

Jiří Neubauer, Marek Sedlačík, Oldřich Kříž

Základy statistiky

Aplikace v technických
a ekonomických oborech
3., rozšířené vydání

- měření a zjišťování
- teoretické modely
- empirické modely
- základy indukční statistiky
- počítačové zpracování dat
- praktické užití statistiky

Jiří Neubauer, Marek Sedlačík, Oldřich Kříž

Základy statistiky

**Aplikace v technických
a ekonomických oborech**

3., rozšířené vydání

Grada Publishing

Upozornění pro čtenáře a uživatele této knihy

Všechna práva vyhrazena. Žádná část této tištěné či elektronické knihy nesmí být reprodukována a šířena v papírové, elektronické či jiné podobě bez předchozího písemného souhlasu nakladatele. Neoprávněné užití této knihy bude **trestně stíháno**.

Jiří Neubauer, Marek Sedláčik, Oldřich Kříž

Základy statistiky

Aplikace v technických a ekonomických oborech

3., rozšířené vydání

Vydala Grada Publishing, a.s.
U Průhonu 22, Praha 7
obchod@grada.cz, www.grada.cz
tel.: +420 234 264 401
jako svou 8204. publikaci

Lektoroval: doc. RNDr. Jaroslav Michálek, CSc.
Recenzovali: prof. RNDr. Petr Ponížil, Ph.D., doc. Mgr. Zuzana Hübnerová, Ph.D.
Odpovědná redaktorka: Věra Slavíková
Počet stran 296
Třetí vydání, Praha 2021
Vytiskla TISKÁRNA V RÁJI, s.r.o., Pardubice

Vydání odborné knihy schválila Vědecká redakce nakladatelství Grada Publishing, a.s.

© Grada Publishing, a.s., 2021
Cover Design © Grada Publishing, a. s., 2021
Cover Photo © Depositphotos/zphoto

Názvy produktů, firem apod. použité v knize mohou být ochrannými známkami nebo registrovanými ochrannými známkami příslušných vlastníků.

ISBN 978-80-271-4484-6 (pdf)
ISBN 978-80-271-3421-2 (print)

Obsah

1	Úvod do statistiky	15
1.1	Historický přehled	15
1.2	Význam a pojetí moderní statistiky	21
1.3	Statistická jednotka a statistický soubor	25
1.4	Statistický znak	28
2	Popisná statistika	33
2.1	Vyjadřovací prostředky statistiky	33
2.2	Základní zpracování číselných dat	36
2.2.1	Neroztříděná data	36
2.2.2	Bodové rozdělení četností	37
2.2.3	Intervalové rozdělení četností	39
2.3	Charakteristiky polohy	43
2.4	Charakteristiky variability	49
2.5	Charakteristiky koncentrace	56
2.6	Kompletní zpracování dat pomocí aplikace STAT1	58
3	Pravděpodobnost	64
3.1	Základy kombinatoriky	64
3.2	Náhodný pokus a náhodný jev	68
3.3	Pravděpodobnost náhodného jevu	72
3.4	Klasická definice pravděpodobnosti	73
3.5	Geometrická definice pravděpodobnosti	77
3.6	Podmíněná pravděpodobnost	80
3.7	Pravidlo o násobení pravděpodobností	81
3.8	Pravidlo o sčítání pravděpodobností	84
3.9	Úplná pravděpodobnost a Bayesův vzorec	88

4	Náhodná veličina	92
4.1	Náhodná veličina	92
4.2	Distribuční funkce náhodné veličiny	93
4.3	Diskrétní náhodná veličina	94
4.4	Spojité náhodná veličina	98
4.5	Charakteristiky polohy	101
4.6	Charakteristiky variability	105
4.7	Charakteristiky koncentrace	107
5	Modely rozdělení pravděpodobností pro diskrétní náhodné veličiny	112
5.1	Poissonovo rozdělení	112
5.2	Alternativní rozdělení	115
5.3	Binomické rozdělení	116
5.4	Hypergeometrické rozdělení	119
6	Modely rozdělení pravděpodobností pro spojité náhodné veličiny	122
6.1	Rovnoměrné rozdělení	122
6.2	Exponenciální rozdělení	124
6.3	Normální rozdělení	127
6.4	Normované normální rozdělení	130
6.5	Logaritmicko-normální rozdělení	133
6.6	Rozdělení některých funkcí náhodných veličin	136
7	Teoretické základy statistiky	141
7.1	Zákon velkých čísel	141
7.2	Součet nezávislých náhodných veličin	142
7.3	Centrální limitní věty	145
7.4	Věty o normálním rozdělení	152
8	Výběrová šetření	156
8.1	Druhy výběrového šetření	156
8.2	Náhodný výběr a výběrové charakteristiky	159
8.3	Výběrová rozdělení	160
8.4	Populace, výběr a statistické usuzování	163
9	Odhady charakteristik základního souboru	166
9.1	Bodové odhady parametrů	166
9.2	Intervalové odhady parametrů	171
9.3	Intervalové odhady parametrů normálního rozdělení	173
9.4	Intervalový odhad střední hodnoty pro výběry velkého rozsahu	185
9.5	Intervalový odhad parametru alternativního rozdělení	191

10	Testování statistických hypotéz	196
10.1	Pojem hypotézy a podstata testování hypotéz	196
10.2	Jednovýběrové testy hypotéz	203
10.2.1	Testy parametrů normálního rozdělení $N(\mu, \sigma^2)$	203
10.2.2	Test střední hodnoty pro velký výběr	205
10.2.3	Test parametru alternativního rozdělení pro velký výběr	206
10.2.4	Jednovýběrové testy v aplikaci STAT1	207
10.3	Dvouvýběrové testy hypotéz	211
10.3.1	Testy pro nezávislé výběry ze dvou normálních rozdělení	211
10.3.2	Test shody dvou středních hodnot pro velké nezávislé výběry	216
10.3.3	Test shody dvou středních hodnot pro závislé výběry (párový test)	217
10.3.4	Test shody dvou parametrů alternativního rozdělení pro velké nezávislé výběry	219
10.3.5	Dvouvýběrové testy v aplikaci STAT1	220
10.4	Testy hypotéz o rozdělení základního souboru	224
10.4.1	Grafické metody	224
10.4.2	Test nulové šikmosti a nulové špičatosti náhodné veličiny	225
10.4.3	Test normálního rozdělení – C-test	227
10.4.4	χ^2 -test dobré shody	230
10.4.5	Kolmogorovův-Smirnovův test	233
11	Analýza závislostí	237
11.1	Vícerozměrná data	237
11.2	Charakteristiky statistické vazby	240
11.3	Vícerozměrný náhodný vektor	245
11.3.1	Diskrétní náhodný vektor	246
11.3.2	Spojité náhodný vektor	248
11.3.3	Podmíněné rozdělení a nezávislost	250
11.3.4	Charakteristiky náhodného vektoru	251
11.4	Dvourozměrné normální rozdělení	257
11.4.1	Test významnosti korelačního koeficientu	258
11.5	Test nezávislosti v kontingenční tabulce	260
11.6	Regresní analýza	264
11.6.1	Lineární regresní model	266
11.6.2	Odhady a testy v lineárním regresním modelu	267
11.6.3	Regresní přímka	272
11.6.4	Regresní přímka procházející počátkem	275
11.6.5	Regresní logaritmická křivka	276
11.6.6	Regresní hyperbolická křivka	276
11.6.7	Regresní parabolická křivka	277
11.7	Analýza rozptylu	280
	Použité zdroje	291
	Rejstřík	293

O autorech

doc. Mgr. Jiří Neubauer, Ph.D. (*1975)

Vystudoval Přírodovědeckou fakultu Masarykovy univerzity v Brně. Dizertační práci v doktorském studijním oboru aplikovaná matematika obhájil v roce 2006 na Přírodovědecké fakultě Ostravské univerzity v Ostravě. Od roku 2008 absolvoval odborné stáže postupně na Institute of Statistics, Graz University of Technology, Graz, Rakousko, na Department of Statistics, Faculty of Science, University of Malta, Malta, na University of Maribor, Maribor, Slovinsko a na Western Connecticut State University, Danbury, Spojené státy americké. V současné době pracuje na Univerzitě obrany jako zástupce vedoucího Katedry kvantitativních metod. Věnuje se problematice analýzy časových řad se zaměřením na vícerozměrné modely a detekci změn v náhodných procesech. V pedagogické oblasti se věnuje výuce základních statistických metod. Podííl se na řešení výzkumných projektů v rámci své specializace. Publikuje v domácích i zahraničních časopisech.

doc. RNDr. Marek Sedlačík, Ph.D. (*1975)

Vystudoval Přírodovědeckou fakultu Masarykovy univerzity v Brně. Dizertační práci ve studijním oboru Obecné otázky matematiky obhájil roce 2006 na Přírodovědecké fakultě Masarykovy univerzity v Brně, rovněž rigorózní práci v oboru Statistika a analýza dat obhájil na MU v Brně. Od roku 2008 absolvoval odborné stáže postupně na Institute of Statistics, Graz University of Technology, Graz, Rakousko, na Department of Statistics, Faculty of Science, University of Malta, Malta, na National University of Public Service, Budapest, Maďarsko a na Laurea University of Applied Sciences, Helsinky, Finsko. V současné době pracuje na Univerzitě obrany jako vedoucí Katedry kvantitativních metod. Věnuje se problematice mnohorozměrných statistických metod se zaměřením na klasifikační techniky. Garantuje a vede výuku v rámci akreditovaného studia. Podííl se na řešení vědeckých projektů v rámci své specializace. Publikuje v domácích i zahraničních časopisech.

RNDr. Oldřich Kríž (*1945)

Vystudoval Přírodovědeckou fakultu Palackého univerzity v Olomouci. V roce 1993 absolvoval specializované studium statistiky na Fakultě informatiky a statistiky Vysoké školy

ekonomické v Praze a v roce 1997 absolvoval licenční studium Počítačové zpracování dat při kontrole a řízení jakosti na Fakultě chemicko-technologické Pardubické univerzity. Od roku 2004 působil na katedře ekonometrie Fakulty ekonomiky a managementu Univerzity obrany v Brně. Ve výzkumné oblasti řešil úkoly v souvislosti s distančním vzděláváním statistiky a podílel se na řešení projektů v oblasti senzorické analýzy potravin. Publikoval v domácích i zahraničních časopisech. Je autorem a spoluautorem řady didaktických titulů. V současné době spolupracuje s Katedrou kvantitativních metod externě.

Úvodní slovo recenzentů

Zájemcům o statistiku se dostává do rukou už třetí vydání statistické učebnice *Základy statistiky s podtitulem Aplikace v technických a ekonomických oborech*. Po prvním vydání v roce 2012 následovalo v roce 2016 druhé vydání, rozšířené o 11. kapitolu obsahující problematiku analýzy závislostí, zejména využití kontingenčních tabulek a regresní analýzu. Nyní autoři knihu opět upravili a rozšířili v novém, třetím vydání.

Výpočetní technika sňala ze zájemců o statistiku potřebu úmorného počítání a různé statistické programy umožňují snadno a rychle data zpracovat. Na trhu existuje velké množství statistických programů. Bohužel převážně komerčních, které nutí zájemce o statistiku do poměrně velkých počátečních investic. Ze statistických nástrojů, které jsou zdarma, vyniká zejména programovací jazyk R, který je na jedné straně velmi komplexní, ale na druhé straně pro začátečníka obtížně zvládnutelný. V této situaci autoři *Základů statistiky* zvolili jako prostředí pro statistickou analýzu dat tabulkový procesor MS Excel. Ten na jednu stranu trpí problémy, jako je nedodržování standardů, chaos v ovládní, na které jsme u programů této firmy zvyklí, ale ač není zdarma, je velmi rozšířený. I když Excel není primárně určen pro statistické zpracování dat, obsahuje dostatečné množství statistických funkcí použitelných pro aplikaci základních statistických metod. Začátečník tedy může začít pracovat v MS Excel a teprve až zjistí, že už není začátečníkem a investice do komerčního balíku se mu vyplatí, protože narazil na limity Excelu, může začít uvažovat o použití specializovaného statistického software. Pro usnadnění vytvořili autoři Excelovou aplikaci STAT1, ve které jsou základní statistické metody připraveny tak, aby byly snadno použitelné i pro začátečníky. Výhodou knihy je velké množství podrobně řešených příkladů, na kterých si čtenář může ověřit, že vysvětlené metody dobře pochopil a dokáže pomoci nich data správně zpracovat.

Ve třetím vydání byla doplněna 11. kapitola o část věnovanou analýze rozptylu. Současná podoba učebnice tak pokrývá obvyklou problematiku vyučovanou v základních kurzech statistiky na většině univerzit včetně problematiky analýzy závislostí. Tím byly do knihy doplněny kapitoly, které ji rozšířily tak, že nyní pokrývá celý rozsah metod běžně používaných při analýze statistických dat. Mohu tedy knihu doporučit všem zájemcům, kteří potřebují zpracovat a analyzovat své měření a exaktně otestovat, jestli jejich výsledky mohou být dílem náhody nebo za nimi jsou nějaké skryté souvislosti.

prof. RNDr. Petr Ponížil, Ph.D.

Ústav fyziky a materiálového inženýrství, Fakulta technologická, Univerzita Tomáše Bati ve Zlíně

Váženým čtenářům se do ruky dostává kniha věnovaná základům matematické statistiky trojice autorů Neubauer, Sedlačík, Kříž. Záměrem knihy bylo podat výklad statistiky jednoduchým způsobem, který by akcentoval základy statistické práce a usuzování, přitom však respektoval teoretické základy této vědní disciplíny. Jde již o třetí vydání, což potvrzuje úspěšnost takového počínu.

Trpělivý čtenář po přečtení této monografie bude mít vhled do historického vývoje statistiky z první kapitoly, bude rozumět základním pojmům a vztahům v teorii pravděpodobnosti uvedených v kapitole 3, bude schopný rozlišovat mezi empirickými charakteristikami datového souboru a teoretickými charakteristikami náhodných veličin nebo vektorů z kapitol 2, 4 a 11, bude znát základní typy rozdělení pravděpodobnosti z kapitol 5 a 6. Navíc, díky prostudování kapitol 8, 9, 10 a 11 čtenář zvládne odhadovat parametry vybraných rozdělení jak bodově tak intervalově, dále pro něj bude snadné ověřovat hypotézy o parametrech a rozděleních náhodných veličin a vektorů na základě náhodného výběru z jejich rozdělení. Oproti předchozímu vydání jsou v kapitole 11 věnované analýze závislostí zahrnující témata jako je korelační, regresní analýza a kontingenční tabulky nově přidán odstavec o analýze rozptylu (ANOVA).

Úspěšnému nabytí slibovaných dovedností vychází vstřícná struktura celého textu. Potřebná teorie je vyložena postupně, pečlivě a srozumitelně. Namísto důkazů, které pro některé čtenáře nemusí být přínosem, je často teorie aplikována v řadě ukázkových příkladů s řešením. Škála oborů těchto příkladů jednoznačně dokládá šíři užití uvedených statistických metod. Každá kapitola je navíc doplněna souborem příkladů k procvičení s výsledky. Tímto je možné ihned si ověřit získané vědomosti. Čtení knihy je také příjemněeno názornými obrázky a grafy.

Neoddělitelnou součástí statistického zpracování datového souboru je v současnosti práce s vhodným softwarem. O to více si cením pečlivého vysvětlení postupů statistického testování a odhadování v MS Excelu. Používané pojmy v české verzi tohoto běžně dostupného softwaru nejsou v současnosti úplně v souladu s užívanou statistickou terminologií. Navíc lze v učebnici nalézt v některých pasážích i srovnání postupů z více statistických softwarů jako je Statistica, Matlab a MS Excel. Pro výpočty na PC byla autory vytvořena volně stažitelná aplikace STAT1, ve které jsou připraveny ukázky postupů tak, aby přesně odpovídaly uvedeným postupům v knize. Věřím, že kniha Základy statistiky přinese radost svým čtenářům při objevování kouzla statistického usuzování a šíři jeho uplatnění.

doc. Mgr. Zuzana Hübnarová, Ph.D.

Ústav matematiky, Fakulta strojního inženýrství, Vysoké technické učení v Brně

Předmluva

Autoři knihy, kterou právě držíte v ruce, se výuce statistiky na vysoké škole věnují již řadu let. Ve snaze předat svoje zkušenosti studentům, ale i dalším uživatelům statistiky, vytvořili předložený text. Hlavním cílem při jeho přípravě byla snaha podat výklad statistiky jednoduchým způsobem, který by akcentoval základy statistické práce a usuzování, přitom však přiměřeně respektoval exaktní pozadí této vědní disciplíny. Skutečnost, že autoři nabízejí 3. vydání publikace, je snad zprávou o tom, že se tento záměr podařil. Proto je kniha vhodná především pro studium základů statistiky zejména na fakultách s ekonomickým nebo technickým zaměřením.

Statistika je vědní disciplína, která je vybudovaná na třech pilířích: popisné statistice, teorii pravděpodobnosti a teorii náhodné veličiny. Abychom dobře porozuměli smyslu statistiky a jejímu praktickému uplatnění, musíme pochopit podstatu pravděpodobnosti, neboť všechny závěry, ke kterým statistika svými metodami a prostředky dojde, neplatí s exaktní matematickou přesností, ale mají vždy platnost pouze s jistou pravděpodobností – hovoří se o spolehlivosti. Slovo „pouze“ nepředurčuje statistice význam menší než matematice, ale jiný než matematice. Statistika je totiž disciplína velmi praktická a zabývá se všemi takovými reálnými situacemi, ve kterých se potřebujeme opřít o neznámé informace. Ty jsou zjednodušeně řečeno zatím skryté v tzv. teoretických modelech, popisujících tzv. náhodné veličiny. Odkrývání neznámých informací v nejrůznějších reálných situacích nám umožní popisná statistika, která pracuje s naměřenými nebo zjištěnými daty a informace o nich shrne do tzv. empirického modelu.

Na první pohled je zřejmé, že vyjmenované pilíře tu „pravou“ statistiku ještě netvoří. Vztah mezi teoretickým a empirickým modelem přímo souvisí s filozofií statistiky. Taková statistika má totiž induktivní charakter a zabývá se tím, jak odhadnout ty vlastnosti teoretického modelu, které nás zajímají a přitom je neznáme, pomocí modelu empirického. Výklad celé problematiky v této učebnici je proto založen na vybudování základních pojmů a vztahů, srozumitelném popisu základních metod a je protkán řadou řešených příkladů. Teoretické základy statistiky se opírají zejména o vlastnosti normálního rozdělení, centrální limitní větu a závislost mezi jevy.

Potřeba zpracovávat pozorování či měření, shrnout data a postihnout, co říkají, zanedbat nepodstatné detaily a odhalit společné vlastnosti je přítomná v mnoha vědních oborech. Proto nachází metody matematické statistiky praktické uplatnění v široké řadě oblastí, při

řešení nepřehledného množství praktických problémů. Neustále se proto i v textu zdůrazňuje nejdůležitější atribut statistiky, a tím je praktická a reálná podoba řešených problémů.

Problematika základů statistiky je v předložené publikaci prezentovaná v 11 kapitolách. V první kapitole jsou po jednoduchém historickém přehledu zavedené základní pojmy užívané v celém titulu. Obsahem druhé kapitoly je popisná statistika, která tvoří jeden pilíř statistiky jako vědní disciplíny. V této části se čtenář už prvně setkává s počítačovou podporou prostřednictvím aplikace STAT1, která slouží k popisu empirických dat a vytvoření empirického modelu. Třetí kapitola se zabývá pravděpodobnostními základy statistiky, tedy druhým pilířem statistiky jako vědní disciplíny. Třetí pilíř je vybudován a popsán ve čtvrté, páté a šesté kapitole, kde je zaveden pojem náhodné veličiny a popsány jsou zde nejdůležitější teoretické modely náhodné veličiny. V sedmé a osmé kapitole jsou vybudované teoretické základy statistiky a také je zde popsán statistické usuzování s využitím praktického propojení všech třech pilířů. Velmi praktickou a sěžejní část titulu tvoří devátá a desátá kapitola, kde je vybudován aparát pro bodové a intervalové odhady parametrů nejčastěji používaných teoretických modelů, resp. pro testování nejčastějších statistických hypotéz. Ve druhém vydání přibyla v jedenácté kapitole rovněž velmi praktická část o analýze závislostí, zejména o měření závislostí v kontingenční tabulce, resp. o modelování a užití regresních modelů. Ve vydání, které máte v ruce, je nově přidána analýza rozptylu (ANOVA), což je další nástroj pro zkoumání vztahu mezi vysvětlovanými a vysvětlujícími proměnnými.

Součástí dnešního moderního světa je užití výpočetní techniky ve všech sférách života. Statistika je, právě s ohledem na práci s datovými množinami, přímo předurčena k využití počítačů. Na tuto skutečnost reaguje i přístup pedagogů k výuce statistiky na mnoha fakultách. Výklad a řešení praktických úloh je v této učebnici přímo podporován elektronickou aplikací STAT1, která pracuje v excelovském prostředí a umožňuje každému studentovi interaktivně vnímat popsání statistické metody. U vybraných příkladů je použití této aplikace podrobně popsáno. Aplikaci je možné použít i na analýzy vlastních dat. Aplikaci STAT1 spolu se statistickými tabulkami najdete na adrese <http://k101.unob.cz/stat1/>.

Rádi bychom poděkovali kolegovi doc. RNDr. Jaroslavu Michálkovi, CSc. z Univerzity obrany v Brně, který se ujal úlohy lektora, knihu pozorně přečetl a doporučil několik úprav. Dík patří též recenzentům prof. RNDr. Petru Ponížilovi, Ph.D. z Fakulty technologické Univerzity Tomáše Bati ve Zlíně a doc. Mgr. Zuzaně Hübnerové, Ph.D. z Fakulty strojního inženýrství Vysokého technického učení v Brně za jejich podnětné nápady a připomínky.

V Brně 20. 6. 2021

Autoři

Kapitola 1

Úvod do statistiky

První kapitola této knihy věnujeme úvodnímu seznámení se statistikou. Představíme si statistiku jako vědní disciplínu, která se vyvinula z původních starověkých sčítání obyvatel a majetku až k současnosti. Dozvíme se, co si vlastně máme představit pod pojmem *statistika* a jakou roli hraje statistika v moderní společnosti. Začneme si budovat odborný slovník a zavedeme si základní pojmy, abychom se v odborném prostředí domluvili a také rozuměli tomu, co se kde ve „statistickém jazyku“ píše či mluví.

1.1 Historický přehled

Slovo statistika pochází z italského *stato*, původně s významem *stav*, od konce středověku také *státní území*, resp. *stát*. Jako první jej patrně použil Girolamo Ghilini (1589–1669) v práci *Ristretto della civile, politica, statistica e militare scienza* (*Shrnutí civilní, politické, statistické a vojenské vědy*), ve které shromáždil různé znalosti té doby o státu, o jeho obyvatelích, životě, právu, obchodu i výrobě, náboženství i armádě. Především v tomto smyslu se potom slovo *stato* rozšířilo i do jiných jazyků, např. ve tvaru *state*, *Staat*, *état*, *estato*.

Podněty pro vznik statistiky

První historické zmínky o činnostech, které z dnešního pohledu připomínají statistiku, pocházejí už ze starověku. Záznamy o *sčítání obyvatel a majetku* můžeme najít už v písemnostech starých Babyloňanů z období před rokem 3800 př. n. l. Historicky nejstarším směrem ovlivňujícím také vznik statistiky byla existence prvních městských států v 3. a 2. tisíciletí př. n. l. ve starověkých civilizacích, jakými byly Egypt, Čína, Mezopotámie, Palestina, Řecko nebo Řím. Se vznikem městských států vzniká také potřeba jejich správy, se kterou jsou spojené nemalé náklady, proto se zvyšuje výběr daní. K určení jejich výše je ale nezbytné mít číselné údaje o území, obyvatelstvu, zemědělství, obchodu, řemeslech apod. Tyto informace se získávají zejména na základě soupisu obyvatelstva a dalších šetření, která mají z dnešního pohledu charakter statistických šetření.

Jednu z prvních zmínek o statistickém šetření nalezneme také v Bibli, kde je ve Starém zákoně ve 4. knize Mojžíšově informace o sčítání provedeném Mojžíšem po odchodu izraelského národa z Egypta a obsahuje konkrétní počty bojovníků, oddílů a velitelů. Později také nařídil sčítání lidu motivované vojensky král David.

Velké sčítání lidu zavedli také ve starověkém Římě v 5. století př. n. l. Sčítání měli na starosti vysocí úředníci, nazývaní *cenzoři*. Sčítání (cenzy) se konala každých pět let a zjišťovaly se nejen počty obyvatel a jejich majetek, ale také např. počet otroků.

Podobné průzkumy se postupně rozšiřovaly i na další evropské země až do období středověku. Od 16. století byly zřizovány církevní matriky, které se na dlouhou dobu staly základním zdrojem informací o obyvatelstvu.

Tři kořeny statistiky

Vlastní termín *statistika* se začal používat až v 18. století v Německu pro označení *nauky o státu*. Tato vědecká disciplína se začala rozvíjet v 16. století na univerzitách v Itálii a později také právě v Německu, proto se jí říká *univerzitní statistika*. Tehdejší statistické studie obsahovaly především údaje o evropských státech – geografické, politické, ekonomické a další. Na rozdíl od dnešní statistiky neobsahovaly mnoho čísel, většina zaznamenaných údajů měla charakter slovní.

Jedno z prvních státovědných děl *O vládě a správě v různých královstvích a republikách* vyšlo v roce 1562 v Benátkách a napsal je Francesco Sansovina. Přesně o sto let později uveřejnil Ludwig von Seckendorff svou státovědnou knihu *Německý knížecí stát*. Na jejich práce navazuje nejvýznamnější teoretik statistiky v německé jazykové oblasti Gottfried Achenwall (1719–1772). Byl profesorem statistiky na univerzitě v Göttingenu a autorem populární učebnice statistiky, která byla předepsána pro přednášky statistiky i na Karlově univerzitě v Praze.

V Anglii mezitím vznikl zcela jiný okruh statistiky, a to takzvaná *politická aritmetika*, která vycházela z údajů o narozeních a úmrtích, a na tomto základě se pokoušela pozorovat a srovnávat informace o obyvatelstvu za delší časové úseky. Tyto průzkumy vycházely z údajů tehdejších církevních matrik a na jejich základě se snažily odvodit některé obecně platné zákonitosti (např. že se rodí obecně více chlapců než děvčat).

Její nejvýznamnější představitelé jsou William Petty (1623–1687) a John Graunt (1620 až 1674). Petty je považován za předchůdce moderní statistiky i klasické politické ekonomie. Jeho nejvýznamnější dílo *Pět esejí o politické aritmetice* bylo vydáno posmrtně (1960). Graunt byl obchodník a zabýval se především demografií. Napsal první ucelenou demografickou studii s poněkud pochmurným názvem *Přírozená a politická pozorování založená na seznamech zemřelých* (1662).

V 18. století se toto zaměření statistiky začalo prosazovat i v Německu a obě statistické školy se začaly vzájemně ovlivňovat a postupně sblížovat. Statistika začala ve větší míře používat čísla a přestala se zabývat pouze popisem státních pozoruhodností. Postupně začala pronikat i do jiných vědeckých disciplín, aby se nakonec prosadila jako samostatná věda.

Nezávisle na statistice se od 17. století začala rozvíjet ještě jiná teoretická disciplína, která vznikla jako součást matematiky – *teorie pravděpodobnosti*. Zatímco statistika zkoumá hromadné jevy, teorie pravděpodobnosti se naopak zabývá jevy individuálními, jedinečnými. Pravděpodobnost je chápána jako číselné ohodnocení šance – naděje, že sledovaný konkrétní jev nastane. Ve skutečnosti však statistika a teorie pravděpodobnosti představují dva pohledy na stejný problém. Každý hromadný jev je totiž tvořen jednotlivými jevy individuálními, a naopak opakováním individuálního jevu získáme jev hromadný. V současné době nelze

teorii pravděpodobnosti a statistiku od sebe oddělit – teorie pravděpodobnosti je považována za teoretický základ statistiky.

Rozvoj teorie pravděpodobnosti byl zpočátku inspirován hlavně hazardními hrami. Za její počátek se považuje slavná výměna dopisů mezi matematiky Blaisem Pascalem (1623–1662) a Pierrem de Fermatem (1601–1665) zahájená roku 1654. Šlo jim tehdy o otázku, jak spravedlivě rozdělit bank mezi hráče, jestliže série hazardních her musela být předčasně přerušena. Tehdy rozvíjené teorii pravděpodobnosti dnes říkáme *klasická pravděpodobnost*. Mezi další osobnosti, které se věnovaly teorii pravděpodobnosti, patří švýcarští matematici (bratři) Jacob Bernoulli (1656–1705) a Johann Bernoulli (1667–1748), francouzští matematici Abraham de Moivre (1667–1754), Pierre Simon Laplace (1749–1827) a také Siméon Denis Poisson (1781–1840), se kterým se setkáme v 5. kapitole – viz *Poissonovo rozdělení pravděpodobnosti*, které je vhodné pro popis jevů s nízkou pravděpodobností jevu při značném rozsahu výběrového souboru. Významný příspěvek k teorii chyb předložil také vynikající německý matematik Carl Friedrich Gauss (1777–1855), který přispěl k formulování tzv. *normálního rozdělení pravděpodobnosti* – viz 6. kapitola.

Statistika jako nová věda

Postupným splýváním nauky o státu, politické aritmetiky a teorie pravděpodobnosti vznikla v 18. a 19. století statistika jako samostatná vědní disciplína, která popisovala hromadné jevy v nově vznikajících vědách – přírodních, technických i ekonomických. Statistika tohoto období se zabývala především popisem zkoumaných hromadných jevů, proto se také nazývá *popisná – deskriptivní statistika*. Metodou statistických průzkumů byla vyčerpávající šetření prováděná podle zásady: čím více údajů získáme, tím přesnější budou závěry. Toto pravidlo ve statistice převládalo až do konce 19. století.

Významnou osobností nové statistiky byl belgický matematik Adolphe Jacques Quételet (1796–1874), který je zakladatelem prvního národního statistického úřadu (1841) v Evropě. Mimo jiné se věnoval rozsáhlému sběru dat o lidské populaci a prezentoval svůj pojem „průměrného člověka“ jako centrální hodnoty, kolem které se měřené tělesné míry shlukují podle Gaussovy křivky – viz 6. kapitola. V té souvislosti zavedl také pojem index tělesné hmotnosti používaný dodnes pro stanovení míry obezity a známý pod zkratkou BMI (body mass index). Naznačil tak budoucí směřování statistiky k normálnímu rozdělení, střední hodnotě a rozptylu. Pomohl rovněž zavést statistické techniky do kriminalistiky, pomocí statistické analýzy porozuměl Quételet vztahu mezi zločinem a ostatními sociologickými faktory.

Na přelomu 19. a 20. století však dochází ve vývoji statistiky k zásadní změně. Začala éra *matematické – induktivní statistiky*, která na základě teorie pravděpodobnosti umožňuje získat kvalifikované závěry – odhady o sledovaném jevu i z malého dostupného vzorku údajů. Nové statistické postupy otevřely možnosti pro nejrůznější typy průzkumů, ve kterých se z vlastností části usuzuje na chování celku. Na bázi induktivní statistiky vznikly také extrapoláčnické – prognostické metody, které na základě znalosti dat z minulosti umožní vytvořit kvalifikovaný odhad chování v budoucnosti.

Těžiště rozvoje induktivní statistiky se do značné míry přesunulo do anglo-americké oblasti a je spojeno především se jménem anglického statistika sira Ronalda Aylmera Fishera (1890–1962), který stál u vzniku mnoha dnes obvyklých metod statistické analýzy. Je pova-

žován za zakladatele teorie plánování experimentů v biologickém a zemědělském výzkumu. Významných výsledků dosáhl i další anglický statistik William Sealy Gosset (1876–1937), který pracoval jako chemik v irském pivovaru Guinness a tam vymyslel postup, který umožnil provádět z malých výběrů použitelné závěry, přinejmenším však poznat, jak posuzovat vypovídací hodnotu takových výběrů. Gosset se pod svá průkopnická díla podepisoval pseudonymem „Student“, protože jeho firma mu publikování výsledků pod vlastním jménem zakázala.

Další významní představitelé anglické statistické školy byli Francis Galton (1822–1911) a Charles Pearson (1857–1936), kteří položili základy zkoumání závislostí mezi hromadnými jevy. K rozvoji matematické statistiky přispěli také ruští matematici: Pafnutij Lvovič Čebyšev (1821–1894), Andrej Andrejevič Markov (1856–1922) a Andrej Nikolajevič Kolmogorov (1903–1987), který je považován za zakladatele moderní teorie pravděpodobnosti.

U nás dosáhly pozoruhodných výsledků dvě osobnosti. Prvním byl profesor Jaroslav Janko (1893–1965). Svou celoživotní činností velmi významně přispěl k rozvoji matematicko-statistických metod, k jejich nanejvýš užitečnému uplatnění ve výzkumu a praxi, a zapsal se tak do historie matematické statistiky u nás. Známá jsou jeho díla *Jak vytváří statistika obrazy světa a života*, *Základy statistické indukce a Statistické tabulky*. Druhým byl profesor Jaroslav Hájek (1926–1974), kterého lze považovat za nejvýznamnějšího českého statistika v historii české matematiky. Jeho odborné aktivity byly zaměřené na neparаметrické statistické metody.

Současná statistika

Statistika dnes představuje vědní disciplínu se širokým praktickým uplatněním. Používá se zejména jako důležitý nástroj získávání informací ve veřejných sférách našeho života, ale i jako důležitý nástroj řešení nejrůznějších odborných problémů, zejména technických, přírodovědných, ekonomických, vojenských, sociálních. Je tomu tak proto, že moderní statistika využívá všech postupů a metod, které během svého dlouhého vývoje vytvořila nebo si osvojila. Používá jak prvky klasické popisné statistiky, založené na analýze hromadných dat, tak i prvky moderní matematické statistiky, postavené na teorii pravděpodobnosti. Proto statistiku vnímáme nejen jako nástroj poznání (velký nepřehledný soubor dat dokáže nahradit několika výstižnými charakteristikami), ale také jako nástroj rozhodování v neurčitosti (na základě vlastnosti vzorku usuzuje na vlastnosti celého souboru, popř. z informací o minulosti předvídá vývoj v budoucnosti).

Velký význam pro rozvoj a využití statistických metod měl nástup výpočetních technologií, zejména osobních počítačů. Počítač vítězí nad člověkem především v těch úkonech, které jsou pro člověka tradičně nejdélnější – třídění, vyhledávání a výpočty s velkým množstvím dat. Počítačům jsou vlastní také možnosti tabulkového zpracování a grafického vyjadřování. Mezi nejznámější profesionální statistické programy se širokým portfoliem metod a technik patří Statistica, SPSS, SAS, Statgraphics, Minitab a další, z českých produktů QCExpert. Pro potřebu výuky statistiky využívá řada škol i tabulkový kalkulátor MS Excel, který patří k základní výbavě osobního počítače. Naše učebnice bude podporovaná jednoduchou aplikací STAT1, vytvořenou právě v excelovském prostředí.

Statistika byla zpočátku využívána spíše ve vědách přírodních (fyzika, chemie) a technických, v posledních letech však zaznamenává úspěch také v disciplínách humanitního charakteru, například v psychologii, sociologii, pedagogice, ale také v ekonomii, která původně