

Data Intensive Analysis, exercises for day 1

Seija Sirkiä

2016-11-11

1. Create a vector of 100 random numbers with `runif(100)` and assign it to a variable. Calculate the following:
 - i) the maximum of the first 5 numbers
 - ii) how many are greater than 0.5
 - iii) the mean of squares of those between 0.25 and 0.75
 - iv) the sum of every second element
2. Make a factor representing coin tosses out of the random number vector from the previous exercise: the ones below 0.5 should become “heads” and the others “tails”.
3. Learn how to use the function `sample` and find yourself Finnish Lotto numbers for the coming Saturday (7 numbers out of the numbers 1 through 39). Also recreate the 100 coin tosses factor this way.
4. Bring a dataset called `Toothgrowth` in to use. Get familiar with it in all the ways you know, including its documentation. Are the columns vectors or factors? Should they be something else? Calculate the mean growth length for all the six groups. (Note! There are plenty of ways to do this last part with the tools covered so far, but none of them are necessarily very good)
5. Read in the data from one or more of these sources:
 - <https://github.com/CSC-IT-Center-for-Science/data-stat-course/raw/master/datasets/weather-kumpula.csv>
 - http://www.hel.fi/hel2/tietokeskus/data/helsinki/ksv/hki_liikennemaarat.csv
 - <http://www.hameenlinna.fi/Kaupunki-info/Tilastotietoa/Avoin-data/>, take the 2014 data (several files, zipped)or find another data source, or create one from scratch, or all of these. Explore the data you get.
6. Change the format of the traffic data from wide to long.
7. Working with the Hämeenlinna accounting data, find the sum of invoices per account (Tili) from the invoice data, and combine it to the accounts plan data (kirjaussuunnitelma tili) by account number. Pay attention to mismatches.