Výuka Statistiky v České Republice II.

Sborník prací semináře

STAKAN

konaného ve dnech

23.–25. května 2003

v Bystřici pod Hostýnem

Výuka statistiky v České republice II.

Sborník prací semináře STAKAN
zorganizovaného Českou statistickou společností
a Slovenskou štatistickou a demografickou spoločnosťou
za podpory KPMS MFF UK
ve dnech 23. – 25. května 2003 v Bystřici pod Hostýnem

Praha 2004

ISBN 80-239-4086-4

Obsah

PŘEDMLUVA

Z podnětu České statistické společnosti, Slovenské štatistické a demografické spoločnosťou a s podporou výzkumného záměru MSM 113200008 byl ve dnech 23. – 25. května 2003 v Bytřici pod Hostýnem uspořádán metodický semináře STAKAN věnovaný problémům univerzitní výuky aplikované statistiky pro posluchače nematematických oborů. Toto je sborník sestavený ze zvaných přednášek a sdělení přednesených na semináři. Účast padesáti univerzitních učitelů z celého Česko-Slovenska a mimořádně živá diskuse jen potvrdily, že zmíněné problémy jsou aktuální z hlediska nejen vyučujících, ale dotýkají se také velice citlivě pivotální výuky na jednotlivých fakultách (kultura statistických analýz diplomních experimentů, řádné pochopení role stochastických modelů, atd.).

Okruh otázek, které by měly být nejen diskutovány, ale postupně i řešeny, byl poměrně jasně vymezen již na předchozích seminářích pořádaných jak Českou statistickou společností tak Slovenskou štatistickou a demografické spoločnosťou a katedrou matematické statistiky a teorie pravděpodobnosti na Matematicko fyzikální falultě Univerzity Karlovy v Praze. Jedná se především o otázky typu:

- *Jak překonat rozdíly v matematickém vzdělání posluchačů?*
- *Do jaké míry lze obětovat matematickou přesnost při nutném důrazu na aplikace?*
- *Může osvícený nematematik vyučovat statistiku pro své specializované posluchače sám?*
- *Jaké je optimální využití statistického softwaru při výuce?*
- *Jak zamezit vzniku slepé důvěry ve schopnost počitače nezpochybnitelně analyzovat statistická data „automaticky“?*
- *Jaký softwarwe by měl být využíván?*
- *Jaká jsou specifika potřeb jednotlivých oborů (biologie, demografie, ekonometrie, geologie, chemie, atd.)?*

- *Jak hodně potřebujeme k výuce statistiky základy teorie pravděpodobnosti, nebo dokonce teorie míry?*
- *Do jaké míry a jakým způsobem začít se statistickým myšlením již na střední škole?*

Na tyto a některé další otázky odpovídají příspěvky zařazené do sborníku, který představujeme. Využíváme též tuto příležitost, abychom poděkovali recenzentům, především Dr. Ivanu Saxlovi, za pečlivé přečtení rukopisu.

V Praze 17. listopadu 2004
Jaromír Antoch, Gejza Dohnal a Josef Štěpán

ABECEDNÍ SEZNAM ÚČASTNÍKŮ SEMINÁŘE STAKAN III.

Jiří Anděl, Jaromír Antoch, Jitka Bartošová, Josef Bednář, Anna Bérešová, Dagmar Bínová, Petr Blahuš, Dagmar Blatná, Marie Budíková, Gejza Dohnal, Lucie Doudová, Zdeněk Fabián, Ľudmila Gregušová, Petr Hebák, Jan Hendl, Miroslav Holík, Jan Hurt, Jozef Chajdiak, Ivan Janiga, Eva Jarošová, Daniela Jarušková, Jan Kadlec, Zdeněk Karpíšek, Jan Klaschka, Marie Kletečková, Aleš Linka, Jiří Mrověc, Helena Nešetřilová, Karol Pastor, Jan Picek, Zoja Pourová, Marie Prášilová, Zdeněk Půlpán, Patrícia Rexová, Zdeněk Roth, Vladimír Rytíř, Ivan Saxl, Jelena Skibová, Jan Strouhal, Pavel Stříž, Alžběta Svobodová, Lenka Šimonová, Josef Štěpán, Hana Štiková, Antonín Tkadlec, Jiří Trešl, Josef Tvrdík, Jan Ámos Víšek, Petr Volf a Jiří Žváček.

UČEBNÍ POMŮCKA PRO PRÁCI S EKONOMICKÝMI ČASOVÝMI ŘADAMI NA FM VŠE JH

Jitka Bartošová a Monika Uhlířová

Fakulta managementu VŠE, Jarošovská 1117/II, Jindřichův Hradec
barto-ji@fm.vse.cz, uhlir-mo@fm.vse.cz

Pro studenty Fakulty managementu Vysoké školy ekonomické v Jindřichově Hradci je velmi důležité dokázat se orientovat ve statistice. Umožní jim to rychlé vyhodnocování dat a vytvoření vlastního názoru na danou situaci. K procvičení a pochopení všech závislostí a vztahů plynoucích z časových řad se využívá program Statgraphics for Windows. K tomu je ale potřeba velké množství dat, a proto na fakultě managementu byla k tomuto účelu vytvořena databáze Fred. Najdeme v ní mnoho skutečných časových řad setříděných podle různých kritérií. Studenti si jednoduše vyberou vhodnou časovou řadu, uloží ji ve Statgraphicsu a mohou ji statisticky zpracovat. Program Fred usnadní, zrychlí a zpřístupní statistiku i „nestatistikům" a zvyšuje rovněž motivaci při výběru dalších předmětů souvisejících se statistickými metodami.

1. Úvod

Na Fakultě managementu, stejně tak jako na ostatních fakultách Vysoké školy ekonomické Praha, musí každý student projít základními kurzy statistiky A a B. Tyto dva předměty jsou posléze ukončeny soubornou zkouškou. Jedná se tedy o velice důležité kurzy, které jsou posluchači fakulty řazeny k nejobtížnějším a nejvíce časově náročným na přípravu. Převážně celá statistika B je zaměřena na studium ekonomických časových řad. Jejich vyhodnocování probíhá pomocí programu Statgraphics for Windows, který dokáže zpracovat regresní a korelační analýzu, analýzu časových řad atd.

Současná společnost je přehlcena různými informacemi a někdy je opravdu složité nalézt ty konkrétní, které hledáme. Programy, které zjednoduší a urychlí přístup k potřebným datům, jsou velice příznivě přijímány. K takovým programům také patří databáze Fred, kterou vytvořil student Fakulty managementu Stanislav Kříž v rámci interního grantového projektu. Jejím úkolem je získání, analýza a zpřístupnění nejrůznějších (hlavně makroekonomických) ukazatelů. Programy Statgraphics

Klíčová slova: Časová řada, elektronická databáze, statistický program, statistika, učební pomůcka.

for Windows a Fred by tedy měly patřit k učebním pomůckám ve všech ekonomicky a statisticky orientovaných kurzech vyučovaných na fakultě.

2. Statgraphics for Windows

Program Statgraphics vyvinula firma Manugistics v USA (je tedy v angličtině) a v České republice jej distribuuje firma Computer Equipment.

Systém je rozdělen na tyto části:

- Base - poskytuje základní statisticskou analýzu,
- Advanced Regression Models - podává již podrobnější regresní analýzu,
- Advanced Multivariate Methods - provádí vícerozměrné analýzy,
- Time Series Analysis - analyzuje časové řady,
- Quality Control - řídí jakost,
- Experimantal Design - pro navrhování experimentu.

Hlavní výhodou tohoto systému je využívání statistické grafiky.

Ovládání programu Statgraphics je velmi snadné. Statistické metody jsou rozděleny do jednotlivých nabídek podle typu analýzy a typu proměnných. Jednoduše se zde uloží data (nejlépe z MS Excel - pak je systém považuje za číselnou řadu) přes funkce *Ctrl+C* a *Ctrl+V* a pak se s nimi již dá pracovat podle potřeby.

Pro analýzu časových řad se využívá nabídka *Special - Time Series Analysis* a pro určení regresní analýzy hlavně *Relate - Simple, Polynomial, Multiple Regression*. Ke zjištění korelace (síly závislosti) slouží *Describe - Numerical Data - Multiple Var. Analysis - Corelations* a pro analýzu rozptylu použijeme *Compare - Analysis of Variance*. Grafy najdeme v nabídce *Plot*, popřípadě vždy u jednotlivých analýz existuje ikona *Graphic Options*.

Tento program se používá velmi často na různých fakultách vysokých škol, které se zabývají alespoň okrajově statistickými analýzami. Mezi hlavní důvody patří mimo jiné snadné použití, přehledná nabídka, jasně formulované výsledky, úhledné grafické výstupy a možnost použití a uložení mnoha časových řad.

3. Databáze FRED

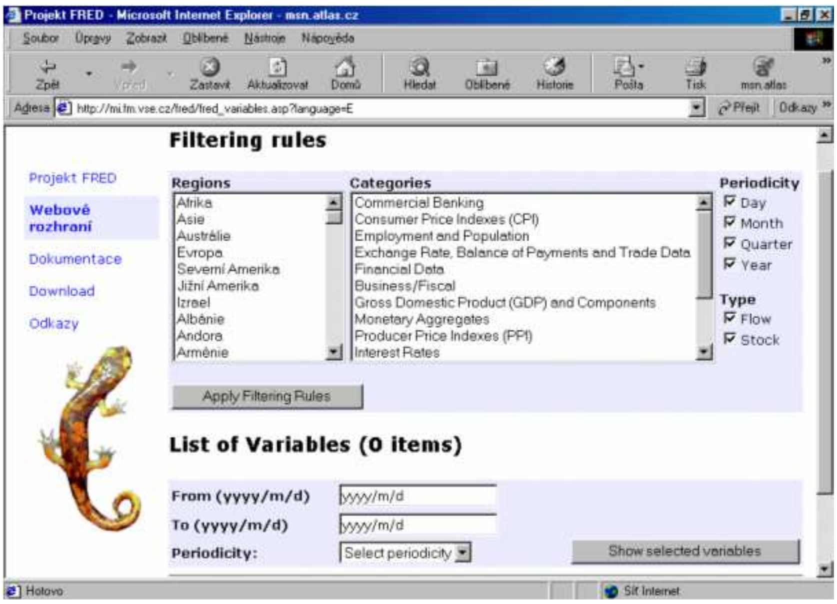
3.1. Seznámení

Programem Fred se zde budeme zabývat trochu podrobněji, protože vznikl jako produkt interního grantového projektu „Fred“ na Fakultě managementu Vysoké školy ekonomické v Jindřichově Hradci. Tato databáze je jednoduše přístupná z webových stránek fakulty, dostaneme se na ni přímo z adresy:
`http://mi.fm.vse.cz/fred/fred.asp`

nebo ji lze stáhnout jako samostatnou aplikaci
`http://mi.fm.vse.cz/fred/download/fredinst.zip`
Úkolem projektu bylo získání, analýza a zpřístupnění nejrůznějších (hlavně makroekonomických) ukazatelů. V současné době zpřístupňuje na internetu některá ekonomická data států USA (cca 400 časových řad), získaná z dalšího grantového projektu, která jsou umístěna na adrese `http://www.stls.frb.org/fred`. Samostatná aplikace obsahuje malou databázi časových řad s několika ukázkovými řadami. V dalším textu se budeme zabývat pouze databází zpřístupněnou přes internet.

Databáze umožňuje:

- přehledný výpis seznamu časových řad,
- vyhledávání dle nejrůznějších kritérií,
- ukládání uživatelského výběru časových řad,
- výpis zvolených časových řad v jednotném formátu (tj. vzájemný převod periodicit denní - měsíční - čtvrtletní - roční),
- komunikaci s ostatními programu pomocí schránky - např. import či
- export dat z MS Excelu, export dat do textového souboru, HTML, MS Word, MS Excel, ... ,
- tvorbu vlastních časových řad.



Obrázek 1. Ukázka vzhledu webovské stránky

3.2. Práce s programem

Verze na internetu je uvedena pod názvem WEB FRED. Je to zjednodušená verze pro práci s časovými řadami, umožňující základní vyhledávání a výpis časových řad v požadované periodicitě. Nejprve se každý musí rozhodnout, zda použít českou nebo anglickou verzi, pak klikne na vybranou položku a rovnou se dostane do systému Fred. Na obrazovce se objeví seznam regionů a kategorií, ze kterých si každý může vybírat jednotlivé časové řady. Data jsou v této verzi zadána do regionu „Spojené státy americké“.

Dále si z nabídky vybereme region, zvolíme kategorii (např. obchodní bankovnictví, indexy spotřebitelských cen, zaměstnanost a populace, měnové kurzy, obchodní bilance, obchodní data, finanční data, obchod/fiskální data, hrubý národní produkt a jeho komponenty, měnové agregáty...), periodicitu a klikneme na položku „*Zobrazit seznam*“. Tím se nám ukáže seznam proměnných (a v závorce celkový počet proměnných) podle parametrů, které jsme zadali. Pak si zvolíme přesně časovou řadu, kterou chceme zobrazit, potvrdíme jí a dáme „*Zobrazit řadu*“.

4. Využití databáze Fred při statistických analýzách

4.1. Charakteristika

Databáze Fred nám poskytuje mnoho potřebných časových řad při procvičování statistických analýz. Někdy je velmi obtížné nalézt reálné řady použitelné v programu Statgraphics. Přesně k tomu slouží projekt Fred, který maximálně zjednodušuje a zpřístupňuje studentům data potřebná při procvičování statistických a ekonomických teorií v praxi.

Pro zpřístupnění této databáze široké studentské i pedagogické veřejnosti Fakulty managementu byla vytvořena v rámci interního grantového projektu č. 21/02 s názvem „Práce s elektronickou databankou ekonomických časových řad“ didaktická pomůcka ve formě návodu v elektonické podobě. Pomůcka je rozdělena do tří částí. První část se zabývá popisem práce s elektronickou databází FRED, druhá popisuje strukturu této databanky a ve třetí je návod na vyhledávání v databázi a nalezená data jsou zde použita při řešení několika vzorových statistických úloh. Vytvořená učební pomůcka je umístěna na internetu na adrese `http://grumpy.fm.vse.cz/iga2102`.

Zjednodušený postup při současném využití obou programů vhodných k výuce statistiky (Fred a Statgraphics for Windows) má tyto fáze:

- výběr časové řady v databázi Fred,
- uložení této řady (Ctrl+C , Ctrl+V) do Statgraphicsu,
- práce s daty v programu Statgraphics.

4.2. Příklad použití

a) Výběr časové řady: Na adrese http://mi.fm.vse.cz/fred/fred.asp, v české verzi, si např. vybereme region Spojené státy americké, kategorii Commercial Banking (tj. Obchodní bankovnictví), periodicitu měsíční, tokovou veličinu a dáme „*Zobrazit seznam*". Vypíše se nám 18 proměnných, mezi nimiž můžeme nalézt např. celkovou investiční činnost ve všech komerčních bankách, celkové úvěry a leasingy v komerčních bankách, celkové spotřebitelské úvěry atd. Z této nabídky si zvolíme např. celkové úvěry a leasingy v komerčních bankách a dostáváme se na tabulku s dalšími informacemi. Po potvrzení se nám vypíše konkrétní časová řada.

b) Uložení řady do Statgraphicsu:

- tuto časovou řadu celou označíme,
- uložíme přes Ctrl+C,
- vložíme do MS Excel (Ctrl+V),
- zkopírujeme do programu Statgraphics.

Tento postup nám přinese jistotu, že Statgraphics bude vybranou časovou řadu považovat za číselná data a umožní nám s ní jednoduše pracovat.
c) Práce s daty ve Statgraphicsu: Nechť např. máme při práci s časovou řadou zadány tyto úkoly:

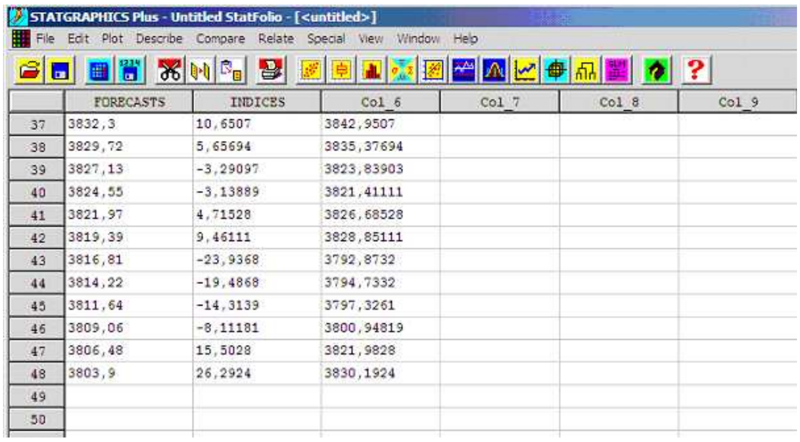
1. Určení typu sezónnosti.
2. Očištění řady od sezónnosti metodou empirických sezónních indexů (rozdílů).
3. Odhad parametrů nejvhodnějšího trendu pro očištěné hodnoty.
4. Odhad budoucího vývoje.
5. Provedení testu sezónnosti.

1. Určení typu sezónnosti: Ve Statgraphicsu tedy máme uloženou řadu úvěrů a leasingů v komerčních bankách od roku 1998 do roku 2000 s měsíční periodicitou. Následuje tento postup: Z hlavního menu vybereme *Special - Time-Series Analysis - Seasonal Decomposition*. Do políčka *Data* uložíme proměnou a vyznačíme měsíční periodicitu a do políčka *Seasonality* napíšeme 12. Dále v nebídce *Seasonal Decomposition* klikneme na *Graphic Options* a vybereme graf *Trend-Cycle* (klouzavé průměry). Z tohoto grafu jsme zjistili, že se jedná o konstantní sezónnost, a proto pravým tlačítkem myši se dostaneme na volbu *Panel options* a zvolíme aditivní model, který podle výše uvedených zjištění je charakteristický pro tuto časovou řadu.

2. Očištění řady od sezónnosti: Postupujeme tedy dále - jsme stále v nabídce *Special - Time-Series Analysis- Seasonal Decomposition*, ale nyní přejdeme na *Save Results* a uložíme si následující řady:

- Trend Cycle (klouzavé průměry),
- Seasonal Indeces (sezónní indexy v
- Seasonally Adjusted Data (očištěné hodnoty).

3. Odhad parametrů nejvhodnějšího trendu pro očištěné hodnoty: Přejdeme do nabídky *Special - Time-Series Analysis - Forecasting*, opět vložíme do *Data* časovou řadu, vyznačíme měsíční data a do *Seasonality* uložíme číslo 12. V nabídce *Tabular Options* vyznačíme *Model Comparisons* a pravým tlačítkem myši vybereme různé trendy. Podle nejmenších hodnoty střední kvadratické chyby (MSE) zjistíme, že nejvhodnějším modelem je Brownovo lineární exponenciální vyrovnání s hodnotou $\alpha = 0,1675$. (Pro lepší orientaci uvádíme obrazovku počítače s výsledky v programu Statgeraphics for Windows.)



	FORECASTS	INDICES	Col_6	Col_7	Col_8	Col_9
37	3832,3	10,6507	3842,9507			
38	3829,72	5,65694	3835,37694			
39	3827,13	-3,29097	3823,83903			
40	3824,55	-3,13889	3821,41111			
41	3821,97	4,71528	3826,68528			
42	3819,39	9,46111	3828,85111			
43	3816,81	-23,9368	3792,8732			
44	3814,22	-19,4868	3794,7332			
45	3811,64	-14,3139	3797,3261			
46	3809,06	-8,11181	3800,94819			
47	3806,48	15,5028	3821,9828			
48	3803,9	26,2924	3830,1924			
49						
50						

Obrázek 2. Výstup z programu Statgraphics for Windows

```
Models
(A) Linear trend = 2986,27 +23,3357 t
(B) Quadratic trend = 3080,45 + 8,4647 t + 0,40192 t^2
(C) Exponential trend = exp(8,00883 + 0,00677734 t)
(D) Simple exponential smoothing with alpha = 0,9999
(E) Brown's linear exp. smoothing with alpha = 0,8325
Estimation period
Model MSE      MAE      MAPE      ME          MPE
------------------------------------------------------
(A)   3090,7   46,2105  1,34534   2,14742E-13 -0,0193847
(B)   1546,33  29,9435  0,884828  3,78956E-13 -0,0120428
```

```
(C)  2530,43  40,6915  1,18002   0,412778    -0,0099849
(D)  881,051  24,727   0,714944  23,3605      0,684489
(E)  327,687  13,8188  0,409005  2,05542      0,0678715
```

4. Odhad budoucího vývoje (na další rok, tedy 12 měsíců): Použijeme nabídku *Special - Time-Series Analysis - Forecasting*. Za vstupní data zvolíme časovou řadu SADJUSTED, měsíční data a do políčka *Number of forecasts* napíšeme číslo 12. Dále zvolíme Brownovo lineárně exponenciální vyrovnání, uložíme předpovědi přes *Save Results* a vyznačíme *Forecasts*. Okno s výsledky zrušíme, zkopírujeme si indexy ze sloupce INDICES pod sebe a sečteme sloupce INDICES a FORECASTS. Tím získáme předpovědi, které zahrnují i sezónní složku.
5. Provedení testu sezónnosti: Zvolíme *Compare - Analysis Of Variance - Multifaktor Anova*. Do vstupního políčka vložíme časovou řadu a vytvoříme dva sloupce s rokem a měsícem, které uložíme do políčka *Factors*. Následně zvolíme v Tabular Options položku ANOVA TABLE. Dostaneme takovýto výstup:

```
Analysis of variance for Col_1 - Type III Sum of Squares
---------------------------------------------------------
Source     Sum of squares  DF Mean Square F-Ratio P-Value
---------------------------------------------------------
MAIN EFFECTS
  A: Col_7      1,92813E6   2  964064,0   562,35  0,0000
  B: Col_8       254838,0  11   21367,0    13,51  0,0000
RESIDUAL          37715,3  22   1714,33
---------------------------------------------------------
TOTAL(CORR.)    2,22068E6  35
---------------------------------------------------------
```

Protože p-hodnota = 0 je menší než α, $\alpha = 0,05$, zjištění provádíme obvykle na 5-ti procentní hladině významnosti), hypotézu H0 (o nezávislosti) na této hladině významnosti zamítáme.V tomto modelu tedy existuje sezónnost, která je významná a nelze jí vynechat.

5. Závěr

Fakulta managementu v Jindřichově Hradci se snaží podporovat všechny iniciativy vycházející od studentů a jednou z nich byl i projekt Fred, jehož výsledky by se měly zařadit do běžné výuky studentů druhého ročníku, kteří musí povinně projít oběma základními kurzy statistiky a složit z nich soubornou zkoušku. Databáze Fred spolu s programem Statgraphics velmi zjednodušuje práci s časovými řadami ekonomických dat (včetně regresní analýzy), zrychlením a usnadněním vyhledávání vhodných dat a tedy jejich zpřístupněním budoucím ekonomům. K usnadnění práce s touto databází a k urychlení jejího zařazení do učebního procesu slouží nově vytvořená didaktická pomůcka „Práce s elektronickou databankou ekonomických časových řad“, kterou najdete na adrese `http://grumpy.fm.vse.cz/iga2102`.

Literatura

[1] Artlová M., Matušů M. a Kozák J. (1999) *Statgraphics Plus for Windows*. Vysoká škola ekonomická v Praze, Praha, ISBN 80-7079-227-2.

[2] Helísek M. (2002) *Makroekonomie: Základní kurz*. Melandrium, Slaný, ISBN 80-86175- 25-1.

[3] Jarošová E. (1994) *Statistika B: Řešené příklady*. Vysoká škola ekonomická v Praze, Praha, ISBN 80-7079-328-7.

[4] Koschin F. (1992) *Statgraphics aneb statistika pro každého*. Grada, Praha, ISBN 80-85424-70-3.

[5] Mach M. (2001) *Makroekonomie II pro magisterské (inženýrské) studium. Část 1, 2.* Melandrium, Slaný, ISBN 80-86195-18-9.

ROBUSTNÍ METODY JAKO DOPLNĚK VÝUKY ZÁKLADNÍHO KURZU STATISTIKY

Jitka Bartošová

Fakulta managementu VŠE, Jarošovská 1117/II, Jindřichův Hradec
barto-ji@fm.vse.cz

Výuka statistiky je na Fakultě managementu Vysoké školy ekonomické v Jindřichově Hradci organizována jednak formou povinných kurzů (jako dvousemestrální „Základní kurz statistiky“) ve druhém ročníku a jednak formou povinně a volně volitelných kurzů ve vyšších ročnících. Jsou to jednosemestrální kurzy navazující na povinný základní kurz, které přispívají k rozšíření a další specializaci statistických znalostí studentů FM VŠE. Tyto předměty jsou zaměřeny jednak na zvýšení teoretických znalostí studentů a jednak na rozšíření škály statistických nástrojů vhodných pro budoucí ekonomy. K takovýmto předmětům patří také nově navrhovaný povinně volitelný předmět „Metody odhadů z ekonomických dat“, který je zaměřen na výběr vhodné metody zpracování konkrétního datového souboru především z hlediska robustnosti odhadovaných charakteristik

1. Úvod

Statistické předměty jsou na Fakultě managementu Vysoké školy ekonomické zařazeny do výuky od druhého ročníku studia. Jejich výuka je postavena na znalostech matematických metod a na potřebách ekonomických aplikací. Cílem výuky statistických předmětů na ekonomických fakultách je vybavení studentů dostatečným arzenálem statistických metod sloužících k tvorbě stochastických modelů ekonomických procesů. K tomu, aby byli studenti vybaveni statistickými metodami v dostatečné šíři a hloubce, nestačí absolvovat povinný základní kurz statistiky. Z toho důvodu je tato základní nabídka výuky statistiky na Fakultě managementu doplněna o nabídku povinně a volně volitelných statisticky orientovaných předmětů.

2. Výuka statistických předmětů na FM VŠE

2.1. Povinný základní kurz statistiky

Statistika je na Fakultě managamentu Vysoké školy ekonomické v Jindřichově Hradci, stejně jako na ostatních fakultách VŠE, povinným předmětem a její výuka je zařazena do 2.ročníku. Cílem výuky je zvládnutí základních statistických

Klíčová slova: Ekonomická fakulta, odhad, robustní metoda, výuka statistiky.

metod v takovém rozsahu a struktuře, které jsou potřebné při běžném stochastickém modelování ekonomických procesů.Obsah předmětu je rozdělen do dvou jednosemestrálních kurzů - Statistika A a Statistika B.

Náplň kurzu A s názvem „Pravděpodobnost a statistika“ tvoří zpracování dat a jejich grafické prezentace, rozdělení četností, charakteristiky souboru, operace s náhodnými jevy a jejich pravděpodobnostmi, systém náhodných veličin, rozdělení některých spojitých a nespojitých náhodných veličin, úvod do statistické indukce, intervalové odhady, testování statistických hypotéz, testy shody parametrů v několika souborech, některé neparametrické testy a analýza rozptylu.

Do kurzu B s názvem „Statistické metody“ jsou zařazena taková témata, jako je analýza závislostí mezi ekonomickými jevy, regresní a korelační analýza dvou a více proměnných, elementární vlastnosti časových řad, jejich modelování, analýza sezónní složky a popis náhodné složky, adaptivní přístupy k modelování, predikce a korelace v časových řadách, srovnávání ekonomických jevů pomocí indexů a rozdílů a základy statistické přejímací a regulační kontroly.

Výuka probíhá v týdenních cyklech formou přednášek a cvičení (2+1). Teoretické pojmy jsou aplikovány ve slovních úlohách s ekonomickou tematikou. Cvičení probíhají v počítačových učebnách a při řešení úloh jsou využívány statistické softwarové produkty Statgraphics for Windows a MS Excel.

2.2. Povinně a volně volitelné statistické předměty

Na povinný kurz ve vyšších ročnících navazuje celá řada teoreticky i prakticky orientovaných volitelných statistických předmětů. Patří k nim např. „Praktikum ke statistice“, „Pravděpodobnostní modely“, „Stochastické modely“, „Základy ekonometrické analýzy“, „Management jakosti“, „Metody manažerského rozhodování“, „Prognózování pro manažery“ a „Analýza dat v manažerském rozhodování“. Cílem těchto předmětů je prohloubit a rozšířit základní znalosti statistických metod o další specializované statistické postupy s přihlédnutím k potřebám ekonomické praxe.

Do skupiny povinně volitelných statistických předmětů byl pro příští semestr navržen další (dvoukreditový) předmět s identem MMH231 a názvem „Metody odhadů z ekonomických dat“. Předmět je vhodný pro studenty, kteří absolvovali základní kurz statistiky, tedy od 3. ročníku výše. Výuka by měla probíhat formou přednášek a cvičení (1+1) a měla by být zakončena zápočtem. Součástí požadavků na udělení zápočtu je úspěšné zpracování zadaného úkolu formou seminární práce. Předmět je zaměřen na metody výběrů a určování výběrových charakteristik s přihlédnutím k jejich robustnosti. Jeho obsah doplňuje a rozšiřuje sortiment metod statistické indukce probíraných v základním kurzu. Obsah předmětu tvoří:

- Pořádkové statistiky a jejich rozdělení.
- Testy a odhady založené na pořádkových statistikách.
- Neparametrické odhady.
- Robustní odhady parametru polohy.
- Robustní odhady parametru měřítka.
- Metody výběru dat.
- Klasické výběrové statistiky a jejich vlastnosti.
- Lineární modely.
- Odhady v lineárních modelech.
- Robustní odhady v lineárních modelech.
- Volba vhodné metody, plánování experimentu.

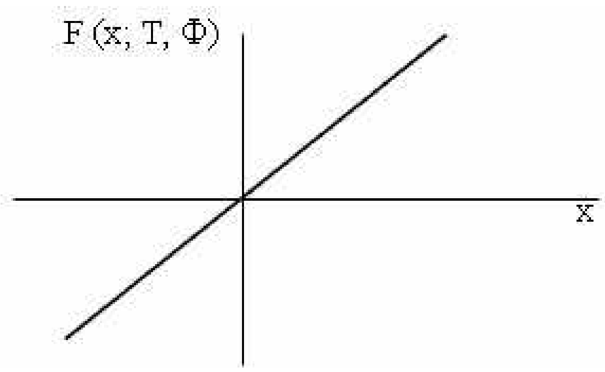
3. Výuka robustních metod na ekonomických fakultách

3.1. Teoretická část výuky

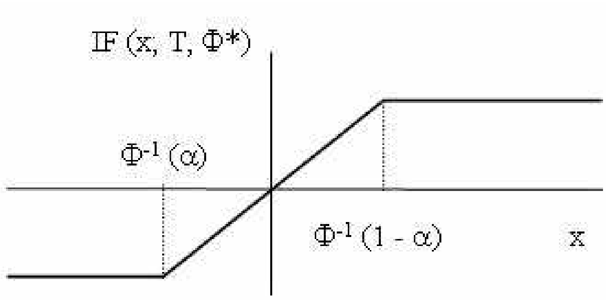
Teorie odhadů, tak, jak je probírána v základním kurzu statistiky, vychází z principu stability, tj. z předpokladu, že malé chyby v modelu by měly jen v malé míře ovlivňovat závěry statistických analýz dat. Mnohé běžně používané statistické metody jsou však velmi citlivé na malé odchylky od předpokladů o matematických modelech. Takovým příkladem jsou metody, které jsou optimalizovány vzhledem k předpokladu o normalitě rozdělení. Reálná ekonomická data obvykle předpoklad o normalitě nesplňují. Často obsahují odlehlá pozorování, která mohou vznikat jednak velkými nepřesnostmi při měření případně hrubými chybami při pozorování nebo přenosu dat, jednak se mohou v reálném datovém souboru skutečně vyskytovat. Na takový datový soubor pak můžeme nahlížet jako na náhodnou směs „dobrých“ a „špatných“ pozorování. Také zaokrouhlováním a slučováním dat do tříd, které se běžně při zpracování reálných ekonomických souborů používá, vnáší do modelu nepřesnosti. Proto je vhodné při jeho vyhodnocování použít některou z robustních metod, které jsou necitlivé na malé odchylky od předpokladů o modelu.

Z robustních metod odhadů parametru polohy a měřítka, se kterými by se ekonomové mohli seznámit, je nejznámější a zároveň nejjednodušší na vyjádření třída L-odhadů, proto se v nově navrhovaném volitelném předmětu zaměříme pouze na tyto odhady a jejich vlastnosti. Mezi jejich hlavní výhody patří snadná interpretovatelnost a jednoduchost výpočtu (pokud ovšem máme k dispozici dostatečně rychlý algoritmus pro třídění dat). Studenti budou v nově navrhovaném předmětu seznámeni s typickými zástupci L-odhadů parametru polohy a měřítka, jako je např. α-useknutý průměr, α-winsorizovaný průměr, výběrový medián, Gastwirthův odhad, mezikvartilové rozpětí a další.

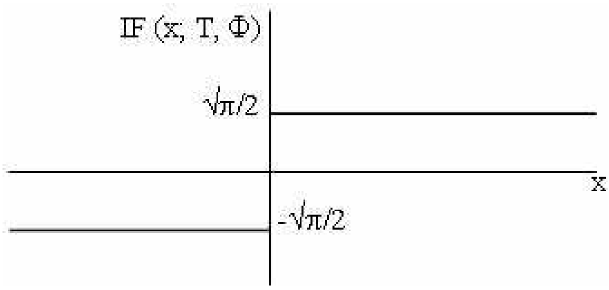
Protože s robustními vlastnostmi odhadů je úzce spojen pojem influenční (vlivové) funkce, která měří vliv přidání dalšího pozorování s libovolně velkou hodnotou do rozsáhlého souboru dat na výsledný odhad parametru, považuji znalost tohoto pojmu za důležitou pro pochopení pojmu robustnost odhadu. Navíc pomocí hodnot influenční (vlivové) funkce lze také vyjadřovat asymptotický rozptyl odhadů, což nám umožňuje porovnávat mezi sebou každé dva odhady z hlediska jejich vydatnosti prostřednictvím podílu jejich asymptotických rozptylů. Studenti by si také měli uvědomit, jak vypadá influenční (vlivová) funkce některých známých odhadů a zároveň by měli vidět souvislost mezi neohraničeností influenční (vlivová) funkce a nerobustností odhadu. Tyto skutečnosti lze úspěšně demonstrovat např. na rozdílech ve tvaru influenčních (vlivových) křivek pro aritmetického průměru, pro α-useknutý průměr ($0 < \alpha < 0,5$) a pro medián, jak je vidět z následujících obrázků, kde je jako model použito standardní normální rozdělení Φ, které je symetrické kolem nuly.



Obrázek 1. Influenční funkce pro výběrový průměr



Obrázek 2. Influenční funkce pro -useknutý průměr



Obrázek 3. Influenční funkce pro výběrový medián

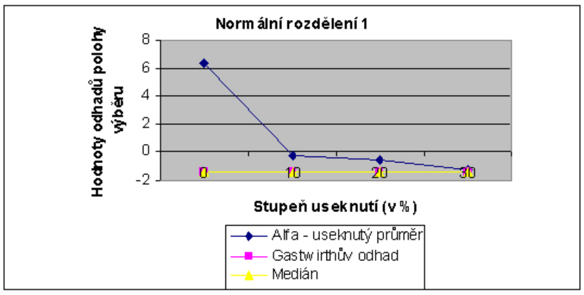
Dalším důležitým pojmem, se kterým by se měli studenti vysokých škol ekonomického zaměření seznámit, je míra robustnosti odhadů (jejich necitlivosti na „malá" narušení modelu), která má buď lokální nebo globální charakter. Měli by se dovědět, že lokálními měrami robustnosti (definovanými pomocí influenční (vlivové) funkce) je míra citlivosti na výskyt hrubých chyb γ*, bod zamítání ρ* a míra citlivosti lokálního posunutí λ*, i to, že pokud $\gamma * < \infty$, nazýváme odhad B-robustním, což znamená, že přidáním dalšího pozorování s libovolně velkou hodnotou nezpůsobíme selhání odhadu. Z globálních měr robustnosti by měli znát význam bodu zvratu a asymptotického bodu zvratu odhadu pro konečné výběry s_n*, který nás informuje o tom, jaký podíl pozorování v souboru můžeme libovolně zaměnit, aniž by došlo k úplnému selhání odhadu. Studenti by se také měli seznámit v nově navrhovaném předmětu s hodnotami měr robustnosti u nejznámějších typů odhadů a měli by pochopit souvislost mezi těmito hodnotami a tvary influnčních křivek.

3.2. Praktická část výuky

Praktické procvičování teoretických poznatků bude probíhat jednak na simulovaných datech a jednak na reálných ekonomických datech, která jsou na Fakultě managementu v dostatečném množství k dispozici díky existenci elektronické databáze Fred. Na ukázku této praktické výuky byl zde simulován malý soubor dat s normálním rozdělením. Vygenerovaná data byla dodatečně uměle kontaminována několika odlehlými hodnotami z množiny 5, 10, 50, 100. U souborů byl postupně zvyšován stupeň kontaminace, takže v souboru „Normalní rozdělení 1" jsou pouze dvě vybočující hodnoty (stupeň kontaminace $s = 10\%$), v souboru „Normalní rozdělení 2" jsou čtyři ($s = 20\%$) a v souboru „Normalní rozdělení 3" je šest těchto odlehlých hodnot ($s = 30\%$). Takto simulované datové soubory jsou zachycena v následující tabulce. V další tabulce jsou zachyceny hodnoty některých odhadů a na grafech je ukázán vliv stupně kontaminace odlehlými hodnotami na tyto odhady.

Normální rozdělení 1	Normální rozdělení 2	Normální rozdělení 3
param: (0,10)	param: (0,10)	param: (0,10)
-19,6898	-19,6898	-19,6898
-12,8709	-12,8709	-12,8709
-11,6522	-11,6522	-11,6522
-9,3214	-9,3214	-9,3214
-7,87647	-7,87647	-7,87647
-6,10104	-6,10104	-6,10104
-5,34613	-5,34613	-5,34613
-4,8941	-4,8941	-4,8941
-3,28805	-3,28805	-3,28805
0,530675	0,530675	0,530675
2,60198	2,60198	2,60198
3,13702	3,13702	3,13702
5,67075	5,67075	50
10,3687	10,3687	50
11,2954	50	50
11,8238	50	100
50	100	100
100	100	100

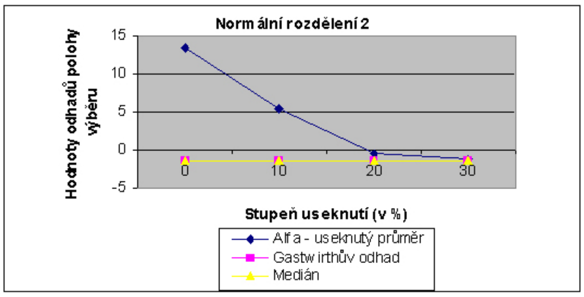
Tabulka 1. Vygenerované soubory kontaminované odlehlými hodnotami.



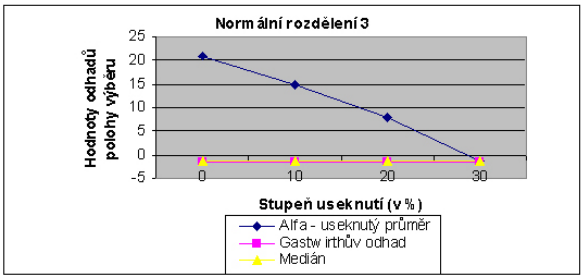
Graf 1. Průběh α-useknutého průměru, mediánu a Gastwirthova odhadu

Výběrové charakteristiky	Normální rozdělení1	Normální rozdělení2	Normální rozdělení3
0% usek. průměr	6,3549	13,4038	20,8461
10% usek. prům.	-0,217933	5,27355	14,8422
20% usek. prům.	-0,51966	-0,51966	7,87639
30% usek.prům.	-1,20977	-1,20977	-1,20977
Medián	-1,37869	-1,37869	-1,37869
Gastwirth. odhad	-1,440682	-1,440682	-1,440682
Dolní kvartil	-7,87647	-7,87647	-7,87647
Horní kvartil	10,3687	10,3687	50,0
33,33% kvantil	-6,10104	-6,10104	-6,10104
66,67% kvantil	3,13702	3,13702	3,13702
Směrod. odchylka	27,7672	36,6712	42,4496
Mezikvart. rozpětí	18,2452	18,2452	57,8765

Tabulka 2. Některé typy odhadů



Graf 2. Průběh α-useknutého průměru, mediánu a Gastwirthova odhadu



Graf 3. Průběh α-useknutého průměru, mediánu a Gastwirthova odhadu

4. Závěr

Předložený článek je zaměřen na některé problémy výuky statistických předmětů na fakultách s ekonomickým zaměřením. Poukazuje na skutečnost, že teorie odhadů, tak, jak je vyučována v povinném základním kurzu statistiky, je postavena na předpokladu, že malé chyby v modelu mohou pouze v malé míře ovlivňovat závěry statistických analýz dat, avšak ve skutečnosti mnoho běžně používaných statistických metod je na tyto malé odchylky od předpokladů velmi citlivých. Tyto problémy jsou patrné především u velice často používaných odhadů, které jsou optimalizovány vzhledem k předpokladu o normalitě rozdělení, jako je např. aritmetický průměr. Proto je vhodné doplňovat povinnou výuku ještě některým volitelnými předmětem který by podal studentům základní informace o robustních metodách odhadů. V článku je popsán cíl, obsah a metody výuky nově navrhovaného povinně volitelného předmětu na FM VŠE s názvem „Metody odhadů z ekonomických dat", který podává studentům informace o robustních a neparametrických metodách odhadů, o měrách robustnosti, významu influenční funkce, tvarech influenčních křivek některých známých odhadů,

Literatura

[1] Anděl J. (2002) *Základy matematické statistiky*. Preprint MFF UK, Praha.

[2] Antoch J. a Vorlíčková D. (1992) *Vybrané metody statistické analýzy dat*. ACADEMIA, Praha, ISBN 80-200-0204-9.

[3] Hampel F.R. (1971) *A general qualitative definition of robustness*. Ann. Math. Statist. **42**, 1887–1896.

[4] Hampel F.R. (1982) *Robuste Schatzungen: Ein anwendungsorentierter Ueberblick*. Biometrical Journal **22**, 3–21.

K VÝUČBE ŠTATISTIKY A VEDECKEJ PRÁCE V KURIKULE VYSOKOŠKOLSKÉHO ŠTÚDIA V ODBORE OŠETROVATEĽSTVO

Anna Bérešová a Mária Kovárová

LF UPJŠ, Ústav sociálnej medicíny, Trieda SNP 1, 040 66 Košice
beresova@central.medic.upjs.sk

1. Úvod

Ošetrovateľský výskum je neoddeliteľnou súčasťou odboru Ošetrovateľstvo. Táto skutočnosť sa deklaruje najrôznejšími spôsobmi vo všetkých učebniciach ošetrovateľstva. Sestry sa s týmto faktom zoznamujú a učia sa ho teoreticky už od základných stupňov profesionálnej prípravy. Problémom je, že väčšina z nich si skutočný výskum, vedeckú prácu, v praxi ani nevie predstaviť. Ich znalosti z oblasti metodológie výskumu a štatistiky, s ktorými prídu na fakultu, sú nedostatočné. So skutočným výskumom sa v klinickej a komunitnej praxi stretávajú len ojedinele. Posilnenie kompetencií a praktických zručností sestier v tejto oblasti je jednou zo základných úloh ich vysokoškolského vzdelávania a znamená, akým spôsobom a v akom rozsahu sa v jednotlivých ročníkoch sestry zoznamujú s výskumnou metodológiou a základmi štatistiky.

Tejto úlohy sa zhostila v plnom rozsahu LF UPJŠ v Košiciach, kde univerzitné vzdelávanie sestier prebieha už jedenásty akademický rok v bakalárskej forme. Do roku 2002 bolo prijímaných na túto dennú formu štúdia ročne v priemere 15 študentov. Výrazným medzníkom v štúdiu na našej fakulte sa stal školský rok 2001/2002. Fakulta získala akreditáciu na zahájenie magisterského štúdia v ošetrovateľstve v 5-ročnej dennej forme a od školského roku 2003/2004 bude zahájené vzdelávanie aj externou formou.

Naším prioritným úsilím je viesť pregraduálnych poslucháčov k schopnosti porozumieť základom výskumu a štatistiky v ošetrovateľstve. Výučba sa realizuje formou jednohodinovej prednášky a dvoch hodín cvičení. Základy výskumu v ošetrovateľstve sú predmetom výučby počnúc prvým ročníkom štúdia.

Program výučby predmetu je zostavený do štrnástich celkov, v ktorých popri stručnej teórii dominuje praktické demonštrovanie a precvičovanie nielen jednotlivých krokov výskumu, ale aj precvičovanie základných štatistických výpočtov a

---

Kľúčové slová: Základy výskumu, výučba štatistiky, odbor ošetrovateľstvo, SWOT analýza.

jednoduchých testov. V obsahovej náplni sa zameriavame na používanie deskriptívnych metód, pozornosť upriamujeme na porozumenie kvantitatívnej a kvalitatívnej analýze dát. Princípy výberov metódy zberu dát v ošetrovateľstve, jednotlivé formy výskumných štúdií a plánov sú organickou súčasťou predmetu. Osobitný dôraz kladieme na výber hodnotiacich meracích nástrojov využívaných v ošetrovateľstve. Významné miesto v ošetrovateľskom výskume v súčasnej dobe má validizácia hodnotiacich mierok pri získavaní porovnávateľných a primeraných údajov od rôznych skupín klientov. (viď príloha)

Z tohto dôvodu každý študent musí zvládnuť podstatu spoľahlivosti a validity mierok. Pozitívom je, že podstatná časť denných študentov ovláda aspoň základné zručnosti práce s PC. Študenti sú schopní používať počítačový softvér EXCEL, a to nielen pri tvorbe vlastnej databázy, ale aj pri deskriptívnej a induktívnej analýze dát, zlepšuje sa aj ich zručnosť grafickej prezentácie vlastných prác. To sa však nedá konštatovať u študentov externých, kde na štúdium sú prijímané hlavne sestry z praxe a vo vekovej kategórii, ktorá nemala umožnený prístup k výuke informatiky.

Popri nesporných pozitívach je potrebné uviesť aj úskalia, ktoré musíme prekonávať, a s ktorými sa nesporne stretávame všetci vo výučbe predmetov zameraných na štatistiku či vedecký výskum. Štatistika a výskum v ošetrovateľstve, je pre našich študentov téma, ktorá je nielen veľkou neznámou, ale stretávame sa s názormi, že témou zbytočnou. Študent musí zvládnuť nový a veľmi špecializovaný slovník, niekoľko komplexných procesov a metód ešte pred tým, než začne spracovávať svoju tému, zhromaždí relevantnú literatúru, aplikuje literárne poznatky na svoju výskumnú úlohu. Pregraduálni študenti musia veľmi dobre ovládať prácu s PC, orientovať sa v knižnici, vedieť kriticky myslieť a čítať odborné texty a mať primerané zručnosti v hodnotení podkladových materiálov.

Cieľom nášho príspevku je poukázať nielen na pozitíva ale najmä problémy, s ktorými sa vo výučbe predmetu Vedecký výskum stretávame. Formou SWOT analýzy sme zhodnotili naše doterajšie skúsenosti vo výučbe.

1.1. S – silné stránky

Za silné stránky považujeme to, že počas 12 rokov predmet osciloval medzi 1.–3. ročníkom. V ďalšom za silnú stránku považujeme, že sa predmet podarilo ukotviť už v 2. ročníku, nakoľko započať výučbu až v 3. ročníku sa javí neskoro. Dôvodom je, že v tomto ročníku, v letnom semestri, poslucháči spracovávajú diplomovú prácu v bakalárskom stupni vzdelávania.

Ďalšou silnou stránkou je precvičovanie praktických príkladov z praxe v rámci cvičení. Príprava malých vedeckých projektov k zápočtu z vedeckej práce výrazne napomáha poslucháčom pri výbere témy na diplomovú prácu, pretože tento môže v nej vyústiť v konečnom štádiu.

Najvýraznejšou a pre nás najčitateľnejšou silnou stránkou je výrazné zlepšenie vedomostí a hlavne zručností vo vedeckej práci a aplikácii štatistických postupov v nej, už v 3.ročníku štúdia.

1.2. W – slabé stránky

Najslabšou stránkou je málo výučbových hodín. V počiatkoch vyučovania tohto odboru na našej fakulte bol obnos hodín 2 hodiny prednáška a 2 hodiny cvičenia. Postupom času súčasne s implementovaním nových predmetov do výuky, obnos hodín na predmet je už len 1/2. Hrozí tendencia znížiť počet hodín cvičení na jednu hodinu, pri zachovaní 1 hodiny prednášky.

Tým pádom je menej priamej výučby, čo je nutné vykrývať nepriamou výučbou - konzultáciami s poslucháčmi, čo sa na druhej strane prejavuje zvýšenou časovou náročnosťou na vyučujúceho.

Ďalšie slabé stránky, ktoré sa v desaťročnom období výuky predmetu objavili, možno zhrnúť do niekoľkých bodov.

1. Poslucháči nemajú predchádzajúce skúsenosti s vedeckým výskumom ani so základnou štatistikou.
2. Sú to obavy študentov, či zvládnu tento predmet a nároky, ktoré budú na nich kladené.
3. Zaradenie predmetu v niektorých školských rokoch až v treťom ročníku.

1.3. O – výzvy

Veľkou výzvou do ďalšieho skvalitňovania výuky predmetu je fakt, že v roku 2002 sa prvý raz zapojili do fakultnej ŠVOČ aj poslucháčky - bakalárky. Vedenie fakulty aj hodnotiaca komisia podali návrh, aby si poslucháči ošetrovateľského smeru vytvorili samostatnú sekciu pre Ošetrovateľstvo.

Ukazuje sa ochota u poslucháčov pracovať na vedeckých problémoch katedry (projekt Zdravé starnutie a NEXT) a klinických pracovísk. Poslucháči sú v základoch pripravení pokračovať vo vedeckej práci aj v praxi

1.4. T – hrozby

Vidíme ich predovšetkým pre učiteľov. Vo výuke štatistiky u poslucháčov, ktorí nemajú dostatočné matematické vedomosti zo stredných škôl, je treba vedieť povedať základ – esenciu. Našou základnou tézou je „neísť hlboko do matematiky“. Nemôžeme zasypať poslucháčov matematickými dôkazmi, lebo to odrádza poslucháčov podrobnejšie sa venovať predmetu. Hľadaním, uvádzaním a precvičovaním praktických príkladov aj z ošetrovateľskej praxe, sa snažíme jednoducho, po „ľudsky“ podať preberanú tému. Poslucháči v rámci samostatnej práce v skupinách sami hľadajú témy, odborné články v časopisoch, v ktorých sa nachádzajú aj štatistické zhodnotenia, grafické vyjadrenia skúmanej problematiky. Na týchto príkladoch sa snažíme demonštrovať základy štatistiky.

Všetky nároky a zručnosti, ktoré majú poslucháči zvládnuť, je dobré podporiť vytvorením bezpečného a štruktúrovaného výučbového prostredia. To znamená, že učiteľ jasne definuje ciele predmetu, študenti vedia čo a kedy sa od nich očakáva, ako budú hodnotení. Na dosahovanie cieľov vo výučbe jednotlivých kapitol sa používajú alternatívne spôsoby. Dobrá príprava výučbových materiálov a aplikácia štatistiky a výskumu na ošetrovateľskú prax predstavujú výzvy aj pre študentov. Študenti by mali vnikať do metód výskumu a do štatistiky interaktívnym spôsobom. Konečným cieľom je schopnosť študenta porozumieť dôležitosti štatistiky, vedeckého výskumu a jeho úlohe v ošetrovateľskej praxi.

Zavedením takého prístupu sa zvyšujú nároky na čas v príprave učiteľov a v ďalšom vyvstáva požiadavka na tvorbu nových výukových pomôcok.

2. Záver

Desaťročné skúsenosti ukazujú, že investícia do pregraduálnej prípravy v oblasti metodológie výskumu v ošetrovateľstve postupne prináša svoje ovocie. Rekognoskácia sestier významne v zrastá v zdravotníckej, ale aj vedeckej obci. Pregraduálne vzdelávanie sestier v oblasti vedeckého výskumu a štatistiky nepochybne zobralo vietor z plachiet pregraduálnemu vzdelávaniu lekárov, našu fakultu nevynímajúc.

Stále zlepšujúca sa kvalita výstupov, a to nielen formálna ale aj obsahová, analytická a interpretačná, nadobúda geometrický vývojový trend. Absolventi štúdia sa aktívne zapájajú do vedeckej činnosti svojich pracovísk a sú tiež aktívnymi prispievateľmi do odborných časopisov. Súčasná orientácia výskumu smeruje viac k ošetrovateľskej praxi v porovnaní s jeho začiatkami. Poslucháči bakalárskeho a následne aj magisterského štúdia sa učia umeniu a vede ošetrovateľskej praxe už z iného pohľadu a v inom kontexte, ako to bolo v začiatkoch zavádzania tohto typu štúdia na Slovensku. Tento proces podľa nášho názoru výrazne ovplyvnili aj existujúce kurikulá pregraduálneho univerzitného vzdelávania sestier.

3. Príloha

Harmonogram prednášok z predmetu *Základy výskumu v ošetrovateľstve* pre 3. ročník ošetrovateľstva – bakalárske štúdium

1. týždeň: Vedecký výskum v ošetrovateľstve.
   - Miesto vedeckého výskumu v ošetrovateľstve.
   - Vedecký prístup.
   - Zdroje poznania.
   - Výskumné metódy.
   - Etické aspekty vedeckého výskumu.
2. týždeň: Vedeckovýskumný proces.
   - Hlavné fázy (preliminárna, plánovacia, empirická, analytická, prezentácia výsledkov) a základné kroky (formulácia problému, formulácia hypotéz, plánovanie výberu, pilotný výskum, zber dát, analýza a interpretácia výsledkov) výskumného procesu.
   - Časová organizácia výskumného projektu.
3. týždeň: Premenné v experimente a ich kontrola.
   - Definovanie pojmov: experiment a premenná.
   - Kategorizácia premenných.
   - Vnútorná validita experimentu.
   - Základné spôsoby kontroly nežiadúcich premenných.
4. týždeň: Preliminárne výskumné kroky.
   - Výber a definovanie výskumného problému.
   - Zdroje problémov.
   - Kritériá na zhodnotenie výskumných problémov.
   - Hypotézy - charakteristika, formulácia.
5. týždeň: Výskumné plány v ošetrovateľstve.
   - Experimenty, kváziexperimenty, neexperimentálne výskumné plány.
   - Longitudinálne a transverzálne štúdie.
6. týždeň: Princípy výberov.
   - Výber populácie a výber vzorky.
   - Základné požiadavky na výber vzorky (náhodnosť, reprezentatívnosť).
   - Základné typy výberov.

7. týždeň: Metódy zberu dát v ošetrovateľstve.
    - Sebaposudzovacie techniky, škály a štandardné sebaposudzovacie mierky.
8. týždeň: Ďalšie metódy zberu dát v ošetrovateľstve.
    - Biofyziologické metódy.
    - Observačné metódy.
    - Q-metodológia.
9. týždeň: Kritériá na výber a hodnotenie meracích nástrojov.
    - Základné pojmy (celkové skóre, pravdivé a chybové skóre).
    - Chyby merania.
    - Reliabilita a validita mierok
10. týždeň: Kvalitatívna analýza dát.
    - Kontrola a hodnotenie výskumného materiálu.
11. týždeň: Kvantitatívna analýza.
    - Deskriptívna a frekvenčná štatistika.
    - Variabilita.
    - Testovanie rozdielov medzi 2 skupinami.
    - Testovanie rozdielov medzi 3 a viacerými skupinami.
12. týždeň: Kvantitatívna analýza - náročnejšie štatistické metódy.
    - Analýza rozptylu.
    - Regresná analýza.
    - Faktorová analýza.
13. týždeň: Využitie výpočtovej techniky vo výskume.
    - Výber vhodných štatistických programov a príprava dát na analýzu.
14. týždeň: Komunikácia vo výskumnom procese.
    - Interpretácia a prezentácia výsledkov výskumu.

Náplňou praktických cvičení je aplikácia teoretických poznatkov získaných v rámci prednášok.

TVORBA STUDENTSKÝCH PROJEKTŮ V SYSTÉMU SAS NA PEF ČZU

Dagmar Bínová

Česká zemědělská univerzita v Praze, Provozně ekonomická fakulta, katedra statistiky, Kamýcká 129, 165 21 Praha 6

binova@pef.czu.cz

Příspěvek je zamyšlením nad možností využití prvků projektového vyučování při výuce statistického softwaru – konkrétně systému SAS – na PEF ČZU, neboť i při vysokoškolské výuce má jít především o aktivní tvůrčí přístup k řešení vybraných témat.

1. Úvod

Na univerzitách, kde se student odborně specializuje, je nutné věnovat pozornost nejen oboru samotnému, ale i zpracování množství kvantitativních i kvalitativních údajů. Cílem předmětů zaměřených na výuku statistiky je mimo jiné i využití statistického softwaru. Tím by nemělo být jen naučit studenty daný software ovládat, ale i použít jej pro hodnotnou statistickou analýzu a především naučit je správně interpretovat její výsledky.

2. Cíl

Částečné odpovědi na otázky „Jak docílit výše uvedených pedagogických cílů?“ a „Jaký způsob výuky zvolit?“ jsou obsahem tohoto příspěvku. Cílem není uvést přesný návod, ale zamyslet se nad současnou praxí a přinést pohled na výhody a nevýhody vyzkoušených postupů vyučování.

3. Metodika

Pedagogickým pracovníkům se nabízí několik možností, jak seznamovat a učit studenty pracovat s odbornými programy a systémy. Obecně se dost často využívá osvědčeného přístupu, kdy výuka spočívá v detailním a dokonalém popisu struktury programu, vysvětlení položek menu a operací funkčních kláves. Jiný postup je založen na tom, že vyučující zmíní, názorně vysvětlí a předvede jen vybrané nejzákladnější nabídky programu, a potom předvede práci s programem na zajímavém jednoduchém příkladu, v němž jsou použity i složitější prvky.

Klíčová slova: Statistický software, systém SAS, projekt, projektové vyučování, interpretace.

Velmi přínosné využívání informačních prostředků je při práci v tzv. projektovém vyučování. Je to jedna z forem výuky založená na tom, že se v rámci vyučování řeší hlouběji a komplexněji jeden problém, jehož definování a způsob řešení se chápe jako projekt. Náměty projektu mohou vymyslet učitelé sami nebo je mohou navrhnout studenti. Je však důležité, aby téma bylo dostatečně přitažlivé. Vzhledem ke svému rozsahu, počtu použitých prostředků, šíří záběru a různorodostí forem stojí někde mezi řešením složitějších nebo školních úloh a skutečnou praxí, popř. alespoň modelováním skutečnosti. Od souboru jednotlivých nebo na sebe navazujících běžných školních úloh se projekt liší hlavně časovou náročností, motivací studentů, organizací práce, různorodostí nebo šíří témat pokrývajících projekt. Forma projektového vyučování nutí k větší samostatnosti studentů a odpovědnosti za výsledky práce, umožňuje využít individuálních znalostí, dovedností, zájmů, fantazie, schopností a zkušeností jednotlivých studentů. Oproti ostatním formám vyučování dává studentům relativní volnost spojenou s požadavkem na samostatnost a kolektivnost. Výsledky práce studentů není často jednoduché pedagogicky hodnotit.

4. Hodnocení získaných zkušeností

Na PEF ČZU je v současné době možné realizovat výuku statistického softwaru v bakalářských a magisterských studijních oborech ve dvou vědeckých statistických programech. Prvním je program Statistika, který je k dispozici v české lokalizované formě, druhým je SAS Systém, jež není lokalizován. Trh samozřejmě nabízí více programů a balíků obsahující statistické funkce, ale především z finančních důvodů jich nelze využívat větší počet. V rámci individuální výuky a sociologického výzkumu je na PEF ČZU zakoupeno i několik licencí programu SPSS.

Pro potřeby ekonomicko-manažersky zaměřených studijních oborů je do výuky statistiky zařazováno využití statistického softwaru jen informativně – formou jedné či dvou vyučovacích lekcí v programu Statistika. Děje se tak především díky jazykovému vybavení a možnosti intuitivního ovládání (prostředí je velmi podobné programu MS Excel, jež je studenty na PEF velmi často využíván k tvorbě odborných projektů). Pokud mají studenti těchto oborů zájem, pak si mohou v závěru studia zvolit přípravný Předdiplomní statistický seminář, v němž se mohou naučit více.

Studijní obory zaměřené na informační technologie, projektování IS a systémové inženýrství jsou ve výuce statistiky profilovány na využívání složitějšího systému SAS. Studenti těchto oborů mají po prvním ročníku studia absolvovány základní kurzy programování a využívání výpočetní techniky, čili mají k používání obtížnějšího softwaru větší předpoklady.

5. Dosavadní pojetí vyučování

Výuka práce v systému SAS se realizuje ve druhém semestru daného kurzu a je založena na tom, že studenti během prvního semestru (výuka statistiky je u všech studijních oborů povinná a je rozdělena do dvou semestrů, v druhém a třetím ročníku) získali znalosti o základech teorie pravděpodobnosti, popisné statistice, o metodách statistické indukce a o jednoduché regresní a korelační analýze a umějí vypočítat jednoduché úlohy pomocí kalkulačky.

Dosavadní způsob práce výuky systému SAS se uskutečňuje tak, že vyučující zmíní, názorně vysvětlí a předvede vybrané základní nabídky programu, a potom předvede práci s programem na zajímavém jednoduchém příkladu (jednotlivé kroky si student zapisuje do poznámek a zároveň je realizuje na svém počítači). Studenti pak uvedené postupy s jednoduchou modifikací procvičí na příkladě ze sbírky (většina úloh je na základě dat malých souborů). Při případných problémech při procvičování mohou studenti klást otázky a učitel působí v roli asistenta, který upřesňuje chyby v postupu a opravuje.

Ke konci semestru si student zvolí téma projektu a pracuje na něm ve svém volném čase. Jeden týden před termínem zkoušky musí být projekt odevzdán k hodnocení. Obhajoba projektu se uskutečňuje u ústní zkoušky.

Výhodou tohoto postupu práce je předem daná struktura cvičení, možnost psaní písemných opakovacích testů založených na ručním výpočtu (nutí studenty zopakovat si látku z předchozího semestru), dostatek času na volbu tématu projektu a na výběr metod a postupů v systému SAS.

K nevýhodám této pedagogické práce se řadí těžko zvládané ústní i písemné interpretace výsledků, dost často i špatného chápání zadání jednotlivých úloh ve sbírce (student se zaměřuje hlavně na to, aby zvládl výpočetní náročnost) a z toho plynoucí nedostatečně naformulované cíle a závěry zjištěné z použitých analýz v projektu. Pro učitele je těžké projekt zhodnotit a nalezené chyby při obhajobě práce nedávají dostatečný prostor k nápravě (pouze za cenu opakování zkoušky).

6. Zapojení prvků projektového vyučování

Z výše uvedených negativních důvodů bylo ve vybraných studijních skupinách změněno pojetí výuky více směrem k projektovému vyučování. Následující odstavce ilustrují jaké jsou postupy této práce a výhody a nevýhody tohoto nově zvoleného přístupu.

Výuka je koncipována tak, že projekt je stěžejním prvkem ve vyučování praktických cvičení (na přednáškách se přednášející věnuje složitějším postupům jako je regresní a korelační analýza a analýza časových řad, a dále pak i základům analýzy kvantitativních dat). První cvičení je založeno na praktickém seznámení studentů se zdroji dat na internetu. Během prvních dvou cvičení musí student předvést jaké údaje budou sloužit jako datová základna jeho projektu a nahlásit téma projektu. Podstatné je, aby si student vybrat oblast, která je mu blízká a rozuměl (alespoň částečně) metodice získání zvolených údajů (musí se tedy jednat o reálná data).

Přípravy na další cvičení vyučující založí také ve formě projektu, osvědčuje se rozsáhlejší datová základna ze skutečného statistického šetření (např. výsledky Agrocenzu 2000 od ČSÚ). Tak vyučující může vyložit i metodiku získání dat a studenti si lépe ujasní podstatu řešené úlohy.

Úvod každé hodiny je tvořen seznámením s řešeným tématem, důvod jeho zařazení do hodiny a diskusí o tom, kde lze tyto statistické postupy uplatnit. Každá hodina má tedy své téma, jehož řešení se studentům předvede na jim už známé datové základně. To předpokládá, že studenti mají databázi k dispozici (žádná data nepřepisují) a všechny kroky práce si sami zkoušejí. Vzhledem k tomu, že systém SAS nabízí několik alternativních postupů jak použít některé statistické metody, je dobré využít této možnosti. Je to totiž velice efektivní metoda pro nacvičování ústní interpretace výsledků (napoprvé výsledky interpretuje vyučující a s každým dalším alternativním krokem jsou už výsledky okomentovány studenty). Druhá polovina cvičení je věnována praktickému procvičení předvedených postupů, kdy podkladem jsou na začátku studentem vybrané údaje. Během procvičování působí učitel jako asistent a konzultuje se studenty řešení jejich projektu.

V první a druhé třetině semestru probíhá kontrola splnění použití předepsaných povinných statistických metod v projektu. Na konci semestru probíhá ústní obhajoba projektu, která je předmětem zápočtu. Pro získání zápočtu je nutné též napsat dvě písemné práce, které jsou zaměřené na písemnou interpretaci výstupů ze systému SAS (výstupy jsou získány použitím procvičených statistických metod a ze známých proměnných „semestrové databáze“). Opravené písemné práce pak slouží jako podklad ke správné interpretaci studentských projektů.

Mezi výhody tohoto vyučovacího přístupu se řadí užší spolupráce učitele se studentem (konzultuje se vhodnost dat, struktura projektu, interpretace jednotlivých postupů atd.) a znalost řešené problematiky ještě před obhajobou projektu. Další výhodou je, že ve výuce jsou používaná reálná data a tak se vyučující nevyhne tomu, aby použil i složitější prvky analýzy (různé transformace a úpravy), které student může také uplatnit ve své práci. Největší předností této vyučovací práce je prostor pro interpretace a komentáře výsledků, srozumitelné stanovení cíle práce a vyzdvihnutí podstatných a logicky zdůvodnitelných závěrů. Obhajoba projektů je většinou příjemným vyvrcholením společné práce učitele a studenta.

Mezi nevýhody pro uplatnění této práce se studenty lze zařadit nedostatečnou znalost metodiky sbírání dat (zejména u časových řad se metodika v různých letech mění a může být i zdrojem výkyvů, které se mohou jevit jako náhodné, ačkoli tomu tak ve skutečnosti není). Další „nevýhodou" je časová náročnost, neboť studenti jsou nuceni často konzultovat svůj projekt i mimo cvičení a vyžadují tedy dostatek konzultačních hodin. Osvědčuje se vypsat konzultační hodiny hned po každém takto pojatém cvičení (s ohledem na rozvrh studijní skupiny), což je někdy při předem daném rozvrhu nemožné. Čím větší je časový odstup konzultace od cvičení, tím méně mají studenti potřebu konzultovat a preciznost práce klesá. Též časové vytížení učitele pro čtení a opravování písemných komentářů je značné (nehledě na tvorbu příprav). Při práci ve velké studijní skupině efektivita této práce klesá a ztrácí smysl (maximum je tak 15 studentů).

7. Shrnutí a diskuse

Je již téměř jisté, že komunikační schopnosti studentů na odborné úrovni nejsou nijak excelentní, vždyť jen přečíst popřípadě naformulovat zadání úlohy je pro ně nelehkým úkolem. Je proto nutné, aby vyučující se studenty během vytváření projektů spolupracoval, sledoval způsoby volby tématu (téma by mělo být studentovi blízké), rozhodování ve výběru metod, studování podkladů (informací) a především logickou a věcnou interpretaci výsledků.

Jak je z výše uvedených řádek zřejmé, vždy lze vyučovací hodinu vhodně modifikovat a různým způsobem zatraktivnit, je to jen otázkou snahy a píle pedagoga. Pokud je však skupina studentů příliš velká nebo má cvičení v brzkých ranních či pozdních večerních hodinách, pak efektivita takto vynaložené práce přichází nazmar. Motivace studentů a dobrá práce pedagoga je jen jedním z mnoha faktorů, které ovlivňují kvalitu vyučování.

8. Závěr

V praktickém životě nezáleží jen na správně zvolené metodě statistické analýzy, ale i na jejím vyhodnocení a adekvátní interpretaci. Právě slovní vyjádření zjištěných výsledků bývá pro studenty největším oříškem k rozlousknutí (ačkoli se tato dovednost pro studium na vysoké škole tiše předpokládá). Proto by se vyučující měl při výuce zaměřit i na zvládnutí tohoto problému.

Zapojení projektového vyučování do výuky systému SAS je velkou šancí pro řešení některých nedostatků v kvalitě odevzdávaných studentských prací a i zvýšení přitažlivosti tohoto předmětu. Na druhou stranu to však vyžaduje vyšší zapojení pedagoga při asistenci v hodinách (i mimo ně) a při průběžném hodnocení práce na projektu. Nehledě pak i velkou časovou náročnost procvičování interpretací, které by mělo být jak písemné tak i ústní.

Literatura

[1] Černochová M., Komrska T. a Novák J. (1998) *Využití počítače při vyučování: Náměty pro práci dětí s počítačem*. Portál, Praha, ISBN 80-7178-272-6.

[2] Slavík J. a Novák J. (1997) *Počítač jako pomocník učitele: Efektivní práce s informacemi ve škole*. Portál, Praha, ISBN 80-7178-149-5.

ŠPECIFIKUM VÝUČBY ŠTATISTIKY ŠTUDENTOV SOCIÁLNYCH VIED

Ľudmila Gregušová

Katedra manažmentu a informatiky, Akadémia Policajného zboru v Bratislave, Sklabinská 1, 835 17 Bratislava
gregusov@minv.sk

Nutnosť a potreba tvorivého využívania štatistických metód neustále rastie nielen v hospodárskej praxi a prírodných vedách, ale najmä pri riešení problémov spoločnosti vo vedách sociálnych. Príspevok poukazuje na problémy výučby štatistiky pre „neštatistikov“, za akých možno považovať študentov sociálnych vied. Autorka na základe viacročnej praxe sa zaoberá niektorými možnosťami skvalitnenia a zefektívnenia výučby štatistiky s využitím Excelu.

1. Úvod

Skúsenosti pedagógov vyučujúcich štatistiku všeobecne ukazujú, že stredné školy nepripravia svojich absolventov dostatočným matematickým aparátom, potrebným ku štúdiu štatistiky. Študenti našej vysokej školy tu veru nie sú žiadnou svetlou výnimkou. Preto musíme vychádzať z potrieb minimalizovania matematického aparátu štatistiky a väčšieho priblíženia štatistických metód študentom využitím výpočtovej techniky. Zavádzanie nových technológií do vzdelávania však nenapreduje tak rýchlo, ako by sme chceli či potrebovali. Problém nesúvisí len s materiálnou stránkou veci, hoci ceny hardvéru a príslušného softvéru tiež nie sú zanedbateľné. Jeho jadro je predovšetkým v nutnosti prestavby a celkovej regulácie pedagogických procesov. Treba mať na mysli cieľ vzdelávania a tento dosiahnuť, už nie však po starom, ale postupným inovovaním a skvalitňovaním ako obsahovej stránky predmetu, tak i metodického prístupu k výučbe.

2. Metodológia výučby štatistiky

Preto v záujme ďalšieho skvalitňovania vysokoškolskej i vedeckej prípravy policajných manažérov je nielen vhodné učiť štatistiku s použitím počítačov, ale je nutné nadobudnúť aj potrebné praktické skúseností, podnety k inovácii, k ďalšiemu rozvoju výučby štatistiky a jej pedagogických metód. Jednou z nich je aj projektové vyučovanie, umožňujúce spojenie potrebnej štatistickej teórie a praktickej činnosti, ktoré sú nutné k štatistickému spracovaniu vedeckých problémov. Preto je potrebné viesť

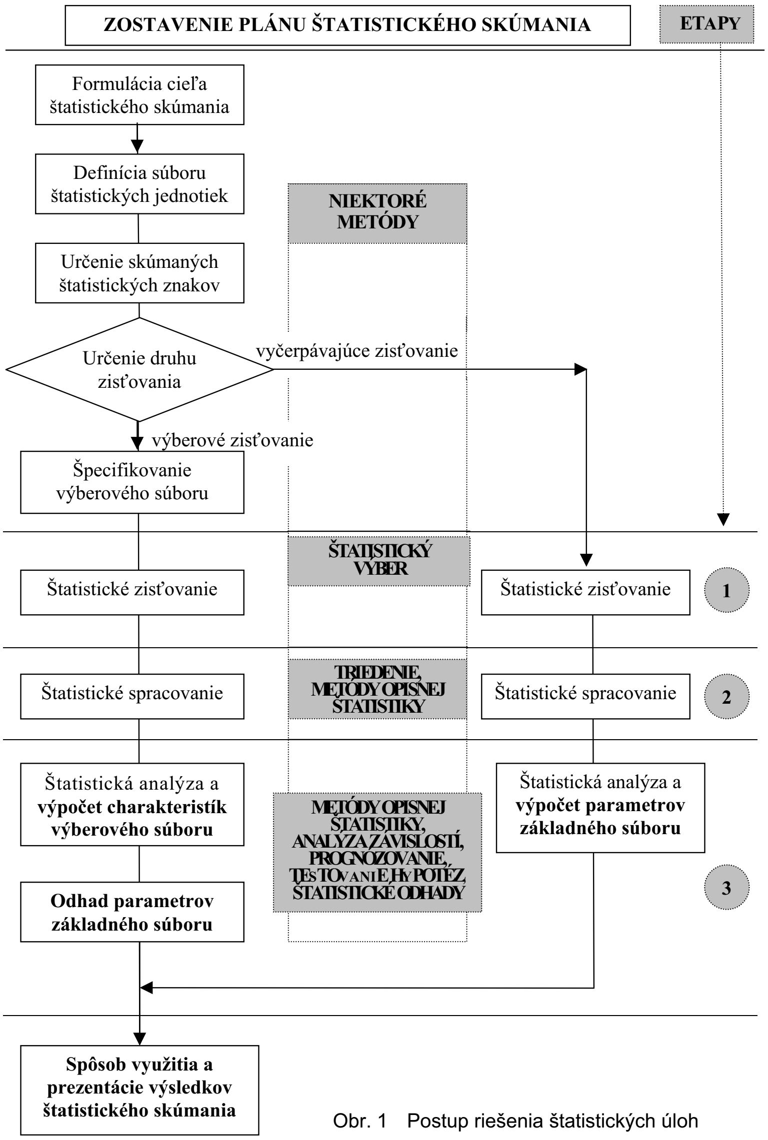
Klíčové slová: Projektové vyučovanie, počítace vo výučbe, statistické metódy, výpočtová štatistika.

študentov k praktickým aplikáciám výpočtovej štatistiky, umožňujúcim analýzu a odhalenie zákonitostí skúmaných hromadných javov vo forme projektov.

Uplatnenie novej organizácie výučby a nových vyučovacích metód vplýva nielen na vzťah jedinca k učebnému predmetu, na jeho motiváciu či aktivizáciu, ale hlavne na rozvoj samostatného tvorivého i logické myslenia študentov. Nejde iba o jednoduché rozšírenie obzoru študentov a prezentáciu štatistického aparátu, ale o to, ako priviesť študentov „nematematikov“ k analytickej práci, ktorá je dôležitou súčasťou pri riadení a rozhodovaní. A keď používam termín „nematematikov“, mám na mysli práve takých študentov, ktorých matematika nikdy veľmi nelákala. Prišla som k názoru, že tu pomôže jedine moderná výpočtová štatistika. Práve ona poskytuje pomerne dostatočné množstvo prostriedkov, ktoré nám umožňujú využiť výsledky štatistických analýz pri riadení a rozhodovaní pomerne „bezbolestne pre nematematikov“. Na využitie takýchto nástrojov je potrebné vedieť:

- Formulovať problém.
- Vybrať vhodnú metódu riešenia.
- Overiť vhodnosť, či predpoklady použitia zvolenej metódy.
- Zadať zoradené vstupné údaje.
- Interpretovať výstupné výsledky podľa použitej metódy.
- Prezentovať riešenie vo vhodnej forme.

Ako príklad si dovolím uviesť jeden z výborných projektov, týkajúci sa ankety k názorom na trest smrti. Avšak každé minimalizovanie matematického aparátu podmieňuje maximalizovanie pozornosti pedagóga a potrebu konzultovať hore uvedenú postupnosť krokov, čo vedie k pomerne úzkej individuálnej spolupráci pedagóga i študenta. A sme pri koreni problému. Odkedy pôsobím na APZ, a už je to desať rokov, snažím sa o to, aby predmet Základy štatistiky splnil svoj cieľ. Tým cieľom je utvoriť u študentov ucelený súbor vedomostí, zručností a schopností v získavaní a využívaní štatistických informácií a metód pri riadení a rozhodovaní v bezpečnostných službách. Avšak bez využívania výpočtovej techniky sme vždy zostávali len kdesi na polceste. Pri objeme 30 vyučovacích hodín (1/1) v jedinom semestri to ani inak nešlo. Pozitívne zmeny prichádzali pomaly a skutočne nastali až pred dvoma rokmi. Predstavujú ich nasledujúce momenty vo vývoji predmetu:



Obr. 1 Postup riešenia štatistických úloh

- Presun z 1. a 2. ročníka do vyšších ročníkov.
- Predmet sa stal povinne voliteľným.
- Projektové vyučovanie.
- Zavedenie Excelu do výučby.
- Využitie informačných systémov.

Voliteľnosť predmetu otvorila cestu k individuálnemu prístupu pri výučbe tých študentov, ktorých neodradila matematika. To, že predmet je vo vyšších ročníkoch znamená, že mnohé projekty sa kryjú s problematikou diplomových prác. Študenti sú nielen vyzretejší, ale majú zmysel pre samostatné spracovanie reálnych údajov pri riešení projektov. Zavedenie Excelu do výučby bolo po zrelej úvahe všetkého za i proti. Excel je široko dostupný, známy a považovaný za súčasť počítačovej gramotnosti. Existuje v národnej (príp. českej verzii) a iteračným spôsobom možno meniť a nastaviť voľby v dialógových oknách a tak presne dosiahnuť požadovanú presnosť, úpravu či zobrazenie. Samozrejme, že treba mať na mysli aj určité nevýhody Excelu, keďže nie je špecializovaným štatistickým softvérom a teda nemá všetky štatistické nástroje (o chybách lokalizácie radšej nehovoriac). Avšak v záujme ďalšieho skvalitňovania vysokoškolskej aj vedeckej prípravy policajných manažérov je uvedený aparát rozhodne namieste. (Jednak nevyžaduje znalosť angličtiny, zvláštnej obsluhy ba ani nové finančné investície.) Navyše umožňuje využívanie pracovných listov. Tieto vychádzajú z reálnych údajov, prezentujú jednotlivé vzorové riešenia (na záver), vhodne motivujú a dostatočne ilustrujú riešené úlohy. Prv než sa pozrieme na možnosti využívania informačných systémov, pozrime sa na štatistické metódy, využívané v procese riešenia štatistických úloh. Na ObrázkU 1 vidíme jednotlivé etapy štatistického skúmania s príslušnými metódami.

3. Štatistické skúmanie v Policajnom zbore

Vzhľadom k tomu, že v rezorte Ministerstva vnútra existuje vlastné štatistické skúmanie tzv. Evidenčno-štatistický systém kriminality (EŠSK) so základnými štatistickými jednotkami:

- Trestný čin.
- Známy páchateľ.

Mnohí študenti sa a priori (ale mylne) domnievali, že obsahom predmetu budú informácie o štandardných štatistických výstupoch a zostavách EŠSK. (Obdobne boli ochotní akceptovať informácie o ŠEDN - Štatistickej evidencii dopravných nehôd). Keďže tieto informácie už mali na stredných policajných školách, mysleli si, že majú „vystaráno“. Až pomocou Obrázku1 pochopili, čo znamená aplikovať štatistické metódy, čo možno pomocou štatistiky odhadnúť, a s akou „presnosťou“. Aby nám však štatistika pomáhala pri manažérskom rozhodovaní, je potrebné ešte čosi navyše. Vedieť si poradiť so skutočnými problémami, riešenými vhodnými štatistickými metódami, a vedieť si medzi tými metódami vybrať, ako už bolo spomínané pri využívaní nástrojov výpočtovej štatistiky.

4. Možnosti vedeckého zhodnotenia informačného potenciálu rezortu

Ak spomíname súčasné možnosti v EŠSK, či v ŠEDN, pričom využívame štandardné alebo neštandardné štatistické výstupy, predsa sme len do istej miery viazaní na to, čo oba systémy sledujú a čo sledovať schopné sú. Buduje sa však nový Automatizovaný informačný systém polície (AISP), ktorý je navrhnutý ako systém integrovaného prístupu k informačným systémom. Súčasný stav budovania AISP dovoľuje odpovedať na niektoré závažné otázky sledovania kriminality. Napr.: Aké bude EŠSK v AISP? Aké budú nové štatistické jednotky? Ako sa zmení vyhľadávanie štatistických informácií v AISP? AISP umožní komplexné využívanie registrovaných údajov, čo zvýši kvalitu podkladov pre riadiacu, rozhodovaciu a analytickú činnosť. Ovplyvní tiež možnosti štatistických analýz a výskumu. AISP bude poskytovať nové štatistické jednotky a znaky ako napr.: Udalosť, Skutok, Osoba, Vec, Organizácia, Skupina, Proces, Spis, Objekt a pod. Nové podmienky výskumu spočívajú v tom, že kriminalita už nemusí byť vyhodnocovaná len s prihliadnutím na osobu páchateľa a jeho charakteristiky, ale aj s prihliadnutím na osobu poškodeného, prípadne oznamovateľa a pod. U poškodeného je možné sledovať aj viktimologické vlastnosti a tieto štatisticky vyhodnocovať vo vzťahu k páchanej trestnej činnosti. Keďže AISP je objektovo-orientovaný systém, každý objekt v ňom bude definovaný svojimi atribútmi a taktiež prípadnými vzťahmi k iným objektom systému. AISP nezhromažďuje predpripravené štatistické informácie, ako je tomu napríklad v EŠSK alebo v ŠEDN. V AISP je štatistika vykonávaná nad údajmi zhromaždenými za účelom informačných služieb, ktoré systém poskytuje. Čo však z uvedeného vyplýva pre pedagóga?

Ukazuje sa, že nastala akási nová éra podmienok na využitie štatistiky v policajnej praxi. Keďže doterajšie štatistické výstupy predstavovali spracovanie známych základných štatistických súborov, o ktorých sa dali získať všetky dostupné informácie použitím metód opisnej štatistiky, čo nevyžadovalo (a ani to nebolo možné), robiť pravdepodobnostné interpretácie záverov. Nové štatistické jednotky so svojou údajovou základňou o jednotlivých štatistických znakoch umožňujú výrazné rozšírenie štatistických výstupov s možnosťami vyhľadávania objektov. Tým sa ponúka možnosť budovania si vlastnej údajovej základne pre výskumné účely. Vyhľadávanie podľa atribútov rýchleho výberu umožní tvorbu vlastných štatistických súborov.

Pri splnení podmienok náhodného výberu a skúmaním takto vzniknutých výberových súborov bude možné v rámci štatistickej inferencie robiť závery o neznámych základných súboroch (napr. súbor vlastníkov neregistrovaných zbraní, alebo súbor drogovo-závislých dosiaľ netrestaných páchateľov a pod.) a to s dopredu volenou spoľahlivosťou. Na to všetko však treba študentov náležite pripraviť.

5. Tematický plán výučby

Tematický plán výučby predmetu, pomocou ktorého sme sa snažili pripravovať našich študentov vhodným pragmatickým spôsobom, ukazuje nasledujúca Tabuľka 1.

č.	Názov témy	P	S	Cv	PZ	Spolu
1.	Úvod do štatistiky	2	2			4
2.	Teória pravdepodobnosti a náhodné premenné	2			2	4
3.1	Štatistický výber a triedenie	2				2
3.2	Triedenie pomocou tabuľkového procesora EXCEL, kontingenčné tabuľky a grafy				2	2
4.	Charakteristiky štatistických súborov a opisná štatistika v EXCELI	2			2	4
5.	Základy korelačnej a regresnej analýzy, štatistické závislosti a grafy v EXCELI	2			2	4
6.	Dynamika štatistickej štruktúry a prognózovanie	2			2	4
7.1	Využitie štatistických metód v policajnej praxi, stav a perspektíva	2				2
7.2	Spracovanie výskumu a využitie štatistických programov v policajnej praxi				2	2
8.	Súhrnné zhrnutie, zápočtová písomka			2		2
	Spolu	14	2	2	12	30

Tabuľka 1. Tematický plán výučby predmetu

6. Záver

Za desať rokov prešiel predmet Základy štatistiky určitým vývojom s pozitívnymi výsledkami, pričom sa ukazuje, čo asi možno vo výučbe štatistiky v blízkej i vzdialenej budúcnosti očakávať:

- Nutnosť tvorivého využívania štatistických metód bude aj v budúcnosti naďalej rásť a nebude to bez počítačov a príslušného softvéru.
- Taktiež vzrastie potreba efektívneho využitia informačno - komunikačných technológií, začínajúc počítačovými sieťami, a nekončiac informačnými systémami.
- Potrebné zmeny neprinesie len výučba a výchova študentov a doktorandov, ale v neposlednom rade tiež príprava pedagógov a zavedenie nových vyučovacích metód.

7. Zhrnutie

Cieľom príspevku bolo:

1. Predstaviť možnosti výučby, využitia a aplikovania štatistiky, rešpektujúc konkrétnu úroveň nadobudnutých znalostí vyššej matematiky.
2. Naznačiť vývojové trendy, ktoré obohacujú možnosti štatistiky a poskytujú prostriedky na využitie štatistických analýz pri riadení a rozhodovaní.
3. Ukázať vhodné spojenie možností štatistickej teórie s informačným potenciálom rezortu MV.

Literatúra

[1] Antoch J. (2002) *Vybraná kritéria pro volbu programového vybavení pro (statistickou) analýzu dat*. In: *Výuka statistiky v České Republice I*, Sborník prací semináře 29.11.2002 v Praze, Matfyzpress, 8–12.

[2] Bartošová J. (2002) *Příspěvek k výuce statistických předmětů na FM VŠE*. In: *Výuka statistiky v České Republice I*, Sborník prací semináře 29.11.2002 v Praze, Matfyzpress, 8–12.

[3] Gregušová Ľ. (2002) *Projektové vyučovanie štatistiky na počítačoch*. In: *Zborník referátov medzinárodnej konferencie Aplimat, 7.–8. februára 2002*. SjF STU, Bratislava, 187–192.

[4] Gregušová Ľ. (2002) *Štatistika v policajnej praxi*. APZ, Bratislava.

[5] Chajdiak J. (2002) *Štatistika v Exceli*. STATIS, Bratislava.

[6] Kubanová J. a Linda B.: *Jak učit statistiku s MS Excel*. In.: *Zborník konferencie Matematická štatistika a numerická matematika a ich aplikácie, Kočovce 14.–18. 6. 1999*. SvF STU 1999, Bratislava, 147–151.

[7] Stankovičová I. a Ďurčíková A. (1999) *Projektové vyučovanie predmetu základy štatistiky na FM UK*. In: *Zborník konferencie Matematická štatistika a numerická matematika a ich aplikácie*. STU SF, Bratislava, 162–165.

[8] Tvrdík J. (1998) *Excel, statistika, lokalizace a zmatek*. Informační Bulletin České statistické společnosti **9(2)**, 13–20.

STATISTIKA NA AMERICKÝCH STŘEDNÍCH ŠKOLÁCH

Jan Hendl

FTVS UK, Katedra kinantropologie a humanitních věd,
José Martího 31, 162 52 Praha 6
hendl@ftvs.cuni.cz

1. Úvod

V roce 2000 přibližně třičtvrtě milionů studentů v USA využilo možnost navštěvovat na střední škole (high school) kurz Avanced Placement (AP) v 35 různých oborech. Tyto studijní programy jsou podporovány Úřadem pro college (College Board), který iniciuje návrh programů a přípravu učitelů. Ročně projde kolem 60 000 učitelů doškolovacími kurzy pro jednotlivé obory. AP kurzy jsou výsledkem spolupráce mezi středními školami a colleges, tedy školami, které jsou mezistupněm mezi střední školou a universitami. AP kurzy zajišťují, že student dostane příležitost prohloubit si středoškolské vědomosti ve směru svých zájmů nebo předpokládaného uplatnění. AP statistika je novějším přírůstkem do palety takových kurzů. První zkoušky z tohoto kurzu absolvovali studenti v květnu 1997. Za prvních pět let existence kurzu se zúčastnilo zkoušek z AP statistiky celkem 124 120 studentů.

V našem příspěvku objasníme filosofii AP statistiky a obsah kurikula, a uvedeme příklady úloh, které studenti řeší při závěrečném přezkoušení.

AP statistika je ekvivalentní jednosemestrálnímu úvodu do statistické analýzy na vysoké škole a je hodnocena příslušným kreditem na škole typu college. Na většině středních škol se látka AP statistiky probírá celý rok. Předpokladem pro kurz je zkouška z algebry. Filosofie kurzu se opírá o doporučení NCTM standardů (NCTM 2000) pro výuku matematiky a obsahové linie pro statistiku. Zdůrazňuje se využití elektronických kalkulátorů se statistickými funkcemi a grafikou. Mezi nejvíce používané pomůcky patří učebnice Davida Moora (viz literatura).

Účelem AP statistiky je uvést studenty do hlavních konceptů a metod sběru, analýzy dat a statistického usuzování, aniž by se využívaly koncepty vyšší matematiky. Se studenty se probírají čtyři tématické okruhy:

1. Zkoumání dat: zkoumání konfigurací a odchylek od konfigurací.
2. Plánování studie: rozhodování co a jak měřit (získávání dat).

Klíčová slova: Výuka statistiky na amerických středních školách, kurz AP (Avanced Placement).
Poděkování: Rád bych poděkoval řadě kolegů za podnětné, byť ne vždy souhlasné diskuse.

3. Anticipace konfigurací: konstruování a zkoumání modelů pomocí pravděpodobností a simulací.
4. Statistická inference: testy a odhady, práce s modely.

Prioritou je učení na příkladech a projektech a schopnost komunikace. Výuka se řídí zásadami, které zdůrazňují:

- Zvládnutí spíše základních konceptů než vštípení vzorců.
- Umění psaní zpráv.
- Využívání technologií.
- Projektovou výuku.

Studenti při výuce analyzují data pomocí elektronických kalkulaček se statistickými a grafickými funkcemi nebo na počítači, provádějí experimenty ve třídě, realizují individuální nebo skupinové projekty, nebo na počítači simulují různé pravděpodobnostní modely. V kurzech se setkávají studenti s nejrůznějšími zájmy, mnozí z nich nemají zájem o běžnou matematiku nebo nejsou v matematice nijak silní.

2. Přehled hlavních témat v kurzech AP statistiky

Uvádíme hlavní témata, ze kterých jsou studenti přezkušovány na konci AP statistiky v květnu každého roku. U jednotlivých témat popisujeme vybrané metody nebo koncepty, kterým se má výuka zvláště věnovat. Upořádání témat je určeno cílem představit obsah kurzu, ale neurčuje jejich pořadí, jak se na škole probírají.

2.1. Zkoumání dat

A. Interpretace grafického zobrazení rozdělení jednorozměrných (krabicový graf, graf lodyhy a listu, histogram, kumulativní četnosti, graf trendu)

1. Centrální tendence a rozptýlenost.
2. Shluky a mezery.
3. Odlehlé hodnoty a další neobvyklé vlastnosti.
4. Tvar rozdělení.
5. *Hledání vyhlazené křivky rozdělení (fakultativní).

Rozpracování témat:

– Měření a rozptýlenost hodnot, grafy lodyhy a listu, krabicové grafy, histogramy, grafy trendu, zjištění centrálních hodnot a charakteristik rozptylu, křivky hustoty, normální rozdělení a výpočty hustoty, hodnocení „normality“.

– Sběr dat a zkoumání jejich struktury: rovnoměrné rozdělení, zvonovité rozdělení, bimodální rozdělení a určení nejlepší metody grafické prezentace dat.
– Určování odchylek od základního tvaru a hodnocení citlivosti interpretace k odstranění vybraných bodů (hodnot).
– Porozumění povaze asociace versus příčinnost při statistické interpretaci.

B. Shrnutí rozdělení jednorozměrných dat

1. Měření centrální tendence: medián, průměr.
2. Měření rozptýlenosti: rozpětí, interkvartilové rozpětí, směrodatná odchylka, rozptyl.
3. Měření polohy: kvartily, percentily, z-skóry.
4. Využití krabicového grafu.
5. Efekt změny jednotek na souhrné statistiky.

Rozpracování témat:

– Porozumění náhodnému chování.
– Používání počítače nebo kalkulačky při zobrazování rozdělení a určování 5ti-bodovém shrnutí, uspořádání dat a vytvoření histogramů.
– Určení relativních četností, výpočet rozptylu a standardní odchylky z definice a pomocí výpočetní techniky.
– Znalost vlastností směrodatné odchylky.
– Určení vztahu mezi mediánem a průměrem ve vztahu k hustotě rozdělení s různými tvary.
– Definice normálního rozdělení a zkoumání jeho tvaru s různou hodnotu parametru průměru a rozptylu.
– Výpočet plochy po křivkou pomocí výpočetní technologie a určení vztahu k z-skórům.
– 68 – 95 – 99,7 pravidlo pro normální rozdělení.
– Definice rozdílností mezi spojitým a diskrétním rozdělením ve vztahu k funkci rozdělení.

C. Porovnání rozdělení jednorozměrných dat (krabicové grafy, zrcadlově uspořádané grafy lodyhy a listu)

1. Porovnání centrální polohy a rozptýlenosti: rozptyl uvnitř skupin a mezi skupinami.
2. Porovnání shluků a mezer v datech.
3. Porovnání odlehlých hodnot a neočekávaných vlastností.
4. Porovnání tvarů.

Rozpracování témat:

- Porovnání množin dat použitím různých grafických metod a zkoumání mezer a shluků s cílem určit možné asociace.
- Porozumění efektu odstranění odlehlých hodnot a posouzení potřebnosti takové akce.

D. Zkoumání dvojrozměrných dat

1. Analyzování konfigurací v bodovém grafu.
2. Korelace a linearita.
3. Přímka zjištěna metodou nejmenších čtverců.
4. Reziduální graf, odlehlé hodnoty a vlivné body.
5. Transformace k dosažení linearity: logaritmování a mocninné transformace.

Rozpracování témat:

- U dvojrozměrných dat porozumění tomu, že asociace neimplikuje příčinnost.
- Pochopení vztahu mezi korelačním koeficientem a koeficientem determinance; výpočet r pomocí vzorce a počítačové technologie.
- Interpretace korelace a regrese.
- Výpočet reziduálních hodnot a grafů s cílem posoudit asociace v datech a odchylky od základní předpokládané konfigurace.
- Prokládání dat v bodovém grafu.
- Porozumění tomu, že nulová hodnota korelačního koeficientu nemusí znamenat absenci asociace v datech.
- Zkoumání logaritmické a mocninné transformace dat a jejich vliv na vztah.
- Zkoumání regresní přímky jež prochází počátkem (speciální případ: $y = bx$).

E. Explorace kategoriálních dat

1. Marginální a bodové četnosti, asociace.
2. Podmíněné relativní četnosti a asociace.

Rozpracování témat:

- Kategoriální data: analýza dvojrozměrných tabulek, Simpsonův paradox.

2.2. Plánování studie

A. Přehled metod sběru dat

1. Census.
2. Výběrové šetření.
3. Experiment.
4. Observační studie.

Rozpracování témat:

– Sběr dat z dobře navrženého plánu pro demonstraci validních informací k posouzení hypotéz.

B. Plánování a realizace šetření

1. Charakteristiky dobře navrženého a provedeného šetření.
2. Populace, výběr a náhodný výběr.
3. Zdroje chyb v šetření.
4. Jednoduchý náhodný výběr.
5. Stratifikovaný náhodný výběr.

Rozpracování témat:

– Porozumění povaze parametru populace a statistice jako jeho odhadu, získaného pomocí výběru a používání správného značení obou charakteristik.
– Procvičování vhodného plánu: Určení zkoumané otázky, proměnných, které se budou měřit, aby bylo možné posoudit zkoumané otázky a určení metod sběru dat.

C. Plánování a provedení experimentu

1. Charakteristiky dobře navrženého a provedeného experimentu.
2. Ošetření, kontrolní skupina, experimentální jednotka, náhodné přiřazení a replikace.
3. Zdroje systematické chyby a rušivého efektu, včetně efektu placeba a zaslepení.
4. Úplně znáhodněný plán.
5. Znáhodněné bloky, včetně párového plánu.

Rozpracování témat:

– Návrh experimentů: komparativní experimenty, znáhodnění, jak náhodně přiřadit do skupiny, plán výběru, systematická chyba, variabilita, výběrové rozdělení, observační studie versus experimentální studie.
– Zvládnutí slovníku experimentálního plánu: jednoduchý náhodný výběr, dobrovolná odpověď, nepokrytí populace, nedodání odpovědi, systematické chyba v odpovědi, stratifikovaný náhodný výběr, bloková plán, populace, výběr, placebo efekt, zaslepení, párované dvojice, atd.

D. Zobecnění výsledků z observačních studií, experimentů a šetření

2.3. Anticipování konfigurací

A. Pravděpodobnost jako relativní četnost

1. „Zákon velkých čísel".
2. Pravidla pro sčítání a násobení pravděpodobností a nezávislost.
3. Diskrétní náhodný proměnné a jejich pravděpodobnostní rozdělení, včetně binomického a geometrického rozdělení.
4. Simulace of pravděpodobnostního rozdělení, včetně binomického a geometrického rozdělení.
5. Průměr (očekávaná hodnota) a směrodatná odchylka náhodné proměnné a lineární transformace náhodné proměnné.

Rozpracování témat:

– Přiřazení pravděpodobností a pojem výběrového prostoru.
– Používání výpočetní technologie pro simulaci pravděpodobnostních rozdělení jako: binomické nebo geometrické rozdělení.
– Základní principy počtu pravděpodobnost i: komplementární události, spojení a průnik jevů a pravidla počtu pravděpodobnosti, podmíněná pravděpodobnost, jaké hodnoty mohou mít pravděpodobnosti. Bayesova formule.
– Důraz na rozpoznání rozdílu mezi nezávislými jevy a jevy s nulovým průnikem a metody výpočtu pomocí stromového grafu.
– Použití tabulek a výpočetní technologie pro určení percentilů normálního rozdělení, použití technik simulace pro modelování výběrových rozdělení.

B. Kombinace nezávislých náhodných proměnných

1. Koncepty nezávislosti versus závislosti.
2. Používání tabulek normálního rozdělení.
3. Normální rozdělení jako model měření.

C. Výběrové rozdělení

1. Výběrové rozdělení výběrových relativních četností.
2. Výběrové rozdělení výběrového průměru.
3. Centrální limitní teorém.
4. Výběrové rozdělení rozdílu dvou nezávislých relativních četností.
5. Výběrové rozdělení rozdílu dvou nezávislých výběrových průměrů.
6. Simulace výběrových rozdělení.

Rozpracování témat:

- *Normální aproximace binomického rozdělení s korekcí na spojitost, centrální limitní teorém, lineární kombinace nezávislých normálně rozdělených proměnných.
- Binomické rozdělení, výpočet pravděpodobností, průměr a rozptyl.
- Normální aproximace pro binomické rozdělení a relativní četnosti a četnosti.
- Rozdělení výběrových průměrů a centrální limitní teorém.
- Kontrolní grafy.

2.4. Statistické usuzování

A. Konfidenční intervaly

1. Koncept konfidenčního intervalu.
2. Konfidenční interval pro relativní četnost při velkém výběru.
3. Konfidenční interval pro průměr při velkém výběru.
4. Konfidenční interval pro rozdíl dvou relativních četností při velkých výběrech.
5. Konfidenční interval pro rozdíl dvou průměrů při velkých výběrech (nezávislé, závislé).

Rozpracování témat:

- Konfidenční intervaly a jejich chování, používání simulací pro modelování chování intervalů spolehlivosti a posuzování jejich chování při různých datových konfiguracích.

B. Statistické testy významnosti

1. Logika testu významnosti, nulová a alternativní hypotéza, p-hodnoty, jednostranný a dvoustranný test, koncept chyby I. a II. druhu, koncept síly testu.
2. Test pro relativní četnost při velkém výběru.
3. Test pro průměr při velkém výběru.
4. Test shody dvou relativních četností při velkých výběrech.
5. Test shody dvou průměrů při velkých výběrech (nezávislé, závislé).
6. Chi kvadrát test dobré shody, homogenity relativních četností a nezávislosti v kontingenční tabulce.

Rozpracování témat:

– Testy významnosti, chyba I. A II. druhu, síla testu, zneužívání statistického testu.
– Důraz na předpoklady testovací procedury včetně non-normálních populací.
– Procvičování formulace zprávy o provedení testu.
– Volba rozsahu výběru.
– Velikost chyby.
– Dvourozměrná tabulka a testování pomocí chi-kvadrát testu.

C. Speciální případy normálně rozdělených dat

1. t-rozdělení.
2. t-test pro jeden výběr.
3. t-testy pro dva výběry (nezávislé, závislé).
4. Inference pro směrnici regresní přímky.

Rozpracování témat:

– Párový t-test.
– t-testy pro dva nezávislé výběry se spojenými a nespojenými rozptyly.
– F-test a jeho robustnost vzhledem k poruše předpokladů.
– Inference pro regresi.
– Jednoduchý model lineární regrese, odhad regresních parametrů, konfidenční intervaly pro odhad přímky a predikční interval.
– *Volitelná témata*: analýza rozptylu pro regresní model a statistické usuzování pro korelaci a mnohorozměrnou regresi, model ANOVA, ANOVA tabulka a její vztah k F-testu a t-testu, dvoufaktoriální ANOVA.

3. Realizace výzkumného projektu

Cílem AP kurzu je rozvinout schopnosti studentů provádět vlastní výzkum. Proto mají během studia realizovat vlastní projekty řešení reálného problému. Data musejí být získána podle dobře navrženého plánu, aby se zajistila validita závěrů. Výzkumný plán by měl ukázat, že student umí zvolit vhodné metody sběru dat, aby bylo možné uplatnit metody statistického usuzování. Student má demonstrovat, že zvládl příslušné koncepty kurikula AP statistiky. Projekty se mají zadávat během celého roku a různou formou, od krátkých prací až po delší podrobnější projekty.

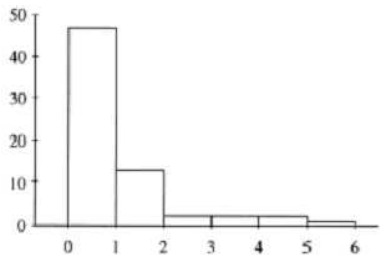
4. Přezkoušení

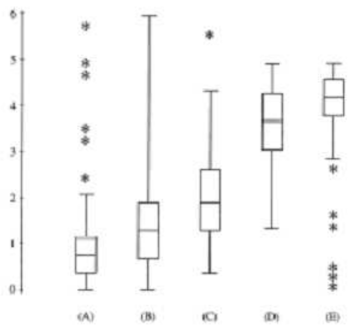
Pro každý AP kurz škola, která poskytuje AP kurzy a zavázala se dodržovat podmínky AP kurzů, obdrží z centra úlohy pro závěrečné zkoušky, které se uskutečňují ve společném termínu. Zkouška trvá přibližně 3 hodiny a skládá se ze dvou částí:

- Část 1: 40 otázek s volitelnou odpovědí (50% z celkového počtu bodů).
- Část 2: Otázky s volnou odpovědí (50% z celkového počtu bodů):
  – 5 krátkých otázek (75% druhé části);
  – 1 podrobnější studie (25% druhé části).

Příklady v závěrečné zkoušce pro část 1: Celkový skór z přezkoušení se upravuje na náhodné uhádnutí. Student má zvolit jednu z pěti možných odpovědí.

1. Histogram na obrázku ukazuje množinu hodnot měření. Který box graf ukazuje stejnou množinu hodnot?



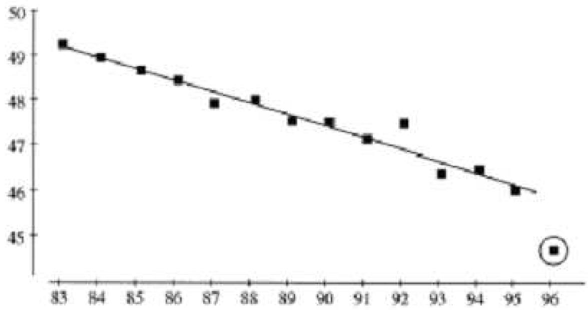


2. Z populace se náhodně vybral a množina o rozsahu 10. Výběr má rozptyl nula. Jaké z dále uvedených tvrzení musí platit?
   I. Populace má rozptyl roven nule.
   II. Výběrový průměr je roven výběrovému mediánu.
   III. Všech 10 hodnot má stejnou numerickou hodnotu.
   (A) I pouze
   (B) II pouze
   (C) III pouze
   (D) I a II
   (E) II a III

3. Učitel vyučuje statistiku ve dvou třídách. Ranní třída má 25 studentů a jejich průměr z výsledků z prvního testu byl 82. Odpolední třída má 15 studentů a jejich průměr byl 74. Jaký je průměr výsledků, jestliže učitel zkombinuje výsledky z obou tříd?

   (a) 76
   (b) 78
   (c) 79
   (d) 80
   (e) Průměr nelze spočítat, protože nejsou k dispozici individuální výsledky pro studenty.

4. Na obrázku je vidět, že byla vypočtena metodou nejmenších čtverců přímka hodnotami procent získaných zápasů sportovního týmu škol v letech 1983 až 1995.



Kroužkem je označeno procento získaných zápasů v roce 1996. Které z následujících tvrzení správně popisuje, jak hodnota z roku 1996 změní korelační koeficient a polohu přímky získané metodou nejmenších čtverců, jestliže se znovu vypočítají pro všechny boy včetně roku 1996.

(a) Bod z roku 1996 zvětší sklon přímky a zesílí korelační koeficient.
(b) Bod z roku 1996 zvětší sklon přímky a zeslabí korelační koeficient.
(c) Bod z roku 1996 přiblíží přímku k horizontále a zesílí korelační koeficient.
(d) Bod z roku 1996 přiblíží přímku k horizontále a zeslabí korelační koeficient.
(e) Bod z roku 1996 nebude mít žádný efekt na sklon přímky, protože sleduje také klesající trend.

5. Tabulka ukazuje část pravděpodobnostního rozdělení pro náhodné proměnné X a Y. Jestliže X a Y jsou nezávislé a pravděpodobnost jevu $P(X = 3, Y = 4) = 1/16$.

x	$P(X = x)$	y	$P(Y = y)$
1	1/6	1	?
2	2/3	2	1
3	?	3	1
–	–	4	?

Jaká je pravděpodobnost $P(Y = 1)$?

(a) $1/8$
(b) $1/6$
(c) $1/4$
(d) $3/8$
(e) $1/2$

6. Náhodný výběr dvou pozorování se vybral z populace, která je normálně rozdělená s průměrem 100 a směrodatnou odchylkou 5. Která u uvedených pravděpodobností nejlépe aproximuje pravděpodobnost, že součet obou pozorování je větší než 221?

(a) $0,0015$
(b) $0,0250$
(c) $0,0500$
(d) $0,4500$
(e) $0,9985$

Příklady otevřených otázek v závěrečné zkoušce pro část 2:

1. Záchranná služba v Coloradu, chce poznat chování turistů, když se ztratí, aby navrhla lepší strategie pro jejich vyhledání. 200 turistů, náhodně vybraných z těch, kteří si zažádali o vstupenku, odpověděli na otázku, zda půjdou raději

dolů, nahoru z kopce nebo zůstanou na místě, jestliže se ztratí. Každý z nich byl klasifikován, zda je zkušený nebo méně zkušený.

	Směr nahoru	Směr dolů	Zůstane na místě
Nezkušený	20	50	50
Zkušený	10	30	40

2. Firma na výrobu automobilových baterií uvažuje implementaci dvou plánů kontroly kvality QC. Jestliže výroba funguje správně, váha baterií je normálně rozdělená se specifikovanou střední hodnotou a směrodatnou odchylkou S.

   – QC (A) považuje baterii za špatnou, jestliže její váha je pod 2S od specifikovaného průměru.
   – QC (B) považuje baterii za špatnou, jestliže je vzdálena více než 1,5 interkvartilového rozpětí pod dolním kvartilem.

   Předpokládejme, že je proces v pořádku.

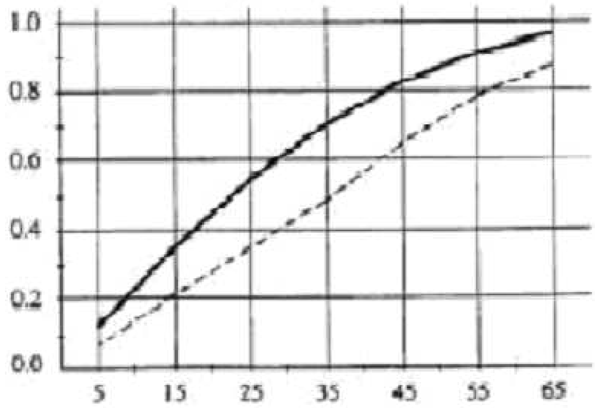
   (a) Jaká část baterií bude vyřazena plánem A?
   (b) Jaká je pravděpodobnost, že aspoň jedna ze dvou baterií bude vyřazena plánem A?
   (c) Jaká část baterií bude vyřazena plánem B?

3. Tabulka obsahuje kumulativní pravděpodobnosti populace ve Spojených státech od roku 1900 do roku 2000. Například 0,344 resp. 34,4 procent populace bylo mladší nebo rovno 15ti letům v roce 1900, zatímco pouze 0,209 resp. 20,9 procent je mladší nebo rovno 15 letům v roce 2000.

Věk	1900	2000
5	0,121	0,066
15	0,344	0,209
25	0,540	0,344
35	0,700	0,480
45	0,822	0,643
55	0,906	0,781
65	0,959	0,870

Data o rozdělení věku v populaci

Následující graf ukazuje kumulativní relativní četnosti vynesené proti věku v letech 1900 a 2000 (plnou čárou je znázorněn rok 1900). Tabulka a graf se mají využít pro srovnání věkové struktury v letech 1900 a 2000.



Proveďte:

(a) Aproximujte medián věku pro každé rozdělení.
(b) Aproximujte interkvartilové rozpětí pro každé rozdělení.
(c) Použitím výsledků z částí (a) a (b), popište pro historické pojednání jednou větou nebo dvěma větami srovnání rozdělení věků v letech 19000 a 2000.

Literatura

[1] Moore D. a McCabe G. (1998) *Introduction to the Practice of Statistics, 3rd ed.* W.H. Freeman, New York.

[2] Moore D. (1994) *The Basic Practice of Statistics*. W.H. Freeman, New York.

[3] Yates D., Moore D. a McCabe G. (1998) *The Practice of Statistics*. W.H. Freeman, New York.

WWW stránky pro AP statistiku

[1] `http://www.collegeboard.org/ap/statistics`

[2] `http://wash.cr.k12.ia.us/academics/math/apinfo/states`

[3] `http://forum.swarthmore.edu/epigone/apstat-l`

NCTM standardy 2000

[1] `http://standards-e.nctm.org`

ANALÝZA ROZDELENIA PRIEMYSELNÝCH PODNIKOV V RÁMCI PREDMETU EKONOMICKO-ŠTATISTICKÉ ANALÝZY

Jozef Chajdiak

FHI EU, Katedra štatistiky, Dolnozemská cesta 1, 852 35 Bratislava
chajdiak@statis.biz

Predmet Ekonomicko-štatistické analýzy sa učí na Fakulte hospodárskej informatiky ako voliteľný predmet pre študentov štvrtého a piateho ročníka. Predmet nadväzuje na základné štatistické a ekonomické predmety z predchádzajúcich ročníkov. Hlavnou úlohou výučby predmetu je, aby si študenti prehĺbili svoje schopnosti samostatne analyzovať konkrétne súbory údajov z národohospodárskej praxe a robiť praktické závery na základe získaných analytických výsledkov. Vzhľadom k tomu, že analýzy sa realizujú na konkrétnych súboroch údajov, nezanedbateľnou časťou procesu výučby je aj získanie znalostí o bezprostredných číselných hodnotách analyzovaných skutočností.

V rámci predmetu Ekonomicko-štatistické analýzy sa používa niekoľko súborov údajov. Jedným z nich je súbor anonymizovaných podnikových údajov z priemyslu Slovenskej republiky za rok 2000. Súbor obsahuje údaje o výnosoch spolu, pridanej hodnote, hospodárskom výsledku pred zdanením, osobných nákladoch, pasívach, vlastnom imaní a priemernom evidenčnom počte zamestnancov. K dispozícii sú tiež údaje o právnej forme organizácie, druhu vlastníctva a oddieloch OKEČ.

Vlastná analýza sa sústreďuje na vypracovanie tabuliek rozdelenia podnikov do tried špecifikovaných zvolenými hodnotami príslušných ukazovateľov pričom frekvencie výskytu sa určujú na základe počtu podnikov patriacich do príslušných tried (klasická frekvenčná tabuľka) a tiež príslušných objemov absolútnych ukazovateľov a z týchto objemov vypočítaných hodnôt relatívnych ukazovateľov (rozšírená frekvenčná tabuľka). Okrem vypracovania troch rozšírených frekvenčných tabuliek, študenti majú úlohu otestovať zhodu rozdelenia s normálnym rozdelením a snaživejší študenti aj nájsť vhodnejšieho rozdelenia pri nezhode s normálnym rozdelením.

Druhým okruhom úloh je výpočet základných opisných štatistík a ich interpretácia.

Tretiu skupinu úloh tvoria úlohy analýzy rozptylu, kde premenou reakcie sú určené ekonomické ukazovatele a faktorom vplyvu na ich variabilitu je identifikačná premenná právna forma so svojimi úrovňami. Okrem výpočtu ANOVA tabuľky, sa overuje splnenie predpokladov použitia modelu, analyzujú sa kontrasty a zostrojuje

Klíčové slová: Výuka štatistiky pre odbor ekonómov.

sa tabuľka opisných štatistík pre jednotlivé úrovne faktorovej premennej a súbor spolu. Súčasťou je grafická prezentácia výsledkov, hlavne formou grafov priemerov a box-plotov.

Súbor poskytuje aj ďalšie možnosti analýzy. Prakticky veľmi zaujímavý je postup analýzy pomocou Excelu, časti Pivot Table.

STAKAN 2003

VÝUKA PŘEDMĚTŮ SE STOCHASTICKÝM ZAMĚŘENÍM NA FSI VUT V BRNĚ

Zdeněk Karpíšek, Pavel Popela a Josef Bednář

Vysoké učení technické v Brně, Fakulta strojního inženýrství, Ústav matematiky, Technická 2, 616 69 Brno

krpisek@um.fme.vutbr.cz

Autoři příspěvku se již řadu let věnují výuce statistických předmětů a předmětů, které využívají matematické stochastické modely, na Fakultě strojního inženýrství Vysokého učení technického v Brně. Jde o předměty v základním kurzu, na odborných inženýrských specializacích, vlastním specializovaném oboru Matematickém inženýrství, v doktorském studiu a kurzech celoživotního vzdělávání.

1. Poměrně nedávná historie z let 1959 až 1990

Nejprve několik slov ke starší historii výuky těchto předmětů. Po různých variantách výchovy technických inženýrů vznikla samostatná Strojní fakulta VUT v roce 1959. Období od tohoto roku do zhruba roku 1980 je charakterizováno pouze sporadickými pokusy o začlenění stochastického modelování do základního kurzu matematiky. Šlo pouze o občasnou iniciativu některých vyučujících, neboť jejich zaměření a striktní osnovy více neumožňovaly. Jiná situace byla v oboru strojírenská technologie, kde byly základy statistických metod obsahem speciálního předmětu a také součástí státní závěrečné zkoušky. Naopak v rámci vědecké přípravy procházela základním školením ze statistiky většina tehdejších interních i externích aspirantů. Jejich výuka však probíhala formou individuálních konzultací na základě požadavků školitelů a byla determinována nepříliš dostupnou literaturou. Postupně se ve větší míře dostaly statistické metody do výuky v tzv. postgraduálních kurzech, které fakulta zajišťovala na základě objednávky firem a ústavů.

Situace výuky pravděpodobnosti a statistiky v letech 1980 až 1990 byla v dobrém smyslu slova poznamenána v roce 1980 zavedením nového předmětu Matematika IV v rozsahu 3/3 pro 2. ročník studia, jehož jedinou náplní byly zmíněné metody. Šlo o důsledek unifikace osnov na všech strojních fakultách v tehdejší ČSSR, ovšem se snížením hodinové dotace ostatních tří matematických předmětů v základním kurzu. Navíc byl tento předmět zaveden pouze na pětiletých (konstrukčních) oborech a nikoli na čtyřletých (technologických a automatizačních) oborech, kde naopak

---

Klíčová slova: Výuka pravděpodobnosti a statistiky, statistický software.

Poděkování: Rádi bychom poděkovali za podporu grantu VVZ MŠMT CEZ: J22/98 261100009.

velmi chyběl. Naše tehdejší katedra zabezpečovala mimo tohoto předmětu ještě dva statistické předměty na specializacích (strojírenská technologie a slévárenská technologie), školení ve vědecké přípravě na fakultě z vybraných partií matematické statistiky a speciální předměty stochastického charakteru v tzv. postgraduálních kurzech technických kateder.

2. Od roku 1990 po současnost

Období od roku 1990 až po současnost je významně jiné, jak dokládají následující fakta. Po různých turbulencích a hledáních v roce 1990, kdy se objevily možná dobře míněné snahy omezit výuku stochastických předmětů na doporučenou či nepovinnou formu, se podařilo i za cenu postupného snížení hodinové dotace z 3/3 na 2/2 nejen udržet předmět Matematika IV, ale navíc jej rozšířit v souvislosti se zánikem čtyřletého studia na všechny obory na fakultě v rámci jednotného kurzu matematiky v prvním stupni magisterského studia. Tímto předmětem prochází v posledních letech ročně zhruba 500 studentů. V souvislosti se zahájením prezenčního a později i kombinovaného bakalářského studia byl nejprve vytvořen podle oborů diferencovaný předmět Statistika a pravděpodobnost, který nyní v jednotné formě pro všechny obory má rozsah 2/1. Tento předmět ročně absolvuje cca 220 studentů. Třetím předmětem v základním kurzu je nepovinný předmět Statistický software v rozsahu 0/1, jejž absolvuje ročně cca 130 studentů. Výuce statistických programů pro PC na naší fakultě je níže věnován zvláštní odstavec. V popisovaném období se podařilo významně rozšířit výuku statistiky také na technických specializacích, tj. ve druhém stupni magisterského studia (ve 4. a 5. ročníku). Jde celkem o sedm současných předmětů, jimiž ročně projde zhruba 160 studentů. Všechny výše zmíněné předměty jsou uvedeny v Tabulce 1.

Specifická situace je na vlastním pětiletém oboru Matematické inženýrství našeho ústavu. Obor, který vznikl díky příznivé situaci na fakultě a nadšení několika pracovníků ústavu v roce 1993, vychovává inženýry, v jejichž profilu dominuje orientace na matematické modelování a řešení technických i netechnických úloh a problémů. Obor vychází z dlouholeté vědeckovýzkumné spolupráce a konzultační činnosti pracovníků ústavu s pracovišti Fakulty strojního inženýrství a dalších fakult VUT i jiných univerzit, Akademií věd ČR a výzkumnými ústavy, a dalšími institucemi a firmami. Je akreditován na naší fakultě a zajišťován ve spolupráci s Katedrou aplikované matematiky Přírodovědecké fakulty MU v Brně. Mimo základních strojírenských předmětů projdou studenti základními matematickými předměty, šesti předměty stochastického charakteru, dále předměty souvisejícími nebo využívajícími stochastické modely (optimalizace, fuzzy množiny, obrazová analýza), předměty zaměřenými na numerické metody a funkcionální ana lýzu, solidními základy programování a specializovanými kurzy. Diplomové práce jsou pak orientovány na

Název předmětu	Ročník	Rozsah ZS	Rozsah LS	Pro obor
Matematika IV	2.		2/2 z, zk	Magist. stud.
Stat. a pravděp.	2.	2/1 z, zk		Bakal. stud.
Stat. software	2.		0/1 z	Magist. stud.
Stat. analýza	4.		2/1 kl	Materiálové inž.
Stat. analýza	5.		2/0 z	Tech. prostř.
Apl. statistika	4.	2/1 z, zk		Řízení jakosti
Technika exp.	5.	2/1 z, zk		Řízení jakosti
Analýza inž. exp.	5.		2/1 z, zk	Fyz. inž.
Apl. mat.	3.		2/1 z	Fyz. inž.
AM - zprac. dat	5.		2/2 z, zk	Konstr. strojů

Tabulka 1. Předměty v základním studiu na FSI VUT.

konkrétní matematické modely inženýrských úloh a vybrané teoretické problémy, včetně tvorby programů pro PC a realizace výpočtů. Ve všech pěti ročnících je dohromady zhruba 110 studentů a z cca 20 absolventů ročně má více než polovina diplomovou práci obsahově zaměřenou na přímé či podpůrné stochastické modely. Předměty se stochastickým zaměřením uvedeného oboru jsou uvedeny v Tabulce 2.

Název předmětu	Roč.	Rozsah ZS	Rozsah LS	Kredity	Typ cv.
Pravděp. a stat. I	2.		3/2 z,zk	6	C2a
Pravděp. a stat. II	4.	3/2 z,zk		6	C2a
Stoch. procesy	4.		3/1 z, zk	4	C2a
Stoch. modelování	4.		0/2 kl	3	C1
Pravděp. a stat. III	5		2/2 kl	4	C2a
Analýza inž. exp.	5.		2/2 z, zk	5	C2a

Tabulka 2. Obor Matematické inženýrství na FSI VUT.

V každém akademickém roce pokračuje v postgraduálním doktorském studiu oboru Matematické inženýrství 5 až 7 našich absolventů, někteří další na jiných ústavech naší fakulty. V celém současném postgraduálním doktorském studiu na FSI jsou zařazeny 4 předměty stochastického charakteru, které absolvuje ročně cca 80 studentů. Tyto předměty jsou uvedeny v Tabulce 3.

Poznamenejme, že bližší informace o výše uvedených předmětech (garanti, návaznost, osnovy, požadavky aj.) může najít zájemce na Internetu [3].

3. Další „stochastické" aktivity

Mimo vlastní výuku jsou partie popisné statistiky, základů teorie pravděpodobnosti, matematické statistiky, stochastických procesů a jejich aplikací také součástí státních závěrečných zkoušek na oborech Matematické inženýrství a Řízení jakosti, jimiž projde ročně asi 45 studentů. Rovněž při státních doktorských zkouškách na FSI se s otázkami na aplikace těchto partií potká každoročně zhruba 25 doktorandů a při přijímacích zkouškách do postgraduálního doktorského studia se snaží odpovědět každoročně od 70 do 90 uchazečů na otázky spojené s teorií pravděpodobnosti, statistikou a jejich aplikacemi. V postgraduálním oboru Matematické inženýrství byly již obhájeny 4 disertační práce, jejichž obsahem je ve větší či menší míře modelování problémů s neurčitými informacemi metodami stochastického charakteru. Poměrně rozsáhlá je také konzultační činnost specialistů Ústavu matematiky při přípravě diplomových a disertačních prací na dalších ústavech FSI a jiných fakultách.

Název předmětu	Ročník	Rozsah ZS	Rozsah LS
Statistická analýza	1.	20 h	
Dynam. a vícerozm. stoch. modely	1.		20 h
Empirické modely	1.		20 h
Plánování experimentu	1.		20 h

Tabulka 3. Postgraduální doktorské studium na FSI VUT.

Za zmínku stojí také přednáška a část semináře na Univerzitě třetího věku VUT, které již druhým rokem přibližují studentům-seniorům možnosti aplikací a aspekty použití statistických metod jak ve výzkumu a výrobě, tak i v každodenním životě.

4. Statistický software

Souběžně se základním kurzem pravděpodobnosti a statistiky (předmět Matematika IV) ve 2. ročníku magisterského studia je nyní na naší fakultě vyučován volitelný (nepovinný) předmět Statistický software. Tento předmět je zaměřen na seznámení se statistickými systémy Statgraphics a Statistica, které se používají ve výuce a výzkumu na odborných ústavech fakulty a dále na seznámení se statistickými funkcemi tabulkového procesoru Excel.

V roce 1992 byla výuka statistického softwaru přímo začleněna do cvičení předmětu Matematika IV. Díky tehdejší hodinové dotaci 3/3 probíhala asi čtvrtina cvičení na počítačových učebnách na software Statgraphics 5.1, jehož verze DOS byly postupně modernizovány až na Statgraphics plus 2.4 for WIN. Z důvodu následného zkrácení dotace na rozsah 2/2 vznikl v roce 1996 volitelný předmět Programový systém Statgraphics. V roce 2002 jsme vybavili jednu počítačovou učebnu pro výuku výkonným systémem Statistica 6.0. Protože jsou studenti seznamováni s oběma systémy, byl tento předmět přejmenován na Statistický software. Vzhledem k rozsahu 0/1 je jeho výuka v 7 dvouhodinových lekcích a obsahově i časově navazuje na předmět Matematika IV. Přestože je předmět Statistický software nepovinný, zapisuje si ho více než čtvrtina studentů, protože zde mohou vypracovat povinnou aplikačně zaměřenou semestrální práci z předmětu Matematika IV. Využití statistického softwaru ve výuce statistiky přineslo:

- zefektivnění výkladu pojmů pravděpodobnosti a statistiky,
- vizualizaci složitých pojmů a možnosti hlubšího pochopení na základě interaktivních změn parametrů,
- využití systému na rutinní výpočty otevřelo prostor pro zkvalitnění výkladu modelování a interpretace získaných výsledků,
- osvojení základních postupů při zpracování pozorovaných dat.

Přesto si autoři myslí, že v základním kurzu by měl software sloužit pouze jako podpora a neměla by se statistika učit jen na počítači, protože existuje nebezpečí redukce chápání statistických metod na pouhé mačkání kláves.

5. Zkušenosti a poznámky závěrem

Zdá se, že výuka a výchova studentů na FSI VUT směrem ke stochastickým modelům je na tom tak dobře, že už může být jenom hůře. Pro dokreslení příznivé skutečnosti lze totiž uvést, že z celkem 75 předmětů zabezpečovaných v současné době Ústavem matematiky na FSI, je 21 předmětů přímo zaměřeno na pravděpodobnostní a statistické metody a 23 předmětů s danými metodami více či méně souvisí. Některé zkušenosti a poznatky z našeho ústavu však nejsou tak optimistické.

U mnoha studentů se ve výuce nedaří získat skutečný zájem o metody a tím méně pak o stochastický způsob myšlení. V nižších ročnících berou studenti tyto předměty jako povinnost a svůj názor mění obvykle až s přibližujícím se koncem studia, přesněji s diplomovou prací. Situace je lepší na našem oboru Matematické inženýrství, ale studenti tohoto oboru povětšinou vědí, proč si jej vybrali a někteří dokonce odhadli, co je čeká. Dokonce i nemalá část studentů postgraduálního studia na FSI nenachází příliš velké zalíbení ve statistických metodách, i když je za potřebné považují. Není se možno příliš divit, neboť jim někteří školitelé zařazují do studijního plánu tyto předměty jako tzv. osvětové. K výraznému zájmu ze strany těchto doktorandu však dochází v souvislosti se zpracováním měření a modelováním nedeterministických problémů při přípravě disertace, kdy někteří získávají ke statistice apod. velmi kladný vztah. Doufejme, že jim vydrží i po úspěšné obhajobě disertační práce a následné vědecké či praktické činnosti. V úspěšnosti výuky matematických předmětů v základním kurzu se u studentů neblaze projevuje nepříliš velká současná úroveň matematických znalostí a dovedností ze střední školy. Ta se pak promítá formou prerekvizit do výuky pravděpodobnosti a statistiky, i když si žádný z našich mladých vyučujících po rychle získaných pedagogických zkušenostech nedovolí konstruovat stochastické modely v inženýrském studiu jinak nežli na intuitivním základu. Dochází tak k rozporu s časovými dotacemi předmětů a snahou spíše studenty pro statistiku získat než odradit (získání považujeme za svou svatou povinnost). Své také sehrává studijní literatura. V současnosti stále chybí dostupná a výkladově přístupná učebnice typu kalkulu. Na WWW stránkách se sice postupně objevují kvalitní české učební texty z pravděpodobnosti a statistiky, jejichž hypertextové provedení dokládá, že autoři dostali za publikaci buď hodně zaplaceno [4] anebo spíše propadli radosti z dobře vykonané práce [5]. Na naší fakultě pokrýváme výuku stochastických předmětů v základním kurzu klasickým učebním textem a učební pomůckou pro kombinované bakalářské studium [1,2].

Zvláštní problematiku má výuka pravděpodobnosti a statistiky na PC. K výše uvedeným informacím pouze dodejme, že ve stochastických předmětech má bezesporu přednost hotový, solidní a stabilní profesionální statistický software pro PC. Tyto požadavky však díky nevyrovnanosti, nedostatečnému pokrytí statistických úloh a popisu metod i ovládání oblíbený Excel zdaleka nesplňuje, avšak v některých případech je dobrou pomůckou. Po mnohaletých zkušenostech se nám osvědčil Statgraphics (jak verze DOS, tak i WIN) a v současnosti také Statistica 6.0. Ve vybraných specializacích navíc používáme S-Plus 6.0 a QCExpert a statistické knihovny Matlabu a MathCadu. Nejen naše zkušenosti se statistickým softwarem však ukazují, že jde ve výuce pouze o pomůcku. Aby se student seznámil s probíranými metodami důvěrněji, budou výpočty na kalkulačce asi ještě nějaký čas nenahraditelné.

Další rychle se rozvíjející oblastí je využívání statistických apletů a specializovaného software z Internetu - viz např. [6]. Jde o velmi účinné pomůcky pro výuku, avšak jejich využívání mimo případných autorských práv také brání dostupnost, rychlost komunikace a potřeba práce na síti v reálném čase. Problémů s výukou stochastických předmětů je mnoho. Zvláštní pozornost si na technické fakultě zasluhuje např. metodika, jak do výuky začlenit poznatky z aplikovaného výzkumu, spolupráce s dalšími pracovišti a aplikace statistiky v praxi. Díky pochopení vedení naší fakulty a řady kolegů z našeho ústavu i z jiných pracovišť fakulty (někteří totiž prošli naším školením ze statistických metod) se nám ve výuce leccos podařilo, v něčem tápeme a hledáme. Výuka matematiky a tím také předmětů se stochastickým zaměřením mívá obvykle tyto možné výsledky: kuchařka, řemeslo a aplikace, způsob myšlení a tvůrčí práce. Snažíme se o dosažení alespoň druhého stupně. Závěrem zmiňme, že při výuce stochastických předmětů jsou široce využívány poznatky řady projektů řešených na odboru a zejména pak výsledky VVZ MŠMT CEZ: J22/98 261100009. Zajímající se studenti jsou do řešení projektů zapojování, a tak je docilováno propojování výuky s výzkumem.

Literatura

[1] Karpíšek Z. (2002) *Matematika IV - Statistika a pravděpodobnost. Učební text*. FSI VUT v CERM Brno, Brno.

[2] Karpíšek Z., Popela P. a Bednář J. (2002) *Statistika a pravděpodobnost. Učební pomůcka - studijní opora pro kombinované studium*. FSI VUT v CERM Brno, Brno.

[3] `http://www.fme.vutbr.cz/studium/zakl.html`

[4] `http://www.quantlet.com/mdstat/mmstat.html`

[5] `http://home.zcu.cz/ friesl/Vyuka/Odkazy.html`

[6] `http://www.stat.sc.edu/rsrch/gasp/`

INTERNETOVÁ VÝUKA STATISTIKY A ANALÝZY DAT

Aleš Linka, Miroslav Brzezina, Petr Volf a Jiří Týř

Technická univerzita v Liberci, Hálkova 6, 461 17 Liberec

ales.linka@vslib.cz, miroslav.brzezina@vslib.cz, petr.volf@vslib.cz, jirka@dst.cz

Studium statistiky a analýzy dat je běžně studenty považováno za obtížné, protože vyžaduje širokou paletu dovedností, jež zahrňují jak kvantitativní a grafické pochopení podstaty problému tak i dobré matematické schopnosti. Atraktivní a velmi výkonnou inovací klasického přístupu k výuce statistiky a analýzy dat je přístup využívající Intranet nebo Internet.

1. Úvod

Spojení matematiky a internetu, statistiky a internetu jsou v posledních letech stále častější. Příklady jsou např. datové zdroje na Webu, distribuce dat prostřednictvím Webu, online časopisy, elektronické učební texty i další učební pomůcky. Rovněž může využívat velké množství software přístupného přes Web, ale i software přímo spustitelný přes webovské pohlížeče. Nové technologie přináší prospěch nejen studentům, ale i učitelům, i když přístup obou skupin k výuce prostřednictvím Internetu může být rozdílný. Z pohledu studentů lze hlavní přínosy shrnout do následujících bodů:

- rychlý a snadný přístup k metodologii a datům přes webovské prohlížeče
- interaktivní příklady, které jsou jednou z nejrychlejších metod učení
- snadný a plynulý přechod z učebny k domácí práci i s případnou plnou podporou statistického software

Z pohledu učitelů pak hlavní užitek internetové technologie spočívá:

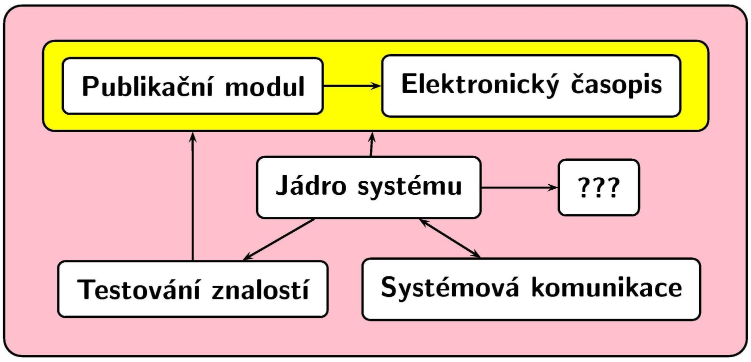
- v rychlém a snadném vysílání metodologie a dat přes prohlížeče k uživateli
- v uživatelsky přívětivém prostředí
- v mocném a flexibilním nástroji pro šíření výsledků výzkumu

Klíčová slova: Výuka pomocí Internetu, statistika, analýza dat.

Poděkování: Příspěvek vznikl s podporou projektu FRVŠ 378/2003, FRVŠ 374/2003 a MSM 245100303.

2. Využití Webu ve výuce statistiky a analýzy dat

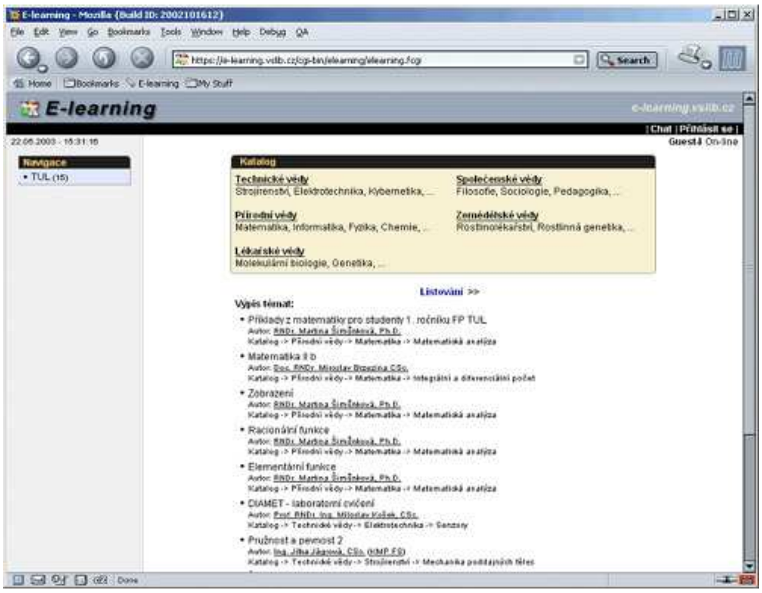
Užití Webu ve výuce statistiky a analýzy dat je velmi různorodé, od interaktivních příkladů a softwarových demonstrací až po kompletně webovské kurzy bez použití pera a papíru, jako je například ten, který nabízí K. L. Norman z Univerzity of Maryland [1]. Elektronické výukové texty bychom mohli rozdělit do několika základních skupin. Do první skupiny patří texty, které vznikly převedením knihy nebo jen některé její části do elektronické podoby, jež byla umístěna na Internet. Zde se můžeme setkat s texty v různých formátech. Mezi nejoblíbenější patří formát PDF. Soubory v tomto formátu lze s pomocí webovského prohlížeče snadno zobrazit, máme-li jako plugin prohlížeče nainstalován program Adobe Acrobat Reader. Poněkud pracnější je inovace obsahu textů v tomto formátu, neboť zde je nutností nástroj, který z inovovaného zdrojového textu opět vyrobí soubor ve formátu PDF a samozřejmě nový soubor opět musíme umístit na Internet.



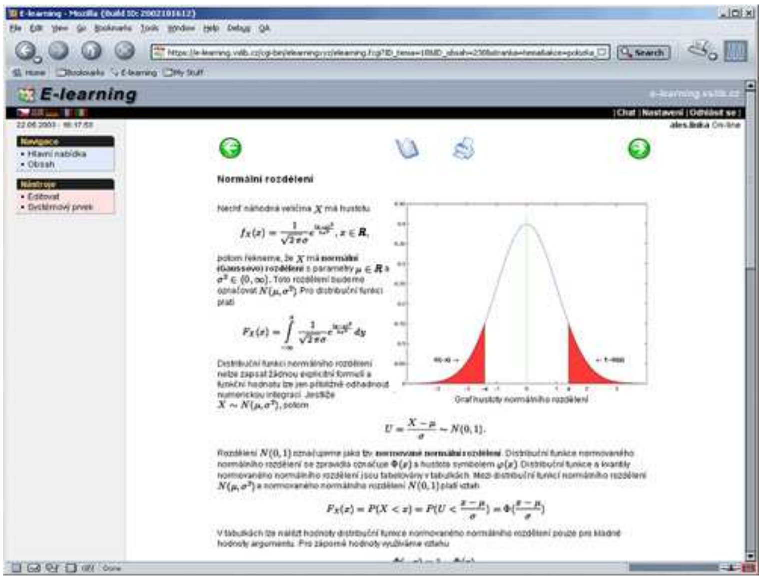
Obrázek 1. Struktura univerzitního e-learningového systému na TU v Liberci

V současnosti se na Webu ve velké míře objevují učební texty hypertextové a to nejen ty, které byly převedeny do hypertextové podoby ze zdrojového textu v Latexu pomocí konvertoru latex2html nebo generovány ze zdrojového textu ve Wordu, ale i ty, jež obsahují různé dynamické prvky. Příkladem hypertextové učebnice s dynamickými prvky je elektronická učebnice HyperStat [5], která obsahuje řadu didakticky užitečných dynamických prvků nebo odkazů na tyto prvky ve formě appletů, Java a CGI skriptů.

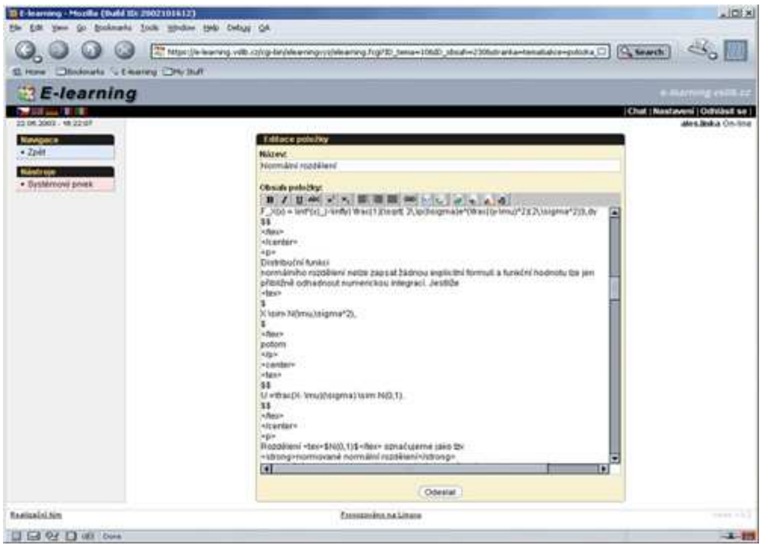
Pokročilejší formou jsou elektronická skripta MM*Stat [3], jež se skládají z výkladu, doplňujících informací, několika druhů příkladů, cvičení a doplňkových partií. Interaktivní příklady využívají statistický software Xplore [4], který lze pomocí Java appletů používat přes Web.



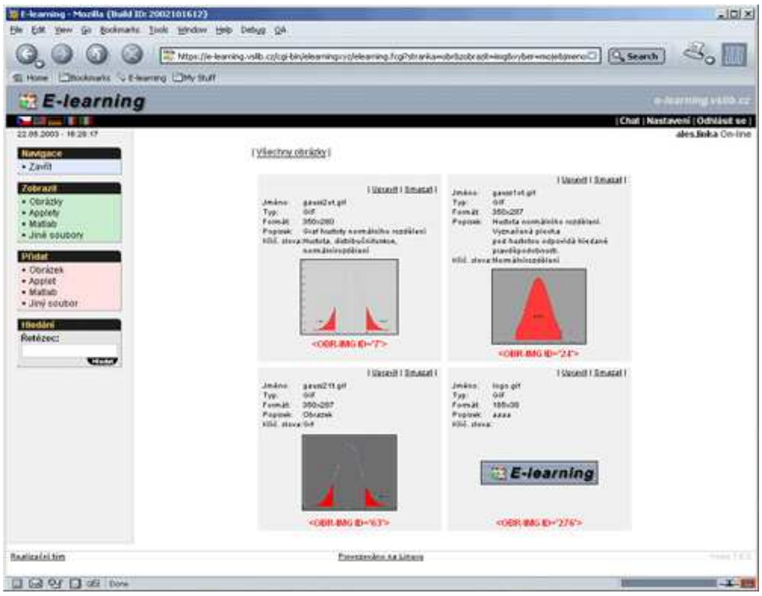
Obrázek 2. Vstupní stránka do publikačního modulu UES.



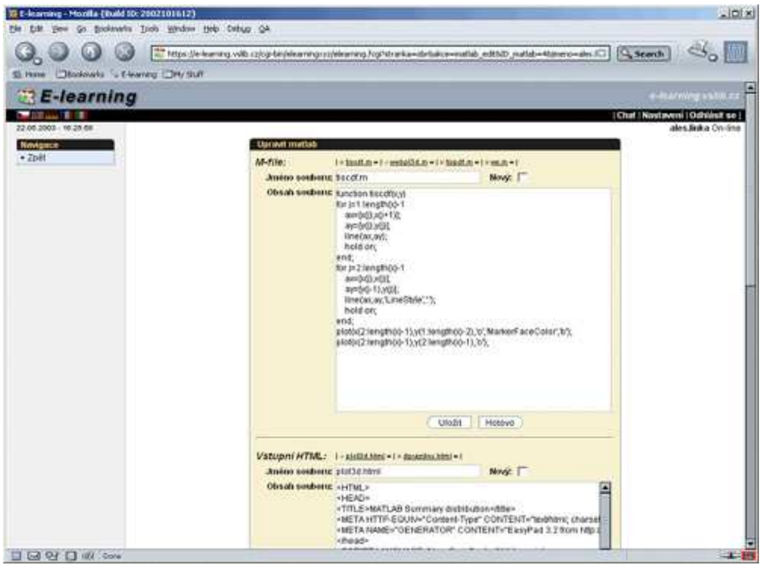
Obrázek 3. Úkázka publikovaného textu v UES.



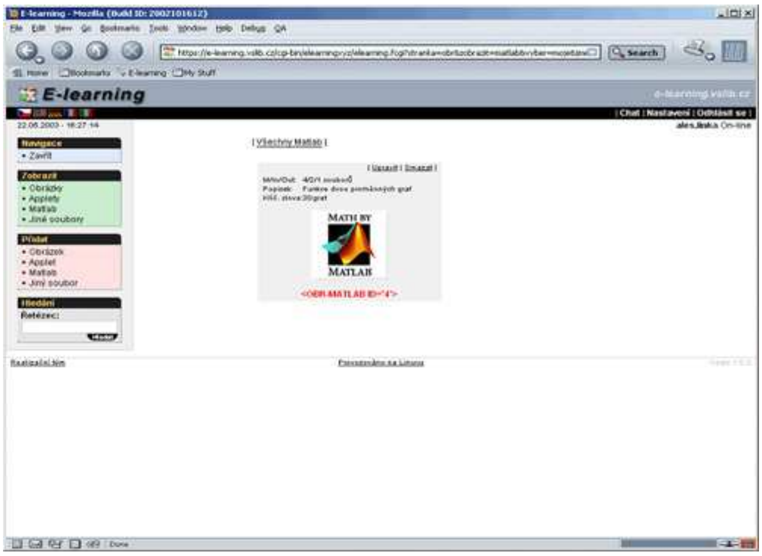
Obrázek 4. Vstup do módu pro vkládání a editaci textu v USE



Obrázek 5. Vstup do módu pro vkládání dynamických aplikací a obrázků.



Obrázek 6. Editační okno pro aplikace Matlab Web Server.



Obrázek 7. Seznam vytvořených aplikací pro Matlab Web Server

Mezi statistické programy, které mají svůj webovský interface patří i volně šiřitelný statistický software R. Rweb je webovský interface pro R (Robert Gentleman and Ross Ihaka), který ze vstupního okna převezme R script, spustí R, zpracuje script v dávkovém režimu a vrátí výstup ke zobrazení ve webovském prohlížeči (viz [2]). Zdrojový kód R je téměř shodný s S a Splus. Ten, kdo má nějaké zkušenosti s těmito balíky, nebude mít s používáním R problém.

WebStat 3.0 [9] je statistický software pro analýzu dat na Webu. WebStat načítá soubory v textovém formátu nebo soubory z Excelu z lokálního PC nebo přes webovské rozhraní. Rovněž lze použít simulovaná data z širokého škály rozdělení. Výsledky analýzy jsou předány ve formě HTML stránky nebo v tiskové podobě.

Celá řada dalších softwarových balíků má webovské rozhraní, např. Matlab, Statistica, Splus nebo Mathematica. Pořízení a provozování příslušných Webů ovšem není nijak levnou zaležitostí a navíc příslušné aplikace pro zařazení do interaktivních učebnic je zpraviedla nutné teprve vytvořit.

3. Univerzitní e-learningový systém na TUL

V rámci několika projektů probíhal v minulých letech na Technické univerzitě v Liberci vývoj některých prvků pro internetovou výuku, ať již jsou to elektronické interaktivní učebnice, např. Linka *et al.* [6], Militký [8] nebo databáze příkladů z matematiky (Brzezina [1]).

V současné době probíhá již déle než rok vývoj vlastního komplexního systému pro e-learning, viz Linka *et al.* [7]. Tento systém je od samého začátku koncipován pro využití více univerzitami. Obecná struktura systému je znazorněna na Obrázku 1. V systému byl ukončen vývoj „Publikačního modulu“. Tento modul umožňuje publikování a editování odborných textů ve formátu HTML s využitím vlastních speciálních tagů. Tyto speciální tagy slouží pro vkládání multimediálních prvků, jako jsou obrázky, applety, dynamické aplikace Matlab Webu atd. Jednoduchý editor pro vkládání a editaci aplikací Matlab Webu je součástí systému. Speciálními tagy neslouží jen pro vkládání výše uvedených obrázků, appletů a Matlab Web aplikací, ale i pro dynamicky generované matematické vzorce, které jsou generované ze zdrojovém textu jazyka TeX.

Systém mohou spravovat uživatelelé, kteří mají přidělena odpovídající práva (např. úpravu témat, úpravu systémových prvků, administraci uživatelů atd.). Prostředí celého systému je koncipováno jako vícejazyčné. Jazykové mutace je možné spravovat přes webovské rozhraní, ke kterému mají přístup pouze uživatelé se speciálními právy. Dnes je systém dostupný v pěti jazycích (čeština, angličtina, němčina, francouzština a italština).

Nyní také probíhá vývoj testovacího a zkušebního modulu, který bude určen pro ověření znalostí. Modul bude umožňovat vkládat testy, editovat je a následně generovat podle zadaných pravidel. Testy budou generovány buď náhodně samotným systémem nebo vybírány registrovaným uživatelem. Při vkládání testů lze vybírat z pěti typů testů ("single choice testing", "multi choise testing", odpověď číslem, slovem, větou). Výsledek testů se bude ukládat do profilu uživatele a na jejich základě si uživatel bude volit další postup testování znalostí. Pro univerzálnost bude možno testy generovat i do písemné podoby.

Literatura

[1] Brzezina M., *Příklady z matematiky*. http://zombie.kap.vslib.cz/matematika

[2] Banfield J., *Rweb*. http://www.math.montana.edu/Rweb/

[3] Härdle W., Rönze B., *MM*Stat*. http://www.quantlet.com/mdstat/scripts/mmcze/java/start.html

[4] Härdle W., Rönze B., *Xplore*. http://www.xplore-stat.de/index_js.html

[5] Lane D. M., *HyperStat online textbook*. http://davidmlane.com/hyperstat/

[6] Linka A, Volf P. a Bajzík V., *Řízení jakosti v textilu*. Odkaz na adrese: http://www.ft.vslib.cz/index.cgi?sou=study/studijni_pomucky.htm&

[7] Linka A, Volf P. a Bajzík V., *Univerzitní e-learningový systém*. Odkaz na adrese: http://e-learning.vslib.cz

[8] Militký J., *Elektronická příručka zpracování dat v Matlabu*. Připraveno k publikaci na Univerzitním e-learningovém systému.

[9] *WebStat 3.0* http://www.webstatsoftware.com/

STATISTIKA NA ČESKÉ ZEMĚDĚLSKÉ UNIVERZITĚ V PRAZE

Helena Nešetřilová

Česká zemědělská univerzita v Praze, Provozně ekonomická fakulta, katedra statistiky, Kamýcká 129, 165 21 Praha 6

nesetrilova@pef.czu.cz

Cílem tohoto příspěvku je podat rámcovou informaci o výuce statistických předmětů na české zemědělské univerzitě v Praze. česká zemědělská univerzita má čtyři fakulty a Institut tropů a subtropů. Vzhledem k tomu, že na každé fakultě je profil absolventů poněkud jiný, liší se i rozsah a zaměření statistických kurzů na jednotlivých fakultách. Aby bylo zřejmé, kolik studentů v rámci příslušné fakulty statistické kurzy, vyučované katedrou statistiky, absolvuje, je v následující tabulce uveden počet studentů v denním studiu na jednotlivých fakultách ve školním roce 2002/2003.

Fakulta	Počet studentů
Provozně ekonomická	4524
Agrobiologie	2026
Lesnická	1434
Technická	1273
Institut tropů a subtropů	84
Celkem	9541

Tabulka 1. Počet denních studentů ČZU ve školním roce 2002/2003.

Katedra statistiky, která zajišťuje výuku statistiky na Provozně ekonomické, Agrobiologické a Technické fakultě a Institutu tropů a subtropů, vyučuje v současné době 17 povinných a 34 volitelných kurzů. Základní kurz statistiky, který je povinný pro všechny studenty, je na největší Provozně ekonomické fakultě obvykle dvousemestrový (rozsah 2/2), na ostatních fakultách jednosemestrový. Jeho zařazení do curricula se liší, podrobnější informace jsou shrnuty v následujících tabulkách.

Pokud je základní statistický kurz dvousemestrový (Provozně ekonomická fakulta), vyučují se v něm obvykle v prvním semestru základy pravděpodobnosti, popisná statistika, bodové a intervalové odhady parametrů základního souboru, základní parametrické testy založené na výběrech z normálně rozdělené populace včetně jednoduchých modelů analýzy rozptylu a vybrané na distribuci nezávislé testy.

Klíčová slova: Výuka statistiky v oblasti zemědělství.

Fakulta	Rozsah	Semestr	Kurz
Technická	2/1	2	Statistika
Agrobiologie	2/2	3	Biometrika
ITS	2/2	3	Statistika

Základní (povinný) kurz statistiky

Provoz a ekonomika	2/2	5	Statistika I
		6	Statistika II
Podnikání a administrativa	2/2	3	Statistika I
		4	Statistika II
Systémové inženýrství	2/2	2	Statistika I
		3	Statistika II
Informatika	2/2	3	Teorie pravděpodobnosti
		4	Matematická statistika I
		5	Matematická statistika II
Veřejná správa a reg. rozvoj	2/2	3	Základy statistiky

Situace na fakulě provozně ekonomické.

Druhý semestr je věnovám především regresní a korelační analýze a časovým řadám. Menší část druhého semestru je vyhrazena základní informaci o metodách vícerozměrné statistické analýzy a také indexní analýze. V základním kurzu se studenti učí pracovat s programem STATISTICA a pomocí tohoto softwaru také zpracovávají individuální projekty, kterými je zakončen každý semestr výuky statistiky. Kromě povinného základního kurzu, který je zařazen do bakalářské části studia, si mohou studenti v magisterské části studia vybrat další statistický kurz z nabídky volitelných předmětů. Vzhledem k poměrně velkému počtu volitelných kurzů uvádím jenom výběr některých z nich:

- Předdiplomní statistický seminář.
- Finanční a pojistná matematika.
- Statistická analýza časových řad.
- Matematická statistika III.
- Mezinárodní zemědělská statistika.
- Statistické metody v marketingu a obchodu.
- Statistická analýza výběrových šetření.
- Statistical Seminar (je vyučován v angličtině).
- Financial and Insurance Mathematics I, II (kurzy jsou vyučovány v angličtině).
- Seminář výpočetní statistiky.

Z volitelných kurzů se samořejmě otevírají pouze ty, na které se přihlásí dostatečný počet studentů. Oblibě mezi studenty se těší mj. Předdiplomní statistický seminář, který si obvykle studenti volí v okamžiku, kdy zjistí, že budou statistiku ke zpracování diplomové práce skutečně potřebovat. Takto motivováni se ke statistice znovu vracejí a mají možnost (ba dokonce povinnost) konzultovat s vyučujícími způsob jak získat a posléze zpracovat svá vlastní data a jak interpretovat získané výsledky.) V průběhu magisterského studia se studenti v rámci Předdiplomního semináře i některých jiných volitelných předmětů mohou seznamit (kromě STATISTICA) i s dalším statistickým softwarem (SAS, SPSS). To, že má katedra možnost používat několik softwarových balíků, není způsobeno výjimečnými finančními možnostmi katedry statistiky, ale tím, že celá zemědělská univerzita má zakoupenu multilicenci (SAS) nebo software byl původně zakoupen pro řešení konkrétního grantového projektu, na kterém se katedra spolupodílela (sociologický výzkum, SPSS).

Studenti, kteří ve studiu pokračují v postgraduální formě, se mohou ke statistice vrátit ještě jednou. Statistické předměty pro studenty postgraduálního doktorského studia jsou už obvykle zaměřeny na specifickou skupinu statistických metod, která odpovídá odborné orientaci malých skupinek doktorandů a mají proto mnohem neformálnější průběh. Opět pouze výběr některých tématických okruhů (kurzů):

- Ekonomické časové řady.
- Vícerozměrné statistické metody.
- Statistické metody pro agrobiologii.
- Statistické procedury v SAS.
- Statistika pro sociology.
- Vybrané metody aktuárské matematiky.

Na všech úrovních výuky, od základních kurzů až po postgraduální kurzy, se nám osvědčily individuální studentské projekty. Pomáhají totiž rozvoji statistického myšlení, schopnosti, formulovat problémy jiných oborů pomocí statistiky a zpětně hledat interpretace statistických výsledků v těchto oborech. Tomu, jak jsou tyto projekty na čZU organizovány, je věnován příspěvek Ing. Dagmar Bínové, na který bych chtěla na tomto místě odkázat. Rozvoj statistického myšlení souvisí i s citem pro vhodné použití jednotlivých metod. Celá řada metod je navíc vázána na jisté teoretické předpoklady, které (ve větší či menší míře) nebývají v praxi splněny. Získat cit pro to, kdy lze porušení těchto předpokladů tolerovat a kde je naopak podstatné a tudíž netolerovatelné, je velmi obtížné. Studentské projekty by měly studentů m pomoci přemýšlet i o této stránce statistiky. Užitečná je v této souvislosti knížka 1, která je používána na některých statistických seminářích.

Literatura

[1] Rand R. Wilcox (2001) *Fundamentals of Modern Statistical Methods*. Springer.

ZMENY TRENDU V SEZÓNNYCH ČASOVÝCH RADOCH. APLIKÁCIA NA DEMOGRAFICKÉ DÁTA

Karol Pastor

KTPaMŠ, FMFI UK, MLYNSKÁ DOLINA, 842 48 Bratislava
pastor@fmph.uniba.sk

Príspevok sa zaoberá identifikáciou zmien trendu sezónneho časového radu. Pri hľadaní odľahlých pozorovaní využíva klasickú dekompozičnú metódu. Postup je demonštrovaný na mesačných časových radoch živoa mŕtvonarodených v SR. Analýza je doplnená interpretáciu príčin zmien, medzinárodným porovnaním a zmiešaným regresným modelom.

1. Úvod

Cieľom tohoto príspevku nie je ponúknuť nové metódy, ako skôr na jednom konkrétnom príklade z praxe ukázať postup pri analýze časových radov a metodicky zdôvodniť jednotlivé kroky. Príspevok zároveň poukazuje na možnosti využitia štatistiky v demografických analýzach. Nemenej zaujímavý je samotný analyzovaný časový rad počtu narodených v SR od r. 1979, preto budeme venovať pozornosť aj interpretáciám zmien v trende. Podobný priebeh má i rad narodených v ČR, i keď najmä v poslednom desaťročí príčiny jednotlivých zmien môžu byť iné.

Vývoj v demografických časových radoch má mnoho pravidelných aj nepravidelných výchyliek. Sezónne výchylky majú svoje biologické i sociálno-kultúrne príčiny. Dôsledkom prirodzenej variability demografických javov sú náhodné výchylky, ktoré sa spravidla dajú modelovať pomocou Poissonovho, binomického alebo normálneho rozdelenia. Popritom môže existovať množstvo ďalších nepravidelných výchyliek, ktoré sú spôsobené vonkajšími zásahmi, ale i vedomými zmenami v správaní sa populácie. Na druhej strane, analýza týchto rušivých momentov môže pomôcť identifikovať náhle zmeny vo vývoji, odhaliť ich príčiny a prípadne i kvantifikovať ich účinok. O to pôjde aj v tomto príspevku. V prvom kroku je preto potrebné zistiť trend a pravidelné odchýlky, v druhom kroku analyzovať signifikantné nepravidelnosti a ich možné príčiny.

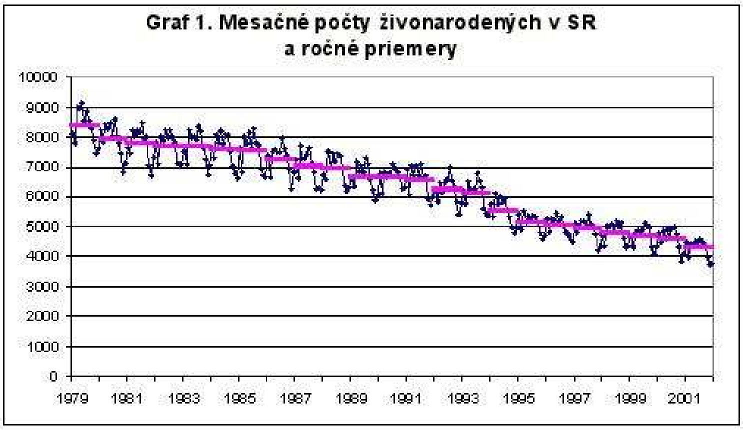
2. Analyzované dáta

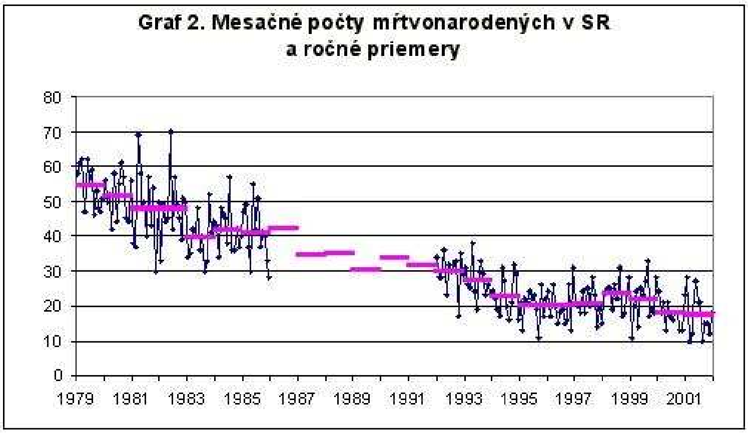
Dáta použité v tejto práci pochádzajú zo ŠÚ SR. Ide o mesačné počty živonarodených (*ZN*) a mŕtvonarodených (*MN*) v SR za roky 1979-2001, t.j. 23 rokov, čiže 276

Kľúčové slová: Sezónne časové rady, zmena trendu, demografia.

mesiacov (Graf 1, Graf 2). Za ten čas sa viackrát menila koncepcia spracovávania a publikácie dát. Preto mesačné počty mŕtvonarodených v rokoch 1986 - 1991 chýbajú (ročné existujú), v publikáciách FSU sa nenachádzajú a ŠÚ SR ich nemá k dispozícii.

Pre spomínané účely by boli vhodnejšie relatívne dáta, očistené aj od vplyvu vekovej štruktúry, také detailné dáta sa však nezisťujú. Naviac, boli by zaťažené pomerne veľkou náhodnou chybou. Na druhej strane, počas krátkodobých zmien sa veková štruktúra príliš nemení a preto stačí pracovať s absolútnymi počtami demografických udalostí. V prípade potreby použijeme ich prepočet na štandardizované mesiace s dĺžkou$365,25/12 = 30,4375$ dňa.





V záverečnej časti článku sú údaje o pôrodnosti porovnávané s indexom reálnej mzdy [4]. Aj tieto údaje pochádzajú zo ŠÚ SR, sú však k dispozícii len ako štvrťročné, pričom reálna mzda je porovnávaná s úrovňou reálnej mzdy príslušného obdobia predchádzajúceho roku, počnúc od roku 1990 (r. 1990 = 100 %).

3. Modely SARIMA a intervenčná analýza

Jedna z moderných štatistických metód, ktorá pri takýchto úlohách prichádza do úvahy, je intervenčná analýza na základe Boxovej-Jenkinsonovej metodológie. Táto metodológia poskytuje pomerne dobré modely na krátkodobé prognózy. Základy spomínaných metód možno nájsť napr. v učebniciach [1], [3], [5], resp. v práci [7]. Všeobecný zápis modelu SARIMA $(p, d, q)x(P, D, Q)_s$ s dĺžkou periódy s má tvar

$$\varphi(B)\Omega(B^s)(1-B)^d(1-B^s)^D Y_t = \theta(B)\tau(B^s)\varepsilon_t$$

kde B je operátor posunutia späť, $B\ Y_t = Y_{t-1}$ a $\varphi(B)$, $\theta(B)$, $\Omega(B^s), \tau(B^s)$ sú príslušné polynomické operátory stupňa p, q, P, Q a ε_t je biely šum. Ak v časovom rade $\{Y_t\}$ v čase T nastala porucha známeho typu, tak v uvedenom modeli miesto Y_t bude premenná Z_t, spĺňajúca niektorý z nižšieuvedených vzťahov (alebo analogický)

- aditívna porucha (additive outlier)

$$Z_t = Y_t + \omega P_{T,t}$$

- posun úrovne (level shift outlier)

$$Z_t = Y_t + \omega S_{T,t}$$

- inovačná porucha (innovational outlier) napr.

$$Z_t = Y_t + \omega\ \varphi(B)^{-1}\theta(B)S_{T,t}.$$

Tu $P_{T,t}$ je jednotkový impulz (pulse funktion) v čase T,

$$P_{T,t} = \begin{cases} 1, & t = T, \\ 0, & t \neq T, \end{cases}$$

a $S_{T,t}$ je jednotkový skok (step function) v čase T,

$$S_{T,t} = \begin{cases} 1, & t \geq T, \\ 0, & t < T, \end{cases}$$

pričom platí $(1-B)\ S_{T,t} = P_{T,t}$.

Analýzou radov s prítomnou poruchou sa zaoberá intervenčná analýza, a v niektorých štatistických balíkoch je aj softvérovo spracovan [7]. Pomocou intervenčnej analýzy je možné vyhodnotiť účinok nepravidelných zmien v modeloch SARIMA, ak poznáme typ zmeny a jej odozvy a okamih T, v ktorom nastala. V prípade, že typ ani okamih nepoznáme, treba ich najprv identifikovať napr. pomocou klasickej dekompozičnej metódy. Intervenčná analýza zlyháva i vtedy, ak chýbajú niektoré dáta (ako napr. v prípade skúmaného radu mŕtvonarodených).

Časový rad ZN, ktorý skúmame v tomto príspevku, je možné modelovať pomocou procesu SARIMA $(2,1,1)x(1,0,1)_{12}$. Pre účely článku sa však lepšie hodí dekompozičná metóda, z nej vzchádzame v ďalšej časti.

4. Trendová zložka

V demografických časových radoch sezónne odchýlky sú spravidla úmerné veľkosti veličiny. To platí aj v prípade ZN a MN. Použijeme preto multiplikatívny model (t.j. s proporcionálnou sezónnosťou). Ak by sa ukázala potreba, model je možné kombinovať a zaviesť ešte ďalšie, inštrumentálne premenné, na odlíšenie rôznych podmienok vývoja (napr. rôzna legislatíva), napr.

$$y_t = T_t\, S_t\, I_t\, E_t,$$

kde $I_t = a_0$ (pred zmenou) resp. a_1 (po zmene), pričom y_t je hodnota radu v čase t, a T_t, S_t a E_t sú trendová, sezónna a reziduálna zložka. Pri praktických výpočtoch sa niekedy pracuje s aditívnou reziduálnou zložkou, čo však spravidla vedie len k minimálnym rozdielom.

Trend možno modelovať parametrickou krivkou (napr. priamkou, exponenciálou), neparametrickou (napr. kĺzavými priemermi) alebo po častiach konštantnou funkciou (priemery za jednotlivé roky). Hlavnou výhodou parametrickej krivky sú jednoduché prognózy, hlavnou nevýhodou nezanedbateľné skreslenie ostatných zložiek. Nehodí sa na rady s nepravidelným priebehom trendu. Na skúmanie sezonality sú vhodnejšie kĺzavé priemery, ktoré lepšie eliminujú nepravidelnosti, a pritom zachovávajú trend. Nevýhodou po častiach konštantnej funkcie je nadhodnocovanie resp. podhodnocovanie sezónnej zložky na začiatku a na konci cyklu. Výhodou je zachovanie nepravidelností (ak nám o ne ide) a predovšetkým, metóda je pouiteľná i v prípade chýbajúcich údajov.

Na modelovanie radu ZN sme použili kĺzavé priemery (chronologický priemer dĺžky 13) i schodovitú funkciu. Rad má značné nepravidelnosti. Priemerný ročný úbytok je 2218 detí, v skutočnosti sa takáto ani podobná hodnota nevyskytuje. "Prirodzený" ročný úbytok je 500-1500 detí (okolo 1,5 %), v ôsmich rokoch je "vyvolaný" úbytok nad 3000 detí. Sú to roky (aj s hypotetickými príčinami a medziročným poklesom)

- 1980 - pokles o 5,1 % (zdraženie detského ošatenia),
- 1986-87 - pokles o 3,3 resp. 3,6 % (dôsledok liberalizácie zákona o UPT),
- 1989 - pokles o 3,8 % (politická neistota ?),
- 1992 - pokles o 5 % (o.i. cenová liberalizácia od 1.1.1991),
- 1994-95 - pokles o 9,4 resp. 7,4 % (reforma prídavkov na deti platná 1.1.1993-1.7.1994), úbytok spolu 9000,
- 2001 - pokles o 7,3 % (významné zvýšenie regulovaných cien r. 2000).

K interpretácii možných príčin mimoriadneho poklesu pôrodnosti v uvedených rokoch sa vrátime v časti 7.

5. Sezónna zložka

Sezónne indexy sa počítajú ako priemer (aritmetický, alebo správnejšie geometrický) hodnôt S_tE_t v jednotlivých mesiacoch. Sezónne indexy pre rad ZN aj MN pre rôzne modely trendu, pre štandardizované aj neštandardizované mesiace sú v Tabulke 1. Významnosť sezónnej zložky možno testovať viacerými spôsobmi. Malo by sa pritom pracovať so štandardizovanými mesiacmi, aby sa odlíšila sezonalita spôsobená zmenou intenzity javu od sezonality spôsobenou rôznou dĺžkou mesiaca. Prepočet na štandardizované mesiace vykonáme tak, že pôvodné dáta prenásobíme koeficientami $365,25/(12 * d_j)$, kde d_j je počet dní daného mesiaca, t.j. číslami

$$0,981855; 1,08705; 0,981855; 1,01458; ...$$

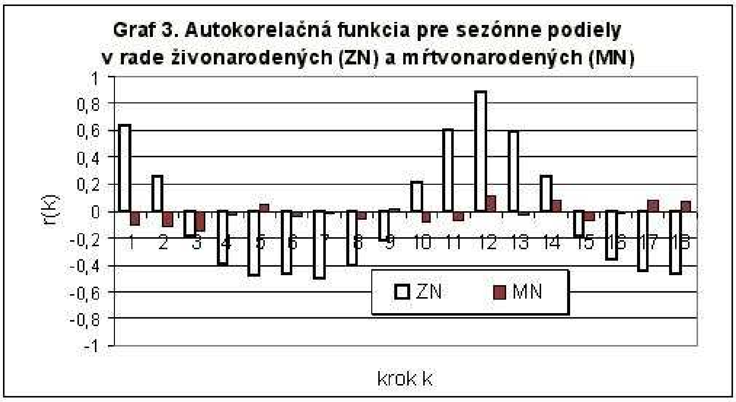
Priestupné roky sa obyčajne neberú do úvahy.

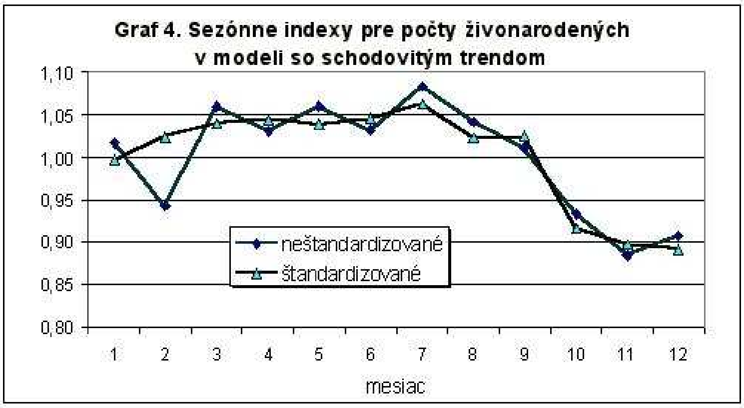
Prvým kritériom významnosti sezonality je autokorelačná funkcia r_k radu $y_t/T_t = S_tE_t$,, t.j. y_t očisteného od trendovej zložky. Pritom možno využiť Bartlettovu aproximáciu (viď napr. [1], s.104)

$$\sigma(r_k) \cong \sqrt{\frac{1}{n}(1 + 2\sum_{j=1}^{k_0} r_j^2)}, \qquad k > k_0.$$

Ak napr.$r_k < 2n^{-1/2}$ pre $k > 0$, rad možno považovať za realizáciu bieleho šumu. Vysoké hodnoty $r_{\alpha k}$ pre niektoré k a $\alpha = 1, 2, ...$ svedčia o významnej sezónnosti s periódou (dĺžkou) k. V prípade radu ZN je $r_{12} = 0,88$, čo svedčí o vysoko významnej sezónnosti, v prípade MN je $r_{12} = 0,12 < 0,14 = 2{\cdot}204^{-1/2}$, čiže sezónnosť je nevýznamná (Graf 3).

Iný spôsob, ako testovať významnosť sezónnych indexov, je analýza rozptylu. Ak rozdiely medzi triedami (mesiacmi) sú významné, sezonalitu treba brať do úvahy. (V multiplikatívnom modeli by sa správne malo pracovať s $ln(S_t.E_t)$, rozdiel je však spravidla zanedbateľný). Pre ZN sú štandardizované sezónne indexy vysoko významné, $P = 3 \cdot 10^{-114}$, nadpriemerné hodnoty sú v marci až septembri s maximom v júli ($S_7 = 1,06$), podpriemerné sú v októbri až decembri (minimum v decembri $S_{12} = 0,89$). Sezónne indexy pre schodovitý trend i pre kĺzavé priemery vychádzajú takmer zhodné. Pre MN sú nevýznamné ($P = 0,35$). Pre časový rad podielov $MN/(ZN + MN)$, čo je pozorovaná miera mŕtvorodenosti, vychádza $P = 0,73$, čiže sezóna nemá nijaký vplyv na to, či sa dieťa narodí ako mŕtve.





Keďže mŕtvorodenosť má charakter náhodnej chyby a predstavuje iba zlomok živorodenosti, sezonalita všetkých narodených je rovnaká ako u živonarodených. Osobitne sme ju neskúmali.

	kĺzavé priemery		schodovitý trend		schodovitý trend	
	živonarodení		živonarodení		mŕtvonarodení	
mesiac	neštand.	štandard	neštand.	štandard	neštand.	štandard
1	1,005	0,986	1,017	0,997	1,043	1,024
2	0,933	1,013	0,943	1,024	0,921	1,001
3	1,050	1,030	1,060	1,039	1,038	1,019
4	1,022	1,036	1,030	1,044	0,984	0,998
5	1,054	1,034	1,059	1,039	1,113	1,092
6	1,030	1,044	1,031	1,045	1,009	1,024
7	1,085	1,064	1,083	1,062	1,069	1,049
8	1,045	1,026	1,042	1,022	1,016	0,997
9	1,017	1,031	1,011	1,025	0,916	0,929
10	0,942	0,924	0,933	0,915	0,961	0,944
11	0,895	0,907	0,884	0,896	0,895	0,907
12	0,922	0,905	0,908	0,891	1,035	1,016

Tabulka 1. Sezónne indexy pre vyšetrované modely, SR 1979–2001.

6. Reziduálna zložka

Kvalitu modelu možno posudzovať pomocou rôznych kritérií. Je žiadúce, aby výsledná reziduálna zložka (v prípade multiplikatívneho modelu jej logaritmus) sa čo najviac ponášala na biely šum. Informáciu o významnosti autokorelácie rezíduí poskytuje napr. Durbinov - Watsonov test s testovacou charakteristikou $DW \in \langle 0, 4 \rangle$,

$$DW = \frac{\sum\limits_{t=2}^{n} (e_t - e_{t-1})^2}{\sum\limits_{t=1}^{n} e_t^2}.$$

V prípade radu ZN (t.j. živonarodených) je $DW = 1,73$ v modeli s kĺzavými priemermi a $DW = 1,50$ pre model so schodovitým trendom, čo svedčí o miernej autokerelácii, čiže o istej zotrvačnosti jednostranných odchýlok od modelu. Pre rad MN v modeli bez sezónnosti je autokorelácia nevýznamná, $DW = 2,22$.

Zhodu modelovaných a pozorovaných údajov možno najlepšie overiť pomocou chí-kvadrát testu dobrej zhody (pre absolútne počty). Nulová hypotéza predpokladá, že rozdiely sú iba náhodné, v praxi sa však pre demografické časové rady spravidla zamieta. To platí najmä pre tie časové rady, ktoré nemajú vyslovene biologickú (t.j. náhodnú) povahu. V prípade radu ZN je kritickou hodnotou $\chi^2(0,05; 275) = 314,7$ a hodnota testovacej charakteristiky pre model so schodovitým trendom je 839,8, pre model s kĺzavými priemermi je vyrovnaný rad kratší, $\chi^2(0,05;263)$ = 301,8 a testovacia charakteristika je 656,5. Rozdiely sú teda nenáhodné. Najväčší príspevok do testovacej charakteristiky chí-kvadrát v modeli so schodovitým trendom má rok 1990, vysoké príspevky majú aj roky 1979, 1987 a 1994. Hodnotu príspevku väčšiu ako 10 má 20 pozorovaní, najodľahlejšie z nich sú október a november 1990 (o 432 resp. 391 detí viac oproti očakávaniu). Model by bolo možné formálne vylepšiť pomocou umelých premenných, pre účely tohoto článku to však nie je potrebné.

V prípade MN je kritická hodnota $\chi^2(0,05;203)$ = 237,2, hodnota testovacej charakteristiky pre model bez sezónnosti je 230,9. Znamená to, že model bez sezónnosti je dobrý model pre MN $(P = 0,087)$.

7. Najvýznamnejšie nepravidelnosti a ich interpretácia

Ako bolo uvedené vyššie, v rade ZN je viacero nepravidelností rôznych typov, ktoré nemožno označiť za náhodu. Sú to jednak krátkodobé izolované výchylky, (krátkodobé odchýlky od trendu), ďalej atypické roky s odlišným priebehom, v priebehu ktorých mohlo dôjsť k zmene trendu, a napokon roky s mimoriadnym poklesom, a vtedy podnet na zmenu treba hľadať na jar predchádzajúceho roku. Uvedieme ich zoznam aj s možnými príčinami.

- Na jar roku 1979 boli v ČSSR realizované cenové opatrenia, ktoré sa týkali najmä detského ošatenia a energie všetkého druhu. Antinatalitný účinok týchto opatrení sa prejavil (spolu s nárastom potratovosti) už v druhom polroku 1979 a najmä r. 1980 a, ako dokladá Koubek [6], hlavne vo vekovej skupine 25 - 34 rokov a na plodnosti v druhom a treťom poradí.
- Na jeseň 1986 bol schválený zákon o umelom preruení tehotenstva s účinnosťou od 1.1.1987. Zrušenie interrupčných komisií viedlo k zvýšeniu hrubej miery potratovosti asi o 34 % oproti r. 1985. Tým by sa dal vysvetliť pokles pôrodnosti v druhej polovici r. 1987.
- Zníženie pôrodnosti r. 1989 (o 3,8 %) mohlo by vyvolané narastajúcou neistotou a napätím v politickej klíme po demonštráciách roku r. 1988 (v marci v Bratislave, na jeseň v Prahe) a očakávaním politických zmien. V prvej polovici roku 1990 sa ešte rodilo významne menej detí ako obvykle, avšak v druhej polovici významne viac, čo možno interpretovať ako dôsledok upokojenia a eufórie na jar 1990.
- Od 1.1.1991 sa v ČSFR rozbehla ekonomická reforma, súčasťou ktorej bola liberalizácia cien spotrebného tovaru, zrušenie mladomanželských pôžičiek a útlm hromadnej bytovej výstavby. Počiatočný kadodenný nárast cien základných potravín vyvolal paniku medzi obyvateľstvom a krátkodobé zvýšenie potratovosti. Účinky cenového šoku na pôrodnosť podľa okresov SR sú študo-

vané v práci [10]. R. 1992 poklesla pôrodnosť o 5 %, v "bohatších" okresoch ešte viac.

- Podľa reformy prídavkov na deti účinnej od 1.1.1993 do 30.6.1994, nárok na prídavky mali iba rodiny s celkovým mesačným príjmom 16800.- Sk bez ohľadu na počet členov domácnosti. Reforma najviac postihla rodiny so stredným príjmom a viacerými demi. Pôrodnosť na Slovensku r. 1994 klesla oproti predošlému roku o 9,4 % a r. 1995 o ďalších 7,4 %, pričom v druhom polroku 1995 sa už stabilizovala [10].
- Na začiatku r. 2000 počet narodených opäť nakrátko mierne vzrástol oproti trendu, čo pravdepodobne súvisí s magickým dátumom (1.1.2000).
- Hlboký pokles r. 2001 (o 7,3 % oproti r. 2000) začal už od novembra 2000. Za príčinu možno označiť súbor ekonomických opatrení schválených v decembri 1999 a januári 2000, menovite zvýšenie regulovaných cien palív a energie, nájomného, cestovného vo verejnej doprave atď.

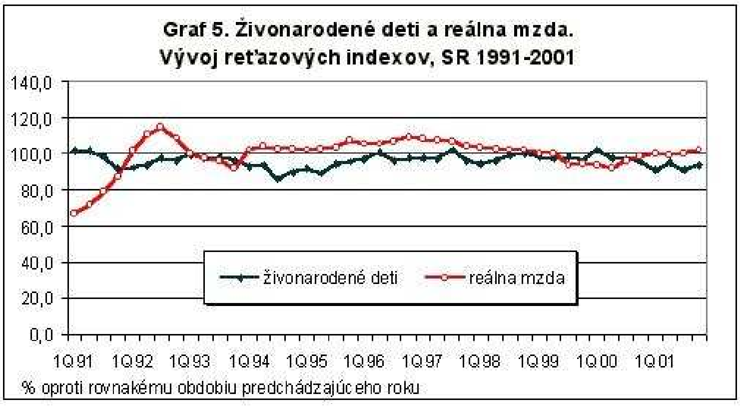
8. Závislosť od vývoja reálneho príjmu

Jestvuje viacero indícií, že pôrodnosť je ovplyvnená ekonomickou situáciou rodín, preto by bolo vhodné porovnať časový rad živonarodených s časovým radom relevantných ekonomických ukazovateľov, napr. s vývojom reálnej mzdy. Takéto rady sú však k dispozícii len ako štvrťročné (za obdobie od r. 1990), preto sme do tejto podoby upravili aj údaje o živonarodených.

Pri porovnávaní dvoch časových radov sa treba vyvarovať zdanlivej korelácii. Tá hrozí napr. vtedy, ak oba rady majú monotónny trend, alebo tiež vtedy, ak oba majú významnú vlastnú sezónnu zložku. Reálne mzdy sú najvyššie na konci roka a najnižšie na začiatku, zatiaľ čo pôrodnosť je najvyššia v lete, t.j. v treťom štvrťroku, a najnižšia v štvrtom. Preto miesto absolútnych hodnôt, resp. bázických indexov sme porovnávali reťazové indexy, počítané vždy oproti rovnakému obdobiu predchádzajúceho roku. Tým by mal byť vplyv zdanlivej korelácie odstránený. Takto vzniknuté štvrťročné rady reťazových indexov sú na grafe 5.

Vzájomná korelačná funkcia $r_k = r(y_t; x_{t-k})$ medzi príslušnými reťazovými indexami živonarodených y_t a reálnej mzdy x_t je na grafe 6. Sú tam vyznačené aj intervaly spoľahlivosti pre r_k za platnosti hypotézy o nezávislosti, pričom bola použitá aproximácia $s(r_k) = n^{-1/2}$. Vzájomná korelačná funkcia dosahuje (významné) lokálne maximum pre k = 3 ($r_3 = 0,44$) a k = 14 ($r_{14} = 0,65$) (t.j. 3/4 a 3 a pol roka). V prvom prípade to zodpovedá okamžitej reakcii, keď napr. po ekonomickom šoku ľudia okamžite revidujú svoje reprodukčné plány a správanie, vďaka čomu o 3/4 roka klesne pôrodnosť. Interpretácia druhého lokálneho maxima je náročnejšia. Jedna z možností je, že ide skôr o frekvenciu ekonomických šokov.

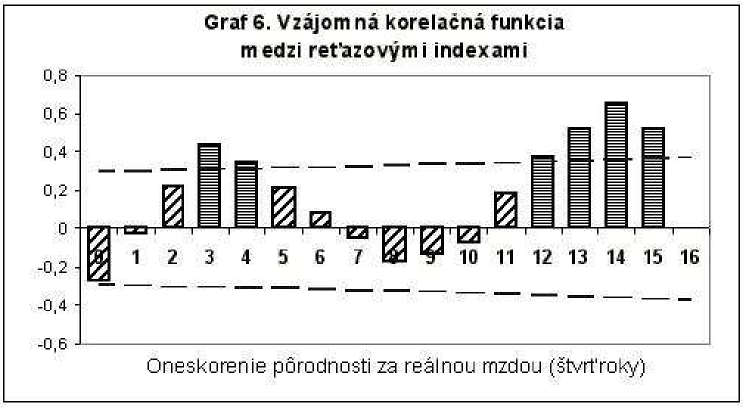
Podľa inej interpretácie na pokles reálnej mzdy reagujú ľudia okamžite, ale na vzrast len opatrne, s odstupom 2-3 rokov (čo sa prejaví zhruba o tri a pol roka).



Regresná priamka

$$\hat{y}_t = 0,1576x_{t-3} + 80,031$$

o.i. vyjadruje, že kontšantnému indexu reálnej mzdy zodpovedá asi 4,2 % medziročný úbytok pôrodnosti. Koeficient determinácie tejto rovnice (R^2 = 0,19) nie je veľký, to však neprekvapuje, pretože pôrodnosť závisí aj od iných faktorov. Nevýhodou tohoto modelu však je veľká autokorelácia rezíduí (DW = 0,9), čo je spôsobené prirodzenou zotrvačnosťou oboch premenných.



Lepšie štatistické vlastnosti má autoregresný model radu pre (štvrťročné) reťazové indexy živonarodených AR(1) resp AR(2). Zmeny v počte živonarodených najlepšie vysvetľuje zmiešaný model

$$\hat{y}_t = 31,36 + 0,5668\, y_{t-1} + 0,272\, x_{t-3} - 0,172\, x_{t-4},$$

kde R^2 = 0,38, DW = 2,2. Na udržanie pôrodnosti na rovnakej úrovni konštantná reálna mzda nestačí. Z uvedenej rovnice možno odvodiť, že pri dlhodobo nemennej reálnej mzde sa dosiahne stabilný medziročný úbytok narodených 4,5%. Náhly pokles reálnej mzdy o 10 % má za následok úbytok živonarodených o ďalších 2,72 % po tri štvrte roku. Analogické modely je možné zostaviť pre pôrodnosť časovo posunutú o 3 a pol roka po zmene reálnej mzdy.

Koeficient determinácie v skúmaných rovniciach nie je vysoký, preto závislosť pôrodnosti od priemerného reálneho príjmu by sa nemala preceňovať. To aj preto, že reálna mzda sa týka všetkých zamestnancov národného hospodárstva, a pôrodnosť iba vybranej časti. Naviac, pri dlhodobých úvahách už nemožno abstrahovať od vekovej štruktúry. V každom prípade však získané výsledky potvrdzujú hypotézu o vplyve ekonomických faktorov na úroveň pôrodnosti.

9. Záver

Na záver pripomíname, že zatiaľ čo miera mŕtvorodenosti nezávisí od sezóny, sezónne indexy pre rad živonarodených s maximom v júli a minimom v decembri sú vysoko významné a počas skúmaného obdobia sa nemenili. Je zaujímavé, že za rovnaké obdobie zistila Fialov [2] v ČR iný typ sezónnosti s vrcholom v marci - apríli. V prehľade [8] Lam a Miron uvádzajú, že podobný typ sezónnosti ako v ČR je typický aj pre severnú Európu, kým sezónnosť s vrcholom v lete (ako v SR) je typická pre južnú Európu. V USA naproti tomu pôrodnosť má vrchol v auguste - septembri. Rozdiely v sezónnom profile sa pripisujú najmä podnebiu. Otvorenou otázkou zostáva, či sezónne rozdiely sú ovplyvnené viac biologickými alebo sociálno-kultúrnymi faktormi.

Analýza rezíduí (nepravidelných výchyliek) potvrdzuje citlivosť pôrodnosti na vonkajšie podnety, ako sú napr. zmena prídavkov na deti, cenové šoky, politická neistota, zmeny populačnej klímy. Neistota v spoločnosti má odkladný účinok. Náhly pokles reálnych príjmov má za následok významný a pretrvávajúci pokles pôrodnosti. Na negatívne impulzy pôrodnosť reaguje takmer okamžite (s odstupom 9 mesiacov), ale na pozitívne len opatrne. Je teda jednoduchšie pôrodnosť znížiť ako zvýšiť.

Postup demonštrovaný v príspevku možno využiť aj pri výuke metód analýzy časových radov.

Literatura

[1] Cipra T. (1986) *Analýza časových řad s aplikacemi v ekonomii.* SNTL/Alfa, Praha.

[2] Fialová L. (1995) *Sezónnost demografických událostí v českých zemích v 17.-20. století*. Demografie **37(1)**, 9–21.

[3] Hindls R., Kaňoková J. a Novák I. (1997) *Metody statistické analýzy pro ekonomy*. Managment Press, Praha.

[4] Chajdiak J. (2002) *Vývoj hospodárstva SR do júla 2002*. In: *Zborník príspevkov Štatistické metódy v praxi. 11. slovenská štatistická konferencia SŠDS, 11.–13. 9. 2002 Nitra*. SŠDS, Bratislava, 101–107.

[5] Chajdiak J., Komorník J. a Komorníková M. (1999) *Štatistické metódy*. STATIS, Bratislava.

[6] Koubek J. (1990) *Vliv populační politiky na plodnost v Československu*. Demografie **32(3)**, 193–203.

[7] Kozák J., Hindls R. a Hronová S. (2000) *About new concepts of seasonal adjustment of time series*. Acta Oeconomica **6**, Applications of Mathematics and Statistics in Economy, EFUMB Banská Bystrica, 67–79.

[8] Lam D. a Miron J.A. (1994) *Global Patterns of Seasonal Variation in Human Fertility*. Human Reproductive Ecology **709** of Annals of the New York Academy of Science, 9–28.

[9] Pastor K. (2000) *Analýza sezonality v demografických časových radoch - pôrodnosť.* In: *Zborník príspevkov Štatistické metódy v praxi. 11. slovenská štatistická konferencia SŠDS, 11.–13. 9. 2002 Nitra*. SŠDS, Bratislava, 191–195.

[10] Pastor K. (2000) *Economic Determinants of Fertility Changes, Slovak Republic Data 1990-1996*. Acta Oeconomica **6**, Applications of Mathematics and Statistics in Economy, EFUMB Banská Bystrica, 133–142.

PATŘÍ STATISTIKA JEN NA VYBRANÉ ŠKOLY?

Vladimír Rytíř

FaME UTB, Mostní 5139, 760 01 Zlín

vladimir.rytir@email.cz

1. Úvodem

K mému krátkému zamyšlení mne vedl nejen článek pana Hendla z Bulletinu ČStS *NCTM standardy jako výzva pro Českou statistickou společnost* [1] či již dávno vyřčená poznámka pana Hindlse, že „středoškolák neví, co je to statistika“, ale i osobní praktické zkušenosti s výukou této discipliny.

2. Otázky pilotního průzkumu

Co se týká výuky statistiky na středních školách, provedl jsem pilotní průzkum na VOŠ v Uherském Hradišti, Fakultě managementu a ekonomiky UTB ve Zlíně a VŠE v Praze. Otázky zněly:

1. Jaký byl typ, název a místo Vaší střední školy, kterou jste navštěvoval(a)?
2. Probíhala na této škole výuka statistiky?
3. Pokud ano, v jakém rozsahu?
4. Byla statistika vyučována jako samostatná disciplina či v rámci jiných předmětů?

3. Dílčí výsledky

Z odpovědí 71 studentů VOŠ v Uherském Hradišti vyplynulo, že 47 studentů nemělo statistiku vůbec, 11 pouze velmi povrchně v rámci jiných předmětů (matematika, účetnictví) a zbytek – tj. 13 – mělo statistiku po dobu jednoho roku. Se znalostmi matematiky jsou na tom studenti této VOŠ obdobně.

O mnoho lépe na tom ale nebyli ani studenti ostatních dvou škol. Jen ve stručnosti:

- na FaME UTB ve Zlíně nemělo statistiku v předchozích vzdělávacích stupních z 93 studentů vůbec 41, pouze omezeně 25 a zbytek 27 jako samostatný předmět;

Klíčová slova: Kde a jak se učí statistika.

Poděkování: Rád bych poděkoval všem, kteří mi umožnili konat mé „průzkumy“.

- na VŠE ze 353 studentů vůbec 222, pouze omezeně 36 a 95 jako samostatný předmět.

4. Závěry z dílčích výsledků?

Největší nedostatky vidím ve výuce – a to nejen statistiky – na základních a středních školách, chybné práci různých akreditačních komisí, tvorbě „osnov“ a jejich následném dogmatickém plnění a o slabé úrovni pedagogů.

Absolvovaná škola	Bez statistiky	V jiném kurzu	Samostatný kurz	DÍLČÍ SOUČTY
VOŠ	47	11	13	71
FaME UTB	41	25	27	93
VŠE	222	36	95	353
CELKEM	310	72	135	517

Tabulka 1. Pilotní průzkum na vybraných školách.

Je překvapivé, že ne každé gymnázium či obchodní akademie mají ve svých učebních plánech statistiku zařazenou. Je proto podle mě naivní si myslet, že během studia na vysoké škole student dohoní vše, co bylo zanedbáno na nižších stupních.

Literatura

[1] Hendl J. (2002) *NCTM standardy jako výzva pro Českou statistickou společnost*. Informační Bulletin České statistické společnosti **13(1)**, Praha, ISSN 1210-8022.

PRAVDĚPODOBNOST VE STAROVĚKU A STŘEDOVĚKU

Ivan Saxl

Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, Katedra pravděpodobnosti a matematické statistiky, Sokolovská 83, 186 75 Praha 8 – Karlín
ivan.saxl@mff.cuni.cz

1. Empirické počátky

Počátek teorie pravděpodobnosti je obecně spojován se jmény Blaise Pascala a Pierra Fermata, kteří v roce 1654 na popud Gombauda (nebo Gombaulda) rytíře de Méré řešili ve své korespondenci jisté problémy týkající se hry v kostky. Jak tomu však obvykle bývá, zakladatel vědecké disciplíny či objevitel klíčového poznatku má řadu předchůdců a předřešitelů. Je potom dovršitelem určité etapy poznání, spíše pokrývačem zastřešujícím existující stavbu, než kopáčem jejích základů. Proto je rozumné se ptát, jaké byly skutečné počátky teorie pravděpodobnosti a hledat je v dobách, kdy vznikala jiná odvětví matematiky, jako geometrie, algebra, numerická matematika. M.G. Kendall, jeden z předních vědců minulého století v oboru pravděpodobnosti a statistiky, píše: „*Mathematics never leads thought, but only expresses it*“. Z jakého poznání se tedy odvozovala teorie pravděpodobnosti a proč se tak opozdilo, že k jeho vyjádření dochází až v době, kdy jiné oblasti matematiky jsou již značně rozvinuté? Přitom právě nejistota provází skoro všechny naše činnosti, uskutečnění jejichž cílů je nanejvýš možné, nikdy však jisté. Těmto otázkám je i v současnosti věnována poměrně velká pozornost a odpovědi na ně nejsou ani zdaleka hotové.

Počátky pravděpodobnosti jako empirické vědy možná spíše hledáme než nalézáme v hrách, z nichž patrně nejstarší jsou hry v kostky. Prvním hracím nástrojem byla zřejmě hlezenní kost (*astragalus*) kopytnatců, viz Obrázek 1a, která je dosti nepravidelným šestistěnem a byvši hozena, může zaujmout čtyři možné polohy. Ve vykopávkách z doby před 40 000 lety je počet nalezených hlezenních kůstek v lidských sídlištích nápadně vysoký, z čehož archeologové usuzují, že byly používány ke hrám.

Nejstarším typem hry možná byla pouhá ekvilibristika spočívající v nadhazování a chytání kůstek hřbetem ruky. Na egyptských malbách z doby I. dynastie (3500 let př. Kr.) se kostka objevuje již jako nástroj v deskových hrách spočívajících

Klíčová slova: Historie pojmu pravděpodobnosti, hry v kostky, karetní hry, náhoda ve věštbách, pravděpodobnost v bibli a talmudu.

Poděkování: Tato práce vznikla za podpory grantu MSM 113200008 a grantu GAČR č. 201/03/0946.

v posouvání figurek. Dochovaly se i celé hrací soupravy, viz Obrázek 1b, pro hry senet a Psi a šakali, viz Obrázky 2-3, (Egypt, 1900 př. Kr.) a pro tzv. Královskou hru z Uru (hra nalezená v sumerském městě Uru se datuje zhruba do 2650 př. Kr., kostky jsou ve tvaru pyramid). Senet i Psi a šakali jsou obdobou vrhcábů, které se z nich patrně vyvinuly; k určení počtu kroků pěšců se používaly buď hlezenní kůstky nebo tyčinky s jednou označenou stranou. Nejvyšší počet bodů získával hod, při němž padly všechny (čtyři?) značené strany. O pravidlech her existují pouze dohady.

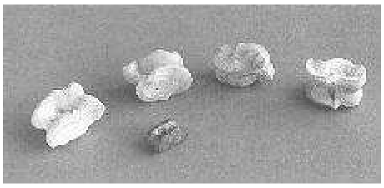
Řekové předpokládali, že kostky ($\pi\varepsilon\sigma\sigma\omega\iota$, hra sama pak $\pi\varepsilon\sigma\sigma\varepsilon\iota\alpha$) mají své počátky v Řecku. Za vynálezce byl považován jeden z hrdinů trojské války Palomédés[1]. V Iliadě sice není explicitně zmiňován, kostky samotné však ano, a to v souvislosti s Patroklem, který údajně jako chlapec málem v hněvu zabil svého spoluhráče v kostkách. V této době již kostky patrně byly používány ve hře sázkového typu o peníze. Za uvedení stojí kostkám věnovaný úryvek z Hérodota (kniha 1, odst. 94):

„*Lýdové se řídí podobnými obyčeji jako Řekové, kromě toho že prostituují své děti ženského pohlaví. První z lidí, pokud víme, používali mincí ražených ze zlata a stříbra a první začali pěstovat obchod. Lýdové sami tvrdí, že jsou jejich vynálezem hry, které se nyní pěstují u nich i u Řeků. vynašli prý je v té době, kdy kolonizovali Tyrsénii. Vypravují o tom toto: Za časů krále Atya, syna Manova, nastala po celé Lydii veliká nouze o potraviny. Lýdové zprvu trpělivě vyčkávali, a když neustávala, hledali nějakou odpomoc. Každý z nich prý vymýšlel něco jiného. A tak tehdy vynašli hru v kostky, vrhcáby, s míčem a všechny jiné hry kromě pessů, jejichž vynález si Lýdové nepřivlastňují. Když hry vynalezli, používali je proti hladu tak, že jeden den hráli od rána do večera, aby nemohli hledat potravu, a druhý den přestali hrát a jedli. Takovým způsobem to vydrželi osmnáct let*“.

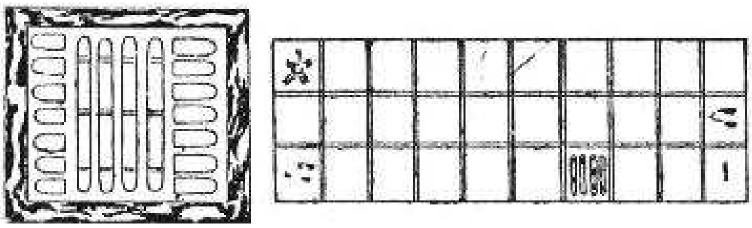
Jednotlivé strany kůstek měly bodové hodnoty 1, 3, 4, 6, (pravděpodobnosti hodu podle dochovaných napodobenin jsou 1/10 pro 1 a 6 a 4/10 pro strany zbývající). Jedničce se říkalo „pes“ a figura 1, 3, 4, 6 při hodu čtyřmi kostkami se nazývala Venušin vrh či Venuše (pravděpodobnost přibl. 1/26). O Euripidovi se tradovalo, že hodil „čtyřicítku“ – patrně čtyři čtyřky (pravděpodobnost přibližně 1/39).

Římské kostky již byly kamenné, často zdobené; kamenné kostky však byly nalezeny i v Egyptě a keramické v Indii. Z Egypta jsou i kostky opatřené reliéfy Osirise, Isidy a dalších bohů; byly zřejmě užívny k liturgickým účelům. K obecnému rozšíření hry v kostky dochází až za Ptolemaiovců (kolem 300 př. Kr.).

[1] *Palomédés, syn Naupliův*, universální řecký vynálezce, kromě kostek (určených k zabavení před Trójou se nudících vojáků) vynalezl také maják, míry a váhy, stavění stanů a skoro celou alfabetu (sudičky pouze β, τ a samohlásky), někdy ovšem leccos z toho i Hermes. Později trochu splývá s Prométheem. V trojské válce byl Řeky ukamenován díky lsti Odysseově a jeho otec, poté co se marně dožadoval omluvy za synovu smrt, navštěvoval manželky u Tróje pobývajících válečníků a přemlouval je k nevěře.



Obr 1a. Hlezenní kůstky, vpředu bronzová kopie.
Obr 1b. Hrací stolek pro senet z hrobu faraóna Tutanchamona (1347 - 1339 př. Kr.)



Obr 2. Deska z doby III.dynastie (2600 - 2700 př.Kr) s pěšci a tyčinkami.

Obr 3. Moderní kopie hracích plánků senetu a Psů a šakalů.

K produkování náhody se používaly také tyčinky (délky kolem 7 cm a průřezu 1 cm^2) s čísly 1, 2, 5, 6, řídčeji též 3, 4 nebo s odpovídajícím počtem vrypů, a to v Anglii kolem počátku našeho letopočtu a též u Mayů. Tacitus v *Germanii* vedle zmínky o velké náklonnosti jejích obyvatel ke hře v kostky také popisuje použití tyčinek opatřených runami k věštění.

Zvýšení počtu hracích poloh se nejprve dosahuje zbroušením hlezenních kůstek (ty se však potom rychle obehrály) nebo také nařezáním jiných zvířecích kostí, vynětím morku a obroušením do přibližně krychlového tvaru: takové kostky pak byly duté. Kostky jiných tvarů jsou spíše kuriozity. Z Řecka pochází kostka – pravidelný dvacetistěn, nalezeny byly také čtrnáctistěny, osmnáctistěny a devatenáctistěny. Číslování kostek bylo zpočátku proměnlivé, ale kolem roku 1400 př. Kr. v Egyptě se ustálilo na dnešním standardu se součtem 7 na protilehlých stěnách.

O popularitě kostek v Řecku svědčí to, že byly častým motivem umělecké tvorby, jak ukazuje slavná Sódatésova kůstka, viz Obrázek 4, i keramika dvou žen hrajících v kostky, viz Obrázek 5. Dostaly se dokonce i na platidla, jak je vidět na minci korintské.

Rozšíření hry v kostky v Římské říši za vlády císařů bylo značné: dosvědčují to i nástěnné mozaiky hráčů nalezené na stěnách domů v Pompejích, viz Obrázek 6. Suetonius ve svých *Životech římských císařů* píše o Božském Augustovi:

„*Výtky kostkářství se nijak nelekal: hrál v kostky prostě a nepokrytě pro své rozptýlení ještě jako stařec, a nejenom v měsíci prosinci, nýbrž i v jiné sváteční i všední dny*“.

Tamtéž z dopisu Tiberiovi:

„*Povečeřel jsem, milý Tiberie, s obvyklými společníky... Při večeři jsme hráli, jak se na staré pány sluší, včera i dnes. Pro házení kostek jsme stanovili, že kdo hodí psa nebo šestku, přispěje do banku za každou kostku po denáru, celý bank že připadne tomu, kdo hodí Venuši*“.

Z Augustova dopisu dceři:

„*Posílám ti dvě stě padesát denárů, tolik, kolik jsem rozdal každému ze spoluhodovníků, chtěli-li by si spolu při večeři zahrát v kostky nebo sudou-lichou*“.

O Božském Claudiovi píše Suetonius takto:

„*V kostky hrál nesmírně horlivě. Napsal dokonce knihu o této dovednosti. Hrával na projížďkách, maje vůz i hrací desku tak upraveny, aby nevznikl při hře zmatek*“.

Jiné prameny pak sdělují, že s oblibou hrával levou rukou proti pravé.

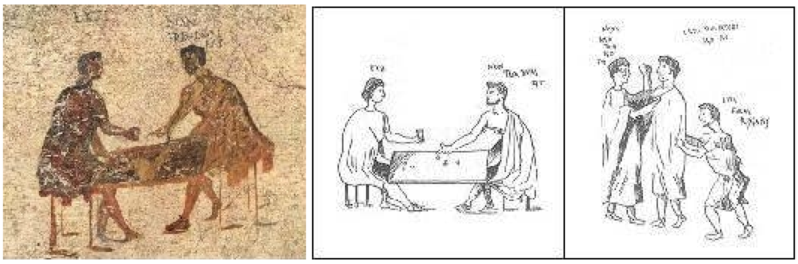
U Lukiana ve fiktivní besedě Króna, Diova svrženého otce, s jeho knězem, bůh na otázku co je v jeho moci, když jinak všechno řídí Zeus, odpoví:

„*Věci opravdu nemalé. Ledaže se ti zdá být nicotné mít zdar v kostkách, aby tobě ukazovaly vždy šestku, když ostatním padá jednička*“.

O něco dále v Krónových zákonech je nařízeno: „*Všichni ať hrají o ořechy; bude-li někdo hrát o peníze, ať je druhý den bez jídla*“.

Obrázek 4. Sódatésova kůstka, řecká keramika z 3. stol. př. Kr.

Obrázek 5. Ženy hrající v kostky (keramika kolem 300 př. Kr.) a korintská mince (6. Stol. př. Kr.)



Obrázek 6. Freska z Pompejí a schématické znázornění dvou dalších fresek ze stejného místa s nápisy ve slangové latině: „*Exsi*“ a „*Non tria dvas est*“ [Končím. To nejsou tři, ale dvě.] na první, „*Noxsii amii tria iigo fui*“, „*Urtii. Pillatorii hgo tui*“ a barmanovo „*Itis foras rixsatis*“ na druhé [To mě urážíš, příteli, jsou to tři. Podvodníci jako ty si zaslouží ránu. Vypadněte, když se chcete prát.]. Jedná se patrně o vtip, protože hráči hrají třemi kostkami.

U Lukiana najdeme také vyprávění o mladíkovi, který se zamiloval do Praxitelovy Venuše:

„*Když se mu podařil dobrý vrh, zvláště když mu padl obraz samotné bohyně, žádné dvě kostky stejné, zbožňoval bohyni a dělal si naději, že jeho vášeň bude* *odměněna. Když hodil špatně, proklínal všechny kolem a byl tak zdrcen, jakoby utrpěl těžkou osobní ztrátu*".

Ještě úryvek z Propertia:

„*Když jsem potřeboval Venuši, vždy ti zatracení psi se na mne lepili*".

Konečně také Iuvenalis (I. satira):

Kdy s takovou vášní se hazardně hrálo?
K stolům, kde v kostky se hraje, se nejde s hubeným měšcem;
má-li se opravdu hrát, to přistaví celičkou truhlu.
K jakým tu dochází bitvám, když účetní převezme roli
zbrojnoše. Je to snad bláznovství pouhé, když sto tisíc ve hře
promarní člověk, jenž otroku v zimě chce tuniku upřít?

Silně rozšířena byla hra v kostky také ve starověké Indii. Nepoužívaly se však hlezenní kůstky, ale oříšky *vibhidaka* – plody stromu *Terminalia bellerica*. Od prohry v kostkách se odvíjí hlavní dějová linie eposu *Mahábharáta* (písemná forma ze 4. stol. po Kr., předpokládaná doba vzniku od 4. stol př. Kr.) a kostky se objevují i ve vedlejším příběhu o králi Nálovi, v němž je zmíněno ještě výběrové šetření. Přitom herní um je prezentován jako početní znalost. Hra v kostky je také námětem několika vyprávění v Somadévově (1035–1085) Oceánu příběhů, kde je rovněž zmíněna existence veřejných heren.

Nikde však nebyla nalezena žádná explicitní zmínka o relativní četnosti vrhů určitých čísel či jejich kombinací, naopak bodování na hlezenních kůstkách přiřazuje nejnižší bodovou hodnotu vrhu s nízkou relativní četností. Stejnou četnost všech vrhů krychlové kostky patrně také nikdo nepředpokládal, resp. o tom není v literatuře žádná zmínka. Jako důvod bývá někdy uváděno, že kostky byly nedokonalé a proto rovnost četností nemohla být zpozorována. Testování dochovaných kostek to však přesvědčivě nepotvrdilo a navíc i naše současné kostky mají k dokonalosti také hodně daleko. Nedokonalý tvar patrně nebyl příčinou neuvažování o četnostech bodových hodnot.

2. Vztah společnosti ke hrám, církevní a státní zákazy

Patrně nejstarší svědectví negativního poměru společnosti ke hrám v kostky lze najít v Rgvédu, nejstarším z védských textů, obsahující nářek hráče v kostky, který svou vášní všechno zničil; v závěru mu sluneční bůh přikazuje, aby se napravil. V Bibli se kritika hráčských vášní nevyskytuje, byly totiž explicitně zakázány a hra v kostky se rozšířila až v římských dobách; proto je námětem některých pasáží v talmudu (viz níže).

V Evropě se hra v kostky udržela v masové oblibě od římských dob až do renesance, kdy byla zčásti vytlačena kartami. Ke kritice a zákazům však docházelo jak ze strany církve, tak i státu. Římská republika hraní kostek zákonem omezila na dobu Saturnálií (kolem našich vánoc, nepracovali ani otroci, dávaly se dárky). Sv. Cyprián z Kartága pronáší kolem roku 240 kázání „*De Aleatoribus*" a další slavné kázání „*Contra Alearum Ludos*" je z roku 1423 od sv. Bernarda ze Sieny. Císař Fridrich II. vydává r. 1232 „*lex de aleatoribus*" a svatý Ludvík (IX.) zakazuje svým úředníkům nejen kostky, ale i šachy, návštěvu hostinců a smilstvo – vše v jedné řadě – a kostky se v celém svém království nesmějí ani vyrábět. Sálský císař Otto I. v r. 952, trevírský koncil v letech 1227 a 1238 i koncil ve Worcesteru v r. 1240, všichni vydávají zákaz hraní kostek duchovními.

Je ovšem třeba zdůraznit, že všechny tyto zákazy se vztahovaly na hru o peníze a měly na zřeteli s ní související průvodní jevy. Ty jsou explicitně vyjmenovány v Povídce odpustkáře v Chaucerových *Canterburských povídkách*:

Kristovy hřeby! Pro Kristovy údy
a jeho krev a srdce, u všech všudy.
Máš tři a pět, a sedm padlo mně.
Kristova muko! Hraješ-li falešně
já do srdce ti vrazím tuhle dýku.
Hle, to jsou plody kostek hazardníků:
klam, zloba, vražda, proklínání kleté.

Nejednalo se tedy o hříšnost zkoušení náhody či zaměstnávání Prozřetelnosti rozhodováním o nedůstojných záležitostech. Však také – patrně s vědomím marnosti potlačovat nudu a hravost – vymýšlí biskup Wibold z Cambray kolem r. 960 hru v kostky pro duchovní, v níž lze „vyhrávat" ctnosti. Křižáci účastnící se III. křížové výpravy již mají v předpisech přímo vymezena pravidla pro hraní kostek, neboť nudící se voják by snadno mohl podlehnout daleko horším neřestem: bojovníci nižší úrovně než rytíři nesměli hrát o peníze vůbec a rytíři ani duchovní nesměli prohrát více než 20 šilinků za 24 hodin.

Ocenění her jako prostředku rozptýlení lze najít také v již zmíněném Chaucerovi v *Povídce fráterově*. Dorigena se trápí nepřítomností svého manžela Arviraga válčícího v Anglii:

Přátelé . . . i jinam vodili ji podél řek
a k zábavám u tichých studánek,
do libých míst, kde nebyla sama,
kde tančili a hrál se šach a dáma.

Ještě úryvek ze zprávy poslané Margery (24. 12. 1484) manželu Johnu Pastonovi z *Pastonských dopisů* (korespondence norfolkské rodiny z let 1420 až 1503, bohatý zdroj informací z období válek dvou Růží):

Necht' Vás potěší, že jsem poslala Vašeho nejstaršího syna k lady Morleyové, aby se dověděl jakými činnostmi pro ukrácení chvíle se v jejím domě bavili o prvních vánocích po smrti lorda jejího manžela; a ona mi řekla, že nebyly žádné maškarády *ani hlučné rozpravy, ale pouze hry u stolu, vrhcáby a šachy a karty; jen tyhle zábavy svým lidem dovolila a žádné jiné.*

Série zákonů z počátku vlády Eduarda III., zakazující některé hry, k nimž Jindřich VIII. přidal i kostky a karty, již neměly za cíl zlepšení mravů, ale měly obrátit pozornost mužů k mužným sportům, zvláště pak bojového charakteru. Tedy nikoliv kuželky, biliár a kostky, nýbrž lukostřelba byla povolenou a králem žádanou zábavou.

Ze všech výše citovaných zdrojů však bohužel získáme jen velmi mlhavou představu o pravidlech těchto tu zakazovaných, jinde doporučovaných her. Římané hráli obvykle se čtyřmi hlezenními kůstkami, se třemi a méně často se dvěma šestibokými kostkami. Dvě kostky se však vyskytují v citovaném úryvku ze Chaucera a lze na ně najít odkazy i jinde. Montmort[2] ve své knize z roku 1708 píše o hře zvané „le quinquenove" rošířené ve Francii a hře zvané „hazard" běžné pouze v Anglii; obě hry se hrály pouze s dvěma kostkami. Hry však asi byly podobné: základem nejspíš bylo hození tzv. sázky (main), tj. čísla mezi 5 a 9. Při úspěšném hodu hráč pokračoval ve hře, při neúspěchu hrál další. Větší bere bylo patrně pravidlem. Slovo „hazard" bylo s velkou pravděpodobností dovezeno do Evropy křižáky. Jeho etymologie je dosud předmětem sporů. Jedno vysvětlení je odvozuje od *al zhar*, což je kostka, druhé od *asar*, značící obtížné. Další hypotéza odvozuje slovo ze jména syrské pevnosti El Azar (Hazait, Hazar), o níž se v souvislosti s hrou v kostky zmiňuje Godefroy de Bouillon[3]. Slovo se začalo používat ve dvanáctém až třináctém století, nejstarší nalezené použití literární je ve Waceově[4] *Roman de Brut* (1154) a v *Erekovi a Enidě* Chrétienna de Troyes (asi 1160).

3. Počátky kombinatorické pravděpodobnosti a náhoda ve věštbách

Kombinatorika se objevila daleko dříve v Asii, nejstaršími dosud nalezenými a spolehlivými prameny jsou především sútry. *Bhagabati Sútra* (kolem 300 př. Kr.) obsahuje počty permutací k prvků z n pro $k = 1, 2, 3$ a stejně tak počty kombinací. Početní operace jsou použity k řešení problémů typu „jaký je počet filosofických soustav, jež lze zformulovat kombinací daného počtu základních axiomů", „jaký je

[2] *Pierre Rémond de Montmort* (1678–1719), francouzský matematik, známá je kniha *Essay d'analyse sur les jeux de hazard* (1708, rozšířené vydání 1713), v níž řeší řadu kombinatorických a pravděpodobnostních úloh.

[3] *Godefroy de Bouillon* (1061–1110), vůdce první křižácké výpravy a kníže jeruzalémský. Zmíněný text zní: *A Hazait s'en ala ung riche mendement, et l'apiel-on Hazait pour le fait proprement que ly dés fu fais et poins premierement.* [Do Hazait jelo skvělé poselstvo a nazývá se Hazait právě proto, že tam byly původně vyráběny a tečkami značeny kostky.]

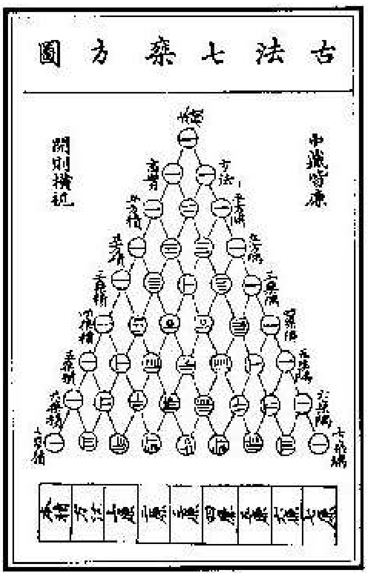
[4] *Wace z Jersey* (†po 1171), básnicky zpracoval „*Historia regum Britannniae*" od Geoffrey Monmoutha (1100?–1154), popisující legendární osídlení britských ostrovů Aeneovým pravnukem Brutem a představující hlavní zdroj zromantizované artušovské legendy.

počet smyslových podskupin sestavitelných z pěti smyslů“ či „jaké podsoubory lze vytvořit z daného počtu mužů a žen“. Pingala se v *Čanda Sútře* (3. stol. př. Kr.) zabývá prosodií, určuje počet vznikajících kombinací různých písmen a explicitně popisuje konstrukci Pascalova trojúhelníku (meruprastara). *Sthananga Sútra* (kolem 2. stol. př. Kr) jmenuje permutace a kombinace mezi dlouho studovanými problematikami, vedle teorie čísel, aritmetiky, geometrie, zlomků a rovnic.

Binomická čísla $C_r(n)$ byla tedy v Indii známá již velmi dávno, podle autora z 10. stol. byl znám i jejich součet $(1+1)^n = \sum_{r=0}^{n} C_r(n) = 2^n$. U Bhaskary II (1114–1185) se vyskytují permutace s opakováním i bez něj a rovněž odpovídající kombinace.

Kombinatorika se patrně vyskytuje i ve védách (12. až 9. stol. př. Kr.), konkrétně v jejich čtvrtém díle – *Atharvavédě*, jež obsahuje především magická zaříkadla. Interpretace védských sbírek je však obtížná a není vždy zřejmé, nakolik je oprávněná.

V Číně se Pascalův trojúhelník z binomických čísel až do $n = 8$ vyskytuje v knize *Jaspisové zrcadlo čtyř prvků* od Ču Š’ťia (asi 1270 – asi 1330), viz Obrázek 7, a do $n = 6$ u Ťia Sien (kolem 1100), vždy v souvislosti s algoritmy pro odmocňování. Rovněž v arabské matematice 12. a 13. stol. byla většina kombinatorických pravidel již známa a používána v různých odvětvích matematiky, Pascalův trojúhelník patrně znal již al-Karadži (953–1029).



Obrázek 7. Pascalův trojúhelník v knize *Jaspisové zrcadlo čtyř prvků* (1303) a titulní strana *Le Sorti*.

V Evropě znal dvouprvkové kombinace údajně již Boethius (475–525), s prvním systematickým kombinatorickým výpočtem se se však trochu kuriózně setkáváme až u již výše jmenovaného Wibolda z Cambray, v jeho zbožné hře popsané kronikářem Baldericem[5] v 11. stol. Mnich hodil třemi kostkami a dostal jednu z kombinací s opakováním, kterých je $C'_3(6) = 56$. Tím si vylosoval jednu z 56 ctností Wiboldem pečlivě vypočtených a očíslovaných, kterou pak musel po 24 hodiny praktikovat.

Další příklad správného výpočtu počtu kombinací s opakováním pro $r = 2$ se objevuje v již zmíněném kázání sv. Bernarda ze Sieny, kde je věta: „. . . *a jako je tento misál sestaven z 21 písmen, tak ve hře v kostky je 21 hodů*“. Kazatel má na mysli hod dvěma kostkami, tedy počet možných kombinací je $C'_2(6) = 21$.

Správná hodnota $C'_2(6)$ se vyskytuje v některých anglických středověkých písních. *Chaunce of the Dyse* má 56 slok, každá odpovídá jedné kombinaci s opakováním tří čísel ze šesti, na př. pro 6, 5 a 3 uvádí:

Merkur jenž dává výmluvnost
Ti při narození byl tak nakloněn
Že velký díl ti dal umění
Lidská srdce si podrobit
A o všech věcech správně promluvit.
Proto svými slovy vládni tak moudře,
jak srdce si může přát a jazyk vyslovit.

A z jiné písně na 6, 5, 3:

Máš šest, pět, tři,
Tvé přání se může vyplnit.
Však má-li se to vskutku stát,
ve dne v noci se zločinu střez.

Písně byly užívány k pokoutnímu věštění. Hod čtyřmi kostkami byl běžně užíván v řeckých i římských chrámech, pes byl nepříznivý, Venuše naopak, pro ostatní případy byly sestaveny tabulky. Každý hod byl pojmenován jménem některého boha. Výjimečně se užívalo i pěti kostek. Podle překladu Jamese Frazera (v komentáři k Pausaniovým *Cestám po Řecku*) zní nápis na stěně chrámu v Malé Asii:

1.3.3.4.4=15 Hod Zachránce Dia
Jedna jednička, dvě trojky, dvě čtyřky
Záměr, o němž přemýšlíš vyplň směle
Jdi do toho, bohové ti dávají příznivé znamení
Nevzdávej se. Žádná smůla se tě nedotkne.

[5] *Balderic le Rouge* (†1097), biskup v Noyonu a Tournai, autor latinské kroniky *Chronicon Cameracense et Atrebatense* (Kronika Cambrai a Arrasu), zahrnující celé merovejské i karolinské období a končící až v roce 1090.

Není zřejmé, zda na pořadí čísel záleželo. Pokud ne, pak pravděpodobnost této číselné kombinace je přibližně 0.08 pro hlezenní kůstky, užívány však byly i kostky.

Tažení losu za účelem věštby se vyskytovalo i v jiných podobách. Jak již bylo zmíněno, Tacitus popisuje věštění u Germánů v knize *Germania*: početné nařezané větvičky byly popsány runami, pak naházeny na bílou látku. Kněz po modlitbě vybral tři z nich a vyložil význam takto vylosovaných run.

Populární bylo také věštění z Vergilia: *Aeneida* byla otevřena na náhodné stránce, poslepu vybrán řádek a interpretován. M. G. Kendall připomíná (patrně apokryfní) historku z moderních dob. Margaret Bonfieldová, první žena v britské vládě (1928), se údajně radila s Vergiliem před obnovením podpor v nezaměstnanosti a nalezla slavnou větu, jíž Aeneas začíná Didoně líčit útrapy trojské války: *„Infandum, regina, iubes renovare dolorem. . .“* [Nevýslovný bol, královno, obnoviti velíš. . .]. Ostatně u křesťanů byla a bezpochyby dosud je k podobným účelům používána bible.

Všechny tyto pochybné obřady se přenesly i do středověku, přestože církev od časů Konstantina Velikého podobné praktiky zakazovala. Kronikář Balderic zmiňuje řešení sporu mezi biskupy z Autunu, Arrasu a Poitiers o tělo sv. Leodegara (616–678) právě losem (údajně vyhrál biskup z Poitiers, dnes však jsou ostatky tohoto poněkud sporného světce uloženy v St. Maixent a v Ebreuil). Také John Wesley (1703–1791), zakladatel metodistické církve, popisuje podrobně ve svém *Deníku*, jak se spolu s přítelem – po modlitbách a dlouhém rozvažování – vyhledal pomoc Prozřetelnosti v záležitosti své perspektivní ženitby tažením lístků s nápisy „*Ožeň se*“, „*Letos ne*“ a „*Už na to nemysli*“ (po vytažení třetího lístku s úlevou pronesl „*Tvá vůle se staň!*“).

K věštění byly také používány karty. Francesco Marcolini v r. 1540 publikuje *Le Sorti* [Kniha osudu], viz Obrázek 7, v níž si každý s balíčkem karet může najít svůj osud. Na 50 otázek (zčásti společných pro obě pohlaví, zčásti rozlišených) je připraveno 4500 odpovědí-předpovědí-rad, v nichž pomocí devíti karet si každý najde tu sobě souzenou. Kdo není zvědavý na osud, může pomocí kostek také skládat hudbu. Bylo vydáno několik knih, na nichž se snad podíleli také C. P. E. Bach, W. A. Mozart a J. Haydn, a které umožňovaly pomocí dvou kostek „skládat“ dobové tance. Zřejmě se jednalo o systémy na sebe vhodně navazujících tanečních úryvků. Ve všech podobných zábavách ovšem po pravděpodobnosti není ani stopy.

Pokud se jedná o výše uvedená kombinatorická čísla, správných hodnot bylo nejspíš dosaženo pozorným vypsáním všech kombinací a jejich sečtením. S prvním odůvodněním se setkáváme v básni *De vetula* [0 stařence], původně připisované Ovidiovi, dnes shoda na Richardovi de Fournival (1190–1260), humanistovi a kancléři katedrály v Amiensu. V básni je pasáž o kostkách a jsou počítány permutace tří hodů, viz Obrázek 8. Odpovídající pasáž zhruba obsahuje následující výklad:

„*Když všechna čísla jsou stejná, je možností 6, když dvě jsou stejná a třetí jiné, je případů 30, protože každý pár může být zvolen šesti způsoby a třetí číslo pěti, a* *když jsou tři různá, možností je 20, protože třicetkrát 4 je 120, ale každá možnost může vyniknout šesti způsoby. To dává 56 možností. Ale když jsou všechna stejná, pak každé číslo je možné jen jednou, když jsou dvě stejná a jedno jiné, pak jsou možné tři způsoby a když jsou všechna různá, způsobů je 6.*“

Převedeno do dnešního zápisu: nejprve se spočítá kombinace $C_1(6) = 6$, pak $C_2(6) = 15$ a násobíme dvěma, protože dvakrát se musí vyskytovat první i druhé číslo z vybrané dvojice – takže 30, konečně $C_3(6) = 20$. A nyní v druhé a třetí skupině provedeme permutace $P'_{r,s,t}(3)$, tj. $P'_3(3) = 1$, $P'_{1,2}(3) = 3$ a $P'_{1,1,1}(3) = 6$. Po sečtení dostáváme správný výsledek 216. Báseň obsahuje také tabulku, v níž jsou vypočteny četnosti trojic dávajících stejné součty – těch je celkem 16, počínaje 3 a konče 18, a u každého součtu je pak uvedena jeho četnost. Zábavné je, že báseň byla přeložena do francouzštiny H. Cocherisem a vydána v roce 1862, překladatel však odvození zřejmě neporozuměl: uvádí, že možných součtů tří čísel je celkem 16 a že se vyskytují v různých počtech, vlastní zdůvodnění však v textu vynechal.

Stařenka tedy už zřejmě chápe smysl binomických koeficientů a umí s nimi zacházet, problém počtu, kolikrát se který součet může vyskytovat, je vyřešen a pravděpodobnosti výskytu by bylo možno určit. A přece se tak nestalo. Zhruba o 200 let později se s méně jistým pochopením lze setkat v komentáři k Dantovu *Očistci*, k veršům ze 6. zpěvu (překlad O. F. Bablera):

Když pak po hře v kostky zástup rozejde se,
ten, který prohrál, zbude rozmrzelý
a znovu zkouší, kostkou v ruce třese,
však s druhým ostatních jde průvod celý.

V originále je ovšem *il giuco della zara*, tj. přesně „hazardní hra“ se slovem hazard již zaběhnutým a kostka není explicitně zmíněna ani v třetím verši *ripetendo le volte e tristo impara*, v němž si hráč opakuje „vrhy“.Komentátor Benvenuto d'Imola vysvětluje, že kostky mají čtvercové stěny a že při hodu třemi kostkami se číslo může objevit vícekrát, jako třeba při součtu 3 to jsou tři jedničky, při součtu 4 dvě jedničky a dvojka, že stejně jsou na tom čísla 17 a 18 a že i všechna ostatní čísla mohou být vytvořena více způsoby. Výklad pravidel hry a pokus o matematickou analýzu jejího průběhu je již v prvních komentářích k Dantovi od Jacopo Giovanni della Lana z doby kolem 1324–1328, avšak komentáře byly publikovány až od roku 1477.

A tak ještě téměř o 400 let později se Galileo Galilei v krátkém traktátu *Sopra le Scoperte dei Dadi* (mezi 1613 a 1623, anglický překlad v knize [1]) na žádost nejmenovaného hráče pouští do vysvětlení poznatku, že při hře třemi kostkami se součet 10 vyskytuje častěji než 9. *Stařenčina* tabulka ukazuje, že pravděpodobnost devítky je 25/216=0.1157, zatímco desítky 27/216=0.125, rozdíl je tedy 1/108. Galileo vypočítává způsoby, jak dostat jednotlivé součty, odvodí jejich počet – 216 –

a sestaví celou tabulku. Hráčská praxe tazatele musela být značná, když si tak malého rozdílu všiml.

Alespoň velmi stručně o kombinatorice v kartách, které se v Evropě začaly šířit kolem začátku 15. stol. Jejich původ není jasný, drobné zmínky jsou z konce 14. stol., doložena je výroba karetní soupravy pro zábavu šíleného francouzského krále Karla VI. v roce 1393. Také sv. Bernard ze Sieny ve svém výše zmíněném kázání zmiňuje „*karty, na nichž jsou namalovány postavy*“, explicitně „*králové, královny, sluhové a rytíři*“ – patrně se jednalo o tarokové karty. Moderní hra má 56 obyčejných karet ve čtyřech barvách, 21 karet trumfových a jednoho žolíka, tedy celkem 78. Existuje však i stará velká verze s 35 kartami trumfovými a 6 žolíky, tj. celkem 97 karet. Při následujícím zjednodušení došlo k vypuštění jedné z figur a počet se ustálil na 52. Příčinou pomalého vytlačování kostek kartami a vrhcáby byla patrně větší komplikovanost těchto her i jejich obtížná kombinatorika, navíc karetní soubory byly ve srovnání s kostkami nesrovnatelně dražší.

O jejich počtu lze vymýšlet různá zdůvodnění, ale ta mohou být ošidná. O 52 kartách lze konstatovat, že odpovídají počtu týdnů v roce, dále že očíslováním spodka, svrška a krále čísly 11, 12, 13 a sečtením všech dostaneme 364, což s jednotkou za žolíka je počet dnů v roce. A když sečteme písmena Ace, Two, Three, ..., Jack, Queen, King, francouzsky As, Deux, ..., Roi, německy As, Zwei, ..., König (s ch jako jedním písmenem) i česky eso, dvě, ..., spodek, dáma, král, vždy vyjde 52. Avšak prvních karet nebylo 52, ale 56, a jejich počet odpovídá kombinatorickému číslu $C'_3(6) = 56$; také počet trumfů je $C'_2(6) = 21$ (ve velké hře $C'_4(4) = 35$ odvozené z hlezenních kůstek). Možná je to také náhoda, možná se tvůrci karetních her vyznali v kostkové kombinatorice.

Ve 14. a 15. století byly tedy kombinatorické metody již vymyšleny a byl pochopen jejich význam pro strategii her v kostky apod. Konkrétně, že počet způsobů, jak dosáhnout daného součtu, určuje jeho snadnost či obtížnost, tj. že např. při hodu třemi kostkami lze součty mezi 8 až 13 získat každý více než dvaceti způsoby, tedy s nadějí rovnou či větší než 1:10. Známá byla také úměrnost mezi počtem realizací a počtem dosažených jednotlivých případů. Rozdíly mezi jednotlivými autory vznikaly pouze chybným započtením permutací. Do jaké míry tyto výsledky byly známy obecně, lze dnes těžko posoudit: patrně však nedostatečně.

4. Pravděpodobnost v tóře a v rabínské literatuře

Hlavní díla rabínské literatury jsou často nazývána *ústní tórou* na rozdíl od *psané tóry*, kterou tvoří pět knih Mojžíšových. *Mišna* (napsaná kolem roku 200) obsahuje rozhodnutí rabínů a diskuse nad příkazy *Tóry, gemara* je soubor diskusí rabínů nad mišnou a spolu s ní tvoří Talmud, který se vyskytuje ve dvou verzích – jerusalémské (kolem 400) a babylónské (427–560) – a představuje nejen soubor náboženských předpisů, ale také celé trestní a občanské právo.

V této literatuře chybějí vzorce, zhusta se však projevuje pochopení pro to, že náhoda může být také nestranným soudcem, ať už Božím přičiněním nebo z povahy uskutečňovaného procesu. Nápadná je zřetelná opozice ke zvyklostem okolních národů. Zatím co kostky i hlezenní kůstky byly široce využívány v pohanských chrámech i ke hrám, u Židů bylo jejich užití ve svatyních zakázáno pod trestem smrti a hry byly buď stíhány nebo bylo aspoň hráči opovrhováno s odkazem na tóru: „*Když vejdeš do země, kterouž Hospodin, Bůh tvůj, dává tobě, neuč se činiti podlé ohavnosti národů těch. Nebude nalezen v tobě, kdož by vedl syna svého nebo dceru svou skrze oheň, ani věštec, ani planétník, ani čarodějník, ani kouzelník, ani losník, ani zaklínač, ani hadač, ani dotazující se mrtvých. Nebo ohavnost jest Hospodinu, kdožkoli to činí. . .* “, Deuter. X, 9–12.

Naproti tomu byly náhodné procesy široce využívány v liturgii i pro nalezení práva, neboť „*Los pokojí svády, a mezi silnými rozeznává*“, Přísl. XVIII. 18 (podle komentáře jsou silní ti, kteří mají mezi sebou velký spor). Nejvíce rozšířeno bylo losování z urny. Druhé hlavní užití bylo vyjádření Boží vůle, např. při dělení majetku, dědictví, opět s odkazem na *Tóru*: „*Avšak losem ať je rozdělena země, vedle jmen pokolení otců svých dědictví vezmou*“, Numeri XXVI, 55. V těchto případech byla představa náhodnosti odmítnuta, zvláště bylo-li losování přímo na příkaz Boží. Příkladem je dělení země mezi izraelské kmeny, popsané v knize Josue, kap. XIV až XVI.

Jiným příkladem losování v liturgii byl obřad obětování dvou kozlů na Den Smíření. Jeden z kozlů losem určený byl obětován Hospodinu, druhý vyhnán do pouště podle příkazu v Lev. XVI, 7–10. Losy musely být stejné, ze dřeva, kovu či kamene, označeny „pro Hospodina“ a „pro Azazela“ a vloženy do urny. Pak velekněz sáhl do urny, potřásl jí, do každé ruky uchopil jeden los a ten v pravé ruce položil na zvíře, jež měl po pravici, druhý na zvíře po levici. Šťastným znamením bylo, když Hospodinu určený kozel byl po pravici.

Náhodným procesem byly děleny také služby v chrámu. Ve starých dobách se výběr řešil během k okraji oltáře: kdo se první přiblížil na čtyři lokty (cca 2m), vyhrál a mohl uklidit popel po zápalných obětech. Když však jednou došlo ke srážce a zranění, byla zavedena náhodná volba. Kněží utvořili kruh a dohodli se na poměrně vysokém čísle, 80, 100 nebo 1000. Pak každý dal před sebe ruku s jedním, dvěma či třemi prsty (ne však palcem) a bylo provedeno odpočítání. Počítání podle prstů bylo nezbytné proto, že bylo zakázáno počítat lidi přímo, muselo to být podle nějakých předmětů, částí majetku apod. Tato povinnost vychází z I. knihy Samuelovy XV, 4: Saul shromáždí lid a počítá jej podle jejich ovcí. Ty jsou však explicitně zmíněny pouze ve *Vulgatě* („*. . . praecipit Saul populo et recensuit eos quasi agnos. . .* “), zatímco v *Kralické bibli*, současném překladu ani v *Septuagintě* není o ovcích ani slovo a i uvedený text lze chápat, že byli počítáni jako ovce a ne podle nich. Talmudická tradice však přímé počítání lidí s odvoláním na uvedený citát nedovoluje.

Náhodné metody bylo užíváno také k dělení masa mrtvých zvířecích obětí mezi kněze o svátcích. Každý odevzdal nějaký svůj předmět a nezaujatá osoba předměty rozložila na připravené kusy masa. Kněží si pak vzali kus se svým předmětem.

Vztah ke hráčům her byl rovněž přesně stanoven, přičemž byli rozlišováni podle toho, zda hrají kostkami či se skořápkami nebo slupkami z granátových jablek. Klíčový pojem byl Ashmakta – důvěra, a rozuměl se jím vztah ke slibu, že se ve hře za jistých okolností něco zaplatí, ale slib nebyl často dodržován a ani se to nevyžadovalo. Příkaz pak zněl: „*Tito jsou pak nevolitelní: hráč v kostky, lichvář, chovatel holubů, prodavači v rok odpočinku*". Volitelnost se vztahovala na úřad soudce či na vystupování jako svědek, rokem odpočinku byl každý sedmý. Diskuse kolem tohoto příkazu je v mišně i gemaře složitá, např. se uvádí, že nevolitelnost se vztahuje pouze na ty, kteří se vyjmenovanými činnostmi výhradně živí atd. Obecně však důvodem k zatracení hráčů bylo, že očekávají výhru aniž sami co dají, nijak nepřispívají k obecnému blahu a odvracejí se od studia *Tóry*. Hra a využívání náhody nejsou tedy obecně zakázány, problém však představuje bezpracný zisk, který je staven na roveň loupeži.

V *Talmudu* je uveden také jeden příklad užití losu v soudním řízení, totiž dělení dědictví mezi bratry s odkazem na výše vzpomenuté dělení země mezi židovské kmeny popsané v knize Josue.

Zajímavé jsou příkazy na řešení situací, do nichž náhodnost vstoupila sama, tedy bez lidského přičinění. Rabíni v *Talmudu* pro ně nabízejí vzorová řešení.

I. Ve městě je devět řeznictví s košer masem, jedno s masem nečistým. Na ulici je nalezen kus masa neznámého původu. Lze je považovat za košer podle „zákona většiny", neboť je košer s pravděpodobností 9:1 (pravidlo říká, že oddělení je oddělení z většiny). Jiná situace však vznikne, když si někdo donese maso domů a nepamatuje si, zda je koupil v košer řeznictví pro sebe nebo v nečistém pro svého psa. Pravděpodobnost je pak 1:1, a maso je nutno považovat za nečisté.

II. Většina však má být přesně vypočitatelná, jako v předchozím případě nebo alespoň výrazně převládající. Jako příklad je uvedeno, že soudní písaři jsou v převážné většině zákonů znalí, takže libovolný písemný dokument od soudu má být považován za správně vyhotovený, i když existenci zákonů neznalého písaře nelze zcela vyloučit.

III. Další příklad aplikuje násobení pravděpodobnosti: Muž zemře a zanechá po sobě těhotnou ženu. Jaká je pravděpodobnost, že porodí živého chlapce a tedy potenciálního dědice? Předpokládejme, že chlapci i děvčata se rodí se stejnou četností, takže poměr pohlaví je 1:1. Vyskytují se však s jistou pravděpodobností, řekněme a, také potraty. Pak ovšem pravděpodobnost narození živého chlapce je $0.5(1-a) < 1/2$.

IV. Ještě jeden příklad na násobení pravděpodobnosti. Nevěrné ženě je zakázáno žít s manželem a ten se s ní může beztrestně rozvést. To však neplatí, pokud muž prohlásí, že není panna a byla mu nevěrná před sňatkem. Zde se objevuje dvojí pochybnost. Panenství mohla ztratit v době zasnoubení se svým budoucím mužem nebo s někým jiným, to je 1:1. Bez ohledu na to, s kým to bylo, je zde možnost 1:1, že byla znásilněna, což není důvod k rozvodu. Takže pravděpodobnost zákonného rozvodu je jenom 1/4.

Z mystických spisů připomeňme alespoň spis *Sefer j'círa* [Kniha stvoření], vzniklý patrně před 3. stol. po Kr. Podle jedné legendy se jedná o zápis *ústního zákona* sděleného Hospodinem Mojžíšovi na Sinaji, podle jiné byl zapsán Abrahamem, který jej napsal po třech letech přemýšlení a s Hospodinem přikázanou pomocí Sema. V knize je popisováno stvoření světa z čísel a písmen hebrejské abecedy, jež ovšem nejsou pouhými jazykovými symboly, ale existují samostatně mimo lidskou mysl. Stvoření je pak proces formování těchto písmen a jím vzniká reálný svět. Znalost a pochopení textu knihy ve spojení s dobrým úmyslem umožňuje stvořit Golema či jiného živého tvora (protože však dokonalosti stvoření Božích nelze dosáhnout, je Golem němý).

22 písmen hebrejské abecedy je považováno za kameny stvoření. První kapitola rozebírá skupinu božího jména JHV a jejích $P(3) = 6$ kombinací značících šest směrů – nahoru, dolů, sever, jih, východ, západ. Druhá kapitola popisuje tvorbu párů, v komentářích je odvozen počet variací $V_2(22) = 462$. Ve čtvrté kapitole je pak naznačen výpočet permutací $P(n)$ (kámen znamená písmeno, dům pak slovo): „*Dva kameny tvoří dva domy, tři stavějí šest domů, čtyři stavějí dvacet čtyři domy . . . sedm staví 5400 domů, a jdi a dál počítej*“. V komentářích se ještě objevuje problém čtyřech písmen, z nichž dvě jsou stejná, tj. $P'_{2,1,1}(4) = 12$.

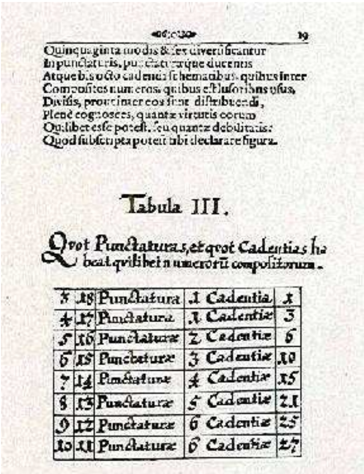
Kombinatorika i elementární aritmetika pravděpodobnosti se tedy v židovské literatuře objevují velmi časně a mají praktický charakter. Je zvláštní, že zatímco zde byla využívána k počítání písmenných skupin s vysokým výchozím počtem 22, v evropském prostředí skupiny o pouhých 6 nebo 4 členech odpovídající kombinatorice kostek byly zkoumány až téměř o tisíc let později.

5. Období předcházející vzniku teorie pravděpodobnosti

Jedním z nejdůležitějších impulsů k Pascalovu zájmu o problematiku pravděpodobnosti byla dlouho známý a poněkud umělý problém, nazývaný dnes *úlohou o rozdělení sázky*. Vyskytuje se v italských rukopisech ze 14. a 15. století, O. Ore (současný americký matematik) našel nejstarší zmínku v rukopise z roku 1380. Obvykle však bývá spojován se jménem L. Pacioli[6], viz Obrázek 8, který jej uvádí ve své knize z roku 1494 takto: dva hráči míčové hry se dohodnou, že domluvenou

[6] *Luca Pacioli* (1445–1517), italský matematik, autor několika knih, učil na universitách v Pise, Florencii, Bologni a jinde, přítel Leonarda da Vinci.

sázku získá ten, kdo první dosáhne šesti bodů. V okamžiku, kdy stav je 5:3, musejí hru ukončit. Jak se mají o sázku rozdělit? Pacioliho řešení, že to mají udělat v poměru již dosažených bodů, je zřejmě nesprávné. Jak upozornil N.F. Tartaglia[7] v roce 1556, v případě přerušení za stavu 1:0 by odtud vyplývalo, že celou sázku bere vítěz první hry dřív, než druhý hráč vůbec začne hrát.



19

Quinquaginta modis & sex diversificantur
In punctaturis, punctaturaque ducentis
Atque bis octo cadentiis schematibus, quibus inter
Compositos numeros, quibus est lusoribus usus,
Divisis, prout inter eos sunt distribuendi,
Plene cognosces, quantae virtutis eorum
Quilibet esse potest, seu quantae debilitatis:
Quod subscripta potest tibi declarare figura.

Tabula III.

Qvot Punctaturas, et qvot Cadentias habeat qvilibet numerorum compositorum.

3	18	Punctatura	1 Cadentia	1
4	17	Punctatura	1 Cadentiae	3
5	16	Punctaturae	2 Cadentiae	6
6	15	Punctaturae	3 Cadentiae	10
7	14	Punctaturae	4 Cadentiae	15
8	13	Punctaturae	5 Cadentiae	21
9	12	Punctaturae	6 Cadentiae	25
10	11	Punctaturae	6 Cadentiae	27

Obrázek 8. Tabulka a text z básně *De vetula* a portrét Luca Pacioli

Obrázek 9. N. F. Tartaglia a G. Cardano

[7]*Niccolo Fontana Tartaglia* (1499–1557), matematický samouk, živil se vyučováním nižších partií matematiky, autor řady matematických knih.

Tartaglia navrhuje řešení, podle nějž hráči dosáhnuvší počtů A, B z dohodnutého počtu her Z se dělí v poměru $\big(Z + (A - B)\big) : \big(Z - (A - B)\big)$, tedy v uvedeném případě v poměru $2 : 1$. Ani toto řešení však není správné. Ostatně Tartaglia mu sám příliš nevěřil a komentuje je slovy: „*. . . řešení tohoto problému je záležitost spíše právnická než matematická, takže ať se už rozdělení provede jakkoliv, vždy se najde důvod ke sporům*“. O dva roky později G.F. Peverone[8] v prakticky identickém případě dochází k dělení $6 : 1$, což je již hodně blízko správnému řešení. Především se však při řešení neopírá o odehranou minulost, nýbrž o dosud neodehranou budoucnost, tj. respektuje pravidlo, že dělení musí vycházet z toho, kolik který hráč potřebuje vyhrát her, aby sázku získal celou a jak je to pravděpodobné bez ohledu na již prokázané „kvality“ či přesněji přízeň Náhody. Tuto zásadu formuloval Cardano[9] v tzv. *principu úměrnosti*: vyhraná sázka má být úměrná počtu způsobů, kterými lze vyhrát. Z ní vyplynulo správné řešení $7 : 1$, podané o sto let později B. Pascalem a P. Fermatem.

Logické důsledky této jinak málo praktické úlohy, která se dnes vyskytuje pouze v historických pracích, jsou však značné. *Při řešení nejisté úlohy v budoucnosti máme uvažovat pouze to, co se může stát, a nikoliv to, co se již stalo*. O tom, že tato zásada není obecně používána, se v běžném životě přesvědčujeme více než často.

Zbývá odpověď na otázku položenou na začátku: proč se teorie pravděpodobnosti tak opozdila za ostatními oblastmi matematiky?. Bývají navrhovány čtyři obecné důvody: nedostatečně zvládnutá kombinatorika, pověrčivost hráčů (to se ovšem vztahuje pouze k představě, že hry jsou exklusivní cestou k pravděpodobnostnímu myšlení), nedostatečná představa o náhodnosti jako důležitém elementu dění, morální či náboženské přehrady bránící připuštění náhodnosti dějů (z důvodů výše zmíněných pomíjíme nedokonalost kostek, ostatně spornou).

Problém náboženství nelze podceňovat, podle křesťanské věrouky všechno dění je v rukou Božích. S protiargumentem, že pak by vymizely hry i náhoda a štěstí se vyrovnal sv. Tomáš Akvinský poukazem na univerzální zákony a dílčí zákony a že se sice nic nemůže vymknout ze zákona universálního, ale u zákonu dílčího se to stát může a pak mluvíme o náhodnosti. Hugenotovi de Moivrovi dokonce pravděpodobnostní zákony vyjadřují pevný řád Vesmíru a vedou k uznání díla Stvořitelova. Mimo teologické úvahy to náhoda má dokonce horší. Spinoza v r. 1677 píše, že „*událost může být považována za náhodnou jedině ve vztahu k našim nedostatečným*

[8] *Giovanni Francesco Peverone* (1509–1559), italský matematik, autor knihy *Due brevi e trattati. . .* (1558), věnované posloupnostem, řešení rovnic, základům geometrie a také pravděpodobnosti.

[9] *Girolamo Cardano* (1501–1576), význačný italsky matematik a polyhistor, autor několika knih: v *Ars Magna* probírá algebru komplexních čísel a řešení kubických rovnic, dvě encyklopedie přírodních věd „obsahují ode všeho trochu, od kosmologie po konstrukci strojů, od užitečnosti přírodních věd po ďábelský vliv démonů, od zákonů mechaniky po kryptologii“. Jeho kniha *Liber de Ludo Aleae* věnovaná hrám a pravděpodobnosti (sám byl vášnivý hráč) však byla publikována až 1663.

znalostem“ a u d’Alemberta v textu z r. 1750 najdeme: „*Přesně vzato neexistuje žádná náhoda, jedině její ekvivalent: naše nevědomost, díky níž my sami jsme její příčinou*“. Stejné názory najdeme ostatně i u Laplace. Ten si představoval universální bytost nevyčerpatelné inteligence vědoucí o Vesmíru v každém okamžiku vše a pojem pravděpodobnosti vůbec nepotřebující. Pro nás je však nutný „*zčásti díky nevědomosti, zčásti díky znalosti. Víme, že ze tří či více jevů by se měl státi pouze jediný, z ničeho však nemůžeme usoudit, který z nich to bude. V tomto stavu nerozhodnosti je nemožné ohlásit výsledek s jistotou*“.

Naštěstí jsme dnes ovlivněni moderní biologií: variabilita druhů i jejich skupinového i individuálního vývoje je jen obtížně popíratelná a výrazný vliv náhodných procesů je rozpoznán. Stejně je tomu v kvantové fyzice, zachycující řadu náhodných procesů na nás zcela nezávislých, např. radioaktivní rozpad. Proto jsme na pojem náhodnosti jako charakteristického rysu života a všeho dění již zvyklí a nedovedeme si představit, že tomu někdy bylo jinak. Lze tedy dosti oprávněně soudit, a zde lze opět ocitovat myšlenku z úvodu o matematice pouze vyjadřující naše poznání, že hlavní příčinou pozdního zrodu pravděpodobnosti byla obecná nepřipravenost společnosti i jednotlivců k porozumění jejím koncepcím, neschopnost je začlenit do existujícího myšlenkového systému a také jakýmkoliv způsobem využít. Za místo jejího vzniku je patrně správné považovat Itálii 14. až 16. století, ale její myšlenky a postupy, ve společnosti nepřipravené je přijmout, zapadaly okamžitě po svém zrodu. Zásluhou západní Evropy, jmenovitě Francie 17. století bylo, že znovu oživené pravděpodobnostní problémy již dokázaly vzbudit společenský zájem.

Literatura

[1] David F.N. (1998) *Games, Gods and Gambling, A history of probability and statistical ideas*. Dover Publ., Inc., Mineola, New York.

[2] Hasofer M. (1966) *Random mechanisms in talmudic literature*. Biometrika **54**, 316–321.

[3] Juškevič P. (1978) *Dějiny matematiky ve středověku*. Academia, nakladatelství ČSAV, Praha.

[4] Kansara N.M. (2000) *Vedic sources of the ‘Vedic mathematics’*. Sambodhi XXIII. Plně reprodukováno na adrese: `http://vmacademy.com/kenneth/articles/kansara/kansara.htm`

[5] Kendall M.G. (1956) *The beginnings of probability calculus*. Biometrika 43, 1–14.

[6] Kendall M.G. (1957) *A note on playing cards*. Biometrika **44**, 260–262.

[7] Kendall M.G. (1960) *Where shall the history of statistics begin?* Biometrika **47**, 447–449.

[8] Majstrov L.E. (1967) *Teorija verojatnostej*. Nauka, Moskva.

[9] Ore O. (1960) *Pascal and the invention of probability theory*. Am. Math. Mon. **67**, 409–419.

[10] Rabinovitch N.L. (1969) *Probability in Talmud*. Biometrika **56**, 437–441.

[11] Rabinovitch N.L. (1970) *Combinations and probability in rabbinic literature*. Biometrika **57**, 203–205.

[12] Shafer G. (1993) *The early development of mathematical probability*. In: *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, I. Grattan-Guinness Ed., Routledge, London, 1293–1302.

[13] Sefer Yetzirah `http://www.wbenjamin.org/saadia.html` (Commentary by Saadia ben Joseph (al-Fayyumi).

Poznámka: Části 1–3 a 5 se opírají především o práce [1] a [5–7], část 4 pak o práce [2, 10, 11, 13], hojně byla využívána i literatura dostupná na Internetu, např. [4, 13].

STEREOLOGIE: INTERPRETACE VÝBĚRŮ GEOMETRICKÝMI PROSTŘEDKY

Ivan Saxl a Viktor Beneš

Univerzita Karlova v Praze, Matematicko-fyzikální fakulta, Katedra pravděpodobnosti a matematické statistiky, Sokolovská 83, 186 75 Praha 8 – Karlín
ivan.saxl@mff.cuni.cz, viktor.benes@mff.cuni.cz

1. Úvod

Když kupujete půlku chleba, vyberete si tu, v níž jsou dobře patrné dutiny vhodné velikosti rovnoměrně rozmístěny na celém řezu. Podobně zvolíte dílec Nivy, který má na celém řezu hodně plísňových oblastí. Přitom usuzujete, že to, co vidíte na jednom dvourozměrném řezu, platí pro celý objem, tj. že chléb je správně vykynutý a sýr dobře vyzrálý. Vaše úvaha je příkladem stereologického odhadu, při němž z výběrového souboru - rovinného řezu - usuzujete na vlastnosti celého trojrozměrného objektu.

V obou případech se jedná o objekty, které mají dvě složky - tzv. matrici (těsto, bílá sýrová hmota) a rozptýlenou složku či fázi (dutiny, plíseň) - a vy hodnotíte tři aspekty: a) zda sekundární fáze je rozmístěna všude a rovnoměrně, b) zda izolované složky fáze mají přijatelnou a nepříliš rozdílnou velikost, c) zda celková plocha řezu zabraná fází není příliš velká ani příliš malá.

V podobné situaci se nachází pracovník v laboratoři, zkoumající vzorek materiálu, v němž druhou fází mohou být zpevňující částice nebo naopak trhlinky či technologií způsobené defekty, lékař hodnotící biopsií odejmutý sloupec tkáně se zdravými či patologickými buňkami, zemědělský výzkumník prohlížející vzorek půdy s dutinami a kořenovými systémy rostlin.

2. Odhady z geometrických výběrů

Dostáváme se k definici: Stereologie je oddíl stochastické geometrie (tj. té části teorie pravděpodobnosti a matematické statistiky, jejímž předmětem nejsou čísla, ale geometrické objekty) zabývající se odhady vlastností populací geometrických objektů z jejich výběrů pořízených geometrickými prostředky - tj. řezy, průniky a projekcemi (stručný přehled pravidel geometrického výběru je v [1]).

Klíčová slova: Geometrický výběr, objemový podíl, plošný podíl, čárový podíl, invariantní funkcionály, stereologické rovnice.

Poděkování: Tato práce vznikla za podpory grantu MSM 113200008 a grantu GAČR č. 201/03/0946.

Obecně nastávají dvě zásadně rozdílné situace. Studovaný objekt je deterministický (zdravý kontra patologický orgán - ledvina, plíce, kost a její dřeň, porušená strojní součástka) a náhodný je systém výběrů probíhající podle vhodně zvoleného, obvykle několikastupňového plánu či projektu - *projektová stereologie, design stereology*. Druhý případ nastává, když sledovaná fáze (neboť o tu se vesměs jedná) má vlastnosti s dostatečnou přesností popsané nějakým stochastickým modelem. Jednoduchým příkladem modelu mohou být koule v jistém rozmezí náhodně se měnícího poloměru rozmístěné rovnoměrně náhodně v prostoru (tj. jejich středy jsou realizací Poissonova bodového procesu). Právě tento speciální případ tzv. *Booleova modelu* je vhodný pro popis pórů v chlebu. Na poloze a orientaci rovinného řezu potom jen málo záleží, všechny řezy dávají velmi podobné výsledky a předložený problém řešíme *modelovou stereologií*.

Obě větve stereologie řeší tři typy problémů. *Globální stereologie* zjišťuje intenzity geometrických charakteristik zkoumaného systému jako celku (objem a plocha povrchu alveol na jednotku objemu plic, relativní objem - objemový podíl - zpevňující fáze v materiálu) včetně přednostních orientací v systému se vyskytujících ploch a čar. *Lokální stereologie*, se zabývá jednotlivými izolovanými složkami fáze, tj. jejich rozměry a tvary. Konečně *stereologie prostorového rozmístění* se pokouší popsat, jak jsou jednotlivé prvky systému v prostoru rozmístěny (přitahují se a proto shlukují nebo naopak se odpuzují a pokoušejí se zaujmout quasi-pravidelné prostorové rozmístění).

Charakteristickým znakem většiny geometrických výběrů je, že má nižší dimenzi než objekt, ze kterého byl pořízen. Pro průniky (řezy) objektu Z_k dimenze k s lineárním podprostorem (přesně se používá termínu r-rovina) F_r dimenze r v prostoru dimenze d platí obvykle $\dim(Z_k \cap F_r) = r + k - d$. Takže rovinné řezy (profily) konečného trojrozměrného objektu jsou omezené plošky, hranice objektů vytvoří v rovině uzavřené křivky, hranám odpovídají body a vrcholy nejsou zachyceny vůbec. Jestliže těleso není konvexní, může mít řez jednoho tělesa i několik izolovaných složek (řezem činky mohou být dva kruhy a jedna elipsa, řezem žebříku s n příčli celkem až $(n + 2)$ elips). Dimenze projekce je nezbytně vždy nejvýše rovna dimenzi lineárního podprostoru, do nějž promítáme. V tomto příspěvku bude pozornost věnována pouze globální stereologii. Dříve než budou uvedeny obecné vztahy, všimněme si několika konkrétních historických úloh v nichž jsou vlastnosti výběrového souboru dobře patrné.

a) *Buffonův čtverec.* Renesanční i osvícenská šlechta si ráda krátila dlouhou chvíli hrou *na čtverce*, spočívající v házení mince poloměru r na podlahu s čtvercovými dlaždicemi (o straně a, tj. ploše $|A| = a^2$).

Předpokládejme nejprve, že uvnitř čtverce je další dostatečně velká oblast A' plochy $|A'|$ a poloměr mince je zanedbatelný vzhledem k rozměrům této oblasti a tedy i čtverce. Jaká je pravděpodobnost P, že mince, která dopadne do čtverce, zasáhne i oblast A'? Buffonovo správné řešení bylo $P = |A'|/|A|$. Pokud bychom nechtěli zanedbávat poloměr mince, museli bychom požadovat, aby *střed mince* x ležel jak ve čtverci, tak v oblasti A'.

Hru lze chápat i jako stereologickou úlohu. Nechť $\{x_1, \ldots, x_n\} \subset A$ je soubor bodů. Sledujeme průniky $x_i \cap A',\ i = 1, \ldots, n$ za podmínky $x_i \cap A \neq \emptyset$ a přiřazujeme jim hodnoty $p_i = 1$, když $x_i \cap A' \neq \emptyset$ a $p_i = 0$ jinak. Potom $n^{-1} \sum_{i=1}^{n} p_i = m/n$ je odhadem plošného podílu A' v A, který si označíme $A_A(A') = |A'|/|A|$. V duchu stejné symboliky označíme podíl m/n počtů bodů zasáhnuvších A', resp. A, symbolem $P_P(A')$, a píšeme $A_A(A') = P_P(A')$ (jedná-li se o obecné tvrzení o rovnosti plošných a bodových podílů, argument A' se vynechává). Z dvourozměrných oblastí A, A' jsme tedy vybrali dva 0-dimenzionální soubory a poměr jejich mohutností je odhad plošného podílu $|A'|/|A|$ resp., pokud změříme hranu čtverce, dostaneme odhad plochy $m|A|/n$ i zcela nepravidelné oblasti A'.

Zobecnění na trojrozměrný prostor je zcela bezprostřední. Házením koulí do krychle objemu V a registrováním, zda jejich středy protínají vymezenou oblast V' bychom stejným postupem odhadli objemový podíl $V_V(V') = P_P(A')$, ve zkráceném zápise $V_V) = P_P$. Musíme ovšem zůstat jenom u úvahy, protože k praktické realizaci takovéto hry bychom potřebovali čtyřrozměrný prostor.

Při hře na čtverce se ovšem častěji sledoval případ, kdy mince protnula okraj. K tomu nedošlo, pokud její střed ležel ve čtverci o straně $a-2r$ a ploše $A_0 = (a-2r)^2$ (pro jednoduchost píšeme dále X místo $|X|$). Dopadl-li střed mince do zbývající části čtverce, mohly nastat dva případy: i) hranice (okraj) mince protnula ve dvou bodech buď jenom vertikální, či jenom horizontální spáry a střed ležel v okrajovém pásu o ploše $A_2 = 4r(a - 2r)$ nebo ii) mince protnula jak vertikální, tak horizontální spáru celkem ve čtyřech bodech a střed ležel v jednom ze čtyř čtverečků kolem vrcholů čtverce o celkové ploše $A_4 = 4r^2$. Pravděpodobnosti jednotlivých případů jsou $P_0 = A_0/A = (1 - 2q)^2,\ P_2 = A_2/A = 4q(1 - 2q),\ P_4 = A_4/A = 4q^2$, kde $q = r/a$. Poměr pravděpodobností $P_4/P_2 = (q^{-1} - 2)^{-1} \doteq q$ pro $q \ll 1$.

Na hru se také můžeme podívat jinak. Znovu budeme házet minci (tak, aby její střed ležel ve čtverci) celkem n-krát a počítat průsečíky jejího obvodu se spárami. Mince má obvod $2\pi r$ a tento obvod protíná spáry v jednom případě dvakrát, ve druhém čtyřikrát, ve zbývajících případech je neprotíná vůbec, přitom celková délka obvodů všech mincí je $L = 2\pi rn$. Očekávaný (teoretický) počet průsečíků tedy bude $n(2P_2 + 4P_4) = 8nr/a$ a odtud počet průsečíků na jednotkovou délku obvodu mincí je $N_L^m = 4/(\pi a)$. Délka spáry připadající na čtverec je $2a$ a tedy délková intenzita spár na jednotku plochy je $L_A^s = 2/a$. Dosazením dostaneme vztah $L_A^s = \frac{\pi}{2} N_L^m$.

Předpokládejme nyní, že všechny mince dopadly do oblasti plochy W, takže délková intenzita jejich obvodů je $L_A^m = 2\pi rn/W$. Střední délka spár na ploše W je $2W/a$, takže intenzita průniků spár s obvody mincí bude $N_L^s = \frac{8nr}{a} / \frac{2W}{a} = \frac{4nr}{W} = \frac{2}{\pi} L_A^m$, takže dostaneme symetrický vztah $L_A^m = \frac{\pi}{2} N_L^s$.

Vztahy pro L_L^s a L_L^m jsou stereologické relace. Máme dva systémy čar v rovině - pravidelný systém spár neznámé délky a soubor obvodů náhodně položených mincí známého poloměru r, tj. náhodný testovací systém délky $L = 2\pi rn$. Naměřili jsme jistý počet průsečíků, řekněme M, které představují 0-dimenzionální geometrický výběr z čárového systému spár provedený testovacím systémem a, ekvivalentně, 0-dimenzionální výběr z testovacího systému provedený čárovým systémem. Z jejich počtu dostaneme nestranný odhad $M/(4rn)$ plošné intenzity spár L_A^s. Při házení se ovšem nesmíme pokoušet zasáhnout spáru, jako hráči při hře, ale musíme minci vrhat nejlépe poslepu.

V druhém případě máme náhodný systém čar - obvodů mincí - a testujeme jej známým pravidelným čárovým systémem spár. Opět dostaneme nestranný odhad intenzity neznámého čárového systému obvodů mincí. Kombinací výsledků získaných natáčením a posouváním testovacího systému či jeho prodloužením se výsledný odhad bude zpřesňovat; u původní úlohy s podlahovými spárami je to ovšem nemožné.

Na Internetových strankách lze najít mnoho interaktivních simulačních programů a sledovat jak pomalu se s rostoucím počtem hodů jehlou na síť rovnoběžek zpřesňuje odhad čísla π. Programy jsou založeny na následující historické úloze.

b) *Buffonova úloha o jehle.* V této úloze z 18. století je jehla délky l házena na síť rovnoběžných čar vzdálených o $d > l$. Otázka zní, jaká je pravděpodobnost, že jehla protne některou čáru. Tato pravděpodobnost je rovna poměru délky průmětu jehly do normály k čarám ke vzdálenosti čar. Při izotropní orientaci jehel je tato délka rovna (ϕ označíme úhel mezi jehlou a normálou k čarám)

$$\frac{2}{\pi} \int_0^{\pi/2} l \cos\phi \mathrm{d}\phi = \frac{2}{\pi} l,$$

takže hledaná pravděpodobnost je rovna $P = 2l/(\pi d)$.

Stejně jako jsme v případě hry na čtverce použili obvodů mincí jako testovacího souboru pro systém spár, nyní můžeme použít jehel jako testovacího systému k odhadu délkové intenzity rovnoběžek. Očekávaný počet průsečíků bude $2nl/(\pi d)$, takže na délce nl bude systémem rovnoběžek indukována intenzita průsečíků $N_L^j = 2/(\pi d)$. Přitom na jednotkovou plochu připadá intenzita čar systému $L_A^s = 1/d$, takže opět stejně jako ve hře na čtverce dostaneme vztah $L_A^s = \frac{\pi}{2} N_L^j$.

Při inverzní úloze hrají roli testovacího systému rovnoběžky. Jestliže jehly dopadly na oblast plochy W, je jejich délková intenzita $L_A^j = nl/W$ a snadno dostaneme $L_A^j = \frac{\pi}{2} N_L^s$, kde N_L^s je intenzita průsečíků na rovnoběžkách.

Všimněme si, že ve vztazích nevystupují explicitně ani strany čtverců či vzdálenosti rovnoběžek, ani průměry mincí či délky jehel. Skutečně na nich vůbec nezáleží, rozhodující jsou pouze délkové intenzity čárových systémů a intenzity průsečíků vzájemně indukovaných, tj. uskutečněné geometrické výběry z testovaného systému provedené systémem testovacím a naopak. Pro obecný případ byl vztah odvozen poprvé Croftonem v roce 1869 a platí pro libovolný testovací i sledovaný soubor. Pokud je jeden z nich anizotropní, musí být druhý izotropní (tak, jako tomu bylo v obou rozebíraných případech).

Zcela analogicky můžeme řešit úlohu o jehle také v trojrozměrném prostoru (potřebujeme ovšem beztížný stav a virtuální roviny místo čar). Dostaneme $S_V^r = 2N_L^j$, kde $S_V^r = 1/d$ je plošná intenzita rovnoběžných rovin vzdálených o d a N_L^j je intenzita průsečíků jehel s rovinami (geometrický výběr z rovin provedený jehlami), tj. jejich střední počet na jednotku délky souboru jehel. Při inverzní úloze představují jehly rozházené v oblasti obsahu W čárový systém s délkovou intenzitou $L_V^j = nl/W$. Testovací roviny z něj vyberou $N_L^j L_V^j W$ průsečíků, jež na ploše $S_V^r W$ testovacích rovin obsažených v oblasti obsahu W vytvoří bodový soubor intenzity $P_A^r = \frac{N_L^j L_V^j W}{S_V^r W} = \frac{1}{2} L_V^j$, takže opět symetricky $L_V^j = 2P_A^r$. V obou případech se opět jedná o zcela obecné vztahy, umožňující odhadovat plošnou intenzitu ploch v prostoru z intenzity bodů jimi indukovaných na čárovém testovacím systému a naopak délkovou intenzitu čárového systému v prostoru plošným testovacím systémem.

V předcházejících úvahách jsme ovšem nevěnovali pozornost jedné skutečnosti. Všimněme si např. vztahu $A_A(A') = P_P(A')$, kde estimátor m/n je podíl počtů bodů zasahujících A, resp. A'. Obecně jsou m, n závislé náhodné proměnné a tedy $\boldsymbol{E}\frac{m}{n} \neq \frac{\boldsymbol{E}m}{\boldsymbol{E}n}$. Přesnou formu všech předchozích vztahů uvedeme v odst. 4.

3. Charakteristiky konvexních těles

Na příkladech jsme viděli, že geometrické výběry poskytují možnost odhadovat plochy obrazců i délky jejich hranic, stejně jako objemy trojrozměrných těles a plochy jejich povrchů. Stačí to pro charakterizování těles a obrazců? Jaké vlastně máme požadavky na veličiny popisující tělesa a obrazce? Další myslitelnou charakteristikou by mohly být např. střední hodnoty obsahů ploch a délek projekcí do izotropních trsů a svazků přímek a rovin. Dozvěděli bychom se z nich něco něco podstatně jiného?

Rozumný soubor požadavků na charakteristiky $\phi(K)$ konvexních těles K, který se dá vhodně rozšířit i na jejich konečná sjednocení, je následující: $\phi(\emptyset) = 0$ a $\phi(K)$

se nemění při pohybu K, tj. $\phi(K) = \phi(gK)$, kde g je element grupy G eukleidovských pohybů. Dále je $\phi(K)$ monotonní, tj. $\phi(K') \leq \phi(K)$ pro $K' \subseteq K$, takže část tělesa nemůže mít větší hodnotu charakteristiky než celek. Konečně musí být $\phi(K)$ aditivní, takže platí $\phi(K \cup K') = \phi(K) + \phi(K') - \phi(K \cap K')$, tj. při sjednocení dvou konvexních těles musíme charakteristiku jejich průniku odečíst. Charakteristiky splňující tyto vlastnosti nazýváme invariantní aditivní funkcionály a lze dokázat, že v d-rozměrném prostoru tvoří vektorový prostor dimenze $d + 1$ (Hadwigerův charakterizační teorém, který platí i pro konečná sjednocení kompaktních množin). Jinými slovy, v trojrozměrném prostoru lze nalézt čtveřice invariantních aditivních funkcionálů $\phi_0(K), \ldots, \phi_3(K)$ takových, že pro každý invariantní aditivní funkcionál $\phi(K)$ platí $\phi(K) = \sum_{i=0}^{3} c_i \phi_i(K), 0 \leq c_i < \infty$ pro $i = 0, \ldots, 3$. První z těchto funkcionálů je Eulerova charakteristika a v uvažovaném případě konvexní množiny je roven jedné; mění se při přechodu od jedné konvexní množiny k jejich sjednocení a pokud jsou množiny disjunktní, udává jejich počet. Dalšími funkcionály jsou v trojrozměrném prostoru např. objem $\nu_3(K)$, obsah hranice (povrch) $\nu_2(\partial K)$ a střední šířka $H_1(K)$, tj. střední délka ortogonálních projekcí tělesa do trsu izotropních přímek. Ta se obvykle se značí $w(K)$, zde je pro jednotnost značení projekčních charakteristik použit méně běžný symbol H_1; často se používá i v technice a nazývá se pak střední Feretův průměr nebo klešťový průměr, neboť její odhad pro malé tělísko bychom dostali jako střední hodnotu údajů mikrometru při různých natočeních tělíska.

V rovině lze konvexní obrazec popisovat pomocí plochy $\nu_2(K)$ a dále buď délkou hranice $\nu_1(\partial K)$, nebo střední šířkou $h_1(K)$, které musejí být navzájem úměrné. Platí totiž, že obvod je π-násobkem střední šířky, což je dvourozměrná verze věty Cauchy. Její trojrozměrnou analogií je vztah mezi plochou hranice $\nu_2(\partial K)$ a střední plochou rovinného průmětu $H_2(K)$: $\nu_2(\partial K) = 4H_2(K)$. Platnost těchto vztahů zřejmá pro kouli a kruh je zcela obecná, takže střední plocha projekce jednotkové krychle je rovna $\frac{3}{2}$ a střední šířka jednotkového čtverce je $4/\pi$. Konečně pro konvexní množinu na přímce (interval) máme následující charakteristiky: délku $\nu_1(K)$ a opět počet (identicky rovný jedné).

Všechny charakteristiky jsou shrnuty v následující tabulce. Střední šířku v rovině, tj. izotropní projekci do svazku přímek, značíme pro odlišení od prostorového případu malým písmenem, tj. $h_1(K)$. Některé isotropní aditivní funkcionály včetně všech pravě vybraných mají další důležitou vlastnost, totiž $\phi(\alpha K) = \alpha^j \phi(K)$, kde αK je skalární násobek („α-násobná zvětšenina“ K) a j je nezáporné celé číslo; $\phi(K)$ jsou potom homogenní funkcionály stupně j. Z Hadwigerova charakterizačního teorému plyne, že když $\phi(K), \phi'(K)$ jsou dva monotónní aditivní invariantní funkcionály se stejným stupněm homogenity, pak $\phi(K) = c\phi'(K)$ (příkladem jsou obě uvedené verze věty Cauchy). Obecně v $\mathbb{R}^d$ je $j = d$ pro $\nu_d(K)$, $j = d - 1$ pro $\nu_{d-1}(\partial K)$ a $j = 1$ pro všechny střední šířky.

$d = 3$	$d = 2$	$d = 1$
objem $\nu_3(K)$	plocha $\nu_2(K)$	délka $\nu_1(K)$
plocha hranice $\nu_2(\partial K) = 4H_2(K)$	délka hranice $\nu_1(\partial K) = \pi h_1(K)$	$\nu_0(K) = 1$
střední šířka $H_1(K)$	$\nu_0(K) = 1$	–
počet $\nu_0(K) = 1$	–	–

Tabulka 1. Soubor homogenních invariantních aditivních funkcionálů konvexních těles v $\mathbb{R}^d,\ d = 1, 2, 3$.

4. Základní stereologické rovnice

Pomocí integrální geometrie [2,3] lze odvodit Croftonův teorém vzájemně svazující funkcionály konvexního tělesa a jeho průniků lineárními podprostory F_r, tzv. r-rovinami (tj. rovinami F_2, přímkami F_1, body F_0):

$$\phi(K) = c \int_{F_r \uparrow K} \phi'(F_r \cap K) \mathrm{d}\mu(F_r \uparrow K) = c\mu(F_r \uparrow K) \boldsymbol{E} \phi'(F_r \cap K),\ r = 0, 1, 2,$$

kde ϕ, ϕ' jsou invariantní aditivní funkcionály, $\mu(F_r \uparrow K)$ je míra všech F_r, které mají s K neprázdný průnik (zasahují K) a konstanta c závisí na typech funkcionálů ϕ, ϕ' a na r, ne však na K. Zcela jednoduchý je případ obsahu (plochy) v $\mathbb{R}^2$, tj. $\phi = \nu_2(K)$. Konvexní obrazec K protneme sítí rovnoběžných přímek $F_1(x)$ vzdálených o $\mathrm{d}x$ procházejících body x jejich společné normály n. Ty, které zasahují K, mají bod x v intervalu $[x_1(K), x_2(K)]$ délky $h_1(K|n) = \nu_1([x_1(K), x_2(K)])$, který je průmětem obrazce K do normály n. Plochu pak aproximujeme jako sumu součinů délek tětiv $\nu_1(F_1(x) \cap K)$ a jejich vzdáleností $\mathrm{d}x$ a od sumace přejdeme k integrálu

$$\nu_2(K) = \int_{x_1}^{x_2} \nu_1(F_1(x) \cap K) \mathrm{d}x = h_1(n, K) \boldsymbol{E} \nu_1(F_1(x) \cap K), \tag{1}$$

což je formule nahoře uvedeného typu s tím rozdílem, že protínající přímky nejsou izotropní, ale mají pevnou orientaci danou normálou n. Konstanta úměrnosti c je v tomto případě rovna jedné.

Délku projekce $h_1(K|n)$ však při aplikaci vztahu (1) znát nemusíme, neboť obrazec lze protnout takovým svazkem rovnoběžných přímek (ekvivalentně rozšířit integrační obor), že interval $[x_1, x_2]$ délky h tímto svazkem na normále n vytknutý obsahuje celý interval $[x_1(K), x_2(K)]$. Přímky, které K neprotínají, do sumace ani k hodnotě integrálu nepřispívají, a pokles střední hodnoty je plně kompenzován tím, že $h > h_1(K|n)$. Tak dostáváme místo rovnice (1)

$$\nu_2(K) = h \boldsymbol{E} \nu_1(F_1(x) \cap K). \tag{2}$$

Přirozeně ovšem je účelné volit intervaly $[x_1(K), x_2(K)]$ a $[x_1, x_2]$ co nejbližší, tzn. že ze souboru hodnocených řezů právě jen oba krajní řezy (vzdálené o h) mají s K prázdný průnik.

Zobecnění na trojrozměrný objem $\nu_3(K)$ je analogické, lze však měřit jak délky tětiv $\nu_1(F_1(x) \cap K)$ tak plochy rovnoběžných rovinných řezů $\nu_2(F_2(x) \cap K)$ generovaných souborem rovnoběžných přímek, a jejich střední hodnoty násobit projekčními mírami H_2 (plocha vytknutá souborem přímek v jejich normálové rovině; musí obsahovat ortogonální projekci K do této roviny) resp. H_1 (interval vytknutý rovnoběžnými rovinami na jejich normále; musí obsahovat ortogonální projekci K do této normály). Dostaneme pak

$$\nu_3(K) = H_1 \boldsymbol{E} \nu_2(F_2(x) \cap K) = H_2 \boldsymbol{E} \nu_1(F_1(x) \cap K). \tag{3}$$

Postup lze zřejmě použít i v případě, že těleso K je uvnitř nějakého jiného (obvykle neprůhledného) tělesa libovolného tvaru. Stačí pouze zajistit, aby těleso K leželo celé uvnitř testované oblasti (obě okrajové roviny resp. soubor obvodových přímek jej nezasahují).

Často řešená úloha je následující. Konvexní obrazec K obsahuje jiný konvexní obrazec $K' \subset K$; chceme odhadnout poměr jejich ploch. Přímky protínající obrazec K protínají i K', takže pro obě plochy $\nu_2(K)$, $\nu_2(K')$ můžeme použít rovnici (1). Vydělením obou rovnic dostaneme pro poměr (tzv. plošný podíl K' v K, standardní značení $A_A(K', K)$)

$$\begin{aligned} \frac{\nu_2(K')}{\nu_2(K)} &= \frac{\boldsymbol{E}\nu_1(F_1(x) \cap K')}{\boldsymbol{E}\nu_1(F_1(x) \cap K)} = \boldsymbol{E}\alpha_x \frac{\nu_1(F_1(x) \cap K')}{\nu_1(F_1(x) \cap K)} \\ &= \boldsymbol{E}_w L_L(F_1 \cap K', F_1 \cap K), \end{aligned} \tag{4}$$

kde váha $\alpha_x = \nu_1(F_1(x) \cap K)/\boldsymbol{E}\nu_1(F_1(x) \cap K)$, $L_L(F_1 \cap K', F_1 \cap K)$ je lineární podíl tětiv K' na tětivách K a $\boldsymbol{E}_w$ značí, že střední hodnota je vážená. Vztah se podstatně zjednoduší, jestliže $\nu_1(F_1(x) \cap K)$ je konstantní a $\alpha_x = 1$. Jestliže např. K je čtvercové pozorovací okénko a $F_1(x)$ jsou přímky rovnoběžné s jeho stranou a, je $\nu_1(F_1(x) \cap K) = a$ pro všechny $F_1(x)$ zasahující K a vztah (4) nabude tvaru

$$A_A(K', K) = \boldsymbol{E} L_L(F_1 \cap K', F_1 \cap K). \tag{5}$$

Zcela identickým postupem lze odvodit analogický vztah pro trojrozměrný objekt K' obsažený v krychli K a protínaný rovinami $F_2(x)$ rovnoběžnými s jejími stěnami či přímkami $F_1(x)$ rovnoběžnými s jejími hranami nebo body $F_0(x)$ zasahujícími K. Objemový podíl K' v K je

$$\begin{aligned} V_V(K', K) &= \boldsymbol{E} A_A(F_2 \cap K', F_2 \cap K) = \boldsymbol{E} L_L(F_1 \cap K', F_1 \cap K) \\ &= \boldsymbol{E} P_P(F_0 \cap K', F_0 \cap K), \end{aligned} \tag{6}$$

kde $P_P(F_0 \cap K', F_0 \cap K) = \nu_0(x \cap K' | x \cap K \neq \emptyset)$. Všechny uvedené vztahy zůstávají v platnosti i v případě, kdy K' je např. sjednocení konečného počtu konvexních částic a vztah (6) bývá nepřesně slovně vyjadřován jako rovnost objemového, plošného, čárového a bodového podílu. Ve skutečnosti se jedná o rovnost obecně vážených středních hodnot jednotlivých podílů.

Při řešení historických úloh jsme dospěli ke vztahům mezi systémem testovacích čar a dalším čárovým systémem v rovině či systémem ploch v prostoru. Podobně jako lze v rovině testovat jeden čárový systém jiným čárovým systémem, lze v prostoru testovat jeden systém ploch jiným systémem ploch. Na průniku $F_2(x) \cap K$ jsme dosud sledovali plochy $\nu_2(F_2(x) \cap K)$ obsahující informaci o objemu $\nu_3(K)$. Podobně $\nu_1(\partial(F_2(x) \cap K))$ obsahuje informaci o povrchu K, musíme však sledovat průniky K systémem rovin různé orientace. Analogií vztahu (3) je

$$\nu_2(\partial K) = \frac{4}{\pi} H_1 \boldsymbol{E} \nu_1(\partial(F_2(x) \cap K)), \tag{7}$$

tj. hranice ∂K indukuje v rovinách F_2 uzavřené křivky $\partial(F_2(x) \cap K)$, jejichž střední délka je úměrná plošnému obsahu ∂K.

Obecně, mějme v určité omezené oblasti soubor ploch P s plošnou intenzitou $S_V(P)$ (střední hodnota plošného obsahu P v jednotce objemu) a testovací systém Q s plošnou intenzitou $S_V(Q)$ v téže oblasti. P v Q indukuje čárový systém $P \cap Q$ délkové intenzity $L_A^Q(P)$. Pak $\frac{4}{\pi} L_A^Q(P)$ je rovno $S_V(P)$, pokud alespoň jeden ze souborů P, Q je izotropní. Naopak také $S_V(Q) = \frac{4}{\pi} L_A^P(Q)$, kde $L_A^P(Q)$ je čárová intenzita $P \cap Q$ v P. Jedná se opět o velmi obecné stereologické vztahy podobné těm, které jsme dostali pro dva čárové systémy v rovině ($L_A = \frac{\pi}{2} N_L$) a pro čárový systém s plošným v prostoru ($S_V = 2N_L,\ L_V = 2P_A$).

Můžeme tedy konstatovat, že geometrické výběry obsahují řadu informací o vlastnostech objektu K, z nějž byly pořízeny. Typ informace je určen dimenzí d prostoru, do nějž je objekt vnořen a v němž je výběr uskutečněn, dále dimenzí k objektu K a konečně dimenzí souboru r uskutečňujícího výběr. Jestliže $k = d$, pak bodový výběr obsahuje informaci o Lebesgueově míře $\nu_d(K)$, výběr čárovým systémem obsahuje informaci o $\nu_d(K)$ a také o míře hranice $\nu_{d-1}(\partial K)$ (to už je ovšem míra Hausdorffova), výběr uskutečněný nadplochou (nadrovinou) dimenze $d-1$ obsahuje kromě všech předcházejících informací také informaci o střední šířce $H_1(K)$. Pokud provádíme výběr ze souboru objektů a chtěli bychom určit jejich počet v analyzované oblasti, musí být obecně výběr d-rozměrný, jinak je nezbytné přijmout nějaké předpoklady o tvarech zkoumaných objektů. Jestliže je dimenze $k < d$, bodový výběr nepřináší žádnou informaci, protože pravděpodobnost, že objekt zasáhne, je nulová téměř jistě. Požadovaná dimenze souboru provádějícího výběr se pak pro všechny vlastnosti zvyšuje, např. při $k = d - 1$ (kruhy rozmístěné v $\mathbb{R}^3$ nebo úsečky v $\mathbb{R}^2$) podávají informaci o funkcinálech $\nu_k(K)$ soubory dimenze $r \geq 1$.

Pro ilustraci těchto pravidel předpokládejme, že analyzujeme soubor trojrozměrných částic v oblasti objemu V. Na rovinných řezech tímto souborem budou řezy částicemi (profily). Plochy profilů obsahují informaci o objemech částic (jejich plošný podíl je odhadem objemového podílu částic), obvody profilů informují o celkovém obsahu hranic částic a počet profilů kombinuje informace o šířkách a počtech částic. Pro odhad počtu částic bychom museli určit jejich počet N v nějaké menší oblasti objemu ΔV a odhad jejich celkového počtu (za předpokladu, že jsou rozmístěny rovnoměrně náhodně v celé oblasti) by byl $NV/\Delta V$. Analogicky můžeme soubor testovat systémem přímek: pak délky tětiv budou obsahovat informace o objemech částic a jejich počet bude kombinovat informace o obsahu hranic částic a jejich počtu. Jakákoliv informace o střední šířce částic je ztracena. Konečně bodový soubor může podat pouze kombinovanou informaci o objemech a počtu částic, konkrétně o jejich objemovém podílu. Obecné pravidlo tedy zní: čím je vyšší dimenze souboru uskutečňujícího výběr, tím více informací získáme. Proti němu ovšem stojí snaha o minimální časovou i pracovní náročnost, takže pokud nás zajímá pouze objemový podíl, dáváme přednost výběrům bodovými soubory. Všechna tato poněkud komplikovaná pravidla jsou důsledkem skutečnosti, že geometrické výběry nejsou prosté, nýbrž vážené vlastnostmi částic: o bodovém výběru rozhoduje objem částice, o čárovém výběru míra plochy její hranice a o výběru nadplochou střední šířka částice. Tato základní vlastnost geometrických výběrů byla podrobně rozebrána v práci [1].

Detailní odvození vztahů, diskusi a příklady použití lze nalézt v připojeném seznamu literatury.

Literatura

[1] Saxl I. a Beneš V. (2002) *Geometrický výběr*. In: *Výuka statistiky v České republice I*. Matfyzpress, Praha, 79–89.

[2] Santaló L.A. (1976) *Integral Geometry and Geometric Probability*. Addison-Wesley, Reading (Mass.).

[3] Saxl I. (1989) *Stereology of Objects with Internal Structure*. Academia, Prague & Elsevier, Amsterdam.

[4] Stoyan D., Kendall W.S. a Mecke J. (1995) *Stochastic Geometry and its Applications*. J. Wiley & Sons, New York.

[5] Weibel E. (1995) *Stereological Methods, Vol. 2*. Academic Press, London.

DOPLNĚK TIŠTĚNÝCH KNIH: INTERNET

Pavel Stříž

FaME UTB, Mostní 5139, 760 01 Zlín
`striz@fame.utb.cz, pavel.striz@email.cz`

Motto: Uměním není tvorba, když k tomu mám *skoro* vše.
Umění je tvorba, když k tomu nemám *skoro* nic.
Takovou tvorbu však nemá každý *tolik* rád.

Možná bych mohl začít i nadpisem *Doplněk Internetu: Tištěné materiály*. Mé zkušenosti jsou takové, že na Internetu jsem vždy našel o trochu více než jsem byl schopen „vydolovat“ z naší knihovny a zdrojů mně jinak dostupných.

1. Úvodem : Co nás na Internetu nesmí překvapit

Internet nám přináší řadu výhod, a to nejen k výuce statistiky, ale především nám zpřístupňuje různé druhy informací (levně?). Odborníci a akademici, kteří řeší stejné či podobné problémy, mohou myšlenky a výstupy svých prací prezentovat na Internetu skoro okamžitě. I přes to vše vidím nevychytané mouchy, na kterých se pracuje a jejich řešení několika lidem značně usnadní život.

Upozorním především na mezery v citování, souběžné prezentování datových souborů na Internetu a některá specifika datových a popisných souborů určených k publikování.

V této souvislosti si vždy vybavím – pro potřeby citací – státní normy `ČSN ISO 690` a `ČSN 01 0185` a jak s těmito normami zápasí studenti u diplomových prací. Citování zdrojů na Internetu (více o tom snad v samostatném článku) je značně nestabilní záležitostí. Dvě třetiny odkazů nefunguje a druhá polovina je po *drobných* zásazích jiná než byla minulý rok (minulý týden; no nepřehánějme, včera ve `23:59:57` hodin). A navíc budu čtenáři tvrdit, že by měl přes Internet publikovat své datové soubory. Zachovejme však zatím chladnou rozvahu.

2. Mám nápad. Jak začít?

Vezměme v potaz, že chci prezentovat své statistické výstupy na Internetu a zároveň chci prezentovat některé své datové soubory – např. tabulky, ze kterých mohu tvořit statistické propočty a kreslit grafy.

Klíčová slova: Data v publikacích, soubory s daty, knihovny dat, zdroje na Internetu.
Poděkování: Rád bych poděkoval Dr. Vladimíru Rytířovi za podnětné diskuze.

Není úplně špatný začátek podívat se po internetovém světě a zjistit, jak tento *problém*[1] řeší ostatní. Zeptáte-li se mě, pak vám doporučím své oblíbené internetové odkazy, internetové rozcestníky. Tím třeba nic nezískáme, ale určitě toho mnoho nepokazíme.

Za anglicky psané portály a rozcestníky bych zmínil domovskou stránku pana Hosseina Arshama (7i). Z českých rozcestníků bych zmínil odkazy České statistické společnosti (15i), statistického sysla (6i) nebo odkazy pana Vošvrdy (VosvrdaWeb; 8i). Za bližší průzkum určitě stojí i portál a rozcestník Caramba (11i) atd.

Otázka: Kdo mi zaručí, že *tato URL* pojede i za několik let?
Odpověď: Bohužel v této chvíli nikdo. Je to citování „na vlastní nebezpečí".

3. Domovské stránky (Home pages)

Datové soubory najdeme na řadě místech. Zaměřím se na osobní stránky odborníků a akademiků. Důvod je jednoduchý. Takového odborníka či akademika považuji za základní statistickou jednotku. Firmy, společnosti atd. jsou už něco více. Struktura osobní stránky může být např. takováto:

- *Představení* (Introduction)
- *Rodina* (Family)
- *Výuka* (Teaching)
- *Výzkum* (Research)
- *Internetové odkazy* (Links)

4. Knihy a články

Výuka a výzkum nás zajímají přednostně. Jak jsou tyto položky nejčastěji organizovány již přeskočím. Avšak jednou z podpoložek jsou bezesporu tištěné a publikované materiály.

Většina knih a učebnic je natolik hodnotných a cenných, že stojí za to je na Internetu alespoň představit, když už ne celé prezentovat. Seznámení s knihou se děje zhruba v takovém scénáři:

- *Soubory s daty* (Data files)
- *Přivítání* (Welcome)
- *Předmluva, poděkování, osnova* (Preface, Acknowledgments, Table of Contents)
- *Soubory s daty, materiály ke stažení* (Data files, Downloads)

[1] Většinu problémů řeší souběžně více odborníků. Pravda někdy pod jinými názvy.

- *Komentáře a recenze* (Comments on, Reviews)
- *Rejstříky, literatura atd.* (Lists, References etc.)
- *Opravenka* (Errata)
- *Objednání knihy* (Ordering the book)

U internetových prodejců knih je struktura trochu jiná a také poskytované informace se liší kniha od knihy. Největšího knižního rozpětí dosáhli zatím na Amazonu; internetovém obchodu (9i).

5. Datové soubory

Vhodně uložená data v souborech jsou klíčem k rychlému a *bezbolestnému* přístupu. Pokud jsou soubory ve formátu `.txt`, `.dat`, bez přípony nebo obecně zapsány v `ASCII` formátu není (obvykle) žádný problém.[2]

Problém nastává (desetinné `. / ,` a formáty rozdělování řádků a sloupců {„ "; " "; ' '; ` `; ␣; `Tab`; `Enter` nebo `: / ;` a další znaky} beru jako perličku):

- U popisů cizojazyčných (obecně: v jiném než v `ASCII` formátu).
  Řešení: Tyto popisy dáváme do zvláštních souborů nebo článků.
- U souborů ne přímo určených k bádání (formáty zvukové, obrazové).
  Řešení: Musíme tyto formáty upravit do formy, se kterou umíme již pracovat; zmínil bych o tom snad více v navazujícím článku.
- U souborů komprimovaných nám již známým (statistickým) balíkem.
  Řešení: Nějakou cestou bych měl program získat a soubor v něm otevřít (pokud zrovna nemám chuť zkoumat metodu komprese). Nebo využít jiného programu a funkce Import/Export a převést si to do zvoleného formátu. Je nutné vědět, který program umí co otevřít.[3]
- A jeden z nejhezčích problémů nastává u dat, která jsou zabudována přímo do programů s jistým, nejčastěji ukázkovým, algoritmem.
  Řešení: Někdy stačí příkazy programovacího jazyka jen vymazat a vytvořit si

[2] Rád bych vyjmenoval datové formáty (programy jsou chráněny příslušnými zákony), které používá internetové vydavatelství (12i) vedle `.txt`, `.dat` (`ASCII` struktura) a `.xls` (*Microsoft Excel*). Jsou to `.dta` (*StataQuest*), `.mtp`, `.mtw` (*MiniTab*), `.sav` (*SPSS*), `.xpt`, `.tpt` (*SAS*), `.sf3` (*Statgraphics for Windows*), `.jmp` (*JMP-IN*), `.wrk` (*StatIT*), `.83*` (*TI83*), bez přípony (*S-Plus*) a `.mc4` (*MicroCase*). Jedinečnost (u knih `ISBN`) by se měla nějakým způsobem vztahovat i na soubory.

[3] Pokud by se čtenář dostal do problémů s jinými formáty (příponami souborů), a který kam patří, může mě kontaktovat. Pokusil bych se poskytnout pomocnou ruku. Pamatuji si z dřívější internetové četby odkaz na datový soubor mě tehdy neznámého formátu. Po jisté době testování všech možných i nemožných programů jsem zjistil, že jsem ztracen. Téměř po dvou letech (úplně jinou cestou) jsem zjistil, že je to formát malého programu vytvořený učitelem pro jeho studenty. V našich podmínkách bychom řekli program na jednosemestrové použití. Program byl přístupný na síti, ale jen pro zmíněné studenty jistých studijních kruhů. . . (všemocné to lokální sítě).

vlastní datový soubor v `ASCII` formátu, pokud máme k dispozici zdrojový program.[4] A někdy stačí trochu umu a použít programování k extrakci těchto dat při spuštění upraveného programu.

6. Komentář datových souborů

Vrátím se k prvnímu bodu předchozího odstavce. Takto by mohl vypadat popis datového souboru určeného k publikování:

- *Název (Name)* Někteří stihnou přečíst jen název souboru, proto se snažíme být výstižní, jak to také bývá dobrým zvykem u názvů článků a kapitol knih.
- *Druh dat (Type)* Zda-li se jedná o simulované hodnoty, časové řady, nějaké druhy experimentů, sportovní výsledky apod.
- *Velikost (Size)* Počet pozorování a proměnných. Celková velikost v bytech bude výhodná u rozsáhlých databází a u souborů s kompresí či komprimací.
- *Stručný popis (Descriptive abstract)* Snažíme se podat dostatek informací a výstižně popsat, co dále následuje. Doporučení je ne více než 10 řádků.
- *Zdroje dat a informační prameny (Sources)* Důraz je kladen na publikovaná data. Nesmíme zapomenout, že na knihy a články se čtenáři odkazují a na práce navazují. Autoři datových souborů přesně neví, co všechno budou zájemci o tento datový soubor potřebovat a co budou chtít vědět. Proto se snažíme být co možná nejpřesnější.
- *Popisy proměnných (Variable descriptions)* Tvůrce souboru by měl podat podrobný popis proměnných a jednotek dat. Měl by následovat návod jak se souborem pracovat, zda-li nechybí nějaké hodnoty apod.
- *Doplňující poznámky (Special notes)* Zde je místo na podrobný popis všeho, co by mělo být ještě o datech řečeno a co by mohlo být přehlédnuto.
- *Co se za daty skrývá (Story behind the data)* Autora tohoto článku nejoblíbenější pasáž ke čtení. Vždy mi to připomíná zadání slovních úloh (nejen z matematiky na ZŠ a SŠ). Zde by tvůrce datového souboru měl sdělit vše, na co by se případný čtenář mohl zeptat. Pokud by tato část měla být značně dlouhá a vyčerpávající, je vhodné k tomu připsat samostatný článek. [5]
- *Pohled pedagogický (Pedagogical notes)* Tato část je určena učitelům a řešitelům. Autor datového souboru by měl vysvětlit k čemu tato data sbíral a proč je

[4] Výčet nemůže být úplný, ale rád bych zmínil (programy jsou chráněny příslušnými zákony): `.bas` v programovacím jazyce *BASIC*, `.c` v jazyce *C*, *C++*, `.java` v objektovém programovacím jazyce *JAVA*, `.m` u programu *MATLAB*, ale i vlastní `www` stránky v `.htm` a `.html` formátu. Nevyjímaje `.tex` formát typografického sázecího systému *TEX*.

[5] This and the next section should be fairly concise. If you find them getting too long – it's time to write a full "Datasets" article! (10i → /v1n1/datasets.html)

prezentoval. Prezentace *správného* modelu a řešení (pokud je *známé*) není určitě na závadu.

- *Literatura (References)* Zde by měly být uvedeny použité zdroje a zdroje, které by mohly být nápomocny a mohly by „zoufalého“ řešitele popostrčit směrem vpřed.

- *Odesílatel (Submitted by)* Základní údaje o odesílateli dat a zvláště kontakt na něj – možnost se dotazovat, problém prodiskutovat či jen upozornit na drobnou chybičku.

Upraveno z přílohy internetového článku `10i->/v1n1/datasets.html` z roku 1993.

Otázka: Kdo mi zaručí, že soubory budou (obsahově) stejné?
Odpověď: Nikdo. Je to vizitka autorů a příslušných vydavatelství.

7. Články s datovými soubory

Jak by to tedy mohlo fungovat? Po článku by měl následovat odkaz na datový soubor a jeho popis. Uvedu příklady. Články můžeme nalézt na (10i → /jse_index.html) a datové soubory s popisy na (10i → /datasets/jse_data_archive.html). Články tady spravuje vydavatel a zároveň editor.[6]

8. Knihy s datovými soubory

Tato část je velmi důležitá, protože si stále nedokáži představit výuku statistiky bez nějakých podkladů. A už vůbec ne bez alespoň nějakých dat (ať už simulovaných, reálných atd.).

Snahou by mělo být data u knih prezentovat nejen v tištěné formě, ale alespoň na disketě, CD-ROM či DVD. Ideální by bylo mít tyto data spravovaná na Internetu vydavatelem knihy. Struktura by mohla vypadat asi takto:

Přípony souborů			Umístění	Charakteristika a druh souborů dat
`.txt`	`.xpt`	`.bas`	Kapitola 1	Vzorové úlohy a jejich vyřešení.
`.xls`	`.wrk`	`.c`	Kapitola 2	Cvičení [Kontrolní výsledky].
`.sf3`	`.sav`	`.java`	Kapitola 3	Vzorové případové studie a jejich vyřešení.
`.mtw`	`.jmp`	`.html`	Kapitola 4	Cvičné případové studie [Výsledky].

[6] Čtenáři bude možná známější sbírka ResearchIndex (1i). Nalezneme zde soubory `.pdf`, `.ps`, `.dvi`, `.djvu` a image (stáhnout lze do `.gifu`). Komprimované soubory jsou v `.gz` a `.z`. Literatura k mému článku na ROBUSTu 2002 je přístupná právě z této internetové knihovny. Technicky a prakticky nemám žádný citovaný zdroj v originální papírové formě.

`.dat` `.mc4` `.tex`	Kapitola 5	Další materiály: Meze se nám nekladou.
⋮	⋮	⋮

Alespoň jeden datový formát by měl být v `ASCII` nebo ve formátu, s jehož otevřením a přečtením by čtenář neměl mít problémy.

8.1. Ukázky z anglických materiálů

Uvedu jeden příklad, na kterém jsem se cvičil ve statistických propočtech před svojí bakalářskou prací. Cvičil jsem se v četbě ACF, PACF a v ARIMA modelování. Stále totiž zastávám názor, že praktické zkušenosti statistika jsou někdy mnohem více než znalost všech dostupných algoritmů.

Příklad je z knihy Hyndmana a kol. (2i), [3]. Byl to běh mužů na 400 metrů na Olympijských hrách (Tabulka 1; 1, 2; nebo z http://www.olympic.org).

Výstup sice není nejlepší, ale je dostatečně vhodný pro následné diskuze (otázky lékaře, sportovce a statistika o tom, kde jsou možné hranice *modelů*).

V krátkém časovém horizontu je i dosti přesný. Moje předpověď na rok 2000 byla 43,1 sekund (Tabulka 1; 4; z hlediska smyslu nemají další desetinná místa v článku význam). Skvělý Michael Johnson vyhrál v čase 43,18 sekund. Intervalové odhady seděly (Tabulka 1; 5, 6). Radost to byla veliká, statistická.

Ovšem zpět do reality. Pan Hyndman upravil mezi roky 1999 a 2003 svůj internetový datový soubor. Nejen že změnil rozdělovač sloupců z tabulátoru na mezeru, ale zvláště upozorním na roky 1908, 1980 a 1992. Jednou to vypadá, že bylo měřeno na jedno a dvě desetinná místa a podruhé měřeno jinak. Z hlediska měření je v tomto rozdíl.

Sportovní nadšenec tuší, že hodnoty z roku 2003 jsou správné, úplné. Pokud bych však vytvořil drobný program na vstupní soubor roku 1999, v roce 2003 budu zklamán a budu muset změnit citaci a malinko předělat svůj program. To se týká souboru na Internetu. Tuším, že v různých vydáních knih není problém: cituji vydání, ze kterého data přebírám.

Soubor *olympic.dat* z roku		Předpovědi modelu				Data vydavatelství (12i)		
1999	2003	rok	stř.	dol.	hor.	©	čet.	vydání
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1896 → 54.2	1896 ␣ 54.2	2000	43,1	40,5	45,7	2004	10	1. 27
1900 → 49.4	1900 ␣ 49.4	2004	43,2	40,6	45,8	2003	9	2. 17
1904 → 49.2	1904 ␣ 49.2	2008	42,6	39,9	45,3	2002	11	3. 4
1908 → 50	1908 ␣ 50.0	2012	42,6	39,9	45,4	2001	12	4. 11
1912 → 48.2	1912 ␣ 48.2	2016	42,6	39,8	45,4	2000	8	5. 6
1920 → 49.6	1920 ␣ 49.6	2020	42,8	39,9	45,7	1999	6	6. 5
1924 → 47.6	1924 ␣ 47.6	2024	42,3	39,4	45,2	1998	8	7. 1
1928 → 47.8	1928 ␣ 47.8	2028	42,4	39,4	45,4	1997	2	8. 1
1932 → 46.2	1932 ␣ 46.2	2032	42,5	39,4	45,5	1996	1	9. 1
1936 → 46.5	1936 ␣ 46.5	2036	42,7	39,6	45,9	1995	0	
1948 → 46.2	1948 ␣ 46.2	2040	42,3	39,1	45,5	1994	0	
1952 → 45.9	1952 ␣ 45.9	2044	42,5	39,2	45,7	1993	2	
1956 → 46.7	1956 ␣ 46.7					1992	1	
1960 → 44.9	1960 ␣ 44.9	Poznámka:				1991	1	
1964 → 45.1	1964 ␣ 45.1	Jedná se o 95%-ní inter-				1990	2	
1968 → 43.8	1968 ␣ 43.8	valové odhady.						
1972 → 44.66	1972 ␣ 44.66					Poznámka:		
1976 → 44.26	1976 ␣ 44.26					Celkem je nabízeno		
1980 → 44.6	1980 ␣ 44.60					73 knih a publikací		
1984 → 44.27	1984 ␣ 44.27					s datovými soubory.		
1988 → 43.87	1988 ␣ 43.87							
1992 → 43.5	1992 ␣ 43.50							
1996 → 43.49	1996 ␣ 43.49							

Tabulka 1. Soubor `olympic.dat` (2i → /data/olympic.dat; velikost souboru je 256 bytů) z různých let (1, 2); jeden z ARIMA modelů – předpověď s intervaly spolehlivosti (3, 4, 5, 6) a zajímavosti z internetového knihkupectví (12i; 7, 8, 9): jedna z možných budoucích cest citování a uchovávání dat z knih a článků na Internetu.

8.2. Ukázky z českých materiálů

Zmíním dva české zdroje s odkazy na statistické učebnice. Prvním zdrojem jsou cvičebnice s datovými soubory na Internetu. Čtenář si může projít materiály o ekonomických časových řadách [1, 2] a může si stáhnout některá data od jednoho z autorů, od pana Arlta (5i).

Druhým zdrojem je učebnice od pánů Melouna a Militkého [5] s navazující Sbírkou úloh (s disketou) [4], ze které odcituji jedno ze cvičení. Tedy příklad se vším všudy tak, jak to student může mít jedině jen rád (název a zadání úlohy, hodnoty a kontrolní výsledky).

Úloha V6.20 *Obsah kyseliny sírové v nitrační směsi dvěma metodami.*
Pro stanovení obsahu kyseliny sírové v nitrační směsi byla připravena série umělých vzorků. Vzorky byly analyzovány jednak s použitím titrátoru DL 40 RC, jednak manuelně. Předpokládá se, že obě metody poskytují výsledky se shodným rozptylem. Rozhodněte, zda obě metody poskytují srovnatelné výsledky.

x	y	x	y	x	y
75,073	75,072	75,499	75,407	75,993	76,299
74,349	74,418	75,366	75,316	75,531	75,909
74,527	74,497	75,469	75,511	75,954	75,977
74,866	74,898	75,558	75,490	76,131	76,103

Data. Stanovení obsahu titrátor x [%], manuelně y [%].

Kontrolní hodnoty:
$r = 0,9718, D = 94,44\%, y = -3,59(6,06, A) + 1,048\ (0,080) \times x, 1\ o, 1\ e$

/floppy/Data.zip/Excel/6kapitol/V6.xls V620x: 75,073 74,349 74,527 74,866 75,499 75,366 75,469 75,558 75,993 75,531 75,954 76,131

/floppy/Data.zip/Excel/6kapitol/V6.xls V620y: 75,072 74,418 74,497 74,898 75,407 75,316 75,511 75,49 76,299 75,909 75,977 76,103

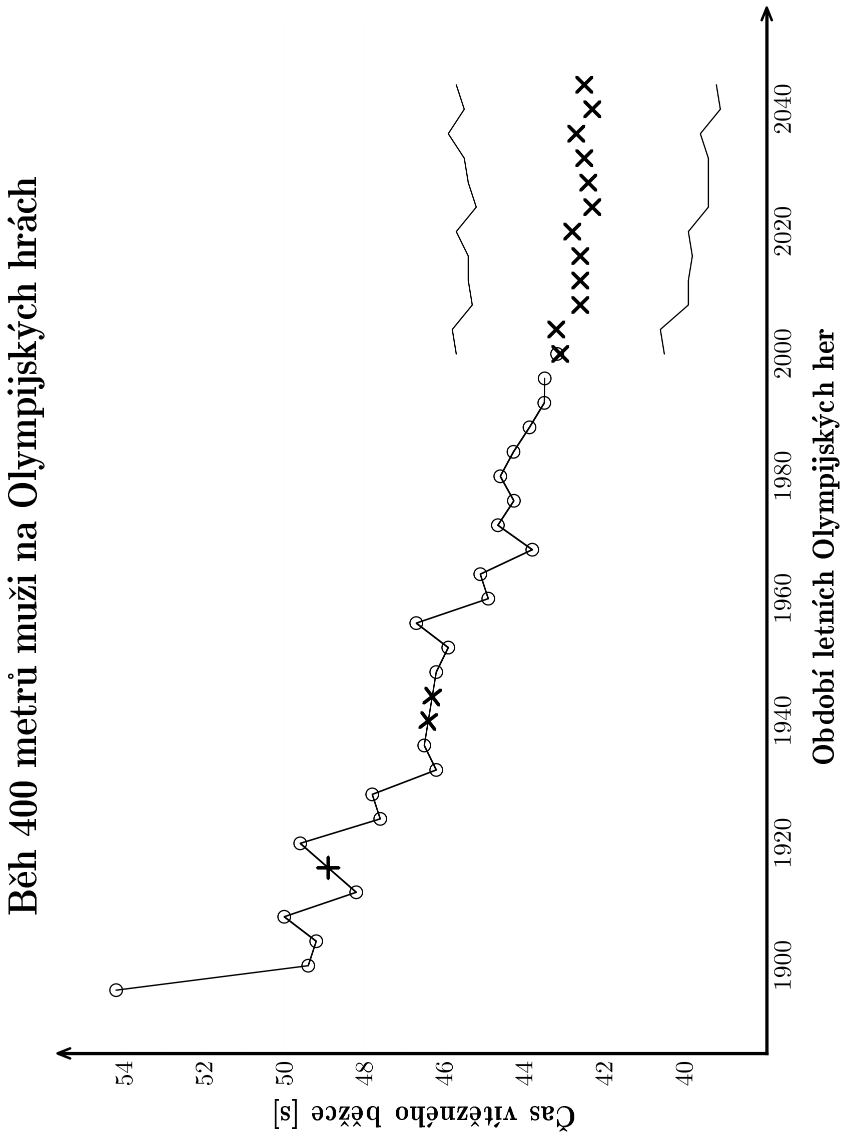
Zdroj: Přepis je použit ze cvičebnice pánů Melouna a Militkého [4, str. 190]. Datový soubor lze nalézt na disketě ke cvičebnici [4] a na Internetu (4i). Metodu řešení lze nalézt v učebnici (české statistické Bibli *nestatistiků* pánů Melouna a Militkého [5]. *Otázka: Kolik peněz chtít za zpřístupnění takových dat?*

Odpověď: Záleží na autorech, spravovatelích databází a vydavatelích knih.

9. Internetové knihovny s datovými soubory

Většina z nás má své oblíbené sbírky. Jednou z mých oblíbených je sbírka při Carnegie Mellon University (3i). Vedle této sbírky zmíním rozcestník u StatSci (14i). Datové soubory nalezneme i u většiny známých programů.

Kvůli čemu je tedy tento poprask? Ve statistice například kvůli genetice a finančním časovým řadám a obecně k ukládání rozsáhlých databází *v podstatě* čehokoliv (dolování v data miningu je dobrý příklad meziodvětvového bádání). Už si nelze (ne že bychom dříve mohli) vytisknout více Ecommerce knih nebo delších materiálů (13i; např. o genetice: Human Genome Project; `*hgp*.*`).



Graf 1. Ukázka ARIMA modelování a jeho 95% intervaly spolehlivosti.
Cvičení 5.7 (2i), [3]. Údaje let 1916, 1940 a 1944 chybí, protože tehdy se bádalo jinak.

10. Pohled snad směrem vpřed

Autor promlouvá sám k sobě: Rád bych citoval internetové články a knihy svých oblíbených autorů.

Autor si odpovídá: Internet to umožňuje. Jen my všichni musíme z části Internetu udělat knihu, ve které nemizí písmenka a nemění se stránky, když už jsem se jednou rozhodl *toto* internetové místo citovat k jistému datu.

Dočasným řešením je pro mě ukládání `www` stránek, které používám ve své práci a článcích, včetně tohoto, na diskety, CD-ROM, DVD,. . .

11. Závěrem : Kdy mohu přestat být klidný?

Na tento článek se pokusím navázat článkem o datových souborech, které tvoří rozsáhlé databáze, ale ten začne vznikat až ve chvíli, kdy autor bude mít po ruce více než jen modem a připojení přes Český Telecom.

Například z takového připojení stáhnout 10 MB, 10 GB (nyní mohu být drobně neklidný) či 10 TB (a nyní ještě o trochu více) je slušná noční můra každého (ne)statistika.[7]

11.1. Internetové zdroje : Odkazy byly funkční k 31. srpnu 2003.

(1i) `http://citeseer.nj.nec.com`
(2i) `http://go.to/forecasting`
(3i) `http://lib.stat.cmu.edu/datasets`
(4i) `http://meloun.upce.cz/add_on/meloun/c3/c3disk.htm`
(5i) `http://nb.vse.cz/~arlt/hlavni/odbzdr.htm`
(6i) `http://ssysel.hyperlink.cz`
(7i) `http://ubmail.ubalt.edu/~harsham`
(8i) `http://vosvrdaweb.utia.cas.cz/TEST/pages/data.htm`
(9i) `http://www.amazon.com`
(10i) `http://www.amstat.org/publications/jse`

[7] A hned tu máme nový problém: Už nebude stačit ani DVD. Co potom? Náznak řešení měl ukázat tento článek: Ukládání na Internetu na pevných discích serverů (podle Huber-Wegmanova systému dělení jsou na poslední příčce distribuované knihovny dat (Distributed Data Archives) s 10^{15} bytů; řádově to jsou petabyty; 1 PB $= 2^{50}$ bytů $\approx 10^{15}$ bytů).

Rozloučím se statistickým humorem: Odborného asistenta na pitevně se táží, zda-li tuší co značí počet kulek v těle oběti. Již všeznalý odborný asistent ví, zná a odpovídá:

Když je v těle oběti jedna kulka, vrahem byl pravděpodobně profesionální odstřelovač.

Když jsou v těle oběti 2, 3 či 4 kulky, vrahem byl pravděpodobně profesionální zabiják.

A když je v těle oběti více kulek, vrahem byl jistě profesionální statistik.

(11i) `http://www.caramba.cz`
(12i) `http://www.duxbury.com` → `Data Library; Online Book Companions`
(13i) `http://www.gutenberg.org` → `Index of all texts`
(14i) `http://www.statsci.org/datasets.html`
(15i) `http://www.statspol.cz/odkazy/index.htm`
(16i) `http://www.tenmax.com/teleport/pro/home.htm`

Literatura

[1] Arlt J. a Arltová M. (1997) *Příklady z analýzy ekonomických časových řad*. VŠE v Praze, Praha, ISBN 80-7079-056-3.

[2] Arlt J., Arltová M. a Rublíková E. (2002) *Analýza ekonomických časových řad s příklady*. VŠE v Praze, Praha, ISBN 80-245-0307-7.

[3] Hyndman R.J., Makridakis S. a Wheelwright S.C. (1998) *Forecasting: Methods and Applications, 3rd edition*. John Wiley & Sons, Inc., New York, ISBN O-471-53233-9.

[4] Meloun M. a Militký J. (1996) *Statistické zpracování experimentálních dat. Sbírka úloh*. Univerzita Pardubice, Pardubice, ISBN 80-7194-075-5.

[5] Meloun M. a Militký J. (1998) *Statistické zpracování experimentálních dat v chemometrii, biometrii, ekonometrii a v dalších oborech přírodních, technických a společenských věd. 2. vydání*. EAST PUBLISHING, Praha, ISBN 80-7219-003-2.

STAKAN 2003

JAKOU STATISTIKU UČIT NESTATISTIKY?

Josef Tvrdík

Přírodovědecká fakulta OU, 30. dubna 22, 701 03 Ostrava
tvrdik@osu.cz

Po téměř čtyřiceti létech potýkání se statistikou a třinácti létech jejího vyučování stále nenalézám spolehlivé odpovědi na celou řadu zdánlivě jednoduchých otázek spojených s výukou nestatistiků. Někdy se sice zdá, že jasná odpověď se nabízí, ale velmi často takovou odpověď nová zkušenost zpochybní. Některé takové otázky a pokusy o alespoň částečné odpovědi na ně jsou uvedeny v příspěvku.

1. Úvod

Otázka v názvu příspěvku obsahuje dva skryté předpoklady:

(i) existuje více statistik,

(ii) je jen jedna kategorie nestatistiků.

Zatímco první předpoklad je v pořádku (pokud je někdo přesvědčen, že existuje jen jedna statistika, stejně ji nestatistikům nelze vnucovat celou, takže rozhodně se musí z ní vybrat a má tedy smysl se ptát, co vybrat), druhý předpoklad evidentně splněn není. Nestatistici, kterých by se výuka statistiky měla/mohla týkat, jsou velmi různorodá skupina. Jednu krajnost zmiňuje Komenda [3] - porodní asistentky s jednosemestrálním kursem v bakalářském studiu, o druhém extrému píše Víšek [10] - ekonomové ve dvousemestrálním kursu na Institutu ekonomických studií FSV UK navazujícím na důkladně vyučovanou matematiku a následovaným třemi semestry ekonometrie. Výběr vyučovaných témat a zejména způsob výuky by měl být volen právě podle toho, kdo je vyučován. Názory a poznámky v dalším textu vycházejí především ze zkušeností z výuky v jednosemestrálním kursu pro nematematické bakalářské obory na Přírodovědecké fakultě, ve dvousemestrálním kursu pro studenty informatiky a postřehů z dlouholeté spolupráce s lékaři, biology, psychology a učiteli na statistickém zpracování jejich dat. Pokud někdo zjistí, že tyto poznámky mají o trochu širší platnost, budu příjemně překvapen.

Klíčová slova: Výuka statistiky, Excel versus jiný statistický software, data pro výuku, elektronické distanční vzdělávání.

Poděkování: Rád bych poděkoval řadě kolegů za podnětné, byť ne vždy souhlasné diskuse.

2. Neodbytné otázky

2.1. Koho učit a jak učit efektivně?

Zdálo by se, že odpovědi na tyto otázky jsou všeobecně známé, o vyučování statistiky vychází řada publikací, i u nás je kateder vyučujících statistiku mnoho, jejich počet se nesnižuje [11], o výuce statistiky se diskutuje na seminářích STAKAN i jiných, na kterých se schází pravidelně několik desítek statistikých kantorů. Proč ale značná část lidí, kteří statistiku potřebují pro svou práci a obrací se o pomoc, je naprosto nevybavená jak základními znalostmi, tak často ani představou, co vlastně lze statistikou řešit? Zřejmě ty potřebné adepty statistiky naše kursy buď minuly nebo výuka nezanechala dostatečnou stopu. Úroveň (ne)znalosti klientů často nezávisí na tom, zda a jaký kurs statistiky absolvovali, dokonce možná ani na tom, kdo je učil. Povzbudivé je snad jen to, že pokud výsledky statistické analýzy svých dat opravdu potřebují, většinou jsou schopni si nutný výsek znalostí docela rychle doplnit. Opravdová motivace dělá divy, které v základním kursu pro nestatistiky, zejména je-li zařazen do nižšího ročníku, pozorujeme velmi zřídka.

2.2. Napsat učební text nebo naučit?

Učebních textů o základech statistiky jak u nás, tak po světě, je snad až nadbytečné množství a přitom statistická vzdělanost absolventů kursů (bohužel včetně absolventů mých kursů) je často překvapivě děravá. Zřejmě je jednodušší napsat úvodní učební text o statistice než základům statistiky naučit studenty. Sám jsem také přispěl k inflaci úvodních učebních textů skripty [6], která byla celkem vlídně přijata jak kolegy, tak i většinou studentů a dokonce brzy vyprodána. O to větší je pak můj zármutek, když např. o testování hypotéz student u zkoušky sice podrobně převypráví úvodní příklad s detaily příběhu, ale k obecnějšímu pohledu na základní myšlenku testování hypotéz, k jehož snadnému pochopení měl úvodní příklad posloužit a který ve skriptech samozřejmě následuje, se už student nepropracuje. To pak člověk přestává věřit Komenskému, že výuka má jít od jednoduchého ke složitějšímu.

2.3. Čím začínat?

Zde jsou dvě možnosti donedávna v kursech pro nestatistiky u nás zhruba stejně četné: buď začít základy pravděpodobnosti nebo popisnou statistikou. Začínat od teorie pravděpodobnosti znamená nebezpečí, že odstrašíme všechny matematiku nemilující posluchače (mezi nestatistiky to bývá většina) a u ostatních přispíváme k vytvoření představy, že statistika je matematika a data vlastně nepotřebuje. Naopak, začneme-li deskriptivními metodami, studenti od začátku vidí praktickou stránku statistické analýzy dat. Ale je tu riziko, že studenti podlehnou prvnímu dojmu, jak statistika je snadná a pak buď podcení nebo odmítají následující myšlenkově náročnější témata nutná pro statistickou indukci. I přes tyto nevýhody však začít popisnými metodami považuji v kursu pro nestatistiky za vhodnější a jsem velmi rád, že tento názor začíná být mezi statistickými kantory většinový.

2.4. Hodně nebo důkladně?

V obvyklém jednosemestrálním nebo dvousemestrálním kursu pro nestatistiky je vždy nutno řešit dilema: Předložit povrchně širší paletu statistických metod (důkladněji to nelze stihnout) a doufat, že absolventi kursu si pochopení souvislostí později při aplikacích prohloubí (pokud to budou potřebovat)? Nebo raději méně metod, ale klást větší důraz na pochopení základních konceptů a doufat, že své znalosti statistických metod absolventi si rozšíří (pokud to budou potřebovat)?

Preferuji druhou možnost, cílem kursu by mělo být především u studentů rozvíjet dovednost „koukat se“ na data a tedy i na svět pomocí rozumně užívaných metod statistiky. K tomu patří z pravděpodobnosti naučit alespoň tolik, aby studenti pochopili, co je kvantil. Je to důležité při interpretaci počítačových výstupů, ať tato interpretace neskončí jen bezmyšlenkovitým rozhodováním podle mechanického pravidla. Dále naučit studenty vnímat rozdíl mezi parametrem a jeho odhadem, vysvětlit intervalový odhad, základy testování hypotéz a naučit je statisticky hodnotit vztah dvou veličin. Když toto zvládnou, je naděje, že další věci se dokáží naučit třeba i sami, bude-li to nutné.

2.5. Reálná nebo generovaná data?

Řešení úloh z oblasti, která je studentům blízká a pochopitelná, je nepochybně silnou motivací. Ale data z úloh „ze života“ mají často nevýhodu v tom, že u nich nebývají splněny předpoklady potřebné pro aplikaci jednoduchých metod statistické analýzy, které chceme ukázat. Dokonce ani data a úlohy doporučované v učebnicích nemusí být to nejlepší, co studentům předvést. Anděl [1] podrobně rozebírá v jedné učebnici doporučovaný příklad (Galileův pokus) a ukazuje, že to, co autoři doporučují, je plno neodpustitelných chyb. U dat uměle vygenerovaných sice máme zaručeno splnění potřebných předpokladů, ale motivace je slabší, studenti chtějí spojení s reálným světem. Nabízí se užívat pro výuku kombinaci: umělá data + „reálná“ pohádka, ale ani toto řešení není ideální. Když to nepřiznáme, je to podvod, který se většinou stejně provalí. Když tu pohádku přiznáme, je ono motivující spojení s realitou slabší.

2.6. Excel nebo specializovaný statistický software?

To je další z otázek, na kterou nenacházím jednoznačnou odpověď. Pokud se ve výuce užívá specializovaný statistický software, pak hrozí, že po skončení kursu se s ním studenti už nikdy nepotkají. Navíc se v základním kursu využije z drahého statistického paketu jen jeho malá část. A užívat volně šiřitelný software, např. R, má zase nevýhody v jeho obtížnějším ovládáním pomocí řídicího jazyka, nikoliv pro studenty obvyklým „klikáním". Při volbě běžně dostupného Excelu pokryjeme většinu výpočtů pro základní kurs (kromě pár typů statistických grafů), většinou dostaneme i numericky správné výsledky, ale musíme studenty naučit je najít, tzn. nevěřit informacím jak v dokumentaci statistických funkcí Excelu, tak textům v jejich výstupu. Tabulkové procesory ve svých počátcích nebyly pro statistické zpracování dat navrhovány, rozhodně to nebylo jejich hlavním cílem. Statistické funkce, které byly během dlouhé historie do Excelu postupně dodávány, nejsou navrženy nijak obratně ani v originální anglické versi (nekonsistentní návrh funkcí a jejich rozhraní, redundatní a nepřehledný výstup). A česká lokalizace Excelu přímo hrozí vyvolat „morální" úhonu studentů. Nápovědy ke statistickým funkcím a doplňku Analýza dat zřejmě překládal někdo, kdo o statistice patrně nic neslyšel ani z velké dálky. Výsledek jeho práce, zejména užívané české pojmy, sice může statistika pobavit, ale začátečníkovi moc neposlouží. Ve výstupech se průměru říká střední hodnota, počet stupňů volnosti je označen jako rozdíl (?!), v regresi jsou směrodatné odchylky odhadů parametrů označeny jako chyba střední hodnoty atd. Podrobněji se lze dočíst o chybách a nedostatcích ve statistických funkcích Excelu a doplňku Analýza dat např. v [4,5,7,8]. Po létech váhání mezi těmito volbami jsem došel ke kompromisnímu přístupu – v kursu pro studenty informatiky užívat obojí. Počítání v Excelu připravuje studenty do drsého praktického života (podobně jako Cimrmanův běh s kufrem) a paralelní seznámení se statistickým softwarem (NCSS) jim dá možnost vlastního posouzení jeho výhod a současně i zkušenost, že ani běh v módním barevném statistickém tričku není bez námahy a bez rizika upadnutí. O volbě statistického software nejen pro výuku kvalifikovaně píše Antoch [2], ale žádná jednoduchá univerzálně platná doporučení pro výběr nejlepšího statistického programového vybavení pochopitelně ani tam nenajdete.

2.7. Elektronické distanční vzdělávání?

Navzdory zkušenosti vyslovené na záčátku této kapitoly (úroveň neznalosti často nezávisí na kantorovi) jsem skeptický k možnostem distančního vzdělávání v úvodním kursu pro nestatistiky. Připadá mi, že přece jen přímou výukou lze studentům pochopení některých obtížných pojmů usnadnit nebo přinejmenším „polidštit". Navíc u distančního vzdělávání je o mnoho pomalejší zpětná vazba mezi studenty a kantorem, o které tak krásně a stručně píše Komenda [3] - kantor musí „vzít rozum do hrsti a mít oči na stopkách“. U distanční výuky i při velkém úsilí kantora je korekce způsobu výuky tak pomalá, že v jednosemestrálním kursu je její dopad na studenta skoro nemožný. To není dohad, ale vlastní zkušenost z několikaleté výuky základního kursu statistiky v bakalářském kombinovaném (distančním) studiu oboru Aplikovaná informatika na naší katedře. Velmi rád bych se v hodnocení možností e-learningu mýlil. Pak by statističtí kantoři mohli jít penze nebo se věnovat aplikacím. První by prospělo mé chátrající tělesné schránce, druhé snad vážnosti statistiky a úrovni jejích aplikací.

3. Závěr

Pokud je čtenář nespokojen s příliš skeptickými postoji autora, stručností, neúplností či povrchností probíraných otázek a odpovědí, může si napravit chuť přečtením Zvárova článku [10], případně i dalších příspěvků z tohoto pro statistického kantora pozoruhodného a inspirujícího sborníku. Nicméně se domnívám, že ani důkladné prostudování tohoto sborníku nás nezbaví neodbytných otázek o výuce statistisky a pracného hledání odpovědí.

Literatura

[1] Anděl J. (2003) *Statistické modely*, Statistika **83(2)**, 1–17.

[2] Antoch J. (2002) *Vybraná kriteria pro volbu programového vybavení pro (statistickou) analýzu dat*. In: *Výuka statistiky v České republice I*, Matfyzpress, Praha, 1–7.

[3] Komenda S. (2002) *Statistika pro nestatistiky*. In: *Výuka statistiky v České republice I*, Matfyzpress, Praha, 65–68.

[4] Pastor K. (2002) *Intervaly spoľahlivosti pre pravdepodobnosť zriedkavých javov a ich použitie v praxi*. In: *Zborník prispevkov konference PRASTAN 2002*, Slovenská štatistická a demografická spoločnosť, Bratislava, 49–56.

[5] Pastor K. (2001) *Testovánie hypotéz v Exceli*. In: *Zborník prispevkov 10. medzinárodný seminár Výpočtová štatistika*, Slovenská štatistická a demografická spoločnosť, Bratislava, 106–108.

[6] Tvrdík J. (1998) *Základy statistické analýzy dat*. Přírodovědecká fakulta Ostravské university.

[7] Tvrdík J. (1998) *Excel, statistika, lokalizace a zmatek*. Informační Bulletin České statistické společnosti *9(2)*, 13–20.

[8] Tvrdík J. (2001) *Excel ve výuce statistiky*. In: *Zborník prispevkov PRASTAN STAKAN 2001*. Slovenská štatistická a demografická spoločnosť, Bratislava, 136–140.

[9] Víšek J.Á. (2002) *Jak je či jak by měla být vyučována statistika pro ekonomy*. In: *Výuka statistiky v České republice I*, Matfyzpress, Praha, 95–102.

[10] Zvára K. *Zkušenosti s výukou statistiky pro nematematiky*. In: *Výuka statistiky v České republice I*, Matfyzpress, Praha, 107 – 114.

[11] Žváček J. Katedry vyučující statistiku, `http://www.statspol.cz/odkazy/katedry/ceskestatisticke.htm`

VÍTĚZNÁ OCHRANO MORAVY ZŮSTAŇ MATKOU LIDU SVÉMU

Po ničivé a svým rozsahem největší válečné smršti, která se během třiceti let několikrát přehnala Evropou (1618 – 1648), soustředili věřící svoji naději do úcty mariánské. A tak ve 2. polovině 17. století vznikají mariánské poutní místa nebývalou měrou. Procesí putují k Boží rodičce na mnohé posvátné hory a kopce, opředené často zbožnými tradicemi a pověstmi o zázračných událostech a uzdraveních.

K takovým posvátným horám patří i Svatý Hostýn – nejnavštěvovanější mariánské poutní místo na Moravě a po Velehradě snad nejpamátnější vůbec. Již po tři století je hostýnská hora se svojí mariánskou svatyní cílem tisíců poutníků z blízkého okolí i vzdálenějších krajů, kteří zde nacházejí útočiště ve svých strastech a potřebách.

Hostýnské vrchy, nazývané podle památné hory Hostýna (735 m n. m.) jsou nejzápadnějším výběžkem moravských Karpat. Hostýn tvoří dva vrcholy. Severnější s rozhlednou je nejvyšším místem hory, zatímco druhé temeno (718 m n. m.) u západního okraje je zastavěno poutním chrámem a dalšími objekty. Hora Hostýn s charakteristickou siluetou kostela je již zdaleka viditelná díky rovinaté krajině úrodné Hané, která se západně od Hostýnských vrchů široce rozprostírá kolem řeky Moravy.

Někteří autoři opírajíce se o pověsti a tradici, ale spíše svým zbožným přáním, rádi by vyšperkovali historii Hostýna nejen „zázrakem“ s Tatary ve 13. století, ale nejraději by posunuli počátky křesťanské tradice až do doby velkomoravské k sv. Cyrilu a Metodějovi, ba dokonce až do slovanského dávnověku, kdy zde měla být uctívána modla Hostýna či Radegasta. Hostýn je však poprvé spolehlivě historicky doložen teprve v 16. století. Tím však není řečeno, že nejstarší osídlení tohoto místa nesahá hluboko před jeho psané dějiny. Pravěké osídlení hostýnské výšiny dokládají četné architektonické nálezy. Plastičtější a úplnější obraz pravěkého a časně historického osídlení Hostýna může však v budoucnu podat jen systematický archeologický výzkum.

Mezi stovkami poutních míst v Evropě i jinde ve světě stěží najdeme obraz podobný obrazu hostýnskému – „Matku s blesky“. Panna Maria s Ježíškem, který metá na nepřátele křesťanů blesky. Tento originální výjev není původní, vznikl teprve na základě barokní pověsti o Tatarech, natolik je však srostlý s pojmem Svatého Hostýna, že jedno od druhého nelze oddělit. Hostýnskou pověst nemáme doloženu v soudobých pramenech, poprvé se o ní zmiňují až po polovině 17. století jesuitští historiografové Bohuslav Balbín a Jiří Crugerius. Nebudeme proto daleko od pravdy, když vznik pověsti o zázračném zachránění Moravanů proti moři Tatarů položíme až do doby rekatolizace po třicetileté válce.

Za vpádu Mongolů, lidově zvaných Tatarů, na Moravu v roce 1241, hledali lidé pro sebe a svůj nejskromnější majetek útočiště v lesích a na horách. Takovým útočištěm se mohl zřejmě stát i Hostýn, zvláště když zde již existovaly mohutné valy pravěkého osídlení. Traduje se, že Moravané utekli před Tatary na Hostýn a za navršenými záseky statečně odráželi jejich útoky. V horkém létě dolehl však na obránce nedostatek vody. V nejvyšší tísni se prý obrátili s důvěrou k Matce Boží, na jejiž přímluvu vytryskl hned pod temenem hory pramen. Nato Tataři vidouce marnost svého úsilí obrátili se k Olomouci.

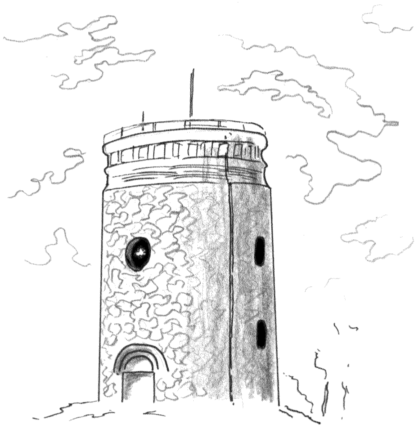
Boj křesťanů proti Tatarům v roce 1241 nebyl tak důležitý po stránce historické věrohodnosti, jako spíše symbolem boje křesťanské Evropy proti tureckému nebezpečí, které po pádu Cařihradu roku 1453 až do bitvy u Vídně v roce 1683, smrtelně ohrožovalo evropskou civilizaci. V době, kdy se rodí hostýnská pověst o Tatarech, objevuje se desetitisícová turecká armáda na východní Moravě. V roce 1663 Turci, lidově opět nazývaní Tataři, třikrát vpadli na naše území a napáchali na moravsko-uherském pomezí nedozírné škody, kruté násilí a nepředstavitelné utrpení zejména prostému venkovskému lidu. Někteří badatelé právě v tomto tureckém vpádu viděli skutečný zdroj vzniku hostýnské pověsti.

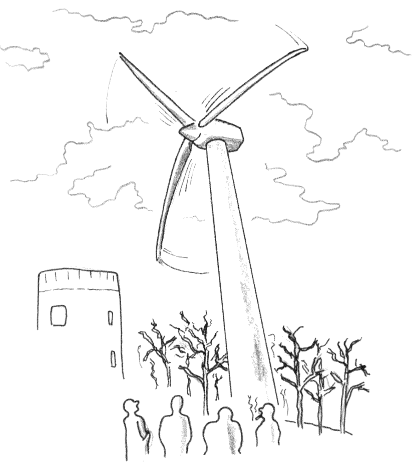
Neodmyslitelnou součástí každého poutního místa je také pramen zázračné uzdravující vody. Pramen je nejen důležitým symbolem, ale i skutečným občerstvením znavených poutníků a provází ho také uzdravující pověst. Na Hostýně tryská takový pramen ze skály na západním svahu přímo pod chrámem. Podle této tradice vytryskl při tatarském ležení v roce 1241 k záchraně obležených křesťanů a víra v zázračnou moc hostýnské vody má ve zbožném moravském lidu hluboké kořeny podnes.

Ať je náš pohled na Svatý Hostýn jakýkoliv, přesto můžeme přijmout pověst o Tatarech jako symbol hledání nadpřirozené pomoci tam, kde naše prostředky byly již dávno vyčerpány. Vždyť tuto zkušenost jsme si mohli ověřit již tolikrát i v dobách minulých a lidstvo si ji nese i v našem tisíciletí.

VÍTĚZNÁ OCHRANO MORAVY

ZŮSTAŇ MATKOU LIDU SVÉMU





(eds.) Jaromír Antoch, Gejza Dohnal a Josef Štěpán

Výuka statistiky v České republice II.

Sborník prací semináře STAKAN
zorganizovaného Českou statistickou společností
a Slovenskou štatistickou a demografickou spoločnosťou
za podpory KPMS MFF UK
ve dnech 23. – 25. května 2003 v Bystřici pod Hostýnem

Vydala

Česká statistická společnost

Vydání první

Praha 2004

ISBN 80-239-4086-4