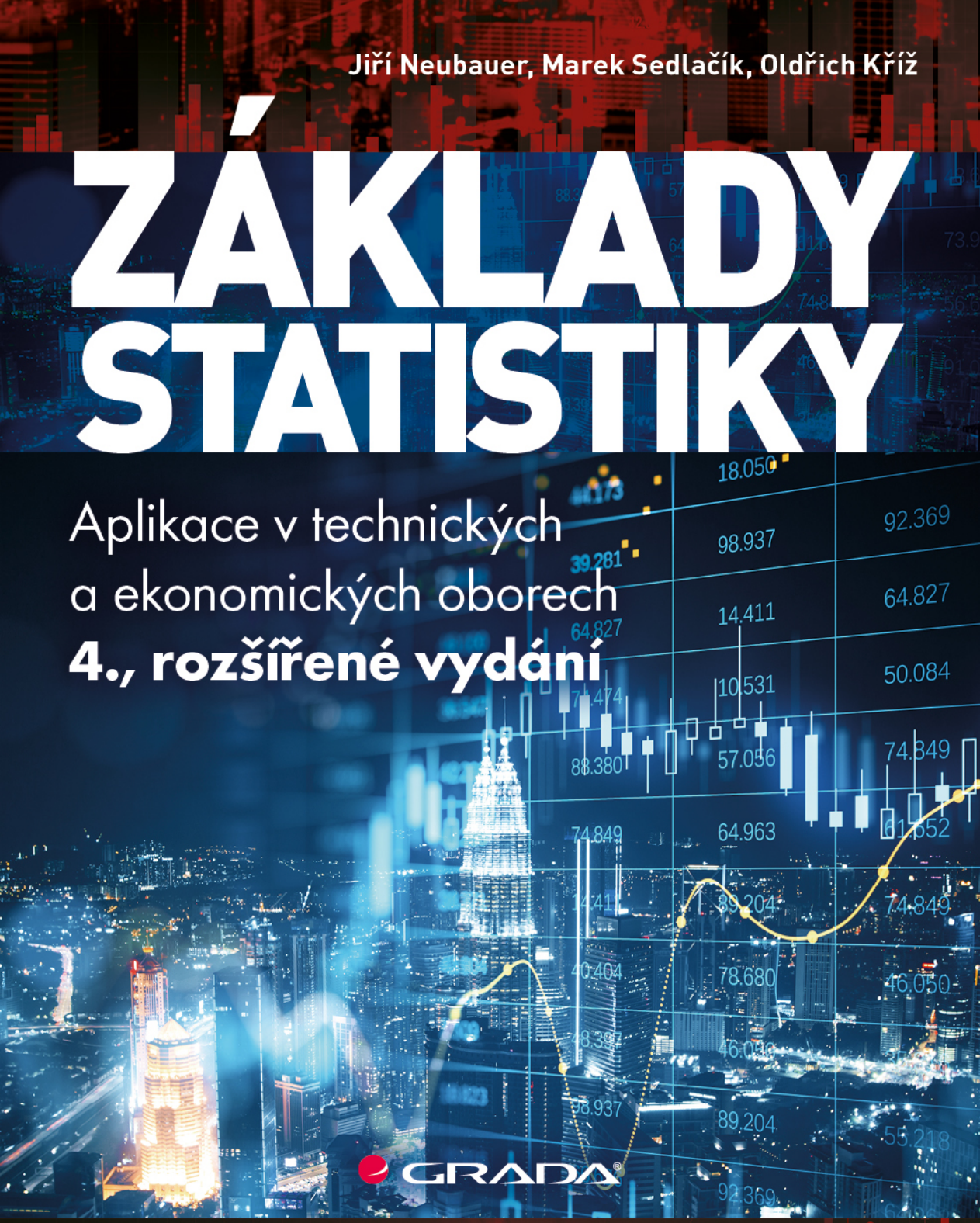
Jiří Neubauer, Marek Sedlačík, Oldřich Kříž

ZÁKLADY STATISTIKY

Aplikace v technických a ekonomických oborech

4., rozšířené vydání



GRADA®

měření a zjišťování | teoretické modely | empirické modely
základy induktivní statistiky | počítačové zpracování dat
praktické užití statistiky

GRADA®

Jiří Neubauer, Marek Sedlačík, Oldřich Kříž

ZÁKLADY STATISTIKY

Aplikace v technických
a ekonomických oborech
4., rozšířené vydání

Grada Publishing

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.

ISBN 978-80-271-8221-3 (pdf)
ISBN 978-80-271-5581-1 (print)

Obsah

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. Podílí 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. Podílí 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 mohu 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, Sedlačíka 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 mohu 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 závislostí, 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 znát, 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 nepřeberného 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 popsané jsou zde nejdůležitější teoretické modely náhodné veličiny. V sedmé a osmé kapitole jsou vybudované teoretické základy statistiky a také je zde popsané statistické usuzování s využitím praktického propojení všech třech pilířů. Velmi praktickou a stěžejní 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 popsané statistické metody. U vybraných příkladů je použití této aplikace podrobně popsané. 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 dovolují 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.

V Brně 31. 7. 2025 Autoři

Kapitola 1

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 sbližovat. 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 extrapolační – 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 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 nejzdlouhavější – 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 nejrůznějších 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 *biomedicínském 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 četných reforem 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 Quételetův 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 založený 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 záznamy 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.

1.2 Význam a pojetí moderní statistiky

V současné době se pojem *statistika* používá v různých významech, v různých souvislostech a také s ohledem na různé praktické situace. V praktickém životě se můžeme setkat se čtyřmi různými významy, které spolu souvisí. Statistikou se rozumí:

a) vědní disciplína, která se zabývá sběrem, zpracováním a vyhodnocováním statistických údajů,
b) číselné i nečíselné údaje nebo souhrn údajů o hromadných jevech,
c) praktická činnost, která vede k získání informací – údajů o zkoumaných jevech,
d) instituce, která provádí praktickou statistickou činnost nebo tuto činnost řídí.

Abychom si udělali korektní obrázek o tom, co budeme pod pojmem statistika rozumět a v jakých souvislostech či situacích budeme tento pojem používat, podívejme se na následující podkapitoly.

Hromadná pozorování a hromadné jevy

Při studiu statistiky budeme vycházet, jak už bylo zmíněno, z teorie pravděpodobnosti, kterou si blíže popíšeme ve 3. kapitole. Základním pojmem pravděpodobnosti jsou tzv. *náhodné pokusy*, tj. takové pokusy, jejichž výsledky nelze předem stanovit. Pro výsledky jednotlivých náhodných pokusů zavedeme označení *náhodné jevy*. Pro *statistické pozorování* – někdy se také hovoří o *statistickém šetření* – jsou typické hromadné jevy. Přívlastkem *hromadný* zdůrazňujeme, že se statistika zabývá pouze takovými náhodnými jevy, které se v prostoru a čase mohou mnohokrát opakovat nebo se vyskytují ve velkém počtu případů. To tedy znamená, že jevy jedinečné (neopakovatelné) statistika do svého zkoumání nezahrnuje.

Hromadné jevy jsou tedy výsledky hromadných pozorování, která se uskutečňují v podstatě dvěma způsoby:

a) jako *výsledky opakovaných pokusů* – tj. za stálých podmínek opakujeme náhodný pokus a po každém pokusu zaznamenáme jeho výsledek; např. $35\times$ opakovaně měříme koncentraci určité látky v roztoku, $60\times$ opakovaně měříme hodnotu elektrického proudu v obvodu, $14\times$ opakovaně měříme vzdálenost dvou bodů v terénu apod.,
b) jako *výsledky pozorované na velkém počtu jednotek* – tj. na všech (mnoha) jednotkách, které máme k dispozici, provedeme měření nebo zjištění hodnoty a všechny takto získané hodnoty si poznamenáme; např. změříme dobu reakce na jistý podnět u 15 řidičů, změříme výkon 23 atletů ve skoku do dálky z místa, zjistíme měsíční příjem u 80 zaměstnanců, zjistíme názor 150 vysokoškoláků na bulvární deník apod.

Pokud jde o vyjadřování výsledků pokusů, hovoříme často o *obměnách (variantách)*. Pro statistiku je obvyklé dvojí vyjadřování obměn – *číselné a slovní*. Např. při vážení rohlíku vyjádříme výsledek, tj. hmotnost rohlíku, ve tvaru 47,8 g (vyjádření číselné: 47,8), při zjišťování výsledku zkoušky z ekonomie vyjádříme výsledek ve tvaru „C“ (vyjádření slovní: dobře). Způsobům vyjadřování výsledků náhodných pokusů se ale budeme ještě dále věnovat podrobněji (viz podkapitolu 1.4).

Při popisu výsledků hromadných pozorování stojí za povšimnutí dvě jejich formy – *měření* a *zjišťování*. Při měření zpravidla získáváme výsledky v číselné podobě jako hodnoty z měřicího přístroje. Hodnoty jsou vyjádřené v určitých jednotkách – fyzikálních, chemických či jiných. V souladu s matematickými a odbornými pravidly je lze také vzájemně převádět. Např. při měření rychlosti auta dostaneme 83,7 km/h, při měření výšky postavy novorozence

dostaneme 49 cm, při měření tvrdosti vody dostaneme 1,8 mmol/l, při měření velikosti proudu dostaneme 12,5 mA, při měření obsahu tuku v mléku dostaneme 1,48 g/l atd. Vlastní zpracování celých množin takovýchto číselných informací – dat – už provádíme bez jednotek (viz kapitolu 2 – Popisná statistika).

Při *zjišťování* získáváme výsledky v číselné nebo slovní podobě jako hodnoty získané z předem definované množiny obměn. Někdy také hovoříme o *popisu* sledovaných objektů. Např. při průzkumu v obchodu zjistíme počet zákazníků u jedné pokladny: 6, při prověřování školních výsledků zjistíme počet bodů z písemného testu u jednoho studenta: 28, při průzkumu kvality pracího prášku zjistíme názor jedné zákaznice: velmi dobrý, při předvolebním průzkumu zjistíme preferenci jednoho voliče: strana B atd.

Zdroje statistických dat

Při řešení konkrétního problému reálného světa se setkáme často s potřebou provést statistické šetření, jehož výsledkem jsou statistická data. Podle typu konkrétního problému bude zdrojem takových dat experiment, dotazování, výkaznictví, pozorování či tzv. sekundární data.

Experimentem budeme rozumět cíleně prováděnou činnost zpravidla za účelem ověření vlivu určitého faktoru na zkoumaný ukazatel. Např. budeme experimentem ověřovat vliv nové technologie výroby na jistou vlastnost výrobku, vliv použitého hnojiva na objem rostlinné produkce, můžeme testovat výrobek na nové podmínky užití, v rámci experimentu můžeme sledovat chování zkoumaných osob v různých modelových situacích apod.

Dotazování je jednoduchý způsob získávání statistických dat, který se provádí písemně (dotazníky, internetové dotazníky) nebo ústně (osobně, telefonicky, ve skupinách). Takto je možné získat informace hromadného charakteru od tzv. *respondentů*, tj. osob náhodně určených k dotazování. Např. vedení střední školy může prostřednictvím dotazníku získat informace o názorech na výuku toho či onoho předmětu, vedení podniku může získat informace o jazykových schopnostech svých pracovníků apod. V některých případech bývá účelné dotazování organizovat anonymně. Na principu dotazování jsou založené také tzv. *ankety*, které však nelze považovat za reprezentativní šetření. Vyplnění anketního dotazníku je totiž naprosto dobrovolné, proto získaný obraz o řešeném problému může být pouze orientační. Např. vydavatel časopisů se takto bude zajímat o zájem čtenářů o jednotlivé rubriky, výrobce nápojů si takto může zjistit názory na kvalitu jeho limonád apod.

Výkaznictví je možné vnímat jako specifickou formu dotazování. Výkazy slouží ke sledování ekonomické činnosti různých subjektů. Jejich odevzdávání a vyhodnocování řídí ČSÚ na základě zákona č. 89/1995 Sb., podle kterého mají ekonomické subjekty tzv. *zpravodajskou povinnost*. Na této formě statistického šetření se podílí také jednotlivá ministerstva a jejich odborné orgány.

Při *pozorování* se obvykle sleduje chování lidských subjektů v různých situacích prostřednictvím smyslů – sledování, ochutnávání, poslouchání, čichání apod. Výsledek pozorování je zpravidla subjektivní a závisí na osobě pozorovatele a na okamžiku, kdy je pozorování prováděno. Např. se formou pozorování provádí tzv. *senzorické analýzy*, kdy se prostřednictvím ochutnávek hodnotí nápoje a potraviny. Podobně lze takto hodnotit vůni sledovaného parfému.

Všechny výše uvedené formy statistického šetření využívaly tzv. *primární data*. V některých případech je možné využít i *sekundární data*, tj. data, která byla získána za jiným účelem v minulosti (například v rámci jiného průzkumu). Sekundární data lze získat z různých

tištěných i elektronických materiálů (statistické ročenky, firemní materiály, novinové zdroje, počítačové databáze, datové nosiče, apod.).

Vztah pravděpodobnosti a matematické statistiky

Ještě jednou se vraťme k pravděpodobnosti. I když počátky pravděpodobnosti jsou spojené s řešením často zajímavých problémů z oblasti hazardních her, v současné době nejčastější aplikace počtu pravděpodobnosti směřují do oblasti statistiky. Okolo nás existuje mnoho věcí, jevů, událostí, které nelze předvídat – jsou důsledkem náhody. Otázkami náhody a náhodných dějů se zabývají dvě matematické disciplíny: teorie pravděpodobnosti a matematická statistika.

Teorie pravděpodobnosti je matematická disciplína, jejímž východiskem je zkoumání náhodných pokusů. Při náhodném pokusu není výsledek jednoznačně určen jeho počátečními podmínkami. Náhodnost určitého pokusu je teoreticky spojena s nedostatečnou znalostí těchto počátečních podmínek. Náhoda však neznamená subjektivní nevědomost, nastoupení každého náhodného jevu lze prostřednictvím matematického aparátu[1] číselně „ocenit“, tedy přiřadit mu pravděpodobnost. Teorie pravděpodobnosti je tedy tou částí matematiky, která přináší do života matematický aparát pro počítání s náhodnými událostmi. Je tak teoretickým základem pro další disciplíny, které s náhodou pracují, jako je teorie náhodných veličin a matematická statistika. Proto jsou užitečné také modely různých rozdělení pravděpodobností (např. binomický, normální, exponenciální atd. – viz kapitoly 5 a 6).

Matematická statistika je naproti tomu věda, která zahrnuje studium dat vykazujících náhodná kolísání, ať už jde o data získaná pečlivě připraveným pokusem provedeným pod stálou kontrolou experimentálních podmínek v laboratoři, či o data pocházející přímo z terénu. Statistika se tedy zabývá získáváním informací z empirických dat, jejím principem je učinit na základě vzorku závěr o celku. Předpokládá, že data obsahují nepřesnosti a nejistoty, které jsou způsobeny náhodnými vlivy. Matematickou statistiku tvoří soubor metod pro zpracování hromadných dat, v nichž se závěry vyvozují na základě teorie pravděpodobnosti. Právě těmto úkolům statistiky (a také v těchto souvislostech) se budeme věnovat v dalších částech této učebnice – viz kapitoly 7, 8, 9 a 10.

Součásti matematické statistiky

Jak jsme už naznačili, v rámci hromadných pozorování provádíme měření nebo zjišťování sledované veličiny u velkého počtu jistých objektů. Výsledkem pozorování jsou potom *hromadná empirická data*, která v sobě zahrnují (spíše skrývají) řadu informací o sledované veličině. Tyto informace však „na první pohled“ nejsou zřejmé, data totiž představují neuspořádanou, až chaotickou „horu“ údajů a nelze z nich prakticky žádné informace vyčíst. Proto je třeba data nejprve zpracovat a informace v nich obsažené získat. Zpracováním empirických dat se zabývá *popisná statistika* (viz dále kapitolu 2). Využívá k tomu různé *tabulky* a *grafy*, které pomáhají objevit významné vlastnosti sledované veličiny. Hovoříme často o tabulkovém či grafickém vyjádření rozdělení četností. Některé tabulky poskytují zdrojová data pro tvorbu grafů. Dalším prostředkem popisu hromadných empirických dat jsou tzv. *číselné charakteristiky*, které vyjadřují určité vlastnosti sledované veličiny jediným číslem. K určení takových čísel použijeme

[1] Axiomatická teorie pravděpodobnosti publikovaná v roce 1933 A. N. Kolmogorovem je založená na teorii míry, alternativní bayesovská teorie publikovaná v roce 1955 E. T. Jaynesem je založená na klasické logice pro případ výroků, jejichž pravdivostní hodnota není jen 0 nebo 1, ale leží mezi těmito hodnotami.

jen elementární matematické operace. Cílem popisné statistiky je tedy zpřehlednění informací obsažených v datovém souboru.

Další součástí matematické statistiky jakožto vědního oboru je tzv. *matematická statistika* v užším slova smyslu, která se systematicky zabývá (zejména pomocí teorie pravděpodobnosti) matematickými metodami vhodnými pro analýzu statistických dat. Obecně má deduktivní povahu, předmětem našeho zájmu je vždy určitý celek, tzv. základní soubor (viz podkapitolu 1.3), ale cesta, kterou se k němu dostaneme, má naopak výhradně induktivní[2] charakter. Důležitými součástmi matematické statistiky jsou:

a) *Teorie odhadu* – zabývá se určováním odhadů neznámých parametrů základního souboru pomocí hromadných empirických dat získaných náhodným výběrem (viz podkapitolu 8.1) a studuje různé přístupy k získání bodových a intervalových odhadů (viz podkapitoly 9.1 a 9.3).
b) *Testování statistických hypotéz* – zabývá se statistickými procedurami pro ověřování hypotéz o základním souboru a o srovnávání více souborů z různých hledisek pomocí hromadných dat získaných náhodným výběrem (viz kapitolu 10 – Testování statistických hypotéz).
c) *Statistická predikce* – zabývá se statisticky kvalifikovanými odhady budoucího vývoje sledované veličiny na základě její současné dynamiky.

Na závěr výkladu o významu a pojetí moderní statistiky ještě připojme jednu zásadní myšlenku. V minulosti se statistika často ztotožňovala s pouhým zjišťováním, sumarizací a publikováním zjištěných údajů. V současné době lze předpokládat, že moderní statistika má všechny atributy vědní disciplíny schopné v podstatně větším měřítku respektovat potřeby kvalifikovaných rozhodovacích procesů. Proto nezapomeňme: statistiku nelze ztotožňovat s pouhým elementárním zpracováním údajů! Statistiku musíme spojovat s ohledem na její výrazně praktický charakter s širokou škálou metod a technik, které umožňují kvalifikované rozhodování na bázi kvantitativních informací o praktickém problému.

Příklady k procvičení

1. Rozhodněte, zda je možné definované jevy považovat za jevy hromadné:
   a) hrubý měsíční příjem učitelů na středních školách v ČR,
   b) počet dětí v českých rodinách,
   c) počet nezaměstnaných v Jihomoravském kraji v září 2011,
   d) denní tržba v prodejně,
   e) počet dosažených gólů konkrétním hráčem za zápas v hokejové lize 2011/2012,
   f) rychlost připojení k internetu u vlastního počítače.
2. Posuďte, jakým způsobem je možné u popsaného věcného problému získat statistická data:
   a) vliv použitého krmiva na živé přírůstky sledovaných prasat,
   b) denní spotřeba vody v domácnosti,
   c) měsíční tržba v soukromém obchodu,
   d) názor na úroveň základních služeb mobilního operátora,
   e) hodnocení světlého výčepního piva z českých pivovarů,
   f) vliv druhu benzínu na výkon motoru,
   g) porovnání cenové hladiny v několika supermarketech.

[2] Při induktivním způsobu myšlení nalézáme při zkoumání jednodušších konkrétních případů pomocí abstrakce jejich společnou obecnou zákonitost – v induktivní statistice to probíhá tak, že z vlastností výběrového souboru budeme usuzovat na vlastnosti základního souboru.

3. Je anketa reprezentativní statistické šetření?
4. Vyjmenujte některé hromadné jevy:
   a) z oblasti vaší profesní činnosti,
   b) z oblasti vaší zájmové činnosti,
   c) z oblasti veřejného zájmu (zdroje: noviny, rozhlas, televize, internet).

Řešení.
1. a) ano; b) ano; c) ne; d) ano; e) ne; f) ano;
2. a) experiment; b) pozorování; c) výkaznictví; d) dotazování; e) pozorování; f) experiment; g) pozorování;
3. není.

1.3 Statistická jednotka a statistický soubor

V podkapitole 1.2 jsme uvedli, že úkolem statistiky je provádět hromadná pozorování a sledovat hromadné náhodné jevy. Protože statistika je věda velmi praktická, budeme hromadná pozorování provádět na reálných objektech nebo subjektech, které jsou z určitého konkrétního důvodu předmětem našeho zájmu. Pozornost proto budeme věnovat nejprve *statistickým jednotkám* a jejich jednoznačnému vymezení, potom si vysvětlíme pojmy *základní soubor* a *výběrový soubor*.

Definice 1.3.1 Jednotlivé objekty nebo subjekty, které jsou při statistickém zkoumání sledované, se nazývají *statistické jednotky*. Každá statistická jednotka musí být jednoznačně vymezena, aby nemohlo dojít k dvojímu nebo jinak zkreslenému výkladu zjištěných údajů. Statistické jednotky se vymezují z hlediska:
- věcného,
- prostorového,
- časového.

Příklad 1.3.2 Statistickými jednotkami mohou být:
– osoby, lépe řečeno jisté kategorie osob – novorozenci, žáci, voliči, zaměstnanci podniku, důchodci, pacienti. . . ,
– věci a předměty – výrobky, stroje, budovy. . . ,
– organizace – podniky, úřady, školy, obce. . . ,
– zvířata – psi, ryby, sloni. . . ,
– rostliny nebo plody – pšenice, růže, břízy, jablka. . . ,
– události, jevy – sportovní výkony, poruchy, meteorologické jevy. . . .

Příklad 1.3.3 Proveďte věcné, prostorové a časové vymezení těchto statistických jednotek:
a) všechna osobní auta projíždějící v úterý mezi 14. a 16. hodinou 110. km dálnice D1 směrem na Brno;
b) všechna děvčata ze 6. tříd znojemských základních škol v červnu roku 2012;
c) všichni kapři v jihočeském rybníku Bezdrev v listopadu 2010;
d) všechny 50gramové rohlíky z týdenní produkce pekaře Ječmínka v týdnu od 12. do 17. 3. 2012; uvažujme dále, že výroba těchto rohlíků bude za nezměněných podmínek (stejná mouka, stejná voda, stejná teplota pecí, stejná směna. . .) pokračovat i v dalším období;
e) všechny hypotetické výsledky výkonového testu u jednoho volejbalisty – výskok s dohmatem odrazem snožmo z rozběhu – v tréninkové hale v období letní tréninkové přípravy 2011.

Řešení. Vymezení statistických jednotek v jednotlivých případech:

a) věcné: osobní vozy, prostorové: 110. km dálnice D1 směr Brno, časové: konkrétní úterý v daném 2hodinovém intervalu,
b) věcné: žákyně 6. tříd, prostorové: ZŠ ve Znojmě, časové: červen 2012,
c) věcné: kapři, prostorové: rybník Bezdrev, časové: listopad 2010,
d) věcné: 50gramové rohlíky, prostorové: pekařství Ječmínek, časové: období týdne od 12. do 17. 3. 2012, resp. období od 19. 3. 2012 dále,
e) věcné: výskok s dohmatem odrazem snožmo z rozběhu, prostorové: tréninková hala, časové: léto 2011.

Definice 1.3.4 Množina statistických jednotek stejného typu a shodného vymezení tvoří *statistický soubor*. V rámci statistického šetření budeme rozlišovat dva typy souborů:

- *základní soubor (populace)* – množina všech shodně vymezených statistických jednotek,
- *výběrový soubor (výběr, vzorek)* – podmnožina základního souboru, tedy vybraná část populace.

O základním souboru se také hovoří jako o statistickém souboru, který je předmětem našeho zájmu a o jehož vlastnostech se mají činit závěry. Může být buď *reálný*, pokud všechny jeho statistické jednotky reálně existují, nebo *hypotetický*, který je sice obecně definován, ale při statistickém zkoumání reálně jeho statistické jednotky neexistují nebo jich existuje pouze část – viz příklad 1.3.5. Z charakteru základního souboru bezprostředně vyplývá, že může mít *konečný* rozsah nebo může být *nekonečný*.

Při přípravě statistického šetření je jedním z nejdůležitějších požadavků na kvalitní analýzu *homogenita* základního souboru. Ta je z valné části zabezpečená pomocí shodného vymezení všech jednotek v základním souboru. Je třeba mít na paměti, že chyby ve vymezení základního souboru se potom přenáší na výběrový soubor, se kterým se dále pracuje. To však vede k nespolehlivým až chybným závěrům – viz příklad 1.3.6.

Počet jednotek základního souboru je vesměs velký. Připomeňme si, že v začátcích statistiky bylo cílem vyčerpávající šetření, neboli celý základní soubor. Až teprve matematická statistika přinesla možnost provádět pouze výběrová šetření a namísto celé populace pracovat se vzorkem. To, že se dává dnes přednost výběrovému šetření před vyčerpávajícím šetřením, má hned několik důvodů:

- důvody ekonomické – výběrové šetření šetří čas i peníze; zejména u rozsáhlých souborů by měření nebo zjišťování na všech statistických jednotkách nebylo časově možné nebo by bylo velmi drahé (např. vážení 85 000 kusů rohlíků),
- důvody technické – při prováděném měření se statistická jednotka může znehodnotit (např. při degustaci masové konzervy se musí konzerva otevřít, čímž se znehodnotí),
- důvody praktické – v situacích, kdy je základní soubor zcela nebo z části hypotetický, je potom měření prakticky nemožné (např. sportovní výkony není možné nekonečně mnohokrát opakovat).

S metodami pořizování výběrových souborů se seznámíme později (viz podkapitolu 8.1).

Příklad 1.3.5 V příkladu 1.3.3 jsme provedli věcné, prostorové a časové vymezení popsaných statistických jednotek. Definujte ve stejných situacích základní a výběrový soubor.

Řešení.

a) Základní soubor: hypotetický – množina všech osobních vozů, které daným místem v dané době projíždí (jejich počet však konkrétně stanovit před měřením nelze, teoreticky budeme základní soubor považovat za nekonečný), výběrový soubor: např. 80 náhodně vybraných vozů;
b) základní soubor: reálný – množina všech žákyň 6. tříd na všech ZŠ ve Znojmě v červnu 2012 (jejich počet je 345, základní soubor má tedy rozsah konečný), výběrový soubor: např. 25 náhodně vybraných žákyň;
c) základní soubor: reálný – množina všech kaprů v rybníku Bezdrev v listopadu 2010 (jejich počet reálně stanovit nelze, z praktických důvodů lze však předpokládat, že počet je konečný!), výběrový soubor: např. 50 náhodně vybraných kaprů;
d) základní soubor: reálný i hypotetický – množina všech rohlíků vyrobených v pekařství Ječmínek v týdnu od 12. do 17. 3. 2012 (85 000 kusů) a množina dalších rohlíků, jejichž výroba bude dále pokračovat (základní soubor je tedy nekonečný), výběrový soubor: např. 120 náhodně vybraných kusů;
e) základní soubor: hypotetický – fiktivní množina všech možných výskoků s dohmatem odrazem snožmo z rozběhu (jejich počet je možné si pouze teoreticky představit, reálně neexistuje), výběrový soubor: např. 10 provedených výskoků s dohmatem během jednoho tréninku.

Příklad 1.3.6 Rozhodněte, zda konstruované základní soubory jsou s ohledem na sledovanou veličinu homogenní a rozhodnutí zdůvodněte:

a) všechna osobní auta projíždějící 110. km dálnice D1 v úterý odpoledne ve směru na Brno i na Prahu – sledovat budeme rychlost,
b) všichni žáci 6. tříd znojemských základních škol v červnu roku 2012 – sledovat budeme výsledky testu z fyziky a výkony v běhu na 60 m,
c) všichni kapři v jihočeském rybníku Bezdrev v listopadu 2010 a 2011 – sledovat budeme hmotnost kapra,
d) všechny 50gramové rohlíky z produkce pekaře Ječmínka v období od 12. do 31. 3. 2012 – sledovat budeme hmotnost rohlíku.

Řešení.

a) Provoz na dálnici v obou směrech je obecně odlišný – za tohoto předpokladu je prostorové vymezení chybné, základní soubor je tedy nehomogenní,
b) při sledování vědomostí žáků ZŠ se u chlapců a děvčat obecně neočekávají různé výsledky – za tohoto předpokladu je věcné vymezení správné, základní soubor je homogenní; při sledování sportovních výkonů u chlapců a děvčat (obecně u mužů a žen) existují biochemické, anatomické, fyziologické a jiné odlišnosti zdůvodňující různé sportovní výkony – za tohoto předpokladu je věcné vymezení chybné, základní soubor je nehomogenní,
c) podmínky chovu ryb ve dvou různých letech obecně shodné být nemohou – za tohoto předpokladu je časové vymezení chybné, základní soubor je nehomogenní; reálné a smysluplné je však definovat dva různé homogenní základní soubory, v každém roce jeden, s cílem porovnání hmotnosti kaprů v obou letech,
d) uvažujme, že výroba probíhá v celém období za zcela shodných podmínek – za tohoto předpokladu je časové vymezení správné, základní soubor je homogenní.

Příklady k procvičení

1. Proveďte věcné, prostorové a časové vymezení následujících statistických jednotek se sledovanou vlastností a posuďte, zda je základní soubor homogenní:
   a) rychlost aut jedoucích denně po 15. hodině v Brně, ulice Provazníkova;
   b) výška smrků v oblasti Březník NP Šumava v roce 2010;
   c) napětí v elektrické síti v domácnostech na sídlišti Jižní Svahy ve Zlíně v září 2011;
   d) výkon v běhu na 12 minut absolventů Střední policejní školy v Praze v roce 2012.
2. Pro statistické zjišťování prováděné prostřednictvím formuláře určete, s jakými statistickými jednotkami toto zjišťování pracuje a vymezte je věcně, prostorově a časově:
   a) v případě daňového přiznání,
   b) v případě povinného ručení motorového vozidla.
3. Řešte následující úkoly:
   a) Posuďte, zda je z pohledu statistiky předmětem našeho zájmu základní soubor nebo výběrový soubor.
   b) Jak musí být vymezené statistické jednotky v základním souboru?

Řešení.

1. a) Všechna auta jedoucí daným místem, Provazníkova ulice Brno, každý den po 15. hodině, nehomogenní (protože hustota provozu se obecně každý den liší);
   b) všechny smrky v dané oblasti, Březník, 2010, homogenní;
   c) všechna hypotetická měření v elektrické síti, domácnosti v domech na sídlišti, září 2011, homogenní (ale za předpokladu shodné technologie měření);
   d) všechny hypotetické výsledky běhu na 12 minut absolventů SPŠ Praha, 2012, homogenní.
2. a) Plátce daně, vypočítaná částka daně, FÚ Olomouc, duben 2012;
   b) majitel motorového vozidla, výše pojistky, ČSOB pojišťovna, 2012.
3. a) Základní soubor;
   b) věcně, prostorově a časově shodně.

1.4 Statistický znak

Už bylo řečeno, že statistiku budeme využívat při řešení konkrétních a praktických problémů z reálného života. Z formulace problému nám musí být jasné, jaké výrobky, osoby, zvířata, rostliny, události, organizace jsou předmětem našeho zájmu, ale také o jaké jejich vlastnosti se budeme zajímat. V další části se tedy zaměříme právě na tyto vlastnosti, tzv. *statistické znaky*, a to jak z pohledu jejich vyjádření, tzv. *hodnot* a *obměn*, tak také z pohledu statistických operací s nimi.

Definice 1.4.1 Vlastnosti, které u statistických jednotek budeme v rámci statistického šetření sledovat, nazýváme *statistické znaky* neboli *statistické proměnné*. Různé hodnoty, kterých může statistický znak nabývat, nazýváme *obměny* neboli *varianty*. Podle způsobu vyjádření hodnot dělíme statistické znaky na:

- *číselné – měřitelné*,
- *slovní – kategoriální*.

Podle typu vztahů mezi hodnotami a obměnami budeme rozlišovat statistické znaky:

- *metrické – měřitelné*,
- *ordinální – pořadové*,
- *nominální – jmenovité*.

V rámci statistického šetření (měření, pozorování) vyjádříme míru sledované vlastnosti u každé jednotky statistického souboru prostřednictvím tzv. *hodnoty* statistického znaku. Aktuální – naměřené hodnoty proměnné jsou *data*. Počet naměřených hodnot odpovídá rozsahu výběrového souboru. Hodnotu znaku ve smyslu vyjádření různého stupně dané vlastnosti označíme jako *obměnu – variantu* statistického znaku. Počet obměn je zpravidla menší, nejvýše roven rozsahu souboru. Např. k otázce zavedení školného na VŠ se vyjádřilo 73 studentů 1. ročníků ze 2 fakult jedné VŠ: souhlasím – 25 studentů, nesouhlasím – 34 studentů a je mi to jedno – 14 studentů; výsledky průzkumu jsou vyjádřené 3 obměnami (souhlasím, nesouhlasím, je mi to jedno), celkový počet naměřených (zjištěných) hodnot je 73.

Statistické znaky lze klasifikovat podle několika hledisek. Především se nabízí hledisko vyjádření hodnot znaku číselně nebo slovně. Lze-li hodnoty znaku vyjádřit číselně, rozumí se tedy *numericky*, jde o znak *číselný* (např. počet žáků ve třídě, spotřeba vody v domácnosti za rok). Pokud se hodnoty znaku vyjadřují slovně, hovoříme o znaku *slovním* (např. barva očí člověka, druh vlastnictví bytu). Slovní proměnné se často označují jako *kategoriální*. Takové členění statistických znaků není samoúčelné, protože pro číselné a slovní znaky budou konkrétní statistické postupy a metody vesměs rozdílné. Při zpracování dat hraje roli také to, zda data představují hodnoty znaku nespojitého (diskrétního) nebo spojitého. *Nespojité* znaky nabývají pouze jednotlivé číselné nebo slovní hodnoty (např. počet dvojchyb tenisty v zápase, počet vadných výrobků v sérii, státní příslušnost studenta VŠ). *Spojité* statistické znaky mohou nabývat libovolných hodnot v rámci určitého intervalu (např. doba čekání na obsluhu v restauraci, obsah tuku v mléku, sladkost limonády). Více se o vztahu mezi spojitými a nespojitými znaky dozvíme později.

Na tuto základní klasifikaci statistických znaků (proměnných) úzce navazuje třídění podle typu vztahů mezi hodnotami a obměnami znaků. Podle tohoto kritéria dělíme proměnné na metrické, ordinální a nominální.

Metrické neboli *měřitelné* proměnné jsou takové proměnné, které nabývají výhradně číselných hodnot a vyjadřují tedy velikost měřené vlastnosti. Jejich další dělení se provádí podle oboru jejich hodnot. Pokud jsou tyto hodnoty vyjádřené pouze kladnými čísly (např. rychlost auta na dálnici, výkon ve skoku do výšky), proměnnou označíme jako *kardinální*. V české literatuře se také používá pojem *poměrová* proměnná. Její každé dvě hodnoty lze porovnávat jak rozdílem, tak i podílem. Je tedy možné stanovit, o kolik jednotek je jedna hodnota větší (event. menší) než druhá, a také kolikrát je jedna hodnota větší (event. menší) než druhá. Druhou skupinu metrických proměnných tvoří takové proměnné, které nabývají kladné i nekladné číselné hodnoty (např. teplota vzduchu ve °C, počet dětí v rodině). Tyto proměnné jsou nekardinální, zpravidla se označují jako *intervalové*. U těchto proměnných lze každé dvě hodnoty porovnávat jen rozdílem, lze tedy stanovit, o kolik jednotek je jedna hodnota větší (event. menší) než druhá. Porovnání podílem není možné zpravidla proto, že množina obměn obsahuje nulu. Ve statistice však rozdíl mezi poměrovou a intervalovou proměnnou nehraje velkou roli. Je totiž zřejmé, že každou nekardinální metrickou proměnnou lze vhodnou (a jednoduchou) transformací převést na proměnnou kardinální. To se prakticky odráží v tom, že pro obě kategorie metrických proměnných se vesměs používají shodné metody statistických analýz. Rozdíl mezi nimi se týká až praktické interpretace získaných výsledků.

Při zpracování metrických dat považujeme většinou odpovídající proměnné za spojité, jako kdyby mohly nabývat kterékoliv hodnoty z číselného intervalu, i když při praktickém měření tomu tak není. Dokonce i u veličin, které principiálně spojité jsou, jako rozměr nebo čas,

musíme při praktickém měření volit konečnou jednotku rozlišení, takže i tyto proměnné se chovají navenek jako diskrétní (nespojité). Přesto při statistickém zpracování budeme většinou užívat pro metrické proměnné postupy matematicky odvozené pro veličiny spojité.

Příklad 1.4.2 Poměrové znaky jsou např. cena benzinu Natural 95, výška dospělého muže, obvod kmene javoru, délka kapra, hmotnost jablka dané odrůdy, počet zaměstnanců podniku, počet členů domácnosti atd. (Všimněte si, že u všech uvedených znaků představuje obor hodnot množinu pouze kladných čísel!)

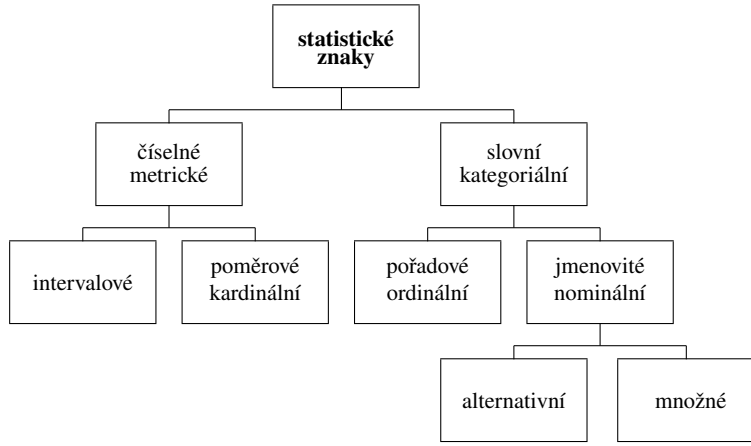
Příklad 1.4.3 Intervalové znaky jsou např. počet chyb v diktátu, chyba při zaokrouhlení ceny nákupu (na celé Kč), počet prodaných televizorů za den, výše kapesného žáka 4. třídy ZŠ, rok narození apod. (U všech uvedených znaků obsahuje množina hodnot znaku nulu!)

Ordinální neboli *pořadové* proměnné jsou slovní (kategoriální) proměnné, u jejichž obměn má smysl jejich uspořádání, lze je tedy jednoznačně seřadit od varianty vyjadřující nejnižší úroveň sledované vlastnosti až do varianty s úrovní nejvyšší, nebo naopak (např. dosažené vzdělání: základní, střední, vysoké). Toho se často využívá k tomu, že slovně vyjádřeným obměnám ordinální proměnné se podle jejich pořadí přiřazují pořadová nebo jiná čísla (stupně, body, procenta apod.), která vyjadřují pořadí slovních variant (např. školní klasifikace: výborně – 1, chvalitebně – 2, dobře – 3, dostatečně – 4 a nedostatečně – 5). Rozdíl dvou hodnot ordinální proměnné potom vyjadřuje rozdíl v jejich pořadí! Proto je důležité pracovat s pořadovými čísly jako s určitou formou kvantifikace těchto obměn a zohledňovat skutečnost, že nelze nikdy stejnou hodnotu u dvou různých statistických jednotek považovat za zcela totožnou (např. dva studenti byli u zkoušky hodnoceni stupněm dobře – z toho však nelze usoudit, že oba studenti mají zcela shodné vědomosti). To je zásadní rozdíl mezi pořadovým a metrickým znakem, který vyvolává potřebu poněkud odlišných metod zpracování.

Příklad 1.4.4 Ordinální znaky s možným oborem hodnot:
- rychlost chemické reakce (+, ++, +++),
- odpověď na konkrétní otázku v sociologickém průzkumu (naprosto souhlasím, spíše souhlasím, mám neutrální postoj, spíše nesouhlasím, naprosto nesouhlasím),
- kategorie vojenských hodností v AČR (mužstvo, poddůstojníci, rotmistři, praporčíci, nižší důstojníci, vyšší důstojníci, generálové),
- pořadí v soutěži výcviku psích plemen (1. pořadí, 2. pořadí, pracovní, nehodnocený),
- kategorie kuřáků (počet vykouřených cigaret za den: do 5, do 10, do 20, nad 20).

Nominální neboli *jmenovité* proměnné jsou slovní (kategoriální) proměnné, které nelze vzájemně porovnávat, tedy u jejichž obměn nelze stanovit žádné pořadí. Lze pouze stanovit shodu nebo neshodu v hodnotě znaku u každých dvou statistických jednotek (např. druhy masa: vepřové, hovězí, telecí, kuřecí, jiné). Někdy se hodnoty nominálního znaku mohou vyjádřit také čísly, nemají však žádný kvantitativní význam (např. čísla tramvají: 1, 6, 7 a 12). Data odpovídající hodnotám nominálního znaku vždy vyžadují zpracování specifickými metodami.

Příklad 1.4.5 Nominální znaky s možným oborem hodnot:
- dominance ruky (pravák, levák),
- barva výrobku (červený, modrý, jiný),



Obr. 1.1 *Klasifikace statistických znaků*

– rodinný stav muže (svobodný, ženatý, rozvedený, vdovec),
– cizí státní občanství (Slovensko, Polsko, Rakousko, Německo, jiné),
– kategorie přívlastkových vín (kabinetní, pozdní sběr, výběr z hroznů, výběr z bobulí, výběr z cibéb, ledové a slámové).

Třetím kritériem třídění proměnných je hledisko počtu obměn. Smysluplné je to prakticky jen pro slovní proměnnou: pokud nabývá pouze dvou obměn, mluvíme o *alternativním* znaku (např. pohlaví: muž, žena; odpověď na otázku v referendu: ano, ne), je-li počet obměn větší než dvě, jedná se o znak *množný* (např. fakulty UO v Brně: FVL, FVT a FVZ; dosažené vzdělání: ZŠ, SŠ, VŠ-Bc, VŠ-Mgr, VŠ-Ph.D.). Z čistě praktických důvodů, bez matematického pozadí, se alternativní proměnná někdy vyjadřuje jako numerická proměnná s obměnami 0 a 1. Jedna její obměna se označí číslicí 1 (zpravidla ta, která nás v dané souvislosti více zajímá) a druhá potom číslicí 0. Hovoří se o nulajedničkové proměnné. Pro tyto proměnné je vyvinutá řada metod, které využívají právě jistého zjednodušení prostřednictvím dvojice obměn. Proto se také množné proměnné někdy převádějí na alternativní proměnné, a to tak, že se více obměn spojí do jedné varianty, která nás s ohledem na účel zkoumaného problému více zajímá, a zbývající obměny se spojí do druhé obměny (např. dosažené vzdělání: nižší = ZŠ + SŠ a vyšší = Bc + Mgr + Ph.D.).

Na alternativní proměnnou lze, pokud je to s ohledem na zkoumaný problém užitečné, převést i jakoukoli metrickou proměnnou. Při takové „transformaci“ měřitelné proměnné na proměnnou slovní musíme však počítat s určitou ztrátou informace obsažené v původních datech (např. hrubý měsíční příjem lékařů: všechny hodnoty převedeme do dvou kategorií, příjem pod a nad částkou 35 tis. Kč).

Klasifikaci statistických znaků je možné zjednodušeně vyjádřit schématem – viz obrázek 1.1.

Statistický znak nabývá vždy slovních nebo číselných hodnot a je zjišťován u každé statistické jednotky statistického souboru. Jestliže ve statistickém souboru pracujeme jen s jedním znakem (s jednou proměnnou), říkáme, že se jedná o *jednorozměrný soubor*. Zkoumáme-li současně dva nebo více znaků, jde o dvourozměrný, resp. obecně *vícerozměrný soubor* (např.