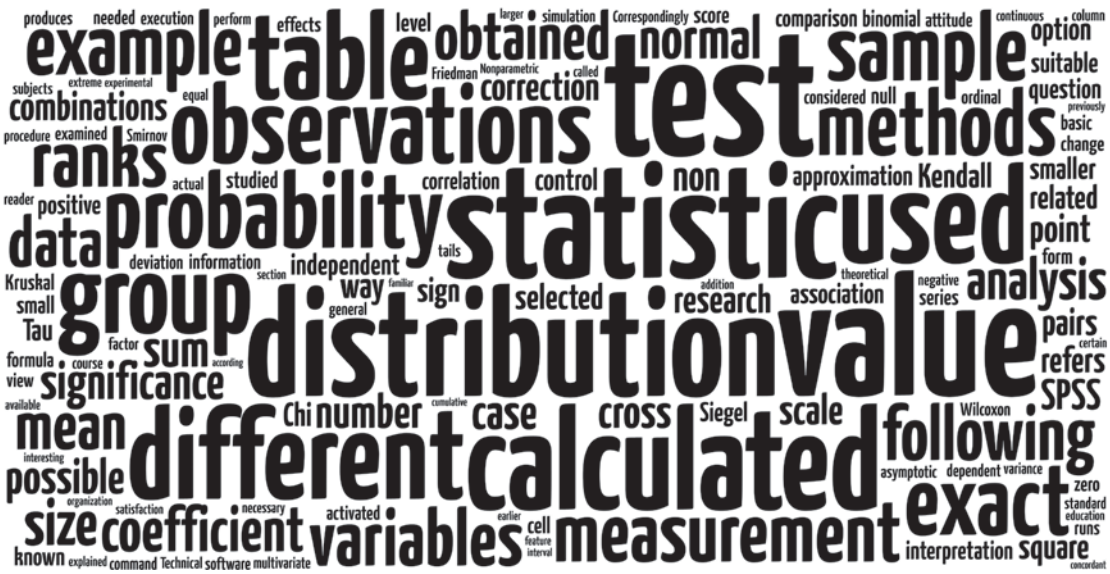
Section I

Basics of Nonparametric Statistics



Brief Glossary for a Beginner Researcher

!	Read: Factorial. For 5, the factorial is $5! = 1 \times 2 \times 3 \times 4 \times 5 = 120$, see Chapter 3, Section 3.1.
$\binom{n}{k}$	The number of combinations. Read: "*n* choose *k*." The number of combinations is computed by $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. See Chapter 3, Section 3.1 therein.
$\sum_{i=1}^{k}$	Read: Sum when *i* changes from 1 to *k*. A technical expression for summing up the numbers. *i* can refer to cases, sections, columns, rows, or other factor that is not wanted to mark open $1+2+\cdots+i+(i+1)+(i+2)+\cdots+k$. Instead of *i*, there can be another index like *j*, and instead of *k*, there can be another sign like *n*, *r*, or *c*.
Additive	Summable. For example, probability is in some cases additive; the probability of getting heads or tails is the probability of heads + the probability of tails = 0.5 + 0.5 = 1.
Approximation	The value of a statistic approaches the true value; however, it does not reach it. See also "Asymptotic."
A posteriori	From Latin: "Afterward." Knowledge that is gained empirically. Comparing groups (e.g., in Kruskal–Wallis test) is based on empirical observations of what is the mean of group observation ranks in the sample—paired comparison is thus a post hoc method that happens a posteriori. See "A priori" and "Post hoc."
A priori	From Latin: "Beforehand." Knowledge based on reasoning. For example, Jonckheere–Terpstra test is based on an experimental design in which it is known that a bigger effect should induce a bigger response (i.e., e.g., more medicine induces a greater relief). In this test, the paired comparison is a post hoc method that happens a priori. See "A posteriori" and "Post hoc."
Asymptotic	Approximates to a certain distribution or value, but never reaches it. "A test value reaches the Normal distribution asymptotically" means that test value is approximately Normal when the sample size increases.
Binomial distribution	From Greek: *bi* = two and *nomos* = area, province. If a variable can have only two values (like girl/boy, succeeded/failed, head/tail) its occurrence follows a binomial distribution in a random case. See section 3.1.

Central limit theorem (CLT) It can be presented nonmathematically in the following way: "the sum of random variables begins to approximate Normal distribution when the number of random variables becomes high enough." An important argument for the approximation of test statistics based on Normal distribution.

Continuity correction A correction factor that is used to combine continuous and discrete distributions (like binomial distribution) that brings discrete distribution closer to a continuous one.

Conservative test A test is conservative when it rejects null hypothesis only when it is certain. For example, in the pair-wise comparisons of Kruskal–Wallis or Friedman tests, a conservative method is used in fixing the *p*-value. This happens by correcting *p*-value with two times the number of all single comparisons, that is, with the factor $k(k-1)/2$ when it would be enough to correct them by a smaller correction.

Expected value It is often marked with the symbol *E*. A theoretical value of a cell that is computed from marginal values of a table. *E* is the value which is expected in a cell if the null hypothesis is valid. In statistical testing, the observed frequency is compared with the expected one. See "Observed frequency."

Extreme More unlikely, a rarer. A quality, for example, in cross-tabulation to produce a rarer—and therefore more unlikely—value for the test statistic. When these more extreme tables occur, the test statistics get bigger values and, therefore, reject the null hypothesis more certainly than a more usual or normal result.

Independent sample A sample with independent observations and groups that are formed from these observations. In its simplest form, there can be two groups of independent observations. These groups can be sexes, organizations, etc. See "Paired sample."

Multivariate method Methods that can be used when analyzing multiple variables simultaneously.

Multinomial distribution Greek: *multi* = many and *nomos* = area, province. If the variable gets more than two values (like, e.g., when throwing a dice), its occurrence follows a multinomial distribution. See Chapter 3, Section 3.2 therein; compare also "Binomial distribution."

Notation A way to write symbols. Different symbols, that is, different notations for the same formulae are used in different books.

Observed frequency Also: Observed count. It is often marked with the symbol O. Amount of observations in cells of a cross-table. See "Expected value."

Observed *p*-value Marked p or p_{obs}. Probability of the value of a test statistic in its own distribution, also significance. It is often more meaningful to express observed probability (e.g., p 0.023) than the traditional significance level at three levels (e.g., $p < 0.05$). See "Significance level."

Paired sample A sample type where several measurements are made of the same objects (or somehow unified objects). In its simplest form, it can be used to examine the same case twice in a before–after type measurement. The paired sample can also be formed by, for example, twins, spouses, or otherwise related units. See "Independent sample."

Parameter From Greek: "side measure." In many analyses, various statistics are computed or estimated. These statistics are called parameters. Parameters are the moving particles of the model, that is, the statistics or coefficients with different values depending on the data, the amount of variables, model, and selected variables. Parameters are estimated, and statistics and coefficients are computed.

Post hoc From Latin: "Afterward." Assessment that happens afterward. In analysis of variance and its nonparametric counterparts, the test statistic tells only the main result; after the testing—*post hoc*—the groups are compared more precisely.

Power of a test The power to reject the null hypothesis when it is wrong and should be rejected.

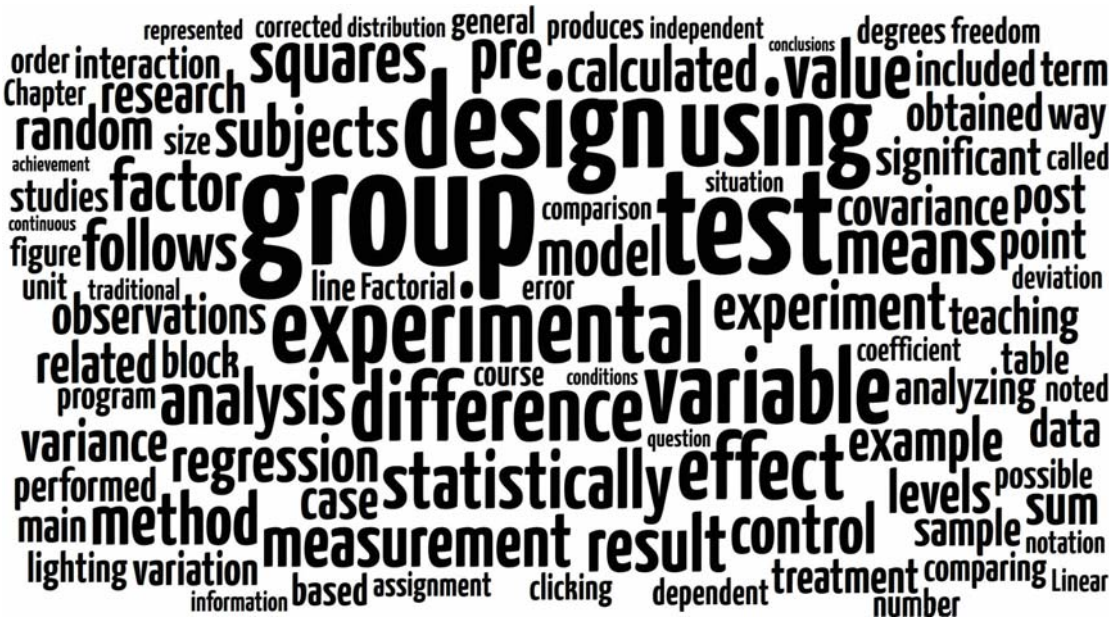
Robustness A method is robust if it produces the approximately correct result, even though the assumptions are not met. An assumption can be, for example, that the observations are Normally distributed. A robust method can compute approximately the correct result, even though the assumption of Normality is not true for the variable.

Significance level It is marked with α for example, $\alpha = 0.05$. Usually, the preset highest level of probability with which the null hypothesis can be rejected. Probability for risk of rejecting null hypothesis if it was correct. Traditionally, there are three significance or risk levels: 5% ($\alpha = 0.05$), 1% ($\alpha = 0.01$), and 0.1% ($\alpha = 0.001$). See also "Observed *p*-value."

Syntax Software language. Software executes commands made with certain syntax. For example, in SPSS software, the syntax is hidden behind the icons, but in reality, it listens to syntaxes that are chosen by clicking an icon.

Section II

Basics of Experimental Research in Human Sciences



CHAPTER 8

Introduction to Experimental Research

Goals of the Chapter

1. To orientate the concepts and symbols used in the experimental research.
2. To be able to discriminate between different types of experimental- and non-experimental designs.
3. To be able to recognize central threats to validity and to know techniques to reduce those.
4. To be able to use techniques for random assignments.

A major part of the so-called scientific research has throughout its history been aimed at mastering the interesting challenges that questions of cause and effect have posed to statistical inference. As one may remember,[1] the English philosopher David Hume already presented in the 18th century the principles that are still highly significant for present-day science. According to him, one may only speak about *cause and effect relations* when, first, two events relate to each other. To use a modern term, one can say that there is *covariation* between two variables, which means that a change in the presumed cause factor brings about a change in the presumed effect factor (Trochim & Landis, 1982). Second, a cause and effect relation requires that the factor, presumed to function as a cause, *temporally precedes* the factor, presumed to function as an effect. Third, a continuous relation must exist between the variables, so that there are *no plausible alternative explanations* for the observed results. Of these factors, the last one is essential when validity is concerned as Cook and Campbell (1979; see also Campbell & Stanley, 1963) have described very well in their book addressing

Cause and effect relation when
- two events covariate with each other
- two factors follow each other temporally
- no third explanation to be reckoned with available

[1] See Section VI of Volume 1.

quasi-experimental designs. The internal validity of a piece of research has, after all, traditionally been defined by asking whether "one measures what was supposed to be measured"; when the results of the research can be explained by a factor, the existence of which was not taken into account when planning the research, one has not measured "what was supposed to be measured" and, thus, the research cannot be valid. The threats to the validity of research are addressed in more detail in what follows; each of the research frames is able to answer, certain of the challenges brought about by the threats to validity.

When watertight knowledge concerning cause and effect relations is sought, experimental designs are the best option for the researcher. In many cases, experimental designs have brought about the biggest advancements in Western science during the last 300 years. In a field such as medicine, it is hard to imagine that a new treatment could be introduced if it wasn't tested "scientifically," in practice, using experimental methods. A major part of the modern day- or theoretical matters of physics, such as the temperature, the boiling temperatures, the relations between pressure and volume, the functioning of the motor of a vehicle, and aerodynamics, were originally discovered using the methods of experimental research. Experimental designs have been an ideal of sorts for science and still *knowledge obtained using experimental methods is valued as the most reliable sort of knowledge, especially when the researcher wishes to know about cause and effect* or to draw conclusions that are as watertight as possible from the experimental results. Still, it must be noted that not even the most complete experimental design is able to answer all the challenges. At the same time, it must be noted that, according to a methodological purist, the weakest designs (the so-called "pre-experimental designs") do not offer almost any kind of protection against false or incomplete conclusions.

Experimental research is the ideal method to use when the researcher wishes to find out about cause-effect relations.

In what follows in Chapter 9, the most central experimental designs will introduce. The designs are grouped according to their reliability, so that genuine experimental designs are addressed first, quasi-experimental designs second, and the weakest, so called pre-experimental designs, last. After that, the techniques of the analysis of variance, that is, the methods used to analyze information provided by the experimental setting will be introduced. Before that, some relevant preliminary things are addressed, such as, the central terms and symbols within the experimental approach (Section 8.1), central threats to validity (Section 8.2), and techniques for random assignments (Section 8.3) in this chapter.

It is fair to tell that this section is greatly inspired by William Trochim's (2006b) material; while reading that material, I got some kind of eureka in analyzing experimental designs. Especially illuminating was his treat for the quasi-experimental designs. I warmly suggest the beginner reader to dig into his material also.

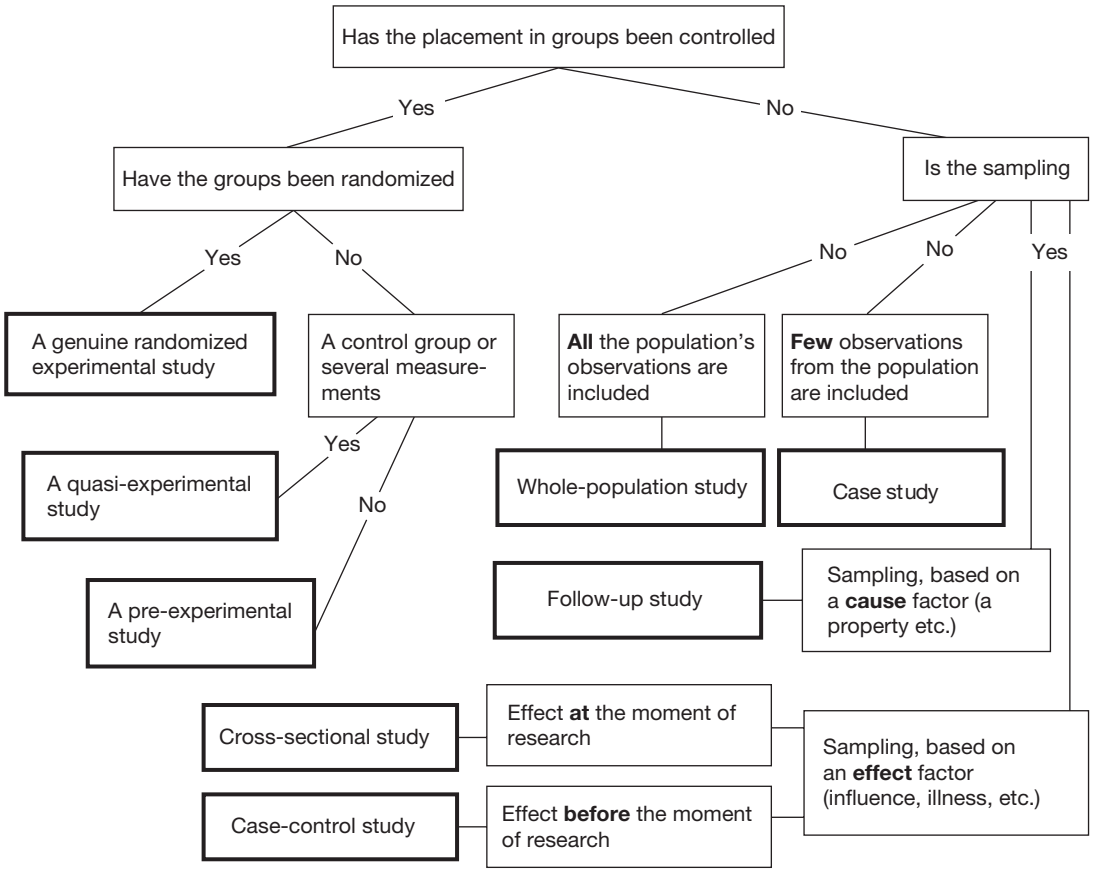


Figure 8.1 Types of Research Grouped According to Sampling Type (Modifying Sarna, 2009, p. 17)

8.1 Central Terms and Symbols

8.1.1 Types of Research

On the next page, different types of research are illustrated (Figure 8.1). From the point of view of the acquisition of information, research studies can be divided into *experimental* studies and observational *survey* studies. Sarna (2009, p. 17) sums up the difference between the two by stating that in experimental studies, the researcher actively divides the units of observation into different groups (typically experimental and control groups). The researcher also manipulates things that are related to experimental settings and circumstances. In survey studies, the researcher is more passive and he cannot directly influence the way the observational units are divided into different groups; in survey research, the sampling methods and ways of classifying variables that are chosen define the groups that are used.

A case study is a good research method, but it is impossible to draw inferences concerning the cause-effect relations using it.

Let it be noted that it is not always sensible to value one type of research as being better or worse. The reason for carrying out research determines which research types are sensible to use. From the point of view of understanding a phenomenon and describing it profoundly, *a case study*, belonging to the qualitative research tradition, is often an excellent type of research.[2] On the other hand, from the point of view of stating a generalization or a cause and effect relation, a case study is not a sensible option. On the other hand, one can state, based on computations using the binomial test, that it is possible to draw statistically significant inferences that are based on only *four* cases and clinically significant inferences based on three cases.[3] Thus, it has become popular during the previous years to combine several cases or small research studies into a so-called meta-analysis study.[4] As noted before, from the point of view of making generalizations or drawing inferences concerning cause and effect relations, the genuine experimental design is the overwhelmingly best option.

Experimental designs are also essentially connected to how different types of research are valued from the point of view of a systematic literature review.[5] As a refreshment of memory, different types of research are presented below in an order of importance (for more details see Goodman et al., 1996; Goodman, 1993; Partanen & Perälä, 1997; Mäkelä et al., 1996):

Types of research designs, in order of credibility from the point of view of cause and effect:

1. Large randomized and controlled experiments[6]
2. Small randomized and controlled experiments
3. Non-randomized studies with a control group[7]
4. Non-randomized studies with a historical control group
5. Cohort study
6. Case-control study
7. Cross-sectional study
8. Register study
9. A series of consecutive cases
10. A single case study

The experimental designs are weighted herein most heavily and the case studies most lightly. Nonetheless, let it be remembered that the order is based on the cause and effect relation–not any other criterion related to the reliability, profoundness or extensiveness of knowledge.

---

[2] Methods of qualitative research are introduced in Section III of Volume 1.
[3] For more details, see Chapters 2 and 3.
[4] Of meta-analyses, see closer Chapter 41 (41.2), Section VI of Volume 1.
[5] A systematic literature review has been addressed in Section I of Volume 1.
[6] The first two can be called true experiments (*randomized controlled trial, RCT*).
[7] Research types 3 and 4 are *quasi-experiments*.

Section III

Basics of Multilevel Modeling

CHAPTER 11

Introduction to Multilevel Modeling

Goals of the Chapter

1. To orient oneself to multilevel modeling of a relevant dataset.
2. To gain a general view of the special features related to multilevel analysis.
3. To understand the importance of preliminary examination of dataset for further analyses.

11.1 Multilevel modeling as a Part of Family if Regression Analysis

Multilevel modeling is part of the family of regression analysis.[1] Regression analysis for its part, belongs to a large group of various methods known as multivariate methods which are used to analyze several variables at once. This section is not an introduction to multivariate analysis—a beginning reader will gain a great deal of insight into this matter from Section II of Volume 2. Section II of Volume 2, and particularly Chapter 9 (9.2) therein, will provide a good understanding about regression analysis for a beginner. To reduce the need to browse, some of that text has been included in this main section as an orientation to actual multilevel modeling.

multilevel modeling
or
hierarchical linear
modeling...

A methodology for previously mentioned questions, which, in the human sciences context, has usually been called either *multilevel modeling*[2] (Goldstein, 1986; 1987; 1991; 1995; 2003; among

[1] See Chapter 9 (9.2), Section II of Volume 2.

[2] The terms multilevel modeling and hierarchical linear modeling are used interchangeably during the text. Mainly, however, the multilevel modeling as emphasized in the header of the Section. Note that Goldstein used the British spelling "modelling" in his texts, while Raudenbush and Bryk used "modeling." It seems that the American version has gained more popularity in the literature. Note, hence, also the minor linguistic deliberance: I am using the

linear mixed models
or
some other name

others) or *hierarchical linear modeling* (HLM) (Bryk & Raudenbush, 1987; Raudenbush, 1988; Raudenbush & Bryk, 2002, among others), has been developed in the 1980s. It is also known as *linear mixed models* (Garson, 2013a; SPSS 16 and later versions, among others), *mixed effects models* or *random effects models* (in biometrics), *random-coefficient regression models* (in economic sciences) or *covariance components models* or *variance component analysis* (in statistics) (for more detail on sources, see Raudenbush & Bryk, 2002, p. 5–6).

11.2 Basic Assumptions

Basic requirement for the dataset:
*Screened as all multivariate datasets

Perhaps, however, it might be wise to remind the reader that before the actual analyzes, the researcher should examine their dataset and variables carefully. The main principle is that one cannot make a good cake with spoiled ingredients; if the data are faulty or biased, the results cannot be trusted. A good rule of thumb for analyzing the dataset is the *Garbage In—Garbage Out* (GIGO) principle; one cannot bake an excellent cake by using rotten eggs. Naturally, it is assumed that the data is successfully encoded and there are no mechanical mistakes. Occasional typos or other random errors in a large dataset do not cause for the study to fail.

*no systematic errors

Systematic error is a larger problem; perhaps some of Jill's results are on John's row or a part of the attitude variable is, for some incomprehensible reason, in the wrong column. Mistakes such as these may be created unintentionally; if the dataset has been worked on in Excel software, for example, such mistakes should be carefully eliminated. A systematic error can be searched for in SPSS software, using, for example, the *Frequencies, Explore* or *Descriptives* selection in the *Descriptives* submenu of the *Analyze* main menu.[3] These basic issues are discussed in some more detail in Section I of Volume 2.

*clustered structure in the dataset

Each method also has certain assumptions, which are discussed further with the methods themselves.[4] For example, for the results of the variations of the family of traditional regression analysis (e.g., GLM, ANOVA and ANCOVA)[5] to be exact, it is usually

American spelling for "modeling" instead of the British spelling "modelling" deliberately even though, otherwise, the text follows the British system.

[3] See in detail in Section I of Volume 2 and Chapter 2 therein.

[4] A more advanced reader may benefit from the list compiled by David Garson about how to search for different assumptions. The list can be found at http://www.statisticalassociates.com/assumptions.pdf (last accessed on October 8, 2016).

[5] GLM, ANOVA, and ANCOVA are also part of the ANOVA family, which is discussed in Chapter 10 (10.2 and 10.3), Section II of Volume 2, and in Section II of this volume. Some basics also come up in this chapter. If these

required that the data be a random sample of the population and the observations be independent of each other. However, *in many applications* where information is gathered in everyday life—such as from schools or classrooms, various organizations, sports clubs, or by using stratified sampling—*the data is clustered or nested*. In practice, this means that the data has a certain hierarchical and internal or "nested" structure: Individual students belong to a certain class, classes belong to a certain school, schools are part of a certain municipality (or whoever organizes the education), and so on. Correspondingly, individual workers belong to a certain work community (such as a ward in a hospital), the work community forms part of the larger organization (a hospital), and the organization might also belong to a larger whole (a health care district or central hospital). The term *nested* describes the situation well: The young in a particular nest are more similar to each other than the young randomly picked for the study from various nests.

11.3 General Characteristics of the Data Suitable for the Multilevel Modeling

It is obvious that students/workers reached in the same school/work community are more similar to each other than would have been to students/workers picked completely at random. The obvious explanation for this is that all the students in a school have been affected by the same circumstances, atmosphere, rules, and culture. The situation might be completely different in the neighboring school. If pupils in a certain class of a certain school are studied, it is observed that the students within a class are more similar to each other than the other students in the school—not to even mention, the randomly selected students. In addition to the influence of the school culture, the students in a class are united by factors such as the teacher, among others.

Clustered dataset is characterized by the following general features:

characters of a clustered data

1. The dataset has a natural multilevel (hierarchical) structure
2. Dataset has both individual and group level variables
3. The observations are dependent of each other

Malin (2009) has excellently encapsulated what happens if this structure is not taken into account in the analysis:

1. the coefficients related to the model are probably faulty.
2. the standard errors of the estimates are probably faulty.

terms seem unfamiliar for the reader, it is warmly recommended to read the introductory texts to the aforementioned chapters.

3. statistical results are probably faulty.
4. conclusions are probably faulty.
5. accurate knowledge cannot be obtained about the associations between the data structure and the studied phenomenon.

In the clustered school data, an interesting question is how much a teacher or a school can explain the variation in student level; this is known as the teacher and school effect. Another interesting question is how individual variation differs in various schools, organizations, or countries with regard to certain interesting background factors when the upper level homogenizing influence is accounted for. After all, it is possible that in some schools or countries, the parents' socioeconomic background factors can strongly explain the learning results, whereas in other schools or countries, they are not such an important factor.

11.4 Software and related Literature

Multilevel modeling requires specific software

It does not make sense to perform the calculations related to multilevel modeling manually. Instead, there are several software packages developed especially for that. The best-known ones may be HLM and MLwiN. A free student's version of the first one is available online.[6] It is also possible to perform multilevel modeling with several well-known general software packages (such as SPSS, SAS, and STATA). The advantage of these is the ability to use other analysis options in addition to multilevel modeling with the dataset. Third, some software packages that specialize in structural equation modeling (SEM, see Section V)—such as LISREL, MPLUS, and STREAMS—can perform multilevel modeling. The advantages of the last mentioned are the multilevel examinations related to longitudinal dataset and the multilevel modeling of latent variables.

There are good sources for further reading

Of course, it is not possible to enter into the nuances of multilevel modeling in this short presentation. Lots of literature has been written about the subject, and the interested reader is encouraged to become familiar with the list of sources in, for example, Garson (2013a). The seminal works of Goldstein (1995, 2003), Hox (1995, 2002), Raudenbush and Bryk (2002), and Snijders and Bosker (2002) are naturally worth reading. If one gets over the basic mathematical notation and the notation related to regression analysis, the books lead to depths of the matter. One practical and approachable work is, among others, Heck and Thomas (2000), which introduces both the traditional multilevel modeling and the multilevel modeling of latent variables from the perspective of

[6] http://www.ssicentral.com/hlm/index.html (last accessed on October 15, 2016).

Section IV

Basics of SEM and Path Modeling in AMOS Environment

CHAPTER 19

Background and History of SEM

Goals of the Chapter

1. To gain a general view of the history of SEM (structural equation modeling) and path modeling.
2. To understand the connection of SEM and family of factor analysis.
3. To gain a general view of the software used in SEM analysis.
4. To become familiar with the essential literature in the field.

The last 10 years have seen the phenomenal increase in interest in the family of factor analysis, especially in educational research. In Figure 19.1, one sees that in the ERIC database with more than one million scientific papers, the hits of the key word '"factor analysis (general)' were in a decreasing trend from 1973 on until 2003. After that threshold year, the trend of *exploratory factor analysis* (EFA), as well as the "newcomers" *confirmatory factor analysis* (CFA) and *structural equation modeling* (SEM), have been in an increasing expansion in writing. Maybe the newcomers have opened new perspectives for the old family of methods, who knows.

In Chapter 8, Section II of Volume 2, the difference between EFA and CFA is handled. The main difference between EFA and CFA is whereas in EFA, the researcher is looking for a model or a theory to be explained among the combinations of variables, i.e. *exploring the data*, in CFA, a pre-existing model or theory is examined; it is confirmed, whether or not the model in question receives support from the data.[1] CFA is also called as SEM.[2] In

[1] An attentive reader may recall that statistics related to performing a SEM analysis are presented initially in Chapter 13 (13.2), Section II of Volume 1. In this section, those statistics and their interpretation is discussed in more detail.
[2] In principle, an SEM software (for example, AMOS, EQS, LISREL, MPLUS) can be used to analyse traditional regression models, measurement models,

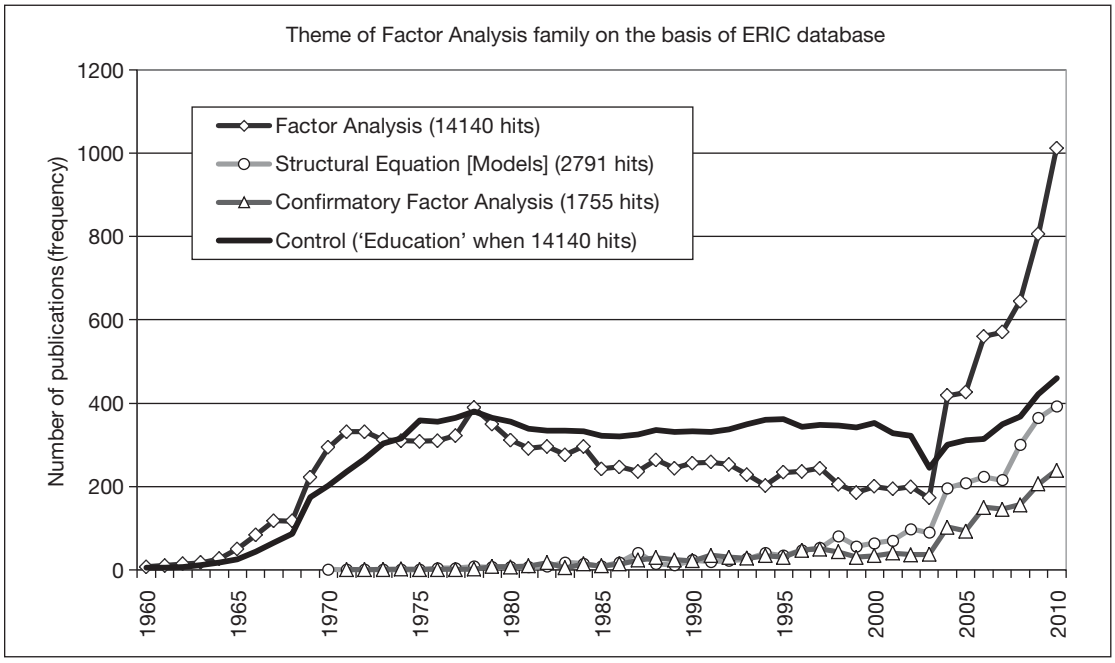


Figure 19.1 Historical Trends of Factor Analysis Family in ERIC Database

what follows, the SEM is used as a general name, if there is no other reason to use CFA.

Early warning on laborious notation and statistics…

A specialized language and notational system have been developed around SEM; these differ considerably from those of other analysis methods. The language in question is used, among other contexts, in program syntax and outputs; for this reason, discussing it is necessary. On the other hand, SEM analysis has inspired many researchers to contribute to the development of the method; an exceptionally large number of different tests and statistics exist for judging how well the models fit the data. These points will be covered in more detail in section 21.2.2, on the theory and concepts related to CFA.

Historical note

SEM began to develop strongly in the late 1960s, when the Swedish researcher Karl G. Jöreskog became interested in factor models that involve constraining the loading parameters of the

structural equation models and path models. Of these, the measurement models are strictly speaking confirmatory factor models and, thus, their analysis should be called CFA. Analysis of the entity of the latent variables and their connection, that is, analysis of structural equation models is called SEM analysis. These terms are, however, often used in a rather flexible manner. Some consider SEM as a subcategory of CFA. On the other hand, others consider SEM a parent category that includes both CFA and path models. In Chapter 8 (8.3), Section II of Volume 2 with CFA, the terms are used somewhat interchangeably.

factors.[3] Prior to this, Sewall Wright had developed the idea of path models; Wright already published a path model in 1918 and later developed it further. The path model contained the characteristics that are currently considered typical of structural equation models (Bollen, 1989a, p. 6). In the early seventies, Jöreskog (1973)–simultaneously with Keesling (1972) and Wiley (1973)—developed the LISREL model structure, which, in turn, enabled the regression-, the path-, and multilevel representations (Leskinen, 1987, p. 3; Bollen, 1989a, p. 6). Other software packages were developed as well. After the LISREL software,[4] such well-known software packages as EQS (Bentler, 1995), AMOS[5] (Arbuckle, 2011), and MPLUS (Muthén & Muthén, 1998–2007)[6] have been developed for building and analyzing the models. In addition, the SAS software includes its own procedure in performing structural equation modeling (SAS CALIS).

Software

Traditional sources on the SEM analysis include Jöreskog & Sörbom (1988; 2002b), Bollen (1989a) and Bentler (1995). Other valuable sources[7] for SEM analysis are, for example, Bartholomew, Knot & Moustaki (2011), Brown (2006), Byrne (1989; 2001; 2006), Hayduk (1987; 1996), Jöreskog *et al.* (2003), Kaplan (2001), Kline (2005), Loehlin (2004), Maruyama (1998), Schumacker & Lomax (2004), Raykov & Marcoulides (2006), and Thompson (2004). The reader also benefits from familiarizing with Internet materials produced by David Garson (2012d).

Selected additional literature

---

[3] Jöreskog had already published his first writings on the topic in the late 1960s (Jöreskog, 1967; 1969; 1970).

[4] While writing this, the latest version is LISREL 9.2 (Jöreskog & Sörbom, 1999a; 1999b; Jöreskog et al., 2003).

[5] In the literature, one sees an abbreviation Amos (*Analysis of Moment Structures* with low case letters) for the software. However, in what follows, the capital letters are used in abbreviation (AMOS), because of the uniformity in expressions with other software and programs in the book. While writing this (Autumn 2016), the latest version is IBM SPSS AMOS 23—the numbering system began to follow the SPSS system strictly from AMOS 7 to AMOS 16. The screen capturers are taken from version 22.

[6] Professor Petri Nokelainen and Dr Markus Mattsson especially suggested noting Bengt and Linda Muthén's MPLUS program, because its strength is the analysis of variables of nominal and ordinal scales. This may be valuable, especially in context of human sciences, where one often works with variables measured with the Likert scale. One may remember (see Section I and section 7.1 therein), that the Likert type of scale is an *ordinal* scale, though many analyses require at least an *interval* type of scale in the measurement. While writing this, the latest version of the program is MPLUS 7.4. At the site of the MPLUS program http://www.statmodel.com/references.shtml (last accessed October 16, 2016), there is good amount of relevant literature related with the program and SEM as a whole.

[7] All the experts who read the manuscript suggested their own favorite books—sometimes these were different than what I have used. This list is a combination of those.

CHAPTER 20

Research Questions and Assumptions in SEM

Goals of the Chapter

1. To gain a general view of the types of research problems to which SEM can give answers.
2. To know the assumptions behind SEM.

20.1 When SEM Is Appropriate

Performing SEM requires an underlying theory, whose correctness or probability is investigated...

If the theoretical model and the information contained in the data are contradictory, the model is unlikely true.

SEM can be used in situations where a theory on the variables' relationships exists. The aim of the analysis is to examine the plausibility of the theory, based on the dataset at hand (or to "ask the data" whether the theory receives support from it). The question is investigated using the correlation or covariance matrix. The basic idea is that if the theory postulates that certain variables are related, these variables should also be correlated more strongly than variables that are expected to be unrelated. According to Ullman (2001, p. 657), the basic technical question of SEM is: Will the population level covariance matrix, derived from the theory, be congruent with the covariance matrix observed in the sample? The same idea can, of course, also be phrased differently: The observed covariance matrix that is produced on the basis of the data is compared with the theoretical covariance matrix specified in the theory. If these matrices differ considerably, it can be said that the model is unlikely and, thus, poor.

SEM can be used to perform an extensive investigation into the variables' relationships. This method can be used to investigate, for instance, longitudinal or to even test extremely complicated relationships between variables. It is particularly effective when one wants to study the relationships between latent variables and not

Section V

Basics of Survival Analysis

CHAPTER 23

Introduction to Survival Analysis

Goals of the Chapter

1. To orientate oneself with the concepts used within the family of survival analysis
2. To familiarize oneself with the relevant further literature

Survival analysis (or *event history analysis, duration analysis,* and *transition analysis*) has become an established method in medical studies that typically examines things, such as how long a person lives after being diagnosed with cancer. In the context of the studies in human sciences, "survival" is relational and the analysis can be thought of as an event history analysis, although in some cases, like in dropout studies, individuals actually drop out of a course and are, therefore, in the analysis comparable to "dead."

An "event" does not always need to be negative or passive like "death," "dropping out of a course," "unemployment," or "divorce." It can also be positive and active, like "gaining peace after a conflict," "ratifying a decision," "making an innovation after receiving a funding decision," "enrolling onto a course after making a payment," "getting married," etc. Even though an event can be positive, it is more or less a convention of speaking of a *hazard function* $H(t)$, where t refers to time. This will be handled in what follows.

Survival or event history analyses are divided into three large groups: nonparametrical, semiparametrical, and parametric methods. Under these groups, there are four main ways to make Survival Analysis (Garson, 2009a). *Nonparametric* models make no assumptions about the distribution of the hazard function or which factors of the covariates explain it, even though it could be graphically illustrated. The nonparametric methods are divided into two categories: (a) *life tables* which are used to illustrate statistical Life Expectancy (*period and cohort life tables*) or the expectancy of some other interesting event within a certain time frame and (b) the *Kaplan–Meier survival analysis* (Kaplan & Meier, 1958), which is used for actual survival analysis when time is the only meaningful independent

Nonparametric models:

- Life tables
- Kaplan-Meier Survival Analysis

Semiparametric models:

- Cox regression

Parametric models:
- Weibull models
- Poisson regression

factor. In the *semiparametric* methods, there is no assumption of the distribution of the hazard function, but there are strong assumptions on how the covariates (like sex or age) affect the hazard function. The most central method of these is (c) *Cox regression* urges the reference (like the Kaplan–Meier): (Cox, 1972), which is probably the most highly used method in survival analysis. If, however, the researcher is able to determine the shape of the hazard function and how the covariates affect it, *parametric* methods can be used. The most commonly used parametric methods are (d) *Weibull models*, although there are others as well. Most parametric methods are suitable for situations where the *passing of time is of interest* (hence the name *duration models*). If the *number of events* is of the focus (*count models*), (e) *Poisson regression* is the most commonly used method. The two lastly mentioned—parametric models—are not handled within this book. A good view of Poisson regression can be found, for example, from Garson's material (Garson, 2012c) online.[1]

There is quite a large amount of literature relating to survival analysis. Garson (2012a; 2012b; 2012c; 2013b) has combined a good number of them in his material—some can also be found in this section. For a beginning reader, it is worthwhile reading Garson's material in any case. Event history analysis (Garson, 2012a),[2] life tables and Kaplan-Meier method (Garson, 2012b),[3] and Cox regression (Garson, 2013b)[4] can be found online.

Some basic readings regarding survival analysis are, for example, Blossfeld, Golsch, & Rohwer (2007), Blossfeld, Hamerle, & Mayer (1989), Box-Steffensmeier & Jones (2004), Hosmer & Lemeshow (1999), Kalbfleisch & Prentice (1980), Klein & Moeschberger (1997), and Yamaguchi (1991).

Like with many other methods in this book, a small dataset is also constructed in the context of survival analysis and it will be used for making the analysis first with a spreadsheet program and later on with SPSS program. However, Garson (2013b) suggests that the STATA software would be the most suitable software for executing Cox regression. Information about this can be found, for example, from Cleves et al. (2008) and from the manual itself (StataCorp, 2005).

In what follows, Chapter 24 presents special issues relating to LT (24.1) and computing LE (24.2). The latter can be useful, for example, for health economics and a planner for counties. Chapter 25 presents actual survival analysis methods: Kaplan-Meier method (25.1) and Cox regression (25.2).

---

[1] aiempi ei toimi. tämä toimii: http://www.statisticalassociates.com/loglinearanalysis.htm (last accessed October 16, 2016).

[2] parempi: http://www.statisticalassociates.com/parametricsurvival.htm (last accessed October 16, 2016).

[3] parempi: http://www.statisticalassociates.com/lifetables_kaplanmeier.htm (last accessed October 16, 2016).

[4] http://www.statisticalassociates.com/coxregression.htm (last accessed October 16, 2016).

CHAPTER 24

Life Tables

Goals of the Chapter

1. Get to know when and how life tables are used.
2. Be able to form a life table and to interprete it.
3. Be able to use SPSS in analyzing a life table.

Researchers and planners who are interested in projecting the population into the future by age and sex use demographic projection techniques to find out so-called survival rates. These are derived from life tables or census data and are used to calculate the number of people that will be alive at the future in time. Here, the concept of "being alive" can be understood broadly; the technique is not bound to population planning alone. Generally, the *life tables* are used to compute the statistical life expectancy (period and cohort life tables) or the expectancy of any interesting event during a certain time. In the context of life tables, the concept of "actuarial methods" (e.g., Sarna, 2009) or "actuarial studies" (e.g., Garson 2012a) are in use.

In what follows, two examples of how to use life tables to analyze Life tables and to compute population life expectancy are given. section 24.1 will consider a more general use of life tables in the context of experimental studies (though not restricted to experimental settings). The following Section 24.2 will present analysis of the population life tables.

24.1 Using Life Tables in Traditional Research

The first example of a life tables and related notation is a traditional life expectancy analysis with a small hypothetical dataset.

The same dataset will be used later on when presenting Kaplan–Meier method and Cox regression (Chapter 25).

24.1.1 Suitability

Life tables are traditionally used to measure mortality, survivorship, and the life expectancy of a population at varying ages. "Survivorship" can be understood broadly; in the example to come, it is examined how a course aimed at controlling oneself relates to violent behavior. The research question is: How many of the participants "survived" without a violent burst after a course for anger control?

24.1.2 Assumptions

The basic assumptions in life tables and life expectancy analysis is that those who have dropped out of the data (or "died") are no different than those present in the data and that time is the only interesting independent factor explaining the change.

Assumptions for Life Tables and Survival Analysis

Using life tables—as well as Kaplan–Meier method and Cox regression—has the following requirements:

1. a clearly and specifically determined start and end time
2. the endpoint gets two values (yes/no; here violence/no violence)
3. the start time can vary chronologically
4. the time period of the observation can vary
5. the portion of intervals that are short and did not end in an event ("censored", i.e., living and missing) should not be too big. Otherwise, the algorithm for computing the life expectancy can produce too low a value (under estimation). In addition
6. the distributions of the observation periods should be, approximately, equal in each group, and the portions of censored observations should be approximately equal

All the assumptions actualize in the example case, even though there are more censored cases in the test group than there are in the control group. The constructs and computation formulae of life tables are considered in more detail in the context of the SPSS output in section 24.1.4.

24.1.3 Briefly on Theory

24.1.3.1 Data

A simple (hypothetical) experimental longitudinal study concerning anger management is used as an example throughout the section. In the dataset, there are three variables: *COURSE*, which tells us whether the person attended a course of self-control (1)