PSY 9556B (Feb 12) Growth Mixture Modeling

Let's start with a brief introduction to other "clustering" techniques

- Imagine a data file that you would to conduct an exploratory factor analysis (EFA)
  - The data has variables in columns and subjects in rows
  - The objective is to find a set of factors to explain the variables
- It is possible to transpose the data file so that the variables are in rows and subjects in columns
  - You could now perform an EFA of subjects to determine if people fall into categories/prototypes/classes/types /clusters/profiles
  - This procedure is referred to as Q-technique (see Little, p. 227-228)
- A somewhat similar procedure is cluster analysis
- An alternative procedure is profile analysis using the multivariate approach to repeated measures (see chapter in Tabachnick and Fidell)
- These techniques aim to identify meaningful groups of subjects similar on the variables of interest

An Example of Profiles from a Latent Class Analysis

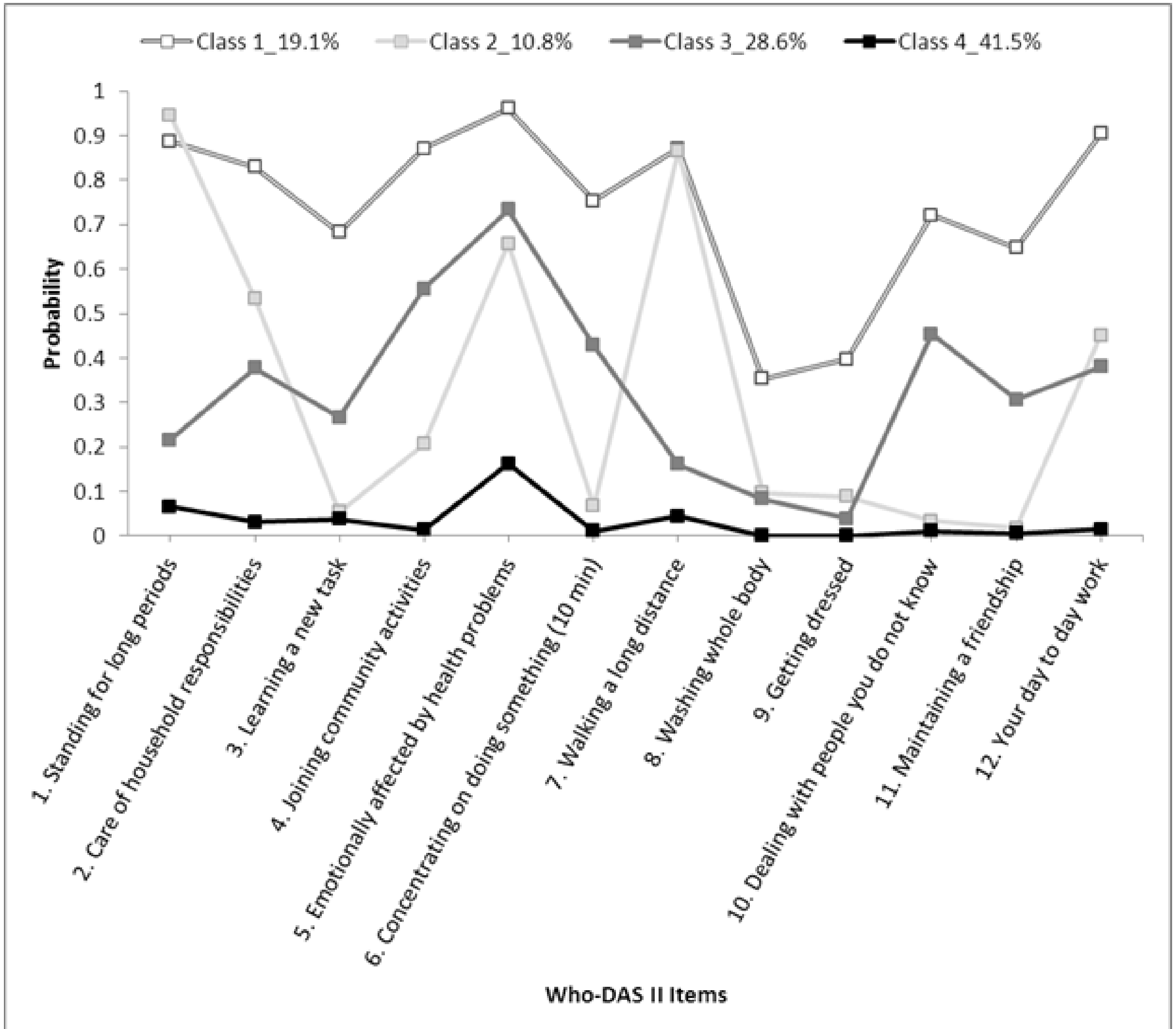
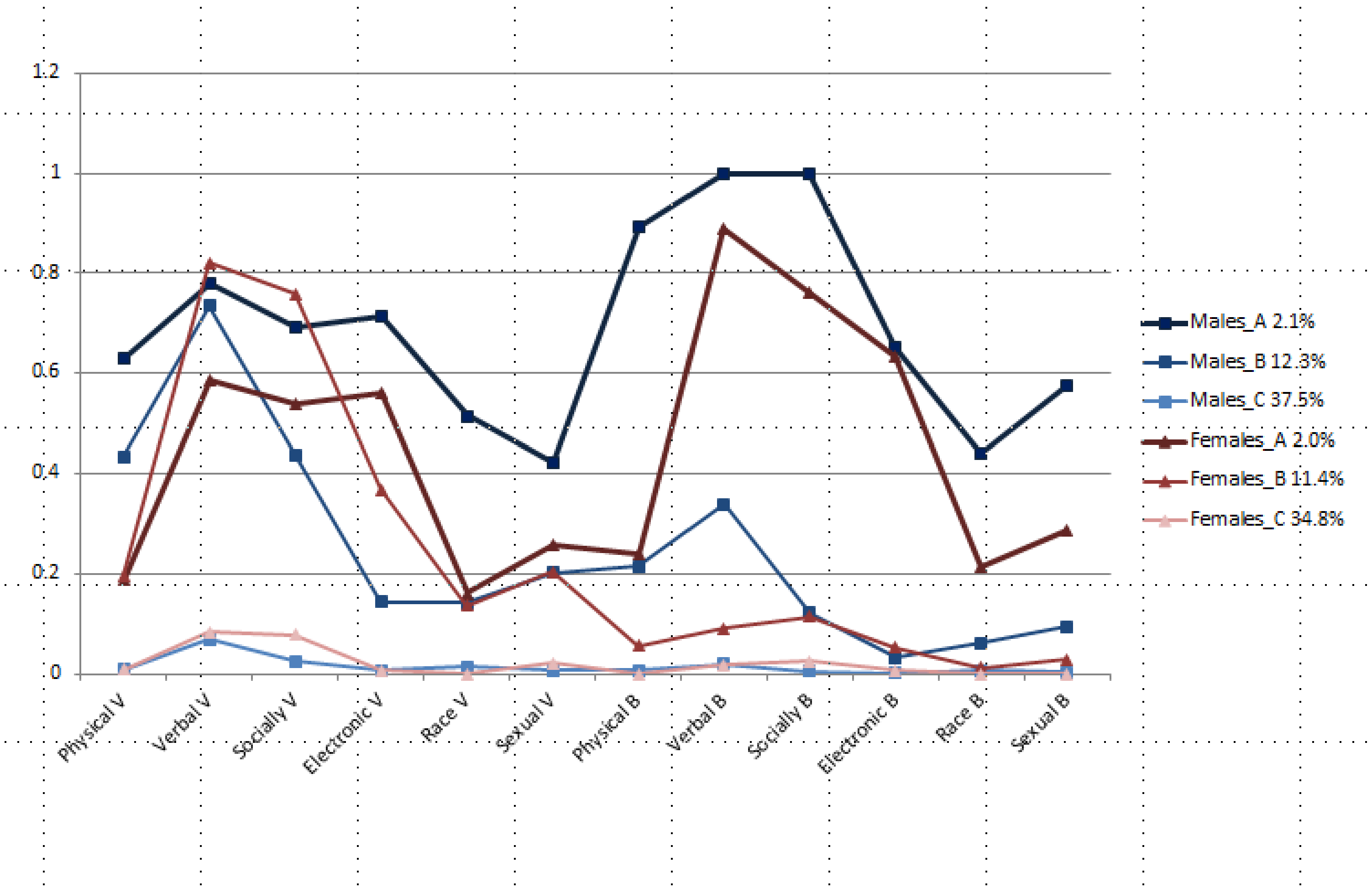


Figure 1. Latent Class Analysis Profiles

Another Latent Class Analysis



Another Latent Class Analysis

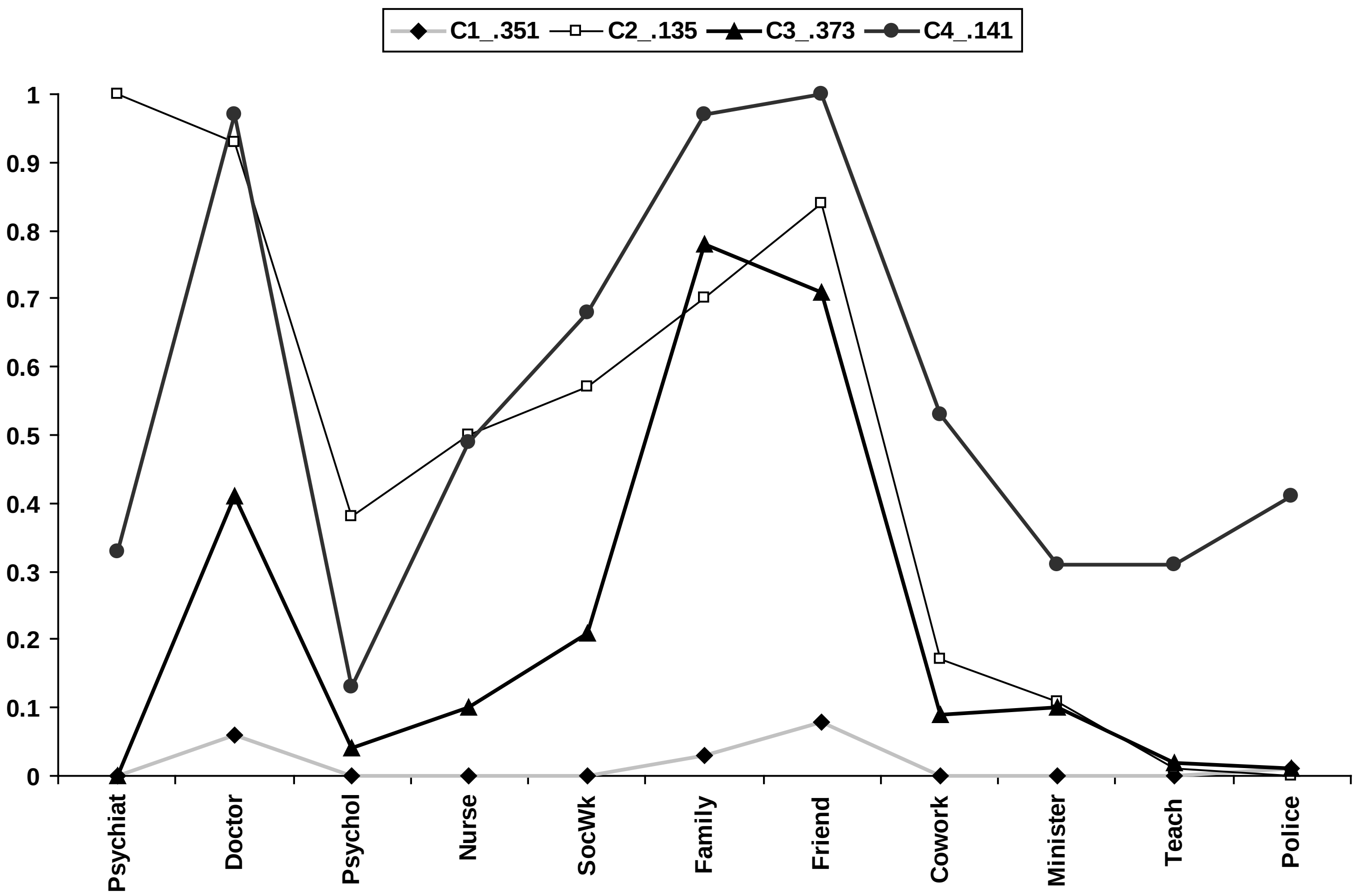


Figure 4. Classes obtained from Latent class analysis. Proportions indicated in legend above. Responses refer to “problems with your emotions, mental health, use of alcohol or drugs, or experiences of violence?” These classes could be labeled as (1) No Use of Services (2) Psychiatrist/Doctor/Friend/Family/Social Worker (3) Friend/Family (4) Friend/Family/Doctor/Coworker/Nurse.

Alternative Way of Plotting Results

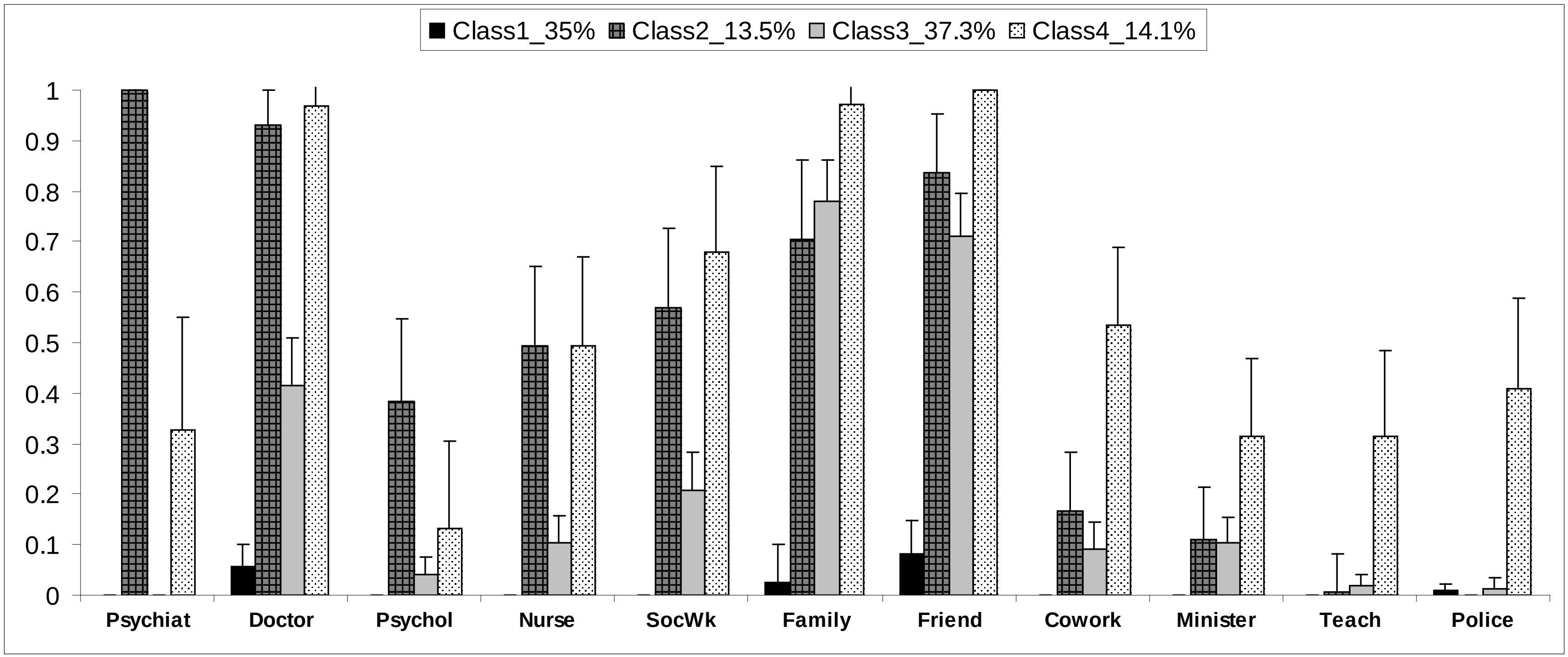
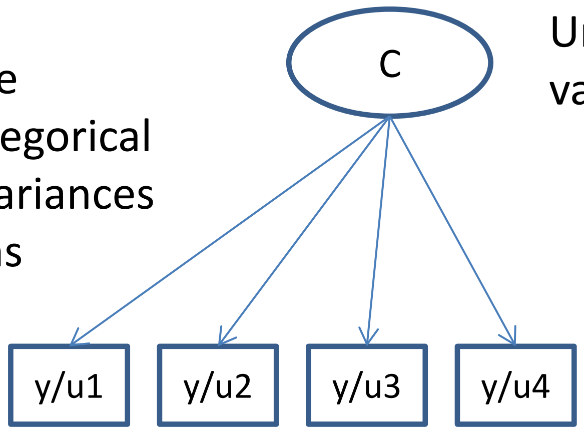


Figure 4. Classes obtained from Latent class analysis. Proportions indicated in legend above. Responses refer to "problems with your emotions, mental health, use of alcohol or drugs, or experiences of violence?" These classes could be labeled as (1) No Use of Services (2) Psychiatrist/Doctor/Friend/Family/Social Worker (3) Friend/Family (4) Friend/Family/Doctor/Coworker/Nurse.

Latent Class Analysis

- LCA uses a modelling approach to derive the classes (categorical items)
- Latent profile analysis (continuous items)
- A specific number of classes tested in steps
  - Start with one class and see if model improves in fit by adding a second class

Item parameters are probabilities for categorical items and means/variances for continuous items



Unordered latent class variable with K categories

$$P = \frac{1}{1 + e^{-\tau}}$$

threshold to probability

LCA – Deciding on Number of Classes (applies also to GMM)

- The number of classes to model is determined by comparing models differing in number of classes
- The chi-square difference test based on the likelihood ratio not appropriate for LCA
- An approximation by Lo, Mendell, and Rubin (2001) has been proposed
- Mplus program provides bootstrap likelihood ratio test (BLTR) as a test to compare the increase in model fit by adding a class.
- Can also use Akaike's Information Criterion (AIC; Akaike, 1987) and the Bayesian Information Criterion (BIC; Schwartz, 1978)
- Entropy, indicates the precision of classification (Magidson & Vermunt, 2002)
- Theory
- Number of individual in classes

LCA – Reporting Analyses for Selection of Number of Classes

Table 3. LCA Analyses

Fit Index and Classification Indices	Number of Classes 1	2	3	4	5
Loglikelihood	-2948.022	-2385.174	-2294.951	-2248.947	-2217.669
AIC	5920.045	4820.348	4665.901	4599.893	4563.338
Adj. BIC	5931.192	4843.572	4701.202	4647.271	4622.792
Lo-Mendell-Rubin LRT p-value		1111.684 p = .0000	178.200 p = .0028	90.863 p = .0516	61.776 p = .1006
Bootstrapped LRT p-value		1125.696 p = .000	180.447 p = .000	92.008 p = .000	62.555 p = .000
Entropy		0.876	0.826	0.846	.850
Number of people in each category		190, 247	87, 162, 198	86, 48, 123, 190	107, 38, 61, 51, 190

See figure, slide 2

EFA vs. LCA Example

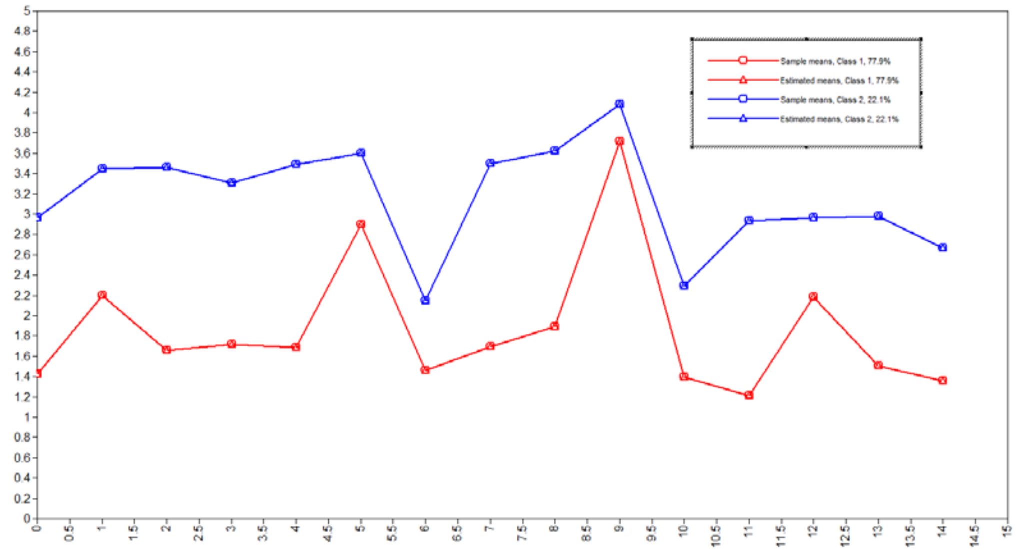
Table 1

Three-factor solution and two-class latent class model

Variables	3 Factor Varimax Rotated Loadings			2-Class Model Means	
	I	II	III	Low Class n = 590	High Class n = 165
Abandonment	.60	.41	.13	1.66	3.46
Mistrust	.66	.23	.27	1.71	3.31
Emotional Deprivation	.78	-.05	.16	1.68	3.49
Defectiveness	.74	.29	-.01	1.22	2.94
Social Isolation	.72	.18	.15	1.70	3.50
Dependence	.49	.53	-.18	1.39	2.30
Vulnerability to Harm	.65	.31	.19	1.36	2.67
Enmeshment	.19	.72	.07	1.46	2.15
Failure	.66	.43	-.22	1.42	2.97
Entitlement	.44	-.05	.58	2.18	2.96
Insufficient Self-Control	.64	.14	-.02	2.20	3.44
Subjugation	.62	.54	.02	1.50	2.98
Self-Sacrifice	.03	.63	.42	2.90	3.60
Emotional Inhibition	.67	.13	.15	1.89	3.61
Unrelenting Standards	.01	.15	.84	3.71	4.08

EFA vs. LCA Example

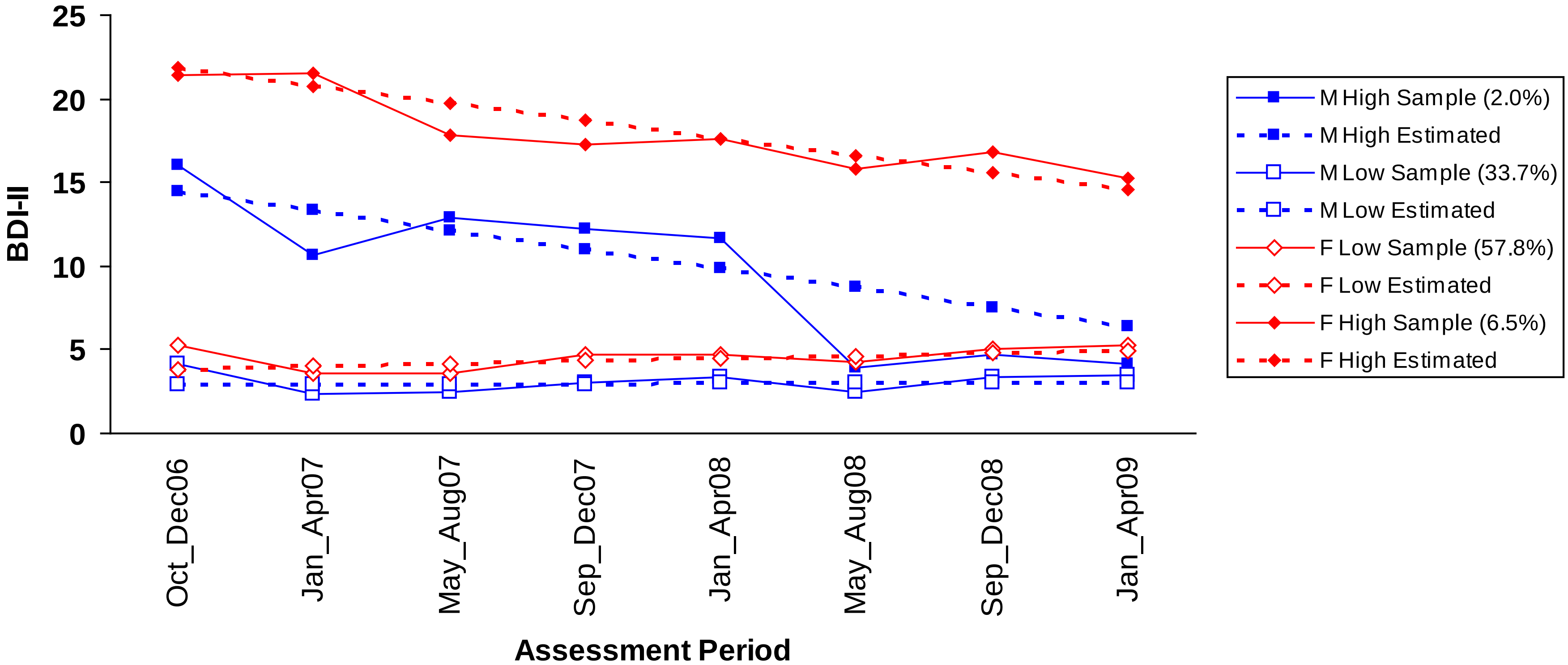
```
Title: YOUNG SCHEMA;
data: File is c:\Paul\mplus\schemascales.dat;
    Format is 1F4.0, 1F1.0, 15F8.2;
LISTWISE = ON;
variable: names are studid gender failure iscl aban mist edep
ssac enme sali einh usta depe defe enti subj vuln;
missing = blank;
auxiliary = studid gender;
usevariables are failure iscl aban mist edep
ssac enme sali einh usta depe defe enti subj vuln;
classes = c (2);
analysis: type = mixture;
starts = 100 10;
plot:
type is plot3;
series = failure iscl aban mist edep
ssac enme sali einh usta depe defe enti subj vuln (*);
output: tech1 tech11;
savedata:
file is c:\paul\mplus\classmember3;
save= cprobabilities;
```



EFA vs. LCA Example

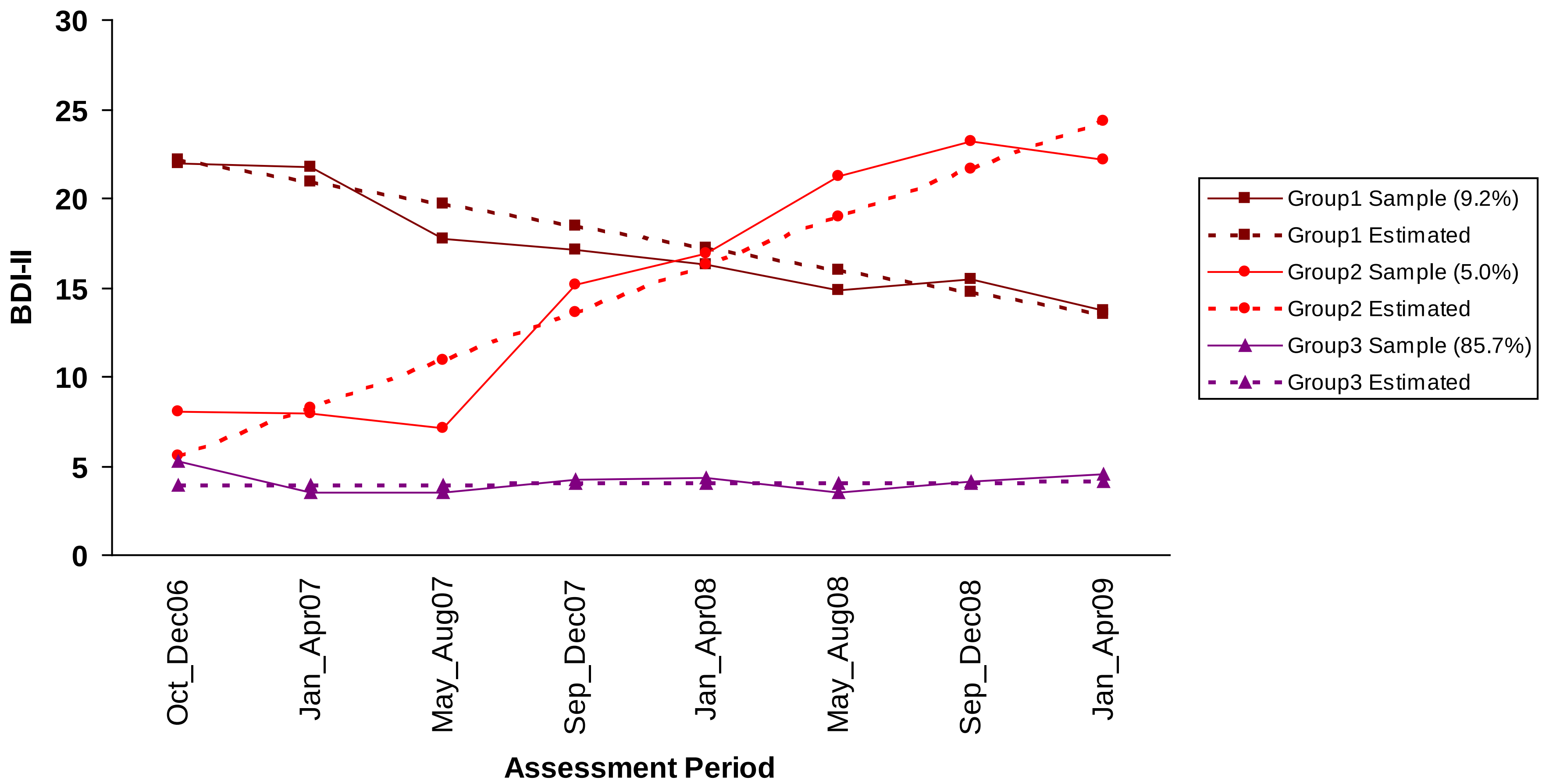
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Latent Class 1				
Means				
FAILURE	1.420	0.061	23.275	0.000
ISCL	2.196	0.073	29.996	0.000
ABAN	1.658	0.051	32.608	0.000
MIST	1.712	0.079	21.749	0.000
EDEP	1.684	0.083	20.332	0.000
SSAC	2.900	0.057	50.940	0.000
ENME	1.458	0.032	45.864	0.000
SALI	1.696	0.072	23.684	0.000
EINH	1.893	0.090	21.052	0.000
USTA	3.713	0.058	64.124	0.000
DEPE	1.391	0.033	41.651	0.000
DEFE	1.215	0.046	26.515	0.000
ENTI	2.184	0.053	41.492	0.000
SUBJ	1.503	0.070	21.336	0.000
VULN	1.357	0.052	25.891	0.000
Variances				
FAILURE	0.744	0.076	9.843	0.000
ISCL	1.177	0.080	14.634	0.000
ABAN	0.892	0.082	10.825	0.000
MIST	0.819	0.077	10.686	0.000
EDEP	1.036	0.094	11.047	0.000
SSAC	1.255	0.059	21.442	0.000
ENME	0.514	0.044	11.773	0.000
SALI	0.912	0.084	10.802	0.000
EINH	1.115	0.103	10.852	0.000
USTA	1.690	0.071	23.813	0.000
DEPE	0.445	0.038	11.696	0.000
DEFE	0.573	0.066	8.656	0.000
ENTI	1.026	0.060	17.097	0.000
SUBJ	0.513	0.070	7.355	0.000
VULN	0.513	0.054	9.423	0.000
Latent Class 2				
Means				
FAILURE	2.967	0.251	11.822	0.000
ISCL	3.443	0.163	21.120	0.000
ABAN	3.457	0.323	10.695	0.000
MIST	3.306	0.174	18.982	0.000
EDEP	3.486	0.227	15.350	0.000
SSAC	3.600	0.114	31.519	0.000
ENME	2.145	0.143	15.045	0.000
SALI	3.498	0.261	13.386	0.000
EINH	3.621	0.183	19.796	0.000
USTA	4.078	0.132	30.994	0.000
DEPE	2.295	0.166	13.824	0.000
DEFE	2.936	0.314	9.340	0.000
ENTI	2.964	0.128	23.197	0.000
SUBJ	2.979	0.175	17.012	0.000
VULN	2.667	0.196	13.600	0.000
Variances				
FAILURE	0.744	0.076	9.843	0.000
ISCL	1.177	0.080	14.634	0.000
ABAN	0.892	0.082	10.825	0.000
MIST	0.819	0.077	10.686	0.000
EDEP	1.036	0.094	11.047	0.000
SSAC	1.255	0.059	21.442	0.000
ENME	0.514	0.044	11.773	0.000
SALI	0.912	0.084	10.802	0.000
EINH	1.115	0.103	10.852	0.000
USTA	1.690	0.071	23.813	0.000
DEPE	0.445	0.038	11.696	0.000
DEFE	0.573	0.066	8.656	0.000
ENTI	1.026	0.060	17.097	0.000
SUBJ	0.513	0.070	7.355	0.000
VULN	0.513	0.054	9.423	0.000

Growth Mixture Modeling

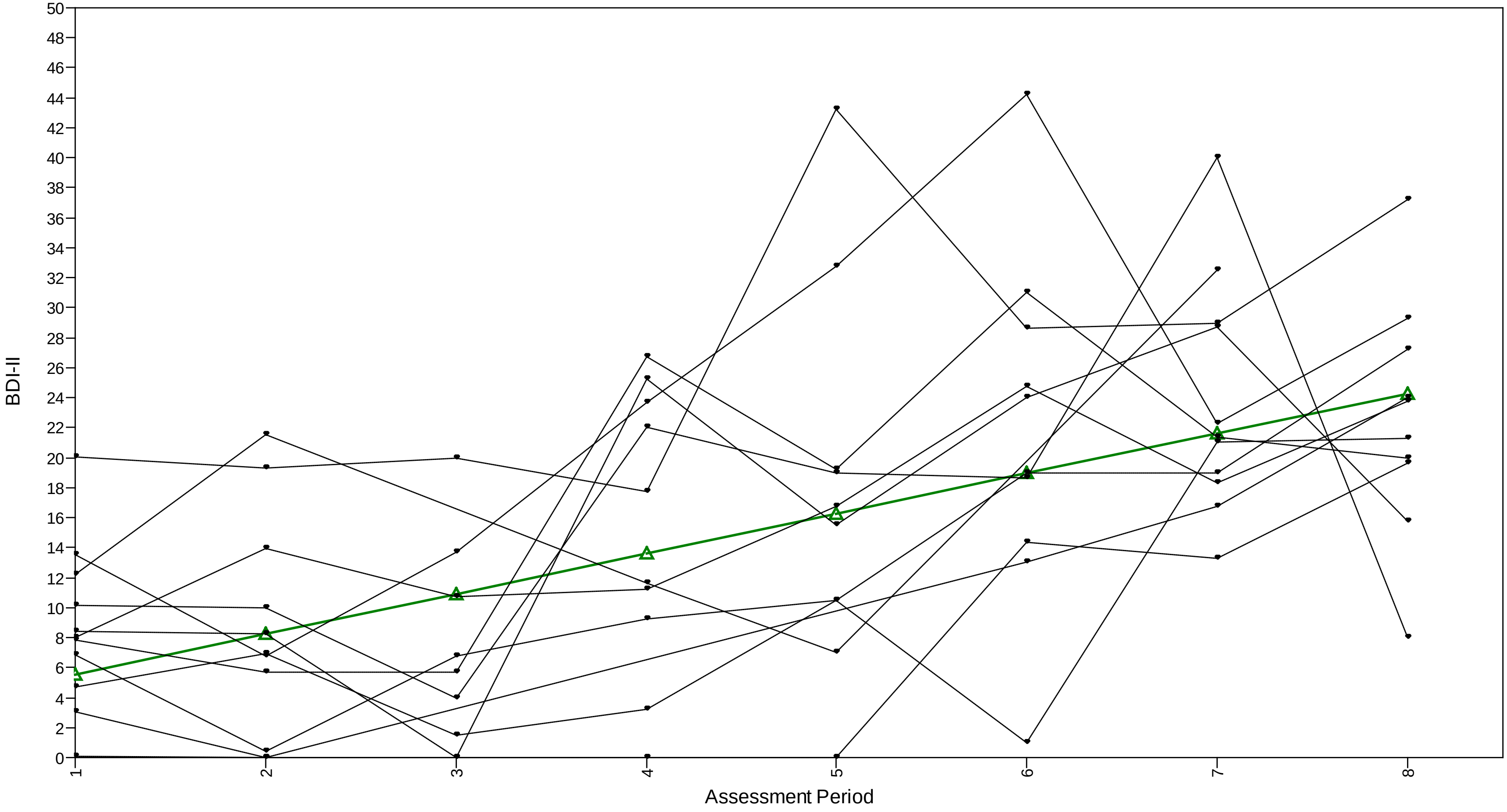



```
VARIABLE: NAMES ARE dep1-dep8;
CLASSES = cg (2) c (2);
KNOWNCLASSES = cg (g = 0 g = 1);
```

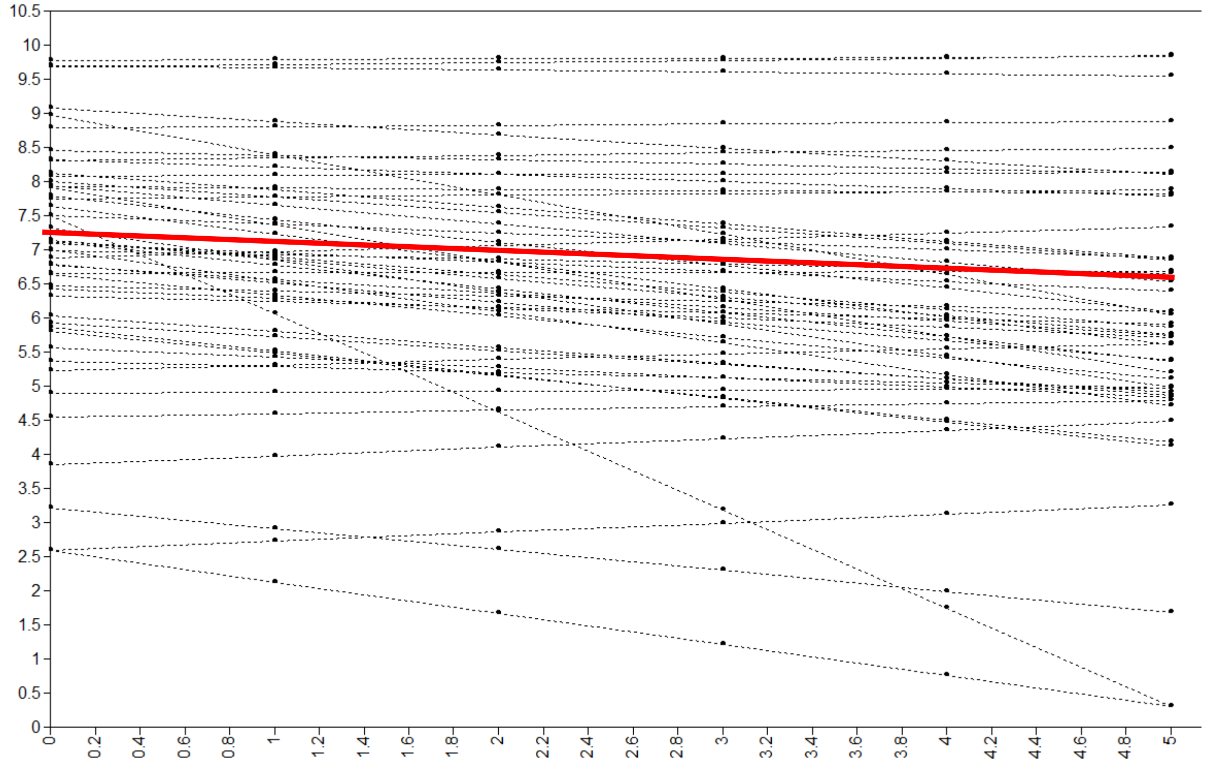
Growth Mixture Modeling



Growth Mixture Modeling



Growth Mixture Modeling: Another Example
Start with a LGM Model




```
USEVARIABLES ARE w1enjun w6enjun w11enjun w16enjun w21enjun w26enjun;
model:
I S | w1enjun@0 w6enjun@.2 w11enjun@.4 w16enjun@.6 w21enjun@.8 w26enjun@1;
plot:
type=plot3;
series = w1enjun(0) w6enjun(1) w11enjun(2) w16enjun(3) w21enjun(4) w26enjun(5);
output: sampstat residual stdyx tech4 modindices;
```

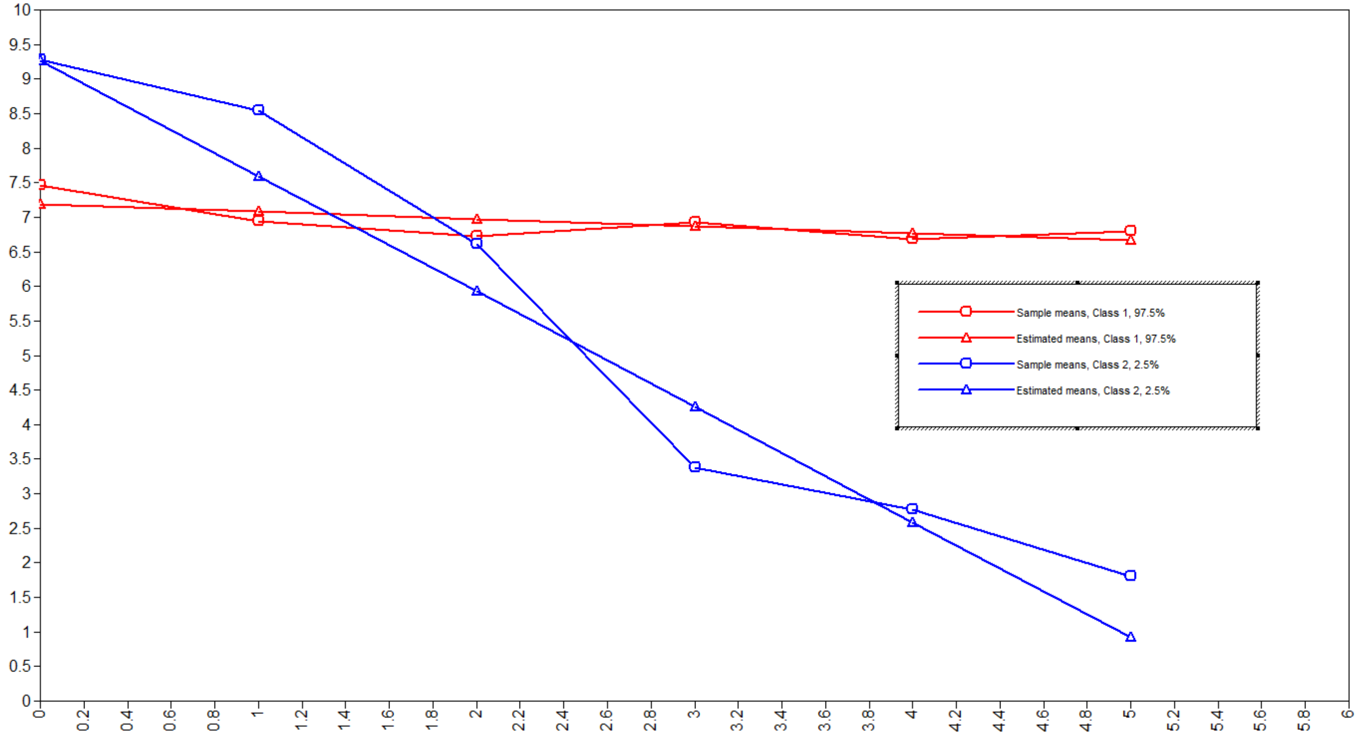
LGM Output

```
MODEL FIT INFORMATION

Number of Free Parameters                        11

Loglikelihood

          H0 Value                        -3779.045
          H1 Value                        -3745.014

Information Criteria

          Akaike (AIC)                     7580.091
          Bayesian (BIC)                   7624.349
          Sample-Size Adjusted BIC         7589.443
            (n* = (n + 2) / 24)

Chi-Square Test of Model Fit

          Value                              68.063
          Degrees of Freedom                     16
          P-Value                            0.0000

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                            0.089
          90 Percent C.I.                     0.068  0.111
          Probability RMSEA <= .05            0.002

CFI/TLI

          CFI                                 0.957
          TLI                                 0.960

Chi-Square Test of Model Fit for the Baseline Model

          Value                            1239.482
          Degrees of Freedom                     15
          P-Value                            0.0000

SRMR (Standardized Root Mean Square Residual)

          Value                               0.078
```

```
 S         WITH
    I                    -0.580      0.252     -2.301      0.021

 Means
    I                     7.241      0.100     72.383      0.000
    S                    -0.724      0.112     -6.442      0.000

 Variances
    I                     2.948      0.289     10.192      0.000
    S                     2.338      0.348      6.718      0.000
```

GMM 2 Classes

```
USEVARIABLES ARE w1enjun w6enjun w11enjun w16enjun w21enjun w26enjun;
classes = c(2);
ANALYSIS: type = mixture;
starts = 40 16;
MODEL:
%overall%
I S | w1enjun@0 w6enjun@.2 w11enjun@.4 w16enjun@.6 w21enjun@.8 w26enjun@1;
plot:
type=plot3;
series = w1enjun(0) w6enjun(1) w11enjun(2) w16enjun(3) w21enjun(4) w26enjun(5);
output: sampstat stdyx tech1 tech4 tech7 tech11 tech12 tech14;
```



GMM Two Classes

```
     MAXIMUM LOG-LIKELIHOOD VALUE FOR THE UNRESTRICTED (H1) MODEL IS -3745.014

RANDOM STARTS RESULTS RANKED FROM THE BEST TO THE WORST LOGLIKELIHOOD VALUES

Final stage loglikelihood values at local maxima, seeds, and initial stage start numbers:

           -3746.051  253358           2
           -3746.051  364676           27
           -3746.051  650371           14
           -3746.051  76974            16
           -3746.051  107446           12
           -3746.051  285380           1
           -3746.051  68985            17
           -3746.051  637345           19
           -3746.051  902278           21
           -3746.051  347515           24
           -3746.051  392418           28
           -3746.051  207896           25
           -3759.097  851945           18
           -3759.097  246261           38
           -3759.097  462953           7
           -3759.097  127215           9


THE BEST LOGLIKELIHOOD VALUE HAS BEEN REPLICATED.  RERUN WITH AT LEAST TWICE THE
RANDOM STARTS TO CHECK THAT THE BEST LOGLIKELIHOOD IS STILL OBTAINED AND REPLICATED.


THE MODEL ESTIMATION TERMINATED NORMALLY
```

GMM 2 Classes

```
Number of Free Parameters                         14

Loglikelihood

          H0 Value                         -3746.051
          H0 Scaling Correction Factor        1.6304
            for MLR

Information Criteria

          Akaike (AIC)                      7520.101
          Bayesian (BIC)                    7576.429
          Sample-Size Adjusted BIC          7532.004
            (n* = (n + 2) / 24)



FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODEL

    Latent
   Classes

       1        402.70103          0.97506
       2         10.29897          0.02494


FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASS PATTERNS
BASED ON ESTIMATED POSTERIOR PROBABILITIES

    Latent
   Classes

       1        402.70107          0.97506
       2         10.29893          0.02494


CLASSIFICATION QUALITY

     Entropy                         0.969
```

GMM 2 Classes

Latent Class 1

S WITH				
I	-0.172	0.235	-0.732	0.464
Means				
I	7.183	0.106	67.879	0.000
S	-0.515	0.107	-4.821	0.000
Variances				
I	2.799	0.338	8.291	0.000
S	0.951	0.282	3.376	0.001
Residual Variances				
W1ENJUN	2.056	0.255	8.064	0.000
W6ENJUN	1.540	0.235	6.555	0.000
W11ENJUN	2.348	0.367	6.396	0.000
W16ENJUN	0.989	0.119	8.336	0.000
W21ENJUN	0.879	0.180	4.877	0.000
W26ENJUN	0.691	0.117	5.919	0.000

Latent Class 2

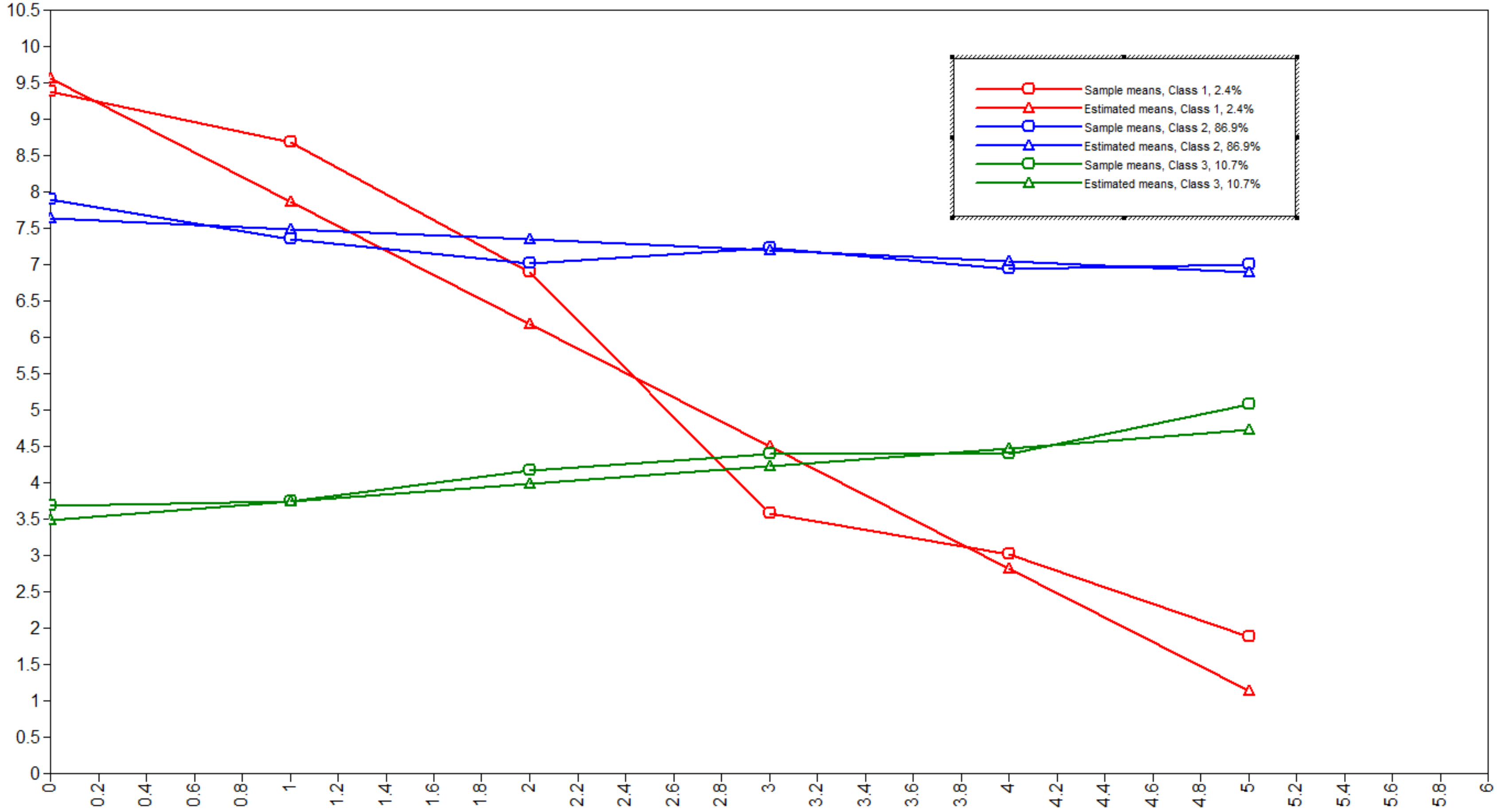
S WITH				
I	-0.172	0.235	-0.732	0.464
Means				
I	9.266	0.541	17.125	0.000
S	-8.348	1.006	-8.296	0.000
Variances				
I	2.799	0.338	8.291	0.000
S	0.951	0.282	3.376	0.001
Residual Variances				
W1ENJUN	2.056	0.255	8.064	0.000
W6ENJUN	1.540	0.235	6.555	0.000
W11ENJUN	2.348	0.367	6.396	0.000
W16ENJUN	0.989	0.119	8.336	0.000
W21ENJUN	0.879	0.180	4.877	0.000
W26ENJUN	0.691	0.117	5.919	0.000

GMM 2 Classes

```
TECHNICAL 11 OUTPUT

     Random Starts Specifications for the k-1 Class Analysis Model
        Number of initial stage random starts                  40
        Number of final stage optimizations                    16



     VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 1 (H0) VERSUS 2 CLASSES

          H0 Loglikelihood Value                        -3779.045
          2 Times the Loglikelihood Difference             65.990
          Difference in the Number of Parameters                3
          Mean                                             -6.979
          Standard Deviation                               25.801
          P-Value                                          0.0004

     LO-MENDELL-RUBIN ADJUSTED LRT TEST

          Value                                            62.529
          P-Value                                          0.0005
```

```
TECHNICAL 14 OUTPUT

     Random Starts Specifications for the k-1 Class Analysis Model
        Number of initial stage random starts                  40
        Number of final stage optimizations                    16

     Random Starts Specification for the k-1 Class Model for Generated Data
        Number of initial stage random starts                   0
        Number of final stage optimizations for the
           initial stage random starts                          0
     Random Starts Specification for the k Class Model for Generated Data
        Number of initial stage random starts                  40
        Number of final stage optimizations                     8
     Number of bootstrap draws requested                   Varies


     PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR 1 (H0) VERSUS 2 CLASSES

          H0 Loglikelihood Value                        -3779.045
          2 Times the Loglikelihood Difference             65.990
          Difference in the Number of Parameters                3
          Approximate P-Value                              0.0000
          Successful Bootstrap Draws                            5
```

GMM 3 Classes



GMM 3 Classes

```
     MAXIMUM LOG-LIKELIHOOD VALUE FOR THE UNRESTRICTED (H1) MODEL IS -3745.014

RANDOM STARTS RESULTS RANKED FROM THE BEST TO THE WORST LOGLIKELIHOOD VALUES

Final stage loglikelihood values at local maxima, seeds, and initial stage start numbers:

           -3724.148  573096           20
           -3724.148  285380           1
           -3724.148  364676           27
           -3724.148  372176           23
           -3724.148  399671           13
           -3724.148  27071            15
           -3724.148  207896           25
           -3724.148  107446           12
           -3724.148  902278           21
           -3724.148  195873           6
           -3724.148  68985            17
           -3742.891  253358           2
           -3746.051  569131           26
           -3746.051  939021           8
           -3752.868  153942           31

1 perturbed starting value run(s) did not converge.



THE BEST LOGLIKELIHOOD VALUE HAS BEEN REPLICATED.  RERUN WITH AT LEAST TWICE THE
RANDOM STARTS TO CHECK THAT THE BEST LOGLIKELIHOOD IS STILL OBTAINED AND REPLICATED.
```

GMM 3 Classes

```
Number of Free Parameters                        17

Loglikelihood

          H0 Value                        -3724.148
          H0 Scaling Correction Factor       1.5092
            for MLR

Information Criteria

          Akaike (AIC)                     7482.296
          Bayesian (BIC)                   7550.694
          Sample-Size Adjusted BIC         7496.750
            (n* = (n + 2) / 24)


FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASSES
BASED ON THE ESTIMATED MODEL

    Latent
   Classes

       1          9.81451          0.02376
       2        358.94485          0.86912
       3         44.24064          0.10712


FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASS PATTERNS
BASED ON ESTIMATED POSTERIOR PROBABILITIES

    Latent
   Classes

       1          9.81451          0.02376
       2        358.94482          0.86912
       3         44.24067          0.10712


CLASSIFICATION QUALITY

     Entropy                         0.879
```

```
Latent Class 1

 S        WITH
    I                  0.603      0.212      2.844      0.004

 Means
    I                  9.556      0.462     20.686      0.000
    S                 -8.424      0.993     -8.480      0.000

 Variances
    I                  1.199      0.232      5.173      0.000
    S                  0.609      0.257      2.366      0.018

Latent Class 2

 Means
    I                  7.639      0.133     57.293      0.000
    S                 -0.744      0.120     -6.175      0.000

Latent Class 3

 Means
    I                  3.492      0.598      5.843      0.000
    S                  1.232      0.404      3.053      0.002
```

GMM 3 Classes

VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 2 (H0) VERSUS 3 CLASSES

H0 Loglikelihood Value	-3746.051
2 Times the Loglikelihood Difference	43.805
Difference in the Number of Parameters	3
Mean	2.832
Standard Deviation	9.517
P-Value	0.0023

LO-MENDELL-RUBIN ADJUSTED LRT TEST

Value	41.508
P-Value	0.0030

PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR 2 (H0) VERSUS 3 CLASSES

H0 Loglikelihood Value	-3746.051
2 Times the Loglikelihood Difference	43.805
Difference in the Number of Parameters	3
Approximate P-Value	0.0000
Successful Bootstrap Draws	10

GMM 4 Classes

```
Number of Free Parameters                          20

Loglikelihood

          H0 Value                          -3717.004
          H0 Scaling Correction Factor         1.4087
            for MLR

Information Criteria

          Akaike (AIC)                       7474.007
          Bayesian (BIC)                     7554.476
          Sample-Size Adjusted BIC           7491.012
            (n* = (n + 2) / 24)


FINAL CLASS COUNTS AND PROPORTIONS FOR THE LATENT CLASS
BASED ON THE ESTIMATED MODEL

    Latent
   Classes

       1         53.59245          0.12976
       2          8.82657          0.02137
       3          9.49459          0.02299
       4        341.08638          0.82588
```

```
TECHNICAL 11 OUTPUT

     Random Starts Specifications for the k-1 Class Analysis Model
        Number of initial stage random starts                  40
        Number of final stage optimizations                    16


     VUONG-LO-MENDELL-RUBIN LIKELIHOOD RATIO TEST FOR 3 (H0) VERSUS 4 CLASSES

          H0 Loglikelihood Value                        -3724.148
          2 Times the Loglikelihood Difference             14.288
          Difference in the Number of Parameters                3
          Mean                                              7.526
          Standard Deviation                               15.775
          P-Value                                          0.2197

     LO-MENDELL-RUBIN ADJUSTED LRT TEST

          Value                                            13.539
          P-Value                                          0.2322


    PARAMETRIC BOOTSTRAPPED LIKELIHOOD RATIO TEST FOR 3 (H0) VERSUS 4 CLASSES

          H0 Loglikelihood Value                        -3724.148
          2 Times the Loglikelihood Difference             14.288
          Difference in the Number of Parameters                3
          Approximate P-Value                              0.0200
          Successful Bootstrap Draws                          100
```