

Jak sbírat výzkumná data

Metodiku sběru dat v didaktice by měla pokrýt samostatná příručka, zde jsou uvedeny pouze techničtější otázky přímo související s otázkami jejich dalšího zpracování v programu R. Data zásadně zapisujeme do tabulkového procesoru, nikdy ne do editorů typu Word či jiný.

Ačkoli jsme často zvyklí zapisovat ve tvaru ukázaném v tabulce 3, je taková vhodná až pro výsledné prezentování dat – ve fázi sběru dat a pro potřeby uchovávání primárních dat je ve skutečnosti velmi nežádoucí.

třída	výsledky pretest	výsledky posttest
8.A	25.6%	60%
7.B	24%	52%

Tabulka 3: Nemáme možnost komentovat jednotlivé hodnoty/měření.

Je dobrým zvykem zapisovat data tak, že sloupce představují určité kategorie, zjištění a každý řádek pak jedno samostatné měření/pozorování jako můžeme vidět v tabulce 4.

#	třída	typ	score	datum	učitel	škola
1	8.A	pretest	25.6%	13-10-2015	Konečná	ZŠ Vidlička
2	7.B	pretest	24%	13-10-2015	Vopěnka	ZŠ Vidlička
3	8.A	posttest	60%	30-10-2015	Konečná	ZŠ Vidlička
4	7.B	posttest	52%	30-10-2015	Konečná	ZŠ Vidlička

Tabulka 4: Správná podoba tabulky primárních dat.

Stává se, že potřebujete určité měření komentovat – například bylo málo dětí a měli byste pozorovat jindy, učil suplent a výsledky tím mohou být ovlivněné atp. U tabulky 3 použijete komentář k buňce, v horším případě si buňku barevně obarvíte a

můžete si být jisti, že v okamžiku zpracovávání dat budete mít x komentářů které nelze nijak rozumně zpracovávat (nejsou filtrovatelné), v pestrobarevné tabulce kde význam jednotlivých barviček (rovněž nefiltrovatelných) můžete jen odhadovat.

U doporučené podoby tabulky snadno přidáme další sloupec „komentář“ nebo „vyloučit ze zpracování“ a třeba pomocí automatického filtru tyto řádky odfiltrujeme. Všimněte si prvního sloupce – „#“. Protože Calc čísluje jednotlivé řádky, tak nás většinou nenapadne přidat si sloupec s vlastním číslováním, což se může ukázat jako velmi užitečné.

[Kontingenci tabulky vysvětlit]

[zde by se asi hodily vývojové diagramy pro postup jak to udělat]

Souhrn: při zápisu primárních dat používejte vždy strukturu naznačenou v tabulce 4, zásadně nepoužívejte funkce komentářů a barevného značení buněk, vždy si doplňkové informace uložte do dalšího sloupce(ů). Dávejte si pozor na funkci „slučování“ buněk, její užití v primárních datech signalizuje chybný návrh struktury tabulky.

Před statistickým zpracováním data neřadíme, resp. pořadí záznamů nemá vliv na použité metody či analýzy. U časových řad je nutné uvádět „timestamp“, prosté spolehnutí se na pořadí řádků je nebezpečné. Z každého řádku dat musí být patrné „kdo, kdy, kde a co“ zjistil. Pravidlo *kdo-kdy-kde-co* je výbornou pomůckou pro kontrolu správného návrhu tabulky pro ukládání primárních dat.

Vypsali jste všechny bezobratlé z učebnice uvedeného vydavatele a oponent vám tvrdí, že žabronoška v této učebnici navzdory vašemu výzkumu není vůbec uvedená – budete procházet znovu celou učebnici – ne, protože si můžete rovnat záznamy podle pořadí zápisu, tak snadno zjistíte mezi kterými zvířaty se zmínka o žabronošce nachází a stačí vám počítat několik stránek.

PŘESNOST A SPOLEHLIVOST Hodnota 2cm není totéž co 2,0cm. V první případě říkáme že jsme měřili na dětském pravítku se stupnicí jen v celých centimetrech, a měřený objekt je v intervalu 1,5-2,5 cm, v druhém případě jsme měli běžné pravítko rozlišené i na milimetry, a měřený objekt byl v intervalu 1,9-2,1cm. Druhý údaj je tedy výrazně přesnější. Spolehlivost je pak vyjádření nkolik naše měření odpovídá skutečnosti, jak pečlivě jsme odečetli údaje na pravítku. [lépe]

Kolik dat potřebujeme nasbírat?

To je poměrně obtížná teoretická otázka a existuje několik výpočtů pro stanovení potřebné velikosti vzorku. Upřímně řečeno je ale málokdo skutečně používá ¹⁰, pro potřebu diplomové práce plně vyhovuje koncept „plovoucího průměru“.

```
plot(cumsum(trida8A.vek) / seq(along=trida8A.vek),
     type='l', xlab='Číslo měření', ylab='Plovoucí průměr')
abline(h=mean(trida8A.vek), lty=2, col='blue')
```

[doplnit graf výstupu]

Modrá vodorovná linka ukazuje aritmetický průměr výsledný, klikatá černá čára jeho kolísání v průběhu zvyšujícího se počtu měření. Uvedený příklad je poněkud nesmyslný, věk dětí ve třídě pochopitelně zjistíte u všech dětí, tudíž víte kdy přestat sbírat data. Správnější by bylo sledovat plovoucí průměr všech osmáků do vašeho výzkumu zahrnutých, ale především věk žactva osmých ročníků asi není hlavním výzkumným cílem a plovoucí průměr budeme sledovat u proměnné, kterou se snažíme vysvětlit - tedy třeba úspěšnost v didaktickém testu dotčených osmáků. Ustálení průměru této úspěšnosti je dobrým signálem přijatelné velikosti testovacího vzorku.

¹⁰ Myslím tím, že respektuje jejich výsledek, nikoli že je spočítá, a uvede nyní již statisticky podpořenou informaci o nedostatečné velikosti vzorku. Takové sdělení je úsměvné – jediným správným postupem je zvětšit datový soubor.

Je možné užít plovoucí průměr i u neparametricky rozložených dat? Ano – plní zde stejnou funkci. Aritmetickému průměru se u neparametrických rozložení vyhýbáme abychom zamezili jejich dezinterpretaci, znamená to, že samotný průměr je nějak chybný.