1 ZÁKLADY ŠTATISTIKY

Úspešné rozhodovanie v ekonomickej praxi nie je možné bez aspoň minimálneho využívania štatistiky a štatistických metód. Vysvetlenie pojmu štatistika nie je jednoduché, hoci v súčasnosti takmer každý človek niečo o štatistike vie a má o nej nejakú predstavu. Názory na štatistiku sú však veľmi rôzne. Na jednej strane sa štatistika preceňuje, čo je dané predovšetkým tým, že jedným z dôležitých nástrojov štatistiky je matematika ako všeobecne uznávaná vedná disciplína. Na druhej strane sa štatistika podceňuje, čo je spôsobené možnými zámernými skresleniami požadovaných a poskytovaných zistení. Štatistika je v skutočnosti veľmi dôležitý a ničím nezastupiteľný nástroj objektívneho poznávania reálneho sveta a jeho zákonitostí.

V tejto kapitole sa bližšie oboznámime s pojmom štatistika a aspoň stručne s jej vývojom od prvých úradných zisťovaní až po dnešnú modernú štatistiku. V časti 1.2 objasníme základné štatistické pojmy, ako sú štatistická jednotka, štatistický znak a štatistický súbor. Časť 1.3 je venovaná etapám štatistického skúmania, počas ktorých sa získavajú, spracovávajú a vyhodnocujú informácie o vlastnostiach hromadných javov a procesov. Časti 1.4 a 1.5 sa zaoberajú konkrétnymi metódami zobrazenia kvalitatívnych a kvantitatívnych dát. Východiskom spracovania teoretickej časti textu tejto kapitoly boli publikácie autorov Hendl, 2014; Cyhelský, Kahounová, Hindls, 2010; Pacáková a kol., 2009; Kotlebová, 2017; Bakytová, 1990, pričom detailnejšie sa touto problematikou zaoberajú napríklad Řezánková, Löster, Šulc, 2019; Meloun, Militký, 2012; Hair a kol., 2006 a ďalší.

1.1 Štatistika a jej vývoj

Samotný pojem štatistika sa v súčasnosti používa vo viacerých významoch: ako vedná disciplína, praktická činnosť, štatistický úrad, ale aj štatistické údaje alebo charakteristiky. K vysvetleniu pojmu štatistika prispeje krátky pohľad do histórie (detailnejšie napr. Hald, 2007, 2003 a 1998; Žák, 2006; Lee, 2016; Stigler, 1999 a 1986; Bakytová, 1990).

Úradné zisťovania

Najstaršie písomné dokumenty o využívaní štatistiky pochádzajú z oblasti Sumeru. Sú to záznamy o počte osôb, počte kusov domáceho dobytku a úrode v konkrétnych časových intervaloch.

Staroveká ríša, mezopotámskymi mestskými štátmi počínajúc, bola finančne úplne závislá od úspešného výberu daní, či už išlo o pracovné povinnosti obyvateľstva, naturálne alebo peňažné dávky. Na tieto účely existovali prepracované „štatistické metodiky“

a z mnohých písomných pamiatok sme schopní vytvoriť si predstavu o systéme vtedajšej štátnej správy.

Zvlášť prepracovaný systém evidencie, tzv. cenzus, mal staroveký Rím. Cenzus sa pravidelne vykonával v republikánskom období; pôvodne išlo o súpis nehnuteľného majetku, neskôr aj otrokov a dobytka, na základe ktorého sa vypočítala výška dane pre jednotlivých občanov. Asi najznámejší sa dnes javí jeden z posledných pravidelných cenzov, uskutočnený za cisára Augusta okolo prelomu letopočtu, ktorý je známy v dejinách ako časové určenie narodenia Ježiša Krista.

Obdobie raného stredoveku prinieslo do strednej a západnej Európy všeobecný rozvrat a negramotnosť bola bežným javom aj medzi panovníkmi. Určité ohniská vzdelanosti udržovala cirkev, ktorá bola schopná viesť evidenciu svojho majetku a jeho zmien. Členovia cirkevných rádov boli zamestnávaní aristokraciou ako „štatistici".

Stredovekými „štatistickými ročenkami" boli vrchnostenské urbáre, z ktorých okrem evidencie príjmov poddaných bolo možné vyčítať aj rozsah pozemkového vlastníctva šľachty a cirkvi. Mestské daňové knihy a daňové zoznamy obsahovali zoznam daňovníkov spolu s ich nehnuteľným majetkom a výškou odvedených daní. V 14. storočí sa v Európe objavujú prvé cirkevné matriky.

Univerzitná štátoveda

16. storočie so sebou prinieslo nový pohľad na svet a človeka, rozvoj filozofie a začiatky moderných vedných disciplín. Je to obdobie veľkých spoločenských zmien. Spoločnosť sa viac diferencuje a z potreby jej efektívneho riadenia vzniká aj potreba jej skúmania. V duchu opisnej štatistiky vyšlo v Benátkach v roku 1562 jedno z prvých štáto-vedných diel *O vláde a správe v rôznych kráľovstvách* od Francesca Sansoviny.

V roku 1589 použil Talian Girolamo Ghilini ako prvý termín š*tatistika*, v pôvodnom význame išlo o stav štátu. Štatistika mala vtedy význam súhrnu znalostí o významných štátnych záležitostiach. Celé storočia bol termín štatistika používaný miesto termínu štátoveda najmä v Nemecku.

Existovala tzv. *univerzitná štatistika*, ktorá spočívala v cykle prednášok na stredovekých univerzitách, na ktorých profesori verbálne opisovali obyvateľstvo, územie, obchod, peňažníctvo, armádu atď. jednotlivých štátov.

Politická aritmetika

V Anglicku v 17. storočí vznikol iný okruh štatistiky, tzv. *politická aritmetika* (dnes hovoríme o demografii), ktorá vychádzala z údajov o narodených a zomrelých a pokúšala sa na ich základe skúmať vývoj stavu obyvateľstva v dlhších časových obdobiach. Išlo vlastne o prvé skúmania časových radov.

Zakladateľom tejto disciplíny bol John Graunt (1620 – 1674) a Wiliam Petty (1623 – 1687), ktorí ako prví považovali demografické javy za hromadné javy. Graunt odhalil pomer medzi počtom mužov a žien v populácii a stabilný pomer medzi počtom narodených chlapcov a dievčat. Zostavil úmrtnostné tabuľky na základe skúmania vymierania jednotlivých vekových skupín. Ďalším reprezentantom anglickej školy bol Edmund Halley (1656 – 1742), ktorý na konci 17. storočia pri konštrukcii úmrtnostných tabuliek prepojil štatistiku s teóriou pravdepodobnosti.

Postupy skúmania Graunta, Pettyho a ďalších boli nazvané politickou aritmetikou nielen preto, že jedna z kníh Pettyho sa volala *Politická aritmetika*, ale tiež preto, že používali čísla, jednoducho aritmetiku pri skúmaní a charakterizovaní hromadných javov.

Medzi politickými aritmetikmi a univerzitnými štatistikmi dochádzalo v ďalšom období k ostrým výmenám názorov, ale aj k vzájomnému obohacovaniu znalostí.

Teória pravdepodobnosti

Ak odhliadneme od prác indických a čínskych matematikov, ktorí sa riešením kombinatorických úloh zaoberali už v staroveku, môžeme položiť základy tejto matematickej disciplíny do Európy 16. storočia.

Najstaršia práca, venovaná špeciálne problémom, ktoré dnes zahŕňame do *teórie pravdepodobnosti*, je spis Hieronyma Cardana (1501 – 1576) *De ludo aleæ*, datovaný rokom 1526. Teóriou pravdepodobnosti sa zaoberal tiež Galileo Galilei (1564 – 1642), keď v roku 1632 uskutočnil rozbor astronomických meraní pomocou teórie pravdepodobnosti.

Za skutočný počiatok teórie pravdepodobnosti je však považovaná korešpondencia, ktorú v roku 1654 viedli Blaise Pascal (1623 – 1662) a Pierre de Fermat (1601 – 1665) o problémoch, s ktorými sa na Pascala obrátil rytier de Méré. Išlo o úlohu rozdelenia stávky a úlohu o kockách, konkrétne otázku, koľkokrát treba hodiť jednou kockou alebo dvomi kockami, aby šanca, že padne aspoň jedna šestka, respektíve dve šestky, bola nadpolovičná.

Pascal bol všestranne nadaný a svoj pomerne krátky život rozhodne nepremárnil – v šestnástich rokoch publikoval prácu o kužeľosečkách, v dvadsiatich sa stal tvorcom Pascaliny – počítačového stroja, v rokoch 1648 – 1653 sa zaoberal skúmaním atmosférického tlaku, o rok neskôr už pracoval s pravdepodobnosťou. De Fermat vyštudoval právo, bol vynikajúcim znalcom klasických i živých jazykov a matematike sa venoval len zo záľuby. Dnes je považovaný za zakladateľa teórie čísel a spoluzakladateľa teórie pravdepodobnosti.

Ďalším z mužov, ktorí prispeli k zrodu novej matematickej disciplíny, bol Christian Huygens (1629 – 1695). Aj keď vzdelaním právnik, zaoberal sa prírodnými vedami: fyzikou, astronómiou a tiež matematikou. Stal sa duchovným otcom strednej hodnoty (aj keď tento pojem nepoužíval) a klasickej definície pravdepodobnosti, napriek tomu, že nehovoril o „pravdepodobnosti“, ale o „očakávanej výhre“.

Prvou známou prácou o teórii pravdepodobnosti je dielo *Ars conjectandi* od švajčiarskeho matematika Jacoba Bernoulliho (1654 – 1705). S teóriou pravdepodobnosti je ďalej spojených množstvo osobností, ako francúzsky matematik Abraham de Moivre (1667 – 1754), anglický duchovný Thomas Bayes (1702 – 1761), švajčiarsky matematik Leonhardt Euler (1707 – 1783), francúzski matematici Pierre Simon de Laplacea (1749 – 1827) a Simeon Denis Poisson (1781 – 1840), nemecký matematik Karl Friedrich Gauss (1777 – 1855) a ruskí matematici Pafnutij Lvovič Čebyšev (1821 –

2 OPIS JEDNOROZMERNÝCH ŠTATISTICKÝCH SÚBOROV

Štatistické spracovanie dát pomocou tabuliek a grafov uľahčuje ich vizuálnu analýzu a celkové posúdenie dátovej konfigurácie. Na ďalšie spracovanie však potrebujeme dátový súbor podrobnejšie opísať. Preto sa počítajú rôzne číselné, resp. opisné charakteristiky, ako sú priemerná hodnota, variabilita, šikmosť a špicatosť rozdelenia početností skúmanej číselnej premennej, ktoré vyjadrujú základné vlastnosti dát.

V tejto kapitole sa budeme podrobnejšie zaoberať najskôr charakterizovaním úrovne a menlivosti hodnôt číselného znaku v štatistickom súbore pomocou charakteristík polohy a charakteristík variability (časti 2.1 a 2.2). Časť 2.3 sa zameriava na ďalšie vlastnosti štatistického súboru, ktorými sú tvar rozdelenia z hľadiska súmernosti, resp. koncentrácie triednych početností v určitej triede alebo jej okolí. Touto problematikou sa zaoberá množstvo domácich i zahraničných publikácii autorov (pozri napr. Řezánková, Löster a Šulc, 2019; Anděl, 2019; Terek, 2017; Ross, 2017; Sodomová, 2016; Cyhelský, Kahounová a Hindls, 2010; Pacáková a kol., 2009; Hindls, Hronová a Seger, 2002; Wonnacot, 1993; Bakytová, Ugron a Kontšeková, 1975 a pod.).

2.1 Charakteristiky úrovne a polohy

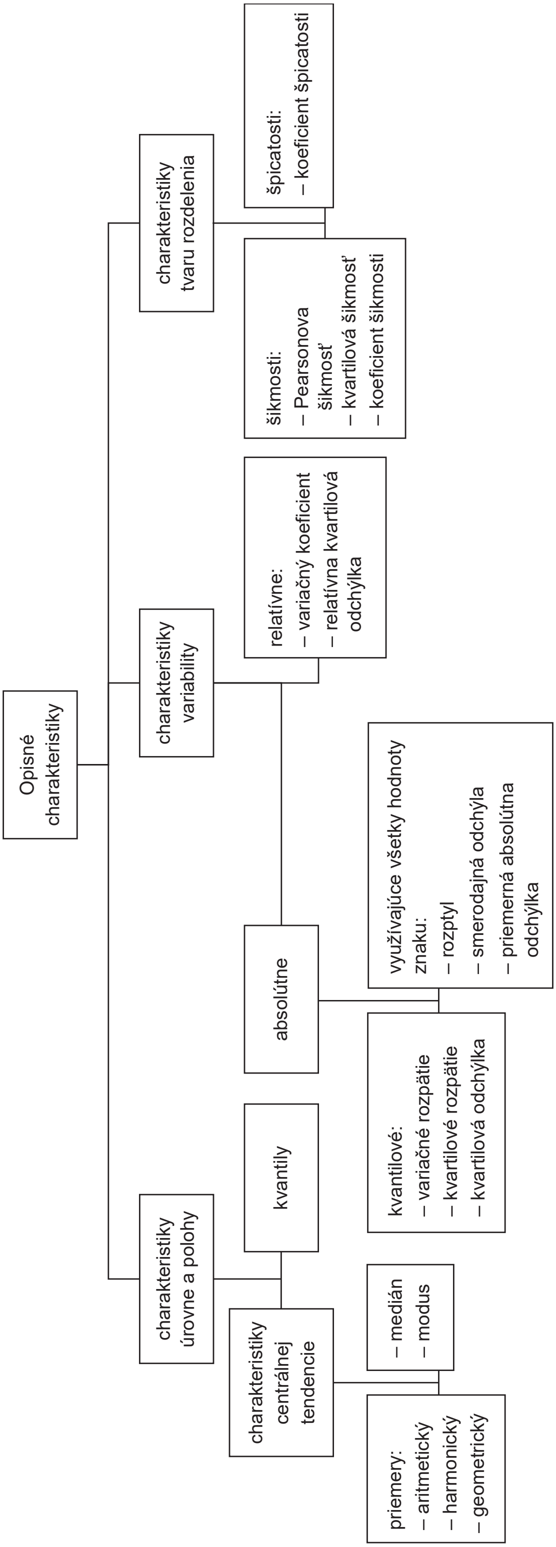
Úroveň (veľkosť, hladinu) kvantitatívnej premennej X, ktorá nadobúda hodnoty x_i, $i = 1, 2,..., n$, z ktorých jedna je minimálna x_{min} a jedna je maximálna x_{max}, charakterizuje hodnota x_{str}, pre ktorú platí:

$$x_{min} \leq x_{str} \leq x_{max}$$

Hodnota x_{str} sa zvyčajne nazýva *charakteristika centrálnej tendencie* (úrovne a polohy). Medzi najpoužívanejšie charakteristiky centrálnej tendencie patria *aritmetický priemer*, *medián* a *modus*. Rozhodnutie, ktorú z týchto charakteristík použijeme pri opise dát, závisí od cieľa a predpokladov analýzy. Prvým obmedzením je škála, v ktorej sú znaky merané. Napríklad aritmetický priemer sa používa iba za predpokladu použitia metrickej škály dát. Ak sú však dáta symetricky rozdelené, všetky tieto charakteristiky sú približne rovnaké. Uvedieme si základné zásady použitia jednotlivých charakteristík úrovne a polohy.

Aritmetický priemer je vhodné používať:

- ak ide o dáta získané v metrickej škále,
- ak je rozdelenie približne symetrické,
- ak chceme použiť štatistické testy.



Obrázok 2.1 **Klasifikácia opisných charakteristík**

Zdroj: Vlastné spracovanie.

Medián je vhodné používať:
- ak ide o dáta v ordinálnej alebo metrickej škále,
- ak chceme poznať stred rozdelenia dát,
- ak rozdelenie dát je silne zošikmené,
- ak dáta obsahujú odľahlé pozorovania.

Modus je vhodné používať:
- ak ide o dáta v nominálnej, ordinálnej alebo metrickej škále,
- ak má rozdelenie viac vrcholov,
- ak chceme získať o rozdelení iba základný prehľad,
- ak chceme poznať najčastejšie sa vyskytujúcu hodnotu.

2.1.1 Priemery

Priemer je vo všeobecnosti definovaný ako funkcia všetkých nameraných hodnôt danej premennej. Existuje viac priemerov (aritmetický, geometrický, kvadratický, harmonický, chronologický), pričom my sa najviac zameriame na aritmetický priemer, pretože jeho použitie v praxi je veľmi časté.

Aritmetický priemer

je podiel súčtu všetkých hodnôt znaku x_1, x_2,... x_n a ich počtu, teda rozsahu štatistického súboru n. Označujeme ho symbolom $\bar{x}$.

Jeho výpočet závisí od toho, či máme k dispozícií netriedené alebo triedené údaje štatistického znaku. V prípade netriedeného znaku používame *jednoduchý aritmetický priemer*, ktorý počítame podľa vzťahu:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2.1}$$

Príklad 2.1

Pri kontrole kvality určitého výrobku sme náhodne vybrali desať výrobkov a zistili sme ich životnosť. Zistené údaje (v hodinách) sú tieto: 40, 50, 20, 30, 60, 50, 60, 70, 50, 40.

Vypočítame aritmetický priemer životnosti týchto desiatich výrobkov.

Riešenie

V tomto prípade ide o netriedený štatistický znak malého rozsahu (n = 10), takže pri výpočte je možné využiť jednoduchý aritmetický priemer:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{40 + 50 + 20 + 30 + 60 + 50 + 60 + 70 + 50 + 40}{10} = 47$$

3 ■ ZÁKLADY TEÓRIE PRAVDEPODOBNOSTI

Táto kapitola obsahuje výklad teórie pravdepodobnosti v rozsahu, ktorý je nutný na pochopenie ďalších kapitol učebnice. Sústreďuje sa na tieto problémové okruhy: výklad pojmov náhodný pokus a náhodná udalosť, definície pravdepodobnosti náhodných udalostí, vysvetlenie pojmu náhodná premenná a jej zákona rozdelenia pravdepodobnosti, definíciu a vlastnosti distribučnej funkcie a funkcie hustoty, charakteristiky náhodnej premennej a rozdelenia pravdepodobnosti niektorých diskrétnych a spojitých náhodných premenných, využité v ďalších kapitolách učebnice. Výklad je prispôsobený znalostiam matematiky študentov ekonomických fakúlt vysokých škôl. Teória pravdepodobnosti v podobnom rozsahu a spôsobe výkladu je súčasťou mnohých učebníc, napríklad Hátle a Likeš (1972), Bílková, Budinský a Vohánka (2009), McClave, Benson a Sinicich (2018). Podrobnejší a matematicky exaktnejší výklad tejto matematickej disciplíny môžu záujemci nájsť napríklad v učebných textoch Horáková a Huťka, (2010), v knihách Dupač a Hušková (2005) a Anděl (2019).

3.1 Náhodné udalosti a vzťahy medzi nimi

Teória pravdepodobnosti je teoretickým základom mnohých moderných vedných disciplín. Je základom aj dôležitej časti štatistiky, a to štatistickej indukcie. Pojem pravdepodobnosti úzko súvisí s pojmom náhodného pokusu a **náhodnej udalosti**.

V prírodných vedách, napríklad vo fyzike alebo chémii, poznáme mnoho pokusov, pri ktorých realizácia určitých podmienok vedie k jednoznačnému výsledku – napríklad ručička magnetky sa v blízkosti vodiča vychýli pri zapojení prúdu, kameň vypustený z ruky padne na zem v dôsledku pôsobenia gravitačnej sily a pod. Môžeme však uviesť aj množstvo príkladov, keď výsledok pokusu nie je jednoznačný. Napríklad voda, ktorú zahrejeme na 100 °C, v závislosti od tlaku môže, ale nemusí vrieť. Ak tlak nemeriame, resp. nesledujeme, výsledok pokusu sa v tomto prípade javí ako náhodný v závislosti od pôsobenia neznámeho činiteľa.

Pri sledovaní mnohých prírodných, ale najmä spoločensko-ekonomických javov obyčajne poznáme len niektoré z množstva činiteľov, od ktorých závisí výsledok ich pôsobenia. Keďže ich vplyv nevieme predvídať, dôsledky ich pôsobenia sa javia ako náhodné. Pôsobenie neznámych činiteľov môže viesť k rôznym výsledkom sledovaného javu pri opakovaných realizáciách základných podmienok. Napríklad pri použití rovnakého osiva a pri rovnakých dávkach určitého hnojiva (základné podmienky) sa môžu hektárové výnosy na rôznych poliach líšiť v dôsledku odlišných pôdnych a poveternostných podmienok. Pri rovnakých základných podmienkach výrobného procesu môže byť rôzna kvalita vyrobených výrobkov, o nový výrobok na trhu môže byť väčší alebo menší záujem, ako na základe známych podmienok na trhu očakávame a pod.

Náhodný pokus

Realizáciu základných podmienok, ktoré nevedú k jednoznačnému výsledku, označujeme ako náhodný pokus.

V teórii pravdepodobnosti má pojem náhodného pokusu širší význam, než v akom sa obyčajne používa. Nespája sa len s činnosťami, ktoré majú výlučne experimentálny charakter. Za náhodný pokus sa považuje napríklad zhotovenie výrobku, zavedenie nového výrobku na trh, uzatvorenie poistnej zmluvy, hod hracej kocky, náhodný výber štatistickej jednotky zo základného súboru a pod. Vždy však ide o *opakovateľné pokusy*.

Náhodná udalosť

Udalosť, ktorá pri realizácii náhodného pokusu v závislosti od náhody môže, ale nemusí nastať, nazývame náhodnou udalosťou.

Pri realizácii uvedených náhodných pokusov môžeme pozorovať napríklad tieto náhodné udalosti:

- vyrobí sa prvotriedny výrobok,
- vyrobí sa nepodarok,
- o nový výrobok na trhu je mimoriadny záujem,
- na hracej kocke hodíme šesť bodov,
- na hracej kocke hodíme párne číslo,
- nastane poistná udalosť,
- zo súboru uchádzačov o zamestnanie vyberieme uchádzača so základným vzdelaním.

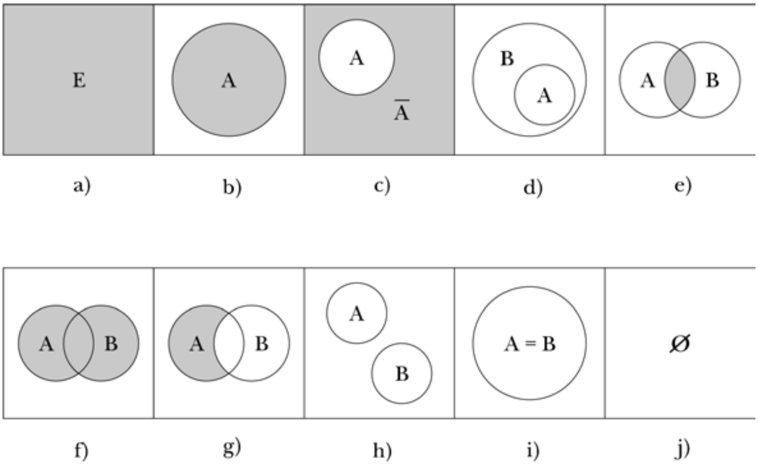
Je zrejmý rozdiel medzi náhodnou udalosťou A – *na hracej kocke hodíme šesť bodov* a náhodnou udalosťou B – *na hracej kocke hodíme párne číslo*. Udalosť A je bezprostredný, *elementárny výsledok* hodu kockou, kým pre udalosť B sú priaznivé tri možné výsledky – padnú dva body, štyri body alebo šesť bodov. Ak nastane udalosť A, teda kockou hodíme šesť bodov, určite nastane aj udalosť B. Hovoríme, že udalosť A *je časťou udalosti* B.

Elementárne náhodné udalosti

Náhodné udalosti, z ktorých pri náhodnom pokuse určite práve jedna nastane a k žiadnej z nich neexistuje náhodná udalosť, ktorá by bola jej časťou, voláme elementárne náhodné udalosti alebo elementárne výsledky náhodného pokusu.

Ak množinu všetkých elementárnych náhodných udalostí označíme E, náhodnú udalosť môžeme definovať ako ľubovoľnú podmnožinu množiny E.

Každá náhodná udalosť je podmnožinou množiny elementárnych náhodných udalostí *E*.



Obrázok 3.1 **Vzťahy medzi náhodnými udalosťami**

Zdroj: Vlastné spracovanie, Pacáková a kol. 2009.

Vzhľadom na to, že náhodná udalosť je množinou elementárnych výsledkov náhodného pokusu, sú vzťahy medzi náhodnými udalosťami analogické ako vzťahy medzi množinami, čo znázorňuje obr. 3.1.

a) Udalosť E, ktorá pri realizácii náhodného pokusu určite nastane, nazývame **istá udalosť**. Tvoria ju všetky elementárne náhodné udalosti (obr. 3.1a).
b) Každá náhodná udalosť A je nejakou podmnožinou množiny elementárnych udalostí E (obr. 3.1b).
c) Udalosť $\overline{A}$ je **opačnou udalosťou** k udalosti A. Nastane práve vtedy, ak nenastane udalosť A (obr. 3.1c).
d) Udalosť A **je časťou** udalosti B, $A \subset B$ v prípade, že ak nastane udalosť A, určite nastane aj udalosť B (obr. 3.1d).
e) Udalosť, pri ktorej nastávajú súčasne obidve udalosti A aj B, sa nazýva **prienik** alebo **súčin** a označujeme ju $A \cap B$ (obr. 3.1e).
f) Udalosť, ktorá znamená, že nastane aspoň jedna z udalostí A alebo B, nazývame **zjednotenie** alebo **súčet** udalostí A, B a značíme $A \cup B$ (obr. 3.1f).
g) **Rozdielom** udalostí A a B je taká udalosť $A - B$, ktorá nastáva vtedy, ak nastane udalosť A a súčasne nenastane udalosť B (obr. 3.1g).
h) Udalosti A a B, ktoré nemôžu súčasne nastať, voláme **nezlučiteľné** alebo **disjunktné** udalosti (obr. 3.1h).