Crossed Random Effects for Other (Not Longitudinal) Repeated Measures Designs

- Topics:
  - **ANOVA for repeated measures (RM)**
  - MLM for repeated measures

RM ANOVA works well when…

- Experimental stimuli are ***controlled*** and ***exchangeable***
  - Controlled → Constructed, not sampled from a population
  - Exchangeable → Stimuli vary only in dimensions of interest
  - …What to do with non-exchangeable stimuli (e.g., words, scenes)?

- Experimental manipulations create ***discrete conditions***
  - e.g., set size of 3 vs. 6 vs. 9 items
  - e.g., response compatible vs. incompatible distractors
  - …What to do with *continuous* item predictors (e.g., time, salience)?

- One has ***complete data***
  - e.g., if outcome is RT and accuracy is near ceiling
  - e.g., if responses are missing for no systematic reason
  - …What if data are not missing completely at random (e.g., inaccuracy)?

Motivating Example: Psycholinguistic Study Designs

- Word Recognition Tasks (e.g., Lexical Decision)
  - Word lists are constructed based on targeted dimensions while controlling for other relevant dimensions
  - Outcome = RT to decide if the stimulus is a word or non-word (accuracy is usually near ceiling)

- Tests of effects of experimental treatment are typically conducted with the person as the unit of analysis…
  - Average the responses over words within conditions
    - Contentious fights with reviewers about adequacy of experimental control when using real words as stimuli
    - Long history of debate as to how words as experimental stimuli should be analyzed… F_1 ANOVA or F_2 ANOVA (or both)?
    - F_1 only creates a "Language-as-Fixed-Effects Fallacy" (Clark, 1973)

ANOVAs on Summary Data

Original Data per Person

	B1	B2
A1	Item 001 Item 002 ..... Item 100	Item 101 Item 102 ..... Item 200
A2	Item 201 Item 202 ..... Item 300	Item 301 Item 302 ..... Item 400

Person Summary Data

	B1	B2
A1	Mean (A1, B1)	Mean (A1, B2)
A2	Mean (A2, B1)	Mean (A2, B2)

"F1" Within-Persons ANOVA on *N* persons:

$$RT_{cp} = \gamma_0 + \gamma_1 A_c + \gamma_2 B_c + \gamma_3 A_c B_c + U_{0p} + e_{cp}$$

"F2" Between-Items ANOVA on *I* items:

$$RT_i = \gamma_0 + \gamma_1 A_i + \gamma_2 B_i + \gamma_3 A_i B_i + e_i$$

Item Summary Data

	B1
A1, B1	Item 001 = Mean(Person 1, Person 2,... Person *N*) Item 002 = Mean(Person 1, Person 2,... Person *N*) ..... Item 100
A1, B2	Item 101 = Mean(Person 1, Person 2,... Person *N*) Item 102 = Mean(Person 1, Person 2,... Person *N*) ..... Item 200
A2, B1	Item 201 = Mean(Person 1, Person 2,... Person *N*) Item 202 = Mean(Person 1, Person 2,... Person *N*) ..... Item 300
A2, B2	Item 301 = Mean(Person 1, Person 2,... Person *N*) Item 302 = Mean(Person 1, Person 2,... Person *N*) ..... Item 400

Choosing Amongst ANOVA Models

- **F1** Within-Persons ANOVA on **person** summary data:
  - Within-condition ***item*** variability is gone, so items assumed fixed
- **F2** Between-Items ANOVA on **item** summary data:
  - Within-item ***person*** variability is gone, so persons assumed fixed

- Historical proposed ANOVA-based resolutions:
  - **F'** → quasi-F test with random effects for both persons and items (Clark, 1973), but requires complete data (**uses least squares**)
  - **Min F'** → lower-bound of F′ derived from F1 and F2 results, which does not require complete data, but is **too conservative**
  - **F1 x F2 criterion** → effects are only "**real**" if they are significant in **both F1 and F2 models** (*aka*, death knell for psycholinguists)

  - But neither model is complete (two wrongs don't make a right)…
  - **MLM to the rescue?**

Crossed Random Effects for Other (Not Longitudinal) Repeated Measures Designs

- Topics:
  - ANOVA for repeated measures
  - **MLM for repeated measures**

Multilevel Models to the Rescue?

Original Data per Person

	B1	B2
A1	Item 001 Item 002 ..... Item 100	Item 101 Item 102 ..... Item 200
A2	Item 201 Item 202 ..... Item 300	Item 301 Item 302 ..... Item 400

Pros:

- Use all original data, not summaries
- Responses can be missing at random
- Can include continuous predictors

Cons:

- **Is still wrong (is ~F1 ANOVA)**

Level 1: $y_{ip} = \beta_{0p} + \beta_{1p}A_{ip} + \beta_{2p}B_{ip} + \beta_{3p}A_{ip}B_{ip} + e_{ip}$

Level 2: $\beta_{0p} = \gamma_{00} + U_{0p}$

$\beta_{1p} = \gamma_{10}$

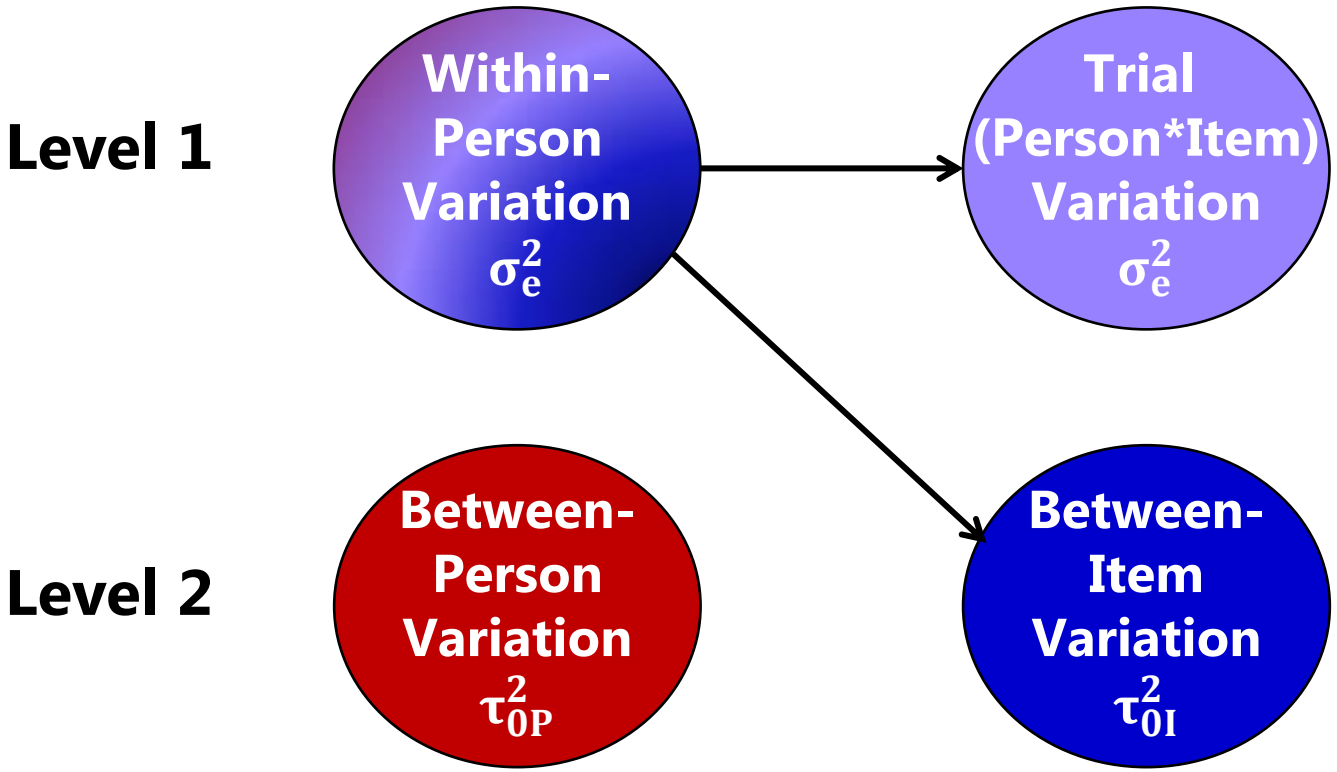
$\beta_{2p} = \gamma_{20}$

$\beta_{3p} = \gamma_{30}$

Level 1 = Within-Person Variation (Across Items)

Level 2 = Between-Person Variation

Multilevel Models to the Rescue?



Empty Means, Crossed Random Effects Models

- **Residual-only model:**
  - $RT_{tip} = \gamma_{000} + e_{tip}$
  - Assumes no effects (dependency) of subjects or items

- **Random persons (or "subjects") model:**
  - $RT_{tip} = \gamma_{000} + \mathbf{U_{00p}} + e_{tip}$
  - Models systematic mean differences **between persons**

- **Random persons and items model:**
  - $RT_{tip} = \gamma_{000} + U_{00p} + \mathbf{U_{0i0}} + e_{tip}$
  - Also models systematic mean differences **between items**

A Better Way* of (Multilevel) Life

Random effects over **persons** of **item** or **trial** predictors can also be tested and predicted.

- **Multilevel Model with *Crossed* Random Effects:**

$$RT_{tip} = \gamma_{000} + \gamma_{010}A_i + \gamma_{020}B_i + \gamma_{030}A_iB_i + U_{00p} + U_{0i0} + e_{tip}$$

t trial
i item
p person

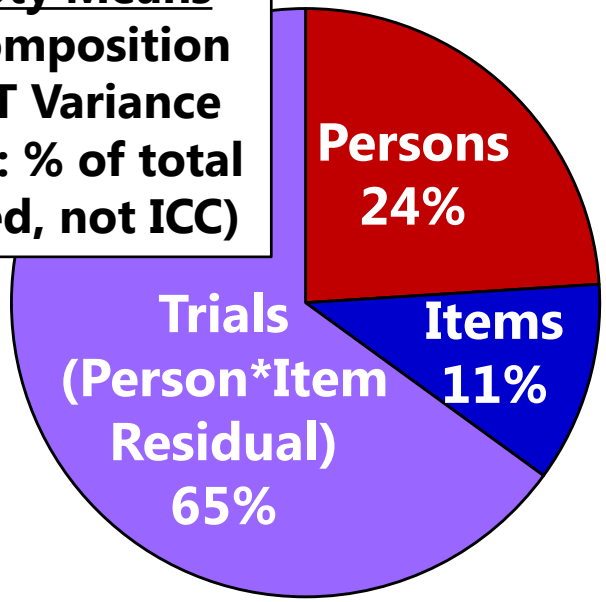
- Both **subjects** and **items** as random effects:
  - Subject predictors explain between-subject mean variation: τ^2_{00P}
  - Item predictors explain between-item mean variation: τ^2_{0I0}
  - Trial predictors explain trial-specific residual variation: σ^2_e
  - * Except for *fierce debates about whether all item predictors need random slopes no matter what (aka, "maximal models")*

Example 4: Psycholinguistic Study

(Locker, Hoffman, & Bovaird, 2007)

- Crossed design: 38 subjects by 39 items (words or nonwords)
- Lexical decision task: RT to decide if word or nonword
- 2 word-specific predictors of interest:
  - A: Low/High Phonological Neighborhood Frequency
  - B: Small/Large Semantic Neighborhood Size

Empty Means Decomposition of RT Variance (note: % of total is used, not ICC)



Model and Results

$$RT_{tip} = \gamma_{000} + \gamma_{010}A_i + \gamma_{020}B_i + \gamma_{030}A_iB_i + U_{00p} + U_{0i0} + e_{tip}$$

Pseudo-R^2:

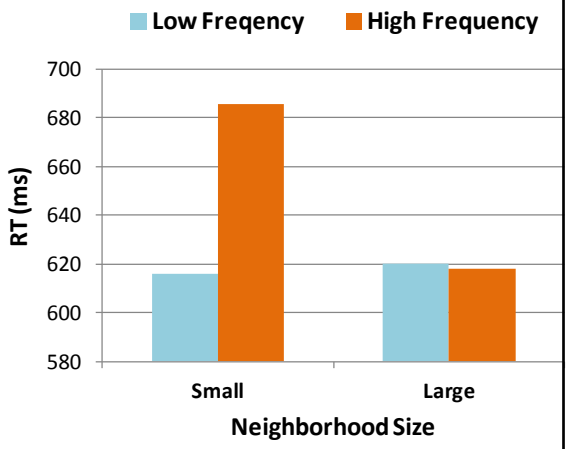
Residual ≈ 0%

Subjects ≈ 0%

Items ≈ 30%*

Total R^2 ≈ 3.3%

***Significant item variability remained**



Tests of Fixed Effects by Model

	A: Frequency Marginal Main Effect	B: Size Marginal Main Effect	A*B: Interaction of Frequency by Size
F_1 Subjects ANOVA	F (1,37) = 16.1 p = .0003	F (1,37) = 14.9 p = .0004	F (1,37) = 38.2 p < .0001
F_2 Words ANOVA	F (1,35) = 5.3 p = .0278	F (1,35) = 4.5 p = .0415	F (1,35) = 5.7 p = .0225
F' min (via ANOVA)	F (1,56) = 4.0 p = .0530	F (1,55) = 3.5 p = .0710	F (1,45) = 5.0 p = .0310
Crossed MLM (via REML)	F (1,32) = 5.4 p = .0272	F (1,32) = 4.6 p = .0393	F (1,32) = 6.0 p = .0199

Ch. 12 Simulation: Type 1 Error Rates

Condition			Models					
Item Variance	Subject Variance		**1: Both Random Effects**	2: Random Subjects Only	3: Random Items Only	4: No Random Effects	5: F1 Subjects ANOVA	6: F2 Item ANOVA
Item Effect:								
2	2		**0.03**	0.09	0.03	0.09	0.09	0.03
2	10		**0.05**	0.14	0.05	0.12	0.15	0.05
10	2		**0.04**	0.32	0.04	0.31	0.32	0.04
10	10		**0.05**	0.31	0.05	0.29	0.33	0.05
Subject Effect:								
2	2		**0.04**	0.04	0.12	0.11	0.04	0.12
2	10		**0.05**	0.05	0.34	0.34	0.05	0.36
10	2		**0.04**	0.03	0.12	0.09	0.03	0.12
10	10		**0.06**	0.06	0.34	0.31	0.05	0.37

Model Items as Fixed → Wrong Item Effect

Condition			Models					
Item Variance	Subject Variance		1: Both Random Effects	**2: Random Subjects Only**	3: Random Items Only	4: No Random Effects	**5: F1 Subjects ANOVA**	6: F2 Item ANOVA
Item Effect:								
2	2		0.03	**0.09**	0.03	0.09	**0.09**	0.03
2	10		0.05	**0.14**	0.05	0.12	**0.15**	0.05
10	2		0.04	**0.32**	0.04	0.31	**0.32**	0.04
10	10		0.05	**0.31**	0.05	0.29	**0.33**	0.05
Subject Effect:								
2	2		0.04	0.04	0.12	0.11	0.04	0.12
2	10		0.05	0.05	0.34	0.34	0.05	0.36
10	2		0.04	0.03	0.12	0.09	0.03	0.12
10	10		0.06	0.06	0.34	0.31	0.05	0.37

Model Subjects as Fixed → Wrong Subject Effect

Condition			Models					
Item Variance	Subject Variance		1: Both Random Effects	2: Random Subjects Only	**3: Random Items Only**	4: No Random Effects	5: F1 Subjects ANOVA	**6: F2 Item ANOVA**
Item Effect:								
2	2		0.03	0.09	0.03	0.09	0.09	0.03
2	10		0.05	0.14	0.05	0.12	0.15	0.05
10	2		0.04	0.32	0.04	0.31	0.32	0.04
10	10		0.05	0.31	0.05	0.29	0.33	0.05
Subject Effect:								
2	2		0.04	0.04	**0.12**	0.11	0.04	**0.12**
2	10		0.05	0.05	**0.34**	0.34	0.05	**0.36**
10	2		0.04	0.03	**0.12**	0.09	0.03	**0.12**
10	10		0.06	0.06	**0.34**	0.31	0.05	**0.37**

Conclusions

- A RM ANOVA model may be less than ideal when:
  - Stimuli are not completely exchangeable within conditions
  - Item conditions (predictors) are not discrete
  - Missing data may result in bias, a loss of power, or both

- ANOVA is a special case of a more general family of multilevel models (with nested or crossed effects as needed) that can offer additional flexibility:
  - Useful in addressing statistical problems →
    - Dependency, heterogeneity of variance, unbalanced or missing data
    - Examine predictor effects pertaining to each source of variation more accurately given that all variation is properly represented in the model
  - Useful in addressing substantive hypotheses →
    - Examining individual differences in effects of experimental manipulations