# Non-Normal Growth Mixture Modeling

Bengt Muthén & Tihomir Asparouhov
Mplus
www.statmodel.com

bmuthen@statmodel.com

Presentation to PSMG May 6, 2014

# Overview

1. A new GMM method
   - Examples of skew distributions
   - Normal mixtures
   - Introducing mixtures of non-normal distributions
   - Cluster analysis with non-normal mixtures
   - Non-normal mixtures of latent variable models:
     - GMM of BMI in the NLSY multiple-cohort study
     - GMM of BMI in the Framingham data
     - Math and high school dropout in the LSAY study
     - Cat's cradle concern
   - Disadvantages and advantages of non-normal mixtures
   - Mplus specifications
2. A new SEM method: Non-normal SEM
   - Path analysis
   - Factor analysis
   - SEM

References:

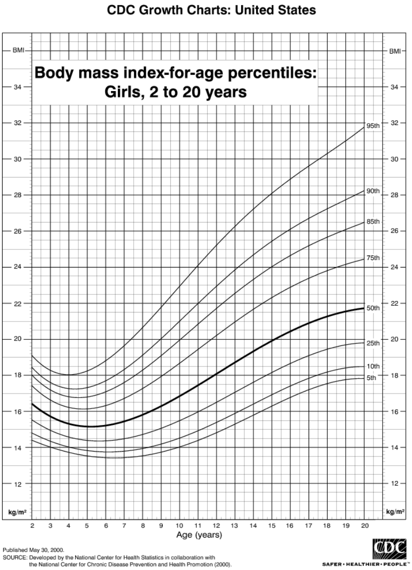
Asparouhov & Muthén (2014). Non-normal mixture modeling and SEM. Mplus Web Note No. 19. - More to come

# Examples of Skewed Distributions

- Body Mass Index (BMI) in obesity studies (long right tail)
- Mini Mental State Examination (MMSE) cognitive test in Alzheimer's studies (long left tail)
- PSA scores in prostate cancer studies (long right tail)
- Ham-D score in antidepressant studies (long right tail)

# Body Mass Index (BMI): $kg/m^2$

Normal $18 < BMI < 25$, Overweight $25 < BMI < 30$, Obese $> 30$



# NLSY Multiple-Cohort Data Ages 12 to 23

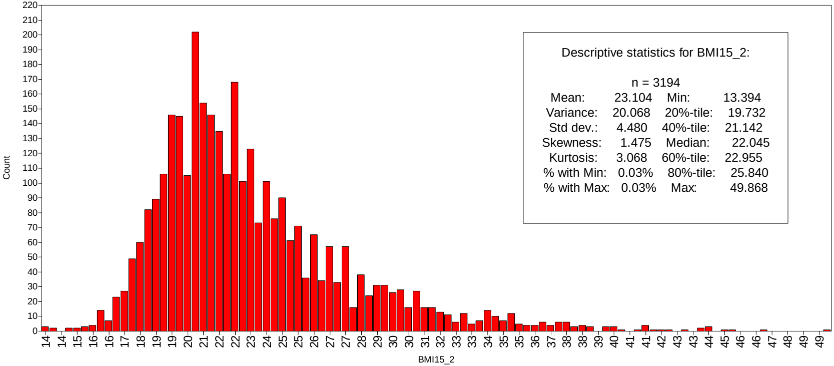
Accelerated longitudinal design - NLSY97

| | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1997 | 1,165 | 1,715 | 1,847 | 1,868 | 1,709 | 613 | | | | | | |
| 1998 | | 104 | 1,592 | 1,671 | 1,727 | 1,739 | 1,400 | 106 | | | | |
| 1999 | | | 108 | 1,659 | 1,625 | 1,721 | 1,614 | 1,370 | 65 | | | |
| 2000 | | | | 57 | 1,553 | 1,656 | 1,649 | 1,597 | 1,390 | 132 | | |
| 2001 | | | | | 66 | 1,543 | 1,615 | 1,602 | 1,582 | 1,324 | 109 | |
| 2002 | | | | | | | 1,614 | 1,587 | 1,643 | 1,582 | 1,324 | 106 |
| 2003 | | | | | | | 112 | 1,497 | 1,600 | 1,582 | 1,564 | 1,283 |
| Totals | 1,165 | 1,819 | 3,547 | 5,255 | 6,680 | 7,272 | 8,004 | 7,759 | 6,280 | 4,620 | 2,997 | 1,389 |

NLSY, National Longitudinal Survey of Youth

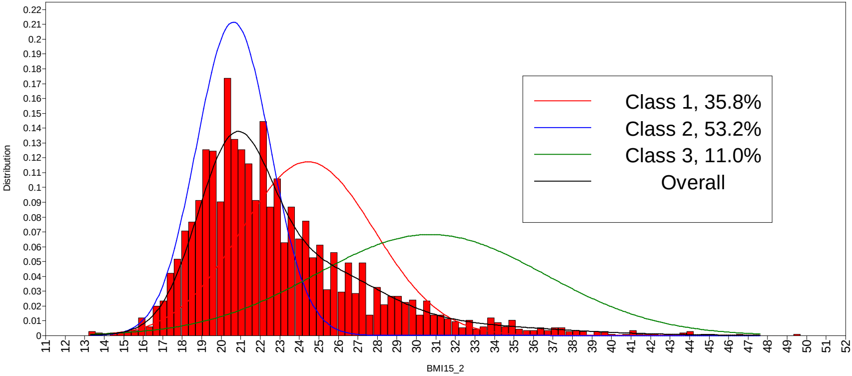
Source: Nonnemaker et al. (2009). Youth BMI trajectories: Evidence from the NLSY97, Obesity

# BMI at Age 15 in the NLSY (Males, $n = 3194$)



# Mixtures for Male BMI at Age 15 in the NLSY

- Skewness = 1.5, kurtosis = 3.1
- Mixtures of normals with 1-4 classes have BIC = 18,658, 17,697, 17,638, 17,637 (tiny class)
- 3-class mixture shown below

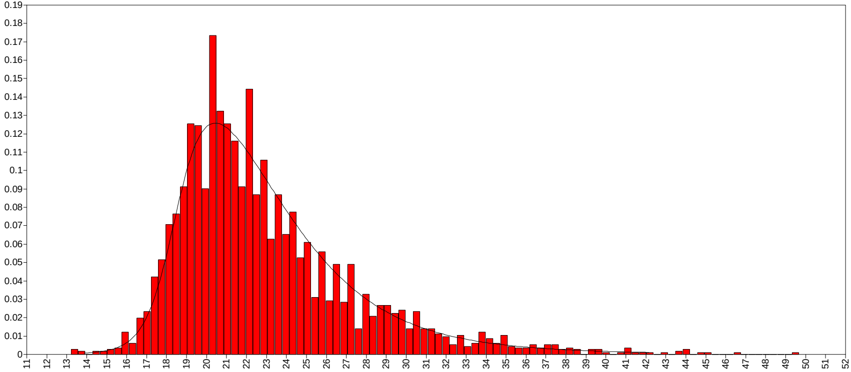


# Several Classes or One Non-Normal Distribution?

- Pearson (1895)
- Hypertension debate:
  - Platt (1963): Hypertension is a ”disease” (separate class)
  - Pickering (1968): Hypertension is merely the upper tail of a skewed distribution
- Schork et al (1990): Two-component mixture versus lognormal
- Bauer & Curran (2003): Growth mixture modeling classes may merely reflect a non-normal distribution so that classes have no substantive meaning
- Muthén (2003) comment on BC: Substantive checking of classes related to antecedents, concurrent events, consequences (distal outcomes), and usefulness
- Multivariate case more informative than univariate

# What If We Could Instead Fit The Data With a Skewed Distribution?

- Then a mixture would not be necessitated by a non-normal distribution, but a single class may be sufficient
- A mixture of non-normal distributions is possible



# Introducing Mixtures of Non-Normal Distributions in Mplus Version 7.2

In addition to a mixture of normal distributions, it is now possible to use

- Skew-normal: Adding a skew parameter to each variable
- T: Adding a degree of freedom parameter (thicker or thinner tails)
- Skew-T: Adding skew and df parameters (stronger skew possible than skew-normal)

References

- Azzalini (1985), Azzalini & Dalla Valle (1996): skew-normal
- Arellano-Valle & Genton (2010): extended skew-t
- McNicholas, Murray, 2013, 2014: skew-t as a special case of the generalized hyperbolic distribution
- McLachlan, Lee, Lin, 2013, 2014: restricted and unrestricted skew-t

# Skew T-Distribution Formulas

Y can be seen as the sum of a mean, a part that produces skewness, and a part that adds a symmetric distribution:

$$Y = \mu + \delta |U_0| + U_1,$$

where $U_0$ has a univariate t and $U_1$ a multivariate t distribution. Expectation, variance ($\delta$ is a skew vector, $\nu$ the df):

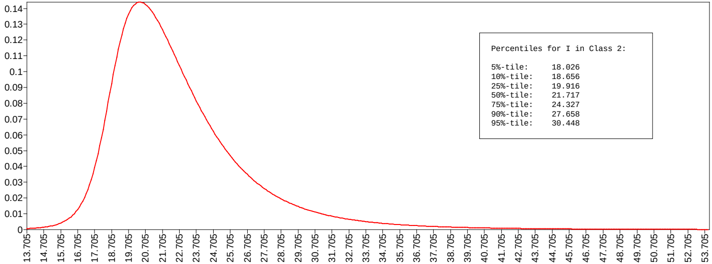
$$E(Y) = \mu + \delta \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \sqrt{\frac{\nu}{\pi}},$$

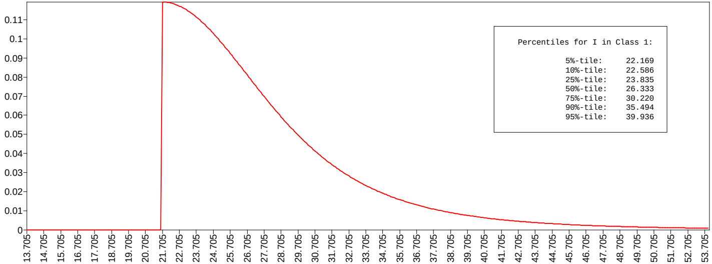
$$Var(Y) = \frac{\nu}{\nu-2}(\Sigma + \delta\delta^T) - \left( \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \right)^2 \frac{\nu}{\pi} \delta\delta^T$$

Marginal and conditional distributions:

- Marginal is also a skew-t distribution
- Conditional is an extended skew-t distribution

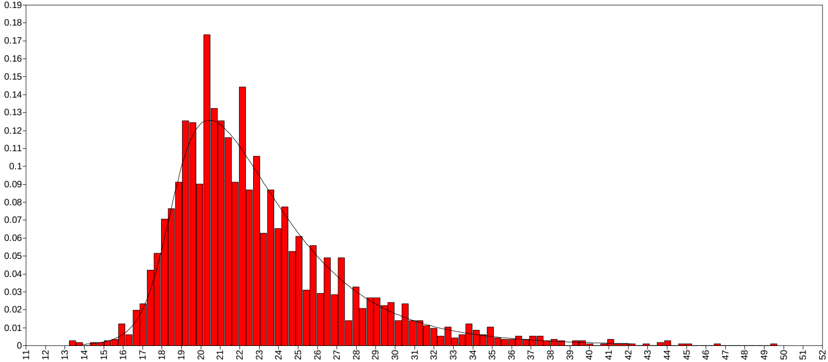
# Examples of Skew-T Distributions



# BMI at Age 15 in the NLSY (Males, $n = 3194$)

- Skewness = 1.5, kurtosis = 3.1
- Mixtures of normals with 1-4 classes: BIC $= 18,658$, $17,697$, $17,638$, $17,637$ (tiny class). 3-class model uses 8 parameters
- 1-class Skew-T distribution: BIC $= 17,623$ (2-class BIC $= 17,638$). 1-class model uses 4 parameters



# Mixture Modeling of the Australian Institute of Sports Data: BMI and BFAT ($n = 202$)

- AIS data often used in the statistics literature to illustrate quality of cluster analysis using mixtures, treating gender as unknown
- Murray, Brown & McNicholas forthcoming in Computational Statistics & Data Analysis: ”Mixtures of skew-t factor analyzers”:
  How well can we identify cluster (latent class) membership based on BMI and BFAT?
- Compared to women, men have somewhat higher BMI and somewhat lower BFAT
- Non-normal mixture models with unrestricted means, variances, covariance

Table : Comparing classes with unknown gender

Normal 2c:
LL = -1098, # par.'s = 11, BIC = 2254

| | Class 1 | Class 2 |
|---|---|---|
| Male | 78 | 24 |
| Female | 0 | 100 |

Normal 3c:
LL = -1072, # par.'s = 17, BIC = 2234

| | Class 1 | Class 2 | Class 3 |
|---|---|---|---|
| Male | 85 | 1 | 16 |
| Female | 1 | 14 | 85 |

Skew-Normal 2c:
LL = -1069, # par.'s = 15, BIC = 2218

| | Class 1 | Class 2 |
|---|---|---|
| Male | 84 | 18 |
| Female | 1 | 99 |

T 2c:
LL = -1090, # par.'s = 13, BIC = 2250

| | Class 1 | Class 2 |
|---|---|---|
| Male | 95 | 7 |
| Female | 2 | 98 |

Skew-T 2c:
LL = -1068, # par.'s = 17, BIC = 2227

| | Class 1 | Class 2 |
|---|---|---|
| Male | 95 | 7 |
| Female | 2 | 98 |

PMSTFA 2c:
BIC = 2224

| | Class 1 | Class 2 |
|---|---|---|
| Male | 97 | 5 |
| Female | 5 | 95 |

# Cluster Analysis by "Mixtures of Factor Analyzers" (McLachlan)

Reduces the number of $\mu_c$, $\Sigma_c$ parameters for $c = 1, 2, \ldots C$ by applying the $\Sigma_c$ structure of an EFA with orthogonal factors:

$$\Sigma_c = \Lambda_c \, \Lambda_c' + \Theta_c \tag{1}$$

This leads to 8 variations by letting $\Lambda_c$ and $\Theta_c$ be invariant or not across classes and letting $\Theta_c$ have equality across variables or not (McNicholas & Murphy, 2008).
Interest in clustering as opposed to the factors, e.g. for genetic applications.

# Non-Normal Mixtures of Latent Variable Models

Models:

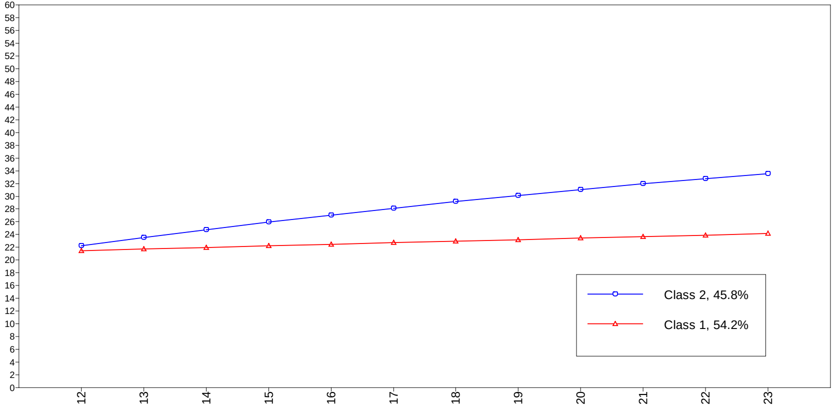
- Mixtures of Exploratory Factor Models (McLachlan, Lee, Lin; McNicholas, Murray)
- Mixtures of Confirmatory Factor Models; FMM (Mplus)
- Mixtures of SEM (Mplus)
- Mixtures of Growth Models; GMM (Mplus)

Choices:

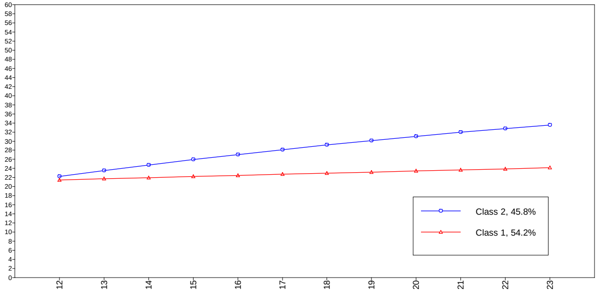
- Intercepts, slopes (loadings), and residual variances invariant?
- Scalar invariance (intercepts, loadings) allows factor means to vary across classes instead of intercepts (not typically used in mixtures of EFA, but needed for GMM)
- Skew for the observed or latent variables? Implications for the observed means. Latent skew suitable for GMM - the observed variable means are governed by the growth factor means

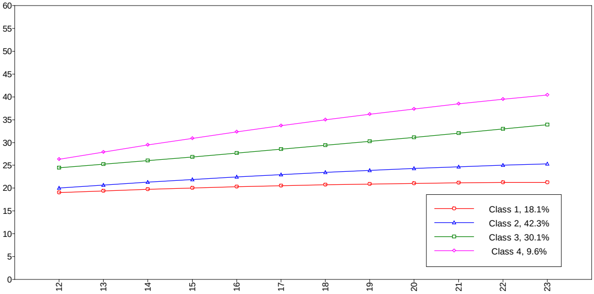
# Growth Mixture Modeling of NLSY BMI Age 12 to 23 for Black Females ($n = 1160$)

- Normal BIC: 31684 (2c), 31386 (3c), **31314 (4c)**, 31338 (5c)
- Skew-T BIC: 31411 (1c), **31225 (2c)**, 31270 (3c)

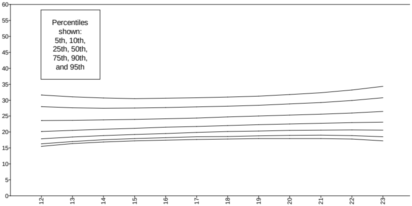


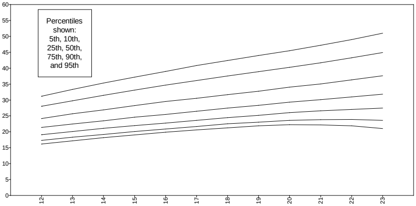
# 2-Class Skew-T versus 4-Class Normal



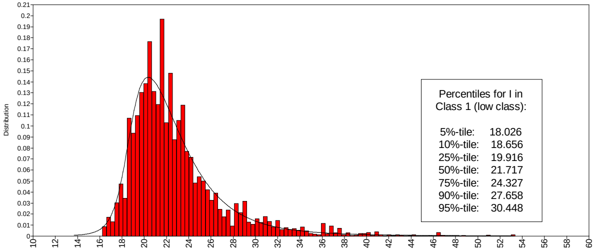



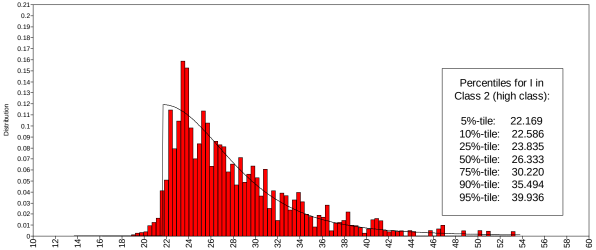
# 2-Class Skew-T: Estimated Percentiles (Note: Not Growth Curves)



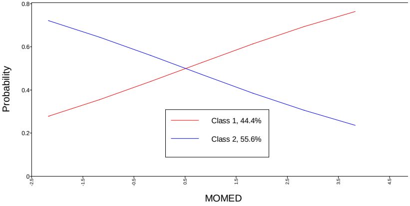



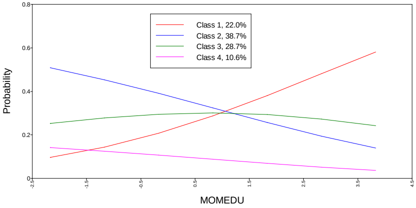
# 2-Class Skew-T: Intercept Growth Factor (Age 17)



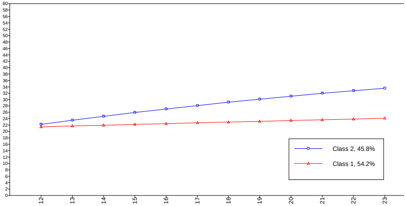



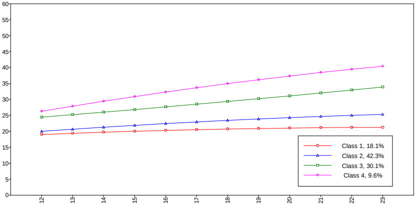
# Regressing Class on a MOMED Covariate ("c ON x"): 2-Class Skew-T versus 4-Class Normal




Recall the estimated trajectory means for skew-t versus normal:



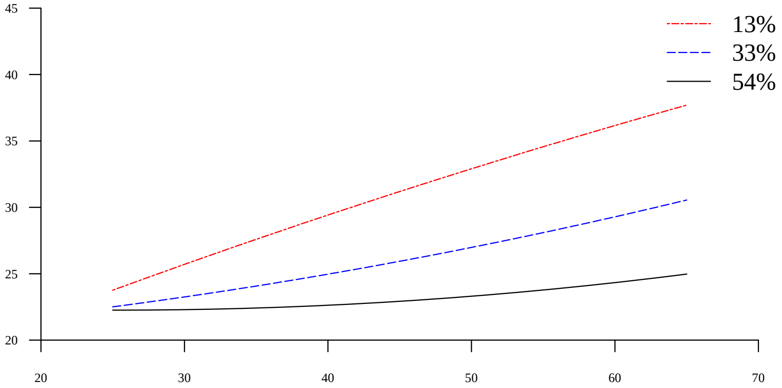
# Framingham BMI, Females Ages 25 to 65

- Classic data set
- Different age range: 25 to 65
- Individually-varying times of observations

Quadratic growth mixture model

- Normal distribution BIC is not informative:
  - 16557 (1c), 15995 (2c), 15871 (3c), 15730 (4c), 15674 (5c)
- Skew-T distribution BIC points to 3 classes:
  - 15611 (1c), 15327 (2c), **15296 (3c)**, 15304 (4c)

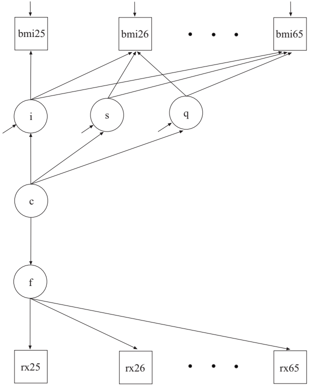
# Framingham BMI, Females Ages 25 to 65: 3-Class Skew-T



# BMI and Treatment for High Blood Pressure

- BMI = $kg/m^2$ with normal range 18.5 to 25, overweight 25 to 30, obese $> 30$
- Risk of developing heart disease, high blood pressure, stroke, diabetes
- Framingham data contains data on blood pressure treatment at each measurement occasion
- Survival component for the first treatment can be added to the growth mixture model with survival as a function of trajectory class

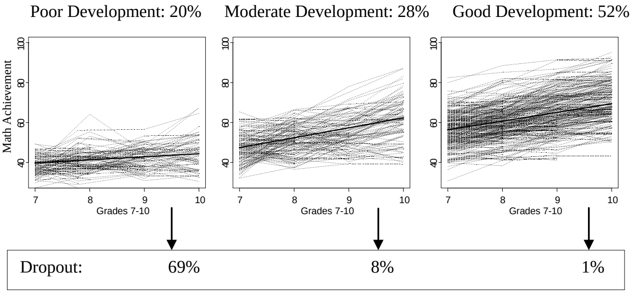
# Parallel Process Model with Growth Mixture for BMI and Discrete-Time Survival for Blood Pressure Treatment




Treatment probabilities are significantly different for the 3 trajectory classes (due to different f intercepts) and in the expected order. Probability plots can be made for each class as a function of age

- Health and Retirement Study (HRS) data
- Different age range: 65 to death
- GMM + non-ignorable dropout as a function of latent trajectory class
- Zajacova & Ailshire (2013). Body mass trajectories and mortality among older adults: A joint growth mixture-discrete-time survival model. The Gerontologist

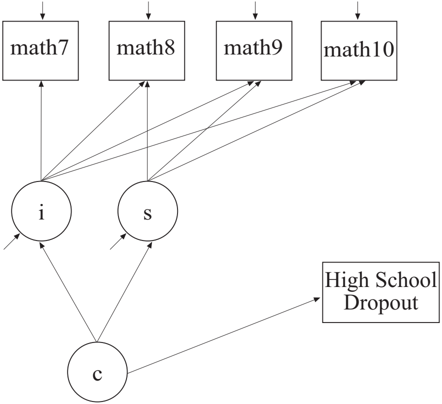
# Growth Mixture Modeling: Math Achievement Trajectory Classes and High School Dropout. An Example of Substantive Checking via Predictive Validity



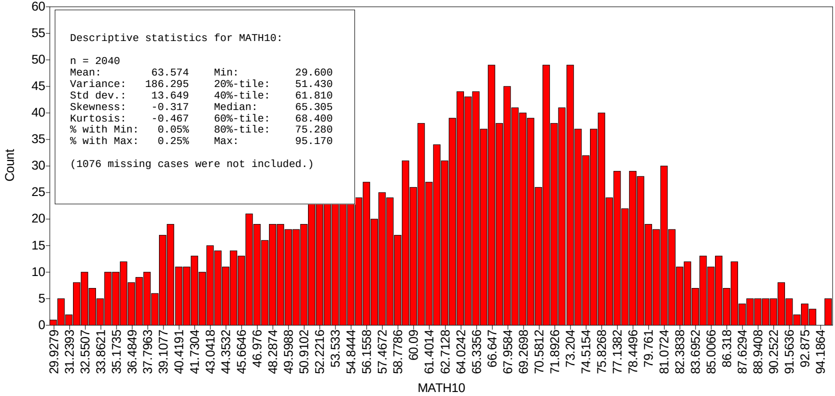

Source: Muthén (2003). Statistical and substantive checking in growth mixture modeling. Psychological Methods.

- Does the normal mixture solution hold up when checking with non-normal mixtures?

# Growth Mixture Modeling: Math Achievement Trajectory Classes and High School Dropout



# Growth Mixture Modeling: Math Achievement Trajectory Classes and High School Dropout



# Growth Mixture Modeling: Math Achievement Trajectory Classes and High School Dropout

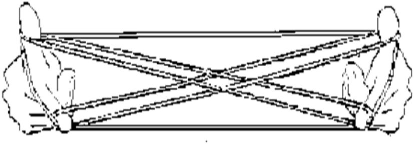
Best solutions, 3 classes (LL, no. par's, BIC):

- Normal distribution: -34459, 32, **69175**
- T distribution: -34453, 35, 69188
- Skew-normal distribution: -34442, 38, 69191
- Skew-t distribution: -34439, 42, 69207

Percent in low, flat class and odds ratios for dropout vs not, comparing low, flat class with the best class:

- Normal distribution: 18 %, OR = 17.1
- T distribution: 19 %, OR = 20.6
- Skew-normal distribution: 26 %, OR = 23.8
- Skew-t distribution: 26 %, OR = 37.3

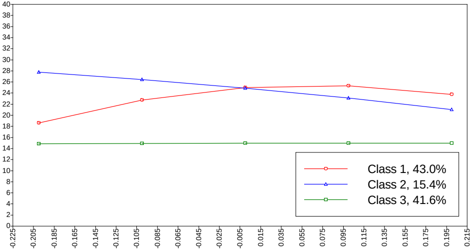
# Cat’s Cradle Concern



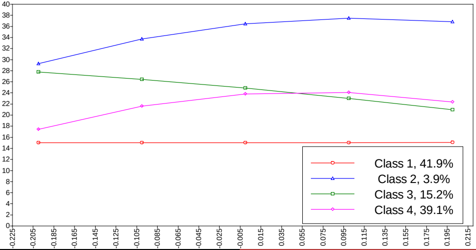
Source: Sher, Jackson, Steinley (2011). Alcohol use trajectories and the ubiquitous cat’s cradle: Cause for concern? Journal of Abnormal Psychology.

# Cat's Cradle Concern: A Simulated Case ($n = 2,000$)

Data generated by a 3-class skew-t. 3-class skew-t, BIC=43566:




4-class normal - cat's cradle with a high/chronic class, BIC=44935:



# Disadvantages of Non-Normal Mixture Modeling

- Much slower computations than normal mixtures, especially for large sample sizes
- Needs larger samples; small class sizes can create problems (but successful analyses can be done at n = 100-200)
- Needs more random starts than normal mixtures to replicate the best loglikelihood
- Lower entropy
- Needs continuous variables
- Needs continuous variables with many distinct values: Likert scales treated as continuous variables may not carry enough information
- Models requiring numerical integration not yet implemented (required with factors behind categorical and count variables, although maybe not enough information)

# Advantages of Non-Normal Mixture Modeling

Non-normal mixtures

- Can fit the data considerably better than normal mixtures
- Can use a more parsimonious model
- Can reduce the risk of extracting latent classes that are merely due to non-normality of the outcomes
- Can check the stability/reproducibility of a normal mixture solution
- Can describe the percentiles of skewed distributions

# Mplus Specifications

DISTRIBUTION=SKEWT/SKEWNORMAL/TDIST in the ANALYSIS command makes it possible to access non-normality parameters in the MODEL command

- Skew parameters are given as $\{y\}$, $\{f\}$, where the default is $\{f\}$ and class-varying. Having both $\{y\}$ and $\{f\}$ is not identified
- Degrees of freedom parameters are given as $\{df\}$ where the default is class-varying
- $df < 1$: mean not defined, $df < 2$: variance not defined, $df < 3$: skewness not defined. Density can still be obtained
- Class-varying $\{f\}$ makes it natural to specify class-varying $f$ variance
- Normal part of the distribution can get zero variances (fixed automatically), with only the non-normal part remaining

# Mplus Input Example: NLSY BMI 2-Class Skew-T GMM

```
VARIABLE:    NAMES = id gender age_1996 age_1997 race1 bmi12_2
             bmi13_2 bmi14_2 bmi15_2 bmi16_2 bmi17_2 bmi18_2 bmi19_2
             bmi20_2 bmi21_2 bmi22_2 bmi23_2 black hisp mixed c1 c2 c3
             c1_wom c2_wom c3_wom momedu par_bmi bio1_bmi bio2_bmi
             bmi_par currsmkr97 bingedrnk97 mjuse97 cent_msa
             liv2prnts adopted income hhsize97;
             USEVARIABLES = bmi12_2 bmi13_2 bmi14_2
             bmi15_2 bmi16_2 bmi17_2 bmi18_2
             bmi19_2 bmi20_2 bmi21_2 bmi22_2 bmi23_2;
             USEOBSERVATIONS = gender EQ -1;
             MISSING = ALL (9999);
             CLASSES = c(2);
ANALYSIS:    TYPE = MIXTURE;
             STARTS = 400 80;
             PROCESSORS = 8;
             DISTRIBUTION = SKEWT;
             ESTIMATOR = MLR;
```

# Mplus Input Example, Continued

```
MODEL:      %OVERALL%
            i s q |bmi12_2@-.5 bmi13_2@-.4 bmi14_2@-.3
            bmi15_2@-.2 bmi16_2@-.1 bmi17_2@0 bmi18_2@.1
            bmi19_2@.2 bmi20_2@.3 bmi21_2@.4 bmi22_2@.5 bmi23_2@.6;
            %c#1%
             i-q;
             i-q WITH i-q;
            bmi12_2-bmi23_2(1);
            %c#2%
            i-q;
            i-q WITH i-q;
            bmi12_2-bmi23_2(2);
OUTPUT:     TECH1 TECH4 TECH8 RESIDUAL;
PLOT:       TYPE = PLOT3;
            SERIES = bmi12_2-bmi23_2(s);
```

Skew and Df Parameters

| | | | | |
|---|---|---|---|---|
| **Latent Class 1** | | | | |
| I | 6.236 | 0.343 | 18.175 | 0.000 |
| S | 3.361 | 0.542 | 6.204 | 0.000 |
| Q | -2.746 | 1.399 | -1.963 | 0.050 |
| DF | 3.516 | 0.403 | 8.732 | 0.000 |
| **Latent Class 2** | | | | |
| I | 4.020 | 0.279 | 14.408 | 0.000 |
| S | -0.875 | 0.381 | -2.296 | 0.022 |
| Q | 3.399 | 1.281 | 2.653 | 0.008 |
| DF | 3.855 | 0.562 | 6.859 | 0.000 |

# Mplus Output Example, Continued

| Technical 4 Output: Estimates derived from the model for Class 1 | | | |
|---|---|---|---|
| Estimated means for the latent variables | | | |
| | I | S | Q |
| 1 | 28.138 | 10.516 | -2.567 |
| Estimated covariance matrix for the latent variables | | | |
| | I | S | Q |
| I | 48.167 | | |
| S | 25.959 | 40.531 | |
| Q | -21.212 | 32.220 | 113.186 |

# Mplus Output Example, Continued

Estimated correlation matrix
for the latent variables

| | I | S | Q |
|---|---|---|---|
| I | 1.000 | | |
| S | 0.588 | 1.000 | |
| Q | -0.287 | 0.476 | 1.000 |

Estimated skew for the
latent variables

| | I | S | Q |
|---|---|---|---|
| 1 | 6.653 | 3.437 | -1.588 |

# Single-Class (Non-Mixture) Applications with Non-Normal Distributions

Non-Normal SEM with Skew-normal, t, and skew-t distributions:

- Allowing a more general model, including non-linear conditional expectation functions
- Chi-square test of model fit
- Percentile estimation of the factor distributions

ML robustness to non-normality doesn't hold if residuals and factors are not independent or if factors don't have an unrestricted covariance matrix (Satorra, 2002).
Asparouhov-Muthén (2014):

> *There is a preconceived notion that standard structural models are sufficient as long as the standard errors of the parameter estimates are adjusted for failure of the normality assumption, but this is not really correct. Even with robust estimation the data is reduced to means and covariances. Only the standard errors of the parameter estimates extract additional information from the data. The parameter estimates themselves remain the same, i.e., the structural model is still concerned with fitting only the means and the covariances and ignoring everything else.*

$$Y = \nu + \Lambda\eta + \varepsilon$$

$$\eta = \alpha + B\eta + \Gamma X + \xi$$

where

$$(\varepsilon, \xi) \sim rMST(0, \Sigma_0, \delta, DF)$$

and

$$\Sigma_0 = \begin{pmatrix} \Theta & 0 \\ 0 & \Psi \end{pmatrix}.$$

The vector of parameters $\delta$ is of size $P + M$ and can be decomposed as $\delta = (\delta_Y, \delta_\eta)$. From the above equations we obtain the conditional distributions

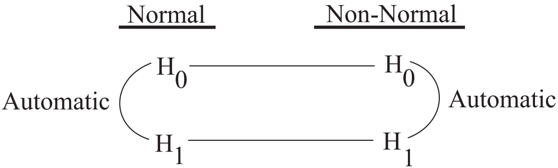
$$\eta | X \sim rMST((I-B)^{-1}(\alpha + \Gamma X), (I-B)^{-1}\Psi((I-B)^{-1})^T, (I-B)^{-1}\delta_\eta, DF)$$

$$Y | \eta \sim rMST(\nu + \Lambda\eta, \Theta, \delta_Y, DF)$$

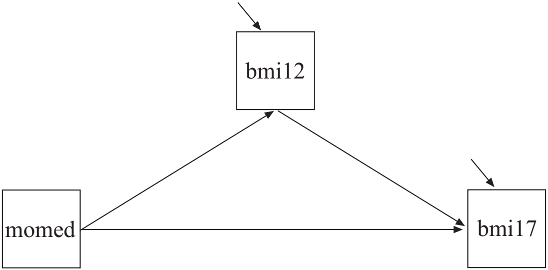
$$Y | X \sim rMST(\mu, \Sigma, \delta_2, DF)$$

Adding skew and df parameters to the means, variances, and covariances of the unrestricted H1 model

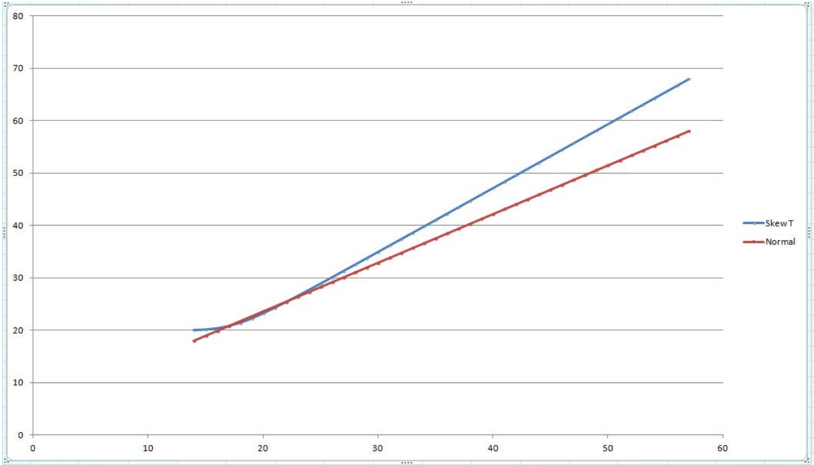
SEM Likelihood-Ratio Chi-Square Testing



# Path Analysis Model



# Regression of BMI17 on BMI12: Skew-T vs Normal



- Regular indirect and direct effects are not valid
- Modeling non-normality and non-linearity needs the more general definitions based on counterfactuals

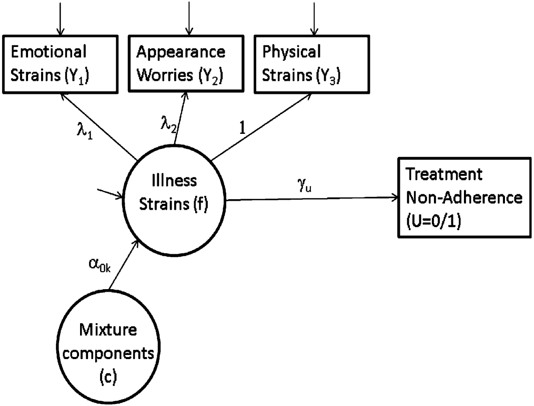
The key component of the causal effect definitions, $E[Y(x, M(x^*)|C = c, Z = z]$, can be expressed as follows integrating over the mediator $M$ (C is covariate, Z is moderator, X is "cause"):

$$E[Y(x, M(x^*)) \mid C = c, Z = z] =$$

$$\int_{-\infty}^{+\infty} E[Y|C = c, Z = z, X = x, M = m] \times f(M; E[M|C = c, Z = z, X = x^*) \, \partial M$$

Muthén & Asparouhov (2014). Causal effects in mediation modeling: An introduction with applications to latent variables. Forthcoming in Structural Equation Modeling.

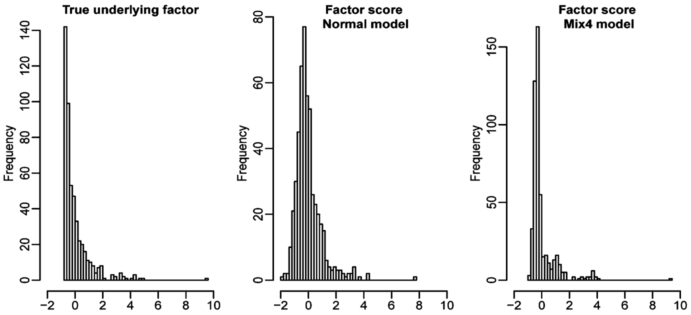
Wall, Guo, & Amemiya (2012). Mixture factor analysis for approximating a nonnormally distributed continuous latent factor with continuous and dichotomous observed variables. Multivariate Behavioral Research.




(c)

# Non-Normal Factor Distribution

Figure 6 of Wall et al. (2012):




Factor distribution can be more parsimoniously specified as skew-t than their mixture of normals. Percentiles are obtained for the estimated distribution.