Advances in Mixture Modeling And More

Bengt Muthén & Tihomir Asparouhov

Mplus
www.statmodel.com

bmuthen@statmodel.com

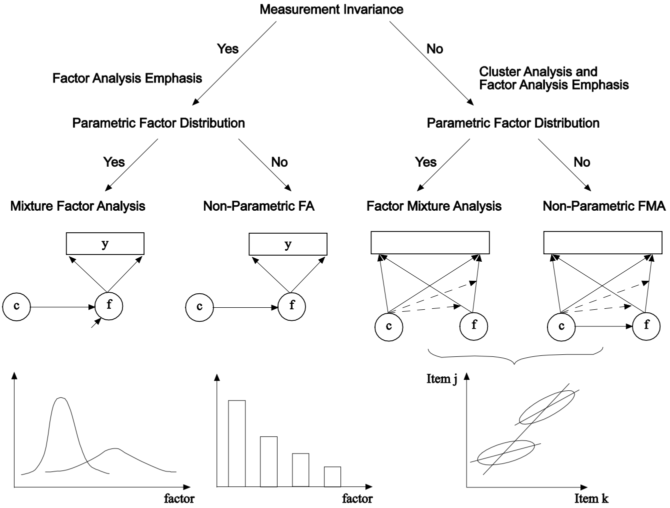
Keynote address at IMPS 2014, Madison, Wisconsin,
July 22, 2014

Overview of Latent Variable Models

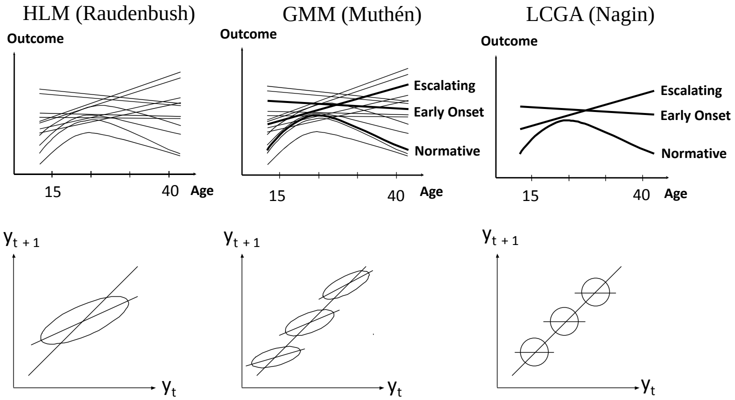
	Continuous latent variables	Categorical latent variables	Hybrids
Cross-sectional models	Factor analysis, SEM	Regression mixture analysis, Latent class analysis	Factor mixture analysis
Longitudinal models	Growth analysis (random effects)	Latent transition analysis, Latent class growth analysis	Growth mixture analysis

Source: Muthén, B. (2008). Latent variable hybrids: Overview of old and new models. In Hancock, G. R., & Samuelsen, K. M. (Eds.), Advances in latent variable mixture models, pp. 1–24. Charlotte, NC: Information Age Publishing, Inc.

Characterizing Mixture Models



Characterizing Growth (Mixture) Models



Growth Mixture Modeling (GMM) References

- Muthén, B. & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. Biometrics
- Muthén, B., Brown, C.H., Masyn, K., Jo, B., Khoo, S.T., Yang, C.C., Wang, C.P., Kellam, S., Carlin, J., & Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. Biostatistics
- Muthén, B. & Asparouhov, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In Fitzmaurice, G., Davidian, M., Verbeke, G. & Molenberghs, G. (eds.), Longitudinal Data Analysis. Boca Raton: Chapman & Hall
- Muthén & Asparouhov (2014). Growth mixture modeling with non-normal distributions. Submitted

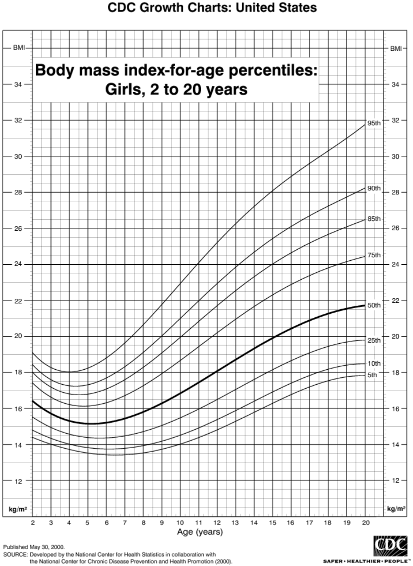
2. Today's Focus: Skewed Distributions

Examples:

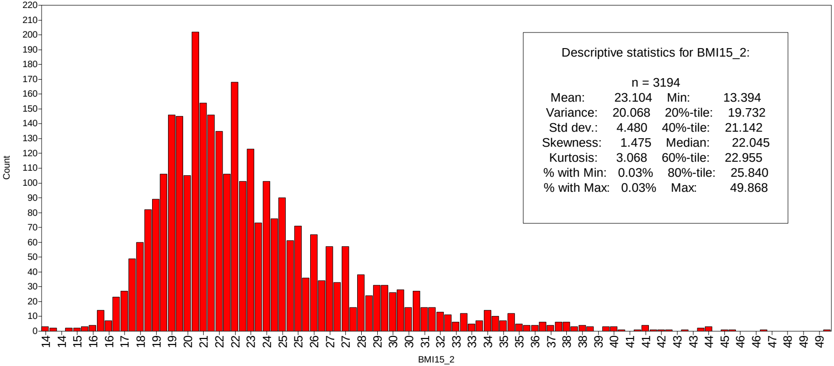
- Body Mass Index (BMI) in obesity studies (long right tail)
- Mini Mental State Examination (MMSE) cognitive test in Alzheimer's studies (long left tail)
- PSA scores in prostate cancer studies (long right tail)
- CD4 cell counts in AIDS studies (long right tail)
- Ham-D score in antidepressant studies (long right tail)

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

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

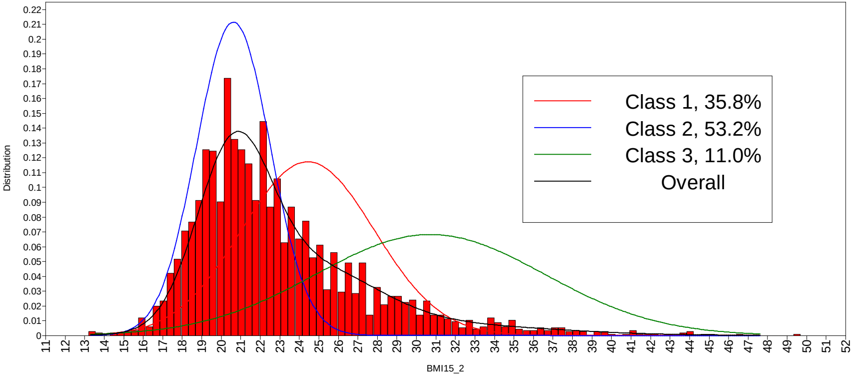


BMI at Age 15 in the NLSY (Males, $n = 3194$)



3. 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

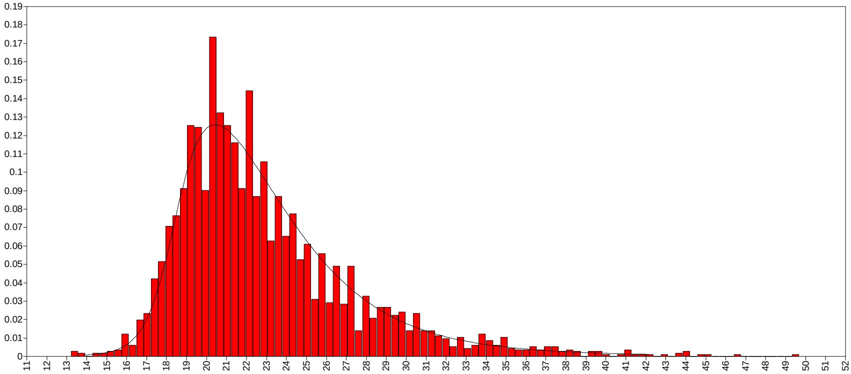


Several Classes or One Non-Normal Distribution?

- 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 & Schork (1988): 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



4. 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

- T: Adding a degree of freedom (df) parameter (thicker tails)
- Skew-normal: Adding a skew parameter to each variable
- 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 (δ is a skew vector, ν the df):

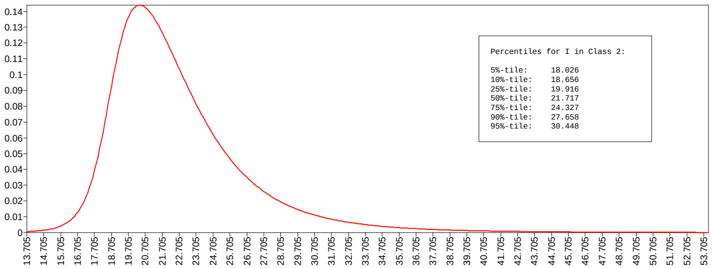
$$E(Y) = \mu + \delta \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \sqrt{\frac{\nu}{\pi}},$$

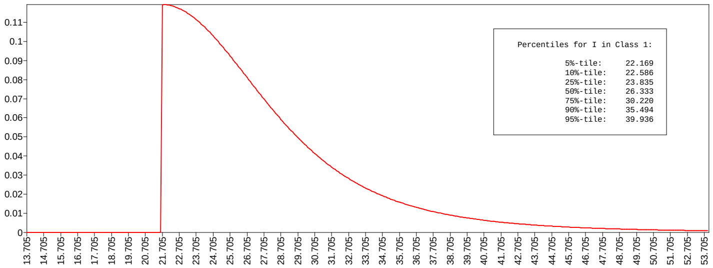
$$Var(Y) = \frac{\nu}{\nu-2}(\Sigma + \delta\delta^T) - \left(\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \right)^2 \frac{\nu}{\pi} \delta\delta^T$$

Marginal and conditional distributions:

- Marginal is also a skew-t distribution
- Conditional is an extended skew-t distribution; conditional expectation function not always linear; conditional variance not always homoscedastic

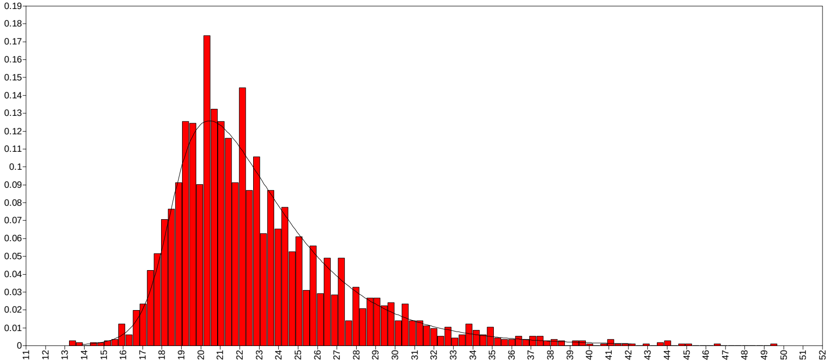
Examples of Skew-T Distributions





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



4.1 Cluster Analysis with Non-Normal Mixtures
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 μ_c, Σ_c parameters for $c = 1, 2, \ldots C$ by applying the Σ_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 Λ_c and Θ_c be invariant or not across classes and letting Θ_c have equality across variables or not (McNicholas & Murphy, 2008).
Interest in clustering as opposed to the factors, e.g. for genetic applications.
(EFA mixtures not yet available in Mplus for non-normal distributions, but can be done using "EFA-in-CFA".)

Models:

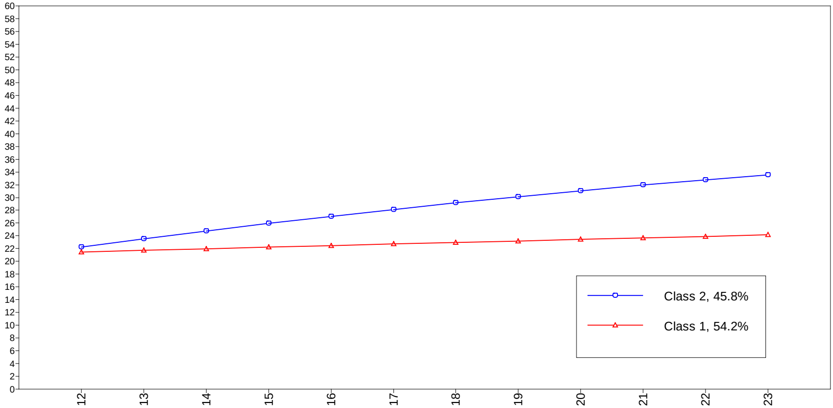
- Mixtures of Exploratory Factor Models (McLachlan, Lee, Lin; McNicholas, Murray)
- Mixtures of Confirmatory Factor Models; FMM (Mplus)
- Mixtures of SEM (Mplus)
- Mixtures of Growth Models; GMM (Mplus)

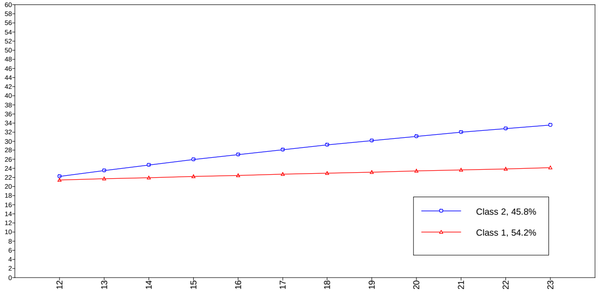
Choices:

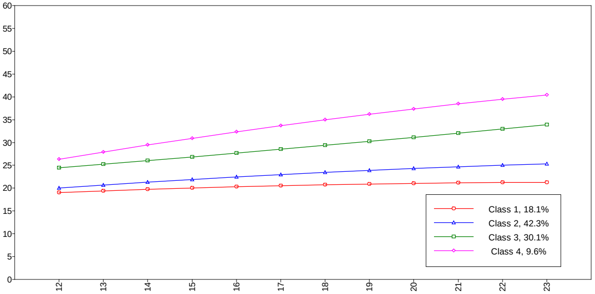
- Intercepts, slopes (loadings), and residual variances invariant?
- Scalar invariance (intercepts, loadings) allows factor means to vary across classes instead of intercepts (not typically used in mixtures of EFA, but needed for GMM)
- Skew for the observed or latent variables? Implications for the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

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

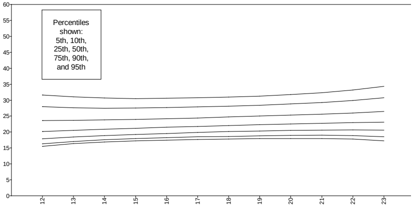
- Normal BIC: 31684 (2c), 31386 (3c), **31314 (4c)**, 31338 (5c)
- Skew-T BIC: 31411 (1c), **31225 (2c)**, 31270 (3c)

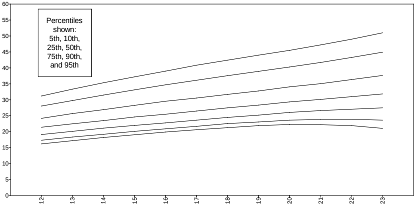



Class 2, 45.8%
Class 1, 54.2%

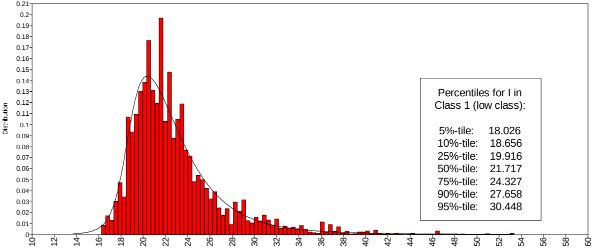

Class 1, 18.1%
Class 2, 42.3%
Class 3, 30.1%
Class 4, 9.6%

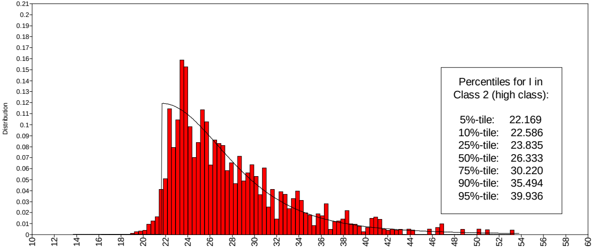
2-Class Skew-T: Estimated Percentiles (Note: Not Growth Curves)



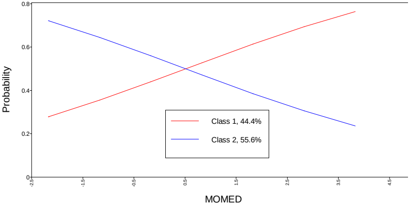


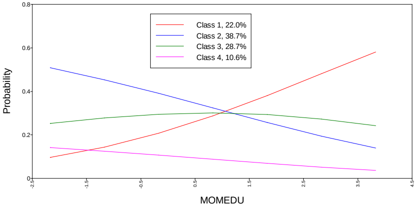
2-Class Skew-T: Intercept Growth Factor (Age 17)



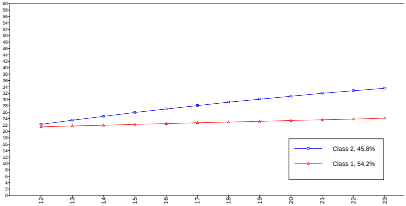


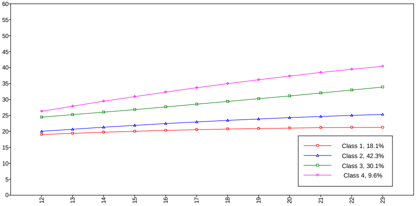
Regressing Class on a MOMED Covariate ("c ON x"): 2-Class Skew-T versus 4-Class Normal





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





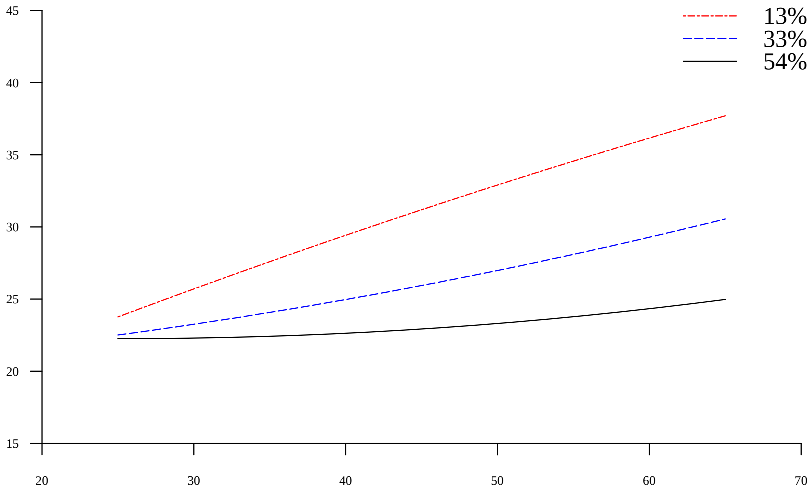
4.4 GMM of BMI in the Framingham Data: Females Ages 25 to 65

- 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)

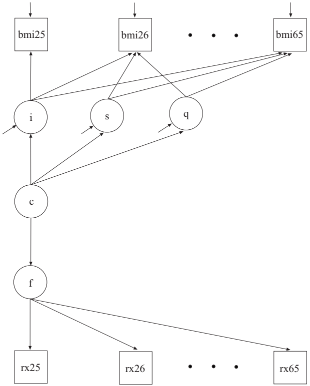
Framingham BMI, Females Ages 25 to 65: 3-Class Skew-T



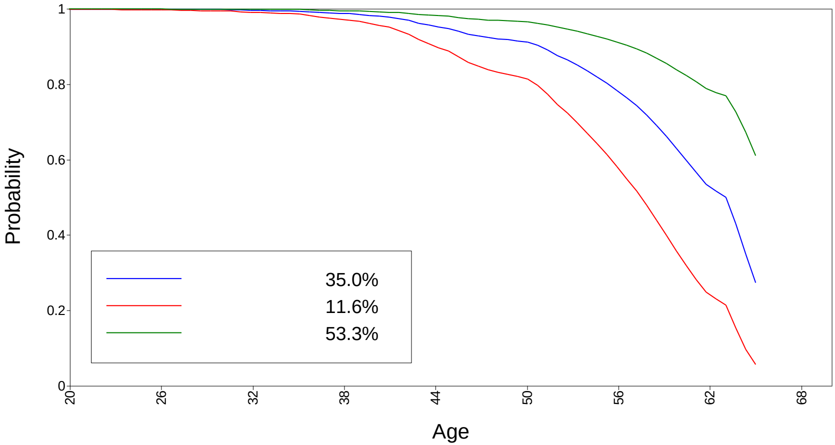
BMI and Treatment for High Blood Pressure

- 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

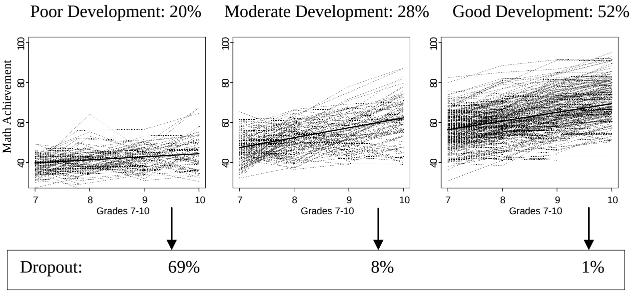
Parallel Process Model with Growth Mixture for BMI and Survival to First Blood Pressure Treatment



Parallel Process Model with Growth Mixture for BMI and Survival to First Blood Pressure Treatment: Survival Curves for the Three Trajectory Classes



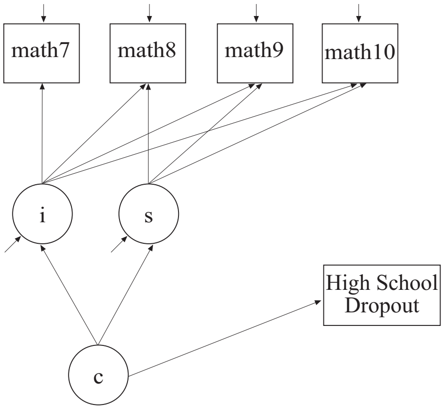
4.5 Growth Mixture Modeling: Math Achievement Trajectory Classes and High School Dropout. An Example of Substantive Checking via Predictive Validity



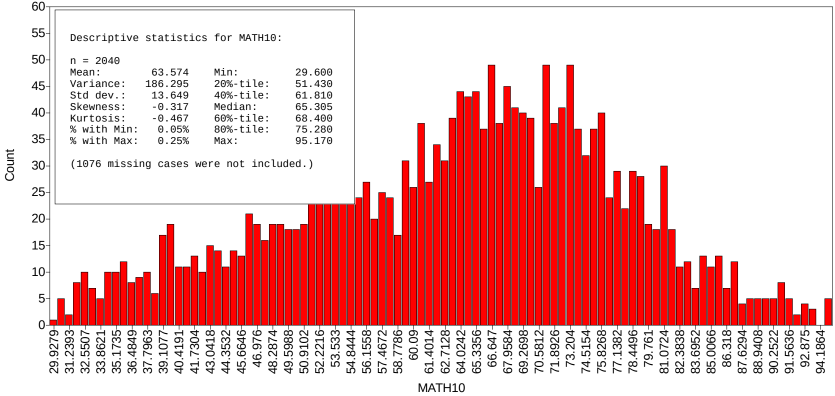
Source: Muthén (2003). Statistical and substantive checking in growth mixture modeling. Psychological Methods.

- 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

4.7 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

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

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

- Allowing a more general model, including non-linear conditional expectation functions and heteroscedasticity
- Chi-square test of model fit using information on skew and df
- Missing data handling avoiding the normality assumption of FIML
- Mediation modeling allowing general direct and indirect effects
- Percentile estimation of the skewed factor distributions

Asparouhov & Muthén (2014). Structural equation models and mixture models with continuous non-normal skewed distributions. Forthcoming in Structural Equation Modeling.

ML Estimation: Normality Violations in SEM

- ML estimate robustness to non-normality in SEM
- SEs and chi-square can be adjusted for non-normality (sandwich estimator, "Satorra-Bentler")
- MLE robustness doesn't hold if residuals and factors are not independent (Satorra, 2002) as with residual skew $\neq 0$
- 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

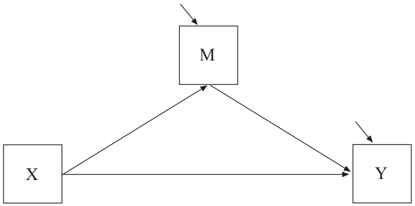
$$\Sigma_0 = \begin{pmatrix} \Theta & 0 \\ 0 & \Psi \end{pmatrix}.$$

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

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

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

5.1 Path Analysis of Firefighter Data ($n = 354$)

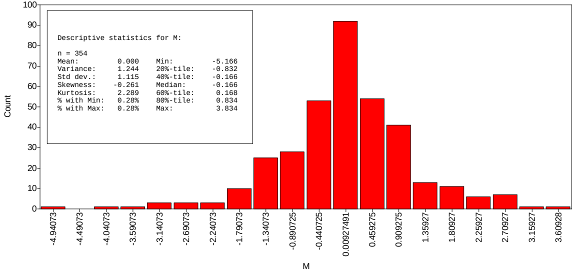


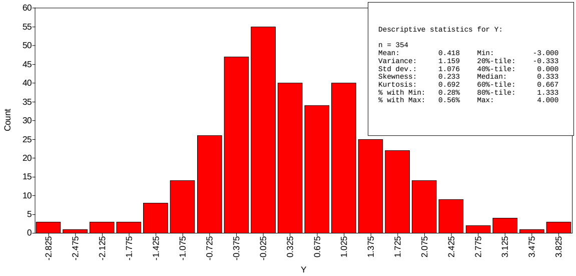
- X represents randomized exposure to an intervention
- M is knowledge of the benefits of eating fruits and vegetables
- Y is reported eating of fruits and vegetables

Sources:

- Yuan & MacKinnon (2009). Bayesian mediation analysis. Psychological Methods
- Elliot et al. (2007). The PHLAME (Promoting Healthy Lifestyles: Alternative Models Effects) firefighter study: outcomes of two models of behavior change. Journal of Occupational and Environental Medicine

Firefighter M and Y Distributions





Summary Comparison of Models for the Firefighter Example

Distribution	LogL	Number of parameters	BIC
Normal	-1058	7	2157
t-dist	-1045	8 (adding df)	**2137**
Skew-normal	-1055	9 (adding 2 skew)	2162
Skew-t	-1043	10 (adding df and 2 skew)	2144

- The df parameter of the t-distribution is needed to capture the kurtosis, but skew parameters are not needed
- The t-distribution allows for heteroscedasticity in the Y residual as a function of M; the conditional expectation functions are linear; usual indirect effect valid

Comparing Normal and T-Distribution Estimates

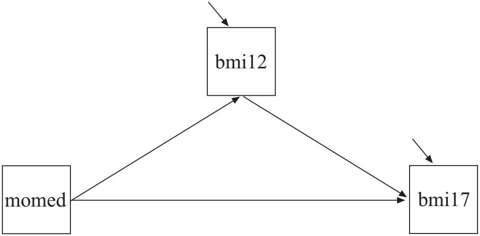
Normal distribution - regular SEM

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
M ON				
X	0.397	0.119	3.346	0.001
Y ON				
M	0.142	0.051	2.755	0.006
X	0.108	0.116	0.926	0.354
Intercepts				
Y	0.418	0.056	7.417	0.000
M	0.000	0.058	0.000	1.000
Residual Variances				
Y	1.125	0.085	13.304	0.000
M	1.203	0.090	13.304	0.000
New/Additional Parameters				
INDIRECT	0.056	0.026	**2.127**	0.033

T-distribution

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
M ON				
X	0.371	0.110	3.384	0.001
Y ON				
M	0.119	0.059	2.003	0.045
X	0.134	0.115	1.161	0.246
Intercepts				
Y	0.384	0.055	6.940	0.000
M	0.005	0.053	0.093	0.926
Residual Variances				
Y	0.872	0.088	9.963	0.000
M	0.829	0.092	9.006	0.000
Skew and Df Parameters				
DF	7.248	1.851	3.915	0.000
New/Additional Parameters				
INDIRECT	0.044	0.026	**1.718**	0.086

5.2 Path Analysis Mediation Model for BMI ($n = 3839$)

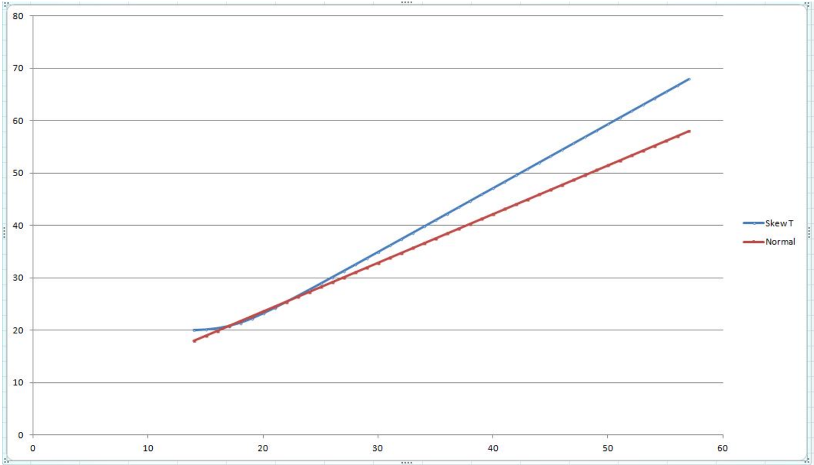


bmi12 skewness/kurtosis = 1.34/2.77
bmi17 skewness/kurtosis = 1.86/5.29

Summary Comparison of Models for the BMI Example

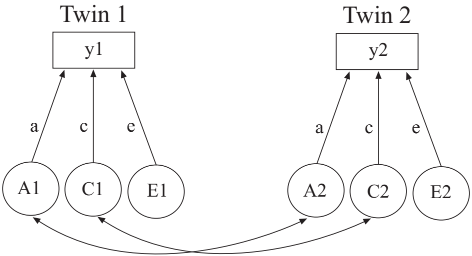
Distribution	LogL	Number of parameters	BIC
Normal	-11207	7	22471
t-dist	-10789	8 (adding df)	21642
Skew-normal	-10611	9 (adding 2 skew)	21295
Skew-t	-10423	10 (adding df and 2 skew)	**20928**

Regression of BMI17 on BMI12: Skew-T vs Normal



Skew-T versus Normal Results

- The skew-t model can be further improved by letting the BMI skew be a function of mother's education
- The modified skew-t model shows that the direct effect is 85% of the total effect while the indirect effect is only 15% of the total effect
- The normal model finds that the direct effect is ignorable relative to the indirect effect
- These drastically different results illustrate the modeling opportunities when we look beyond the mean and variance modeling used with standard SEM



- How will heritability estimates be affected?
- For which variables should the skew parameters be applied?
- How should growth mixtures be incorporated: Different heritability for different trajectory classes; is there genetic influence on trajectory class membership?

And More

Bengt Muthén & Tihomir Asparouhov
Mplus
www.statmodel.com
bmuthen@statmodel.com