

PSY532, PSY232 – Praktický úvod pro používání statistického programu R**podzim 2018****VYUČUJÍCÍ****Mgr. Stanislav Ježek, PhD.** – garance kurzujezek@fss.muni.cz, 549494616, konzultační hodiny: středa 13:00 – 15:00, FSS 2.47**Mgr. Vít Gabrhel** – přednášky, seminářevit.gabrhel@mail.muni.cz, konzultace po dohodě emailem**CHARAKTERISTIKA KURZU**

Tento kurz si klade tři hlavní cíle. Tím prvním je snížení míry závislosti (v ideálním případě úplná nezávislost) na placeném statistickém softwaru, který může být z finančních důvodů nedostupný a který uživateli poskytuje relativně omezené možnosti práce. S tím souvisí snaha vzbudit zájem o programovací jazyk s nepřebernými analytickými možnostmi a vitální komunitou globálního charakteru. V neposlední řadě půjde o osvěžení si běžných procedur v rámci statistické analýzy dat, jakož i o prohloubení dosavadních znalostí v této oblasti.

NÁVAZNOSTI KURZU

Doporučeným předpokladem tohoto kurzu jsou znalosti statistiky v rozsahu kurzu Statistická analýza dat II. - PSY252 (https://is.muni.cz/el/1423/podzim2015/PSY252/psy252_sylabus_p2015_v1.02.pdf) nebo Statistická analýza dat - SOC108, resp. SOC708 (https://is.muni.cz/auth/el/1423/podzim2011/SOC108/op/SOC108-708Statisticka_analyza_dat-sylabus2011.pdf).

ORGANIZACE KURZU

Počet kreditů:	4
Ukončení kurzu:	zápočet
Přednášky:	1 vyučovací hodina týdně
Semináře:	1 vyučovací hodina týdně

INFORMAČNÍ SYSTÉM MU

V informačním systému MU části Studijní materiály k předmětu PSY523 umísťujeme podklady k tématům uvedeným v sylabu především v podobě prezentací, odkazů na další zdroje informací a seminárních materiálů. Na stejném místě jsou též k dispozici pokyny nezbytné k plnění písemných úkolů, popř. doplňky k tomuto sylabu.

KOMUNIKACE S VYUČUJÍCIMI

Preferovaným komunikačním kanálem pro osobní komunikaci týkající se kurzu je email. Používejte prosím „PSY532“ nebo „PSY232“ v předmětu svých emailů. Urychlíte tím jejich vyřízení.

V případě dotazů týkajících se látky kurzu, využívejte prosím v maximální možné míře facebookové skupiny „*Statistika, metodologie, psychometrika*“ (<https://www.facebook.com/groups/461796387316423/>) a „*Kvantitativní kroužek*“ (<https://www.facebook.com/groups/433938776716307/>). Je pravděpodobné, že odpověď na Váš dotaz by mohl zajímat i Vaše spolužáky, a byla by proto škoda uzavřít takovou komunikaci do soukromí emailů. Všichni vyučující kurzu skupinu spravují, pravidelně sledují a přispívají do ní.

K důležitým ohlášením používáme hromadný email studentům.

POŽADAVKY NA UKONČENÍ KURZU**Cvičení**

V průběhu semestru mají studenti jako jednotlivci za úkol zpracovat sadu **až 10 cvičení z R** vycházejících z látky probrané na hodinách. Cílem cvičení je upevnit nabytou znalost. Cvičení budou k dispozici v rámci virtuální třídy (R101: A practical guide to using R as your everyday statistical tool) na platformě DataCamp (<https://www.datacamp.com>). Studující, kteří si ponechají kurz zapsaný po konci období změn v zápisu kurzů, obdrží na svoji školní emailovou adresu pozvánku do virtuální třídy.

Každý úspěšně odevzdaný skript bude oceněn **10 body**. Celkově tak lze za skripty získat až **100 bodů**. **Nejzazší možný termín pro odevzdání cvičení je nedělní půlnoc v rámci týdne, kdy bylo dané cvičení zadáno.** Například pro cvičení k tématu „5. Manipulace s daty“, zadané v pondělí 22. 10. 2018, je to neděle 28. 10. 2018 23:59. Výjimkou je téma „3. Vkládání dat“, kdy je možné cvičení odevzdat do neděle 7. 10. 2018 23:59.

Prezentace vybraného balíčku

Součástí závěrečné hodiny kurzu bude prezentace vybraného balíčku (kupř. pro analýzu v bakalářské či diplomové práci) frekventanty kurzu či týmy frekventantů (bude stanoveno na základě počtu účastníků kurzu). Splnění úkolu je honorováno **10 body**.

Studenti utvoří dvoučlenné týmy a do 1. 11. 2018 si v rámci těchto týmů zvolí jeden balíček (package), který v rámci tématu PREZENTACE VYBRANÉHO BALÍČKU představí ostatním. Balíček by se měl v ideálním případě týkat analýzy, kterou studenti zamýšlejí (např. v bakalářské či diplomové práci) využít. Každý balíček může být prezentován pouze jednou. Horní časový limit pro prezentaci jednoho příspěvku je **7 minut**. Doporučená struktura je 5 minut na prezentaci balíčku a 2 minuty na diskuzi k příspěvku.

Informace o členech týmu a balíčku zamýšlenému k prezentaci vyplňte do 1. 11. 2018 na následujícím odkazu:
<https://docs.google.com/spreadsheets/d/1Hn8iM1emK4ySwK6gWljb1RZYojqu604MBptIhYfEFIA/edit?usp=sharing>

Účast na hodinách

Účast na hodinách je dobrovolná, docházka proto nebude vyžadovaná.

Zápočet

Kurz je zakončen zápočtem uděleným za účast na hodinách, vypracované analýzy a prezentace vybraného balíčku.

Omluvy

Víte-li o tom, že budete mít ze závažných důvodů (kupř. nemoc) problémy s dodržením některého z termínů (např. pro odevzdání skriptů), informujte nás o tom co nejdříve.

Celkové hodnocení

K udělení zápočtu je třeba získat alespoň **60 % z maximálního počtu bodů**, který tvoří součet bodů za vypracované skripty (*80 bodů základ a 20 bodů jako bonus*) a prezentaci (*10 bodů*), tj. **54 bodů**. Celkové hodnocení se počítá z 90 bodů jako základu a bude používat následující stupnici:

Z = 54 a více bodů.

N = 53 a méně bodů.

ÚČAST NA VÝZKUMU V RÁMCI KURZU

V rámci účasti kurzu mohou být studenti požádáni o účast na výzkumech jak pro zlepšení kvality kurzu samotného, tak i pro další účely spojené s výzkumným zaměřením katedry. Účast v těchto výzkumech je ryze dobrovolná a nemá vliv na úspěšné ukončení kurzu či získané hodnocení.

Prvním příkladem takového výzkumu je ověření pozornosti při čtení sylabu kurzu. Pošlete prosím do pátku 21. 9. 2018 email s Vámi vybraným memem vyučujícímu kurzu s nejnižším dosaženým stupněm vzdělání na uvedenu emailovou adresu v tomto sylabu.

PROBLÉMOVÉ OKRUHY

1. PROGRAMOVACÍ JAZYK R A PRÁCE S NÍM

- a) Konceptuální představení jazyka R
- b) Instalace R a Rstudio, orientace v prostředí Rstudio, library and packages
- c) Základní operace – aritmetika, přiřazování hodnot objektům, zjišťování třídy objektů
- d) Efektivní psaní kódu

Doporučená literatura k tématu:

Muenchen, R. A. (2011). *R for SAS and SPSS Users*. Springer: New York.

Norman, M. (2011). *The Art of R Programming*. No Starch Press: San Francisco.

Coding etiquette – <https://ourcodingclub.github.io/2017/04/25/etiquette.html>

Getting started with R and RStudio – <https://ourcodingclub.github.io/2016/11/13/intro-to-r.html>

Google's R Style Guide – <https://google.github.io/styleguide/Rguide.xml>

R basics, workspace and working directory, RStudio projects – http://stat545.com/block002_hello-r-workspace-wd-project.html

RStudio IDE Cheat Sheet – <https://www.rstudio.com/wp-content/uploads/2016/01/rstudio-IDE-cheatsheet.pdf>

Base R Cheat Sheet – <https://www.rstudio.com/wp-content/uploads/2016/05/base-r.pdf>

2. DATOVÉ OBJEKTY

- a) Vector
- b) Matrix
- c) Factors
- d) Data Frames
- e) Lists

Doporučená literatura k tématu:

Muenchen, R. A. (2011). *R for SAS and SPSS Users*. Springer: New York.

Navarro, D. J. (2014). Learning statistics with R: A tutorial for psychology students and other beginners. Available online: <http://www.fon.hum.uva.nl/paul/lot2015/Navarro2014.pdf>

Norman, M. (2011). *The Art of R Programming*. No Starch Press: San Francisco.

Cvičení:

Introduction to R

3. VKLÁDÁNÍ DAT

- a) Import dat formátu .csv
- b) Import dat formátu .txt
- c) Import dat formátu .xlsx
- d) Import dat formátu .sav

Doporučená literatura k tématu:

Muenchen, R. A. (2011). *R for SAS and SPSS Users*. Springer: New York.

Navarro, D. J. (2014). Learning statistics with R: A tutorial for psychology students and other beginners. Available online: <http://www.fon.hum.uva.nl/paul/lot2015/Navarro2014.pdf>

Data Import : : CHEAT SHEET – https://ugoprotocol.github.io/ugo_r_doc/data-import.pdf

Cvičení:

Importing Data in R (Part 1)

Importing data from statistical software packages

4. ČIŠTĚNÍ DAT

- a) Messy data, tidy data
- b) Práce s časovým formátem dat a s textovými proměnnými
- c) Missing values, outliers, errors

Doporučená literatura k tématu:

Muenchen, R. A. (2011). *R for SAS and SPSS Users*. Springer: New York.

Navarro, D. J. (2014). Learning statistics with R: A tutorial for psychology students and other beginners. Available online: <http://www.fon.hum.uva.nl/paul/lot2015/Navarro2014.pdf>

Norman, M. (2011). *The Art of R Programming*. No Starch Press: San Francisco.

Cvičení:

Cleaning Data in R

5. MANIPULACE S DATY

- a) tbl
- b) select, mutate
- c) filter, arrange
- d) summarise, %>%
- e) group_by

Doporučená literatura k tématu:

Data Wrangling with dplyr and tidyr Cheat Sheet –

<https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf>

Cvičení:

Data Manipulation in R with dplyr

6. EXPLORACE DAT

- a) Kategorická data
- b) Numerická data

Doporučená literatura k tématu:

Muenchen, R. A. (2011). *R for SAS and SPSS Users*. Springer: New York.

Navarro, D. J. (2014). Learning statistics with R: A tutorial for psychology students and other beginners. Available online: <http://www.fon.hum.uva.nl/paul/lot2015/Navarro2014.pdf>

Norman, M. (2011). *The Art of R Programming*. No Starch Press: San Francisco.

Cvičení:

Exploratory Data Analysis

7. ÚVOD DO GRAFIKY

- a) Bar chart
- b) Histogram
- c) Boxplot
- d) Error bar
- e) Scatterplot
- f) Pirate plot

Doporučená literatura k tématu:

Kabacoff, R. I. (2011). *R in Action Data analysis and graphics with R*. Manning: Shelter Island.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. Available online: <http://moderngraphics11.pbworks.com/f/ggplot2-Book09hWickham.pdf>.

Data Visualization with ggplot2 Cheat Sheet – <https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf>

Cvičení:

Data Visualization with ggplot2 (Part 1)

8. SROVNÁNÍ SKUPIN

a) t-test (jednovýběrový, dva nezávislé výběry, opakovaná měření), Studentův a Welchův t-test

b) Analýza rozptylu (ANOVA)

Doporučená literatura k tématu:

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. Sage: UK.

Lakens, D. (n.d.). *The perfect t-test*. Staženo dne 8. 10. 2016. Dostupné online na: <http://daniellakens.blogspot.cz/2015/05/the-perfect-t-test.html>

Navarro, D. J. (2014). *Learning statistics with R: A tutorial for psychology students and other beginners*. Available online: <http://www.fon.hum.uva.nl/paul/lot2015/Navarro2014.pdf>

Verzani, J. (2014). *Using R for Introductory Statistics* (2. edice). Chapman & Hall: Boca Raton.

Cvičení:

Intro to Statistics with R: Student's T-test

Intro to Statistics with R: Analysis of Variance (ANOVA)

9. KORELACE, JEDNODUCHÁ REGRESE

a) Korelace (Pearsonovo r, Spearmanovo ρ , Kendallovo tau)

b) Jednoduchá lineární regrese

Doporučená literatura k tématu:

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. Sage: UK.

Navarro, D. J. (2014). *Learning statistics with R: A tutorial for psychology students and other beginners*. Available online: <http://www.fon.hum.uva.nl/paul/lot2015/Navarro2014.pdf>

Cvičení:

Intro to Statistics with R: Correlation and Linear Regression

10. VÍCENÁSOBNÁ LINEÁRNÍ REGRESE A LOGISTICKÁ REGRESE

a) Vícenásobná lineární regrese

b) Logistická regrese

Doporučená literatura k tématu:

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. Sage: UK.

Navarro, D. J. (2014). *Learning statistics with R: A tutorial for psychology students and other beginners*. Available online: <http://www.fon.hum.uva.nl/paul/lot2015/Navarro2014.pdf>

Verzani, J. (2014). *Using R for Introductory Statistics* (2. edice). Chapman & Hall: Boca Raton.

Cvičení:

Multiple and Logistic Regression

11. HIERARCHICKÉ A MIXED MODELY

- a) Úvod do hierarchických a mixed modelů
- b) Linear mixed-effect models
- c) Generalized linear mixed-effect models
- d) Repeated Measures

Doporučená literatura k tématu:Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. Sage: UKCvičení:

Hierarchical and Mixed Effect Models

12. PREZENTACE VYBRANÉHO BALÍČKU

Krátká prezentace vybraného balíčku (kupř. pro analýzu v bakalářské či diplomové práci) frekventanty kurzu či týmy frekventantů (bude stanoveno na základě počtu účastníků kurzu).

ČASOVÁ OSNOVA

Termín	Téma
24-09-18	PROGRAMOVACÍ JAZYK R A PRÁCE S NÍM
01-10-18	DATOVÉ OBJEKTY (16:00 – 17:30)
01-10-18	VKLÁDÁNÍ DAT (18:00 – 19:30)
15-10-18	ČIŠTĚNÍ DAT
22-10-18	MANIPULACE S DATY
29-10-18	EXPLORACE DAT
05-11-18	ÚVOD DO GRAFIKY
12-11-18	SROVNÁNÍ SKUPIN
19-11-18	KORELACE, LINEÁRNÍ REGRESE
26-11-18	VÍCENÁSOBNÁ LINEÁRNÍ REGRESE, MEDIACE A MODERACE
03-12-18	HIERARCHICKÉ A MIXED MODELY
10-12-18	PREZENTACE VYBRANÉHO BALÍČKU

LITERATURA

Base R Cheat Sheet – <https://www.rstudio.com/wp-content/uploads/2016/05/base-r.pdf>Coding etiquette – <https://ourcodingclub.github.io/2017/04/25/etiquette.html>Data Visualization with ggplot2 Cheat Sheet – <https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf>

Data Wrangling with dplyr and tidyr Cheat Sheet –

<https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf>Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. Sage: UK.Getting started with R and RStudio – <https://ourcodingclub.github.io/2016/11/13/intro-to-r.html>Google's R Style Guide – <https://google.github.io/styleguide/Rguide.xml>Harrington, D. (2008). *Confirmatory Factor Analysis*. Oxford University Press.Kabacoff, R. I. (2011). *R in Action Data analysis and graphics with R*. Manning: Shelter Island.

Lakens, D. (n.d.). *The perfect t-test*. Staženo dne 8. 10. 2016. Dostupné online na:

<http://daniellakens.blogspot.cz/2015/05/the-perfect-t-test.html>

Muenchen, R. A. (2011). *R for SAS and SPSS Users*. Springer: New York.

Navarro, D. J. (2014). Learning statistics with R: A tutorial for psychology students and other beginners. Available online: <http://www.fon.hum.uva.nl/paul/lot2015/Navarro2014.pdf>

Norman, M. (2011). *The Art of R Programming*. No Starch Press: San Francisco.

R basics, workspace and working directory, RStudio projects – http://stat545.com/block002_hello-r-workspace-wd-project.html

RStudio IDE Cheat Sheet – <https://www.rstudio.com/wp-content/uploads/2016/01/rstudio-IDE-cheatsheet.pdf>

Verzani, J. (2014). *Using R for Introductory Statistics* (2. edice). Chapman & Hall: Boca Raton.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Available online: <http://moderngraphics11.pbworks.com/f/ggplot2-Book09hWickham.pdf>