

# Biometry. Lecture 1

Alexey Shipunov

Minot State University

January 11, 2012

# Outline

- 1 Course in general
  - Description
  - Grading
  - Course schedule
- 2 Why do we need statistics
  - What is statistics
  - Data
  - Samples

# Outline

- 1 Course in general
  - Description
  - Grading
  - Course schedule
- 2 Why do we need statistics
  - What is statistics
  - Data
  - Samples

# Course in general

## Description

# Course description

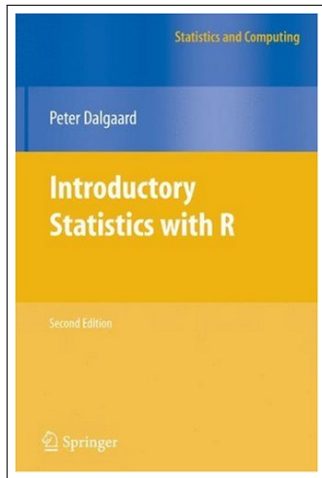
Course will cover introductory statistic concepts in a form designed specifically for biology majors, its goal is to strengthen Biology and Chemistry students statistical knowledge and abilities. It is a practical, software-based examination of the concepts of sampling, hypotheses testing (non-parametric and parametric), descriptive statistics, contingency, correlation, analysis of variation, linear models and basic multivariate techniques. Only biological, real-world data will be used. Course will concentrate on underlying principles, applicability and practical use of methods covered. R statistical environment will be used as a main software tool. The course will be accepted as substitute for Math 240.

- what is data and how to process it;
- what are statistical hypotheses;
- how to get answers from one-, two- and multidimensional data

# Instructor

- Dr. Alexey Shipunov
- Office: Moore 229
- Office Hours: Wednesdays and Fridays, 9 a.m. to 12 a.m.
- Phone: 858-3116
- E-mail: `alexey.shipunov@minotstateu.edu`

# Details



**Lectures** Mondays (TBA),  
Wednesdays and  
Fridays, 8 a.m. to  
8:50 a.m., Moore 213

**Laboratories** Mondays, Moore 213

**Textbook** Introductory Statistics  
with R (Peter  
Dalgaard, 2nd ed.,  
Springer)

# Course in general

## Grading

# Exams

- Four **equal** exams are given during the semester.
- Only the **three** best exams contribute to the final grade.
- Missed exams count zero points. There are **no make-up** exams.

# Labs

- Receiving zero points for **more than one** laboratory results in a failed course.
- Grading of laboratories is based on reports.
- Written reports are prepared and finished during laboratory sessions and sent via e-mail or passed to the instructor right after the particular laboratory session.
- It is expected that you have reviewed the lecture contents before you come to lab.

# Absence

There are five legitimate reasons for absence:

- 1 emergency situations,
- 2 attested medical conditions,
- 3 military duty,
- 4 participation in MSU sports events,
- 5 dependent sick leave.

Absence from exams or laboratories needs to be announced to the instructor in advance **via e-mail**. I strongly recommend attending lectures regularly. Statistically, students who achieved best grades are **always attend lectures**.

# Lecture tests

- At the end of **every** lecture I will give **one** short test question to answer.
- The question will require 1–3 min to answer and respectively, will give from 1 to 3 points (depending on the complexity).

# Points

A total of 600 points can be earned and are distributed as follows:

- Three best exams: 300 points
- Lecture tests: 60 points total
- Laboratory: 240 points (20 points per lab)

Grading points may vary between exams, tests, and labs.

# Letter grades

- $A \geq 90\%$
- $B \geq 80\%$
- $C \geq 70\%$
- $D \geq 60\%$
- $F < 60\%$

A minimum of one letter grade will be deducted from the grade for academic dishonesty / plagiarism.

# Course in general

## Course schedule

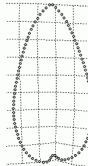
# Tentative course sequence

- Data and data processing
- How to process data: R basics
- Types of data
- One-dimensional data, descriptive statistics
- Contingency tables
- Correlation
- Regression
- ANOVA
- Multidimensional data, data mining

# Course Web site

© Shipunov, A. Biometry [Electronic resource]. 2012—onwards.  
Mode of access: [http://ashipunov.info/shipunov/school/biol\\_299/](http://ashipunov.info/shipunov/school/biol_299/)

## BIOL 299: Biometry



Course materials:

- [Syllabus](#) (PDF, 0.15 Mb)



[Back](#)

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

# Why do we need statistics

## What is statistics

# Definition of Statistics

**Data** Collecting any numerical data, e.g. unemployment rate per state.

**Samples** Working with any subsets (samples) of data, like voting polls.

**Tools** Procedures used to analyze data, such as ANOVA or chi-square statistic.

**Science** Science that develops mathematical procedures to describe data.

In all, statistics is about data.

# Why do we need statistics

## Data

# Small data

- Small data is often self-explanatory.
- Experiments with cognition show that it is easy to operate with 5-9 objects in mind.
- Visual inspection gives an average value close to 2.

2 3 4 2 1 2 2 0

# Uniform data

```
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 2 2 2 2 2
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
2 2 2 2 2 2 2 2
```

- Visual inspection again gives an average value close to 2.
- Uniform data could be (relatively) big, but understandable without special tools.

# Real data

Data from Shipunov et al., *in review*

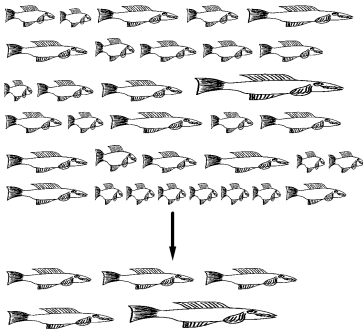
```
88 22 52 31 51 63 32 57 68 27 15 20 26 3 33 7
35 17 28 32 8 19 60 18 30 104 0 72 51 66 22 44
75 87 95 65 77 34 47 108 9 105 24 29 31 65 12
82
```

- However, in most cases biological data is much more complicated.
- Therefore, we will need specific (statistical) tools even for preliminary description of data.

# Why do we need statistics

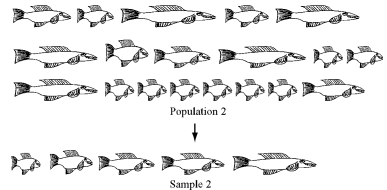
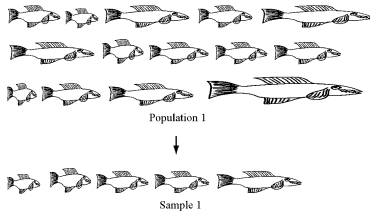
## Samples

# Sampling



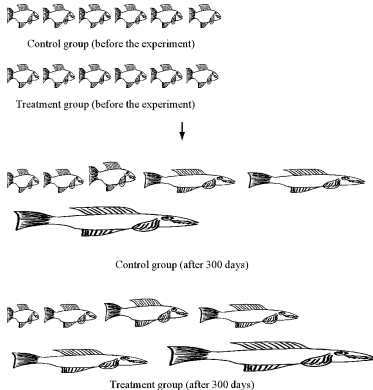
- Biologists often work with large numbers of objects and therefore need to sample (subset) initial population.
- However, the sample may not necessary be a good representative of a population.
- Only statistical tools will help to determine the reliability of the sample.

# Comparing two populations



- Even samples chosen at random from two different populations may not necessary be different.
- Only statistics will help to recognize “true” difference from “false” difference.

# Experiments



- Biologists often conduct experiments. However, natural variation among individuals within a sample may obscure any effect of an experimental treatment.
- Again, only careful examination of samples with appropriate tools will make results of experiment robust.



break:

# Installing R

# Final question (2 points)

## Final question (2 points)

What is sampling?

# Summary

Statistics is:

- Gathering data
- Making samples
- Applying tools
- Develop new ways of things above

# For Further Reading



A. Shipunov.

*Biometry* [Electronic resource].

2012—onwards.

Mode of access: [http:](http://ashipunov.info/shipunov/school/biol_299)

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



P. Dalgaard

*Introductory Statistics with R*. 2nd edition.

Springer, 2008.

*Appendix A*.