


STATISTICS WITH SPSS

Introduction and RMs review

Week 1




Myths in social research???

1. It always starts with hypotheses.
 2. Significance tests are always an essential element of statistical analysis.
 3. The level of association or correlation between variables is enough for explaining studied phenomena.
- 

1. It always starts with hypotheses!

NO. NOT ALWAYS

- ▶ Every research should have a problem, which is transferred into a researchable form by a research question.
 - ▶ There are certain forms of RQs.
 - ▶ Only question WHY leads scholars to researches that are based on theory and thus require hypotheses.
 - ▶ There are 2 types of hypotheses – theoretical and statistical
- 

2. Significance tests are always an essential element of statistical analysis

NO. THEY AREN'T!!!

- ▶ Significance tests are part of the inferential statistics.
- ▶ Inferential statistics can be used only when we operate with probability samples.
- ▶ *Inferential statistics = generalize to the whole population – to be discussed later*
- ▶ These tests cannot be used when operating with data from census or quota, specific, snowball samples.

3. The level of association or correlation between variables is enough for explaining studied phenomena

NO, IT IS NOT!!!

- ▶ Measurement of association and correlation is a higher level of description.
- ▶ Although this association can help to understand the phenomenon, they cannot bring answer to question „why“ because why implies finding the causes behind it.
- ▶ However, if we want to explain a phenomenon, first we have to find the association – where there is no probabilistic association, there is not casual link.

Logic of quantitative research

- ▶ Gathered data ⇔ research design
⇔ research question ⇔ explain a problem
- ▶ Standardized data gathered by use of questionnaire and samples.
- ▶ Unit of analysis = individual, groups,...


Aggregate, sample data

- ▶ To find out the status of the individual characteristics the dataset is based on, to identify their characteristics, to describe them based on the characteristics of the dataset, to identify the distribution of them in the sample
- ▶ To find the causes of variability or the relations or links among various characteristics
- ▶ To monitor the trends/changes in the quantified characteristics of the dataset **by use of longitudinal survey**


Why quantitative data analysis?

- ▶ to present and interpret numerical data
- ▶ to determine the relationship between one thing (an independent variable) and another (a dependent or outcome variable) in a population

Surveys

- ▶ Use of quantitative design
 - ▶ Large samples
 - ▶ Use of representative samples
 - ▶ Collection of data in quantifiable form
- 


What are “statistics”?

- ▶ A set of mathematics concerned with understanding and summarizing collections of numbers
 - ▶ A collection of numerical facts
 - ▶ Estimates of population parameters \Rightarrow derived from samples
- 


Two major branches

- ▶ Descriptives statistics
- ▶ Inferential statistics

Quantitative data and statistical analysis

- ▶ Unit of analysis to be studied – how we choose them ⇒ sample design (probabilistic, quota, deliberate, etc.)
 - ▶ Gathering data by observation and measurement
 - ▶ Controlling data ⇒ cleaning data
 - ▶ Gathering first basic information on the dataset ⇒ univariate analysis
 - ▶ Calculating statistics and monitoring time and content changes
 - ▶ Stating the distribution of the phenomena, relational or correlation analysis
 - ▶ Running statistical checking of the hypotheses
 - ▶ Running inferential statistics
 - ▶ Running multilevel analysis
- 

Dataset and ways to select the units of analysis

- ▶ We have to settle the general universe or population.
 - ▶ Criteria for settling our general population are given by our research topic.
 - ▶ We are very rarely able to study the whole general population.
 - ▶ BUT if we do it, we operate with **Census**
 - ▶ Usually we operate with **samples and sample surveys**
 - ▶ Sample should be **representative** for our population.
- 

Representativeness of the sample

- ▶ Selected sample has the same/analogous structure as the general population from the point of view of various known and unknown characteristics \Leftrightarrow **probability sample**
- ▶ Results can be generalized to the whole general population from where the sample was drawn.

Populations versus Samples

- ▶ Aim to characterize and make inferences not about a particular person but about all people or all people having certain characteristics.
- ▶ These groups of interest are called **populations**.
- ▶ Typically, these populations are too large and inaccessible to study.
- ▶ Instead, we study a subset of the group, called a **sample**.
- ▶ In order to make reliable inferences about the population, samples are ideally **randomly selected**.
- ▶ The population properties of interest are called **parameters**.
- ▶ The corresponding measurements made on our samples are called **statistics**.

Concepts and their operationalization

- ▶ When measuring a certain phenomenon we need to know how it is delimited, defined.
- ▶ **We need its concept and definition!!!**
- ▶ We measure different aspects of the phenomena:
 1. Intensity of the characteristic (both for the units of analysis and the external context)
 2. Distance of objects
 3. Dependence (asymmetrical relation) and relationship (symmetrical relation)
 4. Global characteristics of the sample

Variables and Constants

- ▶ Constants are properties that never change
- ▶ Most sociological parameters vary considerably
 - Between individuals (e.g., political preference)
 - Within individuals (e.g., heart rate)
- ▶ Any variable whose variation is somewhat unpredictable is called a random variable

Scales of measurement


Categorical data

There are basically two kinds of data in this groups:

- ▶ **Nominal data**
- ▶ **Ordinal data**

Continuous data is (sometime) referred to as interval data.

Scales of measurement

- ▶ **Interval scale:** values are numerically meaningful, and interval between two values is meaningful.
 - ▶ **Ratio scale:** operates with a 0 value
 - ▶ **Dichotomous variables:** yes/no, belong/not belong response scale
- 

Nominal

Attributes are only named; weakest

Ordinal

Attributes can be ordered

Interval

Distance is meaningful

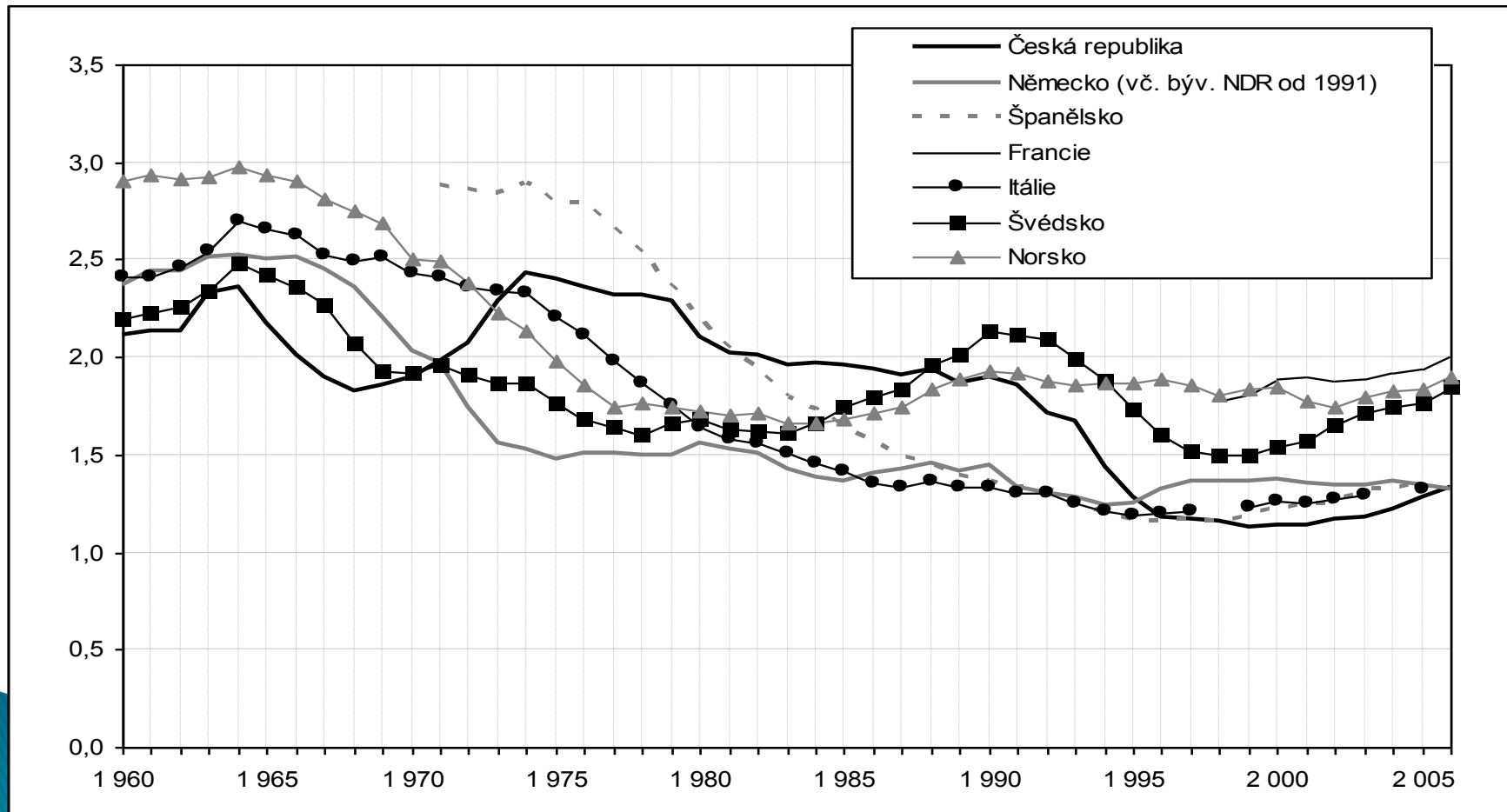
Ratio

Absolute zero

Continuous versus Discrete variables

- ▶ A **continuous** variable may assume any real value within some range

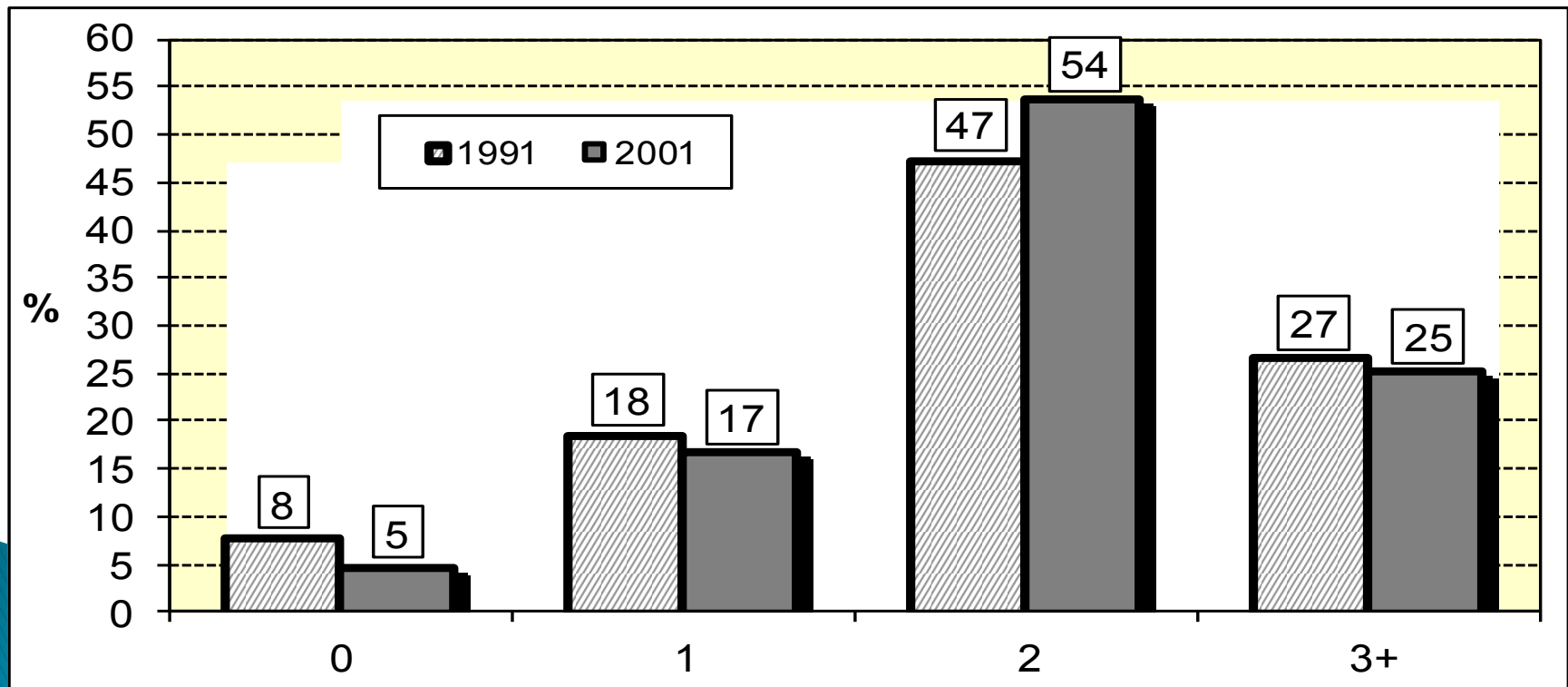
Total fertility rate 1960–2006



Continuous versus Discrete variables

- ▶ A **discrete** variable may assume only a countable number of values (intermediate values are not meaningful).

The number of children ever born to women aged 40–71 in the CR
(1991 and 2001)



Independent versus Dependent variables

- ▶ Studies involve **independent** and **dependent** variables.
 1. The **independent** variable is controlled by the researcher.
 2. The **dependent** variable is measured.
- We seek to detect and model effects of the independent variable on the dependent variable.

How else are they named?




Hypothesis


- ▶ A predicted or expected answer to a research question
- ▶ Example: Education is positively correlated to salary

- ▶ We call the hypothesis that you support (that supports your prediction) the *alternative* hypothesis,
- ▶ We call the hypothesis that describes the remaining possible outcomes the *null* hypothesis.
- ▶ Sometimes we use a notation like H_A or H_1 to represent the alternative hypothesis or your prediction, and H_0 or H_0 to represent the null case.
- ▶ One-tail vs. Two-tails hypotheses

SPSS – a little bit of history

- ▶ SPSS = Statistical Package for the Social Sciences
 - ▶ Was developed in the late 1960's by American political scientists (Norman Nie)
 - ▶ Is not the best or most advanced statistical software, but combines comprehensiveness and accessibility
 - ▶ Still widely used in academic and commercial research
- 

What is SPSS course about?

- ▶ to introduce the basic notions in statistics – population, parameters, descriptive statistics, inferential statistics, sample, variables, etc.
 - ▶ to build and use variables and databases using SPSS
 - ▶ to introduce the basic statistical data analysis methods using SPSS
- 

COMPULSORY LITERATURE

- ▶ Field, A. – *Discovering statistics using SPSS for Windows : advanced techniques for the beginner*, London: Sage Publications, 2009
- ▶ Vaus, D.A. de – *Analyzing Social Science Data*, London: Sage Publications, 2002
- ▶ Hardy, M. A; Bryman, A. *Handbook of data analysis*, London : Sage Publications, 2004
- ▶ Miller, R.L. – *SPSS for Social Scientists*, Houndsmill: Palgrave, 2002

Working datasets

- ▶ European Values Survey dataset – Wave 3 – 2008 – 3 countries: Czech Republic, Finland and France
- ▶ European Social Survey – ESS Cumulative Dataset Rounds 1–7 (data for 4 countries: Austria, Czech Republic, Finland and France)

Course structure

- ▶ 1st part ⇔ lecture
- ▶ 2nd part ⇔ seminar – individual work on PC

Exam dates (to be confirmed)

- ▶ 1st term
- ▶ 2nd term