

<b>B-III – Charakteristika studijního předmětu</b>					
Název studijního předmětu	Data science v R				
Typ předmětu	neprofilující volitelný				
Rozsah studijního předmětu	hodiny/semestr 10/0; 30	doporučený ročník / semestr	2 /	Dvousemestrální předmět	ne
Způsob ověření studijních výsledků	zápočet		Podoba výuky	přednáška+seminář	
Další požadavky na studenta					
Vyučující	prof. RNDr. Jan Hendl, CSc. MUDr. Petr Waldauf, Ph.D.				
Stručná anotace předmětu					
<p>Kurz má naučit uchazeče základní práci se statistickým programem R v grafickém prostředí RStudio. Cílem je získání základních zkušeností a praktických dovedností s rychlým a efektivním zpracováním klinických dat prostřednictvím kolekce balíků tidyverse. V kurzu se budeme zabývat úvodem do statistického zpracování dat v R (deskriptivní statistika, pokročilejší grafická zobrazení dat, základní statistické testy, úvod do multivariantní analýzy). Od uchazečů není vyžadována přechozí zkušenost s programovacím jazykem R. Výuka bude probíhat praktickou formou přímo v RStudio, které si uchazeč nainstaluje na svůj počítač. Budou připraveny výukové skripty a data sety. NUTNÉ MÍT K DISPOZICI VLASTNÍ NOTEBOOK. Přednášená témata zahrnují například: • Úvod do R K čemu datascience, přehled software pro datascience, literatura pro studium, DataCamp, cheatsheets, úvod do R a RStudio, instalace R a RStudio, první příkazy v R • Jak načíst a třídit svoje data: úvod do tidyverse Proč Tidyverse, import dat do R (csv, xlsx) – reader, glimpse, úvod do data wrangling – pipeline, dplyr (select, filter, mutate, arrange, rename, group_by, summarise) • Obrázky do článku i posteru: úvod do vizualizace ggplot (scatter plot, bar plot, box plot, histogram, facets, geom_smooth) • Publikace přímo z RStudio: Úvod do markdown proč markdown, reproducible research, pandoc, html, ms word, pdf • Práce s faktory, strings s datumem a časem factor, forcats, stringer, lubridate • Úvod do funkcí • Relační data wide + long formát dat, co jsou to relační data, jak jsou ukládána data v databázích, left_join, right_join, full_join, anti_join, wide, • Úvod do statistického zpracování dat deskriptivní statistika (gtsummary), t test (párový, nepárový), neparametrické testy, kategoričká data, chi-square test, Fisher test, lineární regrese, logistická regrese, time-to-event analýza, multivariantní lineární, logistická a Cox regrese.</p>					
Studijní literatura					
<b>Povinná:</b> <ul style="list-style-type: none"> <li>• WICKHAM, HADLEY GROLEMUND, GARRETT. <i>R for data science : import, tidy, transform, visualize and model data</i>. Beijing ; Boston ; Farnham ; Sebastopol ; Tokyo: O'Reilly Media, 2016, 492 s. ISBN 978-1-491-91039-9.</li> <li>• Biostatistics With 'R': A Guide for Medical Doctors [online]. Dostupné z: <a href="https://www.bigbookofr.com/">https://www.bigbookofr.com/</a></li> </ul>					
Doplňující informace					
Doplňující informace (nad rámec uváděných údajů)					
Aktivní účast. Semináře, cvičení, samostatné práce.					