Biometry. Lecture 3

Alexey Shipunov

Minot State University

January 27, 2014

Outline

Outline

Outline

Previous final question: the answer

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

Previous final question: the answer

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

```
> sample(data, 10)
```

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 <- scan("http://ashipunov.info/data/data.txt")
> data # check if everything is OK
> summary(data) # descriptive
> precip # internal data: precipitation in main US cities
> sd(precip) # standard deviation (descriptive)
```

Inferential methods: guesses

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

Comparisons

- Are two samples equal (taken from one population)?
- Greater or less?
- Most of comparisons are double, multiple comparisons are statistically dangerous.

Correspondences

- Samples are somehow connected.
- Correspondence does not show neither the strength nor direction of connection.

Correlations

- Samples are somehow related.
- Correlation shows the strength but not the direction of relation.

Regressions

- One sample is a response on the other (response ~ factor model)
- Regression measures both the strength and direction of relation
- Complex regressions are often called **models**

Structure

- Dataset contains groups, subgroups etc. (internal structure)
- Descriptive and inferential

Experimental design

- The special kind of statistics, *experimental design*, helps to plan experiments and choose a right strategy of sampling.
- Last but not least, statistics will help in:
  - Preparation of reports
  - Making reviews
  - Responding to reviews
- In general, inclusion of statistical part will make research or proposal more competitive

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

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?

Data
Entering data into R

R flexibility

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

R flexibility

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

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

R: which 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)
> fix(a)
```

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)
> de(b)
```

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

Loading external data: working with folders

Before the start, 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
```

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

Summary

- Descriptive methods *show*, inferential methods *prove*
- `getwd()` to check working folder
- `setwd()` to change working folder

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.