MSc in Official Statistics

STAT6103: Statistical Computing – 2016

Outline

Location

University of Southampton, Building 39, Social Statistics Research Centre
Room 3013 (Seminar Room)

Lecturer

Dr. Pedro Silva
Building 39 - Short term visitors room
Phone: +5521 21424957
Email: pedronsilva@gmail.com

Aims of Module

This course aims to prepare participants to input, verify, organise, modify, combine, analyse and present data using a range of computing and statistical methods implemented in the general purpose statistical package SAS. It also aims to introduce some basic ideas of statistical computing, such as numerical methods used to obtain summary statistics, iterative methods for solving equations, and simulation. Focus is not on detailed explanation of statistical analysis methods, but several of these methods will feature as examples during the course.

Learning Outcomes

By the end of this course you will be able to:

- enter and manipulate data within SAS;
- perform standard statistical analyses using SAS and obtain the resulting output;
- find additional information on how to perform advanced statistical analyses using SAS and then undertake such analyses.

Key Skills

Proficient statistical computing.
Ability to work independently in retrieving, organising and analysing data from surveys, censuses, and administrative sources.
Though the course is based on SAS statistical software, the skills developed are transferable to other statistical software such as SPSS, STATA, R or S-PLUS with little effort.

Resources

You will require access to the S3RI's computer lab in Building 39 and to SAS software, which is available on the University's computer workstations. This software is currently available for download to your own computer for use with your studies.

Timetable

Start	End	Activity
		Monday - 8 February
09:30	10:00	Registration
10:00	10:30	0. Course introduction
10:30	12:00	1. Introduction to SAS
12:00	13:00	Lunch
13:00	14:30	2. SAS Basics
14:30	15:00	Break
15:00	16:00	Computer workshop 1
16:00	17:30	3. SAS Data Management
		Tuesday - 9 February
09:30	11:00	4. SAS Data Management
11:00	11:30	Break
11:30	12:30	Computer workshop 2
12:30	13:30	Lunch
13:30	15:00	5. SAS Procedures for Statistical Analysis: Linear Regression & More
15:30	16:00	Break
16:00	17:00	6. SAS Procedures for Statistical Analysis: Linear Regression & More
		Wednesday - 10 February
09:30	11:00	7. Advanced Data Management
11:00	11:30	Break
11:30	13:00	8. Interactive statistical analysis: SAS IML Studio
13:00	14:00	Lunch
14:00	15:30	Computer workshop 3
15:30	16:00	Break
16:00	17:00	9. Statistical computing in SAS
		Thursday - 11 February
09:30	11:00	10. SAS IML – Interactive Matrix Language
11:00	11:30	Break
11:30	13:00	11. SAS Macros
13:00	14:00	Lunch
14:00	14:30	Discussion of coursework
14:30	15:30	Computer workshop 4
15:30	16:00	Break
16:00	17:00	12. SAS SQL
		Friday - 12 February
09:30	11:00	13. SAS SQL
11:00	11:30	Break
11:30	12:30	Computer workshop 5
12:30	13:00	Course evaluation & finish

Teaching and Learning Methods

The module is based on lectures, computer lab sessions and individual study.
Key concepts and ideas are presented in the lectures.
Computer lab sessions are used to enable students to gain practical experience using the statistical software used in the module and carry out exercises designed to reinforce the ideas presented in the lectures.
Individual study is also proposed to enable students to consult some of the additional literature made available for the module, thus increasing the breadth and depth of coverage for selected topics.

Syllabus

Introduction to SAS

Overview; Running SAS for Windows; SAS menu and windows; SAS Editor; SAS Log; SAS Output; SAS Explorer; SAS Results; SAS Help; Reading data into SAS; Saving data; Printing;

SAS Basics

Basics of the SAS language: DATA and PROC steps. SAS data sets: data and metadata portions. Variable types: numeric and character. Documenting your data: naming, labelling and formatting.

SAS Data Management

Checking data; Sorting data; Basic data management – creating new variables, recoding variables, using functions; Reshaping data; Combining datasets; Extracting parts of data (case and variable selection); Summarising data; Running analyses on subsets of cases: (a) analysing two or more groups separately – by statement; (b) case selection – filtering.

SAS Procedures for Statistical Analysis: Linear Regression & More

Model and other specifications required for analysis; description of main computations performed by procedure; understanding the core output and results for each procedure; getting help and details about computations performed; computational methods used; categorical versus numeric predictor variables; useful output datasets produced; checking assumptions; simple graphs for diagnostics.

Advanced Data Management

Compressed datasets. Reading data from other statistical sources or from non-standard sources and file formats; writing and exporting data for use by other software; report writing.

Interactive statistical analysis: SAS Insight

Interactive statistical analysis using SAS IML Studio: opening datasets; exploratory graphs; model fitting; data transformations; tools for case identification and highlighting; fitting Generalized Linear Models and exploring diagnostic plots and tables.

Statistical computing in SAS

Numerical methods for summary statistics: means, variances, regression coefficients.
Random numbers and simulation: simulating probability distributions from models.
Sampling and simulation from empirical distributions by bootstrapping.
Solving equations using iterative methods: Newton-Raphson method.

SAS IML – Interactive Matrix Language

SAS data step programming versus SAS IML; SAS/IML Matrices: types of matrices (numeric vs character), assigning values to matrices, printing / viewing contents of matrices; Elements of SAS/IML Language: modules – definition and execution, control statements, command statements, call statements and functions, assignment statements and operators; Subscripting Operations: selecting submatrices, performing row and column reduction; Processing SAS Data Sets: creating matrices from SAS data sets, creating SAS data sets from matrices, adding row and column labels to matrices; Developing a SAS/IML Program: creating modules as subroutines and functions, iterative processing, conditional processing.

SAS Macros

SAS macro facility; macro variables; macros; macro parameters; repetition and conditional generation of statements; interface with data; data dependent code generation.

SAS PROC SQL

Introduction to the relational data model. SQL language. SAS PROC SQL description and comparison with SAS DATA step. Retrieving data from a single table: sorting, filtering, modifying, summarizing and grouping data. Retrieving data from multiple tables. Creating and updating tables.

Indicative reading list

Delwiche, L. D. and Slaughter, S. J. (2008). The Little SAS Book: A Primer. SAS Institute Inc.

Bailer, A. J. (2010). Statistical Programming in SAS. SAS Institute Inc.

Spencer, N. H. (2004). SAS Programming: the One-Day Course. Chapman & Hall CRC.

Der, G. and Everitt, B. S. (2008) A Handbook of Statistical Analyses Using SAS, third edition. CRC Press.

Cody, R. P. and Smith, J. K. (2005). Applied Statistics and the SAS Programming Language, fifth edition.

Blackboard Site

All students are automatically enrolled on the 'MSc in Official Statistics Student Area' blackboard site. If you do not have access to the site, please let your tutor know so that she can enrol you. There is also a discussion board on Blackboard so that you can communicate with your fellow students.

To submit an electronic version of the coursework assignment for the 'Turnitin' plagiarism software, you must access the specific module on Blackboard: '12-13-Statistical Computing-30179 ' and upload the assignment through the 'Assignments' folder according to the instructions listed below. Only students submitting coursework assignments will gain access to this site. If you don't have access to the site, please let your tutor know.

NOTE: the uploading of an electronic version of the coursework assignment does NOT replace the two hard-copies of the assignment that are to be sent to Claire Woodley using the coursework hand-in form available at:
http://www.southampton.ac.uk/demography/postgraduate/taught_courses/msc_official_statistics.page?

Instruction for uploading assignments:
To submit coursework electronically via the TurnitinUK plagiarism device, login to the Blackboard unit for this module and select the Assignments link from the left-hand menu. Find the 'assignment' and click View/Complete. There will be a series of screens to complete, and full instructions. Please note the following points:

1. During the submission process you will see a receipt screen confirming that your assignment has been submitted. This will show the title of your paper and will also provide a Paper ID. You should make a note of this Paper ID for future reference in case of any problems.
2. Following your submission through Blackboard you will receive an email confirmation that your coursework has been received. This usually takes about an hour but may take up to 24 hours depending on how busy the system is. You should retain a copy of this email which will act as a receipt for your electronic submission. If you do not receive this email then you should contact the module coordinator as soon as possible.

Assessment

Assessment is 100% coursework. The deadline for submission of the coursework is **Friday, 18 March 2016** with the post mark required by the preceding Thursday. Please send 2 hard copies of the assignment to:

Moffstat programme administrator
University of Southampton
Building 58, Room 2037
Highfield, Southampton
SO17 1JB

Students who fail to meet the deadline will receive a 5% reduction in their mark for every working day up to 5 working days. After 6 working days, the standard policy is to award zero for the coursework assignment.

Procedure for coursework extensions

If students want to request an extension, they must have medical or personal circumstances to justify the late submission of assessed coursework (medical evidence must be substantiated by a doctor's note). They should initially fill out a coursework extension request form available at the above website and send it to their Personal Tutor, Prof. Paul Smith, who will, where appropriate, agree to the extension of the deadline after discussion with the module co-ordinator and examination officer. Under no circumstances will extensions be allowed beyond a period of two weeks.

Resit arrangements

No mark below 35% is allowed for any module in the programme. All students receiving below the minimum mark will have to re-sit the module according to School of Social Sciences policy.

Academic integrity

The University places the highest importance on the maintenance of academic integrity in the conduct of its affairs, and the Academic Integrity Statement for Students can be found in the University Calendar available online at http://www.calendar.soton.ac.uk/sectionIV/academic-integrity-statement.html. Please familiarise yourself with what is expected of you in this regard by reading through this statement. Your attention is drawn particularly to Appendix 1, which outlines those things which you must seek to avoid, including cheating and plagiarism. A very useful set of interactive guides is available at http://www.studyskills.soton.ac.uk. These aim to help you gain a better understanding of academic integrity and develop your skills so that your assessed work does not accidentally plagiarise the work of others.

You must take particular care in using sources in essays and in your dissertation. Remember that plagiarism includes not only verbatim copying but also direct paraphrasing of a source. Verbatim quotes from a source should always be in quotation marks, with the source indicated, and should be used only occasionally in an essay or other report. Detailed advice on appropriate referencing in essays and dissertations is given in the Division of Social Statistics "Guidelines on writing essays".

Students are encouraged to discuss and exchange ideas, since this is an important part of the educational process. However, it is not acceptable that you read and gain ideas for your coursework from another student's finished work. Copying includes using another student's computer program, output or graphics. If academic integrity is deemed to have been breached, there are a range of penalties that may be applied.

If you are unsure about what is and is not permitted, ask - we will be happy to explain and discuss.

Troubleshooting

Please get in touch with your personal tutor, Prof Paul Smith, should any difficulties arise with this module.