T? – Find one that gives close to "simple" structure to the factor loadings
- "Simple structure" means that each observed variable only loads on one factor. In an AMOS notation, each observed variable only has one arrow going to it from a factor. Simple structure allows you to say, for example, "observed variable 1 is a direct measure of only factor 1", instead of "variable 1 is direct measure of factor 1, 2, and 3"
- What about if simple structure is not found by orthogonal rotation?
- NOTE: implicitly orthogonal rotation keeps the assumption that $Var(\mathbf{f}) = I$, i.e. the factors are uncorrelated.
- Maybe the assumption that $Var(\mathbf{f}) = I$ should be dropped. This leads to **oblique rotations** of the factor loadings.
- We can search for
$$\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Lambda}^{**}}\mathbf{\Phi}\hat{\mathbf{\Lambda}}^{**\prime} + \hat{\mathbf{\Psi}}$$
so that $\hat{\mathbf{\Lambda}^{**}}$ has "simple structure" and $\mathbf{\Phi} \neq I$

Slide 15

Matching items to factors - Examine the rotated loadings

Throughout it is necessary to keep in mind whether the items loading on a given factor share some conceptual meaning.

- What is a BIG loading?
  - Common rule of thumb, if the absolute value of the standardized loading is $> .3$, the variable is relevant for the particular factor.
  - In 1994 Cudeck, R. and O'Dell, L. "Applications of standard error estimates in unrestricted factor analysis: Significance Tests for factor loadings and correlations" *Psychological Bulletin* argue that the $> .3$ cutoff is too simplistic and encourage sig testing.
  - Now possible to get standard errors from SAS Proc factor when using method = ml and the se option.
  - Additional work on calculated standard errors is Ogasawara (1998) "Standard errors for rotation matrices with an application to the promax solution". And there is a specialized program available from Michael Browne's website (at Ohio State University) called CEFA that can calculate these standard errors. All of these results are based on asymptotic results, thus it is still not clear how useful they are for problems with small sample sizes.

Slide 16

Matching items to factors - Continued

- Examine the communality. Rule of thumb, if the communality is less than .10 then the observed variable may be deleted. A communality $< .10$ means that less than 10% of the variability in the observed variable is explained by all of the common factors.
- Items with strong loadings on multiple factors (4 options)
  - allow the item to load on more than one factor
  - choose the item to load on the factor it has the largest loading with
  - choose the item to load on the factor it conceptually goes best with
  - drop the item
- Cross Validation. If the sample size is large enough, an intuitive way to gain an idea about the stability of the factor structure is to split the sample randomly into two equal parts and then fit the model to each part. If the factor structure is similar, this tends to increase our confidence in the genuineness of the factors.

Variance explained by factors

In EFA the sum of the total variance explained by k factors is same for any orthogonal rotation of the k factors. The difference is how the variance is spread out across the factors.

On the other hand, when an oblique rotation is used, summarizing the variance explained by the factors is tricky because we need to talk about the variance explained after adjusting for the other factors (since they are correlated).

Slide 17

School Subjects example from BSMG

Possible to read just the correlation matrix into SAS for analysis.

```
data a (type = corr);
input _type_ $ _name_ $ gaelic english history math algebra geometry;
cards;
mean . 0   0   0   0   0   0
std  . 1   1   1   1   1   1
n    . 220 220 220 220 220 220
corr gaelic       1 .44 .41 .29 .33 .25
corr english    .44   1 .35 .35 .32 .33
corr history    .41 .35   1 .16 .19 .18
corr math       .29 .35 .16   1 .59 .47
corr algebra    .33 .32 .19 .59   1 .46
corr geometry   .25 .33 .18 .47 .46  1 ;
run;
proc contents data = a; run;

proc factor data = a;
var gaelic english history math algebra geometry;
run;
```

options include method = prinit, method = uls, method = ml. By default with no method stated, SAS performs principal component analysis. Other options available are regarding the rotation method which can be used with all of the estimation methods. They are rotate=varimax and rotate=promax

Slide 18

Slide 19

School Subjects example from BSMG - PCA output

proc factor data = a rotate = promax;
var gaelic english history math algebra geometry; run;

```
The FACTOR Procedure
                               Initial Factor Method: Principal Components
                                   Prior Communality Estimates: ONE
                       Eigenvalues of the Correlation Matrix: Total = 6  Average = 1

                               Eigenvalue    Difference    Proportion    Cumulative

                          1    2.72868347    1.59989129        0.4548        0.4548
                          2    1.12879218    0.51350073        0.1881        0.6429
                          3    0.61529145    0.01248259        0.1025        0.7455
                          4    0.60280886    0.08029449        0.1005        0.8459
                          5    0.52251437    0.12060470        0.0871        0.9330
                          6    0.40190967                      0.0670        1.0000

                          2 factors will be retained by the MINEIGEN criterion.
                                          Factor Pattern

                                              Factor1         Factor2

                                gaelic          0.66080         0.44447
                                english         0.68846         0.28977
                                history         0.51636         0.63955
                                math            0.73562        -0.41702
                                algebra         0.74187        -0.37276
                                geometry        0.67817        -0.35410
                                    Variance Explained by Each Factor

                                       Factor1         Factor2
                                     2.7286835       1.1287922
                               Final Communality Estimates: Total = 3.857476

            gaelic         english         history            math         algebra        geometry
        0.63421803      0.55795080      0.67565080      0.71504052      0.68931666      0.58529884
```

Slide 20

School Subjects example from BSMG - PCA OUTPUT

```
                                        Prerotation Method: Varimax

                                      Orthogonal Transformation Matrix
                                                      1               2

                                        1         0.77371         0.63354
                                        2        -0.63354         0.77371

                                           Rotated Factor Pattern

                                              Factor1         Factor2

                                gaelic          0.22967         0.76254
                                english         0.34909         0.66037
                                history        -0.00567         0.82196
                                math            0.83335         0.14340
                                algebra         0.81015         0.18160
                                geometry        0.74904         0.15568

                                    Variance Explained by Each Factor
                                       Factor1         Factor2
                                     2.0865260       1.7709497
                               Final Communality Estimates: Total = 3.857476

            gaelic         english         history            math         algebra        geometry
        0.63421803      0.55795080      0.67565080      0.71504052      0.68931666      0.58529884
```

Slide 21

School Subjects example from BSMG - PCA OUTPUT

Rotation Method: Promax (power = 3)

Inter-Factor Correlations

	Factor1	Factor2
Factor1	1.00000	0.36730
Factor2	0.36730	1.00000

Rotated Factor Pattern (Standardized Regression Coefficients)

	Factor1	Factor2
gaelic	0.09286	0.75757
english	0.23917	0.62522
history	-0.16759	0.86862
math	0.85268	-0.01983
algebra	0.82064	0.02525
geometry	0.76115	0.01045

Variance Explained by Each Factor Eliminating Other Factors

Factor1	Factor2
1.7940097	1.4883579

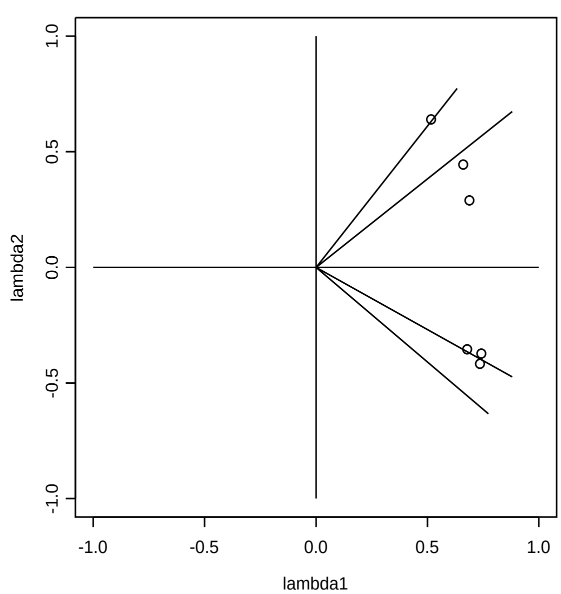
Variance Explained by Each Factor Ignoring Other Factors

Factor1	Factor2
2.3691177	2.0634659

Final Communality Estimates: Total = 3.857476

gaelic	english	history	math	algebra	geometry
0.63421803	0.55795080	0.67565080	0.71504052	0.68931666	0.58529884

Slide 22



Slide 23

School Subjects example from BSMG - ML OUTPUT

proc factor data = a **method = ml** rotate = promax;
var gaelic english history math algebra geometry; run;

```
The FACTOR Procedure
                                 Initial Factor Method: Maximum Likelihood

                                    Prior Community Estimates: SMC

            gaelic          english         history           math          algebra         geometry

        0.30168715       0.29583987      0.20564352      0.41522406      0.41174029      0.29450749


                      Preliminary Eigenvalues: Total = 2.93847068  Average = 0.48974511

                                Eigenvalue    Difference    Proportion    Cumulative

                           1    3.17047301    2.54848714        1.0790        1.0790
                           2    0.62198587    0.72927492        0.2117        1.2906
                           3    -.10728905    0.06189654       -0.0365        1.2541
                           4    -.16918559    0.08194532       -0.0576        1.1965
                           5    -.25113091    0.07525173       -0.0855        1.1111
                           6    -.32638264                      -0.1111        1.0000

 Significance Tests Based on 220 Observations
                                                                                Pr >
                                  Test                    DF    Chi-Square     ChiSq

                      H0: No common factors               15      308.5731    <.0001
                      HA: At least one common factor
                      H0: 2 Factors are sufficient         4        2.1801    0.7027
                      HA: More factors are needed
```

Slide 24

School Subjects example from BSMG - ML OUTPUT

```
 Eigenvalues of the Weighted Reduced Correlation Matrix: Total = 5.71554705   Average = 0.95259117
                                Eigenvalue    Difference    Proportion    Cumulative

                           1    4.57559833    3.43564960        0.8006        0.8006
                           2    1.13994873    1.05025464        0.1994        1.0000
                           3    0.08969409    0.06384279        0.0157        1.0157
                           4    0.02585130    0.03607174        0.0045        1.0202
                           5    -.01022044    0.09510453       -0.0018        1.0184
                           6    -.10532497                      -0.0184        1.0000
                                              Factor Pattern
                                             Estimate/StdErr
                                                  Factor1         Factor2

                                  gaelic          0.55798         0.42461
                                                  0.07342         0.08368

                                  english         0.56901         0.28574
                                                  0.06148         0.08579

                                  history         0.39239         0.44952
                                                  0.07933         0.09364

                                  math            0.73802        -0.27943
                                                  0.05948         0.06620

                                  algebra         0.71847        -0.20868
                                                  0.04950         0.08511

                                  geometry        0.59488        -0.13331
                                                  0.05347         0.08464
```

Slide 25

School Subjects example from BSMG - ML OUTPUT

```
                    Rotation Method: Promax (power = 3)

       Rotated Factor Pattern (Standardized Regression Coefficients)
                             Estimate/StdErr

                                 Factor1         Factor2

             gaelic              0.06268         0.66737
                                 0.07879         0.09821

             english             0.19141         0.51810
                                 0.08863         0.09333

             history            -0.08658         0.63580
                                 0.05532         0.08912

             math                0.81101        -0.04511
                                 0.07027         0.05589

             algebra             0.73469         0.02597
                                 0.07277         0.06756

             geometry            0.57413         0.06505
                                 0.07394         0.07831
```

Slide 26

School Subjects example from BSMG - OUTPUT From MPLUS - ML

```
  DATA:
    FILE IS "C:\Mplus\ph5482\schoolsubjects.txt";
    TYPE IS FULLCORR;
    NOBSERVATIONS = 220;

  VARIABLE:
    NAMES ARE gaelic english history math algebra geometry;
    USEVARIABLES ARE gaelic english history math algebra geometry;

  ANALYSIS:
    TYPE IS EFA 1 3;
    ESTIMATOR IS ML;
    ITERATIONS = 1000;
    CONVERGENCE = 0.00005;
---------------------------------------------------------------------
INPUT READING TERMINATED NORMALLY

SUMMARY OF ANALYSIS Number of groups
1 Number of observations
220

Number of dependent variables                                   6
Number of independent variables                                 0
Number of continuous latent variables                           0
Observed dependent variables
  Continuous
   GAELIC      ENGLISH     HISTORY     MATH        ALGEBRA     GEOMETRY

           EIGENVALUES FOR SAMPLE CORRELATION MATRIX
                  1             2             3             4             5          6
      1         2.729         1.129         0.615         0.603         0.523      0.402
```

Slide 27

MPLUS - ML results for fitting 1 factor

```
RESULTS FOR EXPLORATORY FACTOR ANALYSIS

           EXPLORATORY ANALYSIS WITH 1 FACTOR(S) :

           CHI-SQUARE VALUE              52.682
           DEGREES OF FREEDOM                 9
           PROBABILITY VALUE             0.0000

           RMSEA (ROOT MEAN SQUARE ERROR OF APPROXIMATION) :
           ESTIMATE (90 PERCENT C.I.) IS  0.149 ( 0.111  0.189)
           PROBABILITY RMSEA LE 0.05 IS    0.000

           ROOT MEAN SQUARE RESIDUAL IS        0.0987

           ESTIMATED FACTOR LOADINGS
                  1
              --------
 GAELIC         0.500
 ENGLISH        0.539
 HISTORY        0.350
 MATH           0.726
 ALGEBRA        0.729
 GEOMETRY       0.615

           ESTIMATED RESIDUAL VARIANCES
              GAELIC        ENGLISH       HISTORY       MATH          ALGEBRA     Geometry
              --------      --------      --------      --------      --------   ----------
      1         0.750         0.710         0.878         0.473         0.468       0.621
```

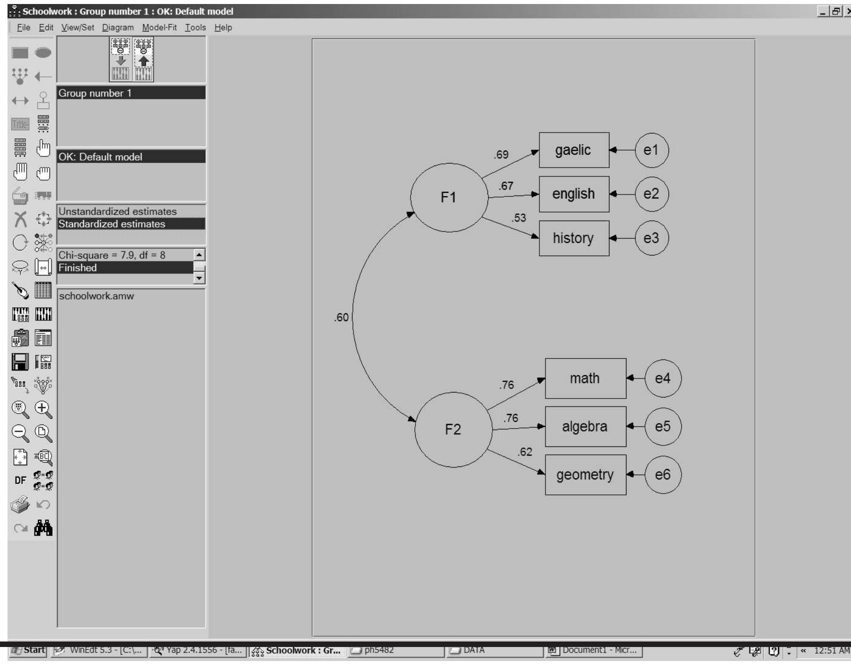
Slide 28

MPLUS - ML results for fitting 2 factors

```
           EXPLORATORY ANALYSIS WITH 2 FACTOR(S) :

           CHI-SQUARE VALUE               2.233
           DEGREES OF FREEDOM                 4
           PROBABILITY VALUE             0.6928

           RMSEA (ROOT MEAN SQUARE ERROR OF APPROXIMATION) :
           ESTIMATE (90 PERCENT C.I.) IS  0.000 ( 0.000  0.077)
           PROBABILITY RMSEA LE 0.05 IS    0.854
           ROOT MEAN SQUARE RESIDUAL IS        0.0140

           VARIMAX ROTATED LOADINGS                    PROMAX ROTATED LOADINGS
                  1             2                            1             2
              --------      --------                    --------      --------

 GAELIC         0.660         0.236       GAELIC         0.683         0.033
 ENGLISH        0.550         0.321       ENGLISH        0.533         0.167
 HISTORY        0.591         0.083       HISTORY        0.648        -0.113
 MATH           0.169         0.771       MATH          -0.031         0.805
 ALGEBRA        0.217         0.716       ALGEBRA        0.040         0.726
 GEOMETRY       0.213         0.571       GEOMETRY       0.077         0.566

           PROMAX FACTOR CORRELATIONS
                  1             2
              --------      --------
      1         1.000
      2         0.524         1.000


           ESTIMATED RESIDUAL VARIANCES
              GAELIC        ENGLISH       HISTORY       MATH          ALGEBRA      Geometry
              --------      --------      --------      --------      --------   ------------
      1         0.508         0.595         0.644         0.377         0.440       0.628
```

Slide 29

CFA of Schoolwork data using AMOS



Slide 30

CFA of Schoolwork data using MPLUS

```
DATA:
  FILE IS "C:\Mplus\ph5482\schoolsubjects.csv";
  TYPE IS FULLCORR;
  NOBSERVATIONS = 220;

VARIABLE:
  NAMES ARE GAELIC ENGLISH HISTORY MATH ALGEBRA GEOMETRY;
  USEVARIABLES ARE GAELIC ENGLISH HISTORY MATH ALGEBRA GEOMETRY;

ANALYSIS:
  TYPE IS GENERAL;
  ESTIMATOR IS ML;
  ITERATIONS = 1000;
  CONVERGENCE = 0.00005;
model:
f1 by GAELIC ENGLISH HISTORY;
f2 by MATH ALGEBRA GEOMETRY;

OUTPUT:   STANDARDIZED;
-----------------------------------------------------------------
Number of observations                                        220
Number of dependent variables                                   6
Number of independent variables                                 0
Number of continuous latent variables                           2

Observed dependent variables
  Continuous
   GAELIC      ENGLISH     HISTORY     MATH        ALGEBRA     GEOMETRY

Continuous latent variables
   F1          F2

Estimator                                                      ML
```

Slide 31

CFA of Schoolwork data using MPLUS

```
TESTS OF MODEL FIT

Chi-Square Test of Model Fit
          Value                              7.954
          Degrees of Freedom                     8
          P-Value                           0.4379

Chi-Square Test of Model Fit for the Baseline Model
          Value                            314.045
          Degrees of Freedom                    15
          P-Value                           0.0000

CFI/TLI
          CFI                                1.000
          TLI                                1.000

Loglikelihood
          H0 Value                        -1716.947
          H1 Value                        -1712.969

Information Criteria
          Number of Free Parameters             13
          Akaike (AIC)                    3459.893
          Bayesian (BIC)                  3504.010
          Sample-Size Adjusted BIC        3462.813
            (n* = (n + 2) / 24)

RMSEA (Root Mean Square Error Of Approximation)
          Estimate                           0.000
          90 Percent C.I.                    0.000  0.079
          Probability RMSEA <= .05           0.764

SRMR (Standardized Root Mean Square Residual)
          Value                              0.033
```

Slide 32

CFA of Schoolwork data using MPLUS

```
                   Estimates     S.E.  Est./S.E.    Std     StdYX

 F1       BY
    GAELIC             1.000    0.000      0.000    0.688    0.690
    ENGLISH            0.973    0.151      6.434    0.670    0.671
    HISTORY            0.770    0.133      5.806    0.530    0.531

 F2       BY
    MATH               1.000    0.000      0.000    0.762    0.764
    ALGEBRA            1.000    0.116      8.630    0.762    0.764
    GEOMETRY           0.806    0.104      7.757    0.615    0.616

 F2       WITH
    F1                 0.314    0.062      5.075    0.598    0.598

 Variances
    F1                 0.474    0.104      4.568    1.000    1.000
    F2                 0.581    0.103      5.667    1.000    1.000

 Residual Variances
    GAELIC             0.522    0.082      6.381    0.522    0.524
    ENGLISH            0.547    0.081      6.752    0.547    0.550
    HISTORY            0.715    0.081      8.772    0.715    0.718
    MATH               0.414    0.068      6.085    0.414    0.416
    ALGEBRA            0.415    0.068      6.091    0.415    0.416
    GEOMETRY           0.618    0.071      8.681    0.618    0.620

R-SQUARE
   Obs.Variable  R-Square

    GAELIC       0.476
    ENGLISH      0.450
    HISTORY      0.282
    MATH         0.584
    ALGEBRA      0.584
    GEOMETRY     0.380
```

Slide 33

The chi-square test for factor analysis models - EFA

- The chi-square test can also be used to see how well an EFA model with q factors fits the data
- For the EFA model the d.f. are specifically $\frac{p(p+1)}{2} - pq - p + \frac{q(q-1)}{2}$
- In EFA, the model being tested is

  H_0: The q factor model is correct (i.e. q factors sufficiently describe the p dimensional vector)

  H_A: More factors are needed
- From Bartholemew and Knott (1998) "Starting with $q = 1$, we then take successive values in turn until the fit of the model is judged to be adequate. Viewed as a testing procedure this is not strictly valid because it does not adjust the significance levels to allow for the sequential character of the test. It rather depends on regarding the p-value of the test as a measure of the adequacy of the model"
- Because of the tendency for the approach described just above to keep adding more and more factors (since more factors will make the model fit better), BK suggest considering the AIC Akaike's information criterion. This is simply a penalized version of the log likelihood where models with more parameters are penalized.

Slide 34

The chi-square test for factor analysis models - CFA

- For CFA the d.f. are $p(p+1)/2-$(the number of parameters being estimated)
- In CFA, the model being tested is

  H_0: The model being assumed is correct

  H_A: A more complicated model is needed
- Satorra-Bentler adjustment, Bollen-Stine adjustment, Amemiya Anderson result

Things we'll do next - Read all of BSMG Chapter 6 and Kline Chapter 7

- More about CFA
- equivalent models
- Nested models
- Multiple Group