

Skeiðupplýsingar

Skeiðsnummar	5021.20
ECTS	7.50
Heiti	Dátuvísindi og hagfrøði við RStudio
Endamál	At geva eina innleiðing í nútímans dátuvísindi, hagfrøði og hugtakið endurskapandi gransking (“reproducible research”) við nýtslu av framkomnum og brúkaravinarlíggum hentleikum í RStudio. Dátuvísindi og hagfrøði við RStudio ger tað nú munandi lættari at skipa og hagfrøðisliga greina data – samanborið við at brúka R einsamalt – við at tvinna saman R-kodu, tekst, formlar og myndir til snøggar hagfrøði-frágreiðingar. Sum KT-verkfrøðingur, dátugreinari ella granskari fært tú soleiðis innleiðslu í bæði fremstu KT-amboð - og grundleggjandi teori - innan hagfrøðisliga greining av data.
Innihald	Dátuvísindi: innlesing, skipan, broyting, plotting og modellering av snøggum data (“tidy data”). Innleiðing til R og RStudio. Menning og skipan av RStudio-verkætlanum eftir hugtakinum endurskapandi gransking. Skjalfesting og kunning: dynamiskar frágreiðingar við RStudio. Hagfrøði: Sløg av data. Sýnistøka og sýnisbýti. Sannlíkindi og stokastiskir variablar. Diskret býti, m.a. binomial- og poisson-býtið. Kontinuer býti, m.a. normalbýtið og eksponentialbýtið. Hagfrøðilig lýsingartøl og grafar. Punkt- og interval-estimatió, konfidens-interval. Hagfrøðilig modell og test av hypotesum fyri bæði numerisk og flokkaði (kategorisk) data. Korrelatió og linjurøtt regressiún, t-test og chi-kvadrat test, styrki av testum (power), stór og smá sýni. Dømi um brúk av hagfrøði innan teldufrøði og onnur fakøki, t.d. fólkaheilsu, lívfrøði, samfelagsfrøði.
Læru- og undirvísingarmóttir	Fyrilestrar, telduvenjingar við open-source R og RStudio, bólkaarbeiði og orðaskifti.
Próvtøkumáttur	Dømingin byggir á hesar lutir: - tvær smærri kravdar innlatingar, og ein kravd stóruppgáva, sum allar skula latast inn til avtalaðar tíðarfreistir. - ein munnlig próvtøka (telur 100%) uttan fyrireikingartíð við støði stóruppgávu, og spurningum innan teori og smærri innlatingum. Stóruppgávan skal leggjast fram og verjast undir próvtøkuni, og hon verður tí send til próvdómara í góðari tíð undan próvtøkuni. - treyt fyri innskriving til próvtøku: at tann lesandi minst hevur latið inn og fingið góðkent tvær tær fyrstu skrivligu innlatingarnar til avtalaðar tíðarfreistir.
Próvdøming	Uttanhýsis
Próvtalastigi	7-talsstigin
Lestrarlisti	Open-source lærubøkur, sum finnast online (ikki neyðugt at keypa tær): - R for Data Science (http://r4ds.had.co.nz/). - OpenIntro Statistics (https://www.openintro.org/). - Møguliga annað tilfar, sum verður kunnað um og brúkt í undirvísingini.
Fortreytir	Støddfrøði B á miðnámsstigi.
Læruúrtøka	Eftir lokið skeið skal studenturin vera førur fyri at: - brúka RStudio til at innlesa rá-data, og skipa snøgg data til hagfrøðiligar greiningar. - stovna RStudio-verkætlanir, gera og nýta R-kodu til hagfrøðiligar kanningar. - gera snøggar hagfrøði frágreiðingar við talvum og myndum eftir hugtakinum endurskapandi gransking við RStudio. - kenna grundleggjandi hagfrøðislig hugtøk, sum hava við data, sannlíkindi, stokastiskar variablar og sannlíkindabýti at gera, og meta um, hvørjar hagfrøðiligar metodur kunnu koma upp á tal við einum ítøkiligum datasetti.
Stig	Bachelor
Ábyrgd	Hannes Gislason, email: hannesg@setur.fo

Course information

Course number 5021.20

ECTS 7.50

Title Data Science and Statistics with RStudio

Aim To introduce modern data science, statistics and the concept of reproducible research by using advanced - but easy to use - facilities in RStudio. Data Science and Statistics with RStudio makes it now much easier to organize and analyse data – as compared to only using R – to knit code, text, formulas and plots into nice statistical reports. As a computer scientists, data scientists or researcher you will get an introduction to both modern computational tools and basic theory within statistical analysis of data.

Content Data Science: data input, organisation, transformation and modelling of tidy data. Introduction to R and RStudio. Configuration and development of RStudio projects for reproducible research. Documentation and presentation: dynamic reports with RStudio. Statistics: Types of data, Random sampling and sampling distributions. Probability and random variables. Discrete distributions, among them binomial and Poisson distribution. Continuous distributions, among them normal and exponential distribution. Statistical description and graphs. Point and interval estimation, confidence-intervals. Statistical models and hypothesis testing for both numerical and categorical data. Correlation and linear regression, t-test and chi-square test, power of tests, large and small samples. Examples of use of statistics within computer science and other fields, e.g. human health, biology, and sociology.

Learning and teaching approaches Lectures, computer exercises with open-source R and RStudio, group work and discussion.

Assessment method The assessment is based on the following elements:

- mandatory assignments (two small and one large) to be handed in on time.
- an oral examination (counts 100%) without preparation time based on the large assignment and questions within theory and the small assignments. The large assignment will be presented and defended during the examination. Therefore, it will be sent to the censor in good time before the examination.

Note: the two small assignments must be handed in on time and approved to be registered for the examination.

Examination External

Grading scale 7-step marking scale

Literature Open-source online textbooks:

- R for Data Science (<http://r4ds.had.co.nz/>).
- OpenIntroStatistics(<https://www.openintro.org>)
- Possibly other materials introduced during the course.

Prerequisites High school mathematics, B-level.

Learning outcomes As a result of this course, the student will be able to

- use RStudio to input data, and to use tidy data for statistical analysis.
- configure RStudio projects for reproducible research, and produce and use R-code for statistics.
- to produce nice reports including tables and plots obeying to the principles of reproducible research with RStudio.
- demonstrate basic knowledge regarding data, probability, stochastic variables and distributions, and assess which statistical methods can be useful for an actual dataset.

Level Bachelor

Contact Hannes Gislason, email: hannesg@setur.fo