

Biometry. Lecture 5

Alexey Shipunov

Minot State University

February 1, 2016



1 R and data

- Entering data into R
- Overgrown calculator
- Taking data into R



```
> setwd("<working folder>")  
or  
"Change dir"  
in menu!
```

(`getwd()` is for checking the folder, `dir()` checks the folder content)



R and data

Entering data into R



Putting the data file into the folder

```
> download.file("http://ashipunov.info/data/spur.txt",  
+ "data/spur.txt") # downloads file instead of scan it  
> dir("data") # should show your file name (mydata.txt)
```

Please do not type starting "+", it is used to show the line break



Reading data from a text file

```
> download.file("http://ashipunov.info/data/mydata.txt",
+ "data/mydata.txt")
> dir("data")
> file.show("data/mydata.txt") # this is a table!
> read.table("data/mydata.txt", sep=";", head=TRUE)
> f <- read.table("data/mydata.txt", sep=";", head=T)
> f
```

head=TRUE, head=T **and** h=T are the same



Reading data from text file with row names and tab separator

```
> download.file("http://ashipunov.info/data/mydata1.txt",  
+ "data/mydata1.txt")  
> file.show("data/mydata1.txt")  
> read.table("data/mydata1.txt", sep="\t", head=TRUE)
```

If the first row contains one less element, the first column will be treated as row names.

Tab separator is a “big invisible space”.

Try the same approach with `mydata2.txt` and `mydata3.txt`



R and data

Overgrown calculator



Vectorization and brackets

```
> log10(((sqrt(sum(c(2,2))))^2)*2.5) # should be 1
```

Calculation order: from inside to outside. If you want decimal logarithm, use `log10()` command.



The order of calculations

```
> 2+3*5  
is probably what you do not want, but  
> 2+(3*5)  
is!  
> 100/0  
Inf
```

TRUE, FALSE, T, F, NA, NaN, Inf and NULL are **reserved words**



Distributions

```
> rnorm(15)
```

`rnorm()` is one of almost 20 embedded distributions. By default, mean is 0 and sd is 1. You may change it, e.g., `rnorm(15, mean=10)` will give numbers around 10. If you want whole numbers, round them with `round()` command.

Normal distribution is a result of the influence of *multiple independent random factors*, dart game is a good example.



R and data

Taking data into R



R data cycle

- Enter data to spreadsheet (e.g., MS Excel)
- Save it as a text file with separators (preferably semicolons or tabs)
- Load it into R
- Work with it
- If you need to change data, go to spreadsheet and repeat first steps



R and Excel connection

There are two options: through text file or through clipboard. In Excel, make two columns with headings, copy them to clipboard, then:

```
> read.table("clipboard", h=T)
```



R and Excel connection: "readxl" package

Make 1.xlsx in Excel first, close Excel

```
> install.packages("readxl") # this is needed only once
# choose the right mirror, wait for installation
> library(readxl)
# suppose you have file "1.xlsx" in the "data" subfolder:
> read_excel("data/1.xlsx", col_names=FALSE)
```



Save your commands!

- On Windows and Unix/Linux: `savehistory()` command
- Different on Mac: it is best to save all contents of R console
- The best name for the file is probably "20160201.r"



For Further Reading



A. Shipunov.

Biometry [Electronic resource].

2012—onwards.

Mode of access:

http://ashipunov.info/shipunov/school/biol_240



A. Shipunov, and many others.

Visual statistics. Use R!

2016—onwards.

Mode of access: http://ashipunov.info/shipunov/school/biol_240/en/visual_statistics.pdf

