Statistical Analysis, Simulation, and Modeling.
Part II: Statistical Inference, Simulation, and Modeling

Semester-end project assignments

(Cheating in any form will not be tolerated, and will severely be punished when detected)

I. Get four data samples from the files named "datafile1.csv" - "datafile4.csv" (attached to the same email, with which you received this document), and conduct a preliminary (visual) analysis of the data (plot histograms, etc.) Based on the preliminary analysis, select one data sample that, in your opinion, would describe human reaction time to a stimulus.

II. Plot PDFs of the three distributions listed below, and explore the shapes of the PDFs by varying (assigning different values to) their parameters:

1) Lognormal distribution (two parameters, μ and σ):

$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}};\ x > 0;\ \sigma > 0.$

2) Weibull distribution (two parameters, c and b):

$f(x) = \frac{cx^{c-1}}{b^c e^{(x^c/b^c)}}; x > 0;\ c > 0,\ b > 0.$

3) Wald distribution (two parameters, μ and λ):

$f(x) = \sqrt{\frac{\lambda}{2\pi x^3}}\, e^{-\lambda\frac{(x-\mu)^2}{2\mu^2 x}};\ x > 0;\ \mu > 0,\ \lambda > 0.$

III. Use the data sample selected in (I) and obtain parameter estimates for the three models of (II). Assess and write down the accuracy of the parameter estimates obtained by constructing the corresponding CIs.

IV. Consider the three models from (III) and conduct a model selection analysis. Select "the best model" for the data (i.e. the "best" distribution from (II) to describe the human reaction time – the data selected in (I)). Justify your choice of "the best model" (e.g. use Q-Q plots and/or quantitative model selection criteria).

Note that your report must include at least 10 graphs illustrating the analysis. Also note that any notations and numbers resulted from computations that are left unexplained in the report may result in an unsatisfactory grade for the assignment.
Each of the four assignments will be evaluated on a scale from 0 to 25 points.
Reports with signs of plagiarism (e.g. with identical solutions, etc.) will all receive 0 points.