MTH1004 summative coursework: term 2

Date set: Friday 8 March 2024
Submission date: Thursday 28 March 2024 by noon
Return date: Friday 3 May 2024

Instructions

This coursework comprises 15% of the overall module assessment. This is an individual assessment and you must not discuss it with other people. Note also that it is an academic offence to use AI tools to produce answers for you. If markers suspect that you have committed an academic offence then they may require you to attend a viva (oral exam) in order to establish the legitimacy of your work. Please refer to our regulations on academic conduct for details.

http://as.exeter.ac.uk/academic-policy-standards/tqa-manual/aph/managingacademicmisconduct

Conduct the two investigations outlined on the following pages. For each investigation, write a short report of your findings, addressing each part of the investigation in a separate section. The data and R functions needed for the two investigations are contained in the file `MTH1004T2CW.RData`, which you should download from ELE at https://ele.exeter.ac.uk/course/view.php?id=10536.

Submit your answers on ELE by the date shown above. You must submit a Zip archive containing two files: one R script (ending .R or .r) containing the R code that you used to produce your results, and one pdf file. Your R script should be formatted to aid readability, should include brief comments to explain what each section of code does, and should run without errors. Your pdf file should contain your two reports. Each report should comprise at most two sides of word-processed text (e.g. written with Word or LaTeX) and as many figures or tables that you feel is appropriate. No credit will be awarded to additional pages of text. (You do not need to write two sides of text to achieve full marks—the model solutions contain only slightly over one side of text.) You must use the page size A4 and a font size of at least 11 points, lines must be single-spaced and all margins must be at least one inch wide. Do not submit any other files, such as extra R scripts or separate figure files.

Assessment criteria

Approximately 50 marks are available for each investigation. Of these 50 marks, approximately 20 are awarded for your method and numerical results, 20 for your discussion of your results, and 10 for the quality of your presentation of your report. To achieve a pass mark, you will have applied the main inferential methods without major errors, presented your results intelligibly, interpreted your results without major errors, and written an intelligible report. To achieve a first-class mark, you will have selected appropriate inferential methods, applied them accurately, presented your results effectively, interpreted your results accurately and with an appropriate level of detail, and written a clear report.

Investigations

1. Daily rainfall totals at a location were measured (in millimetres, mm) for 40 years. In those 40 years, the daily total exceeded 25 mm on 145 days. The file `MTH1004T2CW.RData` includes a data frame called `Rainfall` which contains the 145 excess amounts (that is to say, if the daily rainfall total was x then the value in the data frame is $x - 25$). The file also contains some R functions which are described at the end of this document and which you may use to answer this question.

   (a) Consider modelling the excess rainfall amounts as realisations of independent and identically distributed random variables. Choose a parametric model for these data from among the distributions discussed in the lecture notes. Then estimate the model parameter (or parameters), estimate the standard errors of your parameter estimates, and assess the realism of your model. Include in your report the formulae for your parameter estimates, the numerical values of the parameter estimates and standard errors, a discussion of your model's realism and appropriate graphical evidence. You do not need to include the derivation of your formulae.

   (b) Now consider modelling the excess rainfall amounts as realisations of independent and identically distributed random variables with a distribution which we shall denote by $\mathrm{M}(\sigma, \gamma)$. In this model there are two parameters, σ and γ, where $\sigma > 0$ and $\gamma < 1/2$. This distribution has expectation $\sigma/(1-\gamma)$ and variance $\sigma^2/\{(1-\gamma)^2(1-2\gamma)\}$. Use these moments and the R functions described on the next page to repeat part (a) for this model and then explain which of the two models you prefer.

   (c) The m-year return level is the daily rainfall total that is exceeded once every m years on average. This may be computed from the models in parts (a) and (b) as the amount, x, for which the probability that the excess rainfall amounts exceed $x - 25$ is $1/(365mp)$, where $p = 0.01$ is the proportion of days on which the rainfall total exceeded 25 mm. Use your preferred model to compute the 10-year return level.

2. A study sought to determine the effectiveness of a new antibiotic for treating a certain type of infection in hospital patients in the UK. The study lasted for one year and included all consenting patients with the infection in two hospitals in England. Patients took the new antibiotic and doctors determined whether the infection cleared within two weeks. The data are contained in the data frame `Antibiotic` in the file `MTH1004T2CW.RData`. Each row of the data frame corresponds to a patient, column `Outcome` indicates whether the infection cleared within two weeks (1 = yes, 0 = no), and column `Hospital` indicates whether the patient attended Hospital A or Hospital B. An existing antibiotic is known to clear infections within two weeks for 70% of the target population.

   (a) First, ignore the hospital variable and assume that the patients are a simple random sample from the target population. Use a point estimate and confidence interval to assess the effectiveness of the new antibiotic.

   (b) Now, suppose that you are told that Hospital A is a general hospital but that Hospital B is a specialist hospital (and so treats only a certain type of patient). Use point estimates and confidence intervals to assess the effectiveness of the new antibiotic, and whether the effectiveness differs between the two hospitals.

(c) Comment on the strengths and weaknesses of the design of the study, and what you conclude about the effectiveness of the new antibiotic. Include in your report the formulae for your confidence intervals, as well as their numerical values.

R functions

The file `MTH1004T2CW.RData` contains functions `dmodel`, `pmodel` and `qmodel` to compute the pdf, cdf and quantiles of $\mathrm{M}(\sigma, \gamma)$ distributions. In particular, `dmodel(y, scale = s, shape = g)` will compute the pdf at $y =$ `y` for $\sigma =$ `s` and $\gamma =$ `g`; `pmodel(y, scale = s, shape = g)` will compute the cdf at $y =$ `y` for $\sigma =$ `s` and $\gamma =$ `g`; and `qmodel(p, scale = s, shape = g)` will compute the `p`-quantile for $\sigma =$ `s` and $\gamma =$ `g`. The file also contains a function, `rmodel`, to simulate random numbers from $\mathrm{M}(\sigma, \gamma)$ distributions. In particular, `rmodel(n, scale = s, shape = g)` will simulate `n` random numbers when $\sigma =$ `s` and $\gamma =$ `g`.