

Assignment Project Exam Help

STAT0023 Computing for Practical Statistics

Richard Chandler (r.chandler@ucl.ac.uk)

<https://tutorcs.com>

Tom Honnor (thonnor@ucl.ac.uk)

Department of Statistical Science, University College London

WeChat: [estutorcs](#)

2018–19 session

Course organisers: Prof Richard Chandler (Room 130, 1–19 Torrington Place, r.chandler@ucl.ac.uk) and Dr Tom Honnor (Room 123, 1–19 Torrington Place, t.honnor@ucl.ac.uk).

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

Course organisers: Prof Richard Chandler (Room 130, 1–19 Torrington Place, r.chandler@ucl.ac.uk) and Dr Tom Honnor (Room 128, 1–19 Torrington Place, t.honno@ucl.ac.uk).

Aims of course: to learn to use R and SAS to input and manipulate data, produce graphics and carry out statistical analyses.

<https://tutorcs.com>

WeChat: cstutorcs

Course organisers: Prof Richard Chandler (Room 130, 1–19 Torrington Place, r.chandler@ucl.ac.uk) and Dr Tom Honnor (Room 128, 1–19 Torrington Place, t.honnor@ucl.ac.uk).

Aims of course: to learn to use R and SAS to input and manipulate data, produce graphics and carry out statistical analyses.

Prerequisites: STAT0004 (formerly STAT1006), STAT0005 (formerly STAT2001), STAT0006 (formerly STAT2002).

WeChat: cstutorcs

Course organisers: Prof Richard Chandler (Room 130, 1–19 Torrington Place, r.chandler@ucl.ac.uk) and Dr Tom Honnor (Room 128, 1–19 Torrington Place, t.honnor@ucl.ac.uk).

Aims of course: to learn to use R and SAS to input and manipulate data, produce graphics and carry out statistical analyses.

Prerequisites: STAT0004 (formerly STAT1006), STAT0005 (formerly STAT2001), STAT0006 (formerly STAT2002).

Workload: One 1-hour lecture and one **compulsory** 3-hour workshop per week during term 2.

- Classes split into two groups for workshops – check your timetable to find which group you're in (Tuesday or Friday).
 - You **MUST** stay in your allocated group.
- *Students should complete workshop exercises in their own time, just like non-assessed exercise sheets for other STATS courses.*

Moodle page:

Entitled *STAT023 Computing for Practical Statistics (18/19)*.

All Portico-approved students should be enrolled automatically: others should contact Prof Chandler to arrange access.

<https://tutorcs.com>

WeChat: cstutorcs

Moodle page:

Entitled STA0023 Computing for Practical Statistics (18/19).

All Portico-approved students should be enrolled automatically: others should contact Prof Chandler to arrange access.

Texts: all materials available via Moodle, including links to additional books for students who would find them useful.

WeChat: cstutorcs

Moodle page:

Entitled *STAT0023 Computing for Practical Statistics (18/19)*.

All Portico-approved students should be enrolled automatically: others should contact Prof Chandler to arrange access.

Texts: all materials available via Moodle, including links to additional books for students who would find them useful.

Assessment: Two assessments, set during the lectures on Monday 4th February and Monday 18th March. Each assessment is worth 50% of the course; there is no examination. The assessments will be a mixture of Moodle quiz (to be taken during the workshops of the weeks beginning 4th February and 18th March) and take-home work. The assessed Moodle quiz questions will be taken from weekly quizzes.

Moodle page:

Entitled *STAT0023 Computing for Practical Statistics (18/19)*.

All Portico-approved students should be enrolled automatically: others should contact Prof Chandler to arrange access.

Texts: all materials available via Moodle, including links to additional books for students who would find them useful.

Assessment: Two assessments, set during the lectures on Monday 4th February and Monday 18th March. Each assessment is worth 50% of the course; there is no examination. The assessments will be a mixture of Moodle quiz (to be taken during the workshops of the weeks beginning 4th February and 18th March) and take-home work. The assessed Moodle quiz questions will be taken from weekly quizzes.

Getting help: use **office hour** and **Moodle discussion forum**.

Assignment Project Exam Help

Lecture 1 Part 1:

<https://tutorcs.com>
Introduction and overview

WeChat: cstutorcs

Software used in the course

- Course covers two different statistical 'packages': R and SAS.
- R is a 'free software environment for statistical computing and graphics' (see <https://www.r-project.org/>).
 - Modern environment based on the S programming language
 - Many standard statistical procedures are implemented directly
 - Programming language provides complete flexibility in defining new procedures, customising and enhancing graphics etc.
 - Currently over 13,500 additional specialist packages available from the Comprehensive R Archive Network (CRAN: <https://cran.r-project.org/>).
 - Increasingly widely used in many sectors, both in research and industry.

Software used in the course

- Course covers two different statistical 'packages': R and SAS.
- R is a 'free software environment for statistical computing and graphics' (see <https://www.r-project.org/>).
 - Modern environment based on the S programming language
 - Many standard statistical procedures are implemented directly
 - Programming language provides complete flexibility in defining new procedures, customising and enhancing graphics etc.
 - Currently over 13,500 additional specialist packages available from the Comprehensive R Archive Network (CRAN: <https://cran.r-project.org/>).
- Increasingly widely used in many sectors, both in research and industry.
- SAS ('Statistical Analysis System') is an older, commercial product dating back to the 1970s:
 - For many years the industry standard in sectors including pharmaceuticals and insurance, currently the dominant commercial software worldwide in "advanced analytics".
 - Has both command language and 'point and click' interface.

Assignment Project Exam Help

- Many other statistical / analytical packages exist: Matlab, Minitab, SPSS, Statistica, Systat, ...
 - But once you've learned R and SAS, many of these others should feel familiar to you
- As well as experience with statistical software packages, this course provides you with **generic programming skills** that are valued by many employers
- Programming concepts more readily illustrated using R, hence **60% of the course uses R** and the remaining **40% uses SAS**.

<https://tutorcs.com>

WeChat: cstutorcs

- We will use both R and SAS by typing commands / programs, rather than using a graphical interface ('point and click'). Some advantages of this are:

- The commands to perform a particular analysis can be saved in a file (known as a **script**), which can be edited later and / or easily re-run.
- You can include **documentation & comments** within a script.
- With a script you can see **whether a mistake has been made**, and where.

'Point and Click' rapidly becomes tedious (and prone to error) when repeating similar tasks.

- We will use both R and SAS by typing commands / programs, rather than using a graphical interface ('point and click'). Some advantages of this are:

- The commands to perform a particular analysis can be saved in a file (known as a **script**), which can be edited later and / or easily re-run.
- You can include **documentation & comments** within a script.
- With a script you can see **whether a mistake has been made**, and where.

'Point and Click' rapidly becomes tedious (and prone to error) when repeating similar tasks.

- **Note:** we cannot teach you every single command that you will need: the course aims to **give you the confidence to find appropriate new commands** for yourself.

Assignment Project Exam Help

Lecture 1 Part 2:

<https://tutorcs.com>

R revision

WeChat: cstutorcs

Assignment Project Exam Help

- R already introduced in STAT1006 (now STAT0004)
- Used in conjunction with RStudio for easy organisation of files etc.

- The “Course overview” tab on the STAT0023 Moodle page contains links to:

- Summaries of STAT1006 material for students wishing to refresh their memories (see “Useful books and online resources” link)

● R and RStudio home pages, for students wishing to install the software on their own computers (see “Obtaining R and SAS for home use”).

- A “quick-start” Introduction to R, summarising what you’re expected to know at the start of this course.

<https://tutorcs.com>

WeChat: cstutorcs

- See analysis of **Galapagos island biodiversity data** on STAT0023 Moodle page (script `Workshop1_Galapagos.r`).

Assignment Project Exam Help

<https://tutorcs.com>

WeChat: cstutorcs

- See analysis of **Galapagos island biodiversity data** on STAT0023 Moodle page (script `Workshop1_Galapagos.r`).

- Script illustrates:

- Use of **comments**, to ensure code is clear and readable
- Use of `<-` to **assign result of an operation to an object**
- **Reading data from file** using `read.table()`
- Use of a **data frame** (`species.data`) to store **collection of variables** (all **numeric** or **integer** here but could also include **character**, **logical** or **factor** variables)
- Some ways to find information about an **R** object e.g. using `names()`, `str()`, `summary()`, `class()`
- Use of `[]` to extract parts of an object according to a logical condition (`big.island`)
- Use of `$` to work with named components of an object.
- **Plotting** different types of R object, with control over labels and formatting.
- **Saving graphics** to files in different formats (PDF, JPEG, PNG, ...)
- Etc. (**more in this week's workshop**)

Assignment Project Exam Help

<https://tutors.com>

WeChat: cstutors

R revision: other things you should know about

Vectors, matrices and arrays: including different types (numeric, character, logical etc.) and extracting subsets using `[]` — either with a logical condition (e.g. `species.data[big.island,]`) or with numeric expressions (e.g. `species.data[16,]`).

Simple statistical procedures: summary statistics (`mean()`, `var()`, `sd()`, `median()`, `min()`, `max()`, `range()`, `quantile()`, `table()` etc.), simple test procedures and confidence interval calculations (`t.test()` for means in one or two groups, `var.test()` for F -test for variances in two groups, `chisq.test()` and `fisher.test()` for testing association in contingency tables)

Simple graphics: scatterplots (`plot()` and `pairs()`), boxplots (`boxplot()`), histograms (`hist()`), bar charts (`barplot()`) and density plots (`density()`).

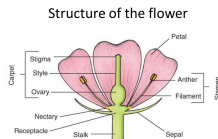
Using the help system: `?` for help about a specific command, `??` to search the help system.

Aims of an exploratory analysis

- 1 To gain a preliminary understanding of structure in a dataset
- 2 To look for possible outliers or data quality problems
- 3 To suggest some initial assumptions (e.g. normality of residuals, constant variance) that may be reasonable as a starting point in subsequent modelling and analysis

- Summary statistics, tables etc. can be helpful but well-designed graphics can often save a lot of work later on

- Example** look at the Fisher Anderson iris data (script `Workshop1_Iris.r` on the STA10023 Module page)



Key principle: message should be clear “at a glance”

Assignment Project Exam Help

Some guidelines:

Labelling: title, axis labels (including units of measurement), legend where necessary — and **choose text size that ensures labels are legible.**

<https://tutors.com>

Scaling: subject to other constraints, choose scales so that data fill up as much of plotting region as possible

WeChat: cstutors

- Other constraints' might include **use of common scales to aid comparison of two sets of data**

Colours / symbols / line types: consider possible loss of quality when photocopying / transporting to Powerpoint / etc. and **never rely exclusively on colour (NB some people are colour-blind: red-green is particularly problematic)**

Assignment Project Exam Help

- Aim to (re)familiarise yourself with the basics of R
- Work through the Galapagos and Iris example scripts
- Ensure that you understand everything — there will be helpers, use them if necessary!
 - Also get used to using the R help system
- Use Moodle quizzes to check that you have understood the material and get ahead with the assessments.

<https://tutors.com>

WeChat: cstutors