DOCTORAL (PHD) STUDIES
COURSE DESCRIPTION

Course title	Field of science	Faculty	Institute
Statistical Data Analysis	Mathematics (N 001)	Faculty of Mathematics and Informatics	Institute of Applied Mathematics
Study method	Number of credits	Study method	Number of credits
Lectures	0	Consultations	1
Individual work	4	Seminars	0

Course summary
Statistical data analysis is widely used in scientific research, government of state and business. The course covers various statistical ideas and approaches including statistical learning. The students learn main statistical principles, approaches, problems and methods of statistical inference, Although the course is designed for students in mathematics, its adapted version suits students of other specialities as well. **Topics:** 1. Data types, scales and structures. 2. Sampling methods. Sampling from finite population. Monte Carlo method and computer simulations. Bootstrap. 3. Exploratory data analysis and visualization. Smoothing and averaging. 4. Classical multivariate statistics: linear models, classification, principal component and factor analysis. 5. Statistical learning. Supervised and unsupervised learning. 6. Principles of Bayesian statistics. Prior selection. Bayesian updating. Markov chain Monte Carlo. 7. Statistical and data analysis sofware. 8. Statistical model. Exponential family of distributions. Generalized linear mixed model. Categorical data and loglinear analysis. Latent variable and multilevel modelling. 9. Graphical models. Structural equation modelling. Causal inference. 10. Bayesian risk and statistical decision theory. 11. Nonparametric and adaptive methods. The bias-variance trade-off. Kernel methods, series expansions, regulia-rization and splines. Reproducing kernel Hilbert space. 12. Model selection problem. Information criteria, cross-validation and bootstrap. 13. Robust statistics. Influence function. 14. Dimensionality reduction and projection pursuit.
Main literature
1. Hastie T., Tibshirani R., and Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction,* 2nd ed., New York: Springer, 2009. *http://www-stat.stanford.edu/~tibs/ElemStatLearn/*
2. Wasserman L. *All of Statistics: A Concise Course in Statistical Inference*, Springer Texts in Statistics, 2004. ISBN: 978-1-4419-2322-6 (Print) 978-0-387-21736-9 (Online). *http://www.bioinfo.org.cn/~wangchao/maa/w.statistic.pdf* *http://read.pudn.com/downloads158/ebook/702714/Larry%20Wasserman_ALL%20OF%20Statistics.pdf*
3. Zaki, M.J. and Wagner, M.JR. *Data mining and analysis*, Cambridge University Press, 2014. ISBN 978-0-521-76633-3 Hardback.
4. Bishop, Christopher. *Pattern Recognition and Machine Learning*, Singapore, Springer, 2006. http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf
5. Ledolter, Johannes. *Data mining and business analytics with R*, Hoboken, New Jersey: John Wiley & Sons, Inc.
6. Čekanavičius V., Murauskas G. *Statistika ir jos taikymai.* I, II, III, Vilnius: TEV, 2000, 2002, 2009.

Consulting teacher	Scientific degree	Pedagogical name	Main publications in the field of science of the last 5 year period
Marijus Radavičius	Dr.	Assoc. Prof.	1. Radavičius M. 2020. A Consistent Estimator of Structural Distribution. *Austrian Journal of Statistics*, 49: 99–105. 2. Radavičius, M., Rekašius, T., Židanavičiūtė, J. 2019. Local symmetry of non-coding genetic sequences. *Informatica*, **30** (3): 553-571.

			3. Radavičius M. 2019. Structural Distribution Estimation. Computer Data Analysis and Modeling: Stochastics and Data Science, Proceedings of the 12th International Conference, Minsk, September 18-21, 2019, pp. 280–284. Publishing Center BSU, Minsk. 4. Radavičius, M. 2016. Hoeffding Type Inequalities for Likelihood Ratio Test Statistic.// Computer Data Analysis and Modeling: Theoretical and Applied Stochastics: Proc. of the Eleventh Intern. Conf., Minsk, September 6-10, 2016, pp.182-184. Publishing center BSU, Minsk. 5. Murauskas, G; Radavičius, M. 2016. Multi-unit assignment problem: FCFS course allocation system data analysis. Lithuanian Journal of Statistics, 55 (1): 70-80 (in Lithuanian).

Approved by the Board of Faculty of Mathematics and Informatics 10/12/2021. Resolution No. (1.5 E) 110000-TPN-42
Board Chairman – assoc. prof. dr. Kristina Lapin