

Informační Bulletin



České statistické společnosti

č. 1. duben 2001, ročník 12

Je úterý, 10. dubna roku 2001. Sedíme v pracovně docenta Antocha a připravujeme Informační Bulletin. Všechno krásně vychází na stránky ... až na první stránku, která je zatím prázdná. Není ani celá stránka, je tu hlavička, logo. Je pozdě večer a Jaromír to vyřešil způsobem jemu vlastním: „Na, vem si to a na tu první stránku něco napiš! Však víš, bude pátek třináctého, každému se přece něco stane, minimálně do něčeho šlápne na ulici ...“. To je přece jasné – pátek, nešťastný to den – o tom se dá napsat spousta věcí. Jenže když tak vzpomínám, na nic konkrétního si nevzpomínám. Dobře, počkám si tedy, bude to už popozítří.

Je pátek, 13. dubna, 2001. Je to Velký pátek. První pátek třináctého v tomhle tisíciletí, které podle řady proroků už vlastně ani nemělo být! Ráno vstávám jako obvykle. S nadějí v duši a s pocitem zodpovědnosti se rozhlížím po něčem neobvyklém. Nic. Proti svým zvyklostem jdu posnídat (třeba se oparím čajem!). Vše probíhá zatím O.K. Cesta do práce. Ulice je nějak podezřele čistá, takže zase nic. Na hygienické stanici, kam jsem dorazil nejprve, mi paní doktorka s úsměvem nabídla kávu. S tou asi také nic nebude. Už se těším do školy, jaký průšvih mne tam čeká. Sotva dosednu za stůl, je to tu! Telefon a v něm: „Tady je oddělení pro vědu a výzkum. Pane docente, ten váš návrh na grant ...“ No to jsem si myslel! Při mé smůle je to zase ztracených několik dní sepisování návrhu a oni ho zamítnou ještě než jsem ho podal! Jenže telefonát pokračuje: „.... je v pořádku, jen je potřeba doplnit na konec dvě věty. Stačí to poslat mejlem.“ Tedy zase nic. Jsou tak hodní, že mne včas upozornili. Nemám já smůlu? Možná odpoledne, až pojedeme na chalupu. Jsou Velikonoce, silnice bude plná, tam se určitě něco semele! Ale ani to nevyšlo.

Je sobota, 14. dubna 2001. Jsem živ a zdrav, venku padá sníh a mně se nic zvláštního nestalo. Jaromíre promiň, jsem prostě smolař! *Gejza Dohnal*

* To samo o sobě je už podezřelé!

Pátek, a k tomu třináctého

Luboš Pick¹

Z podnětu paní redaktorky Dany Mertové z redakce Zemských novin a deníku Slovo jsem se zabýval otázkou, kolik „pátků třináctého“ nastane ve 21. století. Paní redaktorka se pak omezila jen na telefonický rozhovor a informace, které jsem ji sdělil, „využila“ ve svém článku *Za pověřčivostí na pátek třináctého je prý nejistota* otištěném ve Slově ve čtvrtek 12. 11. V úvodu mimo jiné píše: *Během příštího století jich² na lidi ale čeká přesně stodvaasedmdesát, spočítal matematik Luboš Pick z matematického ústavu. Podle něj s číslem třináct nejdou dělat žádné velké čáry: Je totiž liché a dělitelné jen jedničkou a samo sebou.*

Jsem si vědom, že následující výpočty jsou na úrovni řekněme sedmé třídy základní školy, ale když už jsem je (v bláhové představě, že budou paní redaktorku zajímat) sepsal, předkládám je alespoň čtenářům Drbný.

Nadále budu každý měsíc, ve kterém datum třináctého připadá na pátek, nazývat „šťastným měsícem“. (Tento pohled na věc jest subjektivní a nikomu jej nevnucuji. K názvu jsem dospěl díky pozorování zajímavé souhry náhod, že totiž tři pro mne zatím nejdůležitější měsíce, a to říjen 1961, kdy jsem se narodil, listopad 1987, kdy jsem se ženil, a říjen 1995, kdy se narodila má dcera, patří do této kategorie.)

Jako první krok si uvědomíme, že šťastný měsíc je ten a jenom ten, který začíná nedělí.

Nyní zavedeme čtrnáct různých typů letopočtu, a to podle následujícího schématu: jestliže je rok odpovídající danému letopočtu nepřestupný, označíme jej písmenem N , ke kterému přidáme jako index jedno z čísel 1 až 7 podle toho, na který den připadá v tomto roce 1. leden. Analogicky značíme roky přestupné s tím, že místo N píšeme P . Například rok 1998 je typu N_4 , rok 1999 je typu N_5 , rok 2000 je typu P_6 , a konečně rok 2001 je typu N_1 .

Spočítáme, kolik šťastných měsíců obsahuje rok daného typu. Prostým výpočtem dospějeme k následující tabulce:

¹ Redakce Bulletinu ČSS děkuje autorovi a MVS JČMF za laskavé svolení s přetištěním tohoto příspěvku, který byl publikován ve Zpravodaji MVS JČMF v roce 2000.

² tj. pátků s datem třináctého — pozn. redakce

Typ roku	N_1	N_2	N_3	N_4	N_5	N_6	N_7
Počet šťastných měsíců	2	2	1	3	1	1	2

Typ roku	P_1	P_2	P_3	P_4	P_5	P_6	P_7
Počet šťastných měsíců	2	1	2	2	1	1	3

Každý rok má tedy aspoň jeden a nejvýše tři šťastné měsíce. Rok 1998 má šťastné měsíce tři (jmenovitě únor, březen a listopad).

Bude užitečné povšimnout si následujícího faktu: počet dnů obsažených v libovolném 400letém cyklu, který končí rokem dělitelným číslem 400 (například 1601–2000 nebo 2001–2400), je dělitelný sedmi. To plyne z následujícího jednoduchého výpočtu: daný cyklus obsahuje 400 roků, z toho je 303 nepřestupných a 97 přestupných (nezapomeňme, že přestupný rok nastává právě když odpovídající letopočet je dělitelný čtyřmi, ne však pokud je dělitelný stem a zároveň není dělitelný čtyřmi sty). Počet dnů v nepřestupném roce dává při dělení sedmi zbytek 1, v přestupném zbytek 2. Stačí si tedy uvědomit, že číslo $303 + (2 \times 97) = 497$ je dělitelné sedmi. Jinými slovy, každý takový 400letý cyklus začíná stejným dnem, a to pondělkem (jak už víme, rok 2001 je typu N_1).

Nyní si povšimneme, jakým způsobem se střídají typy letopočtů. Začneme-li rokem typu N_1 , následuje rok typu N_2 , pak N_3 , pak P_4 , po něm N_6 , a tak dále. Z následující tabulky plyne, že se cyklus uzavře po 28 letech, pakliže cestou nepotkáme výjimečný rok (tj. přelom století).

N_1	N_2	N_3	P_4
N_6	N_7	N_1	P_2
N_4	N_5	N_6	P_7
N_2	N_3	N_4	P_5
N_7	N_1	N_2	P_3
N_5	N_6	N_7	P_1
N_3	N_4	N_5	P_6

Tento 28letý cyklus obsahuje každý letopočet typu $N_1, \dots, N_7$ třikrát, a každý letopočet typu $P_1, \dots, P_7$ jedenkrát. Každý takový cyklus tudíž obsahuje (viz tabulka) právě 48 šťastných měsíců.

Cyklus 28 let se beze změny zopakuje třikrát (v případě 21. století jde o období 2001–2084). Do konce století nám pak zbývá 16 let, které musíme posoudit samostatně. K tomu nám poslouží první čtyři řádky tabulky s tím rozdílem, že závěrečný letopočet (například 2100) nebude typu P_5 jako v tabulce, ale bude typu N_5 , neboť, dle zmíněné výjimky, tento rok není přestupný. Odtud již s pomocí tabulky vidíme, že období 2085–2100 obsahuje 28 šťastných měsíců. Celkem tedy bude 21. století obsahovat přesně $(3 \times 48) + 28 = 172$ šťastných měsíců, to jest, můžeme se těšit na 172 „pátek třináctého“.

Analogicky s pomocí uvedených tabulek zjistíme, že druhé století 400letého cyklu (např. 2101–2200) obsahuje rovněž 172 šťastných měsíců, třetí

století (2201–2300) pouze 171, ale zato čtvrté století (2301–2400) dokonce 173.

Prostým sčítáním dojdeme k informaci, že každý 400letý cyklus, který je zakončen letopočtem dělitelným 400, obsahuje 688 šťastných měsíců, neboli 688 pátků třináctého.

Pro zajímavost vypočítáme pravděpodobnost výskytu šťastného měsíce. Protože situace se po 400 letech opakuje, stačí vypočítat pravděpodobnost výskytu pátku třináctého v jednom 400letém cyklu. Tato pravděpodobnost je dána zlomkem

$$\frac{688}{4800} = \frac{43}{300}.$$

Není bez zajímavosti, že pravděpodobnost výskytu šťastného měsíce není rovna jedné sedmině. Z předchozích úvah plyne, že je rovna číslu

$$\frac{1}{7} \times \left(1 + \frac{1}{300}\right)$$

a je tedy nepatrně větší než jedna sedmina. Z toho lze učinit jediný možný závěr: *Žijeme prostě ve šťastné době.*

Několik rad do života

Adresát: Kolega Socrates a další

Autor: Martin Betinec, betinec@karlin.mff.cuni.cz

Jaroslav Šterba, sterba@atrey.karlin.mff.cuni.cz

(<http://atrey.karlin.mff.cuni.cz/sterba/barcelona>)

Téma: Text k prodiskutování.

Místo Barcelona, Catalunya, Espanya

Téma: Čas: 1999-2000

Místopis

Barcelona je přístavní město nalézající se na severovýchodě Španělského království ve střední části Katalánska. Přestože historie města sahá až do starověku, převážná většina městských částí vznikla na začátku tohoto století, což lze poznat podle pravoúhlé sítě ulic (láká to ke hraní piškvorek na mapě). Pod celým městem je poměrně hustá síť metra (5 tras) doplněná povrchovou dopravou (autobusy a tramvají). Město je proslavené nejen stavbami Antonia Gaudího. Kvůli olympiádě v roce 1992 zrekonstruovali co se dalo, navíc postavili pátou trasu metra (L2 :-)) a vybudovali moderní sportoviště,

některá z nich jsou přístupná veřejnosti zdarma. Pišqorky se nestaly katalánským národním sportem zřejmě jenom díky tomu, že ideální struktura mapy je narušena několika diagonálními třídami, které jsou ovšem výborným pomocníkem v orientaci. Od východu k západu protíná město nejvýznačnější z nich - Avinguda Diagonal. Po obou stranách jejího západního konce byla vybudována Zona Universitaria - komplex budov Katalánské polytechniky (Universitat Politècnica de Catalunya =: UPC), rozdělený Diagonálou na Campus Nord a Campus Sud. Člověk si musí pamatovat, že jel studovat na jih, a proto je třeba k fakultě Matematiky a Statistiky zamířit [předpokládáme, že jste právě vystoupili z metra na Palau Reial (zelená trasa - L3)] pod azimutem 180 stupňů. Budova Facultat de Matemàtiques i Estadística je v ulici Pau Gargalo 5.

Než začne škola

Střecha nad hlavou

Pokud jedete jako studenti Erasmu, neměli byste mít problém se sháněním ubytování. Teoreticky by mělo stačit napsat mail na Habitatge Jove (info@habitatgejove.com), a oni už se o Vás postarají. Jedná se o organizaci, která má s UPC smlouvu o poskytování koleje studentům Erasmu. Cena koleje se pohybuje prý kolem 27.000 peset na měsíc. Kolej známe jenom jednu a ta je docela daleko od univerzity, až v Badaloně. Ale metro tam jede a člověk má o starost na začátku méně. Jenže ... může se Vám stát, že z Habitatge Jove dostanete odpověď, že je jim to moc líto, ale že všechna místa určená zahraničním studentům jsou rozebrána a že se máte podívat na jejich webovou stránku (www.habitatgejove.com), zejména do míst, kde je inzerce o volných místech. Ještě předtím, než začnete pátrat, zkuste napsat (budete-li na UPC, jinak se to pravděpodobně jmenuje jinak) na instituci zvanou Oficina mobilitat internacional. Tam by Vám mohli být schopni pomoci a jsou případně schopni poslat papír, který ambasáda uznává jako doklad o zajištěném bydlení. No a pro případ, že byste chtěli a nebo byli nuceni hledat sami, tak žádný strach. My oba jsme si našli bydlení přes internetovou stránku Habitatge Jove a byli jsme spokojeni. Ceny podnájmů jsou různé, rozumná cena je mezi 25 až 35 tisíci peset. Jen je třeba se připravit na to, že převážná většina bytů nemá topení ani dvojité okna, takže tam může být v zimě trošku chladno. Exterior znamená okno do ulice, interior značí pohled do dvora.

Officiality

Už je to tak na světě zařízeno, že na některé věci je potřeba nějaké potvrzení. A tím nemyslíme ani tu trošku zařizování u nás na Univerzitě či na fakultě.

Ani velvyslanectví není pro studenty nedobytnou byrokratickou baštou (jen na okraj: za vízum se neplatí). Ale začněme pěkně od začátku.

V Čechách

Jakmile bude jasné, že na Vás vzdálené země čekají s otevřenou náručí, zajděte za paní Klepetářovou a také je nutno na fakultě podepsat finanční dohodu. Paní Klepetářová Vás vybaví dopisem pro ambasádu. Pro získání víza je třeba též výpis rejstříku trestů s ověřeným překladem do Španělštiny a též úředně přeložené potvrzení od doktora. (Ale určitě se informujte přímo na ambasádě, každý rok se to mění). Vyřízení víza netrvá dlouho, v našem případě se jednalo o pár hodin. Pak už jen zaplatit pojištění a koupit lístek. Španělsko čeká.

En Barcelona

Dokud nebudete mít střechu nad hlavou, o nic dalšího se nestarejte. Ale hned poté zajděte na cizineckou policii (A teď přichází to byrokratično; i když podle posledních informací už se situace zlepšuje a do některých front je možno se objednávat přes internet). Je potřeba si vyřídit kartu studenta (Tarjeta de Estudiante). Ještě za Jardy bylo potřeba čekat ve frontě Marokánců žádajících o azyl, fronta byla dlouhá a kdo chtěl být v poledne úspěšný, tak se předtím v noci moc nevyspal. A po dni dni čekání získal informaci či číslo do pořadníku pro podání žádosti. Určitě se zastavte na Campus Nord, v Oficina mobilitat internacional. Dají Vám hromadu prospektů a plán města. A poradí. Všichni studenti Erasmus se chodí ptát asi tam a tak je to přímo studnice informací.

Nákupy

Potraviny lze koupit i za ceny velmi podobné českým. Záleží však na tom kde. Asi nejlevnější jsou Penny Markety (před nedávnem byly přejmenovány na Plus). Protože však nejsou na každém rohu, občas člověk zajde i jinde. Poslouží například Dia - jsou skoro všude a ceny nejsou o tolik vyšší. Jsou tu také další, jako například Spar, E-Consum a jiné, a lze koupit to co u nás a mnohdy i více. A pokud člověk něco hledá a nemůže to najít, je tu El Corte Inglés - obrovský obchodák, kde mají všechno, včetně českého piva. Samozřejmě za patřičné ceny. Zvláštní kapitolu tvoří zelenina a ovoce společně s produkty moře. Toho je tu v jakoukoliv roční dobu víc než dost. Nejlepší je zajít na trh - Mercado/Mercat a vybrat si. Užitečné může být projítí trhu ještě před koupí, neboť ceny se liší, někdy až dvojnásobně.

Jazyky

V Barceloně lze najít značné množství kurzů španělštiny, jen si vybrat. My jsme zpočátku chodili do jednoho intenzivního. „Naše“ jazykovka se jmenovala Institutos Mangold a byla na křižovatce Rambla Catalunya a Gran Via de les Corts Catalanes. A stála za to. Pro jazykovku se nám osvědčila jedna metoda - snažit se postoupit do vyšší úrovně, jakmile se nějaká naskytla. Následujících pár dnů bylo sice těžších, ale na druhou stranu se mluvilo více španělsky. Určitě každého zarazí to množství cizinců z Itálie, Francie či ze zemí mluvících Portugalsky - a všichni se chtějí naučit španělsky. A to mají velkou výhodu: v jejich rodném jazyce se to řekne velmi často skoro stejně a gramatika je budována analogicky. Ale my Češi jsme se nedáme - stačí pár jazykolamů a Brazilci ztratí sebevědomí ... A Catalán neboli Katalánština? Pro porozumění psanému textu a do jisté míry též mluvenému (za předpokladu, že jste se již vypořádali či právě válčíte se španělštinou) jsou dostačující dvě devadesátiminutovky týdně vyučované v severní části univerzitní zóny. Je dobré zkusit alespoň prvních pár lekcí (i za předpokladu, že víte, že tam nebudete moci chodit pravidelně), velmi Vám to ulehčí porozumění psanému textu (zejména učebnicím). Ale nemusíte se nadmíru obávat - katalánština je takovou směsí španělštiny a francouzštiny. Alespoň pro nás. Los catalanes no lo creen.

Pohyb v Barceloně

Městská hromadná doprava má poměrně dobře propracovanou strukturu. Autobusy doplňují síť metra, jízdní řády říkají pouze frekvenci autobusů. Obyčejný lístek stojí 160 peset, deset jízd 825 a měsíční „T-mes“ 5500 peset. Studenti nemají slevy. T-mes se vyplatí hlavně dokud člověk nezná město. Za měsíc za dva jsme poznali, že i do školy se dá dojít pěšky a že celá Barcelona lze přejít tak za 100 minut. Ale ... pokud bych se ještě někdy dostal na půl roku do Barcelony, určitě si koupím kolo (z druhé ruky by nestálo určitě víc než průkazka na MHD na tři měsíce).

Škola

Veškeré věci týkající se studia Vám pomůže vyřešit sekretářka katedry, je velice příjemná a jmenuje se Celia Martín. A nejen to - Martin, když byl nemocný, se dobelhal na sekretariát a tam už si Celia věděla rady. S největší ptí takovou službu asi nepoužijete, ale formality ohledně zápisu budete určitě zařizovat tam. A pravděpodobně s Lupe (prof. Guadalupe Gómez Melis), která se stará o většinu zahraničních studentů. Protože všichni spolužáci jsou doktorandi, nedivte se, že jste nejmladší. Naši soukmenovci byli například

i starší (leckterí měli děti v našem věku) a velice družní jihoameričani (např. Guillermo, un profesor de la salsa y de la tequila). Velice záhy zjistíte, že na katedře studuje jen minimum Katalánců - drtivou většinu tvoří latinoameričané, pro něž je Španělsko branou ke vzdělání. Vyskytují se tu i exempláře ostatních národností, a zřejmě pod vlivem latinoameričanů jsou všichni velice družní; ve škole to platí dvojnásob, zvláště poté, co jsme je nadchli pro socialistickou strategii uvědomělého plnění vytčených úkolů (domácích) pod heslem ¡LOS ALUMNOS DE TODOS LOS PAISES CONJUNTAOS!

Vyučované předměty jsou hodně prakticky zaměřeny, teorie je méně. Většina příkladů má praktickou motivaci, UPC je úzce spjata s průmyslem a tak ani o zajímavá data není nouze. Například na předmětu Operační výzkum jsme měli možnost shlédnout počítačovou simulaci provozu v Barceloně. Zápočtovým úkolem ze stejného předmětu bylo pomocí simulací ověřit model, technologický postup. Z toho důvodu stojí za to určitě znát Matlab, případně C++. Pro statistiku též používají SAS, S-PLUS, znají též SPSS, hojně je využíván Minitab. Kolik předmětů zapisovat je těžká otázka. Jarda, který tam byl v zimě, zapsal čtyři a měl dost času na to chodit i na katalánštinu, Martin se v létě při stejném počtu přednášek a seminářů skoro nezastavil a stále vypracovával zadané domácí úkoly. Ale určitě stojí za to chodit na Navrhování experimentů, Analýzu přežití či Časové řady. Zejména Navrhování experimentů je přednáška, která je současně cvičením i praxí. Veškerý teoretický výklad (a to platí i o ostatních předmětech) je omezen takřka na minimum.

Postgraduální studium se v první fázi, do níž jsme byli jakožto krátkodobí studenti zařazeni, velmi podobá našemu studiu magisterskému - člověk prostě chodí na přednášky a rozšiřuje si obzory. Teprve v dalších letech si vybere téma disertační práce (a vedoucího :-). Ani tamní doktorandské přednášky svojí teoretickou náročností nijak nepřesahují úroveň MFF-áckých magisterských kursů. To, že je UPC, jak již název napovídá, technika, má svou velkou výhodu v důrazu na propojení probírané látky s praxí.

Díky tomu, že na UPC, podobně jako na všech španělských univerzitách, jsou v postgraduálních programech hojně zastoupeni Latinoameričané, se přednášelo ve španělštině. Studijní materiály ovšem zpravidla odpovídají tomu, že dorozumivacím jazykem je katalánština. Popravdě řečeno - není to zase až takový problém, aspoň co se psané formy týče (zatímco u mluveného slova je to nezdárka problém dost zásadní), po absolvování prvních dvou lekcí katalánštiny, kde člověku objasní, kde hledat analogie a kde se naopak skrývají záludnosti, to člověk znající alespoň trochu španělsky zvládne poměrně rychle. Navíc, a to je asi nejdůležitější, se může takřka kohokoli ze spolustudentů, kteří, jak to již z jejich temperament napovídá, jsou velice přátelští

a druhí, zeptat. Sympatické je, že přednášející jsou zvyklí nabízet přednášky předem k ofocení, takže student má poloviční práci s obkreslováním indexů apod., a o to více se může soustředit na problematiku.

Martin absolvoval následující přednášky (kromě nepovinného úvodu do jazyka C, který pro nás neoficiálně zařídila vedoucí katedry a na nějž jsme chodili oba): analýza přežití, metody predikce, simulační prostředí, mnohorozměrná statistika a pokročilé metody navrhování experimentů. Tyto překlady původních názvů jednotlivých kurzů mohou být v českém kontextu poněkud zavádějící, proto nyní v krátkosti poodhalme, co se za nimi skrývalo.

Analýza přežití v tomto smyslu ještě žádné zvláštní překvapení nechystá. Dlužno dodat, že se jedná o základní přednášku věnovanou tomuto tématu, čemuž odpovídá i rozsah probírané látky: definice funkce přežití, její odhady (Kaplan-Meier), porovnávání více výběrů, Coxův model proporčního rizika. Přednášela vedoucí katedry Prof. Guadalupe Gómez ve španělštině, studijní materiály byly k dispozici v obou jazycích. Hodnocení se sestávalo jednak ze závěrečné písemné zkoušky a také z domácích úloh - většinou aplikace probírané látky na konkrétní data. Kdo umí dobře S-PLUS má poloviční práci. Pro nás ostatní to byly zdaleka časově nejnáročnější domácí přípravy.

Metody predikce bychom u nás nazvali asi jednodušeji: časové řady, resp. Box-Jenkinsonova metodologie, tzn. od MA(q), až po SARIMu včetně predikcí. To se dle mého názoru vyplatí slyšet vícekrát, a od techniků sbírajících zajímavá data obzvláště. Opět i zde bylo součástí přednášky domácí zpracování datových souborů s využitím místních programů PEST, TRAMO a SEATS. Přednášel Manuel Martí Recober ve španělštině, materiály v katalánštině. Mimochodem, zapsat si něco, co už teoreticky znáte je kromě jiného výhodný způsob jak se naučit jazyk.

Simulační prostředí byl v podstatě seminář, na který jsme chodili dva statistici a dva informatici a střídali jsme se v referátech (v angličtině :-) týkajících se simulací různých průmyslových a přírodních procesů. Přednášel Josep Casanova střídavě v obou jazycích :-) Materiály převážně v angličtině.

Mnohorozměrnou analýzu přednášela nejprve v první polovině semestru Mónica Bécue a přednesla v podstatě bez jakýchkoli pokusů o důkazy seznam metod (kanonické korelace, faktorovou analýzu, shlukovou analýzu,...). Tradičně i tuto přednesenou kuchařku bylo třeba vyzkoušet na datech. Pro mě byla tato přednáška nejnesrozumitelnější, především z jazykového hlediska. Ve druhé půli semestru převzal štafetu Tomás Aluja a přednášky dostaly nečekaný náboj, když nám určil do dvojic referáty týkající se některé metody zpracování dat (teorie + program), které po jejich přednesení velmi dobře komentoval. Přednášelo se španělsky, s čímž měla první přednášející

původem Francouzka, (a nebo spíše já :-)) poněkud problémy. Materiály byly v angličtině, katalánštině a francouzštině.

V názvu poslední přednášky je matoucí označení „pokročilé“. Ku příkladu přednáška prof. Huškové „Navrhování experimentů“ (vypisovaná na MFF UK) by v tomto duchu mohla nést název „Hodně pokročilé techniky ...“, v případě barcelonském bych raději volil název „Co všechno jsme schopni zjistit pomocí pouhého sčítání“. Zní to možná až příliš ironicky na to, že se mi tato přednáška líbila asi nejvíce. Jednak přednášející napsali dobrou učebnici ve španělštině, jednak to bylo matematicky velmi jednoduché :-), ale hlavně zde bylo možno vyzkoušet vše probírané na zajímavých datech z průmyslové výroby. Pravda, člověk musel dost často hledat ve slovníku (kterého matematika by předtím v Praze napadlo učit se slovíčka typu: *psí konzerva* nebo *pevnost pružiny ve zkrutu*), ovšem klady jednoznačně převážily. Ke zpracování úloh jsme používali program MINITAB. V přednášení se střídali Albert Prat, Pere Grima a Josep Ginebra.

Jardovy přednášky, snad proto, že proběhly v zimním období, byly daleko poklidnější - přece jen má zima na jižní národy zklidňující účinky. A tak, jak již bylo zmíněno, při stejném počtu přednášek měl daleko více času.

Snad nejnavštěvovanější byla přednáška prof. Jaume Barceló, zabývající se základními technikami operačního výzkumu. Dala by se též nazvat všeobecným úvodem do této problematiky; její náplní bylo seznámení jak seznámení s rozličnými výpočetními nástroji (například jazyky Lingo či Lindo), tak ukázka rozličných simulací (velmi zajímavým příkladem simulace byl počítačový model silničního provozu v Barceloně). Podmínkou pro úspěšné zakončení předmětu bylo naprogramování generátoru náhodných čísel, jeho statistické ověření a následná aplikace pro simulaci fiktivního výrobního procesu.

Zpracováním práce na počítači se zabývala též druhá přednáška, která však pro nedostatek zapsaných studentů probíhala pouze formou samostudia a konzultací. Název přednášky „Software de modelització per a programació matemàtica“ zní honosně, úkolem však bylo proniknutí do systému AMPL a následné zpracování tří úloh. (Ale zadání úloh bylo „en català“, takže před vlastním programováním následovalo drobné lingvistické cvičení.)

Analýza kontingenčních tabulek víceméně odpovídala analogické přednášce na MFF vyučované doc. Práškovou, jediný rozdíl byl v softwaru používaném na cvičeních - byť se jednalo jen o pár praktických ukázek, měli jsme možnost získat ideu o tom, jak se používá SAS či SPSS. V přednášení se střídali Erik Cobo a Lúdia Montero.

Na semináři zabývajícím se metodologickými aspekty klinických experimentů vedeném Erikem Cobo jsem se dozvěděl něco o problematice do té doby pro mě neznámé, v závěru následovaly referáty. Nabídku na možnost

odpřednesení referátu v angličtině jsem odmítl, češtinu odmítli oni ... a tak přibyla jedna zajímavá zkušenost navíc.

A pokročilými partii časových řad nás provedl Manuel Martí Recober. Šlo o seminář, který se vyznačoval velmi „rodinným“ přístupem - asi to bylo způsobeno i tím, že jsme byli pouze dva studenti. A tak jsme poměrně rychle zopakovali to, co jsme již znali ze svých domovských fakult (Rebekka z Dortmundu a já z Prahy) a vrhli jsme se do Kálmanova filtru a věcí příbuzných.

Zvyklosti ve škole i mimo ní

Nevykat! Všichni Španělé si tykají, jen výjimečně někdo vyká staré babičce. A další výjimka už je asi jen Juan Carlos. Nezouvat se! Oni taky chodí v botách (často jsou domácnosti bez koberců). Není zvláštností, že si chystají večeři velmi pozdě. Jsou velmi nepřesní, co se času týče. Pokud začíná přednáška oficiálně v 17:30, neznamená to nic víc, než že tam o půl šesté budete stát jako na trní a přemýšlet, zda to neodpadne. Jen klid. Žáci přijdou za deset minut a profesor chvíli po nich. A pokud Vám nějaký Španěl řekne, že se sejdete o půlnoci na začátku Rambly, pak je 0:20 docela slušný čas a dříve přijdou jen ti, co se o půl hodiny spletli. (Asi půjdete na disko, to začne tak ve 2:00 ráno).

Víkendy, sport a kina

Víkend se dá v Barceloně strávit různě. V moři se dá, alespoň podle našich středoevropských měřítek, koupat skoro po celý rok, pokud se zkaží počasí, je zde dostatek jiných možností - kina, muzea, sportovní centra. S návštěvou některých muzeí je lepší počkat na první týden v měsíci - v tomto týdnu je vždy určen den, kdy je vstup zdarma. Nejvíce nás asi nadchlo námořní muzeum a muzeum vědy. Koho zajímá historie, pak je tu muzeum historie města. A pro znalce umění a architektury má Barcelona také mnoho příjemných překvapení.

Kina jsou hojně zastoupena, je zde několik multikin. Cena vstupenek se liší podle dne v týdnu, do některých lze jít na dvě představení za cenu jednoho. Většina filmů je dabována do španělštiny, jen některá kina (např. Icaria Yelmo) hrají v původním znění se španělskými titulky. A samozřejmě jsou i biografy s představeními v katalánštině (pro fajnšmekry - krátká ukázka: „Em dic Bond. Jaume Bond“ :-). Představení často začínají i po půlnoci. Univers - je to karta, kterou si můžete koupit na UPC, umožňuje slevu při návštěvách některých kin, občas jsou dokonce v centru Univers lístky zdarma. Navíc občas pořádají různé nabídky a prezentace různých sportů - my jsme se

tak dostali na plachetnici na moře (na dvě hodiny, do metrových vln a o loď se nás staralo pět amatérů pod vedením instruktora - málem jsme se převrátili pouze jednou).

Nedaleko olympijského stadionu je několik cvičných skal a stěn pro ho-
rolezce. Kdo leze, lezečky s sebou. Část je dokonce v asi 200 metrů dlouhém
tunelu, kterým vede pouze nepoužívaná silnice. Chyty jsou i na stropě. A kdo
se chce podívat někam na pořádné skály, ať vyrazí na Montserrat. Montserrat
je pohoří se světoznámým klášterem. V klášteře je poutníků i turistů tolik,
že se člověk až děsí. Stačí ale jen vyběhnout o kus dál mezi skály a hned je to
něco jiného. S batohem na zádech se lze několik dnů toulat tímto nádherným
pohořím, jehož úpatí je asi 150 metrů nad mořem a nejvyšší skalní jehla má
více než 1200 metrů. Pokud tam pojedete, kupte si mapu - je v ní kresleno
opravdu vše. Dokonce i převisy, pod kterými se dá spát a z knížečky lze vždy
vyčíst, kolik lidí se pod nimi vyspí. A nezapomeňte si vzít dostatek vody -
montserratské studánky jsou plné jen po deštích.

Pár drobností závěrem:

Španělský jazykolam je například „Tres tristes tigres tragaban trigo en un
trigal“. Ale jazykolam bez ř, to není ono. A pokud byste se rádi dozvěděli
i něco navíc, podívejte se na webovou stránku, kterou jsme za tím účelem
připravili.

Martin & Jarda

Excel a „náhodná“ čísla

Petr Máša ¹

Abstrakt: Tento příspěvek se zabývá otázkou generování náhodných čísel
v programu Excel. V textu jsou popsány důvody, proč jsem psal o Excelu
a je rozebrán generátor náhodných čísel používaný v makrojazyce VBA za-
budovaný v MS Excelu 97 či MS Excelu 2000. Též je zmíněno, jak dopadly
další generátory v Excelu.

Úvod

Nejprve si můžeme položit otázku, proč popisovat zrovna generátor náhod-
ných čísel v Excelu, nástroji na prakticky univerzální zpracování tabulek,

¹ pmas7354@artax.karlin.mff.cuni.cz

když existují specializované statistické a matematické balíky, které by měly umět generovat náhodná čísla lépe. Důvod je jednoduchý. Excel je široce rošířený a běžně dostupný – na rozdíl od specializovaného softwaru, který je obvykle velmi drahý a nemá tak široké použití. Představme si například pojišťovnu zabývající se životním pojištěním. Tato pojišťovna zcela jistě zakoupila licenci na specializovaný software pro výpočet životního pojištění. Kromě tohoto softwaru bude mít téměř jistě k dispozici nějaký běžný tabulkový kalkulátor, dejme tomu Excel. Jelikož softwaru nelze nikdy stoprocentně důvěřovat, je občas vhodné si dané výpočty ověřit i pomocí jiného nástroje. V našem případě se nabízí Excel. Též chybějící funkce tohoto specializovaného software (např. simulace, odhadování pravděpodobnosti zruinování pojišťovny, chybějící možnost psaní skriptů) lze obejít s použitím tabulkového kalkulátoru. Při použití generátoru náhodných čísel je ovšem vhodné znát jeho vlastnosti – a to jak v dobrém, tak ve špatném smyslu.

Rád bych tímto způsobem poděkoval Doc. Dr. Jaromíru Antochovi za to, že mě seznámil s problematikou generování pseudonáhodných čísel, za pomoc při přípravě tohoto článku a za užitečné připomínky k předběžným verzím tohoto textu.

Generování náhodných čísel z rovnoměrného rozdělení $R(0,1)$ v Excelu

Generovat náhodná čísla v Excelu můžeme prakticky třemi způsoby. První způsob je použít Analysis Toolpack. Tento nástroj nám sice umožňuje generovat čísla nejen z rovnoměrného rozdělení, ale ze spojitých rozdělení nabízí pouze rozdělení rovnoměrné a normální. Druhá možnost je generovat čísla přímo funkcí v tabulce Rand() (v české lokalizaci příkazem Náhčíslo()). Tyto způsoby mohou být v situaci, kdy potřebujeme větší množství náhodných čísel, např. pro simulace, nevhodné, neboť brzy narazíme na limit počtu řádků tabulky 65535 a počet sloupců 255. Navíc při simulacích obvykle nepotřebujeme skladovat náhodná čísla jako taková, ale jen výsledek nějakého procesu (nějaké simulace jako například odhad průměrného času algoritmu, či již dříve zmiňovanou pravděpodobnost zruinování pojišťovny, apod.), který je ovlivňován právě náhodnými čísly. Na tyto simulace je v Excelu velmi silný a rychlý prostředek – makra VBA (Visual Basic). Můžeme si tak snadno (teoreticky) napsat jednoduchý FOR cyklus, který na základě úmrtnostních tabulek simuluje možný vývoj pojistného plnění ze smluv, které máme uloženy v tabulce. Tuto simulaci můžeme několikrát zopakovat a tak odhadnout zmiňovanou pravděpodobnost zruinování. Tento proces vypadá velmi pěkně, je snadný na naprogramování, vyžaduje minimální znalosti z programování,

a je poměrně rychlý (pokud se při tom nedělají velké přesuny v tabulkách). Má však jednu podstatnou nevýhodu – generátor náhodných čísel.

Jaké generátory Excel používá

První nevýhoda, které si běžný uživatel vůbec nevšimne, je odlišný generátor náhodných čísel v tabulce, v makrech VBA a v Analysis Toolpacku. V nápovědě implementace těchto generátorů popsána není, ale mohu prakticky s jistotou říci, že generátor v makrech VBA a generátor v tabulkách jsou různé. Dále generátor v makrech a generátor v Analysis Toolpacku jsou různé. Na otázku, zdali jsou generátory v tabulkách a v Analysis Toolpacku stejné či různé, mohu vyslovit jen domněnku, že jsou různé. Tato domněnka je podložena strukturálně jinými výsledky testů v baterii DIEHARD.

Pokud si prakticky vyzkoušíme metody generování náhodných čísel, zjistíme, že při použití generátoru v tabulkách, případně v Analysis Toolpacku narážíme na vysoké časové i paměťové² požadavky.

Proto se z výše uvedených důvodů (generování většího množství náhodných čísel, která sama o sobě nemusí být nikam ukládána) budu zabývat hlavně generátorem v makrech. Další nevýhodou je obecně známý fakt u většiny komerčního software, že konkrétní implementace metod (v našem případě jak jsou generována náhodná čísla) není vůbec popsána. Autoři se zřejmě bojí, že by tím ukázali slabiny metody a tím zkazili jméno svého software, avšak z věcného hlediska by tím naopak dali uživateli navíc šanci se případných chybám generátoru vyhnout. Zde je nutno dodat, že prakticky každý generátor pseudonáhodných čísel obsahuje v jistém smyslu systematickou chybu, a je dobré, aby o ní uživatel věděl.

Vraťme se zpět ke generátoru náhodných čísel v makrech VBA. Excel používá lineární (konruentní) generátor typu $X_{n+1} = a * X_n + b \bmod m$, kde $a = 0\text{x}FD43FD = 16\,598\,013$, $b = 0\text{x}C39EC3 = 12\,820\,163$ a $m = 2^{24}$ (m určuje též maximální periodu, které se zde dosahuje)³. maximální periody, tj. 2^{24} . To znamená, že generátor opakuje neustále stejných $2^{24} = 16\,777\,216$ čísel. Běžné generátory dnes mají maximální periodu mnohem delší, např. 2^{32} či 2^{48} , viz Entacher (1997). To může být při simulacích málo, zvláště

² například pro vygenerování vstupu pro baterii testů DIEHARD – 3 milionů náhodných čísel z rovnoměrného rozdělení $R(0,1)$ pomocí Analysis Toolpacku nestačilo přibližně 200MB virtuální paměti

³ Excel ve skutečnosti vrací pouze náhodná čísla od 0 do 1, ale ta generuje s použitím uvedeného vzorečku a výsledek vydělí 2^{24} , z čehož plyne například (minimální) vzdálenost mezi generovanými náhodnými čísly.

pokud na jeden jev spotřebujeme více náhodných čísel a fakt, že do simulace jako náhodná posloupnost vstupují znova táž čísla (při jejich vyčerpání se jde znovu od začátku), nás nutí přemýšlet nad tím, zda výsledek má vůbec smysl.

Zde se opět potvrzuje důležitá myšlenka J. Tvrdíka: „*Excel je přidán ke kritizovaným statistickým programům, neboť je pro statistické výpočty velmi často užíván zejména lidmi potřebujícími statistiku pouze občas a u nich je rozpoznání chybného výsledku ještě méně pravděpodobné než u zkušeného statistika.*“

Další nevýhodou generování náhodných čísel v Excelu (v makrech VBA) je fakt, že nelze nastavit počáteční hodnotu generátoru (tzv. *seed*). Tedy díky tomu prakticky ztrácíme možnost opakovat simulaci se stejnými náhodnými čísly. Tuto možnost nemáme ani v tabulkách. V Analysis Toolpacku tato možnost existuje, avšak čísla se mohou generovat pouze do tabulky (nelze je generovat přímo do nějaké výpočetní procedury).

Výsledky testů náhodnosti

Na testování náhodnosti posloupnosti čísel lze použít například baterii testů DIEHARD (viz Marsaglia (1997)). Tato sada testů realizuje zpravidla představu, jak by se měly chovat některé náhodné jevy (například počet výher v nějaké hře, hodnost náhodné matice apod.). Tyto jevy DIEHARD simuluje tak, že náhodu generuje ze vstupní posloupnosti, tj. z posloupnosti, kterou testuje. U těchto testů je známé pravděpodobnostní rozdělení výsledků a to je porovnáváno se skutečnými výsledky dosaženými na základě simulace. Testuje se vlastně statistická hypotéza o „náhodnosti vstupní posloupnosti“.

Nyní se podívejme, jak dopadl generátor z Excelu v některých testech. Nejprve se podívejme na generátor v makrech VBA. Sada testů uvádí p-hodnotu výsledku na 4–6 desetinných míst, v 8 z 18 testů došlo k naprostému selhání, tj. p-hodnota výsledku byla (v případě vícenásobného opakování se objevila několikrát) 1.0000 resp. 0.0000 (konkrétně to byly testy *birthday spacing*, *binary rank test for 6x8 matrices*, *OPSO*, *OQSO*, *DNA*, *count the 1's test on a stream of bytes* (oba dva) a *minimum distance test*) – jejich popis nechávám na dokumentaci ke zmíněné sadě testů (viz Marsaglia (1997)). K těmto výsledkům udám ještě jednu poznámku – testovaná posloupnost byla délky menší než perioda.

Pokud generujeme náhodná čísla lineárním kongruentním generátorem, bývají nižší bity obvykle méně náhodné, např. pro náš generátor po přenásobení čísel z $[0, 1]$ číslem 2^{24} dostáváme náhodná celá čísla, kde se pravidelně střídají sudá a lichá čísla. Náš generátor v Excelu ovšem vrací čísla z $[0, 1]$

a pokud používáme jen prvních několik cifer, dají se (s trochou dobré vůle při krátkých výběrech) považovat za náhodná. Též v nápovědě popsána metoda⁴ generování čísel od 0 do 100 vzorcem `RAND()*100` nám vybírá „náhodnější“ část pseudonáhodných čísel, což je pozitivní zjištění.

A nyní se podívejme na výsledky generátoru z tabulek. Tento generátor naprosto selhal (viz výše) v testech *OPSO*, *OQSO* a *minimum distance test*.

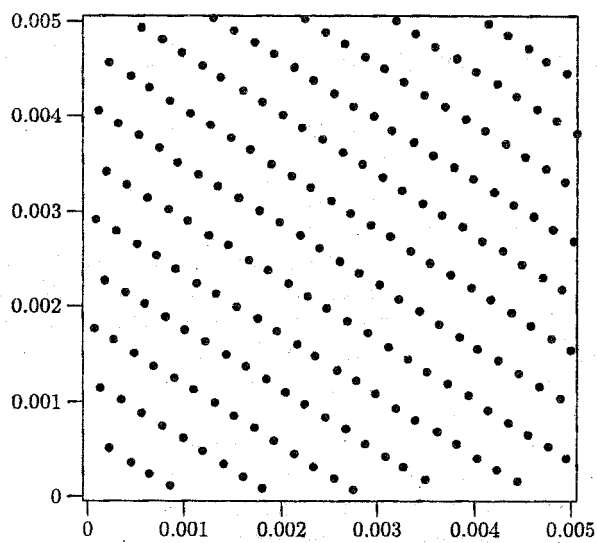
Generátor v Analysis Toolpacku, u nějž by uživatel čekal nejlepší vlastnosti, dopadl kupodivu dost špatně. Tento generátor naprosto selhal v testech *Birthday spacing*, *binary rank test for 31x31 matrices*, *binary rank test for 32x32 matrices*, *Bitstream test*, *OPSO*, *OQSO*, *DNA* a *count the 1's on a stream of bytes*.

Na závěr se vraťme ještě jednou ke generátoru v makrech VBA. Uvedme příklad generování dvojic náhodných čísel při vygenerování celé periody. Na obrázku je vidět, že náhodná čísla se seskupují do nadrovin. Pro generátor náhodných čísel v Excelu jsou tyto nadroviny od sebe hodně vzdálené (na ose x i y jsou uvedeny hodnoty od 0 do 1/200). Pokud nakreslíme tento obrázek pro generátor zabudovaný v programovacím jazyce ANSI C, dosáhneme stejného efektu až při desetinásobném zvětšení obou os, tzn. má nadroviny přibližně desetkrát hustěji ve směru obou os (obdobný obrázek generátoru ANSI C viz Entacher (1997). Podíváme-li se do článku Entacher (1997), zjistíme, že i před přibližně 20–30 lety se používaly lepší generátory (zhlediska počtu nadrovin).

Přidání vlastního generátoru do Excelu

Pokud potřebujeme kvalitní náhodná čísla, je možné si do Excelu napsat vlastní generátor náhodných čísel. Obtížnost je stejná jako napsání makra VBA, tedy několik málo řádek. Tento generátor pak můžeme používat jak v tabulkách, tak v makrech. Pokud si nejsme jisti dobrými vlastnostmi Excelovského ani našeho generátoru a máme-li soubor obsahující dobrá náhodná čísla, můžeme náš generátor zkonstruovat tak, že vrací čísla právě z tohoto souboru.

⁴ týká se sice tabulek, ale lze aplikovat na všechny generátory



Náhodná čísla v Excelu – v makrech VBA

Závěr

Generátor v Excelu v makrech VBA má dost krátkou maximální periodu (konkrétně 2^{24}), které se dosahuje. V testech DIEHARD dopadl tento generátor velice špatně – generovaná čísla tedy moc nenaplní představu o náhodnosti. Nutno dodat, že i jiné často používané generátory pseudonáhodných čísel, např. generátor v jazyce ANSI C, nevykazují o moc lepší výsledky. Dalším zjištěním je, že Excel používá jiný generátor náhodných čísel v tabulkách (tabulkové funkci), resp. v Analysis Toolpacku, a v makrech. Dále je dobré si uvědomit, že pokud generátor náhodných čísel projde sadou testů, neznamená to, že nebude náš konkrétní model používající tento generátor vykazovat systematickou chybu.

Entacher K. (1997). A collection of selected pseudorandom number generators with linear structure.

<http://random.mat.sbg.ac.at/results/karl/server/>

Marsaglia J. (1997). Baterie testů DIEHARD a dokumentace k ní.

<http://stat.fsu.edu/geo/diehard.html>

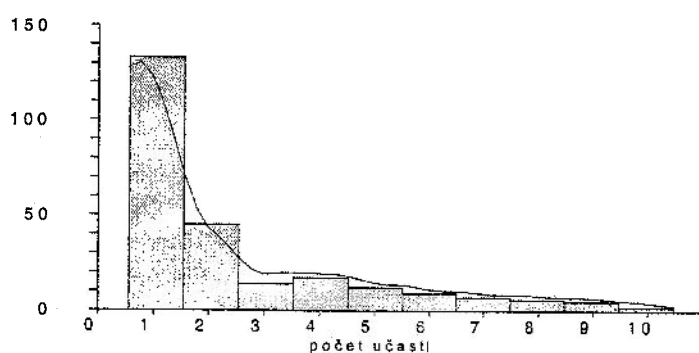
Nápověda k programu Microsoft Excel.

Γvrdlík J.: Věřit statistickému software? Bulletin České statistické společnosti 11, 25–33, 2000.

Netušený objev ze statistické analýzy historie Robustů

Josef Tvrdlík

Součástí Robustu 2000 bylo i sympozium o zdraví jeho účastníků. Asi tři týdny před tím jsme diskutovali s Járrou Antochem, co pro toto sympozium připravit. Shodli jsme se na tom, že nepochybně zajímavé by bylo trochu prozkoumat proces stárnutí dosavadních účastníků. Jára přislíbil pro tento výzkum dodat data, která za dvacet let historie shromáždil v elektronické formě. A opravdu po pár dnech poslal e-mail s množstvím souborů v nejrůznějších formátech. Na první pohled se zdál použitelný jen excelovský soubor, ve kterém byl pečlivě zapsán seznam 250 jmen a vyznačena účast v deseti předchozích ročnících Robustu. Rozdělení počtu účastí vidíte na obr. 1. Jediní dva účastníci bez absence jsou J. Jurečková a J. Antoch a nechyběli ani letos. Na Tomáše Havránka, který jim byl dlouhou dobu po boku při formování historie Robustu, můžeme jen s úctou vzpomenout.



Obr. 1: Rozdělení počtu účastí

Z dalších popisných charakteristik zmíněných dat: průměrný počet účastníků je 60.9, minimum 46 v roce 1980, maximum 83 v r. 1990, průměrný počet účastí jedné osoby je 2.44.

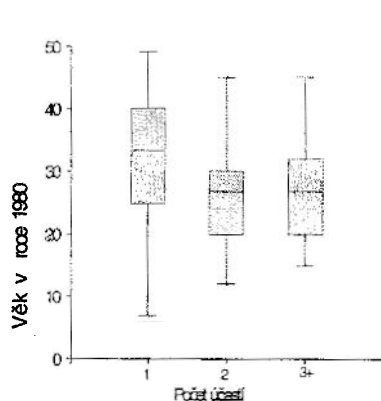
Uvedený datový soubor obsahoval mnoho důležitých informací jako třeba, kdo v Radešíně přednášel v úterý odpoledne a v kolik hodin byla káva, ale bohužel žádný údaj o věku účastníků. Úkol zkoumat proces stárnutí se zdál být nesplnitelný. Opravdový vědecký výzkum však není možný bez odvahy a fantazie. Proto opustíme suchá fakta a vydejme se na cestu dobrodružství

vědeckého poznávání. Je nutno mít cíl a vhodnou vědeckou hypotézu. Obojí máme.

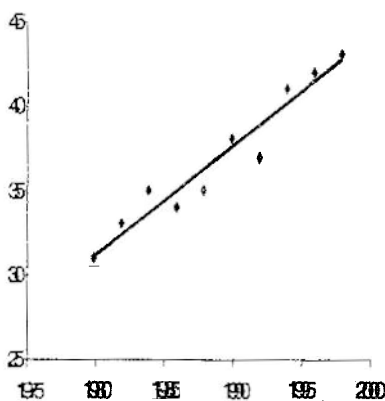
Cíl: Prozkoumat proces stárnutí účastníků Robustu

Hypotéza: Pravidelná účast na Robustu zpomaluje stárnutí

Pohledem na seznam účastníků jsem zjistil, že u řady lidí věk znám nebo jsem schopen ho odhadnout (samozřejmě, že tento odhad je tak zvaně expení). Tak jsem získal věk 114 z 250 dosavadních účastníků a výzkum mohl pokračovat. Už první pohled na data (obr. 2) indikuje, že hypotéza je zdravá. Opakovaná účast na Robustu snižuje věk (dosažený v roce 1980), ale zdá se, že jedno opakování stačí, více účastí nepomáhá k dalšímu snižování.



Obr. 2: Věk podle počtu účastí



Obr. 3: Vývoj průměrného věku účastníků

Průměrný věk v jednotlivých ročnících Robustu a další charakteristiky jsou uvedeny v tab. 1. Povšimněme si, že počet hodnot, ze kterých je počítán průměrný věk, ve všech řádcích přesahuje dvě desítky, takže rozhodně nikde není extrémně malý rozsah výběru.

Graf závislosti průměrného věku na čase je na obr. 3. Tuto závislost můžeme proložit přímkou. Odhad směrnice je $b_1 = 0.645$ (!!!), směrodatná odchylka $s(b_1) = 0.060$, 95%-ní interval spolehlivosti pro β_1 je $(0.508; 0.783)$. Koeficient determinace $R^2 = 0.936$, lineární model vystihuje zkoumanou závislost až překvapivě dobře. Výsledky můžeme bez obav interpretovat: Účastníci Robustu stárnou zhruba o třetinu pomaleji než ostatní populace,

optimisté (na dolním okraji intervalu spolehlivosti) dokonce o polovinu, ale i pesimisté (na horním okraji intervalu spolehlivosti) téměř o čtvrtinu pomaleji, než běží kalendář jiným lidem.

Tab. 1: Průměrný věk účastníků Robustu v jednotlivých ročnících

<i>rok</i>	<i>n</i>	<i>prům. věk</i>	<i>sm.odch.</i>
1980	21	31	5.28
1982	34	33	5.48
1984	44	35	7.06
1986	44	34	6.76
1988	43	35	7.32
1990	52	38	7.99
1992	37	37	7.54
1994	52	41	8.61
1996	44	42	7.83
1998	36	43	8.61

V tomto stadiu jsem považoval výzkum za ukončený, cíle bylo dosaženo, z hypotézy se stává empiricky ověřená vědecká pravda a není důvod váhat s její publikací.

Ale znáte Antocha, nic mu není dost dobré. Přinutil mě podrobněji prohlédnout přílohy e-mailu a v nich nalézt soubor *robust96.dbf*. Po chvíli námahy byla v tomto souboru nalezena položka s rodnými čísly účastníků Robustu 96. Ovšem z 63 účastníků rodné číslo vyplnilo pouze 32. Výzkum však mohl pokračovat na datech opravdu naměřených. Základní charakteristiky věku dosaženého v roce 1980 zjištěného z rodného čísla (veličina *vek80RC*) jsou uvedeny v Tabulce 2 spolu s charakteristikami veličiny *vek80* užívané ve výše uvedeném výzkumu a veličiny *vek80a*, získané z veličiny *vek80*, u níž hodnoty odhadnuté nebo chybějící byly nahrazeny dostupnými hodnotami veličiny *vek80RC*.

Tab. 2: Základní charakteristiky věku zjištěného třemi způsoby

	<i>vek80</i>	<i>vek80RC</i>	<i>vek80a</i>
<i>n</i>	114	32	117
<i>průměr</i>	28.7	27.4	29.0
<i>sm. odch.</i>	9.0	9.7	10.5

Z Tabulky 2 je patrné, že měřením bylo získáno podstatně méně hodnot než odborným odhadem a také to, že základní charakteristiky všech tří

veličin jsou prakticky shodné. Párový t-test pro veličiny vek80 a vek80RC sice přinesl zamítnutí hypotézy o shodě středních hodnot ($n = 29$, difference = -2, $t = -3.35$), ale to nelze brát vážně, neboť statisticky významný byl věcně zcela bezvýznamný rozdíl průměrů (vek80 - vek80a): párový t-test pro vek80 a vek80a: $n = 114$, difference = -0.3, $t = -2.97$. Ale pro Járův klid, aby na tak významném sympoziu nebyly přednášeny výsledky založené na datech získaných jen jakýmsi odhadem, jsem znovu přepočítal průměrný věk účastníků v jednotlivých ročnících, tentokrát z dat naměřených, tj. z veličiny vek80RC a pak i z dat zčásti naměřených, tj. z veličiny vek80a. Zjištěné závislosti se nijak podstatně neliší od té, která je uvedena na obr. 3. Charakteristiky lineárního regresního modelu pro všechny tři závislosti jsou uvedeny v Tabulce 3.

Tab. 3: Souhrnné výsledky odhadů směrnice regresní přímky:

	<i>b1</i>	<i>s(b1)</i>	<i>t</i>	95 % IS		<i>R</i> ²
<i>vek</i>	0.65	0.060	10.8	0.508	0.783	0.94
<i>vek_RC</i>	0.66	0.089	7.4	0.452	0.861	0.87
<i>vek_a</i>	0.65	0.062	10.5	0.507	0.791	0.93

Můžeme konstatovat, že se potvrdily výsledky první etapy výzkumu: *Pravidelní účastníci Robustu stárnou v průměru o třetinu pomaleji než Robust*. Závěr je spolehlivý, neboť nebyla užita žádná simulace, kterou lze dokázat i nemožné.

Možná zobecnění:

- je snadnější data vymyslet než naměřit, proto vymyšlených dat je více,
- z vymyšlených dat jsou spolehlivější výsledky (intervaly spolehlivosti jsou užší, viz tabulka 3),
- vhodně volenou interpretací lze podpořit či zpochybnit téměř cokoliv; jen si musíme vybrat, co se nám více hodí,
- účastníci Robustu nemají úctu k úřadu, nevyplňují rodná čísla a beztak dělají i horší věci,
- opomenutím skrytého cenzorování můžeme získat až nečekaně objevné výsledky.

Valná hromada a volby do výboru

Ve čtvrtek 1.2.2001 proběhla na VŠE v Praze valná hromada České statistické společnosti, na níž byla schválena zpráva o činnosti a zpráva o hospodaření a byl zvolen nový výbor společnosti na funkční období 2001-2003. Bohužel nedošly žádné návrhy na kandidáty do výboru z řad členské základny. Zpráva o činnosti v roce 2000 a o volbách byla zaslána nadřízenému orgánu, Radě vědeckých společností. Odborný referát na valné hromadě přednesla doc. RNDr. Z. Prášková, CSc. z MFF UK v Praze.

Nová podoba www stránky

Doc. J. Žváček připravuje novou podobu www stránky společnosti. Výbor obsáhle diskutoval různé podněty k obsahu stránky, jako seznam statistických pracovišť, seznam odborníků v dané oblasti statistiky (otázka ochoty figurovat na takovém seznamu), odkazy na stránky statistických časopisů, adresy členů (neuvádět explicitně). Stránku si lze prohlédnout na adrese <http://www.statspol.unas.cz>.

Žádáme všechny členy, aby nám posílali jakékoli podněty či připomínky k obsahu a náplni připravované www stránky společnosti.

Bulletin

Dva články z loňského bulletinu byly díky péči prof. ing. J. Jílka, CSc. a doc. RNDr. G. Dohnala, CSc. publikovány již tradičně v příloze časopisu Statistika (č. 2, únor 2001), čímž pokračuje spolupráce, již si Česká statistická společnost velmi váží.

Redakce Informačního bulletinu v zájmu zajištění zajímavé náplně bulletinu uvítá jakékoli příspěvky, které mohou přinést členům užitečné informace, a proto si dovoluje všechny členy oslovit s žádostí o přípravu či získání vhodného materiálu

Spolupráce s ČSÚ

11. ledna 2001 byli členové výboru České statistické společnosti doc. RNDr. J. Antoch, CSc., ing. Z. Roth, CSc. a doc. RNDr. J. Á. Víšek, CSc. přijati předsedkyní Českého statistického úřadu doc. Ing. Marií Bohatou, CSc.,

která se vyslovila pro další pokračování spolupráce ČSÚ s naší společností.
Zároveň bylo dohodnuto uspořádání společného semináře.

Členové výboru na období 2001–2003

Doc. RNDr. Jaromír Antoch, CSc. (MFF UK Praha, předseda)
jaromir.antoch@karlin.mff.cuni.cz

Doc. Ing. Dagmar Blatná, CSc. (VŠE Praha, hospodářka)
blatna@vse.cz

RNDr. Marie Budíková, Dr. (PF MU Brno)
budikova@math.muni.cz

Doc. RNDr. Gejza Dohnal, CSc. (FStr ČVUT Praha, redakce bulletinu)
dohnal@fsik.cvut.cz

Mgr. Miroslav Hartmann (ČSÚ Hradec Králové)
hartmann@gw.czso.cz

RNDr. Marek Malý, CSc. (SZÚ Praha, vědecký tajemník)
marek.maly@szu.cz

Doc. RNDr. Jaroslav Michálek, CSc. (PF MU Brno)
michalek@math.muni.cz

Prof. Ing. Jiří Militký, CSc. (TUL Liberec)
jiri.militky@vslib.cz

Ing. Zdeněk Roth, CSc. (SZÚ Praha)
zdenek.roth@szu.cz

Doc. Ing. Hana Řezanková, CSc. (VŠE Praha)
rezanka@vse.cz

Ing. Josef Tvrdík, CSc. (PřF OU Ostrava)
tvrdik@osu.cz

Doc. RNDr. Jan Ámos Víšek, CSc. (FSV UK Praha, místopředseda)
visek@mbox.fsv.cuni.cz

Doc. Ing. Jiří Žváček, CSc. (Praha, správa www stránek)
jzvacek@atlas.cz

Revizor

RNDr. Jan Klaschka, PhD (ÚI Praha)
klaschka@cs.cas.cz

Akce České statistické společnosti v roce 2001

Statistické dny v Hradci Králové se konají 21.–22. 6. 2001 v prostorách Univerzity Hradec Králové a budou věnovány zajímavým a aktuálním statistickým tématům v příspěvcích jednotlivých účastníků. Cena ubytování je 200 Kč, vložné je 100 Kč. Přihlášky a objednávku noclehu a stravování je potřeba zaslat do 31. května 2001 na adresu organizátora: Prof. RNDr. PhDr. Zdeněk Půlpán, CSc., katedra matematiky, Pedagogická fakulta Univerzity Hradec Králové, V. Nejedlého 573, 500 03 Hradec Králové 3, e-mail: zdenek.pulpan@uhk.cz.

Mezinárodní konference PRASTAN 2001 se bude konat ve dnech 12.–14. 9. 2001 v Kočovcích u Nového Města nad Váhom na Slovensku a její tématické okruhy budou zahrnovat především výuku statistiky a pravděpodobnosti, aplikace statistiky a pravděpodobnosti v praxi, nové trendy ve statistice a pravděpodobnosti, statistický software. Konferenční poplatky jsou: vložné 500 Sk, stravování 400 Sk, ubytování pro účastníky ze Slovenska 300 Sk, pro cizince pravděpodobně vyšší. Závaznou přihlášku včetně objednávky stravy, ubytování a potvrzení o zaplacení účastnického poplatku je nutno zaslat do 31. 8. 2001 na adresu předsedkyně organizačního výboru: Doc. RNDr. Oľga Nánásiová, CSc., Katedra matematiky a deskriptívnej geometrie, Stavebná fakulta STU, Radlinského 11, 813 68 Bratislava, Slovensko, případně e-mailem na adresu tajemníka organizačního výboru: kalina@vox.svf.stuba.sk.



<i>Luboš Pick</i> , Pátek, a k tomu třináctého	2
<i>Martin Betinec</i> , <i>Jaroslav Štěrba</i> , Několik rad do života	4
<i>Petr Máša</i> , Excel a „náhodná“ čísla	12
<i>Josef Tvrdík</i> , Netušený objev ze statistické analýzy historie ROBUSTu	18
<i>Ze Společnosti</i>	22

Informační Bulletin České statistické společnosti vychází čtyřikrát do roka v českém vydání. Předseda společnosti: Doc. RNDr. Jaromír Antoch, CSc., KPMS MFF UK Praha, Sokolovská 83, 186 75 Praha 8, e-mail: jaromir.antoch@karlin.mff.cuni.cz
ISSN 1210–8022

Redakce: Doc. RNDr. Gejza Dohnal, CSc., Jeronýmova 7, 130 00 Praha 3, e-mail: dohnal@fsik.cvut.cz