MULTILEVEL MODELING WORKSHOP

Elizabeth Page-Gould

University of Toronto

WORKSHOP SPONSORS

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

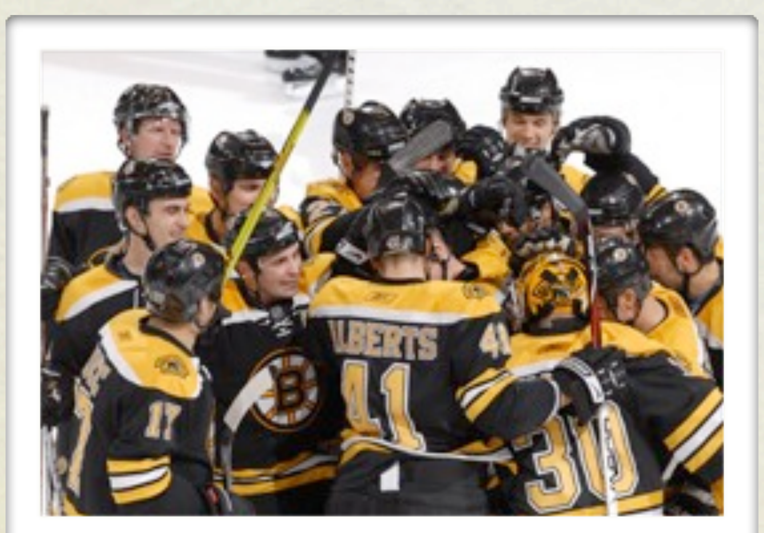
MULTILEVEL MODELING

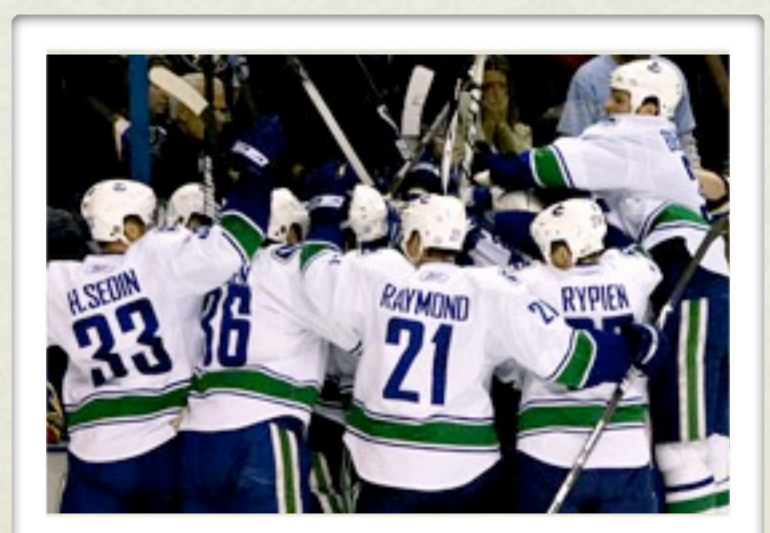
- *A broad class of analyses that deal with hierarchy in your data*

HIERARCHICAL DATA

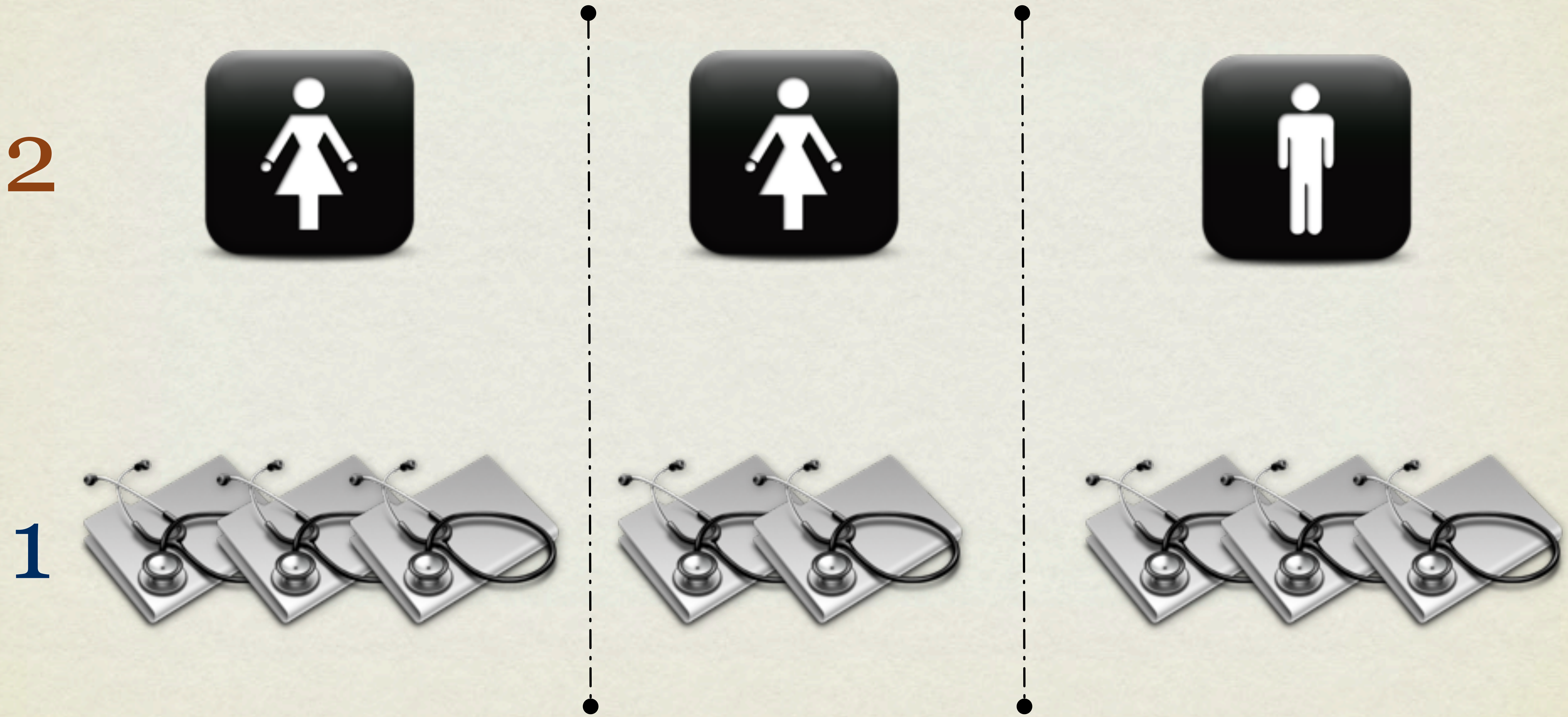
2

1





HIERARCHICAL DATA



MULTILEVEL MODELING

- Broad class of techniques
- 1 face, many names:
    - Hierarchical linear modeling (HLM)
    - 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 HANDOUTS

- Grab a physical copy for note taking now
- Download: http://page-gould.com/mlm/




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

2

California

New York

Massachusetts

1

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
...	...	...	...	...

REGRESSION FUNDAMENTALS

- Descriptive statistics
  - Centrality and Spread
  - Explaining variance: The Grand Prize
- Prediction: Correlation and Covariance
- Simple and Multiple Regression

MEAN

$$\overline{X} = \frac{\sum_{i=1}^{N} x_i}{N}$$

- *Average, expected value*
- Closest midpoint to all data in the sample

CENTRALITY IN LEGISLATOR DATA

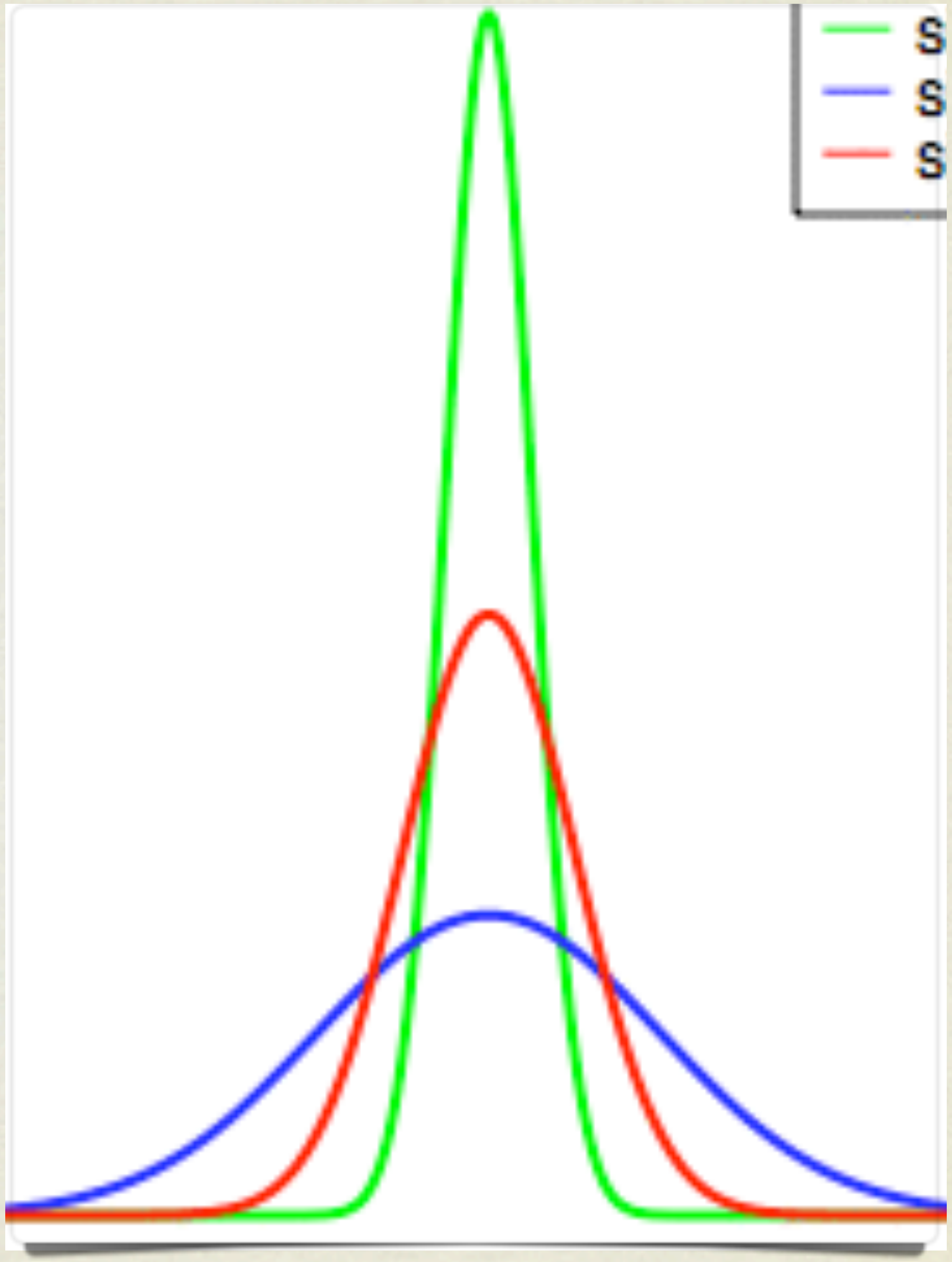
FREQUENCIES VARIABLES=x1 x2 y

/STATISTICS=MEAN MEDIAN MODE.

	Contributions *(money)*	Acreage *(acres)*	% Pro-Tobacco Votes *(voting)*
Mean	12.96	13.01	51.80
Median	4.5	0	59.46
Mode	0	0	1

SPREAD

- *Dispersion of a distribution*
- Variance, Standard Deviation, Range, Interquartile Range



SPREAD IN LEGISLATOR DATA

```
FREQUENCIES VARIABLES=x1 x2 y
 /NTILES=4
 /STATISTICS VARIANCE STDDEV.
```

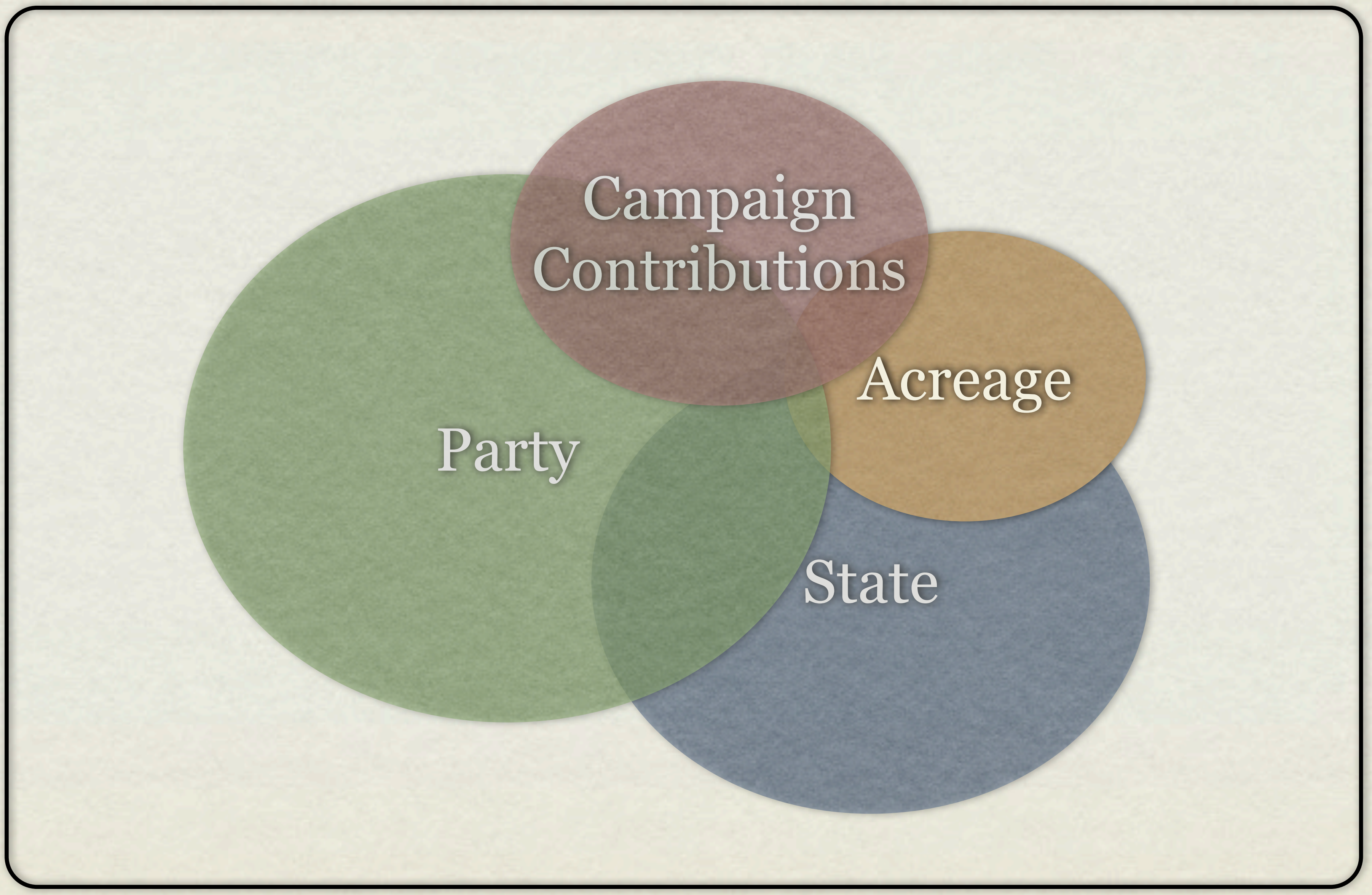
	Contributions *(money)*	Acreage *(acres)*	% Pro-Tobacco Votes *(voting)*
Variance	339.95	1727.42	0.121
Standard Deviation	18.44	41.56	0.348
Interquartile Range	0 - 20.5	0 - 5.8	.2 - .85

VARIANCE

$$\sigma^2 = \frac{\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2}{N-1}$$

VARIANCE

Total Variance in Pro-Tobacco Voting:



EXPLAINING VARIANCE IS THE GRAND PRIZE

- Almost any classical statistic compares:

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

WHAT'S SO SPECIAL ABOUT VARIANCE?

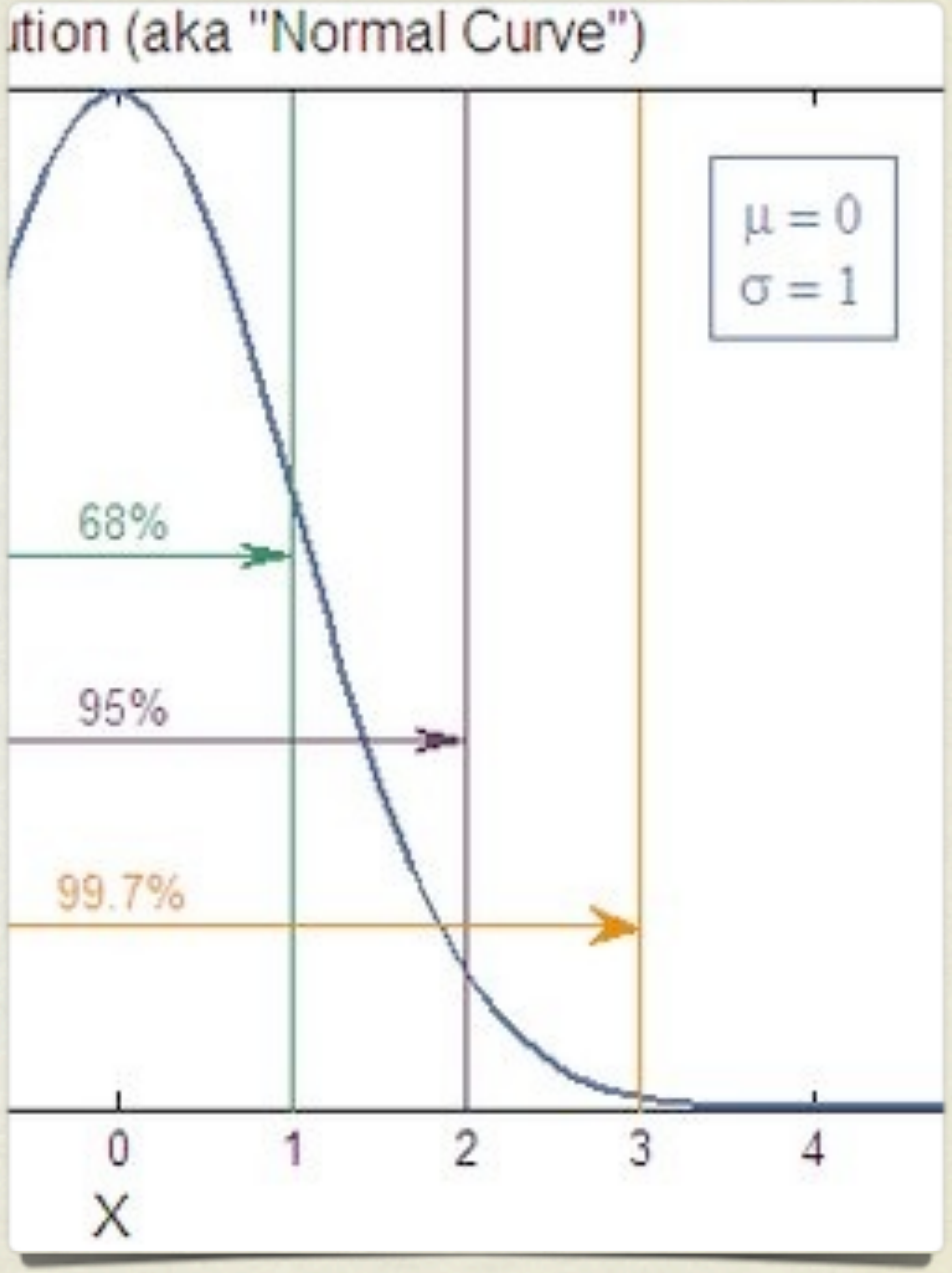
- Classical statistic's "Standard Candle"
  - Standard deviations are the unit of measurement
  - We know the *probability* of observing any deviation

IMPORTANT ASSUMPTIONS OF CLASSICAL STATISTICS

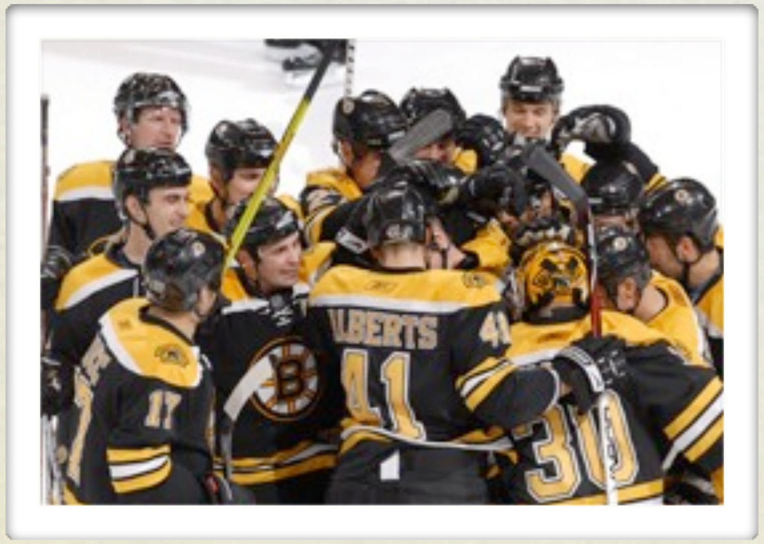
- Assumptions:
  - Data were collected through **random sampling**
  - All data are **normally distributed**
  - **Variance must be equal** across conditions
  - All observations must be **independent**
- If your data violate an assumption:
  - *Transform it if you can, or*
  - *Accept a decrease in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

KEY ASSUMPTION RELATIVE TO MLM:

- **All observations must be independent**



DEPENDENCE IS WHERE IT'S AT



PREDICTION

- Predictive relationships
  - Correlation & covariance
  - Regression

CORRELATION & COVARIANCE

- *How much 2 variables change together*

COVARIATION IN LEGISLATOR DATA

CORRELATIONS VARIABLES=x y

/STATISTICS XPROD.

	Money * Voting	Acres * Voting
Correlation	0.46	0.25
Covariance	$COV = 2.98$	$COV = 3.60$

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})

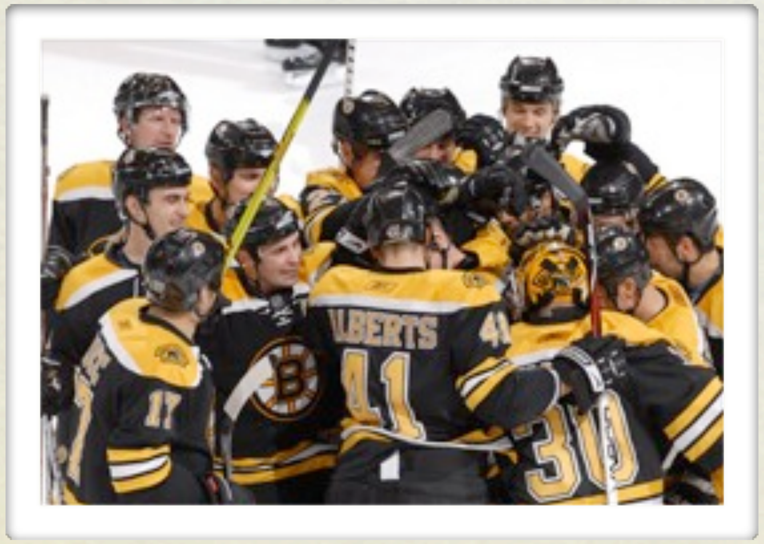
$$COV_{XY} = r_{XY}\sigma_X\sigma_Y$$

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

WHY DO WE CARE?

- Covariation *is* your dependence!

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



WHY DO WE CARE?

- Covariation <u>*is*</u> your dependence!

California

$$COV_{X_{1_{CA}} X_{2_{CA}}}$$

New York

$$COV_{X_{1_{NY}} X_{2_{NY}}}$$

Massachusetts

$$COV_{X_{1_{MA}} X_{2_{MA}}}$$

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

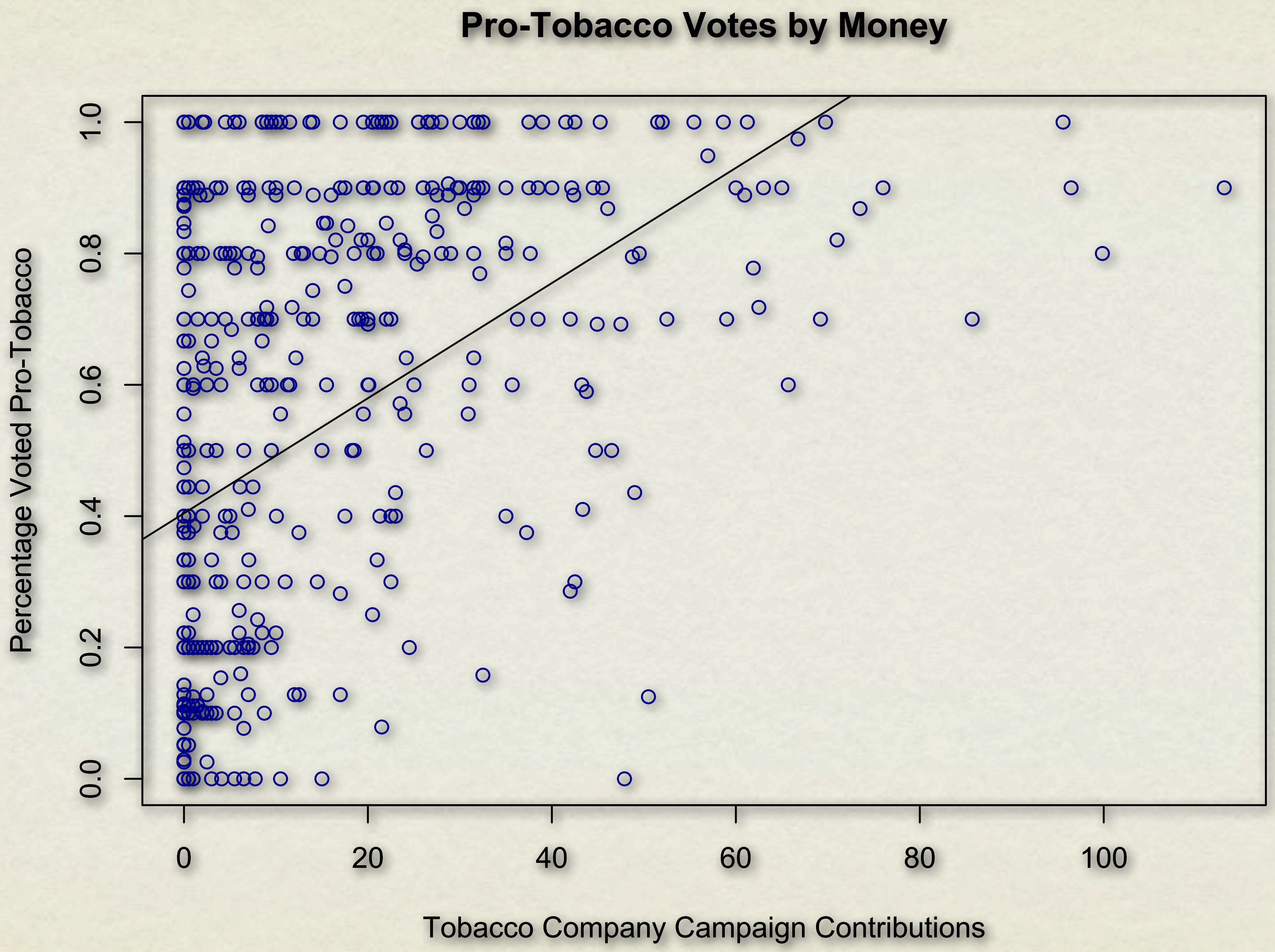
$$\boxed{\hat{y}_i} = \boxed{b_0} + \boxed{b_1} x_i + \boxed{e_i}$$

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

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

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

REGRESSION



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

REGRESSION

$$y_i = \underbrace{b_0}_{\text{Weighted Mean(Y)}} + \underbrace{b_1}_{\text{Influence of X}} x_i + \underbrace{e_i}_{\text{Stuff You Can't Explain}}$$

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

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

REGRESSION

/DEPENDENT y

/METHOD=ENTER x.

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*

$$\widehat{y_i} = b_0 + b_1 x_{1_i} + b_2 x_{2_i} + \left(b_1 x_{1_i}\right)\left(b_2 x_{2_i}\right)$$

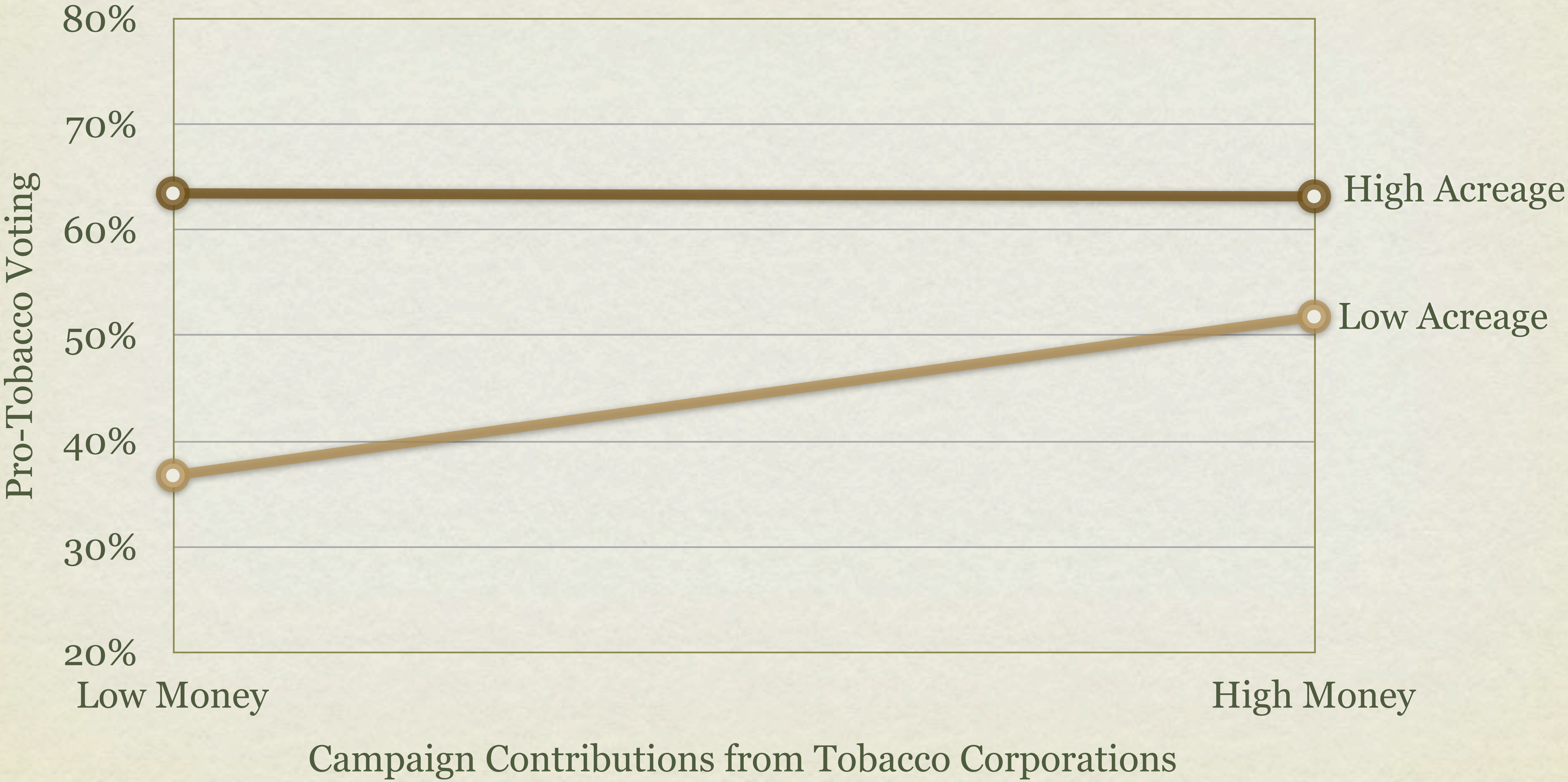
REGRESSION AND LEGISLATOR DATA

```
COMPUTE x1Xx2 = x1*x2.

REGRESSION

/DEPENDENT y

/METHOD=ENTER x1 x2 x1Xx2.
```

Pro-Tobacco Voting=	$b_0 + b_1(\text{Money}) + b_2(\text{Acres}) + b_3(\text{Money*Acres})$
Estimate	= .538 + .009 (money) + .002 (acres) - .00005 (money*acres)

MODERATION IN THE LEGISLATOR DATA

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



BUT WE HAVE TO DO SOMETHING ABOUT THE LEVELS

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

WHAT TO DO ABOUT GROUPS?

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

SOLUTIONS

- Alternatives to MLM:

  1. Aggregate your level 1 variables

  2. Random effects models

- 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 intercept model/random effects ANOVA
  - Random slope models
- What's random about the intercepts and slopes?
  - They are *predicted*
  - So they have *error*

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

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

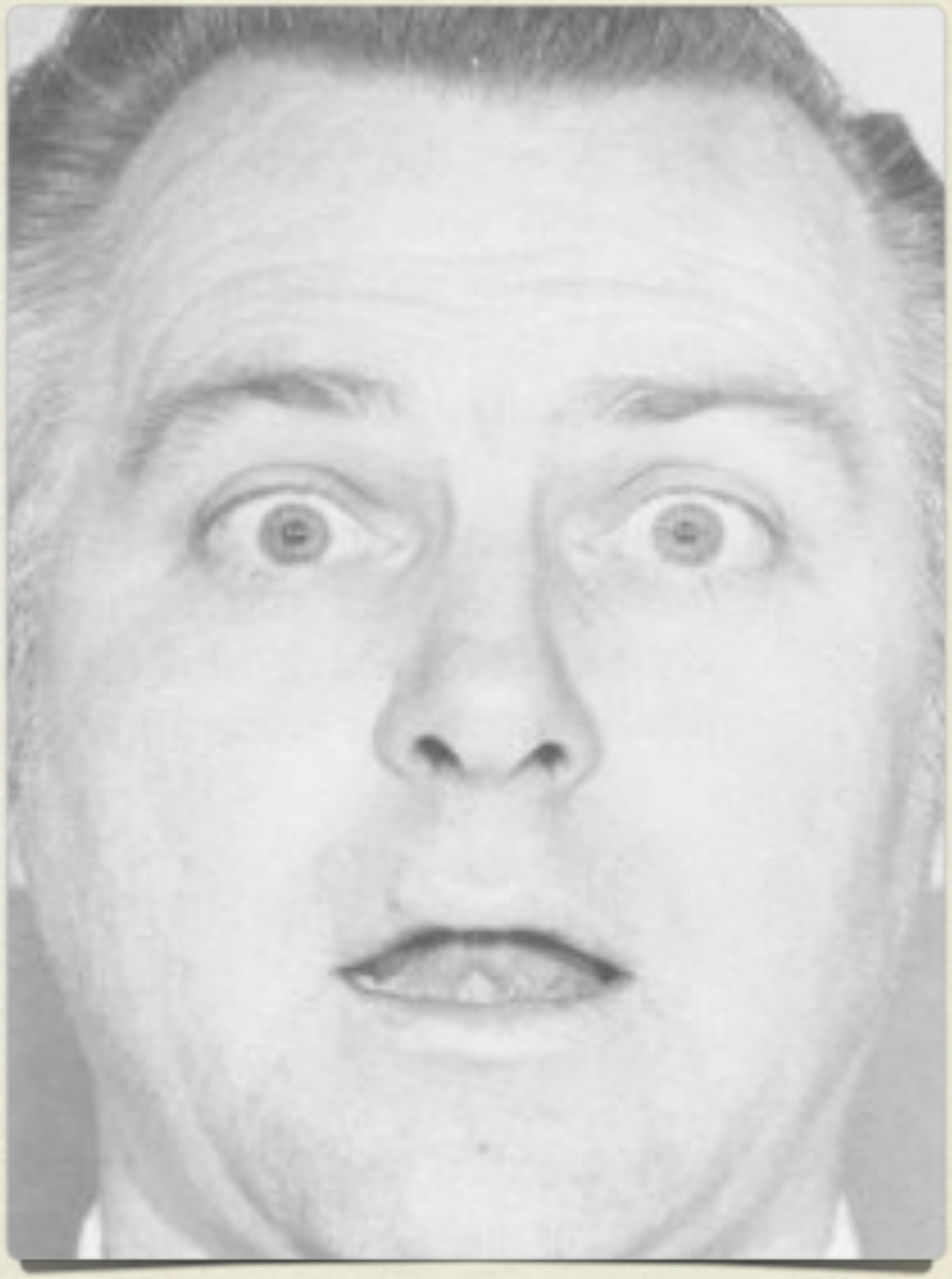
- A unique slope is predicted for each group

WHAT VARIES BETWEEN YOUR GROUPS?

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

WHOA!

- You just learned multilevel modeling!
- Random effects models *are* multilevel models



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: $$\widehat{y_i} = b_0 + b_1 x_i$$

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

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

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

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

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

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 MODEL

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

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

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

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

MULTILEVEL MODEL

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

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

HOW DO I ESTIMATE THE PARAMETERS?

- Thankfully, a computer does it for you using 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:
  - <u>Variance Components</u> - Default in SPSS and SAS, assumes that data points from different groups do not covary
  - <u>Autoregressive</u> - Standard for basic longitudinal designs, assumes that data points next to each other will be highly correlated
  - <u>Unstructured</u> - Default in R, assumes nothing about covariation structure, best for complicated multilevel models, robust against issues like heteroskedasticity
- Great online resource: <u>http://courses.ttu.edu/isqs5349-westfall/images/5349/mixed_covariance_structures.htm</u>

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> - Default in SPSS and SAS and most widely-used method, most akin to a classic ANOVA or regression
        - Note that your *df* will have **decimal** points
    - <u>Between-Within</u> - Only method used by R (and thus the default), more conservative, is robust to complex hierarchical structures

MULTILEVEL MODELING IN SPSS!

- **Random Intercept only:**

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

MULTILEVEL MODELING IN SPSS!

- **Random Intercept & Random Slope:**

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

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)
- 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 variables as necessary to make them normal (if applicable)
   1.2. Dummy coding
   1.3. Centering
   1.4. 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
3. Report results
   3.1. Describing the analysis
   3.2. Reporting the results
   3.3. Effect size & ICC

1.4. 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

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

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

TAKING STOCK: STEP 1

- ☑ Your data are plausible!
- ☑ Your data are normal!
- ☑ Your data is centered!
- ☑ Your data is correctly dummy coded!
- ☑ Your data is correctly organized

2. RUN AN ANALYSIS

2.1. Specify your model

2.2. Analyze the model

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, public policy, 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)
  - Money (Level 1)
  - Will look at both main effects + interaction:

    = Money + Acres + Money * Acres
- Covariates:
  - House (Level 1)

    = House + Money + Acres + Money * Acres

GROUPING VARIABLES

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

GROUPING IN LEGISLATOR DATA

- Level 1: Each *Legislator's* campaign contributions and pro-tobacco voting
- Level 2: Each *State's* tobacco acreage

EFFECTS SPECIFICATION

- Fixed versus random effects
- Covariance matrices
- Method for estimating degrees of freedom

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:
  - Acreage
  - Money
- Random:
  - Intercept for each state

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: *Variance Components*
- Reason:
  - Only estimating a random intercept for each state
  - Assumes that voting patterns of legislators from the same state are simply correlated with each other

DEGREES OF FREEDOM ESTIMATION

- The method of df estimation in a multilevel determines how *df* are estimated
- Easy default: Satterthwaite
  - Most similar to a classical analysis
- If you want to be conservative: Between-within
  - Will give you the lowest degrees of freedom for tests of parameter estimates

DF ESTIMATION IN LEGISLATOR DATA

- Decision: *Satterthwaite*
- Reason:
  - Simple data structure: Only estimating a random intercept for each state
  - We are just using MLM to avoid violating statistical assumptions, and thus want an analysis most similar to classic regression

TAKING STOCK: STEP 2.1.

- ☑ Outcome variable: ____Voting____
- ☑ Predictors & Covariates: House + Money + Acres + Money* Acres
- ☑ Random effects: ____Intercept____
- ☑ “Nesting”/grouping variable: ____State____
- ☑ Covariance matrix: ____Variance Components____
- ☑ Degrees of Freedom: ____Satterthwaite____

2.2. ANALYZE YOUR MODEL!

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

RUN THE MODEL!

- **SPSS:**
  - MIXED voting WITH house money acres

    /FIXED=house money acres money*acres

    /RANDOM=INTERCEPT | SUBJECT(state)

    /PRINT=SOLUTION.

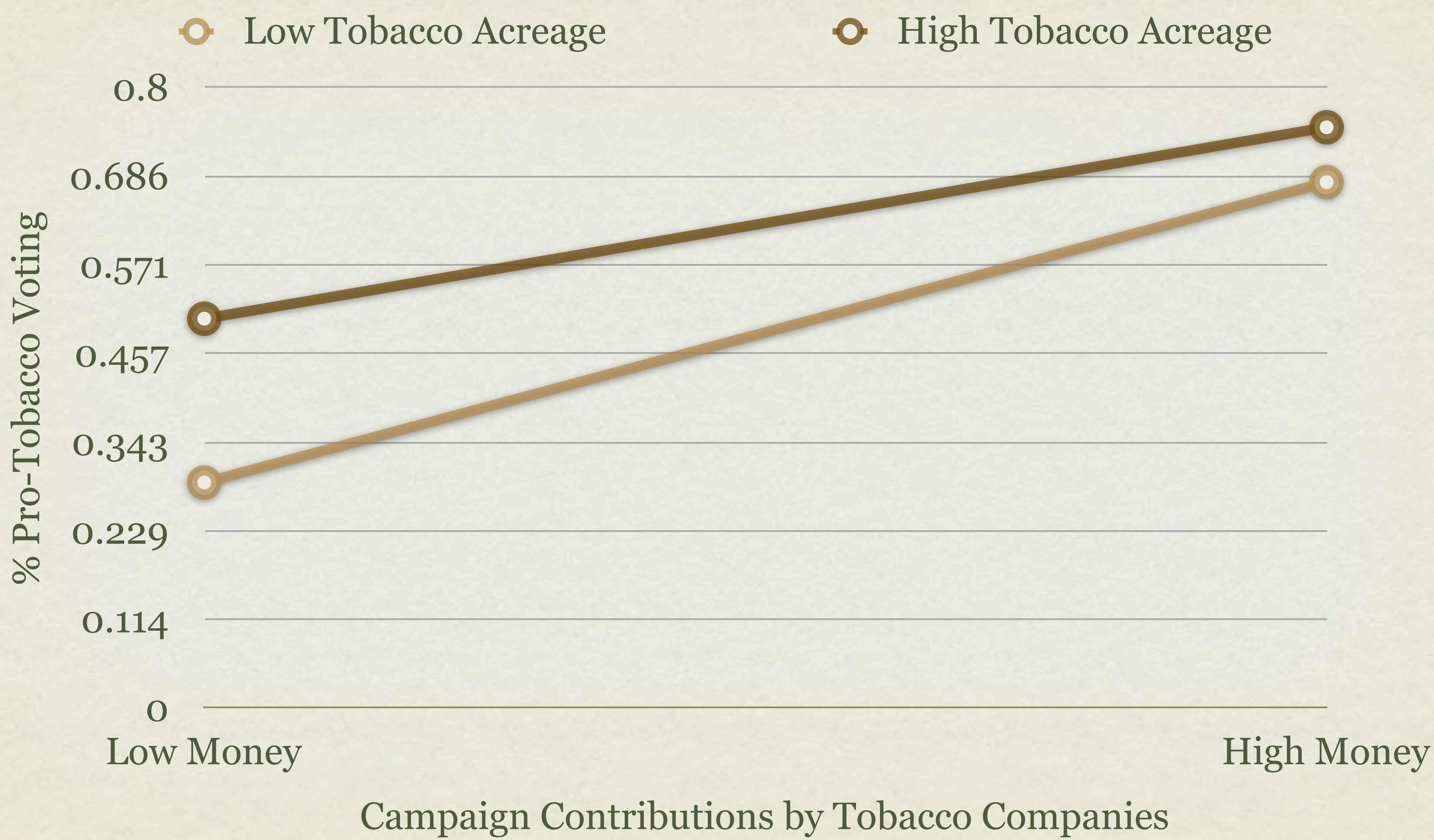
OUTPUT

- Look for:
  - Fixed effects table
  - Random effects table
  - Model evaluation criteria

VISUALIZE!

- Copy estimates into Excel Spreadsheet
- 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

1. Recode your predictor variables
2. Rerun your analysis with the recoded predictors
3. See if the “lower order” effects of non-recoded variables are significant

RECODE VARIABLES

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

SIMPLE SLOPES FOR CATEGORICAL VARIABLES

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

SIMPLE SLOPES FOR CATEGORICAL VARIABLES

- In the omnibus analysis, the house of each legislator is denoted by the variable “*house*”:
  - Senate (*house* = -1)
  - Representatives (*house* = 1)
- To test the simple effects for legislators in each condition:
  - Make two new variables: *senate* and *representatives*
  - Senators get values of: *senate* = 0, *representatives* = 1
  - Representatives get values of: *senate* = 1, *representatives* = 0

SIMPLE SLOPES FOR CONTINUOUS VARIABLES

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

SIMPLE SLOPES FOR CONTINUOUS VARIABLES

- For continuous variable, *money* (prior to centering):
  - M = 12.96
  - SD = 18.44
- To test the simple effects, start with the centered variable, *money*:
  - Make two new variables: *lowmoney* and *highmoney*
  - *lowmoney* = *money* + 18.44
  - *highmoney* = *money* - 18.44

SIMPLE SLOPES IN LEGISLATOR DATA

- Two variables to dummy code:
  - Acres
    - Low Acres = Acres + SD_{Acres}
    - High Acres = Acres - SD_{Acres}
  - Money
    - Low Money = Money + SD_{Money}
    - High Money = Money - SD_{Money}

RERUN ANALYSIS WITH RECODED PREDICTORS

- For an interaction, you need 4 simple slopes tests:
  - 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, $t(521.98) = 9.52$, $p < .001$
  - Significant effect of *money* among states with high tobacco acreage, $t(494.24) = 7.96$, $p < .001$
  - Significant effect of *acres* among legislators with low tobacco company donations, $t(172.58) = 3.08$, $p = .002$
  - No effect of *acres* among legislators with high tobacco company donations, $t(53.02) = 1.45$, $p = .154$

TAKING STOCK: STEP 2

- ☑ You've run your analysis and have the output
- ☑ You've visualized/graphed your results
- ☑ You've calculated simple slopes (if applicable)

3. REPORTING FINDINGS

3.1. Describing the analysis

3.2. Reporting the results

3.3. Effect size & power

- Without Step 3.1., people won't be able to replicate you
- Without Step 3.3., people won't be able to know how likely it is that they *will* replicate you

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: DV, IVs, and covariates
  - *"Pro-tobacco voting was modeled as a function of state tobacco acreage and campaign donations from tobacco companies, controlling for legislative house."*
- 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 for each state …"*
- What covariance matrix and *df* estimation method you used
  - *"… using the variance components covariance structure and the Satterthwaite method of estimating degrees of freedom."*

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
- 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 main effect of tobacco acreage on pro-tobacco voting, $b = 0.0017$, $SE = 0.0007$, $t(91.91) = 2.49$, $p = .015$."
  - "There was a significant main effect of campaign contributions from tobacco companies on pro-tobacco voting, $b = 0.0086$, $SE = .0008$, $t(518.97) = 10.18$, $p < .001$."

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, $b = -0.000046$, $SE = .000012$, $t(474) = -3.82$, $p < .001$."
- 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, $t(521.98) = 9.52$, $p < .001$, and states with high tobacco acreage, $t(494.24) = 7.96$, $p < .001$. However, tobacco acreage only predicted more pro-tobacco voting when campaign contributions from tobacco companies were low, $t(172.58) = 3.08$, $p = .002$, as acreage was unrelated to pro-tobacco voting when campaign contributions from tobacco companies was high, $t(53.02) = 1.45$, $p = .154$."

3.3. EFFECT SIZE AND POWER

- Effect size
- Intraclass correlation

EFFECT SIZE IN MLM

- Unstandardized coefficients & standard errors
- Variance explained

UNSTANDARDIZED COEFFICIENTS & SE

- The significance of each fixed effect
- Unstandardized *b* and its *SE* tells you how reliable your effect is

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*

R^2 IN MLM

- Calculate an R^2 at each level
- Interpretation:
  - Level 1 R^2
    - *Proportional reduction of error when predicting an individual outcome*
  - Level 2 R^2
    - *Proportional reduction of error when predicting a group-level mean*

R^2 IN MLM

- 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

- **SPSS:**

```
MIXED y

/FIXED=INTERCEPT

/RANDOM=INTERCEPT | SUBJECT(w)

/PRINT=SOLUTION.
```

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

$$R_1^2 = 1 - \frac{(.017769 + .078241)}{(.036390 + .092624)}$$

$$R_1^2 = 1 - \frac{.09601}{.129014}$$

$$R_1^2 = 1 - 0.74418 = .25582$$

R^2 FOR LEGISLATOR DATA

$$R_2^2 = 1 - \frac{(.017769 + .078241 / 8)}{(.036390 + .092624 / 8)}$$

$$R_2^2 = 1 - \frac{(.017769 + .00978)}{(.036390 + .011578)}$$

$$R_2^2 = 1 - .5743 = .4257$$

REPORTING R^2

- We calculated the proportion reduction in error of the model for 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 large amount for any given congress person, $R^2_1 = .256$. At the second level, the model reduced prediction error of pro-tobacco voting by a large amount for any given state, $R^2_2 = .426$.

INTRACLASS CORRELATION (ICC)

- *A measure of how dependent observations within a group are on each other*

- If ICC is sufficiently low, then you don't have to use MLM!

- You calculate the ICC **<u>from the baseline model</u>**

Estimate of Level 2 Variance

$$\rho = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_r^2}$$

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 intraclass correlation coefficient was significant, $\rho = 0.28$, $t(525) = 6.74$, $p < .001$, suggesting that the voting behavior of legislators from the same state were not independent and confirming that a multilevel analysis is necessary for these data.*

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

THINGS TO ALWAYS KEEP IN MIND

- Normality of data
- Unstandardized coefficients

NORMALITY

- Normality is important in regular regression, and paramount in multilevel modeling
  - Non-normality usually hurts your power
- If your data are not normal:
  - Skewed: Transform them
  - Heteroskedastic: Use *unstructured* covariance matrix
  - Non-Gaussian: Use a generalized linear model

UNSTANDARDIZED COEFFICIENTS

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

UNSTANDARDIZED VARIABLES

- It totally messes the whole thing up
  - Your slopes and intercepts are *wrong*
  - This has to do with ye ole covariance matrix
- No matter how badly you want standardized coefficients, just don't do it

ADVANCED APPLICATIONS OF MLM

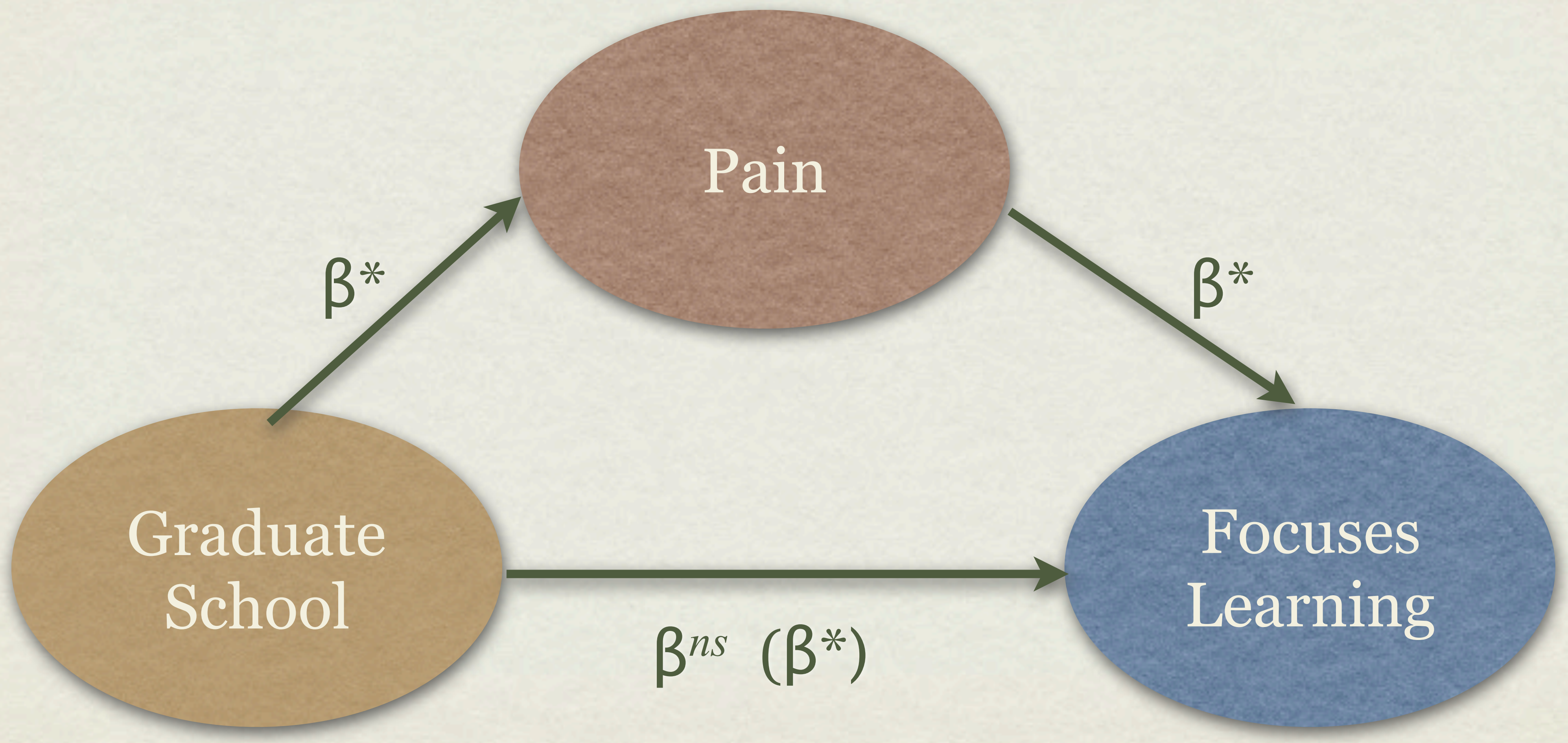
MLM APPLICATIONS

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

MULTILEVEL MEDIATION

- Regular mediation
- Multilevel mediation

REGULAR MEDIATION



REGULAR MEDIATION

- 3 equations:
  - Learning = (Years in Graduate School)
  - (Total Pain Experienced) = (Years in Graduate School)
  - Learning = (Years in Graduate School) + (Total Pain Experienced)

MULTILEVEL MEDIATION

- Similarities
  - Conducting the analysis
- Differences
  - Determining applicability across groups

APPLICABILITY ACROSS GROUPS

- Key Issue:
  - *Does the mediational model apply equally across all your groups?*
- Depends on:
  - Whether you have random slopes for variables in your mediational model

NO RANDOM SLOPES

- Conduct a mediation like normal, except using a multilevel model
  - Make those 3 equations:
    - Outcome = Predictor
    - Mediator = Predictor
    - Outcome = Predictor + Mediator
- Report unstandardized slopes, *b*, and their standard errors instead of standardized slopes
- Note: This is also cool if at least one effect in the indirect path is fixed

ALL RANDOM SLOPES

- Conduct a mediation like normal, except using a multilevel model
- You will need to compute the population covariance, σ_{ab}
  - Kenny, Korchmaros, & Bolger (2003)
  - *Estimates how reliably your mediational model explains the data across your level 2 units*

N-LEVEL MODELS

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

3-LEVEL MODELS IN SPSS

```
MIXED y WITH x1 x2

/FIXED=x1 x2 x1*x2

/RANDOM=INTERCEPT  | SUBJECT(Level2Group * Level3Group)

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

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:
  - Record measurement number (e.g., “time”)
  - Include “time” as a predictor in your fixed effects model
  - Include the slope of “time” as a random effect
  - Use an unstructured covariance matrix and between-within degrees of freedom

GENERAL*IZED* LINEAR MODELING

- Poisson
- Bernoulli

OUTCOME = COUNTS

- Distribution: Poisson
- Example: How many representatives from a state, given its economic productivity

OUTCOME = YES/NO

- Logistic regression
- Distribution: Bernoulli
- Example: Likelihood of being from a particular party

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
  - Example R Syntax:
    - `lmer(y~1+(1|GROUP1)+(1|GROUP2)+x1*x2)`

¡¡THANK YOU!!

- Workshop Sponsors:
  - Association for Psychological Science
  - Society of Multivariate Experimental Psychology
- Statistical Mentors:
  - Joel Greenhouse (Carnegie Mellon)
  - Howard Seltman (Carnegie Mellon)
  - Tom Wickens (UC Berkeley)
- Funding Sources:
  - Connaught Fund
  - Social Sciences and Humanities Research Council of Canada
  - Canada Foundation for Innovation
  - Ontario Ministry of Research & Innovation
  - Harvard Mind/Brain/Behavior Institute
  - National Science Foundation