R Tutorial 2

Instructions:

- Answer all questions.

- Ensure that your findings and results are clearly stated and thoroughly discussed. Please support your arguments using suitable R code with the relevant outputs, interpretations, plots and graphs whenever possible. You should support your argument using appropriate theory that is appropriately referenced.

- The R commands that you use in obtaining your results for all questions must be documented in a R script file. These scripts must be clearly commented. Ensure that any output is clearly stated and interpreted separately from the code as additional comments.

- Include the task name, your name and surname, and your student number in your R script file.

- You MUST label each answer by question number and, where a question has multiple parts, label each part of the question CLEARLY.

- On completion of your assignment, please submit onto RUconnected. If there are any issues uploading onto RUconnected, you may email your submission to: a.langston@ru.ac.za. Please submit your R script file and any other saved data files and plots mentioned in the questions below. Your student number should be included in the name of each file that you submit.

- Each student must complete an individual assignment. You will be assessed based on the quality and/or correctness of the R code, its outputs, and your explanations and interpretations. Acknowledge any help you may have received. Feel free to note any help you may have given to other students in the course.

- This assignment must be submitted by Tuesday, 23 July 2024 by 17:00. Late submissions will be penalized.

- Please note the Rhodes University and the Rhodes University Department of Statistics plagiarism policies.

Questions:

1. Construct the following matrices with R:

   (a) $A = \begin{bmatrix} 7 & 4 & 7 & 4 & 7 \\ 4 & 7 & 4 & 7 & 4 \\ 7 & 4 & 7 & 4 & 7 \\ 4 & 7 & 4 & 7 & 4 \end{bmatrix}$;

   (b) $B = \begin{bmatrix} 16 & 5 & 12 \\ 9 & 22 & 4 \end{bmatrix}$;

   (c) $C = \begin{bmatrix} 5 & 40 & 75 & 110 \\ 10 & 45 & 80 & 115 \\ 15 & 50 & 85 & 120 \\ 20 & 55 & 90 & 125 \\ 25 & 60 & 95 & 130 \\ 30 & 65 & 100 & 135 \\ 35 & 70 & 105 & 140 \end{bmatrix}$.

2. The ice cream flavour preferences for 10 randomly selected high school students is as follows:

   Chocolate Chocolate Vanilla Strawberry Chocolate Vanilla Chocolate Strawberry Vanilla Chocolate

   Create a factor named `flavours` containing these values.

3.

(a) Construct the following data matrix in R and name it `barley.data`

$$\begin{bmatrix} 190 & 8 & 22.0 \\ 191 & 4 & 1.7 \\ 223 & 80 & 2.0 \end{bmatrix}$$

(b) What are the dimensions of `barley.data`?

(c) The columns of the matrix represent the three types of barley, `TypeA`, `TypeB` and `TypeC`, respectively. The rows of the matrix represent the three provinces for which the data was collected, `Navarra`, `Zaragoza`, and `Madrid`, respectively. Using this information, provide the matrix with appropriate row and column names.

(d) Use the appropriate indexing to view the data for `Zaragoza`.

(e) Data was recorded for an additional province, `Valencia`, with measurements of 215 for `TypeA`, 6 for `TypeB` and 2.2 for `TypeC`. Add this additional information to the `barley.data` matrix.

(f) The elements within the matrix represent the number of tonnes of barley produced in a region by type and province. Find the the total barley produced in the region. Find the subtotals for the amount of barley produced by type. Note that the additional information on Valencia should be included in these totals.

(g) Use appropriate indexing to replace all the elements of the `barley.data` matrix that are less than 5 with NA.

4. Consider the following 10 measurements:

$$1.95 \quad 2.86 \quad 3.24 \quad 2.98 \quad 5.56 \quad 2.75 \quad 2.76 \quad 3.38 \quad 3.38 \quad 3.06$$

(a) Create a vector named `x` which contain these elements.

(b) Use the command `length()` on this vector. Interpret this result

(c) Determine the sum of the 3rd, 7th, 9th and 10th component in `x`. Use the appropriate indexing to answer this question.

(d) The command `sort()` sorts the values of a vector in ascending order. Use this command to determine the sum of all measurements, excluding the largest and the smallest one.

5. Solve the following system of equations using R:

$$\begin{aligned} 3x + 4y - 10w + 2z &= 6 \\ 6x - 2y + 3w - 4z &= 3 \\ -4x + 7y + 4w + 3z &= -1 \\ 9x - 3y - 2w + 6z &= 4 \end{aligned}$$

6.

(a) Create the vector `u=(1,2,5,4)` using the `c()` function and the vector `v=(2,2,1,1)` using the `scan()` function.

(b) Provide R code to find which component of `u` is equal to 5.

(c) Provide R code to give the components of `v` greater than or equal to 2.

(d) Find the difference `u−v` and the product `u×v`. How does R perform these operations?

(e) Explain what R does when two vectors of unequal length are multiplied together. Specifically, what is `u×c(u,v)`?

(f) Provide R code to define a sequence from 1 to 10 called `G` and to select the first three components of `G`.

(g) Use R to define a sequence from 1 to 31 with an increment of 2 named `J` and to select the first, third, and eighth values of `J`.

(h) Calculate the scalar product (dot product) of `q=(3,0,1,6)` with `r=(1,0,2,4)`.

(i) Define the matrix `X` whose rows are the `u` and `v` vectors from part (a). Do not use the `matrix()` command.

(j) Define the matrix `Y` whose columns are the `u` and `v` vectors from part (a). Do not use the `matrix()` command.

(k) Find the matrix product of `X` by `Y` and store the result in `W`.

(l) Provide R code that computes the inverse matrix of `W` and then find the transpose of that inverse.