Course unit (module) title	Code
Multivariate statistics with R	

Lecturer(s)	Department(s) where the course unit (module) is delivered
Coordinator: Prof. A. Jakaitienė, PhD **Other(s): Assoc. Prof. N. Bratčikovienė, PhD; Assoc. Prof. T. Žvirblis, 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	Compulsory

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 assignment; 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 placement	Contact hours	Self-study hours	Assignments
1. Acquaintance with the R software Tutor: N. Bratčikovienė				2			**2**	**2**	Self study of W.J.Owen, M.J.Crawley
2. Matrix algebra Tutor: N. Bratčikovienė	2						**2**	**2**	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	3			3			**3**	**3**	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.
4. Eigenvalues and eigenvectors. Tutor: N. Bratčikovienė	3			3			**3**	**3**	Self-study of J. Stewart, T. Day Ch. 8.7, p. 537-544 N. Fieller Ch. 6, p 83-101. Preparation for practice assignments.
5. Probabilities and their properties. Univariate distributions, multivariate normal distribution. Tutor: N. Bratčikovienė	3			3			**6**	**6**	Self-study of Mathai & Haubold. p. 1-56, 61-72, 147-151, 233-241
6. Initial data analysis (IDA). Descriptive statistics. Hypothesis testing. p-value. Normality test. Tutor: A. Jakaitienė	4			3			**7**	**7**	Compilation of a pocket guide of R commands and scripts for a standard data analysis.
7. t-test. ANOVA. Nonparametric hypothesis. Transformations. Chi-squared tests. Sample size and power. Test for Proportions. Tutor: A. Jakaitienė	7			6			**13**	**13**	N. Fieller Ch. 9, p 143-150. M. J. Crawley Ch. 8. p. 344-388. Self-study of the use of G*power program.
8. Linear and non-linear regression models. Tutor: T. Žvirblis	8			12			**20**	**20**	M. J. Crawley Ch. 10. p. 449-497.

9. Factor analysis. Cluster analysis. Tutor: A. Jakaitienė	6						**6**	**6**	To read material in web pages provided by a professor and prepare for the class discussion.
Total	**36**			**32**			**68**	**68**	

Assessment strategy	**Weigh t,%**	**Deadline**	**Assessment criteria**
Test	20	After topic 5	The test consists 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. A test is considered as passed if the score is 5 or higher. There are no retakes during the semester.
Project	30	During the course	In the project, students apply the acquired statistical knowledge by examining freely available data of their choice. The result of the project is evaluated on a scale of 10 points. Assessment strategy of a project is available to students in VU VLE (https://emokymai.vu.lt). A project is considered as passed if the score is 5 or higher. Samples of exam questions and assessment strategies as well criteria for practice assessments, together with all course materials, are available to students through VU VLE (https://emokymai.vu.lt). Failure in a test or project will result in no exam. If you do not attend the exam, the cumulative score is not calculated and the exam report is marked as absent.
Written exam	50	After topic 9	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 VU VLE (https://emokymai.vu.lt). A student is not allowed to take the exam if she/he fails test or project. If a student does not attend the exam, the cumulative score is not calculated and is marked as absent. During the course cycle, there is no retake of test or project. An exam is considered as passed if the score is 5 or higher. In VU VLE (https://emokymai.vu.lt) students will find all course materials.

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
Mathai, Arak M. and Haubold, Hans J..	2018	Probability and Statistics: A Course for Physicists and Engineers		Berlin, Boston: De Gruyter https://doi.org/10.1515/9783110562545
Owen, W. J.	2010	The R Guide.		https://cran.r-project.org/doc/contrib/Owen-TheRGuide.pdf
Torfs, P., & Brauer, C.	2014	A (very) short introduction to R.	Hydrology and Quantitative Water Management Group,	https://cran.r-project.org/doc/contrib/Torfs+Brauer-Short-R-Intro.pdf

			Wageningen University, The Netherlands, 1-12.	
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
David M. Lane, Rice University		Online Statistics education: An interactive Multimedia Course of Study		https://onlinestatbook.com/2/index.html
J. Dadonienė, K. Žagminas, A. Beržanskytė.		Introduction to research methodology		Vilniaus university, 2013. http://www.vu.lt/site_files/LD/Introduction_methodology_2013.pdf

****