

16 Statistické programové systémy

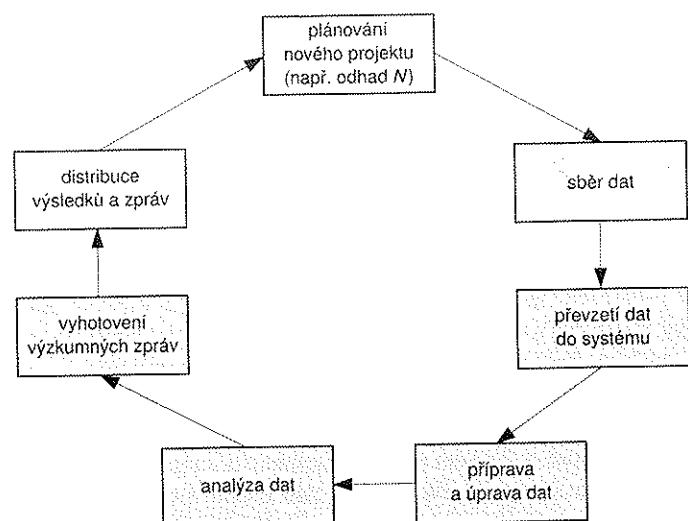
Tvrdí se, že pokud statistická procedura není implementována ve vhodném statistickém programu, jako by neexistovala. V této krátké kapitole upozorníme na některé komerční i volně přístupné statistické systémy, které usnadňují statistickou analýzu dat. Obsáhlé navigační odkazy k materiálům o problematice statistiky i počítacových programů obsahují webové stránky České statistické společnosti (www.statspol.cz). Výskyt zdrojů materiálů o statistice mapuje a komentuje vydatně v českém jazyce internetová stránka doc. Jiřího Žváčka (www.stahroun.me.cz/ssysel). Nalezneme zde kromě jiného odkazy na výpočetní procedury implementované přímo ve webových stránkách internetu.

Některé statistické výpočty umožňují provádět již běžné tabulkové kalkulaátory (např. MS Excel). Statistické programové systémy jsou komplexním nástrojem pro analýzu dat. Ve srovnání s tabulkovými procesory lze uvést jejich následující výhody:

- umožňují zpracovávat velké soubory dat;
- zahrnují větší rozsah neparametrických a parametrických statistických metod;
- obsahují velkou paletu obecných i speciálních grafických prostředků a diagnostiku statistických předpokladů;
- nabízejí ze statistického hlediska rozsáhlejší pomoc uživateli (systémy radí při výběru metod, interpretují výsledky, jsou dodávány vzorové soubory statistických dat);
- některé umožňují výbornou manipulaci s výstupy;
- jsou algoritmicky spolehlivé.

Při výběru statistického systému si všimáme kromě solidnosti firmy, ceny, dobré dokumentace a uživatelsky přívětivého rozhraní pro práci se systémem samozřejmě kvality jednotlivých komponent systému, které se postupně uplatní při práci na výzkumném projektu. Jsou znázorněny v pořadí jejich praktického využití na obrázku 16.1.

Obr. 16.1 Komponenty statistického systému v procesu realizace výzkumného projektu



Komponenta pro **plánování** se využívá při určování potřebného rozsahu výběru a pro specifikaci některých vlastností experimentálních plánů. Taková komponenta např. může obsahovat expertní systém, jenž na základě požadavek doporučí provedení celého projektu.

Komponenta pro **sběr dat** usnadňuje uživateli provádění sběru dat pomocí telefonu nebo internetu.

Komponenta **převzetí dat** do systému umožňuje import dat z jiných programů a databází. Speciálně se to týká dat zaznamenaných v rozšířených programech, jako jsou MS Excel, Access, SQL databáze, Oracle, Informix nebo import z jiných statistických programů.

Část statistického systému pro **přípravu a úpravu dat** slouží pro provedení kontroly a „vyčištění“ dat nebo nutných transformací dat do podoby vhodné k zpracování.

V centru pozornosti je **komponenta pro analýzu dat** s implementovanými statistickými procedurami, jejichž počet jde u větších systémů do stovek. Tato v této komponentě může být zabudován expertní systém pro volbu vhodné statistické metody nebo pro interpretaci výsledků.

Dokonalejší statistické systémy mají propracovanou i komponentu „**portování**“, která z výsledků připraví zprávu v několika různých formátech, kterou uživatel může upravovat podle svých požadavků pro různé verze výzkumných zpráv.

Komponenta pro **distribuci výsledků a zpráv** zajišťuje snadné vydávání zprávy o výzkumném projektu na stránkách internetu apod.

Mezi nejznámější statistické systémy patří STATISTICA (www.statsoft.cz/), SPSS (Statistical programs for Social Sciences, www.spss.cz) a SAS (Statistical Analysis System, www.sas.com). Jedná se o rozsáhlé programové systémy, v nichž nezůstane skoro žádné přání uživatele nesplněno – zvláště pokud dleponuje dostatečnými finančními prostředky. Systém STATISTICA má velkou výhodu, protože je přeložen úplně do češtiny.

Mezi menší systémy patří Statgraphics (www.statgraphics.com) a NCSS (Number Cruncher Statistical System – www.ncss.com). Jedná se o uživatelsky velmi příjemné systémy s dobrými grafickými prostředky. ADSTAT je český produkt firmy TriloByte (www.trilobyte.cz) pro statistickou analýzu dat z oblasti věd přírodních a technických.

• • •

Existuje i řada volně přístupných systémů. Výrobce systému NCSS zpřístupnil hezkou freewareovou verzi NCSS Junior, jejíž možnosti obsahnou základní kurzy statistiky. K němu jako doplněk můžeme používat programový systém EpiInfo (www.cdc.gov/epiinfo) od americké státní zdravotnické organizace CDC (Centers for Disease Control). Jedna z verzí tohoto programu byla přeložena do češtiny. Tento systém se vyznačuje dobře zpracovanou částí pro vstup dat v podobě dotazníku a pro jejich kontrolu.

Zajímavý pro výzkumníky z oblasti psychologie a pedagogických věd je systém OpenStat, jehož autorem je Bill Miller, penzionovaný profesor z Iowa State University. Ten systém OpenStat naprogramoval ve svém volném čase postupně v několika programovacích jazycích (Visual Basic, C++, Turbo Pascal, Delphi Borland) a pro různé operační systémy (Linux a Windows). Užitečná je verze OpenStat3, která pracuje pod operačním systémem Windows 95 a vyšším, naprogramovaná pomocí vývojového systému Delphi Borland verze 5. Pro úpravu programu – jeho počeštění nebo doplnění o další procedury – lze využít volnou verzi vývojového prostředí pro Delphi Borland v. 7. K nalezení aktuálního zdroje lze použít vyhledávač www.google.com.

WinIDAMS (www.unesco.org/idams) je volně dostupný komplexní programový systém pro analýzu dat s mnoha vícerozměrnými procedurami včetně podprogramu pro vytváření regresních a klasifikačních stromů.

Tab. 16.1 Některé programy počítající analýzu síly testu a odhad velikostí výběru

Software	Internet URL	Typ	OS
GPOWER	www.psychologie.uni.trier.de:8000/projects/gpower.html	Freeware	DOS
Power Plant	ftp.per.its.csiro.au/csiro-wa/biometrics/	Freeware	WIN
PS	www.mc.vanderbilt.edu/prevmed/ps/index.htm	Freeware	DOS

Také existuje mnoho komerčních a nekomerčních systémů, jež provádějí různé typy analýzy statistické síly a výpočtu potřebných rozsahů výběru. Tyto systémy se liší rozsahem posuzovaných testů, svojí přesností, jednoduchostí použití a snadnosti, s níž se jim lze naučit. Ve výše jmenovaných systémech jsou moduly, které se touto problematikou zabývají. Mezi volně dostupné systémy také patří GPOWER a PS.

PS je interaktivní program, jenž provádí analýzu síly testu a odhad velikostí výběru. Je komponován pro operační systém Windows. Může se využít pro studie hodnotící dichotomické a spojité míry nebo pro data o přežívání. Studie o dichotomických nebo spojitéch mírách mohou pracovat s nezávislými nebo závislými skupinami. PS program zhodovuje grafy ke zkoumání vztahu mezi silou testu, velikostí výběru a velikostí detekované odchyly od nulové hypotézy. GPOWER má širší záběr než PS co do výběru hodnocených testů, ale je napsán pouze pro operační systém PC DOS. V tabulce 16.1 udáváme pro přehled názvy a adresu přístupu k vybraným systémům z této kategorie programů.

Pro realizaci přístupů z oblasti víceúrovňového modelování lze zakoupit komerční programy jako MIMAS, MLn nebo MIwiN. Ale i s jejich demoverzemi si zájemce dobře vyzkouší první kroky při provádění víceúrovňové analýzy dat. Informace o programech MIMAS a MLn nalezneme na webové stránce multi-level.ioe.ac.uk. Existuje dostatek volně přístupných programů pro realizaci procedur metaanalýzy. Cochranova společnost (Cochran collaboration) usnadňuje zájemcům spolupráci přístupem k databázovým a metaanalytickým prostředkům pro týče medicíny založené na faktech včetně kompletní dokumentace (www.update-software.com/metaview4.htm). Volně dostupný je také program Ralfa Schwarzena META. Makra pro systém SPSS od autora Marka Lipseyho nalezneme na adrese www.wam.umd.edu/~wilsondb/home.html. Na této adrese se také nalézají verze maker pro systémy SAS a STATA a potřebné vzorce pro výpočty v Excelu.

Komu nebudou stačit vyjmenované volně dostupné programy, stáhne si z internetu systém SIMFIT autora W. G. Bardsleye a seznámí se s jeho nejprimitivnějšími možnostmi z oblasti simulací, prokládání dat, statistické analýzy a

fického zpracování nebo začne experimentovat se statistickým programovacím jazykem R (www.r-project.org). Oba tyto systémy jsou spíše určeny pro uživatele s hlubšími matematickými znalostmi.

Programovací prostředí jazyka R zahrnuje:

- efektivní přípravu dat a jejich ukládání,
- soubor operátorů pro práci s vektory a maticemi,
- velký integrovaný soubor prostředků pro základní analýzu dat,
- grafické prostředky pro analýzu a zobrazení dat a
- propracovaný jednoduchý jazyk, který obsahuje podmínky, smyčky, uživatelsky definované rekursivní funkce a prostředky pro vstup a výstup.

Možnost sestavit ze statistických funkcí vhodný prostředek analýzy dat pro řešenou úlohu existuje i v tabulkových procesorech (Excel, TAB602). Rejstřík dostupných statistických funkcí je však mnohem omezenější než u jazyka R. Naproti tomu prostředí tabulkového procesoru zná skoro každý uživatel počítače, takže si pravděpodobnější nalezne rychleji cestu, jak výpočetní možnosti tabulkového procesoru využít i pro statistické úlohy. Kromě několika desítek statistických funkcí (v kapitole 4 jsme poznali výpočet distribučních funkcí) se v Excelu nacházejí i tzv. nástroje, které integrují několik statistických funkcí do jedné procedury. Jedná se například o nástroje pro popisnou jednorozměrnou analýzu dat, mnohonásobnou regresní analýzu, analýzu rozptylu a kovariance. Doplňky pro statistickou analýzu v Excelu se připravují také na komerční bázi. Jedná se například o systémy ANALYSE-IT (www.analyse-it.com) nebo XL STATS (www.man.deakin.edu.au/rodneyc/XLstats.htm).

Doporučujeme, aby se čtenář seznámil se statistickými prostředky a způsobem provádění výpočtů v tabulkovém procesoru a pak provedl srovnání s nějakým statistickým systémem.