Computing for Research I

Homework – Data manipulation and Graphics in R

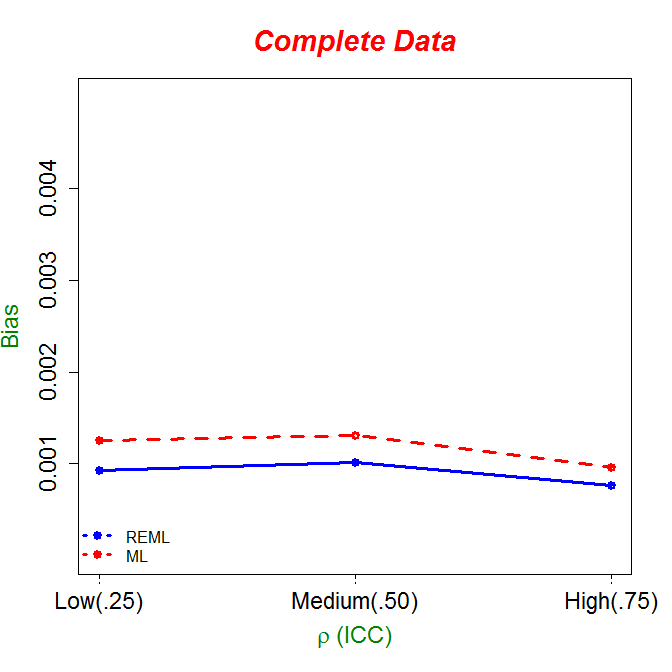
Some investigators in the Pulmonary and Critical Care Division are looking at different pleural fluid tests in order to improve the predictive accuracy of identifying exudative effusions (exudates versus transudates). To be more specific, a pleural effusion is excess fluid that accumulates between the two pleural layers, the fluid-filled space that surrounds the lungs. Excessive amounts of such fluid can impair breathing by limiting the expansion of the lungs during ventilation. A transudate is produced through pressure filtration without capillary injury while exudate is "inflammatory fluid" leaking between cells. A misclassification of the two could lead to misdiagnosis and improper treatment.

For this homework, you will do some exploratory data analyses and generate suggestive graphs using some of the pleural fluid measurements.

1. First, read data set Pleural.xls. You could read it as an .xls using an appropriate package, or transforming it first into a .csv.
2. Give a brief description of the data set, e.g., type of data set, name and number of variables. Save the output in a .txt file.
3. Create a subset retaining only the following variables: *MRN*, *Age,* *Sex*, *True diagnosis*, *pH*, *Protein*, and *Light’s criteria*. Provide some basic summary statistics for each variable, including the percent of missing data.
4. Create side-by-side box plot (or violin plot) using *pH* and *True diagnosis*. Perform an appropriate test to determine if there is a statistical significant difference in the mean pH level between the ‘exudate’ and ‘transudate’ effusions in the *True diagnosis.* First, check the necessary assumptions and provide some graphical display in order to justify the choice of the test.
5. Try reproducing the graph below by paying attention to details. Use the following vectors:

relm <- c(0.00093, 0.00102, 0.00077)

ml<-c(.00126, .00131, .00096)



Notes:

* Considering that you’ll be presenting these results to a clinical audience, please make sure that all your graphs have suggestive titles, axes labels, etc., and an overall appealing look. Except for problem 5, give a short interpretation of your graph.
* Export all graphs as .png files.
* All your answers should be put in one document except for problem 2 where you need to attach a separate .txt file.