Bios 301: Assignment 1

Due Friday, 1 October, 12:00 PM

50 points total.

Submit a single knitr (either `.rnw` or `.rmd`) file, along with a valid PDF output file. Inside the file, clearly indicate which parts of your responses go with which problems (you may use the original homework document as a template). Raw R code/output or word processor files are not acceptable.

1. **Working with data** In the `datasets` folder on the course GitHub repo, you will find a file called `cancer.csv`, which is a dataset in comma-separated values (csv) format. This is a large cancer incidence dataset that summarizes the incidence of different cancers for various subgroups. (18 points)
    1. Load the data set into R and make it a data frame called `cancer.df`. (2 points)
    2. Determine the number of rows and columns in the data frame. (2)
    3. Extract the names of the columns in `cancer.df`. (2)
    4. Report the value of the 3000th row in column 6. (2)
    5. Report the contents of the 172nd row. (2)
    6. Create a new column that is the incidence *rate* (per 100,000) for each row.(3)
    7. How many subgroups (rows) have a zero incidence rate? (2)
    8. Find the subgroup with the highest incidence rate.(3)
2. **Data types** (14 points)
    1. Create the following vector: `x <- c("5","12","7")`. Which of the following commands will produce an error message? For each command, Either explain why they should be errors, or explain the non-erroneous result. (6 points)

        ```
        max(x)
        sort(x)
        sum(x)
        ```

    2. For the next two commands, either explain their results, or why they should produce errors. (4 points)

        ```
        y <- c("5",7,12)
        y[2] + y[3]
        ```

    3. For the next two commands, either explain their results, or why they should produce errors. (4 points)

        ```
        z <- data.frame(z1="5",z2=7,z3=12)
        z[1,2] + z[1,3]
        ```

3. **Data structures** Give R expressions that return the following matrices and vectors (*i.e.* do not construct them manually). (3 points each, 12 total)
    1. $(1,2,3,4,5,6,7,8,7,6,5,4,3,2,1)$
    2. $(1,2,2,3,3,3,4,4,4,4,5,5,5,5,5)$
    3. $\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$
    4. $\begin{pmatrix} 0 & 2 & 3 \\ 0 & 5 & 0 \\ 7 & 0 & 0 \end{pmatrix}$
4. **Basic programming** Let $h(x,n) = 1 + x + x^2 + \ldots + x^n = \sum_{i=0}^{n} x_i$. Write an R program to calculate $h(x,n)$ using a `for` loop. (6 points)