## The Within/Between Problem in Psychology & Academia

Noémi K. Schuurman

March 2023

## The Within/Between Problem in Psychology

Psychological variables may consist of both within-person and between-person variance.

## The Within/Between Problem in Psychology

- Psychological variables may consist of both within-person and between-person variance.
- Statistical associations between such variables will be a combination of the within-person and between-person associations.

## The Within/Between Problem in Psychology

- Psychological variables may consist of both within-person and between-person variance.
- Statistical associations between such variables will be a combination of the within-person and between-person associations.
- ➤ This is a problem when we are interested in either the within-person or the between-person associations or both, rather than the mix.

Awareness of the Within/Between Problem often starts with a simple observation:

The repeated measures of a person may have a different distribution than the scores of different persons.



## Multiple units: Within-unit fluctuations and Between-unit differences

#### N=many data

- cross-sectional data
- panel data
- intensive longitudinal data

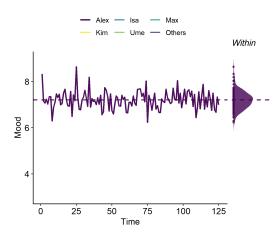
## Multiple units: Within-unit fluctuations and Between-unit differences

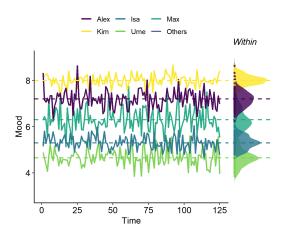
#### N=many data

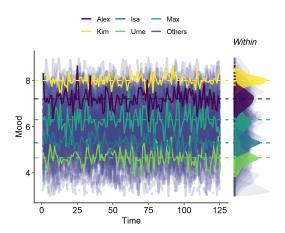
- cross-sectional data
- panel data
- intensive longitudinal data

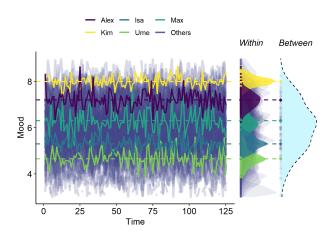
#### N=many data contain, for each variable:

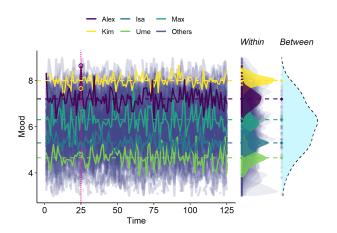
- within-subject fluctuations over time → within-subject variance or state variance
- ▶ time-stable differences between persons → between-subject variance or trait variance

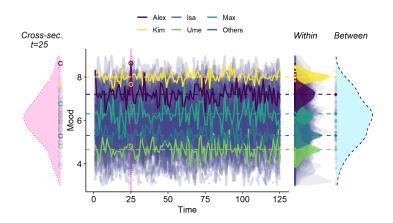


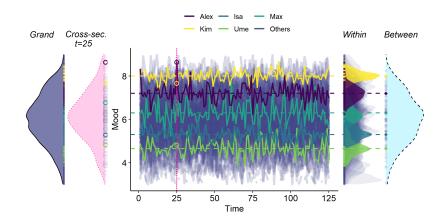












# Within-subject Between-subject differences are *different* variables

"Today Alex feels happier than yesterday."

- ▶ Within-subject differences: People's scores fluctuate over time.
- Within-subject variance: Variance of a person's scores over time.

# Within-subject Between-subject differences are *different* variables

"Today Alex feels happier than yesterday."

- ▶ Within-subject differences: People's scores fluctuate over time.
- Within-subject variance: Variance of a person's scores over time.

"Kim is typically happier than Alex."

- Stable Between-subject differences: People's expected values (mean scores) differ.
- 'Between-subject variance': Variance of the expected values (means) of different people.

# Within-subject Between-subject differences are *different* variables

- "Today Alex feels happier than yesterday."
  - ▶ Within-subject differences: People's scores fluctuate over time.
  - Within-subject variance: Variance of a person's scores over time.
- "Kim is typically happier than Alex."
  - Stable Between-subject differences: People's expected values (mean scores) differ.
  - 'Between-subject variance': Variance of the expected values (means) of different people.
- "Alex is happier today than Kim."
  - Cross-sectional differences in scores: People's scores at a certain point in time differ.
  - 'Cross-sectional variance': Variance of the scores of different people at a specific point in time.

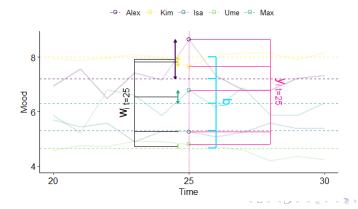
# Within-subject and Between-subject differences are *different* variables

We just capture *both* the within and between variables in our measurements:

## Within-subject and Between-subject differences are *different* variables

We just capture *both* the within and between variables in our measurements:

$$y_{it} = b_i + w_{it}$$
  
Decomposition of a Cross-section

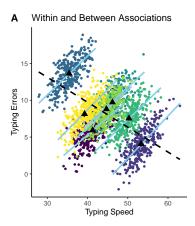


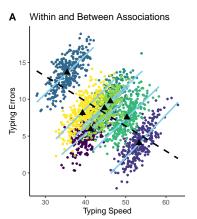
## Variance Decomposition

$$y_{it} = b_i + w_{it}$$
  $\sigma_y^2 = \sigma_b^2 + \mathop{E}\limits_{i}[\sigma_w^2] \ [app]$ 

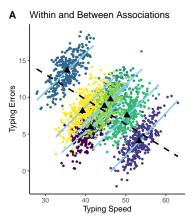
## Variance Decomposition

$$y_{it} = b_i + w_{it}$$
  $\sigma_y^2 = \sigma_b^2 + \mathop{E}\limits_{i}[\sigma_w^2] ext{ [app]}$   $\Sigma_y = \Sigma_b + \mathop{E}\limits_{i}[\Sigma_w]$ 



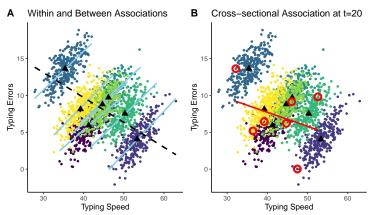


Within: As a person types faster, he/she tends to make more typo's.



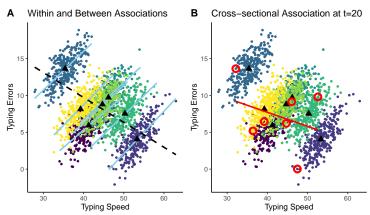
Within: As a person types faster, he/she tends to make more typo's.

Between: People that had a fast average typing speed across trials, tended to make fewer typo's on average across trials.



Within: As a person types faster, he/she tends to make more typo's.

Between: People that had a fast average typing speed across trials, tended to make fewer typo's on average across trials.



Within: As a person types faster, he/she tends to make more typo's.

Between: People that had a fast average typing speed across trials, tended to make fewer typo's on average across trials. Cross-section: People that typed relatively slowly in trial 20, tended to make more typo's in trial 20.

### Variance Decomposition

$$y_{it} = b_i + w_{it}$$

$$\sigma_y^2 = \sigma_b^2 + \underset{i}{E}[\sigma_w^2]$$

$$\Sigma_y = \Sigma_b + \underset{i}{E}[\Sigma_w]$$

If every person has the same within-person covariance matrix:

$$\rho_{y_1,y_2} = \rho_{b_1,b_2} \sqrt{ICC(y_1)} \sqrt{ICC(y_2)} + \rho_{w_1,w_2} \sqrt{1 - ICC(y_1)} \sqrt{1 - ICC(y_2)}$$

where the ICC is the proportion of between-person variance in y.

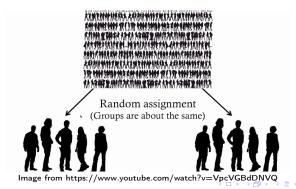
[app]

Use separate n=1 time series models for different persons (e.g., replicated time series design).

- Use separate n=1 time series models for different persons (e.g., replicated time series design).
- Center data before analysis.

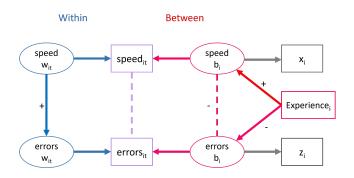
- ▶ Use separate n=1 time series models for different persons (e.g., replicated time series design).
- Center data before analysis.
- Center data during analysis: Latent centering, with a multilevel model.

- 1. Make use of experimental design with random assignment
  - ► All possible between variables are spread out equally over the control and treatment group via random assignment.
  - ► The difference between the two groups then is the *average* within person change that occured in the experimental group.



- 2. Filter out nuisance variable using control variables that are known relate (only) to the nuisance variable
  - Use "overall depression severity" scores as a control variable for associations between variables like "suicidual thoughts" and "sleep problems".
  - Really hard to do well.

#### Control Variables



The within/between problem from the causal inference perspective: Our observed variables are poor proxies/surrogates of the latent within- or between variables of interest.

3. Design your measurement instrument to measure *mostly* one type of variation.

3. Design your measurement instrument to measure *mostly* one type of variation.

3. Design your measurement instrument to measure *mostly* one type of variation.

#### Targeting between (trait):

 Indicate how accurately the statement describes your typical behavior or attitudes.

3. Design your measurement instrument to measure *mostly* one type of variation.

#### Targeting between (trait):

- Indicate how accurately the statement describes your typical behavior or attitudes.
- "I usually like to spend my free time with people".

3. Design your measurement instrument to measure *mostly* one type of variation.

#### Targeting between (trait):

- Indicate how accurately the statement describes your typical behavior or attitudes.
- "I usually like to spend my free time with people".
- Questionnaires with really high test-retest reliability.

3. Design your measurement instrument to measure *mostly* one type of variation.

#### Targeting between (trait):

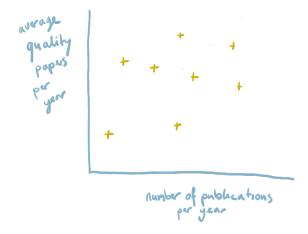
- Indicate how accurately the statement describes your typical behavior or attitudes.
- "I usually like to spend my free time with people".
- Questionnaires with really high test-retest reliability.

#### Targeting within (state):

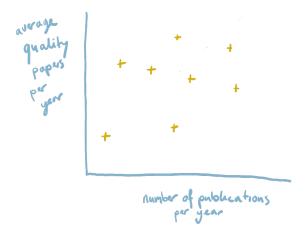
- Indicate how your state deviates from your typical state.
- "Today, I like to spend my free time with people..." → less than usual, the same as usual, more than usual.

## The Within/Between Problem of Academia

## The Within/Between Problem of Academia

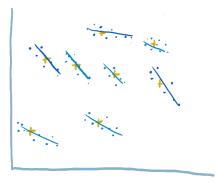


### The Within/Between Problem of Academia



#### Between:

People that tend to publish relatively a lot per year, tend to have relatively high av. quality papers. quality
papers
per
per



number of publications
per year

#### Between:

People that tend to publish relatively a lot per year, tend to have relatively high av. quality papers.

#### Within:

When a person pushes to publish more papers per year, their papers tend to have lower av. quality.