

TECHNICKÁ UNIVERZITA V KOŠICIACH

Strojnícka fakulta

# ZÁKLADY MATEMATICKEJ ŠTATISTIKY

Miriam Andrejiová

Zuzana Kimáková

Košice 2022

Technická univerzita v Košiciach, Strojnícka fakulta  
Katedra aplikovanej matematiky a informatiky

Miriám Andrejiová  
Zuzana Kimáková  
*Základy matematickej štatistiky*

*Recenzenti:*  
doc. RNDr. Denisa Olekšáková, PhD.  
RNDr. Anna Grinčová, PhD.

© Strojnícka fakulta, Technická univerzita v Košiciach  
Tlač: TYPOPRESS - tlačiareň, s.r.o., Pod horou 8, 04016 Košice  
ISBN: 978-80-553-4319-8

# Predhovor

So štatistikou je spojené nie veľmi lichotivé vyjadrenie „*Sú tri druhy klamstva: klamstvo, nehanebné klamstvo a štatistika*“. Tento výrok sa vyskytuje v rôznych obmenách a predstavuje za posledných 100 rokov najčastejšie citovaný výrok o štatistike. Zaujímavosťou je, že nie je jednoznačne určený autor tohoto výroku. Najčastejšie sa spomínajú tri mená: Benjamin Disreali (1804-1881), George Bernard Shaw (1856-1950) a Sir Charles Dilke (1843-1911). Ale aj napriek nepriaznivému vnímaniu štatistiky mnohými ľuďmi, faktom ostáva, že štatistika je neoddeliteľnou súčasťou nášho každodenného života.

Táto vysokoškolská učebnica je určená predovšetkým pre poslucháčov druhého ročníka 1. stupňa vysokoškolského štúdia na Strojníckej fakulte Technickej univerzity v Košiciach vo všetkých študijných programoch. Môže dobre poslúžiť aj študentom iných fakúlt Technickej univerzity v Košiciach.

Učebnica obsahuje základy štatistiky a teórie pravdepodobnosti: popisná štatistika, pravdepodobnosť, náhodné premenné a ich rozdelenia, metódy štatistickej indukcie, regresná a korelačná analýza. Obsahová náplň zodpovedá osnovám predmetu *Štatistické spracovanie dát*. Cieľom učebnice nie je podať vyčerpávajúce množstvo informácií, ale vybudovať dobrý základ v oblasti teórie pravdepodobnosti a metód štatistiky. Teória je doplnená riešenými a neriešenými úlohami, ktoré sú určené na precvičenie jednotlivých tém.

Pri písaní učebnice bolo snahou dodržať pravidlá jednoduchosti a zrozumiteľnosti a poukázať na užitočnosť a nenahraditeľnosť matematickej štatistiky. Po jej zvládnutí by mal čitateľ dokázať správne aplikovať základné štatistické metódy vo všetkých oblastiach technickej praxe.

Veríme, že učebnica bude prínosom a uvítame všetky pripomienky k skvalitneniu jej obsahu. Ďakujeme recenzentom za dôsledné posúdenie tejto učebnej pomôcky. Ich cenné pripomienky, rady a odporúčania prispeli ku zvýšeniu kvality publikácie.

*Autorky*

# Obsah

<b>1</b>	<b>Úvod do štatistiky</b>	<b>5</b>
1.1	Stručné dejiny štatistiky . . . . .	5
1.2	Stručné dejiny teórie pravdepodobnosti . . . . .	11
1.3	Štatistika v súčasnosti . . . . .	15
1.3.1	Štatistika ako veda . . . . .	15
1.3.2	Štatistika a počítače . . . . .	16
<b>2</b>	<b>Deskriptívna štatistika</b>	<b>19</b>
2.1	Základné pojmy . . . . .	19
2.2	Triedenie štatistického súboru . . . . .	23
2.2.1	Jednoduché triedenie . . . . .	24
2.2.2	Intervalové triedenie . . . . .	28
2.3	Grafické zobrazenie . . . . .	31

# Kapitola 1

## Úvod do štatistiky

V súčasnej dobe si len ťažko môžeme predstaviť akýkoľvek výskum bez, hoci len minimálneho využívania štatistických metód. Môže ísť o jednoduché stanovenie početností, meranie rôznych číselných charakteristík, grafické zobrazovanie, testovanie hypotéz, či o náročnejšie viacrozmerné metódy, metódy plánovania experimentov a pod. Význam štatistiky narastal v priebehu dejín ľudstva priamo úmerne s rozvojom výroby, hospodárstva a obchodu. Práve potreby zo života viedli k rozvoju štatistiky a jej metód. Dnes mnohé vedné odbory ako medicína, ekonómia, fyzika, biológia, chémia a mnohé ďalšie prírodovedné a technické odbory bežne aplikujú metódy štatistiky. Na základe štatistiky sa neustále o niečom rozhoduje a toto rozhodovanie ovplyvňuje naše konanie a náš každodenný život.

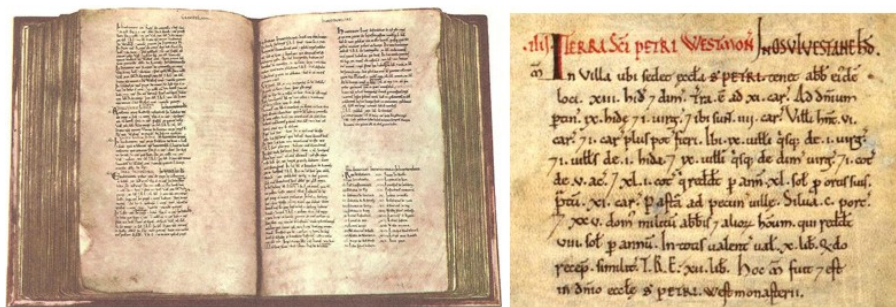
### 1.1 Stručné dejiny štatistiky

Aj napriek tomu, že počiatky metód a postupov modernej štatistiky sú z obdobia 19. a 20. storočia, počiatky štatistiky sú veľmi staré. Pôvodný význam slova *štatistika* pochádza z latinského slova *status*, čo znamená *stav* a v slovnom spojení *status rei publicae* vyjadruje *stav vecí verejnej* alebo *štát*. Pojem štatistika

K najstarším príkladom využitia štatistiky patrí sčítanie obyvateľstva a ich majetku. V starom Babylone (3800 p.n.l.), Číne (3000 p.n.l.) a v Egypte (2500 p.n.l.) sa tieto informácie využívali nielen k zisteniu počtu obyvateľov, ale na ich základe vedeli určiť počet mužov schopných vojenskej služby a predovšetkým mohli stanoviť výšku daní pre každého občana. V 1. tisícročí p.n.l. realizovali sčítanie ľudu aj starovekí Gréci a Rímania. Aténsky politik, reformátor a básnik Solon (630 p.n.l. - 560 p.n.l.) rozdelil všetkých občanov Atén do štyroch kategórií z hľadiska ročného príjmu (bez ohľadu Starovek

na ich pôvod). Rímsky kráľ Servius Tullius (578 p.n.l.-534 p.n.l.) zaviedol majetkový census (sčítanie) a rozdelil obyvateľov podľa majetku do piatich majetkových tried. Prvý rímsky cisár Octavius Augustus (63 p.n.l. - 14 n.l.) nariadil na prelome letopočtu (uvádza sa 2 p.n.l.) asi najznámejšie sčítanie ľudu, o ktorom nájdeme zmienku aj v Biblii.

Stredovek priniesol do veľkej časti Európy úpadok a negramotnosť. Až počas panovania Karola Veľkého (747-814) nastal rozvoj vzdelania, umenia a staviteľstva. Evidencie majetku šľachticov a cirkvi, rovnako aj mestské knihy obsahujúce informácie o majetku, poplatkoch a daniach obyvateľov, patria k cenným informáciám danej doby. Za začiatok oficiálnych štatistík v Anglicku sa považuje kniha *Domesday Book* (*Kniha posledného súdu*, lat. *Liber iudiciarius Angliae*) z roku 1086 (Obr. 1.1). Knihu dal spísať anglický kráľ Viliam I., známy ako Viliam Dobývateľ (1028-1087) a predstavuje najstarší súpis pozemkového majetku v Anglicku. Obsahuje počet obyvateľov a hodnotu majetku, štruktúru obyvateľstva, počet dobytku, súpis majiteľov pôdy a výmeru pôdy.



Obr. 1.1: Domesday Book a záznam z knihy [46]

V 14. storočí sa v Európe objavujú prvé cirkevné matriky, v ktorých sa evidujú všetky svadby, krstiny, resp. úmrtia. Anglický štátnik *Thomas Cromwell*, 1. gróf z Essexu (1485-1540) nariadil, aby duchovní všetkých farností v Anglicku viedli takéto cirkevné matriky od roku 1538. Záznamy neboli úplné, chýbali údaje od ľudí, resp. od farníkov, ktorí neboli ochotní zaplatiť poplatok za daný záznam.

V 16. storočí dochádza k rozvoju filozofie a moderných vedných disciplín, vznikajú mnohé diela, ktoré úzko súvisia s popisnou štatistikou. Nemecký kartograf *Sebastian Münster* (1489-1552) vytvoril v roku 1544 dielo s názvom *Cosmographia*, ktoré popisuje krajiny sveta (napr. Taliansko, Nemecko, Anglicko, Írsko, Francúzsko, Áziu, Afriku a pod.), pričom sú uvedené aj hlavné mestá, história, vládcovia, majetok, zákony a zvyky (Obr. 1.2).



Obr. 1.2: Sebastian Münster a jeho dielo Cosmographia [47]

V roku 1562, taliansky vedec *Francesco Sansovino* (1521-1586) vytvoril prvé tzv. štátovedné dielo s názvom *Del governo ed amministrazione di diversi regni ed repubbliche*, čo v preklade znamená *O vláde a správe v rôznych kráľovstvách*. Sansovini zhrnul vo svojej knihe informácie o 22 štátnych útvaroch, ktoré boli známe v danej dobe (Obr. 1.3).

*Francesco Sansovino*



Obr. 1.3: Francesco Sansovino a jeho dielo [48]

V 17. storočí vzniká v Anglicku tzv. politická aritmetika, ktorej východiskom boli údaje o narodeniach a úmrtiach obyvateľov. Informácie z veľkého počtu matrik v Anglicku a Walesu sa pokúsil spracovať *John Graunt* (1620-1674), ktorý je považovaný za zakladateľa demografie. Vo svojom diele *Natural and Political Observations Made Upon the Bills of Mortality* (*Prírodné a politické postrehy nad londýnskymi súpismi úmrtí*) z roku 1622 pracoval s údajmi o pohlaví, veku a príčine úmrtia (Obr. 1.4). Skúmal pomer medzi počtom mužov a žien v populácii a odhalil stabilný pomer medzi počtom narodených chlapcov a dievčat. Na základe údajov z krstných a úmrtných listov obyvateľov Londýna zostavil prvé úmrtnostné tabuľky <sup>1</sup>. Graunt ve-

*John Graunt*

<sup>1</sup>Úmrtnostná tabuľka poskytuje predovšetkým základné údaje, ktoré vo forme pravdepodobnosti úmrtia alebo strednej dĺžky ďalšieho života popisujú správanie priemerného jedinca danej vekovej kategórie.





*Europäischen Reiche und Völker im Grundrisse* predstavuje predchodcu súčasných štatistických ročeniek štátov (Obr. 1.6).



Obr. 1.6: Gottfried Achenwall a jeho dielo [51]

Anglický ekonóm a autor populačnej teórie Thomas Robert Malthus (1776-1834) v svojom diele *An Essay on the Principle of Population* (*Esej o princípe populácie*) z roku 1798 tvrdil, že počet obyvateľov rastie geometricky a množstvo prostriedkov živobytia lineárne a to môže viesť k sociálno-ekonomickému kolapsu a k večnej chudobe. Malthus bol presvedčený, že veľkosť svetovej populácie musí byť regulovaná chudobou, morom a vojnami.

Thomas R.  
Malthus



Obr. 1.7: T.R. Malthus



Obr. 1.8: A.J.L. Quetelet



Obr. 1.9: W. Farr

Významnou osobnosťou štatistiky bol belgický astronóm, matematik, štatistik a sociológ *Adolphe Lambert Jacques Quetelet* (1796-1874), ktorý stál pri vzniku Londýnskej štatistickej spoločnosti (The Royal Statistical Society) v roku 1834. Quetelet vypracoval zásady moderného sčítania obyvateľstva a zohral dôležitú úlohu pri zavádzaní štatistických metód v sociálnej oblasti. Na základe veľkého súboru údajov, ktoré získal pri sčítaní

Adolphe L.J.  
Quetelet

belgického obyvateľstva, vypočítal priemerného (ideálneho) človeka (*homme moyen*) a vytvoril základ pre priemer, strednú hodnotu a rozptyl.

William  
Farr

*William Farr* (1807-1883) bol britský epidemiológ, ktorý bol považovaný za jedného zo zakladateľov lekárskej štatistiky. Spolu s kolegom, poistným matematikom a ekonómom *Thomasom Rowe Edmundsonom* (1803-1889) zohrali významnú rolu pri určovaní úmrtnosti ako prostriedku na hodnotenie zdravia ľudskej populácie. Kľúčové pre ich inovácie bolo použitie zákona smrteľnosti, o ktorom Edmonds tvrdil, že ho objavil v roku 1832.

Florence  
Nightingale

K priekopníkom lekárskej štatistiky patrí aj anglická ošetrovateľka šľachtického pôvodu *Florence Nightingale* (1820-1910), známa ako *Dáma s lampou*, ktorej zásluhou sa ošetrovanie nemocných stalo uznávaným povoláním. Za pomoci matematiky a metód štatistiky vypracovala štatistickú analýzu chorôb a úmrtnosti v armáde počas Krymskej vojny (v Turecku ošetrovala ranených). Zaviedla reformu v armádnych nemocniciach a za pol roka sa vďaka jej opatreniam znížila úmrtnosť vojakov vo vojenských nemocniciach zo šesťdesiat na dve percentá. Pod jej vedením britská vláda zaviedla systematické zhromažďovanie údajov o počte obyvateľov. V roku 1859 napísala prvú učebnicu moderného ošetrovateľstva a o rok neskôr založila prvú školu pre ošetrovateľky na svete. Pri prezentovaní štatistických informácií používala grafy, pričom vytvorila svoj polárny plošný graf, ktorý ukazoval zmeny v priebehu času (Obr. 1.10).



Obr. 1.10: Florence Nightingale a Diagram príčin [52]

Prelom  
19.-20.stor.

Až do konca 19. storočia je pre štatistiku charakteristické skúmanie a popisovanie hromadných javov, čo v súčasnosti označujeme ako popisná štatistika. Na prelome 19. a 20. storočia vznikajú diela, ktoré tvoria základ modernej štatistiky. Medzi významných autorov patria napríklad *Sir Francis Galton* (1822-1911), *Karl Pearson* (1857-1936), *George Udny Yule* (1871-1951), *Ronald Aylmer Fisher* (1890-1962), *William Sealy Gosset* (1876-1937), *Andrej Nikolajevič Kolmogorov* (1903-1987) a iní. Ich práce a poznatky zna-

menali postupný prechod od popisnej štatistiky k modernejším metódam štatistiky.

V 30-tych rokoch 20. storočia prichádza zmena vo vývoji štatistiky a vzniká indukčná (analytická) štatistika. Objavujú sa metódy, ktoré umožňujú na základe výberu zovšeobecňovať a tvoriť závery o celku. Po druhej svetovej vojne nastal obrovský rozvoj nových štatistických metód a prístupov, ktoré sa vo veľkej miere uplatňujú v mnohých vedných odboroch.

## 1.2 Stručné dejiny teórie pravdepodobnosti

Veľký vplyv na vývoj štatistiky mal vznik teórie pravdepodobnosti. Historickým zdrojom úvah o pravdepodobnosti boli hazardné hry, pričom k najstarším hrám patria hracie kocky.

Archeologické nálezy v severnom Iraku ukazujú, že ľudia tieto hry hrali už v starovekej Mezopotámii na konci 4. tisícročia p.n.l. Najstaršie šesťhranové hracie kocky, podobné dnešným hracím kockám, sú z oblasti Shahr-e Sukhte v juhovýchodnom Iráne. Nálezisko pochádza z doby bronzovej, približne okolo roku 2800-2500 p.n.l. (Obr. 1.11a). Kocky patrili k obľúbeným hazardným hrám aj v starovekom Ríme. Boli vyrobené z rôznych materiálov (napr. drevo, uhlie, kosť, slonovina, meď, mramor) a často nemali pravidelný tvar a veľkosť. Na Obr.1.11b sú kocky z obdobia starovekého Ríma (100-300 rokov n.l.).

*Hracie  
kocky*



Obr. 1.11: Staroveké hracie kocky [53, 54]

Zárodky kombinatoriky môžeme nájsť v Ázii, pričom k najstarším prameňom patria sútry<sup>2</sup>. Bhagabati Sútra (3. storočie p.n.l.) používa kombinatoriku k riešeniu problémov typu *Aký je počet zmyslových podskupín zostaviteľných z piatich zmyslov? Aké podskupiny môžeme vytvoriť z daného počtu mužov a žien?* V Európe sa kombinatorickým úvahám venoval biskup Wibold z Cambray, ktorý okolo roku 960 n.l. vymyslel hru, v ktorej mníši hádzali troma kockami. Biskup vypočítal a očísloval všetkých 56 možných

*Začiatky  
kombina-  
toriky*

<sup>2</sup>Sútra predstavuje krátke pravidlo.

kombinácií s opakovaním, pričom každej priradil istú cnosť. Mních potom pri danej kombinácii musel danú cnosť dodržať 24 hodín. Kombinatorika a kombinatorické pravidlá boli známe v čínskej, resp. arabskej matematike 12. a 13. storočia. Ale aj napriek tomu, že matematici už v staroveku riešili kombinatorické úlohy, tak základy teórie pravdepodobnosti začínajú až v 16. storočí.

V raných počiatkoch teórie pravdepodobnosti existovali dva druhy problémov. Prvý problém súvisel s otázkou *Koľkými spôsobmi môže padnúť istý počet bodiek pri hádzaní dvoma, troma a viacerými kockami?* Druhý problém súvisel s úlohou *Ako spravodlivo rozdeliť stávkku?* V tej úlohe dvaja hráči hrajú sériu hier o istú peňažnú čiastku, t. j. výhru. Túto výhru získa ten hráč, ktorý ako prvý dosiahne počet dopredu určených víťazstiev. Pravdepodobnosť výhry v každej hre je pre obidvoch hráčov rovnaká. Séria hier je predčasne ukončená. Ako má byť spravodlivo rozdelená výhra medzi obidvoch hráčov v závislosti od toho, koľko hier už každý z nich vyhral? Za autora úlohy je považovaný františkánsky mních talianskeho pôvodu *Luca Pacioli* (1447-1517), ktorý v svojom diele *Summa de arithmetica, geometria, proportioni et proportionalita* (*Systematický prehľad aritmetiky, geometrie, pomerov a úmernosti*, 1494) uvádza aj nesprávne riešenie tejto úlohy (Obr. 1.12). Úspešnými riešiteľmi úlohy o rozdelení stávky boli Blaise Pascal a Pierre de Fermat.

*Luca  
Pacioli*



Obr. 1.12: Luca Pacioli a jeho dielo [55]

*Girolamo  
Cardano*

Taliansky lekár a matematik *Girolamo Cardano* (1501-1576), ktorý bol vášnivým hazardným hráčom, napísal v roku 1564 dielo *Liber de Ludo Aleae* (*Kniha o náhodných hrách*). Kniha bola publikovaná až po jeho smrti v roku 1633 a slúžila ako príručka pre hazardných hráčov.

*Začiatok  
teórie  
pravdepodo-  
bnosti*

Za začiatok teórie pravdepodobnosti sa vo všeobecnosti považuje korešpondencia, ktorú v roku 1654 viedli *Blaise Pascal* a zakladateľ teórie čísel a priekopník analytickej geometrie *Pierre de Fermat*. Ich korešpondencia súvisela s problémom rytiera de Méré (1607-1685), ktorý chcel od Pascala zistiť, ako spravodlivo rozdeliť stávkku pri nedokončenej hazardnej hre (úloha

s rozdelením stávky).

Francúzsky matematik *Pierre de Fermat* (1607-1665) sa zaslúžil o rozvoj matematiky v oblasti teórie čísel (známa Veľká Fermatova veta), v teórii pravdepodobnosti, v matematickej analýze a analytickej geometrii.

*Pierre  
de Fermat*

Základné poznatky z teórie pravdepodobnosti a kombinatoriky vyslovil francúzsky matematik, fyzik a filozof *Blaise Pascal* (1623-1662) v svojom diele *Traité du triangle arithmétique* (Pojednanie o aritmetickom trojuholníku) z roku 1654. V aritmetike sa stal známy jeho Pascalov trojuholník a je považovaný za predchodcu diferenciálneho a integrálneho počtu.

*Blaise  
Pascal*



Obr. 1.13: G. Cardano



Obr. 1.14: P. de Fermat



Obr. 1.15: B. Pascal

Ďalším mužom, ktorý významne ovplyvnil formovanie teórie pravdepodobnosti, bol holandský fyzik, matematik a astronóm *Christiaan Huygens* (1629-1695), ktorý v roku 1656 vydal historicky prvý spis venovaný úlohám z teórie pravdepodobnosti *De ratiociniis in ludo aleae* (O počítaní pri hre s kockou). Spis je členený do troch základných častí, pričom vo svojom diele nepoužíva pojem pravdepodobnosť, ale hovorí o očakávanej výhre. Huygens sa zaoberá otázkou, akú cenu má pre hráča stávka predtým, ako sa dozvie výsledok hry a tak vzniká celá teória o náhodných veličinách.

*Christiaan  
Huygens*

Významný vplyv na rozvoji teórie pravdepodobnosti mal švajčiarsky matematik, fyzik a astronóm *Jacob Bernoulli* (1654-1705). V období 1679 až 1685 napísal spis *Ars Conjectandi* (Umenie predpokladať), v ktorom vo štvrtej časti po prvýkrát zavádza a používa pojem pravdepodobnosť. Spis bol vydaný až po jeho smrti v roku 1713 a veľkým prínosom je jeho štvrtá, nedokončená časť, v ktorej sa nachádza prvá formulácia a dôkaz zákona veľkých čísel.

*Jacob  
Bernoulli*

Francúzsky matematik *Abraham de Moivre* (1667-1754) je známy predovšetkým de Moivreovou vetou, podľa ktorej môžeme umocňovať komplexné čísla v goniometrickom tvare. V roku 1718 napísal dielo *The Doctrine of Chances: or A Method of Calculating the Probabilities of Events in Play*

*Abraham  
de Moivre*



(*Doktrína šance alebo metóda výpočtu pravdepodobnosti možných výsledkov v hre*), ktorá je považovaná za prvú učebnicu teórie pravdepodobnosti zameranú na analýzu náhodných javov. V svojej učebnici pokračoval v práci Christiaana Huygensa a Jacoba Bernoulliho.



Obr. 1.16: Ch. Hyugens



Obr. 1.17: J. Bernoulli



Obr. 1.18: A. de Moivre

Thomas  
Bayes

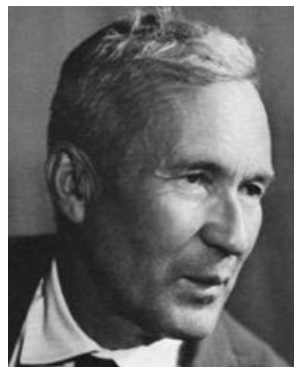
*Thomas Bayes* (1702-1761) bol anglický teológ a matematik, ktorý ako prvý použil indukčnú pravdepodobnosť. Počas svojho života skoro nič nepublikoval, jeho práca bola vydaná až tri roky po smrti a to zásluhou jeho priateľa Richarda Pricea v časopise Royal Society pod názvom *An Essay towards solving a Problem in the Doctrine of Chances* (*Esej na riešenie problému v doktríne šanci*). V svojej práci zavádza základné pojmy a vzťahy súvisiace s podmienenou pravdepodobnosťou a nezávislými javmi, ktoré sú zhrnuté v tzv. Bayesovej vete (resp. Bayesovo pravidlo). Zaujímavosťou je to, že autorom Bayesovho vzorca v takej podobe ako ho poznáme v súčasnosti, je Pierre-Simon de Laplace.



Obr. 1.19: T. Bayes



Obr. 1.20: P.S. de Laplace



Obr. 1.21: A.N. Kolmogorov

K významným osobnostiam vedy na prelome 18. a 19. storočia patrí francúzsky matematik, fyzik, astronóm a politik *Pierre-Simon de Laplace* (1749-1827), ktorý vo veľkej miere ovplyvnil vývoj teórie pravdepodobnosti. Ako prvý sformuloval klasickú definíciu pravdepodobnosti a centrálnu limitnú vetu (1810). Laplace, podobne ako Bayes, sa zaoberal aj podmienenou pravdepodobnosťou a svoju prácu publikoval v roku 1774. V tej dobe ešte nepoznal Bayesove výsledky, dozvedel sa ich až v roku 1781. V roku 1812 vydal dielo *Théorie Analytique des Probabilités* (*Analytická teória pravdepodobnosti*), kde pri budovaní teórie využíva diferenciálny a integrálny počet.

*Pierre-S.  
de Laplace*

K ďalším autorom spojených s teóriou pravdepodobnosti patrí napr. švajčiarsky matematik *Leonhard Euler* (1707-1783), *Simeon Denis Poisson* (1781-1840), ruský matematik *Pafnutij L. Čebyšev* (1821-1894) a iní.

V 20. storočí má významné miesto v budovaní teórie pravdepodobnosti ruský matematik *Andrej Nikolajevič Kolmogorov* (1903-1987), ktorý pracoval aj v oblasti geometrie, matematickej logiky, teórie množín, funkcionálnej analýzy a pod. V roku 1933 vydal knihu *Základy teórie pravdepodobnosti* (ang. preklad *Foundations of Probability Theory*, 1950), pričom definíciu pravdepodobnosti založil na teórii množín a na troch základných axiómach. Svojím dielom položil základy modernej teórie pravdepodobnosti.

*Andrej N.  
Kolmogorov*

## 1.3 Štatistika v súčasnosti

Pojem štatistika získal všeobecný význam zberu a analýzy údajov až začiatkom 19. storočia. Dnes v užšom slova zmysle pod pojmom štatistika rozumieme zvyčajne číselný údaj o nejakej skutočnosti, resp. súhrnné údaje o hromadných javoch. V širšom slova zmysle rozumieme vednú disciplínu, ktorá skúma zákonitosti hromadných javov.

### 1.3.1 Štatistika ako veda

Existuje mnoho definícií pojmu štatistika. Podľa Karla Pearsona (1857-1936), *štatistika je vedecká disciplína zaoberajúca sa zberom, analýzou a interpretáciou získaných údajov z pozorovania alebo z experimentu*. Americký vedec Warren Weaver (1894-1978) definuje *štatistiku ako vedu a umenie, ktoré sa zaoberajú neistými závermi a ktoré používajú čísla, aby ste zistili niečo o prírode a skúsenostiach*. Autor populárnych učebníc Montgomery D.C. uvádza, že *štatistika je veda, ktorá nám pomáha robiť rozhodnutia a vyvodit závery za prítomnosti variability* [Montgomery, 2011]. Podľa Americkej štatistickej asociácie je *štatistika definovaná ako veda založená na učení z dát a na zisťovaní, kontrolovaní a prezentovaní neurčitosti*. Podľa mnohých autorov, *štatistika zahŕňa štúdium objektívnych vlastností skúmaných*

*Štatistika*

*javov a procesov a ich skúmaním umožňuje formulovať všeobecné závery.*

V súčasnej dobe sa štatistika chápe ako *veda zaoberajúca sa zberom, analýzou, interpretáciou a prezentáciou dát získaných z pozorovaní alebo experimentov, ktorá umožňuje prijímať rozhodnutia a formulovať všeobecné závery*. Patrí k integrujúcim vedám, ktoré dávajú iným vedám spoločné metodologické nástroje a postupy.

*Hromadné  
javy*

*Predmetom štatistiky ako vednej disciplíny sú hromadné náhodné javy. Hromadným javom rozumieme jav, ktorý sa za presne definovaných vecných, časových a priestorových podmienok viackrát vyskytuje, resp. opakuje. Hromadným javom môže byť jav, ktorý sa vyskytuje u veľkého počtu objektov (napr. výška žiaka, vek študenta, vzdelanie obyvateľov, počet detí v rodinách a pod.) alebo jav, ktorý je výsledkom veľkého počtu opakovaných pozorovaní istej vlastnosti jedného objektu/prvku (napr. opakované hádzanie hracou kockou, opakované meranie koncentrácie látky v tej istej vzorke roztoku). Skúmaním hromadného javu poznáme podstatu javu, pravidelnosti, vzťahy a jeho vývoj. Bez hromadného pozorovania by sme nemohli o príslušnom jave robiť zovšeobecňujúce závery.*

*Matematická  
štatistika*

*Matematická štatistika je samostatná vedná disciplína, ktorá vychádza z teórie pravdepodobnosti a nachádza sa na rozhraní deskriptívnej štatistiky a odvetví aplikovanej matematiky (napr. lineárna algebra, matematická analýza). Podstatnou zložkou matematickej štatistiky je skúmanie hromadných javov a ich pravidelností, na základe ktorých je potom možné vykonávať úsudky z časti na celok, zo zvláštneho na všeobecné. Môžeme povedať, že základnou metódou matematickej štatistiky je štatistická indukcia, pričom hlavnými nástrojmi sú teória odhadu, testovanie štatistických hypotéz.*

### 1.3.2 Štatistika a počítače

*Štatistický  
softvér*

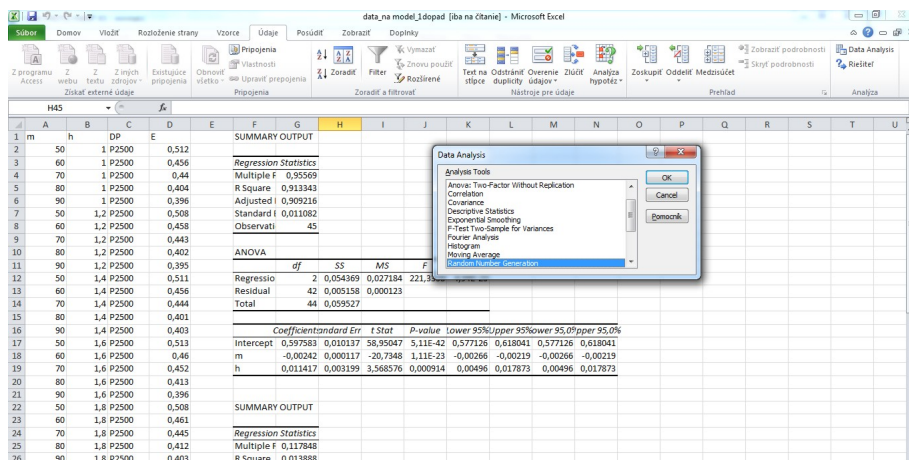
Prístup k štatistike sa významne zmenil s dostupnosťou počítačov. Čoraz častejšie sa na štatistické spracovanie a komplexnú analýzu výskumu využívajú počítače a štatistické softvéry. K nepopierateľným výhodám ich použitia patrí rýchlosť a presnosť spracovania údajov, univerzálnosť, grafické výstupy, množstvo spracovávaných údajov a pod. Ponuka štatistických softvérov je veľmi bohatá. K najznámejším komerčným štatistickým softvérom patrí napr. Statistica, SAS, SPSS. V rozšírenej miere sa na štatistické výpočty používajú aj programy ako Matlab, Maple, Mathematica, Minitab, tabuľkový program Excel alebo voľne šíriteľné programy (napr. R program, Maxima, Octave).

V nasledujúcej časti uvedieme dva programy, ktoré patria k najdostupnejším a populárnym nástrojom štatistiky.



Excel je tabuľkový procesor od firmy Microsoft, ktorý pracuje pod operačným systémom Microsoft Windows a je súčasťou balíka Microsoft Office. Excel má k dispozícii množstvo funkcií, ktoré sú rozdelené do niekoľkých kategórií: matematické, štatistické, logické, vyhľadávacie, finančné, databázové, dátumové, časové a textové a pod. Aj napriek tomu, že to nie je štatistický program, ponúka okrem práce s tabuľkami aj grafické výstupy a základné štatistické spracovanie dát (Obr. 1.22).

Excel



Obr. 1.22: Prostredie EXCEL

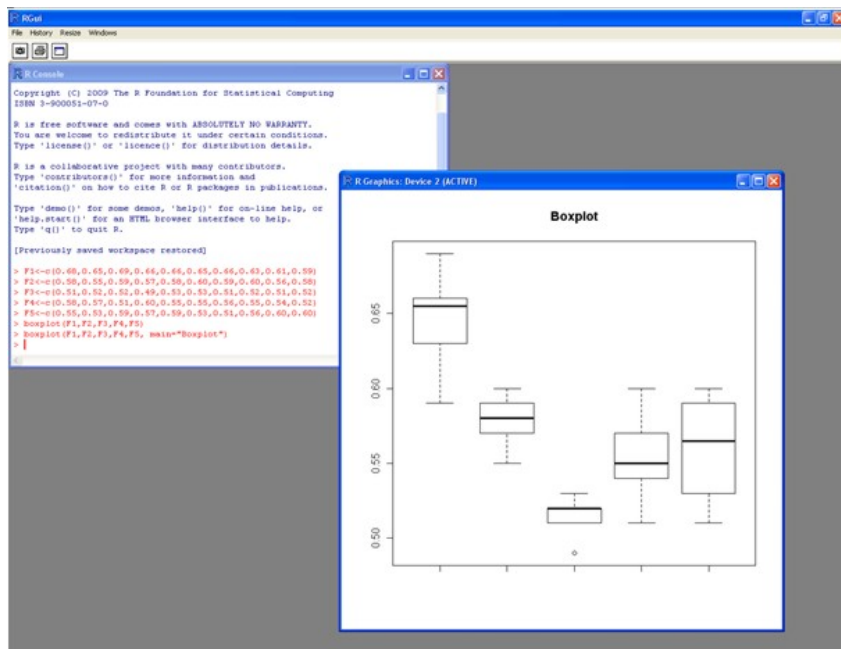
Program R je jazyk a zároveň prostredie umožňujúce manipuláciu, numerické spracovanie, vyhodnotenie a grafické zobrazenie dát. Ide o GNU projekt (od 1997 je jeho súčasťou), ktorý je podobný jazyku a prostrediu S a bol vyvinutý v Bell Laboratories Chambersom J., Beckerom R. a Willkсом A. v roku 1995. Je to voľne šíriteľný nástroj, ktorý je možné používať v Microsoft Windows a tiež v systéme Linux a MacOS X (Obr. 1.23).

Program  
R

Prostredie R poskytuje širokú škálu štatistických metód na analýzu a vyhodnotenie údajov. Priamo v programe R je balík stats, ktorý je implementovaný v základnej ponuke programu. Balík obsahuje viac ako 550 funkcií z rôznych oblastí štatistiky: popisná štatistika, grafické zobrazovanie, testovanie hypotéz, korelačná a regresná analýza, časové rady, zhluková analýza a iné.

Jeho silnou stránkou sú kvalitné obrázky a grafy. Súčasťou základného prostredia je niekoľko štandardných, odporúčaných balíkov (ang. Packages), mnohé ďalšie sa dajú inštalovať a zároveň aj aktualizovať z archívu CRAN na domovskej stránke. Toto oficiálne úložisko nedávno zverejnilo viac ako 16 000 balíkov a mnohé ďalšie sú dostupné prostredníctvom internetu. Výhodou programu R je, že patrí k programom s otvoreným zdrojovým kódom, tzv. OSS (ang. Open Source Software).

Prostredie R sa neustále vyvíja a jeho aktuálnu verziu je možné stiahnuť z domovskej stránky <http://www.r-project.org/>.



Obr. 1.23: Prostredie R

Na záver chceme pripomenúť, že počítače a príslušný softvér nám síce uľahčia prácu a takmer bez problémov spracujú a vyhodnotia dáta, ale bez poznania podstaty danej štatistickej metódy, nie je možný výber vhodnej metódy a rovnako ani analýza získaných výsledkov.

## Kapitola 2

# Deskriptívna štatistika

*Deskriptívna (popisná) štatistika sa zaoberá zisťovaním, spracovaním, analyzovaním a prezentáciou dát pomocou štatistických prostriedkov a metód. Jej úlohou je vhodným spôsobom popísať a zosumarizovať získané údaje. Tvorí základ pre ďalšiu analýzu dát.*

*Deskrip-  
tívna  
štatistika*

### 2.1 Základné pojmy

K základným pojmom popisnej štatistiky patrí hromadný jav, štatistická jednotka, štatistický súbor, rozsah štatistického súboru a štatistický znak.

Pod hromadným javom rozumieme jav, ktorý sa vyskytuje za určitých podmienok opakovane vo veľkom počte a môžeme ho pozorovať alebo získať experimentom (Kapitola 1.3.1). Hromadným javom môže byť aj opakované sledovanie životnosti výrobného nástroja, resp. zisťovanie priemerného veku obyvateľov v krajine a pod. Hromadný jav nadobúda mnoho konkrétnych foriem a skladá sa z mnohých individuálnych javov, ktorých nositeľmi sú štatistické jednotky.

*Hromadný  
jav*

*Štatistická jednotka* je základný prvok (napr. osoba, výrobok, jav, meraná veličina a pod.), ktorý je predmetom skúmania a na ktorom pozorujeme konkrétny prejav určitého hromadného javu. Pri zisťovaní výsledkov zo skúšky z istého predmetu budú štatistickými jednotkami študenti, ktorí sa zúčastnili záverečnej skúšky z daného predmetu.

*Štatistická  
jednotka*

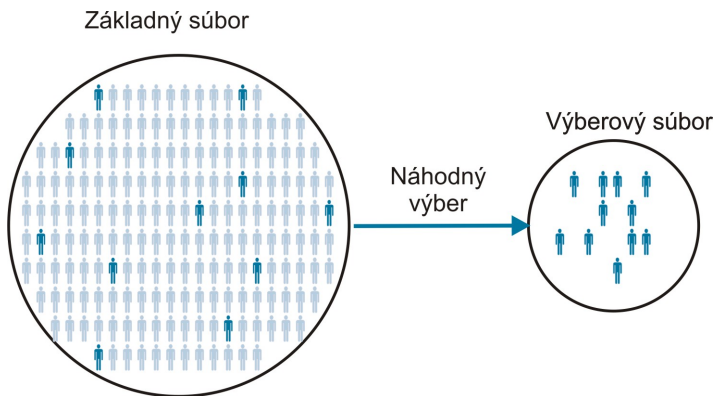
Množinu všetkých štatistických jednotiek, ktoré majú požadované spoločné vlastnosti, nazývame *štatistický súbor*. *Rozsah štatistického súboru* charakterizuje počet štatistických jednotiek v súbore.

*Štatistický  
súbor,  
rozsah*

Súbor všetkých objektov (prvkov), ktoré sú predmetom nášho skúmania, nazývame *základný súbor (populácia)*. V mnohých prípadoch je základný sú-

*Populácia/  
Výber*

bor príliš rozsiahly na preskúmanie všetkých prvkov. V takýchto prípadoch sa používa *výberový súbor* (*výber*, *vzorka*), ktorý je ľubovoľnou podmnožinou základného súboru (Obr. 2.1). O náhodnom výbere hovoríme vtedy, ak o zaradení štatistickej jednotky do výberu rozhoduje len náhoda.



Obr. 2.1: Základný a výberový súbor

*Štatistický  
znak*

*Štatistický znak* predstavuje vlastnosť skúmaného hromadného javu, t.z. je to vlastnosť štatistickej jednotky, ktorá je predmetom štatistického skúmania (hmotnosť súčiastky, životnosť zariadenia, národnosť, študijný program a pod.). Štatistický znak zvyčajne nadobúda v štatistickom súbore niekoľko hodnôt, resp. obmien, preto sa označuje ako premenlivý štatistický znak, resp. *premenná*. Štatistický znak označujeme veľkými písmenami  $X, Y, \dots$  a ich konkrétne hodnoty symbolmi  $x_1, x_2, \dots$  alebo  $y_1, y_2, \dots$ .

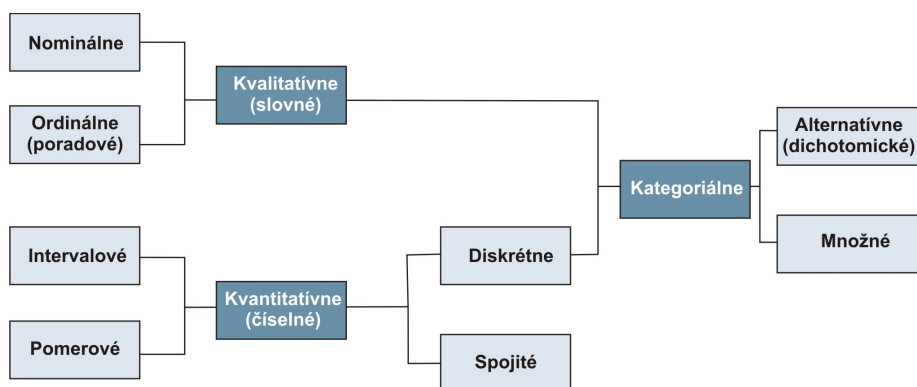
*Druhy  
znakov*

Štatistické znaky môžeme deliť z rôznych hľadísk. Základné delenie rozlišuje dva typy znakov, a to kvalitatívne a kvantitatívne znaky. *Kvalitatívne* (*slovné*, *kategoriálne*) znaky slovné vyjadrujú vlastnosti štatistickej jednotky (napr. kvalita výrobku, spokojnosť zákazníka). *Kvantitatívne* (*číselné*, *numerické*, *merateľné*) znaky číselne vyjadrujú merateľné vlastnosti štatistických jednotiek. Tieto znaky môžeme ďalej rozdeliť na spojité a diskrétne znaky. *Spojité znaky* môžu nadobúdať ľubovoľné hodnoty z ohraničeného alebo neohraničeného intervalu (napr. rozmer súčiastky, pevnosť materiálu). *Diskrétné znaky* môžu nadobúdať len niektoré konkrétne hodnoty (napr. počet kvalitných výrobkov, počet nehôd). Podrobnejšie delenie štatistických znakov je na Obr. 2.2.

*Nominálne  
znaky*

*Nominálne znaky* zodpovedajú kvalitatívnym znakom, ktoré nie je možné navzájom usporiadať. Príkladom môže byť pohlavie respondenta v dotazníkovom výskume (muž, žena), príslušnosť študenta k študijnému programu a pod. O dvoch hodnotách nominálneho znaku vieme iba rozhodnúť, či sa

rovnajú alebo sú rôzne. Hodnoty nominálnych znakov sú zvyčajne vyjadrené slovne, čísla sa používajú ako symbol alebo kód pre jednoduchšie spracovanie a môžu byť ľubovoľné (napr. klasifikácia pracovných úrazov).



Obr. 2.2: Delenie znakov

*Ordinálne (poradové) znaky* predstavujú znaky, ktoré nadobúdajú diskrétne hodnoty a ktoré môžeme jednoznačne usporiadať. Sem patrí napríklad stupeň spokojnosti zákazníka (veľmi spokojný, spokojný, nespokojný, veľmi nespokojný), tepelný pocit zamestnanca na pracovisku (teplé, mierne teplé, neutrálné, mierne chladno, chladno) a pod. Jednotlivé kategórie ordinálnych znakov sú nezlučiteľné a môžeme ich vzájomne usporiadať, ale nevieme určiť o koľko je jedna hodnota väčšia ako druhá. Ordinálne premenné môžeme vyjadriť nielen slovne, ale aj číselne pomocou bodovacej škály.

*Ordinálne  
znaky*

*Intervalové (rozdielové) znaky* sú také znaky, ktorých hodnoty vieme nielen zoradiť, ale vieme určiť o koľko je väčšia, resp. menšia prvá hodnota ako druhá hodnota (napr. Celziova teplota, IQ). Pri týchto znakoch (môžu byť spojité i diskrétne) nevieme určiť, koľkokrát je jedna hodnota väčšia/menšia ako druhá, pretože intervalový znak nemá nulu (nulový bod sa určí dohodou). Príkladom intervalového znaku je aj rok narodenia, pričom dohodnutá nula je začiatok letopočtu.

*Intervalové  
znaky*

*Pomerové (podielové) znaky* sú podobné ako intervalové znaky a opäť môžu byť spojité alebo diskrétne. Od intervalových znakov sa líšia tým, že majú prirodzenú nulu (tzv. absolútna nula), ktorá je pevne daná a má zmysel (napr. vek človeka – absolútna nula je 0 rokov, vzdialenosť od objektu – absolútna nula je 0 m). Zatiaľ čo intervalové znaky môžu byť aj záporné, tak pomerové znaky môžu nadobúdať len kladné hodnoty. Ich hodnoty vieme usporiadať a vieme určiť aj rozdiely medzi hodnotami. Pri pomerových zna-

*Pomerové  
znaky*

koch môžeme zistiť aj to, koľkokrát je jedna hodnota väčšia (resp. menšia) než druhá. Existencia absolútnej nuly umožňuje počítať aj pomery hodnôt (napr. človek, ktorý váži 90 kg, má o 30 kg vyššiu hmotnosť ako človek, ktorý váži 60 kg a v porovnaní s ním je jeden a pol krát ťažší).

Nominálne a ordinálne znaky sú označované ako *kvalitatívne*, intervalové a pomerové znaky ako *kvantitatívne* (merateľné) znaky. V Tab. 2.1 sú uvedené informácie, ktoré získame z rôznych typov štatistických znakov. V Tab. 2.2 sú príklady rôznych typov štatistických znakov.

Tabuľka 2.1: Štatistické znaky – získané informácie

Získané informácie	Kvalitatívne znaky		Kvantitatívne znaky	
	Nominálne	Ordinálne	Intervalové	Pomerové
Otázka	Rovnajú sa?	Väčší, menší?	O koľko?	Koľkokrát?

*Dichotomické/  
Množné  
znaky*

Špeciálnu skupinu kvalitatívnych znakov tvoria tzv. *alternatívne (dichotomické) znaky*, ktoré nadobúdajú len dve možné kategórie (napr. kvalitný a nekvalitný výrobok, ukončené štúdium a neukončené štúdium). *Množné (viackategoriálne) znaky* sú také znaky, ktoré nadobúdajú viac ako dve kategórie (napr. nezávažné, menej závažné, závažné poškodenie výrobku).

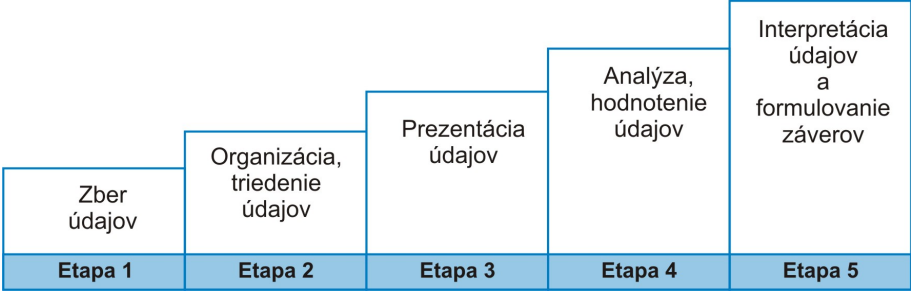
Tabuľka 2.2: Príklady rôznych štatistických znakov

Znak	Príklady
Nominálny	farba očí, krvná skupina, náboženstvo, rodinný stav, pohlavie, národnosť, študijný program, typ strednej školy
Ordinálny	prospech v škole, veľkosť oblečenia, stupeň vzdelania, spokojnosť so službami, stupeň bolesti, stupeň poškodenia
Intervalový	Celziova teplota, inteligenčný kvocient, rok narodenia, časový interval na hodinách
Pomerový	teplota v Kelvinoch, vek človeka (roky), počet gólov, počet detí, hmotnosť (v kg), výška (v cm), čas (v s), krvný tlak (v mmHg), príjem v domácnosti (v €)
Dichotomický	účasť/neúčasť na cvičení, kvalitný/nekvalitný výrobok, strata sluchu/bez straty sluchu, spokojný/nespokojný zákazník

*Štatistické  
skúmanie*

*Štatistické skúmanie* pozostáva zo štatistického zisťovania, štatistického spracovania a analyzovania. *Štatistické zisťovanie* predstavuje získavanie údajov skúmaním vybraných znakov štatistických jednotiek v štatistickom súbore. Výsledkom *štatistického zisťovania* sú údaje, ktoré sú pomocou

štatistického spracovania prehľadne usporiadané, roztriedené a spracované. V poslednej etape sa získané výsledky *analyzujú, vyhodnocujú* a zo získaných výsledkov *formulujú závery*. Na Obr. 2.3 sú znázornené etapy štatistického skúmania, pričom každá etapa používa isté metódy a postupy štatistiky.



Obr. 2.3: Štatistické skúmanie – etapy

2.2 Triedenie štatistického súboru

*Triedenie štatistického súboru* predstavuje prvý krok pri spracovaní údajov. Ide o usporiadanie štatistických jednotiek do skupín, tried. Podľa počtu triediacich štatistických znakov, ktoré môžu byť kvalitatívne alebo kvantitatívne, rozlišujeme jednostupňové a viacstupňové triedenie.

*Triedenie súboru*

*Jednostupňové triedenie* je triedenie podľa jedného triediaceho znaku (napr. priemer súčiastky, vek respondenta, veľkosť mikrotriesky). *Viacstupňové triedenie* je triedenie podľa viacerých znakov (napr. vzdelanie a vek zákazníka, typ dodávateľa a kvalita dodaného tovaru). Napríklad triedenie zamestnancov podľa veku je jednostupňové triedenie. Ak zamestnancov roztriedime aj podľa počtu odpracovaných rokov a typu pracovného zaradenia, tak získame trojstupňové triedenie.

*Druhy triedenia*

V ďalšej časti budeme uvažovať iba jednostupňové triedenie. Ak štatistický znak je diskretný alebo spojitý a pritom nadobúda málo rôznych, často sa opakujúcich hodnôt, tak použijeme *jednoduché triedenie*. Ak štatistický znak je spojitý a nadobúda veľké množstvo rôznych hodnôt, ktoré sa spravidla málo alebo vôbec neopakujú, tak použijeme *intervalové triedenie*.

Pri triedení musíme dodržať *pravidlo úplnosti* (každú štatistickú jednotku zatriedime) a *pravidlo jednoznačnosti* (každú štatistickú jednotku zatriedime iba do jednej triedy).

*Pravidlá triedenia*

Príkladom jednoduchého triedenia je rozdelenie študentov podľa výsledku zo skúšky z daného predmetu (Tab. 2.3).

Tabuľka 2.3: Jednoduché triedenie – hodnotenie z predmetu

Známka	A	B	C	D	E	FX
Počet študentov	15	25	30	35	55	12

Hmotnosť súčiastky je spojité kvantitatívny znak, ktorý nadobúda veľa rôznych hodnôt. Výsledok intervalového triedenia hmotnosti súčiastky (v g) je v Tab. 2.4.

Tabuľka 2.4: Intervalové triedenie – hmotnosť súčiastky (v g)

Hmotnosť	$\langle 1,5; 1,8 \rangle$	$\langle 1,8; 2,0 \rangle$	$\langle 2,0; 2,2 \rangle$	$\langle 2,2; 2,4 \rangle$	$\langle 2,4; 2,6 \rangle$
Počet súčiastok	7	19	25	14	9

### 2.2.1 Jednoduché triedenie

Jednoduché triedenie vzniká v prípade jedného štatistického znaku s malým počtom obmien. Nech  $x_1, x_2, x_3, \dots, x_n$  sú získané hodnoty sledovaného štatistického znaku  $X$ . Každá rôzna hodnota štatistického znaku predstavuje jednu triedu. Nech celkový počet rôznych hodnôt štatistického znaku je  $k$  a nech rozsah štatistického súboru je  $n$ , platí  $k \leq n$ . Zapísaním hodnôt štatistického znaku v poradí, v akom sme ich získali, dostaneme tzv. *prvotnú tabuľku*.

*Variačný rad*

Ak usporiadame hodnoty štatistického znaku podľa veľkosti do neklesajúcej postupnosti (od najmenej po najväčšiu hodnotu), tak získame *variačný rad*. Variačný rad môžeme zapísať v tvare

$$x_{\min} = x_{(1)} \leq x_{(2)} \leq x_{(3)} \leq \dots \leq x_{(n-1)} \leq x_{(n)} = x_{\max}.$$

*Tabuľka početností*

Z variačného radu vytvoríme pomocou čiarkovacej metódy *tabuľku rozdelenia početností* (resp. frekvenčnú tabuľku rozdelenia početností), ktorá slúži ako prvotný názorný spôsob prezentovania výsledkov zisťovania.

*Čiarkovacia metóda*

Podstata čiarkovacej metódy spočíva v tom, že počet čiarok určuje, koľkokrát sa daná hodnota vyskytuje v štatistickom súbore. Pre uľahčenie konečného sčítania čiarok sa odporúča zaznačiť každú piatu hodnotu tak, že predchádzajúce štyri čiarky prečiarkneme. Označenie  $|||| = \equiv$ .



**Príklad 2.1.** *K dispozícii máme záznam počtu chybných výrobkov počas dňa v istej menšej výrobnjej organizácii. Zistený počet nekvalitných výrobkov za obdobie 20 dní je: 0; 2; 0; 3; 1; 4; 1; 3; 5; 2; 0; 1; 3; 3; 2; 2; 1; 4; 3; 2. Úlohou je pomocou čiarkovacej metódy roztriediť dáta.*

### Riešenie.

Štatistický súbor má rozsah  $n = 20$  dní a skúmaný štatistický znak  $X$  je počet chybných výrobkov. Rad hodnôt 0; 2; 0; 3; 1; 4; 1; 3; 5; 2; 0; 1; 3; 3; 2; 2; 1; 4; 3; 2 tvorí prvotnú tabuľku.

Z údajov vieme, že počas 20 dní bolo zistených 6 rôznych hodnôt počtu nepodarkov (0, 1, 2, 3, 4, 5). Iba 3 dni boli také, v ktorých neboli vyrobené chybné výrobky. Počet dní, v ktorých bol zaznamenaný jeden chybný výrobok je 4. Dva chybné výrobky (a rovnako aj 3 chybné výrobky) boli vyrobené počas 5 dní. Počas 2 dní boli vyrobené 4 chybné výrobky a iba jeden deň bol taký, kde bolo 5 nekvalitných výrobkov.

Výsledkom triedenia je tabuľka so šiestimi triedami ( $k = 6$ ), ktorú nazývame tabuľka rozdelenia početností (Tab. 2.5).

Tabuľka 2.5: Tabuľka rozdelenia početností

Trieda ( $j$ )	1	2	3	4	5	6
Počet chybných výrobkov ( $x_j$ )	0	1	2	3	4	5
Čiarkovacia metóda						
Počet dní ( $n_j$ )	3	4	5	5	2	1

Môžeme povedať, že napríklad hodnota štatistického znaku  $x_2 = 1$  sa nachádza v štatistickom súbore 4-krát, čo označujeme ako  $n_2 = 4$ . Táto hodnota sa nazýva absolútna početnosť hodnoty  $x_2 = 1$  sledovaného štatistického znaku.

☑

*Absolútna početnosť  $n_j$*  predstavuje počet štatistických jednotiek, pri ktorých štatistický znak  $X$  nadobúda hodnotu  $x_j$ .

*Absolútne početnosti*

Súčet všetkých absolútnych početností sa rovná rozsahu súboru, t. j.

$$n_1 + n_2 + \dots + n_k = \sum_{j=1}^k n_j = n. \quad (2.1)$$

*Absolútna kumulatívna početnosť*  $N_j$  určuje, koľko štatistických jednotiek v štatistickom súbore má hodnotu štatistického znaku  $X$  menšiu alebo rovnú ako určená hodnota  $x_j$ . Platí

$$n_1 + n_2 + \dots + n_j = \sum_{i=1}^j n_i = N_j. \quad (2.2)$$

Okrem absolútnych početností sa pri triedení používajú aj *relatívne početnosti*.

*Relatívne početnosti*

*Relatívna početnosť*  $f_j$  je podiel absolútnej početnosti  $n_j$  a celkového počtu štatistických jednotiek  $n$ . Platí

$$f_j = \frac{n_j}{n}, \quad j = 1, 2, \dots, k. \quad (2.3)$$

*Relatívna kumulatívna početnosť*  $F_j$  vyjadruje súčet relatívnych početností od prvej po  $j$ -tu triedu a určuje, aká časť štatistického súboru nadobúda hodnotu štatistického znaku menšiu alebo rovnú ako je hodnota  $x_j$ .

Platí

$$F_j = f_1 + f_2 + \dots + f_j = \frac{n_1 + n_2 + \dots + n_j}{n} = \frac{N_j}{n}. \quad (2.4)$$

Relatívne početnosti  $f_j$  a  $F_j$  sa často vyjadrujú v percentách. Hodnota  $100 \cdot f_j\%$  udáva, koľko percent prvkov zo štatistického súboru nadobúda hodnotu  $x_j$ . Hodnota  $100 \cdot F_j\%$  predstavuje, koľko % prvkov zo štatistického súboru má hodnotu menšiu alebo rovnú ako je určená hodnota  $x_j$ .

Ak štatistický súbor má  $k$  rôznych hodnôt (tried), tak platí

$$\sum_{j=1}^k n_j = N_k = n, \quad \sum_{j=1}^k f_j = F_k = 1. \quad (2.5)$$

V Tab. 2.6 je kompletná tabuľka rozdelenia početností.

*Postupnosť triedenia*

*Kroky na zostavenie tabuľky rozdelenia početností:*

1. Určíme rôzne hodnoty štatistického znaku, ktoré tvoria triedy.
2. Roztriedime hodnoty štatistického znaku do správnych tried.
3. Nájdeme absolútne a relatívne početnosti a zostrojíme výslednú tabuľku rozdelenia početností (Tab. 2.6).

Tabuľka 2.6: Tabuľka rozdelenia početností

Trieda	Hodnoty znaku	Početnosť		Kumulatívna početnosť	
		absolútna	relatívna	absolútna	relatívna
$j$	$x_j$	$n_j$	$f_j$	$N_j$	$F_j$
1	$x_1$	$n_1$	$f_1 = \frac{n_1}{n}$	$N_1 = n_1$	$F_1 = \frac{N_1}{n}$
2	$x_2$	$n_2$	$f_2 = \frac{n_2}{n}$	$N_2 = n_1 + n_2$	$F_2 = \frac{N_2}{n}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$k$	$x_k$	$n_k$	$f_k = \frac{n_k}{n}$	$N_k = n$	$F_k = 1$

**Príklad 2.2.** Na zistenie priemernej teploty sme namerali o 14. hodine počas 10 dní hodnoty (v °C): 27; 26; 25; 28; 29; 25; 28; 27; 25; 26. Zostavme kompletnú tabuľku rozdelenia početností. V koľkých dňoch bola nameraná hodnota práve 26°C? Koľko percent meraní malo hodnotu menšiu ako 28°C?

### Riešenie.

Najprv zoradíme hodnoty do variačného radu

$$x_{(1)} = 25 \leq 25 \leq 25 < 26 \leq 26 < 27 \leq 27 < 28 \leq 28 < 29 = x_{(10)}.$$

Zo získaných údajov vieme, že minimálna hodnota teploty je 25°C ( $x_{\min}$ ) a maximálna hodnota 29°C ( $x_{\max}$ ). Rozsah štatistického súboru je  $n = 10$ , pričom máme päť ( $k = 5$ ) rôznych hodnôt (25, 26, 27, 28, 29).

Tabuľka 2.7: Tabuľka rozdelenia početností – denná teplota

Poradové číslo	Trieda (hodnota)	Čiarkovacia metóda	Absolútna početnosť
$j$	$x_j$		$n_j$
1	25		3
2	26		2
3	27		2
4	28		2
5	29		1
$\Sigma$			10

Absolútnu početnosť  $n_j$  získame pomocou čiarkovacej metódy, t.j. do daného stĺpca zapíšeme toľko čiarok, koľkokrát sa hodnota  $x_j$  v štatistickom súbore nachádza (Tab. 2.7).

Relatívnu početnosť  $f_j$  vypočítame podľa vzťahu (2.3), odkiaľ

$$f_1 = \frac{n_1}{n} = \frac{3}{10}, f_2 = \frac{n_2}{n} = \frac{2}{10}, \dots, f_5 = \frac{n_5}{n} = \frac{1}{10}.$$

Kumulatívna absolútna početnosť  $N_j$  určuje, koľko jednotiek v štatistickom súbore má hodnotu menšiu alebo rovnú ako určená hodnota  $x_j$ , t. j.

$$N_1 = n_1 = 3, N_2 = n_1 + n_2 = 3 + 2 = 5, \dots, N_5 = n_1 + \dots + n_5 = 10.$$

Pre kumulatívnu relatívnu početnosť  $F_j$  platí vzťah (2.4), odkiaľ

$$F_1 = \frac{N_1}{n} = \frac{3}{10}, F_2 = \frac{N_2}{n} = \frac{5}{10}, \dots, F_5 = \frac{N_5}{n} = \frac{10}{10} = 1.$$

Výsledok triedenia je v tabuľke rozdelenia početností (Tab. 2.8).

Tabuľka 2.8: Kompletná tabuľka rozdelenia početností – denná teplota

Trieda	Hodnoty znaku	Početnosť		Kumulatívna početnosť	
		absolútna	relatívna	absolútna	relatívna
$j$	$x_j$	$n_j$	$f_j$	$N_j$	$F_j$
1	25	3	$\frac{3}{10} = 0,3$	3	$\frac{3}{10} = 0,3$
2	26	2	$\frac{2}{10} = 0,2$	3+2=5	$\frac{5}{10} = 0,5$
3	27	2	$\frac{2}{10} = 0,2$	5+2=7	$\frac{7}{10} = 0,7$
4	28	2	$\frac{2}{10} = 0,2$	7+2=9	$\frac{9}{10} = 0,9$
5	29	1	$\frac{1}{10} = 0,1$	9+1=10	$\frac{10}{10} = 1,0$

Z tabuľky rozdelenia početností vyplýva, že práve 2 merania mali hodnotu teploty  $26^\circ\text{C}$  ( $n_2$ ). Nameraná teplota bola menšia ako  $28^\circ\text{C}$  v 70% meraní ( $100 \cdot F_3\%$ ).

☑

## 2.2.2 Intervalové triedenie

Intervalové triedenie sa používa predovšetkým v prípade veľkého počtu rôznych hodnôt štatistického znaku, ktoré sa spravidla málo alebo vôbec neopakujú. Príkladom môžu byť namerané hodnoty životnosti istej súčiastky (v hod.): 291; 258; 305; 271; 276; 263; 269; 279; 285; 294; 293; 268; 263; 254; 254; 309; 287; 289; 271; 256; 272; 303; 299; 315; 236; 264; 318. V súbore sa nachádza veľa rôznych hodnôt. Najprehľadnejšie spracovanie

nameraných hodnôt získame tak, ak hodnoty rozdelíme do menších tried, tzv. intervalov.

Intervalové triedenie je triedenie, pri ktorom sa štatistické jednotky v súbore rozsahu  $n$  roztriedia do  $k$  tried,  $k \leq n$ . Označme  $I_j$  ako  $j$ -ty triedny interval, ktorého dolná hranica je  $t_{j-1}$  a horná hranica  $t_j$ . Pri určovaní intervalov je potrebné rozhodnúť, ktorá z hraníc bude patriť do intervalu, t.j. stanoviť typ intervalu  $(t_{j-1}; t_j)$ , resp.  $\langle t_{j-1}; t_j \rangle$ . V nasledujúcej časti budeme pracovať s typom intervalu  $I_j = (t_{j-1}; t_j)$ .

*Intervalové  
triedenie*

Každému triednemu intervalu  $I_j$  priradíme triedny znak  $z_j$ . Triedny znak  $z_j$  predstavuje stred  $j$ -teho intervalu, pre ktorý platí

*Triedny  
znak*

$$z_j = \frac{t_{j-1} + t_j}{2}, j = 1, 2, \dots, k, \quad (2.6)$$

kde  $k$  je počet triednych intervalov.

V prípade intervalového triedenia je potrebné stanoviť počet tried (intervalov) a šírku intervalov. Počet tried nemôže byť veľmi malý, aby sa nestratila podstatná časť informácií o rozdelení hodnôt štatistického znaku v súbore. Na druhej strane, príliš veľký počet tried môže znížiť prehľadnosť triedenia.

Na určenie optimálneho počtu  $k$  triednych intervalov pre štatistický súbor rozsahu  $n$  existuje niekoľko odporúčaných postupov a spôsobov.

*Počet  
intervalov*

Najjednoduchšie pravidlo je

$$k = \sqrt{n} \quad (2.7)$$

alebo tzv. Sturgesovo pravidlo (2.8)

$$k \approx 1 + 3,322 \log n. \quad (2.8)$$

K ďalším možnostiam patria vzťahy (2.9) alebo (2.10)

$$k \leq 5 \log n, \quad (2.9)$$

$$0,55 n^{0,4} \leq k \leq 1,25 n^{0,4}. \quad (2.10)$$

Šírku triedneho intervalu  $h$  môžeme približne vypočítať podľa vzťahu

*Šírka  
intervalu*

$$h \doteq \frac{R_V}{k}, \quad (2.11)$$

kde  $R_V$  je *varičné rozpätie*, pre ktoré platí

$$R_V = x_{\max} - x_{\min}. \quad (2.12)$$

Výsledkom triedenia je tabuľka rozdelenia početností, ktorej schéma je podobná ako v prípade jednoduchého triedenia.

*Postupnosť triedenia*      *Kroky na zostavenie tabuľky rozdelenia početností pre intervalové triedenie:*

1. Určíme počet intervalov  $k$  a šírku intervalu  $h$ .
2. Určíme dolnú  $t_{j-1}$  a hornú hranicu  $t_j$  každého intervalu  $I_j$ .
3. Roztriedime hodnoty štatistického znaku do správnych intervalov (pravidlo jednoznačnosti a úplnosti).
4. Vypočítame stred intervalu pre každý interval.
5. Nájdeme absolútne a relatívne početnosti a zostrojíme výslednú tabuľku.

**Príklad 2.3.** *Pri kontrole kvality výrobku boli namerané odchýlky jedného rozmeru vzoriek výrobku od stanovenej hodnoty (v  $\mu\text{m}$ ): 36,1; 35,2; 33,9; 42,2; 35,6; 39,7; 35,4; 33,5; 39,0; 41,2; 29,7; 34,1; 36,6; 31,5; 37,8; 35,8; 37,0; 39,2; 34,7; 33,6; 40,9; 38,2; 35,1; 40,4; 31,2; 35,6; 40,1; 39,7; 36,2; 32,4; 38,5; 30,7; 32,6; 35,0; 38,1. Zostavme kompletnú tabuľku rozdelenia početností.*

### Riešenie.

Zo získaných 35 údajov vieme, že minimálna hodnota je 29,7  $\mu\text{m}$  a maximálna 42,2  $\mu\text{m}$ .

Počet intervalov  $k$  nájdeme pomocou vzorca (2.10)

$$0,55 \cdot 35^{0,4} \leq k \leq 1,25 \cdot 35^{0,4},$$

odkiaľ  $2,280 \leq k \leq 5,182$ . Odporúčaná dolná hranica počtu tried je 3 a horná 5. V našom prípade volíme optimálny počet tried  $k = 4$ .

Šírku intervalu  $h$  vypočítame približne podľa vzťahu (2.11)

$$h \doteq \frac{R_V}{k} = \frac{x_{\max} - x_{\min}}{k} = \frac{42,2 - 29,7}{4} = 3,125,$$

pričom budeme uvažovať hodnotu 3,2.

Pri špecifikácii triednych intervalov volíme interval tak, aby horná hranica patrila do intervalu. To znamená, že budeme pracovať s intervalom v tvare  $I_j = (t_{j-1}; t_j]$ . Namerané údaje rozdelíme pomocou čiarkovacej metódy tak, aby sa každá hodnota nachádzala práve v jednom triednom intervale. Výsledok intervalového triedenia je v Tab. 2.9.

Tabuľka 2.9: Tabuľka rozdelenia početností – odchýlka rozmeru

$j$	$I_j$	Čiarkovacia metóda	$n_j$
1	(29,6; 32,8)	###	6
2	(32,8; 36,0)	######	12
3	(36,0; 39,2)	######	10
4	(39,2; 42,4)	###	7

Každému triednemu intervalu  $I_j$  určíme triedny znak ako stred intervalu  $z_j$  pomocou vzťahu (2.6), t. j.

$$z_1 = \frac{29,6 + 32,8}{2} = 31,2, \quad z_2 = \frac{32,8 + 36,0}{2} = 34,4,$$

$$z_3 = \frac{36,0 + 39,2}{2} = 37,6, \quad z_4 = \frac{39,2 + 42,4}{2} = 40,8.$$

V nasledujúcom kroku vypočítame ďalšie početnosti  $f_j$ ,  $N_j$ ,  $F_j$  a zostrojíme výslednú tabuľku rozdelenia početností (Tab. 2.10).

Tabuľka 2.10: Kompletná tabuľka rozdelenia početností – odchýlka rozmeru

$j$	$(t_{j-1}; t_j)$	$z_j$	$n_j$	$f_j$	$N_j$	$F_j$
1	(29,6; 32,8)	31,2	6	$\frac{6}{35} = 0,171$	6	$\frac{6}{35} = 0,171$
2	(32,8; 36,0)	34,4	12	$\frac{12}{35} = 0,343$	6+12=18	$\frac{18}{35} = 0,514$
3	(36,0; 39,2)	37,6	10	$\frac{10}{35} = 0,286$	18+10=28	$\frac{28}{35} = 0,800$
4	(39,2; 42,4)	40,8	7	$\frac{7}{35} = 0,200$	28+7=35	$\frac{35}{35} = 1,000$

Z tabuľky môžeme napríklad zistiť, že v súbore sa nachádza 12 vzoriek výrobku, ktorých odchýlka od požadovaného rozmeru je v intervale (32,8; 36,0), čo predstavuje približne 34,30% ( $100 \cdot f_2\%$ ) z celkového počtu vzoriek. Počet vzoriek, ktoré majú odchýlku nanajvýš 39,2  $\mu\text{m}$  je 28, čo v relatívnom vyjadrení predstavuje 80% ( $100 \cdot F_3\%$ ).

☑

## 2.3 Grafické zobrazenie

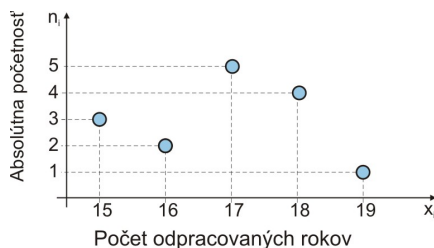
Prezentácia údajov je dôležitou súčasťou každého štatistického skúmania. Jej vhodným výberom sa môže zabezpečiť lepšie pochopenie obsahu a podstaty prezentovaných informácií.

*Grafy* Grafické zobrazenie je účinným nástrojom prezentácie spracovania štatistických údajov. Grafy nie sú síce presné ako tabuľky, ale rýchlejšie a názornejšie môžu poskytnúť predstavu o dôležitých tendenciách a súvislostiach. Pri štatistických analýzach sa používajú rôzne druhy grafov, k najčastejšie používaným patria bodové, spojnicové, stĺpcové a kruhové grafy.

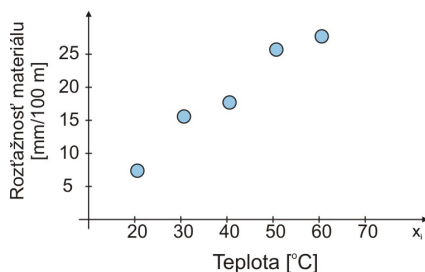
*Zásady tvorby grafov* Poznamenajme, že graf musí byť výstižný, prehľadný, zrozumiteľný a dobre čitateľný. Nesmie chýbať názov grafu, názvy a popis osí, legenda, prípadne vysvetlivky.

*Bodový graf* Bodový graf znázorňuje namerané hodnoty pomocou bodov v pravouhlej sústave súradníc. Body vznikajú ako priesečníky príslušných hodnôt štatistického znaku, ktoré sú zaznamenané na osi  $x$ , a ich absolútnych, resp. relatívnych početností na osi  $y$  (Obr. 2.4).

Špeciálnym prípadom je bodový graf, ktorý sa používa v regresnej analýze pri znázornení závislosti dvoch premenných. V tomto prípade sa na os  $x$  nanášajú hodnoty prvej premennej (nezávislá premenná). Zvislá os  $y$  je určená pre hodnoty druhej, závislej premennej. Takýto graf niekedy nazývame *korelačný diagram*, ktorý v našom prípade znázorňuje závislosť rozťažnosti istého materiálu od jeho teploty zahrievania (Obr. 2.5).



Obr. 2.4: Bodový graf

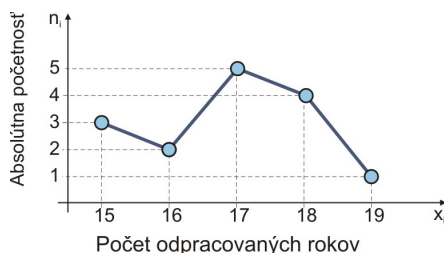


Obr. 2.5: Korelačný graf



*Spojnicový graf* je bodový graf, v ktorom sú susedné body spojené úsečkou. Veľmi často vyjadruje priebeh časového radu, ale slúži aj na zobrazenie absolútnych alebo relatívnych početností. *Spojnicový graf rozdelenia početností* nazývame *polygón rozdelenia početností* (Obr. 2.6). V tomto prípade sa na vodorovnú os znázorňujú triedy (hodnoty) štatistického znaku a na zvislú os ich absolútne, resp. relatívne početnosti.

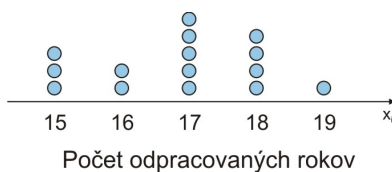
Polygón



Obr. 2.6: Polygón

*Dot plot* je graf, ktorý je alternatívou k polygónu. Na vodorovnú os sa znázorňujú triedy (hodnoty) štatistického znaku a hodnota štatistického znaku je zobrazená ako bod. Body sú umiestnené nad príslušnou hodnotou triedy a ich počet nad danou triedou určuje absolútnu početnosť (Obr. 2.7).

Dot plot



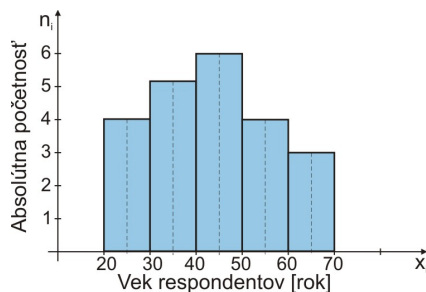
Obr. 2.7: Dot plot

Pre svoju názornosť a jednoduchosť je veľmi rozšírený aj *stĺpcový graf* (vertikálny alebo horizontálny). Každý zo stĺpcov grafu vyjadruje jednu z tried skúmaného štatistického znaku. Výška (resp. dĺžka) stĺpcov odpovedá počtu štatistických jednotiek v danej triede. Špeciálnym prípadom stĺpcového grafu je histogram.

Stĺpcový graf

*Histogram početností* je špeciálny stĺpcový graf, ktorý slúži na zobrazenie rozdelenia absolútnych alebo relatívnych početností (Obr. 2.8). Stĺpce histogramu sú vždy vertikálne a ich výška odpovedá príslušnej početnosti. Šírka stĺpcov odpovedá šírke triedneho intervalu. Ak pospájame stredy horných strán všetkých stĺpcov, tak získame polygón rozdelenia početností. Pri analýze histogramu vieme určiť minimálnu a maximálnu hodnotu a z nich variabilitu skúmaného znaku, stred rozdelenia, počet vrcholov, prítomnosť odľahlých hodnôt, symetriu, strmosť a šikmosť rozdelenia a pod.

Histogram



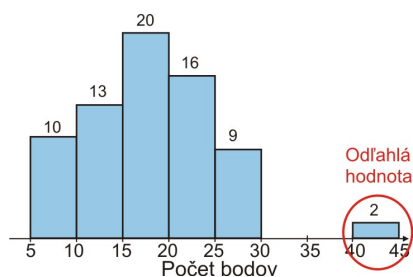
Obr. 2.8: Histogram

### Kroky na vytvorenie histogramu

1. Nakreslíme dve na seba kolmé osi. Na horizontálnu os (os  $x$ ) naznačíme intervaly s danou šírkou.
2. Na vertikálnu os (os  $y$ ) nanesieme absolútnu (resp. relatívnu) početnosť.
3. Zostrojíme obdĺžnik nad každým intervalom s výškou rovnajúcou sa zodpovedajúcim absolútnym (resp. relatívnym) početnostiam.

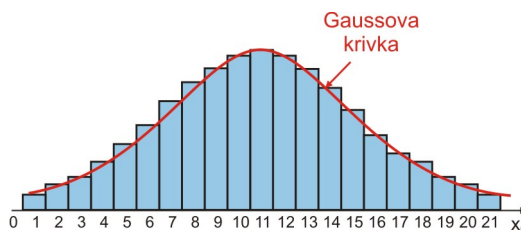
### Odlahlá hodnota

*Odlahlá hodnota* je vo všeobecnosti hodnota, ktorá sa výrazne líši od ostatných hodnôt. Za určitých okolností sa odporúča odlahlé hodnoty vylúčiť zo súboru hodnôt a v analýze dát pokračovať bez týchto hodnôt. Na Obr. 2.9 je znázornený histogram počtu bodov získaných z písomky 70 študentov.



Obr. 2.9: Histogram s odlahlou hodnotou

V prípade, že histogram znázorňujúci rozdelenie početností dostatočne presne kopíruje krivku normálneho rozdelenia (tzv. *Gaussovu krivku*), tak hovoríme, že toto rozdelenie je približne normálne rozdelenie (Obr. 2.10).



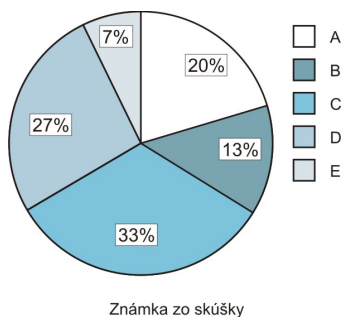
Obr. 2.10: Histogram a Gaussova krivka

*Kruhový (koláčový) graf* patrí do skupiny tzv. plošných grafov. Plocha kruhu predstavuje celý súbor a jednotlivé časti sú znázornené kruhovými výsekmi (Obr. 2.11).

*Kruhový graf*

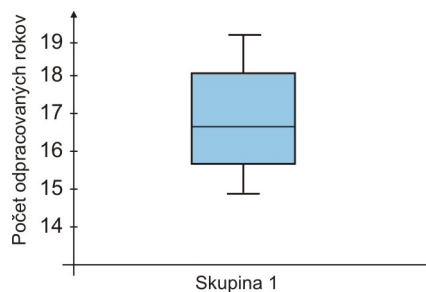
Spojnicový, stĺpcový a kruhový graf ako prvý použil škótsky inžinier, ekonóm a spisovateľ *William Playfair* (1759-1823). V svojom diele *Commercial and Political Atlas* (Komerčný a politický atlas) z roku 1786 použil spojnicový a stĺpcový graf. V teoretickom diele o grafoch *Statistical Breviary* (Štatistický breviár) z roku 1801 použil kruhový graf. Playfair bol veľkým propagátorom grafického zobrazenia. Tvrdil, že grafy sú zrozumiteľnejšie ako číselné tabuľky.

*William Playfair*



Známka zo skúšky

Obr. 2.11: Kruhový graf



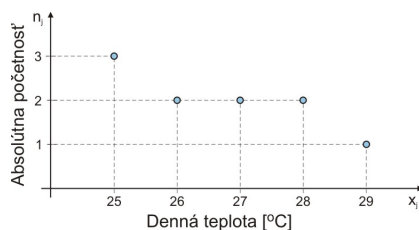
Obr. 2.12: Boxplot

**Príklad 2.4.** V Príklade 2.2 sme sledovali priemernú teplotu počas desiatich dní (v °C). Zostrojme histogram a polygón absolútnych početností.

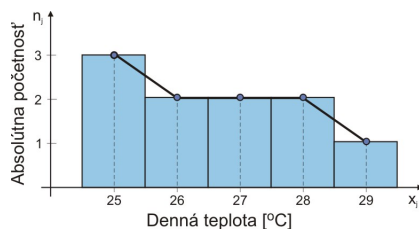
$x_j$	25	26	27	28	29
$n_j$	3	2	2	2	1

### Riešenie.

Bodový graf absolútnych početností, polygón absolútnych početností a histogram sú zobrazené na Obr. 2.13 a Obr. 2.14. ☒



Obr. 2.13: Bodový graf



Obr. 2.14: Histogram s polygónom

*Dalšie  
grafy*

K špeciálnym grafom patrí boxplot (Obr. 2.12), graf sezónnych hodnôt (v časových radoch), sieťové grafy, kartogramy (mapy), kartodiagramy, pyramídy, piktogramy a iné. Kartodiagram vznikne spojením mapy a iného druhu grafu. Vekové pyramídy sa používajú napríklad pri analýzach demografického vývoja zloženia obyvateľov podľa veku.

**Príklad 2.5.** V priebehu mesiaca odpracovalo 33 pracovníkov určitý počet dní:

22; 21; 23; 22; 20; 20; 23; 22; 22; 23; 22; 21; 23; 20; 23; 21; 23; 21; 23; 22; 24; 22; 23; 22; 21; 22; 24; 22; 20; 22; 22; 23; 21.

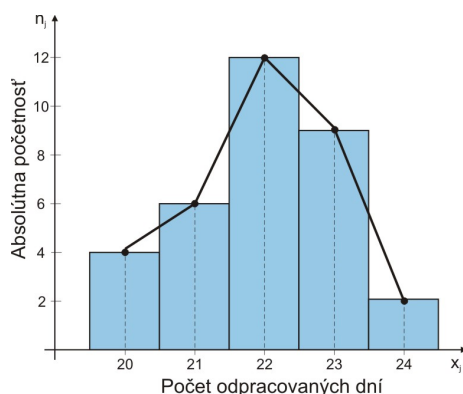
- Zostavme kompletnú tabuľku početností. Koľko percent pracovníkov odpracovalo 21 dní? Koľko pracovníkov odpracovalo menej ako 22 dní?
- Zostrojme polygón a histogram absolútnych početností.

**Riešenie.**

Zo získaných údajov vieme, že minimálny počet odpracovaných dní je 20 ( $x_{min}$ ) a maximálny počet 24 dní ( $x_{max}$ ). Rozsah štatistického súboru je  $n = 33$ , pričom máme 5 rôznych hodnôt (20, 21, 22, 23, 24). Tabuľka početností má 5 tried, t. j.  $k = 5$ . Kvôli väčšiemu rozsahu súboru nebudeme usporiadať hodnoty (nebudeme tvoriť variačný rad), ale priamo použijeme čiarkovaciu metódu. Výsledok triedenia je v nasledujúcej tabuľke početností (Tab. 2.11).

Tabuľka 2.11: Tabuľka rozdelenia početností – počet odpracovaných dní

$j$	$x_j$		$n_j$	$f_j$	$N_j$	$F_j$
1	20		4	$\frac{4}{33} = 0,121$	4	$\frac{4}{33} = 0,121$
2	21		6	$\frac{6}{33} = 0,182$	10	$\frac{10}{33} = 0,303$
3	22		12	$\frac{12}{33} = 0,364$	22	$\frac{22}{33} = 0,667$
4	23		9	$\frac{9}{33} = 0,273$	31	$\frac{31}{33} = 0,940$
5	24		2	$\frac{2}{33} = 0,060$	33	$\frac{33}{33} = 1,000$
$\Sigma$			33	1,0000		



Obr. 2.15: Polygón a histogram – počet odpracovaných dní

Z tabuľky početností vyplýva, že 18,20% ( $100 \cdot f_2\%$ ) pracovníkov odpracovalo 21 dní. Menej ako 22 dní odpracovalo 10 pracovníkov ( $N_2$ ). Polygón a histogram absolútnych početností je na Obr. 2.15.



**Príklad 2.6.** Pri kontrole automatickej linky boli zistené údaje o hmotnosti automaticky balených výrobkov (v g): 40,3; 41,3; 41,4; 40,2; 43,5; 42,6; 41,9; 42,3; 43,1; 44,0; 40,8; 42,9; 41,1; 44,0; 42,8; 41,7; 41,5; 42,5; 42,1; 41,5; 41,9; 41,6. Zostavme kompletnú tabuľku rozdelenia početností. Zostrojme polygón a histogram absolútnych početností.

### Riešenie.

Z údajov ( $n = 22$ ) vieme, že minimálna hmotnosť výrobku je 40,2 g a maximálna 44,0 g. Počet triednych intervalov  $k$  nájdeme pomocou vzorca (2.8)

$$k \approx 1 + 3,322 \log n,$$

odkiaľ  $k \approx 1 + 3,322 \log 22 = 5,46$ . Volíme optimálny počet tried  $k = 5$ .

Šírku intervalu  $h$  vypočítame približne podľa vzťahu (2.11)

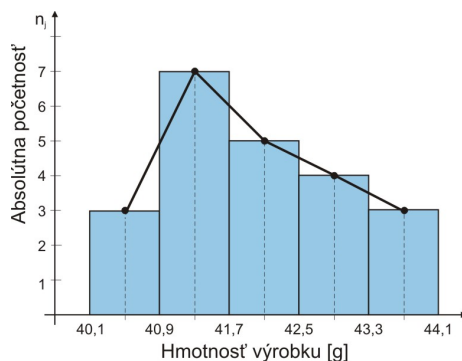
$$h \doteq \frac{R_V}{k} = \frac{x_{\max} - x_{\min}}{k} = \frac{44,0 - 40,2}{5} = 0,76.$$

Výslednú hodnotu zaokrúhlime smerom nahor a budeme uvažovať hodnotu 0,8. Pri špecifikácii triednych intervalov volíme interval tak, aby horná hranica patrila do intervalu. Výsledok intervalového triedenia je v Tab. 2.12.

Každému intervalu  $I_j$  určíme triedny znak ako stred intervalu  $z_j$  podľa vzťahu (2.6). Dostaneme

$$z_1 = \frac{40,1 + 40,9}{2} = 40,5, z_2 = \frac{40,9 + 41,7}{2} = 41,3, \dots, z_5 = \frac{43,3 + 44,1}{2} = 43,7.$$

Polygón a histogram absolútnych početností je na Obr. 2.16.



Obr. 2.16: Polygón a histogram – hmotnosť výrobku

Tabuľka 2.12: Tabuľka rozdelenia početností – hmotnosť výrobku

$j$	$(t_{j-1}; t_j)$	$z_j$	$n_j$	$f_j$	$N_j$	$F_j$
1	$(40,1; 40,9)$	40,5	3	$\frac{3}{22} = 0,136$	3	$\frac{3}{22} = 0,136$
2	$(40,9; 41,7)$	41,3	7	$\frac{7}{22} = 0,318$	10	$\frac{10}{22} = 0,455$
3	$(41,7; 42,5)$	42,1	5	$\frac{5}{22} = 0,227$	15	$\frac{15}{22} = 0,682$
4	$(42,5; 43,3)$	42,9	4	$\frac{4}{22} = 0,182$	19	$\frac{19}{22} = 0,864$
5	$(43,3; 44,1)$	43,7	3	$\frac{3}{22} = 0,136$	22	$\frac{22}{22} = 1,000$

Z tabuľky napríklad vyplýva, že v súbore sa nachádza 7 výrobkov, ktorých hmotnosť je v intervale  $(40,9; 41,7)$ , čo predstavuje približne 31,8%

$(100 \cdot f_2\%)$  z celkového počtu výrobkov. Počet výrobkov, ktoré majú hmotnosť menšiu alebo rovnú ako 43,3 gramov je 19, čo v relatívnom vyjadrení predstavuje 86,4% ( $100 \cdot F_4\%$ ).



## Úlohy

**2.1** Nasledujúce dáta predstavujú počty bodov, ktoré študenti získali z testu z matematiky: 15; 13; 14; 16; 15; 16; 12; 13; 14; 14; 15; 12; 14; 15; 12; 13; 17; 12; 14; 13; 12; 13; 14; 13; 17; 13; 17; 12. Zostavte kompletnú tabuľku rozdelenia početností. Koľko percent študentov získalo práve 14 bodov? Nakreslite histogram a polygón absolútnych početností.

**2.2** Počas 24 mesiacov domácnosť zaznamenala mesačné náklady na elektrinu (v €): 36; 35; 35; 34; 32; 38; 36; 32; 35; 37; 34; 36; 36; 33; 33; 37; 34; 35; 34; 34; 33; 34; 37; 36. Zostavte kompletnú tabuľku rozdelenia početností. Počas koľkých mesiacov boli náklady na elektrinu v domácnosti menšie alebo rovné 35 €? Nakreslite histogram a polygón absolútnych početností.

**2.3** Analýzou 29 vzoriek boli zistené nasledujúce hodnoty istej chemickej látky v roztoku (v mg/l): 23; 22; 22; 21; 19; 25; 23; 19; 22; 24; 21; 19; 23; 20; 20; 24; 21; 22; 21; 21; 20; 21; 24; 21; 23; 20; 22; 23; 23. Zostavte kompletnú tabuľku rozdelenia početností. V koľkých percentách vzoriek boli hodnoty chemickej látky v roztoku menšie ako 22 mg/l? Nakreslite histogram a polygón absolútnych početností.

**2.4** Meraním dĺžky kovových valčekov sa získali tieto hodnoty (v mm): 104; 106; 103; 103; 108; 104; 105; 104; 104; 103; 104; 104; 105; 106; 103; 105; 106; 106; 107; 105; 105; 105; 103; 108; 105; 106. Zostavte kompletnú tabuľku rozdelenia početností. Koľko kovových valčekov malo dĺžku najviac 106 mm? Nakreslite histogram a polygón absolútnych početností.

**2.5** Nasledujúce hodnoty (v cm) boli získané meraním priemeru 30 piestových krúžkov: 14,21; 14,55; 14,38; 14,48; 14,30; 14,53; 14,39; 14,22; 14,54; 14,61; 14,57; 14,71; 14,60; 14,74; 14,62; 14,32; 14,43; 14,74; 14,38; 14,65; 14,47; 14,67; 14,51; 14,72; 14,33; 14,70; 14,40; 14,60; 14,35; 14,71. Intervalovým triedením zostavte kompletnú tabuľku rozdelenia početností. Nakreslite histogram a polygón absolútnych početností.

$$5; \gamma_3 = -0,3362; \gamma_4 = -0,5950; V = 15,2655\%$$

Názov publikácie: ZÁKLADY MATEMATICKEJ ŠTATISTIKY  
Autor: doc. RNDr. Miriam Andrejiová, PhD.  
RNDr. Zuzana Kimáková, PhD.  
Vydanie: prvé  
Vydavateľ: Technická univerzita v Košiciach  
Rok vydania: 2022  
Rozsah: 318 strán  
Náklad: 50 ks  
ISBN: 978-80-553-4319-8