

Homework 1 – Due 11:59 pm CDT, 14 September 2022

The total points on this homework is 125. Out of these 6 points are reserved for clarity of presentation, punctuation and commenting with respect to the code.

1. *This is an exercise with the partial objective of trying to get ideas from the `demo(graphics)` package in R to solve our problem.*

The dataset `student-apt.dat` has data on student scores in aptitude, mathematics, language and general knowledge for students in technical disciplines (group 1), architecture (group 2) and medical technology students (group 3) as indicated in column 1.

- (a) Read in the dataset. Note that column names are not provided in the file. So, please supply them either through the appropriate argument in the `read.table()` function, or by invoking the `names()` function after reading and storing in as a dataframe. Please read the help file on `read.table()` by using the function call `?read.table` for more information. [5 points]
 - (b) Display the observations in a set of pairwise scatter plots of the scores in aptitude, mathematics, language and general knowledge, with color to indicate the groups of the observations. [6 points]
 - (c) Comment on characteristics of the students in the three groups. [2 points]
2. The National Institute of Standards and Technology has a web page that lists the first 5,000 digits of the irrational number π . You can read these digits into R from the website <http://www.itl.nist.gov/div898/strd/univ/data/PiDigits.dat>.
 - (a) Read in the dataset. Note that the file on the website has the first 60 lines which are really different statistics on the data. These 60 lines should be skipped. Look at the `help` function on the `read.table` to see how to `skip` these lines. [7 points]
 - (b) Construct a frequency table of the digits 1 through 9. (Hint: search on terms to get an appropriate function.) [5 points]
 - (c) Construct a bar plot of the frequencies found in part (b). [5 points]
 - (d) Use the chi-square test to test the hypothesis that the digits 1 through 9 are equally probable in the digits of π . What conclusions can you draw? (Hint: use the function `chisq.test()`.) [8 points]
 3. Plot a graph that shows three curves $y = x$, $y = x^2$, and $y = \sqrt{x}$, for x from 0 to 3. Plot a vertical line at 1, 2 and 3, (curves and lines on the same plot). Hint: Decompose this problem into multiple parts: In the first instance, create a vector x consisting of 0 to 3, in increments of 0.01 (say). Create another vector $y_1 = x^2$ and a third vector $y_2 = \sqrt{x}$ and then combine them all to form a dataframe. Use the `plot` function as well as the `lines` function to add lines to an existing plot. Turn in the final plot and also the R code you used for the problem. [15 points]
 4. Consider the dataset available in R called `cars` with the help file that also has more information.
 - (a) Read in the dataset from the file. Call it `cars` (say). [2 points]
 - (b) Attach the dataframe so that the variables in the dataframe are now globally available. [1 point]
 - (c) The speeds are provided in miles per hour. Convert the speeds into feet per second and store the result in an appropriate vector. [5 points]
 - (d) Plot the speed (in feet per second) against the distance (in feet). [4 points]
 - (e) Convert the measurements into the metric system. Note that one mile is equal to 1.6093 kilometres. Store the results in appropriate vectors. [5 points]
 - (f) Detach the dataframe. [1 point]

- (g) Plot the speed (in metres per second) against the distance (in metres). [4 points]
 - (h) Make sure that the plots above are labeled and titled appropriately. Print out the plots using `dev.print()` or otherwise. What can you tell, if anything, looking at the two plots? [2 points]
5. Consider the dataset **pressure** which is in the R software base installation. You may type `help(pressure)` to get more information on this dataset.
- (a) The temperatures are provided on the Celsius scale. Convert them to the Fahrenheit scale and store them in an appropriate vector. [3 points]
 - (b) Create a dataframe consisting of the temperature in the Fahrenheit scale and Pressure. [4 points]
 - (c) Plot temperature against pressure in the Fahrenheit scale. [3 points]
 - (d) Perform a simple linear regression with temperature (in Fahrenheit) against pressure, but with no intercept in the model. Report a summary of the results and plot the fitted line on the plot in (c). Comment. [4 + 2 points]
 - (e) Plot the residuals against the fitted values. Comment. [3 + 2 points]
 - (f) Clearly pressure is not adequate to explain the relationship with temperature. Create another dataframe with four columns, given by temperature in Fahrenheit, pressure, the square of pressure and cubed pressure. [5 points]
 - (g) Use the above to perform multiple linear regression (with intercept) of temperature on the rest. What coefficients are significant? [6 points]
 - (h) On the plot of (c) above, put in the fitted line. Comment on the previous fit and this one. [3 + 2 points]
 - (i) Plot the residuals against the fitted values. Comment. [3 + 2 points]