

Biometry. Lecture 3

Alexey Shipunov

Minot State University

January 26, 2015



- 1 Sampling with R
 - Population and sample
 - Bootstrapping in R
- 2 Overview of statistical methods
 - What is possible to find in the data



- 1 Sampling with R
 - Population and sample
 - Bootstrapping in R

- 2 Overview of statistical methods
 - What is possible to find in the data



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

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



Sampling with R

Population and sample



Population and sample

- (Statistical) population—all research objects
- Sample—subset



Errors of sampling and complete investigation

- Representation: sample may not adequately represent population
- Accuracy: the complete (total) investigation always has less accuracy



Replication

- Replication: Every effect should be researched several times
- Every replicate should be independent!



How many replicates?

- As many as possible
- 30



Randomization

- Randomization: every object should have equal chances to be in the sample
- One of the best ways: introduce order which is knowingly absent in nature



Sampling with R

```
> data <- scan("http://ashipunov.info/data/data.txt")
> data

> sample(data, 15)
> sample(data, 15)
> sample(data, 15)
# Sampling with replacement
> sample(data, 15, replace=T) # same as "replace=TRUE"
> ?sample # calls help which tells what "replace" means
```



Sampling with R

Bootstrapping in R



Bootstrap and jackknife

- **Bootstrap** is the resampling method which replaces part of sample with the subsample of its own (sampling with replacement)
- **Jackknife** is similar but in this case observations will be taking out one by one without replacement

These methods were actively promoted from 1970s (B. Efron, 1979–1982) but only recently started to be used widely



How to bootstrap the mean

```
> data <- scan("http://ashipunov.info/data/data.txt")
> library(bootstrap) # we probably need to install it:
> # install.packages("bootstrap") # note plural and quotes
> result.b2 <- bootstrap(x=spur, 100, function(x) mean(x))
> mean(result.b2$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



Descriptive methods: no guesses

```
> data
# Summary: descriptive
> summary(data)
# it contains six values: two for centers and four for range
>
# Using Student's (t) test for checking
# how reliable is a mean estimation
> t.test(data)
```



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 "20150126.r"



Summary

- `getwd()` to check working folder; `setwd()` to change working folder
- **Replication** and **randomization** are two basic principles of research
- 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 others.

Visual statistics. Use R!

Ongoing translation from Russian.

