

VYJADŘOVÁNÍ VÝSLEDKŮ STATISTICKÉHO ZPRACOVÁNÍ ZDRAVOTNĚ SOCIÁLNÍHO VÝZKUMU

Pravoslav Stránský

Univerzita Karlova, Lékařská fakulta v Hradci Králové

Srovnáme-li různé původní práce publikované v časopisech zaměřených na řešení zdravotně sociálních otázek, zjistíme, že způsob popisu charakteristik sledovaného souboru i způsoby použité k zjišťování rozdílů či souvislostí zkoumaných veličin se značně liší, i když po formální stránce jsou struktura i velikost publikovaných výsledků velmi podobné.

Problémy se týkají v podstatě tří oblastí. První se týká počtu zobrazovaných platných číslic ve výsledcích vypočítaných charakteristik. Do té doby, než se objevily první kalkulačky a po nich počítače, nikdo nepočítal s větší přesností, než byla nezbytně nutná a počet platných číslic byl co nejmenší. Používání kalkulaček svádělo k opsání všech číslic, které byly na displeji. Důvodem byl falešný dojem, že větší počet číslic znamená větší přesnost výsledku. V případě výpočtů prováděných ve velmi rozšířeném programu Excel závisí počet zobrazených číslic na typu dat (standardně je formát buňky Obecný), druhu písma, jeho řezu a velikosti. Při běžném nastavení konfigurace parametrů uvedeného programu se v buňce zobrazí 7 platných číslic, ve skutečnosti je jich uloženo 15.

Hlavní zásadou je, že výsledek nemá obsahovat více platných číslic, než mají vstupní data. Měříme-li výšku v centimetrech, hmotnost v kilogramech, věk v rocích, neuvádíme ve výsledcích desetiny nebo dokonce setiny daných veličin. Počítáme-li z hmotnosti a výšky BMI, uvedeme ho na tři platné cifry, tj. jedno desetinné místo, jak to odpovídá změřené výšce.

V této souvislosti poznámka k platným číslicím. Definice je jednoduchá. První platnou číslicí je první zprava nenulová číslice, ale jak zkušenost ukazuje, její interpretace tak

jednoduchá není, protože se často zaměňuje počet platných číslic s počtem číslic za desetinnou tečkou. Pro ilustraci: číslo 0,00123 je uvedeno na tři platné číslice, stejně jako čísla 123, 12,3 a 1,23.

Takovýto počet uváděných platných číslic souvisí s rozlišovací schopností (rozlišením) použitého přístroje či metody. Velmi často se také používá výraz přesnost měření, což může být zavádějící, protože slovo přesnost se ve spojení s měřením a statistikou používá ještě ve dvou jiných významech. Jeden význam je spojen s variabilitou dat a čím menší je při opakovaných měřeních variabilita, tím je měření přesnější. Bude-li mít měřená veličina Gaussovo rozdělení, potom je mírou přesnosti velikost směrodatné odchylky.

O přesnosti vyjádření výsledku také mluvíme při počítání s přibližnými čísly $\pi = 3,141592$ je přesnější vyjádření Ludolfova čísla než hodnota 3,14. Budeme-li však počítat obsah kruhu, jehož poloměr jsme změřili jako 98 cm (tedy na dvě platná místa), pak výsledek vždy vyjádříme jako $3,0 \text{ m}^2$, i když vypočtená čísla budou hodnoty 3,0171849568, resp. 3,015656.

Zvláštní pozornost počtu platných číslic musíme věnovat vyjadřování výsledků, které jsou poměrem různých čísel, jako jsou například relativní četnosti výskytu pozorovaných jevů, událostí. Počet platných číslic implikuje velikost souboru, ze kterého je poměr počítán. Jestliže test psalo 13 studentů a 8 z nich uspělo, můžeme to vyjádřit tak, že úspěšných bylo 60 %, nikoliv 62 % nebo dokonce 61,5 % (62 % by bylo správné v případě, že by uspělo 80 studentů ze 130, tři platné číslice by byly v pořádku, kdyby test psalo 1 300 studentů a z nich by bylo 800 úspěšných). Přijímáme

proto praktickou zásadu, že při rozsahu souboru v řádu desítek jedinců vyjadřujeme relativní četnost zaokrouhleně na jednu platnou cifru (desítky procent), má-li soubor řádově stovky prvků, výsledek zaokrouhlíme na dvě platné číslice (jednotky procent) atd. Jestliže se v ČR v r. 2005 narodilo 102 211 dětí a z nich bylo 52 453 chlapců, je správné vyjádření odhadu pravděpodobnosti narození chlapce 0,51318 (pět platných číslic implikuje, že číslo bylo získáno ze souboru, ve kterém bylo řádově 100 000 jedinců).

Zvláštní pozornost si zaslouží prezentace výsledků ordinálních dat. Sem patří všechna data získaná z odpovědí na nejrůznější dotazníky, např. na spokojenost s poskytovanou péčí, s kvalitou života, ale i dosažené počty bodů v testech zkoumajících znalost určité látky. Používá-li dotazník např. pětistupňové škály, tak rozdíl mezi druhým a třetím stupněm není stejně velký jako mezi třetím a čtvrtým stupněm, stejně jako rozdíl znalostí při dosažení 60 bodů ve srovnání s 50 body je jiný než mezi 70 a 60 body.

Zásadním důsledkem, který z této skutečnosti vyplývá, je to, že jediná oprávněná operace, kterou lze s ordinálními daty provádět, je třídění. Jako charakteristiku polohy použijeme medián, eventuálně modus a jako charakteristiky variability variační, interdecilové nebo inter-kvartilové rozpětí. Aritmetický průměr a směrodatnou odchylku sice můžeme ze zjištěných čísel spočítat, ale nemají žádný reálný význam a jejich používání svědčí o tom, že autor nechápe podstatu interpretace informace získané z ordinálních dat. Jediným správným přístupem pro zjišťování rozdílů nebo asociací mezi sledovanými ordinálními daty je použití metod neparametrické statistiky. Tu je ostatně nutné použít i u kvantitativních dat, pokud nesplňují některé předpoklady metod parametrických, zejména když musíme zamítnout nulovou hypotézu o jejich normalitě.

** tento příspěvek již vyšel v časopise Kontakt, vol. XI, no. 2, 2009, ISSN 1212-4117, pro svou důležitost je se svolením autora prezentován i zde*

Pravoslav Stránský
str26@seznam.cz