

Informační Bulletin



České Statistické Společnosti

č. 3, prosinec 1997, ročník 8.

Vybraná kritéria pro volbu programového vybavení pro (statistickou) analýzu dat

Jaromír Antoch

Abstrakt: Porozhlédneme-li se po nabídce programového vybavení pro (statistickou) analýzu dat, s překvapením zjistíme, že obsahuje několik set položek. Je to mnohem více než běžný uživatel statistiky předpokládá. Na druhé straně je však třeba poznamenat, že ve větším měřítku se z nich používá pouze velmi malá část. Cílem tohoto krátkého příspěvku je shrnout alespoň základní kritéria, jež by měl uživatel při výběru programového vybavení tohoto typu vzít v úvahu.

Úvod

Je dobře známo, že v uplynulých několika letech se u nás k lepšímu značně změnila situace na trhu, a to nejen programového vybavení. Zbylo v podstatě jediné omezení – peníze, neboť odpadla většina administrativních restrikcí. *Jak a co si však vybrat, aniž bychom své volby brzo litovali?* Ve většině případů hledáme řešení na dlouhou dobu a musíme mít též na paměti, že se jedná o produkty finančně značně nákladné. Zamýšlela se, a každodenně se nad tím zamýšlí, řada lidí. V následujících řádcích proto uvádím alespoň některá vybraná kritéria, jež by potenciální uživatel měl vzít v úvahu. Vycházejí nejenom z osobních zkušeností, ale i ze závěrů, které na základě dlouholetých zkušeností a připomínek svých členů vypracovala pracovní skupina pro hodnocení statistických

systemů německé společnosti pro dokumentaci, informatiku a statistiku. Jejich zodpovězení si, spolu s rekapitulací problémů jež skutečně musí (a bude musit) uživatel každodenně řešit, by mu mohly při výběru pomoci. Je přitom zřejmé, že pro různé uživatele mají jednotlivá kritéria odlišnou důležitost, což je základní důvod, proč nelze ten který programový produkt prohlásit za absolutně nejlepší.

Jak čtenář snadno nahlédne, většina z uvedených požadavků je naprosto samozřejmá. Asi proto se jimi mnohdy neřídíme. Jak se Vás bude snažit přesvědčit téměř každý distributor toho či onoho produktu, je to právě ten jeho systém, který se pro vaše potřeby hodí nejlépe a daným požadavkům nejvíce vyhovuje. Podobně zpravidla dopadnete, obrátíte-li se o radu na tzv. „skalní uživatele“ jednotlivých systémů. Nejlepší proto asi je si vypůjčit ostrou zkušební verzi, *nikoliv demo!*, kterou Vám u většiny produktů na třicet dní zapůjčí každá slušná firma, a trochu si „pohrát“. Poté si sednout někde tiše v koutku, v klidu popřemýšlet a s řadou všetečných dotazů a typickými příklady ze své praxe obejít nejen prodávající, ale i ty uživatele, kteří již na ten který produkt vsadili. No a pak už zbývá to nejjednodušší, nebo nejtěžší(?), rozhodnout se.

1. Všeobecné uživatelské hledisko

- 1.1. Způsob ovládání, tj. dávkové, interaktivní či obojí, řízení příkazy či pomocí menu apod., možnost programování či alespoň tvorby vlastních makropříkazů, kvalita nápovědy, možnost napojení na uživatelovy oblíbené programy, editor, „přátelskost“ chování systému, způsob ošetření chyb, podrobnost indikace chyb, možnosti restartu (zvláště po chybě).
- 1.2. Možnost zápisu souborů potřebných pro monitorování vlastní činnosti, tj. tzv. *skript* a *echo* souborů.
- 1.3. Technická podpora ze strany výrobce a distributora, možnost komunikace s nimi přes Internet.
- 1.4. Existence kurzů pro uživatele, ať již úvodních, specializovaných či na základě objednávky uživatele.
- 1.5. Kvalita manuálů a další dokumentace, způsob presentace a popis použitých statistických algoritmů. Existence manuálů nejenom v elektronické, ale i v tištěné formě.
- 1.6. Podpora jiných jazyků než pouze angličtiny, minimálně pro názvy proměnných je velmi výhodná čeština.
- 1.7. Existence národních uživatelských skupin.

2. Operace s daty a spolupráce s databázemi

- 2.1. Datové modely s nimiž systém umožňuje pracovat.
- 2.2. Popis dat, tj. datové typy, datové struktury, sémantické informace, omezení z hlediska integrity dat, datové slovníky.
- 2.3. Manipulace s daty, dotazovací jazyky, relační operace, možnosti vkládání, opravování a vylučování dat, třídění.
- 2.4. Možnosti pro vstup a výstup dat, formulářově stavěný vstup dat, generátor dotazníků.
- 2.5. Vstup a výstup dat získaných dotazy v databázích, přímý přístup k datovým strukturám jiných programů.
- 2.6. Vstup a výstup dat přes *ODE*, například pro vstup dat a výstup výsledků z/do tabulkového procesoru.

3. Zahrnuté statistické metody

- 3.1. Rozsah a možnosti implementovaných metod.
- 3.2. Správnost, tj. kontrola správné implementace a funkčnosti jednotlivých metod.
- 3.3. Výsledky, způsob jejich prezentace, vysvětlení, popis a jednotnost výpisů.
- 3.4. Kvalita procedur, přesnost algoritmů, jejich validace a existence odkazů na validaci.
- 3.5. Možnosti ovládání, tj. existence standardizovaného ovládacího jazyka a jednotných parametrů pro jeho užívání, možnost jednoduché redefinice předem nastavených (default) hodnot.

4. Grafické nástroje pro statistickou analýzu dat

- 4.1. Rozsah a možnosti implementovaných metod.
- 4.2. Interaktivní grafika.
- 4.3. Výsledky, jejich vysvětlení, popis a jednotnost prezentace.

5. Prezentace výsledků

- 5.1. Numerické výsledky – respektování sémantických informací, možnost manipulace a editování numerických výsledků.
- 5.2. Grafické výsledky – rozlišitelnost, výstupní formáty, možnosti manipulace a editování grafických výstupů, export do různých grafických formátů; export do PostScriptu by měl být samozřejmostí.
- 5.3. Integrovaný výstup textu a grafiky.
- 5.4. Možnost přímé přípravy výstupních zpráv v rámci systému. Možnost napojení na tzv. *desk top publishing*.

6. Technické specifikace

- 6.1. Technické prostředí, tj. vázanost a nároky na hardware, operační systémy či speciální grafické prostředky.
- 6.2. Přenosnost a kompatibilita (minimálně dat a makropříkazů) jak mezi jednotlivými hardwarovými implementacemi, tak mezi jednotlivými verzemi.
- 6.3. Rozšiřitelnost, tj. možnost vývoje vlastních aplikací a programování. Otevřenost architektury, přístup k vnějším knihovnám, možnost volání z jiných programů.
- 6.4. Náklady na instalaci, trénink, udržování a vývoj nových aplikací.
- 6.5. Konfigurovatelnost systému.
- 6.6. Komunikace, možnost nasazení v sítích, přenosy dat a souborů, přístup přes síťový software k distribuovaným databázím, sdílení souborů (nejen datových), sdílení kódů.

7. Existence metod jež potřebujete právě vy ¹ (a to již dlouho)

- 7.1. Robustní metody, Bayesovské metody, plánování experimentů, méně běžná pravděpodobnostní rozdělení apod.
- 7.2. Podpora metod optimálního rozhodování a umělé inteligence.
- 7.3. Automatické ověřování, zda data vyhovují alespoň základním předpokladům kladeným na užité statistické metody.
- 7.4. Mnoho a mnoho dalších metod, jakož i mnoho a mnoho dalších všetečných otázek.

Pár slov na závěr

Jak podotkl kolega Tvrdík na semináři *Analýza dat'97* v Bohdanči, pro různé uživatele mají kritéria zmíněná výše rozdílnou váhu, což je ostatně důvod proč neexistuje jeden univerzální a optimální statistický systém. Nicméně, lidé by neměli zapomínat na to, že jejich zájmy a potřeby se vyvíjejí a mnohdy mění tak rychle, že alespoň jistá otevřenost systému by měla být tím prvořadým požadavkem. Na druhé straně si přiznejme, že nabízené programové produkty se v posledních letech mění mnohem více navenek než uvnitř, tj. spíše na straně Windows všeho druhu a ne na straně pro mnoho z nás mnohem potřebnější, tj. nově implementovaných statistických postupů a algoritmů.

¹Zde si autor vymýšlí v rámci vlastních potřeb

STATA a NCSS z pohledu uživatele

Josef Tvrđík

Příspěvek podobného názvu byl přednesen na konferenci Analýza dat '97 pořádané firmou TriloByte v Lázních Bohdaneč a části textu se shodují s textem, který bude uveřejněn ve sborníku této konference. Článek pro Bulletin ČStS byl napsán na výzvu Jaromíra Antocha ve snaze vyhovět jeho záměru připravit číslo věnované statistickému software. Článek přináší porovnání dvou statistických paketů, které jsou zařazovány do kategorie obecných, tedy nespécializovaných statistických programových systémů. Porovnání vychází z přibližně roční zkušenosti užívání obou paketů v různých úlohách analýzy dat, v případě NCSS je podepřeno i zkušeností z výuky. Tyto zkušenosti snad mohou pomoci v rozhodování o výběru statistického software. Příspěvek je subjektivním pohledem autora, nikoli komerční prezentací porovnávaných paketů.

STATA for Windows 5.0

STATA je mezi obecnými statistickými pakety jedním z nejmladších. Mezi uživatele začal více pronikat až v 90. letech a prosazoval se zejména vysokou hodnotou kritéria [(rozsah funkce)/cena]. Nyní se zdá, že hodnota tohoto kritéria přiblížila ostatním softwarovým produktům.

Celková charakteristika: Universální statistický paket. Dobrá podpora zejména v oblasti data managementu, průzkumových metod, statistického modelování a práce s nominálními veličinami. Otevřený systém s vlastní programovacím jazykem vysoké úrovně. Řada procedur balíku je napsána tímto jazykem. Výhodný může být pro zpracování rozsáhlých, např. epidemiologických dat, kdy je sledováno mnoho jak diskrétních, tak spojitých veličin a součástí úlohy je náročnější předzpracování dat (transformace veličin, sdružování kategorií, spojování dat z více souborů atd.). Hlavní způsob ovládání je pomocí příkazů. Komunikace s programem probíhá způsobem „dostanete jen to, co si explicitně vyžádáte“.

Struktura pracovní obrazovky STATA je na obr.1. Příkazy se zadávají v okně COMMAND. Prostředí Windows usnadňuje vkládání příkazů výběrem z dříve zadaných příkazů v okně REVIEW a ze seznamu proměnných v okně VARIABLES. Některé příkazy (otevření souboru s daty, otevření souboru LOG pro ukládání výsledků, ukládání grafů a pod.) lze zadávat výběrem z menu. Syntaxi příkazu můžeme obecně vyjádřit zápisem

[**by varlist:**] **příkaz** [varlist] [**weight**] [**if exp**] [**in range**] [, **volby**]

Části v hranatých závorkách jsou nepovinné. Pomocí jednotlivých částí je možné vymezit způsob vykonání příkazu, např. **by varlist:** specifikuje, že výpočet má být proveden postupně pro všechny skupiny vytvořené podle hodnot diskretních veličin vymezených v seznamu **varlist**, **if exp** omezuje výpočet na podmnožinu dat splňující podmínku vyjádřenou logickým výrazem **exp**, částí **volby** lze ovlivnit obsah a podrobnost výstupu. Několik příkazů dovoluje i komplikovanější syntaxi, např. předponu **sw** při vytváření regresních modelů postupnou metodou. Příkazem může být i název souboru obsahujícím sekvenci příkazů. Takovou sekvenci lze interaktivně vyzkoušet a pak pohodlně vytvořit z příkazů uložených v okně **REVIEW**.

Zadané příkazy a textová část výsledků se zobrazují vždy v okně **RESULTS**. Pokud máme otevřen soubor **LOG**, pak jdou paralelně i do tohoto trvalého souboru. Pro grafický výstup se automaticky otevře okno **GRAPHS**. Požadujeme-li uložení grafu do nějakého trvalého souboru pro pozdější využití, lze to udělat celkem snadno výběrem operace **SAVE** v menu. Grafy se tedy podobně jako např. v **S-PLUS** zobrazují odděleně od textových výsledků a při vytváření zprávy je nutno je ručně spojovat.

Způsob práce v paketu **STATA**, strukturu a styl zobrazování výsledků můžeme ukázat na několika jednoduchých příkladech. Základní charakteristiky všech veličin z aktuálně otevřeného souboru (vynecháme-li v zadání příkazu seznam veličin, implicitně se předpokládají všechny veličiny):

Summarize

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|----------|------|----------|-----------|----------|----------|
| year | 1248 | 89.97115 | 3.149695 | 85 | 95 |
| month | 1248 | 6.5 | 3.453436 | 1 | 12 |
| area | 1248 | 2.270032 | 1.145403 | 1 | 4 |
| lntsp | 1178 | 4.517814 | 0.424038 | 2.397895 | 5.891644 |
| tsp | 1178 | 100.1868 | 44.73024 | 11 | 362 |

Podrobnější charakteristiky veličiny **lntsp** můžeme dostat po následujícím zadání (příkaz **summarize** lze zkrátit, volba **detail** zruší implicitní volbu):

```
sum lntsp, detail
```

Intercooled Stata 5.0

File Edit Eracs Window Help

Log... Display Results Graph Reggraph Editor Browse More Break

Review

```

sum intsp, detail
sum intsp, detail
clear
sol maximize 200
use "D:\PROMN\151
anova intsp area mon
sum
anova intsp area mon
sum intsp, detail

```

Stata Results

```

Model | 98.0472897 | 14 | 4.14623498 | 31.40 | 0.0000
-----+-----
R-sq.   | 15.9388899 | 3 | 5.17796684 | 99.21 | 0.0000

```

Stata Log

```

. sum intsp, detail
-----+-----
Percentiles
 1| 3.413917 | Smallest
 5| 2.397695 |
10| 2.564946 |
20| 2.033213 |
25| 2.995732 | Largest
 50| 4.49931 |
 75| 4.819221 |
 90| 5.049856 |
 95| 5.219486 |
 99| 5.518422 | 5.891644

```

sum intsp, detail

| | Obs | Sum of Wgt. | Mean | Std. Dev. | Variance | Skewness | Kurtosis |
|----------|------|-------------|----------|-----------|----------|-----------|----------|
| Smallest | 1170 | 1178 | 4.517814 | .4240382 | .1798094 | -.1419869 | 4.157469 |

Command: D:\PROMN

sum intsp, detail

Variables

year
month
area
intsp
tsp

Obr. 1: Pracovní obrazovka paketu STATA.

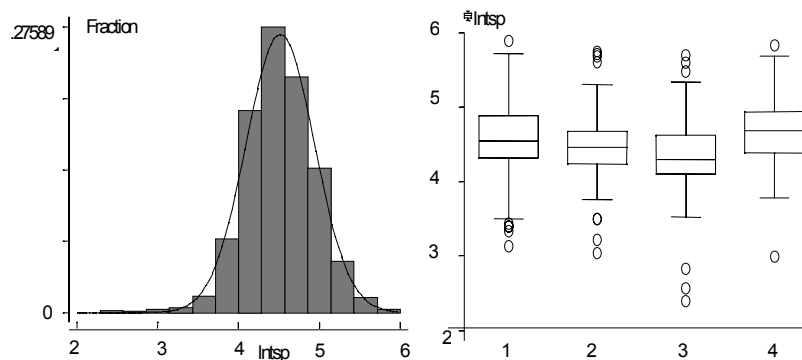
| ----- | | | lntsp | | ----- | |
|-------|-------------|----------|-------------|--|-----------|--|
| | Percentiles | Smallest | | | | |
| 1% | 3.433987 | 2.397895 | | | | |
| 5% | 3.871201 | 2.564949 | | | | |
| 10% | 4.007333 | 2.833213 | Obs | | 1178 | |
| 25% | 4.248495 | 2.995732 | Sum of Wgt. | | 1178 | |
| 50% | 4.49981 | | Mean | | 4.5178140 | |
| | | | Std. Dev. | | 0.4240382 | |
| | | Largest | | | | |
| 75% | 4.804021 | 5.710427 | Variance | | 0.1798084 | |
| 90% | 5.049856 | 5.743003 | Skewness | | 0.1428859 | |
| 95% | 5.209486 | 5.826000 | Kurtosis | | 4.1574890 | |
| 99% | 5.598422 | 5.891644 | | | | |

Grafy popisné statistiky a jejich spojování do jednoho objektu ukazuje sekvence příkazů:

```
graph lntsp, hist bin(14) norm xlab
graph lntsp, box by(area) ylab
```

Grafy se zobrazují v okně GRAPH a výběrem příkazu SAVE z menu je lze snadno uložit do souborů *lhist* a *lbox*. Pak je můžeme spojit do jednoho objektu a ten umístit do zprávy, viz obr. 2:

```
graph using lhist lbox
```



Obr. 2: Grafy popisné statistiky v paketu STATA.

Analýzu rozptylu v modelu s dvěma pevnými faktory zadáme příkazem
anova lntsp area month


```

Number of obs =      1178   R-squared      = 0.2743
Root MSE      = 0.363402   Adj R-squared = 0.2655

```

| Source | Partial SS | df | MS | F | Prob>F |
|----------|------------|------|------------|-------|--------|
| Model | 58.0472897 | 14 | 4.14623498 | 31.40 | 0.0000 |
| month | 42.410252 | 11 | 3.85547745 | 29.19 | 0.0000 |
| area | 15.533899 | 3 | 5.17796634 | 39.21 | 0.0000 |
| Residual | 153.587181 | 1163 | 0.13206121 | | |
| Total | 211.634471 | 1177 | 0.17980839 | | |

STATA vypočte jen to, co žádáte, v tomto případě tedy model bez interakce. Chceme-li úplný model, interakci je nutno zadat v seznamu proměnných:

```
anova lntsp month area month*area
```

```

Root MSE      = 0.3672   Adj R-squared = 0.2501
Number of obs = 1178    R-squared     = 0.2801

```

| Source | Partial SS | df | MS | F | Prob>F |
|------------|------------|------|-----------|-------|--------|
| Model | 59.2700458 | 47 | 1.2610648 | 9.35 | 0.0000 |
| month | 39.4560358 | 11 | 3.5869123 | 26.60 | 0.0000 |
| area | 15.4985645 | 3 | 5.1661882 | 38.31 | 0.0000 |
| month*area | 1.22275610 | 33 | 0.0370532 | 0.27 | 1.0000 |
| Residual | 152.364425 | 1130 | 0.1348358 | | |
| Total | 211.634471 | 1177 | 0.1798084 | | |

Potřebujeme-li skupinové průměry, musíme si je vyžádat pomocí příkazu *table*. V seznamu proměnných se zadají kategoriální veličiny, podle kterých se data seskupují, ve volbách požadovaná statistika, případně i její formát ve výstupu:

```
table month area, c(mean lntsp) format(%9.4f)
```

| month | area | | | |
|-------|--------|--------|--------|--------|
| | 1 | 2 | 3 | 4 |
| 1 | 4.8753 | 4.7056 | 4.6446 | 4.9765 |
| 2 | 4.8588 | 4.7571 | 4.6540 | 4.9405 |
| 3 | 4.7433 | 4.6620 | 4.4979 | 4.8116 |
| 4 | 4.6319 | 4.5347 | 4.3704 | 4.6931 |
| 5 | 4.3942 | 4.3933 | 4.2369 | 4.5396 |

| | | | | | |
|----|--|--------|--------|--------|--------|
| 6 | | 4.2356 | 4.2103 | 4.1015 | 4.4304 |
| 7 | | 4.2436 | 4.2025 | 4.1092 | 4.4028 |
| 8 | | 4.4107 | 4.2904 | 4.1612 | 4.4402 |
| 9 | | 4.4949 | 4.3200 | 4.2135 | 4.5642 |
| 10 | | 4.6705 | 4.5256 | 4.3346 | 4.7016 |
| 11 | | 4.6979 | 4.6020 | 4.3784 | 4.7849 |
| 12 | | 4.6187 | 4.4309 | 4.3986 | 4.7364 |

Uvedené příklady možná vzbuzují dojem, že analýza dat s pomocí paketu, který poskytuje jen to, co si vyžádáme, je nepohodlná. Ale není tomu vždycky tak, neboť ve STATĚ jsou i nástroje, které pohodlí poskytují. STATA má např. jednu velmi užitečnou vestavěnou funkci, která není v jiných paketech nebo jí není možné tak pohodlně zadat - implicitní generování indikátorových proměnných z původních kategoriálních veličin. Indikátory se rozumí $k-1$ dichotomických veličin (s hodnotami $\{0,1\}$), které vzniknou transformací nominální veličiny s k kategoriemi tak, že pro referenční kategorii je hodnota všech indikátorů rovna 0, pro ostatních $(k-1)$ kategorií je hodnota i -tého indikátoru rovna 1, pokud nominální veličina má hodnotu odpovídající i -té kategorii, jinak je hodnota indikátoru rovna 0. Má-li původní veličina chybějící hodnotu, pak hodnota všech indikátorů je *missing* (chybějící). Implicitně je referenční kategorie ta s nejnižším kódem hodnoty původní veličiny, definování jiné referenční kategorie je nutno zadat explicitně a to pak platí až do jeho zrušení, např.

```
char area [omit] 3
char gsov1 [omit] 1
char e14 [omit] 2
```

Implicitní generování indikátorů ilustruje příklad zadání logistické regrese. Předpona *xi:* v příkazu vyvolá generování indikátorů z kategoriálních veličin, před jejichž jménem je prefix *i.* (připomeňme, že v této úloze byly všechny původní regresory kategoriální).

```
xi: sw logistic b26(i.area i.ageg i.a1 i.a4mr i.a3r
i.gsov1 i.f33 i.cusmm i.edulm i.occlm i.g7) i.edulf
i.emplm i.emplf i.occlf i.cusmf i.c17r i.e2r i.bwt1
i.e5 i.e6 i.e8r i.e9r i.e10r i.e11r i.e12r i.e13r
i.e14 i.e15r i.e16 i.f1 i.dens1 i.f3r i.f4 i.f6r
i.f7 i.f21 i.f23 i.htgol i.htcwl i.f31 i.f34r
i.smpm i.traff,
forw pe(.10) pr(.11) lr lock
```

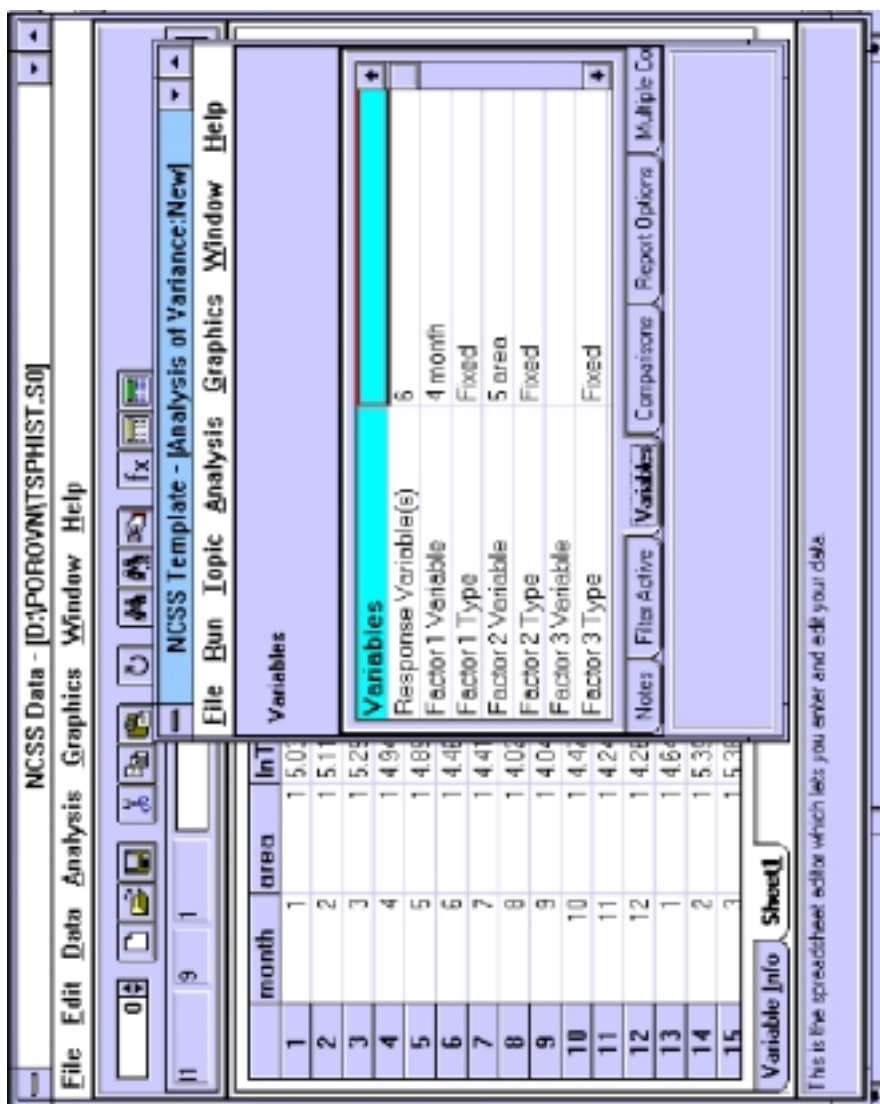
NCSS 6.0

Označení NCSS je zkratka názvu *Number Cruncher Statistical Systems*. Autorem tohoto statistického balíčku je J.L. Hintze, stejně jako známého balíčku SOLO. V NCSS ostatně návaznost na SOLO lze snadno vystopovat, zejména v paletě metod a ve struktuře a organizaci výstupů.

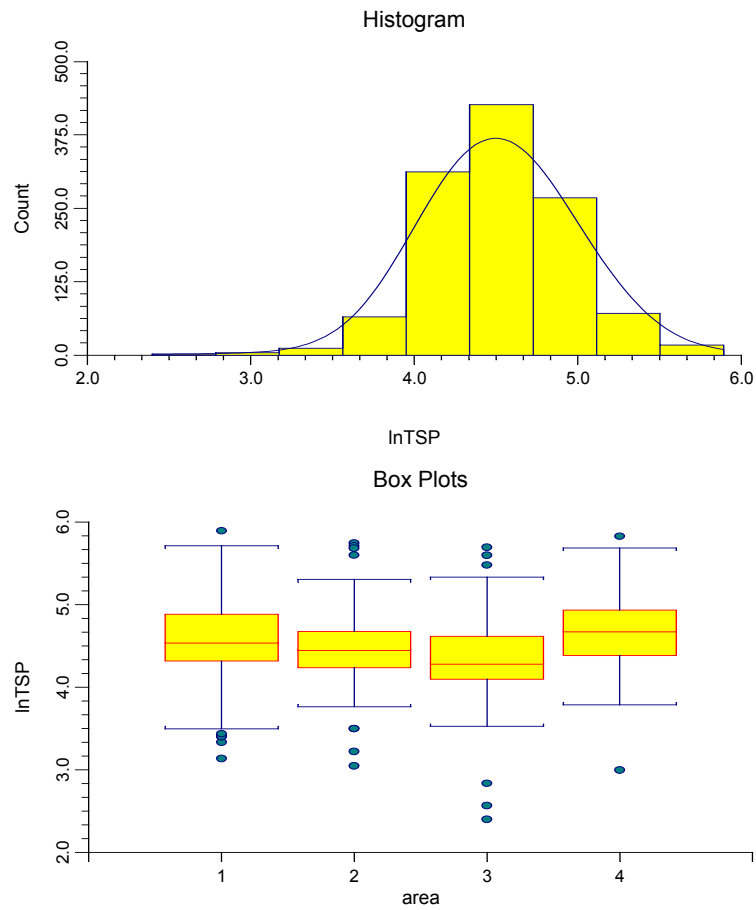
Celková charakteristika: universální statistický balíček, doporučovaný zejména uživatelům-(ne)statistikům. Pokrývá však naprostou většinu požadavků i velmi sofistikované statistické analýzy dat. Ovládání pomocí výběru z menu. NCSS komunikuje stylem „*nabízím, co pravděpodobně můžete nebo máte v dané situaci požadovat, pokud vám to nevyhovuje, musíte to vyjádřit*“. Výsledky (textový i grafický výstup společně) jsou ve formátu RTF (*Rich Text Format*) a tedy snadno importovatelné do běžných textových procesorů.

Struktura pracovní obrazovky balíčku NCSS je na obr. 3. Na obrazovce se překrývají okno se zpracovávanými daty, okno tzv. šablon (*templates*), které umožňuje specifikovat vstupní parametry zvolené analytické procedury, aktuální okno výsledků a LOG okno s výsledky pro trvalé uložení. Mezi okny lze snadno přepínat. Hlavní způsob ovládání je výběr z menu a vyplňování formulářů pomocí myši, v mnohém podobné práci s tabulkovými procesory. Vyplněné šablony lze uložit pro opakované použití. Do LOG souboru se ukládají pouze ty výsledky, které uživatel uloží explicitně, jinak jsou ztraceny a okno aktuálních výsledků je přepisováno následující spuštěnou procedurou. Zadávání transformací veličin a sdružování kategorií je jednoduché, spuštění výpočtu jen pro podmnožinu případů je možné, ale poměrně komplikované. Pokud úloha vyžaduje komplikovanější předzpracování dat, je většinou výhodné toto předzpracování udělat jiným programovým prostředkem a data pak do NCSS importovat. Import a export mnoha běžných formátů dat je součástí NCSS.

Kvalitu grafických výstupů ilustruje obr. 4. Styl textových výstupů ukazují výsledky úlohy analýzy rozptylu s dvěma pevnými faktory, které uvedeny jsou v dalším textu.



Obr. 3: Pracovní obrazovka NCSS.



Obr. 4: Příklad grafů popisné statistiky v NCSS

Pro analýzu rozptylu nabízí NCSS několik procedur. Pro tuto úlohu můžeme zvolit metodu *Unweighted Means*, která však dovoluje pouze úplný model. Pak obdržíme následující výstup:

Analysis of Variance Report

| | |
|-----------|--------------------------|
| Page | 1 |
| Database | D:\CESAR\OVZD\TSPHIST.S0 |
| Time/Date | 18:08:23 05-12-1997 |
| Response | lnTSP |

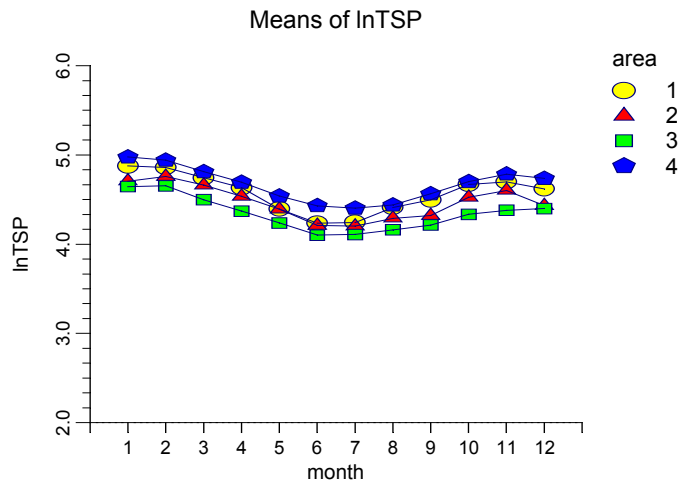
Analysis of Variance Table

| Source Term | DF | Sum of Squares | Mean Square | Prob F-Ratio | Power Level | Apha=0.05 |
|-------------|------|----------------|-------------|--------------|-------------|-----------|
| A (month) | 1 | 39.25019 | 3.568199 | 26.46 | 0.000000 | *1.0 |
| B (area) | 3 | 16.54815 | 5.516051 | 40.91 | 0.000000 | *1.0 |
| AB | 33 | 1.124350 | 0.034071 | 0.2 | 0.999995 | 0.246310 |
| S | 1130 | 152.3644 | 0.134836 | | | |

Total
Adjusted 1177 211.6345
Total 1178

Term significant at alpha = 0.05

Vidíme, že oproti STATĚ v NCSS jsou navíc spočítány i síly testů. Číselné výsledky v tabulce jsou trochu odlišné od těch, které byly spočítány programem STATA. Shodné číselné výsledky bychom dostali, pokud bychom v NCSS zvolili metodu GLM, která navíc i dovoluje specifikovat i tvar modelu, v NCSS je na rozdíl od paketu STATA implicitní volbou úplný model. Součástí výsledků analýzy rozptylu v NCSS jsou rozsáhlé tabulky všech skupinových průměrů a dalších charakteristik (pro značnou délku zde nejsou uvedeny) a také jejich grafické znázornění, jehož příklad vidíme na obr. 5



Obr. 5: Graf skupinových průměrů

Porovnání STATY a NCSS

Charakteristiky pro porovnání obou paketů byly sestaveny s přihlédnutím ke kritériím, která navrhuje Řezanková a Svoboda [2].

| | STATA 5.0 | NCSS 6.0 |
|---|---|--|
| Ovládání | příkazy, s částečnou podporou výběru z menu | menu, šablony |
| Help | vyhledávací | kontextuální |
| Výukový program | samostatný program StataQuest | částečně, součást systému |
| Doprovodná učebnice statistiky | tři samostatné knihy | jen manuál s komentovanými odkazy |
| Import/export jiných formátů dat | Stat/Transfer (zvláštní program), jinak jen ASCII | mnoho, ale některé ne zcela spolehlivé |
| Vstup dat z klávesnice | tabulkový editor | tabulkový procesor |
| Vstup kontingenční tabulky | ano | ano |
| Data Management | kompletní relační operace, snadné zadávání | tabulkový procesor, na rozsáhlých datech obtížné |
| Monitorování dlouhých výpočtů | ne | ano |
| Výstup | text + obrázky zvlášť | RTF |
| Příkazové soubory | ano | jen uložení šablon |
| Přesnost výpočtů | double | double |
| Programování | ano | ne |

Oba porovnávané pakety vyhovují ve většině kritérií. Nepřehlédnutelným nedostatkem paketu STATA je to, že nejsou monitorovány dlouhé výpočty, takže po zadání výpočetně náročné úlohy zůstáváte dlouhé minuty v nejistotě, zda výpočet vůbec pokračuje. U NCSS je nevýhodou velmi omezená podpora opakování dříve zadaných výpočtů a složité zadávání některých operací manipulace s daty.

Závěr

Oba porovnávané pakety patří do kategorie obecných statistických systémů a implementovanými statistickými metodami široce pokrývají běžně užívané postupy analýzy dat. Mohou tedy uspokojit i statisticky náročného uživatele. Pokud nevymezíme hodnotící hledisko, nelze jeden upřednostnit před druhým.

Paket STATA lze doporučit zejména pro pravidelné užívání, kdy ovládání příkazy je pak spíše přednost než obtíž. Výhodou jsou dobré možnosti data managementu i pro rozsáhlá data a otevřenost systému (programování, příkazové soubory pro opakované analýzy). NCSS je vhodný zejména pro uživatele občasně, i pro (ne)statistiky. Pak je výhodou snadné ovládání pomocí menu, pohodlná práce s méně rozsáhlými daty, vysoká grafická kvalita výstupů i jejich snadný import do textových procesorů.

Ačkoliv tabulka v předcházejícím odstavci je zaplněna pozitivním hodnocením ve většině políček, ani jeden z porovnávaných statistických paketů nemůžeme považovat za ideální. Lze dokonce pochybovat, zda ideál je vůbec dosažitelný, mimo jiné i proto, že požadavek na ideální úplnost funkcí paketu a dalších požadovaných vlastností je často v rozporu s požadavkem na jednoduchost ovládání, komunikace a snadnost interpretace výsledků. Nezbyvá, než se smířit s tím, že pro výběr statistického paketu je pořád velmi důležité, *kým a pro jaké úlohy* má být převážně užíván. Statistická analýza dat i s dobrým programovým vybavením je v naprosté většině případů intelektuálně náročná činnost, ve které dovednost ovládání statistického paketu představuje jen menší část požadavků kladených na řešitele úlohy.

Další informace o porovnávaných paketech lze získat na stránkách <http://www.statsol.ie> (NCSS) resp. <http://www.stata.com> (STATA).

Literatura:

- [1] Hintze J. L.: *NCSS 6.0.1 User's Manual I and II*, Number Cruncher Statistical Systems, Kaysville, Utah, 1995.
- [2] Řezanková H., Svoboda K.: *The Ideal Package for Statistical Analysis*, In: COMPSTAT 96, Short Communications and Posters (ed. Prat A., Ripoll E.), 221-222, UPC Barcelona, 1996.
- [3] StataCorp: *STATA release 5 User's Guide*, College Station, Texas, 1997.

S čím dnes na statistiku !?

Jiří Žváček

V tomto příspěvku se zabýváme statistickým softwarem za současného „internetího“ stavu, který stále ještě není stavem konečný a pro potřeby tohoto příspěvku je navíc podstatně zjednodušen. Zvolený pohled je především z hlediska výuky a elementární statistické analýzy a nezabýváme se proto jednotlivými oblastmi statistiky. Je vcelku nesporné, že statistický software se stává oblastí „trhu“ se všemi negativními i pozitivními aspekty. Až na prvním místě jde o peníze a podle zrychlujícího se tempa inovací jich nebude asi málo. I velké a „vznešené“ statistické pakety ustupují od šovinistického „SAS není pro PC“ a snad až komicky působí nabídkově řízený S-PLUS v.4. Stále více roste význam Internetu a „počítačového“ přístupu k výuce daný masovým přístupem k výkonným počítačům v síti a změnou chování studentů i učitelů. Informace o software obsahuje snad každá souhrnná stránka. Stále důležitější (ale i obtížnější) je tedy se na tomto trochu bizarním trhu orientovat.

I. Nabídka

Shareware a specializované pakety se příliš nepoužívají, protože nemají adekvátní podporu. Důvodem je jednak malá důvěra ve spolehlivost výsledků a též neochota se učit nestandardní způsob ovládnání. Nicméně tato možnost existuje. Souhrnné stránky jsou například následující.

- **Kovach** Computing Services Home Page na Internetovské adrese <http://www.kovcomp.com> specializující se na statistický shareware a program MVSP.
- **Simtel** (Keith Petersen) na adrese <http://www.cdrom.com/simtel.net> je obsáhlá sbírka sharewaru. Existuje k ní CD ROM, a dokonce je zrcadlo na VŠE (a přirozeně též v Liberci).

Vhodná ke „stahování“ je **HENSA** na adrese:

<http://micros.hensa.ac.uk/cgi-bin/browser/mirrors/simtelnet/msdos/statstcs>.

Pohrdání sharewarovým softwarem je tak trochu škoda, protože díky Internetu lze snadno získat velmi pěkné produkty jako třeba:

- **Prophet** dostupný na adrese <http://www-prophet.bbn.com/>.
- nebo
- **Kwikstat** a jeho variantu pro Windows se jménem **WINKS**, již lze nalézt na adrese <http://www.texasoft.com/>

Obecné pakety mají ambice pokrýt celou aplikačně zajímavou oblast statistiky. Jedná se většinou o komerční pakety, které mají obvykle vlastní firemní stránku a rozsáhlou podporu firmy, univerzit a zájmových skupin. Význam těchto stránek roste. Jsou na nich publikovány novinky, popisy, lze si stáhnout dema, data, programy či skripty. Paketů jsou stovky. Dnes mají praktický význam pouze ty, které jsou mezinárodně dostupné a mají „internetovskou“ podporu.

- **GAMS** je rozsáhlá stránka o matematickém a statistickém softwaru na adrese <http://gams.nist.gov/>, kde lze nalézt všechno možné.
- Podobně je to u **Dukeů** na adrese <http://www.isds.duke.edu/sites/software.html>.
- **Stata** na adrese http://www.stata.com/support/links/stat_software.html má pěkný seznam komerčních paketů.
- Hodnocení mnoha paketů nalezneme na adrese <http://milkyway.stats.gla.ac.uk/cti/activities/reviews/alphabet.html>.
- Pro software je velmi významná stránka **StatLib** katedry statistiky Carnegie-Mellon University, která obsahuje velmi mnoho zajímavých zdrojů a softwaru. Domovská stránka je <http://lib.stat.cmu.edu/> a má dokonce evropská zrcadla. Zrcadel je poměrně mnoho a mají je nejenom státy, ale i některé katedry (Oxford).

V Evropě se toho příliš mnoho neděje. Komerčně se prosadil i v USA britský **UNISTAT**. Zajímavé jsou některé německé výzkumné projekty, zejména paket založený na Oberonu v **Heidelbergu** a **XploRe** Humboldtovy university, který nabízí řadu pokročilých metody (vyhlazování, neuronové sítě, wavelety apod.).

Z jednotlivých evropských stránek jsou zajímavé zejména archivy.

- Archiv **IASC** v Perugii na adrese <http://fisher.stat.unipg.it/iasc/> obsahující také řadu zrcadel (včetně StatLibu).
- Britský archiv **HENSA** na adrese <http://www.hensa.ac.uk/>, též s celou řadou zrcadel. Pozor, pro zájemce mimo britské ostrovy je přístupný pouze v noci!
- Velmi dobrý rakouský **SUNSITE** Vídeňské univerzity čtenář nalezne na adrese <http://sunsite.univie.ac.at/>.
- **Statlab Heidelberg** na adrese <http://statlab.uni-heidelberg.de/www.html> obsahuje řadu zrcadel významných serverů (včetně StatLibu).

Dále se alespoň krátce zmiňme o systémech, které lze pokládat za klíčové.

- Obří **SAS** je vhodný pro velké instituce, protože je šířeji zaměřen na analýzu dat a více méně předpokládá specializované pracoviště a potřebu vytvářet aplikace tzv. „na míru“. Statistický modul je pouze malou součástí systému a nyní je též nabídkově ovladatelný. Domovskou stránku má na adrese <http://www.sas.com/> a obsahuje například i statistické texty. SAS má další „internetí“ podporu - viz třeba Friendly na adrese <http://www.math.yorku.ca/SCS/StatResource.html#sas> nebo University Wisconsin na adrese <http://www.stat.wisc.edu/computing/sas/>.
- Velmi expanduje **SPSS**, který pohltil mimo jiné Systat, BMDP a Sigma Plot. Domovská stránka s texty a downloady je na adrese <http://www.spss.com/>. Podporu nalezneme též na adrese <http://www.math.yorku.ca/SCS/StatResource.html#spss>. Síla SPSS je především v práci s daty a tabulkami. Nejnovější verze 8.0 má mít zejména vylepšen výstup do MS Office.
- Nejlépe je ovšem hodnocena agresivní **Statistica**, soudě alespoň podle citací z vlastní stránky, jež se nachází na adrese <http://www.statsoftinc.com/>. Statistica uvádí seznam 257 amerických a kanadských univerzit jež ji užívají. Je lokalizována do mnoha jazyků (mimo jiné i čínsky, japonsky, německy, polsky, francouzsky, italsky, ...), přičemž manuály existují i korejsky a rusky!
- **UNISTAT** je plně integrován do prostředí Microsoft Office, což znamená prostředí Excelu s výstupy do Wordu. Domovská stránka s texty a downloady je na adrese <http://www.unistat.com/>. UNISTAT je lokalizován i do **češtiny**. Recenze na verzi 4.0 lze nalézt na adrese http://milkyway.stats.gla.ac.uk/cti/activities/reviews/96_11/excel/unistat.html.
- Možnost využít tabulkové procesory pro výuku je velmi lákavá, o čemž svědčí např. i specializovaná stránka na adrese <http://sunsite.univie.ac.at/Spreadsite/>.
- **STATGRAPHICS** patřil ve své DOSovské podobě k nejoblíbenějším paketům pro běžnou výuku, s nástupem Windows se však pohodlí při ovládání hodně vyrovnalo. Domovská stránka s texty a downloady, obsahující mj. i datové soubory a výukové materiály je na adrese <http://www.manu.com/statgraphics/>.
- **MINITAB** má domovskou stránku s texty a downloady na adrese <http://www.minitab.com/>.

- **Stata** má domovskou stránku s texty a downloady na adrese <http://www.stata.com/>
- Stále populární je **Mathematica** (viz stránka na adrese <http://www.wri.com/>) firmy Wolfram Research, Inc., podpora je též k dispozici například na adrese: <http://www.math.yorku.ca/SCS/StatResource.html#mathematica>.
- Velkou akademickou podporu má **S-PLUS** firmy MathSoft s domovskou stránkou na adrese <http://www.mathsoft.com/splus.html>. Jedná se původně o specializovaný statistický programovací jazyk ve stylu C, který obsahuje mnoho statistických funkcí a procedur. Je velmi vhodný pro vývoj statistických numerických a grafických metod. Používá se zejména na počítačově velmi vyspělých katedrách. Nejnovější verze 4 je dokonce nabídkově řízená, ale i tak zůstává pro masy trochu složitá. Rozsáhlou sbírku procedur a funkcí pro S+ nabízí **MatLib** na adrese <http://lib.stat.cmu.edu/S/>. Další podporu lze nalézt na adrese: <http://www.math.yorku.ca/SCS/StatResource.html#Splus>, sbírku norských odkazů na adrese <http://www.nr.no/stat/engelsk/splus.html>.

Kromě „komerčního“ paketu **S-PLUS** existuje i jeho volná (free) varianta **R**, která je s ním datově i povelově kompatibilní s **S-PLUS**. Byla vyvinuta v Auclandu a je k dispozici na adrese <http://stat.auckland.ac.nz/r/r.html>. Vzhledem k tomu, že se jedná o velmi zdařilý produkt, vážným zájemcům jej vřele doporučujeme (poznámka redakce).

- Velmi zajímavě vypadá i **LispStat**, který je nekomerční. Je založen na Lispu a pracuje v něm mnoho výpočetních statistiků. Nebýt tzv. *polské* notace, jednalo by se patrně o jeden z nejpulárnějších programů pro statistickou analýzu dat. Domovská stránka s texty a downloady je na adrese <http://stat.umn.edu/~luke/xls/xlsinfo/xlsinfo.html>. Má řadu zrcadel, např. <http://sunsite.univie.ac.at/pub/statistics/ucla/xlispstat/umnstat.stat.umn.edu/>. Na bázi Lispu pracuje též graficky orientovaná **ViSta**. Viz též **Friendly** na adrese <http://www.math.yorku.ca/SCS/StatResource.html#lispstat>
- **XploRe** Humboldtovy university je také volný a má velmi pokročilé metody (neuronové sítě, wawelety) na adrese <http://www.xplo-re-stat.de/>. Zajímavé na XploRe je fakt, že si jeho činnost může zájemce „vyzkošet“ přes Internet pomocí Netscape. Zatím sice neexistuje plnohodnotná grafika, ale i přesto se jedná o zajímavý projekt. (XploRe úplně volný není. Lze si sice stáhnout fungující verzi s omezenou dobou použití, poté je však třeba si zakoupit plnohodnotnou licenci. Poznámka redakce.)

- Z komerčních paketů menšího významu je možno jmenovat i **NCSS** na adrese <http://www.ncss.com/>, na kterém je zajímavý výstup do RFT formátu (lze stáhnout demo).

II. Trh

O tržbách a počtech prodaných produktů je málo známo a alespoň nějaké šetření by mělo být provedeno. Jednak na členech České statistické společnosti, jednak dotazy u firem. Něco málo se přesto ví, tak například:

- **SPSS** hovoří o 2 miliónech uživatelů.
- **Statistica** uvádí půl milionu uživatelů (a je dokonce lokalizována do polštiny!).
- **MINITAB** se používá v 60 zemích.
- **SAS** má 4500 zaměstnanců.

Z hlediska použití statistických systémů můžeme rozlišit následující oblasti využití.

Výuka

Je dnes velmi důležitým trhem, protože statistika se učí skoro všude. Jedná se tisíce univerzit a milióny studentů! „Matfyzáků“, na které bývala výuka statistiky orientována, spíše ubylo, páteř dnes tvoří ekonomové. Učitelé si možná povšimli, že úroveň studentstva poklesla, přibylo jich a jsou drzí - ptají se, k čemu to je, a když něčemu nerozumějí, tak viní učitele a předmět. Vzhledem k tomu, že jim dnešní společnost přiznala jistou možnost volby, musejí i statistikové o studenty bojovat. Klíčovou roli přitom hrají počítače (protože jsou symbolem pokroku) a software, který musí být uživatelsky přátelský, atraktivní a stále častěji v národním jazyce. Protože se při výuce zpravidla nelze omezit na jediný paket a univerzity mají dosti heterogenní prostředí, je důležitá i podpora „počítání“ na síti. Příkladnou podporu statistické výuky má **University of Indiana** na adrese <http://www.indiana.edu/~statmath/stat/> či Humboldtova universita v Berlíně.

U nás spíše z finančních a „hardwarových“ důvodů dominují méně náročné pakety, zejména **STATGRAPHICS**, i když by bylo jistě perspektivnější se vrhnout na ambicióznější a pro praxi užitečnější **SPSS** či **SAS**. Většina těchto studentů se stejně musí naučit **Word** a **Excel**, což silně hovoří (alespoň na školách ekonomického typu) pro **UNISTAT**. Pro rozhodování o konkrétním paketu jsou důležité i další aspekty.

- **Výukové materiály.** Stále důležitější je „počítačová“ podpora paketů a výuky. Na Internetu existuje spousta textů, často hypertextových či interaktivních. Existují i specializované stránky věnované výuce statistiky, například:
- **Puranen** na adrese <http://www.helsinki.fi/~jpuranen/links.html>.
- **Friendly** na adrese <http://www.math.yorku.ca/SCS/StatResource.html#courses>, kde lze nalézt hodně materiálu o výuce.
- Mnoho materiálu nabízí Univerzita Heidelberg na adrese <http://www.execpc.com/~helberg/statistics.html#education>.
- Zde lze najít např. i konzultační a výukové odkazy, statistické kurzy na Webu a publikace známých nakladatelství. Konkrétní kurzy a manuály lze nalézt na stránkách kateder a důležité jsou i výukové materiály podpory produktů a samotných produktů.

Data.

Poměrně málo pozornosti se věnuje datům. Je to dáno tím, že ještě nedávno bylo problémem spočítat korelační koeficient pro více než 10 pozorování a většina učitelů i praktických statistiků z této doby „pochází“. Toto se musí bezpodmínečně změnit a studenti musí používat reálná data a rozsáhlé soubory. Je přitom třeba se soustředit na dva okruhy problémů.

- **Odkud brát data?** Studenti již dnes nebudou ťukat do počítače několik cvičných čísel, ale bude nutné ke kurzu data připravit. Zřejmě bude nutné je importovat spolehlivě z mnoha různých produktů a být připraven na obrovské soubory a manipulace s daty. (Okamžitě se nabízí několik vlezlých otázek, například *Kde v dnešní tržní době data vzít a nekrást?*, resp. *Máme k dispozici dostatek metod vhodných pro statistickou analýzu velmi rozsáhlých dat?* Poznámka redakce.)
- **Jak a kde data „skladovat“?** Existuje mnoho tématických souborů a příkladů ke knihám a statistickým paketům. Mnoho datových stránek popisuje Friendly na adrese <http://www.math.yorku.ca/SCS/StatResource.html#Data> a hojně se využívá i datová sbírka StatLib na domovské stránce katedry statistiky Carnegie-Mellon. Speciálně datům je věnována stránka Data Sources (Statistics, WWW Virtual Library)_Davida Rosena, která je částí WWW Virtual Library na adrese <http://www.scs.unr.edu/~cbmr/research/data.html>.

Výzkum

Tuto oblast je třeba pečlivě sledovat, protože zde vznikají zásadní inovace, které se zpravidla velmi rychle promítnou do komerční oblasti. Výzkum se týká hlavně několika vyspělých kateder a omezeného počtu počítačově orientovaných studentů statistiky. Odkazy na katedry lze nalézt na následujících stránkách.

- **StatLib** na adrese <http://www.hensa.ac.uk/ftp/mirrors/statlib/otherplaces/>.
- Nejobsáhlejší seznam kateder nabízí **Duke University** na adrese <http://www.isds.duke.edu/sites/academic.html>.

Určitým vodítkem může též být americký žebříček výuky statistiky za rok 1996 je (viz stránka University Madison).

Matematická statistika

1. **University of California at Berkeley** na adrese <http://elsa.berkeley.edu/>.
2. **Stanford University** na adrese <http://playfair.stanford.edu/>.
3. **University of Wisconsin at Madison** na adrese <http://www.stat.wisc.edu/>.
4. **Cornell University (N.Y.)** na adrese <http://www.cornell.edu/>.
5. **University of Chicago** na adrese <http://galton.uchicago.edu/>.

Toto hodnocení není automaticky směrodatné pro oblast výpočetní statistiky. Některé katedry mají specializované projekty které se systematicky zabývají určitou oblastí nebo mají vhodného specialistu. Zdá se, že za takové lze pokládat (kromě již zmíněné Carnegie-Mellon) následující:

University of Florida má díky Miku Conlonovi výbornou stránku na adrese <http://www.stat.ufl.edu/vlib/statistics.html>, která obsahuje spousty odkazů na další katedry.

- Australská **University of Queensland** má vynikající stránku **Guide on the Web for Statisticians** na adrese <http://www.maths.uq.oz.au/~gks/webguide/index.html>.
- **Institut statistiky a ekonometrie** Humboldtovy university na adrese <http://wotan.wiwi.hu-berlin.de/statistics/index.html> vyvíjí **XploRe**. Jeho součástí je i německý Národní výzkumné centrum pro modelování a simulaci ekonomických procesů.
- **Katedra statistiky university v Glasgově** na adrese <http://www.stats.gla.ac.uk/home.html> provozuje **CTI Center for Statistics** (Britská centra technické pomoci sbírající a redistribuuující tyto „znalosti“).

Na statistických katedrách pro pokročilou výuku tradičně dominuje **S-PLUS**, ale poměrně časté jsou i klasičtější „numerické“ výzkumy ve *Fortranu* a *C++* vycházející z obrovských archivů matematických algoritmů, jako je např. **GASP** na adrese <http://www.stat.sc.edu/rsrch/gasp/>.

Několik kateder se vrhlo na **Lisp-Stat**, jež je k dispozici na adrese <http://stat.umn.edu/~luke/xls/xlsinfo/xlsinfo.html> a který má řadu zrcadel i v Evropě.

Praxe

Zde je třeba vycházet z toho, co od „statistiky“ čekáme. Výstupy ze statistických paketů již dnes nejsou primitivní znakové sestavy, ale posloupnosti textů, tabulek a grafů. Tyto prvky jsou objekty, se kterými můžeme dále manipulovat - upravovat je, zvýrazňovat podstatné atd. za účelem adekvátní prezentace výsledků. U většiny paketů jsou bohužel výsledky v málo přehledné množině „worken“ a tak u pokročilých paketů (**SPSS**, **S-PLUS**) dochází dokonce k automatické konstrukci „obsahu“.

Text.

Dnes již je nutné, aby i statistické pakety pracovaly v textu s fonty a barvami, a uživatel měl možnost výstupy dále upravovat, dopisovat do nich komentáře a vysvětlení apod. Zde je velký pokrok i v podpoře ze strany řady paketů - výstupy jsou adekvátně popisovány a je možno je doplnit o podrobnější vysvětlení - často i z nápovědy. Možnosti UNISTATu jsou v tomto směru asi největší

Tabulky.

Tabulky je nyní nutno prezentovat esteticky a barevně. Tuto možnost tabulkové procesory mají již dávno. U paketů je adekvátní pouze **SPSS**, který chápe tabulku jako speciální objekt a umožňuje kromě běžných operací, jako jsou úpravy textů, fontů či barev a dokonce „přestavbu“ tabulky jednoduchou záměnou řádků a sloupců atd. UNISTAT má díky Excelu běžné operace komplexnější, nicméně jednoduchost konstrukce tabulek a možnosti jejich úprav hovoří spíše pro **SPSS**.

Grafy.

Dnes není problémem zobrazit libovolný graf, takže je dostupný všem.

- **Excel** má vynikající „obchodní“ grafiku a objektový editor, který umožňuje modifikovat každý existující objekt grafu.
- **S-PLUS** má vynikající statistickou grafiku a pokročilý objektový editor, který umožňuje vytvářet nové objekty v již existujícím grafu a modifikovat je. Používá velmi pokročilý grafický systém **Trellis Graphics**, který je důkladně popsán na adrese <http://cm.bell-labs.com/cm/ms/departments/sia/project/trellis/>.
- **Systat** mívá krásné a nápadité grafy.
- **STATGRAPHICS** a **S PLUS** prvky dynamické grafiky.
- **SPSS** se prý ve verzi 8 v oblasti grafiky velmi polepší.

Speciálně matematické (a potažmo i statistické) grafice se věnují v Minnesotě, viz <http://www.geom.umn.edu/> a ve Vídni, viz <http://sunsite.univie.ac.at/Present/edmed96/>. Zdá se však, že budoucnost mají zejména **dynamické** a **aktivní** obrázky. Rozlišit lze následující prostředky.

- Rozhraní pro **on-line** výpočty, umožňující přímo „síťový“ výpočet. Sem patří i přístupy ke specializovaným vyhledávacím strojům a databázím. Různé formy interaktivního přístupu jsou evidentně na vzestupu. Mnoho odkazů v tomto směru viz <http://www.math.yorku.ca/SCS/StatResource.html#Online>.
- **Aktivní prvky**, zejména *Java applety* a programy v *Java scriptu*. Patří sem i *VBscript* a zejména *ActiveX* prvky, které lze programovat i v *Delphi* a *Visual Basicu*. Zdá se, že budoucnost prezentace statistických výsledků leží právě v tomto směru. Tyto malé programky umožňují konstrukci dynamických interaktivních tabulek, grafů a stránek, ve kterých je možno realizovat výpočty, měnit zobrazení atd. Přehled viz <http://www.stat.sc.edu/rsrch/gasp/>. Zmiňme některé typické zástupce:
 - Pěkným příkladem je třeba dynamický *aplet Eurostatu*, ukazující časovou řadu, její predikci a chybu, blíže viz. <http://europa.eu.int/en/comm/eurostat/research/supcom95.21/neural.htm>.
 - Problematikou aktivní grafiky se zabývá mnoho vědců včetně projektu Eurostatu o statistice, multimédiích a nových technologiích na adrese <http://europa.eu.int/en/comm/eurostat/research/supcom95.21/intro.htm>.
 - Přehled dynamických stránek v matematice nalezneme na adrese <http://archives.math.utk.edu/cgi-bin/interactive.html>.
 - Nabídka Duke University je na adrese <http://www.stat.duke.edu/sites/java.html>
 - Pohrát si s dvourozměrným normálním rozdělením lze na domovské stránce IASC na adrese <http://fisher.stat.unipg.it/welcome.html>.

III. Analýza dat

V nabídce základních metod a přesnosti výpočtů se pakety prakticky neliší a nabídka metod silně překračuje potřeby běžné analýzy či výuky. O to větší význam má nápověda a dokumentace, která dnes obvykle má již desítky MB (jenom manuály S-PLUS mají více než 2000 stran v Acrobatu). I z toho důvodu je často nutno volit paket podle použité učebnice.

Většina nabídkově řízených paketů nemá vhodné prostředí pro vývoj nových aplikací, a tím méně nových metod. Nové metody však nalezneme v prostředí **SASu**, **S-PLUS**, **XlispStat** či **XploRe**.

Především pak ale, motem dne by se mělo stát že čas ušetřený na „počítání“ by měl být věnován interpretaci výsledků, která by však měla být pokud možno co nejvíce standardizována.

IV. Prezentace

Konečným cílem statistické analýzy dat je prezentace výsledků. Z tohoto hlediska je možno rozlišit následující formy.

- *Prezentace v rámci statistického systému.* Ať již vezmeme kterýkoliv statistický produkt, vždy je prezentace řádově horší než v „cílových“, tj. textových systémech. Za nejlepší je možno pokládat **SPSS**, který umožňuje objektovou manipulaci s textem, tabulkami a grafy a automaticky vytváří **obsah**. Podobnou, i když mnohem komplikovanější strukturu má **S-PLUS**, kde lze prostřednictvím **Object Browseru** listovat ve většině „objektů“ paketu.
- *Textový procesor.* **UNISTAT** má výstupy přímo ve **Wordu**, **NCSS** v **RTF**. Z ostatních paketů je možný nějaký export, minimálně přes čisté **ASCII**.
- *Prezentační software.* **S-PLUS** podporuje přímý výstup do **Power Pointu**.
- *Systémy podporující výstupy ve formátu HTML.* Stále častěji je cílem prezentace výsledků prostřednictvím Internetu. **SPSS** umožňuje přijatelný export grafů, tabulek i celých výstupů (zdá se nám lepší než z **Wordu**).



Poznámka autora:

Nejnovější „klikovatelná“ verze tohoto příspěvku stránky je na adrese
<http://nb.vse.cz/kstp/win/jirkauvo/sysel/clanky/scimnas.htm>.

Poznámky redakce:

1) Popisy novinek současných verzí systémů **SPSS** a **STATGRAPHICS** lze nalézt v:

- Řezanková H. – Žváček J.: *Poradce, čaroděj a Internet*. Chip, 1997, č. 11.
- Řezanková H.: *Okénka pro statistiky*. Chip, 1998, č.1.