

## Project 2

Τρίτη, 9 Νοεμβρίου 2021 12:24 μμ



Project2...

Konstantinos Fokianos

## DEPARTMENT OF MATHEMATICS & STATISTICS UNIVERSITY OF CYPRUS

DSC 530  
FALL SEMESTER 2021  
PROJECT 2

1. This project represents 25 % of your total grade in the course and focuses on learning some basic elements of R and how it can be used for data analysis. In addition, some theoretical problems will illustrate your basic understanding in probability theory.
2. You can work in groups of 3, at most. Ideally, each student in the group should be coming from a different background (i.e. one student from Business Administration, one student from Computer Science and one student from Math & Stat).
3. Your work will consist of 6 pages and will contain all the names as well as ID numbers of the participants in the group. You are responsible for presenting the relevant programs and their results.
4. You can use any word processor you want but you will send the final work in pdf format by e-mail to [fokianos@ucy.ac.cy](mailto:fokianos@ucy.ac.cy)
5. **DEADLINE: 26/11/2021 UNTIL 20.00. NO ASSIGNMENTS ARE ACCEPTED AFTER THIS DEADLINE.**

1. (10 points) We assume that  $X_1, \dots, X_n$  is a random sample from a Bernoulli distribution with parameter (probability of success)  $p$  and we are interested on estimating the log-odds parameter  $\theta = \log(p/(1-p))$ .
  - (a) Suggest an estimator for  $\theta$  and derive its asymptotic distribution by the arguments used in the class lectures.
  - (b) Compare the asymptotic distribution with the bootstrap distribution of  $\hat{\theta}$ . In particular compare confidence intervals using different  $n$  and  $p$  values and discuss your findings.
2. (5 points) There are two kinds of stroke: ischemic and hemorrhagic. It is known that 13% of strokes are hemorrhagic. A clinical trial is to be conducted in a hospital on patients suffering a hemorrhagic stroke, and  $n = 10$  patients are needed for this trial. When stroke patients first arrive in hospital, it is typically unknown whether the stroke is ischemic or hemorrhagic. Patients are quickly tested to determine this, and entered into the trial if their stroke is hemorrhagic. Let  $N$  be the number of stroke patients tested to achieve  $n = 10$  hemorrhagic patients. Write down the likelihood of  $N$ . Calculate its MLE.

3. (5 points) Given  $X_1, \dots, X_n$  a random sample from the geometric distribution with parameter  $p$

$$\mathbb{P}[X_i = x] = p(1 - p)^x, \quad 0 < p < 1, \quad x = 0, 1, \dots$$

Implement the Newton-Raphson method for obtaining the MLE of  $p$  and compare it with exact answer. In addition study the asymptotic distribution of the MLE by simulation.

4. (5 points) Pick any dataset you prefer using the R package `datasets` (or any other package or web source) and give a short analysis based on the concepts you have been exposed so far. It is of particular importance to explain the phenomenon you study and why the results you obtained are relevant to the study you will consider to pursue.

