

# ST447 Data Analysis and Statistical Methods: Group Project

October 2017

## 1 Aims and Objectives

The main aims of this project are:

1. to enhance your understanding of a certain topic of this course;
2. (in some cases) to develop your coding skills or to facilitate your critical thinking;
3. to give you the experience of writing a small piece of scientific report;
4. to improve your presentation skills;
5. to learn to work as a team (of around 4)<sup>1</sup>.

## 2 Project details

You can pick either of the following, though **slight preference** will be given to the second.

### 2.1 Modern Statistical Methods

Your task is to write a brief description of a modern statistical method, at your own choice, developed since 1960 and widely used in statistics today. Examples of those methods include, for example, Tree methods, Bootstrap, Bagging, Boosting, Empirical likelihood, Neural network/Deep learning, Kernel density estimator, LASSO, Additive models, MCMC, Kalman filter, Support vector machines and etc. You should include:

1. a description of the method, with an illustration using a simple example,
2. the background to the development of the method, and the types of problem(s) that the method attempts to address,
3. a description of how the method works, possibly including the nature and interpretation of any tuning parameters, input and output.
4. you may also include examples of applications of the method, and major contributions to the development of the method.

---

<sup>1</sup>In the “Group Project Referendum” on 10 Oct, you have decided overwhelmingly to support the idea of forming the groups by yourselves. Each group should be ideally and preferably of size 4, and at most of size 5. We reserve the right to renormalise the final marks for different group sizes.

Bibliographic information relating to people who worked on the method with illustrative graphics/pictures is also permitted if you choose. Your work should be suitable for publication in a document on the history of statistics to be given to second year undergraduate students; it should be concise, informative and entertaining if possible. The aim is to test your ability to find sources of information and present work that demonstrates you have understood what you have read.

A dim view will be taken of barely adapted versions of existing descriptions (such as those in Wikipedia) or equations lifted from the literature without proper understanding.

You should not choose a topic which has been systematically covered in this course or any other courses (such as ST443) offered in this term. In case you choose a topic covered in these courses, you are expected to go much further than those included in the lecture notes.

## 2.2 Probability and Statistics in daily life

Your task is to present examples of how the use of probability and/or statistics defies the conventional wisdom in daily life. Some examples include: the birthday problem mentioned in both the lecture and the seminar, and an investigation of how unlikely it is for six people to be killed while cycling in London in a two-week period. A lot of other interesting examples can be found in, for instance, the Significance magazine published by the Royal Statistical Society and American Statistical Association.

<https://www.significancemagazine.com/>

You should include (whereas applicable):

1. a description of the problem,
2. an explanation of why the conventional wisdom fails, or,
3. a solution of the problem,
4. the pros and cons of this solution,
5. you may also include examples of applications of the method to a new (or updated) dataset. If possible, you are encouraged to collect a dataset by yourself.

Your work should be suitable for publication in a book for casual reading to be given to second year undergraduate students; it should be concise, informative and entertaining if possible. The aim is to test your ability to find sources of information and present work that demonstrates your critical thinking.

A dim view will be taken of barely adapted versions of existing descriptions (such as those in Significance) or methods lifted from the literature without proper understanding.

You should not choose a topic which has been systematically covered in this course or any other courses offered in this term. In case you choose a topic covered in these courses, you are expected to go much further than those included in the lecture notes.

## 3 Assessment

This is a group project – you must work collaboratively. You should allocate tasks among the group members in an equitable fashion. The assessed part of this coursework consists of a written report and an oral presentation.

Total (20%): Report (10%), Presentation (10%).

Deadlines: Oral Presentation on 21 Nov 2017 (Tuesday). Report due at **1200 GMT on 7 Dec 2017** (Thursday).

**You are required to let us know your group members and topic selection before the reading week (i.e. by 27 Oct).** In the unfortunate (and perhaps unlikely) scenario, you could still change your topic before the start of Week 7 (i.e., by 6 Nov). Please send me an email containing the aforementioned information (and copy our GTA at H.Feng2@lse.ac.uk).

### 3.1 Oral presentation

Your group will be given a maximum of **10 minutes** (strictly applied) to give your presentation. This is a piece of group work – everyone must contribute to the preparation of material and everyone in the group must speak for at least 2 minutes.

The key to giving a good presentation is careful preparation. You may use whatever props you feel are appropriate. There will be an overhead projector, data projector (connected to a computer on the school network for PowerPoint presentations) and white board in the room. You must make sure that the props do not detract from what you are trying to say.

The only required bit of content is a description of your group project (the modern statistical method chosen by your group or the use of probability/statistics in daily life). Other than that, you are free to include whatever in your project that you think would interest your classmates.

### 3.2 Written report

**Two** copies of the written report should be handed in to Ms Sarah McManus in COL 6.11 in Columbia House before 4pm on 6 Dec 2017 (Wednesday). In the meanwhile, you should send a pdf version of your report via Moodle (details to be announced later).

There is strict page limit for the report (**Maximum 10 A4 sides**, including figures and tables, but **NOT** reference list or relevant R code). You should use an 11 point standard font (for example, times new roman) and 1.5 spacing.

In addition to the **correctness, concreteness, clarity and conciseness** of the report, you should show a **critical awareness** of any weaknesses in the method you discuss or the analysis you present and discuss possible extensions and improvements.