Course plan for
Advanced Applied Statistics
(version 2)

Sune Welling Hansen
Robert Klemmensen
Erik Gahner Larsen

September 30, 2015
1 Course overview

Social scientists study complex phenomena. In order to study these phenomena using quantitative information, they have to use suitable statistical methods that take into account the properties of the research question, research design and the data. The purpose of this course is to give students a detailed understanding of select advanced statistical methods and give them the skills required to implement these methods in either Stata or R, two very popular statistical programs. In many cases, the theoretical and empirical interests of the students result in both interesting and complex topics of study, but using only basic statistical methods would make it impossible to analyze these questions, or would result in less valid (perhaps even invalid) results and conclusions. The main objective of this course is to teach students how they can overcome these problems by providing them with the necessary skills to employ state-of-the-art statistical methods.

Upon completion, students are expected to gather and develop the following knowledge, skills, and competence related aspects:

Knowledge

1. Understand the principles of statistical analysis and the assumptions behind select statistical methods
2. Discuss and understand assumptions behind these methods
3. Evaluate whether these assumptions are correct, or realistic, and correct for breaches with viable solutions (through the advanced methods discussed in the course) for complex problems or data-related issues
4. Interpret results from these select statistical methods

Skills

1. Select appropriate statistical methods for own research questions
2. Analyze and critically assess scientific articles that use the advanced statistical methods taught in the course

Competences

1. Implement advanced statistical methods using either Stata or R
2. Comfortably use statistical software for own analyses
3. Correctly select and interpret the quantities of interest resulting from the application of these methods and link these results to research goals
2 Prior knowledge, mandatory assignments and the exam

Students are expected to have taken at least a course on social scientific method (i.e. be familiar causality, measurement, research design, validity and reliability etc.) and a prior quantitative methods class (i.e. what is a variable, significance testing and confidence intervals, correlation and regression etc.). While having experience with either Stata or R will be to the benefit of the student, it is not a mandatory prerequisite.

In order to be able to take the exam, students have to complete and pass four assignments during the semester. All the assignments are empirical applications, and are graded as passed/not passed. Those who do not pass all four assignments by the date stipulated below will not be able to attend the exam.

The exams consist of a one week take-home written exam (max. 10 pages) graded as passed/not passed or on the 7-point scale. The exam will be an empirical application of one or more theoretical arguments.
<table>
<thead>
<tr>
<th>No.</th>
<th>Topic</th>
<th>Teacher</th>
<th>Date</th>
<th>Time</th>
<th>Room</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction and refresher to basic statistics</td>
<td>Sune</td>
<td>2 September</td>
<td>16-18</td>
<td>U49E</td>
</tr>
<tr>
<td>2</td>
<td>Refresher of linear regression</td>
<td>Robert</td>
<td>9 September</td>
<td>16-18</td>
<td>U29A</td>
</tr>
<tr>
<td>3</td>
<td>How to study causality in non-experimental studies</td>
<td>Sune</td>
<td>16 September</td>
<td>16-18</td>
<td>U29A</td>
</tr>
<tr>
<td>4</td>
<td>Binary logistic regression</td>
<td>Sune</td>
<td>23 September</td>
<td>16-18</td>
<td>U29A</td>
</tr>
<tr>
<td>5</td>
<td>Ordered and multinomial logistic regression</td>
<td>Sune</td>
<td>30 September</td>
<td>16-18</td>
<td>U29A</td>
</tr>
<tr>
<td>6</td>
<td>Time &amp; space: Models for independent pooled cross sections and panel data</td>
<td>Sune</td>
<td>7 October</td>
<td>16-18</td>
<td>U29A</td>
</tr>
<tr>
<td>7</td>
<td>Time &amp; space: Fixed effects and random effects models</td>
<td>Sune</td>
<td>21 October</td>
<td>16-18</td>
<td>U29A</td>
</tr>
<tr>
<td>8</td>
<td>Hierarchical models I</td>
<td>Robert</td>
<td>28 October</td>
<td>16-18</td>
<td>U29A</td>
</tr>
<tr>
<td>9</td>
<td>Hierarchical models II</td>
<td>Robert</td>
<td>4 November</td>
<td>16-18</td>
<td>U29A</td>
</tr>
<tr>
<td>10</td>
<td>Experiments and Causal Inference</td>
<td>Erik</td>
<td>11 November</td>
<td>16-18</td>
<td>U29A</td>
</tr>
<tr>
<td>11</td>
<td>Matching and Propensity Scores</td>
<td>Erik</td>
<td>18 November</td>
<td>16-18</td>
<td>U29A</td>
</tr>
<tr>
<td>12</td>
<td>Regression-Discontinuity Designs</td>
<td>Erik</td>
<td>25 November</td>
<td>16-18</td>
<td>U29A</td>
</tr>
<tr>
<td>13</td>
<td>Instrumental Variable Regression</td>
<td>Erik</td>
<td>2 December</td>
<td>16-18</td>
<td>U29A</td>
</tr>
<tr>
<td>14</td>
<td>Factor analysis</td>
<td>Robert</td>
<td>9 December</td>
<td>16-18</td>
<td>U29A</td>
</tr>
<tr>
<td>15</td>
<td>Communicating results</td>
<td>Robert</td>
<td>16 December</td>
<td>16-18</td>
<td>U29A</td>
</tr>
</tbody>
</table>
4 Overview of labs

Students are required to bring their own laptops to the lab sessions. All details related to the practical parts (data and script) will be uploaded to Blackboard.

<table>
<thead>
<tr>
<th>No.</th>
<th>Topic</th>
<th>Date</th>
<th>Time</th>
<th>Room</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Linear regression</td>
<td>10 September</td>
<td>14-16</td>
<td>U49C</td>
</tr>
<tr>
<td>2</td>
<td>Data simulation and binary logistic regression</td>
<td>24 September</td>
<td>14-16</td>
<td>U49C</td>
</tr>
<tr>
<td>3</td>
<td>Ordered and multinomial logistic regression</td>
<td>1 October</td>
<td>14-16</td>
<td>U49C</td>
</tr>
<tr>
<td>4</td>
<td>Data over time &amp; space</td>
<td>22 October</td>
<td>14-16</td>
<td>U49C</td>
</tr>
<tr>
<td>5</td>
<td>Hierarchical models</td>
<td>5 November</td>
<td>14-16</td>
<td>U49C</td>
</tr>
<tr>
<td>6</td>
<td>Experimental data I</td>
<td>19 November</td>
<td>14-16</td>
<td>U49C</td>
</tr>
<tr>
<td>7</td>
<td>Experimental data II</td>
<td>26 November</td>
<td>14-16</td>
<td>U49C</td>
</tr>
</tbody>
</table>

5 Overview of mandatory assignments

The mandatory assignments are basically graded homeworks and all the information necessary to complete the assignment will be supplied on the hand-out date. Assignments will be handed out through Blackboard, are due the following week and must be submitted electronically through Blackboard, using the examination number of the student as identifier.

<table>
<thead>
<tr>
<th>No.</th>
<th>Topic</th>
<th>Hand-out date</th>
<th>Due date</th>
<th>Feedback by</th>
<th>Resubmission by</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Linear regression</td>
<td>10 September</td>
<td>16 September</td>
<td>23 September</td>
<td>30 September</td>
</tr>
<tr>
<td>2</td>
<td>Logistic regression</td>
<td>1 October</td>
<td>7 October</td>
<td>21 October</td>
<td>28 October</td>
</tr>
<tr>
<td>3</td>
<td>Data over space and/or time</td>
<td>5 November</td>
<td>11 November</td>
<td>18 November</td>
<td>25 November</td>
</tr>
<tr>
<td>4</td>
<td>Experimental data</td>
<td>26 November</td>
<td>2 December</td>
<td>7 December</td>
<td>10 December (!)</td>
</tr>
</tbody>
</table>

Please consult the above table very carefully and plan accordingly. All assignments are designed to evaluate the progress towards the learning outcomes stipulated in the course description.
6 Readings for lectures

The curriculum consists of a number of texts (books, chapters from books, articles and online documents), and the curriculum for each lecture is described below. Texts marked with an asterisk (*) are found in the compendium for the course, and you have to download (or access) the rest via the university library’s website (the exception being texts with a double asterisk (**) which are made available for copying before the lecture.)

Note that books from the Quantitative Applications in the Social Sciences series by Sage (or Little Green Books) - such as Davis (1985) and Fox (1991) - are available online through the university library collection, and hence they are not included in the compendium.

1. Introduction and refresher to basic statistics (Sune)

Review literature and notes from your BA course in statistics. Examples of textbooks:


Make also sure that Stata and/or R are installed on your computer.

2. Refresher on linear regression (Robert)

Required readings:


Recommended readings:

Jasso, G. (1986). 'Is it outlier deletion or is it sample truncation? Notes on science and sexuality’. American Sociological Review, 738–742. (Jasso’s discussion with Kahn and Udry is an entertaining discussion of whether or not to include outliers in an empirical analysis.)
Kahn, J. R. & Udry, J. R. (1986). "Marital coital frequency: Unnoticed outliers and unspeci-
fied interactions lead to erroneous conclusions". American Sociological Review, 734–737.


3. How to study causality in non-experimental studies (Sune)

Required readings:
University Press. Pages 77-83.

Pages 89-94.


4. Binary logistic regression (Sune)

Required readings:
Pages 483-495.

* Rose, L. E. (2002). "Municipal size and local nonelectoral participation: findings from
Denmark, the Netherlands, and Norway". Environment and Planning C: Government and
Policy, 20, 1-23.

Recommended readings:
Cambridge: Cambridge University Press. Chapter 5. (for labs, good source for logistic
regression in R)

labs, good source for logistic regression in Stata)

5. Ordered and multinomial logistic regression (Sune)

Required readings:
Pages 496-503.

of Common Market Studies, 50(1), 88-105.

havior. Evidence from Germany". Political Psychology, 28(4), 471-492.
6. Time & space: Models for independent pooled cross sections and panel data (Sune)

Required readings:


7. Time & space: Fixed effects and random effects models (Sune)

Required readings:


8. Hierarchical models I (Robert)

Required readings:


Recommended readings:


9. Hierarchical models II (Robert)

Required readings:


Recommended readings:


10. Experiments and Causal Inference (Erik)

Required readings:


Recommended readings:


Recommended readings in Danish


11. Matching and Propensity Scores (Erik)

Required readings:

* Gelman, A. & J. Hill (2007). Data Analysis Using Regression and Multilevel/Hierarchical Mod-


**Recommended readings:**


**Recommended readings in Danish**


**12. Regression-Discontinuity Designs (Erik)**

**Required readings:**


**Recommended readings:**


Recommended readings in Danish


13. Instrumental Variable Regression (Erik)

Required readings:


Recommended readings:


Recommended readings in Danish:


14. Factor analysis (Robert)

Required readings:

TBA
Recommended readings:

TBA

15. Communicating results (Robert)
