Part 2

Multilevel Modeling Intro and Code Walk Through

Objectives

- Introduce multi-level model and describe nested data structure
- Understand why we need to account for nested data
- Coding Walk Through

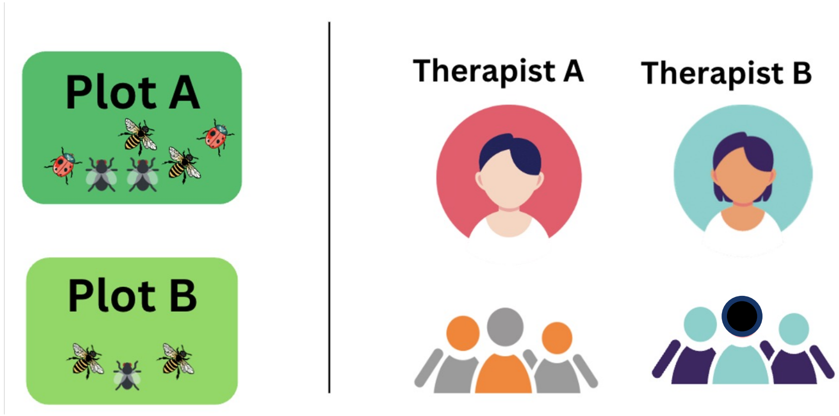
Multi-Level Model

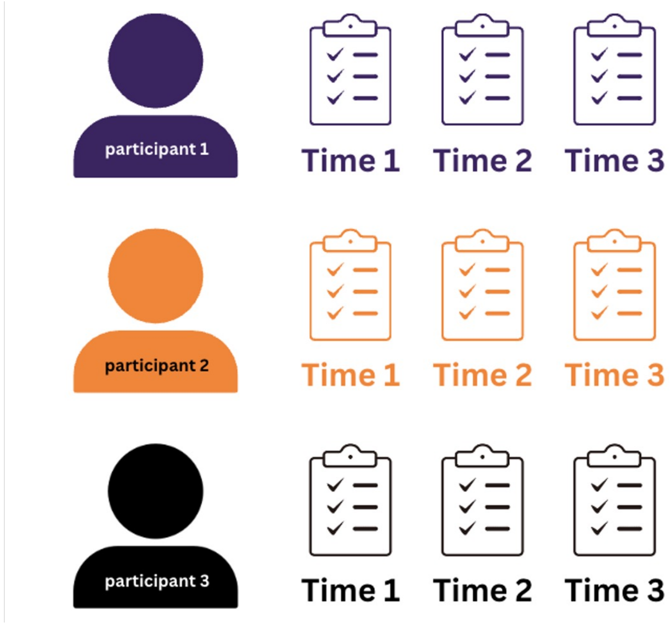
- An extension of the linear model that accounts for situations where we have dependence in our dataset
- A model we use when we have high variability between subjects (that we might not care about)

What are nested data?

- There are multiple observations on a single factor.
- Two common types of nested data structures:

Hierarchical (clustered) **Longitudinal** (repeated measurements)





Nested data are **dependent.**

This **violates** a main assumption (independence) of conventional analytic approaches (e.g., generalized linear model, ANOVA).

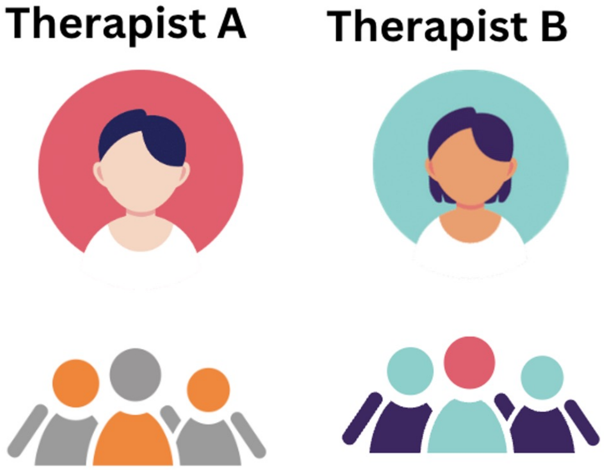
- Adding time/group parameters does not fix this problem

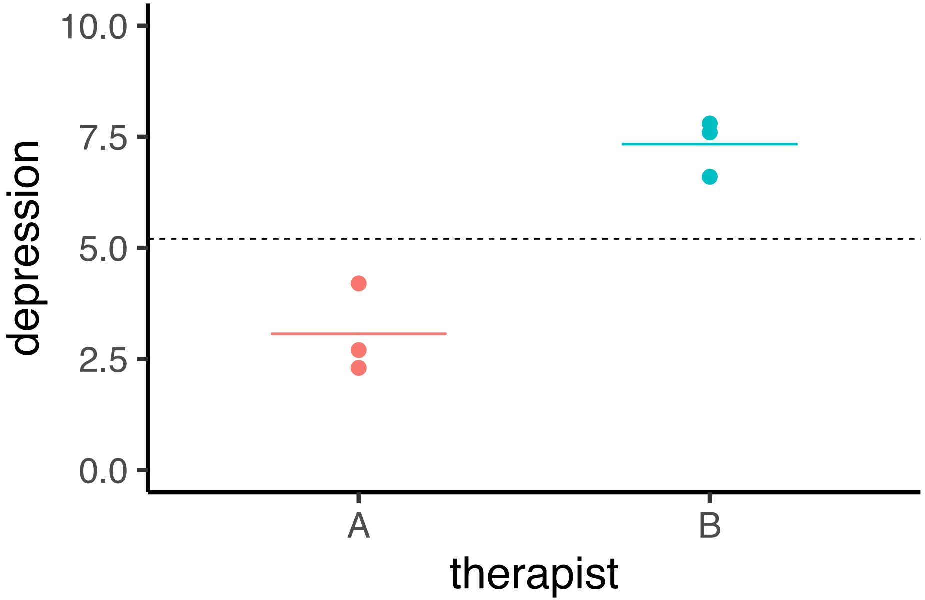
- a repeated measures ANOVA does not parse sources of variability

Why do we care?

- Samples from the same cluster are likely to be more similar than samples from different clusters
- Repeated measures from the same person are typically correlated
- Biased estimates
- This is why use use the multilevel model (also referred to as hierarchical modeling, random effects modeling, mixed-effects modeling, variance components modeling)

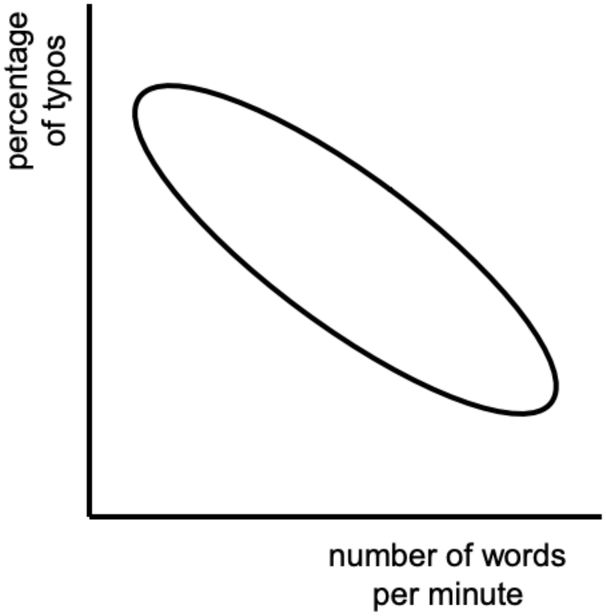
Sources of variability



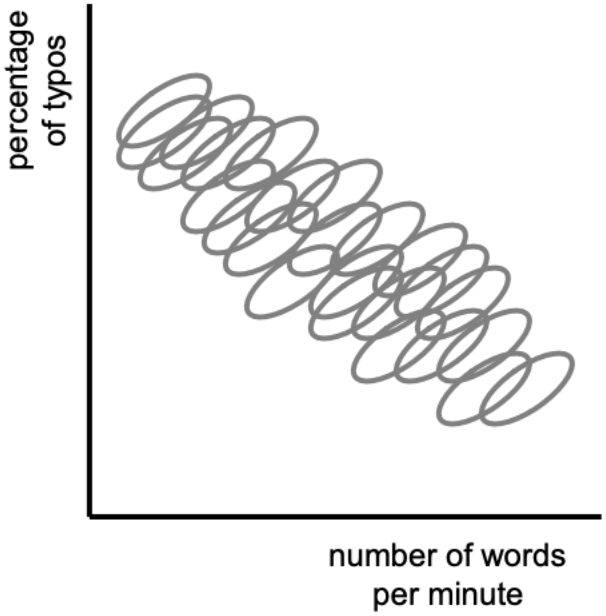


- **Between-group (Level 2)**
  - Clusters vary from each other (e.g., different group means)
- **Within-group (Level-1)**
  - Observations vary within a cluster (e.g., individuals vary from their group mean)

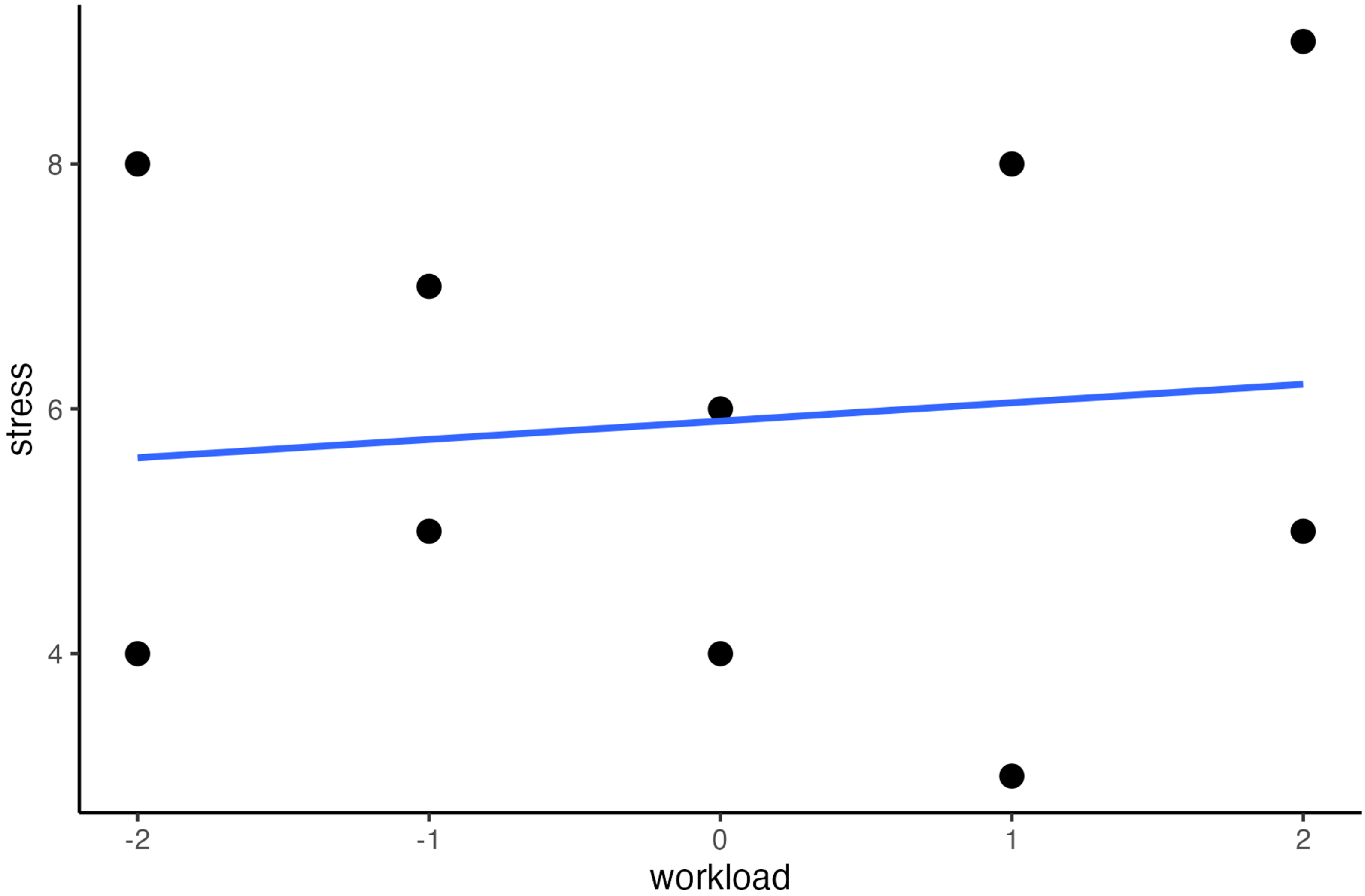
Cross-sectionally



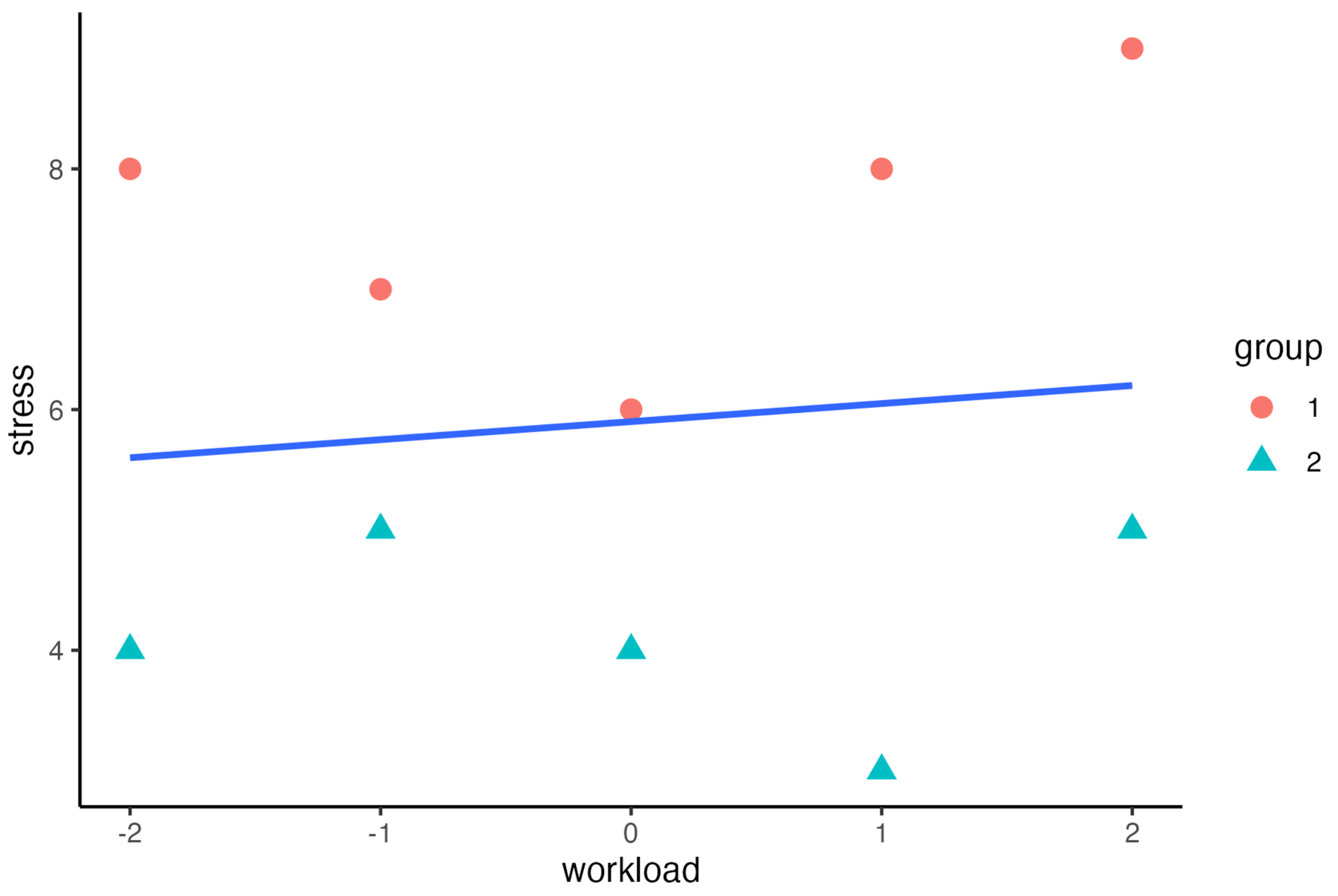
In general



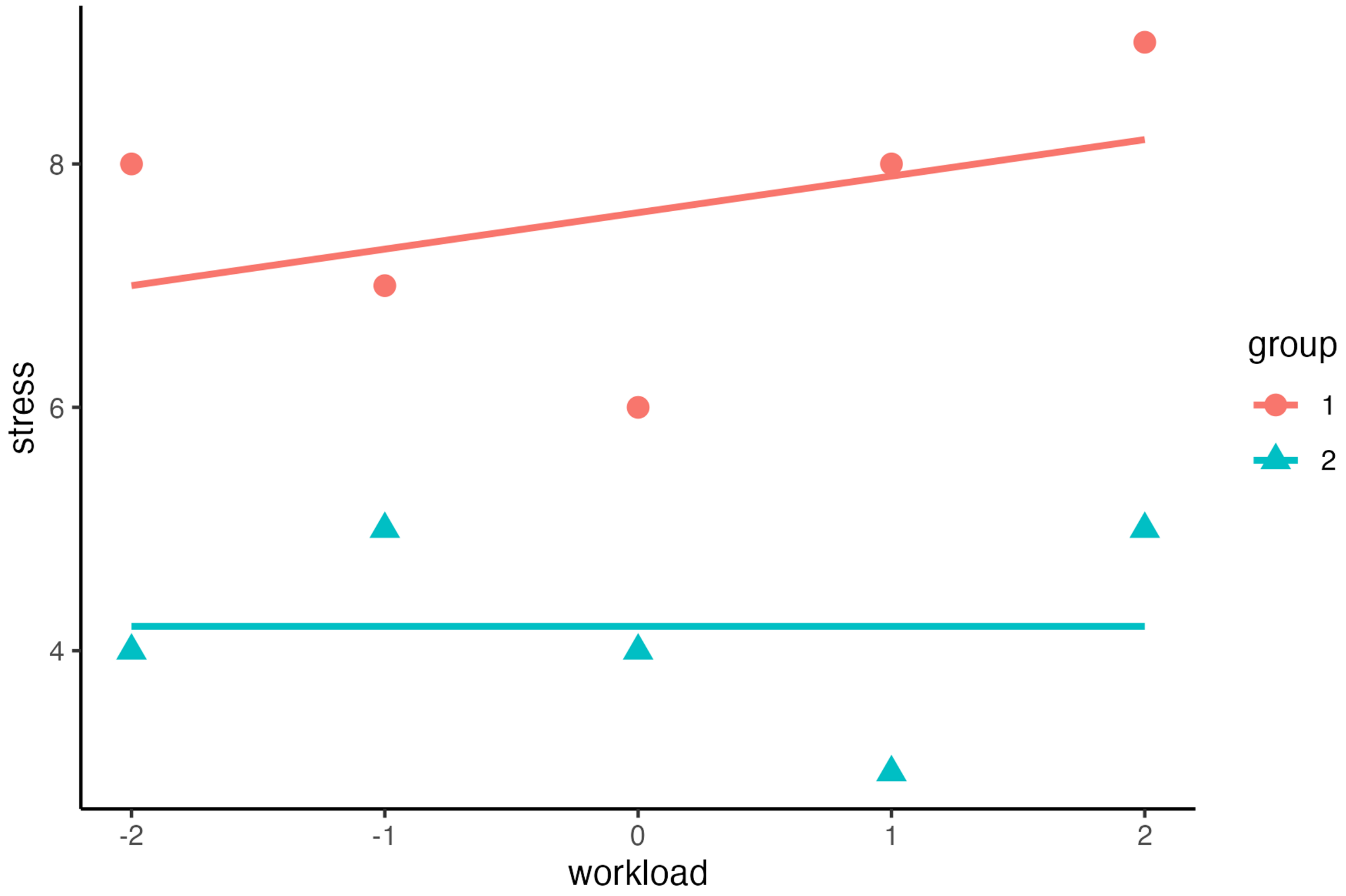
Adapted from Hamaker, E. L. (2012).



stress
8
6
4
-2
-1
0
1
2
workload



stress
8
6
4
workload
-2
-1
0
1
2
group
1
2

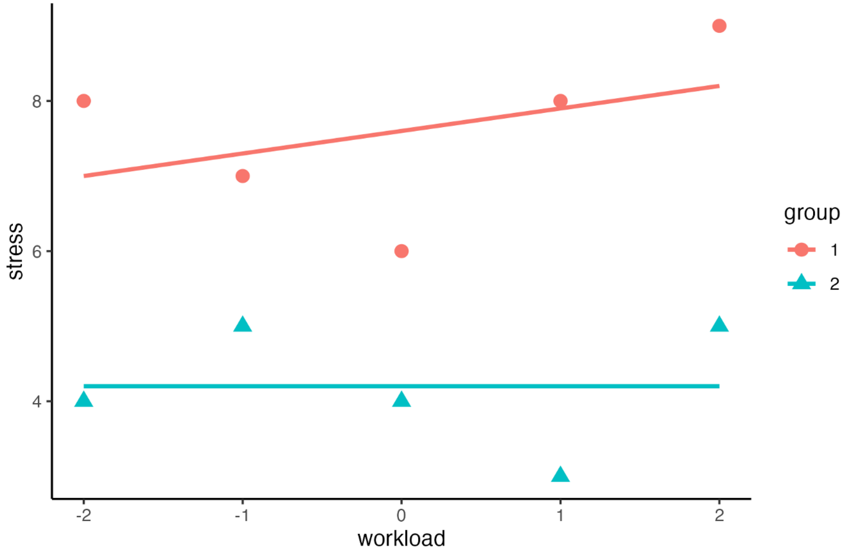

stress
8
6
4
-2
-1
0
1
2
workload
group
1
2

Fixed Effects

- Constants
- Fixed effects have within- and between-group variability too

Random Effects

- Distribution of values that vary across grouping variable
- Slopes and/or intercepts
- We don’t estimate a fixed value for random effects.



Equations Briefly ☺

Outcome variable intercept error

Level 1 $Y_{ij} = \beta_{0j} + e_{ij}$

Level 2 $\beta_{0j} = \gamma_{00} + U_{0j}$ Group deviation from grand mean

grand mean

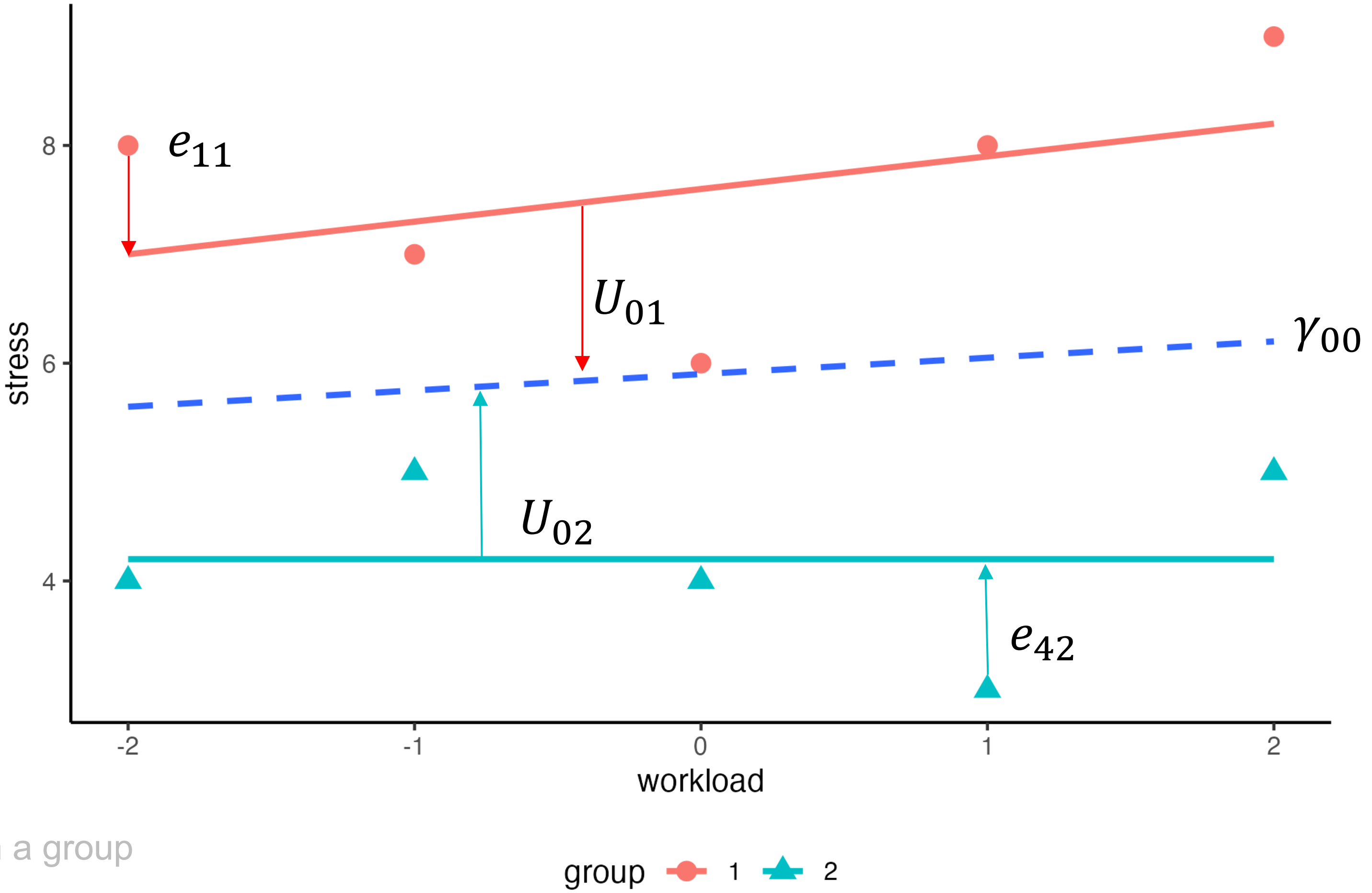
i = individual within a group

j = group

Explanatory variable

$$Y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \ldots e_{ij}$$

$$\beta_{1j} = \gamma_{10} + x_{1ij}$$



$$Y_{ij} = \beta_{0j} + e_{ij}$$
$$\beta_{0j} = \gamma_{00} + U_{0j}$$

i = individual within a group
j = group

Equations Briefly ☺

Level 1 $$Y_{ij} = \beta_{0j} + e_{ij}$$

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

Mixed $$Y_{ij} = \underbrace{\gamma_{00}}_{\text{fixed}} + \underbrace{U_{0j} + e_{ij}}_{\text{random}}$$

General order for model construction

1. Estimate ICC
2. Random-intercept only model (e.g., means model)
    a. includes random effect of grouping variable
3. Add fixed effects (could be at Level 1, 2, or both)
4. Add Random slopes

Intraclass Correlation Coefficient

- Basic description: Describes the amount of dependence in your data
- Slightly more technical description: proportion of variance explained by differences in between-group means
- An ICC = 0 means there is no dependence in your data (this is the situation where you wouldn't need an mlm)
- An ICC = 1 means there is complete dependence in your data (all observations within a group have identical scores)

```
#### ICC ####
ICC1.lme(avoid_people_lag, ID, data=dat) # 0.49 variance is at the btwn person level
```

Code line 178

Random-Effects ANOVA model

- This model has **no predictors**
- Just a **random intercept** for the mean outcome by the grouping variable
- Offers mean level difference in an outcome across groups

```
# Random-intercept only

```{r}
avoid_RE = lme(avoid_people_lag ~ 1,
               random =~ 1|ID,
               data=dat,
               method="ML",
               na.action = na.exclude)
```
```

Adding Fixed Effects

```r
```{r}
avoid_anx1 = lme(avoid_people_lag ~ worried_trait + worried_state,
                random =~ 1|ID,
                data=dat,
                method="ML", # maximum likelihood estimation
                na.action = "na.omit")
```
```

Model assumptions

- Homoscedasticity
  - Can allow for heteroscedastic error variances by using weights argument in lme()
    weights = varIdent(form =~ 1|variablenamehere)
  - Compare model fits w/ and w/o weights argument e.g., anova(model1, model2)
- Linear
  - Could consider growth curve models
  - Squared terms
- $e_{ij} = \ \sim N(0, \sigma^2)$
- Stationary

Code Walk through

We want to know whether someone's **average** (trait-like) **worry** [level 2] and **momentary** (state-like) **worry** [level 1] predicts **avoiding people**.

Time-series data from 79 participants who were asked to complete 4 surveys for 55 days. On average, participants completed a median of 125 surveys.

I have already preprocessed data for long-format, create trait and state variables, and deal with lags.

What to report

- Steps to build model
- Fixed effects, random effects, confidence intervals
- ICCs
- May consider: Degrees of freedom at each level and F, similar to a priori linear contrasts
- If you use any pseudo-R squared values

Considerations

- Centering predictor variables can facilitate interpretation of intercepts across observations/groups & make it easy to decompose within- and between-group effects
- Not centering predictor variables can help if the scale is intuitively meaningful
- Do not group-mean center an outcome variable (this removes between-group variability)

Resources

Videos

- CenterStat. “How many clusters do I need to fit a multilevel model” at https://www.youtube.com/watch?v=aKXcayBhbMc

Articles

- McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review*, *28*, 295-314.
- Hamaker, E. L., & Muthén, B. (2020). The fixed versus random effects debate and how it relates to centering in multilevel modeling. *Psychological Methods, 25*(3), 365-379. https://doi.org/10.1037/met0000239
- Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: an integrative framework for defining r-squared measures. *Psychological Methods*, *24*(3), 309–338. https://doi.org/10.1037/met0000184