

# OPIM 5603 — Statistics in Business Analytics

## Fall 2019, University of Connecticut

### Homework 1

Instructions: Please complete the following questions and submit them as an R script (as an .R file) via the submission link on HuskyCT. You must submit the assignment by the time and due date listed on the course syllabus. Failure to submit a file by the deadline will result in a score of 0 on the assignment.

As with all course material, the problems appearing in this homework assignment are taken from the instructor's real-world experiences, from other courses taught at the University of Connecticut, and from the sources listed in the course syllabus.

Note that R code submitted should work independent of the data that sits in the data structure. For example, suppose there was a vector `r_vec` with the values (1, 2, 6) and the problem asks for you to create R code to create a vector `answer` which doubles each element of `r_vec`. The answer

```
answer ← c(2, 4, 12)
```

would be given no credit. The answer

```
answer ← 2*r_vec
```

would be an appropriate answer.

For all problems, be sure to add comments which will enable the graders to understand your process, as need.

If you have any questions, please submit them via email to the instructor and/or the teaching assistant prior to submitting your solution.

## Problem 1 (50 points)

Create a new R Notebook file called `Homework 1 Solution.Rmd`.

- a. Copy the header from the Lecture 1 in class notes and replace with your name, and remove `number_section`.
- b. Create a section called **Problem 1**.
- c. Create a chunk and execute a series of commands that do the following:
  - Declare a variable `x` and assign it the value of 312 divided by 12.
  - Declare another variable `y` and assign it the value of the variable created in the previous part rounded to the nearest hundredth.
  - There is a function called `ceiling` that returns the value of a variable rounded up. Use the `help` function to learn about the function. Declare a new variable `z` and assign it the value of the variable created in the first part of this problem rounded up to the nearest tenth.
- d. Vectors are objects in R that allow you to store multiple other objects in a variable. We will explore vectors in this problem.

- Execute the line:

```
cityNames = c("Shanghai", "Beijing", "Delhi", "Karachi", "Lagos", "Tianjin")
```

This creates a vector of size 6 that consists of **strings** (text objects).

- Execute the line:

```
cityNames[2]
```

Which city is output and why?

- Look up the population (in millions) and area (in square miles) of the 6 cities. Create two vectors, this time composed of numbers, which store the population and areas, respectively, of the cities. Call these vectors `population` and `area`, respectively.
- Population density is defined as the number of people per unit of area. Create a vector called `population.density` and set it equal to `population` divided by `area`. Print `population.density` to the screen.
- How interesting! R knows how to divide a vector by another vector. But what if the length of the vectors (which you can find by using the function `length`) are not equal? Create a vector called `a` which contain the 6 values 12, 17, 18, 1, 2, 6. Create another vector called `b` which contains the values 2, 3, 4. Execute the command

```
a / b.
```

What do you think R does when you divide vectors that don't have a common length?

- Create a vector called `x.rand` of 100 values, where each value is drawn for a normal distribution with mean 10 and standard deviation 3.

- Suppose we want to divide all values in positions that are odd by 2, and all the values in position that are even by 3. Execute the code and explain why it works.
- e. For each subsequent problem on this assignment, list your answers in R by writing the solutions using sections in the R notebook. Each problem should start with one comment symbol follow by **Problem X**, where **X** is the problem number. Then, for each subsequent part, the problem should be started with two comment symbols followed by **Problem X.y**, where **X** is the problem number and **y** is the subpart. For example, the section associated with this problem would be prefaced with :

```
## Problem 1.e.
```

**Problem 2 (50 points)**

The Gallup organization releases the results of recent polls at its website [www.gallup.com](http://www.gallup.com). Visit this website, read an article of interest, and answer the following questions based on the article chosen. Please be sure to keep your responses to any question limited to three sentences or less.

- a. Summarize the article, and the poll discussed in the article.
- b. Describe the population of interest, or a population of interest, if applicable.
- c. Describe the data that was collected.
- d. Describe a population parameter of interest.
- e. Describe the statistic used to estimate the parameter listed above.
- f. What was the sampling technique used? Discuss if you think this was appropriate, and what the merits of other techniques might be. If the sampling technique is not described in detail, pick which sampling technique you would have utilized and explain why this would be the most appropriate. If many are possible, pick one and elaborate.