Practical 3 – Working with data, vectorising and plotting

Basic Plotting

Q1a. Plot sin(x) for $0<x<12\pi$

Cystic Fibrosis dataset

In the H drive you will find the file `cystfibr.txt` which contains a set of measurements on a set of individuals with cystic fibrosis.

Q2. Take a look at this file using a text editor like Wordpad or Notepad.

Q3. Read the data into a data frame and attach to the data frame.

Q4. Calculate the following

(i) the number of individuals in the dataset;
(ii) the number of variables measured on each individual;
(iii) the names of the variables measured on each individual;
(iv) the mean, median, standard deviation and range of each of the variables (use the `apply` function);
(v) calculate the correlation between each pair of variables? Which pair are the most correlated?

Q5. Create the following data frames and for each one calculate the mean of each of the variables

(i) a new data frame containing just individuals older than 15
(ii) a new data frame containing just individuals with bmp in the interval [70,90]
(iii) a new data frame containing just individuals with fev1 > 30 or rv > 300

Q6. Plot a histogram for the variable height and overlay the density estimate on to the histogram using a blue line (hint : blue is col = 4)

Q7. Plot histograms for height for each sex separately, one above the other. Make sure the x-axis has the same range on both plots.

Q8. Create boxplots for the variables height, weight, bmp, fev1, rv, frc, tlc and pemax, all stratified by sex. Which have evidence of outlying observations

Q9. Use scatterplots between the variables to find any clear relationships between the variables?

Juul dataset

Q10. Read in the data from the file `juul.txt` as a data frame and attach to it.

Q11. Create summaries of the variables in this dataset?

Q12. Which variable has the most missing data?

Q13. How many individuals of each sex are there in the dataset?

Q14. Use the `table` command to create a contingency table of the factors sex and tanner.

Q15. Produce a barplot with a bar for each level of the factor tanner where each bar is split into the two levels of the factor sex.

Q16. Produce another barplot where the roles of the variables sex and tanner are reversed.

Simulation examples

Q17. (plotting) The `rgamma(n,a,b)` function simulates `n` Gamma`(a,b)` rv. Simulate 10000 Gamma `(3,4.2)` rv and make a histogram. Overlay a plot of the density. (`dgamma(x,a,b)` is the Gamma(a,b) density)

Q18. (vectorising) Simulate a Gaussian random walk with 100 steps: `X[1]=0, X[i]=X[i-1]+rnorm(1)`. Plot the walk. Try to vectorise your code.

Q19. What does the following code do? What happened to the `while()` loop?