ÚVOD

CO JE TO STATISTIKA ?

Slovo *statistika* pochází z latinského „*status*“, což znamená stav. Původně se jednalo pouze o stav nějaké země či státu a statistikou se rozuměla činnost spočívající ve zjišťování tohoto stavu. Později se pole působnosti statistiky značně rozšířilo, statistika navíc přestala být pouze praktickou činností a stala se vysoce propracovanou vědeckou naukou. Dnes tato nauka zahrnuje velmi širokou škálu kvantitativních metod umožňujících zjišťovat „stav“ věcí a poměrů v rozličných strukturách. Krom přírodních, společenských a hospodářských poměrů v daném státě lze zjišťovat např. hospodářské poměry v nějaké firmě, stav zásob v obchodním domě, stav vody na českých tocích, stav lesů v České republice apod.

Chceme-li zjistit „stav státu“, provedeme v něm např. sčítání lidu, podobně chceme-li zjistit stav zásob v obchodním domě, provedeme inventuru. Rovněž tak v lese můžeme provést inventuru, té se však v lesnické terminologii říká *inventarizace*. Tato inventarizace by mohla být v principu zaměřena na zjišťování stavu libovolných složek lesního ekosystému. Mohli bychom například zkoumat stav zvěře, lesních plodin, stromů, půdy, lesních cest apod. Středem zájmu lesníka je přitom zpravidla stav zásob dřeva. Popis stavu zásob dřeva spočívá v určení celé řady kvalitativních i kvantitativních údajů jako je např. biologický druh vyskytujících se dřevin a jejich původ, stáří stromů či jejich výška a tloušťka. Zrovna tak bychom ovšem mohli zaznamenat takové údaje jako stupeň opadu jehličí, velikost a počet šišek, kvalitu semen aj.

Zcela obecně pak při tzv. *statistických zjišťováních* zpravidla provádíme popis či měření velikého množství jistých hmotných objektů (např. výrobků, stromů či lidí), čímž získáme veliké množství údajů (většinou číselných). Těmto údajům budeme říkat *hromadná data* (učený název pro hromadu čísel a údajů). Vzhledem k tomu, že získaná data nejsou v surové podobě ničím jiným než chaotickou a neuspořádanou horou údajů, nelze z nich bez dalšího zpracování vyčíst prakticky žádné užitečné informace. A právě takové zpracování hromadných dat, které vede k odhalení informací a zákonitostí v těchto datech skrytých, je předmětem statistiky.

Je přitom rozdíl, získáme-li zpracováním hromadných dat informace pouze o konkrétních prozkoumaných objektech či odhalíme-li nějakou hlubší zákonitost, která se v datech odráží. Například bylo zjištěno, že v letech 1961–1976 se v Československu narodilo 3 910 217 dětí, z toho 2 008 391 chlapců. To je jedna z mnoha informací, které lze získat zpracováním hromadných dat zaznamenaných v matrikách. Z uvedených údajů vyplývá, že mezi narozenými dětmi bylo 51,36%

chlapců a 48,64% dívek. Tento fakt o poměrném zastoupení chlapců a dívek se týká jen a jen konkrétního souboru dětí narozených v uvedeném období. Získaný výsledek však svádí k závěru, že chlapců se obecně rodí více než dívek. To už je ovšem formulace jisté objektivní zákonitosti, která je na získaných hromadných datech ve skutečnosti nezávislá. Tuto zákonitost lze vyjádřit též tak, že narození chlapce je více pravděpodobné než narození děvčete.

Jsme ovšem k vyslovení takové zákonitosti oprávněni a jsou zaznamenaná hromadná data jejím potvrzením či důkazem? Odpověď je bohužel taková, že konkrétní hromadná data nemohou být nikdy zcela stoprocentním potvrzením žádné obecné zákonitosti, a to i tehdy je-li k dispozici značně veliké množství dat. Ono je prostě *možné*, že ve vzorku 3 910 217 dětí narozených v určitém období bude alespoň 2 008 391 chlapců, i když pravděpodobnost narození chlapce a děvčete je stejná. Ukáže-li se ovšem, že je to jen velmi *málo pravděpodobné*, pak se sice možná zmýlíme, pokud ze získaných údajů učiníme závěr o větší pravděpodobnosti narození chlapce, pravděpodobnost takového omylu bude však velmi malá. (Později pomocí prostředků teorie pravděpodobnosti ukážeme, že tato pravděpodobnost je v podstatě zanedbatelná.)

Souhrn metod pro zpracování hromadných dat vedoucích k získání přehledných informací o konkrétních objektech, jejichž popisem a měřením byla data získána, tvoří tzv. *statistiku popisnou* (*deskriptivní*). Metody zpracování hromadných dat sloužící k odhalení obecných zákonitostí, které se v hromadných datech odrážejí, jsou pak předmětem *statistiky matematické*. Metody matematické statistiky jsou přitom metody matematické, založené převážně na teorii pravděpodobnosti.