

Statistical Methods for Data Science

DATA7202

Semester 1, 2021

Assignment 4 (Weight: 25%)

Assignment 4 is due on 14 June 2021 17:00).

Please answer the questions below. For theoretical questions, you should present **rigorous proofs** and **appropriate explanations**. Your report should be visually appealing and all questions should be answered in the order of their appearance. For programming questions, you should present your analysis of data using **Python, Matlab, or R**, as a short report, clearly answering the objectives and justifying the modeling (and hence statistical analysis) choices you make, as well as discussing your conclusions. **Do not include excessive amounts of output in your reports.** All the code should be copied into the appendix and the sources should be packaged separately and submitted on the blackboard in a zipped folder with the name:

`"student_last_name.student_first_name.student_id.zip".`

For example, suppose that the student name is John Smith and the student ID is 123456789. Then, the zipped file name will be `John.Smith.123456789.zip`.

1. **[100 Marks (see details below)]** *Air Secure* wishes to open a number of new service desks, **guaranteeing that in the long run 90% of their customers do not have to wait longer than 8 minutes in a waiting queue before they are served.** Preliminary research by Air Secure showed that on **arrival customers always choose the smallest queue and remain there until served.** This research also investigated the passengers inter-arrival time (in minutes) and the service time. The results are summarized in `data.csv`. The data for the first four passengers are provided below.

```
inter_arrival_time service_time
1.32689531655087 4.47364233325769
0.190173284276185 3.59214966621201
0.101620716736811 4.5189272687414
0.114430104248874 9.97999506103284
```

Perform a Discrete-Event Simulation study in Matlab, Python, or R, to answer the following question.

How many service desks should be *minimally* available to meet the service requirements? Namely, how many service desks should be available such that in the long run 90% of their customers do not have to wait longer than 8 minutes in a waiting queue before they are served. Specifically, run the simulation for $T = 3000$ units of time. Discard the first 30% of the samples and use $N = 50$ batches to estimate the probability that a customer waits less than 8 minutes before entering the service.

Perform a Discrete-Event Simulation study to answer the following question.

- (a) **[15 Marks]** Give the problem summary and describe the project objective.
- (b) **[15 Marks]** Give a specification of variables used in the simulation study. In addition, show a diagram that describes the project dynamics.
- (c) **[40 Marks]** Results and Analysis. Using tables and figures, present a clear outcome of your study. Present the corresponding confidence intervals.
- (d) **[20 Marks]** Formulate your conclusions.
- (e) **[10 Marks]** Appendix. Include all code files used. Explain their interaction and provide a clear and well-commented code.