# Multilevel Data Analysis

**AQ801-3-M & Version 1.1**

Statistical Treatment of Clustered Data – Part A

# Contents & Structure

- Aggregation
- Disaggregation
- The intraclass correlation: Within-group and between group variance, Testing for group differences
- Design effects in two-stage samples
- Reliability of aggregated variables

# Recap:

- In MDA context, what is dependence all about?

# Learning Outcomes

- At the end of this topic, You should be able to:
  - **Discuss the problems arise when multilevel structure of the data is ignored.**
  - **Adjust the sample size based on different sampling method**

# Key Terms You Must Be Able To Use

- **If you have mastered this topic, you should be able to use the following terms correctly in your assignments and exams:**

*Aggregation and disaggregation*
*Shift of meaning*
*Ecological fallacy*
*Intraclass correlation*
*Design effect of a two-stage sample*

## Overview of the Topic

- The topic starts by considering what will happen if we ignore the multilevel structure of the data.
- Are there any instances where one may proceed with single-level statistical models although the data stem from a multistage sampling design
- What kind of errors – so-called ecological fallacies – may occur when this is done?
- The rest of the topic is devoted to some statistical methods for multilevel data that attempt to uncover the role played by the various levels without fitting a full-blown hierarchical liner model.

- First, we describe the intraclass correlation coefficients, a basic measure for the degree of dependency in clustered observations.
- Second, some simple statistics (mean, standard error of the mean, variance, correlation, reliability of aggregates) are treated for two-stage sampling designs.
- To avoid ecological fallacies it is essential to distinguish within-group from between-group regressions.

# Learning Outcome 1.

- **Discuss the problems arise when multilevel structure of the data is ignored.**

# Aggregation

- A common procedure in social research with two-level data is to aggregate the micro-level data to the macro level.
- The simplest way to do this is to work with the averages for each macro-unit.
- There is nothing wrong with aggregation in cases where the researcher is only interested in macro-level propositions, although it should be borne in mind that the reliability of an aggregated variable depends, inter alia (among other things), on the number of micro-level units in a macro-level unit, and thus will be larger for the larger macro-units than for the smaller ones.

# Aggregation

- In cases where the researcher is interested in macro-micro or micro-level propositions, however, aggregation may result in gross errors.
- The **first** potential error is the “shift of meaning” (Firebaugh, 1978; Huttner, 1981)
- A variable that is aggregated to the macro level refers to the macro-units, not directly to the micro-units.

# Aggregation

- The firm average of an employee rating of working conditions, for example, may be used as an index for “organizational climate”.
- This variable refers to the firm, not directly to the employees.
- The **second** potential error with aggregation is the ecological fallacy (Robinson, 1950)

## Aggregation

- A correlation between macro-level variables cannot be used to make ascertains about micro-level relations.
- The percentage of black inhabitants in a neighborhood could be related to average political views in the neighborhood – for example – the higher the percentage of blacks in a neighborhood, the higher might be the proportion of people with extreme right-wing political views.

# Aggregation

- This, however, does not give us any clue about the micro-level relation between race and political view.
- (the *shift of meaning* plays a role here, too.)
- (The percentage of black inhabitants is a variable that means something for the neighborhood, and this meaning is distinct from the meaning of ethnicity as an individual-level variable)

# Aggregation

- The ecological and other related fallacies are extensively discussed by Alker (1969), Diez-Rouz (1998), and Blakely and Woodward (2000).
- King (1997), originally focusing on deriving correlates of individual voting behavior from aggregate data, describes a method for making inferences – within certain boundaries – at the micro level, when data are only available in aggregate form at the macro level.

# Aggregation

- The **third** potential error is the neglect of the original data structure, especially when some kind of analysis of covariance is to be used.
- Suppose one is interested in assessing between-school differences in pupils achievement after correcting for intake differences, and that Figure 3.1 depicts the true situation in the next slide.

# Aggregation

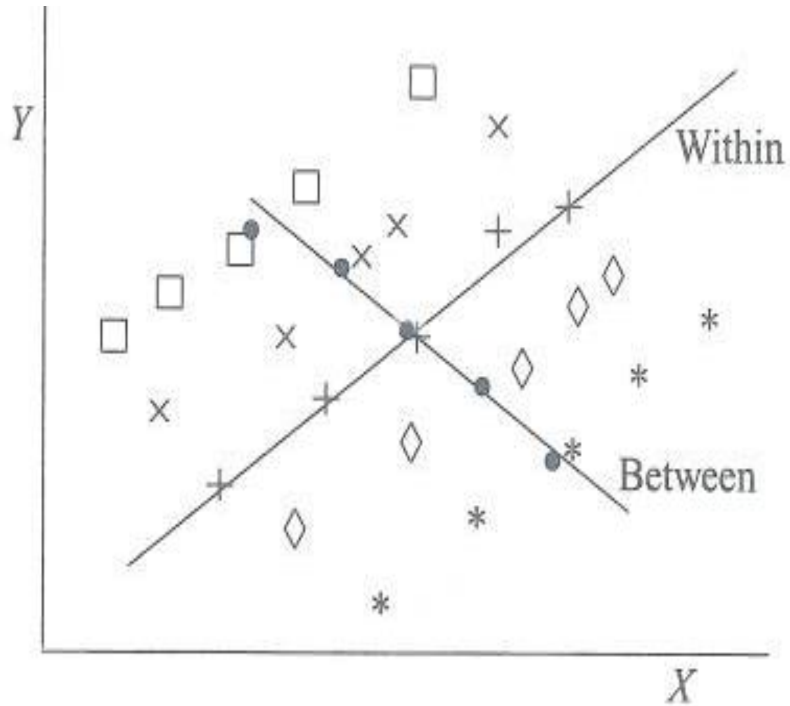


Figure 3.1: Micro-level versus macro-level adjustments.
$(X, Y)$ values for five groups indicated by $*, \Diamond, +, \times, \Box$; group averages by $\bullet$.

# Aggregation

- The figure depicts the situation for five groups, for each of which we have five observations.
- The groups are indicated by the symbols □, X, +, ◊ and *. The five group means are indicated by ●
- Now suppose the question is whether the differences between the groups on the variable Y, after adjusting for differences on the variable X, are substantial.

# Aggregation

- The micro-level approach, which adjusts for the within-group regression of Y on X, will lead to the regression line with positive slope.
- In this picture, the micro-units from the group that have the □ symbols are all above the line, whereas those from the group with the * symbol are all below the regression line.

# Aggregation

- The micro-level regression approach will thus lead us to conclude that the five groups do differ given that an adjustment for X has been made.
- Now suppose we were to aggregate the data, and regress the average $\bar{Y}$ on the average $\bar{X}$.
- The averages are depicted by ●.
- This situation is represented in the graph by the regression line with negative slope.

# Aggregation

- The averages of all groups are almost exactly on the regression line (the observed average $\bar{Y}$ can be almost perfectly predicted from the observed average $\bar{X}$), thus leading us to the conclusion that there are almost no differences between the five groups after adjusting for the average $\bar{X}$.

# Aggregation

- Although the situation depicted in the graph is an idealized example, it clearly shows that working with aggregate data "is dangerous at best, and disastrous at worst" (Aitkin and Longford, 1986, p 42)
- When analyzing multilevel data, without aggregation, the problem described in this section can be dealt with by distinguishing between the within-group and the between-group regressions.

# Aggregation

- The **last objection** to aggregation is that it prevents us from examining potential cross-level interaction effects of a specified micro-level variable with an as yet unspecified macro-level variable.
- Having aggregated the data to the macro level one cannot examine relations such as whether the sentence differential between black and white suspects is different between judges, when allowance is made for differences in seriousness of crimes.

# Aggregation

- Or, to give another example, whether the effect of aptitude on achievement, present in the case of whole-class instruction, is smaller or even absent in the case of ability grouping of pupils within classrooms.

# Disaggregation

- Now, suppose that we treat our data at the micro level.
- There are two situations:
- 1) we also have a measure of a variable at the macro level, next to the measures at the micro level;
- 2) we only have measures of micro-level variables.

# Disaggregation

- In situation (1), disaggregation leads to "the miraculous multiplication of the number of units"
- To illustrate, suppose a researcher is interested in the question of whether older judges hand down more lenient sentences than younger judges.
- A two-stage sample is taken: in the first stage ten judges are sampled, and in the second stage for each judge ten trials are sampled (in total there are 10 x 10 = 100 trials).

## Disaggregation

- One might disaggregate the data to the level of the trials and estimate the relation between the experience of the judge and the length of the sentence, without taking into account that some trials involve the same judge.
- This is like pretending that there are 100 **independent observations**, whereas in actual fact there are only 10 independent observations (the 10 judges)

# Disaggregation

- This shows that disaggregation and treating the data as if they are independent implies that the sample size is dramatically exaggerated.
- For the study of between-group differences, disaggregation often leads to serious risks of committing type 1 errors (asserting on the basis of the observations that there is a difference between older and younger judges whereas in the population of judges there is no such relation).

# Disaggregation

- On the other hand, when studying within-group differences, disaggregation often leads to unnecessarily conservative tests (i.e. type I error probabilities {p-value} that are too low, due to small sample size); this is discussed in detail in Moerbeek et al (2003) and Berkhof and Kampen (2004).

# Disaggregation

- If measures are taken only at the micro level, analyzing the data at the micro level is a correct way to proceed, as long as one takes into account that observations within a macro-unit may be correlated.
- In sampling theory, this phenomenon is known as the design effect for two-stage samples.

# Disaggregation

- If one wants to estimate the average management capability of young managers, while in the first stage a limited number of organizations (say, 10) are selected and within each organization five managers are sampled, one runs the risk of making an error if (as is usually the case) there are systematic differences between organizations.

# Disaggregation

- In general, two-stage sampling leads to the situation that the “effective” sample size that should be used to calculate standard errors is less than the total number of cases, the latter being given here by the 50 managers.
- The formula will be presented in one of the next sections.

## Disaggregation

- Starting with Robinson's (1950) paper on the ecological fallacy, many papers have been written about the possibilities and dangers of cross-level inferences, that is, methods to conclude something about relations between micro-units on the basis of relations between data at the aggregate level, or conclude something about relations between macro-units on the basis of relations between disaggregated data.

## Disaggregation

- Discussions and many references are given by Pedhazur (1982, Chapter 13) Aitkin and Longford (1986), and Diez-Roux (1998)
- Our conclusion is that if the macro-units have any meaningful relation to the phenomenon under study, analyzing only aggregated or only disaggregated data is apt (tend) to lead to misleading and erroneous conclusions.

# Learning Outcome 2

- **Adjust the sample size based on different sampling method**

# Design effects in two-stage samples

- In the design of empirical investigations, the determination of sample sizes is an important decision.
- For two-stage samples, this is more complicated than for simple (one-stage) random samples.

# Design effects in two-stage samples

- This section gives a simple approach to the precision of estimating a population mean, indicating the basic role played by the intraclass correlation.
- Large samples are preferable in order to increase the precision of parameter estimates, that is, to obtain tight confidence intervals around the parameter estimates.

## Design effects in two-stage samples

- In a *simple random sample* the standard error of the mean is related to the sample size by the formula
- $standard\ error = \frac{standard\ deviation}{\sqrt{sample\ size}}$ --(3.16)
- This formula can be used to indicate the required sample size (in a simple random sample) if a given standard error is desired.

## Design effects in two-stage samples

- When using two-stage samples, however, the clustering of the data should be taken into account when determining the sample size.
- Let us suppose that all group sizes are equal, $n_j = n$ for all $j$.
- Then the total sample size is $Nn$.
- The ***design effect*** is a number that indicates how much the sample size in the denominator of (3.16) is to be adjusted because of the sampling design used.

## Design effects in two-stage samples

- It is the ratio of the variance of estimation obtained with the given sampling design to the variance of estimation obtained for a simple random sample from the same population, supposing that the total sample size is the same.
- The design effect of a two-stage sample with equal group sizes is given by
- Design effect = $1 + (n - 1)\rho_I$ --- (3.17)

# Design effects in two-stage samples

- Suppose, for example, we were studying the satisfaction of patients with the treatments provided by their doctors.
- Furthermore, let us assume that $\rho_I$ of 0.30.
- The researchers used a two-stage sample, since that is far cheaper than selecting patients simply at random.

## Design effects in two-stage samples

- They first randomly selected 100 doctors, from each chosen doctor selected five patients at random, and then interviewed each of these.
- In this case the design effect is $1+(5-1) \times 0.30 = 2.2$.

# Design effects in two-stage samples

- The effective sample size, that is, equivalent total sample size that we should use in estimating the standard error (in simple random sampling), is equal to
- $N_{effective} = \frac{N_n}{design\ effect} --- (3.18)$
- in which *N* is the number of selected macro-units.
- For our example we find $N_{effective} = \frac{100x5}{2.20} = 227$

# Design effects in two-stage samples

- So the two-stage sample with a total of 500 patients here is equivalent to a simple random sample of 227 patients.

# Summary of Main Teaching Points

- Aggregation is a procedure in social research with two-level data where it aggregate the micro-level data to the macro level.
- The potential errors are “shift of meaning” ecological fallacy, neglect of the original data structure and is that it prevents us from examining potential cross-level interaction effects.
- Disaggregation leads to “the miraculous multiplication of the number of units

- The degree of resemblance (similarity) between micro-units belonging to the same macro-unit can be expressed by the ***intraclass correlation coefficient***.
- The ***design effect*** is a number that indicates how much the sample size is to be adjusted because of the sampling design used.

- The effective sample size is equivalent to total sample size that we should use in estimating the standard error in simple random sampling.

## Question and Answer Session

# Q & A

# What we will cover next

- **Statistical Treatment of Clustered Data Part B**