_____

Module Outline

1.   Aims and Objectives

The aim is to provide you with an overview of the statistical methods for linear and logistic regression. The module will also involve some practical data analysis using SPSS.

2.   Learning Outcomes

On successful completion of this module you should understand:

- the different techniques involved in fitting regression models;
- how to fit and interpret these models applied to typical datasets arising in Official Statistics.

3.   Key Skills

By the end of this module you will have developed your skills in:

- statistical computing;
- statistical report writing.

4.   Staff

Lecturers

Prof. Peter G.M. van der Heijden (Division of Social Statistics and Demography)

Building 39, Room 3017

Email: p.g.van-der-heijden@soton.ac.uk

Workshop Assistance

Vasiliki Koutra, a current PhD student from the Division of Social Statistics and Demography, will be helping out at the computer workshops providing advice on how to use SPSS and how to interpret the output.

_____

5.   <u>**Teaching and Learning Methods**</u>

The course consists of a series of lectures and integrated computer workshops. The lectures will cover the theoretical aspects of the course, and the computer workshops will involve some data analysis using the techniques introduced in the lectures.

If we forget to mention this during the lectures, the course content is partly based on materials of past and present Southampton colleagues: David Holmes, Fiifi Amoako-Johnson, Natalie Shlomo, Peter Smith, Moshe Feder, John 'Mac' McDonald, James Brown and Nyovani Madise.

6.   <u>**Module Structure**</u>

The module covers the following topics:

- **Linear regression**: scatterplots and correlation, simple linear regression, least squares estimation, ANOVA, residual analysis, statistical testing, multiple regression, model building, handling categorical variables, interactions, transformations, polynomial regression.

- **Introduction to logistic regression**: binary response variables, the basic model, interpreting the model parameters, assessing model fit, model building.

7.   <u>**Assessment Methods**</u>

For those wishing to earn University credits, the assessment will be 100% coursework. This coursework will involve performing a number of analyses with models for continuous and binary responses, and writing an academic report on your findings.

Your submission should be posted to the MOffStat Programme Administrator,
Professional Development Unit, Faculty of Social and Human Sciences, Room 2037, Building 58, University of Southampton, Southampton, SO17 1BJ. The deadline for submission of the assessment is **Tuesday 1st of March 2016**, with the post mark required by the preceding Monday. Send <u>2 copies</u> and attach the coursework hand-in form, available at http://www.southampton.ac.uk/demography/postgraduate/taught_courses/msc_official_statistics.page? It is policy that all assignments must also be submitted electronically via Turnitin. Please note that this is <u>in addition</u> to submission of a hard copy.

_____

Electronic submission of coursework

- To submit coursework electronically via the TurnitinUK plagiarism device, login to Blackboard, select the specific unit for this module (15-16-Regression Modelling-28305), and select the Assignments link from the left-hand menu. Find the coursework and click View/Complete. There will be a series of screens to complete, and full instructions on how to do this are given on the iSolutions webpages at: http://www.southampton.ac.uk/isolutions/computing/elearn/blackboard/student/studentplagiarism.html.

- When you submit electronically you will receive a **submission ID number**, which you **MUST** include on the front cover of your hardcopy submission. You will therefore need to submit electronically **BEFORE** you post your hard copy.

- Turnitin is a plagiarism detection tool, which checks your work against electronic sources and other submissions for the same assignment. You will only be able to submit your assignment once to Turnitin.

- You will also receive an email (maybe up to a day later, but most likely within a few minutes) confirming that you have submitted electronically. You should retain a copy of this email which will act as a receipt for your electronic submission. If you do not receive a submission ID number or an email it means that you have not submitted. If this is the case **you will be penalised**. If you do not receive this email then you should contact the lecturers as soon as possible.

Penalty for late submission

The penalty for unauthorised late submission of coursework is as follows:

University working days late	Mark
1	(final agreed mark)*0.9
2	(final agreed mark)*0.8
3	(final agreed mark)*0.7
4	(final agreed mark)*0.6
5	(final agreed mark)*0.5
More than 5	Zero

_____

For example, if your mark is 63% but you submit your work 3 working days late, then your final mark would be $63 \times 0.7 = 44.1\%$. Note that late submission of **_either_** the hard copy **_or_** the electronic copy will be penalised in this way.

Procedure for coursework extensions

Deadlines are made to be met. If you want to request an extension, you must have medical or personal circumstances to justify the late submission of assessed coursework (medical evidence must be substantiated by a doctor's note). You should initially fill out a coursework extension request form available at

http://www.southampton.ac.uk/demography/postgraduate/taught_courses/msc_official_statistics.page?

 and send it to the programme coordinator who will, where appropriate, authorise the extension of the deadline after discussion with the lecturers and the Examinations Officer. Under no circumstances will extensions be allowed beyond a period of two weeks.

Resit Arrangements

If you find yourself in the unfortunate position of having to resit this module, then you will be set another coursework.

8.   Feedback

'Feedback' refers to any instance in which you receive information about how well you understand the material, how successfully you are progressing in the module, or how to improve your performance. Feedback is continuous and does not refer merely to comments on your assessed work (but certainly includes that). Other instances in which you receive feedback include: seminar/tutorial discussions, question time during lectures, replies to questions during computer workshops, interaction with others in any online discussion forum, replies to email questions you send us, discussions with us during the week.

In line with University policy, feedback on your coursework assessment will be provided within four weeks of the submission deadline, and is likely to be in the form of an individual feedback sheet containing suggestions for improvement.

_____

9.   <u>**Provisional Timetable**</u>

The course will take place in B39 (SSRC). The provisional timetable is as follows:

Monday 25th January 2016

10:00 – 10:15	Introduction to the module
10:15 – 11:00	Lecture: Simple linear regression
11:00 – 11:30	*Coffee*
11:30 – 12:30	Lecture: Simple linear regression (continued)
12:30 – 13:30	*Lunch*
13:30 – 15:00*	Lecture: Confidence intervals and hypothesis testing
15:00 – 15:30	*Tea*
15:30 – 17:00	Computer Workshop 1: Correlation and simple linear regression

Tuesday 26th January 2016

09:00 – 09.45	Lecture: Model checking
09:45 – 10:45	Computer Workshop 2: Model checking
10:45 – 11:15	*Coffee*
11:15 – 12:45*	Lecture: Multiple regression and model selection
12:45 – 13:45	*Lunch*
13:45 – 15:00	Computer Workshop 3: Multiple linear regression and model selection
15:00 – 15:30	*Tea*
15:30 – 17:00*	Lecture: Categorical explanatory variables

Wednesday 27th January 2016

09:00 – 10:30	Computer workshop 4: Categorical explanatory variables
10:30 – 11:00	*Coffee*
11:00 – 12:30*	Lecture: Polynomial regression, non-constant variance and weighted least squares
12:30 – 13:30	*Lunch*
13:30 – 15:00	Computer workshop 5: Quadratic regression and non-constant variance
15:00 – 15:30	*Tea*
15:30 – 17:00*	Lecture: Introduction to logistic regression

_____

Thursday 28th January 2016

09:00 – 10:30*	Lecture: Introduction to logistic regression (continued)
10:30 – 11:00	*Coffee*
11:00 – 12:30	Computer Workshop 6: Logistic regression I
12:30 – 13:30	*Lunch*
13:30 – 15:00*	Lecture: Multiple logistic regression
15:00 – 15:30	*Tea*
15:30 – 17:00	Computer Workshop 7: Logistic regression II

Friday 29th January 2016

09:00 – 10:30	Question and answer session; module evaluation questionnaires; coursework discussion.
10:30 – 11:00	*Coffee*
11.00 -	The Lecture/Computing Room will be available for independent study.

This timetable is subject to change depending on class progress. The lecturer will be around during Friday afternoon if you have any queries about any of the module material.

10.   <u>Prerequisites/Background Reading</u>

There are no pre-requisites for this module. However,

- participants should be familiar with basic descriptive statistics and statistical inference (confidence intervals and hypothesis testing, t-tests, F-tests and likelihood ratio tests).
- no prior knowledge of regression methods will be assumed, but it may be helpful for participants to familiarise themselves with the idea of regression models. A simple starting point would be to read Chapter 13 of Diamond and Jefferies (2001) "Beginning Statistics: an Introduction for Social Scientists". Or read Chapter 7 of the book by Field (2009/2013) for something a little more advanced.
- for the part on logistic regression <u>it is important</u> to have a working knowledge on the natural logarithm function and the exponential function. See, for example, Wikipedia.

_____

11.   <u>Further Reading</u>

The copies of the lecture slides are essential reading! The following is a list of recommended text books in this area. Some of these will be available in the Moffstat library, and most, if not all, of these books are available in the University main library.

1.   Agresti, A. (2007). *An Introduction to Categorical Data Analysis* (2nd edition). Wiley. Covers logistic regression (and many other modelling methods) at an introductory level.

2.   Agresti, A. (2002/2013). *Categorical Data Analysis* (2nd / 3rd edition). Wiley. Covers logistic regression (and many other modelling methods) at an advanced level.

3.   Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis* (3rd edition). New York: Wiley.
A classic text on linear regression. It is rather mathematical though.

4.   Field, A. (2009/2013). *Discovering Statistics using SPSS* (3rd / 4th edition). London: Sage.
<u>Excellent book</u>! Written in a light-hearted (frivolous?) style, this book combines statistical theory and practical application of that theory using SPSS. Good sections on linear and logistic regression. Gentle treatment, nice examples, little maths.

5.   Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression* (2nd edition). Or Hosmer, D. W., Lemeshow, S. and Sturdivant, T, (2013). *Applied Logistic Regression* (3nd edition). Wiley.
This book is widely considered the "bible" of logistic regression analysis. A more advanced introduction to logistic regression than Agresti (2007) but not as comprehensive as Agresti (2002/2013).

6.   Kleinbaum, D. G. and Klein, M. (2010). *Logistic Regression: A Self-Learning Text*. (3rd edition). New York: Springer-Verlag.
An elementary text on logistic regression.

_____

7.   Kleinbaum, D. G., Kupper, L. L., Muller, K. E. and Nizam, A. (2013). *Applied Regression Analysis and Other Multivariate Methods* (5th edition). Pacific Grove, Ca: Duxbury Press.

An elementary text on linear regression.

8.   Kutner, M. H., Nachtsheim, C. J. and Neter, J. (2004). *Applied Linear Statistical Models* (5th edition). McGraw-Hill.

Mammoth! It covers linear and logistic regression and much more on statistical modelling.

9.   Weisberg, S. (2005). *Applied Linear Regression* (3nd edition). New York: Wiley.

A text on linear regression, but not as comprehensive as Draper and Smith (1998).

You are encouraged to report to the Library (and to the lecturers) books that you are finding difficult to get hold of – this will alert the Library to a potential excess demand over supply ….

<div style="border:1px solid black; padding:10px">

Can't find a book in the Library…

The book I want is on loan to someone else. Use WebCat to place a hold on the book even if you think it won't be returned in time (this will notify staff that the book is in demand). If you think more copies of a text may be required, please also tell your course tutor.

WebCat says it's on the shelf but it's not. Start by asking at the Enquiry Desk. If Library staff can't find the book, look again the next day. If you still can't find it, go back to the Enquiry Desk and say that you would like to report the book missing. Library staff will then search again and let you know the outcome.

Other problems using the Library… please ask at the Enquiry Desk for help or contact your Subject Librarian, Harry Gibbs – hgibbs@soton.ac.uk

</div>

A useful electronic resource for material covered during the course (in effect, an electronic statistics textbook) is available online at http://www.statsoft.com/textbook/stathome.html.

12.   <u>Availability of SPSS for Home PCs</u>

For those of you with your own PC or laptop, we would strongly advise you to get hold of a copy of the SPSS software.

To obtain SPSS, go to the iSolutions software distribution service site (http://www.software.soton.ac.uk), type in your user name and password, click on the

_____

"Agree" button at the bottom of the page to agree to the licensing conditions, and then select SPSS from the Applications menu. From the "SPSS availability" page, click on the link "Click here to proceed to the download page". Before downloading the software, click on the information (**i** symbol) corresponding to the item "SPSS v22 for Windows". Read this information carefully. The process gets ever more complicated with each new version of SPSS, but note that as part of the installation process you will require a licence code. If you have difficulty downloading/installing SPSS, please contact iSolutions (serviceline@soton.ac.uk).

## 13.	Workload for the Module

So that you are under no illusions, the University considers that a full-time postgraduate student should spend 1200 hours working (per year). This includes time spent attending lectures, seminars and computer workshops, and time spent studying and being assessed. This works out at **100** hours for a 10 CATS module (which is what STAT6095 is). This will give you an idea of the time you need to devote to this module!

## 14.	Procedures for Problems Associated with the Module

If you have any kind of problem with this module, you should raise the issue with the lecturers in the first instance. If you are not happy with the outcome, you should approach the MSc Official Statistics programme coordinator, Paul Smith. If you are still not happy, you should take up the issue with the Head of Teaching Programmes for the Department of Social Statistics and Demography, who is Dr Nikos Tzavidis.

If you have a major difficulty during the module, e.g. a health problem that prevents you from attending lectures or seriously interferes with your work, you should make sure you obtain the relevant documentation (e.g. a medical certificate), fill in a special considerations form, bring these to the programme coordinator for signature, and then submit these to the Programme Administrator, for filing.

## 15.	Blackboard Unit

To access the Blackboard site use Internet Explorer and enter the web address http://blackboard.soton.ac.uk/, login and select the "MSc Official Statistics Student Area"

_____

site. Select "Individual Units" from the left-hand menu, select the "2015-16 Units" folder, and then select the "STAT6095 – Regression Modelling" folder. Materials for all of the computer workshops, including the worksheets, the datasets, and the solutions, will be made available here. There is also a discussion board on Blackboard so that you can communicate with your fellow students. All students are automatically enrolled on this blackboard site. If you cannot access the site for any reason please let the lecturers know so that you can be enrolled.

16.   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/reports 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 Department of Social Statistics and Demography "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.