Digital Social Research: Statistics Core

<table>
<thead>
<tr>
<th>Academic Year</th>
<th>2017-18, Michaelmas Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day and Time</td>
<td>Thursdays, Weeks 1-8, 9:00 – 11:30</td>
</tr>
<tr>
<td>Location</td>
<td>Lecture Theatre, Ship Street Centre, Ship Street, Oxford, OX1 3DA</td>
</tr>
<tr>
<td>Course Providers</td>
<td>Dr Grant Blank, Oxford Internet Institute, <a href="mailto:grant.blank@oii.ox.ac.uk">grant.blank@oii.ox.ac.uk</a></td>
</tr>
<tr>
<td></td>
<td>Dr Scott Hale, Oxford Internet Institute, <a href="mailto:scott.hale@oii.ox.ac.uk">scott.hale@oii.ox.ac.uk</a></td>
</tr>
<tr>
<td></td>
<td>Dr Jonathan Bright, Oxford Internet Institute</td>
</tr>
<tr>
<td>Teaching Assistant(s)</td>
<td>Thomas Vogl, DPhil Student, Oxford Internet Institute</td>
</tr>
<tr>
<td></td>
<td>Julia Lefkowitz, DPhil Student, Oxford Internet Institute</td>
</tr>
<tr>
<td>Prerequisites</td>
<td>N/A</td>
</tr>
</tbody>
</table>

For comments and enquiries, please e-mail one of the course co-convenors: grant.blank@oii.ox.ac.uk or scott.hale@oii.ox.ac.uk.

Background
Digital Social Research: Statistics Core is one element of the six elements that make up Digital Social Research. A summary of the structure and assessment of Digital Social Research has been included as a separate coversheet.

Overview of Digital Social Research: Statistics Core
Digital Social Research: Statistics Core introduces students to statistics for the social sciences, with an emphasis on application to research on the Internet and society. This course introduces students to the most important types of quantitative social science data: discrete, counted data and continuous data. The course is based on four themes.

1. The focus is on selection and interpretation of statistical techniques, reaching sensible conclusions, figuring out causality, and making decisions, combining graphical, exploratory, and confirmatory approaches in ways that suggest how to improve our understanding in the light of data.

2. This requires hands-on work with data through statistical software; in this course we will use the software package R. All calculations are done using the software, not using hand calculations or calculators. Class lectures and discussions involve use of statistical software. Formative assignments require intensive statistical computing.

3. A hands-on approach to understanding data directs attention away from the formal, theoretical, mathematical properties of statistical estimators, which is sometimes an emphasis in statistics classes. The course emphasizes ability to interpret the substantive significance of graphical and numerical computer output.

4. The strong emphasis on data and use of software leads to a final theme: Data almost never come to researchers in a form appropriate for analysis; they must be converted into a suitable form.
Thus the course teaches common forms of data manipulation and these are incorporated into the formative assignments.

Teaching Arrangements
The course is team taught during Michaelmas term. There is one, two-hour class each week. The format of the sessions includes lectures, student discussions and group work. All students are expected to attend all these sessions. In addition, a weekly two-hour surgery will be led by the teaching assistant to assist students in completing formative assignments. Times and places for the surgery will be set during the first class and posted on the course website on Plato. The surgery is optional but strongly recommended.

Note
Students should note that over the course of the year, small changes may be made to the content, dates or teaching arrangements set out in this reading list, at the course provider’s discretion. These changes will be communicated to students directly and will be noted on the course website.

Assessment
In Digital Social Research: Statistics Core students are assessed on the basis of one summative assessment and weekly formative assignments during Michaelmas term. (Please note that the assessment for this course is different for DPhil students. DPhil students should please refer to the Graduate Studies Handbook for guidance).

Formative Assignments
Weekly formative assignments will be given during class. They will be due later in the week and will be returned the week after they are handed out.

Summative Assessment
One three-hour written exam in Week 9 of Michaelmas Term. The exact date and time are not set by the OII, but will be announced to students as soon as they are set by the Examination Schools. This is usually done a few weeks before the end of the term. The exam can be any time during Week 9, so do not make other travel plans during that week until the date has been finalized.

Any student failing this assessment will need to follow the rules set out in the OII Examining Conventions regarding re-sitting failed examinations.

Assignments
Formative assignments are due before the end of the week when they are handed out. All assignments should follow the format described below.

Computer literacy is a requirement for effective participation in the OII. This is a computer-intensive course. You will need a modern, stable computer running Windows or a Mac, or plan to do all your assignments on the OII computers. You may need to work both at home and at the OII, and transfer files back and forth using a a USB flash drive, via email or a cloud space provider. Take particular care that the disk or drive is no longer being written to when you remove it. Be sure to back up all your class work daily to a removable storage medium. The standard is two full backups of all work. If you have a problem on your personal computer, take your backup and work at the OII.

Computer problems, errors, or crashes are not an acceptable excuse for late or incomplete assignments.

Statistics Assignment format
Formative assignments will be submitted via Plato and instructions will be provided at the first class.
All statistical and graphical output must include the commands that generated the output. Even if the output is correct, you will not receive full credit if you do not include the commands. Output and commands should be single-spaced in a monospace font like Courier New or Consolas. Never use a variable-width font (like Times New Roman) for output or commands. Use a font size so that lines fit the width of a page with 2.5 cm margins. Output should be placed shortly after it is first mentioned.

Key Readings
A reading list is given below for each class. Those items marked with an asterisk (*) are essential reading and MUST be read by all students in preparation for the class. Items which are not marked with an asterisk are additional readings. There are two key texts used throughout the course. They are:

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>Edition</th>
<th>Publisher</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knoke, David Bohrnstedt, George W. Mee, Alisa P.</td>
<td>Statistics for social data analysis</td>
<td>2002</td>
<td>Belmont, CA: Thompson</td>
</tr>
</tbody>
</table>

These and all the readings listed below will be available on reserve at the OII Library at 1 St Giles.

Other Useful Statistics Readings

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>Edition</th>
<th>Publisher</th>
</tr>
</thead>
</table>

Course Topics

0 Distributions and their description
1 Social statistics and statistical software
2 Probability and sampling distributions
3 Testing hypotheses for 1 variable
4 Testing hypotheses for 2 variables
5 Two-way tables
6 Summarizing bivariate relationships
7 Statistical controls: The elaboration model
8 Introduction to linear regression
Week 0: Distributions and their description. Descriptive statistics and graphics.
Tutor(s): Dr Grant Blank

Note: See induction week schedule

This optional class re-acquaints students with less statistics background with different types of distributions of quantitative data and demonstrates how distributions can be summarised using statistics and graphs. This is fundamental material because later work is based on knowledge of these distributions and their summary statistics. This class is optional for students with a stronger background in statistics.

  - Chapter 2

  - Chapter 3

--

Week 0: Introduction to R
Tutor(s): Dr Scott Hale, Dr Jonathan Bright

Note: See induction week schedule

This class introduces the R statistical programming language that will be used in the Statistics Core course. All students should attend, even those who have had previous experience with R. Laptops with the current version of R and RStudio installed are required. Please follow the instructions you were sent to install R and RStudio prior this class. There will limited time scheduled directly before this class to sort out any remaining installation issues.

Week 1: Social Statistics and Statistical Packages.
Tutor(s): Dr Grant Blank, Dr Scott Hale

This introductory class serves dual purposes. The first goal is to familiarize the student with the logic of statistical inquiry; that is, how can we structure an analysis of quantitative data so that we can effectively answer meaningful research questions?

  - Chapter 1

  - Chapter 1

Tutor(s): Dr Scott Hale

The session will introduce principles of probability theory and show how they apply to some key probability distributions such as the normal distribution and the binomial distribution. It will cover the sampling distribution of estimators, a fundamental concept necessary for understanding statistical inference.

* Knoke, David Bohrnstedt, George W. Mee, Alisa P.  
  - Chapters 3.1-3.5

Agresti, Alan Finlay, Barbara  
  - Chapters 2 & 4

Week 3: Testing hypotheses for 1 variable. Null hypothesis. Confidence intervals. Type I and Type II errors.

Tutor(s): Dr Jonathan Bright

This session provides an overview of key statistical concepts involved in hypothesis testing with one variable. The session explains the relevance of null hypotheses and confidence intervals, and ways to avoid inference errors.

* Knoke, David Bohrnstedt, George W. Mee, Alisa P.  
  - Chapters 3.6, 3.8-3.10

Agresti, Alan Finlay, Barbara  
  - Chapters 5 & 6

Cohen, Jacob  
  “The earth is round (p<0.05).” 1994. American Psychologist, 49(12): 997-1003.


Tutor(s): Dr Jonathan Bright

This session continues the discussion about hypotheses testing for cases when two variables are involved. The session explains the meaning of standard errors, and presents different statistical tests to assess the significance of associations between the variables.

* Knoke, David Bohrnstedt, George W. Mee, Alisa P.  
  - Chapters 3.7, 3.8, 4.3

Agresti, Alan Finlay, Barbara  
  - Chapter 7

Tutor(s): Dr Grant Blank

This session introduces analysis of two categorical variables and defines statistical independence. It covers the concept of independence, chi-squared tests of independence, percentaging and interpretation of percentage tables.

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Reference</th>
</tr>
</thead>
</table>
- Chapters 5.1, 5.2 |
- Chapter 8 |

**Week 6: Characteristics of an association: Proportional reduction in error (PRE) statistics**

Tutor(s): Dr Grant Blank

This session moves beyond significance testing to describe other characteristics of an association. It defines proportional reduction in error (PRE) statistics and describes the statistics appropriate for summarizing the association between pairs of variables.

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Reference</th>
</tr>
</thead>
</table>
- Chapter 5.3 |
- Chapter 8 |
- pp. 436-441 |

**Week 7: Statistical controls: The elaboration model. Partial tables and partial coefficients**

Tutor(s): Dr Grant Blank

This session introduces the use of statistical controls via the elaboration model, a three-variable model of the logic of causal analysis in percentage tables. The model uses a third variable as a statistical control to clarify the relationship between two other variables.

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Reference</th>
</tr>
</thead>
</table>
- Chapters 7.1-7.2 |
- Chapter 15 |

Tutor(s): Dr Grant Blank and Dr Scott Hale

This lecture introduces the student to basic bivariate regression or "correlation", and shows how to derive and interpret the numerical measures that are reported from ordinary least squares regression. Concepts covered include regression coefficients, tests for coefficients, residuals, goodness-of-fit, and scatterplots.

* Knoke, David  
  Bohnstedt, George W.  
  Mee, Alisa P.  

- Chapter 6

Agresti, Alan  
Finlay, Barbara

- Chapter 9