CITS5502 Software Processes
Semester 2, 2019
Assignment 2

- Document version number: 0.1.0
- Date: 2019-09-20

This assignment is worth 30% of your final grade and is due on Sunday, 13 October at 11:59pm. It should be submitted via `cssubmit`. All work is to be done individually. The assignment is marked out of 30.

You are expected to have read and understood the University's guidelines on academic conduct. In accordance with this policy, you may discuss with other students the general principles required to understand this project, but the work you submit must be the result of your own effort.

You must submit your work before the submission deadline above. The penalties for late submission are described in the University's guidelines on assessment.

Overview – modelling the production of code

The work for this assignment will be based upon an analysis of a set of data on the effort required to complete software projects, undertaken by students in previous years. You are required to fit two models to three sets of data. You will then need to interpret the fitted models (each has three parameters) in order to answer a number of questions on software processes. In particular, we are looking at the way people acquire knowledge and skill in solving some problem in a given programming language, and whether that transfers to other languages.

Data

The data for this assignment was generated through 14 students writing multiple versions of programs in various languages.

The students had three different *tasks* to do, and repeated each task multiple times. The tasks involved solving two similar problems (Problem 1 and Problem 2) in two different languages (Language A and Language B). Specifically, the tasks were:

- Task P1LA - solve Problem 1 in Language A
- Task P2LA - solve Problem 2 in Language A

- Task P1LB - solve Problem 1 in Language B

For each task, the student re-wrote their program four times, at intervals of at least three days, without reference to the code from their previous attempts. The times (in minutes) taken by the students to code each version of each task are given in the spreadsheet for this assignment. There are (14 students $\times$ 3 tasks $\times$ 4 versions) = 168 data points collected.

What to submit

You should submit:

- A PDF report, the contents of which is detailed below in the "Report contents" section. It should meet the requirements given under "Report guidelines".

  All work which you wish to submit for assessment – including charts, diagrams and tables – should be included in this report.

- A .zip file, containing any Excel spreadsheet or other code used for generating your charts and results. It should include a README file explaining how to reproduce them.

  However, this will only be looked at if the markers have queries about how your results were derived – do **not** put content you wish to be assessed in the spreadsheet or code. If you have created charts using Excel (or some other system) you will need to paste them into your report.

Assignment tasks

1. Averaging data

From the spreadsheet, you should randomly select the data of 7 out of the 14 students, and computer the mean time taken to write each version of each task, giving you 12 means (3 tasks $\times$ 4 versions = 12).

Your subsequent curve fitting task and your report should be based on these 12 data points.

2. Curve fitting

You should choose any two of the following four models and use them to fit curves to the completion times for each task (i.e., for each of your chosen models, there are only 4 data points available for curve fitting, and you need to repeat this curve fitting process for tasks P1LA, P2LA, and P1LB).

The four models available are functions of time (t), and they predict effort (in person-minutes), and have parameters a, b and c which need to be fitted.

- Model M1: Effort = $a + bt + ct^2$

- Model M2: Effort $= \dfrac{a + bct}{bt + 1}$
- Model M3: Effort $= (a - c)e^{-bt} + c$
- Model M4: Effort $= (a - c)(t + 1)^{-b} + c$

Here, $t = 0$, 1, 2 or 3, and denotes each of the four attempts to complete a task; "Effort" is the person-time in minutes required to complete a task. You will need to impose the constraints that $a > 0$, $b > 0$, and $c > 0$, and you may need to apply a stronger constraint on parameter c.

3. Report contents

Your report should contain the following:

Source data (max. $\frac{1}{2}$ page text content):

List the numbers of the 7 random students whose data you are using, and include a table showing the averaged data.

Model fitting (max. $\frac{1}{2}$ page text content):

Show charts of your data and the fitted curves.

Model interpretation (suggested 1–2 pages text content):

Briefly describe which two models you chose for analysis and why. Give an interpretation of what each of the parameters a, b and c signify in the models.

Discuss what software you used for curve fitting.

Model discussion (suggested 2–4 pages text content):

Based on your curve fitting results:

- Describe which model fits the best. Discuss which of the two models seems more *plausible* as a model for the data, and why.
- Discuss the extent to which learning from a previous project carries over to another project of a similar type, but in a different language.
- Discuss the extent to which in solving one project in a language improves performance in solving another project in the same language.
- Discuss the degree to which practice should enable programmers to produce a more predictable estimate of effort.
- Is the learning pattern similar for the three different tasks? Explain your answer.
- What appears to be the minimum time that could be taken to solve any of the 3 problems?

Limitations (suggested $\frac{1}{2}$ to 2 pages text content):

Discuss any limitations of the data or the analysis, and what you would recommend in order to overcome these.

Conclusion ($\frac{1}{2}$ to 1 page text content):

Provide a short conclusion section (half a page or so) summarizing your findings.

Report guidelines

Your report should be in PDF format, and use A4 size pages. It should clearly show your name and student number.

The font for body text should be between 9 and 12 points. The report should contain numbered headings, with useful heading titles. Any diagrams, charts or tables used must be legible and large enough to read. All pages (except possibly the cover, if you have one) should be numbered. If you give scholarly references, you may use any standard citation style you wish, as long as it is consistent. Cover sheets, diagrams, charts, tables, bibliographies and reference lists do not count towards any page-count maximums.

It is expected that the curve-fitting charts be generated using software of your choice (MS Excel, or a programming language such as Python, R, Matlab, etc). You are welcome to include other diagrams or charts if you wish – they can be drawn by hand (as long as they are clear and legible) and scanned for inclusion in the submitted report, or they can be drawn with tools like MS Word, MS Excel or draw.io.

It is suggested that reports be kept to under 10 pages of textual content.

Submission

Submit your assignment as a PDF report via `cssubmit`. Also submit, as a zip file, any code and/or spreadsheets you used to perform your analysis, and instructions on how your results and charts can be replicated.

Assessment

Each component of the report will be assessed on whether

- it is clear and logically laid out
- it meets the requirements given
- it describes and justifies any assumptions made.