

Informační Bulletin



České Statistické Společnosti

č. 2, listopad 1998, ročník 9

Padesát let Mezinárodní biometrické společnosti

Marek Malý

V roce 1997 uplynulo 50 let od založení Mezinárodní biometrické společnosti (International Biometric Society, IBS), jedné z významných organizací sdružujících statistiky celého světa. Při této příležitosti bych chtěl v následujících řádcích připomenout některá historická i současná fakta svázaná se společností a s jejím časopisem *Biometrics*.

IBS označuje sama sebe za společnost zaměřenou na matematické a statistické aspekty biologie a jako potenciální členy uvádí zejména biology, matematiky a statistiky. Jejím účelem je rozvoj a podpora biologických věd prostřednictvím vývoje kvantitativních teorií a aplikací, vývoje a rozšiřování efektivních matematických a statistických technik.

Termín biometrie představuje, doslova vzato, spojením slov $\beta\iota\omicron\varsigma$ a $\mu\epsilon\tau\rho\nu$ měření živého, ve skutečnosti má poněkud užší záběr ([14]). Chápe se dnes zpravidla jako aplikace statistických metod na studium numerických dat založených na pozorování biologického fenoménu ([15]). Často se jeho záběr rozšiřuje i na oblast lékařskou a biomedicínskou a na tvorbu metod i metodologii aplikací.

Není bez zajímavosti si připomenout, že termíny ‚biometrie‘ ‚biometrický‘ se začaly dostávat do povědomí odborníků prakticky před sto lety v souvislosti s prací K. Pearsona (1857-1936), F. Galtona (1822-1911) a W.F.R. Weldona (1860-1906) ([18]), kteří posléze v roce 1901 založili časopis *Biometrika*, částečně také v reakci na postoj Britské královské společnosti, že se matematika a biologie mají vyvíjet odděleně. Sám K. Pearson později zjistil, že jím vymyšlený termín byl již dříve použit Christophem Bernoullim (1782-1863) v roce 1841. Pearsonovým předchůdcem na poli biometrie byl též Pierre Charles-Alexandre Louis (1787-1872), který se snažil zavádět početní metody v

medicínské oblasti. Dnes je za nejstarší známý výskyt termínu ‚biometrie‘ považováno jeho užití ve významu výpočtů o živých organismech Williamem Whewellem (1794-1866) při komentování jisté práce A. Queteleta (1796-1874) v dopise z roku 1831 ([8]).

I když největší rozmach biometrie spadá až do druhé poloviny našeho století, vedl např. George W. Snedecor (1882-1974) již v roce 1915 na Iowa State College kurs ‚Biometric methods of interpreting agricultural data‘ ([7]) a kapitulu o biometrii lze nalézt třeba ve Feldmanově knize ([9]), kde je přepracováním verze z prvního vydání z roku 1923. Biometrie stojí na pomezí dvou oborů, řečeno s Armitagem ([1]), není ani jedním, ani druhým, ale (což je nejdůležitější) je obojím. Význam interdisciplinárních oborů stále roste, a proto musíme ocenit prozíravost generace zakladatelů biometrické společnosti a časopisu.

Okolnosti založení IBS a jejího časopisu a další vývoj byly v literatuře několikrát popsány (viz zejména ([4]), ([6]) a nejnověji ([2]). Stručně připomeňme, že od roku 1938 existovala biometrická sekce Americké statistické společnosti (American Statistical Association, ASA). Její vedoucí Chester I. Bliss (1899-1979), vzděláním entomolog ([5]), byl v roce 1945 iniciátorem zrodu nového časopisu, *Biometrics Bulletin*, jehož editorem se stala Gertrude M. Cox (1900-1978). Časopis byl v prvním roce zasílán necelému tisíci členů sekce. Bliss se pak nejprve pokoušel ze sekce vytvořit novou americkou společnost, ale posléze převážila myšlenka transformace v mezinárodní organizaci. Po předběžných jednáních tak byla IBS založena na první Mezinárodní biometrické konferenci v září 1947. Jejími prvními představiteli se stali R. A. Fisher (1890-1962) jako prezident, J. W. Hopkins (pokladník) a Ch. I. Bliss (sekretář). *Biometrics Bulletin* se v tomto roce změnil v *Biometrics* a stal se oficiálním orgánem nové společnosti, do jejíhož vlastnictví přešel v roce 1950. V *Biometrics* byly původně uveřejňovány i zprávy ze společnosti, které nakonec od 80. let našly prostor v samostatném několikastránkovém čtyřročně vydávaném informačním zpravodaji, *Biometric Bulletin*. Pro zajímavost uvádíme, že v roce 1948 stálo roční členství ve společnosti včetně předplatného 4 dolary a v roce 1957 7 dolarů. V tomto roce měla společnost 1410 registrovaných členů, mezi nimiž byl jeden z Československa ([4]); *Biometrics* odebíralo dalších 890 nečlenů společnosti.

Již v základní koncepci IBS bylo individuální členství kombinováno s myšlenkou regionálních a národních skupin, od roku 1951 k tomu přistoupila možnost, aby se členy společnosti staly i organizace (např. univerzity či jejich části, farmaceutické firmy, atp.). IBS je dnes jednou z mála společností, které přenášejí některé své funkce a pravomoci na oblastní podskupiny. Ještě v roce 1947 byla založena první z nich pro severovýchodní americký region (ENAR),

záhy ji následoval v roce 1948 briský a australsko-asijský region. Údaje o mnoha dalších lze nalézt v ([6]). Ke konci roku 1996 měla společnost kolem 6700 členů z 96 zemí, 16 regionálních a 176 národních skupin ([17]), přičemž nejvíce členů bylo z oblasti USA (2700), Německa (900), Spojeného království (400), Japonska (280), Rakouska a Švýcarska (270) a Francie (220). Mezinárodní biometrické konference, z nichž první stála u zrodu společnosti, se nyní konají v dvouletých intervalech a kromě toho pořádají regiony své vlastní akce, jejichž podoba se liší oblast od oblasti.

Předseda České statistické společnosti ing. Z. Roth, CSc. se zúčastnil biometrického semináře rakousko-švýcarské sekce IBS (založené v současné podobě v roce 1962), který se konal ve dnech 22.-26. září 1997 ve Vídni. K hlavním tématům tam patřily statistické metody v zemědělství a lesnictví, v molekulární genetice, ve zdravotnictví, analýza klinických studií, biometrické metody statistického zajišťování kvality v potravinářství, statistické vzdělávání nestatistiků. V kuloárech semináře se mj. diskutovalo o případném rozšíření rakousko-švýcarské sekce společnosti na středoevropský region zahrnující navíc Polsko, Českou republiku, Slovensko a Slovinsko. S tím by souvisela i změna stávajícího jednacího jazyka, kterým je samozřejmě němčina, na angličtinu. Otázkou ale zůstává, kolik lidí by bylo schopno či ochotno platit členský příspěvek spojený s předplatným *Biometrics*. Po roce se navíc zdá, že myšlenka rozšíření už není tak aktuální.

Výročí založení *Biometrics* a vzniku Mezinárodní biometrické společnosti bylo připomenuto několika akcemi, v roce 1998 např. konferencí pořádanou v Edinburgu u příležitosti 50 let od založení britské sekce. Na 18. biometrické konferenci v Amsterdamu v roce 1996 byla výročí věnována přednáška Petera Armitage, předního britského statistika, který je společností svázán téměř celých 50 let; na začátku 70. let byl jejím prezidentem a později i editorem *Biometrics*. Armitage v ([1]) přibližuje s vlídným pohledem historická fakta týkající se celé společnosti a britského regionu zvláště, když popisuje dominantní roli R. A. Fishera a přínos a průkopnické úsilí dalších, jako byli F. Yates (1902-1994), M. G. Kendall (1907-1983), S. S. Wilks (1906-1964). Říká, že ‚společnost je naštěstí pořád více motivována reálnými aplikacemi než matematickou teorií‘, i když si též plně uvědomuje problémy spojené s tou částí členské základny, která má blíže k biologii. Mnoho oblastí kvantitativní biologie není mezi členy zastoupeno a *Biometrics* se postupně stala pro biology málo atraktivní, články plně soustředěné na aplikace či obsahující převahu biologických informací se postupně vytratily, když bylo recenzenty často požadováno vypuštění ‚biologické vaty‘. Takovéto problémy s vyvážeností mezi biologii a statistikou se jako červená nit táhnou vlastně celou existencí společnosti a jejich reflexe lze nalézt v mnoha materiá-

lech, např. již u zakladatele Ch. I. Blisse či v zápise z jednání výboru britského regionu (*Biometrics* 29, 1973, str. 423).

Zatímco P. Armitage se domnívá, že uvedené problémy lze překonat, v článku P. Jolicoeura [13] zaznívá silně kritický hlas, který příliš východisek nenachází. Základní námitkou je, že IBS ani její časopis *Biometrics* již vůbec neplní svou prvotní funkci společné platformy, na níž se setkávají a dorozumívají biologové a statistici. Členy společnosti je čím dál tím méně biologů a časopis je tak matematizovaný a specializovaný, že je nesrozumitelný nejen pro většinu biologů, ale též pro mnoho biometrů. Jolicoeur připomíná, že podobným vývojem prošel též časopis *Biometrika* (viz též [16]). Dále říká, že statistik by měl natolik znát biologii, aby porozuměl otázkám biologa, a biolog by měl natolik znát statistiku, aby porozuměl odpovědím statistika. Statistické metody musí mnohdy být adaptovány pro účely konkrétní aplikace a ve skutečnosti je bohužel jen málo lidí obeznámeno jak se statistikou, tak s požadovanou aplikační oblastí. Zde by právě měla spolupráci napomáhat biometrická společnost. Jolicoeur dále obšírně rozebírá, jaké reálné podoby může nabývat práce biologa a statistika na společném tématu, a poukazuje na značnou nevyváženost jejich vztahu když tvrdí, že statistici se mnohdy cítí být pouze využíváni bez patřičné kompenzace ze strany biologa, ale že často je opak pravdou - biologové jsou statistiky považováni jen za potenciální zdroj dat. Jolicoeur dále uvádí, že biolog může výrazně přispět i na poli matematickém, ačkoli není zvykem jej v této souvislosti citovat, a že závěry biologa a statistika bez znalosti problematiky mohou být dokonce protichůdné. Jako další zdroj problémů uvádí, že na biometra nemusí být v době drastických škrťů v rozpočtu peníze, protože může být vedením považován za toho, kdo neumí dobře ani biologii ani statistiku. V závěru své stati Jolicoeur z pozice francouzsky mluvícího biometra kritizuje současný stav, kdy je stále větší procento prací publikováno v angličtině.

Mnohé aspekty této uvedené kritiky si IBS nepochybně uvědomuje, jak již bylo zmíněno výše, jiné se zdají být poněkud příliš subjektivní. Komunikační problémy mezi dotýčnými obory jistě existují; z různých pohledů se jich dotýkají např. Hyams [11] (v podnětné, i když humornější podobě v souvislosti s biostatistickými konzultacemi) a Finney ([10]) (v otázce recenzování článků v biologických časopisech biometry).

Již do roku 1953 se datují snahy, aby IBS vydávala vedle *Biometrics* ještě druhý časopis, *Acta Biometrica*, který by byl zaměřený více na kvantitativní biologii a méně na teoretickou statistiku a navíc by preferoval jiné jazyky než angličtinu. Tehdy se nepodařilo vyřešit základní otázky jako ediční politiku, financování a vztah obou časopisů, nicméně o více než 40 let později se protíváha k časopisu *Biometrics* přece jen objevila. V roce 1996 začala IBS ve spolu-

práci s ASA vydávat nový časopis - *Journal of Agricultural, Biological, and Environmental Statistics* (JABES). Editor tohoto časopisu deklaruje jako hlavní účel přispívat k rozvoji a užití statistických metod v zemědělských vědách, biologických vědách včetně biotechnologií a ve vědách týkajících se životního prostředí ([12]). Je požadováno, aby publikované články měly okamžitý a praktický přínos pro statistické konzultanty a vědce pracující v aplikovaném výzkumu ve výše zmíněných oblastech.

Závěrem bychom chtěli uvést, že již citovaný P. Armitage se podílí jako editor na dvou významných publikacích s biostatistickou a biometrickou tematikou, které vydává nakladatelství Wiley. *Sborník Advances in Biometry* [2] je specifickou oslavou prvních 50 let Mezinárodní biometrické společnosti: na 500 stranách rozebírá 30 předních odborníků oblasti biometrie, kterými se zabývají, přičemž hlavní pozornost je věnována oblastem statistické metodologie, které hrály důležitou roli v biometrických aplikacích, a oblastem biologie a medicíny, které měly z těchto aplikací prospěch. Zatímco tento sborník se trochu ohlíží do minulosti, druhé dílo, *Encyclopedia of Biostatistics* [3], je dílem pro budoucnost. Jde o zcela novou publikaci referenčního, slovníkového typu, která byla vydána na jaře letošního roku a se svými skoro 5000 stranami rozloženými do 6 dílů by se mohla stát jedním z důležitých citačních pramenů a vhodným doplněním díla S. Kotze a N. L. Johnsona *Encyclopedia of Statistical Science*. Z prvních ohlasů se zdá, že jednotlivá obsáhlá hesla jsou velmi dobře zpracována.

I Mezinárodní biometrická společnost se snaží držet krok s dobou, a tak nedávno kromě nového loga zřídila také www stránku <http://www.tibs.org>, na které lze získat základní informace o společnosti, členství a vydávaných časopisech. Jsou zde také odkazy na příbuzné společnosti. Své stránky mají i oba časopisy, *Biometrics* (<http://stat.tamu.edu/Biometrics>) a *JABES* (<http://www.amstat.org/publications/jabes>).

Literatura

- [1] Armitage, P. (1996). *The Biometric Society - 50 years on*. *Biometric Bulletin* **13** (4), 3-4.
- [2] Armitage, P., David, H. A., eds. (1996). *Advances in Biometry: 50 years of the International Biometric Society*. J. Wiley & Sons, New York.
- [3] Armitage, P., Colton, T., eds. (1998). *Encyclopedia of Biostatistics*. J. Wiley & Sons, New York.
- [4] Bliss, Ch. I. (1958). *The first decade of the Biometric Society*. *Biometrics* **14**, 309-329.
- [5] Cochran, W. G., Finney, D. J. (1979). *Chester Ittner Bliss, 1899-1979*. *Biometrics* **35**, 715-717.

- [6] Cox, G. M. (1972). *The Biometric Society: The first twenty-five years (1947-1972)*. *Biometrics* **28**, 285-311.
- [7] Cox, G. M., Homeyer, P. G. (1975). *Professional and personal glimpses of George W. Snedecor*. *Biometrics* **31**, 265-301.
- [8] David, H. A. (1995). *First (?) occurrence of common terms in mathematical statistics*. *Am. Statist.* **49**, 121-133.
- [9] Feldman, W. M. (1935). *Biostatistics*. 2nd ed. Charles Griffin & Co., London.
- [10] Finney, D. J. (1997). *The responsible referee*. *Biometrics* **53**, 715-719.
- [11] Hyams, L. (1971). *The practical psychology of biostatistical consultation*. *Biometrics* **27**, 201-211.
- [12] Johnson, D. E. (1997). *An update on the status of JABES*. *Biometric Bulletin* **14** (2), str. 3.
- [13] Jolicoeur, P. (1996). *Biometry, a meeting ground between biologists and statisticians?* *Biometric Bulletin* **13** (4), str. 2.
- [14] Komenda, S. (1991). *Statisticky o statistice*. Informační bulletin CSStS **2** (3), 3-8.
- [15] Last, J.M., ed. (1995). *A dictionary of epidemiology*, 3rd edition. Oxford University Press, Oxford.
- [16] Malý, M. (1994). *Zamyšlení nad několika staršími statistickými publikacemi*. *Statistika* **31** (6), 260-265.
- [17] McGrath, Ch. (1998). *The International Biometric Society (IBS)*. Annual Report on International Statistics, vol. **5**, 89-90.
- [18] Pearson, E. S. (1967). *Studies in the history of probability and statistics XVII. Some reflections on continuity in the development of mathematical statistics, 1885-1920*. *Biometrics* **54**, 341-355.

Možnosti SPSS for Windows 7.5 - Student Version

Hana Skalská

V tomto sdělení bych ráda statistickou veřejnost informovala o produktu, jehož konkrétní stopu jsem zachytila - přiznám, po téměř tříletém pátrání a zejména pomocí náhody - až koncem loňského roku, kdy vyšel již ve verzi 7.5. Nějakou dobu trvala nezbytná korespondence a cesta a konečně zásilka dorazila. Po vyzkoušení možností si dovoluji informovat případně další zájemce na těchto stránkách.

Možnosti systému programů SPSS for Windows jsou v naší literatuře popsány [například 2, 5], většina statistiků je zná ze své praxe. Poslední verze 8.0 [4] je v distribuci u nás asi od března 1998. Využití SPSS (který je považován za standard pro statistickou analýzu dat v řadě vědních oblastí) pro účely výuky je však omezené pro jeho vysokou cenu. SPSS for Windows má, zejména pro svou jednoduchost obsluhy i kvalitu výpočetních procedur, pravděpodobně většinu atributů software který je vhodný pro výuku i prakticky zaměřených studentů. Investice do pořízení dostatečného počtu licencí je však obtížná nejen pro malou školu, ale jeho nákup ve větším počtu instalací je problematický i na velkých školách, jak o tom svědčí například [1]. Proto jednou ze zajímavých možností pro vybavení je program **SPSS 7.5 for Windows Student Version** [3]. Fakt, že byla uvedena do prodeje tato studentská verze 7.5, která má zachovánu řadu možností verze plné, může znamenat očekávaná podstatná vylepšení dalších budoucích plných verzí, z toho plynoucí určité znehodnocení této verze studentské, a to již v blízké budoucnosti. Nicméně investice do studentského software není závratná (tam, kde to umožňuje současný technický stav učebny), může zajistit podstatně vyšší kvalitu výuky v porovnání se současnými možnostmi. Navíc pro hlubší zájemce z řad studentů je dostupná cenově.

Program SPSS 7.5 for Windows Student Version pracuje pod systémem Windows 95 (případně Windows NT 3.51 nebo vyšším). Nemůže být současně nainstalován a spuštěn na jednom počítači s plnou verzí SPSS na pevném disku, nebo pouštěnou na tomtéž počítači ze sítě. Program je dodán na 10 disketách nebo CD nosiči. Po instalaci zabírá 25 MB na pevném disku.

Požadované technické parametry počítače: disketová (3,5") nebo CD mechanika, grafická karta nejméně VGA s rozlišením 640x480 nebo vyšším, procesor alespoň 80486, minimálně 8 MB RAM paměti a nejméně dalších 25 MB volného prostoru na pevném disku pro alokaci prostoru virtuální paměti k ukládání dočasných souborů.

Studentská verze programu má v porovnání s plnou verzí SPSS for Windows, modulem Base, především tato zjednodušení:

- Datový soubor může mít nejvýše 50 proměnných a 1500 pozorování
- Nelze spouštět ani vytvářet syntaktické příkazy a dávkové soubory
- Není možné spojovat soubory
- Nelze počítat agregované charakteristiky
- Help zahrnuje pouze procedury, dosažitelné z této verze
- Nelze doinstalovat žádné další výpočetní moduly. (*Plnou verzi Base systému SPSS lze rozšiřovat o vybrané další moduly, například: Professional Statistics, Advanced Statistics, Tables, Trends, Categories, Exact Tests, Conjoint, Missing Value Analysis - dle potřeb uživatele.*)

Na druhé straně, oproti plné verzi SPSS for Windows modulu Base:

- obsahuje studentská verze **navíc** činnost Exponential Smoothing z modulu Trends 7.5.

Kromě instalačních disket (nebo CD nosiče), 11 datových souborů (při dodávce na disketách to jsou soubory: Employee data, First US Congress, GSS91 political, GSS91 social, Inventor, Real estate, Relationship closeness, Tomato, United States colleges, Western Electric study, World95), je součástí dodávky manuál o rozsahu 111 stran. Tento manuál zahrnuje:

- Stručný průvodce programem a popis helpu. Obsahuje přehled možností SPSS Base Student Version pro analýzu dat a grafické zobrazení. Nejsou vysvětlovány výsledky výpočetních procedur.
- Popis editoru datového pole
- Popis vytvoření a editace grafu, jeho vložení do jiné aplikace
- Změny proměnné překódováním, vytvoření nové proměnné
- Popis práce se syntaktickými příkazy.
- Popis zadání výpočtu statistických procedur. Tomu je věnováno pouze 16 stran, ale postačí pro pochopení, jak provést analýzu dat. Analýzu je možné provést kromě na celém souboru také na jeho podmnožinách, vymezených dalším znakem.

Ve studentské verzi jsou **obsaženy všechny výpočetní procedury plné verze modulu Base**, tj. tyto statistické procedury: četnosti (absolutní, relativní, kumulativní), popisné statistiky, explorační techniky, četnostní tabulky (mnohorozměrná třídění), jednovýběrové a dvouvýběrové testy shody průměrů, párový t test, ANOVA, jednoduchý faktorový zobecněný lineární model, lineární a nelineární regrese, shluková a diskriminační analýza, faktorová analýza, neparametrické testy a **navíc: exponenciální vyrovnání časové řady**. **Není možné** agregování údajů, ani spojování souborů.

Ke grafickému zobrazení výsledků lze vybrat ze 16 typů grafů (stejně možnosti jaké má plná verze), z nichž většina má další varianty, takže celkový počet možností vytvořených grafů je daleko větší.

Velikou výhodou studentské verze jsou zachované solidní možnosti práce s datovým polem. Kromě definování nového souboru, možnosti importu a exportu z jiných formátů nebo do jiných formátů (například .sys, .xls - Excel, .dbf, .tab, .w* - Lotus), čtení dat z ASCII souboru ve volném nebo pevném formátu, lze definovat nové proměnné. Nově definované proměnné lze zadat pomocí více než 100 funkcí aritmetických, statistických, datumových nebo pracujících se znaky, které mohou být kombinovány do transformačních výrazů aritmetickými, logickými nebo relačními operacemi. Tento proces je z uživatelského pohledu velmi jednoduchý. Spolu s možností rozdělení datového souboru na podmnožiny podle dalšího znaku a s možností výběru podmnožin dat pro analýzu dobře naznačuje, jakým nástrojem je plná verze SPSS Base. Výstupní soubor lze uložit ve formátu .html. Grafické výstupy lze editovat a ukládat ve formátech: .jpg, .wmf, .bmp, .cgm, .eps, .tif, .pct.

Porovnání s plnou verzí SPSS for Windows je pro studentskou verzi velmi příznivé (instalace vyzkoušena na počítači s procesorem Pentium 75, 32 MB RAM), neboť:

- Jsou funkční všechny činnosti které usnadňují práci se souborem, se zadáváním analýz a prezentací výsledků. Pouze nelze provádět agregaci dat, nelze spojovat soubory. Statistický průvodce (Statistical Coach) popisuje možnosti analýz a nahrazuje částečně velmi stručný manuál (při používání však občas došlo ke kritické chybě systému).
- Je shodná práce s navigátorem výstupu.
- Objekty výstupního souboru lze přenášet postupně přes "Copy → Edit → Vložit jinak" do dalších aplikací.
- Jsou zachovány rozsáhlé možnosti editování grafů i dalších výstupních objektů (tabulek).
- Grafy je možné exportovat mimo jiné také do formátů .jpg, .bmp, .tif, .html.
- Pomocí technologie pivočních prvků je možné přeskupit výslednou tabulku (obr. 1). V tabulce na obrázku jsou tři pivoční prvky, jejichž přemístováním se tabulka nově uspořádá. To umožňuje provést definitivní uspořádání výstupu na již vytvořené tabulce a usnadňuje práci s prezentací výsledku. Uživatel si vybere variantu, která nejvíce vyhovuje jeho představě o zobrazení výsledku. Není nutné znova provádět výpočty ani převádět tabulku pro další manipulaci například do tabulkového procesoru. Vzhled tabulky (například formáty čísel a písma, použité statistiky, texty, šířky sloupců) lze editovat.

Obr.1: Ukázka změny uspořádání výstupní tabulky, přetažením ikony jednoho ze tří pivotních prvků: Statistics.

a) Původní výstup (použita data WORLD95.sav) - lze rolovat a zobrazit celou tabulku

The screenshot shows two windows. On the left is a 'Report' window displaying a pivot table titled 'Birth rate per 1000 people'. The table has four rows representing regions: OECD, East Europe, Pacific/Asia, and Africa, with a fifth row for 'Middle East'. Each row contains three columns: 'Mean', 'N', and 'Std. Deviation'. On the right is the 'Pivoting Trays1' dialog box, which shows a 'Layers' tray containing 'Row', 'Column', and 'Statistics' elements. The 'Statistics' element is currently positioned in the 'Row' tray.

Region	Mean	N	Std. Deviation
OECD	12,852	21	1,858
East Europe	13,429	14	1,284
Pacific/Asia	26,271	17	11,886
Africa	42,000	19	6,412
Middle East	32,706	17	8,275

Výstup upravený přesunutím aktivního prvku Statistics do sloupců:

The screenshot shows the same two windows as above, but with the 'Statistics' element moved from the 'Row' tray to the 'Column' tray in the 'Pivoting Trays1' dialog box. The 'Report' window now displays a pivot table where the 'Region or economic' categories are in the first column, and the statistical measures ('Mean', 'N', 'Std. Deviation') are in the subsequent columns. The 'Total' row is added at the bottom.

Region or economic	Mean	N	Std. Deviation
OECD	12,852	21	1,858
East Europe	13,429	14	1,284
Pacific/Asia	26,271	17	11,886
Africa	42,000	19	6,412
Middle East	32,706	17	8,275
Latin America	26,805	21	6,804
Total	25,822	109	12,381

volby při výběru statistického software pro výuku, v podmínkách většiny škol. Jedná se vlastně o knižní vydání. Není to ideální řešení. Proto na škole, nejlépe na učebně, by měla být současně alespoň jedna instalace plné verze včetně všech dalších modulů.

Pro prakticky zaměřenou výuku lze tuto studentskou verzi považovat za více než vyhovující, protože je dostupná v našich podmínkách i pro domácí potřeby studentů (možnost analýzy vlastních dat pro závěrečné nebo seminární práce na domácím počítači). Protože jsou zachovány principy ovládání plné verze, lze na plnou verzi bez obtíží přejít, takže studenti jsou lépe připraveni i na eventuální požadavky budoucího zaměstnavatele. Přestože původní určení pro studentské užití je trochu jiné než námi zamýšlené, rozhodli jsme se k nákupu této verze pro celou učebnu a jejím využití nejen k výuce statistiky, ale také pro potřeby dalších předmětů, které mají statistickou metodologii ve svém vědním základu. Zároveň ji doporučíme studentům jako základní literaturu.

Nejvýraznější odchylky studentské verze oproti plné verzi:

- výrazná redukce rozsahu analyzovaného datového souboru,
- nemožnost používání syntaktických příkazů
- nemožnost získat protokol pracovního postupu kterým analýza probíhala

nejsou příliš omezující pro výuku ani pro většinu běžných analýz, které studenti budou provádět. Tato omezení jsou bohatě vyvážena výhodami (dostupnost, kvalitní a spolehlivé statistické procedury, možnost seznámit se s tímto rozšířeným výpočetním prostředím).

Poměr efekt/cena je při srovnání s plnou verzí (k použití v základních kurzech statistiky) jistě velmi vysoký. Cena za licenci na 3 instalace plně funkčního produktu SPSS Base 7.5 (bez dalších modulů) při uplatnění slevy 10 % pro školy, odpovídá přibližně ceně na nákup 30 studentských verzí, přičemž funkční omezení není pro potřeby naší výuky závažné.

Na závěr, pro možnost porovnání, uvádím vybrané položky z dostupného ceníku, který byl platný pro ČR v březnu 1998. V položce **Obnova** je uvedena roční sazba, placená podle zvlášť uzavírané smlouvy o automa-tickém upgrade. Tuto smlouvu lze uzavřít minimálně na tři roky, první rok je bezplatný. Po dobu trvání smlouvy má uživatel zajištěnou aktualizaci těch modulů, na které uzavřel smlouvu o obnově. Může se však dostat do situace, že po aktualizaci pouze některých modulů nelze používat celý, původně nakoupený systém protože verze po upgrade "nespolupracuje" se staršími moduly.

Ceny SPSS for Windows v. 7.5 (březen 1998) - plná verze			
Popis sestavy	Počet instalací	Cena (Kč)	Obnova(Kč)
Pouze modul Base	3	72770	13100
Pouze modul Base	10	177870	32020
Modul Base + 7 dalších modulů	1	139940	56680

Jednotlivé moduly (Tables, Professional statistics, Advanced statistics, Exact test, Categories, Conjoint, Trends, Missind value analysis)	Každý z modulů při 1 instalaci	17930	3230
	každý z modulů při 10 instalacích	98620	17750
Dále existuje možnost zakoupit tzv. akademickou licenci na pět let, s každoroční nájmní smlouvou. Jedna fakulta zaplatí pro potřeby výuky a výzkumu, pro neomezený počet instalací (s možností dalšího pronájmu studentům a zaměstnancům za cenu, kterou si sama stanoví), za plný systém SPSS cca 8000 US\$, při sdružení více fakult (např. jedné univerzity) se cena sjednává - jako příklad bylo uvedeno sdružení dvou fakult: 10000 US\$.			
SPSS for Windows 7.5 - Student Version			
Každá jednotlivá instalace 65 US\$.			
Lze uplatnit množstevní slevu cca 15 % při nákupu 15-20 kusů			

Literatura:

1. Malý M.: Seminář ke 100 letům státní statistiky a statistický den v Liberci. Informační bulletin ČSS, 1997, č. 2, s. 20-22.
2. Řezanková H., Žváček J.: Poradce, čaroděj a internet. SPSS 7.5 for Windows. CHIP, č.11, 1997, s. 144 -146
3. SPSS 7.5 for Windows. Brief Guide. SPSS Inc., Prentice Hall, Inc., Simon&Schuster/A Viacom Company, New Jersey, 1997
4. SPSS Base 8.0. Applications Guide. SPSS Inc., Chicago, 1998
5. Žváček J.: Na statistické frontě klid. SPSS for Windows od verze 5 do verze 6.1. CHIP, č. 7, 1995, s. 96 - 99

EXCEL, STATISTIKA, LOKALIZACE a ZMATEK

Josef Tvrđík

Následující řádky se zabývají některými statistickými funkcemi a moduly doplňku Analýza dat v české lokalizaci populárního tabulkového procesoru Microsoft Excel for Windows 95, verze 7a. Příspěvek byl přednesen na Statistických dnech pořádaných Českou statistickou společností a Jihočeskou univerzitou v Českých Budějovicích ve dnech 16. a 17. září 1998.

Cílem příspěvku není znevažovat užitečnost statistických nástrojů dodávaných jako součást Excelu. Statistické funkce v něm implementované i doplněk Analýza dat jsou samozřejmě užitečné nástroje, které ve spojení s běžnými funkcemi tabulkového procesoru mohou usnadnit řešení řady praktických úloh analýzy dat. Svým rozsahem pokrývají naprostou většinu statistických metod, které jsou vyučovány v uživatelsky orientovaných kursech základů statistiky. Problematický je však někdy způsob implementace a zejména české lokalizace produktu. Vzhledem k velkému rozšíření Excelu a také vzhledem k tomu, že jeho statistické nástroje jsou často užívány především lidmi bez větších zkušeností v analýze dat, mohou být právě nedostatky implementace a lokalizace příčinou nesprávných úsudků a zmatení.

Od této chvíle si představte, že autor těchto řádků se ujímá role šaška, třeba si ho představte v šaškovské čepici. Ne proto, že následující text je tak veselý, ale proto, že chce ukázat na věci, které nejsou v pořádku a při tom náprava těchto nepořádků je na jiných a mocnějších. Pokud zrovna mezi mocnější nepatříte, tak alespoň dobrou zábavu při dalším čtení.

Názvy některých statistických funkcí i jednotlivých modulů doplňku Analýza dat jsou lokalizovány, tj. přeloženy do češtiny. Výsledkem jsou mimo jiné i názvy funkcí SMODCH a SMODCH.VÝBĚR. Toto počestění názvů funkcí stěží výrazně pomůžce českému uživateli. V menu doplňku Analýza dat nalezneme položku „*Pohyblivý průměr*“, ačkoli užívaný český termín je *klouzavý průměr*.

Mezi statistickými funkcemi nalezneme i několik užitečných funkcí pro výpočet hodnot distribučních funkcí a kvantilů často užívaných rozdělení. Jedna z nich se jmenuje NORMDIST a z jejího helpu (texty citací helpů jsou ohraničeny oddělovači *nápověda*., *konec nápovědy*, některé texty jsou zkráceny, krácení je vyznačeno tečkami, zvýraznění tučným písmem je dílem autora-šaška) se dočteme následující:

nápověda: _____

NORMDIST

Vrací **kumulativní normální rozdělení** se zadanou střední hodnotou a směrodatnou odchylkou. Tato funkce má ve statistice velmi široké použití, včetně testování hypotéz.

Syntaxe

NORMDIST(x; průměr; směrod_odch; kumulativní)

X je **hodnota, pro niž počítáme rozdělení.**

Průměr je **aritmetický průměr rozdělení.**

Směrod_odch je směrodatná odchylka rozdělení.

Kumulativní je logická hodnota, která určuje tvar funkce. Pokud kumulativní je PRAVDA, NORMDIST vrací kumulativní distribuční funkci; je-li NEPRAVDA, vrací **pravděpodobnostní míru.**

Poznámky

....

Pokud průměr = 0 a směrod_odch = 1, NORMDIST vrací standardní normální rozdělení, NORMSDIST.

Příklad

NORMDIST(42;40;1,5;PRAVDA) se rovná 0,908789

_____ *konec nápovědy.*

Funkce NORMDIST jen stěží může vracet „*kumulativní normální rozdělení*“, ale z popisu lze vytušit, že tím je míněna hodnota *distribuční funkce* nebo *hustoty* (nikoli „*pravděpodobnostní míra*“) normálního rozdělení podle toho, jakou zadáme hodnotu posledního vstupního parametru „*kumulativní*“. Druhý parametr je vysvětlen jako „*aritmetický průměr rozdělení*“, což patrně vzniklo chybným překladem anglického termínu *mean*, který měl být přeložen jako *střední hodnota*. Nicméně se dočteme, že pro „*standardní normální rozdělení*“ (česky se říká *normalizované normální rozdělení*) můžeme použít funkci NORMSDIST, která funguje zcela podle našeho očekávání, $NORMSDIST(1.96) = 0.975002$. Podobně řádně se chová i inverzní funkce NORMSINV, neboť pro zadanou hodnotu distribuční funkce vrátí správnou hodnotu kvantilu, např.

$NORMSINV(0.025) = -1.95996$.

Povzbuzeni tímto úspěchem zkusíme ještě kvantily *t*-rozdělení, které očekáváme pod funkcí s názvem TINV, její druhý parametr je počet stupňů volnosti. K našemu překvapení však zjistíme, že $TINV(0.025, 500) = +2.248171$, ačkoli bychom očekávali hodnotu blízkou -1.96, tj. blízkou tomuto kvantilu normovaného normálního rozdělení. Na další pokus můžeme nalézt hodnotu kvantilu podobnou očekávané alespoň co do absolutní

hodnoty, $TINV(0.05,500) = +1.964718$. Lehce znepokojeni nahlédneme do helpu funkce TINV a dočteme se:

nápověda: _____

TINV

Vrací **inverzní funkci k funkci TDIST** pro dané stupně volnosti.

Syntaxe

TINV(prst; volnost)

Prst je **pravděpodobnost daného dvojstranného t-rozdělení**.

Volnost je počet stupňů volnosti.

Poznámky:

Pokud není některý z argumentů numerický, vrací funkce TINV chybovou hodnotu #HODNOTA!.

Pokud je prst < 0 nebo pokud je prst > 1, vrací TINV chybovou hodnotu #NUM!.

Pokud není argument volnost celé číslo, je na celé číslo převeden.

Pokud je volnost < 1, vrací TINV chybovou hodnotu #NUM!.

Funkce TINV se počítá jako $TINV=p(t < X)$, kde **X je náhodná proměnná, která doprovází t-rozdělení**.

Funkce TINV používá **opakující se techniku propočítávání funkce**. Se zadanou pravděpodobnostní hodnotou se funkce TINV opakuje dokud není výsledek přesný na $\pm 3 \times 10^{-7}$. Pokud funkce TINV nedosáhne požadovaného výsledku po 100 opakováních, vrací funkce chybovou hodnotu #N/A.

Příklad:

TINV(0,054645;60) se rovná 1,96

_____ *konec nápovědy.*

Některé formulace z nápovědy nás možná pobavily, některé trochu znervózňily, např. „pravděpodobnost daného dvojstranného t-rozdělení“. Co to vůbec je *pravdě-podobnost* nějakého *rozdělení* a co se může skrývat pod „dvojstranným“ t-rozděle-ním? Nicméně je jasné, že klíčem k pochopení je zjistit, k jaké že to funkci je funkce TINV inverzní a zde je uvedeno, že k funkci TDIST. Z helpu funkce TDIST zjistíme toto:

nápověda: _____

TDIST

Vrátí hodnotu **distribuční funkce t** Studentova rozdělení. V případě, že neznáme směrodatnou odchylku základního souboru, je ji možno odhadnout pomocí výběrové směrodatné odchylky t. T-rozdělení je používáno při **hypotetickém testování malých vzorků dat**.

Syntaxe

TDIST(x; volnost; strany)

X je číslo, pro které hledáme hodnotu distribuční funkce.

Volnost je celé číslo, označující počet stupňů volnosti.

Strany určuje, zda se jedná o **jednostranné** či **dvoustranné rozdělení**. Pokud je parametr strany = 1, vrací TDIST hodnotu funkce **jednostranného** rozdělení. Pokud je parametr strany = 2, vrací TDIST hodnotu funkce **dvostranného** rozdělení.

Poznámky:

Pokud není argument numerický, vrací funkce TDIST chybovou hodnotu #HODNOTA!

Pokud je volnost < 1, vrací TDIST chybovou hodnotu #NUM!.

Argumenty volnost a strany jsou převáděny na celá čísla.

Pokud argument strany nabývá jiných hodnot než 1 nebo 2, vrací TDIST chybovou hodnotu #NUM!.

Funkce TDIST se počítá jako $TDIST=p(x < X)$, kde X je náhodná proměnná, která doprovází t-rozdělení.

Příklad:

TDIST(1,96;60,2) se rovná 0,054645

_____ *konec nápovědy.*

Naše dilema se nijak nezmenšilo, podle nápovědy se obě funkce počítají stejně, $TDIST=p(x < X)$ a $TINV=p(t < X)$, obě funkce mají být zřejmě nějaké pravděpodobnosti. Navíc TDIST jsou vlastně funkce dvě, vybíráme jednu z nich zadáním hodnoty jejího třetího vstupního parametru „strany“. Ke které z nich je TINV inverzní? Naštěstí z uvedených příkladů a nápověd můžeme my, co jsme v České statistické společnosti, usoudit téměř s jistotou, že platí následující vztah:

$$TINV(\alpha, n) = t_n(1 - \alpha / 2),$$

kde $t_n(1 - \alpha / 2)$ je $(1 - \alpha / 2)$ -kvantil t -rozdělení s n stupni, takže nezáporné hodnoty kvantilů umíme pomocí funkce TINV vyčíslit a to, že t -rozdělení je symetrické, snad není nutné připomínat, takže na kvantily t -rozdělení se umíme dostat i v Excelu.

Nápovědu ke statistickým funkcím však bude lépe brát trochu s nadhledem, viz např. nápovědu k funkci LINREGRESE:

nápověda: _____

LINREGRESE

Metodou nejmenších čtverců spočte přímku, která nejlépe odpovídá zadaným datům, a vrací pole, které tuto přímku popisuje. Vztah pro tuto přímku je:

$$y = m_1x_1 + m_2x_2 + \dots + b \text{ nebo } y = mx + b$$

kde závislá hodnota y je funkcí nezávislých hodnot x . Hodnoty m jsou koeficienty odpovídající každé z hodnot x , b je konstanta. Všimněte si, že y , x a m mohou být vektory. Pole, které je výsledkem funkce LINREGRESE, je $\{m_n, m_{n-1}, \dots, m_1, b\}$. LINREGRESE může také vracet další regresní statistiky.

Syntaxe

LINREGRESE(známá_y; známá_x; konstanta; stats)

Známá_y je množina hodnot y , které **již znáte podle vztahu $y = mx + b$** .

....

Známá_x je nepovinná množina hodnot x , které **již můžete znát podle vztahu $y = mx + b$** .

.....

Konstanta je logická hodnota udávající, zda má být konstanta b rovna 0.

Pokud konstanta je PRAVDA nebo není uvedena, b se počítá běžným způsobem.

Pokud konstanta je NEPRAVDA, b se položí rovno 0 a hodnoty m se upraví tak, aby platilo $y = mx$.

Stats je logická hodnota udávající, zda chceme zjistit další regresní statistiky.

Pokud stats je PRAVDA, LINREGRESE vrací další regresní statistiky, takže výsledné pole je

$\{m_n, m_{n-1}, \dots, m_1, b; sen, sen-1, \dots, se_1, seb; r_2, sey; F, df; ssreg, ssresid\}$.

.....

Dodatečné regresní statistiky jsou:

Statistika Popis

....

F F statistika, nebo pozorovaná hodnota F. F statistiku lze použít pro rozhodnutí, zda vztah mezi závislými a nezávislými proměnnými není nahodilý.

df Stupně volnosti. **Pomocí stupňů volnosti lze nalézt kritické hodnoty F ve statistické tabulce.** Porovnáním hodnot z tabulky s F statistikou, kterou vrací funkce LINREGRESE, lze určit úroveň spolehlivosti modelu.

konec nápovědy.

Jisté rozpaky vzbuzují i některé moduly doplňku Analýza dat. Tak např. *Popisná statistika* poskytuje následující výsledky:

vaha		delka	
stř. hodnota	99.3956	stř. hodnota	923
chyba stř. hodnoty	2.743841	chyba stř. hodnoty	2.885173
medián	99	medián	110
modus	101	modus	110
sm. odchylka	26.17458	sm. odchylka	27.5228
rozptyl výběru	685.1084	rozptyl výběru	757.5043
špičatost	0.194895	špičatost	0.294819
šikmost	0.164807	šikmost	0.140619
rozdíl max-min	131	rozdíl max-min	146
minimum	40	minimum	45
maximum	171	maximum	191
součet	9045	součet	10073
počet	91	počet	91
věrohodnost (95.0%)	5.451115	věrohodnost (95.0%)	5.731895

Pomiňme, že při zadání více než jedné veličiny dostaneme výstup, ve kterém jsou tabulky s jednorozměrnými statistikami zřetězeny vedle sebe i s opakujícími se názvy statistik, takže výstup je zbytečně rozsáhlý a tím i nepřehledný. Zvědavost vzbuzuje statistika označená „*věrohodnost*“, dokonce tak velkou, že jsem sáhl po manuálu (User's Guide, Microsoft Excel, version 5.0). Tam však je jen odkaz na on-line help a v něm jsem našel jen toto:

nápověda: _____

Věrohodnost-hladina významnosti

Chcete-li ve výstupní tabulce zobrazit řádek pro hladinu spolehlivosti střední hodnoty, zaškrtněte toto pole. Do textového pole zadejte požadovanou hladinu spolehlivosti. Např. při hodnotě 95% se počítá hladina spolehlivosti pro střední hodnotu s významností 5%.

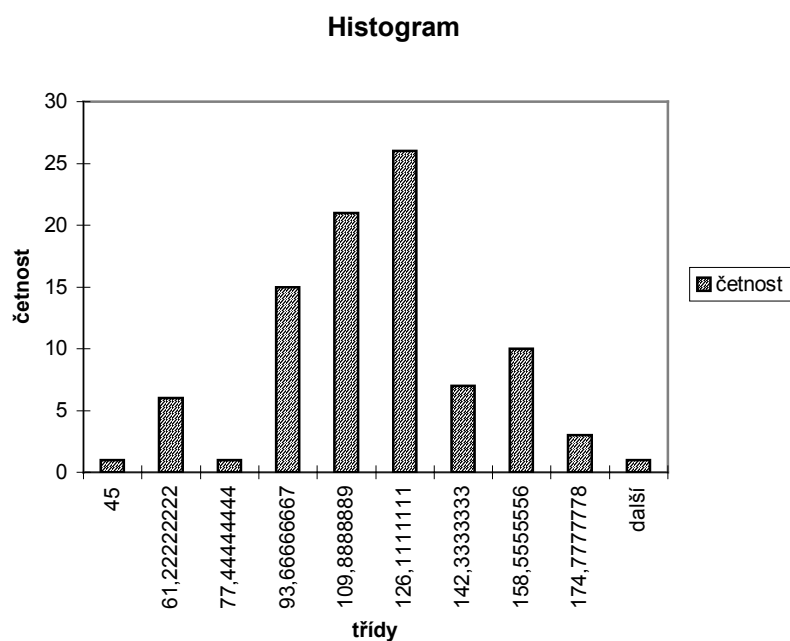
_____ *konec nápovědy.*

Z několika příkladů ale zjistíme, že asi platí vztah

$$\frac{\text{věrohodnost}}{\text{chyba stř. hodnoty}} = t_{n-1}(1 - \alpha / 2),$$

ale mám pochybnosti, že by se v nějaké české knize pro charakteristiku vyhovující tomuto vztahu používal termín *věrohodnost*.

Často užívaným modulem doplňku Analýzy dat je nepochybně *Histogram*. S využitím implicitního nastavení vstupních parametrů můžete dostat následující obrázek: Drobné vady na kráse histogramu lze přehlédnout. Sloupce nejsou nad celou šířkou intervalů a počet významných číslic v popisu pod sloupci je nesmyslně velký. To patrně lze napravit vhodnější volbou vstupních parametrů. Závažnějším nedostatkem však je, že hodnoty popisující středy sloupců (středy jednotlivých intervalů) nejsou hodnoty odpovídající středu, ale pravému okraji intervalu.



Roztomilosti nalezneme i v modulech pro běžné statistické testy. Např. dvouvýběrový *t*-test poskytne následující výstup:

Dvouvýběrový t-test s rovností rozptylů

	Soubor 1	Soubor 2
stř. hodnota	111.9219	107.7778
Rozptyl	734.0097	831.0256
Pozorování	64	27
Společný rozptyl	762.3514	
Hyp. rozdíl st. hodnot	0	
Rozdíl	89	
t stat	0.654039	
P(T<=t) (1)	0.257387	
t krit (1)	1.662156	
P(T<=t) (2)	0.514773	
t krit (2)	1.986978	

Domnívám se, že pro uživatele rozlišujícího mezi jednostranným a oboustranným testem je výstup redundantní, uživateli mezi těmito variantami nerozlišujícímu tato redundance stejně nepomůže. Můj zájem vzbudila statistika označená jako „*rozdíl*“. Skutečnost, že platí $rozdil = n_1 + n_2 - 2$ (tedy je roven počtu stupňů volnosti) svádí k domněnce, že lokalizátor (to není nadávka, to je označení člověka provádějícího lokalizaci) zkratku *df* interpretoval jako anglické *difference* a pak pěkně přeložil do češtiny.

Smyslem tohoto povídání nebylo demonstrovat, že jsem šikovný chlapec, kterému stačí 30 let statistické praxe a trocha obezřetnosti k celkem spolehlivému řešení jednoduchých úloh analýzy dat v Excelu. Právě teď svou šaškovskou čepici odhazují v dál a po uvedených příkladech si troufám tvrdit, že Excel a zejména jeho česká lokalizace má *negativní vliv na uživatele statistiky* v tom, že činí aplikace statistiky obtížnější, než je nutné, a kromě toho kazí slovník uživatele a uživatelovo konceptuální vnímání statistiky jako oboru. Domnívám se, že tento negativní vliv na uživatele statistiky působí proti cílům, které má Česká statistická společnost a tudíž by se tato společnost měla ozvat a pokusit se vyvolat tlak na nápravu. Nalezené nedostatky jsou dvojího druhu. Jednak se týkají designu a implementace programu, za ty zřejmě zodpovídá Microsoft Corporation, jednak se týkají české lokalizace. Ale o tom, kdo je za ni odpovědný, se můžeme jen dohadovat, neboť v informaci o aplikaci Microsoft Excel není o autorech české lokalizace ani zmínka.

Českobudějovické statistické dny

Marek Malý

Ve dnech 16.-17. září 1998 uspořádala Česká statistická společnost ve spolupráci s Jihočeskou univerzitou v Českých Budějovicích Českobudějovické statistické dny, jejichž hlavním tématem byla aplikace statistiky v přírodních vědách.

Program prvního dne byl zahájen úvodní přednáškou předsedy České statistické společnosti ing. Z.Rotha, CSc., která se zabývala tím, jak statistické modely odpovídají skutečnosti a do jaké míry v sobě zahrnují chybu pozorování. Je důležité, aby byl statistický model adekvátní reálné situaci a aby ji vhodně, i když aproximativně popisoval. Ing. Roth poukázal speciálně na situace, kdy se cílový jev měří pomocí několika jiných zástupných jevů, které se zkoumaným jevem souvisejí či jsou jeho součástí. Na příkladě IQ ilustroval rozdíl mezi cílovým jevem a k jeho popisu použitými náhodnými proměnnými, které zobrazují přidružené (korelované) jevy.

Hlavní zvanou přednášku na téma „Explained variation in Cox and logistic regression“ proslovil Prof. Dr. Michael Schemper, vedoucí oddělení klinické biometrie v Institutu pro aplikaci počítačů v medicíně (Institut für Medizinische Computerwissenschaften) univerzity ve Vídni (viz [www stránku http://www.akh-wien.ac.at/imc/biometrie/stat.htm.en.htm](http://www.akh-wien.ac.at/imc/biometrie/stat.htm.en.htm)). Prof. Schemper je velmi známým odborníkem v oblasti biometrie a zejména modelů přežívání. Je mimo jiné členem výboru Mezinárodní společnosti pro klinickou biostatistiku a do loňského roku byl též předsedou rakousko-švýcarské sekce Mezinárodní biometrické společnosti. Ve své přednášce prof. Schemper vyšel z analogie měř vysvětlené variability, které se užívají v logistické analýze a analýze přežívání, s koeficientem determinace v zobecněném lineárním modelu. Popsal a kategorizoval mnoho různých návrhů takových měř a uvedl kritéria pro jejich volbu. Věnoval se též otázce malých výběrů a konfidenčních intervalů a celou problematiku ilustroval na praktických příkladech. Pro zájemce o podrobnější informaci uvádíme, že určitá část jeho přednášky vycházela z článků

– Schemper, M. (1990). *The explained variation in proportional hazards regression*. *Biometrika* **77**, 216-218. Correction: *Biometrika* **81**, 631 (1994).

- Mittlböck, M., Schemper, M. (1996). *Explained Variation for Logistic Regression*. *Statistics in Medicine* **15** (19), 1987-1997
- Schemper, M., Stare, J. (1996). *Explained Variation in Survival Analysis*. *Statistics in Medicine* **15** (19), 1999-2012.

V polední přestávce navštívili účastníci statistických dní Akademii věd ČR, kde díky péči doc. RNDr. Pavla Kindlmana, CSc. a dalších pracovníků shlédli prezentaci několika výzkumných projektů. Program poté pokračoval přednáškami účastníků, které pokrývaly velmi široké spektrum problémů, jak dokumentuje jejich následující přehled:

- RNDr. Petr Šmilauer (Jihočeská univerzita České Budějovice): *Recent developments in multivariate analysis of ecological data*
- Ing. Jan Honner (Český statistický úřad České Budějovice): *Úkoly státní statistiky v regionálním pohledu*
- Ing. Milan Palát, CSc. (Mendelova zemědělská a lesnická univerzita Brno): *Využití statistických metod při výzkumu lesních ekosystémů*
- Ing. Eva Jarošová, CSc. (VŠE Praha), RNDr. Marek Malý, CSc. (Státní zdravotní ústav Praha): *Analýza faktorů ovlivňujících dobu do vzniku komplikací diabetu*
- Ing. Josef Tvrdlík, CSc. (Ostravská univerzita): *Excel, statistika, lokalizace a zmatek*

Poslední přednáška, ač veseleji laděná, upozornila na velmi závažný problém naprosto nevhodného používání statistické terminologie včetně zavádění nových překladů již ustálených pojmů v české lokalizaci programu Excel. Program prvního dne byl uzavřen zdařilým společenským večerem.

Druhý den akce probíhal v počítačové učebně a byl věnován prezentaci a krátkému praktickému vyzkoušení programových balíků STATISTICA (RNDr. M. Kletečková) a UNISTAT (Ing. M. Palát, CSc.). Českobudějovických statistických dní se zúčastnilo přes 25 odborníků z velmi různorodých vysokých škol, ústavů a úřadů z Brna, Českých Budějovic, Hradce Králové, Jindřichova Hradce, Ostravy, Prahy, Ústí nad Labem a Vídně. O nelehkou organizační stránku semináře se postarali dr. M. Kletečková, dr. I. Dostálková, dr. V. Petrášková a ing. M. Maršík ze zemědělské, pedagogické a biologické fakulty českobudějovické Jihočeské univerzity.

Česká statistická společnost hodlá pokračovat v tradici statistických dní a pořádat i další jedno- i vícedenní semináře. Výbor společnosti proto uvítá podnětné návrhy témat i jakoukoli další vhodnou iniciativu z řad členské základny. Bude-

me rádi, pokud příští akce budou ještě početněji navštíveny a budou tak moci lépe přispět k výměně informací v rámci statistické obce.

<i>Marek Malý</i> , Padesát let Mezinárodní biometrické společnosti.....	1
<i>Hana Skalská</i> , Možnosti SPSS for Windows 7.5 - Student Version.....	7
<i>Josef Tvrdlík</i> , Excel, statistika, lokalizace a zmatek.....	13
<i>Marek Malý</i> , Českobudějovické statistické dny	21

IX. VÝROČNÍ SCHŮZE ČLENŮ ČESKÉ STATISTICKÉ SPOLEČNOSTI

se bude konat ve čtvrtek, 4. února 1999 od 13.00

v budově VŠE, nám. W. Churchilla 4, Praha 3.

Na programu bude zpráva předsedy, zpráva o hospodaření společnosti, volby výboru společnosti a odborný referát.

Informační Bulletin České statistické společnosti vychází čtyřikrát do roka v českém vydání.
 Předseda společnosti: Ing. Zdeněk Roth, Csc., SZÚ Praha, MSP Šrobárova 48, 100 42 Praha 10, e-mail: szumsp@earn.cvut.cz. ISSN
 1210 – 8022
 Redakce: RNDr. Gejza Dohnal, Csc., Jeronýmova 7, 130 00 Praha 3, e-mail: dohnal@fsik.cvut.cz.