

interaktivní statistická analýza dat

A. Zaydler*, J. Tvrdek**, E. Ettlerová *

*SVT UK při LF UK v Hradci Králové

**RHS Ostrava

Myšlenka vypracovat software pro statistickou analýzu dat vznikla na semináři ROBUST 84 ve Sřavonicích.

Původní idea udělat rozsáhlý statistický systém, především pro statistiky prošla složitým vývojem, který byl ovlivněn především kapacitně. Velký vliv na zaměření a charakter systému měl také nástup osobních počítačů a skutečnost, že dnes většina profesionálních statistiků má přístup k rozsáhlému statistickému software implementovanému na osobních počítačích. Problematika statistického zpracování na počítačích řady SMEP tím však vyřešena není.

Proto jsou základní ideje systému INSTA tyto:

Systém je určen k použití především pro uživatele nestatistiky případně pro použití při výuce. Je řešen tak, aby byl zajištěn v rámci technických možností počítačů SMEP co největší uživatelský komfort, mj. je využíván i systém SVS /SMEP video systém/, který umožňuje použít při komunikaci přímou adresaci kurzoru na obrazovce s možností bližšího vysvětlení požadované odpovědi v HELPU. Systém INSTA je implementován na počítačích řady SMEP pod operačním systémem DOS RV /t.j. počítače SM4/20, SM520t, SM5200.M1, SM50/50.M1, SM3420, M16-20/ a je naprogramován v Fortranu 77.

Základní charakteristiky systému INSTA jsou tyto:

- systém je schopen pracovat jak s kvalitativními tak s kvantitativními daty
- komunikace je dělána tak, aby počet dotazů byl co nejmenší. Tam, kde je to možné, jsou nasazovány počáteční /default/ hodnoty.
- Kvalitativní data je možné zadávat pro analýzu přímo ve formě kontingenční tabulky nebo tuto tabulku vypočítat z datové matice kvantitativních dat
- Kvantitativní data jsou ve tvaru datové matice, jejíž řádky tvoří jednotlivé objekty a sloupce jednotlivé měřené veličiny. Tato data je možno zadávat pro analýzu přímo nebo načítat několika způsoby z předem připravených souborů
- systém umí zacházet s neúplnými objekty t.j. s objekty, které nemají naměřeny všechny veličiny /tzv. missingy/
- načtená data jsou uložena spolu s dalšími informacemi do systémového souboru základním programem, který zároveň umožňuje tato data editovat, transformovat a ukládat informace o rozdělení datové matice na subpopulace /tzv. podmínka/

- systém má dvě úrovně odfiltrování odlehých hodnot /hodnot pod nebo nad zad^{anou} mezí/. Tyto hodnoty je možné buď považovat za nenaměřené pro účely konkrétního výpočtu bez jejich změny v datech nebo je možné je změnit na nenaměřené v datech
- systém má bohaté možnosti editace a transformace dat
- výpočty jsou prováděny z dat v systémovém souboru
- datovou matici lze dělit do subpopulací /např. pro účely t-testu a pod./tzv. podmínkami
- podmínky umožňují rozdělit objekty podle hodnot nějaké proměnné a to buď podle konkrétních hodnot /kodů/ nebo podle intervalů /mezí/. Takto vzniklé primární rozdělení je možné ještě dále modifikovat, takže je například možné provádět analýzu pro všechny objekty, které nemají v dané veličině naměřenou hodnotu a pod.
- systém umožňuje výpočet základních charakteristik pro celý soubor i pro subpopulace, t-test, jednofaktorovou analýzu rozptylu, korelace a lineární regresi ~~regresí~~ a analýzu kontingenčních tabulek
- systém bude dále rozšiřován jak o další výpočetní tak organizační možnosti.

První verze systému bude k dispozici v průběhu září t.r. a bude distribuována SVT UK při LF UK v Hradci Králové. Ažtoři, uvítají spolupráci, jsou schopni dohodnout se s případnými zájemci a poskytnout jim veškeré informace k tomu, aby mohli k systému přidat další programy.

Adresa pro korespondenci: RNDr. Karel Zaydlar
 SVT UK při LF UK
 Šimkova 870
 500 38 Hradec Králové