

DALYKO (MODULIO) APRAŠAS

Dalyko (modulio) pavadinimas	Kodas
ĮVADAS Į DUOMENŲ ANALIZĘ PROGRAMA „R“	

Dėstytojas (-ai)	Padalinys (-iai)
Koordinuojantis: Rimantė Gaižauskaitė – paskaitos 32 val., pratybos 32 val.	Gyvybės mokslų centras, Biomokslų institutas, Vilniaus universitetas, Saulėtekio al. 7, Vilnius

Studijų pakopa	Dalyko (modulio) lygmuo	Dalyko (modulio) tipas
I pakopa (bakalauro studijos)	1/1	Pasirenkamasis

Igyvendinimo forma	Vykdyimo laikotarpis	Vykdyimo kalba (-os)
Auditorinė, paskaitos, pratybos	Ruduo	Lietuvių kalba

Reikalavimai studijuojančiajam	
Išankstiniai reikalavimai: 1. Anglų kalba. 2. Baziniai kompiuterinio raštingumo įgūdžiai. 3. Biologijos pagrindai.	Gretutiniai reikalavimai (jei yra): 1. Statistikos pagrindai.

Dalyko (modulio) apimtis kreditais	Visas studento darbo krūvis	Kontaktinio darbo valandos	Savarankiško darbo valandos
5	133	Paskaitos – 32 Pratybos – 32	69

Dalyko (modulio) tikslas: studijų programos ugdomos kompetencijos		
Kurso tikslas - suteikti studentams pagrindus apie R programavimo kalbą ir jos taikymą duomenų analizei. Šiame kurse bus nagrinėjami duomenų tvarkymo, vizualizavimo ir statistinės analizės metodai. Kurso pabaigoje studentai įgis vertingų įgūdžių analizuoti ir interpretuoti įvairių sričių duomenis, kurie leis jiems priimti duomenimis pagrįstus sprendimus.		
Dalyko (modulio) studijų siekiniai	Studijų metodai	Vertinimo metodai
1. Susipažinti su programa „R“, jos papildiniais bei pagalbinėmis programomis. Lavinti praktinius darbo programa „R“ įgūdžius ypatingą dėmesį skiriant temoms, kaip duomenis išikelti/išsaugoti, pertvarkyti, susisteminti, atvaizduoti grafiškai ir išanalizuoti bei kaip interpretuoti gautus rezultatus ir pateikti analizės ataskaitą. 2. Įgyti teorinių žinių apie duomenų analizės principus bei susipažinti su įvairiais analizės metodais.	Paskaitos. Darbas praktinių užsiėmimų metu. Savarankiškas darbas. Individualus mokomosios medžiagos studijavimas	Atsiskaitymas už pratybas. Savarankiškas darbas. Egzaminas

Paskaitų ir pratybų temos	Kontaktinio darbo valandos							Savarankiškų studijų laikas ir užduotys	
	Paskaitos	Konsultacijos	Seminarai	Pratybos	Laboratoriniai darbai	Praktika	Visas kontaktinis darbas	Savarankiškas darbas	Užduotys
1. Duomenų analizės principai. Tyrimas. Tyrimo ciklas. Duomenys. Duomenų mokslas. Duomenų tipai ir matavimo skalės. Duomenų struktūros. Duomenų analizės eiga. Duomenų analizės tipai. Atkuriamumas ir atkartojamumas. Tinkamo statistinio metodo nustatymas, prielaidų tikrinimas, modelio parinkimas, modelio tinkamumo tikrinimas ir išvadų formulavimas.	32			32			64	69	Temų analizė savarankiškai naudojant pateiktą mokomąją medžiagą ir rekomenduojamus informacijos šaltinius. Pratimai programomis „R“, „RStudio“ bei kitomis.

<p>2. Programa R. Duomenų analizės ir vaizdinimo sistema <i>R</i>. <i>R</i> sintaksė. Matematinų skaičiavimų galimybės. <i>R</i> objektai ir jų klasės. Funkcijos ir operatoriai. <i>R</i> duomenų struktūros bei kiti programos kodo elementai. Bazinių programos <i>R</i> galimybių išplėtimas: <i>R</i> paketai ir jų valdymas.</p> <p>3. Programa RStudio. Grafinė vartotojo aplinka <i>RStudio</i>. Pagrindiniai programos langai, kortelės, meniu juostos ir mygtukai. <i>RStudio</i> įskiepai („<i>add-ins</i>“) <i>RStudio</i> projektai. <i>R script</i> (.<i>R</i>), ir Quarto (.qmd) dokumentai.</p> <p>4. Dinaminių dokumentų ir ataskaitų rengimas. <i>Markdown</i> ir <i>Quarto</i> sintaksė. Dokumentų pateikimas <i>Microsoft Word</i>, tinklalapio (HTML), PDF (angl. <i>Portable Document Format</i>) ir kitais formatais.</p> <p>5. Duomenų parengimas ir analizė. <i>Tidyverse</i> sistema ir jos principai. Duomenų įkėlimas ir išsaugojimas. Duomenų parengimas analizei: pertvarkymas, perskaičiavimas, perkodavimas, duomenų tipo bei duomenų lentelės formos keitimas ir kitos procedūros. Darbo su įvairiais duomenų tipais (pvz., skaitiniais, tekstiniais, kategoriniais) specifika. Duomenų apibendrinimas ir aprašomosios statistikos. Grafinio atvaizdavimo principai. Paketas <i>ggplot2</i>. Statistinės išvados: hipotezių tikrinimas ir pasikliautiniai intervalai. Hipotezių apie kelių grupių vidurkius tikrinimas taikant įvairius dispersinės analizės metodus (ANOVA, Two – Way ANOVA, Repeated Measures ANOVA). Koreliacija ir tiesinė regresija. Logistinė regresija. Pagrindinių komponentų analizė (PCA).</p>									
Iš viso	32			32			64	69	

Vertinimo strategija	Svoris proc.	Atsiskaitymo laikas	Vertinimo kriterijai
Praktiniai darbai	40	Užduotys turi būti atliktos iki numatyto termino.	Jei visos užduotys atliktos laiku, skiriami 4 balai. Jei užduotis neatlikta iki nustatyto termino, šis skaičius mažinamas 0,25 balo už užduotį. Jei užduotis neatlikta iki sesijos pradžios, skaičius mažinamas 0,5 balo už užduotį. Neatlikus 5 ar daugiau užduočių, neleidžiamalaikyti egzamino.
Du praktinių įgūdžių patikrinimai	30	Pirmas atsiskaitymas praėjus duomenų tvarkymo ir vizualizavimo temas (lapkričio mėnesį). Antras – praėjus statistinės analizės temas (gruodžio mėnesį).	Praktinių užduočių atlikimas užsiėmimo metu. Praktinę užduotį sudarys 10 punktų, už kiekvieną teisingai atliktą punktą 1 balas.

Egzaminas	30	Sesijos arba paskutinio užsiėmimo metu.	Testo tipo egzaminas. Kiekviena užduotis vertinama balais. Balų suma lygi 100, kuri pervedama į 10 balų skalę. Testas paruoštas ir laikomas VMA. Privaloma išlaikyti. Egzaminas laikomas išlaikytu surinkus 34 balus.
Iš viso	100		Kursas yra atsiskaitytas, jei įvykdyti visi privalomi reikalavimai. Galutinį pažymį sudaro visų jį formuojančių dalių įverčių ir baudos taškų suma. Apvalinama vieno skaičiaus po kablelio tikslumu, tad, pvz., 8,49 → 8.

Autorius	Leidimo metai	Pavadinimas	Periodinio leidinio Nr. ar leidinio tomas	Leidimo vieta ir leidykla ar internetinė nuoroda
Privalomoji literatūra				
Hadley Wickham, Mine Çetinkaya-Rundel, and Garrett Golemund.	2023	R for Data Science	(2 nd edition)	https://r4ds.hadley.nz/
Papildoma literatūra				
Salvatore S. Mangiafico	2015	An R Companion for the Handbook of Biological Statistics		https://rcompanion.org/rcompanion/a_02.html
Susan Holmes, Wolfgang Huber	2019	Modern Statistics for Modern Biology		https://www.huber.embl.de/msmb
Salvatore S. Mangiafico	2016	Summary and Analysis of Extension Program Evaluation in R		https://rcompanion.org/documents/RHandbookProgramEvaluation.pdf
Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani	2023	An Introduction to Statistical Learning (with Applications in R)		https://hastie.su.domains/ISLR2/ISLRv2_corrected_June_2023.pdf
R Core Team	2019	An Introduction to R	Periodiškai atnaujinama	https://cran.r-project.org/doc/manuals/r-release/R-intro.html
Carrie Wright, Shannon E. Ellis, Stephanie C. Hicks and Roger D. Peng	2021	Tidyverse Skills for Data Science		https://jhudatascience.org/tidyversecourse/
Jonas Kristoffer Lindeløv	2019	Common statistical tests are linear models (or: how to teach stats)	Periodiškai atnaujinama	https://lindeloev.github.io/tests-as-linear/#41_one_sample_t-test_and_wilcoxon_signed-rank
Gintautas Dzemyda, Olga Kurasova, Julius Žilinskas	2008	Daugiamaečių duomenų vizualizavimo metodai		Mokslo aidai, Vilnius web.vu.lt/mii/j.zilinskas/DzemydaKurasovaZilinskasDDVM.pdf
Hadley Wickham	2017	Advanced R	(2nd edition)	http://adv-r.hadley.nz/
Hadley Wickham	2014	Companion website for “Advanced R”		http://adv-r.had.co.nz/