

Introduction to Statistics and SPSS

Methodology of Conflict and Democracy Studies

December 3

Aim of this lecture

- Variables and their categories
- Population and sample
- Hypotheses and null hypotheses
- Statistical significance
- Introduction to SPSS
- How to make your own variables

Logic of Statistics

- Deductive logic of research
- What we do:
 - Derive hypotheses from the theory
 - Define variables and operationalize our concepts
 - Collect the data
 - Test the hypotheses using statistical models
 - Provide interpretation and decide whether our hypotheses hold or not
- This all requires more than just a few cases

Variables

- Measurable items that change their values
- Number of cars on highways, maximum daily temperature, local turnout in elections
- Independent (predictor) and dependent (outcome) variables
- Main tool for testing hypotheses

Levels of Measurement

- Completely different categorization of variables than IV and DV
- Categorical:
 - Nominal
 - Ordinal
- Continuous:
 - Interval
 - Ratio

Nominal Variables

- Their values cannot be ordered in a logical way
- Names of towns, names of streets, telephone numbers, colors, species of animals, numbers of players
- Binary variables – nominal variables with just two values
 - Someone is employed or is not employed
 - Citizen either voted in election or did not vote
 - You either attend this lecture or you do not

Ordinal Variables

- Their values can be ranked in a logical way however we cannot tell the exact differences between the values
- School grades, Olympic medals, military ranks, age groups
- Ordinal variables tell us more than nominal variables (ordering values) but less than interval and ratio variables

Interval and Ratio Variables

- Interval:
 - We can order the values and we know the differences
 - Equal intervals on a scale represent equal differences
 - Temperature in Celsius
- Ratio:
 - Same as interval but ratios of values are meaningful
 - They contain a true zero
 - Distance in kilometers, time in seconds, number of books in a library
- In SPSS interval and ratio variables are under the same label (scale)

Continuous or Discrete?

- Scale (interval, ratio) variables can be either:
 - Continuous
 - Discrete
- Depends on whether the values can take any values on a scale
- Success rate in a test (in %)
- Number of kids in families

| | Nominal | Ordinal | Scale (interval, ratio) |
|--|----------|----------|----------------------------|
| Can we logically order the values? | No | Yes | Yes |
| Do we know differences between the values? | No | No | Yes |
| Continuous or discrete | Discrete | Discrete | Continuous / discrete |

Population and Sample

- Population:
 - Includes all possible subjects of a dataset
 - All towns of a country, all university students
- Sample:
 - Includes only part of the cases and it is a subset of the population
 - Important feature – representativeness
 - 1,000 people in a survey
 - Many ways of selection – random and non-random

Population and Sample

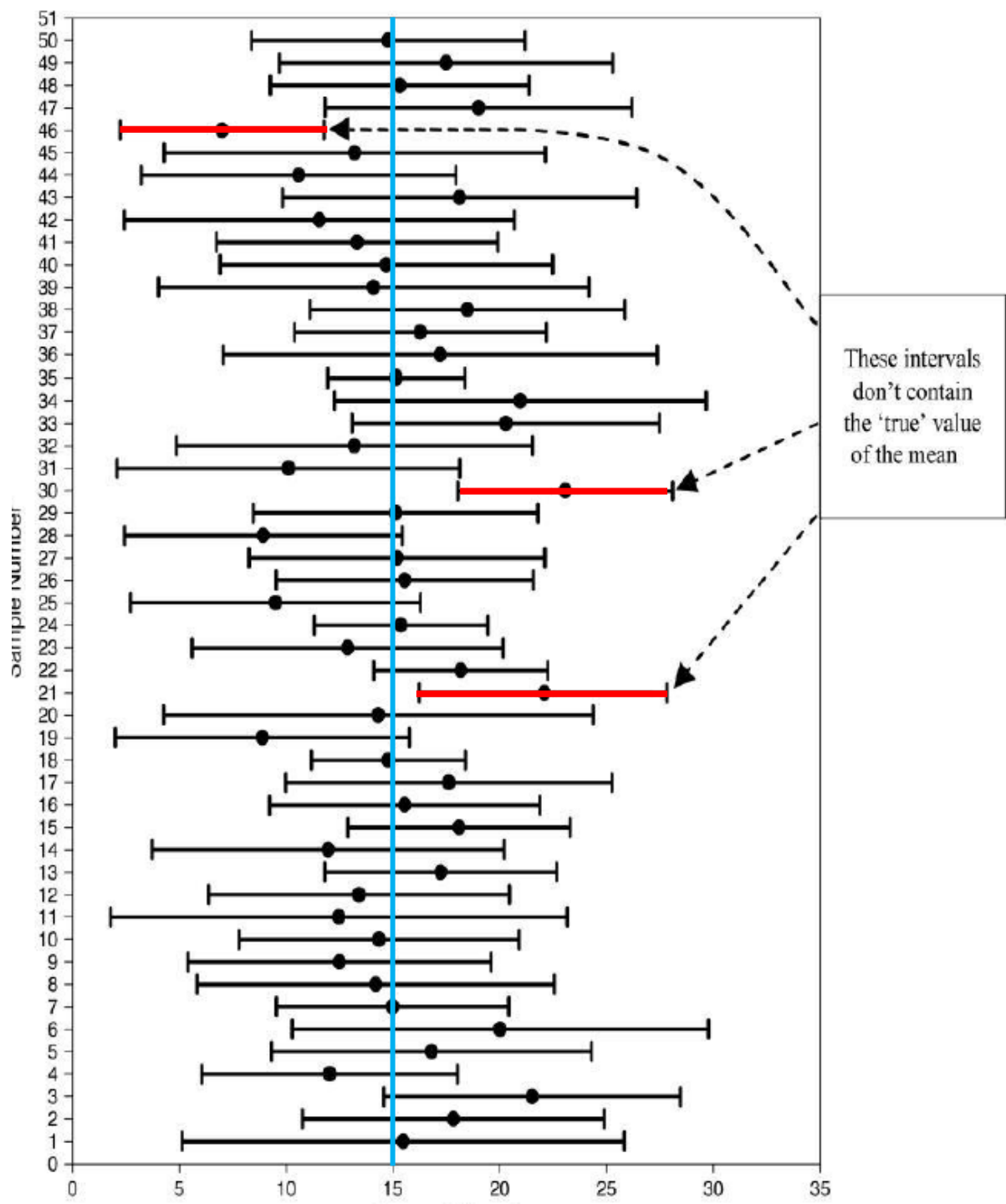
- When your work with population data:
 - You have data for the whole population
 - Your findings apply to the whole population
- When you work with sample data:
 - You have data for the sample only
 - Your aim is to generalize the findings to the whole population
- Nobody cares if 53 per cent of 1,000 survey respondents support Brexit but whether 53 per cent of UK population has this opinion

Hypotheses

- Logical conjecture about the nature of relationships between two or more variables expressed in the form of a testable statement (O’Leary 2004)
- *“Higher unemployment leads to higher frustration of the society”*
- Null hypothesis:
 - Statement about absence of any relationship between independent and dependent variable
 - Every hypothesis has its null hypothesis
- In statistics, all operations test the **null hypotheses**
- After testing the null hypotheses either hold or they are dismissed (what gives support to our hypotheses)

Statistical Significance

- Working with samples is always connected with some sampling error
- Statistical significance allows to estimate whether the found effects are not only random and so they can apply to the whole population
- Levels of significance: 95 %, 99 %, 99.9 %
- Significance and hypotheses testing:
 - If a result is **significant**, we reject the null hypothesis and we gain confidence in our own hypothesis
 - If a result is **not significant**, we keep the null hypothesis and we thus we have no support for our own hypothesis



Statistical Significance

- A statistically significant effect does not necessarily mean that it is also important and meaningful
- A finding that a new medicine reduces body temperature of the patient by 0.01 °C (significant at 99.9 %)
- A finding that a new medicine reduces body temperature of the patient by 1 °C (significant at 99.9 %)