Course unit (module) title	Code
Multivariate statistics with R	

Lecturer(s)	Department(s) where the course unit (module) is delivered
Coordinator: A. Jakaitienė, Prof., PhD **Other(s): N. Bratčikovienė, Assoc. Prof., PhD**	Department of Human and Medical Genetics Vilnius University Faculty of Medicine, M.K. Čiurlionio str. 21, LT-03101, Vilnius

Study cycle	Type of the course unit (module)
Second cycle	Elective

Mode of delivery	Period when the course unit (module) is delivered	Language(s) of instruction
Face-to-face, self-study Lectures, seminars and practice	1st semester	English

Requirements for students	
Prerequisites: Basic concepts of Statistics	**Additional requirements (if any):**

Course (module) volume in credits	Total student's workload	Contact hours	Self-study hours
5	136	68	68

Purpose of the course unit (module): programme competences to be developed

The course introduces statistical methods and underlying concepts for data analysis with a focus on systems biology. The course emphasises modern computational approaches using the statistics software R.
The aim of this course is two-fold: first, students will learn to matrix algebra which will be background for multivariate statistics and mathematical modelling. Second, students will learn standard methods from statistics with application to systems biology datasets of medium complexity. In addition, students should gain a good understanding of the underlying principles and concepts in order to be able to choose from the vast set of available statistical tests and methods and critically employ them.

Learning outcomes of the course unit (module)	Teaching and learning methods	Assessment methods
1.2. Be able to analyse, manage and model statistical data from the field of system biology	Lectures, debates, group discussion and practical assignments	Completion of practical assignments (two tests); Analysis of research paper (written project); Written examination.
2.1. Be able select an appropriate statistics test or model for a given biological domain and problem		
3.2. Perceive statistical methods used in evolutionary processes of biological systems		
3.3. Be able to process advanced statistical data		
3.5. Be able to apply linear algebra to describe evolutionary processes of biological systems		
4.1. Perform duties within the deadlines and goals of a project		
5.1 Be able to work autonomously and as a part of a multidisciplinary team; act honestly and according to ethical obligations		
5.2. Be able to critically analyse their own research quantitative results and know possible ways for improvement		

Content: breakdown of the topics	Contact hours							Self-study work: time and assignments	
	Lectures	Tutorials	Seminars	Exercises	Laboratory work	Internship/work	**Contact hours**	**Self-study hours**	**Assignments**
1. Matrix algebra Tutor: N. Bratčikovienė	4			4			**8**	**8**	Self-study of J. Stewart, T. Day Ch. 8.4 - 8.6, p. 514-530 N. Fieller Ch. 2, p 21-50 and Ch. 7, p. 103-111. M. J. Crawley Ch. 4. p. 159-189. Preparation for practice assignments.
2. Determinants. Orthogonality. Tutor: N. Bratčikovienė	4			4			**8**	**8**	Self-study of J. Stewart, T. Day Ch. 8.3&8.6, p. 505-514 & 531-537 N. Fieller Ch. 4-5, p 59-67. Preparation for practice assignments.
3. Eigenvalues and eigenvectors. Tutor: N. Bratčikovienė	2			2			**4**	**4**	Self-study of J. Stewart, T. Day Ch. 8.7, p. 537-544 N. Fieller Ch. 6, p 83-101. Preparation for practice assignments.
4. Introduction to multivariate regression. Tutor: A. Jakaitienė	2			2			**4**	**4**	Selection of two scientific papers for the analysis in written report. Compilation of a pocket guide of R commands and scripts for a standard data analysis.
5. Parametric and non-parametric hypothesis testing. Tutor: A. Jakaitienė	4			4			**8**	**8**	N. Fieller Ch. 9, p 143-150. M. J. Crawley Ch. 8. p. 344-388. Self-study of the use of G*power program.
6. Linear and non-linear regression models. Tutor: A. Jakaitienė	10		2	10			**22**	**22**	M. J. Crawley Ch. 10. p. 449-497. Preparation for the presentations of written scientific papers overview.

7. Factor analysis. Cluster analysis. Tutor: A. Jakaitienė	6		2	6			**14**	**14**	To read material in web pages provided by a professor and prepare for the class discussion. Preparation for the presentations of written scientific papers overview.
Total	**32**		**4**	**32**			**68**	**68**	

Assessment strategy	**Weigh t,%**	**Deadline**	**Assessment criteria**
Two tests in computer lab. Test 1 Test 2	 15 15	 After topic 3 After topic 7	Two tests consist of various different complexity exercises, those should be completed using R functions or commands. The scoring of each task is presented. Maximum grade of the test is 10 points. The evaluation criteria for the test of the practical assignments are presented to the students in writing one week before the test.
Scientific paper analysis Presentation of scientific paper analysis	15 5	During 5-7 topic	For the scientific paper analysis, students overview and analyse statistical part of one scientific paper. Written project supposed to be presented and discussed with colleagues. Maximum grade for written work and presentation 10 points each. <u>Written project assessment criteria (2 points each):</u> Consistency of presentation; Arguments; Generalization; Correct statistics; Correct citation and references. <u>Presentation assessment criteria:</u> Clear selection of the papers (2 points); Quality of speech (clearity, distinction) (1.5 point); Eye contact with audience (2 points); Quality of visually presented material (1.5 point); Management of questions (quality of answers to the presented questions) (2 points); Management of time (is the time given for presentations used properly) (1 point).
Written exam	50	After Test 2	Test type written exam. The scoring for each exam question is given. Maximum grade of the exam test is 10 points. The evaluation criteria of exam questions are presented to the students in writing at the last lecture.

Author	**Year of public ation**	**Title**	**Issue of a periodical or volume of a publication**	**Publishing place and house or web link**
Compulsary reading				
J. Stewart, T. Day	2015	Biocalculus: Calculus for Life Sciences		Cengage Learning
Nick Fieller	2015	Basics of Matrix Algebra for Statistics with R		Chapman & Hall/CRC The R Series
M. J. Crawley	2013	The R Book.	Second Edition	Willey

Edited J. Ball, V. Bewick and L. Cheek	2005	Medical statistics		https://www.biomedcentral.com/collections/CC-Medical
Power calculation software		G*Power		http://www.gpower.hhu.de/en.html
Optional reading				
D. Bowers.	1996	Statistics from scratch: an introduction for health care professionals.		Wiley
B. Shahbaba.	2012	Biostatistics with R. An Introduction to Statistics Through Biological Data		New York, Springer, http://www.ics.uci.edu/~babaks/BWR/Home.html
G. James et al.	2017	An Introduction to Statistical Learning with Applications in R	ISBN 978-1461471370	http://www-bcf.usc.edu/~gareth/ISL/
Open source environment R		R Project		www.cran.org
J. Dadonienė, K. Žagminas, A. Beržanskytė.		Introduction to research methodology		Vilniaus university, 2013. http://www.vu.lt/site_files/LD/Introduction_methodology_2013.pdf

****