

Výběrovým šetřením se nevyhneme

Jeho pohled na statistiku nejvíce ovlivnil profesor Jaroslav Hájek. „Znalosti, které jsem pod jeho vedením získával, byly neocenitelným přínosem pro mou další práci,” říká Jiří Anděl, vysokoškolský profesor, který svou kariéru spojil s katedrou pravděpodobnosti a matematické statistiky na Matematicko-fyzikální fakultě Univerzity Karlovy v Praze.

Ondřej Vozár, oddělení matematicko-statistických metod a statistické kvality

Kteří lidé nejvíce ovlivnili váš pohled na statistiku a pravděpodobnost?

První dva roky studia na MFF byly společné nejen pro matematické obory, ale i pro matematiku a fyziku. Jednotlivé specializace tedy začínaly ve třetím ročníku. Matematickou statistiku přednášel MgMat. Marcel Josíčko. Výklad založil převážně na Cramérově učebnici, což byl tehdy moderní počin. Navíc pan magistr Josíčko měl hodně zkušeností s praktickými aplikacemi a tím dokázal výrazně oživit přednášenou látku. Dalším výrazným učitelem byl ing. Josef Machek. U něho jsem ostatně psal i svou diplomovou práci. Pan inženýr měl hluboké znalosti statistických metod a některé komplikované výpočty dokázal provést rychleji než osobní počítače. To už je ale zkušenost z pozdější doby. Teorii odhadu a testování hypotéz přednášel prof. Josef Bílý, mimořádně vzdělaný a zdvořilý člověk. Můj pohled na statistiku a pravděpodobnost samozřejmě nejvíc ovlivnil prof. Jaroslav Hájek. Měl jsem to štěstí, že jsem byl jeho prvním aspirantem. Znalosti, které jsem pod jeho vedením získával, byly neocenitelným přínosem pro mou další pedagogickou i vědeckou práci.

Často býváte spojován s legendárním prof. Hájkem, byl jste jeho asistent, když vyučoval v té době velmi moderní kurz matematické statistiky.

MFF jsem absolvoval v roce 1961. Tehdy byl vedoucím katedry statistiky prof. Janko a tajemníkem katedry ing. Machek. V rámci tzv. vědecké přípravy, což byla jedna z forem aspirantury, se mě ujal jako školitel prof. Dupač. Rovnou prohlásil, že mě bude školit jen dva roky, a to během přípravy na odbornou kandidátskou zkoušku. To se už počítalo s tím, že na katedru přejde z Matematického ústavu ČSAV prof. Hájek, který měl prof. Dupače ve funkci mého školitele nahradit. Prof. Hájek se také měl stát vedoucím katedry po prof. Janko, což se také realizovalo. Prof. Hájek ihned převzal základní přednášku z matematické statistiky a ze stacionárních procesů. Přednášku ze statistiky založil na teorii pseudoinverzních matic, což byl tehdy velice moderní postup. Já a Karel Zvára jsme vedli cvičení k této přednášce. Samozřejmě jsme byli nejpilnějšími návštěvníky přednášek, abychom věděli, co vlastně máme cvičit. Ještě bych měl uvést jednu historku z doby, kdy prof. Hájek přebíral funkci vedoucího katedry. Ing. Machek využil příležitosti a po mnoha letech tajemníkování se této nepopulární funkce vzdal. Nově nastupujícímu šéfovi bylo sděleno, že tajemníkem katedry bude Anděl. Při první schůzi katedry prof. Hájek hned na začátku velmi rozčileně prohlásil, že je zdvořilý a jediné možné, aby si vedoucí vybral svého tajemníka sám. On že s dosavadním postupem naprosto nesouhlasí. Viděl jsem, že to není dobrý začátek mé spolupráce s prof. Hájkem, ale on stále zvýšeným hlasem pokračoval. Sdělil, že se rozhodl, že tajemníkem bude Anděl, ale bude to z jeho rozhodnutí a ne z rozhodnutí předchozího vedení katedry. Přednášku ze statistiky pak každým rokem podstatně obměňoval. Snad jen my asistenti jsme věděli, že se vykládá stále stejná látka. Pokud nějaký student propadl a musel tuto přednášku navštěvovat znovu, musel mít pocit, že chodí na úplně jiný předmět. Ještě náročnější bylo jeho pojetí předmětu Stacionární procesy. Velkou část látky si připravoval sám na základě svých vlastních výsledků. Jednou se stalo, že byl zrovna v ne-



prof. RNDr. Jiří Anděl, DrSc.

V letech 1956–1961 studoval Matematicko-fyzikální fakultu Univerzity Karlovy v Praze (MFF UK). Vědeckou přípravu absolvoval pod vedením profesora Jaroslava Hájka, odborníka světového významu v oblasti pořadových testů, teorie výběrových šetření, stochastických procesů a řady dalších oblastí. V roce 1965 obhájil kandidátskou práci, habilitoval se v roce 1972. Docentem byl jmenován v roce 1977. Titul DrSc. získal v roce 1981 a v roce 1986 byl jmenován vysokoškolským profesorem. Celou svou kariéru spojil s katedrou pravděpodobnosti a matematické statistiky MFF UK. V roce 1977 byl pověřen jejím vedením, od roku 1981 do roku 1993 byl jejím řádným vedoucím. Od roku 1993 se stal na MFF UK proděkanem. V roce 1990 byl zvolen prvním předsedou nově vzniklé České statistické společnosti. Ve své vědecké činnosti se zabýval především analýzou časových řad a matematickou statistikou. Vedle desítek odborných statí v předních světových i československých časopisech napsal řadu původních českých učebnic statistiky – Statistická analýza časových řad, Matematická statistika, Statistické metody, Základy matematické statistiky. Dosud aktivně učí a patří mezi nejoblíbenější učitele na MFF UK.

Více na <http://bit.ly/1mTHmDL>.

mocníci a já jsem ho měl v přednášení zastoupit. Navštívil jsem ho a on mi řekl: „Myslím, že platí zhruba následující tvrzení. Tak to ověř a zítra to studentům i s důkazem předneseš.“

Jak přistupoval k výuce?

K výuce přistupoval velmi zodpovědně. O tom ostatně svědčí i různá skripta, která napsal se svými asistenty a kolegy. Občas studentům během přednášky uváděl vědecké problémy. Slíbil, že kdo mu první přinese správné řešení, dostane u zkoušky o jednu otázku méně a místo ní mu bude vyřešení problému klasifikováno známkou výborně. Zadávané problémy bývaly dost obtížné, a tak jen málokdy se stalo, že je některý posluchač vyřešil.

Založil vědeckou školu a vedl disertaci řadě současných významných odborníků a odborníků. Jaký byl jeho způsob vedení aspirantů? Využívají se některé jeho postupy i dosud? Které byste zavedl do současné výuky na MFF UK a výchovy doktorandů?

Prof. Hájek své aspiranty školil v malých skupinkách. Všichni řešili cvičení uvedená v právě probírané knize. Podmínkou bylo, že každý musí pracovat zcela samostatně a řešení nesmí hledat ani v literatuře. Jednou mi prof. Hájek vyprávěl, jak pracoval ve své aspirantuře on. Když studoval nějakou knihu, ve které cvičení nebyla, četl text jen do místa, kde se objevila nějaká matematická věta. Její důkaz si zakryl papírem a vypracovával ho sám. Je jasné, že taková příprava k vědecké práci je velmi účinná, ale nedá se masově použít. Rovněž disertaci musel každý jeho aspirant vypracovat zcela samostatně. Prof. Hájek zadal téma a pak jen velmi kriticky sledoval, zda je postup řešení správný. Když my, jeho aspiranti, jsme se sami stali školiteli, snažili jsme se podobným způsobem školit své vlastní aspiranty. Po roce 1989 se však podstatně změnil vysokoškolský zákon a změnila se i výuka doktorandů. Byly zavedeny předměty vyučované v doktorském studiu a to se tak poněkud přiblížilo magisterskému studiu. Řekl bych, že dnes školitelé více pomáhají svým doktorandům při psaní disertací. Je to dáno i podmínkami, které platí pro doktorské obhajoby. Má-li mít doktorand své výsledky v době obhajoby již publikované, nebo aspoň k publikaci přijaté, musí na disertaci začít pracovat co nejdříve. A to se bez účinné pomoci školitele neobejde.

Vzpomněl byste také na další kolegy a kolegyně?

Někteří z nás, kteří jsme na katedře pravděpodobnosti MFF UK, jsme přímo aspiranty prof. Hájka, mladší kolegové jsou zas našimi doktorandy. Tak uvedu alespoň několik jmen v abecedním pořadí. prof. Dupačová, prof. Hušková, prof. Jurečková a prof. Štěpán (který bohužel nedávno zemřel) byli přímými aspiranty prof. Hájka. Myslím, že jejich mezinárodní reputace je dobře známa a není třeba ji uvádět.

Jak se podle vás změnila statistika v posledních letech, a jak se měnila statistika během vašeho života?

Krátce bych se vyjádřil jen k vývoji matematické statistiky. Podle mého názoru jsme svědky velkého rozvoje zejména abstraktních partií. Stačí porovnat obsah časopisu The Annals of Mathematical Statistics s jeho nástupci The Annals of Statistics a The Annals of Probability. Takových příkladů by se dalo uvést víc. Články popisující konkrétní statistické metody se spíše najdou v časopisech pro techniky, pro lékaře a podobně. Velké změny pochopitelně přinesl nástup počítačů.

Jak váš výzkum a statistiku vůbec ovlivnily počítače?

Přínos počítačů by se dal rozdělit do dvou kategorií. Jednak jde o obecné využití výpočetní techniky počínaje elektronickou poštou, textovými editory, literaturou v elektronické podobě atd., jednak o využití v matematice a statistice. Uvedl bych malý příklad. V předpočítačové době na naší katedře pracovala jednak paní sekretářka, jednak paní výpočtářka. S kolegou Karlem Zvárou jsme dělali průzkum, jak souvisí úspěšnost studia na výsledcích přijímacích zkoušek. Použili jsme na to diskriminační analýzu, napsali vzorečky a požádali paní výpočtářku o provedení výpočtů. Pracovala asi tři dny, pak přinesla výsledky. Kontrola ukázala, že je někde numerická chyba. Tak paní výpočtářka začala znovu. Dospěla k jinému, ale také chybnému výsledku. Ani třetí pokus nepřinesl správný výsledek. Bylo zřejmé, že během dlouhého výpočtu výpočtář s velkou pravděpodobností někde udělá chybu. Když nastoupily počítače, vyměňovali si statistici informace o tom, který program má jaký nedostatek. Dnes už, alespoň pokud jde o základní výpočty, jsou programy dost spolehlivě odladěné. Pokud jde o mne, k usnadnění výpočtů a k jejich kontrole používám programy Mathematica a Maxima, k numerickým statistickým výpočtům pak program R.

Jaký máte postoj k mnohdy neuváženému používání „hotového“ statistického softwaru bez dohledu statistika?

Posluchačům tento problém demonstruji na následujícím příkladu. Provedu simulaci dat pro ilustraci lineární regrese. Vypočtou se odhady parametrů přímky a intervaly spolehlivosti. Všechno funguje tak, jak má. Odhady jsou rozumně blízko skutečným parametrům a intervaly spolehlivosti překrývají dané parametry. Pak se místo hodnot nezávisle proměnné vezmou jejich „pokažené“ hodnoty, jako když i ta nezávisle proměnná je měřena s nezanedbatelnou chybou. Znovu se provedou odhady metodou nejmenších čtverců. Někjaké výsledky vyjdou. Ale odhady jsou daleko od parametrů a intervaly spolehlivosti nemají nic společného s hodnotami parametrů. Teprve použití správné teorie místo metody nejmenších čtverců vede k očekávaným výsledkům. Počítač tedy většinou nějaký výsledek vydá. Ale pokud nejsou splněny základní předpoklady použité metody, je zpravidla výsledek chybný a matoucí.

V České republice jste považován za jednoho z největších odborníků na analýzu časových řad, dlouhá léta jste v této oblasti spolupracoval především s hydrology. Jak se z této pozice díváte na často diskutovaný fenomén „změna klimatu“? Mnozí publicisté a politici přitom velmi rádi odkazují právě na analýzu (nežrádka velmi krátkých) environmentálních časových řad, často bez naprostého porozumění meritů věci.

Podle mého názoru se použité analýzy týkají observačních dat, nikoli dat experimentálních. To ostatně vyplývá ze samé podstaty problému. Proto je třeba na výsledky výpočtů nahlížet jen jako na určité statistické charakteristiky, které nemohou vypovídat nic o kauzální závislosti sledovaných jevů. V oblasti analýzy hydrologických dat byla k dispozici tvrdá kontrola. Pomocí vypočteného modelu byla provedena předpověď průtoků řek a pak se ukázalo, zda tato předpověď byla úspěšná či nikoli. A právě takováto kontrola závěrů, ke kterým dospívají zastánci změny klimatu, mi zatím schází.



Je o vás známo, že se již delší dobu zabýváte průzkumy veřejného mínění. Co vás k tomu přivedlo? Co si myslíte o profesionalitě těchto výzkumů u nás? Prakticky po každých volbách se většinou nestačíme divit, že to vlastně dopadlo úplně jinak, než jak bylo předpovězeno.

Průzkumy veřejného mínění jsou zajímavé samy o sobě. Dají se také použít jako hezká ilustrace při výuce. Navíc to máme takřka v rodině. Můj syn, který také vystudoval matematickou statistiku, pracuje u takové instituce. Jsem přesvědčen, že se používají metody na dobré úrovni. Ostatně společnosti, které takové průzkumy zadávají, by rychle od takové praxe ustoupily, kdyby jim to nepřinášelo správné informace. Záležitost předvolebních průzkumů je trochu komplikovanější. Je třeba si uvědomit, že mnozí voliči mění své názory na poslední chvíli, tedy v době, kdy už je na prezentaci průzkumů vyhlášeno moratorium. Bylo by třeba vzít v úvahu nejen výsledky dotazování, ale i analýzu trendů. A to by veřejnost obtížně přijímala. Přechod od úplného zjišťování k výběrovým šetřením je patrně nevyhnutelný. Došlo k tomu i u vyspělých států. Spíše stojíme před jiným problémem. Naše zákony v tomto směru nejsou dostatečně účinné a jejich vymáhání je mírně řečeno problematické. Stojí za to podívat se do historie, jakou důležitost statistickému zjišťování připisovali vládci. Tak třeba ve starém Římě bylo podání požadované informace povinné a muži, který by se ke sčítání nedostavil, hrozilo, že ještě ten den může být uvržen do otroctví. Nedočetl jsem se, zda se tato hrozba opravdu uskutečňovala a v jakém rozsahu, ale patrně byla dostatečně účinná.

Co si jako statistik myslíte o trendu nahrazovat výběrová zjišťování odhady využívající administrativní data a registry? Není to vlastně návrat k vyčerpávajícím statistickým zjišťováním, kdy místo statistického šetření se data od statistické jednotky (respondenta) získají pomocí povinného hlášení, administrativního úkonu?

Pokud jsou registry spolehlivé, není důvod proti takovému trendu něco namítat. Na druhé straně registry zachytí jen některá statistická data, takže výběrová zjišťování nepůjde ani v budoucnu zcela nahradit.

Jaká je role aplikací pro rozvoj matematické statistiky? Mohl byste zmínit nějakou netradiční aplikaci v oblasti, která nebývá nebo ve své době nebyla považována za vhodnou pro aplikaci statistiky (např. archeologie)? Jakou aplikaci, na které jste spolupracoval, považujete za nejprínosnější?

Aplikace jsou jedním (ale ne jediným) zdrojem podnětů pro další rozvoj matematické statistiky. Třeba prof. C. R. Rao uvádí ve své knize aplikace v archeologii, které ho nepochybně inspirovaly k dalším statistickým výzkumům. Nebo McNemarův test, který byl publikován v časopisu Psychometrika. Jsem přesvědčen, že jeho autor byl inspirován podněty z lékařského výzkumu. Pokud jde o mé vlastní aplikace, tak záleží na tom, co se rozumí slovem nejprínosnější. Třeba výše zmíněné aplikace v hydrologii lze označit za přínosné třeba proto, že některé byly publikovány v předních vědeckých časopisech. Na druhé straně spolupráce s podnikem Škoda v Plzni byla úspěšná, protože zadavateli ušetřila nemalé finanční prostředky. Přitom šlo o standardní aplikaci, o které se, pokud vím, nic nepublikovalo.

Na jedné straně se statistika každý rok rozrůstá o několik desítek metrů časopisů a knih; když se však člověk problémem trochu důkladněji zabývá, exaktní řešení s poctivým vyřešením všech praktických problémů nedohledá. Máte tu samou zkušenost?

Rozhodně není snadné najít zrovna tu metodu, kterou bychom zrovna potřebovali. To platí pro teoretické výsledky zrovna tak jako pro ty aplikační. Nejrychlejší, nejpohodlnější a asi i nejspolehlivější je obrátit se na někoho, kdo to ví. Já jsem měl to štěstí, že jsem se mohl obracet na prof. Dupače, ing. Machka a další. Tito moji učitelé a později kolegové velkoryse přerušili svou práci a snažili se mi pomoci.

Byl jste prvním předsedou České statistické společnosti. Po prvotním nadšení ze znovuoobjeveného spolkového života zažíváme jeho pomalý úpadek. Jakou roli přisuzujete vědeckým společnostem v současné době „volného přístupu k informacím, například přes internet.“

Role vědeckých společností je nemalá. Nejde jen o organizaci vědeckého života, ale i o rozvoj mezilidských vztahů a kontakty mezi kolegy. Podle mého názoru založením České statistické společnosti se podařilo vybudovat dobré vztahy mezi Českým statistickým úřadem, VŠE a MFF. I kdyby žádná další spolupráce neexisto-

vala, již tento fakt stojí za všechnu tu nezbytnou organizační práci. Ostatně třeba takoví lékaři, právníci, lékařníci a další jsou hrdi na své stavovské a odborné společnosti a snaží se zvyšovat jejich prestiž.

Více než 50 let učíte a vaše hodiny jsou pověstné vysokou úrovní a přesností výkladu. Jsou dnešní studenti přicházející na vysoké školy připraveni z matematiky o tolik hůře než dříve, jak se často uvádí ve sdělovacích prostředcích?

Myslím, že matematická připravenost uchazečů přijatých na MFF se v poslední době nijak podstatně nemění. Máme možnost to posoudit na základě testu ze středoškolské matematiky, který studenti 1. ročníku absolvují na úvodním soustředění na Albeři. Někdy se zdá, že výsledky jsou horší a horší, ale pak zas přijde obrát. Rozdíl však může být v pracovní morálce. My jsme tvrdě studovali od první přednášky. Dnešní studenti většinou začnou intenzivně pracovat až na své bakalářské nebo diplomové práci. Téma je obvykle zaujme a vidí, že jsou schopni odvodit nové a přitom zajímavé výsledky.

Dlouhá léta se zabýváte popularizací statistiky, napsal jste řadu čtivých článků v různých časopisech, dlouholetá popularizační práce vyvrcholila knihou *Matematika náhody*, která vyšla též anglicky. Jaké máte zkušenosti s prezentováním aplikací statistiky a pravděpodobnosti na zajímavé praktické problémy (volby, rekordy) laikům a zejména novinářům, rozhlasovým a televizním redaktorům? Můžete uvést nějakou zajímavou zkušenost?

Zmíním se o jedné zkušenosti. Je to už řada let, kdy v televizi probíhala nějaká soutěž, jejíž vítěz si mohl vybrat jednu ze tří cihliček. Ačkoli vypadaly všechny stejně, jedna byla zlatá, druhá stříbrná a třetí bronzová. Zastavil se u mne v Karlíně pan Ulm. Sdělil mi, že už po nějakou dobu si nikdo nevybral zrovna zlatou cihličku a že jim rozhořčení diváci píší, že tam snad ani tu zlatou nedávají. Byl jsem pozván do vysílání, vysvětlil jsem výpočet příslušných pravděpodobností a jejich interpretaci. Všechno bylo v pořádku i proto, že při další hře už zlatá cihlička byla zas vytažena. Já jsem se ale pana Ulma zeptal, proč se obrátil s tím problémem zrovna na mě. Čekal jsem, že mi odpoví, že o mně ví z nějakých propagačních akcí, ale odpověď byla nečekaná. Jednou jel tramvají přes Malostranské náměstí a uviděl nápis Matematicko-fyzikální fakulta. Vystoupil a pana vrátného se zeptal, kdo by mu mohl něco říci o pravděpodobnosti. Pan vrátný se podíval do telefonního seznamu, vyhledal tam seznam osob na katedře statistiky a řekl mu, ať se obrátí na profesora Anděla. A bylo to. Věc má však ještě jeden aspekt. Zjistil jsem, že se nespokojení diváci začali obracet na Československou televizi ve chvíli, kdy poprvé pravděpodobnost nevybrání zlaté cihličky klesla pod naše magické číslo 0,05. Od té doby to občas uvádím ve výuce, když se studenti ptají, proč se používá u testů většinou hladina 0,05.

A co studenti? Již po tisíce let se říká, že jsou stále horší a horší, snaží se proplout za cenu nejmenšího odporu a nemají přirozenou úctu ke svým pedagogům a jsou stále drzejší. Souhlasíte s těmito škarohlídy? Jako dlouholetý proděkan pro studijní činnost jste si s nimi jistě „užil“ dost a dost.

Není pochyb o tom, že dnešní lidé nejsou tak zdvořilí jako bývali dřív. Jako proděkan jsem jednal s mnoha studenty. Často naše setkání začínalo tím, že jsem je poučoval, že mají napřed pozdravit, pak se představit a pak učitele zdvořile oslovit. Většinou to pro ně byla úplně nová informace. Na druhé straně musím konstatovat, že učitelé zdvořilost od studentů ani nevyžadují, jen si mi pak stěžují, že se studenti vůči nim nechovají tak, jak by měli. Většinou však jsem se studenty vycházel přátelsky. Snad jen jeden student z tisíce se snaží zneužít benevolence studijních předpisů a systematicky doslova otravovat ovzduší na fakultě. Obecně mohu říci, že posluchači MFF se chovají lépe než jejich ostatní vrstevníci.

Co byste popřál statistice do nejbližších let? Jaký učební text nebo dosud dostatečně nevyučovanou oblast statistiky v českém jazykovém prostředí postrádáte? V které oblasti nejvíce postrádáte původní učebnici v českém jazyce? Kdybyste měl dostatek času, jakou další učebnici byste napsal? Co byste sdělil nastupující generaci vysokoškolských pedagogů a aplikovaných statistiků?

Statistice bych přál větší počet posluchačů, kteří ji budou studovat. Dosavadní vývoj jejich počtu není příliš optimistický. Z hlediska výuky jsou podle mého názoru všechny hlavní oblasti více či méně pokryty. Zato však



porovnání seznamu přednášek a seznamu dostupné české odborné literatury ukazuje, že učebnic je žalostně málo. Není divu, protože napsání učebnice na rozdíl od vědeckého článku není nijak oceňováno. Připadá mi, že psaní učebnice je pokládáno za koníčka, kterému se věnuje jen nějaký málo vytížený učitel. Studenti jsou od prvního ročníku zvyklí, že si všechny přednášky podrobně zapisují a jen z těchto zápisků se ke zkouškám připravují. Pokud náhodou k nějakému předmětu existuje učebnice, stojí nad ní v rozpacích a netuší, že by její užití mohlo výrazně zefektivnit výukový proces. V časopise *The American Statistician* byl v roce 2007 uveřejněn rozbor toho, jaké statistické metody jsou v lékařských článcích používány. V 91 článcích časopisu *The New England Journal of Medicine* bylo toto pořadí: intervaly spolehlivosti (61), kontingenční tabulky (48), analýza přežívání (39) a pak následovaly další metody. Kdybych se k tomu někdy přinutil, napsal bych učebnici analýzy kategoriálních dat, která by samozřejmě zahrnovala i kontingenční tabulky. Další učebnice v českém jazyce, která schází, by měla být věnována analýze přežívání. Tu by měl napsat někdo jiný, někdo, kdo se touto tematikou zabýval a zabývá. Možná, že k tomu dojde v rámci nového předmětu, který je věnován analýze cenzorovaných dat. Nastupujícím pedagogům bych přál příjemné prostředí na katedrách, hodně úspěchů v pedagogické i vědecké práci a dobré rodinné zázemí. Jako měl pan profesor Hájek. ■