

Biometry. Lecture 4

Alexey Shipunov

Minot State University

January 27, 2016



- 1 Sampling with R
 - Bootstrapping in R
- 2 Overview of statistical methods
 - What is possible to find in the data
 - Descriptive and inferential methods
- 3 Data
 - How to obtain data
 - Entering data into R



- 1 Sampling with R
 - Bootstrapping in R
- 2 Overview of statistical methods
 - What is possible to find in the data
 - Descriptive and inferential methods
- 3 Data
 - How to obtain data
 - Entering data into R



- 1 Sampling with R
 - Bootstrapping in R
- 2 Overview of statistical methods
 - What is possible to find in the data
 - Descriptive and inferential methods
- 3 Data
 - How to obtain data
 - Entering data into R



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

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



Sampling with R

Bootstrapping in R



How to bootstrap the mean

```
> spur <- scan("http://ashipunov.info/data/spur.txt")  
> library(bootstrap) # we probably need to install it:  
> result <- bootstrap(x=spur, 100, function(x) mean(x))  
> mean(result$thetastar) # dollar used to call from inside
```



How to jackknife the mean

```
> result.j <- jackknife(x=spur, function(x) mean(x))  
> mean(result.j$jack.values)
```



Overview of statistical methods

What is possible to find in the data



What is possible to find in the data

- Generalities
- Comparisons
- Associations: correspondences (samples are connected), correlations (related with some strength) and regressions (related with strength and direction)
- Structure



Generalities

- Center and range
- Two ways: descriptive and inferential



Overview of statistical methods

Descriptive and inferential methods



Descriptive methods: no guesses

```
> data
# Summary: descriptive
> summary(data)
# Internal data: precipitation in main US cities
> precip
# Average of precipitation (descriptive)
> mean(precip)
# Standard deviation of precipitation (descriptive)
> sd(precip)
# Even plots (does not help much though)
> plot(precip)
```



Inferential methods: guesses

```
# Using Student's (t) test for guessing mean confidence interval  
> t.test(data)  
# Using Wilcoxon test for median confidence interval  
> wilcox.test(precip)
```



Data

How to obtain data



Observation and experiment

- Observation: minimal influence
- Experiment: direct influence



Problems of observation

- Too many irrelevant factors
- It is hard to minimize the influence



Problems of experiment

- Control group (and possibly single- or double-blind method) are needed
- Measuring of influence



Data

Entering data into R



R is more flexible than any spreadsheet

```
# There is a sequence from 1 to 9  
# Split it in three columns
```



R is more flexible than any spreadsheet

```
# There is a sequence from 1 to 9  
# Split it in three columns
```

```
> matrix(1:9, ncol=3)
```



R: how to find out the command?

```
> corr # Press Tab -- not working!  
> ?correlation # Error!  
> ??correlation # stats:cor.test found  
> ?cor.test  
> RSiteSearch("correlation")  
> example(cor.test) # useful, especially for graphs
```



Using random data

```
> set.seed(0) # this is to make same results on all computers  
> plot(rnorm(1000)) # two new commands  
> set.seed(1); rnorm(10) # you may combine two commands with ";"
```



How to create data object (vector)

```
> a <- c(1, 2, 3, 4, 5)
> a
> b <- 1:5
> b
```

Commands `c()` and `:` create vectors (sequences of numbers)



How to edit an object (vector variant)

```
> e <- edit(a) # on Linux: e <- edit(a, editor="nano")  
> fix(a) # this command changes the object and outputs nothing
```

Commands `edit()` and `fix()` both call external editor. For vectors, they will call the text editor.



How to edit an object (table variant)

```
> data.entry(b) # like (fix) -- changes object  
> de(b) # like edit() -- outputs result
```

These commands both have visual interface and will convert vector into **data frame** (table-like object).

(On Mac, does not work without special efforts).



Loading external data: working with folders

Please create the working folder `biol_240` and `data` folder inside it.

```
> getwd() # shows the name of current folder  
> setwd("c:/biol_240") # use slashes (/) instead of backslashes!  
# For Mac OS X, change "c:" to something appropriate  
> dir() # shows the content of working folder  
> dir.create("data") # will make data folder
```



Placing 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



Weevil experiment: comparing poisons

- Poison on filter paper
- Take the first weevil from a jar; put it on paper; count time; change chemical
- Why is the most effective chemical always the first?



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 "20160127.r"



Summary

- Descriptive methods *show*, inferential methods *prove*



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!

2015—onwards.

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

