

Zpracování dat antropologických výzkumů AAV

Mgr. Patrik Galeta, Ph.D.

galeta@ksa.zcu.cz

www.oba.zcu.cz/personalia/PG.php

ZS 2022

Sylabus je platný ke dni 21. 9. 2022

<http://portal.zcu.cz/wps/portal/predmety/ksa/aav>

Anotace

Kurz AAV je úvodem do kvantitativních analýz v antropologii. Věnujeme se seznámení se specifiky dat kvantitativních výzkumů a se způsoby, jak množství vstupních dat zjednodušit do formy souhrnných číselných údajů, grafů a tabulek. V části kurzu nahlédnete i do problému statistického zobecňování, tj. způsobu jak z vlastností řádově stovek respondentů výběrových šetření odhadnout vlastnosti řádově miliónů lidí z celé populace. Intuitivně se seznámíte s principem intervalů spolehlivosti a testování hypotéz. Pro absolvování kurzu stačí základní matematické znalosti, výuku vedu intuitivní formou. Důraz kladu především na pochopení smyslu analýz.

Rozsah výuky

Přednáška (2 hod.) je každý týden, cvičení (2 hod.) je jednou za dva týdny. Pokud budete souhlasit, bude cvičení probíhat nepravidelně, podle potřeby a na konkrétních termínech bychom se domlouvali průběžně.

Formát prezenční výuky

Během cvičení budete v Počítačové laboratoři FF samostatně pracovat se software pro správu dat (Microsoft® Excel). Výstupem cvičení jsou vyřešené vzorové příklady ve formátu MS Excel. Soubory MS Excel jsou dynamické, tj. změnou vstupních dat se automaticky přepočítají výsledky a budete je tak moci použít na jiné obdobné problémy.

Účast na přednáškách a cvičeních není povinná. Případnou neúčastí ztratíte možnost podílet se na výuce a ovlivnit průběh kurzu, významně se sníží Vaše schopnost pochopit princip probírané látky a zvýší se pravděpodobnost nutnosti naučení se látky nazpaměť bez pochopení souvislostí.

Formát distanční výuky

V případě, že bude výuka probíhat distančně, pak výuka bude probíhat podle rozvrhu v Google Učebně, do které máte přístup výhradně pomocí Google Apps účtu na ZČU (<login>@gapps.zcu.cz).

Požadavky

(1) Domácí příprava

Na přednášky a zejména na cvičení se připravujte. Prostudujte podklady uvedené v harmonogramu výuky. Je to většinou několik stran z učebnic, které jsou v PDF na CW.

(2) Dokončení práce z minulých cvičení

Na cvičení si noste vyřešené úlohy ze všech předchozích cvičení (formát MS Excel), protože ke starším souborům se budeme vracet. Soubory si ukládejte na přenosné disky nebo elektronická úložiště. Uložení na disk počítače v laboratoři není bezpečné, nemusíte každý týden sedět u stejného počítače a vaše soubory může smazat jiný student nebo správce.

Ukončení kurzu

Kurz je ukončen zkouškou. Hodnocení zkoušky sestává ze dvou částí: Závěrečného testu a průběžných Kontrolních testů, a to podle těchto procentuálních vah.

1. **Závěrečný test (75 %):** Závěrečný test hodnotí teoretické a praktické znalosti z celého semestru. Test má písemnou formu a obsahuje 7–12 otázek. Všechny otázky předpokládají volnou slovní odpověď. Hodnotím především pochopení principu problému a schopnost interpretovat výsledky. Důležité je zdůvodnit postup, uvést pouze správný výsledek nestačí. Test budete vyplňovat ručně na papír, počítání je omezeno na minimum. Uspějete, pokud pochopíte souvislosti a budete logicky přemýšlet. K úspěšnému absolvování testu postačí pouze informace z přednášek a doporučené literatury. Závěrečný test má dva opravné termíny.

K testu si přineste pouze psací potřeby, papír, statistické tabulky a kalkulačku. Osobní počítač, mobilní telefon, tablet apod. jako náhradu kalkulačky můžete použít pouze v případě, že nebudete používat ostatní možnosti přístrojů.

2. **Kontrolní testy (25 %):** Kontrolní testy oceňují schopnost praktické práce s daty. Kontrolní testy jsou krátké a jednoduché úkoly vypracované v MS Excel či MS Word, při jejichž řešení si můžete pravidelně ověřovat, zda cvičení dobře zvládáte. Kontrolní test je vždy obdobou úloh řešených na předchozích cvičeních. Splněný test je ten, který je dobře vypočítaný (správné výsledky, správné napsané odkazy na správné buňky listů MS Excel apod.) a je formálně dobře upraven (zarovnání tabulek, barva, formát grafu apod.). Podrobnosti na cvičení.

Kontrolní testy budu zadávat na začátku většiny cvičení. Pro vypracování a odevzdání testů na Portál ZČU bude obvykle stačit okolo 15 minut. Pokud soubor na Portál uložíte vícekrát, platí poslední odeslaná verze. Výsledky hodnocení testů budete mít k dispozici do pátku do 23.59 daného týdne včetně. Za každý správně vyplněný Kontrolní test získáte poměrnou část z celkových 25 %. Pokud například bude během semestru zadáno celkem 10 testů, pak z každého můžete získat 2,5 %. Při odůvodněné a oznámené absenci na cvičení vám individuálně upravím celkový počet testů a poměrný zisk na jeden test se vám tak zvýší (25 % se tak nebude dělit 10, ale např. jen 9, pokud na jedno cvičení nebudete moci přijít). Absenci oznamte a doložte bez odkladu, maximálně do týdne od cvičení, na které jste nepřišli.

Kontrolní testy musíte vytvářet samostatně. Není možné spolupracovat s někým jiným. Nelze rovněž okopírovat aplikaci od někoho jiného a změnit například formát buněk, aby se převzetí "nepoznalo". Pokud test nebude výsledkem Vaší samostatné úvahy, získáte z Kontrolních testů do celkového hodnocení kurzu 0 %, a to bez ohledu na předchozí či následné výsledky Kontrolních testů.

Hodnocení Kontrolních testů má dva opravné termíny, které proběhnou na konci semestru nebo ve zkouškovém období. Opravné testy mají stejnou formu jako ostatní Kontrolní testy, jsou ale rozsáhlejší, osvědčujete v nich znalosti z celého semestru. Výsledky opravných testů nahrazují součet Kontrolních testů zadávaných v průběhu semestru. Do celkového hodnocení kurzu se bude započítávat nejlepší výsledek (tj. buď součet dílčích Kontrolních testů, nebo výsledek u opravných testů). Opravy dílčích Kontrolních testů nebudou.

Doporučujeme, abyste využili mou přítomnost ve výuce a na konzultačních hodinách. Rád Vám ukáži problematické body, které případně bránily uznání úkolu.

Hodnocení kurzu

88,0–100,0 %	Výborně
76,0–87,9 %	Velmi dobře
65,0–75,9 %	Dobře
0,0–64,9 %	Nevyhověl(a)

Kombinované studium

Výuka přednášek probíhá blokově v soboru podle rozvrhu. Pokud se domluvíme, může bloková výuka proběhnout on-line (pro mě preferovaná varianta). Podle uvážení můžete navštěvovat výuku rozvrhovanou pro studenty řádného studia.

Hodnocení zkoušky sestává pouze z testu. Máte k dispozici jeden řádný a dva opravné termíny testu. Ostatní parametry testu jsou stejné jako u testů pro řádné studenty. Termín zkoušky domluvíme v den blokové výuky.

Základní literatura

Doporučuji Vám jednu ze tří výborných učebnic zpracování dat. Ve všech najdete oporu pro většinu témat přednášek. Dvě učebnice jsou anglické, jedna česká.

Agresti A., Finlay B. (2009). Statistical methods for the social sciences. Upper Saddle River: Prentice Hall.

Osvědčený zdroj. Oproti učebnicím biostatistiky (Zar 1999) je určena pro studenty společenských věd. Je proto dobře možné, že bude pro Vás pochopitelnější. K dispozici je i nová edice knihy (Agresti, Finlay 2009).

Zar J. H. (2014). Biostatistical analysis. 5th edition, New Jersey: Prentice Hall.

Výborná učebnice. Její výhodou je, že se Vám pravděpodobně bude zdát přehlednější než Sokal, Rohlf (1995). Učebnice je zřetelně členěna do kapitol, které odpovídají tématům přednášek. Každý problém je dokumentován na vzorových početních příkladech, ve kterých můžete sledovat postup řešení.

Lepš J., Šmilauer P. (2016). Biostatistika. České Budějovice: Nakladatelství Jihočeské univerzity.

Legendy české biostatistiky. Z textu je patrná inspirace Zarovou učebnicí, ale není to pouhý překlad. Text je vhodný i pro nebiology. Recenze knihy jsou na internetu.

Informace ke kurzu

Všechny informace ke kurzu jsou vystaveny na Portále ZČU, sekce CourseWare. Aktuální informace budu posílat elektronicky na adresu, kterou máte uvedenou v Portálu. Co odešlu, máte vědět.

Případné obsahové a technické nedostatky stránek kurzu hlase prosím ihned na adresu galeta@ksa.zcu.cz, pomůžete udržet stránky v použitelném stavu.

Etické normy studia a ochrana duševního vlastnictví jiných autorů

Po celou dobu trvání kurzu jste povinni dodržovat etické normy studia a pracovat v souladu s pravidly ochrany duševního vlastnictví jiných autorů. Bližší informace viz Upozornění studentům vydaného dne 28. 1. 2005 prodekanem pro studijní záležitosti. V případě, že porušíte etické normy, vyhrazujeme si právo test či závěrečnou práci nehodnotit. Nejasnosti a dotazy konzultujte s vyučujícími.

Rozpis témat přednášek a cvičení

1. téma Měření

Jednotka výzkumu, proměnná, slovní a číselné hodnoty proměnných.

Typy proměnných, poměrové, intervalové, ordinální a kategoriální proměnné, spojitá a diskrétní data, kvantitativní a kvalitativní data.

Doporučená literatura

Agresti, Finlay (2009): 11–14

Zar (2014): 1–6

2. téma Popisná statistika I (tabulkové a grafické souhrny dat)

Zápis dat ve formě tabulek, řádky a sloupce tabulky, záhlaví a tělo tabulky, pravidla zápisu hodnot proměnných, číselné kódování slovních hodnot proměnných, desetinná čárka, počet desetinných míst.

Četnostní tabulky kvalitativních a kvantitativních dat, absolutní četnost, relativní četnost, absolutní kumulativní četnost, relativní kumulativní četnost, součet.

Grafické zobrazení struktury kvalitativních dat, sloupcový diagram, výsečový diagram.

Grafické zobrazení struktury kvantitativních dat, histogram, krabicový diagram (*boxplot*).

Sloupcový diagram vs. histogram.

Pojem rozdělení proměnné.

Doporučená literatura

Agresti, Finlay (2009): 31–38

Zar (2014): 6–15

3. téma Popisná statistika II (číselné souhrny dat)

Míry polohy, aritmetický průměr, medián, modus.

Míry rozptylu, rozsah, rozptyl, směrodatná odchylka, koeficient variance.

Grafické vs. číselné způsoby popisu struktury dat.

Zásady správného uvádění výsledků.

Doporučená literatura

Agresti, Finlay (2009): 38–58

Zar (2014): 23–51

4. téma Spojité proměnné a jejich rozdělení

Normální rozdělení a jeho parametry, šikmost, špičatost.

Standardizace dat, standardizované normální rozdělení, statistické tabulky.

Kvantily, percentily.
Odchyly od normality.

Doporučená literatura

Agresti, Finlay (2009): 78–85
Zar (2014): 72–78, 86–89

5. téma Odhady populačních parametrů a intervaly spolehlivosti

Populace, vzorek z populace, reprezentativnost vzorku, náhodný výběr.
Populační parametry a jejich nepoznatelnost, zobecnění vlastností vzorku na populaci, odhad pomocí charakteristik náhodného vzorku, spolehlivost odhadu.
Intuitivní odvození principu intervalu spolehlivosti průměru, konstrukce rozdělení výběrových průměrů, standardní chyba průměru, Studentovo rozdělení.
Přesnost a spolehlivost odhadu pomocí intervalu spolehlivosti, jak velký výběr potřebujeme, vztah velikosti vzorku a přesnosti a spolehlivosti odhadu.
Grafické vyjádření intervalů spolehlivosti, krabicový diagram.
Odvození a konstrukce rozdělení směrodatné odchylky, Pearsonovo rozdělení, interval spolehlivosti pro směrodatnou odchylku.

Doporučená literatura

Agresti, Finlay (2009): 4–6, 85–129
Zar (2014): 18–22, 78–80, 91–93, 111–118

6. téma Testování hypotéz I (jednovýběrové testy)

Základní pojmy testování hypotéz, nulová hypotéza H_0 , alternativní hypotéza H_A , chyba I a II druhu, hladina významnosti, síla testu.
Vztah chyb I a II druhu, velikost vzorku a chyba I a II druhu.
Zásady správného uvádění výsledků testování hypotéz.
Srovnání populačního parametru a charakteristik jednoho vzorku.
Jednovýběrový test pro průměr (*one-sample t-test*). Vliv velikosti souboru.
Jednovýběrový test pro relativní četnosti (*one-sample binomial test*).
Obecný postup při srovnání populačního parametru a charakteristik jednoho vzorku.

Doporučená literatura

Agresti, Finlay (2009): 143–166
Zar (2014): 103–126, 91–98