MLW / KUHeS Statistics and R short course

Session 1 - Practical

Marc Henrion

13 November 2023

Session 1 - Practical

Go to the course website on GitHub:

https://github.com/mlw-stats/R_And_Statistics_Training_Autumn2023/Session1

From here, download the following files:

`btTBreg.csv`
`btTBregHospitals.csv`
`btTBreg_info.txt`

1. Load the `btTBreg.csv` data table into R.
2. The variables `cd41`, `cd42` and `cd41.sk`, `cd42.sk` measure the same variables (`cd4` and `cd4.sk` respectively) in the same individuals at two different time point. This means the data are in wide format. Reformat to long format.
3. Save the reformatted data into a file called `btTBregLong.tab` in such a way that
   i. Columns are tab-separated.
   ii. Column names are saved.
   iii. No row number is saved in the resulting file.
4. Load the `btTBregHospitals.csv` data table. Join the data frames storing `btTBreg.csv` and `btTBregHospitals.csv`.
5. Compute the average patient age and the proportion of male patients for each hospital.
6. Write an R function that computes the following summary statistics, then, using your custom function, compute these for the `bmi`, `cd41`, `cd42` columns:
   i. mean
   ii. median
   iii. inter quartile range
   iv. minimum
   v. maximum
   vi. number of missing values
7. Do the same now, but only for female patients. Repeat for only male patients.