STATISTICS AND DATA ANALYSIS I: INTRODUCTION

ICPSR SUMMER PROGRAM, 2019
June 25-July 19, 3:00-5:00pm

COURSE INSTRUCTOR:

Niccole Pamphilis
University of Glasgow
Niccole.Pamphilis@glasgow.ac.uk

TEACHING ASSISTANT:

Stephen El-Khatib
UC, Riverside
elkh001@ucr.edu

Aurea S. Bolaños Perea
Chico State Department of Political Science
abolanosperea@mail.csuchico.edu

AIMS AND OBJECTIVES:

This workshop provides a basic introduction to statistics, probability, and data analysis. Topics include data acquisition/management, classification, and summarization; basic probability; exploration of common distributions used in statistics; along with confidence intervals and hypothesis testing.

The format for this workshop consists of daily lectures supplemented by sessions in the computing labs. There will be frequent assignments in order to give participants hands-on experience with the methods and techniques covered in the class. Most assignments will involve the use of statistical packages, which will be available for use in computer labs. Participants will be given the opportunity to do their work using the statistical software packages that are most common in their respective academic disciplines, including SPSS, R, and Stata.

The only prerequisite for this course is familiarity with basic algebra. Participants who need a review of basic algebraic concepts are advised to enroll in the ICPSR Summer Program lecture course Mathematics for Social Scientists I along with this workshop.

TEXTBOOKS:

Required:

**Recommended:**


As a companion book to help with coding in R I would recommend the following book:

Fogarty, Brian. 2018. *Quantitative Social Science Data with R: An Introduction*. SAGE Publications Ltd

**COURSE STRUCTURE:**

**Readings:**

Daily readings are outlined below. It is recommended that readings are completed prior to our meetings to allow for questions to be asked during class. Readings are based on the textbook listed above. Though readings from other texts may be useful for additional details or alternative presentations of the material. Copies of books are available in the ICPSR library in Helen New Berry, which can be checked out for short periods of time.

**Daily Meetings:**

Lectures will take place daily and will focus on the material noted in the readings below.

**Labs:**

Lab sessions will be held to provide you with an opportunity to become familiar working with the statistical software used in your field. Lab sessions will also provide an opportunity to work through assignments and ask questions. Lab times and locations will be provided during the 1st week of classes.

**Assessment:**

Four pieces of assessment will be graded and used to determine final grades for the course. The assessments will increase in difficulty as the course progresses. All four pieces will require the use of statistical software, so be sure to leave time to access the computer labs (if needed).
Each person seeking a grade must submit their own piece of work assessment, though participants are encouraged to work together on assignments.

Each piece must be:
- Typed, organized, and stapled
- Answers must be presented in a clearly labelled manner
- Include an appendix with the code you used to produce your final results
- Clearly label your final answers
- Any graphs and visuals that are used should be placed in texts as part of your answer and not as a separate attachment
- Make sure to answer all parts of a question (i.e., prompts to interpret/ explain your answer)

The fourth and final piece of assessment will be a presentation; participants in the course will be encouraged to work in groups for this piece of assessment, though participants who prefer to work independently may discuss this with the instructor during office hours. Information on this piece of assessment will be distributed during week two and the project will be due on the final day of the course.

**GRADING:**

Your final grade will be based on four pieces of assessment. Assessments vary in difficulty and are thus weighted differently to reflect this:

- Assessment 1: 15%
- Assessment 2: 25%
- Assessment 3: 35%
- Project and Presentation: 25%
## SYLLABUS AND READINGS:

### WEEK 1

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Required Reading</th>
<th>Recommended Reading</th>
<th>Suggested Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 June 2019</td>
<td><strong>Introduction:</strong> What are data and statistics</td>
<td>Sirkin Chapter 1</td>
<td>Herzog Chapters 1 and 3</td>
<td>Weiss 1.1 and 2.1-2.2</td>
</tr>
<tr>
<td>26 June 2019</td>
<td><strong>How do we measure the world around us</strong></td>
<td>Sirkin Chapter 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 June 2019</td>
<td><strong>Turning Data into Datasets:</strong> measurement and coding</td>
<td>Herzog Chapter 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28 June 2019</td>
<td><strong>Concept to Operationalisation</strong></td>
<td>Sirkin Chapter 3</td>
<td></td>
<td>LAB DAY 1, Assignment 1</td>
</tr>
</tbody>
</table>

### WEEK 2

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Required Reading</th>
<th>Recommended Reading</th>
<th>Suggested Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>01 July 2019</td>
<td><strong>What are we seeing? What does it tell us:</strong> Univariate Statistics</td>
<td>Sirkin Chapter 4 and 5</td>
<td>Weiss Chapter 3</td>
<td></td>
</tr>
<tr>
<td>02 July 2019</td>
<td><strong>Visualising data:</strong> Graphics and Tables</td>
<td>Herzog Chapter 11 and 12</td>
<td>Barholomew Chapter 1</td>
<td>Assignment 1 DUE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Suggested Reading:</td>
<td>Weiss Chapter 1.3-1.5; 2.3-2.6</td>
<td></td>
</tr>
<tr>
<td>03 July 2019</td>
<td><strong>Generating Summary Statistics and Graphics</strong></td>
<td></td>
<td></td>
<td>LAB DAY 2, Assignment 2</td>
</tr>
<tr>
<td>04 July 2019</td>
<td><strong>Probability:</strong> How do we know if differences matter</td>
<td></td>
<td></td>
<td><em>NO CLASS</em></td>
</tr>
<tr>
<td>05 July 2019</td>
<td></td>
<td>Sirkin Chapter 7</td>
<td>Barholomew Chapter 8 and 9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Suggested Reading:</td>
<td>Weiss Chapter 4</td>
<td></td>
</tr>
</tbody>
</table>
### WEEK 3

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Reading Details</th>
<th>Notes</th>
</tr>
</thead>
</table>
| 08 July 2019  | **Standardising Data:** The Normal Distribution and Z-scores | Required Reading: Sirkin Chapter 8 (through Normality Assumptions)  
Recommended Reading: Bartholomew Chapters 3 and 4  
Suggested Reading: Weiss Chapter 6.1-6.3 and Chapter 7 | Assignment 2 DUE |
| 09 July 2019  | **T-Distribution:** One-Sample T-tests, Critical Values Approach | Required Reading: Sirkin Chapter 8 (One-Sample T-tests)  
Recommended Reading: Weiss Chapter 8  
Suggested Reading: Bartholomew Chapter 10 | Assignment 2 DUE |
| 10 July 2019  | **Determining Differences by groups:** Probability Approach | Required Reading: Sirkin Chapter 9  
Recommended Reading: Weiss Chapter 9.3-9.6 and Chapter 10.1-10.2 | |
| 11 July 2019  | **Differences beyond the Mean:** Variances and Confidence Intervals | Required Reading: Sirkin Chapter 10  
Recommended Reading: Weiss Chapter 10.3 | |
| 12 July 2019  | **Testing:** using R; SPSS; or Stata | | LAB DAY 3  
Assignment 3 Given |

### WEEK 4

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Reading Details</th>
<th>Notes</th>
</tr>
</thead>
</table>
| 15 July 2019  | **Understanding Relationships Part I:** Discrete variables | Required Reading: Sirkin Chapter 6 and 12  
Recommended Reading: Weiss Chapter 13  
Suggested Reading: Bartholomew Chapter 10 | |
| 16 July 2019  | **Measuring relationships:** Strength and Direction | Required Reading: Sirkin Chapter 11  
Recommended Reading: Weiss Chapter | Assignment 3 DUE |
17 July 2019  *Understanding Relationships Part II*: Continuous variables
   Recommended Reading:  
   Weiss Chapter 14

| 18 July 2019 | *Putting it all together* | LAB DAY 4 |
| 19 July 2019 | *Group Presentations*     | Presentations DUE |