



KATHOLIEKE UNIVERSITEIT  
**LEUVEN**

Tim Reeskens@soc.kuleuven.be

# Multilevel Analysis – An Introduction

## A panacea for nested data?



Institute for European Studies – Vrije Universiteit Brussel  
03 februari 2009



## OLS Regression (1)

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

$$e_i \sim N(0, \sigma^2)$$



## OLS Regression (2)

- Example: What is the effect of age on anti-immigrant sentiments across Europe?
- Dependent variable: ethnocentrism
- Independent variable: age
- $i$  = respondents in European Social Survey

$$ethno_i = \beta_0 + \beta_1 age_i + e_i$$

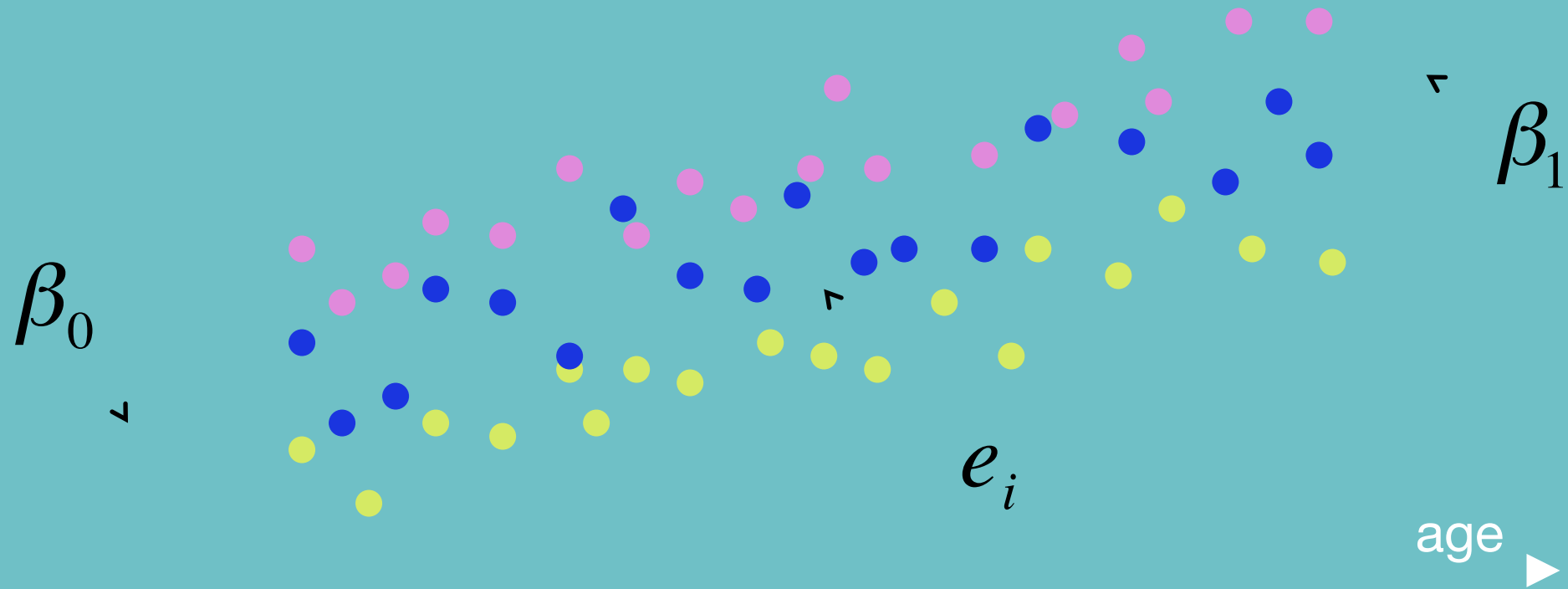
$$e_i \sim N(0, \sigma^2)$$



# OLS Regression (3)

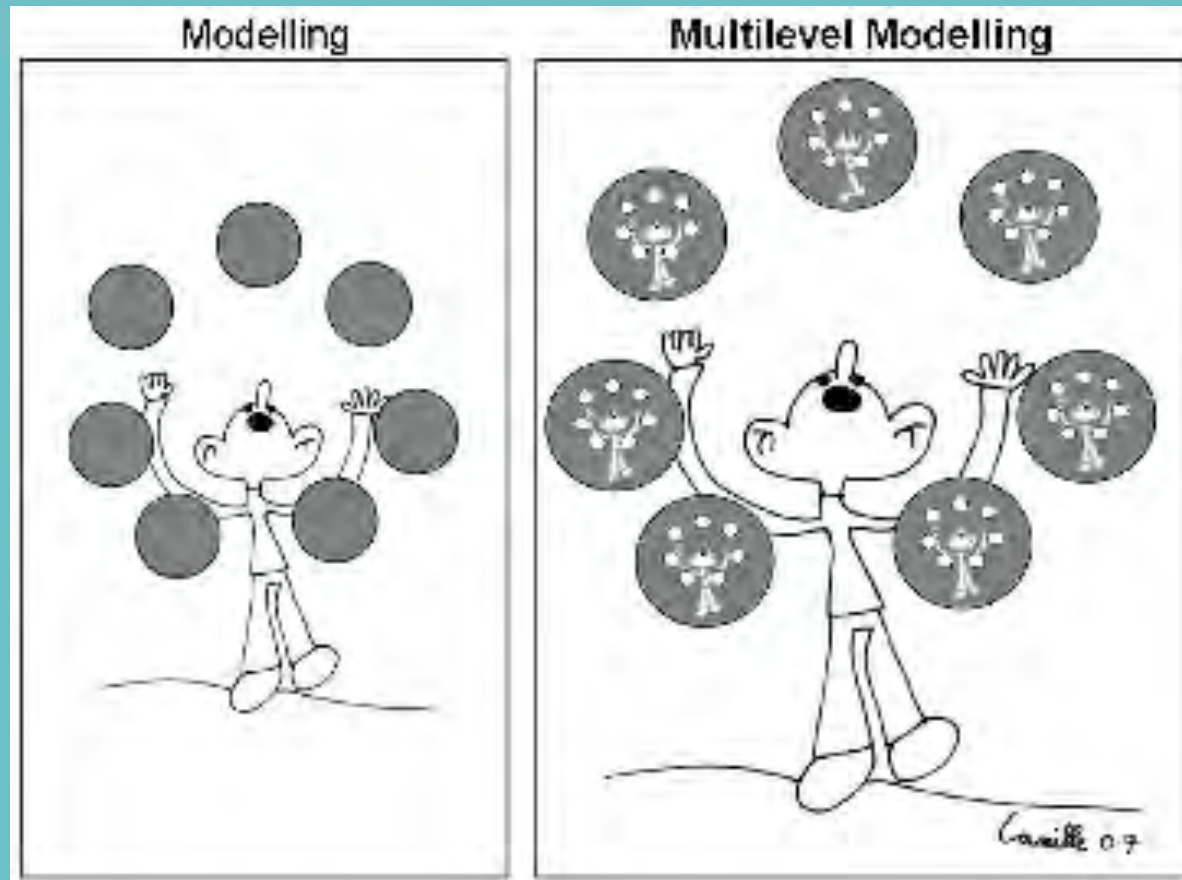
▲ ethnocentrism

Country X  
Country Y  
Country Z





# Nested Data





## Multilevel Regression (1)

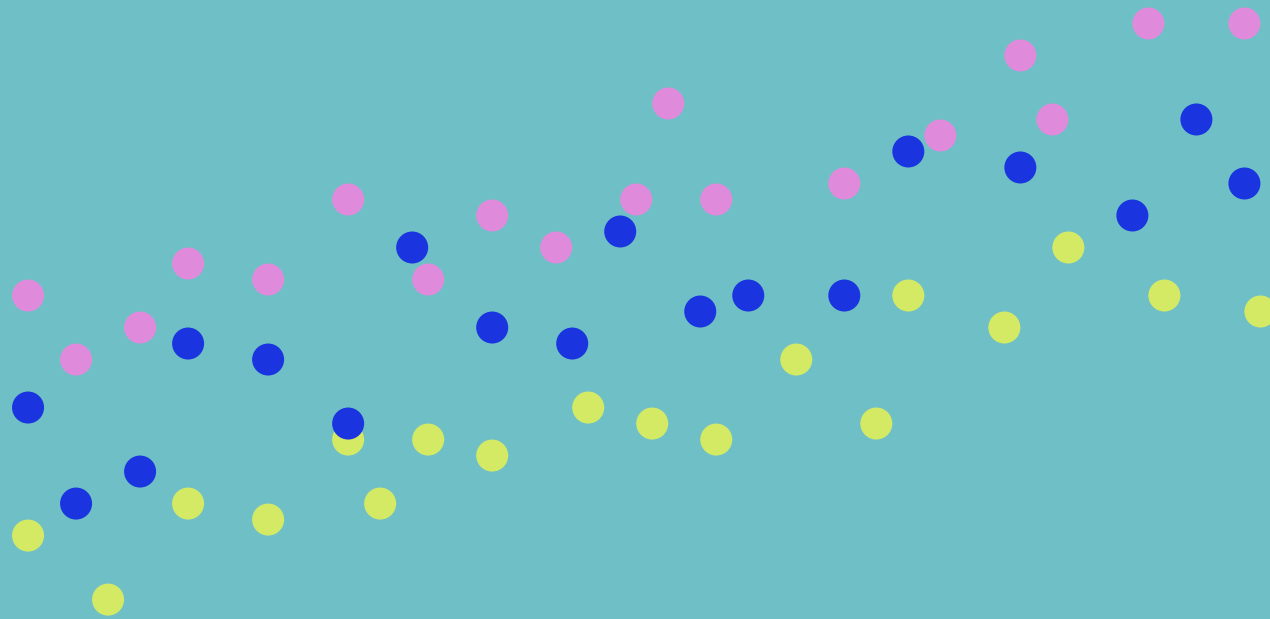
- Key: respondents in nested data structures (like the ESS) aren't independent!
- Ethnocentrism patterns of respondents in country X will be more alike than the general ESS trend line will show!
- Decomposing the variance
  - Between-country variance:  $u_j$
  - Between-respondents variance:  $e_{ij}$



# Multilevel Regression (2)

▲ ethnocentrism

Country X  
Country Y  
Country Z



age ▶



## Random Intercepts

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + u_j + e_{ij}$$

$$e_{ij} \sim N(0, \sigma_e^2)$$

$$u_j \sim N(0, \sigma_u^2)$$



## Intra-Class Correlation

- How much of the variability in ethnocentrism can be attributed by the second level (i.e. countries)?
- Intra-class correlation = 
$$\frac{\sigma_u^2}{(\sigma_u^2 + \sigma_e^2)}$$
- For example,  $\sigma_u^2 = 0.5$  and  $\sigma_e^2 = 4.5$  then  $ICC = 0.5 / (0.5 + 4.5) = 0.1$  or 10 percent that can be attributed to the country level!



## Adding Level-2 Covariates

- I.e. is there less ethnocentrism in countries with a higher level of wealth / in countries with a higher GDP per capita

$$ethno_{ij} = \beta_{0j} + \beta_1 age_{ij} + e_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} GDP_j + u_j$$

- Or  $ethno_{ij} = \gamma_{00} + \beta_1 age_{ij} + \gamma_{01} GDP_j + u_j + e_{ij}$



## Random Slopes Model (1)

- Is there a differential effect of age on ethnocentrism across nations?

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + u_{0j} + e_{ij}$$

$$\beta_{1j} = \beta_1 + u_{1j}$$

$$\text{COV}(u_{0j}, u_{1j}) = \sigma_{u01}$$



## Random Slopes Model (2)

▲ ethnocentrism

$$\sigma_{u01} < 0$$

Country X  
Country Y  
Country Z

age ▶



## Random Slopes Model (3)

▲ ethnocentrism

$$\sigma_{u01} > 0$$

Country X  
Country Y  
Country Z

age ▶



## Random Slopes Model (4)

▲ ethnocentrism

$$\sigma_{u01} = 0$$

Country X  
Country Y  
Country Z

age ▶



## Issues in ML Modeling: Sample

- Statistical analysis: use of (simple) random samples
- Countries in ESS
  - Not randomly drawn
  - “Drawn” out of convenience!
- Is multilevel modeling then appropriate?



## Issues in ML Modeling: Sample Size

- ESS: about 20-25 countries
- Kreft 30/30 rule of thumb
- Power analysis studies on sample size
  - At least 50 groups seem to be preferred
  - Otherwise: parameters are (seriously) biased
- Do we need t-values?
- Substantively instead of statistically significant?



## Conclusion

- Multilevel analysis is a new (complex) technique for a complex social reality
- Applications are very broad:
  - Citizens < n'hoods < municipalities < countries
  - Pupils < classes < schools
  - ...
- Be cautious for underlying assumptions!