BEng Biomedical Engineering
Computational Statistics

Coursework 1 - Practical Examination

Instructions

First, read the following instructions.

This examination consists of 2 questions, worth a total of 100 marks. **Answer both questions.** If you do not manage to complete both questions then please submit the work that you have completed. Each question will involve writing MATLAB code to meet some requirements. For some parts you will be required to provide short written answers. Please type these as comments in your MATLAB script file making it clear which question part they correspond to.

The questions should be completed individually **under examination conditions**. You may not consult with any other students for the duration of the examination. This is an **open-book examination**. You **may** refer to any text books, course materials, source code or other materials available from the internet or KEATS. You **may not** communicate with others using e.g. email, instant messaging or social media sites during the examination.

Before you start, you should download a *zip* file from the link *Coursework 1 download files* under the Assessment section on the module KEATS page. Download the *zip* file and extract the folder and its contents onto your computer. **Write and save all of your work in the examination in this folder**.

Time allowed: 1 hour.

When the invigilator instructs you to stop work please do so immediately. You should then combine all of your solution files into a single *zip* file and submit it through KEATS as instructed.

Assessment

Your coursework will be marked based on the answers you provide and how well the code you write meets the requirements specified.

The overall mark for this coursework will make up 15% of your total mark for this module.

Question 1

A hospital Accident & Emergency (A & E) department is investigating a potential link between the state of consciousness of patients on arrival and their subsequent clinical outcome.

The *Glasgow Coma Scale* (GCS) is a way of numerically grading the conscious state of patients. GCS values range between 3 (deep unconsciousness) and 15 (full consciousness). The GCS values of 20 patients who arrived at A & E on a single day were recorded. For the same patients, their subsequent hospital stay (in days) was recorded along with whether they survived or not. These data are stored in the file *GCS.mat* which contains the following variables:

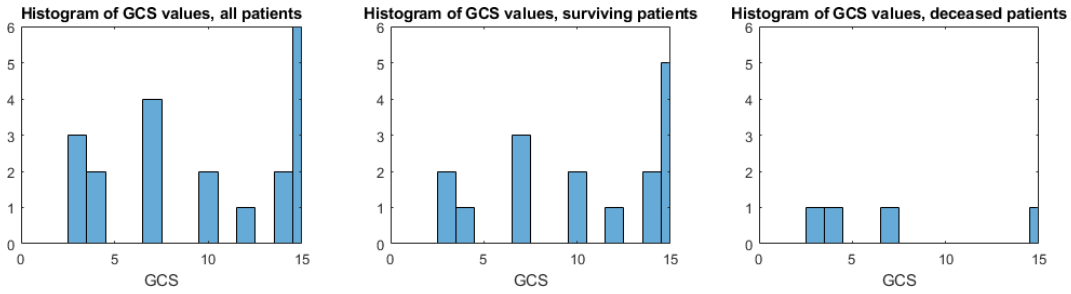
- `gcs`: An array of 20 GCS values, one for each patient.
- `stay`: An array of 20 hospital stay values, in days.
- `survived`: An array indicating survival (=1) or death (=0) of the patients.

Create a MATLAB script called *q1.m* and write all of your answers to the following parts in this file. For written answers, that do not require MATLAB code, write your answers as a comment in the MATLAB script (i.e. start the line with a `%` symbol), clearly indicating which question part it corresponds to.

(a) What type of statistical data are the GCS, hospital stay and survival variables?

[6 marks]

(b) Write MATLAB code to load in the data and split the GCS array into two separate arrays: one for patients who survived and one for those who died. Use the original array and the two new arrays to produce a single figure window containing three histograms. The first should show the distribution of GCS values for *all* patients, the second should show the GCS values for only *surviving* patients and the third should show the GCS values for only *dead* patients. Annotate all plots appropriately, and make sure that the axis limits are the same for all three plots to enable easy comparison. Your figure should look similar to that shown below.



[20 marks]

(c) Use MATLAB to produce a plot that visualises the relationship between GCS and hospital stay. Your plot should clearly show which patients survived and which died. Make sure that the plot is clearly annotated.

[18 marks]

(d) Compute an appropriate measure of correlation between GCS and hospital stay over *all* patients. You can assume that both sets of data are normally distributed. Also compute the probability that the correlation value obtained could have occurred through random variation in the data. Comment on the results.

[6 marks]

Question 2

Researchers would like to find out whether mean tumour perimeter is a good predictor of malignancy. The file *cancer.csv* contains two columns. The first column contains mean tumour perimeter and in the second column there is a diagnosis of whether the tumour is benign (0) or malignant (1).

Create a MATLAB script called *q2.m* and write all of your answers to the following parts in this file.

(a) Load data from the file *cancer.csv*. Display a boxplot showing the categories (0 or 1) on the x-axis and the mean tumour perimeter on the y-axis. Annotate your plot with axis labels and a title.

[15 marks]

(b) The researchers think that a tumour perimeter of 90 is a good threshold to distinguish between malignant and benign tumours. Predict whether the tumours are malignant by applying this threshold to the mean perimeter values. Values smaller than 90 indicate benign (0) and greater than or equal to 90 indicate malignant (1).

[10 marks]

(c) Calculate the number of true positives, true negatives, false positives and false negatives for this test.

[14 marks]

(d) Using the function `calcSeSp` from the case study we covered in class, calculate the sensitivity and specificity of this test and display this information to the command line.

[11 marks]