Informační Bulletin

České Statistické Společnosti číslo 1, duben 1996, ročník 7.

Aktuální poznámka k mírám inflace

Josef Kozák

Míry inflace

Začnu definicemi převzatými z [1] a [2] .

(1) Výchozí mírou inflace je index spotřebitelských cen $I_t, t = 1, \ldots, T$, mající charakter basického cenového indexu, a to v současnosti se základem „prosinec 93 = 1“. [1] Index spotřebitelských cen je kumulativní mírou, protože pro dané t charakterizuje celkovou změnu cenové hladiny ve srovnání s danou výchozí základnou. Protože cenová hladina v čase obvykle roste, má posloupnost indexů spotřebitelských cen zpravidla charakter monotonně rostoucí posloupnosti v čase.

(2) První skupinu odvozených měr představují měsíční míry. Jde v prvé řadě o míru charakteru indexu spotřebitelskch cen s proměnlivým základem „předchozí měsíc = 1“

$$M_t = \frac{I_t}{I_{t-1}}, \qquad t = 2, \ldots, T.$$

Vedle toho se používá též míry typu tempa růstu

$$m_t = 100(M_t - 1), \qquad t = 2, \ldots, T,$$

tj. jde o relativní přírůstek vyjádřený v procentech předchozího měsíce.

[1] V této souvislosti je třeba se zmínit,že takové údaje v dostatečně dlouhé časové řadě vyžadují přepočty „starého“ indexu s výchozí základnou „leden 89 = 1“, což je problematika, jíž je věnována pozornost v zatím nepublikovaném materiálu [3].

(3) Dalšími jsou roční míry. V prvé řadě opět jde o míru typu indexu

$$R_t = \frac{I_t}{I_{t-12}}, \qquad t = 13, \ldots, T,$$

tj. indexu spotřebitelských cen s proměnlivým základem „stejný měsíc minulého roku = 1“; časté je též vyjádření typu tempa růstu

$$r_t = 100(R_t - 1), \qquad t = 13, \ldots, T,$$

tj. jde o relativní přírůstek vyjádřený v procentech stavu stejného měsíce daného roku.

Roční průměry

Měsíčně je v mediích obvykle používána měsíční míra typu tempa růstu. Naopak roční míry téhož typu je používáno poměrně zřídka; v mediích přicházejí v úvahu zpravidla pouze v posledním měsíci daného roku, a to nikoliv ve formě výše uvažovaných charakteristik, nýbrž ve formě průměru ročních měr v jednotlivých měsících daného roku. Této problematice zde chci věnovat pozornost.

Poznámka 1: Při analýze časových řad obvykle vystačíme s časovou proměnnou ve formě posloupnosti pořadových čísel $t = 1, 2, \ldots, T$. Někdy je užitečné pracovat též s posloupností pořadových čísel let $i = 1, 2, \ldots, n$, a posloupností pořadových čísel měsíců $j = 1, \ldots, 12$. Umístíme-li $j = 1$ a $i = 1$ do $t = 1$, platí mezi t_{ji} na jedné straně a (j, i) na druhé straně vztah

$$t = t_{ji} = j + 12(i - 1), \quad j = 1, \ldots, 12, \quad i = 1, 2, \ldots, n.$$

Úvahy o průměrných mírách inflace začněme takto. V prosinci i-tého roku, tj. v bodě

$$t_{12,i} = 12 + 12(i - 1) = 12i, \qquad i = 2, \ldots, n,$$

je k disposici posloupnost měsíčních měr

$$M_{1i}, M_{2i}, \ldots, M_{12,i}, \qquad i = 2, \ldots, n,$$

a posloupnost ročních měr

$$R_{1i}, R_{2i}, \ldots, R_{12,i}, \qquad i = 2, \ldots, n.$$

Vedle výše uvedených charakteristik lze pak míru inflace v i-tém roce vyjádřit jako průměr z uvedených charakteristik; protože se jedná o indexy, je obvyklé uvažovat prosté geometrické průměry M_{pi}, resp. R_{pi} pro $i = 2, \ldots, n$,

jejichž přirozené logaritmy jsou rovny

$$\log(M_{pi}) = \sum_{j=1}^{12} \log(M_{ji})/12, \qquad i = 2, \ldots, n,$$

$$\log(R_{pi}) = \sum_{j=1}^{12} \log(R_{ji})/12, \qquad i = 2, \ldots, n,$$

kde sčítací index j označuje měsíce. Na uvedené průměry typu indexů pak navazují průměry typu temp růstu

$$m_{pi} = 100(M_{pi} - 1), \qquad i = 2, \ldots, n,$$

jako další měsíční míra, resp. průměry stejného typu

$$r_{pi} = 100(R_{pi} - 1), \qquad i = 2, \ldots, n,$$

jako další roční míra.

V této souvislosti je užitečné uvést dva výsledky:

(i) Zřejmě

$$\begin{aligned} 12\log(M_{pi}) &= \log(M_{1i}) + \log(M_{2i}) + \cdots + \log(M_{12,i}) = \\ &= \log(I_{1+12(i-1)}) - \log(I_{1+12(i-1)-1}) + \\ &+ \log(I_{2+12(i-1)}) - \log(I_{1+12(i-1)}) + \\ &+ \cdots + \\ &+ \log(I_{12+12(i-1)}) - \log(I_{12+12(i-1)-1}) \\ &= \log(I_{12+12(i-1)}) - \log(I_{12+12(i-1)-12}), \end{aligned}$$

takže s ohledem na definici roční míry typu indexu máme

$$12\log(M_{pi}) = \log(R_{12+12(i-1)}), \qquad i = 2, \ldots, n,$$

ze kterého vyplývá, že v i-tém roce, $i = 2, \ldots, n$, roční průměr měsíčních měr má stejnou vypovídací schopnost jako prosincová roční míra v daném roce. To je minimálně druhý důvod toho, proč užívat roční míry typu indexu aspoň v posledním měsíci každého roku.

(ii) Zřejmě dále

$$\begin{aligned} 12\log(R_{pi}) &= \log(R_{1i}) + \log(R_{2i}) + \cdots + \log(R_{12,i}) = \\ &= \sum_{j=1}^{12} \log(I_{j+12(i-1)}) - \sum_{j=1}^{12} \log(I_{j+12(i-2)}); \end{aligned}$$

protože však veličina

$$\log(I_{pi}) = \sum_{j=1}^{12} \log(I_{j+12(i-1)})/12, \qquad i = 1, \ldots, n,$$

smí být interpretována jako logaritmus prostého geometrického průměru posloupnosti všech dvanácti indexů spotřebitelských cen v i-tém roce, $i = 1, \ldots, n$, lze roční průměr ročních měr typu indexu vyjádřit též jako

$$R_{pi} = \frac{I_{pi}}{I_{p,i-1}}, \qquad i = 2, \ldots, n,$$

tj. jako podíl sousedních průměrných ročních indexů spotřebitelských cen.

Vážené průměry ?

Po mém soudu má uvedená metodologie následující interpretační nedostatek: v i-tém roce ($i = 2, \ldots, n$) má při kostrukci prostého geometrického průměru I_{pi} index spotřebitelských cen $I_{j+12*(i-1)}$ pro libovolné $j = 1, \ldots, 12$ stejnou váhu. Protože výpočet ročního průměru roční míry je prováděn v prosinci i-tého roku, bylo by po mém soudu správnější počítat vážený geometrický průměr s vahami j v jednotlivých měsících $j = 1, 2, \ldots, 12$, tj. počítat jej podle vzorce

$$R_{vi} = \frac{I_{vi}}{I_{v,i-1}} i, \qquad i = 2, \ldots, n,$$

kde $I_{vi}, i = 1, 2, \ldots, n$, je vážený geometrický průměr indexů spotřebitelských cen v daném roce, kde s ohledem na to, že $\sum_{j=1}^{12} j = 78$, jeho přirozený logaritmus spočítáme podle vzorce

$$\log(I_{vi}) = \sum_{j=1}^{12} j \log(I_{j+12*(i-1)})/78, \qquad i = 1, \ldots, n.$$

Poznámka 2: Pro ilustraci uvádím tabulku převzatou z [3], ve které jsou uvedeny geometrické průměry spočítané dvojím způsobem – prostě i váženě.

$i+77$	I_{pi}	R_{pi}	I_{vi}	R_{vi}
78	0.3401	.	0.3408	.
79	0.3532	1.0383	0.3579	1.0502
80	0.3695	1.0462	0.3696	1.0325
81	0.3723	1.0076	0.3729	1.0089
82	0.3908	1.0498	0.3920	1.0513
83	0.3945	1.0096	0.3947	1.0070
84	0.3987	1.0105	0.4001	1.0136
85	0.4077	1.0226	0.4083	1.0205
86	0.4093	1.0041	0.4096	1.0031
87	0.4097	1.0008	0.4098	1.0006
88	0.4105	1.0020	0.4109	1.0027
89	0.4164	1.0144	0.4172	1.0154
90	0.4557	1.0945	0.4696	1.1256
91	0.7125	1.5635	0.7299	1.5542
92	0.7942	1.1146	0.8070	1.1056
93	0.9573	1.2054	0.9705	1.2026
94	1.0520	1.0989	1.0656	1.0980
95	1.1484	1.0916	1.1600	1.0886

Protože indexy spotřebitelských cen mají charakter monotonně rostoucí posloupnosti v čase, nepřekvapuje, že jejich vážené geometrické průměry I_{vi} jsou v porovnání s prostými geometrickými průměry I_{pi} pro $i = 1, \ldots, 18$, vesměs vyšší. Stojí však za povšimnutí, že taková relace neplatí mezi prostými a váženými ročními průměry ročních měr inflace typu indexu R_{pi} a R_{vi} pro $i = 2, \ldots, 18$: tak například pro rok 1995 nedávno mediálně vychvalovaný výsledek o prostém ročním průměru typu tempa růstu na úrovni 9.16 % by mohl být nahrazen ještě příznivějším výsledkem 8.86 % získaným na bázi váženého geometrického průměru!

Citovaná literatura

[1] Křovák, J., Kociánová, S.: *Využití současných indexů cen pro měření inflace.* Statistika 10/1992, 425–430

[2] Kozák,J.: *Současná česká inflace: reálie a tendence.* Statistika 3/1995, 124-136

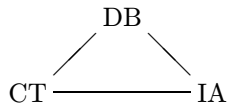
[3] Kozák,J.: *Poznámky k českým mírám inflace.* Nepublikovaný rukopis, leden 1996.

Struktura závislosti

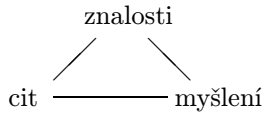
Vzpomínka na Tomáše Havránka (* 11.7.1947 - †17.5.1991)

Marta Horáková

Každý, kdo se během své profesionální dráhy alespoň jednou setkal s doc. RNDr. Tomášem Havránkem, DrSc., nutně se k těmto setkáním v myšlenkách vrací. Já jsem se s ním setkávala přibližně od roku 1982 v článcích, ve skriptech, na seminářích, na přednáškách a na konzultacích. To nejobecnější, co mně z těchto setkání utkvělo v mysli, jsou nápisy CT, IA a DB na třech šuplících v Tomášově pracovně v bývalém SVT ČSAV. Každý z těchto šuplíků obsahoval separáty s obrovským bohatstvím myšlenek z každé z uvedených oblastí. Všechny tyto poznatky zcela jistě spojovaly právě ony cedulky s označením oboru: DB = databáze, IA = umělá inteligence, CT = kontingenční tabulky. Uvažujeme nyní DB jako symbol báze dat popisujících reálný svět hodnotami diskrétních a spojitých veličin, CT jako symbol toho, co jakýmsi způsobem dokáže popsat strukturu závislosti oněch diskrétních a spojitých veličin, a IA jako symbol pravidel, kterými se lze dobře pohybovat mezi fakty. Vzájemnou spolupráci, provázanost a souvislost těchto tří oblastí nyní jednoduše označme



Různá volba naplnění šuplíků DB, CT, IA odpovídá modelům různých oblastí reálného světa. Já se teď věnuji především výchově dítěte s ekvivalentním modelem



Šuplíky v Tomášově pracovně však osahovaly poznatky směřující k tomu, aby tento model mohl být pevnou kostrou pro expertní systémy.

Nově, lépe, radostněji?
(Zamyšlení nad statistickým systémem Statgraphics PLUS for Windows)

Hana Řezanková

Žádostivě jsem očekávala novou verzi oblíbeného statistického systému Statgraphics, tentokrát pro Windows. „Konečně snad budou sloupečky s hloubkou a výseče s výškou, jistě bude perfektní nápověda, v rámci tabulky pro vkládání dat by mohly být snadno vytvářeny odvozené proměnné (možná, že se dají přepočítat při změně původních údajů…).“

Modul základní = Statgraphics po odtučňovací kůře

Instalace základního modulu [2] probíhala ve dvou fázích. Ve verzi 1.01 nebylo možné vybírat proměnné ze zobrazené nabídky – ta sloužila k tomu, aby uživatel mohl jméno přepsat pomocí klávesnice do vymezeného rámečku. Po přeinstalování verzí 1.11 však byla chyba napravena.

Přesto však byly mé pocity značně smíšené. „To jsou perfektní nabídky! Kreslení grafů, popis dat porovnávání, závislosti. Ale co to je v popisu dat? Dělení dat na číselná a kategoriální? Cožpak číselná proměnná počet dětí není kategoriální?“

Otvírám starý známý vzorový datový soubor, obsahující údaje o automobilech z let 1979–1982 (soubor pro pamětníky) a spouštím výpočet pro popis jedné číselné proměnné, kterou vybírám z nabídky. Výsledkem je pouze sdělení počtu hodnot a minimální a maximální hodnoty. „No, bude třeba někde něco zadat. Á, k čemupak je asi tato ikona s listem papíru? Konečně nabídka výpočtů. Zaškrtnu raději všechno, ať vidím, co to dělá. Co se to stalo s oknem výstupů? Samé rámečky s lištami! Tak si nějaký rámeček zvětším, třeba ten druhý se souhrnnými charakteristikami. Hm …, a teď to chci zpátky.“ Ještě zkusím pravé tlačítko – zobrazuje se nabídka souhrnných charakteristik. Pak nastavuji třetí rámeček (kvantily), čtvrtý (tabulka četností), pátý (číslicový histogram), …. „No, některým uživatelům neuškodí procvičit se trochu v ovládání ve Windows. To je přece mnohem důležitější než nějaká statistika!“

„Ikona s grafem bude zřejmě pro kreslení grafů.“ Zajímají mě opět všechny. Okno výstupů se rámečkuje dál. Obsahuje již 7 rámečků s textovými výstupy a 7 rámečků s grafy.

[2] Instalace byla provedena na PC 486DX2/60MHz, 16 MB RAM

Zkouším první ikonu. „Copak asi znamená? Á, výběr proměnných.“ Ťukám na jinou proměnnou. „Co to? V okně výstupů se vše přepočítalo a překreslilo! No, to je báječné! Ale kde jsou předchozí výpočty? Prostě zmizely bez varování . . . “

„Jaképak jsou asi možnosti v oblasti práce s daty?“ Zavírám okno analýz a otvírám tabulkový editor s otevřeným datovým souborem. Označuji první prázdný sloupeček a ťukám na pravé tlačítko myši. Zobrazuje se dialogový rámeček, v němž lze změnit jméno, komentář a typ proměnné. V nabídce Edit nalézám stejnou možnost a navíc možnost generování hodnot (nejen náhodné, ale i vytváření odvozených proměnných) a řazení pozorování.

„Tak tedy zkusím generování hodnot.“ Zobrazuje se dialogový rámeček, který lze využít též jako kalkulačku (výsledky jsou přístupné pomocí tlačítka Display). K dispozici je mnohem více operátorů než ve verzi pro DOS, nové existují např. pro generování náhodných čísel (dokonce ze sedmi rozdělení místo původních devatenácti – to není překlep!), ke zjišťování hodnot distribuční funkce a kvantilů (z celých pěti rozdělení místo původních osmnácti). A prozradím něco dopředu: v chí-kvadrát testu dobré shody jsou k dispozici 4 (slovy čtyři) rozdělení (místo původních osmnácti).

Takže si chci vygenerovat novou proměnnou. Vybírám z nabídky funkci RNORMAL(?,?,?). Na vysvětlení parametrů si volám nápovědu: „RNORMAL(2,3.2,4.9)“. Význam je díky příkladu s desetinnými místy hned jasný. Nač popisovat parametry? Kdo nezná řádně statistiku, ať nepočítá. Ve Windows mám nastavenu desetinnou čárku. „Jakpak asi Statgraphics pozná, co je desetinná čárka a co oddělovač parametrů?“ Nepozná! („Not enough arguments . . . “). Jedinou cestou je přepnout ve Windows zobrazení čísla s desetinnou tečkou.

Spouštím další proceduru, předchozí okno výstupů se schovává do ikonky a objevuje se nové. V reklamních materiálech se píše, že při změně dat (opravě či načtení jiného souboru) se vše přepočítá a překreslí. Měním hodnoty, nic. Schovám okno do ikonky, znova ho otevřu – a je to tady! Vše perfektní. Téměř. V grafu je zakreslena polovina hodnot. Ty, které jsem změnila, jsou mimo původní interval. Přizpůsobit měřítko na osách? To už by uživatelé byli moc rozmazlení. (Nutno však poznamenat, že autoři Statgraphicsu se nad uživateli zželeli a v další verzi tuto chybu opravili.)

Podívám se ještě na regresní analýzu. Zajímá mě nelineární. V názvech doplňkových modulů není, měla by být v základním systému. Jednoduchá, vícenásobná, dost. Nelineární regresní analýza byla zcela vypuštěna! „Ale těch modelů v jednoduché regresi: S-křivka, logistická křivka, . . . “, celkem

12. Jejich vhodnost je porovnávána na základě korelačních koeficientů – nová míra pro nelineární závislost.

Soustředím se raději na grafy. Specializovaných statistických (např. ověřování normality) je o něco více než ve verzi pro DOS. „A co obchodní?“ Zvětšuji rámeček se sloupkovým grafem (lze vytvořit z nabídky grafů i z nabídky pro popis dat). Označuji graf levým tlačítkem myši a tisknu pravé. Popisy, barvy. „Především změnit tu černou kolem grafu (chudák tiskárna!).“ Nelze!!! A navíc žádná hloubka, žádné popisy sloupečků četnostmi. Zmizela i DOSovská možnost kreslení grafů pro případ, kdy v tabulce jsou již četnosti.

Tak si alespoň nakreslím graf nějakého rozdělení, když se obchodní grafy spíše zhoršily. „Ale jak je zadat?“ Přece nijak! Žádné takové grafy již kreslit nelze.

Moduly ostatní – jen pro silné nervy

Po prozkoumání základního modulu jsem se „vrhla“ na moduly ostatní. K dispozici jsem měla následující směsici:

Multivariate Statistics – verze 1.4,
Time Series Analysis – verze 1.2,
Quality Control – verze 1.11
Experimental Design – verze 1.3.

Každý z těchto modulů bylo třeba instalovat samostatně. (Škoda, že jich není více ...). Při každé instalaci dalšího modulu se přepsaly soubory SGWIN.EXE a SGWIN.HLP (číslo verze systému Statgraphics je vázáno na číslo verze naposledy instalovaného modulu).

Ačkoliv samostatný základní systém fungoval celkem bez potíží, po naistalování dalších modulů se vyskytly následující chyby.

1. Po otevření tabulkového editoru a načtení vzorového datového souboru CARDATA.SF se systém vždy samovolně ukončil (tuto chybu jsem obešla tak, že jsem nejprve otvírala datový soubor a pak teprve tabulkový editor).

2. Při načítání vzorových analýz hledal systém vzorová data v neexistujícím adresáři `c:\sgwin\data` (příslušný datový soubor však bylo možné otevřít dodatečně z vhodného adresáře).

3. Práci se systémem nebylo možné korektně ukončit (výjimkou bylo ukončení následující ihned po spuštění). Ukončení bylo nutné provést tak, že jsem se přepnula do správce programů a zadala ukončení Windows. Nebyla

však ukončena práce s Windows, ale se systémem Statgraphics (včetně dotazů na uchování souborů).

Závěr

„Dost bylo hraní, zkusím napsat závěrečnou zprávu. Text a grafy dohromady nedám, tak to zkusím zatím pouze s textem.“ Otevírám okno komentářů. Přepínám se do prvního okna výstupů. Kopíruji text, přenáším ho do okna komentářů. Jeden rámeček, druhý, třetí, ... Není nad jednoduchost. Že nově, to je bez debaty. Zda lépe, to jak v čem. Ale radostněji, to ani náhodou.

Dodatek

Člověk se dosti nerad smiřuje s tím, že něco, od čeho hodně očekával, ho zklamalo. „Určitě je něco chybného na instalačních disketách, vždyť nevím o tom, že by si uživatelé ztěžovali. Určitě ostatním uživatelům vše perfektně funguje.“ Protože mi k recenzi byly původně poskytnuty pouze kopie instalačních disket, použila jsem tentokrát originální. Etikety na disketách oznamovaly že základní systém je ve verzi 1.1, zatímco doplňkové moduly ve verzi 1.0.

Na disku jsem se snažila odstranit všechny soubory z předchozích verzí, i všechny záznamy o Statgraphicsu v souborech systému Windows. Výsledkem instalace bylo, že systém nešel vůbec spustit (instalační program vůbec nepožadoval disketu číslo 3). Obnovila jsem podadresář v adresáři Windows, kde byla původně nainstalovaná 32-bitová podpora. Rozběhl se! Po zadání Open Data File se však systém opět sám ukončil. „Zkusím přihrát další modul, třeba se to zlepší.“ Ačkoli na disketě bylo napsáno „1.0“, systém se hlásí jako verze 1.2 (šlo o zcela stejné verze, jako jsem měla původně). Soubor CARDATA.SF nelze použít vůbec, ostatní soubory však ano. Jména proměnných do zadání výpočtů je třeba zapisovat pomocí klávesnice (nabídka je skutečně pro parádu). „Tak si konečně spustím časové řady.“ Ťukám na položku *Time Series Analysis*. Systém se samovolně ukončuje, já končím také. Navždy.

Poznámka redakce

Autorka svůj slib nesplnila a nainstalovala systém na jiném počítači. Až do třetího doplňkového modulu se žádná z výše uvedených závad neobjevila. Pak zkusila čtvrtý modul (z původního balíku disket). Začaly problémy. Po odinstalování systému a novém nainstalování (i bez čtvrtého modulu) se opět nedalo pracovat ....

Postscriptum o české statistické terminologii.

Zbyněk Šidák

Mnohonásobným zkracováním článku [4] se bohužel stalo, že některá místa v něm připouštějí dvojí výklad, což bych chtěl nyní aspoň dodatečně napravit.

1. V odst. 1 v [4] v souvislosti s češtinářskými poklesky (chybné užívání čárek, slova „standartní", „vyjímka", atd.) jsem napsal pro stručnost odkaz „viz [1]". To ovšem *neznamená*, že by J. Žváček se těchto poklesků dopouštěl, nýbrž přesně *naopak*, že je pranýřuje, v čemž se shodujeme.

2. V odst. 2 v [4] jsem použil slov „hnilobný rozklad" v uvozovkách při připomínce poznámky v J. Andělovi [2]. Uvozovky zde však *neznamenají*, jak by se mohlo zdát, že jde o doslovný citát z J. Anděla [2]; těchto slov použil teprve J. Machek [3], odst. 3. (Problém je v tom, že ve [4] používám uvozovek někdy pro vyznačení citací, někdy však jen pro zvýraznění určitých slov.)

J. Žváčkovi i J. Andělovi se za tato možná nedorozumění omlouvám.

Nyní ještě několik dalších drobných poznámek.

3. V odst. 5 v [4] projevuji svou nelibost nad podivným českým slovem *skór*. Mezitím však v časopise Vesmír (**74** (1995), č. 5, str. 297) vyšel článek, v němž jazykový odborník v podstatě toleruje tento český termín s tím, že prý (citace) „se tento výraz začal objevovat v pracích našich psychologů". Toto konstatování je pravdivé a přitom autor zřejmě neví, že tento výraz se již dlouho objevuje i v matematické statistice, možná i někde jinde.

4. S velkým zájmem sleduji „jazykové koutky" v novinách a časopisech a zdá se mi, že jazykoví odborníci nyní většinou dost tolerují nová slova a nové jevy v češtině, které jsou přinášeny vývojem jazyka. Ačkoliv názory v mém článku [4] mohly snad někdy vyznívat dost striktně, přesto pod jejich povrchem je rovněž jistá míra tolerance, podobně jako u J. Machka, P.S. k článku [3]. Jednak jsem se vždy pokládal za matematika (se zaměřením na statistiku), takže pro mne je nejdůležitější jednoznačná definice, resp. věta, ne už tolik název, termín, jednak jsem smířen s tím, že terminologie se časem mění (viz stejná poznámka v [2] a [3], odst. 13). Ke známému sporu kolem *rozdělení* či *rozložení*, postupné změně termínů *matematická naděje* a *očekávaná hodnota* na *střední hodnotu* (viz též [2]) je možno přidat další,

např. souběžné užívání rovnocenných termínů *σ-algebra jevů* a *jevové pole, nestranný odhad* a *nevychýlený odhad*, *vydatnost* a *eficience*, atd. Hlavně když si jednoznačně rozumíme!

5. Horší je to někdy s termínem *směrodatná odchylka* ve článcích, v nichž nevíme, zdali je ve jmenovateli použito *n* nebo *n-1* (viz též nedávná série článků o tom v IB ČSt Sp). A vůbec nejhorší jsou věty typu „ze zjištěných dat byla vypočtena střední hodnota 23,7 ± 4,2", které se leckdy vyskytují v článcích aplikujících statistiku. Těch 4,2 je *snad* odhad směrodatné odchylky odhadu střední hodnoty, ale jak definovaný a s *n* nebo *n-1* ve jmenovateli? Ale také by to mohl být odhad populační směrodatné odchylky; nebo také celý výraz by mohl znamenat interval spolehlivosti pro střední hodnotu. Takové věty mají daleko k jednoznačnosti a měli bychom je kritizovat a požadovat jasné vysvětlení!

6. Kromě jazykových chyb typu „standartní", „vyjímka" atpod., se vyskytují také chyby matematické a (řekněme) subtilnější, např.: V nepříliš dávné době někteří naši pedagogové byli tak posedlí množinovým pojetím matematiky, že viděli množiny i tam, kde vůbec nebyly. V oné době jsem v učebnicích čítal věty typu „nechť náhodný výběr je množina {5, 12, 7, 3, 5, 9}". Takto pojatý náhodný výběr ovšem *není* množina; v náhodném výběru se čísla mohou opakovat, kdežto v množině nikoliv, číslo buď v ní je nebo není (naivně řečeno, číslo v množině může být nejvýše jedenkrát). Takovéto používání termínů je ovšem opravdu chybné!

7. Znovu bych připomněl, zdali někdo ze znalců historie české statistiky by mi uměl vysvětlit, proč v termínu *směrodatná odchylka* používáme tak nevhodného a vlastně nesmyslného přídavného jména. Slovo *směrodatný* by mělo označovat něco, co udává *směr*, ale tomu tak v našem případě rozhodně není.

Literatura

[1] Žváček, J.: *Kšaft umírající matky – české statistické terminologie.* Informační bulletin České statistické společnosti **5** (prosinec 1994), 20–23.

[2] Anděl, J.: *K problémům české statistické terminologie.* Informační bulletin České statistické společnosti **6** (duben 1995), 4–7.

[3] Machek, J.: *Terminologické úvahy.* Informační bulletin České statistické společnosti **6** (duben 1995), 8–11.

[4] Šidák, Z.: *Znovu k české terminologii: pohled z jiné strany.* Informační bulletin České statistické společnosti **6** (říjen 1995), 11-14.

Ještě k analýze více proměnných.

Jiří Žváček

Vzhledem k rozsahu a jednostrannosti jazykové kampaně cítím potřebu odvést pozornost od nenáviděné „vícerozměrné analýzy“ a vrátit se k původnímu záměru.

Snažil jsem se sice rozzuřit všechny statistiky, zaměřil jsem se však zejména na ty, kteří používají slůvko „multivariační“. Mnohorozměrná analýza mi vcelku nevadí. Jsem dokonce ochoten poskytnout zatím nepoužitý argument: Pod vícerozměrnou analýzou studovali učenci jednoho dnes již zaniklého středoněmeckého státního útvaru i případ, kdy proměnných je nekonečně mnoho (mehr–viel)!

Nepsal jsem ale o správnosti „vícerozměrné analýzy“, pouze jsem uvedl, že „historicky převládá“.

Za chybu v „analiz“ se lidstvu omlouvám. Staré indiánské přísloví říká „Wo man einen Waschbär sieht, gibt es hundert.“ Nalezl jsem v článku jen při letmém pohledu další tři. Možná, že jsem přeci jenom měl dostat korekturu.

Nicméně další argumentaci už nepokládám za tak nespornou.

Vůbec mi nevadí, že „multi“ se častěji překládá „mnoho“. Ve vědě moc demokracie není. Ostatně „variate“ bych také nepřekládál „rozměr“ a nikdy jsem netvrdil, že „vícerozměrná“ je překladem z angličtiny.

O sémantické příbuznosti jsem hovořil proto, že se domnívám, že česká (i statistická) terminologie se odvíjela od německé (v článku „nad Řípem“). Příkladů nalezneme přehršel – není dávno doba, co se střední hodnota Verteilungu nazývala očekávanou. I pojem „multivariate analysis“ patří do těchto dob a odtud i slůvko „historicky“.

Pokládám navíc němčinu za sémanticky podobnou. Zavinil to mj. i Mladýmuž a jeho následníci. Jako příklad může posloužit vládnoucí „pták“. V češtině a němčině je význam stejný, anglicky bych použil v některých situacích kohouta a ruština je vzhledem k ženskému rodu poletavce omezeně použitelná.

Odmítám nařčení, že „rusofilský“ používám v pejorativním smyslu. Je to spíše jistý hold emigrantům, kteří významně přispěli rozvoji česky psané statistiky v předválečném germanofilském intelektuálním prostředí! Ruský název navíc pokládám za výstižnější než anglický i německý, protože vlastně nejde o proměnné ani rozměry.

Ani převládání není náhoda.

Veliký Ocelář kdysi uzákonil tézi o zostřování třídního boje. Pro ty, kteří vyklouzli nedotčeni ideologií, připomínám: *Čím je nepřátel méně a čím více jsou dušeni, tím ostřejších prostředků používají.*

Počty vícerozměrných studentů statistiky na VŠE dnes již o dva řády převyšují počty mnohorozměrných statistiků na MFF, takže budoucnost vícerozměrné analýzy v demokratické společnosti vidím růžově.

Zejména proto, že mnohorozměrnou hájí ti, pro které je i jedna (za určitých okolností) mnoho.

Mám samozřejmě i vážnější důvod pro slůvko „převládá“. Podívejte se do česky psaných článků, recenzí a skript! Pokud budou statističtí koryfejové publikovat anglicky v kybernetických časopisech, nemohou se podivovat, že se česká statistická terminologie vytváří mimo ně!

PS. Volnost americké statistické terminologie mi ostatně jde trochu na nervy a nevidím v ní vzor. Populární Statgraphics, pro většinu studentů vrchol statistiky, již například používá One–Variable a Multiple–Variable analysis (a přidává třetí možnost – Distribution Fitting).

Násilné počešťování také nemám rád. Očekávaná hodnota byla lepší než střední a znovu vyzývám i při této příležitosti k boji proti národní šikmosti a špičatosti (což symetrie a koncentrace?)!

Ordnung sollte sein ...

Ať žije KOČKA

Petr Hebák

Už si přesně nepamatuji, kdy to začalo, ale odborným asistentem jsem už myslím byl, a tím jsem byl jmenován – po třech letech práce pedagogického asistenta na katedře statistiky – za 1 074 Kčs měsíčně hrubého v roce 1965. Rovněž ani nevím, kdy naše tehdejší snažení skončilo, ale jsem si skoro jist, že po roce 1968 jsme už v akci KOČKA nic nového nevytvořili. Formálně uvažováno existuje KOČKA dodnes, protože ji nikdo nikdy nezrušil, ale výsledky její činnosti jsou už dávno zapomenuté a leží snad někde na dně některého z mých šuplíků. Vedení katedry a vážení páni docenti neměli pro tuto naši iniciativu příliš pochopení a celá věc nějak zapadla. Snad se na mne nebudou Jana Kahounová, Honza Seger, Jirka Hustopecký ani Jirka Žváček nijak zlobit, že jsem si dovolil je zde (jen tak a bez titulů a funkcí)

jmenovat a že dokonce míním připomenout důvody i cíle, kvůli kterým tehdy KOČKA vznikla. Nemohu v této souvislosti nevzpomenout na předčasně zemřelého Jirku Kejkulu, který se jako jeden z prvních na katedře statistiky VŠE začal zabývat metodami, které si zatím, vzhledem k terminologickým debatám a zaměření tohoto příspěvku, netroufám jednoznačně pojmenovat. Tento můj častý šachový soupeř, náš kolega a nadaný statistik, by teď musel vysvětlit, proč preferoval pojmenování „vícerozměrné“ statistické metody před pojmenováním „mnohorozměrné“ statistické metody. Neúmyslně jsem možná na někoho zapomněl, ale každopádně jsme se všichni tenkrát marně snažili, aby KOČKA byla úspěšná.

Je na čase říci, co to byla KOČKA. Dali jsme si za cíl navrhnout, dohodnout a nejméně v rámci katedry prosadit a dodržovat, aby se u nejčastěji se vyskytujících pojmů na přednáškách, při cvičeních, při psaní skript a učebnic používala *jednotná statistická terminologie a symbolika.* Říkali jsme si, že je vůči studentům značně nespravedlivé, když různí přednášející či cvičící hledají raději synonyma a skoro až záměrně nestejně pojmenovávají jednoznačně vymezené statistické kategorie, jakož i neobyčejně pestře odlišně označují některé z používaných charakteristik. Není třeba asi vysvětlovat, proč je nešťastné, když se nestatistik obtížně orientuje ve statistické literatuře jen proto, že různí (někdy dokonce i v nestejných kapitolách stejní) autoři používají odlišné názvy a nestejné symboly pro označení stejného pojmu. Připouštěli jsme, že názory na výběr „nejlepšího“ či „nejvhodnějšího“ termínu či symbolu nebudou shodné a bude často nutné smířit se s kompromisy. Ani hledání „nejlepšího“ překladu pojmů „nejčastěji“ se vyskytujících v zahraniční literatuře není cestou, která vede k *objektivně správnému* pojmenování či označení. Někdy stačilo, že členové katedry měli odlišné pocity a názory, od kterých nebyli v zásadě ochotni ustoupit a které jim nedovolovaly se dohodnout. Výsledkem návrhů a dlouhých debat byl materiál, který na několika desítkách stran prezentoval tehdejší názor většiny zúčastněných i výsledky učiněných kompromisů. Z dnešního pohledu bychom celou řadu tehdy dohodnutých „nejlepších“ termínů a symbolů asi vůbec nepřijali či přijali jen s výhradami, o jiných bychom dlouho diskutovali, ale je možné, že na některých bychom se snad i po letech všichni shodli. Ke společným debatám na toto téma již později nikdy nedošlo, a tak snad jedině „uznávané katedrální učební pomůcky“, respektující ve větší či menší míře státní normu a zahraniční zvyklosti, umožňovaly přednášejícím i cvičícím se aspoň částečně shodnout na používané terminologii a symbolice. Ti, co publikovali v SPN či SNTL, byli do určité míry i pod tlakem jazykových i odborných redaktorů, kteří trvali na jistém výkladu státní normy. Pamatuji na mnohé až spory se svojí milovanou manželkou, která dlouhá léta

redigovala statistické učebnice v ekonomické a polytechnické redakci SNTL, zda je správné říkat a psát Rao-Cramérova věta či Raova-Cramérova věta, t test, t-test či test t, zda dvojné či trojné dolní (horní) indexy je tiskárna schopna zvládnout a zda například resp. a popř. jsou výrazy, které vyjadřují totéž. Každopádně existovaly a existují různé argumenty a názory hovořící ve prospěch nestejných názvů či podporující odlišné značení. Už jsem si vůbec nemyslel, že bych se někdy zapojil do diskuse o tom, zda toto či jiné pojmenování nebo označení je lepší než jiné. Smířil jsem se s tím, že zatímco někdy je statistická terminologie a symbolika velmi výstižná a zcela vhodná, jindy je sporná či nešťastná, nebo dokonce jednoznačně nevýstižná a v rozporu s češtinářským výkladem daného slova.Všichni víme, že české překlady jsou jen vzácně lepší než zahraniční originály (i když se i takové případy dají najít). Známe mnoho příkladů ne zcela výstižných překladů do češtiny a je možné uvést i pojmy, jejichž „jazykově správný“ překlad z angličtiny, němčiny či jiného jazyka by nedopadl úplně stejně. Navíc pro některé pojmy překlad do češtiny buď zatím neexistuje, anebo ani není rozumně myslitelný. Rozhodně jsem byl přesvědčen, že už se do žádných argumentací nepustím, ale příspěvky renomovaných statistiků ke statistické terminologii na stránkách časopisu České statistické společnosti a debaty s mladšími kolegy o správnosti či nesprávnosti používání některých termínů a symbolů mnou pohnuly natolik, že jsem se rozhodl nezůstat stranou. Kdysi jsme s Jirkou Hustopeckým v rámci činnosti VÚSEI vydali pět šestijazyčných statistických slovníků. Při jejich přípravě jsme mnohé konzultovali se zahraničními kolegy, kteří často korigovali a opravovali naše představy. Náš překlad do ruštiny, polštiny, francouzštiny, němčiny či angličtiny mnohdy neobstál a dnes bychom jistě zpochybnili nejen některé zahraniční „ekvivalenty“, ale i české „originály“. Nerad bych v čtenáři vzbudil dojem, že míním pouze říkat, že je všechno těžké a obtížné díky pochopitelně odlišným názorům a výkladům. Mám též svou představu o některých „nejvhodnějších“ českých ekvivalentech k pojmům vzniklým v zahraničí, ale o to rozhodně nejde. Jde mi především o podtržení významu **rozumné a respektovatelné dohody** před snahou o nalezení „absolutně správného a zcela výstižného“ překladu. Publikované argumenty ve prospěch určitého termínu jsou mnohdy velmi rozumné, ale rád bych ukázal, že lze najít i protiargumenty, které rovněž nelze označit za úplně nerozumné. Navíc se diskuze soustřeďuje maximálně na pět až deset pojmů a sporných situací je mnohem více. Dohoda by se měla vztahovat především na přednášky, učebnice a články určené nestatistikům, zatímco např. při výzkumu bych byl mnohem tolerantnější. Rozhodně nechci prosadit normu v ošklivém slova smyslu „předpisu otravujícího svobodné přednášení a psaní“, ale hledám

přijatelný kompromis, na jehož základě by se domácí statistici – učitelé možná mohli dohodnout.

Problémy terminologie a symboliky nás provázejí celým statistickým základním i pokročilým kursem. Jistě ani zdaleka nevzpomenu všechny, ale dovolte mi pár příkladů:

– Je vhodný pojem *základní soubor* i pro nekonečné či hypotetické soubory a není naopak nepříjemné, že „vhodnější“ pojem *populace* je v češtině nejméně dvojsmyslný?

– Je lepší označení *proměnná* či *znak* nebo je vhodnější (jeden čas na katedře kritizovaný) pojem *veličina*? Poznamenejme, že Websterův College Dictionary z roku 1995 u slova *variate* uvádí odkaz na „*random variable*“, a nopak pod *random variable* najdeme, že ve statistice *also called variate*. S tímto tedy úzce souvisí pojem *náhodná* či *náhodová veličina* či *proměnná* atd. Kdysi jsme diskutovali na katedře, zda (např. v regresi) je možné mluvit o *nenáhodných* či *pevných* či dokonce *matematických* proměnných. Argumentovalo se tenkrát, že díky chybám měření či zjišťování takové proměnné neexistují, i když je zřejmé, že při výkladu regrese je tento pojem užitečný. Jak jinak vyjádřit, zda *vysvětlující* (asi lepší než *nezávisle*) *proměnné* (asi lepší než *veličiny*) máme zcela či částečně pod kontrolou, nebo zda jsme odkázáni na výsledky *pozorování* či *náhodného pokusu* (*experimentu*). Takto by bylo možné pokračovat. Ve zmíněném slovníku najdeme pod pojmem *variable* dvanáct vysvětlení a je zde mimo jiné poukázáno na blízkou spojitost s pojmem *variabilita*.

– Je opravdu „český“ pojem *variance* věnován výhradně rozptylu, když český překlad ve slovníku nakladatelství ČSAV (K. Hais – B. Hodek) z roku 1985 sice mimo jiné obsahuje **8** STAT *čtverec směrodatné odchylky*, ale nabízí řadu dalších ekvivalentů připomínajících spíše *variabilitu* či *proměnlivost* nebo i *odchylnost*? Zaznamenejme přitom, že slovník spisovné češtiny Ústavu pro jazyk český (Academia 1978) pojem variance nezná. V této souvislosti je možné se ptát, zda *výběrový rozptyl* je dobré značit např. $\mathrm{var}(x)$, var x, $s^2(x)$ nebo s_x^2 a zda *populační rozptyl* je lepší značit S^2 či $D(X)$ nebo raději $V(X)$. Konečně, zda symbol σ^2 není lepší ponechat jen pro označení jednoho z parametrů normálního rozdělení atd.?

– Je „překlad“ anglického *expectation* jako *střední hodnota* aspoň v souladu s interpretací této nejpoužívanější *pravděpodobnostní charakteristiky*? Zmíněný Websterův slovník nás odkazuje na pojem *mathematical expectation*, pod kterým najdeme STATISTICS. „*the summation or integration over all*

values of a variate of the product of the variate and its probability or its probability density also called expectation.“ Není významově lepší *očekávaná hodnota* nebo starší *matematická naděje* atd.?

– Jak je tedy dobré překládat pojem *distribution*? Nemám k dispozici dřívější argumentaci ve prospěch nyní běžně používaného *rozdělení* a staršího *rozložení*. Anglicko-český slovník (K. Hais a B. Hodek) z roku 1984 na str. 564 nabízí několik překladů, mezi kterými jsou sice obě možnosti, ale ani jednou nejde o překlad určený statistikům. O pět pojmů dále je nabídnut statistický překlad *distribution curve* jako *křivka rozložení* a hned *distribuční funkce* jako statistický překlad pro *distribution function*. Slovník spisovné češtiny nám to příliš neulehčí, protože pod slovem *rozdělit* najdeme alternativu rozložit (ve smyslu na části) a naopak. Je *hustota pravděpodobnosti*, jako jedna z forem popisu *rozdělení* (*rozložení*) *spojité náhodné veličiny* a jako *model* „abstraktního chování pravděpodobností“, významově bližší představě o *rozložení* než představě o *rozdělení*? Rozhodně se však nechci pouštět do polemiky, zda modelování pravděpodobnostního chování *spojité náhodné veličiny* více odpovídá myšlence „rozdělování“ pravděpodobností do intervalů nebo argumentaci ve prospěch slova „rozložení“. Každopádně bych však nikdy neřekl, že používání pojmu *rozdělení* „je nevhodný jazykový uzus“.

– Je šťastné značit např. jako χ^2 jedno často používané *teoretické rozdělení* (*rozložení*), naprosto stejně veličinu s tímto rozdělením (rozložením) a s malou modifikací stejně i kvantil tohoto rozdělení (rozložení)?

– Je dobré mluvit kombinovaně anglicky a česky o *diskrétních* a *spojitých veličinách* nebo o *nespojitých* a *kontinuálních veličinách*? Nebylo by lepší zůstat např. u nekombinované české varianty *nespojité* a *spojité náhodné veličiny*?

Všichni víme, že podobný seznam otázek by mohl být mnohem delší a asi by vystačil na samostatnou statistickou publikaci. Zvláště kdybychom začali hledat lepší označení pro některé tradiční pojmy, jako je *regrese* nebo *stupně volnosti*, kdybychom chtěli zabránit odlišnému chápání v různých metodách často používaných pojmů, jako je *faktor, cluster* nebo *komponenta*, či dokonce chtěli předejít nejméně dvojsmyslnému chápání pojmů jako je *odhad, výběr* nebo *statistika*. Mohli bychom přidávat ještě další, jako je otázka nepřeložitelných termínů, problém „do sebe někdy nezapadající“ symboliky či otázka odlišného významu pojmů v různých jazycích.

Dovolte, abych se na závěr svého příspěvku ještě zastavil u tolik diskutovaného pojmu *multidimensional* nebo *multivariate analysis* (resp. *statistical methods*). Je tedy lepší vícerozměrná analýza (vícerozměrné statistické metody) či mnohorozměrná analýza (mnohorozměrné statistické metody)?

Až na příspěvěk Jirky Žváčka všechny ostatní jmenované články preferovaly překlad slova *multi* jako *mnoho* a nikoli jako *více*. Hlavní uváděné argumenty jsou:

- překlad multi = mnoho je mnohem častější než překlad multi = více,
- pojem vícerozměrná analýza je bližší němčině než angličtině,
- překlad více sugeruje otázku „více než co“,
- mnohonásobná korelace či mnohonásobný korelační koeficient nikoho nezarazí,
- **významově** je vhodnější slovo *mnoho*.

Nechci vyslovit žádný kategorický názor, ale dovolte mi uvést jiné argumenty:

- Websterův slovník z roku 1995 na str. 890 pod *multivariate* uvádí: *Statistics. (of a combined distribution) having* **more than one** *variate or variable*,
- zajímavé je sledovat nejen četnosti výskytu překladu mnoho a více, ale i rozlišení situací, ve kterých autoři tohoto i jiných slovníků dali přednost „mnoho“ před „více“ a naopak,
- překlad mnoho sugeruje otázku, zda dva už je mnoho,
- vícenásobná korelace a vícenásobný korelační koeficient též nezarazí,
- významově rozumím pod *multidimensional statistical methods* statistické postupy a metody, ve kterých současně z různých hledisek posuzujeme aspoň dvě proměnné (tedy více než jednu).

Osobně před rozhodnutím, který pojem je „správný“, preferuji dohodu, kterou bychom aspoň pro studenty a nestatistiky dodržovali. V tomto smyslu končím názvem článku „Ať žije KOČKA“.

Pokud by však naše přátelská diskuze pokračovala, doporučuji začít s překladem pojmu *multivariate* a pokračovat snahou o rozlišení třeba *stagewise* od *stepwise* nebo definováním rozdílu mezi *extrémními, odlehlými* a *vybočujícími pozorováními* atd.

Literatura.

[1] Anděl, J.: *K problémům české statistické terminologie.* Informační bulletin České statistické společnosti, č. 2, 1995.
[2] Machek, J.: *Terminologické úvahy.* Informační bulletin České statistické společnosti, č. 2, 1995.
[3] Šidák, Z.: *Znovu k české terminologii: pohled z jiné strany.* Informační bulletin České statistické společnosti, č. 3, 1995.
[4] Žváček, J.:*Kšaft umírající matky – české statistické terminologie.* Informační bulletin České statistické společnosti, č. 4, 1994.

P.S. Poprosil jsem bývalou odbornou redaktorku SNTL paní PhDr. Olgu Mavridisovou, aby laskavě přečetla všechny čtyři výše uvedené příspěvky a vyjádřila se k nim. Prakticky ihned zaznamenala své souhlasné i nesouhlasné poznámky přímo do textu i mimo něj. Nakonec jsem jí ukázal i svůj článek, jakož i (zatím neotištěnou) část odpovědi Jiřího Žváčka a znovu ji požádal o kritické vyjádření. Až po přečtení všeho jsem si vyžádal její souhlas ke zveřejnění všech poznámek ve formě přílohy k mému textu. Namítala, že její poznámky snad ani nejsou publikovatelné, ale ujistil jsem ji, že jistě budou správně pochopeny. Zde jsou v podobě určené k mému zamyšlení:

Část 1 (k článkům [1] až [4]) (kromě asi 20 poznámek přímo do textu kopií článků)

– Autor [4] patrně dráždí už způsobem svého psaní – mě pak podráždil přítomnými chybami, takže u „oxydů“ už jsem to nevydržela, vrátila se zpět a vyznačila to – sorry. Jinak totiž asi nejvíce souhlasím právě s ním,

– nezdá se mi vhodné argumentovat tím, že ten který termín užíval (nebo učil) „pan XY“! Je to tam strašně často,

– multivariate analysis – autor [1] argumentuje proti *více* s tím, že by se muselo dodat „než co“ (o kousek dál už ovšem chybně píše, že prý „o kolik“) – vypadá to spíše jako argument **pro** překlad vícerozměrná (a mimochodem řekla bych, že je to skutečně stejný případ jako „vícefázový“ – a je jedno, že to je v technickém oboru),

– testy – zajímavý problém, zde skutečně pouze lingvistický. Jde o přívlastek a čeština má obvykle přívlastek shodný před jménem – neshodný pak za jménem. Ten shodný je jaksi „těsnější“ – to mluví pro zařazení atributu **před** slovo test. Důvody skladebné pak pro spojovník – jinak by vznikaly

falešné větné dvojice, což nehrozí u testů s přídavným jménem přivlastňovacím, kde je vše zřejmé z koncovky (Studentův test), takže se takto obojí skloňuje; proměnná žádnou koncovku vyjádřit neumí,

– dvoučlověčí názvy – už jsem se vyjádřila dávno, že je pociťuji jako jeden celek, a jako takový by měly mít jen jednu koncovku (i když: zkus si představit, že na prvním místě by byla nějaká česká ženská, tedy – ová; co s tím?),

– ježiši, co se komu nelíbí na *nerovnici*?,

– zajímavé nápady, nadhozené bez řešení, o kterých by se dalo hezky diskutovat, jsou v příspěvku [3] bod 5,

– a okouzlila mne poznámka tamtéž v bodě 9 – nikdy mě nenapadlo o tom uvažovat, nikdo to určitě nebude měnit (taky se zde nenabízí žádná alternativa), ale *směrodatná odchylka* je asi vážně blbost.

Chtěl jsi po mně pět vět. Raději jsem z toho věty nedělala. Nerada totiž počítám. Snad to bude takto stačit. Jinak se měj krásně. (je jich právě pět)

Část 2 (k Tvému článku) (po asi 15 připomínkách či opravách přímo do textu)

Nemám, co bych dodala. Nic zde obsaženého (alespoň podle mne) neuráží. Pod cíl dosažení **rozumné a respektovatelné dohody** se lze jistě jen podepsat.

Část 3 (k odpovědi J. Žváčka)

Moc se mi to líbí. Taky vidím, že pokud se mi zdálo, že autor popuzuje už způsobem svého psaní – neboť reakce byly dosti emocionální – bylo to jeho cílem. Tak z tohoto hlediska by mohl být spokojen. Teď ještě, aby se dosáhlo oné rozumné dohody. Mimochodem, někdy opravdu jedna (či jeden) může být mnoho, ale to je o něčem jiném a do matematiky (potažmo statistiky) bych to netahala. Ať žije KOČKA! Olga.

O Ostravském dnu České statistické společnosti.

Josef Tvrdík

Snad vzniká dobrá a užitečná tradice. Zhruba před rokem byl v Olomouci na Universitě Palackého uspořádán „Moravský den České statistické společnosti“. Dostalo se mu příznivého hodnocení jak od účastníků, tak i v Bulletinu. V pořadí druhá taková akce, tentokrát pod názvem „Ostravský den České statistické společnosti“, byla uspořádána ve spolupráci s Přírodovědeckou fakultou OU 7. února 1996. Jako jeden ze čtyř účastníků obou statistických dnů si dovolím pár poznámek.

Ač hovořím o vzniku tradice, už u druhého statistického dne se nepodařilo zachovat název. Ostrava totiž leží na Moravě jen zčásti. Kromě toho název není důležitý a pořádání podobných dnů se nemusí omezovat jen na Moravu a nejbližší okolí. Za důležitou považuji možnost setkání statistiků a zájemců o aplikace statistiky. Těchto možností je, zdá se, hlavně pro nás „venkovské“ stále málo, a ostatně to asi není jen mimopražský problém (viz článek Petra Volfa v Bulletinu asi před rokem). Proto si myslím, že by bylo užitečné, aby tradice statistických dnů pokračovala. Forma orientovaná spíše osvětově a společensky než vědecky je vhodná a může přilákat zájemce o aplikace, jejichž vzdělávání bychom neměli opomínat. Drobná proměnlivost názvu se může stát součástí tradice statistických dnů.

Pro organizátory bylo potěšující, že se setkali s podporou u vedení fakulty, které umožnilo pozvat k přednášce zahraničního účastníka, prof. Annu Bartkowiak z Wroclawi. Děkan Přírodovědecké fakulty doc. Burian statistický den také úvodním slovem zahájil. Potěšil i zájem o aktivní účast u řady význačných odborníků. Přes absenci několika ohlášených autorů, na poslední chvíli způsobenou jejich nemocí, mohlo tři desítky účastníků vyslechnout sedm zajímavých vystoupení o statistickém software, o aplikacích statistiky a o výuce statistiky na některých fakultách ostravských vysokých škol.

Odborný program statistického dne odstartovala Hana Řezanková pohledem na současný software pro statistiku. Anna Bartkowiak z Institutu informatiky Wroclawské university zaujala přednáškou „Lisp-Stat and its

applications“. Karel Kupka z pardubického Trilobyte přednášel o některých robustních metodách v jazyku (nebo paketu?) S-PLUS. Jiří Militký vyvolal očekávaný ohlas příspěvkem „Transformace dat v regresní analýze“. Petr Veselý z Masarykovy university přednášel o jádrových odhadech a jejich využití ve vyhlazování dat. Závěrečná sdělení v souladu se zvyklostí zavedenou v Olomouci se týkala výuky statistiky v hostitelském městě. Josef Tošenovský informoval o výuce odborníků pro statistickou kontrolu kvality na Strojní fakultě VŠB-TU a Ivan Křivý o výuce statistiky na Přírodovědecké fakultě OU. Devět vytrvalců pak dokončilo program v přilehlé hospodě.

Skladba účastníků odrážela oblasti aplikací statistiky. Kromě učitelů z vysokých škol od Prahy přes Brno a Olomouc po Ostravu byly mezi účastníky i studenti, zastoupení měl průmysl a zdravotnictví. Je trochu škoda, že program tohoto statistického dne zůstal nepovšimnut většinou pracovníků Ostravské university, kteří ve své práci statistiku aplikují. Bohužel na statistiku si mnohý vzpomene až v okamžiku, kdy výsledky statistického zpracování dat je nutno zařadit do textu článku, zprávy či doktorandské práce.

Děkuji všem, kteří přispěli k úspěšnému průběhu statistického dne, zejména přednášejícím, kteří obětovali svůj čas i prostředky, spoluorganizátorům z katedry matematiky a z katedry informatiky a v neposlední řadě i firmě AVROS, jejíž sponsorský dar byl pro atmosféru celého setkání významný.

Zpráva o hospodaření.

Jako každoročně, tak i letos v tomto období podáváme zprávu o hospodaření České statistické společnosti v uplynulém roce. V roce 1995 činily příjmy 16 422,50 Kč. Největší část tvořily příjmy z členských příspěvků 13 803 Kč. Dále byly naším příjmem výnosy z certifikátů ve výši 1 689 Kč, úroky z běžného účtu 80 Kč a vložné na seminář ve výši 850 Kč. Strana příjmů po započtení převodu z roku 1994 ke konci roku 1995 činila 47 055 Kč.

Celkové výdaje v roce 1995 byly 3 928,50 Kč. Podílely se na nich poplatky za vedení běžného účtu ve výši 585 Kč (z toho 501 Kč odměny spořitelně a 85 Kč poplatky za pohyb na účtě), dále výdaje na pohoštění na valné hromadě a semináři (káva, cukr, kelímky) za 514 Kč, nákup obálek 539 Kč,

poštovné 274 Kč, 11,50 Kč pokladní blok, 13 Kč soda a 2 000 Kč odměny za redakční práce na Bulletinu a vedení hospodaření společnosti.

Do roku 1996 bylo převedeno 42 126 Kč, z toho 23 810 Kč na vkladu certifikátu, 18 857 Kč na běžném účtu a 459 Kč v hotovosti.

I když bylo loni vybráno na členských příspěvcích více než v předcházejícím roce v důsledku toho, že řada členů platila zpětně dlužné příspěvky za minulá léta, zůstává stále dost členů, kteří dosud nezaplatili příspěvky za rok 1995. Prosíme proto opětovně o zaplacení dlužných částek společně se zaplacením členského příspěvku za rok 1996, který zůstává v nezměněné výši 60,- Kč. Placení je možné buď přiloženou složenkou nebo osobně doc. Blatné na katedře statistiky a pravděpodobnosti VŠE Praha. Vhodná je i forma placení jednou složenkou za všechny členy Vaší organizace najednou. V tom případě ale prosím o zaslání jmenného seznamu platících, neboť z výpisu na účtu to není možné vyčíst. Připomínáme a opakujeme znovu následující pokyny:

- číslo účtu u České spořitelny je 8024551/0800 (v případě, když nemáte složenku Společnosti);
- jako variabilní symbol uvádějte rodné číslo.

Dáša Blatná

Josef Kozák, Aktuální poznámka k mírám inflace 1

Marta Horáková, Struktura závislosti. 6

Hana Řezanková, Nově, lépe, radostněji? 7

Zbyněk Šidák, Postscriptum o české statistické terminologii. 11

Jiří Žváček, Ještě k analýze více proměnných. 13

Petr Hebák, Ať žije KOČKA. 14

Josef Tvrdík, O Ostravském dnu České statistické společnosti. 22

Dáša Blatná, Zpráva o hospodaření. 23

Informační Bulletin České statistické společnosti vychází čtyřikrát do roka v českém vydání a jednou v roce v anglické verzi. Předseda společnosti: Ing. Zdeněk Roth, CSc, SZÚ Praha, MSP, Šrobárova 48, 100 42 Praha 10, E-mail: roth@szu.cz. ISSN 1210-8022

Redakce: Dr. Gejza Dohnal, Jeronýmova 7, 130 00 Praha 3, E-mail: dohnal@fsik.cvut.cz.