

# Quantitative Methods

## I. Objectives

The course is addressed to those, who are going to prepare empirically oriented MA thesis. Its aim is to help participants to understand basic rules of sampling and statistical inference in order to avoid typical mistakes and traps in applying them to survey data analysis.

Entry requirements for participants: well-grounded elementary statistics course, operating Excel spreadsheet and SPSS package.

In term essay participants will demonstrate their understanding of theory underlying practical methods of dealing with imperfect survey data and elementary technical skills in survey data processing.

## II. Short description of the course

The course deals with two challenges permanently present in making inference about population based on survey data: accuracy and certainty. Both issues will be considered in close to real life situation in which only part of sampled objects are accessible for researcher, causing inference on the population much less certain and precise than in ideal surveys.

The course has two components: theoretical and practical one.

### A. Theory

The first component comprises of introduction to sampling and sample-population inference. Basic concepts will be presented during the lectures covering the following issues:

1. Probability as a language of random experiment description
2. Sampling:
  - i. Sample space, sample statistics,
  - ii. Sample design
  - iii. Sample statistics distribution: bias, dispersion
3. Statistical inference
  - i. Limit theorems
  - ii. Normal distribution, chi-square distribution
  - iii. Interval estimation: trustworthiness and accuracy
  - iv. Estimation accuracy and sample design – design effect
  - v. Complex samples and estimation accuracy
  - vi. Hypotheses testing: decision rule, decision error
  - vii. Significance level, power of the test
4. Survey inference practical issues
  - i. Survey quality
  - ii. Survey cost
  - iii. Sample size and limit theorem applicability
  - iv. Sample „representativeness” misunderstanding
  - v. Assessing consistency of sample and population stratification distributions\*
  - vi. Accuracy and non-response
  - vii. Prediction of referendum outcome\*

## B. Application

In practical component of the course data from GSSR survey data banks will be used in two case studies and will serve as examples of two tasks to be performed in term essay. Necessary computational tools for task completion (Excel, SPSS) will be presented together with case presentations.

The first case deals with effective sample quality assessment. Chi-square statistics is used to express discrepancy between sample and population distribution of basic stratification variables and justify sample quality evaluation.

In the second example quasi-interval of confidence is applied to the interpretation of outcome of imperfect survey in which response rate is much below 100%. Additionally, prediction of population future behavior based on survey responses takes into account lack of definite beliefs of some of respondents answering “don’t know”.

## III. Requirements for completion

Participants will use real survey data from GSSR data base. After simple analysis of selected SPSS data sets they will prepare essay (3000 words), called “Report”, in which statistical theory presented during the lectures will be applied to real survey problem solution. General requirements for the report are described below.

### A. Data

#### 1. Survey selection

From CSS survey data base (kept in COMMON folder) select survey from which select two countries in the same year or two waves of a survey conducted in the same country in different years. Recommended data bases are:

- **(US) General Social Survey 1972-2016**
- **International Social Survey Program 1985-2014**
- **European Social Survey 2002-2014**

You may use data from other sources – consult it in advance with instructor.

#### 2. Variables selection

##### i. For Task 1 (Assessing consistency of sample and population stratification distributions)

Check availability of population distributions of basic stratification variables (gender, age, place of residence, education level, economic activity) in survey documentation or in official, reliable (e.g. census data) sources.

From SPSS data set select 4 stratification variables for which population distributions are available.

Take care of survey data coding with coding used in population distributions – check whether in case of different coding schemes there is “common denominator” for both distributions.

##### ii. For Task 2 (Prediction of referendum outcome)

Select variables which may serve as indicators of peoples’ mass behavior like voting in election or in referendum.

Select indicators related to normative and emotionally important issues for which refuse of an answer (“don’t know”, “difficult to say” responses) indicate attitude as well as other responses. Recommended choice: GSS, ISSP – “abortion legal”.

You may use indicators with more than 2 responses, but remember that they must have predictive potential after their dichotomization.

### 3. Tasks' description

#### iii. Task 1: Assessing consistency of sample and population stratification distributions

1. Select two countries/waves and common set of socio-demographic variables
2. Create two data sets if necessary
3. Find population distributions corresponding to the sample data
4. Recode (if necessary) variables in two samples as to obtain comparable number of categories in each data set (for each country selected) and in corresponding population distributions.
5. Export sample and population distributions of selected variables to Excel spreadsheet.
6. Add necessary formula to the spreadsheet columns and compute chi-square statistics for each variable used in comparison
7. Find chi-square critical value and degrees of freedom for each variable and check if a sample distribution differs from population data significantly or not.
8. Use 0,05 significance level for each comparison
9. Summarize results in a short description of consistency check outcome. Compare countries/waves
10. Attach SPSS syntax in appendix of the chapter.

#### iv. For Task 2: Predicting mass behavior using survey data with non-responses

1. Select at least 5 indicators of an attitude which has obvious and strong relationship to peoples' behavior.  
Recommended choice: 7 questions on abortion legalization.
2. Select two countries or survey waves with data compatible for each indicator.
3. Recode responses to binary variables if necessary.
4. Paste sample data for each indicator (for two countries) to Excel spreadsheet
5. Find response rates for each sample (country)
6. If you will not find actual response rate values in the study documentation, assume 60%
7. Calculate in Excel middle point and the width of the quasi-interval of confidence for population fraction for all binary items in three variants:
  - a. Respondents who answered "don't know" or "difficult to say" will not act (behave, vote, support)
  - b. Respondents who answered "don't know" or "difficult to say" will act like those who answered "Yes" (have value 1 in the binary form of the item)
  - c. Respondents who answered "don't know" or "difficult to say" will act like those who answered "No" (have value 0 in the binary form of the item)
8. Use pessimistic assumption for non-accessible part of the population represented by non-responding people (i.e. assume that fraction of "Yes-sayers" in that part of the population under study lies somewhere between 0% and 100%).
9. Mark these items for which you may formulate ultimate prediction of mass behavior (when quasi-interval for population fraction does not contain 50% point)
10. Compare two surveys in regard to the predictability of the mass behavior on the basis of survey data with non-responses. Which item are and which are not reliable base for prediction?
11. Attach SPSS syntax in appendix of the chapter.