Federal State Autonomous Educational Institution of Higher Education "Moscow Institute of Physics and Technology (National Research University)"

APPROVED
Head of the Phystech School of Biological and Medical Physics
D.V. Kuzmin

Work program of the course (training module)

course: Basic Biostatistics/Базовая биостатистика
major: Applied Mathematics and Physics
specialization: Applied Bioinformatics/Прикладная биоинформатика
Phystech School of Biological and Medical Physics
Chair of Bioinformatics and Systems Biology
term: 1
qualification: Master

Semester, form of interim assessment: 1 (fall) - Exam

Academic hours: 60 AH in total, including:
lectures: 30 AH.
seminars: 30 AH.
laboratory practical: 0 AH.

Independent work: 45 AH.

Exam preparation: 30 AH.

In total: 135 AH, credits in total: 3

Authors of the program:

A.V. Rubanovich, doctor of biological sciences
Y.A. Medvedeva, candidate of biological sciences

The program was discussed at the Chair of Bioinformatics and Systems Biology 04.06.2020

Annotation

The aim of the course is give an idea of the mechanisms of the implementation of genetic information, biostatistics and practical skills in applying bioinformatics and statistical methods for analyzing and interpreting biological data. After completing the course, the student will understand the main methods of assessing statistical significance, methods for allowing for the multiplicity of comparisons, Meta-analysis methods, statistical characteristics of associative tests, ROC analysis, methods for assessing heritability and genetic risks, methods for reducing the number of variables when analyzing large data arrays, data classification methods, the basics of Bayesian data analysis.

1. Study objective

Purpose of the course

Give an idea of the mechanisms of the implementation of genetic information, biostatistics and practical skills in applying bioinformatics and statistical methods for analyzing and interpreting biological data.

Tasks of the course

- give an idea of the basic methods of statistical analysis of biological data;
- make students familiar with modern understanding of statistical populational studies;
- teach how to use the main databases in the field;
- introduce basic algorithms and data formats for statistical genetics and biostatistics.

2. List of the planned results of the course (training module), correlated with the planned results of the mastering the educational program

Mastering the discipline is aimed at the formation of the following competencies:

Code and the name of the competence	Competency indicators
UC-1 Use a systematic approach to critically analyze a problem,and develop an action plan	UC-1.1 Systematically analyze the problem situation, identify its components and the relations between them
	UC-1.2 Search for solutions by using available sources
	UC-1.3 Develop a step-by-step strategy for achieving a goal, foresee the result of each step, evaluate the overall impact on the planned activity and its participants
Gen.Pro.C-1 Gain fundamental scientific knowledge in the field of physical and mathematical sciences	Gen.Pro.C-1.1 Apply fundamental scientific knowledge in the field of physical and mathematical sciences
	Gen.Pro.C-1.2 Consolidate and critically assess professional experience and research findings
	Gen.Pro.C-1.3 Understand interdisciplinary relations in applied mathematics and computer science and apply them in professional settings

3. List of the planned results of the course (training module)

As a result of studying the course the student should:

know:

- the main methods of assessing statistical significance;
- methods for allowing for the multiplicity of comparisons;
- Meta-analysis methods;
- statistical characteristics of associative tests;
- ROC analysis;
- methods for assessing heritability and genetic risks;
- methods for reducing the number of variables when analyzing large data arrays;
- data classification methods;
- the basics of Bayesian data analysis.

be able to:

- use the Internet and reference books on scientific and applied biostatistics character to quickly find the necessary data and concepts;
- compare statistical processing methods and adequately evaluate them applicability;
- apply basic methods of biostatistics in scientific research;
- apply basic methods of biostatistics when working in the laboratory.

master:

- of conducting large scale data arrays;
- of computer analysis of the statistical significance of the results of genetic and medical/biological experiments.

4. Content of the course (training module), structured by topics (sections), indicating the number of allocated academic hours and types of training sessions

4.1. The sections of the course (training module) and the complexity of the types of training sessions

№	Topic (section) of the course	Types of training sessions, including independent work			
		Lectures	Seminars	Laboratory practical	Independent work
1	Biological data structure and descriptive statistics	10	6		10
2	Conjugacy analysis	6	10		12
3	Multidimensional methods	10	4		10
4	Bayesian statistics	4	10		13
AH in total		30	30		45
Exam preparation		30 AH.			
Total complexity		135 AH., credits in total 3			

4.2. Content of the course (training module), structured by topics (sections)

Semester: 1 (Fall)

1. Biological data structure and descriptive statistics

File organization and data management in EXCEL, SPSS and STATISTICA. Descriptive statistics. Some tricks of fast statistical calculations. Statistical hypothesis testing. Exact and Mediated Criteria. I and II type eroors. Multiple comparison testing. Type I error control. Grouping and Simpson's Paradox. Parametric and non-parametric comparison criteria. Variance analysis.

2. Conjugacy analysis

Regression analysis. Residual analysis. Partial correlations and confounders. Contingency analysis of qualitative features. Odds ratio and relative risk. Biomarker statistics. Estimates of the test sensitivity and specificity. ROC analysis.

3. Multidimensional methods

Multiple regression analysis. Methods to reduce the number of predictors. Friedman paradox. Estimates of heritability and genetic risk. The “missing heritability” problem. Factor analysis. The principal component method. Classification methods. Cluster analysis. Discriminant analysis.

4. Bayesian statistics

Limitation of p-values. Reproducibility of experimental results. Bayesian factor. Priors. Statistics in epidemiology. Analysis of large samples. Bayesian frequency estimates of rare events.

5. Description of the material and technical facilities that are necessary for the implementation of the educational process of the course (training module)

Equipment needed for lectures and seminars: computer and multimedia eqipment (projector, sound system).

6. List of the main and additional literature, that is necessary for the course (training module) mastering

Main literature

Provided at the department:

1) Biswas, A., Datta, S., Fine, J. P., Segal, M. R. (eds.) Statistical Advances in the Biomedical Sciences Clinical Trials, Epidemiology, Survival Analysis, and Bioinformatics, 2008, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany

2) Statistical Human Genetics. Edited by Robert C. Elston. Springer Science+Business Media, LLC 2012

Additional literature

Provided at the department:

1) Topics in Biostatistics. Edited by Walter T. Ambrosius2007 Humana Press Inc. 999 Riverview Drive, Suite 208, Totowa, New Jersey07512

2) Agostino Di Ciaccio, Mauro Coli, Jose Miguel Angulo IbaQnez. Advanced Statistical Methods for the Analysis of Large Data-Sets. Springer-VerlagBerlinHeidelberg, 2012

7. List of web resources that are necessary for the course (training module) mastering

1) www.ncbi.nih.nlm.gov
2) www.scopus.com

8. List of information technologies used for implementation of the educational process, including a list of software and information reference systems (if necessary)

For some of the lessons, you will need Zoom. Google Drive to access course materials. The presence of smartphones / laptops during classes is encouraged to participate in interactive exercises.

9. Guidelines for students to master the course

A student who studies discipline must, on the one hand, master a general conceptual apparatus, and on the other hand, must learn to apply theoretical knowledge in practice.

As a result of studying the discipline, the student should know the basic definitions of the discipline, be able to apply this knowledge to solve various problems.

Successful learning requires:

- visits to all classes provided by the curriculum for the discipline;
- conducting the abstract of occupations;
- intense independent work of the student.

Independent work includes:

- reading recommended literature;
- study of educational material, preparation of answers to questions intended for self-study;
- solving problems offered to students in the classroom;

- preparation for performance of tasks of the current and intermediate certification.

An indicator of possession of the material is the ability to answer questions on discipline topics without an outline.

It is important to achieve an understanding of the material being studied, and not its mechanical memorization. If it is difficult to study individual
topics, questions, you should seek advice from the teacher.

Intermediate control of students' knowledge in the form of problem solving in accordance with the subject of classes is possible

SUPPLEMENT

Assessment funds for course (training module)

major:	Applied Mathematics and Physics
specialization:	Applied Bioinformatics/Прикладная биоинформатика Phystech School of Biological and Medical Physics Chair of Bioinformatics and Systems Biology
term:	1
qualification:	Master

Semester, form of interim assessment: 1 (fall) - Exam

Authors:

A.V. Rubanovich, doctor of biological sciences
Y.A. Medvedeva, candidate of biological sciences

1. Competencies formed during the process of studying the course

Code and the name of the competence	Competency indicators
UC-1 Use a systematic approach to critically analyze a problem,and develop an action plan	UC-1.1 Systematically analyze the problem situation, identify its components and the relations between them
	UC-1.2 Search for solutions by using available sources
	UC-1.3 Develop a step-by-step strategy for achieving a goal, foresee the result of each step, evaluate the overall impact on the planned activity and its participants
Gen.Pro.C-1 Gain fundamental scientific knowledge in the field of physical and mathematical sciences	Gen.Pro.C-1.1 Apply fundamental scientific knowledge in the field of physical and mathematical sciences
	Gen.Pro.C-1.2 Consolidate and critically assess professional experience and research findings
	Gen.Pro.C-1.3 Understand interdisciplinary relations in applied mathematics and computer science and apply them in professional settings

2. Competency assessment indicators

As a result of studying the course the student should:

know:

- the main methods of assessing statistical significance;
- methods for allowing for the multiplicity of comparisons;
- Meta-analysis methods;
- statistical characteristics of associative tests;
- ROC analysis;
- methods for assessing heritability and genetic risks;
- methods for reducing the number of variables when analyzing large data arrays;
- data classification methods;
- the basics of Bayesian data analysis.

be able to:

- use the Internet and reference books on scientific and applied biostatistics character to quickly find the necessary data and concepts;
- compare statistical processing methods and adequately evaluate them applicability;
- apply basic methods of biostatistics in scientific research;
- apply basic methods of biostatistics when working in the laboratory.

master:

- of conducting large scale data arrays;
- of computer analysis of the statistical significance of the results of genetic and medical/biological experiments.

3. List of typical control tasks used to evaluate knowledge and skills

During the current control, the student should be able to answer the following questions:

1) Odds ratio and relative risk
2) Multiple regression and the Friedman paradox
3) Methods for assessing publication shift. Funnel Charts.
4) FDR-method of accounting for multiple comparisons
5) Bayesian estimates of the frequencies of rare events

During the class, interactive discussions can take place in the course chats, which will be homework. It is possible to perform patent search as an independent task. Successful completion of all tasks in the course and the implementation of control slices of knowledge gives an advantage on exam.

4. Evaluation criteria

1) Odds ratio and relative risk
2) Multiple regression and the Friedman paradox
3) Methods for assessing publication shift. Funnel Charts.
4) FDR-method of accounting for multiple comparisons
5) Bayesian estimates of the frequencies of rare events
Typical tickets:
Ticket number 1
Strata and the Simpson paradox.
Ticket number 2
Analysis of variance
Ticket number 3
Private correlations and confounders.
Ticket number 4
Biomarker Statistics
Ticket number 5
Sensitivity and specificity of tests
Ticket number 8
ROC analysis

The mark is excellent (10 points) - it is given to a student who has shown comprehensive, systematic, deep knowledge of the curriculum of the discipline, who has an interest in this subject area, has demonstrated the ability to confidently and creatively put them into practice in solving specific problems, and a free and proper substantiation of decisions.

The mark is excellent (9 points) - it is given to a student who has shown comprehensive, systematic, in-depth knowledge of the curriculum of the discipline and the ability to confidently put them into practice in solving specific problems, free and proper substantiation of the decisions made.

The mark is excellent (8 points) - given to a student who has shown comprehensive, systematic, in-depth knowledge of the curriculum of the discipline and the ability to confidently apply them in practice in solving specific problems, correct justification of decisions made, with some shortcomings.

A mark is good (7 points) - it is put up for a student, if he knows the material firmly, sets it up competently and in essence, knows how to apply the knowledge gained in practice, but does not competently substantiate the results obtained.

Evaluation is good (6 points) - it is put up to a student, if he knows the material firmly, sets it up correctly and in essence, knows how to apply this knowledge in practice, but admits some inaccuracies in the answer or in solving problems.

A mark is good (5 points) - it is given to a student, if he basically knows the material, correctly and essentially sets it out, knows how to apply this knowledge in practice, but allows a sufficiently large number of inaccuracies to answer or solve problems.

Grade satisfactorily (4 points) is given to a student who has shown the fragmented, fragmented nature of knowledge, insufficiently correct formulations of basic concepts, violations of the logical sequence in the presentation of program material, but at the same time he has mastered the main sections of the curriculum necessary for further education and can apply knowledge is modeled in a standard situation.

Grade satisfactorily (3 points) - given to a student who showed the fragmented, scattered nature of knowledge, making mistakes in formulating basic concepts, disrupting the logical sequence in presenting program material, poorly masters the main sections of the curriculum required for further education and even applies the knowledge gained in a standard situation.

The rating is unsatisfactory (2 points) - is given to a student who does not know most of the main content of the curriculum of the discipline, makes gross mistakes in the wording of the basic principles and does not know how to use this knowledge when solving typical tasks.

Unsatisfactory mark (1 point) - is given to a student who does not know the main content of the discipline's curriculum, makes gross errors in the wording of the basic concepts of the discipline and does not have any skills to solve typical practical problems.

5. Methodological materials defining the procedures for the assessment of knowledge, skills, abilities and/or experience

During the oral exam, the student is given 30 minutes to prepare. The interview for a student in an oral exam must not exceed one astronomical hour.