Session S:

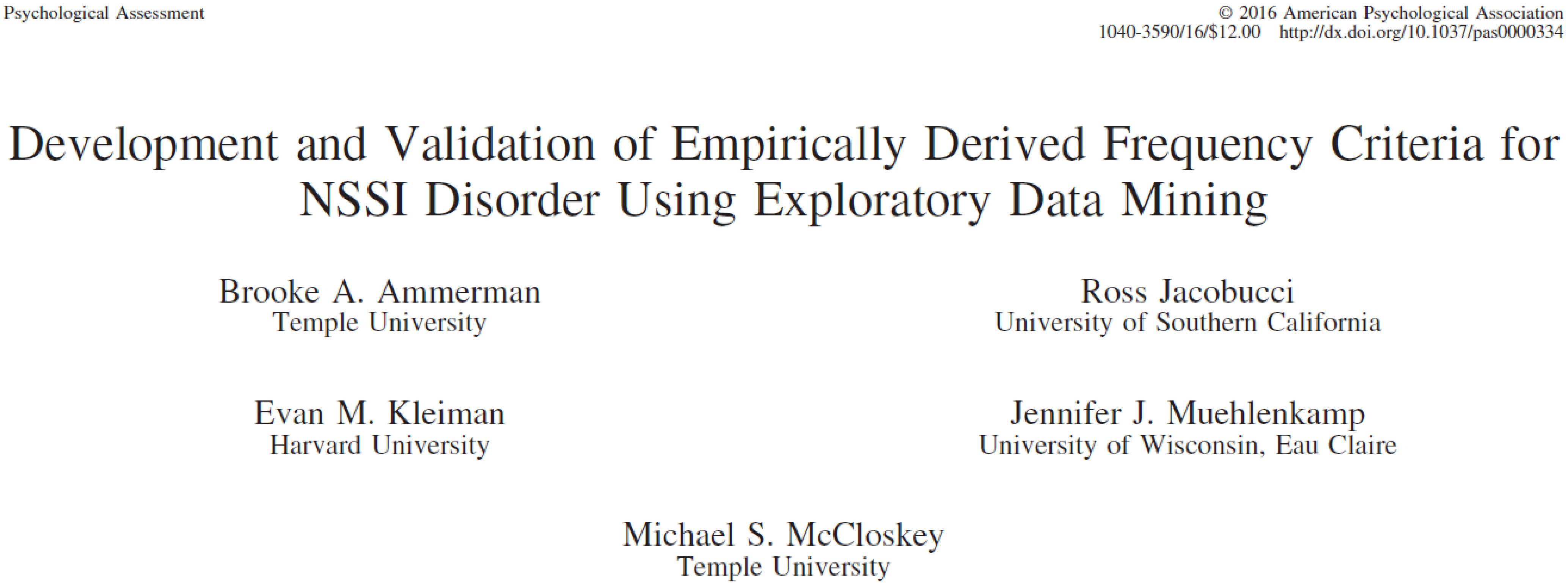
Lecture – SEM Trees

ATI Staff

Arizona State University

Summer 2016

Motivating Example



Psychological Assessment

© 2016 American Psychological Association
1040-3590/16/$12.00 http://dx.doi.org/10.1037/pas0000334

Development and Validation of Empirically Derived Frequency Criteria for NSSI Disorder Using Exploratory Data Mining

Brooke A. Ammerman
Temple University

Ross Jacobucci
University of Southern California

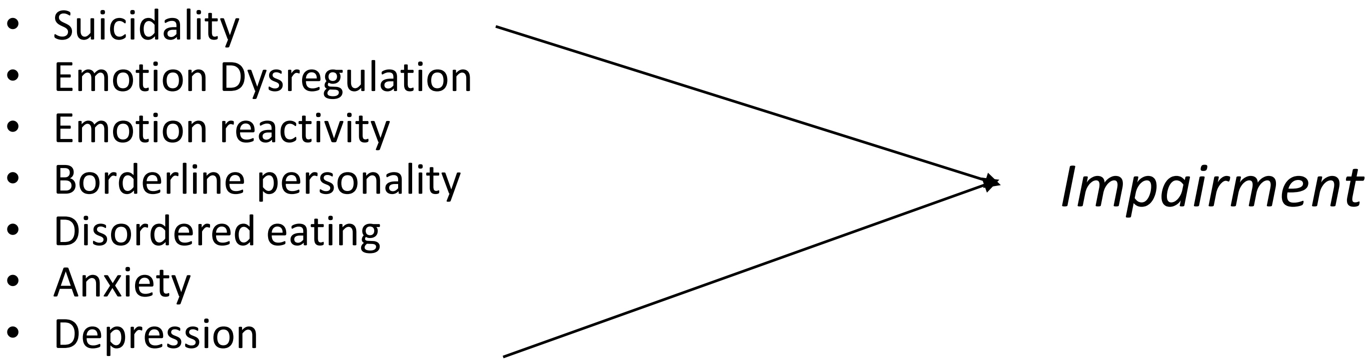
Evan M. Kleiman
Harvard University

Jennifer J. Muehlenkamp
University of Wisconsin, Eau Claire

Michael S. McCloskey
Temple University

Goal of Analysis

- To find clinically meaningful sub groups of non-suicidal self-injury (NSSI)
  - Cutoffs for DSM-V Criteria
  - Current cutoff = 5 NSSI acts in past year
    - Workgroup for disorder stated this is an “arbitrary cutoff”
- Have measure of how many times participants self-injured in last year
  - “Have you ever, intentionally or on purpose, hurt yourself in the following ways, without the intention of killing yourself?”
- Outcome
  - One-factor Factor Analysis consisting of:



    - Suicidality
    - Emotion Dysregulation
    - Emotion reactivity
    - Borderline personality
    - Disordered eating
    - Anxiety
    - Depression

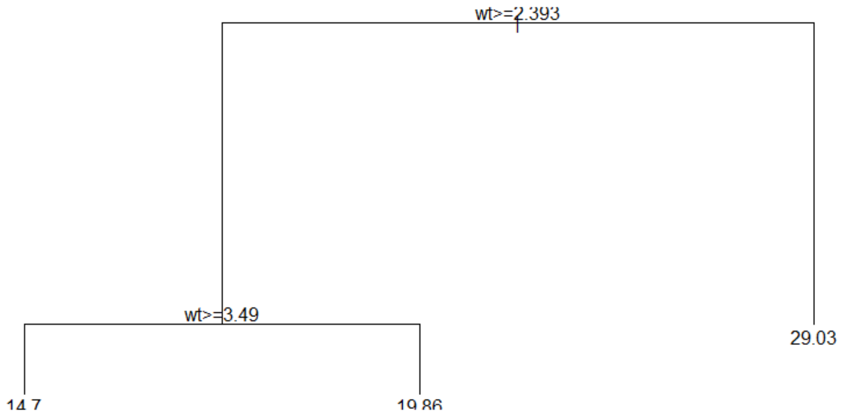
Impairment

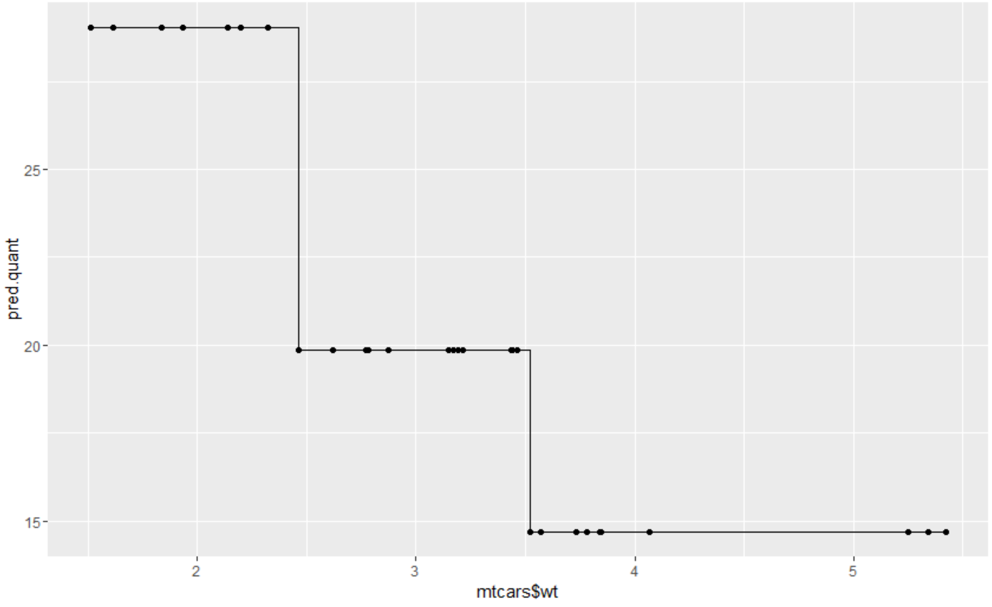
Decision Trees for Groups

Decision Trees for Groups

- As mentioned in previous presentations, Decision Trees (DTs) are non-linear models for predicting categorical or continuous outcomes

```
library(rpart);library(ggplot2); quant.out <- rpart(mpg ~ wt, data=mtcars); plot(quant.out);text(quant.out)

pred.quant <- predict(quant.out); qplot(mtcars$wt,pred.quant,geom=c("step","point"))
```





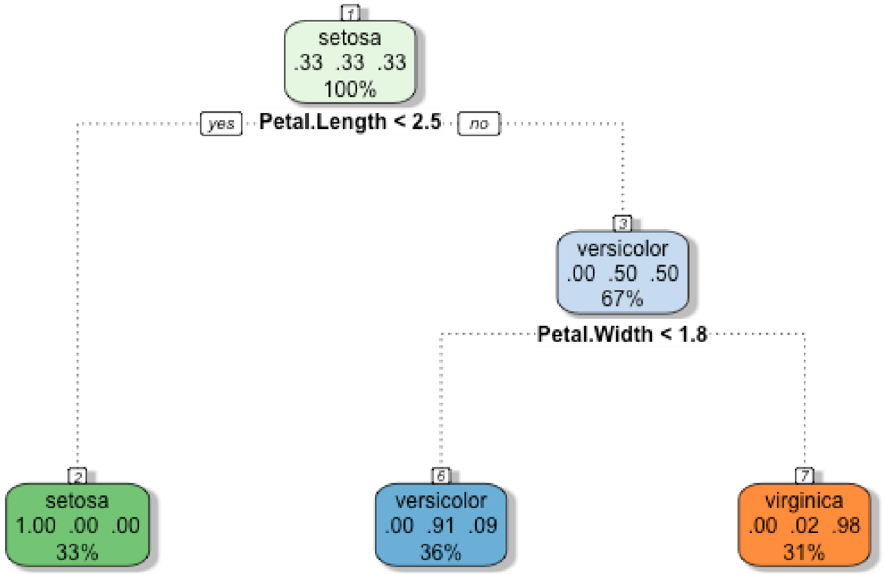
DTs for Groups Cont’d

- Most applications in Data Mining don’t involve people
  - The stepwise prediction for DT applies to groups as well
- Variable(s) you want to create groups with:
  - Predictor (one or more; categorical or continuous)
    - Categorical: Testing whether the levels of the variable are significantly different
    - Continuous: Looking for groups of people based on values of variable(s)
- Variable(s) you care about relationship with groups
  - Outcomes (one or more; categorical or continuous)

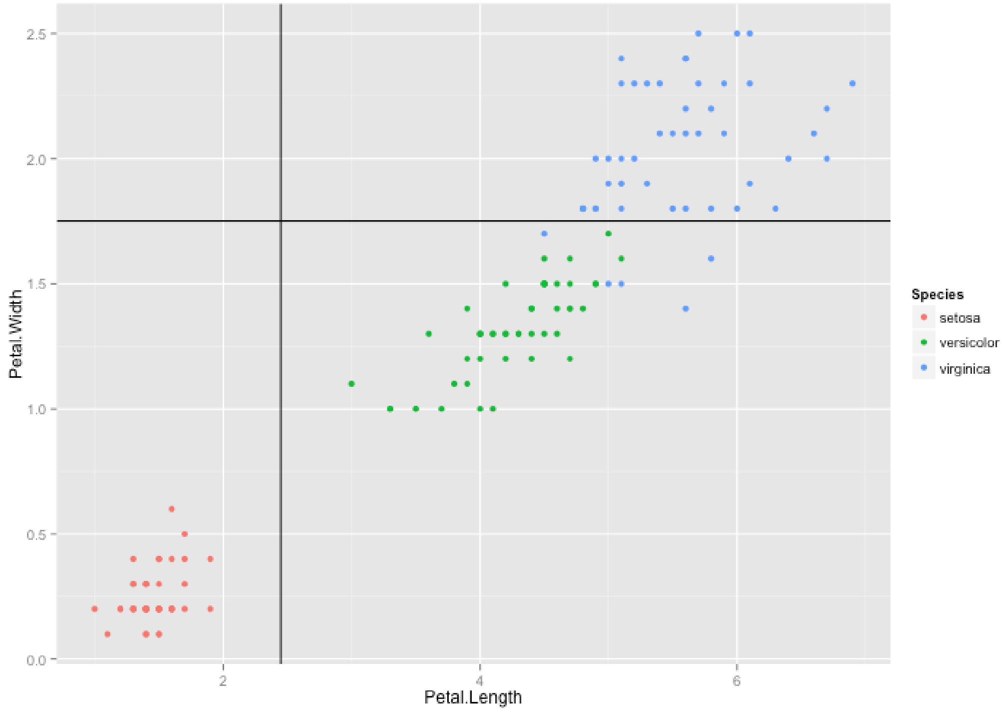
DT for Groups: Traditional Way

- Not people, but flowers!
- **Petals = Predictor**

```
library(rpart);library(ggplot2);library(rattle);attach(iris)
fit2 <- rpart(Species ~ Petal.Length + Petal.Width, data =
iris,control=rpart.control(maxdepth=2))
fancyRpartPlot(fit2)
p <- qplot(Petal.Length,Petal.Width,colour=Species); q <- p +
geom_vline(xintercept=2.45)
q + geom_abline(intercept=1.75,slope=0)
```



Rattle 2014-Dec-04 11:03:19 RJacobucci

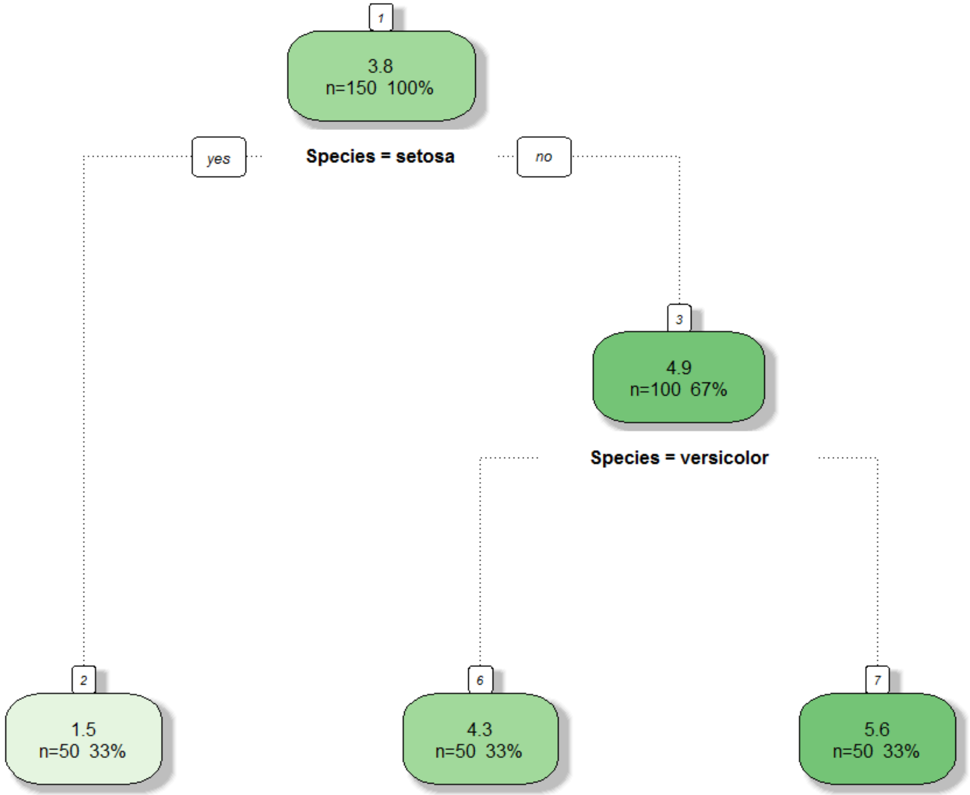


Now, instead of predicting group members, we are searching for groups *within* the predictors.

DT for Groups: Switch Predictors & Outcomes

- **Flower = Predictor**
- Test whether the groups are significantly different

```
library(rpart);library(ggplot2);library(rattle);attach(iris)
fit2 <- rpart(Petal.Length ~ Species, data =
iris,control=rpart.control(maxdepth=2))
```



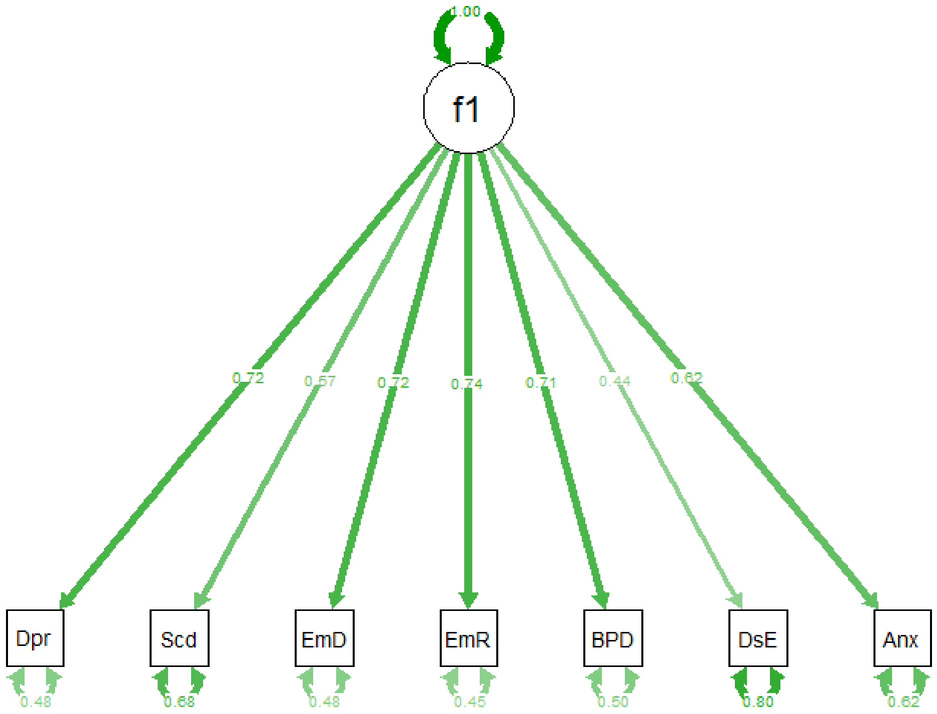
Rattle 2016-Jun-03 06:59:13 RJacobucci

DT for Groups Cont’d

- Many times we have variables that we might want to create subgroups based on
  - Income → find cutoffs for low, middle, high SES groups
  - Have continuous test scores and find cutoff for admission criteria
    - i.e. only accept those > 160 on GRE
- Which group has biggest differences on outcome?
  - Gender or Ethnicity a more “important” grouping in relation to supporting Clinton or Trump?
- Want to find interactions between groups to create further subgroupings
  - Females in low SES prefer Trump while females in high SES prefer Hillary
- Often times limited by only using 1 variable as an outcome
  - Where multivariate DT methods come in
    - E.g. multivariate boosting or SEM Trees

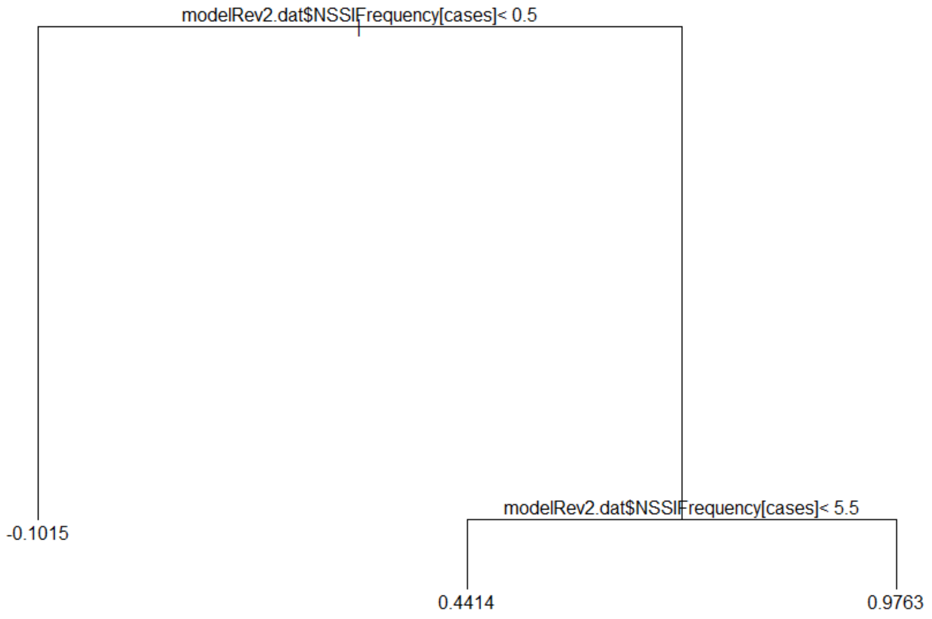
DTs and NSSI

- We want to find subgroups of people based on the number of times they self-injures
- In relation to:
  - A summed score
  - A factor score

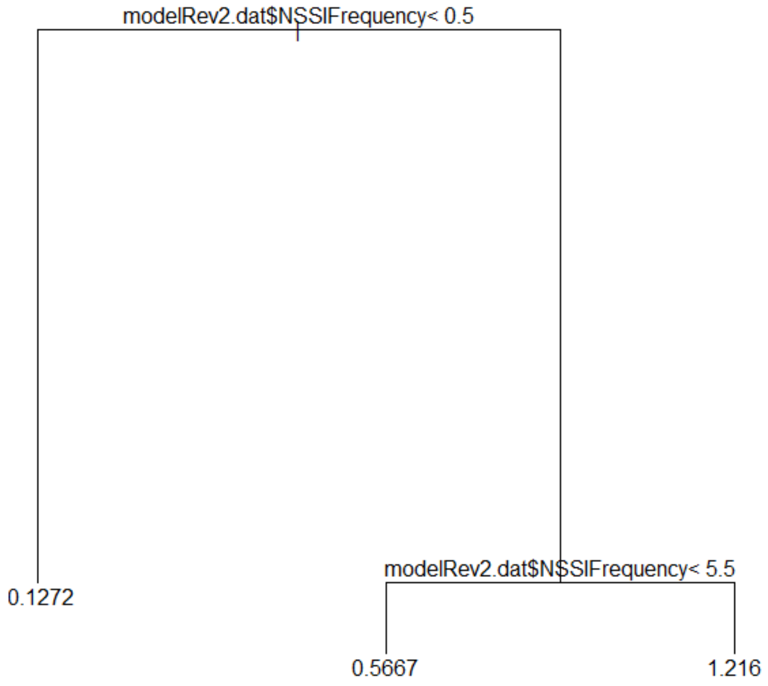


DTs and NSSI cont’d

NSSI → Summed Score


modelRev2.dat$NSSIFrequency[cases]< 0.5
modelRev2.dat$NSSIFrequency[cases]< 5.5
-0.1015
0.4414
0.9763

NSSI → Factor Score


modelRev2.dat$NSSIFrequency< 0.5
modelRev2.dat$NSSIFrequency< 5.5
0.1272
0.5667
1.216

So What Did We Learn?

- Cutoffs for NSSI Frequency Criteria should be:
  - Between 0 & 1
  - Between 5 & 6
- The DT results give us three groups with those cutoffs
- But what does this actually mean?
  - People w/ NSSI value of 0 have a deviation between their actual and predicted factor score that is significantly smaller than between their actual scores and predictions for any other group, including assuming every is in one group
  - More technically, putting people into groups decreases the RMSEA

SEM Trees & NSSI

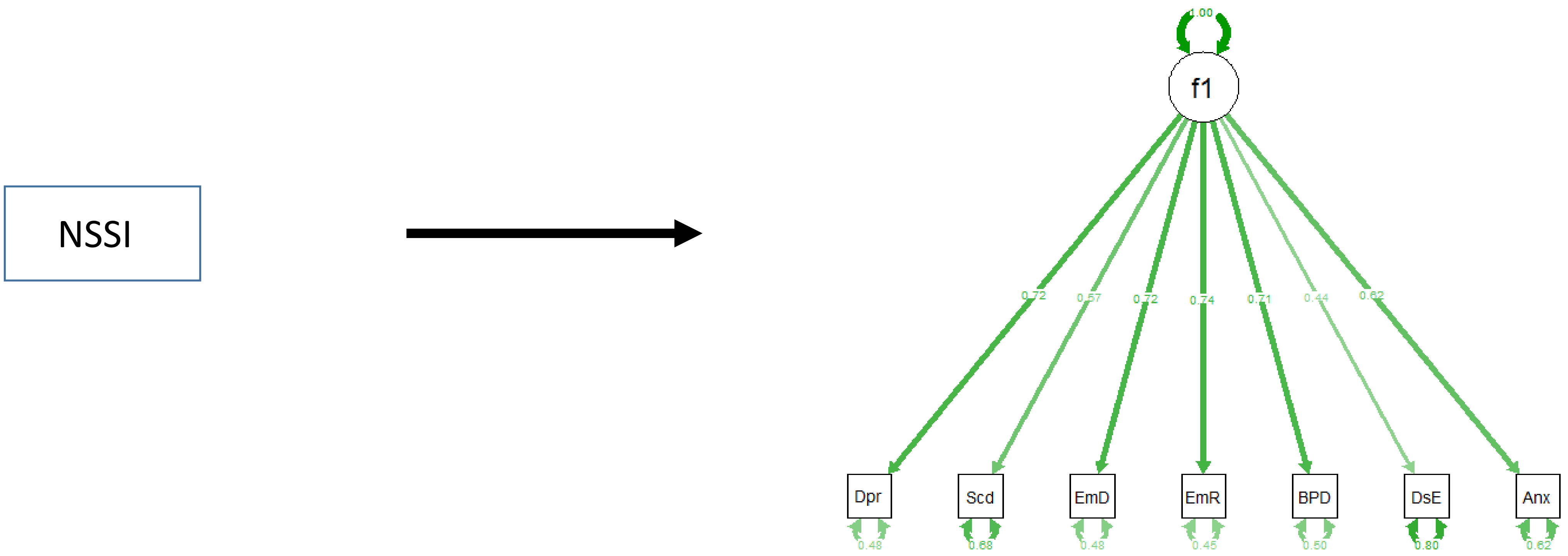
When We Can’t Reduce the Outcome to a Single Value

- Need SEM Trees or other multivariate model (Multivariate Boosting or others)
- Instead of reducing our outcome to a single variable we can use:
    - Means of multiple variables
    - A confirmatory factor model
    - Latent growth model
    - Autoregressive model
    - Etc.

SEM Trees

Predictor

Outcome



SEM Tree Cont’d

- Different than just including a covariate in the model
- In SEM Trees, the covariates predict the *model fit*
    - Not just the latent variable
- In predicting the fit of the model, you are indirectly predicting differences in each of the model parameters
    - i.e. factor variance, mean, loadings ….
    - In other words: If you change the model parameters then you change the model fit

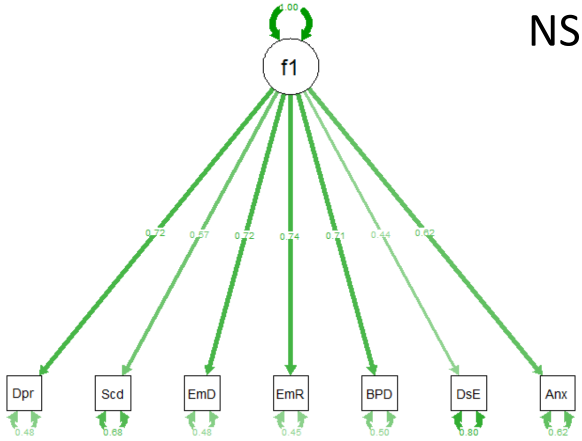
SEM Trees Algorithm

- In our NSSI example, the NSSI variable has integer values ranging from 0 to 1000
- If a covariate is an integer in SEM Trees, the model tests groups at every value:
  - 1. Fit of multiple group model with groups of people with values of 0 and 1-1000
    - Get model fit
  - 1. Fit of multiple group model with groups of people with values of 0 or 1 and 2-1000
    - Get model fit

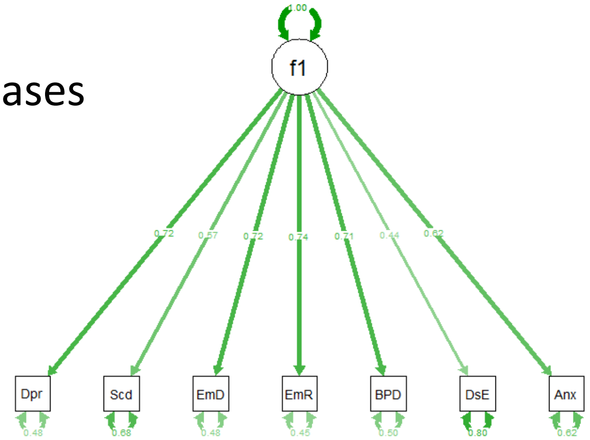
SEM Trees Algorithm Cont’d

- At each tested split, the model becomes

People w/
NSSI = 0

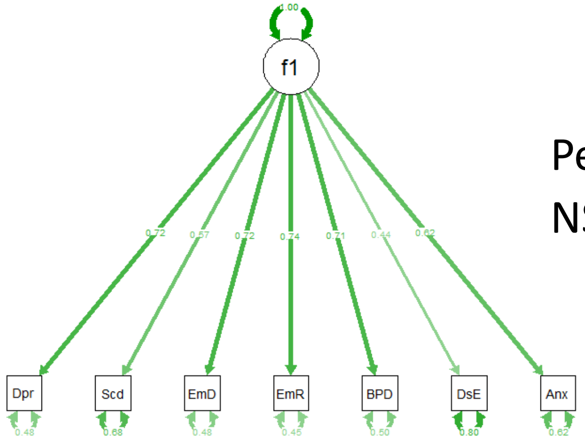


All cases



VERSUS

+



People w/
NSSI > 0

SEM Trees Algorithm Cont’d

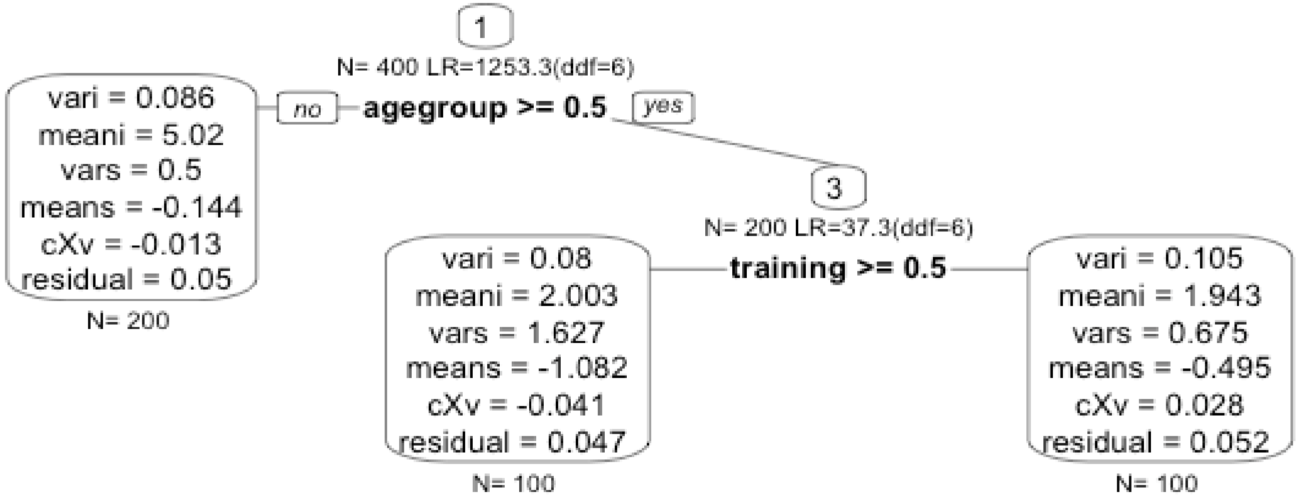
- Fitting the SEM in each group results in a separate fit
  - LL (log-likelihood) for each model
  - Model parameters are now group specific
- This can be compared using the likelihood ratio test
  - Is $LL_{all} <^{*} LL_{NSSI = 0} + LL_{NSSI = 1\text{-}1000}$
    - * Fit will always be better in multiple group model, but about <u>significance</u>
    - Larger LL is better (less negative)
    - 2 times the difference between models is a chi square distribution
      - Df = number of parameters of multiple group model minus parameters of no group model

SEM Trees Algorithm Cont’d

- If the covariate is an integer (& ordinal), tests every possible group in sequence
    - 0 vs. 1-1000, 0-1 vs 2-1000, 0-2 vs 3-1000 etc…
- Same thing if numeric (continuous; i.e. 0.324)
    - Will test every value in sequence, so may be better to round first
- If categorical (factor), one vs. the rest scheme
    - If 5 categories, 15 possible splits
        - Computationally intensive
- Once a best split is determined:
    - Move down one level to start over searching for an additional split
    - Continues until:
        - No longer improves model fit
        - Reaches other stopping criterion
            - Too small N in a node
            - A priori set maximum depth (# of groups)

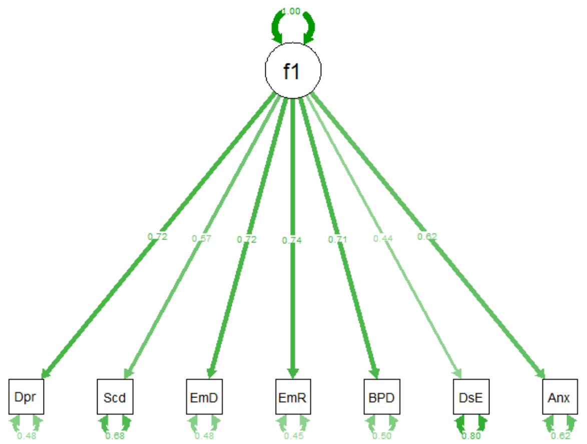
SEM Trees Algorithm Cont’d

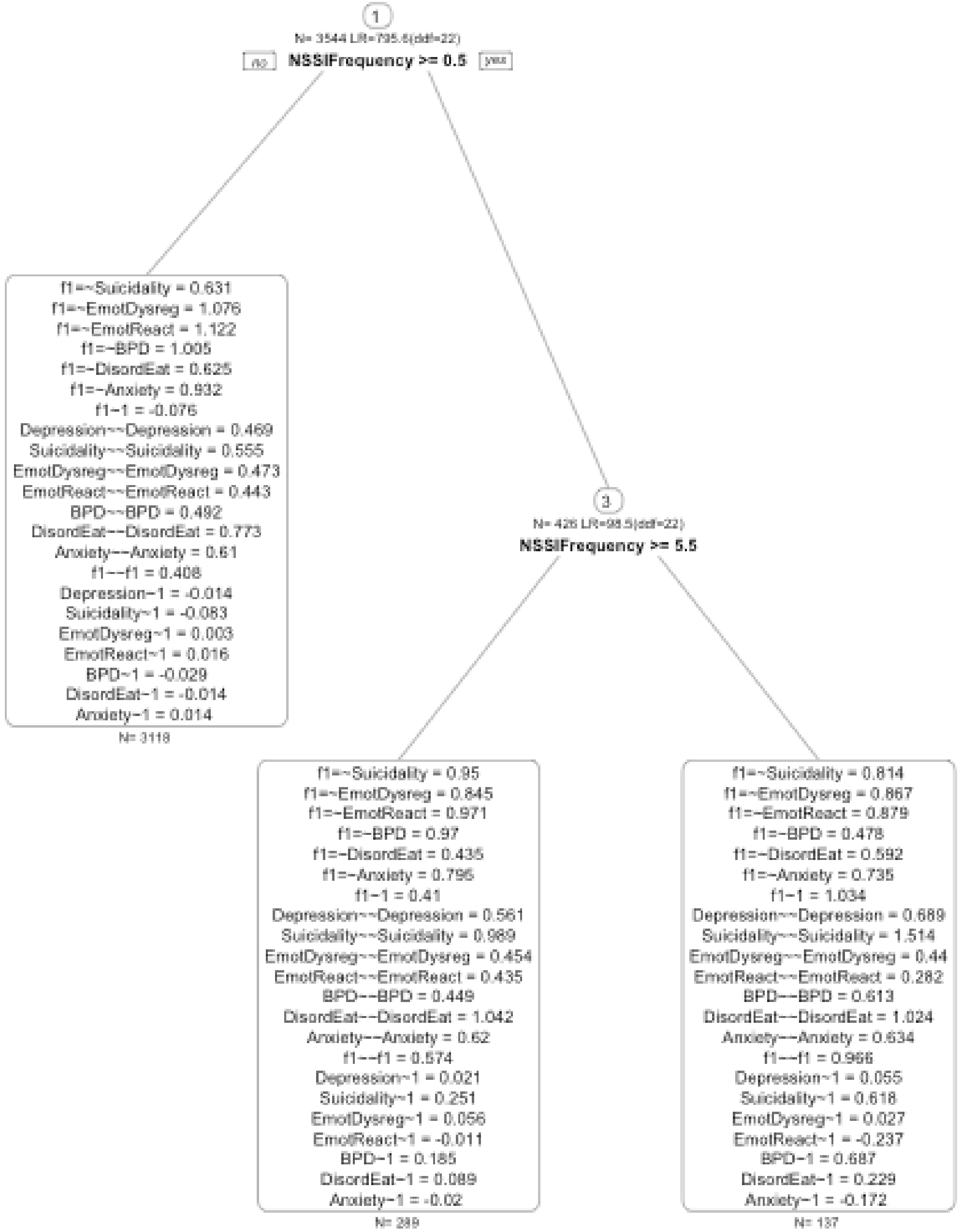
- Just like DT, SEM Trees can create a large tree
- Say the first split is between 0 & 1
- Next level, test splits on other variables
- Example:
  - Second split occurs at agegroup = 1 & between training 0-1
  - No further split for agegroup = 0



SEM Trees & NSSI

- Use a one factor CFA as the outcome
- Came up with the same splits as with DT

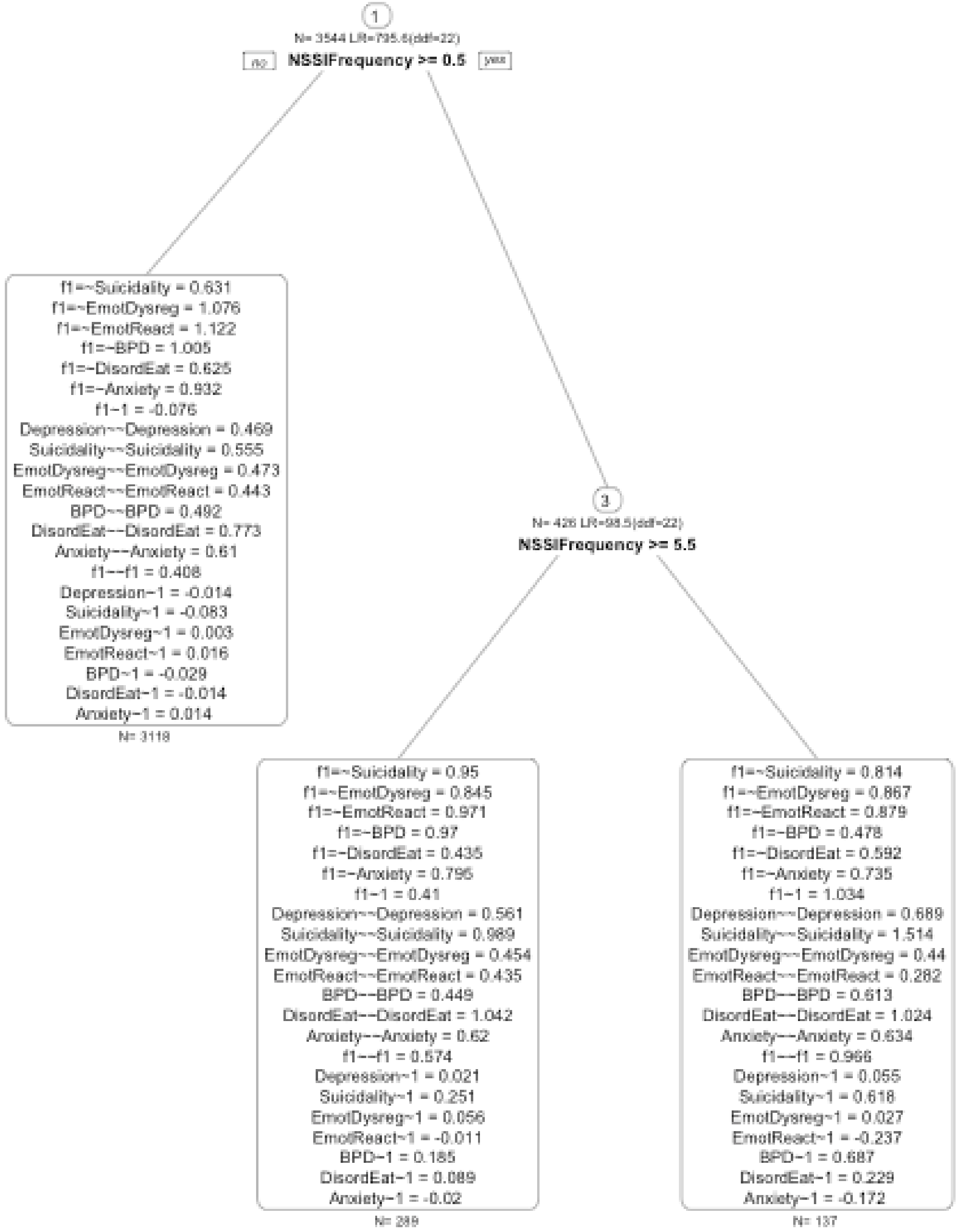




```
N= 3544 LR=795.6(df=22)
NSSIFrequency >= 0.5

f1=~Suicidality = 0.631
f1=~EmotDysreg = 1.076
f1=~EmotReact = 1.122
f1=~BPD = 1.005
f1=~DisordEat = 0.625
f1=~Anxiety = 0.932
f1~1 = -0.076
Depression~~Depression = 0.469
Suicidality~~Suicidality = 0.555
EmotDysreg~~EmotDysreg = 0.473
EmotReact~~EmotReact = 0.443
BPD~~BPD = 0.492
DisordEat~~DisordEat = 0.773
Anxiety~~Anxiety = 0.61
f1~~f1 = 0.408
Depression~1 = -0.014
Suicidality~1 = -0.083
EmotDysreg~1 = 0.003
EmotReact~1 = 0.016
BPD~1 = -0.029
DisordEat~1 = -0.014
Anxiety~1 = 0.014
N= 3118

N= 426 LR=98.5(df=22)
NSSIFrequency >= 5.5

f1=~Suicidality = 0.95
f1=~EmotDysreg = 0.845
f1=~EmotReact = 0.971
f1=~BPD = 0.97
f1=~DisordEat = 0.435
f1=~Anxiety = 0.796
f1~1 = 0.41
Depression~~Depression = 0.561
Suicidality~~Suicidality = 0.989
EmotDysreg~~EmotDysreg = 0.454
EmotReact~~EmotReact = 0.435
BPD~~BPD = 0.449
DisordEat~~DisordEat = 1.042
Anxiety~~Anxiety = 0.62
f1~~f1 = 0.574
Depression~1 = 0.021
Suicidality~1 = 0.251
EmotDysreg~1 = 0.056
EmotReact~1 = -0.011
BPD~1 = 0.185
DisordEat~1 = 0.089
Anxiety~1 = -0.02
N= 289

f1=~Suicidality = 0.814
f1=~EmotDysreg = 0.867
f1=~EmotReact = 0.879
f1=~BPD = 0.478
f1=~DisordEat = 0.592
f1=~Anxiety = 0.735
f1~1 = 1.034
Depression~~Depression = 0.689
Suicidality~~Suicidality = 1.514
EmotDysreg~~EmotDysreg = 0.44
EmotReact~~EmotReact = 0.282
BPD~~BPD = 0.613
DisordEat~~DisordEat = 1.024
Anxiety~~Anxiety = 0.634
f1~~f1 = 0.966
Depression~1 = 0.055
Suicidality~1 = 0.618
EmotDysreg~1 = 0.027
EmotReact~1 = -0.237
BPD~1 = 0.687
DisordEat~1 = 0.229
Anxiety~1 = -0.172
N= 137
```

SEM Trees & NSSI Cont’d

- Can investigate individual parameters to get a better clue to how groups differ
  - Factor mean increased w/ NSSI frequency
  - Most variable means increased
  - Factor variance increased



1
N= 3544 LR=795.6(ddf=22)
no NSSIFrequency >= 0.5 yes

f1=~Suicidality = 0.631
f1=~EmotDysreg = 1.076
f1=~EmotReact = 1.122
f1=~BPD = 1.005
f1=~DisordEat = 0.625
f1=~Anxiety = 0.932
f1~1 = -0.076
Depression~~Depression = 0.469
Suicidality~~Suicidality = 0.555
EmotDysreg~~EmotDysreg = 0.473
EmotReact~~EmotReact = 0.443
BPD~~BPD = 0.492
DisordEat~~DisordEat = 0.773
Anxiety~~Anxiety = 0.61
f1~~f1 = 0.408
Depression~1 = -0.014
Suicidality~1 = -0.083
EmotDysreg~1 = 0.003
EmotReact~1 = 0.016
BPD~1 = -0.029
DisordEat~1 = -0.014
Anxiety~1 = 0.014
N= 3118

3
N= 426 LR=98.5(ddf=22)
NSSIFrequency >= 5.5

f1=~Suicidality = 0.95
f1=~EmotDysreg = 0.845
f1=~EmotReact = 0.971
f1=~BPD = 0.97
f1=~DisordEat = 0.435
f1=~Anxiety = 0.796
f1~1 = 0.41
Depression~~Depression = 0.561
Suicidality~~Suicidality = 0.989
EmotDysreg~~EmotDysreg = 0.454
EmotReact~~EmotReact = 0.435
BPD~~BPD = 0.449
DisordEat~~DisordEat = 1.042
Anxiety~~Anxiety = 0.62
f1~~f1 = 0.574
Depression~1 = 0.021
Suicidality~1 = 0.251
EmotDysreg~1 = 0.056
EmotReact~1 = -0.011
BPD~1 = 0.185
DisordEat~1 = 0.089
Anxiety~1 = -0.02
N= 289

f1=~Suicidality = 0.814
f1=~EmotDysreg = 0.867
f1=~EmotReact = 0.879
f1=~BPD = 0.478
f1=~DisordEat = 0.592
f1=~Anxiety = 0.735
f1~1 = 1.034
Depression~~Depression = 0.689
Suicidality~~Suicidality = 1.514
EmotDysreg~~EmotDysreg = 0.44
EmotReact~~EmotReact = 0.282
BPD~~BPD = 0.613
DisordEat~~DisordEat = 1.024
Anxiety~~Anxiety = 0.634
f1~~f1 = 0.966
Depression~1 = 0.055
Suicidality~1 = 0.618
EmotDysreg~1 = 0.027
EmotReact~1 = -0.237
BPD~1 = 0.687
DisordEat~1 = 0.229
Anxiety~1 = -0.172
N= 137

SEM Trees & NSSI Conclusion

- Agreement in cutoffs across methods
  - Important, as SEM Trees hasn’t been studied that much
- SEM Trees allows for a more comprehensive evaluation
  - While being able to make specific comparisons across groups
- Was able to incorporate more information into the model than if we used mixture models
  - This is not always the case in comparing mixtures and SEM Trees
  - We had a ***very*** informative covariate
- Not that difficult to program and didn’t take long to run
  - Demonstrated in Lab Session T

SEM Trees Options

SEM Trees Options

- Because there is an SEM for each group
  - Should set minimum number of cases per node of tree
    - `min.N` in `semtree.control`
- Large number of covariates?
  - Need to prevent type-1 errors (overfitting)
    - Also, large, uninterpretable tree
  - Can use Bonferroni correction or use cross-validation
  - `bonferroni = TRUE` or `method=“cv”` in `semtree.control`
- Large number of response options for covariate(s)
  - These variables are more likely to be split on
  - Correct with changing `method=“fair”`

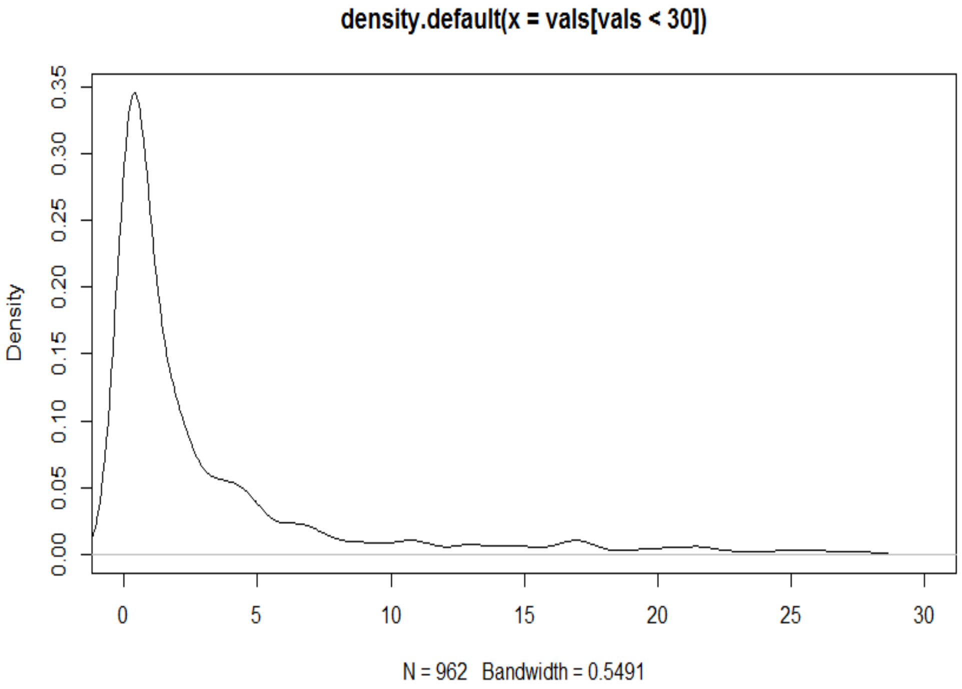
SEM Trees & Invariance

- In many contexts, it is necessary to ensure that groups are measured on the same construct
- This can either be call differential item functioning (IRT) or measurement invariance (SEM)
- This entails constraining certain parameters to be equal across groups
- Most common: constrain factor loadings
- Easy to do in SEM Trees
  - Set `invariance =` parameter names in `semtree.control`
    - Demonstrated in Lab Session T

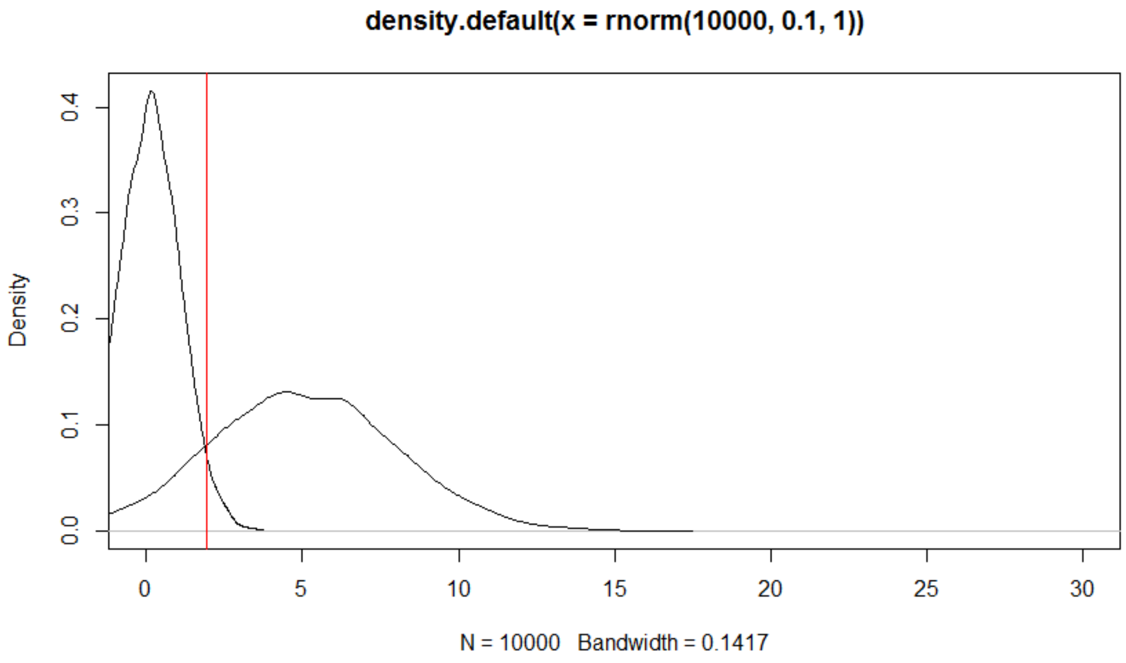
SEM Trees vs. Mixture Models

Mixture Models

- Traditional way of doing things
  - Look for multiple underlying normal distributions underlying NSSI distribution

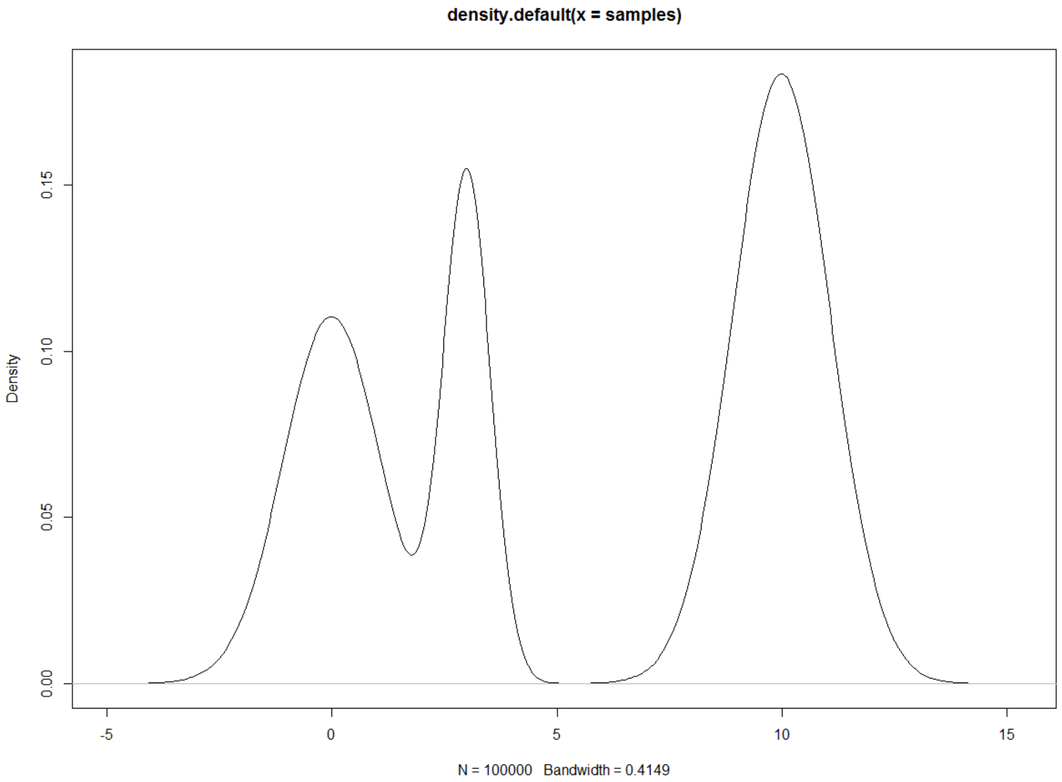


=



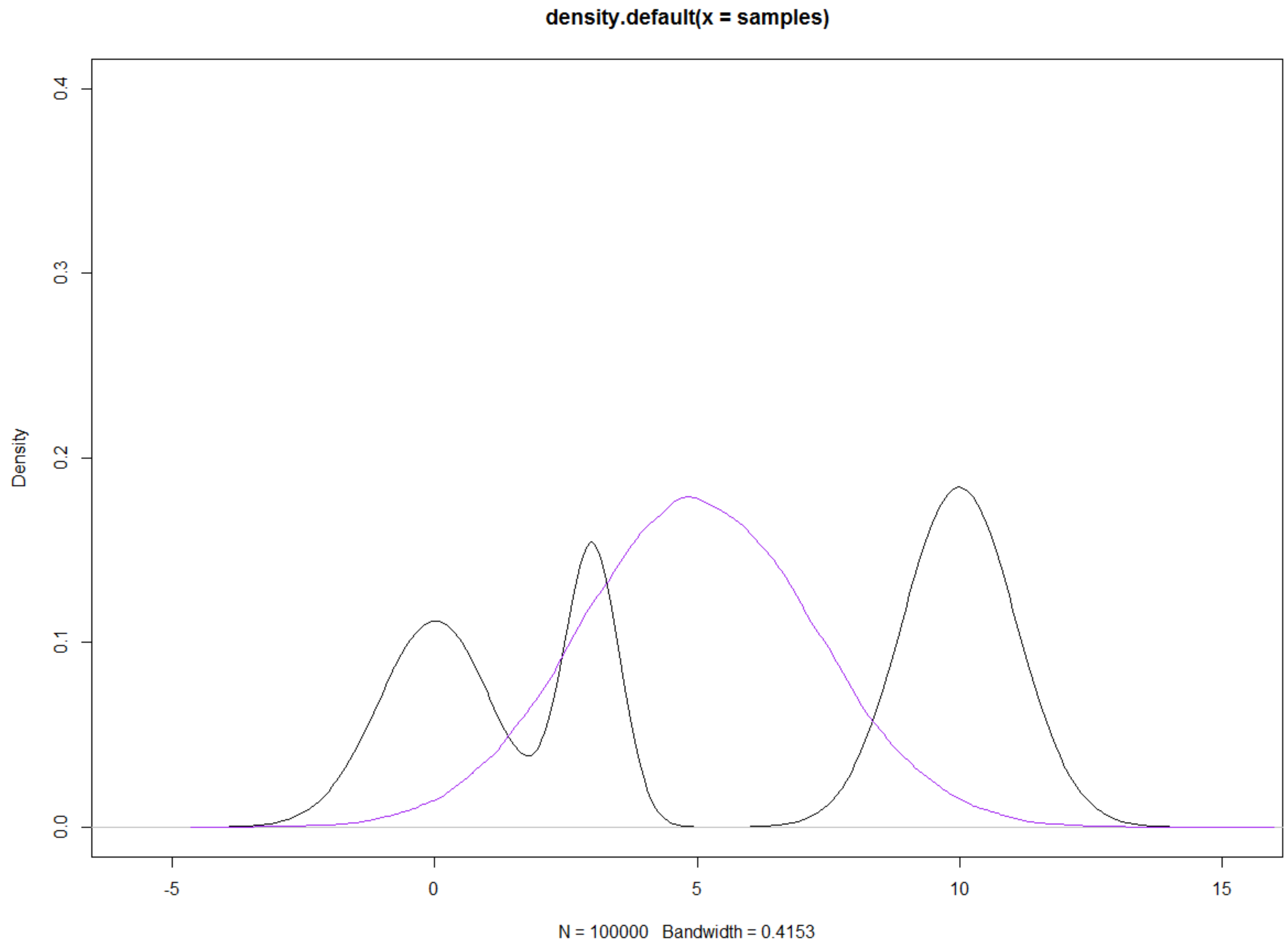
SEM Trees vs. Mixture Models

- Do very similar things, but find groups (classes) a little differently
- Let’s say we have a variable we are trying to find groups in
- Hypothetical distribution for a cognitive score variable:



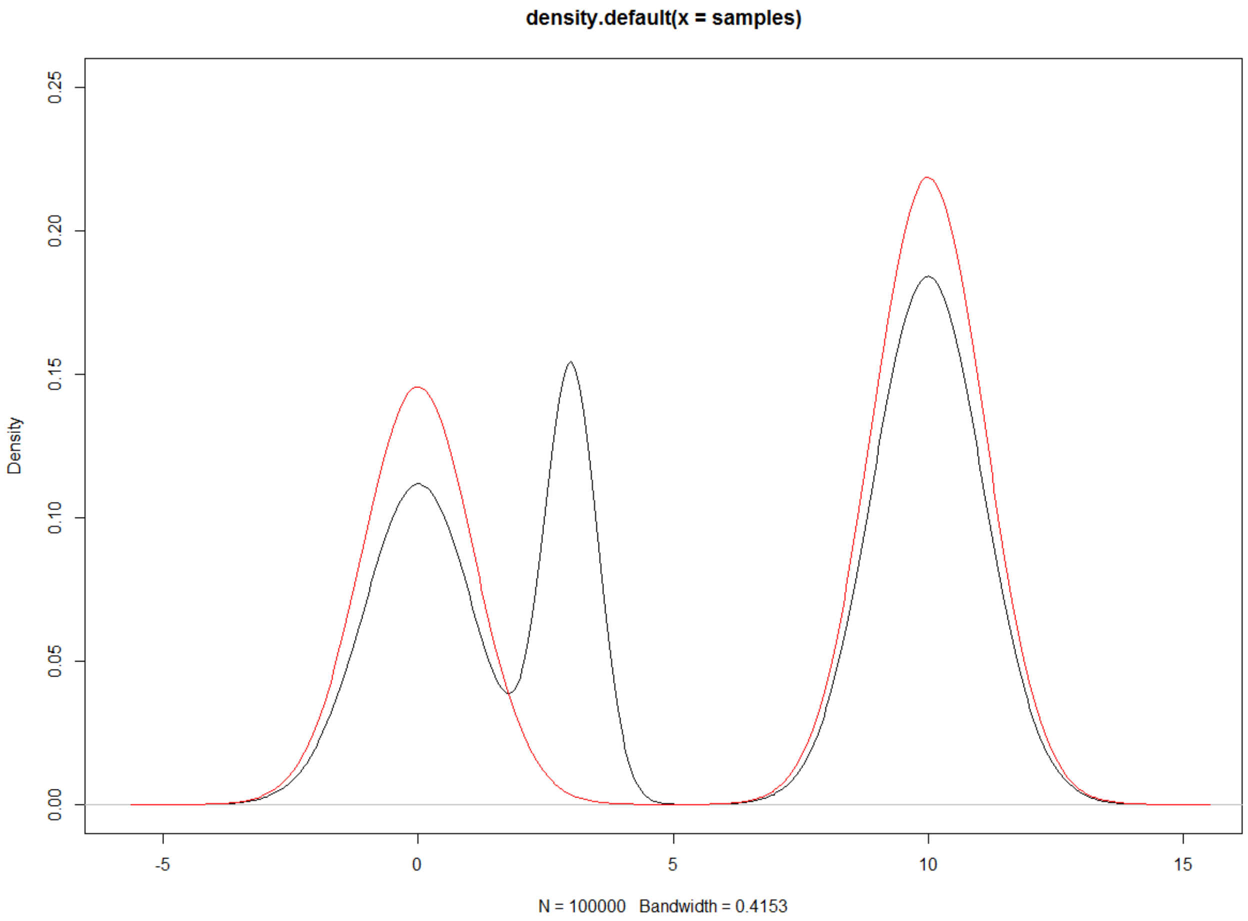
Assume Homogeneity

- Doesn’t do so well



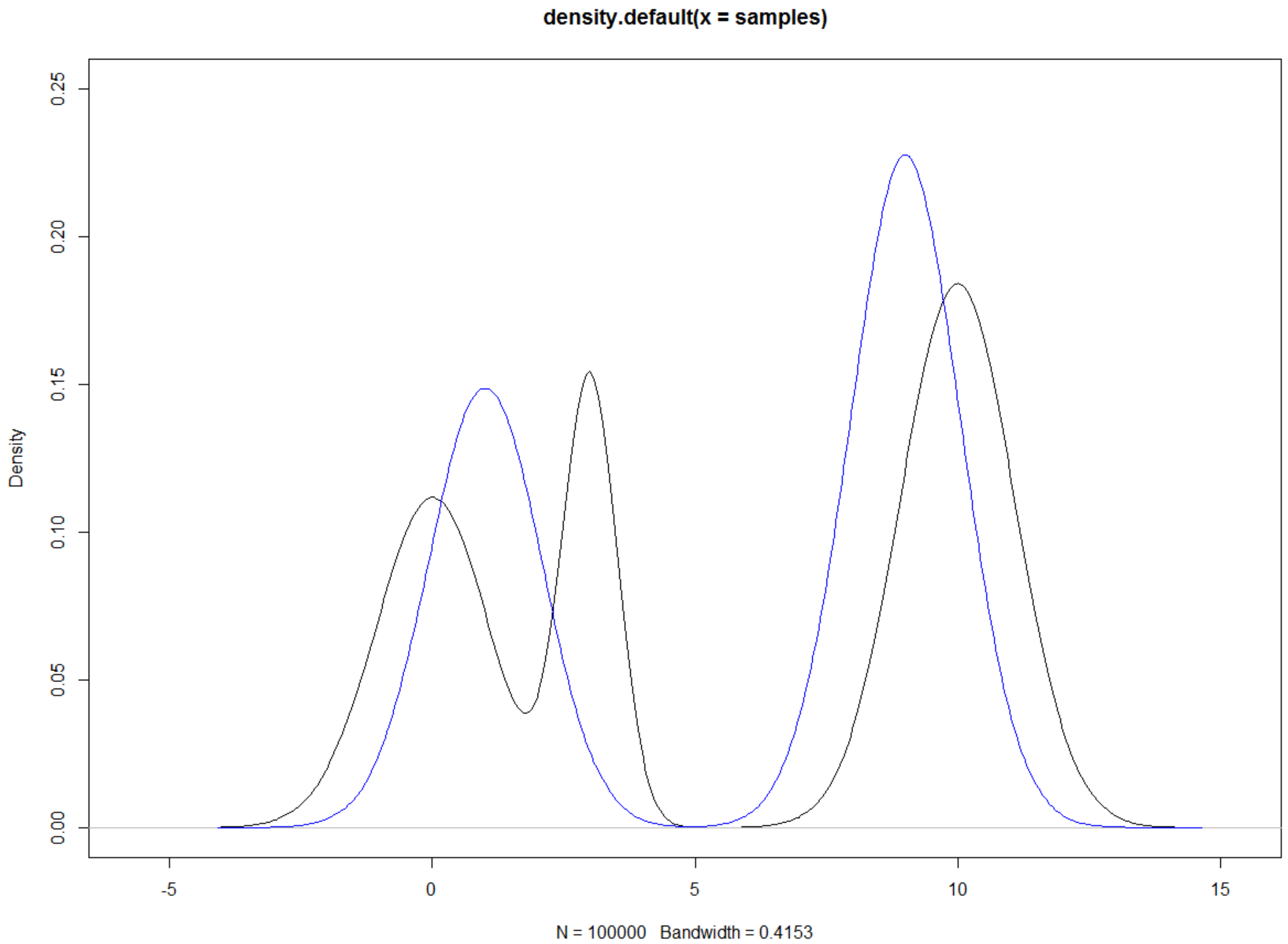
Comparison Example

- Mixture models will attempt to fit this distribution directly
- An example with 2 classes



Comparison Example Cont'd

- SEM Trees will try and fit this distribution
  - But only through splitting on the covariates
  - Meaning SEM Trees is more constrained
  - Doesn't have access to the whole search space
  - For example, the blue line could represent the distribution for a gender variable



Comparison Example Summary

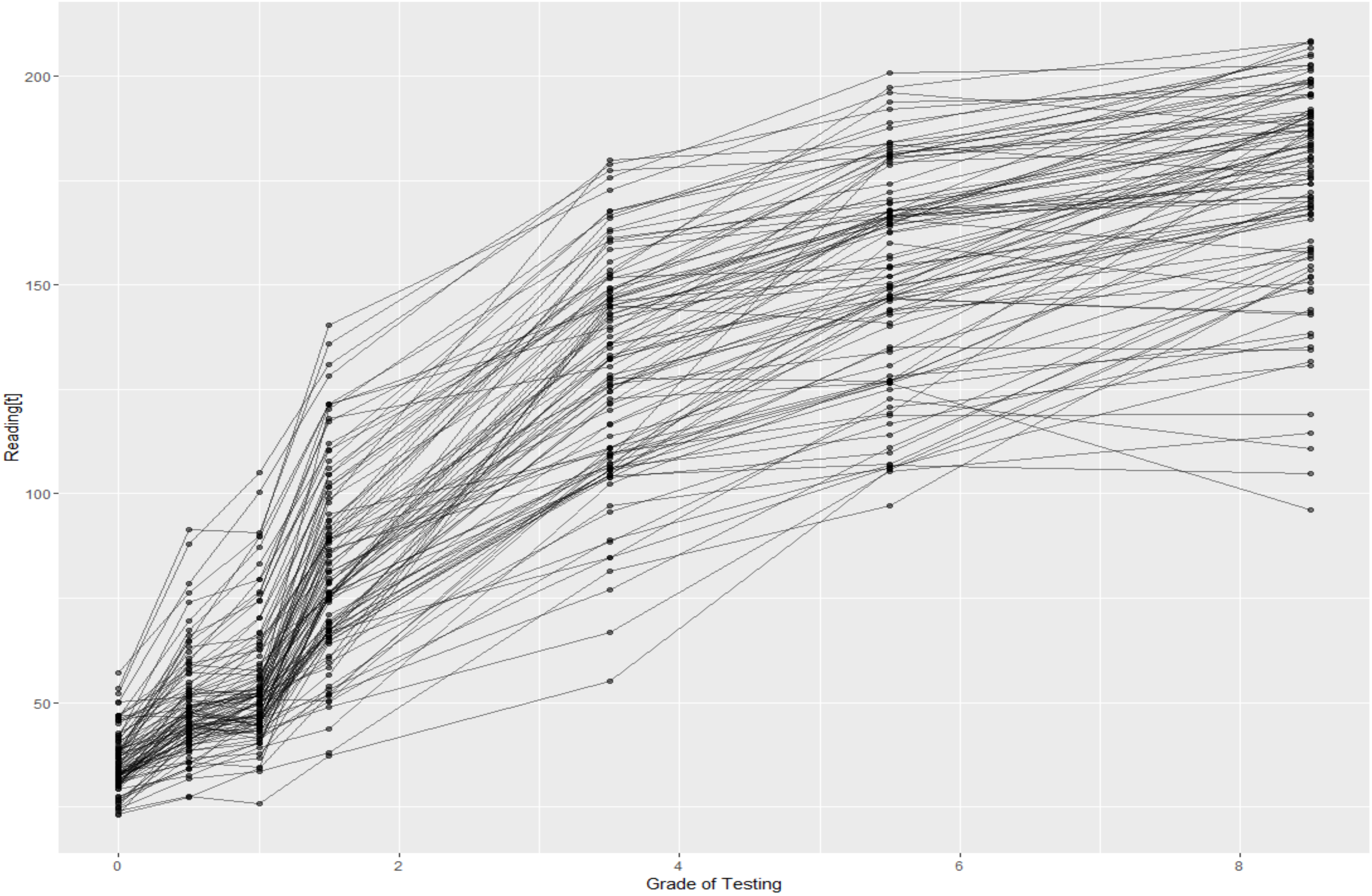
- SEM Trees requires informative covariates to split on
  - Mixtures do not
- SEM Trees may be less likely to overfit
- SEM Trees allows interactions and non-linear prediction
  - In mixtures, the classes are all at the same level, not nested or based on interactions
  - In mixtures, the relationship between class membership and auxiliary variables is constrained to be linear
- Can be very useful to test both methods and compare the resultant groups

SEM Tree vs. Growth Mixture Model

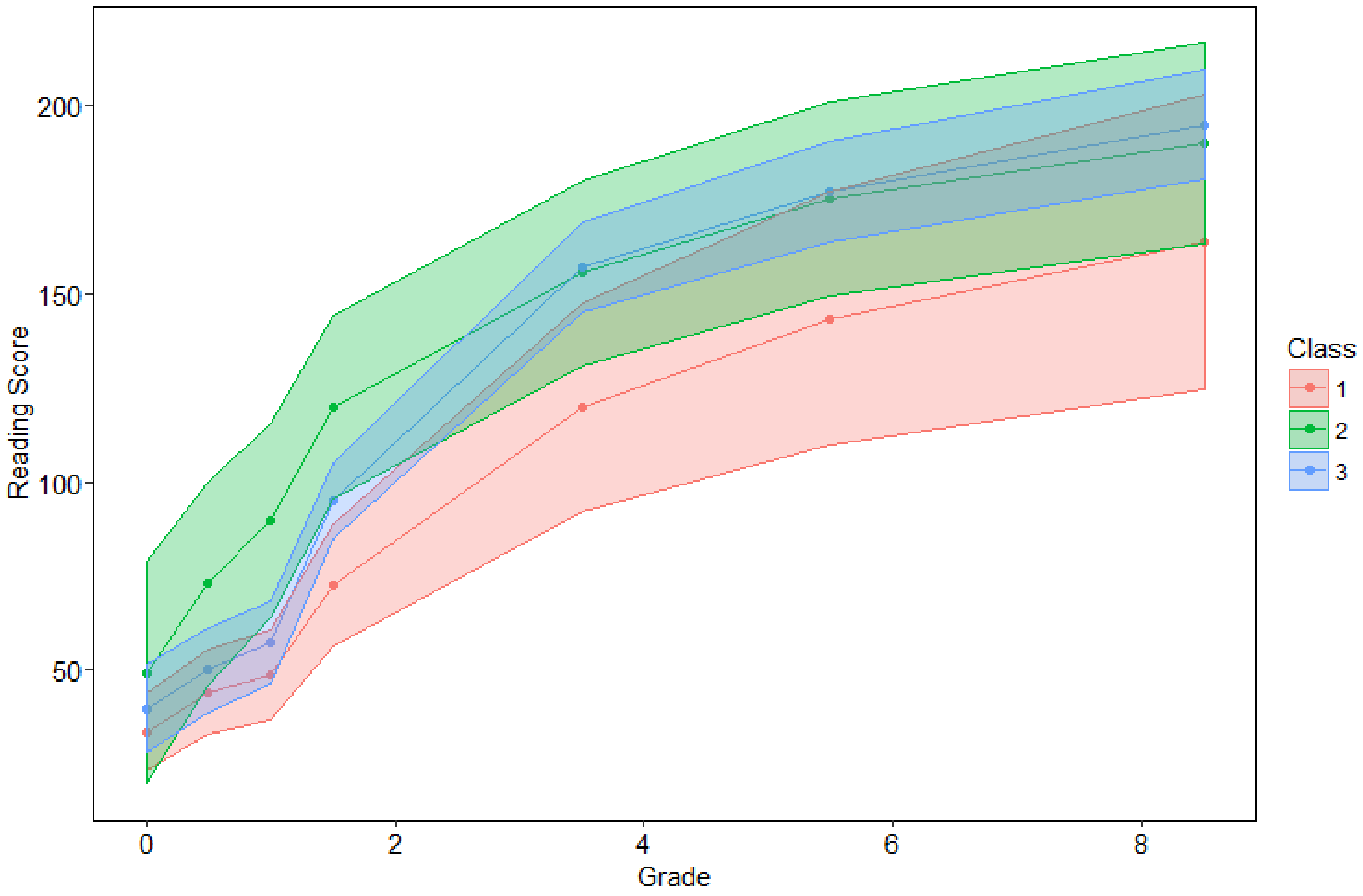
Comparison Example

- Compare differences in trajectories
  - Using a latent growth curve model
  - Change in reading from Kindergarten to 8th grade
  - Early Childhood Longitudinal Study - Kindergarten Cohort (ECLS-K) data
- Questions we are trying to answer
  - Do all children learn at the same rate or are the subgroups?
    - i.e. Late bloomers (start low, but catch up)
    - Those who start high and improve fast
  - Can other measured variables give us insight into these groups
    - Covariates include fine motor skills (*fine*), gross motor skills (*gross*), approaches to learning (*learn*), self-control (*control*), interpersonal skills (*interp*), internalizing behaviors (*int*), and general knowledge (*gk1*)

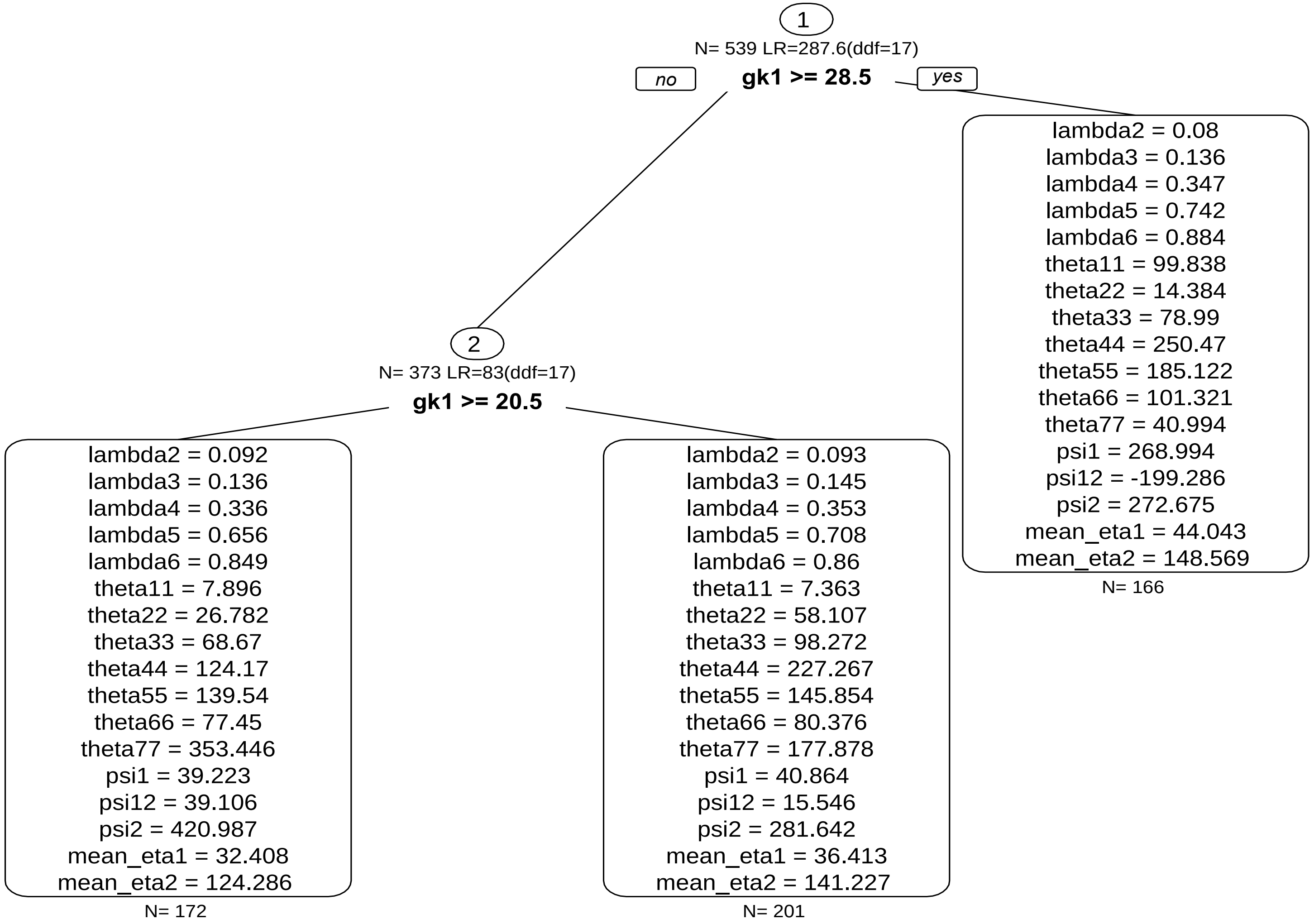
Trajectory Plot



Growth Mixture Model



Growth SEM Trees



1

N= 539 LR=287.6(ddf=17)

no

gk1 >= 28.5

yes

lambda2 = 0.08
lambda3 = 0.136
lambda4 = 0.347
lambda5 = 0.742
lambda6 = 0.884
theta11 = 99.838
theta22 = 14.384
theta33 = 78.99
theta44 = 250.47
theta55 = 185.122
theta66 = 101.321
theta77 = 40.994
psi1 = 268.994
psi12 = -199.286
psi2 = 272.675
mean_eta1 = 44.043
mean_eta2 = 148.569

N= 166

2

N= 373 LR=83(ddf=17)

gk1 >= 20.5

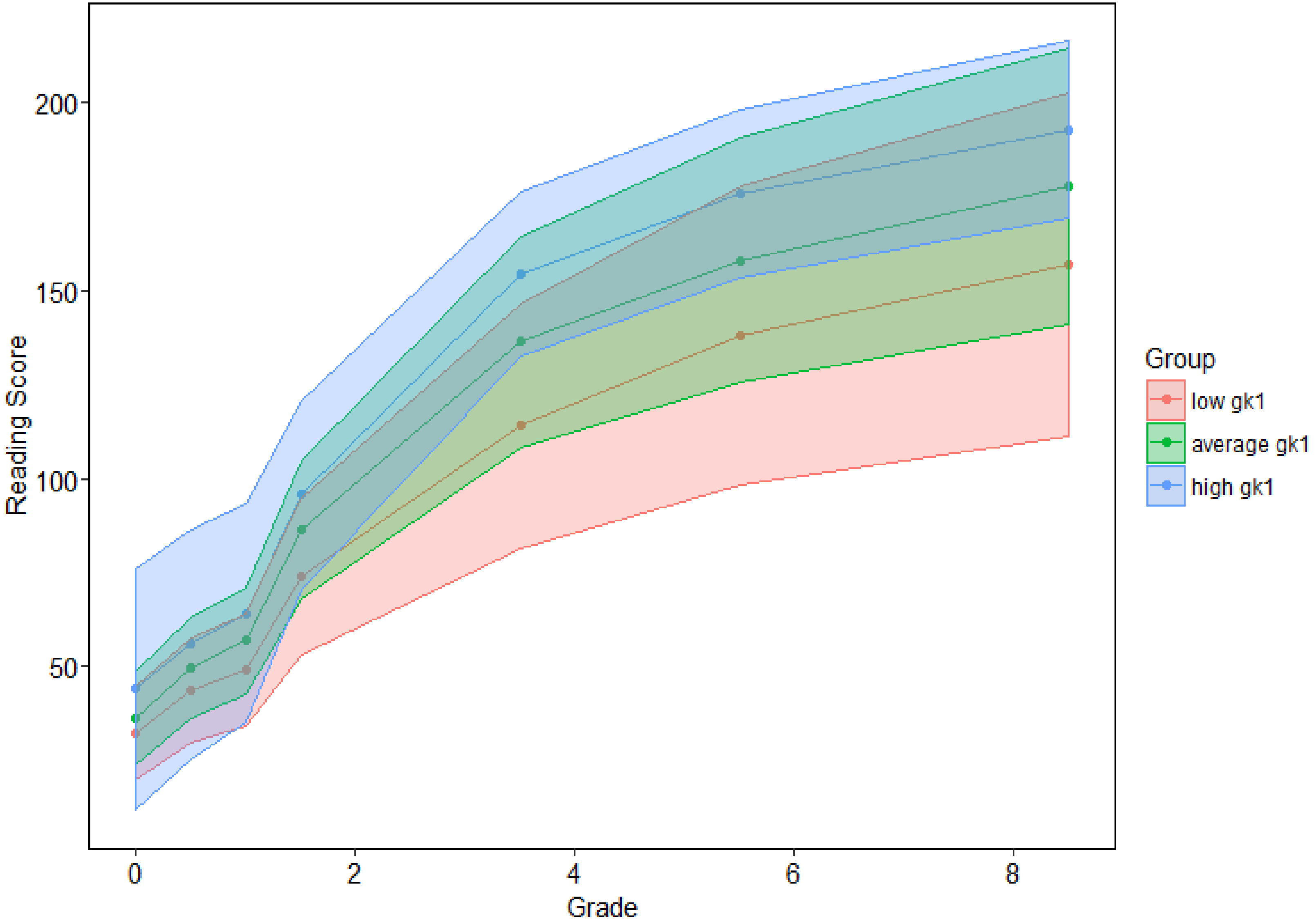
lambda2 = 0.092
lambda3 = 0.136
lambda4 = 0.336
lambda5 = 0.656
lambda6 = 0.849
theta11 = 7.896
theta22 = 26.782
theta33 = 68.67
theta44 = 124.17
theta55 = 139.54
theta66 = 77.45
theta77 = 353.446
psi1 = 39.223
psi12 = 39.106
psi2 = 420.987
mean_eta1 = 32.408
mean_eta2 = 124.286

N= 172

lambda2 = 0.093
lambda3 = 0.145
lambda4 = 0.353
lambda5 = 0.708
lambda6 = 0.86
theta11 = 7.363
theta22 = 58.107
theta33 = 98.272
theta44 = 227.267
theta55 = 145.854
theta66 = 80.376
theta77 = 177.878
psi1 = 40.864
psi12 = 15.546
psi2 = 281.642
mean_eta1 = 36.413
mean_eta2 = 141.227

N= 201

Growth SEM Trees



Comparison Example Summary

- SEM Trees split twice on the general knowledge variable
  - Was a significant auxiliary variable in Mixture Model
- SEM Trees didn’t fit as well
  - More variability within group and overlap across groups
- The Growth Mixture Model had more interesting classes
  - Class 1 caught up to Class 3 in grades 4 and 5
    - Different loadings across groups
  - SEM Trees groups really only differed in Kindergarten

SEM Trees Conclusion

- Has been extended to allow multiple trees
  - SEM Forests – analogous extension from DT → Random Forests
  - Can take a really long time to run (hours/days)
- Very new method
  - Hasn’t been tested to a large extent to set guidelines for best practices
- Can take a long time to run
  - Small model and small # of covariates can be 5 minutes
  - Large model and large # of covariates can be hours
- REMEMBER: Just because you get out a tree, does not mean that it represents real groups
  - Needs to make theoretical sense
  - Could just be the product of overfitting