

Biometry. Lecture 2

Alexey Shipunov

Minot State University

January 22, 2014



- 1 Questions and answers
- 2 R
 - Non-R software
 - Starting with R
- 3 Sampling
 - Population and sample
 - Principles of sampling
- 4 Overview of statistical methods
 - What to search in the data



- 1 Questions and answers
- 2 R
 - Non-R software
 - Starting with R
- 3 Sampling
 - Population and sample
 - Principles of sampling
- 4 Overview of statistical methods
 - What to search in the data



- 1 Questions and answers
- 2 R
 - Non-R software
 - Starting with R
- 3 Sampling
 - Population and sample
 - Principles of sampling
- 4 Overview of statistical methods
 - What to search in the data



- 1 Questions and answers
- 2 R
 - Non-R software
 - Starting with R
- 3 Sampling
 - Population and sample
 - Principles of sampling
- 4 Overview of statistical methods
 - What to search in the data



Previous final question: the answer

What is sampling?



Previous final question: the answer

What is sampling?

- Taking few from many



R

Non-R software



Calculators

- Calculator is almost always embedded into OS
- Too elaborative if we use samples



Spreadsheets

- MS Excel, OpenOffice.org (LibreOffice) Calc, Gnumeric
- Very handy for data input and visualization
- Do not contain advanced and optimized statistical methods
- Are not able to conduct complex calculations



Graphical statistical software

- SPSS and MiniTab
- Have a high diversity of different graphs and plots
- Will fail if you need to repeat the complex procedures with different datasets



Statistical environments

- SAS, S-Plus and R
- Full control: it is possible to implement *every* statistical method
- User should remember commands



R

Starting with R



R history

- Started in 1993 as non-commercial analog of S-Plus
- R is just another implementation of S statistical language developed in AT&T
- In last five years, became a standard for statistical research
- Has more than 5,100 extension packages



R pros and cons

- Extremely flexible, open source
- No GUI: which command?



Installing R



Do something with R (1)

Simple math:

```
> 3+2
```

Do not enter “greater” (>) sign; at the end of each line, type “Enter”.
Be careful with lower/upper case, brackets and quotes!



Do something with R (2)

Plot and average:

```
> plot(1:20)
> mean(1:20)
> 1:20 # What is 1:20?
```

To repeat previous command, use “arrow up”.
“#” is a comment.



Do something with R (3)

Quit R:

```
> q # definition of command  
> ?q # help for command  
> q()
```



Sampling

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



Sampling

Principles of sampling



Replication

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



How many replicates?

- As many as possible
- 30



Sampling with R

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

> sample(data, 15)
> sample(data, 15)
> sample(data, 15)
```



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



Overview of statistical methods

What to search in the data



What to search in data

- Generalities
- Comparisons
- Associations: correspondences, correlations and relations
- Structure



Generalities

- Center and range
- Two ways: descriptive and inferential



Descriptive methods: no guesses

```
> data
# Summary: descriptive
> summary(data)
# Internal data: precipitation in main US cities
> precip
# Standard deviation of precipitation (descriptive)
> sd(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)
```



Final question (2 points)



Final question (2 points)

How to sample 10 items from `data` object?
Write R command.



Summary: Why do we need statistics

- **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!

DMK Press, 2012. Translated from Russian.

