

Multilevel Data Analysis

AQ801-3-M & Version 1.1

Multilevel Theories, Multi-Stage Sampling and Multilevel Models

Contents & Structure

- Dependence as a nuisance
- Dependence as an interesting phenomenon
- Macro-level, micro-level, and cross-level relations

Recap:

- Under which situation, multilevel data analysis method should be applied?

Learning Outcomes

- **At the end of this topic, You should be able to:**
 - **Discuss the concept of dependency**
 - **Differentiate macro-level, micro-level and cross-level relations.**

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:**

Multilevel data structures

Sampling design

Levels

Units

Dependence as a nuisance

Dependence as an interesting phenomena

Multilevel propositions

Introduction

- Phenomena and data sets in the social sciences often have a multilevel structure.
- This may be reflected in the design of data collection: simple random sampling is often not a very cost-efficient strategy, and multistage samples may be more efficient instead.
- This topic is concerned with the reasons why it is important to take account of the clustering of the data, also called their multilevel structure, in the data analysis phase.

- in this case the dependency of observations within clusters is of focal interest, because it reflects the fact that clusters differ in certain respects.
- In either case, the use of single-level statistical models is no longer valid.

Learning outcome 1

- Discuss the concept of dependency

Dependence as a nuisance

- Textbooks on statistics tell us that observations should be sampled independently of each other as standard.
- Thus the standard sampling design on which statistical models are based is simple random sampling: the result of one selection is independent of the result of any other selection, and all single units in the population have the same chances of being selected into the sample.

- Textbooks on sampling, however, make it clear that there are more cost-efficient sampling designs
- One of those cost-efficient sampling designs is the **multistage sample**: the population of interest consists of subpopulations, also called clusters, and selection take place via those subpopulations

- If there is only one subpopulation level, the design is a two-stage sample.
- Pupils, for instance, are grouped in schools, so the population of pupils consists of subpopulations of schools that contain pupils.
- Other examples are: families in neighborhoods, teeth in jawbones, animals in litters, employees in firms, and children in families.

- In a random two-stage sample, a random sample of the primary units (schools, neighborhoods, jawbones, litters, firms, families) is taken in the first stage, and then the secondary units (pupils, families, teeth, animals, employees, children) are sampled at random from the selected primary units in the second stage.
- A common mistake in research is to ignore the fact that the sampling scheme was a two-stage one, and to pretend that the secondary units were selected *independently*.

- The mistake in this case would be that the researcher is overlooking the fact that the secondary units were not sampled independently of each other; having selected a primary unit (e.g. a school) increases the chances of selection of secondary units (e.g. pupils) from that primary unit.

- In other words, the multistage sampling design leads to dependent observations, and failing to deal with this properly in the statistical analysis may lead to erroneous inferences.

- The multistage sampling design can be depicted graphically as in the next slide.
- This shows a population that consists of 10 subpopulations, each containing 10 micro-units.
- A sample of 25% is taken by randomly selecting 5 out of 10 subpopulations and within these – again at random of course – 5 out of 10 micro-units

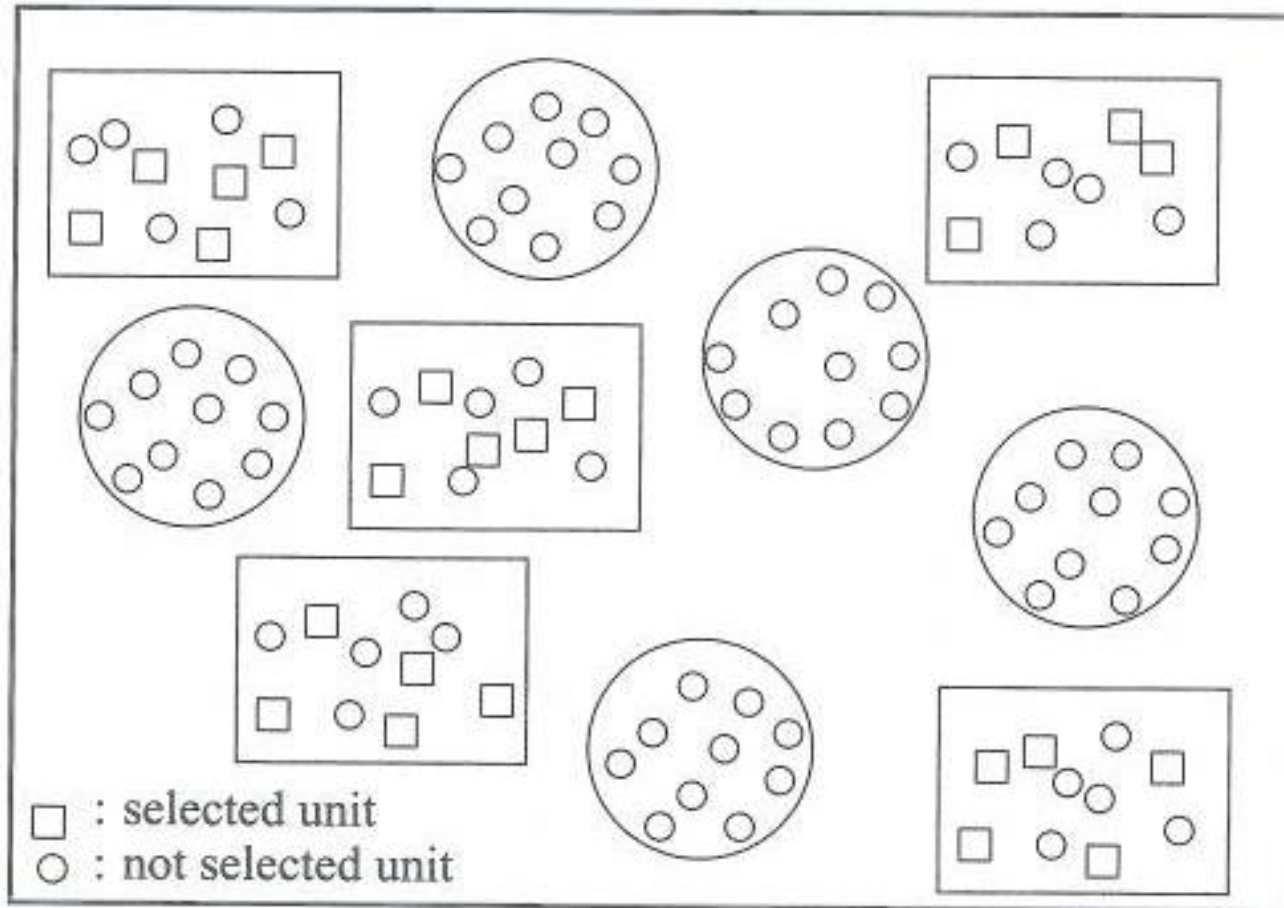


Figure 2.1: Multistage sample.

- Multistage samples are preferred in practice, because the costs of interviewing or testing persons are reduced enormously if these persons are geographically or organizationally grouped.
- Such sample designs correspond to the organization of the social world.

- It is cheaper to travel to 100 neighbourhoods and interview 10 persons per neighbourhood on their political preferences than to travel to 1000 neighbourhoods and interview one person per neighbourhood.

Dependence as an interesting phenomenon

- The previous section implies that, if we want to make inferences on, for example, the earnings of employees in the for-profit sector, it is cost-efficient to use a multistage sampling design in which employees are selected via the firms in which they work.
- A common feature in social research, however, is that in many cases we wish to make inferences on the firms as well as on the employees.

- Questions that we seek to answer may be:
 - Do employees in multinationals earn more than employees in other firms?
 - Is there a relation between the performance of pupils and the experience of their teacher?

- In this case a variable is defined at the primary unit level (firms, teachers) as well as at the secondary unit level (employees, pupils).
- Henceforth we will refer to primary units as *macro-level* units (or macro-units for short) and to secondary units as *micro-level units* (or micro-units for short).
- The micro level is called the lower level and the macro level is called the higher level.

Table 2.1: Summary of terms to describe units at either level in the two-level case.

macro-level units	micro-level units
macro-units	micro-units
primary units	secondary units
clusters	elementary units
level-two units	level-one units

- Examples of macro-units and the micro-units nested within them are presented in Table 2.2 in the next slide.
- It is important to note that what is defined as a macro-unit or a micro-unit depends on the theory at hand.

Table 2.2: Some examples of units at the macro and micro level.

Macro level	Micro level
schools	teachers
classes	pupils
neighbourhoods	families
firms	employees
jawbones	teeth
families	children
litters	animals
doctors	patients
subjects	measurements
interviewers	respondents
judges	suspects

- Teachers are nested within schools, if we study organizational effects on teacher burn-out then teachers are the micro-units and schools the macro-units.
- But when studying teacher effects on student achievement, teachers are the macro-units and students the micro-units.

- The more the achievement levels of pupils within a school are alike (as compared to pupils from other schools), the more likely it is that causes of the achievement have to do with the organizational unit (in this case, the school)
- Absence of dependency in this case implies absence of institutional effects on individual performance.

Table 2.2: Some examples of units at the macro and micro level.

Macro level	Micro level
schools	teachers
classes	pupils
neighbourhoods	families
firms	employees
jawbones	teeth
families	children
litters	animals
doctors	patients
subjects	measurements
interviewers	respondents
judges	suspects

Learning outcome 2

- Differentiate macro-level, micro-level and cross-level relations.

Macro-level, micro-level, and cross-level relations

- Tacq (1986) distinguished between three kinds of propositions: on micro-units (e.g. employees have on average 4 effective working hours per day; boys lag behind girls in reading comprehension), on macro-units (eg. Schools have on average a budget of \$20,000 to spend on resources), or on macro-micro relations (e.g. if firms have a salary bonus system, the productivity of employees will be)

- The use of such a sampling design is quite obvious if we are interested in macro-micro relations, less obvious (but often necessary from a cost-effectiveness point of view) if micro-level propositions are our primary concern, and hardly obvious at all (but sometimes still applicable) if macro-level propositions are what we are focusing on.

- These three instances will be discussed below.
- To facilitate comprehension, following Tacq (1986) we use figures with the following conventions: a dotted line indicates that there are two levels: below the line is the micro-level, above the line is the macro level; macro level variables are denoted by capitals; micro-level variables are denoted by lower-case letters, and arrows denote presumed causal relations.

Multilevel Propositions

- Multilevel propositions can be represented as in Figure 2.2 as below:

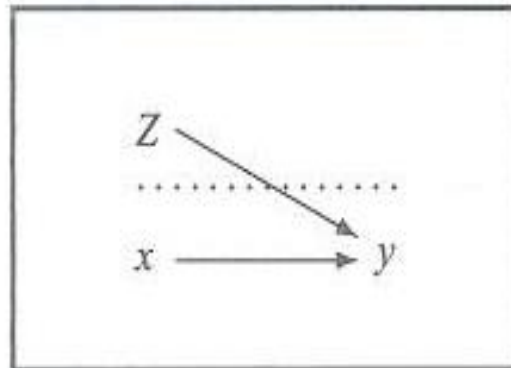


Figure 2.2: The structure of a multilevel proposition.

- In this example we are interested in the effect of the macro-level variable Z (e.g. teacher efficacy) on the micro-level variable y (e.g. pupil motivation), controlling for the micro-level variable x (e.g. pupil aptitude).

Micro-level propositions

- Micro-level propositions are of the form indicated in Figure 2.3 as below:

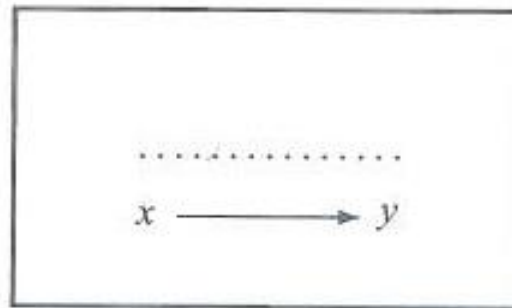


Figure 2.3: The structure of a micro-level proposition.

- In this case the line indicates that there is a macro level which is not referred to in the hypothesis that is put to the test, but which is used in the sampling design in the first stage.
- In assessing the strength of the relation between occupational status and income, for instance, respondents may have been selected for face-to-face interviews by zip-code area.
- This then may cause dependency (as a nuisance) in the data.

Macro-level propositions

- Macro-level propositions are of the form of Figure 2.4 as below

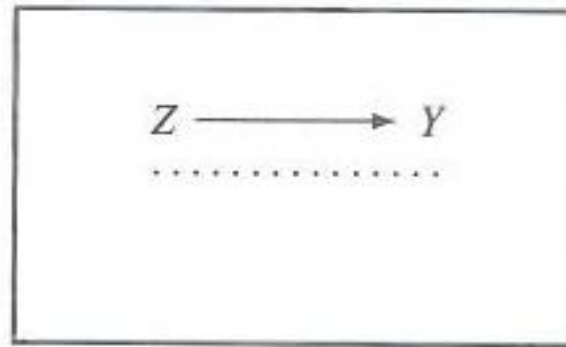


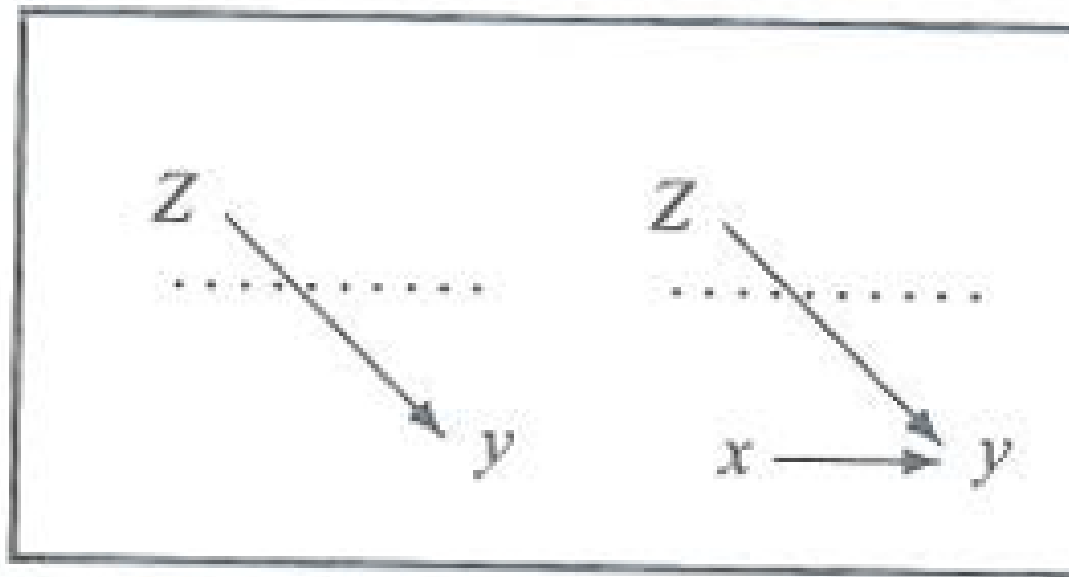
Figure 2.4: The structure of a macro-level proposition.

- The line separating the macro level from the micro level seems superfluous here.
- When investigating the relation between the long-range strategic planning policy of firms and their profits, there is no multilevel situation, and a simple random sample may have been taken.
- When either or both variables are not directly observable, however, and have to be measured at the micro level (eg. organizational climate measured as the average satisfaction of employees), then a two-stage sample is needed nevertheless.

- This is the case a fortiori (for a stronger reason) for variables defined as aggregates of micro-level variables (eg., the crime rate in a neighborhood).

Macro-micro relations.

- The most common situation in social research is that macro-level variables are supposed to have a relation with micro-level variables.
- There are three obvious instances of macro-to-micro relations, all of which are typical examples of the multilevel situation (see Figure 2.5 in the next slide).



- The first case is the macro-to-micro proposition.
- The more explicit the religious norms in social networks, for example, the more conservative the views that individuals have on contraception.
- The second proposition is a special case of this.

- It refers to the case where there is a relation between Z and y , given that the effect of x on y is taken into account.
- The example given in few slides before may be modified to: “for individuals of a given educational level”.

- To put it another way, the relation between Z and y is dependent on X .
- The effect of aptitude on achievement, for instance, may be small in case of ability grouping of pupils within classrooms but large in ungrouped classrooms.
- Next to these three situations there is the so-called emergent, or micro-macro, proposition (Figure 2.6).

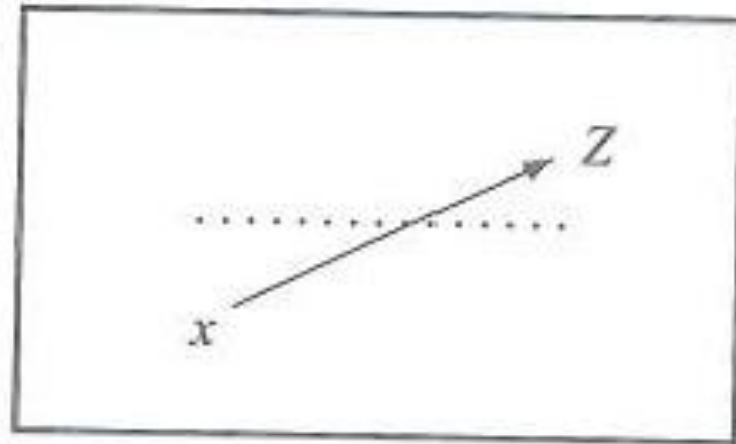


Figure 2.6: The structure of a micro–macro proposition.

- In this case, a micro-level variable x affects a macro-level variable Z (student achievement may affect teachers' experience of stress).
- It is of course possible to form combinations of the various examples given.
- Figure 2.7 contains a causal chain that explains through which micro-variables there is an association between the macro-level variables W and Z (cf. Coleman, 1990).

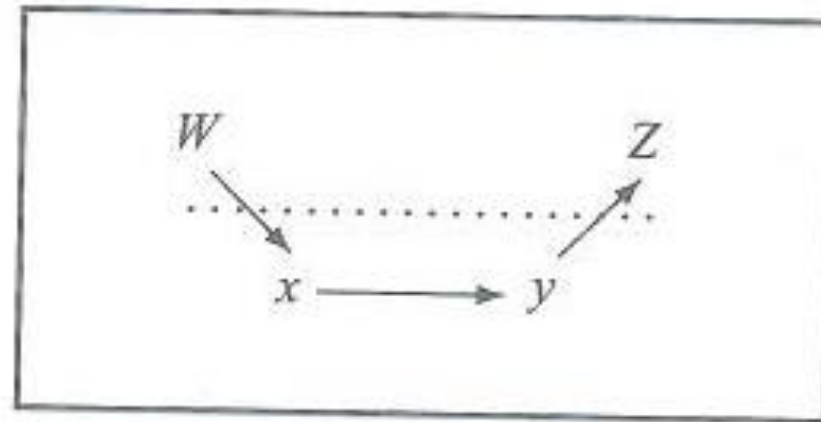


Figure 2.7: A causal macro-micro-micro-macro chain.

- As an example of this chain, we may be interested in why the qualities of a football coach affect his social prestige.
- The reason is that good coaches are capable of motivating their players, thus leading the players to good performance, thus to winning games, and this of course leads to more social prestige for the coach.

- Another instance of a complex multilevel proposition is the contextual effects proposition.
- For example, low social-economic status pupils achieve less in classrooms with a low average aptitude.
- This is also a cross-level interaction effect, but the macro-level variable, average aptitude in the classroom, is now an aggregate of a micro-level variable.

Summary of Main Teaching Points

- Statistical theory assumes that observations should be sampled *independently*. However under two-stage sampling, the observations in each cluster depend on each other. This is called *Dependence as nuisance*.
- Dependence as an interesting phenomenon as one may detect the relationship between group effect and cross level relationship.

Question and Answer Session

Q & A

What we will cover next

- **Statistical Treatment of Clustered Data**