Study designs and statistical methods for current observational studies

Multilevel Analysis

12. June. 2024

Kazuhiro Abe

Project Assistant Professor,
Graduate School of Medicine,
The University of Tokyo

Study designs and statistical analysis in current observational studies

1. Difference-in-Differences
2. Instrumental Variable Analysis
3. Regression Discontinuity

Natural experimental methods (Quasi-experimental methods)

4. Panel Data Analysis / Interrupted Time-Series

5. Propensity Score Analysis (matching, weighting, and adjustment using propensity scores)
6. Adjustment(regression), weighting, stratification, and matching

Multilevel Analysis

Analysis for the study population with the hierarchical structure and nested structure

Fig. Example of hierarchical structure

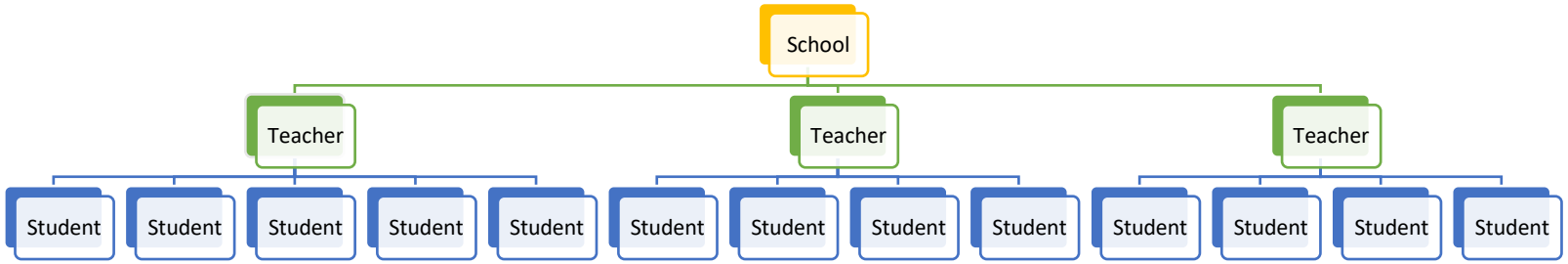


Table. Example of nested structure

id	week	weight
1	1	24
1	2	32
1	3	39
1	4	42.5
1	5	48
1	6	54.5
1	7	61
1	8	65
1	9	72
2	1	22.5
2	2	30.5
2	3	40.5
2	4	45
2	5	51
:	:	:

Student observations within a teacher's classroom may be correlated.

Same as the hierarchical model

Recap: Ordinary Least Square (OLS)

$$y = \beta_0 + \beta_1 x + \mu$$

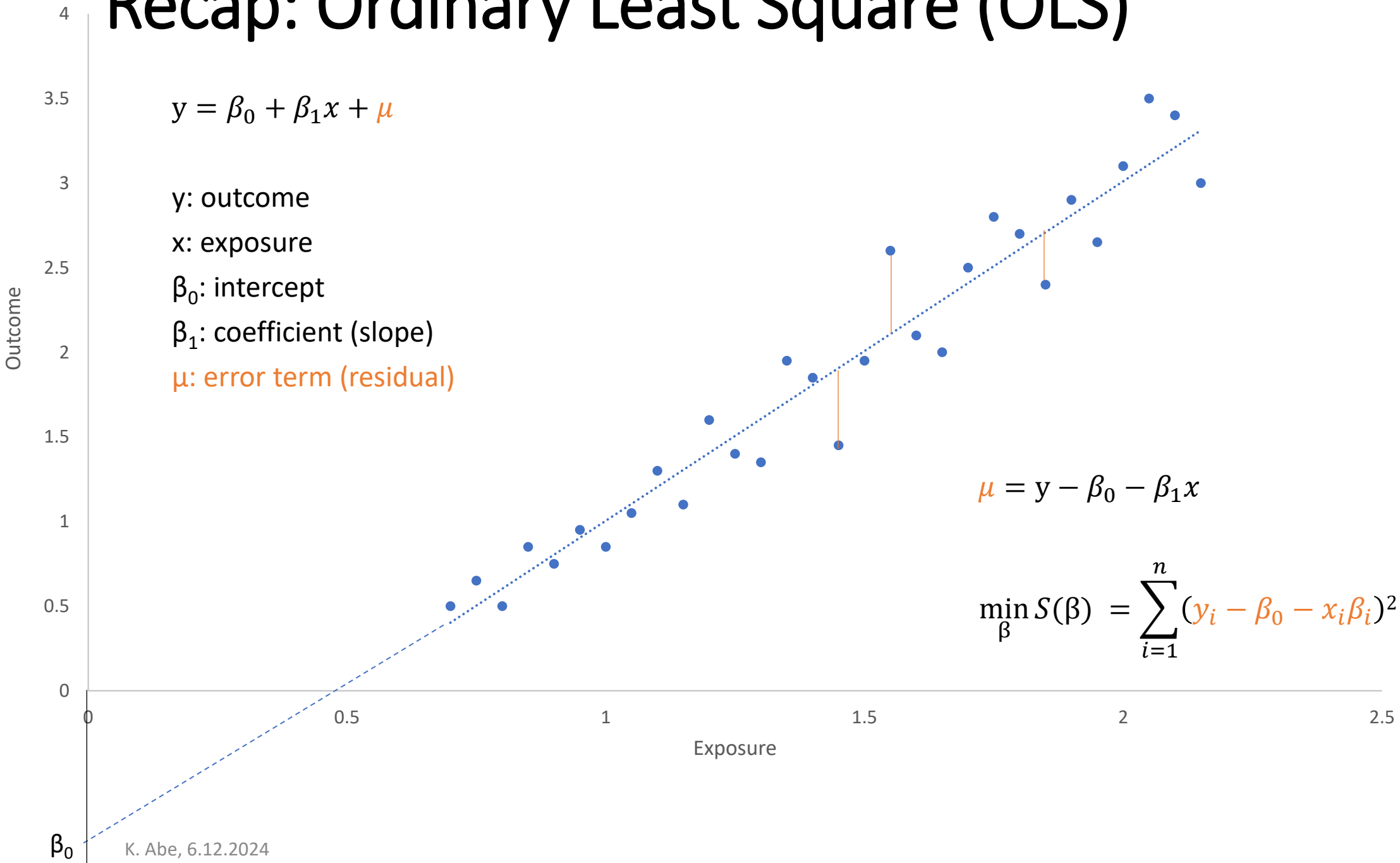
y: outcome

x: exposure

β_0: intercept

β_1: coefficient (slope)

μ: error term (residual)



$$\mu = y - \beta_0 - \beta_1 x$$

$$\min_{\beta} S(\beta) = \sum_{i=1}^{n} (y_i - \beta_0 - x_i \beta_i)^2$$

Gauss-Markov Assumptions

Best Linear Unbiased Estimator (BLUE)
: Efficient, Linear, Unbiased Estimator

Assumption SLR.1 (Linear in Parameters)

In the population model, the dependent variable, y, is related to the independent variable, x, and the error (or disturbance), u, as

$$y = \beta_0 + \beta_1 x + u,$$

where β_0 and β_1 are the population intercept and slope parameters, respectively.

Assumption SLR.2 (Random Sampling)

We have a random sample of size n, $\{(x_i, y_i): i = 1, 2, \ldots, n\}$, following the population model in Assumption SLR.1.

Assumption SLR.3 (Sample Variation in the Explanatory Variable)

The sample outcomes on x, namely, $\{x_i, i = 1, \ldots, n\}$, are not all the same value.

Assumption SLR.4 (Zero Conditional Mean)

The error u has an expected value of zero given any value of the explanatory variable. In other words,

$$E(u|x) = 0.$$

Assumption SLR.5 (Homoskedasticity)

The error u has the same variance given any value of the explanatory variable. In other words,

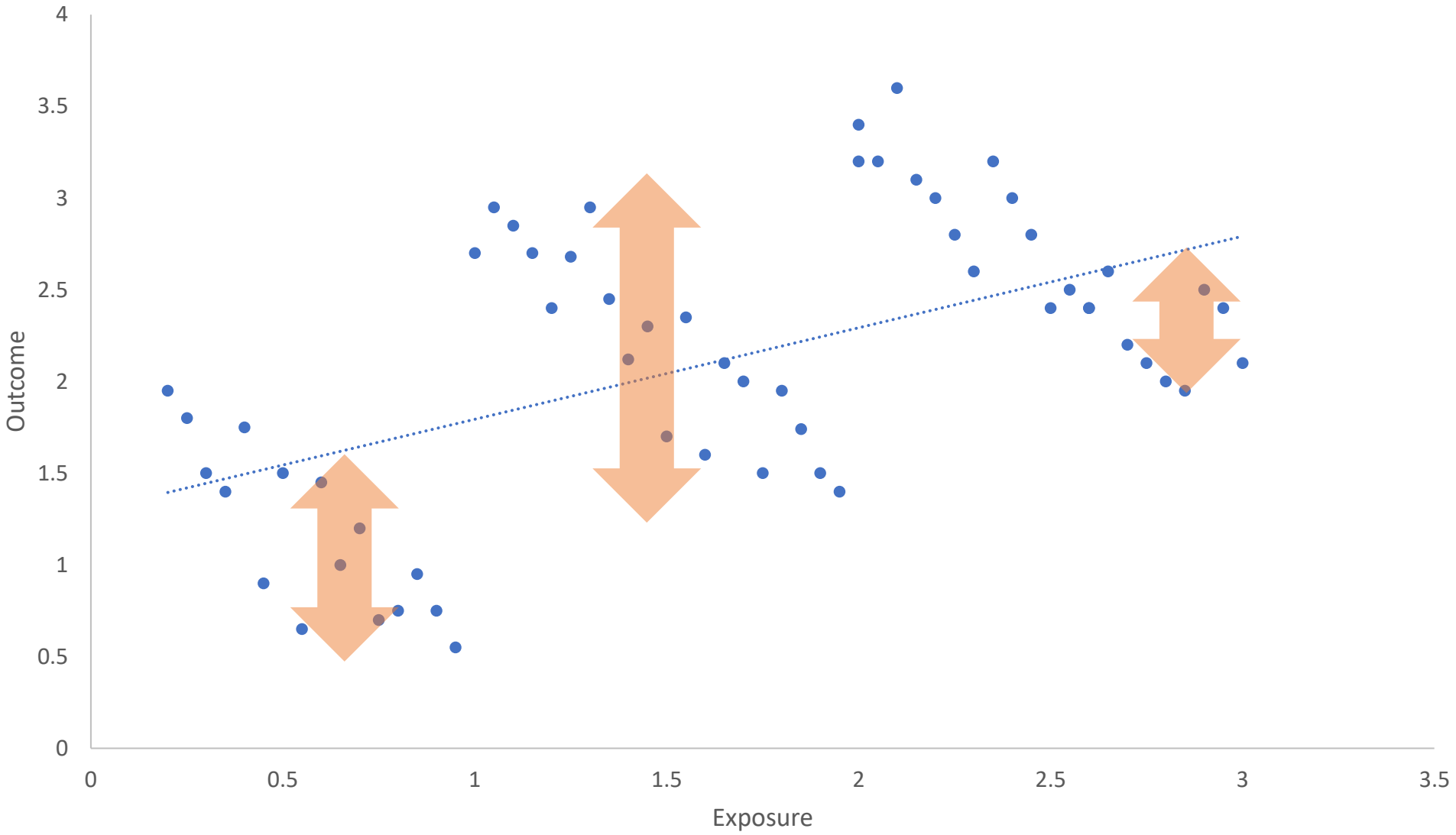
$$Var(u|x) = \sigma^2.$$

Wooldridge J. Introductory Econometrics: A Modern Approach.

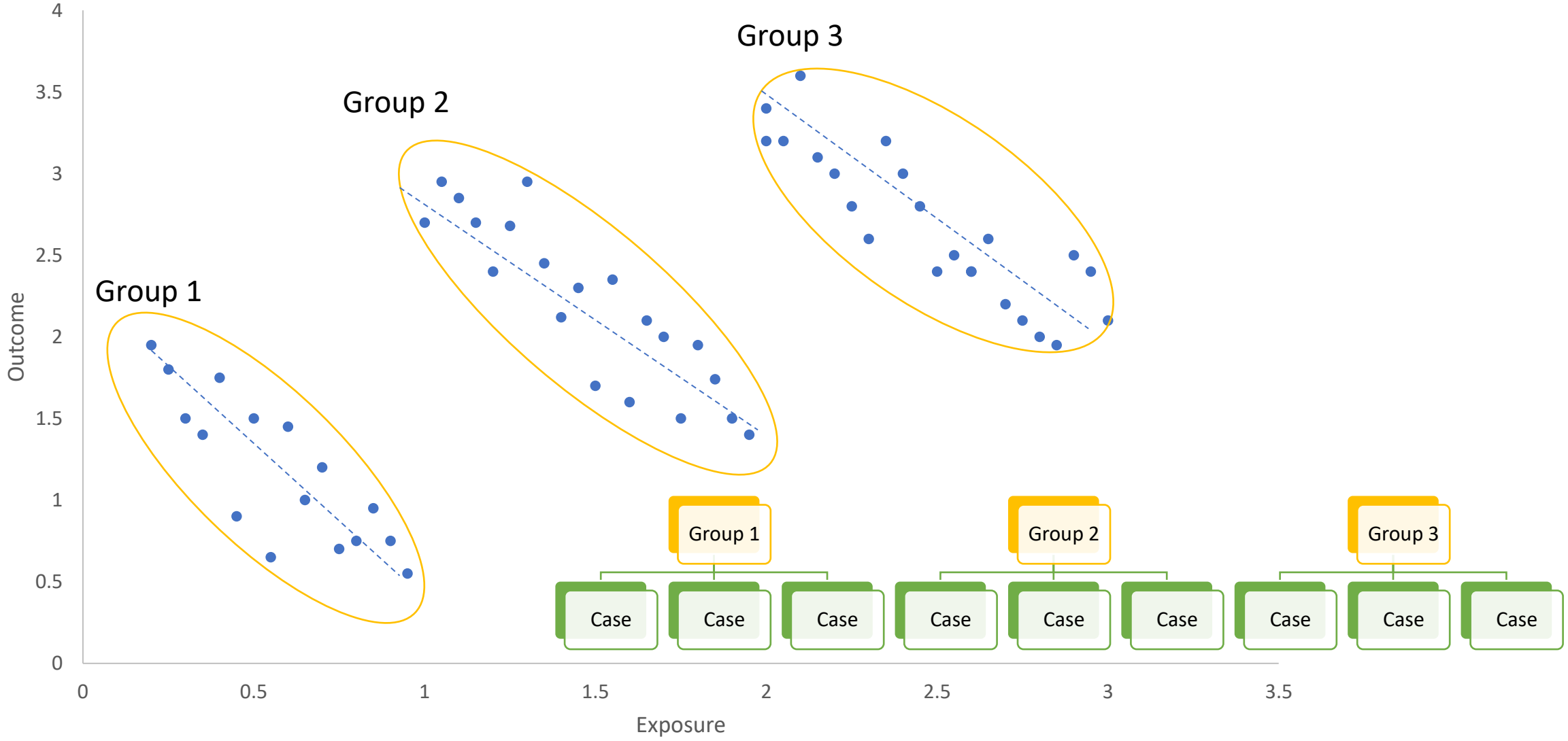
Homoskedasticity

Same variance at any value of exposure

Heteroskedasticity



Multilevel data



Continued only for students