Problem Sets with R Programming

Nith Kosal

June 13, 2021

Problem 1: Vectors

1. Create the vectors:

   (a) (1, 2, 3, ..., 29, 30)

   (b) (30, 29, 28, ..., 2, 1)

   (c) (1, 2, 3, ..., 19, 20, 19, 18, ..., 2, 1)

   (d) (44, 66, 33) and assign it to the name `futureforum`. For parts (e), (f) and (g) look at the help for the function `rep`.

   (e) (44, 66, 33, 44, 66, 33, ..., 44, 66, 33) where there are 20 occurrences of 44.

   (f) (44, 66, 33, 44, 66, 33, ..., 44, 66, 33, 44) where there are 11 occurrences of 44, 10 occurrences of 66 and 10 occurrences of 33.

   (g) (44, 44, ..., 44, 66, 66, ..., 66, 33, 33, ..., 33) where there are 10 occurrences of 44, 20 occurrences of 44, 20 occurrences of 66, 30 occurrences of 33.

2. Create a vector of the value of $e^x cos(x)$ at $x = 3, 3.1, 3.2, ..., 7$.

3. Create the following vectors:

   (a) $(0.1^3 0.2^1, 0.1^6 0.2^3, ..., 0.1^{37} 0.2^{32})$ (b) $(2, \frac{2^2}{4}, \frac{2^3}{4}, ..., \frac{2^{26}}{26})$

4. Calculate the following:

   (a) $\sum_{n=1}^{200}(i^3 + 4i^2)$ (b) $\sum_{n=1}^{25}(\frac{2^i}{i} + \frac{3^i}{i^2})$

5. Use the function `paste` to create the following character vectors of length 34.

   (a) `("Cambodia 1", "Cambodia 2", ..., "Cambodia 34")`. Note that there is a single space between `Cambodia` and the number following.

   (b) `("ffteam1", "ffteam2", ..., "ffteam29")`. Note that there are is no space between `ttteam` and the number following.

6. Execute the following lines which create two vectors of random integers which are chosen with replacement from the integers 0, 1, ..., 999. Both vector have length 244.

```
set.seed(50)
xVec <- sample(0:999, 250, replace=T)
yVec <- sample(0:999, 250, replace=T)
```

Suppose $x = (x1, x2, ..., x_n)$ denotes the vector `xVec` and $y = (y1, y2, ..., y_n)$ denotes the vector `yVec`.

(a) Create the vector $(y_2 - x_1, ..., y_n - x_{n-1})$.

(b) Create the vector $(\frac{sin(y_1)}{cos(x_2)}, \frac{sin(y_2)}{cos(x_x)}, ..., \frac{sin(y_{n-1})}{cos(x_n)})$

(c) Create the vector $(x_1 + 2x_2 - x_3, x_2 + 2x_3 - x4, ..., xn - 2 + 2x_{n-1} - x_n)$.

(d) Calculate $\sum_{i=1}^{n-1} \frac{e^{-x_{i+1}}}{x_i + 10}$

7. Use the vectors `xVec` and `yVec` created in the previous question and the functions `sort`, `order`, `mean`, `sqrt`, `sum` and `abs`.

(a) Pick out the values in `yVec` which are >400.

(b) What are the index positions in `yVec` of the values which are >400?

(c) What are the values in `xVec` which correspond to the values in `yVec` which are >400? (By correspond, we mean at the same index positions.)

(d) Create the vector $(|x_1 - \bar{X}|^{1/2}, |x_2 - \bar{X}|^{1/2}, ..., |x_n - \bar{X}|^{1/2})$ where $\bar{X}$ denotes the mean of the vector $X = (x_1, x_2, ..., x_n$.

(e) How many values in `yVec` are within 200 of the maximum value of the terms in `yVec`? (f) How many numbers in xVec are divisible by 2? (Note that the modulo operator is denoted %%.) (g) Sort the numbers in the vector `xVec` in the order of increasing values in `yVec`. (h) Pick out the elements in `yVec` at index positions 1, 4, 7, 10, 13, ...

8. By using the function `cumprod` or otherwise, calculate
$1 + \frac{2}{3} + \frac{24}{35} + \frac{246}{357} + ... + (\frac{2}{3}\frac{4}{5}...\frac{38}{39})$

Problem 2: Matrices

1. Suppose $A = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 2 & 6 \\ -2 & -4 & 8 \end{bmatrix}$

(a) Check that $A^3 = 0$ where 0 is a 3×3 matrix with entry equal to 0.

(b) Replace the third column of A by the sum of the second and third columns.

2. Create the following matrix B with 15 rows: $B = \begin{bmatrix} 20 & -20 & 20 \\ 20 & -20 & 20 \\ \cdots & \cdots & \cdots \\ 20 & -20 & 20 \end{bmatrix}$

   And than, Calculate the 3×3 matrix $B^T B$. Note: Look at the help for `crossprod`.

3. Create a 6×6 matrix `matA` with every equal to 0. Check what the functions `row` and `col` return when applied to `matA`. Hence create the 6×6 matrix:

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

4. Look at the help for the function `outer`. Hence create the following patterned matrix:

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 6 \\ 3 & 4 & 5 & 6 & 7 \\ 4 & 5 & 6 & 7 & 8 \end{pmatrix}$$

5. Create the following patterned matrices. In each case, your solution should make use of the special form of the matrix — this means that the solution should easily generalize to creating a larger matrix with the same structure and should not involve typing in all the entries in the matrix.

(a)
$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 & 0 \\ 2 & 3 & 4 & 0 & 1 \\ 3 & 4 & 0 & 1 & 2 \\ 4 & 0 & 1 & 2 & 3 \end{pmatrix}$$

(b)
$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 8 & 9 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 9 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{pmatrix}$$

(c)
$$\begin{pmatrix} 0 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 1 & 0 & 8 & 7 & 6 & 5 & 4 & 3 & 2 \\ 2 & 1 & 0 & 8 & 7 & 6 & 5 & 4 & 3 \\ 3 & 2 & 1 & 0 & 8 & 7 & 6 & 5 & 4 \\ 4 & 3 & 2 & 1 & 0 & 8 & 7 & 6 & 5 \\ 5 & 4 & 3 & 2 & 1 & 0 & 8 & 7 & 6 \\ 6 & 5 & 4 & 3 & 2 & 1 & 0 & 8 & 7 \\ 7 & 6 & 5 & 4 & 3 & 2 & 1 & 0 & 8 \\ 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 & 0 \end{pmatrix}$$

6. Solve the following system of linear equations in five unknowns

$$x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 = 7$$

$$2x_1 + x_2 + 2x_3 + 3x_4 + 4x_5 = -1$$

$$3x_1 + 2x_2 + x_3 + 2x_4 + 3x_5 = -3$$

$$4x_1 + 3x_2 + 2x_3 + x_4 + 2x_5 = 5$$

$$5x_1 + 4x_2 + 3x_3 + 2x_4 + x_5 = 17$$

by considering an appropriate matrix equation $Ax = y$. Make use of the special form of the matrix A. The method used for the solution should easily generalize to a larger set of equations where the matrix A has the same structure; hence the solution should not involve typing in every number of A.

Problem 3

1. The following table gives the size of the floor area (ha) and the price ($000), for 15 houses sold in the Canberra (Australia) suburb of Aranda in 1999.

	area	sale.price
1	694	192.0
2	905	215.0
3	802	215.0
4	1366	274.0
5	716	112.7
6	963	185.0
7	821	212.0
8	714	220.0
9	1018	276.0
10	887	260.0
11	790	221.5
12	696	255.0
13	771	260.0
14	1006	293.0
15	1191	375.0

   Type these data into a data frame with column names `area` and `sale.price`.

   (a) Plot `sale.price` versus `area`.

   (b) Use the `hist()` command to plot a histogram of the sale prices.

   (c) Repeat (a) and (b) after taking logarithms of sale prices.

2. The `orings` data frame (DAAG package) gives data on the damage that had occurred in US space shuttle launches prior to the disastrous Challenger launch of 28 January 1986.

The observations in rows 1, 2, 4, 11, 13, and 18 were included in the pre-launch charts used in deciding whether to proceed with the launch, while remaining rows were omitted. Create a new data frame by extracting these rows from `orings`, and plot `total` incidents against `temperature` for this new data frame. Obtain a similar plot for the full data set.

3. For the data frame `possum` (DAAG package)

   (a) Use the function `str()` to get information on each of the columns.

   (b) Using the function `complete.cases()`, determine the rows in which one or more values is missing. Print those rows. In which columns do the missing values appear?

4. For the data frame `ais` (DAAG package)

   (a) Use the function `str()` to get information on each of the columns. Determine whether any of the columns hold missing values.

   (b) Make a table that shows the numbers of males and females for each different sport. In which sports is there a large imbalance (e.g., by a factor of more than 2:1) in the numbers of the two sexes?

5. Create a table that gives, for each species represented in the data frame `rainforest` (DAAG package), the number of values of branch that are NAs, and the total number of cases. [Hint: Use either `!is.na()` or `complete.cases()` to identify NAs.]

6. Create a data frame called Manitoba.lakes that contains the lake?s elevation (in meters above sea level) and area (in square kilometers) as listed below. Assign the names of the lakes using the row.names() function.

Problem 4: Reading the dataset into R

1. Read the dataset in Excel into R.

2. Reading the dataset in cvs file into R.

3. Reading the dataset in Stata file into R.

4. Reading the dataset in SPSS file into R.

Problem 5: Export the dataset and the output

Problem 6: Data transformations

Problem 7: Data transformations