MULTILEVEL MODELING WORKSHOP

Elizabeth Page-Gould

University of Toronto

WORKSHOP SPONSORS

- Association for Psychological Science (APS)
- Society of Multivariate Experimental Psychology (SMEP)

MULTILEVEL MODELING

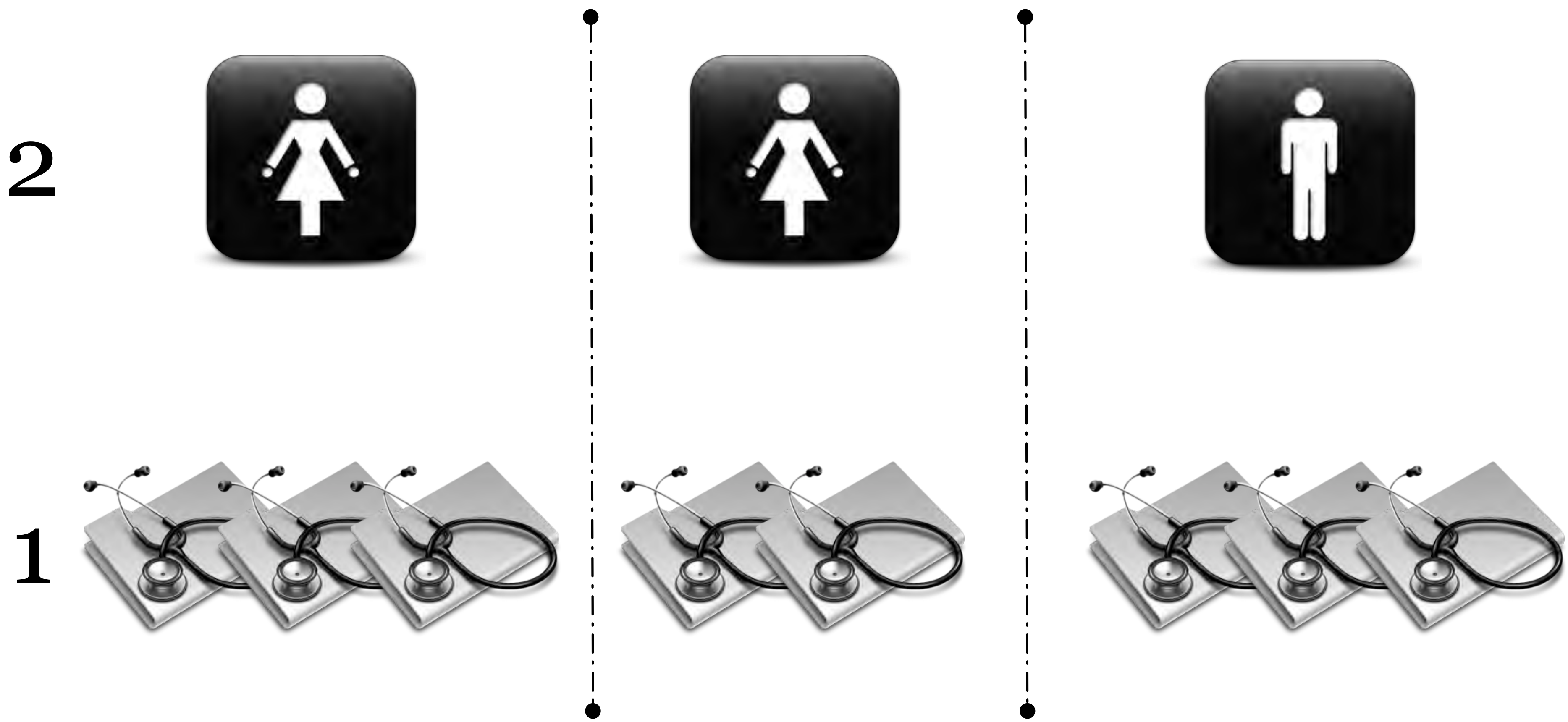
- *A broad class of analyses that deal with hierarchy in your data*

HIERARCHICAL DATA

2

1

HIERARCHICAL DATA



MULTILEVEL MODELING (MLM)

- Broad class of techniques
- 1 face, many names:
  - Hierarchical linear modeling (HLM)
  - General Linear Mixed Model (GLMM)
  - Nested growth curves
  - Mixed models
  - Random effects modeling
  - Random coefficient modeling
  - Covariance components models

WORKSHOP OVERVIEW

- Goal: Immediately use MLM to analyze your data
- Outline:
  - Introduce example dataset
  - Conceptual background
  - Pragmatics: Conducting the analysis & reporting it
  - Advanced applications

GETTING THE MOST OUT OF THIS WORKSHOP

- Review is not just review
- Think about this as a story
- Try to predict where I'm going
- Relate these issues to what you may encounter in your research

SYNTAX AND WORKSHOP RESOURCES

- Download: http://page-gould.com/mlm/aps
- Resources Include:
  - Syntax for all examples in SPSS, R, and SAS
  - PDF of today's slides
  - Some supplementary readings
- Physical Resources: Lecture slides are distributed




BIG TOBACCO GOES TO WASHINGTON

Luke & Krauss (2004)

LEGISLATOR DATA

- Goal:
  - *Identify influences on US Federal Representatives' tobacco-related voting*
- Luke & Krauss (2004) recorded:
  - Percentage of pro-tobacco votes of every representative in the US Congress in 1999
  - Money donated to each representative's campaign by tobacco companies
  - Acreage of tobacco farming in the representative's home state
  - Each representative's home state

LEVELS IN LEGISLATOR DATA

LEVEL 1 VS. LEVEL 2

- Level 1 is the smallest unit of analysis
  - Level 1 datapoints are different in every row
- Level 2 variables are constant for all level 1 variables that are "nested" in it
  - Level 2 variables will be constant across ≥ 2 rows in your data spreadsheet

DATA STRUCTURE

Level 1	Level 2	Level 1	Level 2	Level 1
Legislator ID	State ID	Tobacco Contributions	Tobacco Acreage	% Votes Pro-Tobacco
1	1	21.5	3.04	7.89
2	1	12.0	3.04	12.82
3	1	0	3.04	10.00
4	2	0	5.80	12.82
5	2	11.75	5.80	71.79
...	...	...	...	...

CENTRALITY IN LEGISLATOR DATA

SPSS:

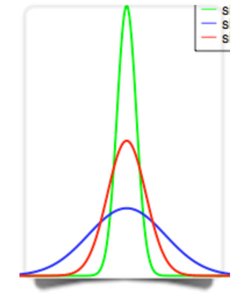
```
FREQUENCIES VARIABLES=x
/STATISTICS=MEAN MEDIAN MODE.
```

R:

```
mean( x, na.rm=T )
median( x, na.rm=T )
names( which.max( table( x ) ) )
```

	Contributions *(money)*	Acreage *(acres)*	% Pro-Tobacco Votes *(voting)*
Mean	12.96	13.01	51.80
Median	4.5	0	59.46
Mode	0	0	100

SPREAD IN LEGISLATOR DATA



SPSS:

```
FREQUENCIES VARIABLES=x
 /NTILES=4
 /STATISTICS VARIANCE STDDEV.
```

R:

```
var( x, na.rm=T )
sd( x, na.rm=T )
quantile( x, probs=c(.25, .75) )
```

	Contributions *(money)*	Acreage *(acres)*	% Pro-Tobacco Votes *(voting)*
Variance	339.95	1727.42	1213
Standard Deviation	18.44	41.56	34.8
Interquartile Range	0 - 20.5	0 - 5.8	20.0 - 84.6

WHAT'S SO SPECIAL ABOUT VARIANCE?

- Classical statistic's "Standard Candle"
  - Standard deviations are the unit of measurement
  - We know the *probability* of observing any deviation

EXPLAINING VARIANCE IS THE GRAND PRIZE

- Almost any classical statistic compares:

$$\frac{\text{Variance Explained by your Independent Variable(s)}}{\text{Unexplained Variance}}$$

PREDICTION

- Predictive relationships
  - Covariance
  - Regression

CORRELATION & COVARIANCE

- *How much 2 variables change together*

COVARIATION IN LEGISLATOR DATA

SPSS:

CORRELATIONS VARIABLES=x y

/STATISTICS XPROD.

R:

cor(x, y, use="complete.obs")

cov(x, y, use="complete.obs")

	Money * Voting	Acres * Voting
Correlation	0.464	0.249
Covariance	2.98	3.60

CORRELATION (r_{XY})

$$r_{XY} = \frac{\sum_{i=1}^{N} z_{x_i} z_{y_i}}{N-1}$$

- *Strength of predictive relationship between X and Y*
- Dimensionless

COVARIANCE (COV_{XY})

$$COV_{XY} = r_{XY}\sigma_X\sigma_Y$$

- *Unstandardized measure of relationship between X and Y*
- Values are in units of “XY”

WHY DO WE CARE?

- Covariation *is* your dependence!

$$COV_{X_{Bruins}Y_{Bruins}} \qquad COV_{X_{Sens}Y_{Sens}} \qquad COV_{X_{Pens}Y_{Pens}}$$

WHY DO WE CARE?

- Covariation *is* your dependence!

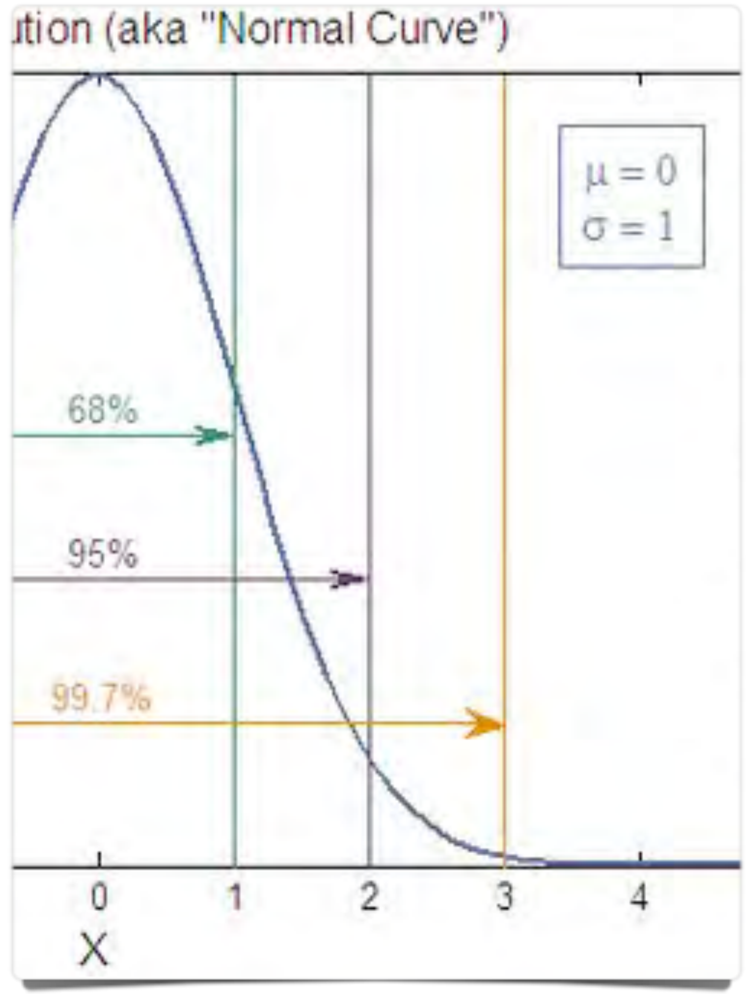
California	New York	Massachusetts
		
$COV_{X_{1_{CA}} X_{2_{CA}}}$	$COV_{X_{1_{NY}} X_{2_{NY}}}$	$COV_{X_{1_{MA}} X_{2_{MA}}}$
 	 	 

IMPORTANT ASSUMPTIONS OF CLASSICAL STATISTICS

- Assumptions:
  - Data were collected through **random sampling**
  - All data are **normally distributed**
  - **Variance must be equal** across conditions
  - All observations must be **independent**
- If your data violate an assumption:
  - *Transform them, if you can, or*
  - *Accept a decrease in statistical power, if you can, or*
  - *Find a test that doesn't require that assumption*

KEY ASSUMPTION RELATIVE TO MLM:

- **All observations must be independent**
  - Specifically: *Residuals are expected to be independent of each other and normally-distributed*



BUT DEPENDENCE IS WHERE IT'S AT

OLD COPING METHODS

- *Groups suck; pretend they don't exist*
  - Use any GLM with no regard for group status
  - Use any GLM with group status as control variable
    - You are still violating assumptions of independence
- *Aggregate*

REGRESSION

- Estimation
- Moderated regression

REGRESSION

Predicted Value

Weighted Average of Y

Influence of X on Y

Stuff You Can’t Explain

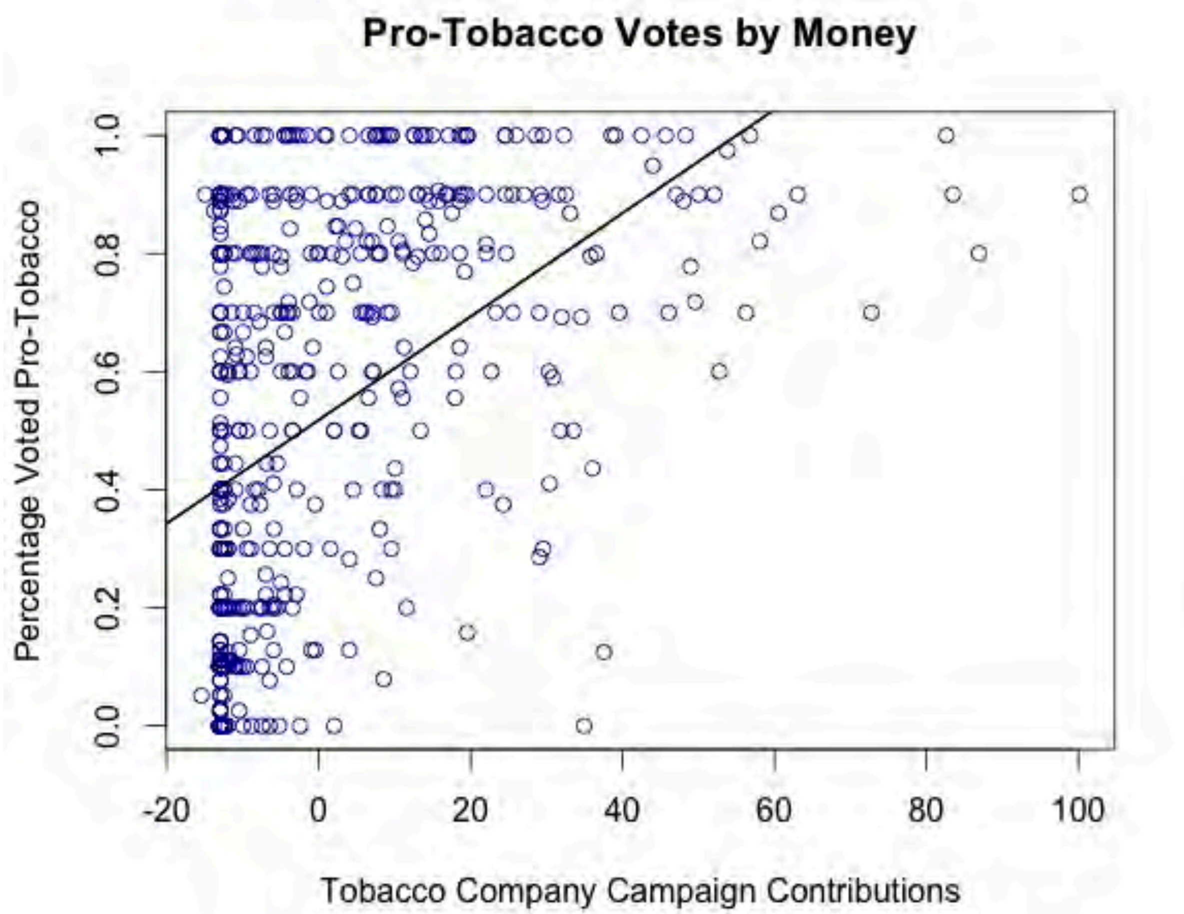
$$\boxed{\hat{y}_i} = \boxed{b_0} + \boxed{b_1} x_i + \boxed{e_i}$$

$$b_0 = \overline{Y} - b_1 \overline{X}$$

$$b_1 = r_{XY} \left(\frac{\sigma_Y}{\sigma_X} \right)$$

$$e_i = \hat{y}_i - y_i$$

REGRESSION



Pro-Tobacco Votes by Money
Percentage Voted Pro-Tobacco
0.0
0.2
0.4
0.6
0.8
1.0
-20
0
20
40
60
80
100
Tobacco Company Campaign Contributions

REGRESSION

Weighted Mean(Y) | Influence of X | Stuff You Can't Explain

$$y_i = \boxed{b_0} + \boxed{b_1} x_i + \boxed{e_i}$$

$$\hat{y}_i = \boxed{b_0} + \boxed{b_1} x_i$$

ESTIMATION

$$y_i = b_0 + b_1 x_i + e_i$$

i = John Kerry

$$y_{Kerry} = .05$$

$$\hat{y}_{Kerry} = .02$$

$$\boxed{e_{Kerry} = .03}$$

REGRESSION AND LEGISLATOR DATA

SPSS:

```
REGRESSION
/DEPENDENT y
/METHOD=ENTER x.
```

R:

```
model <- lm( y ~ x )
summary( model )
```

Pro-Tobacco Voting=	b_0+b_1(Money)
Estimate	= .404 + .009 (money)

MODERATED REGRESSION

- *Regression where the effect of one independent variable depends on another independent variable*
- Allows you to examine ***main effects*** and ***interactions***

$$\hat{y}_i = b_0 + \underbrace{\boxed{b_1 x_{1_i}} + \boxed{b_2 x_{2_i}}}_{\text{Main Effects}} + \underbrace{\boxed{\left(b_1 x_{1_i}\right) * \left(b_2 x_{2_i}\right)}}_{\text{Interaction Effect}}$$

WHAT DOES IT *MEAN?*

- Multiple regression equation:
  - Every "+" represents an additive, main effect
    - The effects of each variable, *independent of its relationship with the other predictors*
  - Every multiplication represents a *dependence between predictors*

$$\widehat{y_i} = b_0 + b_1 x_{1_i} + b_2 x_{2_i} + b_3 x_{1_i} x_{2_i}$$

MODERATED REGRESSION AND LEGISLATOR DATA

SPSS:

```
COMPUTE x1Xx2 = x1*x2.
REGRESSION
/DEPENDENT y
/METHOD=ENTER x1 x2 x1Xx2.
```

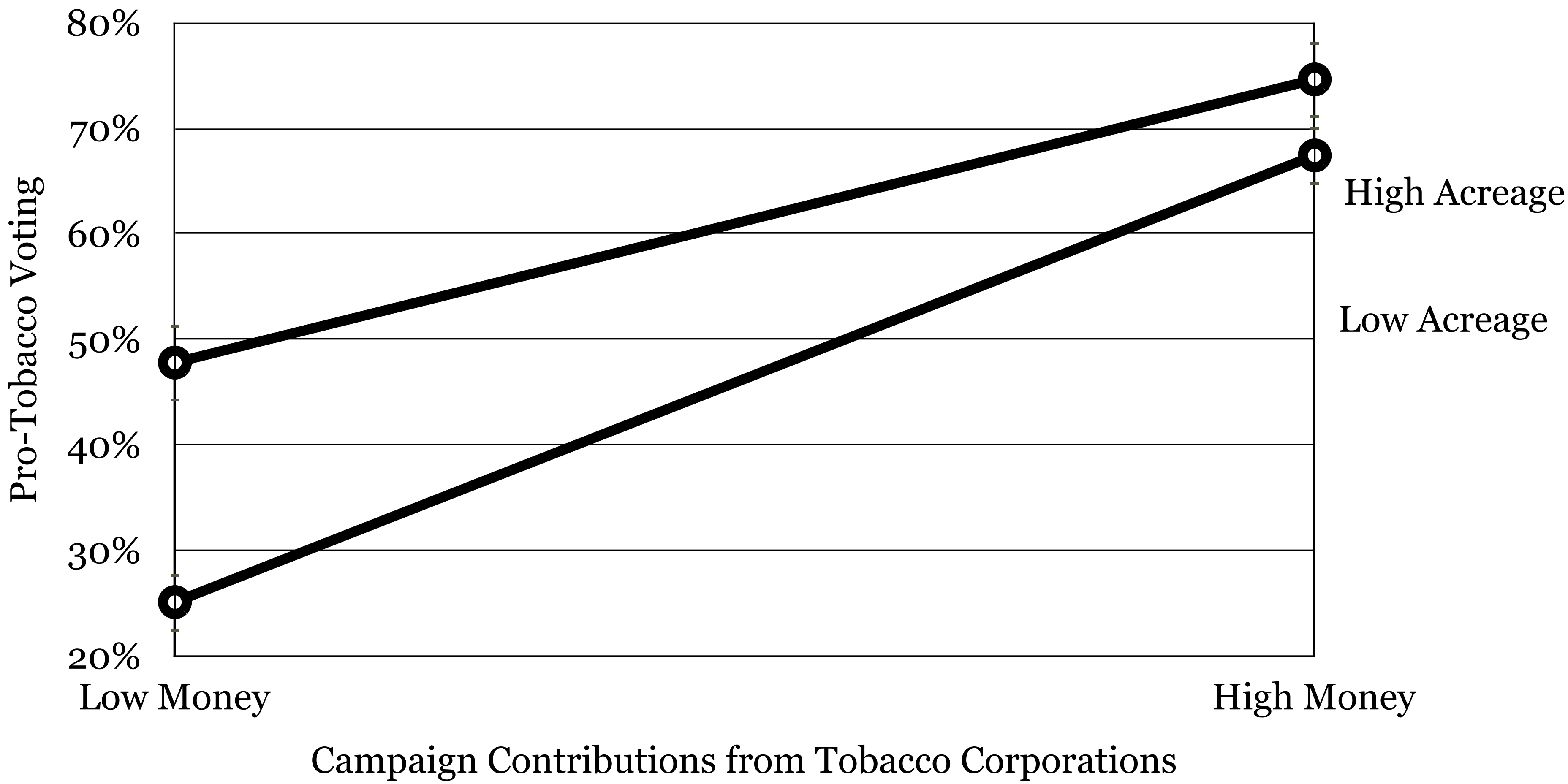
R:

```
moderated.regression <- lm( y~x1*x2)
summary(moderated.regression)
```

Pro-Tobacco Voting=	$b_0 + b_1$(Money)$+ b_2$(Acres) $+ b_3$(Money*Acres)
Estimate	= .537 + .009 (money) + .002 (acres) - .00005 (money*acres)

MODERATION IN THE LEGISLATOR DATA

$$\widehat{y_i} = b_0 + b_{Money} x_{Money} + b_{Acres} x_{Acres} + \left(b_{MoneyXAcres})(x_{Money} \right)\left(x_{Acres} \right)$$



BUT WE HAVE TO DO SOMETHING ABOUT THE LEVELS

- Crux of this workshop:
  - ***Explain all the variance you can***
  - ***Statistical assumptions matter***

WHAT TO DO ABOUT THE GROUPS?

- We *shouldn't* ignore them
  - Ignoring = more unexplained variance
  - Ignoring = inaccurate comparison distributions

SOLUTIONS

- Aggregate your level 1 variables and do your normal statistical thing
- Multilevel Modeling!

AGGREGATED DATA

1. Within each group, calculate the averages of each Level 1 variable
2. Run your analysis with the aggregate variable
    - Each group is your case

NON-AGGREGATED DATA

Legislator ID	State ID	Tobacco Contributions	Tobacco Acreage	% Votes Pro-Tobacco
1	1	21.5	3.04	7.89
2	1	12.0	3.04	12.82
3	1	0	3.04	10.00
4	2	0	5.80	12.82
5	2	11.75	5.80	71.79
…	…	…	…	…

AGGREGATED DATA

State ID	Tobacco Contributions	Tobacco Acreage	Pro-Tobacco Voting
1	4.44	3.04	16.3
2	12.02	5.80	57.2
3	14.58	0.00	36.8
4	33.33	33.00	71.1
5	4.13	0.00	21.1
…	…	…	…

BENEFITS OF AGGREGATING

- All your cases are independent!
  - Use whatever analysis you want
- The aggregated variables will have:
  - Fewer outliers
  - Smaller variance

CONS OF AGGREGATING

- ¡POWER!
  - Your *N* is now the number of groups, not observations
- Changing the *unit* of analysis changes the *meaning*
- Your predictive resolution decreases

DEMO:
REDUCED POWER

RANDOM EFFECTS MODELS

- First form of multilevel modeling
- Types of Random Effects Models:
  - Random-effects ANOVA
  - Random slope models
- What's random about the intercepts and slopes?
  - They are *predicted*
  - So they have *error*
  - The additional error terms are what make them random

WHY AM I TELLING YOU THIS?

- When you run an MLM, you have to declare:
  - Your fixed effects
  - Your random effects

$$\widehat{y_{ij}} = Fixed + Random$$

RANDOM INTERCEPT MODELS

$$\widehat{y_{ij}} = \widehat{b_{0_j}} + b_1 x_{1_{ij}}$$

- = *random-effects ANOVA*
- A unique intercept is predicted for each group

RANDOM SLOPE MODELS

$$\widehat{y_{ij}} = b_0 + \widehat{b_{1_j}} x_{1_{ij}}$$

- A unique slope is predicted for each group

WHAT VARIES BETWEEN YOUR GROUPS?

- Their averages (= *random intercept*)
- Their change (= *random slope*)

WHOA!

- You just learned multilevel modeling!

MULTILEVEL MODELS

MULTILEVEL MODELS!

- Putting it all together
- The equations
- Running a multilevel model

PUTTING IT ALL TOGETHER

- In regression you just estimate the outcome, $\widehat{y_i}$
- In MLM, you estimate parameters on the right side of the equation, too:
  - Intercept: $\widehat{b_0}$
  - Slopes: $\widehat{b_1}, \widehat{b_2}, \ldots$

REGRESSION & MLM

Regression: $$\widehat{y_i} = b_0 + b_1 x_1$$

$$y_i = b_0 + b_1 x_1 + e_i$$

MLM: $$\widehat{y_{ij}} = \hat{b}_{0_j} + \hat{b}_{1_j} x_{ij}$$

$$y_{ij} = b_{0_j} + b_{1_j} x_{ij} + e_{ij} + e_{b_{0_j}} + e_{b_{1_j}}$$

WHY DOES THIS SOLVE OUR PROBLEM?

- All unexplained variance: $\widehat{y_i} - y_i$
- We want to explain more of it by considering groups, $\widehat{y_{ij}} - y_{ij}$
  - Since each group j has its own intercept and/or slope, you are more accurate at predicting $\widehat{y_{ij}}$ for any individual in the group
  - Moreover, you are now accounting for the shared variance among group members

THE EQUATIONS

- Every predicted parameter has an equation that predicts it
- Different Greek symbols are used to differentiate between equations that estimate outcomes (classic regression) and equations that estimate model parameters

MULTILEVEL MODEL

$$\widehat{y_{ij}} = \widehat{b_{0_j}} + \widehat{b_{1_j}} x_{1_{ij}}$$

$$\beta_{0_j} = \gamma_{00} + \gamma_{01} W_j + u_{0_j}$$

$$\beta_{1_j} = \gamma_{10} + \gamma_{11} W_j + u_{1_j}$$

PREDICTED INTERCEPT

$$\widehat{b_{0_j}} = \beta_{0_j}$$

$$\beta_{0_j} = \gamma_{00} + \gamma_{01} W_j + u_{0_j}$$

PREDICTED SLOPE

$$\widehat{b_{1_j}} = \beta_{1_j}$$

$$\beta_{1_j} = \gamma_{10} + \gamma_{11} W_j + u_{1_j}$$

“MULTILEVEL EQUATIONS” FORMAT

$$y_{ij} = \quad Fixed \quad + Random$$

$$y_{ij} = \beta_{0_j} + \beta_{1_j} x_{1_{ij}} + e_{ij}$$

$$\beta_{0_j} = \gamma_{00} + \gamma_{01} W_j + u_{0_j}$$

$$\beta_{1_j} = \gamma_{10} + \gamma_{11} W_j + u_{1_j}$$

“MIXED MODEL” FORMAT

$$y_{ij} = \quad \textit{Fixed} \quad + \quad \textit{Random}$$

$$y_{ij} = \boxed{\left(\gamma_{00} + \gamma_{01}W_j + \gamma_{10}x_{1_{ij}} + \gamma_{11}W_j x_{1_{ij}}\right)} + \boxed{\left(e_{ij} + u_{0_j} + u_{1_j}x_{1_{ij}}\right)}$$

HOW DO I ESTIMATE THE PARAMETERS?

- Thankfully, a computer does it for you
  - Uses an iterative process that minimizes residuals for all estimated parameters
- This process relies on the covariance matrix of individuals within groups
- This process also determines your degrees of freedom

COVARIANCE MATRICES

- Covariance matrix
  - *Assumed relationship among Level 1 data points from the same Level 2 group*
- Most widely used covariance matrices:
  - Variance Components - Default in SPSS and SAS, assumes that data points from different groups do not covary
  - Autoregressive - Standard for basic longitudinal designs, assumes that data points next to each other will be highly correlated
  - Unstructured - Default in R, assumes nothing about covariation structure, best for complicated multilevel models, robust against issues like heteroskedasticity
- Great online resource: http://courses.ttu.edu/isqs5349-westfall/images/5349/mixed_covariance_structures.htm

ESTIMATING DEGREES OF FREEDOM

- The degrees of freedom (df) are *estimated* in MLM based on the iteration process
- Most common df estimation methods in MLM:
  - <u>Satterthwaite</u> - Most widely-used method, because it is akin to a classic ANOVA or regression
    - Note that your *df* will have **decimal** points
    - Default in SAS and only method used by SPSS (and thus the default)
  - <u>Between-Within</u> - More conservative, is robust to complex hierarchical structures
    - Only method used by R (and thus the default)

MULTILEVEL MODELING IN SPSS!

- **Random Intercept only:**

```
MIXED y WITH x

 /FIXED= x

/PRINT= SOLUTION

/RANDOM=INTERCEPT | SUBJECT(group).
```

MULTILEVEL MODELING IN SPSS!

- **Random Intercept & Random Slope:**

```
MIXED y WITH x
 /FIXED= x
/PRINT= SOLUTION
/RANDOM=INTERCEPT x| SUBJECT(group).
```

MULTILEVEL MODELING IN R!

- **Packages:** *nlme* (shown below) or *lme4*
- **Random Intercept only:**

```
mlm.model <- lme( y~x, random=~1|group )

summary(mlm.model)
```

- **Random Intercept & Random Slope:**

```
mlm.model <- lme( y~x, random=~1+x|group )

summary(mlm.model)
```

BENEFITS OF MLM

- ***Theoretical***: More accurately captures reality
- ***Statistical***:
  - Statistical integrity
  - Greater power than aggregating
  - More variance explained!
- ***Pragmatic***: Editors may require it
- ***Tertiary***: It sounds cool

MORE VARIANCE EXPLAINED!

= significance = publication = job security

MLM: STEP-BY-STEP

MULTILEVEL MODELING: STEP-BY-STEP

- Steps in testing a multilevel model
- Demonstrate each step in detail (SPSS & R)
- Interpreting and reporting your results

STEPS TO TEST A MULTILEVEL MODEL

1. Prepare Data

2. Run an analysis

3. Report results

STEPS TO TEST A MULTILEVEL MODEL

1. Prepare Data
   1.1. Inspect your data for plausibility and normality
      1.1.1. Transform non-normal variables (if applicable)
   1.2. Effect-code and Center ALL predictors
   1.3. Organize your data so each level 1 observation has its own row

2. Run an analysis
   2.1. Figure out the specifics of your model
   2.2. Analyze the model
   2.3. Effect size & ICC

3. Report results
   3.1. Describing the analysis
   3.2. Reporting the results

1.2. CENTER AND EFFECT-CODE ALL PREDICTORS

- You <u>must</u> ensure that the zero value for each predictor is meaningful before running the model
  - *See* West, Aiken, & Krull (1996)
  - This <u>only</u> applies to the predictors (including covariates), but not the dependent variable
- Effect-coding
  - *Just like dummy-coding, except you use -1 and 1 instead of 0 and 1*
  - Note: *Contrast-coding would also be acceptable, but always remember that dummy-coding is only for simple effects tests, not omnibus models (Aiken & West, 1991)*
- Centering continuous predictors
  - In MLM, there are two ways to center, by the *grand mean* or the *group mean*

GROUP- V. GRAND-MEAN CENTERING

- Grand-mean Centering: $x_{ij} - \overline{x}$
  - *Interpretation:* This variable now represents each observation's deviation from everyone's norm, regardless of group
- Group-mean Centering: $x_{ij} - \overline{x}_j$
  - *Interpretation:* This variable represents each observation's deviation from their group's norm
- *Easy rule of thumb to live by:*
  - **Group-mean center** all **Level 1 predictors**
  - **Grand-mean center** all **Level 2 predictors**
- ***BUT NOTE:*** You can choose to grand-mean center your Level 1 predictors if you feel that form of centering better represents your research question, but keep in mind that Level 1 and Level 2 effects can become conflated in this case (c.f., Enders & Tofighi, 2007)

CENTERING IN SPSS

- **Grand-mean centering:**

```
AGGREGATE

/OUTFILE=* MODE=addvariables

/variable.grand.mean=MEAN(variable).

COMPUTE variable.grandc = variable -
variable.grand.mean.

EXECUTE.
```

- **Group-mean centering:**

```
AGGREGATE

/OUTFILE=* MODE=addvariables

/BREAK = group

/variable.group.mean=MEAN(variable).

COMPUTE variable.groupc = variable -
variable.group.mean.

EXECUTE.
```

CENTERING IN R

- **Grand-mean centering:**

```
variable.grandc <- variable - mean(variable, na.rm=T)
```

- **Group-mean centering:**

```
group.mean.dataset <- data.frame(group = names(tapply(variable, group, mean,
na.rm=T)), variable.group.mean = as.numeric(tapply(variable, group, mean,
na.rm=T)))

original.data <- merge(original.data, group.mean.dataset)

original.data <- within(original.data, variable.groupc <- variable -
variable.group.mean)
```

1.3. ORGANIZE YOUR DATA

- Every value of a Level 1 variable should have its own row
  - E.g., you studied two partners from 100 married couples
    - You would have 200 rows, with each individual participant having 1 row
  - E.g., you measured reaction times to 50 stimuli in 12 task blocks?
    - You would have 600 rows, with reaction times to each stimulus having 1 row

INCORRECTLY ORGANIZED DATA

Participant	Veggies Day 1	Veggies Day 2	Veggies Day 3
1	3	4	2
2	3	4	5
3	1	2	2
4	3	3	3
5	6	7	7
6	3	4	3
7	4	4	2
8	2	5	4
9	4	3	5

CORRECTLY ORGANIZED DATA

Participant	Day	Veggie Servings
1	1	3
1	2	4
1	3	2
2	1	3
2	2	4
2	3	5
3	1	1
3	2	2
3	3	2

TAKING STOCK: STEP 1

☑ Your data are plausible!

☑ Your data is centered and effect-coded!

☑ Your data is correctly organized

2.1. SPECIFY YOUR MODEL

- Component specification:
  - What is your outcome variable?
  - What are your predictors?
  - What is your grouping variable?
- Effects specification:
  - Fixed versus Random effects
  - Covariance matrix
  - Method for estimating degrees of freedom

COMPONENT SPECIFICATION

- Outcome variable
- Predictors
- Grouping

OUTCOME VARIABLES

- Level 1 Outcomes:
  - Most multilevel modeling in ψ uses Level 1 variables as outcomes
  - Your outcome variable should have a unique value on each row of your dataset
- Level 2 Outcomes:
  - Not impossible, consider level 1 observations to be like test-retest reliability
  - Used more often in sociology, political science, and organizational psych
  - You'll need a lot of groups

OUTCOMES IN LEGISLATOR DATA

- Pro-Tobacco Voting (Level 1)

PREDICTORS

- Think about:
  - Predictors
    - Are your predictions all about main effects, or an interaction?
    - Is each predictor at Level 1 or Level 2?
      - Campaign contributions by tobacco companies - Level 1
      - Acreage of tobacco agriculture in state - Level 2
  - Covariates

PREDICTORS IN LEGISLATOR DATA

- Predictors:
  - Acres (Level 2): Grand-mean center it!
  - Money (Level 1): Group-mean center it!
  - Will look at both main effects + interaction:

    = Money + Acres + Money * Acres
- Covariates:
  - House (Level 1): Effect-code it!

    = House + Money + Acres + Money * Acres

"CROSS-LEVEL" EFFECTS

- A cross-level effect is any effect where the relevant variables come from different levels
  - Typically, the term "cross-level effect" refers to a cross-level MAIN effect
    - *E.g.: Level 2 Acreage is predicting Level 1 Voting*
- **"Cross-level Interaction"**
  - A cross-level interaction is an interaction where the moderating variables come from different levels
    - *E.g.: Level 2 Acreage is interacting with Level 1 Money*
  - When you run a model with a cross-level interaction, **you must specify the Level 1 slope as random** (c.f., Aguinis, Gottfredson, & Culpepper, in press)

GROUPING VARIABLES

- How many levels?
- What is nested in what?

GROUPING IN LEGISLATOR DATA

- Legislators (Level 1) are nested in states (Level 2)

EFFECTS SPECIFICATION

- Fixed versus random effects
- Covariance matrices
- Method for estimating degrees of freedom (SAS only)

FIXED V. RANDOM EFFECTS

- What are your model's random effects?
  - Are you modeling random intercepts only?
  - Are you modeling random intercepts *and* slopes?

FIXED V. RANDOM EFFECTS IN LEGISLATOR DATA

- Fixed (the actual things you want to test):
  - Intercept (b_{oj})
  - Acreage (b_1)
  - Money (b_{2j})
  - Acreage * Money (b_3)
- Random (variance components):
  - Residual variance of legislators within states (e_{ij})
  - Variance of intercepts across states (u_{oj})
  - Variance of slopes of money across state (u_{2j})

$$voting_{ij} = \underbrace{\boxed{b_{0_j} + b_1 * acres_j + b_{2_j} * money_{ij} + b_3 * acres_j * money_{ij}}}_{\text{Fixed Effects}} + \underbrace{\boxed{e_{ij} + u_{0_j} + u_{2_j}}}_{\text{Random Effects}}$$

COVARIANCE MATRICES

- The covariance matrix of a multilevel defines:
  - *How observations from the same group relate to one another*
- Easy defaults:
  - Only modeling a random intercept:
    - Use "*Variance Components*"
  - Repeated-measures data (e.g., diaries):
    - Use *"Autoregressive"* covariance matrix
  - Any complex structure (e.g., both between- and within- random effects):
    - Use "*Unstructured*" covariance matrix

COVARIANCE MATRIX FOR LEGISLATOR DATA

- Decision: *Unstructured*
- Reason:
  - We have both between-group (intercept) and within-group (money) random effects … that's getting kind of complicated
- How to do it:
  - **SPSS:** add "COVTYPE(UNR)" to the "RANDOM" line
  - **R**: default, but the "correlation" option in lme() is how you specify the covariance matrix (type `?lme` on the R command line for more info)

DEGREES OF FREEDOM ESTIMATION

- The method of df estimation in a multilevel determines how *df* are estimated
- We are restricted by our software, right now
  - **SPSS Example:** Satterthwaite
  - **R Example:** Between-within
  - *Note: If you used SAS, you can specify whatever approach you want using the "ddfm=" option on the model line!*

TAKING STOCK: STEP 2.1

- ☑ Outcome variable: ________Voting_________
- ☑ Predictors & Covariates: House + Money + Acres + Money* Acres
- ☑ Random effects: __Intercept + Money Slope__
- ☑ "Nesting"/grouping variable: ________State_________
- ☑ Covariance matrix: ______Unstructured_______
- ☑ Degrees of Freedom: Satterthwaite (SPSS) or Between/Within (R)

2.2. ANALYZE YOUR MODEL!

- Run the model!
- Visualize the output!
- *If* significant interaction:
  - Simple slopes testing

RUN THE MODEL IN SPSS!

- MIXED voting WITH house.effectc money.groupc acres.grandc

  /FIXED= house.effectc money.groupc acres.grandc money.groupc* acres.grandc

  /RANDOM=INTERCEPT money.groupc | SUBJECT(state) COVTYPE(UNR)

  /PRINT=SOLUTION.

RUN THE MODEL IN R!

- mlm.model <- lme(voting ~ house.effectc + money.groupc + acres.grandc + money.groupc * acres.grandc, random=~1+money.groupc|state)

  summary(mlm.model)

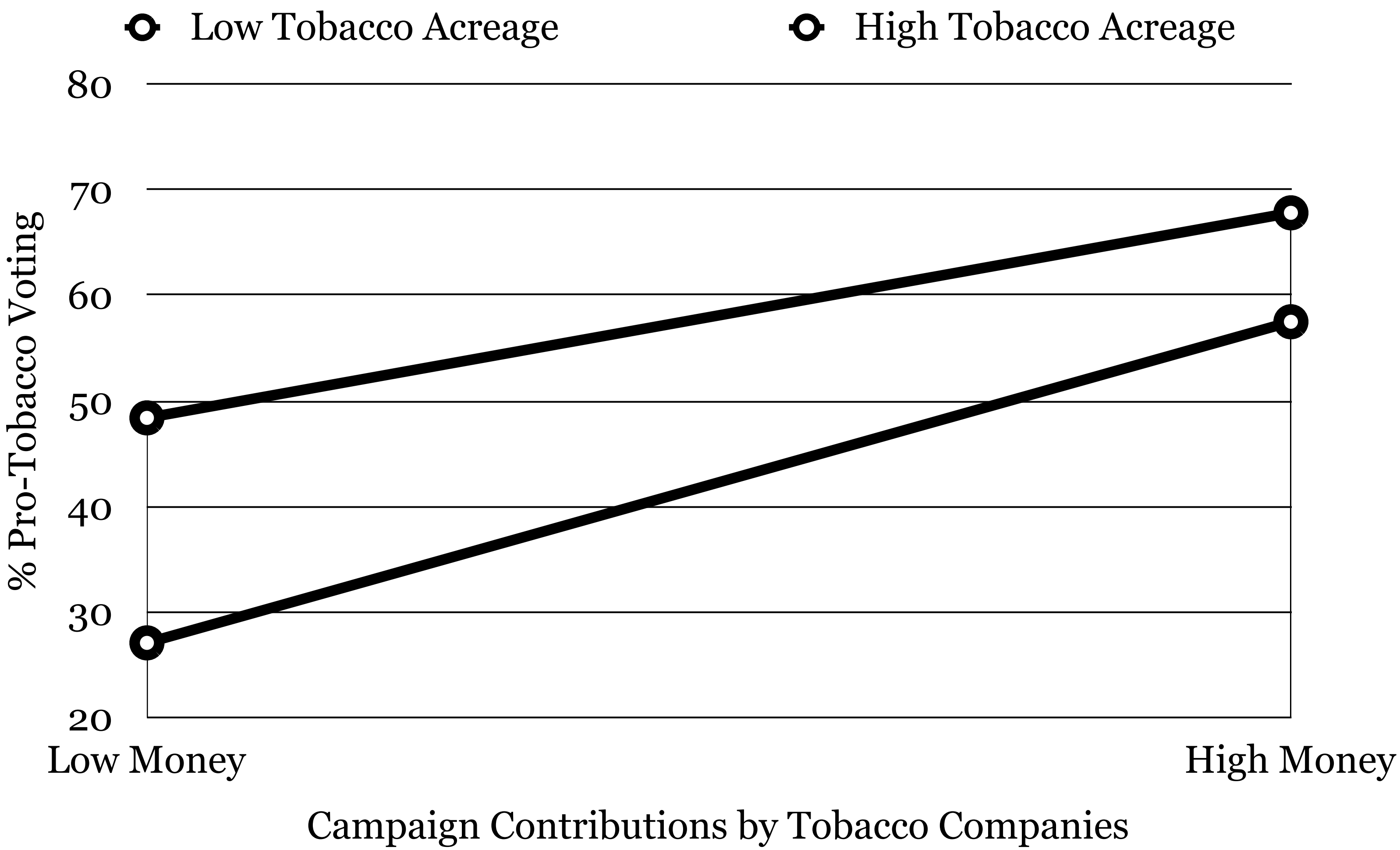
OUTPUT

- Look for:
  - Fixed effects table
  - Random effects table (called “Covariance Parameters” in SPSS)
  - Model evaluation criteria (for model comparisons)

VISUALIZE!

- Use the methods outlined in West, Aiken, & Krull (1996)
- Estimate marginal means by plugging relevant values into equation

PRO-TOBACCO VOTING AS A FUNCTION OF TOBACCO DONATIONS & ACREAGE



SIMPLE SLOPES TESTING

- Aiken & West (1991) outlined a universal method for testing simple effects (see also, West, Aiken, & Krull, 1996)

1. Rescale your predictor variables so that the zero-value represents +/- 1 SD from the mean

2. Rerun your analysis with the rescaled predictors, once for each simple effect test you want to run (ideally, an *a priori* decision)

3. See if the "lower order" effects of the **non-rescaled predictors** are significant

APPLYING AIKEN & WEST (1991) SIMPLE SLOPES METHOD TO MLM

- Deviation from typical method:
  - Use your rescaled variable in the fixed effects
  - Retain the mean-centered variable in the random effects
- Why?
  - You are probing the fixed effects, not the random effects

RECODE VARIABLES

- Two methods, depends on type of variable:
  - Categorical
  - Continuous

SIMPLE SLOPES FOR CATEGORICAL PREDICTORS

- For each categorical variable:
  - Make new variables with names representing each level of the category (e.g., “senate”, “representatives”)
  - Dummy code these new variables so that people in the target condition have a “0” and everyone else has a “1”

SIMPLE SLOPES FOR CONTINUOUS VARIABLES

- Simple-effects coding for continuous variables:
  - For “high” values:
    - Subtract the standard deviation from every CENTERED score
  - For “low” values:
    - Add the standard deviation to every CENTERED score

SIMPLE SLOPES FOR CONTINUOUS VARIABLES

- For the group-centered continuous predictor, *money.groupc*:
  - *SD* = 14.83548
- To test the simple effects, start with the centered variable, *money.groupc*:
  - Make two new variables: *lowmoney* and *highmoney*
  - *lowmoney* = *money.groupc* + 14.83548
  - *highmoney* = *money.groupc* - 14.83548

ALL SIMPLE SLOPES IN LEGISLATOR DATA

- Two variables to re-center around +/- 1 *SD*:
  - Acres
    - low.acres = acres.grandc + $SD_{acres.grandc}$
    - high.acres = acres.grandc - $SD_{acres.grandc}$
  - Money
    - low.money = money.groupc + $SD_{money.groupc}$
    - high.money = money.groupc - $SD_{money.groupc}$

RERUN ANALYSIS WITH RECODED PREDICTORS

- To test 4 simple slopes, you need to run 4 new models:
  - Effect of *Money* for states with low acreage:
    - voting = lowacres + money + lowacres * money
  - Effect of *Money* for states with high acreage:
    - voting = highacres + money + highacres * money
  - Effect of *Acres* for legislators with low campaign contributions:
    - voting = acres + lowmoney + acres * lowmoney
  - Effect of *Acres* for those with high campaign contributions:
    - voting = acres + highmoney + acres * highmoney

LOOK FOR SIMPLE EFFECTS

- Look for simple effects
- For each simple slope analysis:
  - Look at whether the main effect of the non-recoded variable is significant
  - Record the t-statistic and p-value for each

RERUN ANALYSIS WITH RECODED PREDICTORS

- 4 simple slopes tests:
    - Significant effect of *money* among states with little tobacco acreage, $t(474) = 6.65$, $p < .001$
    - Significant effect of *money* among states with high tobacco acreage, $t(474) = 5.16$, $p < .001$
    - Significant effect of *acres* among legislators with low tobacco company donations, $t(48) = 2.95$, $p = .005$
    - Significant effect of *acres* among legislators with high tobacco company donations, $t(48) = 2.31$, $p = .025$

2.3. EFFECT SIZE AND ICC CALCULATIONS

- Effect size
- Intraclass correlation

EFFECT SIZE IN MLM

- Effect size for the whole model
  - *Pseudo* R^2 (Snijders & Bosker, 1994)
- Partial effect sizes for specific model parameters
  - *Semi-partial* R^2 (Edwards, Muller, Wolfinger, Qaqish, & Schabenberger, 2008)

VARIANCE EXPLAINED

- R^2 has slightly different meaning between regression and MLM
- R^2 in normal regression
  - *Percentage of the dependent variable's variance that is explained by the predictor variables*
- R^2 in multilevel modeling
  - *Proportional reduction in prediction error*
  - Called "Pseudo R^2"

Snijders & Bosker, 1994, 1999

PSEUDO R^2 IN MLM

- Calculate a *Pseudo* R^2 at each level
- Interpretation:
    - Level 1 *Pseudo* R^2
        - *Proportional reduction of error when predicting an individual outcome*
    - Level 2 *Pseudo* R^2
        - *Proportional reduction of error when predicting a group-level mean*
- Use same criteria as normal R^2 to identify small, medium, and large *Pseudo*-R^2
    - Cohen (1992): Small $R^2 = 0.02$, Medium $R^2 = 0.13$, Large $R^2 = 0.26$

HOW TO CALCULATE *PSEUDO* R^2

- Run your multilevel model
  - Note the residual & intercept variances
- Run the "baseline model"
  - *Baseline model is a multilevel model with no predictors*
  - Note the residual & intercept variances

BASELINE MODEL IN SPSS

```
MIXED y
/FIXED=INTERCEPT
/RANDOM=INTERCEPT | SUBJECT(group)
/PRINT=SOLUTION.
```

BASELINE MODEL IN R

```
baseline.mlm <- lme( y ~ 1, random=~1|group )

summary( baseline.mlm )
```

R^2 IN MLM

Estimate of Level 1 Variance

$$R_1^2 = 1 - \frac{\left(\sigma_{u_0}^2 + \sigma_r^2\right)_{Comparison}}{\left(\sigma_{u_0}^2 + \sigma_r^2\right)_{Baseline}}$$

Estimate of Level 2 Variance

$n_{\text{observations}}$ within-group

$$R_2^2 = 1 - \frac{\left(\sigma_{u_0}^2 + (\sigma_r^2 / n)\right)_{Comparison}}{\left(\sigma_{u_0}^2 + (\sigma_r^2 / n)\right)_{Baseline}}$$

R^2 FOR LEGISLATOR DATA

$$R_1^2 = 1 - \frac{(.032130 + .074791)}{(.036390 + .092624)}$$

$$R_1^2 = 1 - \frac{0.106921}{.129014}$$

$$R_1^2 = 1 - 0.828755 = 0.171$$

R^2 FOR LEGISLATOR DATA

$$R_2^2 = 1 - \frac{(.032130 + .074791/12)}{(.036390 + .092624\ /\ 12)}$$

$$R_2^2 = 1 - \frac{(.032130 + 0.006233)}{(.036390 + 0.007719)}$$

$$R_2^2 = 1 - 0.8697318 = 0.130$$

SEMI-PARTIAL R²

$$R_\beta^2 = \frac{(df_{\text{numerator}} / df_{\text{denominator}}) * F}{1 + ((df_{\text{numerator}} / df_{\text{denominator}}) * F)}$$

- Where *df* is the denominator degrees of freedom for F
- How to obtain these numbers:
  - **SPSS:** Provides *F* and associated *df* for each effect by default
  - **R:** anova(mlm.model)

(Edwards, Muller, Wolfinger, Qaqish, & Schabenberger, 2008)

PARTIAL R^2 IN LEGISLATOR DATA

$$R^2_{Money} = \frac{1/474 * 53.4474}{1+(1/474 * 53.4474)} = \frac{0.112758}{1.112758} = 0.101$$

$$R^2_{Acres} = \frac{1/48 * 3.5006}{1+(1/48 * 3.5006)} = \frac{0.07292917}{1.07292917} = 0.068$$

$$R^2_{Acres*Money} = \frac{1/474 * 5.187221}{1+(1/474 * 5.187221)} = \frac{0.010943}{1.010943} = 0.011$$

POWER IN MLM

- Power in MLM is rather complicated … check out:
  - Bosker, Snijders & Guldemond (2003)
  - Mathieu, Aguinis, Culpepper, & Chen, 2012
- If you don't want to go all-out, ballparking your power is better than nothing
  - Start with the effect size of interest (either for the whole model or a partial effect)
  - Use guidelines of Cohen (1992) to determine power - available in "Supplemental Readings"
- But the relevant sample size depends on whether your predictors are above Level 1 or not

POWER IN ALL-LEVEL 1 MLM MODELS

- If all your predictors are at level 1 (e.g., money, house)
  - ~30 observations total for a large effect
  - ~70 observations total for a medium effect
  - ~85 observations total for a small effect
- Median Level 1 Sample Sizes = 198 (Dalton, Aguinis, Dalton, Bosco, & Pierce, 2012)

POWER IN MLM MODELS WITH LEVEL 2 PREDICTORS

- If some or all of your predictors are at Level 2 (e.g., acres, gross state product) …
  - ~30 groups for a large effect
  - ~70 groups for a medium effect
  - ~85 groups for a small effect
- Median Level 2 Sample Sizes = 51 (Mathieu *et al.*, 2012)

INTRACLASS CORRELATION (ICC)

- *A measure of how dependent observations within a group are on each other*
- ICCs as low as 0.1 can reflect sufficient clustering to affect linear model estimates
  - If ICC is sufficiently low (i.e., $\rho < .1$), then you don't have to use MLM!
  - Typical ICCs in the literature range from 0.15 - 0.30 (Mathieu *et al.*, 2012)
- You calculate the ICC **<u>from the baseline model</u>**

Estimate of Level 2 Variance

$$\rho = \frac{\sigma^2_{u_0}}{\sigma^2_{u_0} + \sigma^2_r}$$

Estimate of Level 1 Variance

ICC FOR LEGISLATOR DATA

$$\rho = \frac{.036390}{(.036390 + .092624)}$$

$$\rho = .2821$$

- Compare the value of ρ to published significance tables for the correlation coefficient, r, using your Level 1 n to determine significance (*hint*: Google for "calculate significance correlation")
- **Conclusion:**
  - *The voting behavior of legislators from the same state was clustered together*

TAKING STOCK: STEP 2

- ☑ You’ve run your analysis and have the output
- ☑ You’ve visualized/graphed your results
- ☑ You’ve calculated your effect sizes and ICC

3.1. DESCRIBING YOUR ANALYSIS

- What people want to know:
  - The type of multilevel model you conducted (e.g., random intercept? Random slope?)
  - Your "nesting" variable (Level 2 Grouping Variable)
  - Your DV, IVs, and covariates
  - What covariance matrix you used
  - The method of estimating degrees of freedom

DESCRIBING LEGISLATOR ANALYSIS

- Model specification: State ALL your DV, IVs, and covariates, and how they were coded
  - *"Pro-tobacco voting was modeled as a function of state tobacco acreage and campaign donations from tobacco companies, controlling for legislative house. Prior to analysis, state tobacco acreage was centered at the grand mean, campaign donations were centered at the group-mean, and legislative house was effect-coded such that senators were coded with '-1' and representatives received a '1.'"*
- Type of multilevel model conducted
  - *"A 2-level multilevel model was used …"*
- Nesting variable with random effects stated
  - *"… to account for congress people nested within state by estimating a random intercept and random slope of money for each state …"*
- What covariance matrix, *df* estimation, and partial R^2 method you used
  - *"… using an unstructured covariance matrix and the between-within method of estimating degrees of freedom. Effect sizes were estimated with semipartial R^2* (Edwards, Muller, Wolfinger, Qaqish, & Schabenberger, 2008)*"*

3.2. REPORTING YOUR RESULTS

- Statistic
  - The fixed effects for any parameter that you estimated (e.g., b) and its associated standard error, SE
  - The statistic that tests whether the parameter is different from 0 (e.g., t, F) and the associated degrees of freedom
  - Probability of observing that statistic
  - Effect size for that effect
- Visualization
- Results of simple effects testing (if applicable)

REPORTING RESULTS FOR LEGISLATOR DATA

- Main effects - Report *F*- or *t*-values of Fixed Effects:
  - "There was a significant, small main effect of tobacco acreage on pro-tobacco voting, b = 0.002, SE = .001, $t(48) = 2.87$, p = .006, *semi-partial* R^2 = 0.068."
  - "There was a significant, small main effect of campaign contributions from tobacco companies on pro-tobacco voting, b = 0.008, SE = .001, $t(474) = 7.27$, $p < .001$, *semi-partial* R^2 = 0.101."

REPORTING RESULTS FOR LEGISLATOR DATA

- Interaction - Report *F*-value of fixed effect
  - "As shown in Figure 1, campaign donations from tobacco companies significantly moderated the effects of state tobacco acreage on pro-tobacco voting, b = -0.000045, *SE* = .00002, $t(474) = -2.28$, $p = 0.023$, *semi-partial* $R^2 = 0.012$, albeit the effect size for the interaction effect suggested this difference was not practically significant."
- Simple Slopes - Report *t*-values of fixed effects
  - "Simple slopes were examined at one standard deviation above and below the means of both predictors (Aiken & West, 1991). This analysis revealed that tobacco company campaign contributions predicted more pro-tobacco voting among legislators from states with low tobacco acreage, $t(474) = 6.65$, $p < .001$, and states with high tobacco acreage, $t(474) = 5.16$, $p < .001$. State tobacco acreage also predicted more pro-tobacco voting when campaign contributions from tobacco companies were both low, $t(48) = 2.95$, $p = .005$, and when campaign contributions from tobacco companies was high, $t(48) = 2.31$, $p = .025$."

REPORTING MODEL R^2

- "We calculated Pseudo R^2 for the model at each level according to the recommendations of Snijders & Bosker (1994, 1999). At the lowest level, the model reduced prediction error of pro-tobacco voting by a medium amount for any given legislator, $Pseudo\text{-}R^2_1 = .171$. At the second level, the model reduced prediction error of pro-tobacco voting by a medium amount for any given state, $Pseudo\text{-}R^2_2 = .130$. The intraclass correlation coefficient was significant, $\rho = 0.282$, $t(525) = 6.40$, $p < .001$, suggesting that the voting behavior of legislators from the same state were not independent and confirming that a multilevel analysis was necessary for these data."

THINGS TO KEEP IN MIND

THINGS TO KEEP IN MIND

- Research design
- Assumptions of multilevel models
- Importance of unstandardized coefficients in MLM

RESEARCH DESIGN

- Things to always remember:
  - Measure the same variables for every observation
  - Make sure to record the grouping variable
  - Think about your model BEFORE you collect your data
    - Try to make the levels as clear cut as possible

ASSUMPTIONS OF MLM

1. Level 1 residuals are normally distributed around zero
2. Level 2 residuals are multivariate-normal with a mean of zero
3. Level 1 residual variance is homoskedastic both within- and between-groups
4. Level 1 residuals and Level 2 residuals are uncorrelated
    - If your data violate this assumption, you have options (see Culpepper, 2010)
5. Level 1 observations are distinctly classified into Level 2 groups
    - If not, use a cross-classified model (c.f., *Advanced Applications*)

UNSTANDARDIZED COEFFICIENTS

¡¡NEVER STANDARDIZE ALL YOUR VARIABLES BEFORE RUNNING A MULTILEVEL MODEL!!

STANDARDIZED COEFFICIENTS

- It totally messes the whole thing up
  - Your slopes and intercepts are *wrong*
  - This has to do with ye ole covariance matrix
- More general criticism of standardized coefficients:
  - They force the assumption of homoskedasticity on the data
- If you really want an effect size, calculate effect size!

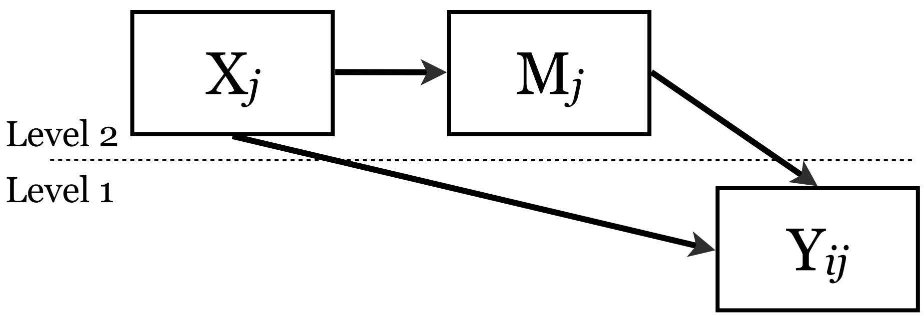
ADVANCED APPLICATIONS OF MLM

MLM APPLICATIONS

- Multilevel mediation
- *N*-level models
- Nested growth curves
- Generali*zed* Linear Modeling
    - Poisson: Count/Frequency data
    - Bernoulli: Logistic Regression
- Bootstrapping multilevel models
- Cross-classification

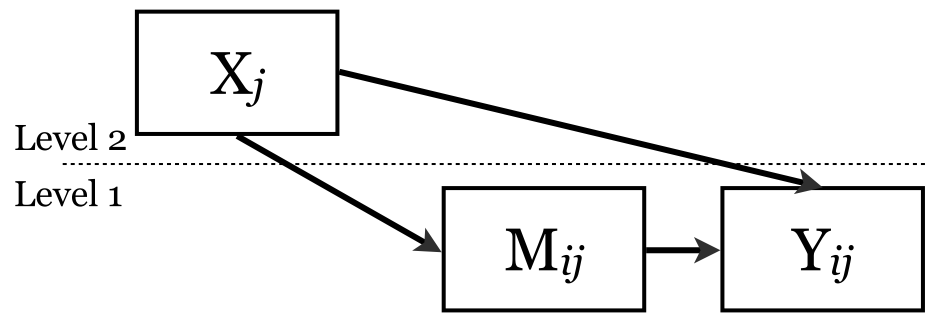
TYPES OF MLM MEDIATION

<u>Normal Causal Steps Approach</u>

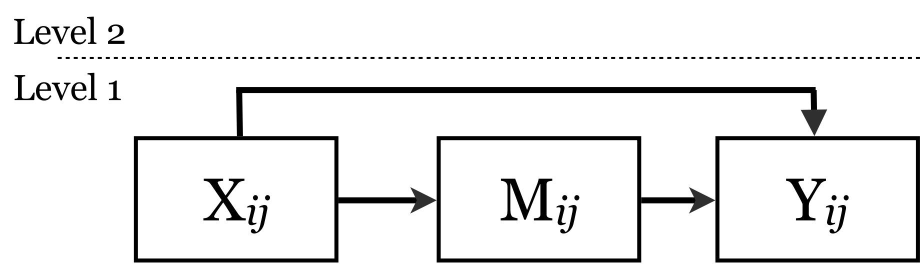


"2-2-1 Mediation"

<u>Causal Steps With Extra Predictors</u>



"2-1-1 Mediation"



"1-1-1 Mediation"

THINGS TO CONSIDER WITH MLM MEDIATION

- *Does my mediational model apply equally to all the groups?*
- *Does my mediational model capture group-level or individual-level processes?*
- Much richer introduction to multilevel mediation:
    - Andrew Hayes' APS Workshop: 1:30 - 3:20 TODAY
        - Room "Wilson A"

THINGS TO CONSIDER WITH MLM MEDIATION

- *Does my mediational model apply equally to all the groups?*
  - This is only an issue if <u>*both*</u> of the slopes between the predictor and the mediator and between the mediator and the outcome variable are random
  - If both these slopes are random, you must calculate the population covariance
  - If so, you will need to compute the population covariance, σ_{ab}
    - *Estimates how reliably your mediational model explains the data across your level 2 units*
    - Kenny, Korchmaros, & Bolger (2003)

THINGS TO CONSIDER WITH MLM MEDIATION

- *Does my mediational model capture group-level or individual-level processes?*
  - *Note:* This is a very active area of current research
- Recommended readings:
  - Zhang, Zyphur & Preacher (2009)
  - Preacher, Zyphur, & Zhang (2010)
  - Preacher, Zhang, & Zyphur (2011)

N-LEVEL MODELS

- Theoretically, you can run models with any number of levels
- Must have sufficient power at the top level

N-LEVEL MODELS

- 2-levels - e.g., children nested in classrooms
- 3-levels - e.g., children nested in classrooms that are nested in schools
- 4-levels - e.g., children nested in classrooms that are nested in schools that are nested in school boards
- 5-levels - e.g., children nested in classrooms that are nested in schools that are nested in school boards that are nested in states
- 6-levels - e.g., children nested in classrooms that are nested in schools that are nested in school boards that are nested in states/provinces that are nested in countries
- ... the limit is decided by your population, data, and empirical design

3-LEVEL MODELS IN SPSS

```
MIXED y WITH x1 x2

/FIXED=x1 x2 x1*x2

/RANDOM=INTERCEPT  | SUBJECT( Level3Group * Level2Group )

/RANDOM=INTERCEPT  | SUBJECT( Level3Group ).
```

3-LEVEL MODELS IN R

```
three.level.mixed <- lme( y~x1*x2, random=~1|Level3Group/Level2Group)

summary(three.level.mixed)
```

NESTED GROWTH CURVES

- You have:
  - Multiple observations from participants who are nested in groups
  - E.g.,: Change in husbands' and wives' health symptoms over time

NESTED GROWTH CURVES

- How to implement:

1. Record measurement number (e.g., “time”)
2. Include “time” as a moderator in your fixed effects model
3. Include the slope of “time” as a random effect
4. Use an unstructured covariance matrix

NESTED GROWTH CURVES IN SPSS

- MIXED y WITH x time

  /FIXED= x time x*time

  /RANDOM=INTERCEPT time | SUBJECT(group*individual) COVTYPE(UNR)

  /RANDOM=INTERCEPT time | SUBJECT(group) COVTYPE(UNR)

  /PRINT=SOLUTION.

NESTED GROWTH CURVES IN R

- growth <- lme(y~x*time, random=~1+time|group/individual)

  summary(growth)

GENERAL*IZED* LINEAR MODELING

- You can conduct multilevel modeling on dependent variables that are not normally-distributed
- Popular cases:
    - Poisson
        - *Used for rare frequency-count outcomes (i.e., integers ≥ 0 and whose expected value is less than 10)*
    - Bernoulli
        - *Used for binary outcomes (i.e., 0 or 1)*

OUTCOME = COUNTS

- Poisson Regression
  - Distribution: Poisson
  - Link Function: Log
- Example: *How many representatives from a state, given its economic productivity?*

POISSON MLM IN SPSS

- GENLINMIXED

```
 /FIELDS TARGET=y
/TARGET_OPTIONS DISTRIBUTION=POISSON LINK=LOG
/FIXED  EFFECTS=x USE_INTERCEPT =TRUE
/RANDOM USE_INTERCEPT =TRUE SUBJECTS =group.
```

POISSON MLM IN R

- library(lme4)

  poisson.mlm <- lmer(y ~ (1|group) + x, family=poisson)

  summary(poisson.mlm)

ICC FOR POISSON REGRESSION

- Generalized linear modeling treats residual variances as constant
  - For Poisson models, the residuals are assumed to be standardized (i.e., $\sigma^2{}_r = 1$)
- As such, ICC and R^2 are calculated with $\sigma^2{}_r = 1$ (Snijders & Bosker, 1999)

$$ICC_{Poisson} = \rho = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + 1}$$

OUTCOME = YES/NO

- Logistic regression
  - Distribution: Binomial with 1 trial (i.e., Bernoulli distribution)
  - Link function: Logit
- Example: *Likelihood of being from a particular party, based on history of pro-tobacco voting*

LOGISTIC MLM IN SPSS

- GENLINMIXED

  /FIELDS TARGET=y

  /TARGET_OPTIONS REFERENCE =0
  DISTRIBUTION=BINOMIAL LINK=LOGIT

  /FIXED EFFECTS=x USE_INTERCEPT =TRUE

  /RANDOM USE_INTERCEPT=TRUE SUBJECTS=group.

LOGISTIC MLM IN R

- logistic.mlm <- lmer(y ~ (1|group) + x, family=binomial)

  summary(logistic.mlm)

ICC FOR LOGISTIC REGRESSION

- Generalized linear modeling treats residual variances as constant
- For logistic models, ICC and R^2 are calculated with $\sigma^2_r = \pi^2/3 = 3.29$ (Snijders & Bosker, 1999)

$$ICC_{Logistic} = \rho = \frac{\sigma^2_{u_0}}{\sigma^2_{u_0} + \frac{\pi^2}{3}}$$

BOOTSTRAPPING MLM

- When to use it:
  - *Your dependent variable is not distributed according to a known probability distribution*
- Solution:
  - *You build your own comparison distribution, from which you will draw tailored p-values*
- How it works:
  - A new comparison distribution is built by drawing samples with replacement of the same size as the original sample from the observed data
    - This process is repeated k times (k = ~5000 is sufficient for stable estimates)
  - The MLM is run on each resampled dataset, and you get a confidence interval for all k parameter estimates

BOOTSTRAPPING MLM IN SPSS

- BOOTSTRAP

  /VARIABLES TARGET=y INPUT= x1 x2

  /CRITERIA NSAMPLES=5000.

- MIXED y WITH x1 x2

  /FIXED=x1 x2 x1*x2

  /RANDOM=INTERCEPT | SUBJECT(group)

  /PRINT=SOLUTION .

BOOTSTRAPPING MLM IN R

- library(boot)
- boot.lmer <- function(data, indices) {

  data <- data[indices,]

  mlm.model <- lmer(y~(1|group) + x1*x2, data=data)

  fixef(mlm.model)

  }
- bootstrapped.mlm <- boot(legislator.data, boot.lmer, 5000)
- boot.ci(bootstrapped.mlm, type="perc", index=1) #iterate through the "indexes" of each parameter in the model

REPORTING THE RESULTS OF BOOTSTRAPPED ANALYSIS

- Additional information:
  - Number of bootstrap resamples/iterations
- Evaluate significance of model parameters with confidence intervals:
  - *APA Style for Confidence Intervals:*
    - 95% CI [lower.bound, upper.bound]
    - E.g., "Within each state, legislators who received relatively more campaign contributions from tobacco companies were more likely to vote pro-tobacco, *b* = 0.008, 95% CI [.006, .010]."
  - You do not need to report *t*-, *F*-, or *p*-values for the slopes if you report the confidence interval

CROSS-CLASSIFICATION

- *When you have more than 1 way you can nest your variables*
- Example:
  - Legislators are nested in both states and parties
  - The hierarchical relationship between states and parties is unclear; they appear to be at the same level
- How to implement:
  - Model a random intercept (or slope) for each group

CROSS-CLASSIFIED MODELS IN SPSS

- MIXED y WITH x1 x2

  /FIXED=x1 x2 x1*x2

  /RANDOM=INTERCEPT | SUBJECT(group1)

  /RANDOM=INTERCEPT | SUBJECT(group2)

  /PRINT=SOLUTION.

CROSS-CLASSIFIED MODELS IN R

- cross.classified <- lmer(y ~ (1|group1)+(1|group2) + x1 * x2)

  summary(cross.classified)

MULTILEVEL MODELLING

- What it is
  - *An extension of regression where parameters (i.e., intercept, slopes) are predicted, in addition to predicting the outcome*
- When to use it
  - *Your data is hierarchical in nature; your observations are not independent*
  - You are <u>not</u> limited by the distribution of your dependent variable, because you can use a generalized linear mixed model or bootstrap your analysis
- How many levels?
  - *When the levels are clear-cut, then however many seem appropriate*
  - *When the lowest-level of observation could be classified into one group OR another (i.e., the category lines are not rigid), then you use a cross-classified model*

¡¡THANK YOU!!

- Questions and feedback:
  - elizabeth.page-gould@utsc.utoronto.ca
  - http://page-gould.com/mlm/aps
- Workshop Sponsors:
  - Association for Psychological Science
  - Society of Multivariate Experimental Psychology
- Other Funding Sources:
  - Social Sciences and Humanities Research Council of Canada
  - Connaught Laboratories Fund

- Some good MLM reference books:
  - **(SPSS-focused)** Bickel, R. (2007). Multilevel analysis for applied research: It's just regression! New York, NY, US: Guilford Press.
  - **(R-focused)** Wright, D. B., & London, K. (2009). Modern regression techniques using R: A practical guide for students and researchers. Thousand Oaks, CA, US: Sage Publications.
  - **(SAS-focused)** Singer, J. D., & Willett, J. B. (2003). Applied longitudinal data analysis: Modelling change and event occurrence. New York, NY, US: Oxford University Press.

NOTES ABOUT DISTRIBUTION OF WORKSHOP MATERIALS

- All workshop materials (slides and syntax) are the original work of Elizabeth Page-Gould, distributed to you with Creative Commons 3.0 International License
  - This means that **YOU MAY**:
    - Freely share, distribute, and even "remix" this work
  - **BUT ONLY UNDER THE FOLLOWING CONDITIONS:**
    - *With attribution:* You provide attribution to Elizabeth Page-Gould with a link to the original workshop materials: http://page-gould.com/mlm/aps
    - *Share alike*: If you alter or remix this work in any way, you must also share your final product with a license that has similar conditions to this one (e.g., to be distributed freely)
    - *Non-commercial:* You may not use these materials for commercial purposes without explicit permission from Elizabeth Page-Gould
- *Note:* I am indebted to Eran Bar-Kalifa (Bar-Ilan University) for identifying two important typos, one of which led to an overestimation of the Level 2 *pseudo-*R^2 in the original slides and workshop. When the correct values were used, the Level 2 *pseudo-*R^2 became a medium-sized effect. This has been fixed as of 2013/05/31. Thank you, Eran!