PSY532, PSY232 – Praktický úvod pro používání statistického programu R **podzim 2016**

VYUČUJÍCÍ

Mgr. Stanislav Ježek, PhD. – garance kurzu, přednášky, semináře
jezek@fss.muni.cz, 549494616, konzultační hodiny: středa 10 – 12, FSS 2.47

Mgr. Hynek Cígler – přednášky, semináře
sirucek@fss.muni.cz, 549498263, konzultace po dohodě emailem, FSS 2.47

Mgr. Vít Gabrhel – přednášky, semináře
vit.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 softwaru, který je na mnoha pracovištích z finančních důvodů nedostupný a který uživateli poskytuje relativně omezené možnosti práce. Za druhé půjde o osvěžení si běžných procedur v rámci statistické analýzy dat, jakož i o prohloubení dosavadních znalostí v této oblasti. V neposlední řadě je cílem vzbudit zájem o programovací jazyk s nepřebernými analytickými možnostmi a vitální komunitou globálního charakteru.

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 umisťujeme 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 VYUČUJÍCMI

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 facebookovou skupinu „*Statistika, metodologie, psychometrika*“ (https://www.facebook.com/groups/461796387316423/). 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

Skripty

V průběhu semestru mají studenti jako jednotlivci za úkol zpracovat **10 skriptů z R** vycházejících z látky probrané na hodinách. Cílem tohoto úkolu je upevnit nabytou znalost. Každý skript bude hodnocen na stupnici **0 až 5 bodů**. Celkově tak lze za skripty získat až **50 bodů**. Skripty vkládejte do odevzdávány **nejpozději v sobotu před následující lekcí.**

Práce lze odevzdávat pouze elektronicky v odevzdávárně ve studijních materiálech předmětu. Jméno vkládaného souboru s prvním seminárním úkolem musí být **S1.doc** a **S2.doc** pro druhý[1], atd., a nic víc; IS k němu automaticky přidá jméno vkládajícího studenta. Komentář/popisek nechejte nevyplněný.

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. 2017 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 15. 10. 2017 na následujícím odkazu: https://docs.google.com/spreadsheets/d/1OHqMIXtxTAxPWMVockNFte3d5M45rzZS-h98pm_Qy7s/edit#gid=0

Účast na hodinách

Účast na hodinách je dobrovolná, docházka proto nebude vyžadovaná, bude ale bodově honorovaná. Za účast na hodině studující získají **1 bod** až do celkové výše **10 bodů**.

Zápočet

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

Omluvy

Omluvy jsou přijímány pouze předem. Pozdější omluvy budou akceptovány, pouze pokud šlo o nepředvídatelné případy. Víte-li o tom, že budete mít ze závažných důvodů problémy s dodržením některého z termínů, informujte nás o tom co nejdříve.

Celkové hodnocení

K udělení zápočtu je třeba získat alespoň **60 % z maximálního počtu bodů** (*70 bodů*), který tvoří součet bodů za vypracované skripty (*50 bodů*), prezentaci (*10 bodů*) a docházku (*10 bodů*), tj. **42 bodů**. Celkové hodnocení bude používat následující stupnici:

Z = 42 více bodů.
N = 41 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.

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

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.

[1] Práce lze odevzdávat ve všech běžných editovatelných formátech textových dokumentů – doc, rtf, odt apod. Prosím, neodevzdávejte práce v obtížně editovatelných formátech, jako je např. pdf, djvu. Jde nám o možnost vkládání komentářů a čitelnost i na mobilních zařízeních.

2. TVORBA DATOVÝCH OBJEKTŮ A MANIPULACE S NIMI

a) Vector
b) Matrix
c) Factors
d) Data Frames

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.

3. VKLÁDÁNÍ A ČIŠTĚNÍ DAT, ZJIŠŤOVÁNÍ ZÁKLADNÍCH INFORMACÍ O DATOVÉM SOUBORU

a) Importování dat do R
b) Čištění dat
c) Popisné statistiky, kontingenční tabulky

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.

4. SROVNÁNÍ DVOU PRŮMĚRŮ A JEDNODUCHÁ ANALÝZA SOUVISLOSTI

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

b) korelace (Pearsonovo r, Spearmanovo rhó, Kandallovo tau)

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

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

5. MNOHONÁSOBNÁ LINEÁRNÍ REGRESE

Korelace, parciální korelace, lineární regrese. Mnohonásobná regrese, pojmy model a parametr, regresní koeficienty b,β, směrodatná chyba regresního koeficientu, mnohonásobný korelační koeficient R, koeficient determinace R2 ,postupná (stepwise) regrese, hierarchická (blockwise) regrese, indikátorové (dummy) proměnné, multikolinearita, supresory, problémy spojené s množstvím prediktorů, grafická kontrola splnění předpokladů regrese, rezidua, odlehlé a vlivné případy. Moderace, mediace a Sobelův test.

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://health.adelaide.edu.au/psychology/ccs/teaching/lsr/

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

6. ÚVOD DO GRAFIKY V R

Bar chart, histogram, boxplot, error bar, scatterplot, 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.

7. ČTECÍ TÝDEN – STUDIJNÍ VOLNO

8. ANALÝZA ROZPTYLU, ANALÝZA KOVARIANCE

Faktoriální analýza rozptylu, model a parametry, fixované a náhodné faktory, hlavní efekty a interakce faktorů, výhody vyváženého designu, kontrasty a post-hoc testy, analýza kovariance, MANOVA.

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://health.adelaide.edu.au/psychology/ccs/teaching/lsr/

9. ANALÝZA KATEGORICKÝCH DAT

a) Neparametrické alternativy pro parametrické testy srovnávající průměry)
b) Binární logistická regrese:

Model a parametry logistické regrese, interpretace regresního koeficientu exp(B), log-likelihood (-2LL), vyjádření shody modelu s daty pomocí R2 Coxe & Snella a Nagelkerka, ověření předpokladů, analýza reziduí a vlivných pozorování.

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://health.adelaide.edu.au/psychology/ccs/teaching/lsr/

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

10. 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).

11. ANALÝZA HLAVNÍCH KOMPONENT, FAKTOROVÁ ANALÝZA

Analýza hlavních komponent, explorační faktorová analýza, konfirmační faktorová analýza. Pravidla pro určení počtu faktorů. Rotace ortogonální a šikmé. Ověření předpokladů FA. Interpretace faktorových matic.

Doporučená literatura k tématu:

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

12. – 13. KONFIRMATORNÍ FAKTOROVÁ ANALÝZA

Teoretický úvod, CFA measurement model, předpoklady použití, práce s chybějícími hodnotami. Demonstrace analýzy skrze balíček lavaan, ukazatele shody modelu s daty jako například RMSEA, CFI, SRMR, AIC nebo BIC

Doporučená literatura k tématu:

Harrington, D. (2008). Confirmatory Factor Analysis. Oxford University Press.

Kline, R. B. (2010). Principles and Practice of Structural Equation Modelling (3rd ed.). The Guilford Press.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. URL http://www.jstatsoft.org/v48/i02/

ČASOVÁ OSNOVA

Termín	Téma
19. 9. 2016	PROGRAMOVACÍ JAZYK R A PRÁCE S NÍM
26. 9. 2016	TVORBA DATOVÝCH OBJEKTŮ A MANIPULACE S NIMI
3. 10. 2016	VKLÁDÁNÍ A ČIŠTĚNÍ DAT, ZJIŠŤOVÁNÍ ZÁKLADNÍCH INFORMACÍ O DATOVÉM SOUBORU
10. 10. 2016	SROVNÁNÍ DVOU PRŮMĚRŮ A JEDNODUCHÁ ANALÝZA SOUVISLOSTI
17. 10. 2016	MNOHONÁSOBNÁ LINEÁRNÍ REGRESE
24. 10. 2016	ÚVOD DO GRAFIKY V R
31. 10. 2016	ČTECÍ TÝDEN – STUDIJNÍ VOLNO
7. 11. 2016	ANALÝZA ROZPTYLU, ANALÝZA KOVARIANCE
14. 11. 2016	ANALÝZA KATEGORICKÝCH DAT
21. 11. 2016	PREZENTACE VYBRANÉHO BALÍČKU
28. 11. 2016	ANALÝZA HLAVNÍCH KOMPONENT, FAKTOROVÁ ANALÝZA
5. 12. 2016	KONFIRMATORNÍ FAKTOROVÁ ANALÝZA I.
12. 12. 20016	KONFIRMATORNÍ FAKTOROVÁ ANALÝZA II.

LITERATURA

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

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.

Kline, R. B. (2010). Principles and Practice of Structural Equation Modelling (3rd ed.). The Guilford Press.

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://health.adelaide.edu.au/psychology/ccs/teaching/lsr/

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

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. Journal of Statistical Software, 48(2), 1-36. URL http://www.jstatsoft.org/v48/i02/

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.