

R for Beginners

DAY 1: Exercises 2

Exercise 2.1. Using the `read.table` function and accessing variables from a data frame with epidemiological data.

The file `BirdFlu.xls` contains the annual number of confirmed cases of human Avian Influenza A/ (H5N1) for several countries reported to the World Health Organization (WHO). The data were taken from the WHO website (www.who.int/en/) and reproduced for educational purposes. Prepare the spreadsheet and import these data into R. If you are a non-Windows user, start with the file `BirdFlu.txt`. Note that you will need to adjust the column names and some of the country names. Use the `names` and `str` command in R to view the data. Print the number of bird flu cases in 2003. What is the total number of bird flu cases in 2003 and in 2005? Which country has had the most cases? Which country has had the least bird flu deaths? What is the total number of bird flu cases per country? What is the total number of cases per year?

Exercise 2.2. Using the `read.table` function and accessing subsets of a data frame with deep sea research data.

If you have not completed Exercise 6 in `Martes1`, do so and import the data from the `ISIT.xls` file. In R, extract the data from station 1. How many observations were made at this station? What are the minimum, median, mean, and maximum sampled depth at station 1? What are the minimum, median, mean, and maximum sampled depth at station 2? At station 3? Identify any stations with considerably fewer observations. Create a new data frame omitting these stations. Extract the data from 2002. Extract the data from April (of all years). Extract the data that were measured at depths greater than 2000 meters (from all years and months). Show the data according to increasing depth values. Show the data that were measured at depths greater than 2000 meters in April.

Exercise 2.3. Using the `write.table` function with deep sea research data.

In the final step of the previous exercise, data measured at depths greater than 2000 meters in April were extracted. Export these data to a new ascii file.

Exercise 2.4. Using the `factor` function and accessing subsets of a data frame with deep sea research data.

Stations 1 through 5 were sampled in April 2001, stations 6 through 11 in August 2001, stations 12 through 15 in March 2002, and stations 16 through 19 in October 2002. Create two new variables in R to identify the month and the year. Note that these are factors. Do this by adding the new variables inside the data frame.