

Jiří Neubauer, Marek Sedlačík, Oldřich Kříž

ZÁKLADY STATISTIKY

Aplikace v technických
a ekonomických oborech
4., 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
4., 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. Automatizovaná analýza textů nebo dat ve smyslu čl. 4 směrnice 2019/790/EU a použití této knihy k trénování AI jsou bez souhlasu nositele práv zakázány.

Jiří Neubauer, Marek Sedlačík, Oldřich Kříž

Základy statistiky

Aplikace v technických a ekonomických oborech

4., 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 10261. publikaci

Lektoroval: doc. RNDr. Jaroslav Michálek, CSc.
Recenzovali: doc. Ing. Luboš Střelec, Ph.D., Ing. Jan Holešovský, Ph.D.
Odpovědná redaktorka: Věra Slavíková
Počet stran 312
Čtvrté vydání, Praha 2025
Vytiskla Tiskárna v Ráji, s.r.o., Pardubice

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

© Grada Publishing, a.s., 2025
Cover Design © Grada Publishing, a. s., 2025
Cover Photo © Golden Dayz/Shutterstock

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

ISBN 978-80-271-8221-3 (pdf)
ISBN 978-80-271-5581-1 (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	Vyjádřovací 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	50
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	125
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	Konstrukce bodových odhadů	171
9.3	Intervalové odhady parametrů	176
9.4	Intervalové odhady parametrů normálního rozdělení	178
9.5	Intervalový odhad střední hodnoty pro výběry velkého rozsahu	190
9.6	Intervalový odhad parametru alternativního rozdělení	195
10	Testování statistických hypotéz	200
10.1	Pojem hypotézy a podstata testování hypotéz	200
10.2	Jednovýběrové testy hypotéz	207
10.2.1	Testy parametrů normálního rozdělení $N(\mu, \sigma^2)$	207
10.2.2	Test střední hodnoty pro velký výběr	209
10.2.3	Test parametru alternativního rozdělení pro velký výběr	210

10.2.4	Jednovýběrové testy v aplikaci STAT1	211
10.3	Dvouvýběrové testy hypotéz	215
10.3.1	Testy pro nezávislé výběry ze dvou normálních rozdělení	215
10.3.2	Test shody dvou středních hodnot pro velké nezávislé výběry	220
10.3.3	Test shody dvou středních hodnot pro závislé výběry (párový test)	221
10.3.4	Test shody dvou parametrů alternativního rozdělení pro velké nezávislé výběry	223
10.3.5	Dvouvýběrové testy v aplikaci STAT1	224
10.4	Testy hypotéz o rozdělení základního souboru	228
10.4.1	Grafické metody	228
10.4.2	Test nulové šikmosti a nulové špičatosti náhodné veličiny	229
10.4.3	Test normálního rozdělení – C-test	231
10.4.4	χ^2 -test dobré shody	234
10.4.5	Kolmogorovův-Smirnovův test	237
11	Analýza závislostí	241
11.1	Vícerozměrná data	241
11.2	Charakteristiky statistické vazby	244
11.3	Vícerozměrná náhodná veličina – náhodný vektor	249
11.3.1	Diskrétní náhodný vektor	250
11.3.2	Spojité náhodný vektor	252
11.3.3	Podmíněné rozdělení a nezávislost	254
11.3.4	Charakteristiky náhodného vektoru	255
11.4	Dvourozměrné normální rozdělení	261
11.4.1	Test významnosti korelačního koeficientu	262
11.5	Test nezávislosti v kontingenční tabulce	264
11.6	Regresní analýza	268
11.6.1	Lineární regresní model	270
11.6.2	Odhady a testy v lineárním regresním modelu	271
11.6.3	Regresní přímka	276
11.6.4	Regresní přímka procházející počátkem	279
11.6.5	Regresní logaritmická křivka	280
11.6.6	Regresní hyperbolická křivka	280
11.6.7	Regresní parabolická křivka	281
11.7	Analýza rozptylu	285
11.7.1	Jednofaktorová analýza rozptylu	285
11.7.2	Metody vícenásobného porovnávání	292
	Použité zdroje	303
	Rejstřík	305

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 prorektor pro vzdělávání a záležitosti studentů. 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 Kříž (*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

Čtenářům je předkládáno rozšířené vydání odborné knihy *Základy statistiky* s podtitulem *Aplikace v technických a ekonomických oborech*, která čtenářům osvětluje základy statistiky s primárním zaměřením na matematickou (induktivní) statistiku s aplikacemi v technických a ekonomických oborech.

Využití statistiky lze nalézt v mnoha vědních oborech, ale lze se s ní setkat i v běžném životě, často s využitím statistických programů, ať již komerčních, tak i volně dostupných. Z tohoto pohledu lze předložený text *Základy statistiky* považovat za aktuální a pro čtenáře za velmi přínosný, neboť se jednoduchým způsobem zaměřuje na výklad statistiky za současného respektování teoretických základů tohoto vědního oboru.

Z obsahového hlediska lze předkládané vydání považovat za komplexní monografii zaměřenou na statistiku a metody statistické analýzy dat. Autoři nejprve čtenáři detailně popisují historický vývoj statistiky, ale především představují základní pojmy a vztahy teorie pravděpodobnosti nezbytně nutné pro pochopení následných částí textu. Ty se pak zaměřují na problematiku výběrových šetření, statistické indukce, analýzu závislostí a analýzy rozptylu, přičemž aktuální vydání bylo rozšířeno o konstrukce bodových odhadů a v rámci analýzy rozptylu i o metody vícenásobného porovnávání. Předkládané rozšířené vydání tak pokrývá prakticky celou problematiku běžně využívaných statistických metod, které jsou mimo jiné standardně vyučovány na ekonomicky a technicky zaměřených univerzitách v základních kurzech statistiky. Pozitivně pak hodnotím zařazení části analýzy závislostí, které jsou především na ekonomicky zaměřených univerzitách vyučovány v navazujících předmětech typu Ekonometrie.

Stejně pozitivně hodnotím fakt, že autoři výklad teorie doplnili četnými obrázky a grafy vztahujícími se k prezentované teorii a také velkým množstvím praktických příkladů. Pro tyto účely bylo využito rovněž výpočetní techniky, která je v současnosti neodmyslitelnou součástí nejen statistických analýz, ale lze se s ní setkat prakticky ve všech sférách běžného života. Praktické příklady v tomto odborném textu jsou pak řešeny v prostředí volně dostupné aplikace STAT1, která pracuje v rámci prostředí tabulkového procesoru MS Excel, čímž je umožněna široká dostupnost a využitelnost nejen při výuce, ale rovněž i v praxi. Čtenáři tak mají možnost nejen samostatně řešit jednotlivé prezentované příklady, ale rovněž mohou aplikaci použít i pro analýzy vlastních dat. Po prostudování knihy *Základy statistiky* tak získají potřebné znalosti z oblasti statistiky a zároveň dovednosti k samostatné práci s využitím aplikace STAT1.

Předkládanou knihu *Základy statistiky* tak mohou doporučit širokému spektru čtenářů, tedy nejen zájemcům o pochopení základů statistiky, ale rovněž uživatelům zpracovávajícím vlastní datové soubory, kdy jim kniha mimo jiné usnadní pochopení a následnou korektní interpretaci získaných výstupů.

doc. Ing. Luboš Střelec, Ph.D.

Ústav statistiky a operačního výzkumu, Provozně ekonomická fakulta, Mendelova univerzita v Brně

Váženým čtenářům je předkládáno již čtvrté, opět rozšířené vydání knihy *Základy statistiky* autorů Neubauera, Sedlačka a Kříže. Od svého prvního vydání v roce 2012 byla kniha díky zájmu čtenářů postupně upravována až do dnešní podoby, což svědčí o její značné oblibě a nesporné kvalitě. Dnes se jedná již o klasickou učebnici statistiky, která je určena zájemcům nejen z řad studentů či praktiků technických a ekonomických oborů, na něž poukazuje podtitul knihy.

Výklad je záměrně podáván názorným a přístupným způsobem tak, aby byly akcentovány zejména principy statistických postupů, což jistě ocení posluchači nematematických oborů. Na druhou stranu se však autoři nedopouštějí žádného přílišného zjednodušení na úkor korektnosti či možnosti upotřebení předkládaných výsledků. Z toho důvodu mohou knihu doporučit i všem čtenářům s hlubším matematickým vzděláním, kterým kniha může sloužit jako ucelený zdroj elementárních pojmů, vztahů a metod pro běžné statistické vyhodnocení dat. Kniha si klade za cíl představit celou teorii názorně, k čemuž přispívá i velká řada ukázkových příkladů s řešením a příjemné grafické zpracování včetně množství přehledných grafů a obrázků.

Monografie zahrnuje témata, která svou strukturou a rozsahem odpovídají obvyklým základním kurzům pravděpodobnosti a statistiky vedeným na většině univerzit v ČR i v zahraničí. V aktuálním vydání došlo k rozšíření 9. kapitoly, zaměřující se na odhady parametrů, o klasické postupy pro určení bodových odhadů, a sice momentovou metodu a metodu maximální věrohodnosti. Doplněna byla i 11. kapitola pokrývající problematiku analýzy závislosti, zejména využití kontingenčních tabulek či korelační a regresní analýzu, kam byly nově zařazeny i metody mnohonásobného porovnávání jako součást post-hoc vyhodnocení v analýze rozptylu.

Statistické zpracování dat je dnes neoddělitelně spojeno s využitím vhodného softwaru, který značně usnadňuje výpočetní části řešených úloh. Ačkoli je na trhu dostupná řada specializovaných statistických programů, mezi nejčastěji používané stále patří MS Excel. Vzhledem k jeho velkému rozšíření se často jedná o první volbu. Je ovšem znáto, že tento software není pro statistické zpracování primárně určen. Používané pojmy jsou zde matoucí a zavádějící, řada běžných metod není implementována. Pro usnadnění vytvořili autoři knihy volně dostupnou Excelovou aplikaci STAT1, ve které jsou připraveny statistické postupy uvedené v knize, případně v jednotlivých částech textu detailně rozebírají výstupy funkcí již v Excelu zavedených. Vše je tak perfektně nachystáno k zahájení vlastní analýzy. Přeji všem čtenářům příjemný zážitek při četbě a dobrou zábavu při objevování obdivuhodných zákonitostí náhody.

Ing. Jan Holešovský, Ph.D.

Ústav matematiky a deskriptivní geometrie, Fakulta stavební, Vysoké technické učení v Brně

Předmluva

Autoři této knihy se výuce statistiky na vysokých školách věnují již řadu let. Na základě svých zkušeností připravili tento text s cílem zpřístupnit statistiku nejen studentům, ale i širší veřejnosti, která s ní přichází do styku. Při tvorbě kladli důraz na srozumitelný výklad základních principů statistické práce a usuzování, přičemž nezanedbali ani teoretické základy oboru. Skutečnost, že kniha vychází již ve 4. vydání, může naznačovat, že se jim tento záměr podařilo naplnit. Publikace je určena především k výuce 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 řešení 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 popsáno 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 třetím vydání byla přidána analýza rozptylu (ANOVA), což je další nástroj pro zkoumání vztahu mezi vysvětlovanými a vysvětlujícími proměnnými. Vydání, které držíte v ruce, bylo rozšířeno o konstrukce bodových odhadů a metody vícenásobného porovnávání.

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 popsáno 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/>.

Autoři si dovoluují vyjádřit poděkování doc. RNDr. Jaroslavu Michálkovi, CSc., který se ujal role lektora, odborně publikaci posoudil a přispěl řadou cenných doporučení a úprav. Poděkování náleží rovněž recenzentům – doc. Ing. Luboši Střelcovi, Ph.D. z Provozně ekonomické fakulty Mendelovy univerzity v Brně a Ing. Janu Holešovskému, Ph.D. z Fakulty stavební Vysokého učení technického v Brně – za jejich odborné připomínky a podnětné návrhy, které významně přispěly ke zkvalitnění textu.

Úvod do statistiky

První kapitolu 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 *Přirozená 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. kapitolu.

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. kapitolu. 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 extrapolací – 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 neparametrické 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 vlastností 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, STATA a další. Velkou oblibu má v současné době statistický program R (volně dostupný jazyk a prostředí pro statistické výpočty a grafiku). 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 cha-

rakteru, například v psychologii, sociologii, pedagogice, ale také v ekonomii, která původně vznikla jako věda sociální, během času se svými metodami přiblížila spíše vědám přírodním. K výraznějšímu rozvoji statistických metod došlo na přelomu 19. a 20. století, a to zejména díky novým objevům ve statistice (zejména nástupu metod matematické statistiky). To vedlo k dalšímu přibližování statistiky reálnému životu a prudkému rozvoji aplikací statistiky v nej-různějších oborech lidské činnosti. Vznikaly tak postupně speciální statistické metody, které tvořily základ speciálních vědních disciplín. Pod názvem *biostatistika*, resp. *biometrika* se např. rozumí aplikace statistiky na biologické problémy, zatímco pro analýzu chemických dat se spíše užívá termín *chemometrie*. Hlavním cílem aplikací statistických metod v *biomedicín-ském výzkumu* je zajistit správnost a odbornost statistického vyhodnocování dat a interpretace získaných výsledků. Používání počítačů k těmto účelům je v dnešní době samozřejmé.

Aplikací statistických metod na ekonomická a sociálně-ekonomická data vznikla samostatná statistická disciplína, *ekonomická statistika*. Předmětem ekonomické statistiky je analýza stavu a vývoje jevů v hospodářské oblasti jako východiska k hospodářskému rozhodování či stanovení hospodářské politiky. Na využití statistických metod je založený průzkum trhu, plánování výroby, prognostika, kontrola kvality výroby, personální politika, výroční zprávy (určené akcionářům). Ještě k vyšší kvalitě ekonomické analýzy vede disciplína označovaná jako *ekonometrie*. Ta představuje syntézu ekonomické teorie, informatiky, matematiky a statistiky. Tato syntéza není však mechanickým spojením ekonomické analýzy s aparátem matematiky a statistiky, resp. elektronickými prostředky, ale jde o propojení vzájemně se podmiňujících vědních disciplín.

Statistika v Českých zemích

Statistika je s historií našeho území spjata již od nepaměti. Důvody jsou zcela praktické a zřejmé. Každý vládce chtěl mít přehled, jaký má majetek, kolik má k dispozici mužů do vojska či od kolika poddaných může vymáhat daně. Ale důvody pro statistické zjišťování byly mnohdy i zcela jiného, humánnějšího rázu. Například za vlády císaře Rudolfa II. v roce 1583 vypukla v českých zemích epidemie moru. V jejím důsledku bylo zahájeno šetření o „zdraví populace“, které mělo zmapovat vznik a rozvoj zhoubných epidemií a umožnit přijímání včasných protiopatření.

Jako významný mezník lze označit datum 13. října 1753, kdy byl vydán *patent císařovny Marie Terezie* o každoročním sčítání lidu. Zdokonalení evidence obyvatel souviselo s rozsáhlou reformní činností Marie Terezie (1717–1780), neboť k provedení čtených reform bylo nutné získat objektivní informace o obyvatelstvu. Za vlády Marie Terezie došlo také k reformě evidence narozených a zemřelých. V této souvislosti byla zavedena i první jednoduchá statistická klasifikace příčin úmrtí.

Jak už víme, první statistický úřad v Evropě byl založen v roce 1841 v Belgii. Řada evropských zemí Queteletův úřad následovala. V roce 1897 byl zřízen *Zemský statistický úřad Království českého*, který se stal prvním skutečně statistickým úřadem na území dnešní České republiky. Poprvé byla soustředěna na jednom místě všechna statistická pracoviště, která až do té doby působila v rámci různých ministerstev a dalších institucí.

Brzy po vzniku samostatného Československa, už v roce 1919, byl založen *Státní úřad statistický* (SÚS) jako nový orgán pověřený celostátními statistickými šetřeními, mezi něž patřilo jako jedno z nejdůležitějších i sčítání lidu. Úřad se v období mezi světovými válkami

rozvíjel, zdokonaloval a rozšiřoval svoji činnost. K tomu přispělo i úzké sepětí se statistickou teorií. Ve 20. a 30. letech 20. století byla téměř polovina kapacity statistického úřadu věnována vědecké a teoretické činnosti.

V období 2. světové války se činnost statistiky v Čechách a na Moravě omezila a odpovídala válečným podmínkám i postavení našeho území. Perzekuována byla řada pracovníků SÚS, někteří z nich byli popraveni (např. předseda úřadu Dr. Jan Auerhan byl 6. 6. 1942 zatčen gestapem a 9. 6. 1942 zastřelen), jiní zemřeli v nacistických věznicích a koncentračních táborech. Bezprostředně po skončení 2. světové války byla činnost Státního úřadu statistického obnovena, s cílem vrátit jej na předválečnou úroveň. Po roce 1948 se československá statistika (zejména v ekonomické oblasti) zaměřovala zejména na úkoly národohospodářské evidence a kontrolu plnění plánu.

Po pádu komunistického režimu v roce 1989 se obnovily předpoklady pro budování objektivní, nestranné a nestranické státní statistické služby. K 1. 1. 1993, se vznikem ČR, převzal *Český statistický úřad* (ČSÚ) všechny kompetence národního statistického úřadu. Jeho úkoly a postavení, stejně jako zásady a úkoly fungování státní statistické služby v ČR, upravil zákon č. 89/1995 Sb., o státní statistické službě, který byl ještě novelizován k 1. 1. 2001. Jeho hlavním úkolem je shromažďovat a zveřejňovat statistické informace o sociálním a ekonomickém rozvoji České republiky a obstarávat statistické informace pro potřeby dalších orgánů státní správy a územní samosprávy. Vedle centrálního pracoviště ČSÚ v Praze existují krajské reprezentace ve všech 14 krajských městech. Prvním předsedou ČSÚ byl čechokanaďan Edvard Outrata (*1936). Mimo oficiální soustavu státní statistiky stojí řada specializovaných komerčních agentur, které se především zabývají statistickými průzkumy (např. marketingovými) pro podnikatelské subjekty, ale jsou také pověřovány úkoly pro státní statistiku.

V současnosti existují orgány statistické služby prakticky ve všech zemích Evropy. Jejich konkrétní podoba a struktura se však může stát od státu lišit, i když v poslední době dochází ke koordinaci státních statistik v rámci všech členských i přidružených zemí EU. Centrálním statistickým orgánem Evropské unie je EUROSTAT, který má sídlo v Lucemburku. Shromažďuje statistické informace o členských zemích Evropské unie, ale také o dalších evropských zemích.

Příklady k procvičení

1. Zjistěte na stránkách ČSÚ (www.czso.cz), jaký je v ČR aktuální počet obyvatel.
2. Jaká instituce zabezpečuje v ČR sčítání lidu?
3. Kdy byl založen Zemský statistický úřad Království českého?
4. Je možné souhlasit s následujícími výroky?
 - a) Začátky statistiky spadají do 18. století.
 - b) Za prvopočátky statistiky lze považovat záznaky o sčítání lidu a majetku ve starověku.
 - c) Pravděpodobnost dnes představuje neoddělitelnou součást statistiky.
 - d) Označení deskriptivní a induktivní statistika představuje z praktického pohledu totéž.
 - e) Stav a vývoj v ekonomické oblasti sleduje disciplína označovaná jako ekonometrie.
 - f) Vrcholný statistický úřad EU je Eurostat.
5. Vyjmenujte některé historické kořeny statistiky.

Řešení.

2. Český statistický úřad;
3. 1897;
4. a) ne; b) ano; c) ano; d) ne; e) ne; f) ano;
5. německá státověda, anglická politická aritmetika a teorie pravděpodobnosti.