

DALYKO (MODULIO) APRAŠAS

Dalyko (modulio) pavadinimas	Kodas
IVADAS 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, Individualių studijų dalykas

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

Reikalavimai studijuojančiajam	
Išankstiniai reikalavimai: 1. Anglų kalba. 2. Baziniai kompiuterinio raštingumo įgūdžiai. 3. Biologijos pagrindai. 4. Statistikos pagrindai.	Gretutiniai reikalavimai (jei yra):

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 – supažindinti studentus su R programavimo kalbos pagrindais ir jos taikymu duomenų analizėje. Studentai išmoks tvarkyti, vizualizuoti ir analizuoti duomenis bei interpretuoti rezultatus, kad galėtų daryti duomenimis pagrįstas išvadas.		
Dalyko (modulio) studijų siekiniai	Studijų metodai	Vertinimo metodai
1. Susipažinti su R programa, jos papildiniais ir pagalbinėmis priemonėmis. Lavinti praktinius įgūdžius: įkelti, išsaugoti, pertvarkyti, susisteminti, vizualizuoti ir analizuoti duomenis, interpretuoti rezultatus ir parengti analizės ataskaitą. 2. Įgyti teorinių žinių apie duomenų analizės principus ir susipažinti su įvairiais analizės metodais.	Paskaitos. Darbas praktinių užsiėmimų metu. Savarankiškas darbas. Individualus mokomosios medžiagos studijavimas	Atsiskaitymas už pratybas. Savarankiškas darbas. Projektas

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. Duomenys. 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.

2. Programa R. Duomenų analizės ir vaizdinimo sistema R. R sintaksė. Matematinų skaičiavimų galimybės. R objektai ir jų klasės. Funkcijos ir operatoriai. R duomenų struktūros bei kiti programos kodo elementai. Bazinių programos R galimybių išplėtimas: R paketai ir jų valdymas. 3. Programa RStudio. Grafinė vartotojo aplinka RStudio. Pagrindiniai programos langai, kortelės, meniu juostos ir mygtukai. RStudio įskiepai („add-ins“) RStudio projektai. R script (.R), ir Quarto (.qmd) dokumentai. 4. Dinaminių dokumentų ir ataskaitų rengimas. Markdown ir Quarto sintaksė. Dokumentų pateikimas Microsoft Word, tinklalapio (HTML), PDF (angl. Portable Document Format) ir kitais formatais. 5. Duomenų parengimas ir analizė. Tidyverse 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) specifiška. Duomenų apibendrinimas ir aprašomosios statistikos. Grafinio atvaizdavimo principai. Paketas ggplot2. Statistinės išvados: hipotezių tikrinimas ir pasikliautiniai intervalai. Hipotezių apie kelių grupių vidurkių tikrinimas taikant įvairius dispersinės analizės metodus (ANOVA, Two – Way ANOVA, Repeated Measures ANOVA). Koreliacija ir tiesinė regresija. Pagrindinių komponentų analizė (PCA).								
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 bala 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.

Projektas	30	Sesijos arba paskutinio užsiėmimo metu.	Studentai savarankiškai atlieka duomenų analizės projektą, naudodami pasirinktą duomenų rinkinį. Vertinami šie aspektai: • duomenų paruošimas ir manipuliacija, • apžvalginė analizė ir vizualizacija, • statistinė analizė bei hipotezių tikrinimas, • analizės ataskaita parengta su <i>Quarto</i>, • rezultatų pristatymas. Kiekviena dalis sudaro 20 % projekto vertinimo. Privaloma pristatyti darbą ir pateikti analizės ataskaitą. Projektas vertinamas 100 balų sistemoje, kuri kovertuojama į 10 balų skalę. Neatsiskaičius projekto, kursas laikomas neišlaikytu.
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 Grolemund.	2023	R for Data Science	(2 nd edition)	https://r4ds.hadley.nz/
Papildoma literatūra				
Vilmantas Gėgžna		Biostatistinės analizės pagrindai	Periodically updated	https://mokymai.github.io/biostatistika/
Susan Holmes, Wolfgang Huber	2024	Modern Statistics for Modern Biology		https://www.huber.embl.de/msmb/
R Core Team	2019	An Introduction to R	Periodically updated	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/
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
Gintautas Dzemyda, Olga Kurasova, Julius Žilinskas	2008	Daugiamatčių 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/