

# Biometry. Lecture 16

Alexey Shipunov

Minot State University

March 9, 2012

# Outline

- 1 Questions and answers
- 2 Two-dimensional statistics
  - Tests for the independence of two variables

# Outline

- 1 Questions and answers
- 2 Two-dimensional statistics
  - Tests for the independence of two variables

# Starting...

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

## Previous final question: the answer

How to decide, which two-sample test is preferable, t-test or Wilcoxon test?

## Previous final question: the answer

How to decide, which two-sample test is preferable, t-test or Wilcoxon test?

- If data is normally distributed or suspected to be normally distributed—t-test
- In all other cases—Wilcoxon test

# Two-dimensional statistics

## Tests for the independence of two variables

# “Classical” sleep data and model formula

```
> str(sleep)
> boxplot(extra ~ group, data=sleep)
> t.test(extra ~ group, data=sleep)
```

sleep is a data in so-called long format, `extra ~ group` is a **model formula** of response ~ factor form.  
“group” should have exactly 2 levels!

# Model formula for leaves data

```
> leaves <- read.table(  
+ "http://ashipunov.info/data/leaves.txt", h=T)  
> leaves12 <- stack(leaves[,1:2])  
> leaves12  
> t.test(values ~ ind, data=leaves12, paired=T)
```

`stack()` converts from short to long form

# Differences in leaf growth

```
> difc <- leaves[,2] - leaves[,1]
> dife <- leaves[,4] - leaves[,3]
> difce <- stack(data.frame(difc, dife))
> boxplot(values ~ ind, data=difce)
> t.test(values ~ ind, data=difce)
```

# Air quality data in May and August

```
> str(airquality)
> air15 <- unstack(airquality[,c(1,5)])
> Normality3(air15)
> boxplot(Ozone ~ Month, data=airquality,
+ subset=Month %in% c(5,8))
> wilcox.test(Ozone ~ Month, data=airquality,
+ subset=Month %in% c(5,8))
```

`unstack()` converts from long to short form  
`%in%` is a selection operator

# Two main questions

- Normal?
- Paired?

# Finishing...

```
> savehistory("20120309.r")
```

# Final question (10 points!)

## Final question (10 points!)

These are points from the first and second exam in one small class:

63, 72, 77, 76, 67, 56, 55, 51, 77, 64

and

87, 86, 76, 79, 54, 60, 97, 80, 73, 97

Both exams were equivalent. Provide a statistical support for the hypothesis that second exam went better. Report commands and all values which support your conclusion.

# Summary: most important commands

- `response ~ factor`—if factor has exactly two levels, this is a model formula for two-sample test
- `%in%`—selection operator

# For Further Reading



A. Shipunov.

*Biometry* [Electronic resource].

2012—onwards.

Mode of access: [http:](http://)

[//ashipunov.info/shipunov/school/biol\\_299](http://ashipunov.info/shipunov/school/biol_299)



P. Dalgaard

*Introductory Statistics with R*. 2nd edition.

Springer, 2008.

*Chapter 5.*