To print higher-resolution math symbols, click the
Hi-Res Fonts for Printing button on the jsMath control panel.

CMSC 478 — Spring 2017 — C. S. Marron
Lab 1: Introduction to R

Arrays, Matrices, and Graphs

Exercise 1: Create the matrix $C = \begin{pmatrix} 6 & 14 \\ 8 & 16 \end{pmatrix}$ by first creating the matrix

$$D = \begin{pmatrix} 1 & 5 & 9 & 13 \\ 2 & 6 & 10 & 14 \\ 3 & 7 & 11 & 15 \\ 4 & 8 & 12 & 16 \end{pmatrix}$$

and then using array indexing to extract C as a subarray of D.

Exercise 2: Use `persp()` to create a perspective plot of the function

$$\frac{\cos(x^2 + y^2)}{x^2 + y^2 + 1}$$

for x and y 100-long sequences from -5 to 5 (create x and y with `seq()`).

Exercise 3: Let A be a real, square matrix. A vector x is an *eigenvector* of A if there exists a real number λ such that

$$Ax = \lambda x,$$

in which case λ is an *eigenvalue* of A corresponding to the eigenvector x. The R function `eigen()` computes the eigenvectors and eigenvalues of a matrix. For example, the command

```
eigenA = eigen(A)
```

creates the list `eigenA` with two components: `eigenA$vectors` is an array containing the eigenvectors of A as columns and `eigenA$values` is an array containing the corresponding eigenvalues.

Create a random three-by-three matrix A as follows:

1. Use `rnorm()` to generate an array of nine random values.
2. Use `matrix()` to form the matrix A from the array of random values.

Compute the eigenvectors x_1, x_2, x_3 and eigenvalues λ_1, λ_2, and λ_3 of A and show that they satisfy the equation $Ax_i = \lambda x_i$.

Note: The operator `%*%` computes a matrix-matrix product; for example,

```
A %*% x
```

computes the product of the matrix A and vector x.

Working with Data

Exercise 4: Download the data file College.csv and load the data in R as `College`. The variables in the data set are described in the Appendix.

1. Are there any missing values? If so, omit the rows with missing values.
2. The first column is just the names of the colleges. These should be treated as row labels, not variables:

   ```
   rownames(College) = College[,1]
   fix(College)
   ```

   However, now the college names appear twice, as row labels and as a column of data. Remove the column of data:

   ```
   College = College[, -1]
   fix(College)
   ```

3. Produce a numerical summary of the variables in the data set. What is the average percentage of Ph.D.'s on the faculty? What is the average cost of books? What is the maximum expenditure per student?

Exercise 5: Use the command `attach(College)` so that you can access the variables of the data set directly (e.g. `Enroll` rather than `College$Enroll`). Produce a pairwise scatterplot matrix of the variables `Top10perc`, `Top25perc`, `PhD`, `S.F.Ratio`, `Expend`, and `Grad.Rate`.

1. Do any of the variables appear to have a positive correlation with graduation rate?
2. Why do `Top10perc` and `Top25perc` appear to be strongly related?
3. Based on the scatterplots, choose any two variables that you think may have an "interesting" relationship. Is the relationship surprising or to be expected?

Exercise 6: What are the mean and median graduation rates? Plot a histogram of graduation rates; do they appear to be approximately normally distributed? Why or why not?

Exercise 7: Say a university is *elite* if the proportion of students coming from the top 10% of their high school classes exceeds 50%. How many elite universities are there? (Don't just count them by hand!)

Hint: The `sum()` function sums the elements of an array. If the array is boolean (TRUE/FALSE), TRUE is counted as 1 and FALSE is counted as 0.

Appendix: College Data Set

Private

Public/private indicator

Apps

Number of applications received

Accept

Number of applicants accepted

Enroll

Number of new students enrolled

Top10perc

New students from top 10\% of high school class

Top25perc

New students from top 25\% of high school class

F.Undergrad

Number of full-time undergraduates

P.Undergrad

Number of part-time undergraduates

Outstate

Out-of-state tuition

Room.Board

Room and board costs

Books

Estimated book costs

Personal

Estimated personal spending

PhD

Percent of faculty with Ph.D.s

Terminal

Percent of faculty with terminal degree

S.F.Ratio

Student/faculty ratio

perc.alumni

Percent of alumni who donate

Expend

Instructional expenditure per student

Grad.Rate

Graduation rate