

Učíme statistiku

Petr Hebák*

Vážení přátelé a kolegové,

pověření úvodním slovem na tomto setkání statistiků ze Slovenska a Čech považuji za velkou čest a za významné ocenění mého více než 37 let trvajících pedagogického působení na katedře statistiky (dnes katedře statistiky a pravděpodobnosti) na Vysoké škole ekonomické v Praze. Skutečnost, že jsem byl požádán předsedou České statistické společnosti o jednu ze vstupních přednášek k diskusi na této sympaticky pojmenované konferenci je pro mne nesmírně zavazující. Od roku 1962, kdy jsem jako student pátého ročníku oboru *Statistika* byl pověřen vedením dvouhodinového cvičení *nestatistiků* druhého ročníku (přesněji dvakrát padesát minut jednou týdně v semestru, který měl patnáct týdnů a v kursu, který byl organizován systémem 2/2 dva semestry po sobě na všech fakultách VŠE), se sice hodně změnilo, ale (pro někoho možná překvapivě) ne tolik, aby to znamenalo zásadnější průlom do výuky základního kursu statistiky. Na druhé straně pozornost, která je věnována výuce statistiky ve významných zahraničních časopisech (za všechny bych jmenoval ty, které mívám možnost pravidelně sledovat *The American Statistician*, *Technometrics*, *The Canadian Journal of Statistics*, *Journal of American Statistical Association*, *Statistical Science* a *Teaching Statistics*), potvrzuje oprávněnost obsahového zaměření tohoto setkání. Patřím mezi ty, kteří se neustále v neformálních debatách na katedře i jinde těmito otázkami zabývají. Je nutné, ale do určité míry obtížné, od sebe oddělit výuku statistiky pro *nestatistiky* a *statistiky*. Omlouvám se proto organizátorům této konference, že se ve svém příspěvku podrobněji dotknu i výuky *statistiky pro statistiky*, protože tuto oblast našeho působení považuji za zcela zásadní pro další rozvoj naší profese. Ve svém příspěvku se pokusím, kromě dlouholeté výuky na VŠE, využít i zkušenosti, které jsem získal při výuce základů statistiky v roce 1992 pro Graduate Business School na University of Pittsburgh, dále při výuce na MBA kursech (nejdříve v letech 1990 a 1991 v Čelákoviciích a od roku 1994 pro Pražskou manažerskou školu) i při dvouměsíčním pobytu ve Švédsku za účelem vedení základního kursu statistiky pro studenty *nestatistiky* z University of Vaxjö. Úmyslně v této vstupní poznámce zvýrazňuji rozlišení *nestatistiků* a *statistiků*, protože mnozí absolventi různých škol sice vystudují a ukončí obor *Statistika*, ale pak se celý život ve svém zaměstnání i v osobních zájmech věnují všemu jinému jen ne statistice. Naopak někteří absolventi jiných oborů jsou významnými funkcionáři či pracovníky ve statistice (např. na Českém statistickém úřadu), ačkoli jejich statistické vzdělání je nepatrné. Naštěstí navíc je zde nemalá a pro rozvoj našeho oboru nesmírně užitečná skupina významných statistických odborníků, jejichž původní zaměření bylo zcela jiné a teprve dodatečně se zaměřili na určitou oblast statistiky.

Ve svém příspěvku bych se rád dotkl celé řady otázek, které se pro větší přehlednost textu pokusím pojmenovat hned na začátku. Některé z nich přesahují obsahové zaměření tohoto setkání a vystačily by na samostatné konferenci, takže jim budu věnovat nestejnou pozornost:

1. Jaké je postavení statistiky v zahraničí a u nás?
2. V čem se statistika za posledních dvacet let změnila a co lze očekávat?
3. Co se od nás statistiků při výuce požaduje?
4. Jsme pedagogové, vědci nebo obojí?
5. Jakou formu výuky statistiky zvolit?

* Katedra statistiky a pravděpodobnosti, VŠE Praha, hebak@vse.cz

5. Jakou formu výuky statistiky zvolit?
6. Jak jsme na tom se statistickou literaturou?
7. Jaký obsahový a časový rozsah výuky statistiky bychom doporučili?
8. Kdo a jak učí statistiku kromě nás?
9. Jaký by měl být obsah základního kursu statistiky pro bakaláře?
10. Jaký by měl být obsah kursu statistiky pro magistry na různých oborech?
11. Jaké máme naděje na obchodních, podnikatelských či MBA kursech?
12. Jak organizovat studium na oboru *Statistika*?
13. Jak organizovat postgraduální studium statistiků?

Ad 1) Jaké je postavení statistik v zahraničí a u nás?

Profesor D. W. Marquardt se už před více než deseti lety (ze své tehdejší pozice prezidenta Americké statistické asociace) velmi zajímavě zamýšlel na úvodních stránkách *JASA* nad významem statistiků. Pečlivě odlišoval *vnější* postoj *ostatních* k nám statistikům od *vnitřního* pohledu, kterým se sami na sebe díváme. Z celého článku sice vyzařovala důvěra v naše schopnosti a znalosti, ale především jistá obava nad budoucností statistiky ve společnosti. Nestačí, že statistika je plnohodnotný a vážený vědecký obor, ale musí být i silným konkurentem jiných oborů při jejich uplatňování ve veřejném životě a musí být i dobrou profesí pro naše absolventy statistiky. Říkal tehdy, že mezi námi naštěstí jsou špičkoví odborníci, kteří úspěšně rozvíjejí teoretické aspekty naší disciplíny a zaručují do budoucna svit naší vědy, ale zároveň varoval a upozorňoval, že nezapojíme-li se aktivně, iniciativně a hlavně prospěšně do centra dění kolem nás, může se snadno stát, že zůstaneme mimo toto dění. Finanční prostředky nezabezpečí úspěšná tvůrčí činnost, ale především smysluplné *veřejné* zapojení a na sebe upozorňující výsledky aplikací.

Článek G. Hahna a R. Hoerla pod názvem *Key Challenges for Statisticians in Business and Industry*, který vyšel v minulém roce v *Technometrics* a sloužil jako základ k velmi rozsáhlé následné diskusi, posunul debatu o budoucím postavení statistiků do zcela nového světla. Autoři varují před *demokratizací* statistiky (= statistika bez statistiků), která již dříve postihla matematiku a některé další obory. Veřejně dostupný statistický software má vážné následky jak pro rutinní používání stávajících statistických metod, tak i pro rozvoj nových. Hoerl říká, že nejvíce alarmující je skutečnost, že špičkoví manažeři nejen kladou otázky na použití statistiky ve své organizaci či firmě, ale sami se účastní výchovy a přípravy nových statistiků. Zdaleka už to není jen tradiční oblast průmyslové statistiky, ale týká se to téměř všeho (finančních transakcí, prodeje a nákupu, objednávek atd.). Mnozí aplikují statistiku bez statistických rad. Kdo je učí? Často někdo jiný než statistici. K zabezpečení tohoto zřetelného trendu v obchodě i průmyslu stačí dostupný software. Rozvíjejí se nové oblasti (neural networks, fuzzy logic, data mining, Bayesian networks) a my statistici často ani nevíme, co se děje. Shrnutí: „*Jiní statistiku nejen používají, učí, ale i rozvíjejí.*“ Z toho autoři vyvozují, že rostoucí popularita statistiky nemusí nutně znamenat rostoucí popularitu statistiků. K zabezpečení širší úlohy statistiků jsou statistické znalosti nutné, ale nikoli dostačující. Statistik už nemůže sedět u počítače a čekat, až mu někdo donese data k analýze. Statistik je součástí týmu a nikoli jen jeho konzultant. Pozici konzultanta snad může mít špičková katedra statistiky na blízké univerzitě. Znalosti matematiky, statistiky a ovládání počítače jsou dnes už samo-zřejmou kvalifikací statistika, ale ani to ještě nemusí stačit. Autoři článku nabádají a vybízejí k odpovědnosti statistiků *za svět statistiky bez statistiků*, a to nejen např. při rozvoji statistického softwaru. Statistici musí přivítat demokratizaci statistiky, pomoci při výuce a při vysvětlování použitelnosti statistických paketů. Musí získat důvěru těch druhých, aby *nás statistiky vzali do hry*. Je třeba být aktivní při rozvíjení nových postupů potřebných pro analýzu velkých datových souborů, takových postupů, které nebude možné zneužít málo znalým uživatelem. Statistici se musí podílet na společném výzkumu

Klíčové je jak to dělat a jak to časově ve výuce zvládnout! Diskuse k článku byla velmi rozsáhlá a ne vždy souhlasná.

Přenesme se však k nám, do našich domácích poměrů. Širokou nestatistickou veřejnost by asi bylo možné (a to asi nejen u nás) rozdělit do několika skupin s pro nás převážně neradostnými představami o významu statistiků a potřebě statistiky. Statistik bohužel u nás není na čele žebříčku popularity. Víme, že tyto názory všichni velmi dobře znáte, ale přesto si je stručně připomeňme (v pořadí od nejbanálnějšího k ideálnímu):

- a) Statistika jsou čísla a statistik je člověk, který je prezentuje jak ON chce. Tím nemá cenu nás (= širokou veřejnost) zatěžovat.
- b) Statistika je možná nějaký vědní obor, ale my (= vedoucí pracovníci, výrobci, manažeři) máme vážnější starosti (než statistická čísla) jako je legislativa, účetnictví, zisk, prodej, finance, organizace práce a další. Statistické úvahy nám příliš nepomohou a navíc statistická data jsou většinou nekvalitní.
- c) Statistika = matematika a s tou my (= obyčejní lidé) nechceme mít nic společného. My (= absolventi jiných oborů) jsme kdysi měli jeden kurs statistiky, který v nás zanechal ty nejhrůznější vzpomínky. Doufáme, že se s ní už nesetkáme a nové vzdělání tohoto typu odmítáme jako neprospěšné.
- d) Statistika = poskytování dat někomu jinému bez návratnosti a my (= dřívější připravovatelé dat do výkazů) z toho nemáme žádný prospěch. Navíc pochybujeme, že takový prospěch vůbec existuje.
- e) Statistiku nepotřebujeme vysvětlovat a statistiky rovněž nepotřebujeme, protože my (= uživatelé Excelu či jiného podobného produktu) si v zásadě vše umíme spočítat sami.
- f) Statistika může možná výjimečně v některých oblastech života, jako jsou výzkumy veřejného mínění, předpovědi na základě trendů, kontrola jakosti nebo snad i jiné oblasti, být užitečná, ale tyto případy mají velmi malou četnost výskytu.
- g) Statistika je vážený vědní obor a zároveň užitečný pomocník při rozhodování. Je vyloučené a chybné nemít statistiku či se obejít bez statistiky. Kvalifikované statistiky my (= ???) nutně do našich pracovních týmů potřebujeme. Rozhodující jsou jejich schopnosti, lidské vlastnosti, statistické znalosti i praktické zkušenosti. Nabídněte nám vhodného absolventa oboru *Statistika*.

Našel jsem zajímavý údaj, že např. v USA v jednom rozsáhlém výzkumu více než 90 % z 1 000 vybraných společností mělo marketingové oddělení se statistickými speciality nebo aspoň s jedním kvalifikovaným statistikem. U těchto vybraných organizací patřily platy statistiků mezi ty nejvyšší.

Na VŠE historicky převládaly dva až extrémní postoje při výchově statistiků. Ze statistického hlediska *minimalistický* postoj (naštěstí ho prosazovali většinou jiní než my) preferoval před statistickými předměty tolik jiných předmětů, že absolvent oboru statistika nakonec toho moc o statistice nevěděl a mnohdy byl zaměřen zcela jinam než na statistiku. Naše časté polemiky s některými kolegy jsou v tomto smyslu až tradiční a někdy jsme marně vysvětlovali, že nechceme zřít na VŠE nový *Matfyz*. Opravdu si nemyslím, že statistiku lze na oboru *Statistika* nahradit něčím jiným a tedy ani velmi užitečným pojišťovníctvím nebo dokonce pouhou znalostí činnosti Českého statistického úřadu. Statistik se přece nemůže při výběru oborových předmětů ke studiu rozhodovat mezi předměty statistickými a nějakými jinými. To neznamená, že student nemá volitelnost, nebo že např. pojišťná statistika by neměla být nabízena. S potěšením proto vítám dnešní (zatím ještě poměrně vzácné, ale nadějně) žádosti některých firem o skutečně kvalifikovaného statistika. Je třeba vytvořit vhodné prostředí a vychovávat kvalitní statistiky. Nejdříve musí však být nadaní, kvalifikovaní a potenciálně užiteční statistici-absolventi, aby vůbec existovala aspoň vzdálená naděje, že veřejnost změní svůj zatím stále převládající neutrální nebo

aspoň vzdálená naděje, že veřejnost změní svůj zatím stále převládající neutrální nebo dokonce negativní postoj k významu statistiky a potřebě statistiků. Přesvědčit mohou jen úspěšné výsledky.

Druhý extrém je *maximalisticky* přehnaný postoj k matematickému aparátu, důkazům vět, vzorcům, algoritmům a dílčím detailům různých tvrzení, postupů a metod. Dejme přednost interpretaci, filozofii, alternativám, individuálním úsudkům před objektivními a prokazatelnými pravdami. Preferujeme statistické myšlení a schopnost nadhledu na základě konkrétních znalostí před mechanickým používáním matematického aparátu. Pociťujeme však velmi mlhavou hranici mezi nezbytnou teoretickou vybaveností každého statistika a nestejným všeobecným vzděláním, statistickým rozhledem s individuálním úsudkem. Bez základních znalostí je nadhled často povrchní a může pak mít blízko k některému z výše uvedených negativních postojů ke statistice. To, že nás odmítají někteří vedoucí pracovníci jiných fakult či oborů, by nás nemělo odradit. Zkušenosti ukazují, že odmítavé postoje často vyplývají z neznalosti. Od nás to však vyžaduje dokonalou nabídku a neustálou trpělivou snahu o spolupráci.

Ad 2) V čem se statistika za posledních dvacet let změnila a co lze očekávat?

Něčeho jsme se dotkli v souvislosti s předchozími úvahami o významu statistiky a statistiků. Určitě největší změnou jsou stále lepší a dokonalejší počítače s kvalitním softwarem a v důsledku zvýšené kapacity paměti počítačů i rozsáhlé databázové systémy. Většina organizací, institucí a firem má velmi dobré počítače, kvalitní software a někdy dokonce i rozsáhlé databáze. Mějme však stále na paměti několik skutečností, které se netýkají jen informační technologie a počítačové revoluce:

- Počítače jsou naši služebníci a nikoli ředitelé.
- Matematika je základ naší vědy, ale respekt mnohých je až přehnaný.
- Naše časopisy (říká G. Hahn a R. Hoerl) přinášejí přesná řešení nerealistických problémů jako protiklad přibližných řešení skutečných a potřebných problémů.
- V minulosti platilo (říká G. Box v roce 1990), že matematika je dobrý sluha špatného pána a je možné, že v budoucnosti to stejné bude platit pro informační technologii.
- Způsob výběru, kvalita dat a kontroly budou mít stále větší význam.
- Vědomě či nevědomě snižujeme (předchozí) nematematické aspekty statistické vědy (= vypovídací schopnost dat), ale na druhé straně je znatelný přechod k nematematické výuce statistiky. Nesmí to být na úkor kvality a obsahu výuky. K tomuto problému se podrobněji vrátíme.

Pro pochopení některých změn ve statistice považuji za nesmírně zajímavý článek B. Efrona z minulého roku ve *Statistical Science* pod názvem R. A. Fisher in the 21st Century. Každá doba (říká Efron) potřebuje své hrdiny a R. A. Fisher je statistickým hrdinou dvacátého století. Vliv Fishera na statistiku 20. století byl zcela zásadní. Ve dvacátém století se statistika vyvinula ... *from a small obscure field into a big obscure field*... Fisherova snaha o jednoznačnost a nespornost statistických úsudků sice kontrastovala s individualistickým postojem např. Savageho či de Finettiho, ale přesto fisherovská filozofie je do určité míry (Efron se to snaží v podrobném členění specifikovat a prokázat) kompromis mezi uznávanou četnostní a dlouho diskutovanou bayesovskou filozofií. Mnozí označují 21. století za *plně bayesovské* a objevují se už učebnice základního kursu statistiky, které jsou psány na subjektivistickém pravděpodobnostním základu. Jiní říkají, že plný rozmach bayesovské statistiky lze očekávat až kolem roku 2010, zatímco někteří jsou ještě skeptičtější. Jako příznivec bayesovského způsobu uvažování musím radostně konstatovat, že dnes (na rozdíl od období nesmiřitelných sporů) již většina statistiků užitečnost bayesovských úvah nezpochybňuje. Bayesovské myšlenky jsou stále častěji zařazovány do výuky základního kursu i pro nestatistiky.

Za nejzásadnější změny ve *statistice pro nestatistiky* lze tedy považovat:

1. Rozvoj informační technologie se všemi důsledky.
2. Rozvoj nových oblastí statistiky (především pro velké datové soubory).
3. Ústup od matematického přístupu k výkladu.
4. Důraz na způsob pořízení údajů, na kvalitu dat a na kontroly.
5. Demokratizaci statistiky jako důsledek dostupného kvalitního softwaru.
6. Zvýšení potřeby statistického myšlení.
7. Zvýšení významu statistiky v oblasti individuálního rozhodování.
8. Růst významu bayesovského přístupu.
9. Větší důraz na zkoumání jevů a procesů ve vzájemných souvislostech.

Ad 3) Co se od nás statistiků při výuce požaduje?

Otázka zní až banálně, ale rozdíly mezi požadavky na výuku statistiky jsou poměrně značné. S oblibou říkám, že *hrajeme na přání posluchačů*, ale ve skrytu duše vím, že to je velice obtížné a mnohdy nesplnitelné. Navíc se od nás většinou nic konkrétního nechce.

Máme-li třeba k dispozici 15 (slovy patnáct) vyučovacích hodin na renomovaném typu studia význačné instituce a požaduje se od nás *jen to užitečné a hlavně pro posluchače použitelné ze statistiky*, určitě to není lehká situace. Představují-li k tomu posluchači náhodnou směs doktorů filozofie, lékařů, právníků, inženýrů techniků a inženýrů ekonomů, kteří v průměru před deseti lety měli jeden kurs statistiky s neurčitým a nespécifikovatelným obsahem, je situace těžší. Není-li navíc k dispozici pro výuku počítač s plátnem a někteří posluchači nemají ani možnost v zaměstnání či doma nahlédnout aspoň do Excelu, zdá se situace už téměř beznadějná. Co se v této chvíli od nás opravdu chce? Máme provést kvalifikovanou a zcela nadhledovou reklamu na možnosti statistiky, máme připravit vše zahrnující, ale velmi stručnou učební pomůcku, anebo máme předvést výsledky úspěšných aplikací z naší letité praxe? Nevím, ale každopádně se jako pedagog v této pozici necítím příliš ve své kůži. Přání tohoto typu však překvapivě nejsou ani tak výjimečná. Pro srovnání v USA mají MBA kursy v průměru kolem 120 hodin s tím, že počítačové učebny jsou neomezeně k dispozici a potřebná literatura je snadno dostupná. Posluchači rovněž před lety už nějaký kurs statistiky absolvovali.

Na bakalářském studiu bývalo zvykem od nás vyžadovat pro všechny studenty nestatistických oborů dvousemestrové kursy typu 2/2 (= dvě hodiny přednášek a dvě hodiny cvičení týdně). Na VŠE po sametové revoluci došlo ke snížení na 2/1 se dvěma zápočty a jednou (tzv. bakalářskou) zkouškou. V zahraničí není příliš zvykem mít cvičení (téměř neexistují), ale naopak je zvykem mít velice kvalitní, rozsáhlé a velmi podrobné učebnice i manuály k příkladům. Důraz je na domácí práci a popř. i práci na konzultacích. U nás na VŠE je jedno devadesátiminutové cvičení jednou za 14 dní, což je podle názoru studentů a katedry žalostně málo. K této problematice se však ještě vrátíme. Fakulty ani katedry od nás nic speciálního nevyžadují, a je tedy v našich možnostech naplnit tento čas podle vlastního uvážení (částečně katedry a částečně konkrétního vyučujícího na přednášce či cvičení). Na většině oborů tímto kursem už statistika končí a jen vzácně si studenti vyberou ještě něco z naší další nabídky. Na magisterském stupni žádná povinná výuka statistiky není a oborové katedry si jen velmi vzácně vyberou něco z našich kursů. Stejně jako v zahraničí je statistika přirozenou součástí povinné části všeobecného ekonomického vzdělání, ale konkrétní požadavky jsou spíše vzácností. V zahraničí převládá statistika na bakalářském i magisterském stupni (na *minor* a *major* stupni).

V posledních letech se nám pozvolně daří pronikat na některé fakulty či obory s konkrétní nabídkou (jako je marketingový výzkum, kontrola jakosti, vyhodnocování experimentů, terénní průzkumy, rozhodování za rizika a nejistoty, kurs statistiky

v angličtině se zaměřením na terminologii). Situace se v tomto smyslu tedy mírně zlepšuje a občas se od nás něco konkrétního chce a my rádi vyhovíme. Z nepochopitelných finančních důvodů však katedry mnohdy brání *své území* a z nejasných obav odmítají cokoli jiného jako nepotřebné.

Ad 4) Jsme pedagogové, vědci nebo oboji?

Tato otázka sem na první pohled nepatří, ale podle mého názoru je užitečná a má své opodstatnění. Dobrá pedagogická práce zahrnuje vše od přípravy přednášek, datových souborů (individuálních či na síti), případových studií, přes vlastní přednášky, cvičení, přípravu a opravení testů a domácích prací, semináře, konzultace, až po učebnice, skripta či jiné studijní materiály. To vše vyžaduje mnoho času a při dostatku studentů i lidí a je pak vcelku lhostejné, zda hlavní *vysvětlovací* čas je především věnován jako u nás cvičením anebo se spíše soustřeďuje do individuálních konzultací. U nás je navíc jakási stálá obava, že studenti své domácí práce i testy opiší či dokonce koupí. Hledají se různé cesty jak tomu zabránit, a to opět vyžaduje více času nebo (i) lidí. Od zahraničí se odlišujeme především tím, že většinou neexistují tak kvalitní a hlavně *na tělo* připravené studijní materiály, které by výrazně snížily potřebu podrobných přednášek a zcela vyloučily potřebu cvičení. Píšeme buď málo, anebo hlavně jinak než kurs či student potřebuje. V USA je (až přehnaně) zvykem, že ke každému kursu statistiky existuje speciální učebnice, manuály s řešením příkladů a i manuály pro asistenty, kteří se studenty přímo pracují. Od studenta se očekává, že si doma na kalkulačce či v počítači (téměř každý má studentskou verzi např. Minitabu či malý software, který byl přiložen k učebnici společně s datovými soubory) vše projde a nejasné přijde konzultovat a diskutovat. Ve Švédsku prodejna objedná příslušnou učebnici a další materiály přímo v USA a domácí studium je pak v angličtině, což nikoho neobtěžuje a pedagogy nezatěžuje psaním. Student pak opravdu nemá potřebu si ve škole příliš psát a zápisy mají spíše charakter drobných poznámek. Píšeme tedy skripta a učebnice moc, málo nebo jinak? V poslední době na VŠE vznikla atmosféra, že skript už je moc a učitelé by se měli raději věnovat vědecké práci a publikovat především jinak.

Tím se dostáváme k tomu, jaké by mělo být ideální rozdělení pracovního času mezi studium, domácí (školní) pedagogickou, domácí (školní) vědeckou práci a veřejnou aktivitu (pedagogický proces, konference, vědecké rady či schůze atd.) u asistenta, docenta či profesora na vysoké škole. Učitel zajisté celý život musí sledovat odbornou literaturu (nově vyšlé knihy, časopisy, příspěvky na konferencích či jiné). Je neustále třeba zlepšovat stávající kursy a připravovat nové. Vše se přece jen vyvíjí, takže ani materiály pro studenty nejsou konstantní a vše se stále inovuje.

Co to je ve statistice vědecká práce a kdo ji dělá? Většinou se vědeckou prací rozumí příspěvek do časopisu, sborníku nebo na konferenci ve formě prezentace určitého postupu z literatury anebo úspěšná aplikace známého postupu. Jen úplná menšina *těch naprosto nejlepších* pod vědeckou prací rozumí skutečný přínos statistické teorii či nějakou výjimečně úspěšnou aplikaci relativně neznámého přístupu k určitému problému. V předchozím textu už bylo řečeno, že tito nejlepší z nás zaručují svit našeho vědního oboru do budoucnosti, ale většina z nás musí náš obor proslavovat jinak. Podle mého názoru je prvním a nejdůležitějším úkolem učitele na libovolné škole práce pedagogická. Musíme učit způsobem, který odpovídá současnému stavu poznání, používat moderní prostředky výuky (naše aula s výstupy přímo z počítače na plátně v kombinaci s dobrým pedagogem je u nás velmi oblíbené a chválené divadlo pro studenty) a musíme se studentům skutečně věnovat. To vše vyžaduje i studium, ale hlavně čas, který pak možná chybí pro jiné činnosti. Osobně zastávám názor, že vědecká činnost pedagoga vysoké školy je nezbytná a užitečná, ale hra na vědu ve formě článků (do nevýznamných publikací či

časopisů) je zavádějící a zatěžující. Student by měl být na prvním místě našeho zájmu, jelikož za kvalitu jeho znalosti my učitelé odpovídáme.

Dovolte mi i poznámku, která se tohoto příspěvku netýká, neboť souvisí s našim finančním ohodnocením, ale nepřímo se dotýká i našeho času, který věnujeme pedagogické práci. Právě mně byla odevzdána diplomová práce pod názvem *Příjmy ve školství*. Nejsou tam překvapivá zjištění, ale srovnání s vyspělými zeměmi i s jinými profesemi u nás je přímo plačtivé. V roce 1998 byl průměrný hrubý měsíční plat ve školství o dva tisíce nižší než byl celkový průměr v ČR. Vysokoškolský učitel měl v roce 1995 plat více než třináctkrát nižší (ve srovnatelné měně) než jeho kolega v Německu. Studie OECD porovnává platy s profesí *zedník* a například v Německu má vysokoškolský učitel asi o 50 % více, zatímco u nás je to asi 14 %. Smutných čísel je mnoho a žádná vláda zatím neschvaluje ke skutečnému zlepšení. O této skutečnosti se zmiňuji jenom z toho důvodu, abychom pochopili, proč valná většina vysokoškolských učitelů hledá druhé, eventuálně i třetí zaměstnání. V USA je plat dobrý, vedlejší činnost je až zakázaná a každý se plně musí věnovat své škole, svému oboru a svému pracovišti. Za to je dobře placen. U nás mnozí přednáší či cvičí na několika místech, konzultují ve firmách, vyhotovují studie na objednávku a hledají další aktivity. Prostě si naši učitelé z pochopitelných důvodů zlepšují svoji finanční situaci. Do popředí se jednoznačně dostává kvantita pedagogické práce před její kvalitou.

Ad 5) Jakou formu výuky statistiky zvolit?

Těžko se mi podaří nějaký vyčerpávající přehled všech možností a kombinací výuky statistiky, ale pokusím se vyjmenovat některé ty nejznámější:

- a) Klasická s cvičením. Typická výuka na VŠE v posledních čtyřiceti letech. Přednášející jmenuje jednu nebo několik knih, podle kterých (s jistými výhradami) snad bude přednášet a jednu sbírku příkladů, podle kterých se bude cvičit a ukládat domácí úkoly. Cvičící má logické zpoždění za přednáškou a často na konci nestačí procvičit vše, jelikož to ještě nebylo přednášeno. Cvičící mnohdy provádí stručný výklad, protože studenti z přednášek nic nevědí.
- b) Nezávislá s cvičením. Modifikace předchozího klasického přístupu, která spočívá v relativní nezávislosti přednášejícího a cvičícího. Na přednáškách a cvičeních se dělají mírně jiné věci a hlavně styl práce a mnohdy i způsob výkladu je zcela odlišný. Tento způsob je sice především vhodný pro pokročilejší kursy na oboru *Statistika*, ale je použitelný i pro základní kurs. Přednášející se snaží o nadhled, filosofii, statistický způsob myšlení a o případné důkazy některých zásadnějších tvrzení a plně se opírá o literaturu. Cvičení je přirozeným doplňkem k přednášce. Může, ale nemusí být na počítačové učebně (např. pro počet pravděpodobnosti to není třeba, zatímco pro vícerozměrné metody je to téměř nutné). Cvičí se to, co nebylo přednášeno, a přednáší se to, co nebylo cvičeno. Časový souzvuk je velmi nutný, aby stejný problém (např. autokorelace) byl řešen současně. Velký počet náročných domácích prací (odevzdávaných pravidelně v průběhu semestru a nikoli až na konci) je pro klasický i nezávislý přístup žádoucí.
- c) Bez cvičení. Typické pro zahraničí, kde jsem měl možnost působit. U nás se týká především jednorázových nebo úžeji zaměřených kursů (např. v rámci MBA nebo nově vznikajících programů). Vyžaduje kvalitní učební pomůcky s velkým důrazem na odpovědnou a samostatnou práci studentů. Dobré kombinovat s kvalitními konzultacemi a s neustálou kontrolou odevzdaných prací.
- d) Bez přednášek. Většinou jde o počítačově orientované kursy, kde předmětem zájmu je určitý produkt (Excel) či statistický paket (u nás především Statgraphics, SPSS, NCSS nebo v poslední době i SAS). Výklad statistiky může, ale nemusí být hlavním cílem kursu.

e) Bez přednášek i bez cvičení. Vyžaduje důkladný přehled požadavků, propracované osnovy a dobrou literaturu. Důraz je na odevzdané práce a popř. i na konzultace.

Osobně preferuji pro specializované kursy (jako jsou vícerozměrné metody, regresní analýza či časové řady) způsob nezávislé kombinace přednášky a cvičení. Kurs regrese (ve kterém máme k dispozici 3/2 týdně jeden semestr) se mně a mé sympatické kolegyni už v tomto pojetí celkem daří. Pro použití tradiční varianty v základním kursu (přednášky + cvičení) je nezbytné mít k dispozici nejméně dvě hodiny cvičení týdně. Přesto bych však raději volil nezávislou formu před klasickou a velký důraz bych kladl na velký počet domácích prací. Typické pro VŠE je, že student v průběhu semestru doma téměř nepracuje. Tento stav je třeba změnit. Spíše než 2/1 bych preferoval 3/0 a opět bych kladl důraz na kvalitní studijní materiály, domácí práce a konzultace. Stále úmyslně nechávám stranou obecný styl výkladu ve smyslu matematického či nematematického přístupu, důrazu na vzorce a způsob jejich prezentace a *konkrétního obsahového* zaměření výkladu. K těmto otázkám se však již brzy dostaneme.

Ad 6) Jak jsme na tom se statistickou literaturou?

Velmi delikátní otázka, protože u nás poměrně hodně vycházejí kvalitní statistické knížky, skripta a k dispozici jsou dnes již mnohem více i zahraniční texty nejrůznějšího typu. Kdysi u nás vyšel překlad známé knihy bratří Wonnacotů, která je typickou americkou učebnicí. Má přibližně tisíc stran, jsou tam velmi přehledné vzorce i výstupy z počítače, výklad je velmi srozumitelný (i přes - ze statistického hlediska - méně kvalitní překlad), je to klasický obsah základního kursu od způsobů zjišťování, deskriptivní statistiky, přes počet pravděpodobnosti, induktivní úsudky, regresi, časové řady, teorii rozhodování, bayesovskou statistiku až po vybrané problémy (zmínky o indexech, kontrole jakosti, speciálních způsobech velkých výběrů, robustní statistice). Podle této knihy by se snadno dalo přednášet i cvičit. Je tam spousta domácích úkolů a mnoho věcí k individuálnímu zamyšlení. Nejde mi vůbec o chválu *právě této knihy* (jistě existují i mnohem lepší), ale o typ vhodné učebnice. Bylo by možné uvést i některé další, například komplet (knihy spolu se dvěma manuály) *Statistical Thinking for Managers* (D. K. Hildebranda a L. Otta) či podobný komplet *Statistical Methods for Business and Economics* (D. L. Harnetta a A. K. Soniho) stojí za zaznamenání. Takových jsou však bez přehánění stovky a je až zarážející, že téměř obsahově stejné knihy mohou v takovém rozsahu opakovaně vycházet. *Pan profesor jednoduše předepíše svou knihu a všichni si ji koupí*. Knihy tohoto typu vycházejí v tak velkých nákladech, že je to z ekonomického hlediska až překvapivé.

Kromě těchto rozsáhlých *kompletních* učebnic lze neustále pozorovat nejrůznější nové tendence a snahy předložit nové texty – *jiné a jednoduché*. Za všechny bych jmenoval alespoň o něco starší (1986) *Statistics, a Conceptual Approach* (K. L. Weldona), která je snahou o základní kurs téměř bez vzorců s mnoha povídavými příklady ze skutečného života, a relativně novou (1996) *Statistics, a Bayesian Perspective* (D. A. Berryho), která je jedním z prvních pokusů o výklad základního kursu v bayesovském pojetí.

U nás zatím převládá tradiční způsob výkladu, i když např. někteří moji kolegové se neustále snaží o vylepšení svých dřívějších textů a alespoň cenou některých již předhonili zahraniční zvyklosti (v USA či Švédsku popisované ucelené učebnice základního kursu stojí kolem 20 až 30 dolarů a úzce specializované statistické knihy jsou drahé, tj. kolem 60 a více dolarů). Neříkám, že píšeme špatně, ale píšeme trochu jinak. Pro studenta-*nestatistika* bývá čtení našich knih mnohdy obtížné, takže bohužel většinou naše knihy nechválí. Je velmi obtížné podle těchto knih vést přednášky na bakalářském či jiném podobném studiu. Nerad bych přehnaně kritizoval a bagatelizoval, protože u nás vyšla celá řada velmi dobrých knih. Stále však mluvíme o učebnici pro *nestatistiky*. Možná, že jednou dojdeme ke švédskému modelu přímého dovozu učebnic s tím, že to třeba bude

jednodušší a levnější, ale osobně bych to zatím nedoporučil. Přece bychom měli být schopni napsat vynikající podobnou učebnici základního kursu, která by pak měla mnoho opakovaných vydání a pro delší dobu by byla vyhovující.

Zásadně jiná situace je při vydávání úzce zaměřených odborných textů. Očekávaných čtenářů je tak málo, že se žádnému nakladatelství (bez dotace či sponzorského příspěvku) nevyplatí odbornou knihu tohoto typu vydat. Takové knihy mohou (musí) být drahé, ale nemělo by to překročit *rozumné* meze. Bolestně nesu skutečnost, že u nás téměř neexistuje dobrá monograficky orientovaná a zároveň uživateli uznávaná kniha ke statistickým metodám obecně a vícerozměrným metodám zvláště. Neznám ani dobrou statisticky orientovanou knihu k marketingovému výzkumu a o monografii k bayesovským přístupům si nechávám zdát. Žádosti o podporu ze strany školy či snahy o získání grantu se zatím míjejí účinkem. Snad jiní mají či budou mít větší štěstí, takže odpusťte, že v této souvislosti mluvím více o sobě.

Ad 7) Jaký obsahový a časový rozsah výuky statistiky bychom doporučili?

Několikrát zmíněných 2/2 základního kursu pro bakaláře po dva semestry znamená (při 13 běžných týdnech v semestru) celkem 104 *vyučovacích* hodin neboli (při 45 minutách vyučovací hodiny) celkově 78 hodin. To je jistě zcela dostatečné za předpokladu, že tím statistika pro studenta nekončí. Mohl by a měl by se s ní v budoucím dalším studiu setkávat v různých podobách podle typu zvoleného oboru. Statistické myšlení by mělo doprovázet nejen jeho další studium. Pro srovnání v USA i ve Švédsku jsou kursy tohoto typu (v rozsahu kolem 50 až 60 hodin) pouze jednosemestrové, ale na ně navazuje další (nejméně dvojnásobný) počet hodin v rámci *major* studia. Přibližný rozsah kolem 65 (např. 3/2) vyučovacích hodin na bakalářském a stejný rozsah na druhém (magisterském, inženýrském či jinak pojmenovaném) stupni bych určitě považoval za rozumný. Proti současnému stavu na VŠE by to znamenalo užitečné zvýšení celkově asi o 30 %. Oba nezávislé kursy by byly ukončené zkouškou, jejíž formu i obsah by též bylo možné zevrubně diskutovat. Otázce zkoušení se zatím zcela vyhýbám. Problém je u nás v tom, že kdybychom jen jednou připustili, že statistika v povinném základu bude jen jeden semestr s tím, že bude potom i na inženýrském stupni, budou sice všichni souhlasit s první částí (=snížení) a pak nám neumožní realizaci té druhé části (=zvýšení). V konečném výsledku by rozsah vyučované statistiky byl nižší, což je pro nás pochopitelně zcela nepřijatelné.

Obsahové zaměření základního kursu se může jevit jako *jasné a bezproblémové*. Nežlobte se, ale nemohu si odpustit zmínku o skvělé knize Ch. Wanga (1993), která má název *Sense and Nonsense of Statistical Inference* (Controversy, Misuse, and Subtlety). Autor zde uvádí řadu nádherných pojednání i příkladů špatného pochopení, nevhodné interpretace a nevhodné prezentace, nešťastného výkladu až elementárních pojmů a postupů. Kniha je určena všem *statistikům* obecně a učitelům statistiky speciálně, ale sloužit může každému, kdo se zajímá o statistiku a snaží se používat statistické nástroje. Autor knihy se sice zajímá především o *testy významnosti*, *statistické zobecňování*, *příčinnost* a *subjektivní úsudky*, ale ve svých úvahách se dotýká téměř všech oblastí statistiky. Každý z nás přece pociťuje mechaničnost mnohých výkladů, zákeřnost různých interpretací, vícesmyslnost řady běžných výroků, omezenou použitelnost doporučovaných postupů a metod, jakož i nedostatky častých tvrzení. Za všechny bych jmenoval rozpor mezi praktickou a statistickou významností, obtíže s hodnocením individuálního vlivu činitelů, nevyjasněné otázky měření asociace ve vztahu ke kauzální interpretaci, široké potíže s náhodností, mechanické hodnocení významu subjektivních úsudků, ale seznam problémů je ve skutečnosti mnohem delší. U nás na katedře se přibližně doporučuje pro první semestr třetinový podíl popisné statistiky, pravděpodobnosti a matematické statistiky, zatímco ve druhém semestru (konkrétní časové rozvržení podle postoje přednášejícího) se vykládá regresní analýza, analýza časových řad a indexní teorie. Přednášející se nechtějí

nechat svázat do předpisu *kolik čeho učit*, takže v tomto směru existují značné rozdíly v pohledu na užitečnost různých témat. K tomuto se ještě podrobněji vrátím.

Ad 8) Kdo a jak učí statistiku kromě nás?

To úzce souvisí s již zmíněnou demokratizací statistiky. V prvním letošním čísle (February 1999) věnoval jeden z nejlepších (pokus mohu posoudit) statistických časopisů *The American Statistician* speciální pozornost základním kursům statistiky. Pod označením: *Special Section – Undergraduate Statistics Education: What Should Change* byly zařazeny dva zásadní příspěvky. J. J. Higgins svůj příspěvek pojmenoval *Nonmathematical Statistics: A New Direction for the Undergraduate Discipline* a i u nás známý R. V. Hogg napsal článek s názvem *Let's Use Continuous Quality Improvement in Our Statistics Programs*. V Higginsově článku se nedejme zmást označením *Undergraduate*, protože ve skutečnosti má autor na mysli (jak v úvodu zdůrazňuje) *Curriculum for Undergraduate Majors*. Zřejmě totiž statistické katedry v USA věnují především pozornost přípravě M.S. a PhD statistiků před běžnou výchovou statistiků i nestatistiků na *minor*, ale i na *major* stupni. Mluví se o 220 univerzitách a colleges, které nabízejí čtyřleté statistické programy (možná něco ve smyslu našeho oboru na VŠE – PH), což představuje asi 10 % všech čtyřletých univerzit a colleges v USA. Přibližně 60 % těchto programů je na nestatistických katedrách nebo na katedrách, které kombinují statistiku s něčím jiným (hlavně s matematikou). Pro srovnání více než 50 % univerzit a colleges mají program v *computer science* a mnohé mají samostatné takto zaměřené katedry.

Pamatuji po roce 1989 ostré střety s představiteli jiných fakult, kdy se nám na akademickém senátu školy a na rektorátu téměř zázrakem podařilo s Jiřím Žváčkem vybojovat nejdříve naši *Fakultu informatiky a statistiky* a posléze i samostatný obor *Statistika a ekonometrie*. Domnívám se, že samostatných statistických programů (oborů) není v České republice ani pět (vím o dvou až třech), takže možná nemáme ani těch amerických 10 %. I my máme zkušenosti se snahou jiných kateder či odborníků z jiných oborů vyučovat *naše* předměty či kurzy. Týká se to nejen matematiků v oblasti počtu pravděpodobnosti, ekonometrů v oblasti matematické statistiky, ale i obchodníků v oblasti marketingových statistických metod, informatiků v oblasti bayesovských sítí a velkých databází i dalších. Například nepochybujeme o schopnostech matematiků učit počet pravděpodobnosti, ale matematici to prostě učí *jinak* než my statistici potřebujeme. K tomu M. Kendall už v roce 1976 říká, že ... *existuje typ matematika, který je nebezpečný našim potřebám anebo přinejmenším brání tomu, abychom se dostali do světa vědy a businessu...*. Mnozí statistici (Gibbons – Freund 1980, Moore 1988 a další) diskutují o tom, zda matematici mají učit statistiku a volají po nezávislosti statistické disciplíny. Statistické myšlení je důležitá složka nezaujatého postoje a vzdělání vůbec, ale přesto na univerzitách nepohlíží na statistiku jako na základní disciplínu ve stejném významu jako na matematiku, historii, biologii či jiné obory. Pravděpodobně to nejlépe vystihl G. Box v *Technometrice* (1990):

„Statistics is, or should be, about scientific investigation and how to do it better, but many statisticians believe it is a branch of mathematics. Universities and government agencies such as the National Science Foundation believe this also and mistakenly categorize it as one of the mathematical sciences. Now I agree that the physicist, the chemist, the engineer, and the statistician can never know too much mathematics, but their objectives should be better physics, better chemistry, better engineering, and in the case of statistics better investigation. Whether in any given study this implies more or less mathematics is incidental.“

Higgins dále říká, že existuje nematematická stránka statistiky a její dobrý výklad může přetáhnout ty, kteří dnes jdou na Computer Science, Engineering nebo jiné kvantitativní oblasti. Naprosté oddělení statistiky a matematiky má ovšem i finanční bariéry, ale ústředním problémem je ...*lack of consensus within the statistics community as to what the undergraduate statistics degree should be about...*

Zajímavá je Higginsova zmínka o inzerátu (únor 1998), který byl podáván Social Science Reference Center pod označením *Statistics: Data Services Specialist (New Position)*. Asi chtěli někoho graduovaného v kvantitativních vědách, ale je to poučení, co si vlastně představovali, že takový statistik všechno umí. Uchazeč o toto místo měl být schopen konzultovat v metodologických statistických otázkách, měl umět SAS, SPSS a Stata, měl ovládat operační systémy Windows, UNIX a Mac, měl být schopen rozvíjet web stránky na internetu, měl mít zkušenosti se sběrem dat různého typu. Ten uchazeč měl umět to co patří do oboru *Computer Statistics* i to co patří do oboru *Statistika*. Zajímavé však je, že inzerát nevyžadoval absolvování statistického oboru či nějaký titul ve statistice. Pravděpodobně autoři tohoto inzerátu ani neměli příliš velké tušení o možných znalostech a dovednostech statistika.

K této problematice se ještě vrátíme v souvislosti s obsahovým zaměřením studia na oboru *Statistika*. Každopádně je naší povinností pečlivě sledovat vše, co uživatelé i zákazníci od nás chtějí, aby tradiční zapojení matematiky nevyústilo v nedostatečnou pozornost věnovanou (dále uvedeným) *nematematickým aspektům statistiky*.

Ad 9) Jaký by měl být obsah základního kursu statistiky pro bakaláře?

Odpověď na tuto otázku nemůže být překvapivá, protože je těžké přijít s něčím zázračně novým. Je možné diskutovat některé koncepční otázky, které se týkají samotného způsobu výkladu (matematické či nematematické pojetí), významu výkladu (znalosti) vzorců a některých algoritmů, jakož i oblast pravděpodobnostního a induktivního způsobu myšlení. Diskutovat je též možné některé dílčí a technické otázky výkladu a zkoušení. Měli bychom zvážit možnost uplatnění myšlenky o preferenci nadhledového výkladu statistických metod před popisem těchto metod samotných, jak jej v posledních letech nabízí celá řada statistiků (např. R. J. Harris ve známé monografii *A Primer of Multivariate Statistics* (1985) v první kapitole pod názvem *The forest before the trees*).

V diskusi na této konferenci by bylo užitečné se vyjádřit aspoň k následujícím otázkám základního kursu statistiky pro nestatistiky bakaláře v rozsahu kolem 65 vyučovacích hodin v jednom semestru uzavřeném zkouškou:

Jako vstup do diskuse zatím nabízím pracovní definici, podle které je *Statistika (vědní) obor, který se zabývá získáváním, zpracováním a analýzou dat pro potřeby rozhodování*. Každý vyučující základního kursu by asi měl mít svůj osobní názor na tyto otázky:

- Jaké znalosti by měl mít student po absolvování základního kursu statistiky?
- Jaké předběžné znalosti z matematiky je třeba vyžadovat?
- Jaké předběžné počítačové znalosti je třeba vyžadovat?
- Jsou cvičení nutná a jak mají být organizována?
- Do jaké míry má smysl odvozovat vzorce a vysvětlovat jejich logiku?
- Měl by seznam vzorců být připraven katedrou?
- Jaký význam by měly mít individuální konzultace?
- Kolik času věnovat způsobu pořízení dat a jejich prezentaci?
- Pravděpodobnost je objektivní nebo (i) subjektivní kategorie?
- Co učit z pravděpodobnosti?
- Jaká je potřeba induktivních úsudků v různých oblastech života?
- Jaký je rozsah teoretického aparátu pro výklad statistické indukce?

- Mají všichni na katedře učit zcela stejný základní kurs?
- Jaký je vhodný stupeň zapojení matematického přístupu do výkladu?
- Jaký má být význam případových studií?
- Jaký by měl být typ a rozsah domácích prací?
- Jaké formy kontroly činnosti studentů používat?
- Jaký by měl být obsah zkoušky a jaká je její nejvhodnější forma?

K této problematice mi dovoluji jednu poznámku, která se týká hranice mezi *vzorci*, které by student měl nejlépe nejen při zkoušce znát z paměti (měl by je považovat za trvalou součást svých statistických znalostí) a mezi *vzorci*, které je vždy možné si v případě potřeby najít. U těch druhých může být vůbec otázkou, zda je třeba s nimi studenty (nestatistiky) při výkladu zatěžovat. Někdy úsměvně na katedře poznamenáváme, že student by si neměl spojit představu o aritmetickém průměru se zmáčknutím klávesy F6 nebo ENTER v příslušném okénku. Zároveň se tím částečně řeší i problém hranice mezi tím, co vypočítáme z *hlavy* nebo na co nám stačí kalkulačka, proti úlohám, které řešíme pouze pomocí počítače.

Osobně zastávám názor, že výklad základních pojmů (příklady úloh, k jejichž řešení může statistik přispět, dělení proměnných podle nejrůznějších hledisek, vysvětlení pojmů jako je populace a výběr, přednosti a nevýhody různých způsobů pořízení dat atd.) je téma až na dvě přednášky. Pak je vhodné na nějakém hodně malém souboru s čísly typu 1, 2, 3, 4, 5 (např. počet členů domácnosti) ukázat výpočet *aritmetického průměru*, *mediánu*, *modu*, *rozptylu*, *směrodatné odchylky* a třeba i *kvartilů*, *variačního rozpětí* a *kvartilového rozpětí* demonstrovat význam a podstatu charakteristik *úrovně* a *variability*. Po vysvětlení klíčového pojmu *rozdělení četností* je možné se zmínit o typickém (*normálním*) rozdělení *jednomodálním* a *symetrickém*. Pojmů by nemělo být příliš (jsou v učebnici a vzorce každý dostal) a hlavně by každý měl pochopit, že různé charakteristiky a nestejné způsoby prezentace dat mají v odlišných situacích proměnlivý význam. Časový, prostorový a věcný aspekt v datech je též třeba vysvětlit. Souběžně na cvičení je vhodné vytvořit vlastní soubor (např. o studentech ročníku) a je nutně třeba provést relativně zevrubnou analýzu dat pomocí počítače (grafy, tabulky, rozdíly, podíly, charakteristika, variabilita, šikmost, extrémní a chybějící hodnoty atd.). Průzkumová analýza dat a její možnosti – to je příležitost pro cvičícího. Tomu všemu dohromady bych nevěnoval více než čtyři týdny (ideálně 3/2 a realisticky 2/2). Pravděpodobnostní myšlení studentům většinou zcela chybí, přičemž je to nezbytný základ dalšího výkladu. Tuto část rozhodně není dobré na jedné straně odbyt několika větami v jedné či dvou přednáškách a na druhé straně nemá cenu uvádět komplikované vzorce např. o vztahu mezi podmíněnými, marginálními a simultánními zákony rozdělení.

Za velmi závažnou a potřebnou považuji diskusi o filozofii výkladu pojmu *pravděpodobnost*. Je naprosto nutné jednoznačně odlišit použitelnost klasické definice v herních, losovacích a výběrových situacích od statistického (četnostního) pojetí u opakovatelných pokusů za přibližně stejných podmínek a od subjektivního pojetí v neopakovatelných rozhodovacích úlohách. Pojmy (*ne*)*slučitelnost* a především (*ne*)*závislost* musí každý chápat v souvislosti s „beztrestným“ sčítáním a násobením pravděpodobností. Otázka logického součtu a součinu není ani zdaleka samoúčelná a je tedy vhodné mít dostatek jednoduchých příkladů (řešených pomocí kalkulačky). Vysvětlení „organického spojení“ mezi statistickým a pravděpodobnostním rozdělením určitě není ztráta času. *Binomické*, *rovnoměrné* a *normální rozdělení* je naprosté minimum. Pochopení spojitosti mezi *očekávanou hodnotou* a *průměrem*, jakož i mezi jinými statistickými a pravděpodobnostními charakteristikami se opět vyplácí. Pojmy pravděpodobnostní a distribuční funkce, kvantily spojitě náhodné veličiny, výjimečné postavení normálního rozdělení a podstata zákona velkých čísel uzavírá výklad pravděpodobnosti, kterému je též možné věnovat nejvýše čtyři týdny (druhou třetinu semestru). Doma by každý student měl spočítat několik desítek

elementárních příkladů (poslouží sbírky či učebnice), protože bez nich si neosvoji pravděpodobnostní přístup, který by měl být užitečný nejen pro výklad statistiky! *Statistická generalizace* (z náhodného výběru na populaci) úzce souvisí s problematikou *odhadu a s testováním hypotéz*. Výklad na přednáškách musí být velmi citlivý a především je třeba „nezahltit“ studenta příliš velkým počtem pojmů. Logika celého přístupu a použitelnost této techniky musí mít přednost před např. *Neymanovou–Pearsonovou větou* či *testem věrohodnostním poměrem*. Neučíme matematickou statistiku, ale ukazujeme některé nejdůležitější možnosti statistické indukce a to není totéž! Nepochopení myšlenky např. *P-hodnot, hladiny významnosti, kritických hodnot, přesnosti a spolehlivosti odhadu*, případný úžas nad existencí *chi-kvadrát, t, či F rozdělení* nebo některých dalších pojmů, to je závažný nedostatek výkladu, zatímco mnohé jiné aspekty či pojmy jsou méně podstatné. Výuku základního kursu uzavírá přehled různých typů úloh, které se týkají vztahů, klasifikace, srovnání, a dále některé typy testů, které též mohou pomoci při jejich řešení (rozsah závisí na času, který máme k dispozici). Celé je to příprava budoucího výkladu statistických metod.

Ad 10) Jaký by měl být obsah kursu statistiky pro magistry na různých oborech?

Samostatnou a možná z hlediska způsobu a formy výkladu nejobtížnější oblastí základního kursu pro nestatistiky jsou statistické metody. Je vůbec otázkou, které to mají být, na jakém oboru, v jakém rozsahu a v jaké struktuře je zařadit. Nadhledový přehled je určitě užitečný pro každého, ale další etapa výkladu je podle mého názoru velice sporná. Každopádně by se tady nemělo projevit osobní zaměření přednášejících ani individuální preference různých oblastí statistiky. Mnohé učebnice z posledních let začínají přehledem a klasifikací statistických metod. Třídění statistických metod je většinou prováděno podle:

- obecného cíle úlohy (srovnávací, klasifikační, extrapoláčn^í či modelovací),
- charakteru závislosti či vztahu (jednostranné a vzájemné vztahy),
- počtu vysvětlovaných a vysvětlujících proměnných a v rámci různých variant podle charakteru proměnných (kvantitativní, alternativní, kategoriální, umělé, transformované, zpožděné),
- záměru klasifikační úlohy (pro popis, vysvětlení či pro rozhodování),
- typu srovnávací úlohy (prostorové, věcné či časové a v rámci různých možností podle počtu posuzovaných jednotek či období),
- charakteru úlohy (statická či dynamická),
- specifik dané oblasti (kontrola jakosti, marketingová rozhodnutí, ekonomické předpovědi, experimentování s modelem).

Někteří pedagogové konkrétní používané testy považují za samostatnou statistickou metodu, zatímco jiní je po obecném výkladu zařazují podle charakteru úlohy a zvolené metody. Patřím spíše k těm druhým. Na VŠE je zvykem druhý semestr základního kursu věnovat výkladu *regresní analýzy, indexní teorie a analýzy časových řad*, zmínit se poměrně stručně o *analýze rozptylu* a o *kontingenční tabulce*.

Na regresní a korelační analýze je možné demonstrovat pedagogické rozpory, které pociťuje asi každý přednášející a jeho cvičící při výkladu určité statistické metody. Pojem závislosti považují v počtu pravděpodobnosti i ve statistice za jeden z nejdůležitějších a nejkomplikovanějších. Na závislost je možné pohlížet z filozofického, matematického, věcného a samozřejmě i z našeho specificky pravděpodobnostního nebo ze statistického hlediska. Regresní analýza je jednou z mnoha možností přístupu k tomto problému. Pro korelační analýzu jsou všechny zúčastněné kvantitativní proměnné náhodnými veličinami, zatímco pro regresní analýzu je výhodné (a většinou nespílitelné), když kvantitativní vysvětlující proměnné jsou pod kontrolou experimentátora. Ve světoznámé monografii

Applied Regression Analysis Drapera a Smithe příklad s portlandským cementem ve vztahu k době tuhnutí betonu se táhne celým textem. Odhadují se různé typy *regresních funkcí*, počítají se rozličné *korelační koeficienty*, demonstrují se speciální postupy jako je *ridge regrese* a otázka možnosti manipulace s hodnotami vysvětlujících proměnných je velkoryse „přehlédnuta“. Otázka oprávněnosti podmínek regresní analýzy, stejně jako skutečnost, že celý příklad má jen třináct pozorování, zůstává mimo pozornost autorů této věhlasné knihy. Nešťastné až nepoužitelné aplikace regresní analýzy se táhnou celou historií této nejoblíbenější, ale bohužel též nejneužívanější statistické metody. Publikace o regresní analýze mívají běžně více než pět set stran, vycházejí samostatné knihy o regresní grafice, regresní diagnostice, nelineární regresi, doplňkových informacích, bayesovské regresi a mnoho dalších.

Jaký rozsah výkladu na přednáškách je tedy vhodný a do jaké míry má smysl se pouštět do komplexní regresní úlohy na cvičení? Stačí, když student pochopí význam samotné regresní funkce, obtíže s výběrem a kontrolou hodnot vysvětlujících proměnných, jakož i interpretaci regresních či jiných koeficientů? Kde je ta vhodná hranice a jaká je nejnadějnější forma výkladu? Především v našich učebnicích chybí příklady **prokazatelně úspěšných** regresních úloh. Počítačové sestavy s grafy a s výsledky různého charakteru není možné vytvářet až na cvičení, kde pozornost může být věnována jen zcela elementárním aspektům problematiky. Například soustředění na regresní přímku je pedagogicky pochopitelné, ale nedostatečné a uživatelsky až zavádějící. Interpretace regresního a korelačního koeficientu je nutná, ale pro *dobrou reklamu metody* je to málo. Diskutujme tuto problematiku a zvažujme vhodný přístup! Netýká se to jen právě regresní a korelační analýzy, týká se to celé řady (především vícerozměrných) statistických metod. Je třeba přesvědčit širokou odbornou veřejnost o užitečnosti našich postupů a nástrojů. Jen tak máme naději rozšířit naše pedagogické i pracovní zapojení do jiných oborů! Nemohu si odpustit poznámku k indexní analýze, která v minulosti ve výkladu základního kursu statistiky na VŠE pro nestatistiky (i statistiky) hrála mimofádně přehnaně významnou úlohu. Analytický význam indexní teorie (různé typy rozkladů a jejich interpretace) začínal postupně převládat nad původním jednoznačně popisným významem indexů. Mimochodem Irwin Fisher tuto skutečnost vyčítal Montgomerymu a dalším autorům logaritmického rozkladu už před více než šedesáti lety. Dnes se u nás na VŠE indexní teorie vyučuje sice v mnohem menším rozsahu než dříve, ale stále ji někteří přednášející věnují větší pozornost než si zaslouhuje. Je to na úkor jiných závažnějších témat.

Byl bych velice rád, kdyby výklad statistických metod byl součástí samostatného kursu na inženýrském (magisterském) stupni s nejružnějšími modifikacemi podle typu a charakteru školy, fakulty či oboru. Preference práce s kategoriálními daty na společensky zaměřených oborech, marketingové metody a techniky na obchodních i manažerských oborech, kontrola jakosti a využití experimentálních dat při řešení technologických a výrobních otázek, otázky predikce a makroekonomického i mikroekonomického modelování či předvídání pro *národohospodáře* i analytiku velkých společností atd. Určitě nelze každého učit všechno a je třeba velmi pečlivě vybírat to (v dané situaci) nejvhodnější. Ve vlastním výkladu preferovat filozofii přístupu a diskusi nad konkrétními výstupy úspěšných úloh před technickými otázkami příslušných algoritmů a před ryze matematickými či statistickými aspekty posuzované metody. Víím, je někdy velmi obtížné nebagatelizovat a příliš nezjednodušovat statistickou stránku výkladu s tím, že rozhodující je interpretace a použití, ale nalezení toho vhodného kompromisu je možné. Samozřejmě, že to vyžaduje hluboké znalosti a zkušenosti vyučujících.

Připomínám, že na naší katedře máme pro všechno dohromady na bakalářském stupni dvousemestrový kurs rozsahu 2/1 týdně se dvěma zápočty a s jednou zkouškou. Na druhém stupni většinou u jiných oborů nemáme nic, takže mé úvahy o magisterském kursu patří zatím spíše do oblasti utopie. K tomu, abychom s takovým kursem prorazili, musíme nejdříve hodně pracovat sami na sobě. Musíme se věnovat prezentaci výmluv-

ných a skutečně úspěšných využití statistiky. V popředí zájmu jsou dnes otázky pořízení dat, využití databází, kvalita dat před kvantitou, nematematické aspekty statistiky ve výkladu, nové formy výuky s využitím rostoucích možností počítačů; naším úkolem je aktivní zapojení do nově vznikajících oblastí statistiky, reakce na kritiku odborníků z jiných profesí a mnohé další. Jen tak máme naději, že nás druzí budou chtít mít ve svých programech, při účasti na grantech, společných projektech a jsme znovu zpátky u významu a postavení statistiky...

Ad 11) Jaké máme naděje na obchodních, podnikatelských či MBA kursech?

Obecně záleží na typu studia. Naše naděje jsou solidní, ale musíme se neustále zlepšovat a přicházet s atraktivní nabídkou! Nikdo neodmítá zařadit *statistiku* jako přirozený základ všeobecného vzdělání (manažera, ekonoma, finančního či daňového poradce), ale k tomuto účelu nám skutečně může být přiděleno jen oněch na začátku zmíněných 15 hodin přednášek. Tak špatné je to však jen výjimečně. K proniknutí do centra dění *nestatistické* výuky bychom měli přijít s něčím novým a možná i částečně neznámým. Již bylo řečeno, že na katedře obchodního podnikání pronikáme s terénními a marketingovými průzkumy, na informatice s bayesovskými sítěmi, na podnikatelské fakultě s kontrolou jakosti, na všech typech studia s rozhodováním za rizika a nejistoty a s anglickou terminologií. Na MBA studiu je vždy zájem o práci s daty při použití Excelu, někdy o prezentaci různých výsledků a analýzu cen, ve vzácných případech o metodu conjoint a korespondenční analýzu, obecně o výzkumy trhu různého typu a možná i jiné oblasti. Záleží na kvalitě a přesvědčivosti naší práce.

Ad 12) Jak organizovat studium na oboru Statistika?

To je velmi zásadní a nesmírně důležitá otázka. S naší nabídkou kursů, jakož i s předpokládanou strukturou předmětů našeho oboru, se podrobně zabýváme v bulletinu. Proti minulosti jsme velmi odvážně vynechali pomlčku, takže místo statisticko-pojistné inženýrství se náš obor bude jmenovat statistické a pojistné inženýrství. S vedoucím katedry velmi často otázky našeho oboru diskutuji a v něčem máme sice odlišné názory, ale nakonec se většinou dohodneme. Pro zajímavost uvádím uspořádání, ve kterém se v první části (A) výhradně soustřeďuji na nabídku naší katedry (statistiky a pravděpodobnosti) a případné nabídky příbuzných kateder (demografie, ekonomické statistiky, ekonometrie a popř. i matematiky a některých dalších) nechávám úmyslně stranou. Druhá část (B) shrnuje některé nové myšlenky a přístup J. J. Higginse z výše citovaného článku o nematematické statistice, což může být možná zajímavou inspirací:

A. Tradiční předměty

Práce s daty	Analýza časových řad
Úvod do statistiky	Vicerozměrné metody
Počet pravděpodobnosti	Neparametrické metody
Matematická statistika	Robustní statistika
Teorie výběrových šetření	Bayesovská statistika
Analýza rozptylu	Statistická kontrola jakosti
Regresní analýza	Kategoriální data

B. Nové předměty

1. Vědecký proces

Měření a pozorování, přirozená variabilita, znáhodnění, náhodný výběr, opakovatelnost, popisná analýza dat (vstupní kurs, který se opírá o experimenty studentů)

2. *Plánování, příprava a organizace výzkumů*

Studium význačných aspektů výzkumů různého typu od přípravné fáze, přes sběr a uložení dat. Vymezení cílové populace, různé typy výběrových schémat, výhody a nevýhody různých způsobů sběru dat (telefonní, poštovní, focus groups, domácí návštěvy), příprava dotazníků, problémy s nonresponse zkreslením.

3. *Plánování a organizace vědeckých experimentů*

Omezení a výhody běžných plánů, identifikace důležitých faktorů a zdrojů variability, příklady z různých oblastí (výroba, medicína, lékařství, zemědělství, životní prostředí).

4. *Statistický software pro práci s daty*

Kapacita paměti pro různé metody, specifika, popisná analýza, manipulace s daty, práce se soubory, makra, programovací možnosti, práce s maticemi, grafika (to může být několik kursů)

5. *Statistická grafika*

Efektivní grafické znázornění kvantitativní informace, vícerozměrná data, grafický software.

6. *Computer Science ve statistice*

Příprava databází, práce s velkými databázemi, počítačové sítě, bezpečnost a ochrana dat, web aplikace (společně s katedrou informatiky)

7. *Statistická komunikace*

Komunikace s pravděpodobnými uživateli, schopnost klást správné otázky, komunikace v oblasti statistických myšlenek, komunikace s médii, typické chyby ve statistice špatné použití statistiky (společně se statistickou laboratoří).

8. *Manažerské principy ve statistice*

Manažerské přístupy v oblasti pořizování dat a při přípravě vědeckých studií, plány projektu, otázky vedení a kontroly, týmová práce (společně s katedrou managementu).

Ad 13) Jak organizovat postgraduální studium statistiků?

Podle mého názoru největší slabina VŠE. Místo specializovaného a vysoce odborného postgraduálního studia se závěrečným titulem M.S. nebo PhD. ve statistice je organizováno doktorské studium s titulem Dr. velmi blízké dřívějšímu studiu s titulem CSc. Povinné předměty, volitelné předměty a hlavně velmi málo statistiky. Zájem o toto studium je velmi malý až téměř žádný. Chyba se stala hned na začátku při vzniku všezaahrnujícího organizačního systému, do kterého zainteresované katedry vstupují jen velmi málo. Projít celým systémem až k titulu je poměrně pracné a asi to neodpovídá výsledné odborné prestiži. Informací o organizačním přístupu škol v zahraničí mnoho nemám, takže zasvěcené diskusní příspěvky by asi byly velmi vítané.