

Introduction to Statistics and SPSS

Methodology of Conflict and Democracy Studies

November 25

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 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 variable with just two values
 - Someone is employed or he/she is not employed
 - Citizen either voted in election or did not vote
 - You either came to this lecture or you did not

Ordinal Variables

- Their values can be ranked in a logical way however we cannot tell 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 have to contain a true zero
 - Distance in kilometers, time in seconds
- In SPSS interval and ratio variables are under the same label (scale)

Continuous or Discrete?

- Continuous (interval, ratio) variables can be either:
 - Continuous
 - Discrete
- Depends 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 | No | Yes |
| Do we know differences between the values? | No | Yes | 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 students of a university
- Sample:
 - Includes only part of the cases and it is subset of the population
 - Important feature – representativeness
 - 1000 people in a survey
 - Many ways of selection – random and non-random

Population and Sample

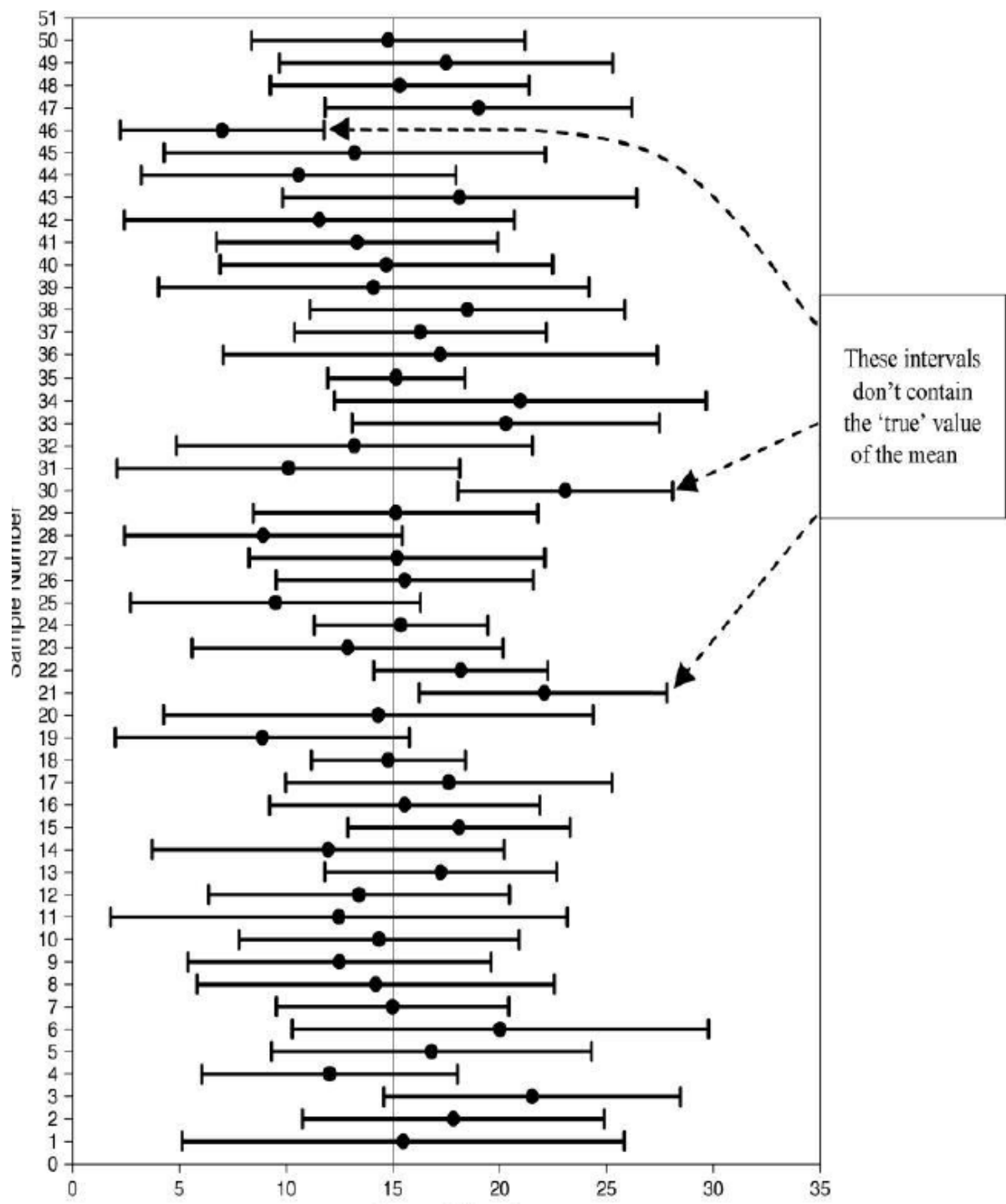
- Working with population data:
 - You have data for the whole population
 - Your findings apply to the whole population
- Working with sample data:
 - You have data for the sample only
 - Your aim is to apply the findings to the whole population
- Nobody cares if 53 per cent of 1,000 survey respondents 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 hypotheses:
 - Statement about absence of any relationship between independent and dependent variable
 - Every hypotheses has its null hypotheses
- In statistics, all operations test the **null hypotheses**
- After testing null hypotheses 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 they can be applied 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 hypotheses
 - If a result is not significant, we hold the null hypothesis and we thus we have no support for our own hypothesis

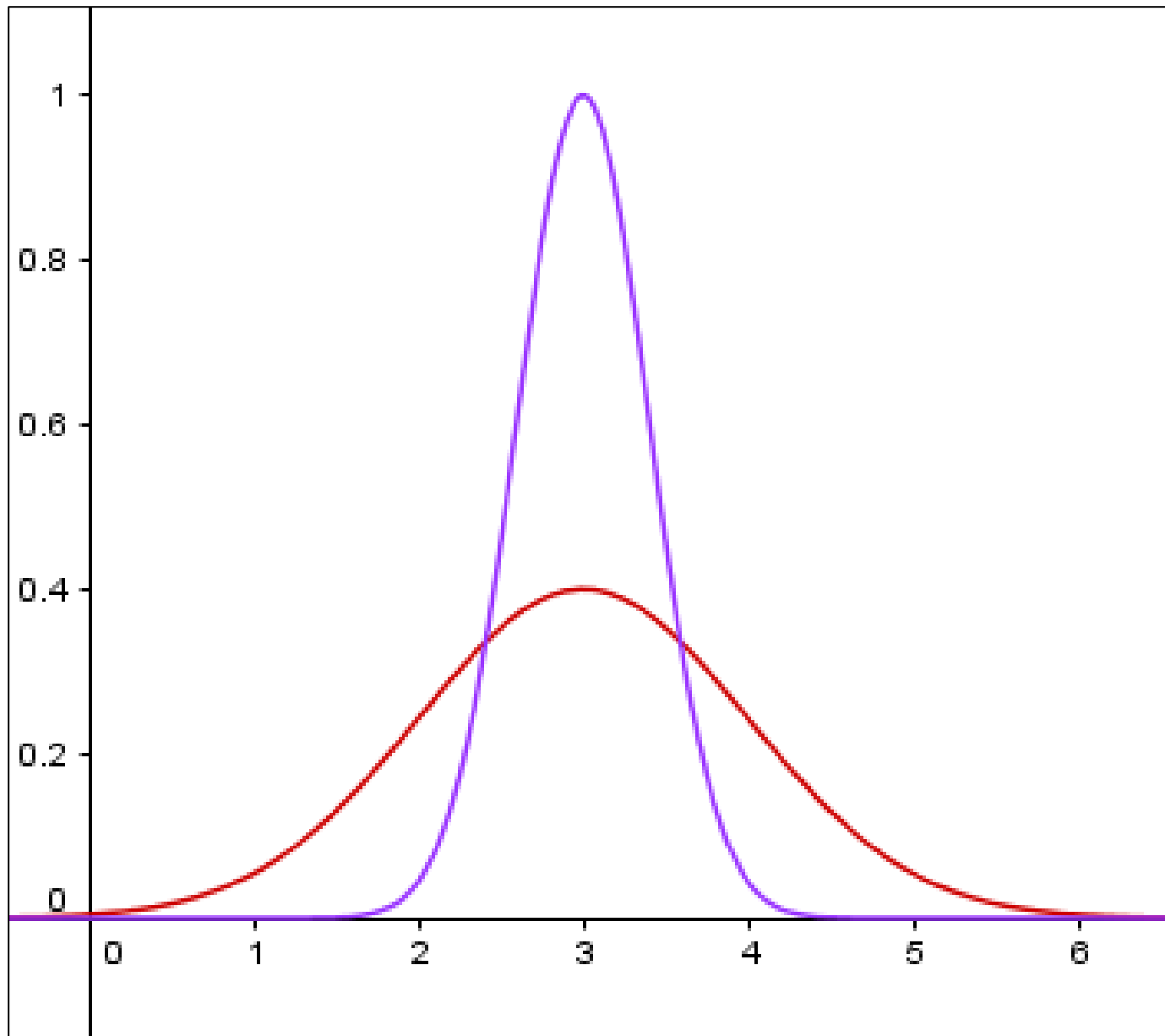


Statistical Significance

- 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 %) Cravath 2006 2010
- A finding that a new medicine reduces body temperature of the patient by 1 °C (significant at 99.9 %)

Descriptive Statistics

- Basic measures to summarize the characteristics of your data
- Various types:
 - Central tendencies – mean, median, modus
 - Dispersion – standard deviation, variance, minimum, maximum
- Not all descriptives are suitable for all types of variables
- We use them to describe and explore your data



| | Variable | Mean | Std. Dev. | Minimum | Maximum |
|-----------------------------------|-----------------|-------------|------------------|----------------|----------------|
| Reelection | Nominal | 0.75 | 0.43 | 0 | 1 |
| Number of grants | Scale | 1.33 | 1.52 | 0 | 10 |
| Grant in election year | Nominal | 0.36 | 0.48 | 0 | 1 |
| Incumbent terms | Scale | 2.19 | 1.21 | 1 | 6 |
| Unemployment | Scale | 19.41 | 13.15 | 0.00 | 94.94 |
| Number of challengers | Scale | 2.18 | 1.38 | 1 | 7 |
| Grants in EUR (per capita) | Scale | 77.74 | 209.62 | 0.00 | 5,331.82 |
| Mayor from governing party | Nominal | 0.41 | 0.49 | 0 | 1 |

How to Obtain Descriptives in SPSS

- Analyze > Descriptive Statistics > Frequencies
- Move variables of interest to the right
- In 'Statistics' choose all measures you require

Statistics

Age of respondent, calculated

| | | |
|----------------|---------|--------|
| N | Valid | 2398 |
| | Missing | 0 |
| Mean | | 49,04 |
| Median | | 49,00 |
| Mode | | 50 |
| Std. Deviation | | 17,561 |
| Minimum | | 15 |
| Maximum | | 90 |
| Sum | | 117591 |

