

Informační Bulletin



České Statistické Společnosti

číslo 2., červenec 1992, ročník 3.

Pár slov o příbuznosti nepodobných.

Stanislav Komenda

Najde se jen málo věcí, v nichž se názory lidí shodnou v souhlase téměř naprostém. Napadají mě dvě. Totiž shoda v tom, že meteorologové se mýlí a statistikové lžou – a co jim navíc přitěžuje – lžou velice nudně.

O tom posledním chtěl bych říci několik slov.

V mnoha různých obměnách a variantách kolují v informačním řečišti člověčenstva charakteristiky typu (citováno podle H. Svobody: *Moderní statistika*): „Jsou tři druhy lži – lži, odsouzeníhodné lži a statistiky“. Vzhledem k tomu, že věc mizerné pověsti nedokázala vyvodit z toho, co se o ní traduje, přiměřené poučení a houževnatě trvá na svém právu pobývat ve společenství věd s pověstí slušnou anebo jen nepatrně pošramocenou, jako jsou například (uvádím namátkou) otorhinolaringologie nebo sexuologie, nabízí se otázka, v čem to vězí. Že by prolhaní statistikové dokázali zfalšovat i statistiky, podle kterých je statistika používána v biomedicínských vědách stále častěji?

Autor těchto řádek je statistik, takže člověk nikoli nezávislý a rozhodně předpojatý. To pochopitelně už předem snižuje hodnověrnost jeho postojů, tvrzení a výpovědí. Učinně tedy gentlemanskou dohodu – spočívající v tom, že on se nebude vyjadřovat úplně jako statistik, a čtenář (pokud se najde někdo ochotný tyhle řádky přečíst) bude jeho slovům alespoň věřit. Řekněme asi tak z jedné třetiny. Pak se uvidí.

Svět, ve kterém žijeme, je plný absurdit. Obecně soudíme, že statistická tabulka je vrcholem nudy. Tomu, kdo si před spaním v posteli listuje ve statistikách nemocnosti, bychom asi ochotně přiznali nárok na psychiatrickou léčbu, a to, že dosud léčen není, si vysvětlujeme domněnkou, že nejbližší léčebna má plně obsazeno.

Člověk, který se nudí, půjde si hrát. Anebo alespoň hře přihlížet. Hrát kopanou nebo se na hraní kopané koukat je pro spoustu lidí spolehlivým prostředkem, jak se zabavit. Hra je pro nás tedy něco jako opak, jako protipól statistiky.

Jenomže – statistik podobně jako účastník hry – oba se odvolávají na stejné. Dovolávají

Typeset by \LaTeX

se toho, čemu, v dojemné shodě říkají – Náhoda. Tabulky úmrtnosti stejně jako ukazatele skóre jsou náhodou nepochybně ovlivňovány. V tomto smyslu jsou tedy Statistika – nudná a truchlivá tetka, a Hra – rozverná holčička, ochotně svolná pobavit a rozptýlit, blízké příbuzné. Možná dokonce sestry. Sestřičky rodné právě tím společným chromozómem Náhody.

Jen tak mimochodem – možná jste četli jednu z nejlepších her Voskovce a Wericha – Osel a stín. A snad máte možnost sáhnout si pro ni do regálu knihovny. V tom případě tak učinite. Je to hra s velkým „H“. Na její úvodní stránce můžete najít věnování: Svě Múze, Karlu Schönbaumovi, vděčně – V+W. V pamětech národního umělce Jana Wericha (Jan Werich vzpomíná ... vlastně potlach, Melantrich, Praha 1982) se najde na straně 46 vysvětlující text „... s Voskovcem jsme napsali hru Osel a stín na Lukianovu anekdotu tak přibližně 2000 let starou. Dal nám k tomu popud náš přítel a kamarád a znamenitý člověk, Karlíček Schönbaum, správně JUDr. Karel Schönbaum, kterému jsme nakonec hru připsali ... Karlíček Schönbaum byl advokát, ale víc ho bavila matematika. Já mám na nočním stolku detektivku – a on měl u postele tlustou knihu s názvem Rovnice vysoké matematiky, a v té si večer čítal jak v detektivce. Nás přivedl na myšlenku Osla a stínu, a všechno německé, co se ve hře vyskytuje, všelijaká německá rýmování, bylo od něj“.

Tolik Jan Werich.

JUDr. Karla Schönbauma, advokáta, který miloval matematiku, zavraždili nacisté v koncentračním táboře. Chromozóm náhody, spojující statistiku a hru, se v této rodině projevil ještě jinak. Karel Schönbaum měl bratra, univ. profesora dr. Emila Schönbauma, který byl předním statistickým odborníkem předválečné republiky, odborníkem v pojistné matematice a také spoluautorem projektu sociálního pojištění (za tuto informaci děkuji dr. Marcelu Josífkovi). Statistika a hra se tedy snést dokážou.

Jistě lze namítnout, že – stejně jako jedna vlaštovka nedělá jaro – nedělá jeden statistik ze statistiky zábavu. To je pravda. Ze statistiky neudělá nic zábavného ani Světová statistická asociace. Jenže o to vůbec nejde. Jde o něco jiného – totiž o to, že statistika pojednává a vypovídá o jevech a okolnostech každodenního života řečí matematiky. Řečí čísel a kategorií, řečí strohou a úspornou. Takže nutně chladnou a nezaujatou.

Potíž je v tom, že chladných a nezaujatých statistických tabulek se chápeme rukama, jejichž vlásečnicemi koluje kromě plazmy a erythrocytů také zaujetí a přání vidět věci v určitém světle. Obvykle předpojatém, přejícím si, aby věci dopadly takhle a nejinak. Aby výsledky šetření vypovídaly příznivě o tomto a odsuzovaly tamto.

Zatímco teorie matic, diferenciální geometrie, topologie a jiné matematické vědy promlouvají k matematikům, obracejí se statistiky k mnohem širšímu obecenstvu. Třeba statistiky lékařské. V tomto se statistika od ostatní matematiky liší. Nelžou statistiky; lhát mohou interpretace statistiky, což také vydatně činí. Statistická teorie je teorie matematická a jako taková je výsadou specialistů. Interpretace statistických výsledků je však přístupná komukoli. Takže nutně i amatérům, toužícím, aby statistiky vypovídaly o dané věci právě tak, jak si oni přejí. A protože svá přání považují za přání ušlechtilá, nevidí důvod, proč by měli své ušlechtilosti přitahovat otěže. Živým organismům jsou vlastní smysly, jejichž pro-

střednictvím se orientují v životním prostředí, což je svého druhu poznávání. Člověk je však jediný biologický druh, který poznává své životní prostředí uvědoměle, natolik cílevědomě, že vyčlenil poznávání jako samostatnou aktivitu. Poznal totiž, že právě v této schopnosti cílevědomě a aktivně poznávat je jeho největší síla. Proces poznávání životního prostředí a konec konců i poznávání sebe sama má svou organizaci. Velice složitou organizaci, kterou nazýváme věda, a která je tvořena hluboce i široce založenou soustavou vědních oborů, mezi nimiž mají svá místa i vědy biomedicínské. Výchozím předpokladem je přitom představa, že jsme schopni poznat, že svět, který nás obklopuje je poznatelný a uchopitelný soustavou obecných zákonů, k nimž můžeme dospět na základě dílčích pozorování a měření v kontrolovaných experimentech. Zákonů, které podají výpověď o souvislostech jevů a veličin.

Základními nástroji biomedicínských věd jsou pozorování, měření a experiment. Ať už se jedná o vážení novorozence nebo rozbor výstupů počítačového tomografu. Na rozdíl od měření fyzikálních, chemických a technických vystupuje v měřeních biomedicínských výrazně do popředí jisté specifikum. Je jím relativní složitost zkoumaných objektů, ježímž důsledkem je relativní komplikovanost výpovědi o těchto objektech. Řečeno jinak, na výsledku biomedicínského měření se zpravidla podílí celá řada faktorů a vlivů, které se vymykají kontrole a jejichž intervence má na výsledku měření zpravidla velmi podstatný podíl. Podíl, který nelze téměř nikdy zanedbat.

Vývoj věd a vědecké metodologie ukázal, že vliv faktorů tohoto druhu lze při formulování zákonů postihnout statisticky. Protože tyto faktory se chovají a projevují stejně jako se chová a jeví za jistých okolností náhoda, a protože statistika je s náhodou v příbuzenském poměru.

V biomedicínské oblasti se toto působení projevuje způsobem, který jsme si zvykli označovat jako biologická variabilita. Jde o jistý projev neurčitosti, nejednoznačnosti měření a tedy i výpovědí, které na základě těchto měření činíme. Nejde o žádné záhady – biologická variabilita prostupuje naší každodenní praxi stejně samozřejmě, jako jí prostupují dýchání a trávení.

V souboru těhotenství, která skončila porodem zdravého novorozence, kolísá délka těhotenství v rozmezí několika týdnů; podobně porodní hmotnosti těchto novorozenců se liší navzájem v rozpětí mnoha set gramů a jejich porodní délka v rozpětí několika centimetrů. Jistě, lze uvést faktory, jako dědičnost, věk rodičky a další, umožňující učinit předpověď uvedených veličin určitější. Rozhodně však ne zcela jednoznačnou. Jistá neurčitost tu zbude vždycky a tím také prostor pro statistiku. To, že se skrze statistiku lze, není důvod pro odsouzení statistiky, ale pro odsouzení lháře – a také ovšem podnět k lepšímu poznání nástroje, s jehož pomocí se tak mnohé dá poznat.

K čtyřicátému výročí založení katedry statistiky Vysoké školy ekonomické v Praze

Lubomír Cyhelský

Vznik katedry statistiky Vysoké školy ekonomické (VŠE) v Praze 1. srpna 1952 je historická událost, která nejenom zahrnuje určitou míru pikantnosti už jenom tím, že k ní došlo více než celý rok před vlastním vznikem VŠE v Praze, založené 1. září 1953, ale tento historický akt se vyznačuje i tím, že se stal důležitým bodem rozvoje dlouhodobé přípravy statistických specialistů na vysokých školách v naší vlasti.

První neúplné vysokoškolské studium, které mělo statistickou povahu, bylo na našem území zahájeno již v roce 1904 pod názvem studium pojistné techniky na Českém vysokém učení technickém (ČVUT) v Praze. Od roku 1921 probíhal v Československu souběžně nepravidelný cyklus přednášek o pojistné matematice a statistice na přírodovědecké fakultě Karlovy univerzity v Praze. Po druhé světové válce byl schválen zákon č. 122/1946 Sb. o statisticko – pojistném studiu, jímž byl nepravidelný cyklus přednášek o pojistné matematice a statistice na přírodovědecké fakultě Karlovy univerzity přeměněn s účinností od začátku školního roku 1946/47 na studium matematické statistiky, pojistné matematiky a ekonometrie, a jímž bylo dosavadní dvouleté studium pojistné techniky na Vysoké škole speciálních nauk (fakulta ČVUT) přetvořeno skupinou pedagogů vedenou prof. Jaroslavem Janko (sté výročí narození této významné statistické osobnosti si připomeneme na sklonku roku 1993) na úplné vysokoškolské studium statisticko – pojistného inženýrství. Toto studium bylo hlavním střediskem výchovy nejkvalifikovanějších statistických inženýrů matematicko – statistického i ekonomicko – statistického zaměření až do školního roku 1951/52, kdy byla Vysoká škola speciálních nauk ČVUT v Praze tehdejšími vládními činiteli zrušena. Studující, kteří ve školním roce 1951/52 dokončili na Vysoké škole speciálních nauk 2. a 3. ročník statistického inženýrství, dostudovali většinou statistické inženýrství na tehdejší ekonomicko – inženýrské fakultě ČVUT. Studenti, kteří v tomto školním roce dokončili na Vysoké škole speciálních nauk 1. ročník statistického inženýrství, byli převedeni na tehdejší hospodářskou fakultu Vysoké školy politických a hospodářských věd (VŠPHV) v Praze. Tato skutečnost byla posledním a zřejmě rozhodujícím podnětem pro vznik katedry statistiky na hospodářské fakultě VŠPHV.

Kromě posledně uvedeného existovaly samozřejmě i jiné podněty, vedoucí ke vzniku katedry statistiky na hospodářské fakultě VŠPHV. Studium hospodářských oborů nebylo a není totiž při jakémkoli pojetí myslitelné bez zvládnutí statistických disciplin. To si uvědomovali na začátku padesátých let především někteří pedagogové na katedře financí hospodářské fakulty VŠPHV. Právě tato katedra měla za úkol, kromě mnohého jiného, i zabezpečení externích vyučujících statistiky. Za tímto účelem vznikl již v roce 1951 na katedře financí kabinet statistiky, jehož jediným členem a zároveň vedoucím se stal odborný asistent Augustin Hlaváček. Hlavním úkolem kabinetu statistiky katedry financí hospodářské fakulty VŠPHV bylo vytvořit v průběhu letního semestru školního roku 1951/52 předpoklady pro založení katedry statistiky na hospodářské fakultě VŠPHV. Účinnou podporu při řešení tohoto úkolu poskytovali hlavnímu organizátorovi přípravných prací A. Hlaváčkovi někteří tehdejší asistenti katedry financí hospodářské fakulty VŠPHV (zejména Miroslav

Tuček), prezident tehdejšího Státního úřadu statistického Dr. František Fajfr a jeho náměstek Ing. František Herbst a tehdejší matematicko–statističtí asistenti Vysoké školy speciálních nauk Ing. František Fabian a Dr. Ing. Jaroslav Hájek.

Samostatná katedra statistiky na hospodářské fakultě VŠPHV vznikla 1. srpna 1952 z kabinetu statistiky katedry financí a byla při svém založení vsutku určitou kuriozitou, protože se při svém vzniku skládala pouze ze dvou interních členů (A. Hlaváčka a autora této poznámky). Externím vedoucím katedry statistiky byl ustanoven náměstek předsedy Státního úřadu statistického Ing. František Herbst. Tento miniaturní kolektiv dvou odborných asistentů a jednoho asistenta měl v průběhu jednoho až dvou měsíců splnit následující bezprostřední úkoly:

- a) zabezpečit zorganizování statistického studijního oboru na hospodářské fakultě VŠPHV včetně zabezpečení výuky nejen statistických, ale i ostatních předmětů ve školním roce 1952/53;
- b) zabezpečit výuku teorie statistiky a ekonomické statistiky na nestatistických studijních oborech hospodářské fakulty VŠPHV ve školním roce 1952/53;
- c) zabezpečit ve školním roce 1952/53 výuku statistiky na právnické fakultě Karlovy univerzity v Praze.

Na základě četných jednání v průběhu srpna a části září 1952 a při velké podpoře tehdejšího děkana hospodářské fakulty VŠPHV prof. JUDr. Vladimíra Sedláka se do začátku školního roku 1952/53 podařilo rozšířit stav interních pedagogů katedry statistiky o Ing. Bohumila Řezníčka, Ing. Eduarda Linka a Ing. Benedikta Kordu. Zároveň byli k externí spolupráci získáni četní statističtí teoretici (např. Dr. Ing. Jaroslav Hájek, prof. Dr. František Egermayer) a někteří významní statističtí praktici (např. Dr. Vladimír Gruzín).

Důsledkem personálního rozšíření katedry statistiky včetně následného příchodu Ing. Jaromíra Waltera na katedru statistiky VŠPHV v lednu 1953 bylo, že se výše uvedené úkoly podařilo ve školním roce 1952/53 se ctí splnit. Mimořádným úspěchem katedry statistiky a velkou odměnou za její úsilí od 1. srpna 1952 až do konce školního roku 1952/53 byl vznik samostatné fakulty statistiky. Dnem 1. září 1953 byla totiž hospodářská fakulta VŠPHV v Praze přetvořena na Vysokou školu ekonomickou v Praze s pěti fakultami, jednou z nichž se stala právě fakulta statistiky. O náplni statistického studia na fakultě statistiky si může čtenář učinit do určité míry úsudek podle následujícího učebního plánu, který byl sestaven podle výkazu o studiu prof. Ing. Václava Čermáka. DrSc., který byl jedním z jeho prvních absolventů. První ročník ekonomicko–statistického inženýrství studoval ve školním roce 1951/52 na Vysoké škole speciálních nauk, druhý ročník ekonomicko–statistického inženýrství studoval ve školním roce 1952/53 na VŠPHV a třetí až pátý ročník specializace statistika zemědělství absolvoval v letech 1953 až 1956 na fakultě statistiky Vysoké školy ekonomické v Praze (kromě zemědělsko–statistické specializace oboru ekonomická statistika se studovaly na fakultě statistiky i specializace průmyslové a obchodní statistiky).

Učební plán studia oboru ekonomická statistika (specializace statistika zemědělství) v letech 1952 až 1956 na Vysoké škole speciálních nauk (1. ročník), VŠPHV (2. ročník) a fakultě statistiky VŠE v Praze:

Povinný předmět	Počet týdenních hodin přednášek/cvičení zápočtů (z) a zkoušek (*) v semestru							
	1.	2.	3.	4.	5.	6.	7.	8.
Matematika	3/2	4/4*		2/1*				
Podnikové početnictví	4/2	4/2*	2/1*	2/1z				
Hospodářský zeměpis	3/-*							
Dějiny hospodářství ČSR		2/-						
Politická ekonomie	4/-	4/-*	2/2	4/2*				
Společenské nauky	2/2z	2/2*	4/2z	4/2*		2/-z		
Právo	3/-*		4/-*					
Ruština	-/2z	-/2*	-/4z	-/4*				
Teorie statistiky			2/2z	4/2*	2/2*			
Dějiny hospod. teorií					2/-	2/-*		
Ekonomická statistika					3/2z	3/2	*	
Základy technologie					4/2z	2/2z		
Národohospod. plánování					2/2z	2/2*		
Finance a úvěr					2/-z	2/-*		
Ekonomika odvětví						3/2	*	
Němčina					-/2	-/2z	-/2	-/2*
Organizace a plánování podniku							2/-z	2/-*
Vlastní náklady a tvorba cen								2/2*
Mechanizace evidence							2/4*	
Vybrané kapitoly ze statistiky							2/-z	2/-z
Statistika zemědělství							2/2z	2/2*
Počet hodin předn/cvič.	19/8	16/10	14/11	16/12	15/10	16/10	8/8	8/6
Počet zkoušek	2	5	2	5	1	5	3	4
Počet zápočtů	2	-	3	1	4	1	3	1

V 9. semestru se konala předdiplomní praxe. V 10. semestru bylo studium zakončeno státní zkouškou z politické ekonomie, společenských nauk a obhajobou diplomové práce, spojenou se zkouškou ze statistiky.

Katedra statistiky hospodářské fakulty VŠPHV v Praze stála na počátku kalendářního roku 1953 před stále náročnějšími úkoly. Brzdou v jejich plnění se stalo vedení katedry, které bylo prováděno na dálku, bez osobní účasti vedoucího katedry při rozhodování závažných i drobných problémů. Proto byl s účinností od 1. února 1953 jmenován nový vedoucí katedry z interních pracovníků. Stal se jím Doc. Ing. Eduard Link, který stál v čele katedry statistiky VŠPHV (od 1. září 1953 se stala katedra statistiky hlavní katedrou fakulty statistiky VŠE Praha) a VŠE do školního roku 1954/55, kdy její vedení převzal Doc. Ing. Benedikt Korda. Na začátku školního roku 1953/54 se katedra statistiky VŠE Praha rozrostla o Ing. Ilju Nováka, Ing. Vladimíra Roubíčka, Ing. Jana Zelinku a Jana Vraného, který však poměrně brzy odešel pracovat do aparátu OSN.

Ve školních letech 1953/54 až 1957/58 bylo hlavním úkolem katedry statistiky nejenom stále zabezpečení kvalitní výuky teorie statistiky, ekonomické statistiky a jiných statistických disciplín na všech nestatistických studijních oborech Vysoké školy ekonomické v Praze, ale zejména zabezpečování zdárného vývoje fakulty statistiky VŠE Praha. Víta- nou pomocí v plnění těchto úkolů byl příchod nových pedagogů. Šlo o Ing. Jaru Kaňokovou (1954), Ing. Metoděje Bauera (1955), Ing. Václava Čermáka (1956), Ing. Jiřího Škabradu (1956) a Josefa Kašpara (1957). Katedru statistiky v tomto období naopak opustil Ing. Bohumil Řezníček, který přešel na Vysokou školu dopravní do Žiliny, a zakládající člen katedry statistiky Augustin Hlaváček, který odešel věnovat se statistické praktické činnosti.

V roce 1958 tehdejší vládní činitelé pod heslem argumentů o údajné úspornosti zrušili nejenom tehdejší fakultu financí a fakultu všeobecně ekonomickou, ale bohužel i fakultu statistiky, a vytvořili z těchto tří fakult jednu fakultu politické ekonomie. Katedra statistiky VŠE nadále plnila jako svou hlavní funkci zabezpečování statistických disciplín na celé VŠE. Někteří její členové se snažili o zachování studijního oboru ekonomická statistika a usilovali o znovuoobnovení fakulty statistického typu. To už jsou ale otázky, které značně přesahují rámec této vzpomínky na založení katedry statistiky 1. srpna 1952.

LITERATURA:

1. Cyhelský L., Kejkula J., *Dvě desítky let katedry statistiky na VŠE v Praze*, Statistika č.6 (1972), 247.
2. Cyhelský L., *Příprava vysokoškolsky vzdělaných ekonomicko – statistických specialistů v ČSSR v letech 1945 – 1985*, Statistika č.5 (1985), 201.
3. Egermayer F., *Úprava statistického studia na našich vysokých školách*, Statistický obzor č.3 (1946), 350.
4. Egermayer F., *První statisticko – pojistní inženýři*, Statistický obzor č.2 (1947), 238.
5. Janko J., *Studium statistiky*, Statistický obzor č.1 (1947), 1.
6. Podzimek J., *Vývoj československé statistiky v událostech a datech 1945 – 1975*, Výzkumný ústav sociálně ekonomických informací, Praha, 1976.
7. Walter J., *Studium statistiky na Vysoké škole ekonomické*, Statistický obzor č.7–8 (1953), 311.

Příprava dat pro počítačové zpracování.

H. Řezanková, J. Žváček

1. Úvod

Na počátku statistické analýzy dat stojí získávání dat a jejich vkládání do počítače. V souvislosti s růstem množství dat, která je třeba zpracovat, se práce s daty stává stále důležitější. Pozornost se soustřeďuje především na kontrolu dat a na uchování dat v komprimovaném tvaru.

Současné programové prostředky poskytují řadu výhod, zejména

- snadné vkládání dat s možností jejich *kontroly* při zadávání,
- možnost *modifikace datových souborů*, jejich spojování či rozdělování, vytváření nových proměnných pomocí výpočtů z hodnot již existujících proměnných, provádění výběrů z datových souborů, ať již náhodně či v závislosti na zadané logické podmínce,
- možnost *generování hodnot*.

2. Uspořádání dat

V programových systémech jsou napozorované či jinak získané údaje zpravidla ukládány jako matice, kde každý sloupec odpovídá určitému znaku a řádek statistické jednotce nebo časovému období. Nejnázornějším způsobem pro vkládání dat je proto forma tabulky.

Kromě vlastních dat lze do tabulek zahrnout popisy sledovaných jednotek (jména pacientů či pracovníků, názvy výrobků, podniků), resp. popis časového období (datum měření apod.). Tyto popisy se rovněž zapisují do určitého sloupce, nejsou však používány pro statistické výpočty – slouží pro usnadnění orientace ve vstupních datech, případně ve výsledných tabulkách a grafech.

V počítačové terminologii se používají následující termíny: sloupce tabulky představují *proměnné*, řádky jsou *pozorování* neboli *případy*.

1.proměnná	2.proměnná	...		

jednotlivá pozorování

Programové systémy pracují s různými *typy proměnných*. Základem jsou vždy dva typy

- *číselný* (pro kvantitativní údaje nebo pro kvalitativní, pokud jsou označeny čísly) a
- *znakový* (pro kvalitativní údaje nebo pro popis sledovaných jednotek).

Kromě toho většina systémů umožňuje pro popisné účely zadávat též

- *datum*, například ve tvaru DD/MM/RR nebo americkém MM/DD/RR, kde DD je den, MM je měsíc a RR je poslední dvojčíslí z roku.

Obvykle je jedna proměnná zapsána v jednom sloupci, tedy jako *vektor*. Existují však situace, kdy lze jako proměnnou nazvat *matici*, například v systému Statgraphics se dvou- rozměrná tabulka četností se ukládá jako proměnná ve tvaru matice.

3. Vstup dat

Pod pojmem vstup dat chápeme zavedení dat do operační paměti počítače nebo do dočasné oblasti na disku, kde jsou data připravena ke zpracování. Rozlišujeme

- vkládání dat pomocí klávesnice,
- vstup dat z již existujících datových souborů,
- vstup výsledků statistických procedur (zejména reziduí, teoretických hodnot, korelační matice, případně vygenerovaných hodnot).

Tuto možnost by měl obsahovat statistický paket.

Vkládání dat se obvykle provádí do operační (vnitřní) paměti počítače, kde se data buď přímo zpracovávají, nebo odkud se ukládají do souboru na disk či disketu (vnější paměť počítače) pro další zpracování, a to buď v textovém (tzv. ASCII tvaru), nebo ve speciálních tvarech. K této činnosti lze použít buď přímo některý statistický programový systém, nebo programy specializované na přípravu dat. Můžeme rozlišit následující základní způsoby vstupu dat (v závorce jsou vždy uvedeny nejčastější příklady programových systémů).

A) Postupné vkládání jednotlivých hodnot (INPUT).

Jednotlivé hodnoty určité proměnné se zadávají postupně, počet pozorování (hodnot) se specifikuje buď před vlastním vkládáním dat, nebo vložením speciálních znaků místo další hodnoty. Tento přístup je nejjednodušší z hlediska naprogramování, ale není vhodný z hlediska uživatele, neboť má za následek těžkopádné kontroly a opravy. Používá se v jednoduchých statistických programech.

B) Vložení všech hodnot jedné proměnné (READ).

Hodnoty se obvykle zadávají do jednoho nebo několika řádků, oddělují se mezerou nebo čárkou. Tato možnost je určena především pro vložení malého počtu hodnot, například pro ověření nějaké metody. Můžeme se s ní setkat například v paketu Statgraphics.

C) Vkládání pomocí celoobrazovkového textového editoru.

V rámci tohoto způsobu existují dva základní formáty pro vstup dat, a to

- *pevný*, v němž všechny hodnoty určité proměnné začínají na stejné pozici na řádku (nemusí být odděleny), a
- *volný*, kdy jsou jednotlivé hodnoty odděleny (například mezerou, čárkou nebo jiným speciálním znakem) a nezáleží na jejich pozici na řádku.

Obvykle se jedno pozorování zapisuje na jeden řádek, jednou z výjimek je editor REVIEW statistického systému SPSS/PC+, který umožňuje vložit pouze 80 znaků na řádek (paket ovšem umožňuje přečíst pozorování z více než z jednoho řádku).

Lze použít libovolný textový editor, je však třeba dbát na to, aby výsledný tvar dat byl tzv. ASCII, to znamená, že textový soubor je bez speciálních znaků (pro typy písma, úpravu stránky k tisku apod.). Pokud je použit specializovaný textový procesor (např. T602), je třeba provést uložení dat formou exportu do tvaru ASCII (v T602 je to ASCII1).

D) Vkládání pomocí specializovaného tabulkového editoru (EDIT, BROWSE).

Součástí některých statistických paketů jsou tabulkové editory, které jsou však značně rozdílné. Spojuje je pouze základní princip – vkládání hodnot do *políček tabulky*, takže na rozdíl od textového editoru není třeba se zabývat oddělovači, ani pozicemi na řádku (přechod mezi sloupečky tabulky se obvykle uskutečňuje pomocí speciální klávesy – např. tabulátoru).

Umožňují zpravidla

- transformace proměnných,
- vypouštění řádků a sloupců,
- operace se sloupci,
- generování sloupců (náhodných čísel).

Jako příklad lze uvést pakety BMDP PC, Statgraphics, SYSTAT či Minitab. Do této skupiny lze zařadit též specializované tabulkové procesory.

E) Vložení všech hodnot jednoho pozorování do formuláře (EDIT).

Tento postup je charakteristický pro databankové systémy, v nichž jsou údaje o statistické jednotce (podniku, výrobku, pracovníku) vkládány do *políček formuláře*, který je rozvržen na jednu nebo více obrazovek. Většinou lze spolu s grafickou úpravou formuláře navrhnout i kontroly pro vstup hodnot do políček.

Obvykle lze tuto formu vstupu kombinovat se vstupem hodnot do tabulky. Je tomu tak nejen v databankových systémech, ale i ve specializovaných systémech pro přípravu dat, například v datovém editoru DataEntry II, který lze zakoupit spolu se statistickým systémem SPSS/PC+.

Pokud jde o *vstup dat z již existujících souborů*, rozlišujeme

F) vstup z vlastních souborů systému (GET, USE), jestliže byla data uchována stejným statistickým paketem, kterým budou zpracovávána, a

G) vstup ze souborů jiných systémů (IMPORT, TRANSLATE), jestliže byl datový soubor připraven v jiném systému, než ve kterém bude zpracováván.

Většina statistických paketů obsahuje alespoň jeden prostředek pro vstup dat (obvykle typu D, tabulkový editor) a navíc možnost importu dat z jiných programových systémů, tj.

- z textových editorů (typ C),
- z tabulkových procesorů (typ D),
- z databankových systémů (typ E a D), případně
- z integrovaných systémů (umožňují vstupy typu C, D i E) a
- z jiných statistických paketů.

Pokud jde o možnosti převádění datových souborů mezi jednotlivými programovými systémy, pak existuje několik základních možností:

- import z jiného systému přímo (pokud existuje),
- import přes některý standardní tvar (ASCII, tvar tabulkového procesoru Lotus 1–2–3 nebo databankového systému dBASE), tzn. – v prvním systému provedeme export do standardního tvaru (pokud existuje) a ve druhém systému import tohoto tvaru, nebo
- v prvním systému zadáme „tisk do souboru“ (bude vytvořen ASCII tvar) a ve druhém systému provedeme import ASCII tvaru,
- použití speciálního programu pro převod dat (např. DBMS/COPY), v němž se zadá typ a název zdrojového souboru a typ a název cílového souboru (data jsou tedy již v požadovaném tvaru a není třeba provádět import).

4. Datové soubory a proměnné

Datový soubor může obsahovat buď jednu nebo více proměnných. U většiny statistických paketů se při zpracování postupuje tak, že se do operační paměti (nebo dočasné oblasti na disku) načte obsah souboru (příkazy GET, USE, IMPORT), a pak lze pracovat pouze s proměnnými, které se vyskytují v tomto souboru. (U těchto paketů ovšem obvykle existuje možnost spojování souborů, takže lze datové soubory upravit.) Existují však i výjimky, například paket Statgraphics si při svém spuštění vytváří seznam proměnných (Data Directory) z aktuálního adresáře na disku či disketě. V tomto případě je možno zpracovávat hodnoty proměnných z různých souborů.

K identifikaci proměnných slouží *jméno* (název), které obvykle zadává uživatel. I když v některých systémech neexistuje omezení na jméno proměnné (typickým příkladem jsou tabulkové procesory), je třeba si uvědomit, že statistické pakety toto omezení mají, a mohou tedy nastat problémy při importu dat.

Obecně platí, že by jméno nemělo přesáhnout délku osmi znaků, mělo by začínat písmenem a mělo by obsahovat pouze písmena, číslice a znak podtržítka (_). Pokud uživatel zadává desetimístná jména (v systému, který to umožňuje), měl by se držet zásady, že se tato jména v prvních osmi znacích liší (obecně by se pro lepší orientaci měla lišit v počátečních znacích – např. MLEKVIROBA, MASOVIROBA, nikoliv obráceně).

U většiny systémů nezáleží na tom, zda se jména zapisují malými či velkými písmeny (lze je zaměňovat). Výjimkou je například paket Statgraphics, kde malé a velké písmeno má jiný význam.

Je třeba si uvědomit, že mezera znamená oddělovač, nesmí se nikdy vyskytnout v žádném jménu (pokud by se vyskytla, byla by tak zadána dvě jména).

V systémech, které umožňují pracovat s proměnnými z různých souborů, se uvádí obvykle název proměnné, složený ze jména souboru (viz dále) a jména proměnné. V paketu Statgraphics jsou tato dvě jména oddělena tečkou, v databankových systémech šipkou ($- >$), která se zapisuje jako dva znaky – pomlčka a relační znaménko „větší než“.

K přesunu dat z operační paměti počítače na disk či disketu slouží obvykle příkaz nebo nabídka **SAVE**, které vyžadují zadání jména či názvu souboru, jež musí splňovat požadavky operačního systému MS-DOS.

Celý název souboru se vždy skládá z maximálně osmiznakového jména a tříznakové koncovky (extenze, rozšíření), která je od jména oddělena tečkou (např. DOPIS.TXT).

Koncovka obecně označuje typ souboru (program, text, data), často také označuje, jakým programovým systémem byl soubor vytvořen. Některé systémy si koncovku doplňují samy a nelze ji měnit (Statgraphics si doplňuje koncovku .ASF), jiné ji sice doplňují, ale je možno zadat jinou (SYSTAT doplňuje koncovku .SYS), u ostatních záleží na uživateli, zda ji vůbec zadá, a pokud ano, tak jakou (systém SPSS).

U jména souboru zpravidla nezáleží na tom, zda se používají malá či velká písmena. Výjimkou je paket Statgraphics, kde je v některých případech požadováno, aby jméno souboru bylo zapisováno velkými písmeny, a navíc může obsahovat pouze písmena a číslice a musí začínat písmenem.

Jestliže se soubor vyskytuje (nebo má vyskytovat) v jiném než aktuálním adresáři na disku či disketě, je třeba u jména souboru uvádět tzv. cestu, která se skládá z označení disku (diskety) a adresáře, například

A:\DATA\VYROBA.SYS

označuje soubor, který je uložen na disketě v mechanice A, v adresáři DATA a má jméno VYROBA s koncovkou SYS.

Výjimkou je opět paket Statgraphics, který neumožňuje pracovat se soubory z jiného než předem vymezeného adresáře pro data (v průběhu práce lze však nastavit kterýkoli jiný adresář).

5. Kontrola dat

Před vlastním statistickým zpracováním je třeba věnovat velkou pozornost kontrole dat, neboť předpokladem správné analýzy jsou správná data. (Je znám případ, kdy po provedení a publikování výsledků sociologického šetření bylo náhodně v datech zjištěno 32 druhů pohlaví.)

Kontrolu dat je možné provádět jednak při vkládání dat do počítače (logická kontrola), jednak pomocí statistického zpracování již vložených dat (statistická kontrola).

Při kontrole dat je třeba vycházet z toho, jaké jsou hlavní příčiny vzniku chyb v datech. Podle zkušeností to jsou

- **nesprávná odpověď** (nesmyslné údaje, chybějící údaje),
- **nesprávné měření** (v ekonomii zejména jiné ceny, nerespektování metodiky, organizační změny),
- **nesprávné kódování**, špatná měrná jednotka, stupnice, identifikace, desetinná čárka místo tečky,
- **chyby záznamu dat**, zejména překlepy, vynechání údajů,
- **chyby přenosu dat**, při čtení médií, uvnitř počítače, v síti.

Klasické metody kontroly dat jako je optická kontrola či opakovaný vstup jsou nejenom pracné, ale dokonce ani nepostihují většinu příčin vzniku chyb v datech.

Logická kontrola umožňuje odhalovat nepřipustné hodnoty. Lze k tomu využít programových systémů určených pro přípravu dat, které v sobě mají zabudovanu možnost nastavení určitých kontrol nebo možnost vkládání hodnot pomocí výběru z předem nastavených možností (u diskrétních proměnných s malým počtem hodnot).

Nejpoužívanější metody jsou

- **metoda intervalu**, vhodná pro číselné údaje. Pro kontrolu je nutno zadat interval, v němž se mohou pohybovat hodnoty proměnné,
- **metoda masky**, vhodná zejména pro kombinované údaje, jako jsou například číselníky, datumy atd., kde je předepsán typ znaku pro určitou polohu,
- **metoda kontrolního čísla**, kdy se k údajům připojuje součet či jiné kontrolní číslo (např. u rodného čísla),
- **metoda implikace**, kdy se vytvářejí logické vztahy mezi některými údaji (např. mezi datem narození a rodným číslem, muž nemůže mít děti atd.).

Logickou kontrolu je možno nejlépe zabezpečit v databankových systémech.

Statistická kontrola umožňuje odhalovat nepravděpodobné hodnoty a kombinace hodnot. Lze například zjišťovat

- **extrémní hodnoty** na základě výpočtů aritmetického průměru a směrodatné odchylky (metoda „šesti sigma“),
- **vzácné kategorie**, tj. vzácně se vyskytující dvojice znaků u kategoriálních dat, a to pomocí vícerozměrných tabulek četností,
- **odlehlá pozorování** na základě regresní analýzy.

6. Práce s chybějícími údaji

Statistika pracuje s velkými soubory reálných dat, kde je poměrně častým jevem, že některé údaje chybí. Obzvláště velký význam má práce s chybějícími údaji tam, kde se jedná o zpracování informace o více proměnných. Typickým případem je práce s kontingenčními tabulkami.

V praxi rozlišujeme dva druhy chybějících pozorování:

- A) Systémové chybějící údaje**, což jsou většinou výsledky operací se vstupními údaji mimo definiční obor operace. Nejčastěji je to neuvedení čísla tam, kde jej očekáváme (místo čísla je tedy prázdný znak), nebo nedefinovaná operace (např. dělení nulou). Tento výsledek se v některých systémech označuje jako **ERROR** (chyba).

Jako chybějící pozorování lze také zavést výsledky aritmetických operací, které nelze v počítači zobrazit. Jsou to také *příliš velká čísla*, například když výsledek násobení převyší zadané (či strojové) maximum.

- B) Uživatelské chybějící údaje**, což jsou údaje mimo definiční obor konkrétní proměnné (např. nesmyslný údaj, omezení výpočtu na data, která pokládáme subjektivně za „možná“ atd., zpravidla označovaných jako **MISSING**).

Problematika chybějících pozorování se řeší v zásadě třemi cestami:

- A) Vypouštěním**, což je nejjednodušší způsob založený na prostém vyloučení **proměnné, pozorování či údaje**.

Ve statistických programových paketech se používají následující možnosti:

- COLUMNWISE**, vypuštění **proměnné** (sloupec v datové matici), která obsahuje chybějící údaj,
- LISTWISE**, vypuštění **pozorování** (řádek v datové matici), které obsahuje chybějící údaj (nejjobvyklejší),
- PAIRWISE**, vypuštění pouze **jednoho údaje** (typický případ je výpočet korelační matice v případě, kdy chybí jeden údaj).

- B) Zavedením zvláštního čísla „missing“**, reprezentujícího chybějící pozorování, pro které je implementována speciální algebra.

Tento způsob práce je typický pro většinu dnešních systémů pro vstup dat. Kromě specializovaných statistických systémů jej zahrnují například i *tabulkové procesory* (v nich se vzhledem k zaměření používá dalšího rozlišení typu chybějícího čísla, a to chyba při výpočtu, označená **ERROR**).

Výhodou tohoto přístupu je, že veškeré výpočty jsou prováděny automaticky s dodržением algebry chybějících pozorování, takže se může stát, že se i chybějící pozorování použije. Výsledek operace s chybějícím číslem totiž nemusí být vždy být chybějící číslo:

$$\begin{aligned} 0 * \text{missing} &= 0, & 0 / \text{missing} &= 0, \\ \text{missing} \wedge 0 &= 1, & 0 \wedge \text{missing} &= 0. \end{aligned}$$

- C) Doplněním chybějících údajů**. Tento způsob práce je také poměrně rozšířen a doplnění chybějících údajů se provádí zejména nahrazením chybějících údajů

- *průměrnou hodnotou proměnné* (metoda *MEANSUB* v *paketu SPSS*),
- *regresním odhadem* na základě známých hodnot jiných proměnných.

7. Kompresce dat

Kompresce, neboli *zhušťování* dat, zahrnuje prostředky, jak zobrazit data v menším objemu beze ztráty informace. Praxe ukazuje, že je velký rozdíl mezi množstvím informace obsažené v datech a mezi potřebným objemem paměti pro jejich uložení. Již velmi jednoduchými prostředky můžeme dosáhnout značné úspory paměti. Je ovšem mnoho cest, jak toho dosáhnout. Můžeme rozlišit

- **programové prostředky**, kam zahrnujeme metody, kterými může programátor zredukovat potřebný objem paměti pro data; tyto prostředky jsou závislé na charakteru dat;
- **automatické kompresní algoritmy**, kam lze zahrnout algoritmy pro automatickou kompresi libovolných dat;
- **metody založené na modelu dat**, ve kterých se data nahrazují vhodným modelem a ukládají se pouze parametry modelu.

První přístup souvisí s vhodným využitím programovacího jazyka, všimněme si proto dvou obecnějších přístupů:

Automatické kompresní algoritmy umožňují automaticky redukovat objem libovolných dat. Takto lze ovšem postupovat pouze v případech, kdy je struktura dat dopředu známa.

Velmi důležitou a často používanou skupinou algoritmů z oblasti kódovacích tabulek jsou algoritmy založené na **proměnlivém kódování**. Tyto algoritmy vycházejí ze dvou myšlenek

- znak není nutno zobrazovat posloupností bitů konstantní délky (např. v ASCII je to 8 bitů, tedy 1 byte paměti), ale lze jej zobrazit různě dlouhými posloupnostmi bitů,
- statistickou analýzou dat zjistíme, jak optimálně pro daná data zakódovat znaky (jak mají být tyto posloupnosti dlouhé).

Podobný přístup byl použit např. při konstrukci známé Morseovy abecedy. Optimální způsob konstrukce takového kódu navrhl Huffman. Je založen na **shlukování zdola** v tabulce četností výskytu znaků v souboru. Algoritmus spočívá v tom, že se vytvoří tabulka rozdělení četností a shlukují se kumulativně vždy dvě nejmenší četnosti.

Huffmanův kód má mnoho dalších předností. Jednou z nich je **vlastnost prefixu**, což znamená, že se žádný kód neopakuje jako začátek jiného kódu. Znamená to, že dekódování může být prováděno postupně, znak je rozpoznán, když se vyskytne v kódovací tabulce.

Huffmanův algoritmus je **dvouprůchodový**, takže proces komprimace může být pomalejší než u jednaprůchodových algoritmů. V současné době je velmi populární jednaprůchodový algoritmus nazvaný podle autorů Ziv–Lempel zdokonalený Welchem (tzv. **LZW**), který je důmyslnou kombinací Huffmanova kódu s procesem „učení se na datech“.

V **metodách založených na modelu dat** se data nahrazují vhodným matematickým

nebo statistickým modelem a výsledky jsou často *irreverzibilní*, nelze již zpětně rekonstruovat zcela přesně původní data.

Pro představu si uvedme tři jednoduché metody

- **interpolační metoda**, vhodná pro pomalu se měnící hladké funkce; můžeme si například pamatovat pouze liché body a sudé získat interpolací dvou sousedních,
- **regresní metoda**, v níž data prokládáme polynomem dostatečně vysokého stupně; parametry určíme např. metodou nejmenších čtverců a v paměti uchováujeme pouze počet bodů a parametry polynomu;
- **FFT**, rychlá Fourierova transformace, ve které si periodickou funkci představíme jako Fourierův rozvoj dostatečně vysokého stupně; metoda je obdobná regresní metodě.

Modelovými metodami můžeme dosáhnout skutečně významné komprese. Jako extrémní příklad se uvádí (Research and Development, March 1988, str.37) komprese 1 : 10 000 při použití **fraktálů** s náhodným doplněním detailů pro kompresi fotografie.

Kompresní algoritmy jsou realizovány v **kompresních programech**. (Dnes je již většina programových produktů dodávána v komprimovaných souborech kvůli úspoře disket.) Důležitým vedlejším efektem kompresních programů je však kromě **úspory paměti** také **ochrana dat před neoprávněným použitím** (komprimované soubory lze zajistit heslem) a **ochrana proti počítačovým virům**.

Některé programové systémy umožňují přímo pracovat s daty v komprimovaném tvaru.

8. Závěr

Nabízí se tedy otázky

- *V jakém programovém systému připravit data?*
- *V jakém tvaru uchovávat datové soubory?*

Pro prvotní vkládání dat by měl být použit některý databankový systém, nejlépe z řady Xbase, což jsou systémy, které respektují standard dBASE III+ (např. dBASE IV a Fox-Pro). Jako důvody lze uvést

- databankové systémy jsou určeny ke vkládání a uchování dat,
- databankové systémy poskytují široké možnosti kontroly dat,
- tvar dat ukládaných systémem dBASE je standardem, s nímž umí pracovat téměř všechny ostatní systémy, a odpadají tedy starosti s převáděním dat pomocí dalších importů a exportů.

Data připravená databankovým systémem by měla být přečtena statistickým paketem, v němž bude prováděno zpracování. Pokud to paket umožňuje, měly by být doplněny popisy proměnných a jejich hodnot, případně chybějících údajů. Měla by být provedena statistická kontrola. Pro opravy a drobné úpravy již existujícího datového souboru postačí tabulkový procesor, který je obvykle součástí statistického balíku.

Data by měla být uchována ve tvaru, který používá příslušný statistický systém a jehož součástí jsou informace o proměnných a jejich hodnotách, případně speciální systémové proměnné. Výhodné je používat pakety, které umožňují ukládat data v komprimovaném tvaru, jako je tomu například u systému SPSS/PC+.

Literatura k některým problémovým okruhům:

Kontrola dat:

- [1] Hála R., *Automatická kontrola správnosti číselných označení*, MAA'86 č.6, str. 216–220.
- [2] Herzmann J., *Chyby ve vstupních datech*, MAA'86 č.3, str.102–104.
- [3] Novák V., *Použití fuzzy množin k obsahové kontrole vstupních dat*, Informačné systémy č.12 (1984), 143–151.

Práce s chybějícími údaji:

- [4] Beale E. M. L., Little R. J. A., *Missing Values in Multivariate Analysis*, JRSS B, 37 (1975), 129–145.
- [5] Čakrt M., *Statistické programy a chybějící hodnoty: varování před nečekanými výsledky*, Sociologický časopis (1990), 94–100.

Komprese dat:

- [6] Czontó J., *Zníženie objemu informácie metódou zhustenia dát*, Informačné systémy č.6 (1977).
- [7] Held G., *Data Compression*, John Willey, Norwich, 1984.
- [8] Pissanetzky S., *Sparse Matrix Technology*, Academic Press, New York, 1984 (rusky Mir, 1988).
- [9] Sciamanda R. J., *Another Approach to Data Compression.*, BYTE (February 1987), 137–142.
- [10] Šturcová J., *Komprézia dát.*, Informačné systémy č.1. (1984).

Nebyl bych tvrdohlavý, kdyby ...

Poznámka k článku J. Anděla *Nebuďte tvrdohlaví*

Jan Klaschka

Dal bych si poradit a nebyl tvrdohlavý, kdybych se mohl spolehnout, že $P(D3|KAK) =$
 1. To však z předpokladů (i) – (iv) vyplývá jen v případě, že moderátor *musí* některé dveře otevřít. (Ať čtu, jak čtu, vidím jen, že jedny otevřel.)

Má-li moderátor také možnost žádné dveře neotevřít, vyžaduje předložené řešení revizi. Předpoklad (iv) je třeba upravit – moderátor otevře v situaci AKK druhé dveře s pravděpodobností q rovnou nejvýše $1 - p$. O $P(D3|KAK)$ můžeme předpokládat, že nabývá nějaké hodnoty r . Z Bayesovy věty pak tentokrát vyplývá, že

$$P(KAK|D3) = \frac{p_2 \cdot r}{p_1 \cdot p + p_2 \cdot r},$$

tedy za předpokladu $p_1 = p_2 = p_3 = \frac{1}{3}$ je

$$P(KAK|D3) = \frac{r}{p + r}.$$

Tato pravděpodobnost je v případě, že $r < p$, menší než jedna polovina. Chce-li soutěžící vyhrát auto s alespoň padesátiprocentní pravděpodobností při libovolných hodnotách p a r , nelze mu zřejmě poradit nic lepšího, než aby si hodil korunou.

V soutěžích, jako je ta, o níž je řeč, nebývá asi snahou pořadatele za každou cenu minimalizovat výhru. Kdybych nicméně byl v roli nepřejícího moderátora a měl tu možnost, volil bych $r = 0$ a $q = 1 - p$ (tedy strategii „zkoušet“ soutěžícího právě tehdy, když ukáže na správné dveře). Mou nejsnazší obětí by byl čtenář Informačního Bulletinu, který by se řídil doporučením vysloveným v názvu komentovaného článku.

Ze světa

Nizozemská statistická společnost

Václav Fidler

Nizozemská statistická společnost VVS (*Vereniging voor statistiek*) byla založena v roce 1946. Na konci minulého roku měla 1445 členů. Každý člen VVS je členem nejméně jedné z osmi následujících sekcí:

sekce	počet členů
matematicko – statistická	441
operačního výzkumu	528
sociálně – společenských věd	371
výpočetní	313
podniková	278
lékařsko – biologická	276
ekonomická	269
zemědělská	178

Jedinou podmínkou členství je placení příspěvků (114 guldenů ročně, včetně členství jedné sekce). Společnost nedostává státní podporu.

V aktivitě VVS lze rozlišit decentrální – to je na úrovni sekcí – a centrální složky. K typické aktivitě sekcí patří půldenní setkání s přednáškami 2–3 krát ročně; počet účastníků bývá 20–50. Občas je takové setkání kombinováno s návštěvou podniku či pracoviště. Některé sekce organizují i vícedenní konference. Například sekce matematicko–statistická organizuje každý rok třídenní konferenci s přednáškami významných zahraničních statistiků; lékařsko–biologická sekce organizuje pro své členy každé dva roky třídenní konferenci s přednáškami účastníků. Většina sekčních aktivit je přístupná i pro nečleny.

Členství ve VVS zahrnuje předplatné na tři časopisy. *Statistica Neerlandica*, vycházející čtyřikrát ročně a obsahující teoreticky i prakticky zaměřené články v angličtině, je zahraniční vizitkou společnosti. Výroba této vizitky spolkně skoro 40% členských příspěvků. Podle některých členů příliš mnoho za teoretický časopis, který je v jejich očích přísný při výběru článků. Diskuse na toto téma zrodila před lety časopis *Kwantitatieve Methoden*. Kvantitativní Metody jsou tištěny na laciném papíru, články jsou v holandštině nebo v angličtině a selekce je mírná. Třetím orgánem je *VVS – Bulletin*: měsíčník informující o aktivitách společnosti a sekcí.

Jinou centrálně organizovanou aktivitou je statistické školení. VVS definovala několik odborných úrovní, vyvinula studijní programy a organizuje zkoušky; vlastní (večerní) výuka je přenechána speciální, pro tento účel zřízené a finančně od VVS oddělené organizaci. Na nejnižší úrovni se jedná o elementární statistiku; na vrcholu je (státem chráněný) titul *Statisticus – VVS*, k jehož získání je třeba složit dílčí zkoušky dosahující univerzitní úrovně. Diplomy VVS jsou vyžadovány pro některé funkce, například v Ústředním statistickém úřadě.

Důležitou každoroční akcí společnosti je Statistický den. Dopolední program tradičně obsahuje plenární přednášku domácího či zahraničního významného statistika, odpoledne následují kratší volné příspěvky v několika paralelních sezeních. Statistického dne se pravidelně zúčastňuje 500 lidí, což svědčí o jeho úspěchu. O výroční členskou schůzi, kterou den končí, je ale zájem poněkud menší: 30–40 členů, převážně členů různých komisí a představenstev sekcí.

Předsednictvo VVS má 7 členů: předsedu (v současné době Prof. Dr. I. W. Molenaar), tajemníka, pokladníka a čtyři členy, kteří jsou předsedy pomocných komisí: koordinační komise (koordinace aktivit sekcí), komise pro vnější styky, komise pro školení a zkoušky, publikační komise.

O seznamech a tak ...

Sémantický obsah pojmu „seznam“ je velmi úzce vymezen. Nicméně šíře jeho společenského významu může být velká. Zvláště když se jedná o seznamy osob. Od seznamů, v nichž by mnozí z nás rádi četli své jméno, přes seznamy, s nimiž řada z nás každý den pracuje a vůči nimž je náš postoj zcela lhostejný až po seznamy, na nichž bychom se nikdy nechtěli objevit. Význam seznamu se navíc často mění s dobou. To souvisí s tím, že každý seznam, kromě výčtu prvků v něm obsažených, implicitně zahrnuje další informaci o svých prvcích. Tato informace je neoddělitelně spjata s kontextem, v němž byl seznam vytvořen. V tomto směru se jedná o pojem statistický. Jeho interpretace zcela podléhá všem nástrahám, stejně jako interpretace výsledků statistické indukce. Použití seznamu v jiném, byť „velmi podobném“ kontextu už může zcela změnit jeho význam. Tím se může stát velmi nebezpečnou zbraní. Jeho (ne)vhodnou aplikací lze napáchat – ať už úmyslně či neúmyslně – řadu škod.

Narodil jsem se v roce 1952. V té době se prý vyvěšovaly ve výkladních skříních seznamy živnostníků a jiných „nebezpečných žvlů“. Já je nepamatuji. Nicméně pamatuji dobře jiné seznamy, vypracovávané k různým příležitostem, například k 1. máji. Od té doby seznamů neustále přibývá. Stačí zajít do nejbližší trafiky a koupit si nějaký. To je jen logický důsledek současné „informační exploze“. Jsou však i seznamy, které se všeobecně považují za užitečné a bez nichž se většina z nás ani neobejde. Telefonní seznamy, seznamy účastníků konferencí, adresáře společností, profesí a další a další. Dokonce si dovolím tvrdit, že neznám žádný zbytečný seznam. Jsou pouze seznamy, které by nemusely být, kdyby Ale to „kdyby“, to je život.

A tak se konečně dostávám k seznamu, o kterém jsem původně chtěl psát a který vyvolal u mnohých z Vás smíšené pocity. K seznamu, který vlastně ani neexistuje. K seznamu, kvůli němuž jeden z našich kolegů zrušil své členství ve společnosti. **Seznam neplatičů.** Jak hrozně to zní! Budeme-li mluvit o „seznamu členů, od nichž dosud nepřišel členský příspěvek“, význam tohoto hypotetického seznamu se poněkud posune (mnohem zábavněji o tomto fenoménu píše ve své knize Helmut Swoboda). Ale stejně, tisknout takový seznam, to se přece nedělá! Ano, opravdu souhlasím, že by se to dělat nemělo. My jsme podobný úmysl ani neměli a opět zdůrazňuji, že tento seznam dosud neexistuje. Nicméně, mezi slušnými lidmi se nedělají ani jiné věci. Například neplnit své závazky, jakým je i placení členského příspěvku. Ten kolega, který byl „hrozbou seznamem neplatičů“ (přiznávám, že nevhodnou) tak pobouřen, navrhuje každého, kdo nezaplatil, prostě vyloučit. A to se dělá? Nejsme přece hokynáři, kteří když nezaplatíte, tak Vám nic nedají. Členství ve společnosti by mělo být něco víc, než jen povinnost platit příspěvky a naopak, společnost není pouze institucí pro vybírání příspěvků.

Seznamem neplatičů už vyhrožovat nebudu a všem, kterých jsem se tím dotkl, všem se hluboce omlouvám.

Gejza Dohnal.