

## PSY 117 - Statistická analýza dat

Jaro 2016

### VYUČUJÍCÍ

- Mgr. Stanislav Ježek, PhD. – garance kurzu, přednášky, semináře  
jezek@fss.muni.cz, 549494616, konzultační hodiny: středa 14 – 16, FSS 2.47
- Mgr. Jan Širůček, PhD. – přednášky, semináře  
sirucek@fss.muni.cz, 549498263, konzultace po dohodě emailem, FSS 2.47
- Mgr. Hynek Cígler – semináře  
hynek.cigler@mail.muni.cz, konzultace po dohodě emailem, FSS 2.47
- Mgr. Vít Gabrhel – semináře  
vit.gabrhel@mail.muni.cz, konzultace po dohodě emailem

### CHARAKTERISTIKA KURZU

Cílem kurzu je seznámit studenty se základy statistiky používané v psychologické praxi a výzkumu a běžném životě. Studenti získají statistickou gramotnost - 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 studiích. V rámci kurzu budou studenti seznamováni paralelně s českou i anglickou terminologií, aby byli po skončení kurzu schopni dále studovat a používat internetové zdroje.

#### *Návaznosti kurzu*

Kurz úzce souvisí s výukou metodologie. Mnoho problémů v metodologii má statistický základ a naopak mnohé problémy či omezení statistiky je potřeba zohledňovat v metodologii. Studijním programem stanovené pořadí absolvování kurzů je statistika – metodologie a v souladu s tím je koncipován obsah těchto povinných kurzů. 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 matematické vzdělání a funkční znalost kancelářského software (zejm. tabulkového kalkulátoru).

### ORGANIZACE KURZU

- Počet kreditů: 5
- Ukončení kurzu: zkouška
- Přednášky: 1 hodina týdně
- Semináře: 1 hodiny týdně

Do seminárních skupin se studenti zapisují prostřednictvím informačního systému MU.

### INFORMAČNÍ SYSTÉM MU

V informačním systému MU části Studijní materiály k předmětu PSY117 umístujeme 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.

### KOMUNIKACE S VYUČUJÍCÍMI

Preferovaným komunikačním kanálem pro osobní komunikaci týkající se kurzu je email. Používejte prosím „PSY117“ v předmětu svých emailů. Urychlíte tím jejich vyřízení.

Emaily adresujte dr. Ježkovi. Pouze v případě záležitostí, které se týkají specificky seminářů, pište přímo vedoucím seminářů.

Pro dotazy týkajících se látky kurzu využijte prosím v maximální míře předmětové diskuzní fórum v IS nebo Facebookovou diskuzní skupinu „Statistika, metodologie, psychometrika“ <http://goo.gl/Mt95eT>. Je pravděpodobné, že odpověď na Váš dotaz by mohla 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

Úspěšné zvládnutí kurzu předpokládá samostatné studium ze základní literatury popsané v tomto sylabu, vyřešení všech úloh k procvičování, které jsou ve studijních materiálech PSY117 a splnění následujících hodnocených aktivit:

### Seminární práce

V průběhu semestru mají studenti za úkol pracovat na níže uvedených úkolech. Jejich realizace je podmínkou pro to, aby se student mohl přihlásit ke zkoušce. Seminární úkoly jsou zadávány na seminářích a v případě potřeby doplňovány informacemi v IS.

Práce se se odevzdávají v *odevzdávací* ve studijních materiálech předmětu. Jméno vkládaného souboru s první seminární prací musí být **S1.doc** (resp. **S2.doc** pro případnou druhou)<sup>1</sup> a nic víc; IS k němu automaticky přidá jméno vkládajícího studenta. Komentář/popisek nechtejte nevyplněný.

Seminární práce jsou hodnoceny na následující škále: **přijata (10b), nepřijata (0b)**. „Nepřijetím“ je míněno vrácení k přepracování. Na přepracování má autor týden ode dne vrácení. Vynikajícím pracím může být uděleno ještě 5 bodů jako bonus. Opravené práce vkládejte do ISu do stejné odevzdávací jako práce původní. Jako jméno vkládaného souboru nyní použijte **S1o.doc**. Přijaté opravené práce budou hodnoceny **5 body**.

Stanovené rozsahy prací jsou záměrné a závazné. Práce nedosahující minima i práce překračující maximum budou automaticky vráceny bez komentářů jako nepřijaté.

### Seminární práce 1: Zamyšlení nad komunikováním statistiky v médiích

Zamyšlení nad způsoby, jimiž je statistika používána 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 statistických sdělení. Věnujte pozornost také induktivním aspektům statistiky. Na internetu je mnoho informací k tomuto tématu; můžete vyhledávat též *misuse statistics* nebo *abuse statistics*. Viz též Utts (2003), Gigerenzer (2002), Wainer (2007), Savage Wainer (2008) ve studijních materiálech, popř. Hooke (1983) v rozšiřující literatuře.

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

Seminární práci lze zpracovat i na jiné téma, např. z oblasti historické provázanosti statistiky a psychologie, vizualizace dat či kontroverzí kolem statistického testování hypotéz. Lze uvažovat i o jiném formátu sdělení, než eseji či úvaze. Pokud chcete zpracovávat nějaké takové téma, učíte prosím písemnou nabídku emailem svému vedoucímu seminární skupiny, a to do **31.3**. Nabídka by měla na několika řádcích obsahovat základní tezi a navrhovaný způsob/formát zpracování. Zvolené téma lze zpracovat pouze po odsouhlasení vedoucími seminární skupiny (nejde o cenzuru, jen chceme předejít zbytečným zklamáním).

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

Termín: **1.5.**

<sup>1</sup> 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.

## Seminární práce 2: Výběrové šetření – skupinový projekt

Záměrem tohoto úkolu je vyzkoušet si realizaci jednoduchého výběrového šetření. Úkolem pětičlenných týmů bude zvolit si jednoduchou výzkumnou otázku (př. výskyt nějakého jevu, názoru v nějaké populaci), pokusit se na ni získat odpověď jednoduchým výběrovým šetřením a celý projekt prezentovat v podobě *posteru* na přednáškovém setkání. Detailní instrukce budou zadány na semináři a vloženy do studijních materiálů.

Termín: **18.5.**

### Průběžné hodnocení

V průběhu semestru budou na semináře zařazeny tři krátké průběžné desetiminutové testy. V každé bude možné získat 10b. Předběžné termíny průběžných testů jsou: **23.3.; 13.4.; 4.5.**

Množství bodů z průběžných písemek, které se bude počítat do celkového hodnocení, se počítá podle následujícího excelovského vzorce: =ZAOKROUHLIT((SUMA(P1;P2;P3)-MIN(P1;P2;P3))\*1,5;0). Tento výpočet eliminuje jedno zakolísání směrem dolů (včetně případné absence).

Na průběžné písemky se náhradní termíny poskytují pouze výjimečně.

### Zkouška

Kurz je zakončen zkouškou. Zkouška má písemnou podobu (papír-tužka) a je možné za ni získat **50b.** K úspěšnému složení zkoušky je nutné získat minimálně **30b.**

Zkouší se v rozsahu látky, který je vymezen v tomto sylabu k předmětu PSY117.

### Celkové hodnocení

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

Celkové hodnocení bude používat následující stupnici

**A: 100 - 92b   B: 91 – 84b   C: 83 – 77b   D: 76 – 71b   E: 70 – 65b   F: 64 a méně bodů.**

### Uznávání dříve absolvovaných kurzů statistiky

1. **Ekvivalentní kurzy dříve absolvované na katedře psychologie FSS** budou uznávány v plné míře. Své žádosti o uznání v takovém případě směřujte 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, průběžné zkoušky) a bude jim za ně započítán plný počet bodů (50b). I nadále jsou však povinni absolvovat závěrečnou zkoušku. Žádosti o uznání v tomto případě zasílejte na mail [jezek@fss.muni.cz](mailto:jezek@fss.muni.cz). V předmětu zprávy použijte „uznani PSY117“ (bez diakritiky). V textu stačí uvést kód absolvovaného předmětu na FSS (typicky SOC108). 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.

3. **Specificky, absolventi kurzu SOC108** mohou žádat o úplné uznání PSY117, pokud ukončili SOC108 s hodnocením „A“.

### 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. Omluvenky za nepřítomnost na seminářích, zejména pro nemoc, odevzdávejte prosím na Studijním oddělení FSS – jeho pracovnice omluvenku zavedou do IS.

### 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. [JH]: *Přehled statistických metod zpracování dat. Analýza a metaanalýza dat. 4. vydání.* Brno: Portál 2012.
- Howitt D., Cramer, D.[HC]: *Introduction to statistics in psychology, 5th.* Pearson, 2011. Aktuálně i 6. vydání z 2013. Online materiály pro studenty na [http://wps.pearsoned.co.uk/ema\\_uk\\_he\\_howitt\\_statspsych\\_5/175/44877/11488739.cw/index.html](http://wps.pearsoned.co.uk/ema_uk_he_howitt_statspsych_5/175/44877/11488739.cw/index.html).
- Howell, David C. [DH] *Statistical methods for psychology. 8th ed.* Belmont, CA: Wadsworth Cengage Learning, 2013.

## Články a výňatky:

- Cohen, J.: The Earth is round ( $p < .05$ ). *American psychologist*, 1994 (49), 12, 997 – 1003.
- Cohen, J.: A Power primer. *Psychological Bulletin*, 1992 (112), 1, 155 – 159.
- 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.<sup>2</sup>
- Baron, J.[1]: Hypotheses testing. Kapitola 7 v J. Baron: *Thinking and deciding*, 4<sup>th</sup>, s. 161- 182. Cambridge: CUP, 2008.
- Baron, J.[2]: Judgement of correlation and contingency. Kapitola 8 v J. Baron: *Thinking and deciding*, 4<sup>th</sup>, s. 183 - 198. Cambridge: CUP, 2008.
- Morgan, G. A. et al: Chapter 2 - Data coding, entry and checking. In G. A. Morgan et al: *SPSS for introductory statistics. Use and interpretation, 2<sup>nd</sup> ed*, 15 – 22. LEA, 2004.
- Savage, S., Wainer, H.: Until Proven Guilty: False Positives and the War on Terror. *Chance*, 2008 (21), 1, 59 – 62.
- Swoboda, H.: *Moderní statistika*. Praha: Svoboda, 1977.<sup>3</sup>
- Tversky, A., Kahneman, D.: Belief in the law of small numbers. In D. Kahneman, P. Slovic, A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*, 23 – 31. Oxford: Oxford University Press, 1982. Více též v přednášce při přebírání Nobelovy ceny na [http://nobelprize.org/nobel\\_prizes/economics/laureates/2002/kahneman-lecture.html](http://nobelprize.org/nobel_prizes/economics/laureates/2002/kahneman-lecture.html).
- 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*, 2008 (52), 1, 70 – 79.
- Urbánek, T., Denglerová, D., Širůček, J.[UDŠ]: *Psychometrika - měření v psychologii*. Praha: Portál 2011.
- Utts, J.: What educated citizens should know about statistics and probability. *American Statistician*, 2003 (57), 2, 74 – 79.
- de Vaus, D.: How to prepare data for analysis. In D. de Vaus, *Analyzing social science data: 50 key problems in data analysis*, 1 – 16. Sage, 2002
- Wainer, H.: The most dangerous equation: Ignorance of how sample size affects statistical variation has created havoc for nearly a millenium. *American Scientist*, 2007, 95 (3), 249 – 256.

## Rozšiřující zdroje

- Abelson, R.: *Statistics as a principled argument*. Hillsdale: LEA, 1995.
- Arbuthnott, J.: An argument for Divine Providence taken from the constant regularity observ'd in the birth of both sexes. *Philosophical Transactions*, 1710 (27), 186 – 190.
- Cowles, M.: *Statistics in psychology. An historical perspective, 2<sup>nd</sup> Ed.* Mahwah: LEA, 2001.
- Field, A.: *Discovering statistics using SPSS, 4<sup>th</sup> Ed.* Sage, 2013. (popř. předchozí vydání)<sup>4</sup>
- Freeman, J., Walters, S.J., Campbell, M.J.: *How to display data*. Blackwell, 2008.<sup>5</sup>
- Gaito, J.: Measurement Scales and Statistics: Resurgence of an Old Misconception. *Psychological Bulletin*, 1986, 87 (3), 564 – 567.
- Gigerenzer, G.: In the year 2054: Innumeracy defeated. In P. Sedlmeier, T. Betsch (Eds.), *etc: Frequency processing and cognition*, 55 – 66. Oxford: Oxford University Press, 2002.

<sup>2</sup> Pokud nepochopíte konstrukci stonkolistu z Hendela...

<sup>3</sup> Starý, ale čtivý text. Dobrý jako popularizační úvod do statistiky. Je v několika výtiscích v knihovně FSS.

<sup>4</sup> Velmi ceněná učebnice. Přeskakuje základy a jde přímo na věc se SPSS. V knihovně je třetí vydání nejméně v 5 výtiscích prezenčně. Hlavní učebnice navazujícího kurzu PSY252.

<sup>5</sup> Vynikající, jednoduchý zdroj o grafickém zobrazování dat.

- Glass, G. V., Hopkins, K. D. [GH]: *Statistical methods in education and psychology, 3rd Ed.* Allyn and Bacon, 1996.<sup>6</sup>
- Good, P. I., Hardin, J. W.: *Common errors in statistics (and how to avoid them)*. Wiley-Interscience 2003.
- Huck, S.: *Reading statistics and research, 6<sup>th</sup>*. Pearson, 2012.
- Lord, F. M.: On the statistical treatment of football numbers. *American Psychologist*, 1953, 8, s. 750-751.
- Morgan, S. E., Reichert, T., Harrison, T. R.: *From numbers to words. Reporting statistical results for the social sciences*. Allyn & Bacon, 2002.
- Nickerson, R. S.: Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological methods*, 2000 (5), 2, 241 – 301.
- Osecká, L., Osecký, P.: *Receptář jednoduchých metod statistické indukce*. Brno, AV ČR 1996.<sup>7</sup>
- Scholten, A. Z., Borsboom, D.: A reanalysis of Lord's statistical treatment of football numbers. *Journal of Mathematical Psychology*, 2009 (53), 69 – 75.
- Stone, J. V.: *Bayes' rule: a tutorial introduction to bayesian analysis*. Sebtel Press, 2013.

### Statistický populár pro rozšíření obzorů

- Mlodinow, L.: *Život je jen náhoda. Jak náhoda ovlivňuje naše životy*. Praha: Slovart, 2009.
- Best, J.: *Damn lies and statistics: untangling numbers from the media, politicians and activists*. Berkeley and Los Angeles: University of California Press, 2001.
- Best, J.: *More damn lies and statistics: how numbers confuse public issues*. Berkeley and Los Angeles: University of California Press, 2004.
- Cowles, M.: *Statistics in psychology – an historical perspective, 2<sup>nd</sup>*. LEA, 2001<sup>8</sup>.
- Huff, D.: *How to lie with statistics*. New York: W.W. Norton & Company, Inc., 1954.
- Hooke, R.: *How to tell the liars from the statisticians*. New York: Marcel Dekker, 1983.<sup>9</sup>
- Salsburg, D.: *The lady tasting tea: how statistics revolutionized science in the twentieth century*. W. H. Freeman and Company, 2001.
- Woolfson, M. M.: *Everyday probability and statistics. Health, elections, gambling and war*. London: Imperial College Press, 2008.

### Internetové zdroje

- Zdroj o zdrojích, online kalkulačkách apod.: <http://statpages.org/>
- Bohatá doplňující stránka k Huck (2012): <http://www.readingstats.com>
- Vynikající zdroj o statistice, spíše pro pokročilé. <http://www.statsoft.com/textbook/stathome.html>
- 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 než specializovaném statistickém software.

Na univerzitách jsou běžně využívány obecné statistické softwarové balíky, u nás SPSS a Statistica. Ty jsou však velmi drahé a pro běžnou akademickou praxi obsahují spoustu nepotřebných funkcí. Univerzální a drahý software lze dnes snadno nahradit použitím malých specializovaných aplikací na internetu. Jejich přehled naleznete např. na <http://statpages.org/>. V současnosti roste popularita otevřeného systému pro statistické výpočty R (<http://www.r-project.org/>).

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

G\*Power (<http://www.psych.uni-duesseldorf.de/aap/projects/gpower/>) je program pro kalkulaci související se silou testu. Zvláště vhodný pro odhad potřebné velikosti vzorku při plánování výzkumu.

<sup>6</sup> Původní hlavní ideový zdroj kurzu, mohu zapůjčit. Výňatky (co nenajdu v [HC]) budou v ISu.

<sup>7</sup> Dostupné ve studijních materiálech

<sup>8</sup> Příběh toho, jak se psychologie a statistika společně vyvíjely.

<sup>9</sup> Jedno až dvoustránková zamyšlení nad statistikou v každodenním životě. Zdaleka není tak stará, jak vypadá.

## PROBLÉMOVÉ OKRUHY

Níže uvedené problémové okruhy jsou součástí přednášek předmětu PSY117. 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, úroveň měření, kvalita měření, organizace dat, kontrola dat. Tvorba datové matice (v Excelu a SPSS), kódování proměnných.

Literatura: [HC] kap 1 (nebo [DH] kap 1); 43–50; [UDŠ] 59–66; [JH] 81-85; Morgan at al; de Vaus

### 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 zešikmené, negativně zešikmené), normální (Gaussovo) rozložení, velikosti oblastí pod křivkou normálního rozložení, Poissonovo rozložení, graf absolutních a relativních četností, sloupcový graf, histogram

Literatura: [HC] kap 2 a 4 (nebo [DH] kap 2 s15-31 + kap 3); [JH] kap 3; [GH] kap. 6; Good, Hardin (2003) 107–125; stonkolist: Emerson, Hoaglin (2003), Gaussova křivka: Swoboda 73–87.

### 3. Míry centrální tendence a variability, transformace

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 ( $T$ ,  $IQ$ ), normalizované skóry, percentily, šikmost, špičatost, krabicový graf s anténami

Literatura: [HC] kap 3 a 5 (nebo [DH] kap 2 s32-62); [JH] kap 3; [UDŠ] 245 – 254; [GH] 94.

### 4. Pravděpodobnost a pravděpodobnostní rozložení

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

Literatura: [JH] kap 4 (121–140), 139-145; [GH] kap. 6 a 9; Swoboda 29 – 33; [HC] kap 15 (nebo [DH] kap 5)

### 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  $\alpha$ .

Literatura: [HC] kap 6, 7, 36 a v druhé půli semestru 10 ; [JH] 247–276; Baron[2]

### 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: [HC] kap 8; [JH] 277 – 290

### 7. Statistická indukce, intervalové odhady

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

Literatura: [HC] kap 9, 11, 37; Hendl 156–181

### 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: [HC] kap 10, 12, 13, 16, 17, 34; Hendl 181–196; 213–230; 245; 417–441; 252–253; 413–419; Osečí 1-36; Cohen 1992, 1994; Arbuthnott; Baron[1]

### 9. Testy pro nominální a ordinální proměnné

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

Literatura: [HC] kap 14,18; Hendl 197–204; 230–244, 307–338

### 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 ( $\eta^2$ ,  $\omega^2$ ), interakce faktorů

Literatura: [HC] kap 19, 20 Hendl 204–206, 347–360

## ČASOVÁ OSNOVA

Datum	Téma přednáškového setkání	Téma semináře	Důležité termíny
24. 2.	Proměnné, zobrazování dat, četnosti, distribuce	tvorba datové matice	Žádosti o uznání
2. 3.	Míry centrální tendence a variability	četnosti, zobrazení rozložení	
9. 3.	Normální rozložení	míry centrální tendence a variability	
16. 3.	Pravděpodobnost	normální rozložení, z-skóry	
23. 3.	Korelace	PP1, pravděpodobnost	
30. 3.	Lineární regrese	korelace	31.3. alternativní téma seminární práce
6. 4.	Pořadové korelace, shrnutí	regrese	
13. 4.	Statistická indukce, intervalové odhady	PP2, pořadové korelace	
20. 4.	Testování hypotéz	intervalové odhady	
27. 4.	Přehledy testů, síla testu	t-testy, volba testu	
4. 5.	Testy pro nominální a ordinální proměnné	PP3, chíkvadrát	

11. 5.	Analýza rozptylu	chíkvadrát pokrač., ANOVA	6.5. seminární práce
18. 5.	Prezentace posterů – 2. seminární úkol. Všechny seminární skupiny dohromady.		