

PSY717 - Statistická analýza dat

VYUČUJÍCÍ

Mgr. Stanislav Ježek, PhD. – garance kurzu, přednášky, semináře

jezek@fss.muni.cz, 549494616, konzultační hodiny: středa 14 – 16, FSS 2.53

Mgr. Jan Širůček – přednášky, semináře

sirucek@fss.muni.cz, 549498263, konzultace po dohodě emailem, FSS 2.45

CHARAKTERISTIKA KURZU

Cílem kurzu je seznámit studenty se základy statistiky používané v psychologickém výzkumu. Studenti získají porozumění základním prvkům statistiky a dovednost je aktivně i pasivně používat. Studenti získají dovednost připravit data pro statistické zpracování, spočítat základní statistiky, otestovat běžné typy hypotéz. Kurz klade důraz i na komunikaci, tj. slovní popis výsledků i schopnost porozumět takto popsaným výsledkům v empirických kvantitativních studiích. V rámci kurzu jsou studenti seznamováni paralelně s českou i anglickou terminologií, aby byli po skončení kurzu schopni dále studovat a používat internetové zdroje.

Návaznosti kurzu

Kurz je úzce souvisí s výukou metodologie. Mnoho problémů v metodologii má statistický základ a naopak mnohé problémy či omezení statistiky je potřeba zohledňovat v metodologii. Doporučené pořadí absolvování kurzů je statistika – metodologie, avšak i při opačném pořadí má student možnost integrovat poznatky obou disciplin. Toto propojení je nezbytné pro vypracování diplomové práce a je součástí požadavků při státní bakalářské zkoušce.

Na statistice stojí také značná část psychometrie a tvoří tak přirozený základ pro studium psychodiagnostiky na magisterském stupni.

Předpokladem pro studium základů statistiky je běžné středoškolské vzdělání.

ORGANIZACE KURZU

Počet kreditů: 5

Ukončení kurzu: zkouška

PROGRAM

Během semestru proběhnou 3 setkání.

Program prvního setkání

Prezentace – vytvoření datové matice, kódování a kontrola dat, popisné statistiky

Zadání seminární práce

Konzultace k problémovým okruhům 1 – 3.

Program druhého setkání

Prezentace – korelace, lineární regrese, pravděpodobnost

Konzultace k problémovým okruhům 4 – 6.

Program třetího setkání

Prezentace – testování hypotéz, anova

Zpětná vazba na seminární práci

Konzultace k problémovým okruhům 7 – 10.

INFORMAČNÍ SYSTÉM MU

V informačním systému MU části Studijní materiály k předmětu PSY717 najdou studenti podklady k tématům uvedeným v sylabu především v podobě prezentací z přednášek a odkazů na další zdroje informací. Na stejném místě jsou též k dispozici pokyny nezbytné k plnění písemných úkolů, popř. doplňky k tomuto sylabu. Studijní materiály jsou společné s prezenčními variantami tohoto kurzu.

KOMUNIKACE S VYUČUJÍCMI

Preferovaným komunikačním kanálem pro osobní komunikaci týkající se kurzu je email. Používejte prosím „PSY717“ v předmětu svých emailů. Urychlíte tím jejich vyřízení. Emaily adresujte dr. Ježkovi.

V případě dotazů týkajících se látky kurzu, využívejte prosím v maximální možné míře předmětové diskuzní fórum. 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 diskuzní fóra pravidelně sledují a přispívají do nich.

K důležitým ohlášením používáme hromadný email studentům.

POŽADAVKY NA UKONČENÍ KURZU

Seminární práce 1: Zamyšlení nad komunikováním statistiky v médiích

Zamyšlení nad způsoby, jimiž je statistika využívána či často zneužívána v médiích (popř. v odborné komunikaci). Úkolem studenta je vyhledat v populárních médiích (noviny, časopisy, lépe serióznější) článek či zprávu, která se opírá o nějaké statistiky (výzkum, průzkumy apod.), dohledat originální zdroj těchto dat (výzkumná zpráva apod.) a zamyslet se nad způsobem prezentace těchto statistik a jeho důsledky. Práce by měla obsahovat krátké shrnutí vybrané zprávy (zejm. závěry), shrnutí statistických metod (ukazatele, zobrazení) použitých v originálním zdroji, úvaha nad tím, co bylo vybráno (a případně pozměněno) do článku a jaké důsledky to má pro význam sdělení. Jádrem úvahy by měly být otázky jako: Jak dobře, přesně shrnuje novinový článek původní výzkumnou zprávu? Nepřehání článek závěry původní studie (přílišné zobecnění, optimismus)? Jsou zmíněny i potenciální problémy originální studie?

Tato seminární práce se zaměřuje na tzv. statistickou gramotnost (statistical literacy), kterou označujeme schopnost aktivně i pasivně komunikovat statistické výsledky. Být statisticky gramotný znamená nenechat se opít každým rohlíkem s pěknými čísly a grafy; zároveň však neodmítat apriori každou prezentaci nějakých statistických sdělení. Na internetu je mnoho informací k tomuto tématu; můžete vyhledávat též *misuse statistics* nebo *abuse statistics*. Viz též Utts (2003) či Gigerenzer (2002) ve studijních materiálech.

Lze využít i internetové noviny a časopisy. Lze použít i zpravodajství v anglickém jazyce.

Závazný rozsah práce je **4000 - 5000** znaků vč. mezer.

Termín: **30. 4.**

Zkouška

Kurz je zakončen zkouškou. Zkouška má písemnou podobu a je možné za ni získat **50b**. K úspěšnému složení zkoušky je nutné získat minimálně **25b**.

Zkouší se v rozsahu látky, který je vymezen v tomto sylabu k předmětu PSY717.

Celkové hodnocení

Celkem lze v průběhu kurzu získat 10b; za závěrečnou zkoušku lze získat 50b.

Celkové hodnocení bude používat následující stupnici

A: 60 - 53b **B:** 52 - 48b **C:** 47 - 42b **D:** 41 - 36b **E:** 35 - 30b **F:** 29 a méně bodů.

Uznávání dříve absolvovaných kurzů statistiky

1. **Kurzy dříve absolvované na katedře psychologie FSS** budou uznávány v plné míře¹. Své žádosti o uznání v takovém případě směrujte sekretárce katedry psychologie.

2. **Kurzy absolvované jinde než na katedře psychologie FSS** budou uznávány tou formou, že jejich absolventi nebudou povinni plnit požadavky na práci v semestru (seminární práce) a bude jim za ně započítán plný počet bodů (10b). I nadále jsou však povinni absolvovat závěrečnou zkoušku. Žádosti o uznání v tomto případě zasílejte na mail jezek@fss.muni.cz. V předmětu zprávy použijte „uznani PSY717“ (bez diakritiky). V textu stačí uvést kód absolvovaného předmětu na FSS. V případě žádosti o uznání kurzů z jiné fakulty či univerzity, uveďte též odkaz na sylaby absolvovaných kurzů, popř. sylaby přímo přiložte. **Žádosti zasílejte do konce 2. týdne semestru.** Později zasláným žádostem nebude vyhověno.

¹ Platí zejména pro studenty, kteří si zapsali PSY454 a již z bakalářského studia mají absolvován kurz PSY117.

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 dodržением některého z termínů, informujte o tom některého z vyučujících co nejdříve.

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.

LITERATURA

Základní zdroje

- Hendl, J.: *Přehled statistických metod zpracování dat. Analýza a metaanalýza dat*. Brno: Portál 2004.
 Osecká, L., Osecký, P.: *Receptář jednoduchých metod statistické indukce*. Brno, AV ČR 1996.²
 Urbánek, T.: K prezentaci výsledků statistických analýz, 1. část. *ČsPsych*, 2007 (51), 6, 601-609.
 Urbánek, T.: K prezentaci výsledků statistických analýz, 2. část. *ČsPsych*, (52), 1, 70-79.
 Cohen, J.: The Earth is round ($p < .05$). *American psychologist*, 1994 (49), 12, 997-1003.²
 Cohen, J.: A Power primer. *Psychological Bulletin*, 1992 (112), 1, 155-159.²
 Utts, J.: What educated citizens should know about statistics and probability. *American Statistician*, 2003 (57), 2, 74-79.
 Good, P. I., Hardin, J. W.: *Common errors in statistics (and how to avoid them)*. Wiley-Interscience 2003.
 Morgan, S. E., Reichert, T., Harrison, T. R.: *From numbers to words. Reporting statistical results for the social sciences*. Allyn & Bacon, 2002.
 Abelson, R. P.: *Statistics as principled argument*. Lawrence Erlbaum Associates, 1995.

Další užitečné publikace v češtině

Základy statistiky jsou staré, takže i poměrně letité knihy obsahují totéž, co ty nové s barevnými obrázky. Základy jsou ve všech knihách stejné, takže si student může vybrat takovou, která je napsána jemu srozumitelným jazykem. Starší učebnice se od novějších liší především podstatně menším důrazem na velikosti účinku, statistickou sílu a použití neparametrických metod³. Počítejte s tím, prosím. Například:

- Svoboda, H.: *Moderní statistika*. Praha: Svoboda, 1977.⁴
 Lamser, V., Růžička, L.: *Základy statistiky pro sociology*. Praha: Svoboda, 1970.

O měření v psychologii pojednává přístupná kniha: Urbánek, T.: *Základy psychometrie*. PsÚ AV ČR a PsÚ FF MU, 2002.

Cizojazyčná literatura

S výjimkou Svobodova textu jsou všechny učebnice v češtině považovány studenty psychologie za obtížně srozumitelné (nikoli neprávem). Vřele doporučujeme studovat přednostně či paralelně z lety ověřených anglicky psaných učebnic. Např.

- Glass, G. V., Hopkins, K. D.: *Statistical methods in education and psychology, 3rd Ed.* Allyn and Bacon, 1996.⁵
 Grimm, L. G.: *Statistical applications for the behavioral sciences, 4th Ed.* Wiley, 1993.⁶
 Moore, D. S.: *The basic practice of statistics, 4th Ed.*, Freeman, 2006⁷.
 Moore, D. S., McCabe, G. P., Craig, B.: *Introduction to the Practice of Statistics. 6th Ed.*, Freeman, 2007.
 Coolican, H.: *Research methods and statistics in psychology, 4th Ed.* Hodder & Stoughton, 2004.

² Dostupné ve studijních materiálech

³ Totéž platí pro rozdíl mezi špatnými a dobrými učebnicemi.

⁴ Studenti si chválí srozumitelnost textu. Je v několika výtiscích v knihovně FSS.

⁵ Můj hlavní zdroj, mohu zapůjčit.

⁶ Populární na katedře psychologie FF UP v Olomouci.

⁷ Profesor David Moore je velmi oceňovaným autorem statistických učebnic. Jako jeden z mála dokáže zjednodušovat a zároveň se vyhnout zavádějícím zjednodušením. Díky tomu ho oceňují nejen studenti, ale i statistikové.

Welkowitz, J., Ewen, R. B., Cohen, J.: *Introductory statistics for the behavioral sciences*, 5th. Harcourt Brace, 2000⁸.

Blaikie, N. : *Analyzing quantitative data*. Sage, 2003⁹.

de Vaus, D.: *Analyzing social science data: 50 key problems in data analysis*. Sage, 2002¹⁰.

Pro mnohé studenty je výhodné studovat paralelně i některou z učebnic používání SPSS. Příkladem vhodných jsou:

Morgan, G. A. et al: *SPSS for introductory statistics. Use and interpretation*, 2nd Ed. LEA, 2004.

Field, A.: *Discovering statistics using SPSS*, 2nd Ed. Sage, 2005. (nebo 1. vydání z r. 2000)¹¹

Pozor na velmi jednoduché příručky k SPSS; ty často nenabízejí žádný vhled do statistiky. Pro většinu uživatelů Windows znalých základů statistiky je používání SPSS snadné a intuitivní, takže pořizování takové příručky není nutné.

Velmi užitečným a přístupným slovníkem statistických a příbuzných termínů v anglickém jazyce je Everitt, B. E., Wykes, T.: *A dictionary of statistics for psychologists*. Arnold, 1999.

Příručka k tomu, jak dělat statistiku v Excelu: Gupta, V.: *Statistical analysis with Excel*. VJ Books, 2002.

Rozšiřující texty

Rozšiřující články a kapitoly jsou k dispozici ve studijních materiálech v ISu.

Tversky, A., Kahneman, D.: Belief in the law of small numbers. In D. Kahneman, P. Slovic, A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases*. Oxford: Oxford University Press, 1982.

Gigerenzer, G.: In the year 2054: Innumeracy defeated. In P. Sedlmeier, T. Betsch (Eds.), *etc: Frequency processing and cognition*. Oxford: Oxford University Press, 2002.

Emerson, J. D., Hoaglin, D. C.: Stem-and-leaf displays. In D. C. Hoaglin, F. Mosteller, J. W. Tukey (Eds.), *Understanding Robust and Exploratory Data Analysis*, 7-32. New York: Wiley, 1983.

Internetové zdroje

Vynikající zdroj o statistice, spíše pro pokročilé. <http://www.statsoft.com/textbook/stathome.html>

Zdroj o zdrojích: <http://en.wikipedia.org/wiki/Statistics>

Online učebnice: <http://en.wikibooks.org/wiki/Statistics>

Software

Pro procvičování statistiky je nutná zejména zručnost v používání tabulkového kalkulátoru, např. MS Office Excel či OpenOffice.org Calc. Většinu základních statistik je možné spočítat v tabulkovém kalkulátoru. Stejně tak správa dat je v malých výzkumných projektech podstatně pohodlnější v tabulkových kalkulátorech.

Na univerzitách jsou běžně využívány obecné statistické softwarové balíky, u nás SPSS a Statistica. Ty jsou však velmi drahé a pro běžnou praxi obsahují spoustu nepotřebných funkcí. Univerzální a drahý software lze dnes snadno nahradit použitím malých specializovaných aplikací na internetu. Jejich přehled naleznete na <http://statpages.org/>.

I když práce se statistickým software není těžištěm tohoto kurzu, doporučujeme studentům, aby se statistickými programy v průběhu kurzu zkusili pracovat a naučili se pracovat s jejich výstupy.

⁸ Stručná, klasika s moderními prvky, dobrá. Nemaže med kolem huby, jde přímo k cíli. Dobré zaměření na sílu testu.

⁹ Učebnice pro odvážné. Autor je trochu narcistně posedlý vlastním novátorstvím, které "vhodně" doplňuje minimum citovaných zdrojů :-). Učebnice doplňuje statistiku o kousky metodologie/epistemologie (co jsou to data, co je kauzalita apod.). Myslím, že není dobré si i tak obtížný úkol pochopení základů statistiky komplikovat žabomyšími epistemologickými spory (což autor dělá). Pokud ale máte tendenci se hodně zamýšlet nad tím, co jsou data, proměnné a proč "ti statistiku děláme", je to dobrý zdroj.

¹⁰ Výtečný zdroj pro lidi, co absolvovali kurz statistiky, právě pracují na diplomce a potřebují poradit s některými praktickými problémy analýzy. Vhodné i jako doplňující zdroj při studiu. Zahrnuje všechny aspekty kvantitativního výzkumu. V každé kapitole popisuje jeden problém (vč. vysvětlení, v čem je vlastně problém) a jeho řešení.

¹¹ Velmi ceněná učebnice. Trochu ale přeskakuje základy, takže bychom ji doporučili spíše k doplňkovému studiu.

PROBLÉMOVÉ OKRUHY

Níže uvedené problémové okruhy tvoří obsah PSY717 (a odpovídají jednotlivým přednáškám v prezenčním studiu PSY117/454). Jedná se o relativně rozsáhlé okruhy, takže některé pojmy a přístupy uvedené v sylabu se na přednáškách objeví pouze v podobě odkazu na literaturu k samostudiu.

1. Proměnné, výzkumný kontext

Data, proměnné, úrovně měření, kvalita měření, organizace dat, kontrola dat. Tvorba datové matice (v Excelu a SPSS), kódování proměnných.

Literatura: Hendl 43–50; Urbánek 12–24

2. Zobrazování dat, četnosti, distribuce

Tabelace dat, šíře intervalů, minimum, maximum, odlehlá hodnota (outlier), absolutní a relativní četnosti (frekvence), kumulativní absolutní a relativní četnosti, rozložení (rozdělení) četností(dat), tvary rozložení (normální, bimodální, uniformní, pozitivně zešíkmené, negativně zešíkmené), normální (Gaussovo) rozložení, velikosti oblastí pod křivkou normálního rozložení, Poissonovo rozložení, graf absolutních a relativních četností, sloupcový graf, histogram

Literatura: Hendl 85–92; 100; 139; Good, Hardin (2003) 107–125; stonkolist: Emerson, Hoaglin (2003), Gaussova křivka: Swoboda 73–87.

3. Míry centrální tendence a variability

Modus, medián, průměr, vážený průměr, vhodnost použití různých měr centrální tendence, (variační) rozpětí, kvartilové rozpětí, směrodatná odchylka (populační, výběrová), rozptyl, vliv přičítání konstanty a násobení konstantou na m a s , z -skóry a další standardní skóry (T , IQ), percentily, šikmost, špičatost, krabicový graf s anténami

Literatura: Hendl 93–99

4. Pravděpodobnost

Pojetí pravděpodobnosti, počítání s pravděpodobnostmi, náhodné jevy, podmíněné pravděpodobnosti, Bayesův teorém, normální pravděpodobnostní rozložení a další běžná rozložení.

Literatura: Hendl 115–127; 134; Swoboda 29 – 33

5. Vztahy mezi proměnnými, korelace

Korelace – Pearsonův, Spearmanův, Kendallův koeficient a jejich vlastnosti. Koeficient determinace, kovariance. Kontingenční tabulka, marginální četnosti. Lineární vztah, monotónní vztah, pozitivní a negativní vztah. Těsnost vztahu. Bodový graf. Parciální a semiparciální korelace. Korelace mezi položkami, Cronbachovo α .

Literatura: Hendl 237–265

6. Lineární regrese

Statistická predikce, lineární vs. nelineární regrese, lineární a kvadratická funkce, odhad, modelování, regrese, reziduum, prediktor, závislá a nezávislá proměnná, zdroje variability, stanovení regresní přímky metodou nejmenších čtverců, regresní rozptyl a reziduální rozptyl, koeficient determinace jako ukazatel úspěšnosti regrese, homoskedascita, mnohočetná (mnohonásobná) regrese, logistická regrese

Literatura: Hendl 266 - 290

7. Statistická indukce, intervalové odhady

Vzorek(výběr), statistiky vs. parametry, estimační charakteristiky popisných statistik, výběrová rozložení, centrální limitní teorém, směrodatná chyba (průměru), výběrové rozložení průměru, relativní četnosti, rozptylu, bodové vs. intervalové odhady.

Literatura: Hendl 146–147, 150–174

8. Testování hypotéz

Statistická(nulová) hypotéza, výzkumná (alternativní) hypotéza, jednostranná vs. oboustranná hypotéza(test); Bayesovský přístup k testování hypotéz vs. Fisher-Pearson-Neymanovský (tradiční) přístup, úroveň(hladina) statistické významnosti, chyba I. a II. typu a jejich pravděpodobnost, (statistická) síla testu,

jednovýběrový t -test, dvouvýběrový t -test (pro nezávislé výběry), párový t -test (z -test), Levenův test, testování korelačního koeficientu, velikost efektu, Cohenovo d , transformace z d na r a naopak.

Literatura: Hendl 175–220; 234; 401–412; 252–253; 413–419; Osečtí 1-36, Cohen 1992, 1994.

9. Testy pro nominální a ordinální proměnné

Parametrické vs. neparametrické testy, znaménkový test, test relativních četností, test dobré shody (χ^2), závislost kategoriálních proměnných (χ^2 , Cramerovo V , koeficient shody pozorovatelů κ), Wilcoxonovy testy (jednovýběrový, dvouvýběrový), Mann-Whitney U .

Literatura: Hendl 191–197; 220–232, 297–302, 304–313, 322–323

10. Analýza rozptylu

Problém s prováděním většího počtu testů, rybaření v datech, Bonferroniho korekce, princip analýzy rozptylu, rozptyl mezi skupinami (SS, MS), rozptyl uvnitř skupin (SS, MS), statistika F (F -test), analýza rozptylu s jedním faktorem (one-way), předpoklady analýzy rozptylu, post-hoc testy (S-N-K, Scheffe, LSD), velikost účinku (η^2 , ω^2), interakce faktorů

Literatura: Hendl 198–200, 337–346, 354–356