Second-level analysis

Multilevel modeling, Robust regression, thresholding and multiple comparison correction

Suhwan Gim

Student of the Cocoan Lab

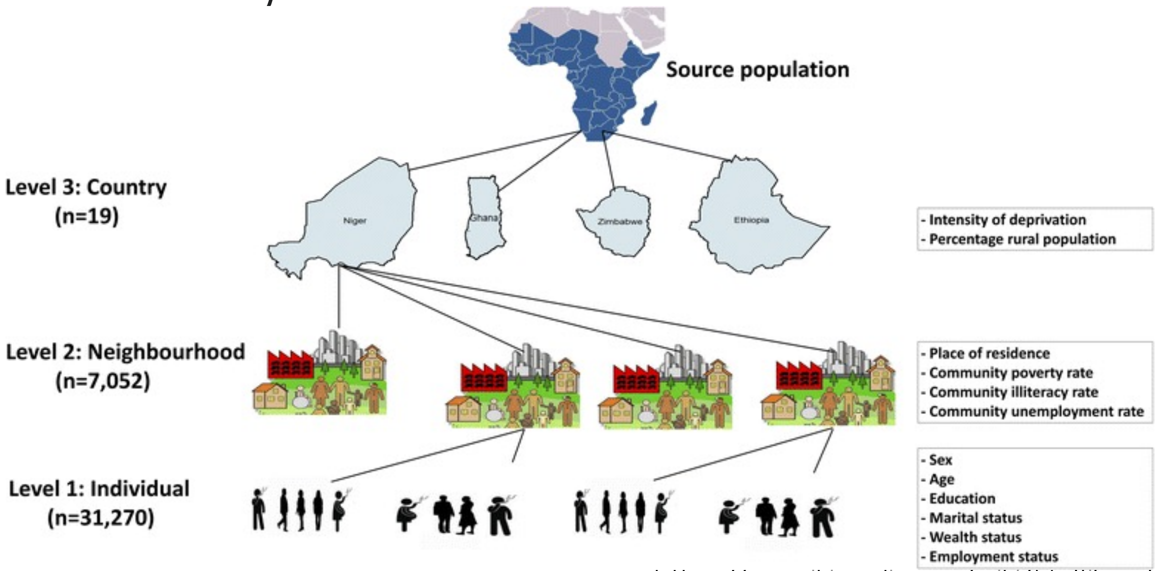
Table of contents

- The second-level analysis (or group-level analysis)
  - Introduction to **multilevel model** (linear-mixed effect model)
  - Why is the second-level analysis needed in fMRI?
  - **Robust Regression Toolbox** in CanlabTools
    - Let's try to perform Robust Regression using the pain datasets (Woo et al. 2014)
- **Thresholding** and **multiple comparison** correction
  - What the family-wise error rate?
  - Why we need to correct multiple comparison ?
  - The methods that can correct a multiple comparison
  - Recommended materials

Multilevel model

(hierarchical linear models, linear mixed-effect model, mixed models, nested data models, random coefficient, random-effects models, or random parameter models)

- What is the multilevel model?
  - Statistical models of parameters that **vary at more than one level**
  - Multilevel models are particularly appropriate for research designs where data for participants **are organized** at more than one level (i.e., nested data)
- The units of analysis are usually individuals (at a lower level) who are nested within contextual/aggregate units (at a higher level)



https://en.wikipedia.org/wiki/Multilevel_model

Figure from Uthman, O. A., Ekström, A. M., & Moradi, T. T. (2016). Influence of socioeconomic position and gender on current cigarette smoking among people living with HIV in sub-Saharan Africa: disentangling context from composition. *BMC public health*, *16*(1), 1-9.

Multilevel model

(hierarchical linear models, linear mixed-effect model, mixed models, nested data models, random coefficient, random-effects models, or random parameter models)

- Benefits of a multilevel approach
  - Correctly account for complex data structures

    : Where single-level model can underestimate each level's variances
  - Incorporate information on group level relationships

    : Where aggregate analysis only examines group level, and individual analysis can ignore groups (or incorrectly treat group effects as individual effects)
  - Link context to the individual

    : How individual relationships are moderated by broader context

https://youtu.be/YLkXP3Edd80

Nested (or hierarchical, multilevel) structure

- If we don't consider the structure of data?

The between-level variability will be ignored

https://towardsdatascience.com/using-mixed-effects-models-for-linear-regression-7b7941d249b
http://people.linguistics.mcgill.ca/~morgan/book/lmem.html

Multilevel model

- How to perform? (Two stage random effects formulation)
  - First, estimate first-level slope and intercept separately

First-level slope　　　　　　firs-level residual

$$Y_{ij} Employtment\ status = \beta_{0j} + \beta_{1j} Education + \beta_{2j} Martial\ Status + \beta_{3j} Wealth\ status + e_{ij}$$

Intercept

  - Second, first-level beta become the dependent variable of next level analysis

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Place\ of\ residence + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + \gamma_{21} Place\ of\ residence + u_{2j}$$
$$\beta_{3j} = \gamma_{30} + \gamma_{31} Place\ of\ residence + u_{3j}$$

Second-level intercept　　　　Second-level variable

Multilevel model

- What is the strength?
  - The between-subject variability will be considered
  - We can explicitly examine the multilevel variable
    - Examine experimental condition effects only by controlling multilevel variable
    - Test how multilevel variable influence the first-level experimental condition effects

Multilevel Model Notation

- Every level-1 b justifies a level-2 equation

$$Y_{ij} = \beta_{0j} + \beta_{1j} SES_{ij} + \beta_{2j} AGE_{ij} + \varepsilon_{ij}$$

- Level 2 equations include random term...

$$\beta_{0j} = \gamma_{00} + u_{0j}$$ Equation for intercept

$$\beta_{1j} = \gamma_{10} + u_{1j}$$ Equation for SES

$$\beta_{2j} = \gamma_{20} + u_{2j}$$ Equation for AGE

Note: If you don't wish to include a random term for any level-2 equation, you don't have to!

Let's try to perform multilevel model
using SEMIC pain calibration data (`glmfit_multilevel` in CanlabTools)