Conducting Longitudinal Data Analysis: Knowing What to Do and Learning How to Do It

SRCD 2019 Professional Development Workshop
Daniel J. Bauer & Patrick J. Curran

Part I

Learning What To Do: Survey of Techniques

©2019 Curran-Bauer Analytics, LLC

Objectives

- Distinguish between different longitudinal data structures
- Clarify the analysis approaches that are associated with each type of data structure and the questions they can answer

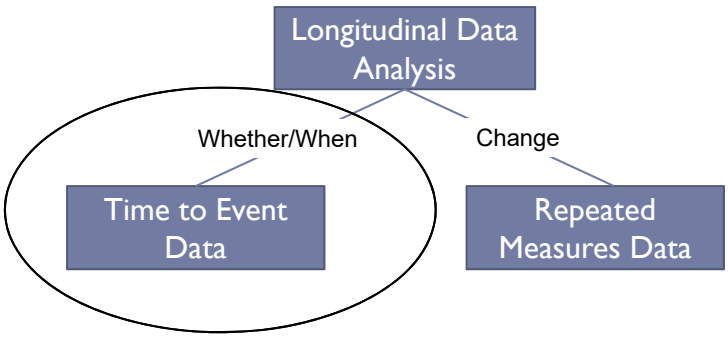
Longitudinal Data

- The term "longitudinal data" and hence also "longitudinal data analysis" is often used to describe different kinds of data structures.
- It is helpful to differentiate these structures because they address different research questions and call for different analysis approaches.

©2019 Curran-Bauer Analytics, LLC

Types of Longitudinal Analyses

- Can first split by research focus into two distinct research questions that lead to very different data structures and analytic methods…



Time to Event Data

- Time to event data is used to evaluate whether and when an event occurs.
- Examples:
  - Does the age of menarche depend on the stability of the family configuration?
  - Do boys use marijuana more often and at earlier ages than girls?

©2019 Curran-Bauer Analytics, LLC

Time to Event Data

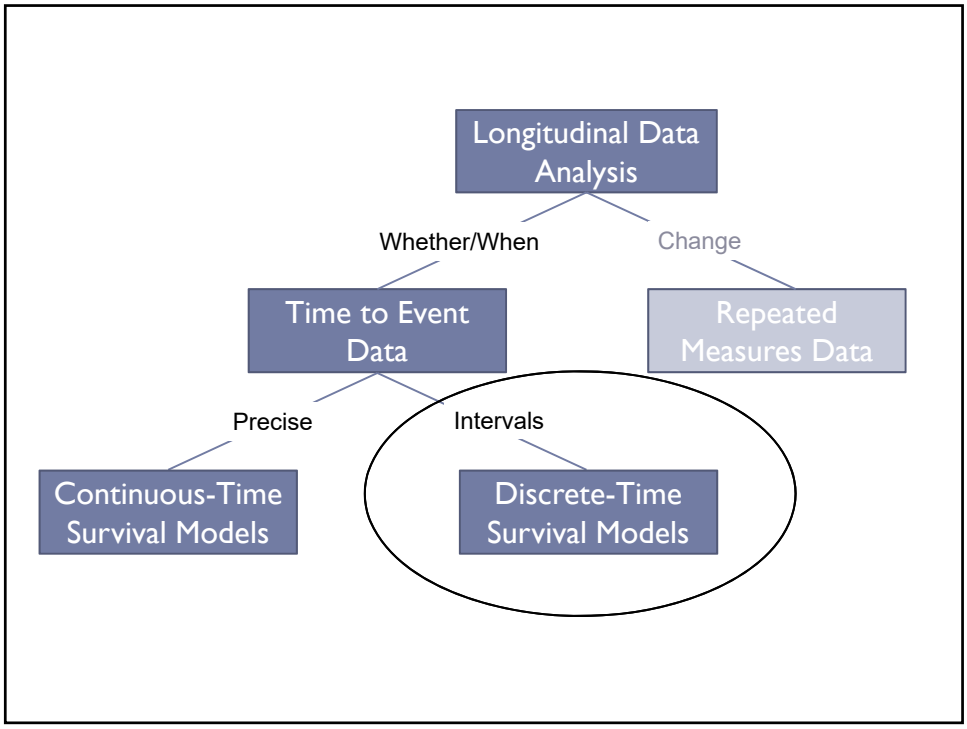
- Sometimes time to event data is obtained based on a single assessment
  - Sample individuals at age 40
  - Ask when first obtained a job, graduated from college, got married, became a parent, etc.
- Such data are subject to retrospective recall bias
- Better time to event data can be obtained using a prospective longitudinal design.
  - Sample individuals every 5 years from 15 to 40
  - At each occasion ask when first obtained a job, graduated from college, got married, became a parent, etc.

Survival Analysis

- Time to event data is best analyzed using survival analysis
  - a.k.a. hazard models, event history analysis, life table analysis
- Two flavors
  - Continuous time survival analysis:
    - Event times are measured in continuous time, such that the timing of the event is known with high precision (e.g., seconds or days) and it is uncommon for events to occur at exactly the same time for any two people.
  - Discrete time survival analysis:
    - Event times are measured on discrete time, that is, the timing of the event is measured coarsely (e.g., months or years), and it is common for events to occur at the same time for multiple people.

Allison (2010); Lee & Wang (2003); Singer & Willett (2003)

©2019 Curran-Bauer Analytics, LLC



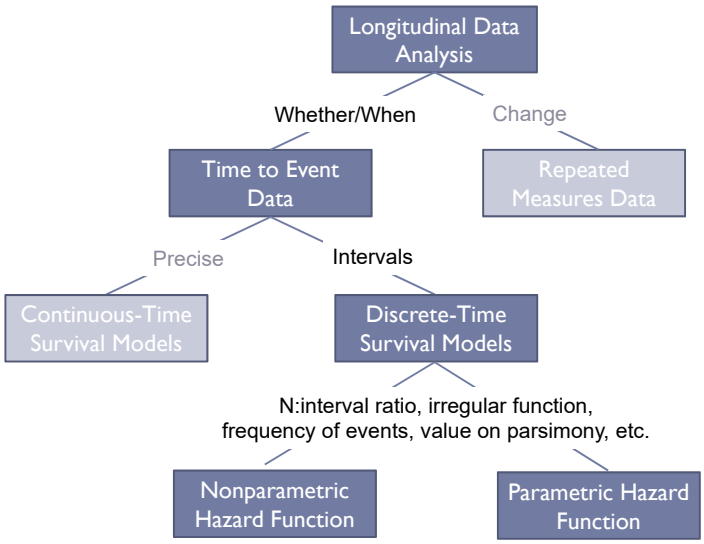
Discrete Time Hazard Models

- In the social and behavioral sciences time is usually measured coarsely and hence discrete time models are most useful.
- Predict the *hazard* of event occurrence
  - Hazard at time t is probability of event occurrence at time t given the event did not already occur at a prior time
- Based on the model for the hazard, one can also compute the *lifetime distribution function* and the *survival function*
  - LDF is cumulative probability of event occurrence over time
  - Survival function is complement of LDF; gives probability of not experiencing the event by time t

©2019 Curran-Bauer Analytics, LLC

Hazard Functions

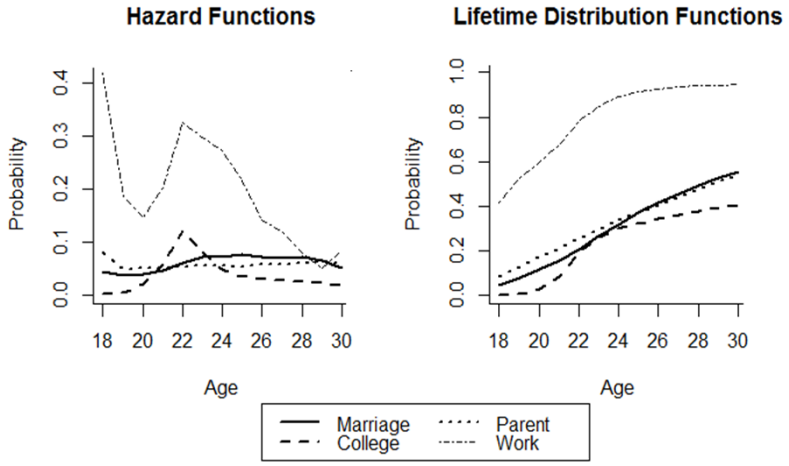
- Most commonly, the hazard function is non-parametric (unstructured)
  - Hazards are uniquely estimated at every time point
- Sometimes can be advantageous to utilize a parametric hazard function (structured)
  - Hazards are constrained to change according to some function



©2019 Curran-Bauer Analytics, LLC

A Nonparametric Discrete-Time Model

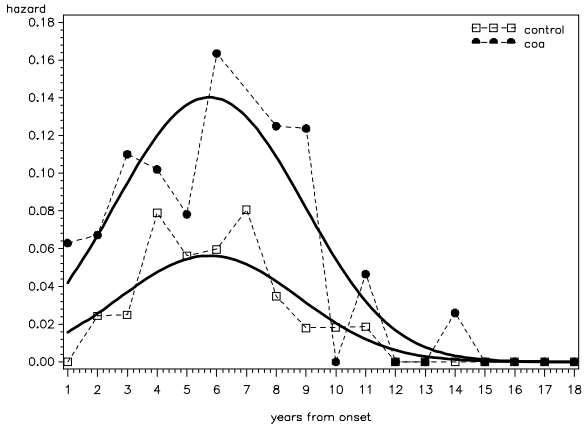
- Examining the timing of role status transitions



Dean, Bauer & Shanahan (2014)

A Parametric Discrete-Time Model

- Hazard for developing an alcohol disorder as a quadratic function of years elapsed since onset of alcohol use

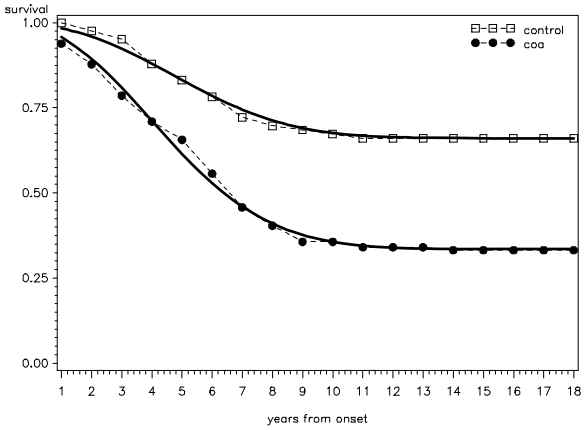


Hussong, Bauer & Chassin (2008)

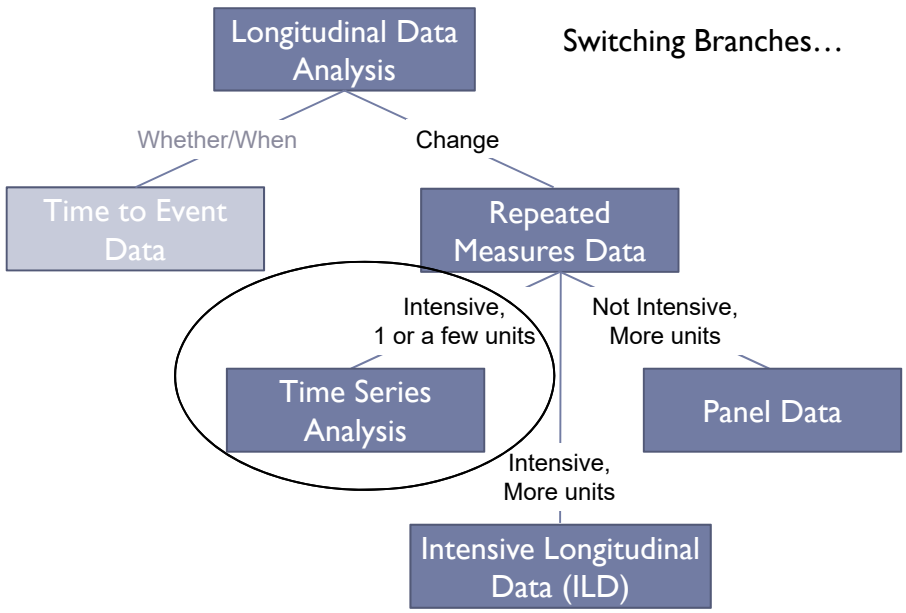
©2019 Curran-Bauer Analytics, LLC

Survival Functions

- Probability of not being diagnosed with an alcohol disorder as a function of years from onset of alcohol use



Hussong, Bauer & Chassin (2008)



©2019 Curran-Bauer Analytics, LLC

Time Series

- Another type of longitudinal data structure is time series data
- Time series data typically consists of a very long sequence of measurements on a single unit

Example Time Series Data

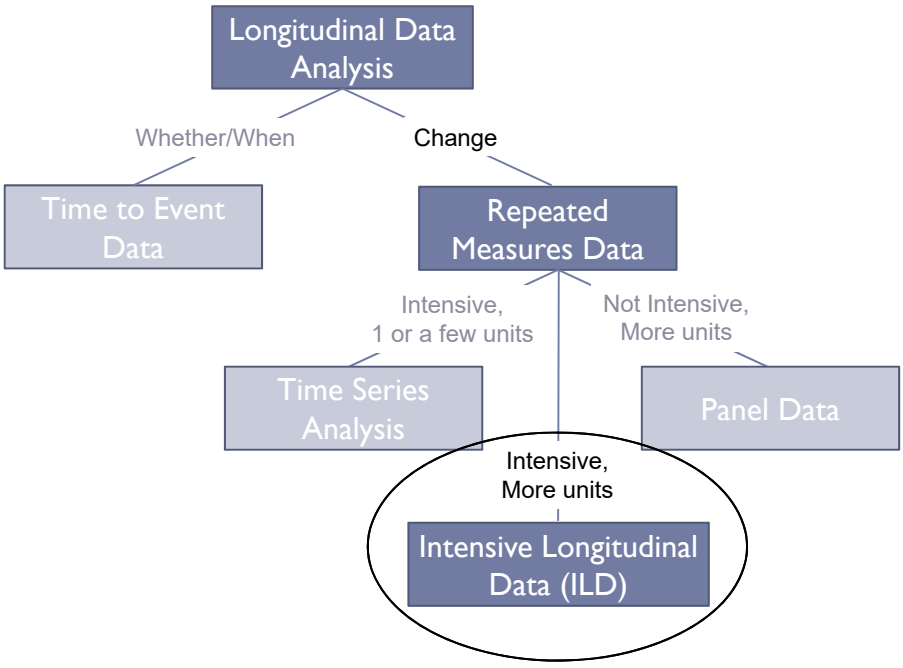
- S&P 500 index over the past 5 years

finance.yahoo.com

©2019 Curran-Bauer Analytics, LLC

Goals of Time Series Analysis

- In building a time series model, the primary goal is prediction / forecasting
  - What do we expect to happen to the S&P 500 tomorrow? Next week? Next month?
- Prior observations are used to predict future observations
  - Common models are autoregressive, moving average, ARMA, ARIMA, etc.
  - These models vary in their assumptions and complexity
- Sometimes interest is also in extracting other information
  - cyclical trend information like seasonal and/or weekly trends, although often try to de-trend data when doing time-series analysis
  - overall amount of variability (volatility)



©2019 Curran-Bauer Analytics, LLC

Intensive Longitudinal Data Analysis

- In the behavioral and health sciences, now common to use experience sampling to obtain time series on many individuals
  - Daily positive and negative affect ratings over 2 months for older adults
  - Daily pain ratings for one month for patients with rheumatic disease
- A time series model sometimes fit to each individual's data, and the parameter estimates then become individual difference variables
  - Standard deviation of a time series indicates intra-individual variability, instability
  - Magnitude of autoregression parameter has been interpreted as "inertia" in emotion research: larger values indicate it takes longer to return to equilibrium

Jahng, Wood & Trull (2008); Wang, Hamaker & Bergeman (2012)

Individual Differences in Time Series

- Pain ratings for patients with rheumatic disease

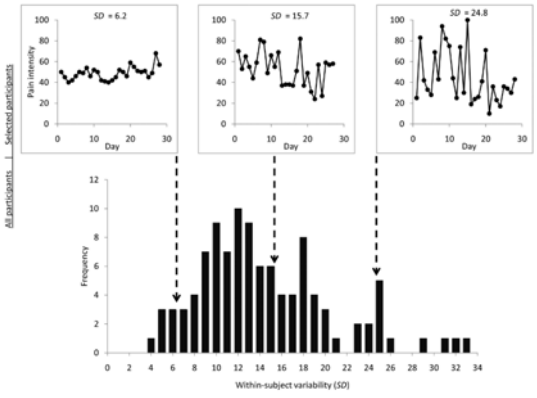
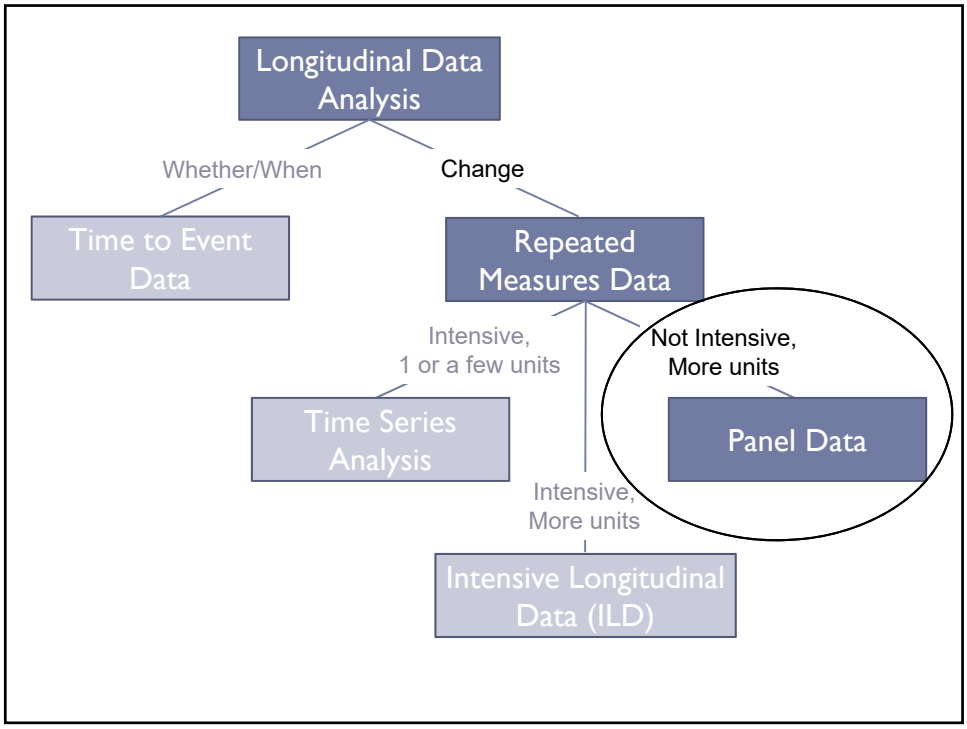


Figure 1.
Study 1: Distribution of within-subject variability in pain intensity across participants, with daily ratings of three selected participants

Schneider et al. (2012)

©2019 Curran-Bauer Analytics, LLC



Panel Data

- Most often, what we mean by longitudinal data is panel data
- Panel data is data collected over a relatively small number of time points on a relatively large number of units
- Examples:
  - Biennial assessments of alcohol and substance use in adolescents and young adults from ages 14 to 30.
  - Monthly assessments of vocabulary production by infants/toddlers from 8 to 24 months of age
- Other similar data structures arise from
  - Randomized clinical trials
  - Accelerated longitudinal designs

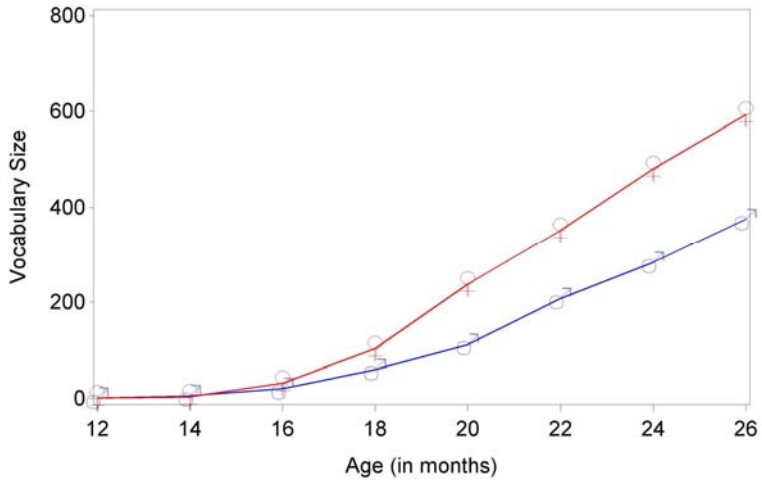
©2019 Curran-Bauer Analytics, LLC

Analysis Goals

- A common goal when collecting panel data is to evaluate change over time
- One can distinguish between mean-level change…
  - On average, how much more quickly does the vocabulary production of girls increase relative to the vocabulary production of boys?
- and individual-level (within person) change…
  - What do individual trajectories of vocabulary production look like?
  - How great are the individual differences in these trajectories?

Example: Vocabulary Development

- Mean-level differences in developmental change

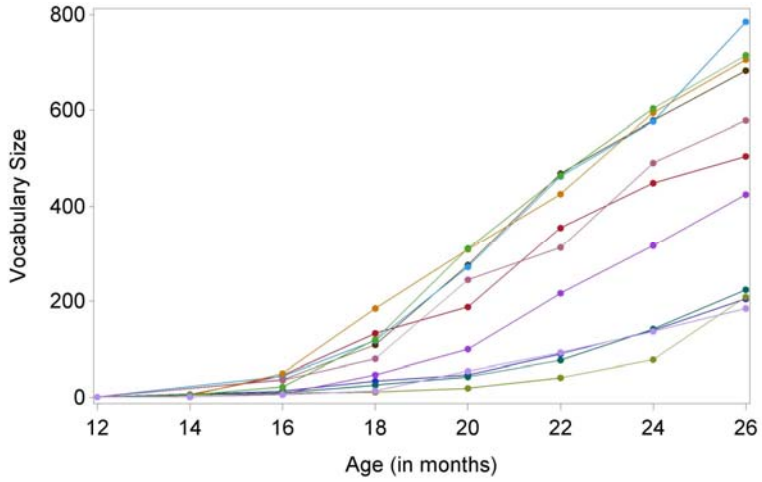


Huttenlocher et al. (1991)

©2019 Curran-Bauer Analytics, LLC

Example: Vocabulary Development

- Individual differences in developmental change

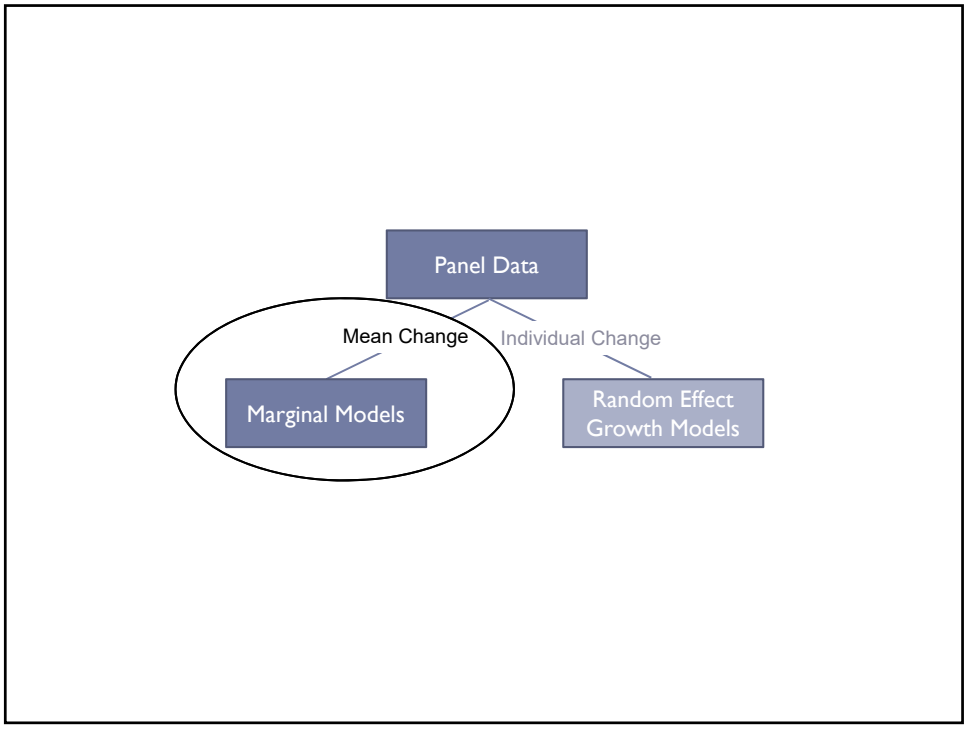


Huttenlocher et al. (1991)

Analytic Techniques

- Models that focus exclusively on mean-level change sometimes called *Marginal Models*
- Those that emphasize individual change often do so through inclusion of *Random Effects*
  - often called "growth models"

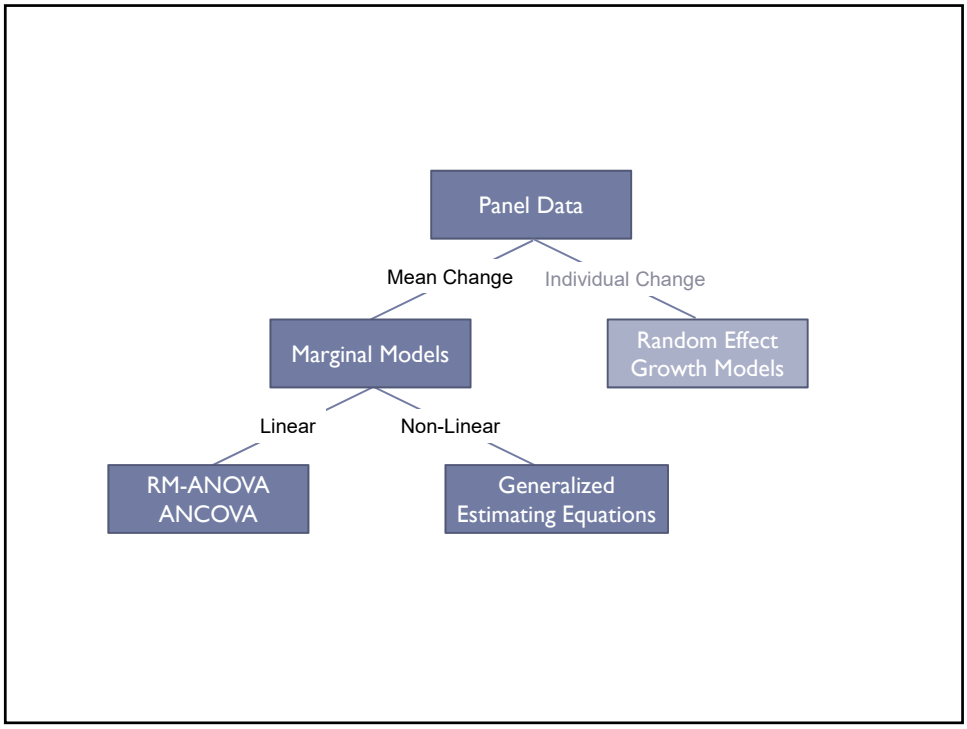
©2019 Curran-Bauer Analytics, LLC



Marginal Models

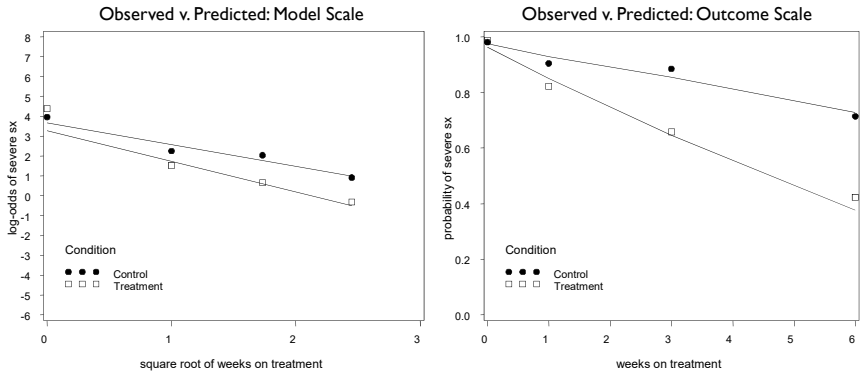
- Modeling approaches that focus on mean change:
  - Repeated measures ANOVA and MANOVA
  - ANCOVA (especially with pre/post data)
  - Generalized Estimating Equations

©2019 Curran-Bauer Analytics, LLC



Example: Treatment of Schizophrenia

- GEE model fit to evaluate efficacy of drug treatment for symptoms of schizophrenia



Hedeker & Gibbons (2004)

©2019 Curran-Bauer Analytics, LLC

Modeling Mean Change

- GEE particularly popular in longitudinal health research because it provides robust inferences without making strong assumptions about individual differences
  - Particularly useful when working with discrete outcomes
- Repeated measures ANOVA and ANCOVA remain popular for within-subjects experimental designs but less so for over-time longitudinal data
  - Except when limited to 2 waves of data…

Limitations of Two Time Points

- Having only two time points makes it difficult to study anything but mean change
- Problem is it is difficult to parse individual change from error

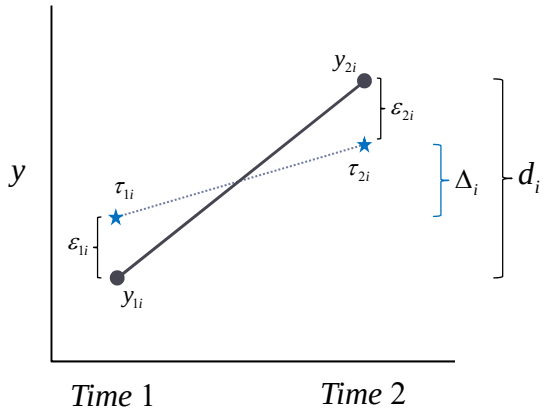
Two waves of data are better than one, but maybe not much better
Rogosa, Brandt, Zimowski (1982)

- Obvious measure of individual change is difference score between two time points, but this is often unreliable

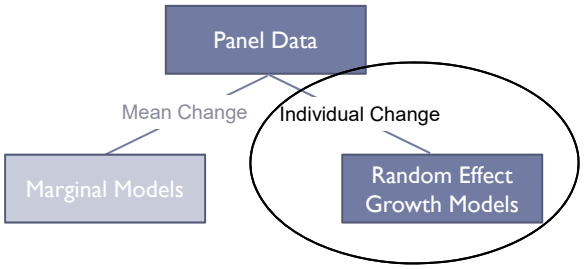
©2019 Curran-Bauer Analytics, LLC

The Trouble with Difference Scores

- True and observed difference for a hypothetical case



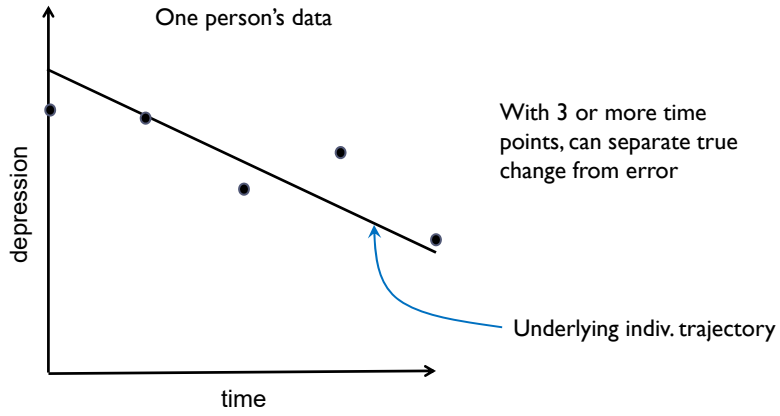
Switching Branches…



©2019 Curran-Bauer Analytics, LLC

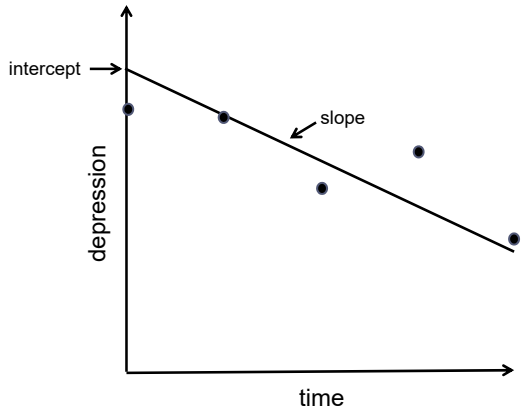
Modeling Individual Change

- Random effects models start by focusing on individual or within-person change



A Growth Curve for One Person

- Can summarize line by two pieces of information
  - the intercept and the slope unique to individual *i*



©2019 Curran-Bauer Analytics, LLC

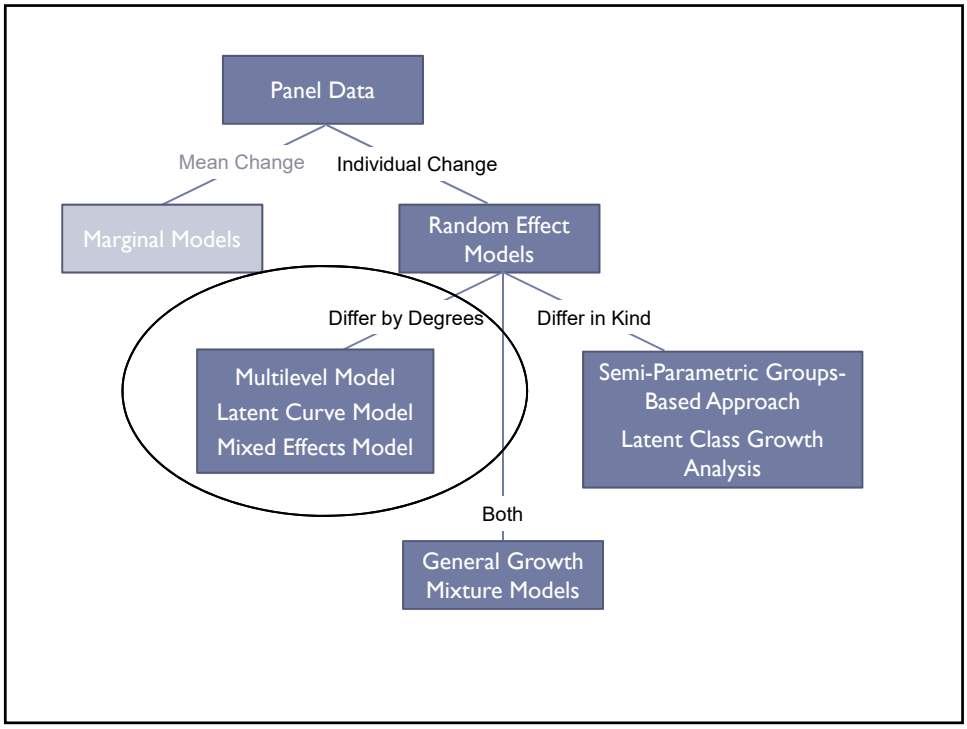
Random Effects

- Intercept and slope coefficient values permitted to vary over individuals
- Variation in growth coefficients described by a distribution
  - Growth coefficients are *random effects* in statistical sense that they come from a probability distribution
- We usually don't literally estimate the coefficients for each person, but rather the parameters of the distribution from which they came
  - This allows us to make inferences to full population of people from which our sample was drawn

Differences by Degree or Kind?

- Key question is how to define the distribution of the random effects
- Modeling approaches differentiate based on assumptions about how individuals are thought to differ from one another
  - Quantitative variation on a continuum: differences of degree
  - Qualitative differences between types of trajectories: differences in kind

©2019 Curran-Bauer Analytics, LLC



Example: Modeling Vocabulary Growth

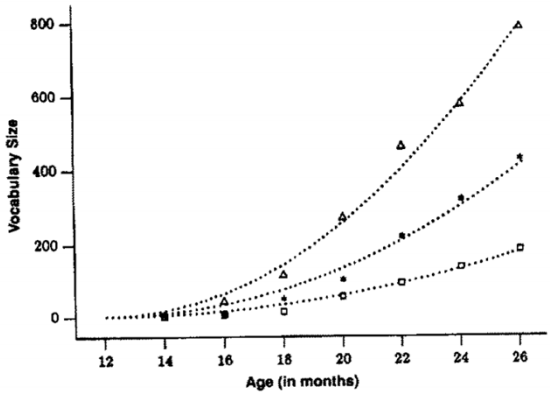
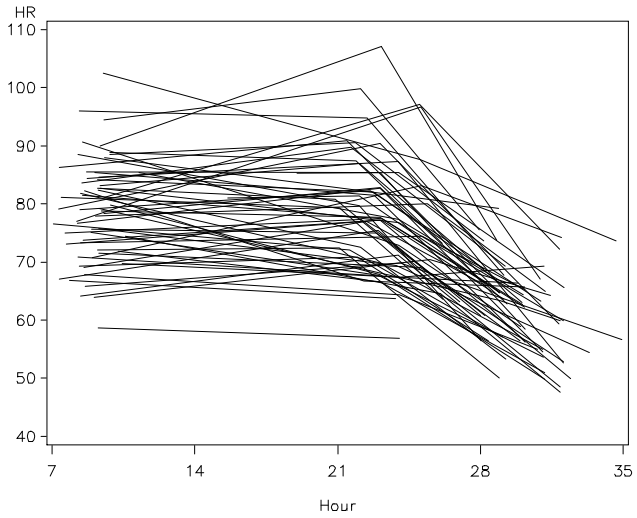


Figure 1. Range of observed versus predicted growth trajectories for Group 1 children. (□ = observed values for Child 11; * = observed values for child 5; △ = observed values for Child 7

Huttenlocher et al. (1991)

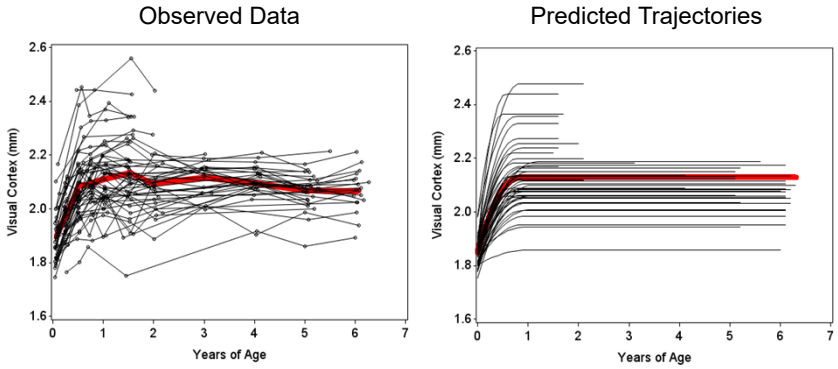
©2019 Curran-Bauer Analytics, LLC

Example: Ambulatory HR Over 24 Hrs



Richman, Pek, Pascoe & Bauer (2010)

Example: Cortical Development

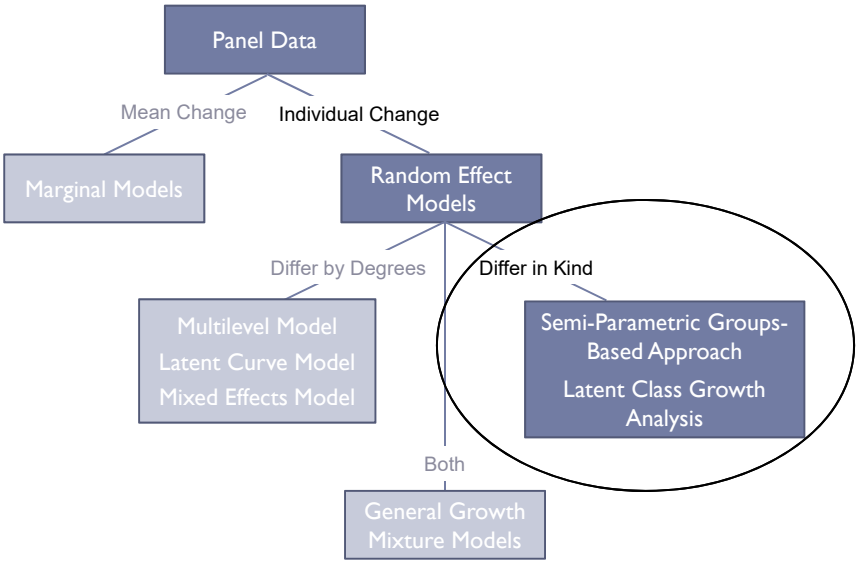


Sheridan, Cohen, Bauer & Lin (in progress)

©2019 Curran-Bauer Analytics, LLC

Conditional Models

- Often goal is not only to see individual differences in change but also to examine differences as a function of a predictor
  - Predictor could be known grouping variable (boys/girls vocabulary)
  - Predictor could be continuous (stress)
- Plot expected trajectories by levels of predictors to visualize differences in change over time
- But what if trajectories might differ as a function of some unknown grouping?



©2019 Curran-Bauer Analytics, LLC

Example: Physical Aggression

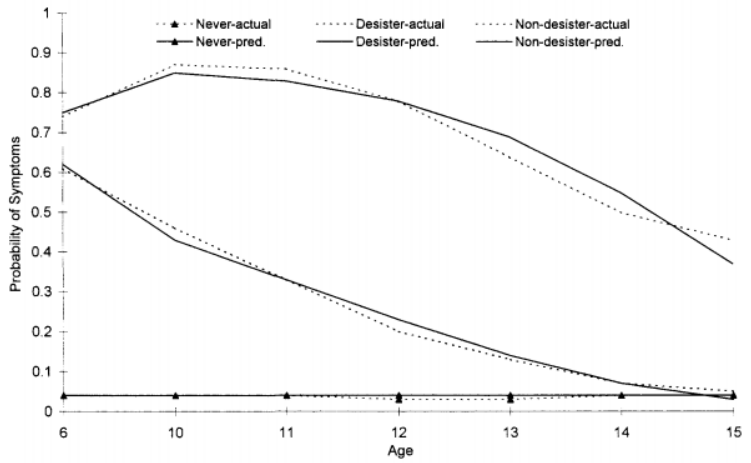
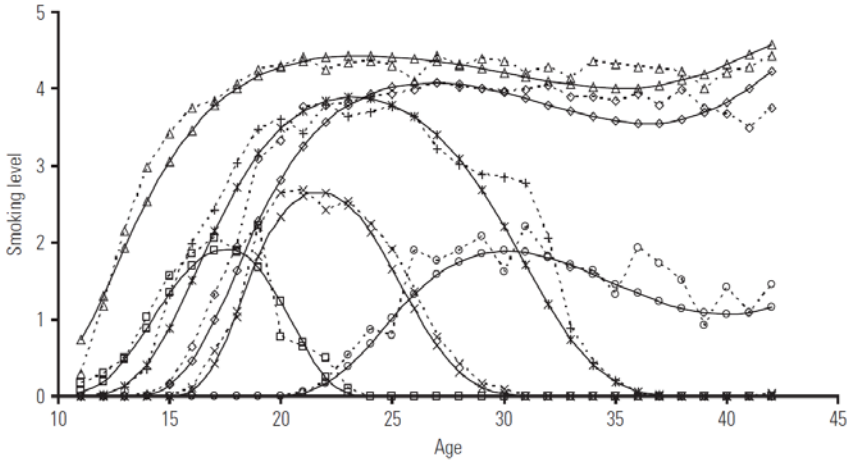


Figure 3. Trajectories of symptoms of physical aggression (Montreal sample). pred. = predicted.

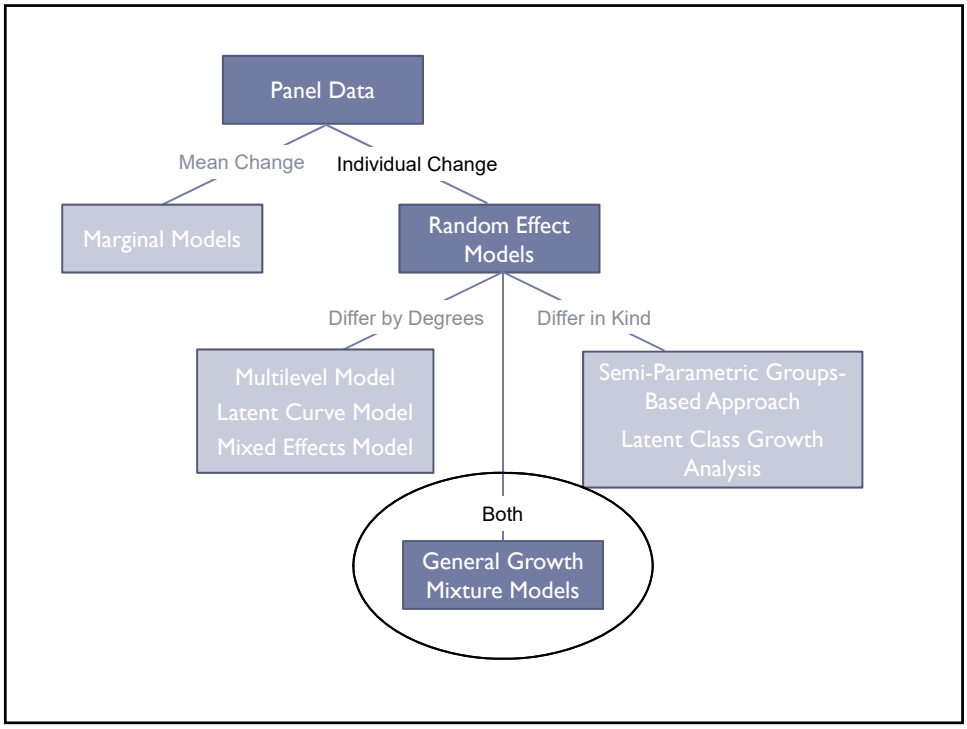
Nagin (1999)

Example: Tobacco Use Trajectories



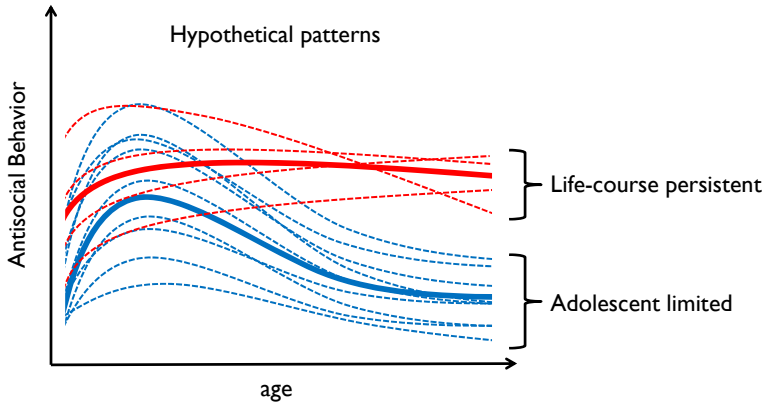
Chassin, Curran, Wirth, Presson & Sherman (2009)

©2019 Curran-Bauer Analytics, LLC



Archetypal Case: Antisocial Behavior

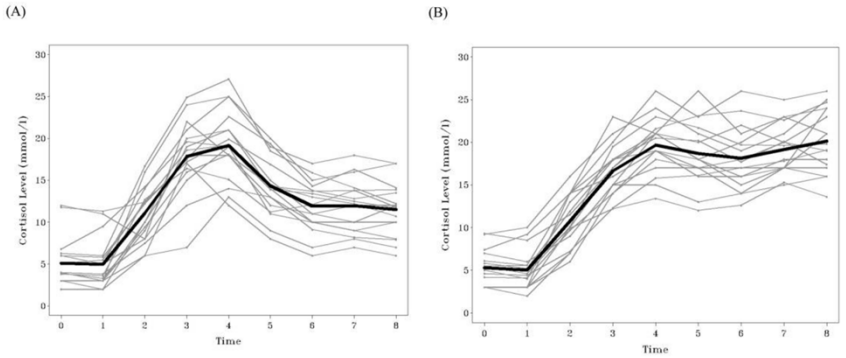
- Moffitt (1993) posited a developmental taxonomy for antisocial behavior, but can imagine variation within each "theme"



Moffitt (1993)

©2019 Curran-Bauer Analytics, LLC

Example: Cortisol Response to Stress



Down-regulation group (Typical)

Chronic stress group

Ram & Grimm (2013)

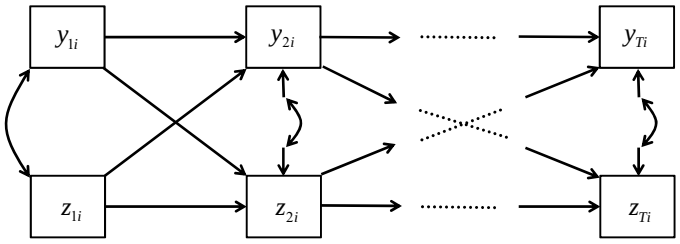
Other Panel Data Models

- There are many, many other panel data models that are useful in a variety of other circumstances
- We'll briefly look at some of these here…
  - Not a comprehensive listing

©2019 Curran-Bauer Analytics, LLC

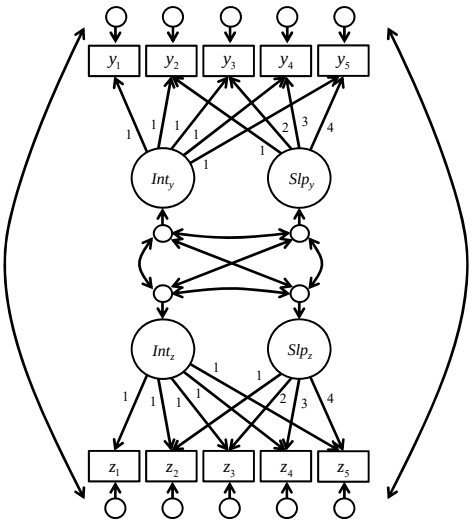
Autoregressive Cross-Lag (ARCL) models

- Evaluate bidirectional effects as they unfold over time



Multivariate Growth Models

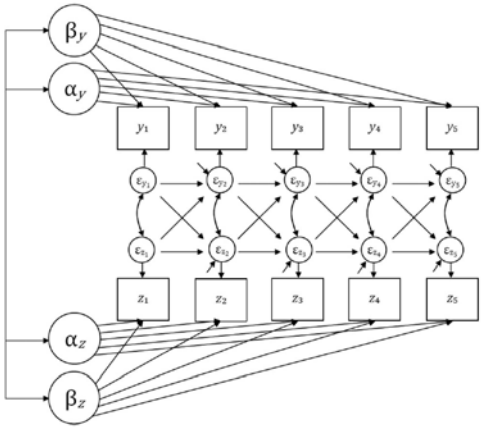
- Examine how trajectory coefficients are related across processes



©2019 Curran-Bauer Analytics, LLC

Models that Combine Growth and ARCL

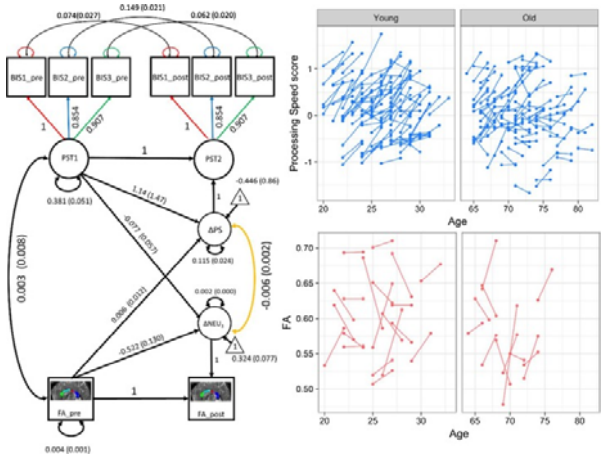
- *LCM-SR* to separate within and between person relationships



Curran et al. (2013)

Latent Change Score Models

- Examine relationships between time-adjacent changes



Kievit et al. (2018)

Time Varying Effects Models

- Evaluate how effect of predictor varies over time

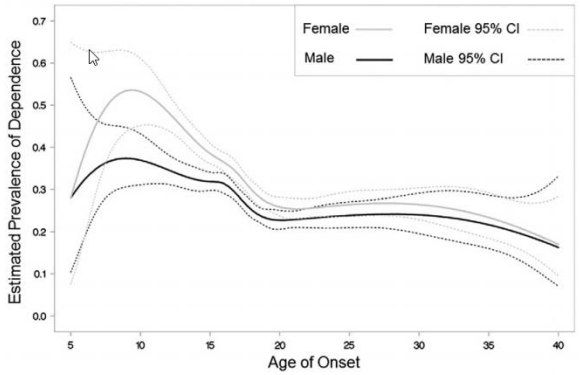
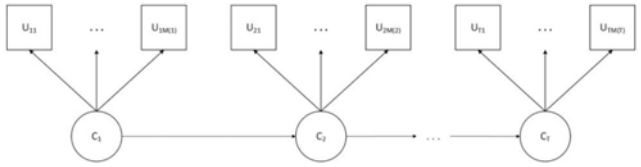


Fig. 2. Rate of nicotine dependence among ever-regular smokers as a function of age of first regular cigarette use, by sex.

Lanza & Vasilenko (2015)

Latent Transition Analysis

- Evaluate stage-like change



	Latent Status			
	Non-Users	Cigarette Smokers	Binge Drinkers	Bingers with Marijuana Use
Transition probabilities (Rows for Time 1, Columns for Time 2):				
Non-Users	**.895**	.042	.062	.002
Cigarette Smokers	.165	**.645**	.002	.188
Binge Drinkers	.115	.000	**.827**	.058
Bingers with Marijuana Use	.062	.000	.000	**.938**

Lanza et al. (2010); Scorza et al. (2015)

©2019 Curran-Bauer Analytics, LLC

Summary by Research Focus / Data Type

Research Focus	Data Type	Model
Whether/when event occurs	Time to event	Precise times: Continuous-time survival analysis Coarse intervals: Discrete-time survival analysis
Predict future observation	Time series	AR, MA, ARMA, etc
Individual differences in volatility, inertia	Intensive longitudinal	Multiple time series analyses
Mean differences in change	Panel (2+ waves)	RM-ANOVA, ANCOVA, GEE
Within-person change	Panel (3+waves)	Quantitative Ind. Differences: Multilevel, mixed effects, latent curve Qualitative Ind. Differences: Latent class growth analysis, Semiparametric groups-based approach Qual Themes + Quant Variations: General growth mixture model

Summary by Research Focus / Data Type

Research Focus	Data Type	Model
Bidirectional effects over time	Panel data	Auto-regressive Cross Lag
Change in one process related to change in another	Panel data	Trajectories: Multivariate growth model Pairs of time points: Latent change score model
Change + bidirectional	Panel data	Latent curve model with structured residuals
Timing of predictor effects	Intensive longitudinal Long-term panel	Time-varying effects models
Progression through stages/sequence	Panel data	Latent transition analysis

©2019 Curran-Bauer Analytics, LLC

Summary

- "Longitudinal data" may refer to a variety of different data structures
- Analytic techniques vary by data structure and are designed to answer different questions
- Surveyed many commonly used techniques, when they are typically applied, and to what ends
- Many exceptions to generalities noted here, and many techniques not covered
- A large and growing literature to keep up with
- So how do you keep up with it?

Part II

Learning How to Do It: Options for Additional Training

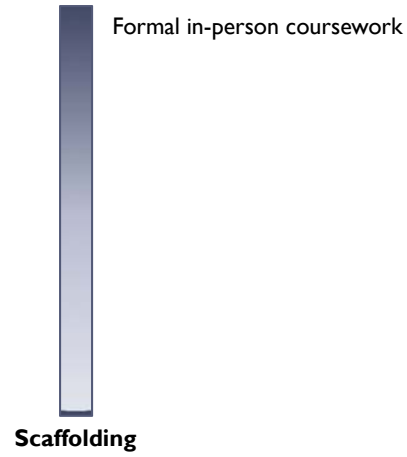
©2019 Curran-Bauer Analytics, LLC

Objectives

- Describe training options available for learning how to conduct longitudinal data analyses of various kinds

Range of Options

- Training options range in level of formality and scaffolding



©2019 Curran-Bauer Analytics, LLC

Coursework

- Large, research focused universities will tend to have some graduate-level coursework
- Example: UNC has a quantitative psychology graduate program with six faculty who regularly teach graduate level courses…

Quantitative Psychology Graduate Teaching Schedule 2019-2020

Term	*Instructor*	*Day & Time*	*Credit Hours*	*Enrollment Cap*	*Enrollment Restrictions*
Fall 2019					
830 (Statistical Methods I)	Harrison	TTh 2:00-3:15	3	25	psych or permission of instructor
846 (Multilevel Modeling)	Bauer	Mon 9:00-11:30	3	35	permission of instructor
791 (Advanced Structural Equation Modeling)	Bollen	Mon 2:00-4:30	3	15	permission of instructor
859 (Item Response Theory)	Thissen	Wed 9:00-11:30	3	15	permission of instructor
848 (Advanced Topics in Quantitative Psychology)	Curran	Wed 1:00-2:00	1	10	permission of instructor
Spring 2020					
831 (Statistical Methods II)	Bauer	TTh 2:00-3:15	3	25	psych or permission of instructor
839 (Quantitative Research Methods)	Curran	Mon 9:00-11:30	3	10	permission of instructor
791 (Machine Learning)	Gonzalez	Wed 9:00-11:30	3	15	permission of instructor
859 (Computational Statistics)	Gates	Tue 9:00-11:30	3	15	permission of instructor
848 (Advanced Topics in Quantitative Psychology)	Bollen	Mon 2:00-4:30	1	10	permission of instructor

Advantages of Coursework

- In depth treatment of topic over a semester
- Hands on examples
- Opportunities to complete assignments and obtain feedback

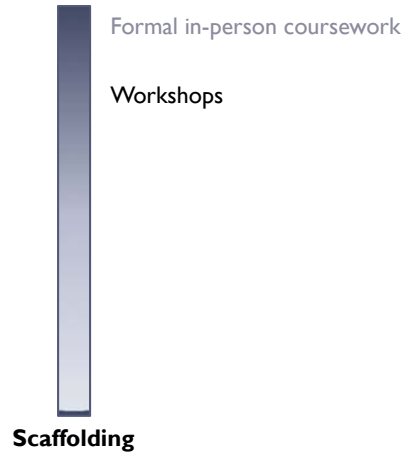
©2019 Curran-Bauer Analytics, LLC

Disadvantages of Coursework

- Accessibility can be limited
  - Are you at a university that offers these classes?
  - Are you eligible to take them?
    - Still a student? If not, does instructor allow others to sit in?
    - Restrictions on enrollment
  - Can you find them?
    - Sprinkled over departments
  - Are they being offered when you need them and can take them?
- Time involvement
  - A semester is a long time to wait if you have an analysis you need/want to do now
  - Homework assignments may not bear much similarity to your situation

Range of Options

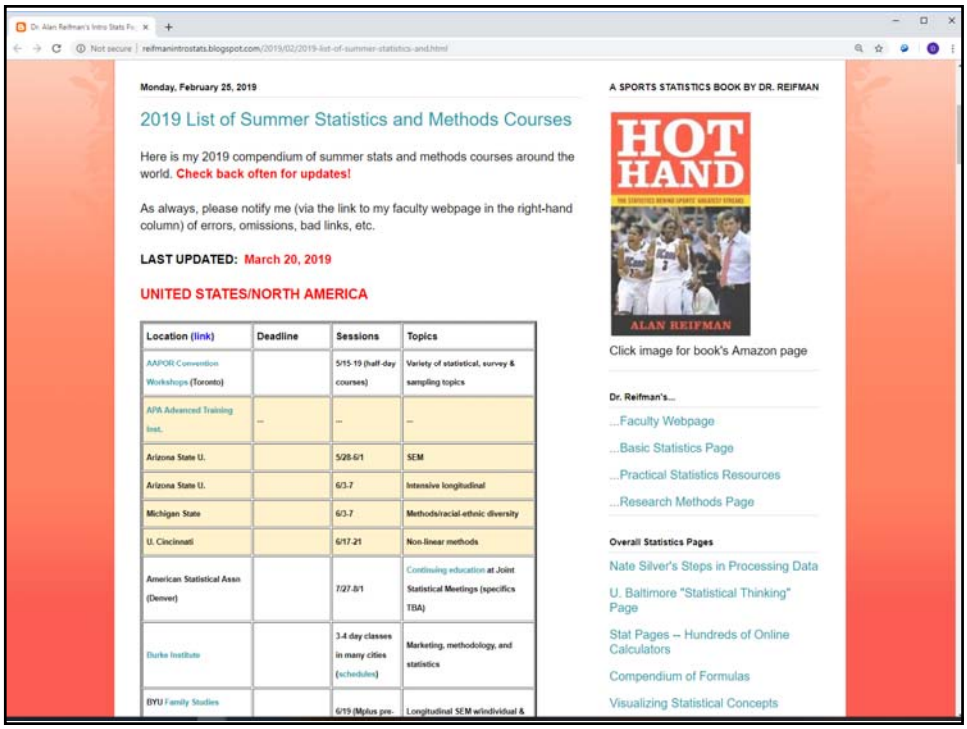
- Training options range in level of formality and scaffolding



©2019 Curran-Bauer Analytics, LLC

Workshops

- Many different groups offer training seminars on longitudinal data analysis techniques
  - Curran-Bauer Analytics
  - ICPSR
  - APA Advanced Training Institute
  - Statistical Horizons
  - Stats Camp
  - Many more…
- Nice listing of these here…
  - http://reifmanintrostats.blogspot.com/2019/02/2019-list-of-summer-statistics-and.html



Monday, February 25, 2019

2019 List of Summer Statistics and Methods Courses

Here is my 2019 compendium of summer stats and methods courses around the world. **Check back often for updates!**

As always, please notify me (via the link to my faculty webpage in the right-hand column) of errors, omissions, bad links, etc.

LAST UPDATED: March 20, 2019

UNITED STATES/NORTH AMERICA

Location (link)	Deadline	Sessions	Topics
AAPOR Convention Workshops (Toronto)		5/15-19 (half-day courses)	Variety of statistical, survey & sampling topics
APA Advanced Training Inst.	--	--	--
Arizona State U.		5/28-6/1	SEM
Arizona State U.		6/3-7	Intensive longitudinal
Michigan State		6/3-7	Methods/racial-ethnic diversity
U. Cincinnati		6/17-21	Non-linear methods
American Statistical Assn (Denver)		7/27-8/1	Continuing education at Joint Statistical Meetings (specifics TBA)
Duke Institute		3-4 day classes in many cities (schedules)	Marketing, methodology, and statistics
BYU Family Studies		6/19 (Mplus pre-	Longitudinal SEM w/individual &

A SPORTS STATISTICS BOOK BY DR. REIFMAN

Click image for book's Amazon page

Dr. Reifman's...

...Faculty Webpage

...Basic Statistics Page

...Practical Statistics Resources

...Research Methods Page

Overall Statistics Pages

Nate Silver's Steps in Processing Data

U. Baltimore "Statistical Thinking" Page

Stat Pages -- Hundreds of Online Calculators

Compendium of Formulas

Visualizing Statistical Concepts

©2019 Curran-Bauer Analytics, LLC

Advantages of Workshops

- In depth treatment of topic
  - 5-day workshop approximately equivalent to a semester-length class
- Training obtained quickly through intensive coverage
- Can take when you need it

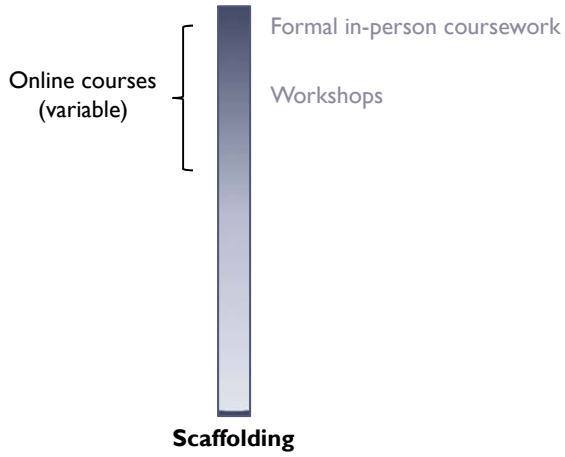
Disadvantages of Coursework

- Cost, although can sometimes be offset
  - Institutional travel awards
  - Awards made by professional organizations
    - SMEP award for students from under-represented groups: https://smep.org/resources/underrepresented-fellowships
  - Write into grants (e.g., K and R awards)
- Travel
  - Can avoid if do online workshop, but remote participation less engaging and effective (in our opinion)
- Assignments/feedback often limited
  - But may be able to obtain individualized consulting on your specific project while in attendance

©2019 Curran-Bauer Analytics, LLC

Range of Options

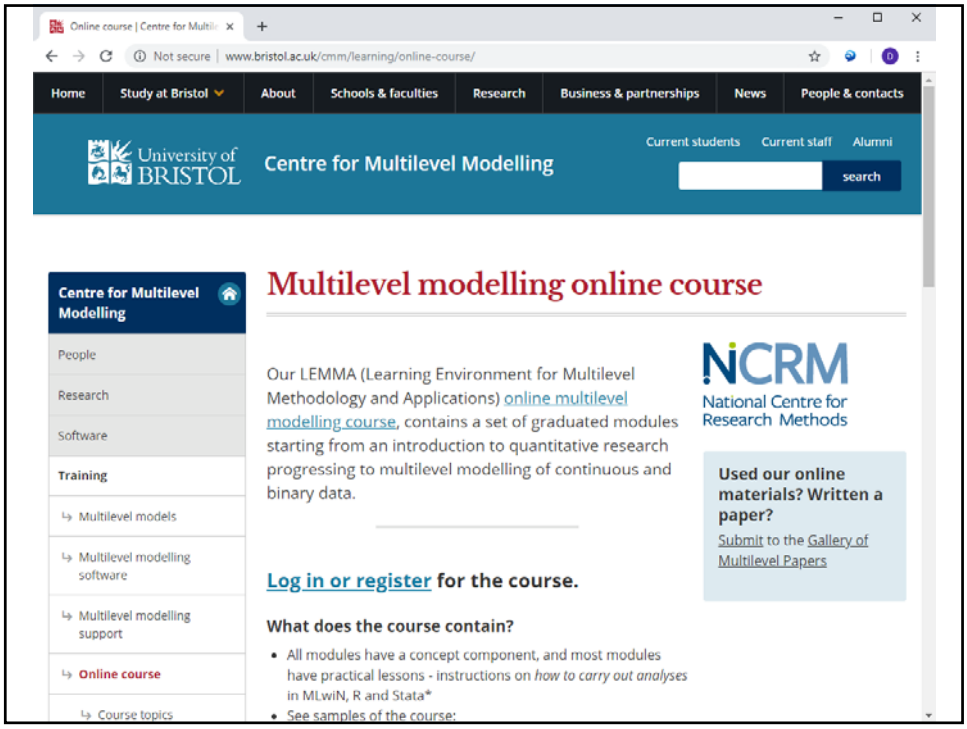
- Training options range in level of formality and scaffolding



Advantages of Online Courses

- Often low cost or free
- Can do whenever you want and in as much depth as you like

©2019 Curran-Bauer Analytics, LLC



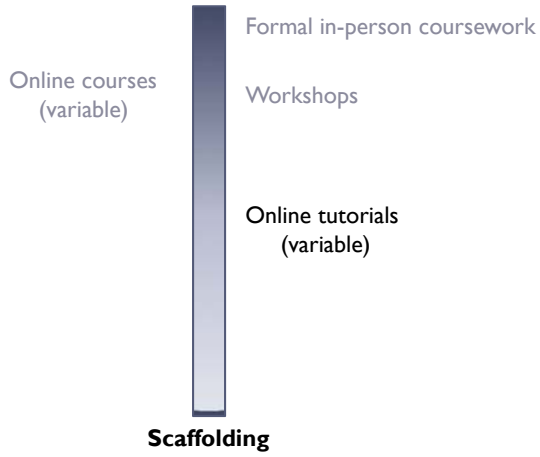
Disadvantages of Online Courses

- Can be hard to find
- Little/no contact with instructors
- Requires a lot of self-discipline

©2019 Curran-Bauer Analytics, LLC

Range of Options

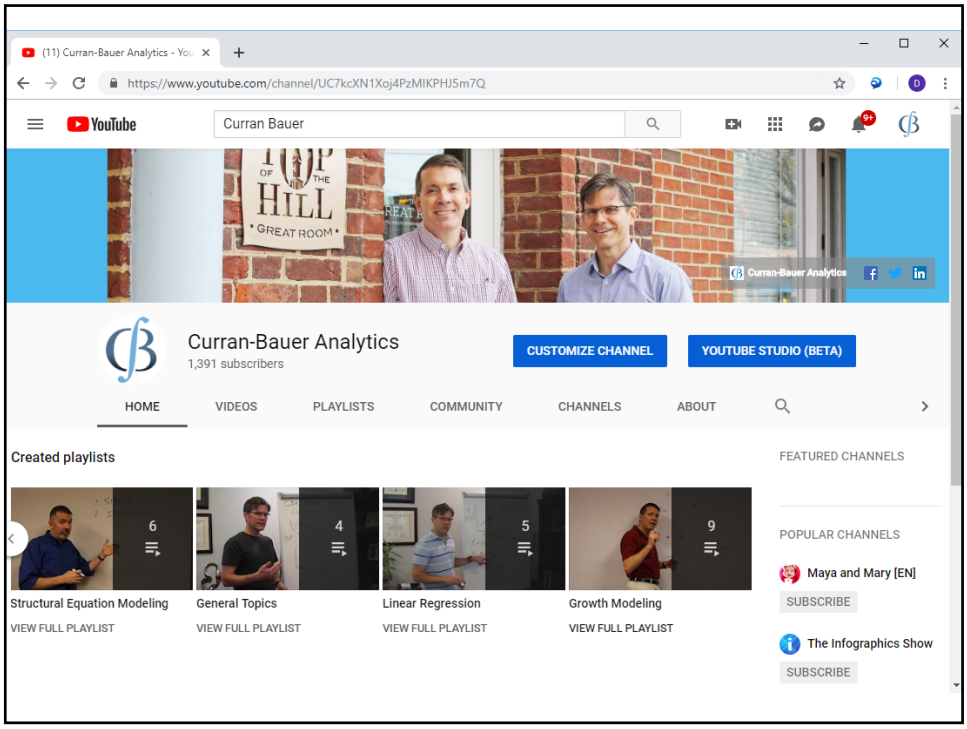
- Training options range in level of formality and scaffolding



Advantages of Online Tutorials

- To the point
  - 15-minute YouTube video tells you exactly what you needed to know
- Can do whenever you want
- May be able to find multiple tutorials on the same topic
  - Obtain diversity of perspectives, see taught different ways

©2019 Curran-Bauer Analytics, LLC



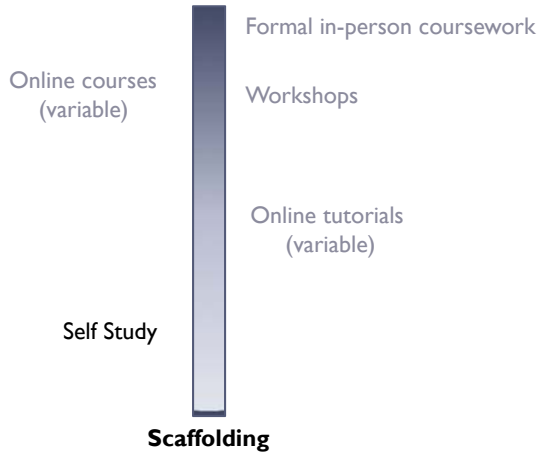
Disadvantages of Online Tutorials

- Coverage typically less in depth
- Little or no contact with instructors
- Often no assignments or opportunities for feedback on application of models
- May focus on concepts or may focus on implementation but often not both

©2019 Curran-Bauer Analytics, LLC

Range of Options

- Training options range in level of formality and scaffolding



Self Study

- Can always read about techniques on your own
  - Textbooks
  - Tutorial articles
  - Applications in the research literature you might emulate

©2019 Curran-Bauer Analytics, LLC

Advantages of Self Study

- Do on your own time, to the extent you want
- Resources generally low-cost
  - Can often obtain from university library
- Pursue whatever topics interest you as you go

Disadvantages of Independent Study

- Textbooks sometimes challenging to navigate on one's own
- Tutorial papers often focus on simplest cases
  - May not match your situation well
- Application papers seldom tell you how to do it or what to watch out for
- Little opportunity to seek clarification when confused
- Requires much discipline to stick to it

©2019 Curran-Bauer Analytics, LLC

Our Opinion

- An all-of-the-above strategy can be useful
- Often a good idea to try to get a sense of what you want to learn from exploring low-cost resources
  - Articles, on-line tutorials/videos, pre-conference workshops
- Then often useful to obtain more formal training
  - Coursework, workshops
  - Build confidence and skills
- Can then engage in self-study activities as needed with greater confidence

Summary

- There are many options for pursuing additional training in quantitative methods
  - Coursework, workshops, online classes and tutorials, textbooks and articles
- These vary considerably in level of formality
- Available resources within each category also vary considerably in quality
  - Ask around before investing time/money

Thanks for your time and attention

©2019 Curran-Bauer Analytics, LLC