

COMP6235 Group Coursework Instructions

Module:	<i>Foundations of Data Science</i>	Lecturers:	<i>ES, HF, MB</i>
Assignment:	<i>Statistics coursework</i>	Weight:	<i>15%</i>
Deadline:	<i>17/11/17</i>	Feedback:	<i>8/12/2017</i>

Instructions

The following coursework is worth **15%** of the assessment of the module. Please note that you are expected to use the statistics package R to carry out this coursework (hence, e.g., plots are expected to have been generated using R).

Download the data set fish.txt about the catch of a hypothetical fishing fleet from

<http://users.ecs.soton.ac.uk/mb8/datascience/fish.txt>

and import it into R. The data set records data for a time period of one day during which one fisherman has fished in a lake. The fisherman has recorded every catch he has made during this time. The data set consists of two rows with X values giving the times at which a fisherman has made a catch and the Y values indicating the size of that catch (i.e. its weight in kg). Using R, your task is to analyse this data set.

In a first step, generate plots that illustrate the distributions of X values (times of catch, the format is hours, fraction of hours on a 24h schedule for the day), Y values (size of catch). Characterise and describe these distributions by their median, and mean values, variances, and additional measures introduced in the introduction to statistics lectures that you think shed light on the shape of the respective distributions. Assuming that the data are a sample from a larger population, give mean values with 95% confidence intervals for both distributions.

In a second step, it is of interest to analyse the dependence between time of catch (X value) and size of catch (Y value). For this purpose, generate a plot that shows the dependence of X on Y and discuss the observed relationship. Then, characterise this relationship by statistical measures that evaluate the co-dependency between X and Y. Analyse the amount of information about Y that is given by knowledge of X.

What is the best time to go fishing at this lake and why?

Write up your finding in a short report of no more than 2 pages/500 words and submit it electronically via handin as one pdf file before 12pm (noon) November 17 2017.

Submission

You must submit the following documents

- **One** pdf document that contains your written report and includes the figures you produced,

The deadline for the submission is 12pm November 17.

Marking Scheme

Quality of the figures (do they meet professional standards, are relationships discussed in the text clearly visible? Are axes labeled properly? Captions?): 20%

Technical content (do you address all questions? Are the answers correct?): 50%

Quality of the writing (formatting, correct spelling and grammar, description of your findings about the data set): 30%