Course title
WA105 Introduction to SPSS

Instructor details
First name, last name: Florian Meinfelder
Department/Unit: Department of Statistics and Econometrics
Institution: University of Bamberg
Full postal address for ECPR correspondence:
Otto-Friedrich Universität Bamberg
Lehrstuhl für Statistik und Ökonometrie
Feldkirchenstraße 21
D-96052 Bamberg
Phone: +49 951 – 863 2741
Fax: +49 951 – 863 2532
E-mail: florian.meinfelder@uni-bamberg.de

Short Bio
Florian Meinfelder is a lecturer at the Department of Statistics and Econometrics at the University of Bamberg, teaching, among others, courses on Bayesian Statistics, statistical analysis with missing data and courses on R programming. His research interest focuses on robust Multiple Imputation methods, as well as on matching algorithms for missing data and causal inference problems.

Prerequisite knowledge
Note from the Academic Convenors to prospective participants: by registering to this course, you certify that you possess the prerequisite knowledge that is requested to be able to follow this course. The instructor will not teach again these prerequisite items. If you doubt whether you possess that knowledge to a sufficient extent, we suggest you contact the instructor before you proceed to your registration.

- Introductory statistics:
  - Basic knowledge of sampling-based inference in general
  - Basic knowledge of the (generalized) linear regression model
  - Basic knowledge of further multivariate methods (ANOVA, cluster analysis, ...)
- Prior experience with spreadsheet data, e.g. analysis in MS Excel, but no prior knowledge in STATA, R, SAS, required!

Short course outline
This course introduces the capabilities of SPSS (the ‘Statistical Package for the Social Sciences’). While SPSS offers many GUI-based functionalities, the emphasis of the course will be on the usage of the syntax interface. By the end of the course users will be familiar with the ‘architecture’ of SPSS, and will be able to write their own analysis syntax, starting with the import of the original analysis data

Disclaimer: the information contained in this course description form may be subject to subsequent adaptations (e.g. taking into account new developments in the field, specific participant demands, group size etc.). Registered participants will be informed in due time in case of adaptations.
set, and finishing with the final analysis relevant for the user’s research. Since many statistical standard analyses will be used as examples, the participants are expected to have at least a vague idea of the methods to be covered, since the focus of the course is on conveying the idea of doing data analysis using a statistical programming tool rather than on the underlying statistical methods to be used for demonstration purposes.

**Long course outline**

*Reproducible Research* has become a ‘hot topic’ in data-based-analysis context for the last couple of years (it should always have been that way, since it is a basic principle of science in general). This objective promotes the usage of statistical programming software such as SPSS, STATA, SAS or R. SPSS is a wide-spread tool for doing data analysis, and while many users are mainly using its graphical user interface, some functionalities are only accessible via syntax commands. Additionally, getting used to work with the syntax interface automatically generates reproducible research (if you follow some simple rules), and thus will help the user to modify existing results (e.g. as requested by the reviewers of a submitted paper).

The course starts with some ‘fundamentals’: Reading data into SPSS, and organizing your analysis syntax, such as commenting code or ‘hiding’ less important syntax in sub-project syntaxes. It is generally desirable to keep your code tidy, and to assign temporary results, code snippets, saved data sets to appropriate directories (part of this is also, for instance, to never modify the original data set), because it will make it easier for others to understand the syntax (and yourself when you are reviewing or modifying your code a couple of weeks after writing it). After becoming familiar with the basic concepts of SPSS, its ‘architecture’, we will produce some simple analyses using the tutorial data sets which are part of the standard installation. Participants will have the opportunity to already work on several exercises during day one of the course. The first day of the course will also focus rather on descriptive statistics, such as creating tables and cross-tabulation, aggregating data over groups, and computing useful statistics per group (means, variances, ranges, etc.). We will also give an introduction to data management and handling at the end of the first day, where we demonstrate the applicability of multiple data windows, and introduce some commands for merging different data sources via identifier variables.

Day two of the course will cover data editing, some more advanced analysis options in SPSS as well as an introduction to the macro language and SPSS’ programmability interface (to Python and R). Data editing is one of the main assets of SPSS, since the meta data window gives an overview of variable and value labels, as well as the measurement level and missing data coding. Recoding and aggregating data will be among the main tasks to be carried out before the actual analysis. After the data-editing part we will continue with the plotting and output options in SPSS, since this is another vital part of research documentation. The more advanced analysis options, such as linear regression or generalized linear regression, will be covered in the afternoon of day two. At this point, we will go back to using the GUI options, as it is much easier to handle high-level analysis in SPSS this way. However, the syntax will always be inserted into our general project syntax, thus upholding the ‘reproducibility doctrine’. Regression apart, this section of the course will primarily deal with multivariate data analysis in general. Additionally, a small digression is planned to introduce the options in SPSS for handling missing data properly. The final part of day two is reserved for some more advanced options which help to integrate SPSS in an even more sophisticated data analysis environment. We do not expect participants to be able to make use of these advanced options immediately, but we want to create a general awareness for the possibilities (and limitations) of SPSS. The macro language in SPSS will have been superficially introduced by then, but in this final part, we will show, how it can be used to organize your syntax a better way by replacing copy&paste code snippets with loop structures and functions. Other features of SPSS which have been added since version 14 comprise the programmability interface to R and Python, as well as the ‘Output Management System’ that allows capturing output as data. While we will present several tutorials and exercises for the topics to be covered on day two, this final part will only consist of examples
but we hope they help even participants who had already been familiar with SPSS before the course to ‘inspire’ new ways of using the software

By the end of the course participants will be able to carry out data analysis using SPSS – ‘from the cradle to the grave’ (i.e. from reading in the raw data up to the final sophisticated analysis), and be able to organize their research project such that others will have no difficulties in understanding how the SPSS analysis had been carried out.

**Day-to-day schedule (Friday 13 February to Saturday 14 February)**

<table>
<thead>
<tr>
<th>Topic(s)</th>
<th>Details</th>
</tr>
</thead>
</table>
| Friday afternoon | • Fundamentals of SPSS and project organization  
• Introductory commands and analyses | See course description |
| Saturday morning | • Data editing  
• Output handling  
• Advanced data analysis with SPSS | See course description |
| Saturday afternoon | • Advanced data analysis with SPSS (ctd.)  
• Further topics: macro language, programmability interface to R and Python | See course description |

**Day-to-day reading list**

| Readings (please list at least the compulsory reading for the scheduled day) |
|-----------------------------|---------------------------------------------------------------|
| Friday afternoon | Any introductory Statistics textbook | Chapters covering hypothesis testing, (generalized) linear regression models, ANOVA, multivariate analysis |
| Saturday morning | [http://www.spsstools.net/spss.htm#GeneralTutorials](http://www.spsstools.net/spss.htm#GeneralTutorials) |  |
| Saturday afternoon | [http://www.spsstools.net/Macros.htm](http://www.spsstools.net/Macros.htm) |  |

**Software and hardware requirements**

SPSS  
Workstations or Notebooks with MS Windows 7 as OS.

**Literature**

Course literature

**Lab requirement**
Workstations or Notebooks with MS Windows 7 as OS.

**Other recommended courses (before or after this course)**
The following other ECPR Methods School courses could be useful in combination with this one in a ‘training track’. NB this is an indicative list.

**Before this course:**

<table>
<thead>
<tr>
<th>Course title</th>
<th>Summer School</th>
<th>Winter School</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**After this course:**

<table>
<thead>
<tr>
<th>Course title</th>
<th>Summer School</th>
<th>Winter School</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>