

Úvod

Využitie štatistických metód a štatisticky orientovaný spôsob uvažovania nie je v biologickej antropológii zaužívaný od počiatku jej vzniku v európskom novoveku, a ani v súčasnosti nepatrí k optimálne rozvinutým oblastiam antropológie. Napriek tomu, že celý rad štatistických metód vznikol v súvislosti s potrebami vedných odborov zaoberajúcich sa človekom (napr. lineárna regresná analýza vznikla v súvislosti s odhadovaním výšky postavy človeka na základe dĺžky kostí jeho skeletu, Pearson (1899)). Čiastočne je dôvodom po storočia pretrvávajúci typologický prístup k štúdiu človeka, v dôsledku čoho sa na dlhý čas zakonzervoval rasový koncept, ktorý brzdil akékoľvek štúdium ľudskej rozmanitosti a prispôbivosti. Antropologické štúdie devätnásteho storočia, napísané často na mnoho desiatkach až stovkách strán, do najmenšieho detailu popisovali niekoľko ľudí či niekoľko ľudských kostí, a bez akýchkoľvek údajov o spôsobe vzniku vzorky vyvodzovali d'alekosiahle závery napr. o „rasových“ rozdieloch (cf. Broca, 1862). Aj niektorí zakladatelia a priekopníci biologickej antropológie v storočí dvadsiatom sa využitiu štatistiky bránili. Dokonca i znamenitý Aleš Hrdlička (1869–1943), americký antropológ českého pôvodu, zakladateľ a dlhodobý editor dnes celosvetovo uznávaného odborného časopisu *American Journal of Physical Anthropology*, štatistike neprial a nepodporoval použitie pokročilých štatistických metód v článkoch publikovaných v tomto časopise (Ashley Montagu, 1944). Situácia sa začala meniť v polovici 20. storočia, kedy sa biologická antropológia začala orientovať ekologickým smerom a skúmať ľudské populácie a variabilitu ľudských vlastností v nich i medzi nimi ako výsledok dlhého procesu prispôbovania sa rôznym podmienkam prostredia (Mielke a kol., 2011, str. 3–22). Populačné vzorky, spôsob ich vzniku a ich vlastnosti (populačné mikroevolučné procesy) sa stali pre interpretáciu výsledkov a vyvodzovanie záverov zásadné, rovnako tak ako sa štatistika stala dôležitým nástrojom zabezpečujúcim platnosť (štatistickú validitu) antropologických záverov. Už prvá česko-slovenská učebnica antropológie má kapitolu zoznamujúcu s podstatou základov štatistického uvažovania (Suchý, 1967), nasledovaná vznikom niekoľkých ďalších prác o štatistike a bioštatistike, vhodnej pre antropológov, biológov a lekárov (napr. Zvára, 1999, 2001; Komenda, 2000). Aplikácia štatistických metód je dnes súčasťou takmer každej odbornej práce v biologickej antropológii, pričom pokroky v rozvoji štatistických metód umožňujú výber vhodnej metódy prispôbenej skúmaným javom a komplexné usporiadanie štatistických porovnaní. Výučba základných štatistických metód patrí medzi štandardné súčasti štúdia biologickej antropológie. Napomáha adeptom naučiť sa pristupovať k riešeným problémom koncepčne od samotného začiatku, t.j. počiatočného nápadu a návrhu výskumu až po rozhodovanie o závažnosti nimi zistených výsledkov. Ako v prípade každého nástroja aj v prípade štatistických metód platí, že pri správnom použití dobre spĺňajú svoj účel, ale pri nesprávnom použití môžu výsledky výskumu v neznámej miere strácať svoju platnosť.

Dnešný antropologický výskum sa bez uplatnenia štatistického uvažovania nezaobíde. I keď sa väčšina biológov, lekárov ani antropológov štatistikmi nikdy nestanú (až na výnimočné personálne únie), výsledky výskumov i možnosti ich publikácie na správnom usporiadaní a štatistickom spracovaní dajú silno závisia. Kľúčovým pre obojstranne úspešnú spoluprácu preto je na strane týchto odborov porozumieť základom štatistických postupov, vedieť sa vyjadrovať aspoň základnou štatistickou terminológiou a vedieť štatistikovi vysvetliť svoj výskumný zámer; na strane štatistika potom pochopiť podstatu zamerania jednotlivých odborov, charakter ich metodológie a typické problémy, s ktorými sa po štatistickej stránke stretávajú – a vytvoriť si vzťah obojstranného porozumenia a prospešnosti. Jedným z cieľov tejto knihy je podporiť kladný prístup mladých ľudí k využívaniu štatistických metód v antropológii a ďalších odboroch a umožniť ich aplikáciu v prostredí programu \mathbb{R} , univerzálneho výpočtového prostredia, voľne dostupného všetkým (R Development Core Team, 2013).

Táto kniha nie je ani „kuchárskou knihou“ (pretože matematická štatistika nie je knihou receptov alebo návodov), nie je ani „teoretickou matematickou štatistikou“ (pretože teória je vysvetlená prostredníctvom príkladov pochádzajúcich z reality praxe biologickej antropológie) a nie je ani „učebnicou prostredia \mathbb{R} “ (pretože program \mathbb{R} je v knihe použitý len ako nástroj, prostredníctvom

ktorého sú riešené reálne situácie). Čím teda kniha je? **Kniha je unikátnou kombináciou sily teoretického základu matematickej štatistiky implementovaného v prostredí \mathbb{R} s cieľom pochopiť a riešiť praktické situácie z biologickej antropológie (ako aj biológie a medicíny).** Prečo práve \mathbb{R} ? Pretože \mathbb{R} naozaj stojí za to. Nie je to len „užívateľsky neprispôsobivý“ komerčný klikací softvér slúžiaci na riešenie vybraných problémov z aplikácií s obmedzeným množstvom použiteľných štatistických metód (pokiaľ sa nejaká metóda v takomto softvéri nenachádza, je nutné ju hľadať inde). Je to voľne šíriteľné komplexné prostredie slúžiace na štatistické výpočty, prostredie schopné prispôbiť sa požiadavkám užívateľa. Je to interaktívne prostredie s príťažlivou statickou, ako aj animačnou grafikou. Je to prostredie, v ktorom je možné riešiť nekonečné množstvo praktických, ako aj teoretických štatistických problémov. \mathbb{R} je „všetko v jednom“ a naozaj stojí za to!

Kniha je napísaná na základe dlhoročných skúseností prvého autora z prednášania bioštatistických predmetov na štyroch univerzitách (Univerzita Komenského, University of Vienna, University of Glasgow a Masarykova univerzita) pre študentov antropológie, medicíny, aplikovanej matematiky a štatistiky. Je jeho víziou komplexného prístupu k matematickej štatistike pozostávajúceho z dvoch trojpiliérových systémov

1. (a) matematicko-štatistická teória, (b) implementácia v \mathbb{R} a (c) aplikácia na reálne situácie a
2. (a) plánovanie bio-medicínskych štúdií, (b) databáza a dátový manažment a (c) interpretácie.

Je členená do ôsmich kapitol. Prvá kapitola Vedecké štúdie obsahuje informácie o plánovaní bio-medicínskych štúdií, ktoré by mal vedieť prakticky používať každý vedec pracujúci s reálnymi dátami. Zahŕňa témy ako typy vedeckých štúdií, teória a prax náhodného výberu, presnosť merania, znáhodnenie a zaslepenie, pravidlá tvorby bio-medicínskych databáz („Achilovej päty“ výskumu), plánovanie rozsahu náhodného výberu („piliéra“ plánovania) a pod. Druhá kapitola podáva prehľad filozofie pojmov „model rozdelenia pravdepodobnosti“ a „štatistický model“, z ktorých sú postavené „pevné základy“ výskumu na reálnych dátach. Nachádza sa v nej prehľad vybraných, najčastejšie sa vyskytujúcich modelov ako model binomického rozdelenia, Poissonovho rozdelenia, multinomického a súčinnového multinomického rozdelenia, negatívne-binomického rozdelenia, model normálneho rozdelenia a pod. Parametre týchto modelov sú podkladom „štatistického výkladového slovníka“, ktorý slúži na interpretácie výsledkov. Tretia kapitola Charakteristiky polohy a variability a štatistická grafika obsahuje metódy základnej číselnej a grafickej exploratórnej analýzy. Charakteristiky polohy a variability zjednodušene charakterizujú dáta. Štatistická grafika je „obrazom v ráme“, prostredníctvom ktorého sú dáta a výsledky štatistických analýz prezentované. Štvrtá kapitola Testovanie hypotéz komplexne zahŕňa najdôležitejší „pojmotvorný štatistický aparát“, bez znalosti ktorého nie je možné robiť žiadnu štatistickú analýzu, t.j. ide o „piliér“ štatistických analýz a ich interpretácií. Ďalšie štyri kapitoly sú uceleným prehľadom najčastejšie používaných štatistických metód. Ide o systematiky prehľad jednovýberových, dvojjvýberových a viacvýberových praktických situácií. Tieto kapitoly sú základom štatistickej inferencie, prostredníctvom ktorej sú zovšeobecnené a interpretované výsledky analýz. Teoretický úvod vo väčšine prípadov obsahuje (1) teoretické predpoklady, (2) matematickú definíciu hypotéz, (3) testovaciu štatistiku (Waldovu testovaciu štatistiku, testovaciu štatistiku pomerom vierohodnosti alebo skóre testovaciu štatistiku), (4) jej rozdelenie za platnosti nulovej a alternatívnej hypotézy, (5) definíciu kritického oboru a silofunkcie, (6) definíciu p-hodnoty, (7) definíciu empirických intervalov spoľahlivosti, (8) algoritmus výpočtu minimálneho rozsahu súboru a (9) odvodenie vybraných typov testovacích štatistík. Po teoretickom úvode nasledujú riešené didaktické príklady na simulovaných alebo reálnych dátach (vysvetľujúce vlastnosti testov alebo demonštrujúce ich použitie) nasledované **riešenými reálnymi príkladmi, ktorých riešenia sú štrukturované do odsekov špeciálne vyvinutých pre túto knihu** – (1) slovná a matematická formulácia hypotéz, (2) testovacia štatistika, (3) zamietacia oblasť, (4) empirický dvojstranný interval spoľahlivosti, (5) štatistický záver, (6) slovný záver a (7) antropológický slovný záver. **Informácie o dátach sú prezentované v štrukturovaných odsekoch tiež špeciálne vyvinutých pre túto knihu** – (1) hodnotený súbor, (2) súbor dát, (3) popis premenných, (4) biologické súvislosti a

(5) ciele. Zadania príkladov vhodné pre antropológov (biológov alebo lekárov) su podfarbené sivou farbou. Matematicky náročnejšie kapitoly sú označené hviezdíčkou. Definície, vety a dôležitá teória sú zvýraznené z ľavej strany hrubou vertikálnou čiarou. Riešenia príkladov v \mathbb{R} sú zvýraznené iným typom písma, jednotlivé riadky kódu sú očíslované a poznámky odlíšené sivou farbou. Kniha spolu obsahuje 244 príkladov (približne 90 % riešených), 1612 riadkov \mathbb{R} -kódu, 72 obrázkov, 38 tabuliek a 24 antropologických dátových súborov. Súčasťou knihy je aj tabuľkový prehľad použitých a naprogramovaných funkcií a ich definície a trojjazyčný slovník štatistických pojmov (slovenčina, čeština a angličtina). ??

Zodpovednosť jednotlivých autorov:

1. doc. PaedDr. RNDr. Stanislav Katina, PhD. – vytvoril koncepciu knihy, navrhol štrukturované riešenia príkladov a štrukturované informácie o dátach, vyriešil a naprogramoval príklady v \mathbb{R} a napísal text všetkých kapitol v L^AT_EX-u, vytvoril tabuľky funkcií s ich definíciami, vytvoril trojjazyčný slovník štatistických pojmov, register, zoznam grafov, tabuliek, matematických symbolov a skratiek;
2. doc. RNDr. Miroslav Králík, PhD. – napísal antropologické úvody ku kapitolám 6–8, antropologické slovné závery k príkladom, ako aj niektoré podkapitoly prvej kapitoly (Štúdie kostrových pozostatkov a Zisťovanie presnosti merania), pripravil dáta a štrukturované informácie o nich;
3. Mgr. Adela Hupková – pripravila dáta a štrukturované informácie o nich, pripravila obrázky k dátam, preložila text napísaný Miroslavom Králíkom do slovenčiny a napísala ho v L^AT_EX-u, spolupodieľala sa na vytvorení tabuliek funkcií s ich definíciami a tvorbe trojjazyčného slovníka štatistických pojmov.

Stanislav Katina by sa rád poďakoval:

- študentke Mgr. Ivete Selingerovej za korekcie kapitol 2 a 4;
- školiteľovi a neskôr kolegovi doc. RNDr. Františkovi Štulajterovi, CSc. za nespočetné hodiny strávené nad teoretickými detailami matematickej štatistiky;
- RNDr. Alene Šefčákovej, PhD., doc. RNDr. Milanovi Thurzovi, CSc., RNDr. Silvii Bodoríkovej, PhD., RNDr. Eve Neščákovej, PhD., RNDr. Petrovi Velemínskemu, PhD., RNDr. Jane Velemínskej, PhD., doc. RNDr. Jaroslavovi Brůžkovi, PhD. a doc. RNDr. Miroslavovi Králíkovi, PhD. za dlhoročnú spoluprácu pri riešení antropologických problémov a diskusie, ktoré sa stali základom širšieho pohľadu na realitu antropologických aplikácií a
- v neposlednom rade prof. RNDr. Ivanke Horovej, CSc. za vytvorenie príjemného prostredia na Útave matematicky a štatistiky MU pri písaní tejto knihy.

Prípravu a vydanie tejto publikácie podporil projekt *Formování mezinárodního týmu pro výzkum evoluční antropologie moravských populací* z Operačního programu Vzdělávání pro konkurenceschopnost.



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ