CSC3031  
UNIVERSITY OF EXETER  
College of Medicine and Health  
Applied Data Science  
January 2022  
Module Convenor: TJ McKinley  
Duration: 2 HOURS + 30 MINS UPLOAD TIME  
Word count: Not applicable

This is an **OPEN BOOK** exam.

**Additional Materials needed:** You will need R installed, plus a copy of the USPE.rds, VADeaths.rds and cancer.csv files.

**Extra instructions**:

* Please answer **ALL questions**.
* Please write up your answers using R Markdown, and submit the final compiled HTML, PDF or Word document to eBART within the specified time limit. If your work does not compile, then please submit the Markdown (.Rmd) or an R script (.R) file instead.
* Please use your **CANDIDATE NUMBER** and **NOT** your name on any answer sheets.
* Make sure all figures are well-formatted with neat axis labels.
* In addition to the marks below, a further **6 marks** (in total) are available for the formatting of your code.

# Questions

1. You are given a numerical vector x. Using vectorised operators, write a piece of code that extracts all elements of x that are whole numbers. **[2 marks]**
2. Using for loops, write a function to calculate the element-wise multiplication of two matrices , where and are input matrices of the same size, and is the output matrix. Each element of must be the product of the two corresponding elements of and . Arguments in your function should be checked for consistency. **[10 marks]**
3. Let and be two vectors, each containing random numbers sampled uniformly in the region . Hence and similarly . If is the number of the samples where , then we can approximate the number as .

* The function runif(n, -1, 1) generates a vector of random numbers uniformly in the region . Write a function in R that takes an integer input argument , and approximates by using the method described above. Input arguments must be checked for consistency. **[8 marks]**

1. The USPE dataset consists of United States personal expenditures (in billions of dollars) in the categories: food and tobacco, household operation, medical and health, personal care, and private education for the years 1940, 1945, 1950, 1955 and 1960. Place the USPE.rds file into your working directory and load this into R using the command:

* USPE <- readRDS("USPE.rds")
* Produce a plot showing how expenditure (in billions of US dollars) changes over time with respect to the different categories of spending. **[7 marks]**.

1. The data set VADeaths contains death rates, which are measured per 1,000 population per year. They are cross-classified by age group (rows) and population group (columns). The age groups are: 50–54, 55–59, 60–64, 65–69, 70–74 and the population groups are Rural/Male, Rural/Female, Urban/Male and Urban/Female.

* Place the VADeaths.rds file into your working directory and load this into R using the command:
* VADeaths <- readRDS("VADeaths.rds")
  1. Convert this data set into “tidy” format, with separate columns for urban/rural and sex. **[4 marks]**
  2. Create mean death rates in rural and urban groups for each age-group (averaged over males and females). Present this as a table with age down the rows and rural / urban along the columns. **[5 marks]**

1. A dataset on mortality rates from cancers per 100,000 people is available in the data set cancer.csv that can be downloaded from ELE. These data are from the U.S. Surveillance, Epidemiology and End Results (SEER) program. This data frame contains the following columns:
   * Age: values for age categories as defined by the SEER program;
   * Year: year of estimates;
   * Female: age-adjusted mortality rate per 100,000 population (AAR) for females;
   * Male: age-adjusted mortality rate per 100,000 population (AAR) for males.
   1. Read this dataset into R and produce a summary. Make sure that each variable is encoded correctly so that the summary table is informative. **[4 marks]**
   2. Plot the AARs over time for males, stratified by age. What does the plot suggest about how AARs are changing over time and with respect to age? **[6 marks]**
   3. Plot the AARs over time for both sexes, stratified by age. What does the plot suggest about how AARs are changing over time and with respect to age and sex? **[8 marks]**
   4. The percentage change in AAR between years and , denoted , is specified as: where and are the AARs for years and respectively. Calculate the percentage change in overall AAR between the years 2000 and 2019, stratified by sex and age-class. Based on your estimates, how much have the overall AARs changed over time and are these changes consistent across sex and age-classes? **[10 marks]**

**Total: [70 marks]**

**END OF PAPER**