Multilevel Models

Philippe Rast

PSC 204B
UC Davis, Winter 2018

Topics

Single-level Regression:

- Week 1 Linear Regression (G&H: 3,4)
- Week 2 Multiple Regression
- Week 3 Violation of Assumptions
- Week 4 Logistic Regression and GLM (G&H: 5, 6)
- Week 5 Over-fitting, Information Criteria and Model comparison (McE: 6)
- Week 6 Regression inference via simulations (G&H: 7–10)

Multilevel Regression:

- Week 7 Multilevel Linear Models (G&H: 11–13)
- Week 8 Multilevel Models (G&H: 14, 15)
- Week 9 Multilevel Models & Bayesian Inference
- Week 10 Fitting Models in Stan and brms (G&H: 16, 17 / McE: 11)

Overview

A Bigger Picture

- Historical context
- Development of statistical models
    - Interplay among what we can do in terms of modeling and what we have in terms of data
- ▷ Interrelated
- Focus on longitudinal methods
    - Short overview of where we came and where we might be heading

Development of Longitudinal Models in Psychology

Years	Popular Methods (Population ↔ Individual)	Characteristic
1950	R-Technique … P-Technique	Cattell's Data Box
1960	ANOVA	Experimental Clusters
1970	Two-Stage	Few waves few datapoints
1980	Mixed-effects $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$ and variants	Increasing N few waves
1990	Marginal e.g. $h(\boldsymbol{X}_i\boldsymbol{\beta})$	Large N (e.g. HRS)
2000		Intensive Measurements
2010	Time-series e.g. $\alpha + \beta y_{t-1} + e_t$ dynamic models	Time Series continuous flow

Development of Longitudinal Models in Psychology

Years	Popular Methods (Population ⟷ Individual)	Characteristic
1950	R-Technique — P-Technique	Cattell's Data Box
1960	ANOVA	Experimental Clusters
1970	Two-Stage	Few waves few datapoints
1980	Mixed-effects $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$ and variants	Increasing N few waves
1990	Marginal e.g. $h(\boldsymbol{X}_i\boldsymbol{\beta})$	Large N (e.g. HRS)
2000	Time-series e.g. $\alpha + \beta y_{t-1} + e_t$ dynamic models	Intensive Measurements
2010		Time Series continuous flow

Development of Longitudinal Models in Psychology

Years	Popular Methods (Population ↔ Individual)	Characteristic
1950	R-Technique … P-Technique	Cattell's Data Box
1960	ANOVA	Experimental Clusters
1970	Two-Stage	Few waves few datapoints
1980	Mixed-effects $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$ and variants	Increasing N few waves
1990	Marginal e.g. $h(\boldsymbol{X}_i\boldsymbol{\beta})$	Large N (e.g. HRS)
2000		Intensive Measurements
2010	Time-series e.g. $\alpha + \beta y_{t-1} + e_t$ dynamic models	Time Series continuous flow

Development of Longitudinal Models in Psychology

Years	Popular Methods (Population ⟷ Individual)	Characteristic
1950	R-Technique … P-Technique	Cattell's Data Box
1960	ANOVA	Experimental Clusters
1970	Two-Stage	Few waves few datapoints
1980	Mixed-effects $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$ and variants	Increasing N few waves
1990	Marginal e.g. $h(\boldsymbol{X}_i\boldsymbol{\beta})$	Large N (e.g. HRS)
2000		Intensive Measurements
2010	Time-series e.g. $\alpha + \beta y_{t-1} + e_t$ dynamic models	Time Series continuous flow

Development of Longitudinal Models in Psychology

Years	Popular Methods (Population ⟷ Individual)	Characteristic
1950	R-Technique … P-Technique	Cattell's Data Box
1960	ANOVA	Experimental Clusters
1970	Two-Stage	Few waves few datapoints
1980	Mixed-effects $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$ and variants	Increasing N few waves
1990	Marginal e.g. $h(\boldsymbol{X}_i\boldsymbol{\beta})$	Large N (e.g. HRS)
2000	Time-series e.g. $\alpha + \beta y_{t-1} + e_t$ dynamic models	Intensive Measurements
2010		Time Series continuous flow

Development of Longitudinal Models in Psychology

Years	Popular Methods (Population ⟷ Individual)	Characteristic
1950	R-Technique … P-Technique	Cattell's Data Box
1960	ANOVA	Experimental Clusters
1970	Two-Stage	Few waves few datapoints
1980	Mixed-effects $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$ and variants	Increasing N few waves
1990	Marginal e.g. $h(\boldsymbol{X}_i\boldsymbol{\beta})$	Large N (e.g. HRS)
2000		Intensive Measurements
2010	Time-series e.g. $\alpha + \beta y_{t-1} + e_t$ dynamic models	Time Series continuous flow

Development of Longitudinal Models in Psychology

Years	Popular Methods (Population ↔ Individual)		Characteristic
1950	R-Technique	P-Technique	Cattell's Data Box
1960	ANOVA		Experimental Clusters
1970		Two-Stage	Few waves few datapoints
1980		Mixed-effects $\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$ and variants	Increasing N few waves
1990	Marginal e.g. $h(\boldsymbol{X}_i\boldsymbol{\beta})$		Large N (e.g. HRS)
2000			Intensive Measurements
2010		Time-series e.g. $\alpha + \beta y_{t-1} + e_t$ dynamic models	Time Series continuous flow

Some thought about...

- Nomothetic research: Search for general laws
- Aggregation: Typical approach to obtain nomothetic information
- Idiographic research: Focus on particularities of individual
- Ergodicity: Assumption that distribution of Variables in the population reflects distribution of variables in individual.

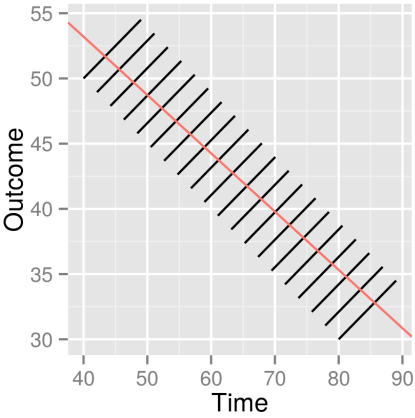
Aggregate $\neq$ Nomothetic

Aggregation

- Moments (Means, variances, proportions, correlations)
- Best guess for the aggregate
- ▷ Average across individuals

Nomothetic

- Pertains to extraction of general laws
- Individual parameters that are common among group/population
- ▷ All individuals increase over time



No Ergodicity? No Problem!

Idiographic does not mean that we can't obtain "general laws" or generalize.

- Modeling individuals may yield common parameters.
- Commonality is the nomothetic part
  - ▷ Exploration of degree of commonality

Approximate nomothetic information with aggregation

- aggregation only yields nomothetic information if ergodicity holds

Nomothetic vs Idiographic

Assumption

- Results obtained at the population level reflect to some degree with-person processes
- Generalization from population to individual are meaningful

  Are these assumptions tenable?

Statement of Problem:

- Psychology focuses on individual variation between cases
- Results are commonly generalized to variation and explanation
  - in given populations
  - within individuals in these populations
- Ergodic theorem confines generalizability
- Ergodicity is hardly ever met in psychology

Nomothetic vs Idiographic

Consequence

- Results obtained from *inter*individual variation yield different results from studies based in *intra*individual variation
- we cannot blindly base statements about processes that take place within people on results that were obtained with standard large sample analyses.

Ergodicity

Assumption that distribution of Variables in the population reflects distribution of variables in individual.

- All population moments (e.g., means, variances, covariances) must be identical to the corresponding within-person moments
- Structural changes of time must be absent
- All developmental processes are by definition non-ergodic
- All within-person moments must be identical across individuals
- all subjects have to conform to the same statistical model

Approaching the Individual

Type of Information

Aggregate $\longleftrightarrow$ Idiographic

Regression models

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$$

Mixed Effects

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

Mixed Effects Location and Scale

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i \text{ with } g(\sigma^2_{\nu_i}) = \boldsymbol{u}'_i\boldsymbol{\zeta} \qquad \text{and } g(\sigma^2_{\epsilon_{ij}}) = \boldsymbol{w}'_{ij}\boldsymbol{\tau}$$

Approaching the Individual

Type of Information

Aggregate $\longleftrightarrow$ Idiographic

Regression models

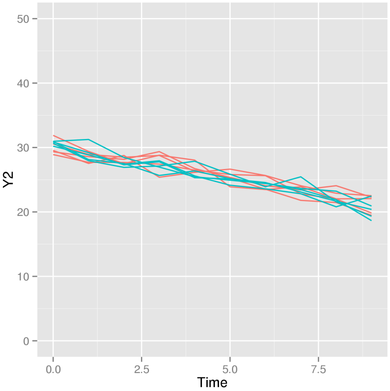
$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Mixed Effects

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

Mixed Effects Location and Scale

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i \text{ with } g(\sigma^2_{\nu_i}) = \boldsymbol{u}$$



Approaching the Individual

Type of Information

Aggregate ⟷ Idiographic

Regression models

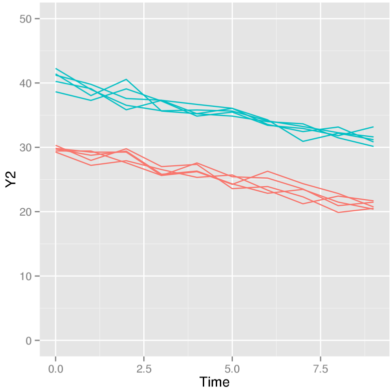
$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Mixed Effects

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

Mixed Effects Location and Scale

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i \text{ with } g(\sigma^2_{\nu_i}) = \boldsymbol{u}'_i$$



Approaching the Individual

Type of Information

Aggregate ⟷ Idiographic

Regression models

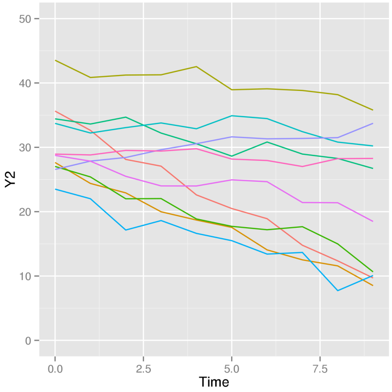
$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Mixed Effects

$$\boxed{\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i}$$

Mixed Effects Location and Scale

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i \text{ with } g(\sigma^2_{\nu_i}) = \boldsymbol{u}_i'$$



Approaching the Individual

Type of Information

Aggregate ⟷ Idiographic

Regression models

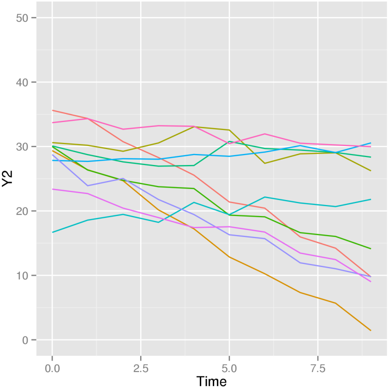
$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Mixed Effects

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

Mixed Effects Location and Scale

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i \text{ with } g(\sigma^2_{\nu_i}) = \boldsymbol{u}_i'\boldsymbol{\zeta}$$



Approaching the Individual

Type of Information

Aggregate ⟵ ⟶ Idiographic

Regression models

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$$

Mixed Effects

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

Mixed Effects Location and Scale

$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$ with $g(\sigma^2_{\nu_i}) = \boldsymbol{u}_i'\boldsymbol{\zeta}$ and $g(\sigma^2_{\epsilon_{ij}}) = \boldsymbol{w}_{ij}'\boldsymbol{\tau}$

Approaching the Individual

Type of Information

Aggregate $\longleftrightarrow$ Idiographic

Regression models

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Mixed Effects

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$$

Mixed Effects Location and Scale

$\boldsymbol{Y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i$ with $g(\sigma^2_{\nu_i}) = \boldsymbol{u}_i'\boldsymbol{\zeta}$ and $g(\sigma^2_{\epsilon_{ij}}) = \boldsymbol{w}_{ij}'\boldsymbol{\tau}$

...in a Nutshell

- Data are more intense and offer more information at individual level
- Methods need to be developed and explored that take advantage of individual information
- Move from aggregation to idiographic models (Castro-Schilo & Ferrer, 2013; Hamaker, 2012; Molenaar, 2004)

Features of Intensive Designs

- Observe change that occurs at different time scales
- ▷ Integration of different developmental trajectories within individuals

Introduction: Intensive Measurement Design

Classic multiwave design:

- Multiwave data do not have information on smaller scale
- ▷ Learning within waves: Taste of intensive designs
- Difficult to obtain micro level information
- No information on day-to-day variability
- Aimed at "slow" developmental processes

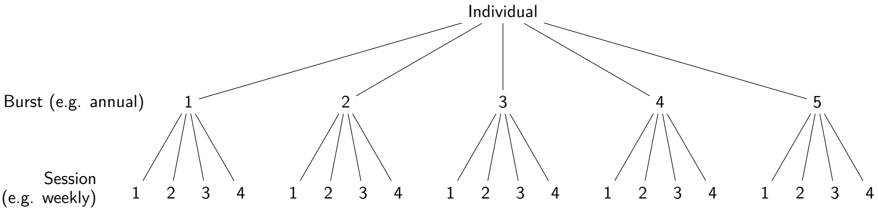
Ecological Momentary Assessments (EMA) & daily diary

- Measurement at small scale (hours, days)
- All information on micro level
- Aimed at short-term changes

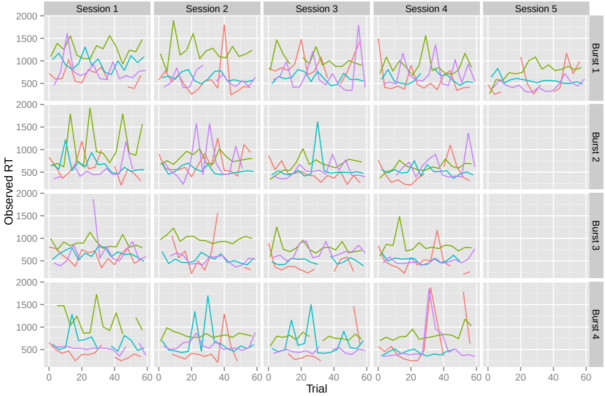
Introduction: Intensive Measurement Design

Intensive measurement designs (Nesselroade, 1991)

- Combination of multiwave design and EMA methods
- Multiple scales
- Suited for measuring change/variation in short and long term



Introduction: Intensive Measurement Design



- Variability is not noise but carries information*
  - ▷ systematic dynamic patterns of covariation

e.g. Cognition, affect, language use, perceived control† etc.

---

*Brose & Röcke, 2013; Fiske & Rice, 1955; Ram & Gerstorf 2009; Woodrow, 1932 etc.

†Eid & Diener, 1999; Eizenman, et al. 1997; van Geert et al. 2002; Rast et al. 2012; Siegler 1994 etc.

Typical Approach to Obtaining Index of IIV

Common approach to obtain within-person variability

- Extract IIV and compute an index (e.g. iSD)
- ▷ Step 1: De-trend data, print out residuals, compute index
- ▷ Step 2: Index$_i$ is used as predictor or dependent variable

Typically:

- Information about individual means is not retained
  - Especially problematic with heteroscedastic error terms
- Dependency among iM and iSD is not modeled
  - Interrelation between random effect terms of the means (location) part and the within-person variance (scale) is not independent.

Mixed Effects Location Scale Model

Alternative approach:
Mixed Effects Location Scale Model

(Hedeker et al., 2008; Rast et al. 2011, 2012, 2014, submitted)

- Model mean structure (location) and variability (scale) of the response
- Permits the use of explanatory variables for both
  - between-person variance as well as for
  - individual differences in within-person variance
- All correlations across both levels
  - interdependencies are all maintained
- All parameters are estimated **simultaneously**

Mixed Effects Location Scale Model

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i,$$

- ▷ Standard: $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2_\epsilon \boldsymbol{\Psi}_i)$

LSM: IIV may fluctuate between individuals (i) and across time (j)

$$\sigma^2_{\epsilon_{ij}} = g(\mathbf{W}'_{ij}\boldsymbol{\tau} + \mathbf{V}'_{ij}\mathbf{t}_i).$$

  - $\boldsymbol{\tau}$ defines the average WP variance: τ_0 (intercept), τ_1 (slope)
  - Time-varying covariates $\mathbf{W}_{ij}$ for the fixed and $\mathbf{V}_{ij}$ for the random effect to influence the within-person variance estimate

- ▷ Different error distribution for each individual at each occasion

Mixed Effects Location Scale Model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i,$$

- ▷ Standard: $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2_\epsilon \boldsymbol{\Psi}_i)$

LSM: IIV may fluctuate between individuals (i) and across time (j)

$$\sigma^2_{\epsilon_{ij}} = g(\mathbf{W}'_{ij}\boldsymbol{\tau} + \mathbf{V}'_{ij}\mathbf{t}_i).$$

  - $\boldsymbol{\tau}$ defines the average WP variance: τ_0 (intercept), τ_1 (slope)
  - Time-varying covariates $\mathbf{W}_{ij}$ for the fixed and $\mathbf{V}_{ij}$ for the random effect to influence the within-person variance estimate

- ▷ Different error distribution for each individual at each occasion

"It is such variation from sitting to sitting, or from day to day, here designated by the term, 'quotidian variation,' that is to be considered. [...] .The responses on different days clearly are not all of the same category; they belong to different statistical populations." (Woodrow, 1932, pp. 246)

Application: Daily Reports of Stress and Affect

$$\text{Level 1:} \qquad y_{ij} \sim N(\mu_{ij}, \sigma^2_{\epsilon_{ij}}) \qquad \text{for } j = 1, ..., 7$$

$$\mu_{ij} = \beta_{0i} + \beta_{1i} \times \text{Session}_{ij}$$

$$\sigma^2_{\epsilon_{ij}} = \exp(\tau_{0i} + \tau_{1i} \times \text{Stress}_{ij})$$

$$\text{Level 2:} \qquad \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\tau} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_{\boldsymbol{\beta}} \\ \mu_{\boldsymbol{\tau}} \end{bmatrix}, \begin{bmatrix} \sigma^2_{\boldsymbol{\beta}} & \sigma_{\boldsymbol{\beta\tau}} \\ \sigma_{\boldsymbol{\tau\beta}} & \sigma^2_{\boldsymbol{\tau}} \end{bmatrix} \right)$$

$$\text{Hyperpriors:} \qquad \begin{bmatrix} \mu_{\boldsymbol{\beta}} \\ \mu_{\boldsymbol{\tau}} \end{bmatrix} \sim N(\mathbf{a}, \mathbf{B})$$

$$\begin{bmatrix} \sigma^2_{\boldsymbol{\beta}} & \sigma_{\boldsymbol{\beta\tau}} \\ \sigma_{\boldsymbol{\tau\beta}} & \sigma^2_{\boldsymbol{\tau}} \end{bmatrix}^{-1} \sim Wish(\mathbf{R}, k).$$

Rast, Hofer & Sparks (2012)

Application: Daily Reports of Stress and Affect

$$\text{Level 1:} \quad y_{ij} \sim N(\mu_{ij}, \sigma^2_{\epsilon_{ij}}) \quad \text{for } j = 1, ..., 7$$

$$\mu_{ij} = \beta_{0i} + \beta_{1i} \times \text{Session}_{ij}$$

$$\sigma^2_{\epsilon_{ij}} = \exp(\tau_{0i} + \tau_{1i} \times \text{Stress}_{ij})$$

$$\text{Level 2:} \quad \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\tau} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_{\boldsymbol{\beta}} \\ \mu_{\boldsymbol{\tau}} \end{bmatrix}, \begin{bmatrix} \sigma^2_{\boldsymbol{\beta}} & \sigma_{\boldsymbol{\beta}\boldsymbol{\tau}} \\ \sigma_{\boldsymbol{\tau}\boldsymbol{\beta}} & \sigma^2_{\boldsymbol{\tau}} \end{bmatrix} \right)$$

$$\text{Hyperpriors:} \quad \begin{bmatrix} \mu_{\boldsymbol{\beta}} \\ \mu_{\boldsymbol{\tau}} \end{bmatrix} \sim N(\mathbf{a}, \mathbf{B})$$

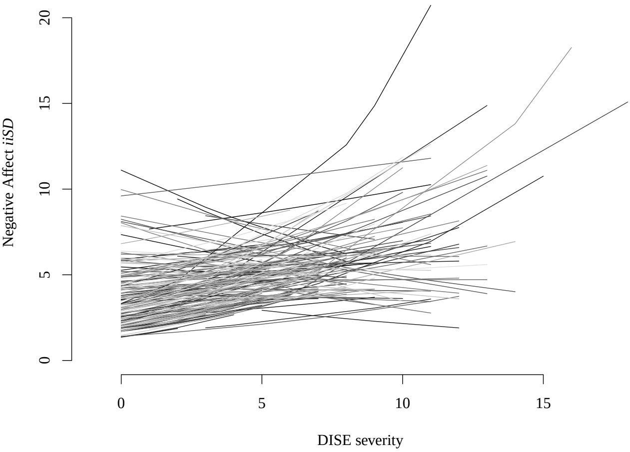
$$\begin{bmatrix} \sigma^2_{\boldsymbol{\beta}} & \sigma_{\boldsymbol{\beta}\boldsymbol{\tau}} \\ \sigma_{\boldsymbol{\tau}\boldsymbol{\beta}} & \sigma^2_{\boldsymbol{\tau}} \end{bmatrix}^{-1} \sim Wish(\mathbf{R}, k).$$

Rast, Hofer & Sparks (2012)

Results from Rast, Hofer and Sparks (2012)

Reactivity of negative affect: iSD by stress (DISEsev)

	Estimate	95% C.I.
β_0	22.94**	[21.9, 24.0]
β_1	-0.41**	[-0.56, -0.24]
$\sigma^2_{\beta_0}$	40.50**	[30.8, 52.5]
$\sigma^2_{\beta_1}$	0.13*	[0.04, 0.34]
r_{β_1,β_0}	$-.29^{ns}$	[-.69, .17]
r_{τ_0,β_0}	.80**	[.66, .90]
r_{τ_1,β_0}	$-.69$**	[-.84, -.48]
r_{τ_0,β_1}	$-.09^{ns}$	[-.58, .42]
r_{τ_1,β_1}	$.14^{ns}$	[-.41, .66]
r_{τ_1,τ_0}	$-.75$**	[-.87, -.55]
τ_0	2.55**	[2.28, 2.83]
τ_1	0.12**	[0.06, 0.17]
$\sigma^2_{\tau_0}$	1.27**	[0.80, 1.86]
$\sigma^2_{\tau_1}$	0.03**	[0.02, 0.06]



Application: LSM for Intensive Measurement Designs

Level 1:

$$
\begin{aligned}
y_{ijk} &\sim N(\mu_{ijk}, \sigma^2_{\epsilon_{ijk}}) \\
\mu_{ijk} &= \alpha_{0jk} + \alpha_{1jk}\text{Session}_{ijk} + \alpha_{2jk}\text{Burst}_{jk} + \beta\text{Session}_{ijk}\text{Burst}_{jk} \\
\sigma^2_{\epsilon_{ijk}} &= \exp(\tau_{0k} + \tau_{1k}\text{Session}_{ijk} + \tau_{2k}\text{Burst}_{ijk} + \lambda\text{Session}_{ijk}\text{Burst}_{jk})
\end{aligned}
$$

Random effects between individuals *and* between bursts within individuals.

Hyperpriors:

$$
\begin{bmatrix} \sigma^2_{\alpha_0} & \sigma_{\alpha_0\alpha_1} \\ \sigma_{\alpha_1\alpha_0} & \sigma^2_{\alpha_1} \end{bmatrix}^{-1} \sim \textit{scaled-Wishart}\left(\mathbf{R1}, k1\right)
$$

$$
\begin{bmatrix} \sigma^2_{\boldsymbol{\alpha}} & \sigma_{\boldsymbol{\alpha\tau}} \\ \sigma_{\boldsymbol{\tau\alpha}} & \sigma^2_{\boldsymbol{\tau}} \end{bmatrix}^{-1} \sim \textit{scaled-Wishart}\left(\mathbf{R2}, k2\right)
$$

Rast & MacDonald (2014)

Design Considerations for Optimizing Power to Detect Individual Differences in Change

Motivation

Statistical power is fundamental for addressing most research questions

- Makes sure that we are able to reject null hypothesis in favor of alternative hypothesis
- Projects may fail to succeed due to insufficient statistical power

▷ Standard to include power analyses in research proposals

As researchers we have some options to obtain sufficient power†

▷ Focus on longitudinal study designs to detect individual differences in change

†Cohen (1988); Hansen, Collins (1994); MacCallum, Browne, Sugawara (1996); Maxwell (1998); Maxwell, Kelley, Rausch(2008); Muthén, Curran (1997), Satorra, Saris (1985) etc.

Overview

Design considerations in longitudinal studies to detect individual differences in change:

- Latent Curve Model
- Growth Rate Reliability (GRR)
- Effects of design on GRR
- Relation of GRR to statistical power

Also

- Effect of statistical test on power
- Multi- vs. single-parameter tests

- Recommendations

Statistical Power

What is statistical power (π)

- Probability of rejecting null hypothesis in favor of alternative
- β: Type II error (fail to reject null hypothesis)
- $\pi = 1 - \beta$
- $\pi \geq .80$ (Cohen, 1988)

Factors that influence power

- Statistical significance α
- N
- Effect sizes
- Measurement error
  - ▷ Design
  - ▷ Statistical Test

Reliability & Change

Basic question:

- If the effect is present, how likely is it, that we are going to detect it?
- As discussed before, a number of parameters influence power.
- Prominent:
    - Effect size (signal)
    - Error (noise)
- Classical Test Theory:
    - $\text{Reliability} = \frac{\text{Signal}}{\text{Signal}+\text{Noise}}$

Reliability & Change

Related:

- Reliability of difference scores

Difference scores are (even today!) considered, by some, to be flawed

- ▷ "..., we believe that difference scores are particularly likely to have psychometric flaws" Furr & Bacharach (2014), p. 160
- However: Reliability is not a matter of believe

Longstanding debate:

- \- Lord (1956, 1963); Cronbach & Furby (1970)
- \+ Rogosa (1982, 1995), Rogosa and Willett (1983)

Current state: Most of the time difference scores are reliable – under the condition that there *is* change

Reliability & Change

- Rogosa (1994):

of true change and the variance of the difference of the errors. For parameter configurations that require all individuals to grow at about the same rate, the low reliability of the difference score properly reveals that you can't detect individual differences that ain't there.

Reliability & Change

- Debate about difference scores: 1950's to 1990's

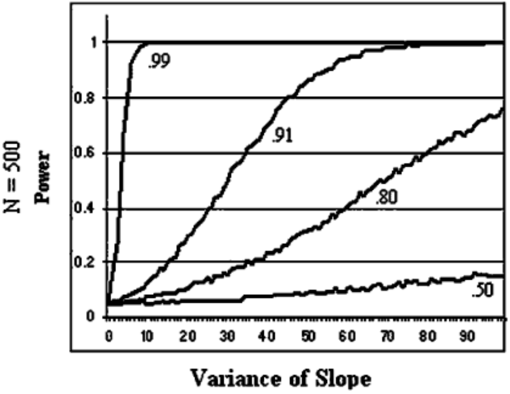
- Similar "debate" in longitudinal models
  - How well can we detect individual difference in change?
  - Number of papers: Hertzog et la. (2006, 2008, 2010)

Finding: Most existing longitudinal studies do not have sufficient power to detect either individual differences in change or covariances among rates of change.

Quote: The authors âĂIJpersuade LGCM [latent growth curve model] users not to rest on substantive findings, which might be invalid because of inherent LGCM lack of power under specific conditionsâĂİ (von Oertzen et al. 2010, p. 115)

Reliability & Change

e.g. Figure 3, Hertzog et al. (2008) Evaluating the Power of Latent Growth Curve Models to Detect Individual Differences in Change, *SEM, 15*, 541–563



Reliability & Change

- Not typically what we see.
- Significant variance in slope is *very* common.
- Why these results?
- Striking similarities with difference score debate

- Revisited papers and ran own simulation studies[‡]

[‡] Rast, P.,& Hofer, S. M. (2014). Longitudinal design considerations to optimize power to detect variances and covariances among rates of change: simulation results based on actual longitudinal studies. *Psychological Methods, 19*, 133âĂŞ54.

Latent Curve Model aka Multilevel Model

- Modeling individual differences (in change) in SEM notation

General expression for a time-structured latent curve model

$$\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon},$$

- $\mathbf{y}$ response vector for person i and w occasions where $\mathbf{y}' = [y_{i1}, y_{i2}, ..., y_{iw}]$
- $\boldsymbol{\Lambda}$ is the $(w \times p)$ factor loadings matrix with p parameters
  - Typically $\boldsymbol{\Lambda}_{w,2}$ reflects time structure
- $\boldsymbol{\eta}$ captures change characteristics as sum of fixed and random components $(\boldsymbol{\alpha} + \mathbf{z})$
- $\boldsymbol{\epsilon}$ contains vector of w residual terms

Standard assumptions ($\mathbf{y} = \boldsymbol{\Lambda}\boldsymbol{\eta} + \boldsymbol{\epsilon}$):

- $E(\boldsymbol{\epsilon}) = 0$
- $COV(\boldsymbol{\eta}, \boldsymbol{\epsilon}) = 0$
- $E(\mathbf{y}) = \boldsymbol{\mu}$
- $E(\boldsymbol{\eta}) = \boldsymbol{\alpha}$
- $COV(\boldsymbol{\eta}, \boldsymbol{\eta}) = \boldsymbol{\Psi}$
- $COV(\boldsymbol{\epsilon}, \boldsymbol{\epsilon}) = \boldsymbol{\Theta}$

Mean structure for **y**:

$$\boldsymbol{\mu} = \boldsymbol{\Lambda}\boldsymbol{\alpha}$$

Covariance structure:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Psi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}.$$

Covariance matrix of the random coefficients $\mathbf{\Psi}$:

$$\mathbf{\Psi} = \begin{bmatrix} \sigma_I^2 & \\ \sigma_{SI} & \sigma_S^2 \end{bmatrix}$$

- σ_I^2 captures individual differences in intercept
- σ_{IS} covariance among intercept and slope
- σ_S^2 captures individual differences in slope
- Typically unstructured

$\mathbf{\Lambda}$ defines the loadings (i.e., intercepts and slopes)

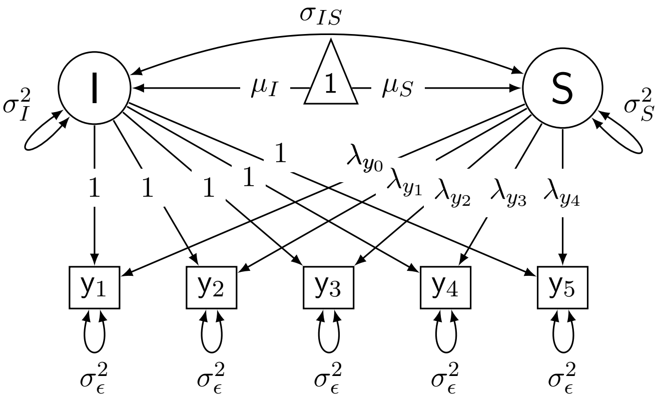
$$\mathbf{\Lambda} = \begin{bmatrix} 1 & \lambda_0 \\ 1 & \lambda_1 \\ \vdots & \vdots \\ 1 & \lambda_w \end{bmatrix}.$$

Time scale goes into $\mathbf{\Lambda}$

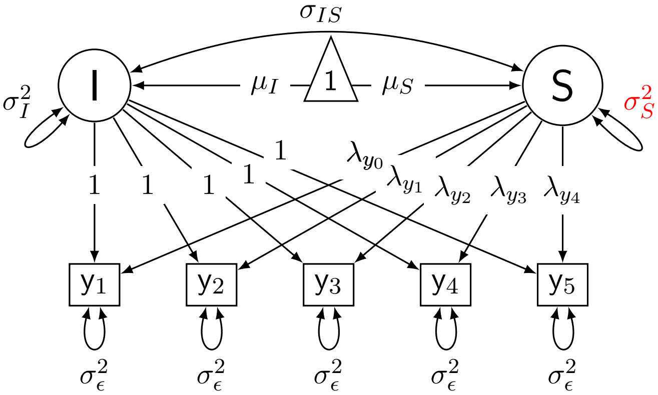
Also, $\mathbf{\Theta}$ may take different error structures (eg. Grimm & Widaman, 2010).

- Standard assumption $\mathbf{\Theta} = \sigma_\epsilon^2 \mathbf{I}$.

Graphical Representation



Graphical Representation



Design Considerations

How well can σ_S^2 be detected?

- Reliability of the growth rate (Willett, 1989)

$$\text{GRR} = \frac{\sigma_S^2}{\sigma_S^2 + \left[\frac{\sigma_\epsilon^2}{\mathsf{SST}}\right]}$$

$$\mathsf{SST} = \sum(\lambda_w - \overline{\lambda})^2$$

GRR:

- σ_S^2
- σ_ϵ^2
- SST; Sum of squared deviation of time points about their mean

! GRR is *not* an index of measurement reliability

SST and Study Design

SST$= \sum(\lambda_w - \overline{\lambda})^2$:

- Sum of squared deviations of time points about their mean
- Influences the effect of σ_ϵ^2 in GRR
- Captures study design
- Keeps GRR constant for different scale transformations

Ingredients of SST

- Number of measurement occasions
- Study length
- Spacing and interval between measurements

SST is potentially the only feature we (researchers) can influence

Influence of SST on σ_ϵ^2

Measurement error is expressed in σ_ϵ^2

- SST can considerably reduce effect of σ_ϵ^2 on GRR

For example: Repeated measurements

- at occasions: 0, 2, 5

▷ SST $= (0 - 2.3\overline{3})^2 + (2 - 2.3\overline{3})^2 + (5 - 2.3\overline{3})^2 = 12.6\overline{6}$

i.e. $\frac{\sigma_\epsilon^2}{12.6\overline{6}}$

- at occasions: 0, 2, 5, 7

▷ SST $= (0 - 3.5)^2 + (2 - 3.5)^2 + (5 - 3.5)^2 + (7 - 3.5)^2 = 29$

i.e. $\frac{\sigma_\epsilon^2}{29}$

Influence of SST on σ_ϵ^2

Measurement error is expressed in σ_ϵ^2

- SST can considerably reduce effect of σ_ϵ^2 on GRR

For example: Repeated measurements

- at occasions: 0, 2, 5
  - ▷ SST $= (0-2.3\overline{3})^2 + (2-2.3\overline{3})^2 + (5-2.3\overline{3})^2 = 12.6\overline{6}$

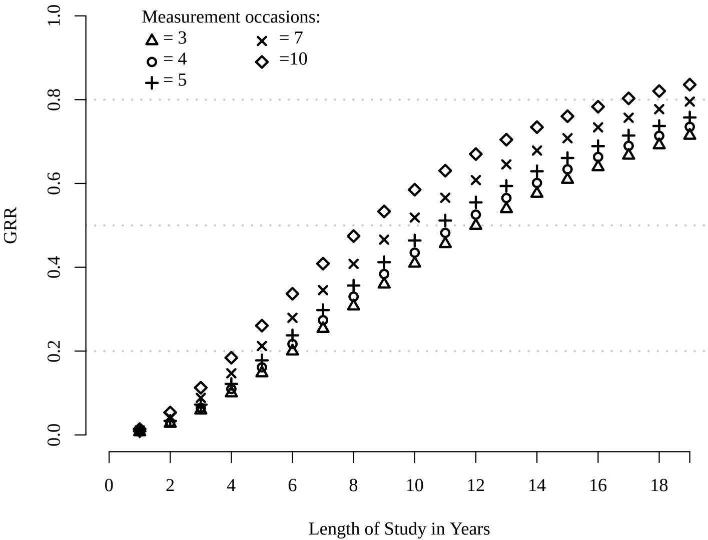
i.e. $\frac{\sigma_\epsilon^2}{12.6\overline{6}}$

- at occasions: 0, 2, 5, 7
  - ▷ SST $= (0-3.5)^2 + (2-3.5)^2 + (5-3.5)^2 + (7-3.5)^2 = 29$

i.e. $\frac{\sigma_\epsilon^2}{29}$

"...with sufficient waves added, the influence of fallible measurement rapidly dwindles to zero" (Willet, 1998, p. 598)

Study Length and Number of Measurement Occasions



Rast & Hofer (2014), Figure 1

SST and Spacing of Measurement Occasions

For example: Four repeated measurement across 7 years but unequally spaced

- At occasions: 0, 2, 5, 7
  - ▷ SST $= (0-3.5)^2 + (2-3.5)^2 + (5-3.5)^2 + (7-3.5)^2 = 29$
- At occasions: 0, 1, 6, 7
  - ▷ SST $= (0-3.5)^2 + (1-3.5)^2 + (6-3.5)^2 + (7-3.5)^2 = 37$

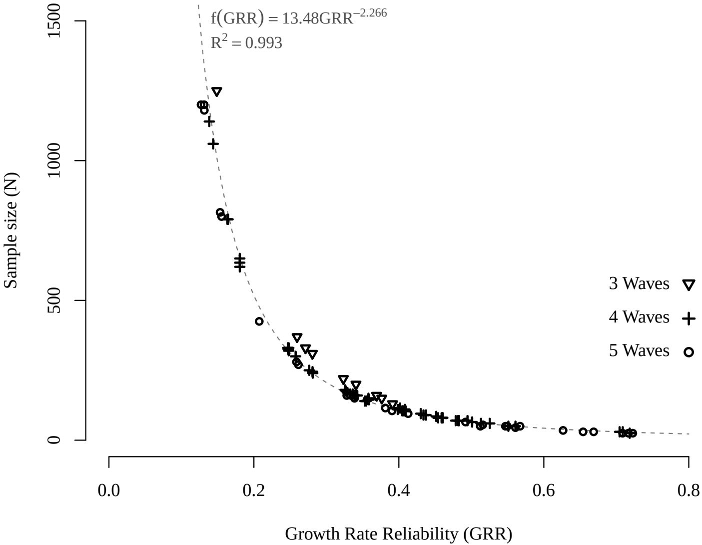
Under otherwise equivalent conditions (σ_S^2, σ_ϵ^2) we obtain different reliability estimates for the growth rates:

$$\frac{\sigma_S^2}{\sigma_S^2 + \left[\frac{\sigma_\epsilon^2}{37}\right]} > \frac{\sigma_S^2}{\sigma_S^2 + \left[\frac{\sigma_\epsilon^2}{29}\right]}$$

Relation of GRR to Statistical Power

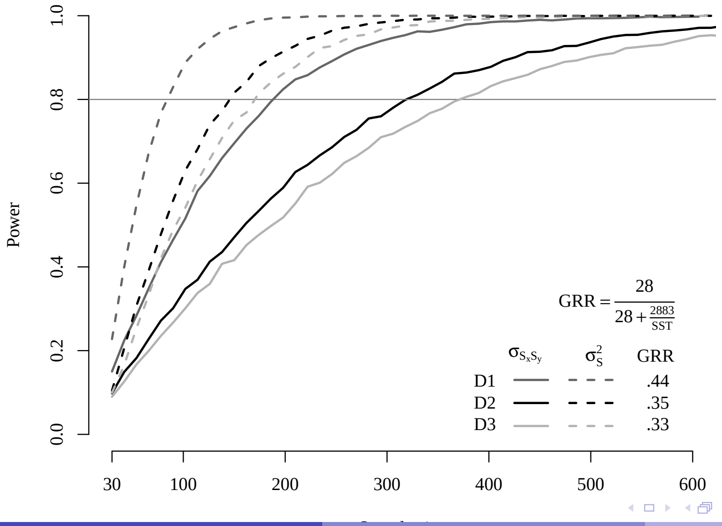
- We now have seen how different design and effect size aspects influence GRR
- How does that all relate to power?
- how does that "look" in actual data?
- What is the functional relation among power and reliability

Relation of GRR to Statistical Power



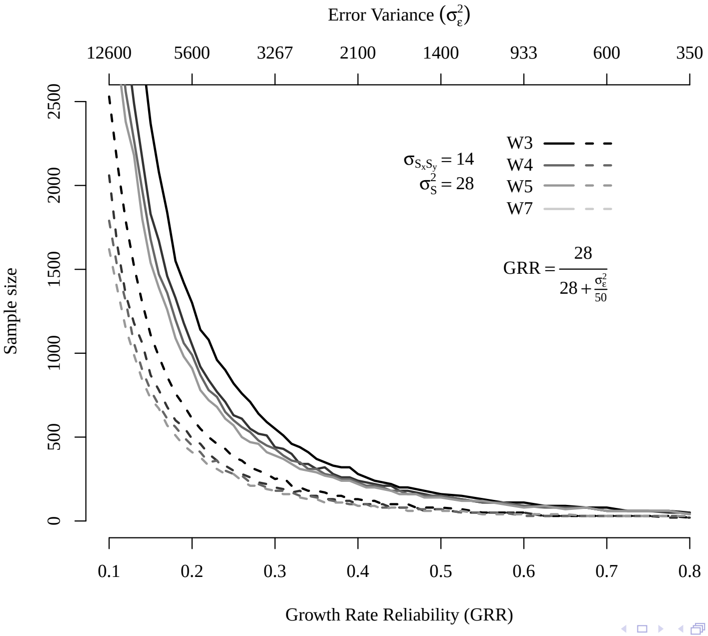
Rast & Hofer (2014), Figure 4

Effect of Time Interval on N



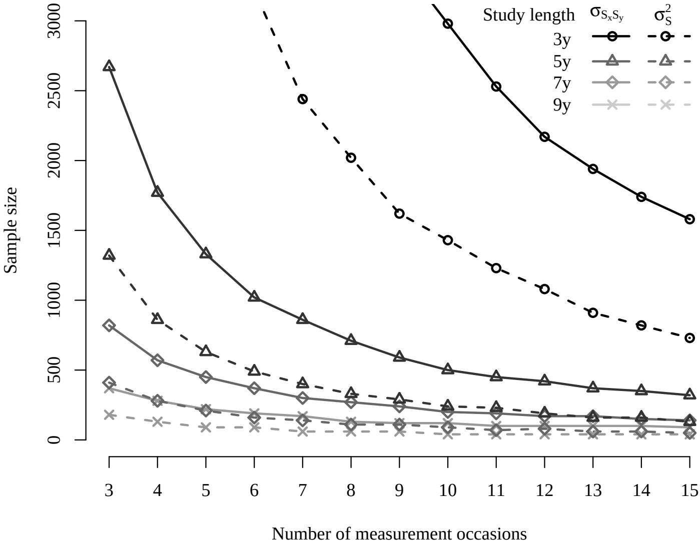
- $D1 = 0, 1, 9, 1$
  $\mathsf{SST}_{D1} = 82$
- $D2 = 0, 3.3, 6.$
  $\mathsf{SST}_{D2} = 55.4$
- $D3 = 0, 4.9, 5.$
  $\mathsf{SST}_{D3} = 50$

Rast & Hofer (2014), Fig



Rast & Hofer (2014), Figure 10

Unique Contribution of Duration and Number of Waves



Rast & Hofer (2014), Figure 7

Statistical Test

Power depends on the type of statistical test

- Single-parameter
- Multiparameter

In case of testing for individual differences in slopes/change

- effect of covariance on multiparameter test

Single- and Multiparameter Tests

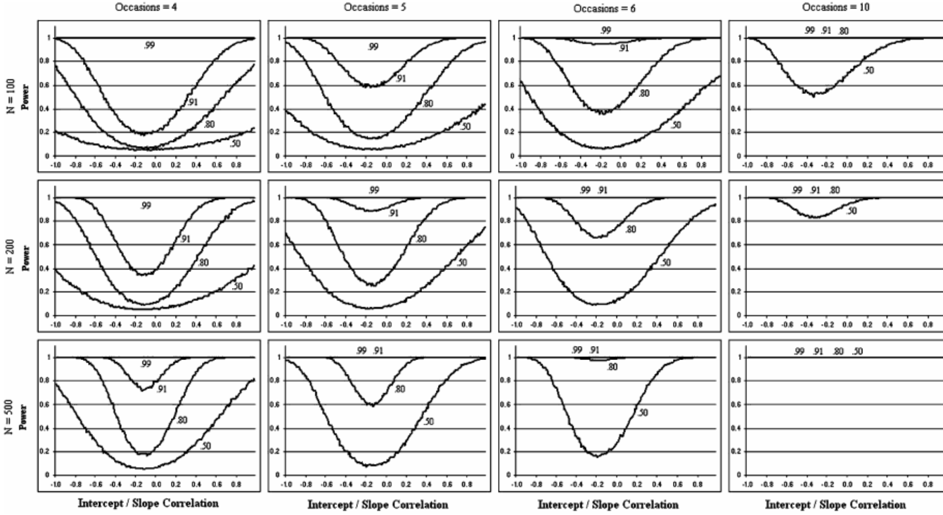
- Single-parameter test (e.g. Wald test) tests significance of σ_S^2 via $\hat{SE}$ of $\hat{\sigma}_S^2$
- Multiparameter test (e.g. LR) of σ_S^2 is based on constrained vs. unconstrained model:

$$\begin{bmatrix} \sigma_I^2 & \\ \sigma_{SI} & \sigma_S^2 \end{bmatrix} \text{ vs. } \begin{bmatrix} \sigma_I^2 & \\ 0 & 0 \end{bmatrix}$$

- ▷ Multiparameter: 2 df as covariance is included
- ▷ Single-parameter: 1 df

Effect of covariance

For the most part, we are not interested in $COV(I, S)$ - however covariance can influence power of multiparameter test: (Hertzog et al., 2008, Figure 1)



Effect of covariance

Likelihood ratio test (LR)

- $LR = 2(L_1 - L_0)$
- Maximized log-likelihood values for
  - unrestricted L_0
  - restricted L_1 models

LR test for growth variance σ_S^2 expressed in expected variance

- Unrestricted model:
  $\sigma_I^2 + 2\lambda\sigma_{IS} + \lambda^2\sigma_S^2$
- Restricted model with $\sigma_S^2 = 0$:
  $\sigma_I^2 + 0 + 0$

Size of LR determines statistical significance.

- LR is approximately χ^2 distributed
- Large LR, with respect to df, yields statistical significance
- Small LR yields no statistical significance

This implies

- $\sigma_I^2 + 2\lambda\sigma_{IS} + \lambda^2\sigma_S^2$ vs. $\sigma_I^2 + 0 + 0$
- $2\lambda\sigma_{IS} + \lambda^2\sigma_S^2$

▷ Is the sum of all growth effects $>$ 0?

- For a given λ there is a covariance σ_{IS} that nullifies the growth effects

$$\begin{aligned} 2\lambda\sigma_{IS} + \lambda^2\sigma_S^2 =& 0 \\ 2\lambda\sigma_{IS} =& -\lambda^2\sigma_S^2 \\ \sigma_{IS} =& -\frac{\lambda^2\sigma_S^2}{2\lambda} \\ \sigma_{IS} =& -\lambda\frac{\sigma_S^2}{2} \end{aligned}$$

In correlation metric:

$$r = -\lambda\frac{\sigma_S^2}{2\sqrt{\sigma_I^2\sigma_S^2}}$$

Given that

$$r_{IS} = -\lambda \frac{\sigma_S^2}{2\sqrt{\sigma_I^2 \sigma_S^2}}$$

is weighted by λ, power to detect σ_S^2 in different designs will be minimized by different r_{IS}

As r_{IS} approaches $-\lambda \frac{\sigma_S^2}{2\sqrt{\sigma_I^2 \sigma_S^2}}$

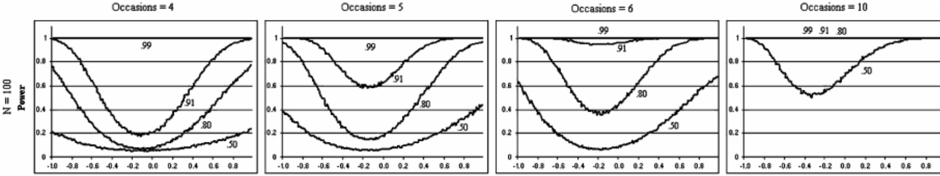
- Power that is drawn from covariance effects decreases for LR
- At minimum, power equals that of single parameter test
  - All information from covariance is canceled out and LR test only relies on difference among σ_I^2

worst multiparameter test is just as powerful as single-parameter test

t best multiparameter test is *much* more powerful than single-parameter test

e.g. Hertzog et al. 2008

- $\sigma_I^2 = 100$
- $\sigma_S^2 = 50$
- Unit scale $\lambda = 0, 0.11, 0.22, 0.33, 0.44, 0.56, 0.67, 0.78, 0.89, 1.00$



$r_4 = \frac{-0.33\times 50}{2\sqrt{100\times 50}} = -.11 \quad r_5 = -.15 \quad r_6 = -.19 \quad r_{10} = -.35$

Discussion and Recommendations

- Study design can have enormous impact on sample size requirements

- Design is very important in studies with short duration
  - Intensive designs in early phases?
  - More frequent measurement
  - Beware of unwanted effects such as warm-up and retest

- Perform power analysis under different design conditions!
  - Making generally valid statements about power and design is very difficult

- Don't waste power on poor tests