

**TECHNICKÁ UNIVERZITA V KOŠICIACH**

Strojnícka fakulta

**ŠTATISTICKÉ METÓDY  
V PRAXI**

Miriam Andrejiová

Edícia vedeckej a odbornej literatúry

Košice 2016

Technická univerzita v Košiciach, Strojnícka fakulta

Miriam Andrejiová

Štatistické metódy v praxi

Recenzenti:

doc. RNDr. Viktor Pirč, CSc.

RNDr. Anna Grinčová, PhD.

© Strojnícka fakulta, Technická univerzita v Košiciach

Tlač: EQUILIBRIA, s.r.o., Košice

ISBN: 978-80-553-3078-5

# Predhovor

V súčasnej dobe čoraz výraznejšie rastie dopyt po štatistických informáciách a po výsledkoch kvalitných štatistických analýz. Rovnako aj úspešné rozhodovanie v mnohých oblastiach nášho života a technickej praxe nie je možné bez aspoň minimálneho využívania štatistiky a štatistických metód. Preto nemôžeme ani v najmenšom pochybovať o význame štatistiky.

Táto vysokoškolská učebnica je určená predovšetkým pre poslucháčov prvého ročníka druhého stupňa vysokoškolského štúdia Strojníckej fakulty Technickej univerzity v Košiciach v študijnom programe Inžinierstvo kvality produkcie, Bezpečnosť technických systémov a Strojárske technológie. Rovnako dobre ale môže poslúžiť aj študentom študujúcim iný študijný program, resp. poslucháčom iných fakúlt Technickej univerzity v Košiciach.

Učebnica obsahuje základy štatistiky a vybrané štatistické metódy: metódy štatistickej indukcie, závislosť kvalitatívnych a kvantitatívnych premenných, úvod do analýzy časových radov a štatistických metód riadenia kvality. Mnoho ďalších štatistických metód (napr. dvojfaktorová analýza rozptylu, metóda plánovania experimentov, viacrozmerné rozhodovacie metódy, zhuková analýza, faktorová analýza a pod.) nie sú obsahom učebnice. Cieľom učebnice nie je podať vyčerpávajúce množstvo informácií, ale vybudovať dobrý základ v oblasti štatistických metód.

Počítač a kvalitný softvér, ktorý dokáže zrealizovať efektívne a rýchlo štatistické vyhodnotenie, sa stáva súčasťou nášho života. To ale neznamená, že nie je potrebná znalosť štatistiky. Práve naopak. Na pochopenie počítačových výstupov je nutné poznanie podstaty použitých štatistických metód, pretože zlá interpretácia výsledkov a záverov môže viesť k nesprávnym rozhodnutiam.

V dôsledku toho sú riešenia vybraných úloh realizované bez použitia štatistického softvéru. Pri riešení úloh budeme zaokrúhľovať všetky medzivýsledky a výsledky na tri, resp. na štyri desatinné miesta, preto sa môže stať, že pri

následnom porovnaní s počítačovým výstupom sa budú výsledky mierne líšiť. Grafické výstupy sú realizované v prostredí R, v programe OriginLab a pomocou tabuľkového procesora Excel. Aj keď niektoré kritické hodnoty potrebné pri testovaní štatistických hypotéz je možné získať pomocou štatistických programov, resp. on-line štatistických aplikácií, sú v závere učebnice pre lepšiu orientáciu pri riešení úloh pripojené skrátené základné štatistické tabuľky. Pre úplnosť je prvá kapitola doplnená aj o stručné základy z teórie pravdepodobnosti a o vybrané rozdelenia náhodných premenných.

Snahou pri písaní učebnice bolo dodržať pravidlá jednoduchosti a zrozumiteľnosti a poukázať na užitočnosť a nenahraditeľnosť štatistických metód v praxi. Po ich zvládnutí by mal čitateľ dokázať správne aplikovať štatistické metódy pri analyzovaní a vyhodnotení štatistických údajov vo všetkých oblastiach technickej praxe.

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

Autorka

Učebnica vznikla za podpory KEGA 072TUKE-4/2014.

# Obsah

<b>1</b>	<b>Základy štatistických metód</b>	<b>11</b>
1.1	Popisná štatistika . . . . .	12
1.1.1	Základné pojmy . . . . .	12
1.1.2	Triedenie štatistického súboru . . . . .	14
1.1.2.1	Jednoduché triedenie . . . . .	14
1.1.2.2	Intervalové triedenie . . . . .	17
1.1.3	Číselné charakteristiky . . . . .	18
1.1.3.1	Charakteristiky polohy . . . . .	19
1.1.3.2	Charakteristiky variability . . . . .	21
1.1.3.3	Vybrané charakteristiky šikmosti a špicatosti . . . . .	24
1.1.4	Grafické zobrazenie . . . . .	26
1.2	Základy teórie pravdepodobnosti . . . . .	32
1.2.1	Základné pojmy . . . . .	32
1.2.2	Náhodná premenná a jej číselné charakteristiky . . . . .	35
1.2.2.1	Diskrétna náhodná premenná . . . . .	36
1.2.2.2	Spojité náhodná premenná . . . . .	37
1.2.2.3	Číselné charakteristiky náhodných premenných . . . . .	39
1.2.3	Rozdelenia diskretných náhodných premenných . . . . .	41
1.2.3.1	Dvojbodové rozdelenie . . . . .	41
1.2.3.2	Binomické rozdelenie . . . . .	41
1.2.3.3	Hypergeometrické rozdelenie . . . . .	42
1.2.3.4	Poissonovo rozdelenie . . . . .	42
1.2.4	Rozdelenia spojitých náhodných premenných . . . . .	43
1.2.4.1	Rovnomerné rozdelenie . . . . .	43
1.2.4.2	Exponenciálne rozdelenie . . . . .	44
1.2.4.3	Normálne rozdelenie . . . . .	45
1.2.4.4	Ďalšie rozdelenia . . . . .	48
<b>2</b>	<b>Metódy štatistickej indukcie</b>	<b>51</b>
2.1	Základný a výberový súbor . . . . .	51

---

2.2	Teória odhadu . . . . .	52
2.2.1	Bodový odhad . . . . .	53
2.2.2	Intervalový odhad . . . . .	54
2.2.3	Intervaly spoľahlivosti na odhad parametrov normálneho rozdelenia . . . . .	55
2.2.3.1	Odhad strednej hodnoty, ak poznáme rozptyl . . . . .	55
2.2.3.2	Odhad strednej hodnoty, ak nepoznáme rozptyl . . . . .	55
2.2.3.3	Odhad rozptylu, ak nepoznáme strednú hodnotu . . . . .	56
2.3	Testovanie hypotéz . . . . .	59
2.3.1	Základné pojmy testovania hypotéz . . . . .	59
2.3.2	Jednovýberové testy o parametroch normálneho rozdelenia . . . . .	62
2.3.2.1	Test strednej hodnoty, ak poznáme rozptyl . . . . .	63
2.3.2.2	Test strednej hodnoty, ak nepoznáme rozptyl . . . . .	64
2.3.2.3	Test rozptylu základného súboru . . . . .	65
2.3.3	Dvojvýberové testy o parametroch normálneho rozdelenia . . . . .	68
2.3.3.1	Test zhody rozptylov dvoch nezávislých súborov . . . . .	69
2.3.3.2	Test zhody stredných hodnôt dvoch nezávislých súborov . . . . .	70
2.3.3.3	Test zhody stredných hodnôt dvoch závislých súborov . . . . .	72
2.3.4	Testy odľahlých hodnôt . . . . .	76
2.3.4.1	Grubbsov test . . . . .	77
2.3.4.2	Dixonov test . . . . .	77
2.3.5	Testy dobrej zhody . . . . .	80
2.3.5.1	$\chi^2$ -test dobrej zhody . . . . .	80
2.3.5.2	Kolmogorovov test . . . . .	84
2.3.5.3	Kolmogorovov-Smirnovov test . . . . .	87
2.3.6	Neparametrické testy . . . . .	91
2.3.6.1	Znamienkový test . . . . .	91
2.3.6.2	Jednovýberový Wilcoxonov test . . . . .	93
2.3.6.3	Dvojvýberový Wilcoxonov test . . . . .	96
2.3.7	Testovanie normality . . . . .	99
2.3.8	Testovanie homogénosti viac ako dvoch základných súborov . . . . .	103
2.3.8.1	Bartlettov test . . . . .	103
2.3.8.2	Cochranov test, Hartleyov test . . . . .	104
2.4	Analýza rozptylu . . . . .	109

2.4.1	Základné pojmy analýzy rozptylu . . . . .	110
2.4.2	Jednofaktorová analýza rozptylu . . . . .	110
2.4.3	Metódy mnohonásobného porovnávania pre jednofakto- rovú analýzu rozptylu . . . . .	117
2.4.4	Kruskal-Wallisov test . . . . .	121
<b>3</b>	<b>Závislosť kvantitatívnych a kategoriálnych premenných</b>	<b>129</b>
3.1	Závislosť kvantitatívnych premenných . . . . .	129
3.1.1	Regresná analýza . . . . .	129
3.1.1.1	Jednoduchý lineárny regresný model . . . . .	131
3.1.1.2	Nelineárne regresné modely . . . . .	139
3.1.2	Korelačná analýza . . . . .	149
3.1.2.1	Koeficient korelácie, koeficient determinácie . . . . .	149
3.1.2.2	Index determinácie, index korelácie . . . . .	154
3.1.2.3	Spearmanov koeficient poradovej korelácie . . . . .	156
3.1.3	Viacnásobná lineárna regresia . . . . .	161
3.2	Závislosť kategoriálnych premenných . . . . .	168
3.2.1	Kontingenčná a asociačná tabuľka . . . . .	168
3.2.2	Posúdenie závislosti kategoriálnych premenných . . . . .	170
3.2.3	Koeficienty (miery) kontingencie . . . . .	173
<b>4</b>	<b>Úvod do analýzy časových radov</b>	<b>177</b>
4.1	Časové rady, základné pojmy . . . . .	177
4.1.1	Analýza časových radov . . . . .	178
4.1.2	Prognóza časového radu . . . . .	179
4.2	Základné spracovanie časového radu . . . . .	180
4.2.1	Grafická analýza časového radu . . . . .	180
4.2.2	Agregácia hodnôt . . . . .	181
4.2.2.1	Agregácia hodnôt intervalových časových radov	182
4.2.2.2	Agregácia hodnôt okamihových časových radov	183
4.2.3	Kalendárna úprava . . . . .	185
4.2.4	Základné charakteristiky časového radu . . . . .	186
4.3	Dekompozícia časového radu . . . . .	192
4.4	Analýza trendovej zložky . . . . .	193
4.4.1	Popis trendovej zložky pomocou regresnej analýzy . . . . .	194
4.4.1.1	Lineárne alebo linearizovateľné modely trendu . . . . .	195
4.4.1.2	Modifikovaný exponenciálny trend . . . . .	199
4.4.1.3	Logistický trend . . . . .	200
4.4.1.4	Gompertzov trend . . . . .	202



4.4.2	Metóda kľzavých priemerov . . . . .	207
4.4.3	Exponenciálne vyrovnanie časového radu . . . . .	217
4.5	Analýza sezónnej zložky . . . . .	221
4.6	Verifikácia vhodnosti modelu . . . . .	228
4.6.1	Grafická analýza . . . . .	228
4.6.2	Absolútne a relatívne miery presnosti . . . . .	229
4.6.3	Index determinácie . . . . .	233
4.6.4	Durbin-Watsonova charakteristika . . . . .	233
4.6.5	Koeficient autokorelácie . . . . .	234
<b>5</b>	<b>Úvod do štatistických metód riadenia kvality</b>	<b>237</b>
5.1	Základné pojmy . . . . .	237
5.2	Tradičné nástroje riadenia kvality . . . . .	239
5.2.1	Kontrolný formulár . . . . .	239
5.2.2	Vývojový diagram . . . . .	240
5.2.3	Diagram príčin a následkov . . . . .	240
5.2.4	Histogram . . . . .	241
5.2.5	Korelačný diagram . . . . .	243
5.2.6	Pareto diagram . . . . .	243
5.2.7	Regulačný diagram . . . . .	248
5.3	Klasické regulačné diagramy . . . . .	249
5.3.1	Podstata klasických regulačných diagramov . . . . .	249
5.3.1.1	Nenáhodné zoskupenia . . . . .	251
5.3.2	Regulačné diagramy meraním . . . . .	252
5.3.2.1	Diagram $\bar{x} - R$ . . . . .	253
5.3.2.2	Diagram $\bar{x} - s$ . . . . .	258
5.3.2.3	Diagram $\tilde{x} - R$ . . . . .	259
5.3.2.4	Diagram $x - R_k$ . . . . .	260
5.3.3	Regulačné diagramy porovnávaním . . . . .	265
5.4	Spôsobilosť výrobného procesu . . . . .	265
5.4.1	Index spôsobilosti pre znak kvality s normálnym rozdelením . . . . .	267
5.4.1.1	Index spôsobilosti $C_p$ . . . . .	267
5.4.1.2	Index spôsobilosti $C_{pk}$ . . . . .	268
5.4.1.3	Index spôsobilosti $C_{pm}$ . . . . .	269
5.4.1.4	Index spôsobilosti $C_{pm}^*$ . . . . .	269
5.4.1.5	Index spôsobilosti $C_{pmk}$ . . . . .	270
5.4.1.6	Metodika VDA . . . . .	270
5.4.2	Indexy spôsobilosti pri inom ako normálnom rozdelení . . . . .	274

---

5.4.3 Výkonnosť výrobného procesu . . . . .	276
<b>Dodatok A Štatistické tabuľky</b>	<b>279</b>
<b>Dodatok B Program R</b>	<b>305</b>
<b>Literatúra</b>	<b>315</b>

# Kapitola 1

## Základy štatistických metód

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. Vedné odbory ako medicína, ekonómia, fyzika, biológia, chémia a mnohé ďalšie prírodovedné a technické odbory bežne aplikujú uvedené metódy. Na základe štatistiky sa neustále o niečom rozhoduje a toto rozhodovanie ovplyvňuje naše konanie a náš každodenný život.

Pôvodný význam slova *štatistika* pochádza z latinského slova *status*, čo znamená *stav* a v slovnom spojení *status rei publicae* vyjadruje *stav veci verejnej* alebo *štát*. Slovo *statistik* sa začalo používať až okolo roku 1750. Zaviedol ho nemecký Gottfried Achenwall (1719-1772) a týkal sa výhradne analýzy dát o štáte. Pojem štatistika získal všeobecný význam zberu a analýzy údajov až začiatkom 19. storočia. Dnes 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.

*Predmetom štatistiky* ako vednej disciplíny sú *hromadné náhodné javy*, ktoré sa za presne definovaných podmienok vecných, časových a priestorových viackrát vyskytujú, resp. opakujú. Skúmaním hromadného javu poznáme podstatu javu, pravidelnosti, vzťahy a jeho vývoj. Bez hromadného pozorovania by sme nemohli o príslušnom jave robiť zovšeobecňujúce závery.

V dnešnej dobe, v súvislosti s rozvojom výpočtovej techniky, sa na štatistické spracovanie a komplexnú analýzu výskumu používajú predovšetkým počítače. Najaktuálnejšia ponuka programov pre oblasť štatistických výpočtov je široká: od najznámejších komerčných softvérov (napr. Statistica, Statgraphics, SAS, SPSS a i.), cez matematické programy Matlab, Maple, Mathematica, tabuľkový program Excel, až po voľne šíriteľné programy (tzv. Open Source Software), ako R program, Maxima, Octave a iné. K výhodám používania počítačovej techniky a štatistického softvéru, resp. vhodnej aplikácie patrí rýchlosť a presnosť spracovania údajov, univerzálnosť, kvalitné grafické výstupy, množstvo spracovaných údajov a pod. Počítač a príslušný softvér nám prácu pri štatistickom spracovávaní síce výrazne zjednodušia, ale znalosť algoritmu výpočtu sa nahrádza v mnohých prípadoch (v závislosti od softvéru) iba schopnosťou spustiť analýzu na počítači. Je nutné si uvedomiť, že bez poznania podstaty danej štatistickej metódy, nie je možný výber vhodnej metódy a ani následná analýza získaných výsledkov.

## 1.1 Popisná štatistika

*Popisná (deskriptívna) štatistika sa zaoberá zisťovaním, spracovaním, analyzovaním a prezentáciou dát pomocou štatistických prostriedkov a metód. Úlohou štatistického zisťovania je zozbierať štatistické údaje o skúmaných hromadných javoch alebo procesoch. Výsledkom štatistického zisťovania sú údaje, ktoré sú pomocou štatistického spracovania prehľadne usporiadané, roztriedené a spracované. V poslednej, najdôležitejšej etape sa získané výsledky analyzujú, vyhodnocujú a zo získaných výsledkov formulujú závery.*

### 1.1.1 Základné pojmy

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

*Štatistická jednotka* je základný prvok, na ktorom pozorujeme konkrétny prejav určitého hromadného javu (napr. osoby, výrobok a pod.). Množinu všetkých štatistických jednotiek, ktoré majú požadované spoločné vlastnosti, nazývame *štatistický súbor*. *Rozsah štatistického súboru* charakterizuje počet štatistických jednotiek v skúmanom štatistickom súbore. *Štatistický znak* predstavuje vlastnosť hromadného javu, ktorá je predmetom štatistického skúmania (hmotnosť

súčiastky, veľkosť rázovej sily, životnosť zariadenia a pod.).

V praxi sa štatistické znaky často delia podľa spôsobu vyjadrenia do dvoch základných skupín: 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 diskrétné 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 dopravných nehôd). Všeobecné delenie znakov je na Obr. 1.1.



Obr. 1.1: Delenie znakov

*Nominálne znaky* zodpovedajú kvalitatívnym znakom a majú diskrétny, neusporiadateľné hodnoty. Príkladom nominálneho znaku je typ použitého dopravného pásu (nový, opotrebovaný, skladovaný), pohlavie respondenta v dotazníkovom výskume (muž, žena) a pod. *Ordinálne (poradové) znaky* predstavujú znaky, ktoré nadobúdajú diskrétny hodnoty a ktoré môžeme jednoznačne usporiadať. Sem patrí napríklad stupeň spokojnosti zákazníka (veľmi spokojný, spokojný, nespokojný, veľmi nespokojný), tepelný pocit zamestnanca na pracovisku (teplo, mierne teplo, neutrálne, mierne chladno, chladno) a pod. Ordinálne premenné môžeme vyjadriť nielen slovné, ale aj číselne pomocou bodovacej škály. Špeciálnu skupinu kvalitatívnych znakov tvoria tzv. *alternatívne*

(*dichotomické*) znaky, ktoré nadobúdajú len dve možné kategórie (kvalitný, nekvalitný výrobok) a *množné* (*viackategoriálne*), ktoré nadobúdajú viac ako dve kategórie (nezávažné, menej závažné, závažné poškodenie výrobku). *Intervalové znaky* umožňujú nielen zoradenie objektov, ale aj kvantifikáciu a porovnanie rozdielov medzi nimi (o koľko je väčšia, resp. menšia prvá hodnota ako druhá hodnota, napr. životnosť výrobku, odpracovaná doba). *Pomerový znak* umožňuje určiť, koľkokrát je jedna hodnota väčšia (resp. menšia) než druhá (napr. veľkosť napínacej sily). Nominálne a ordinárne znaky sú označované ako kvalitatívne, intervalové a pomerové ako kvantitatívne znaky.

### 1.1.2 Triedenie štatistického súboru

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

*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.

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).

#### 1.1.2.1 Jednoduché triedenie

Nech  $x_1, x_2, x_3, \dots, x_n$  sú získané hodnoty sledovaného štatistického znaku. Každá rôzna hodnota štatistického znaku predstavuje jednu triedu. Nech celkový počet rôznych hodnôt štatistického znaku je  $k$  a nech rozsah štatistic-

kého súboru je  $n$ , platí  $k \leq n$ . Zapísaním hodnôt štatistického znaku v poradí, v akom sme ich získali, dostaneme tzv. *prvotnú tabuľku*. Ak usporiadame hodnoty štatistického znaku podľa veľkosti od najmenej po najväčšiu hodnotu, tak získame *variačný rad*. Variačný rad môžeme zapísať v tvare

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

Z variačného radu vytvoríme pomocou čiarkovacej metódy *variačnú tabuľku početností*, ktorá slúži ako prvotný názorný spôsob prezentovania výsledkov zisťovania (Tab. 1.1).

Tabuľka 1.1: Variačná tabuľka 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	$x_2$	$n_2$	$f_2 = \frac{n_2}{n}$	$N_2 = n_1 + n_2$	$F_2 = \frac{N_2}{n}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$k$	$x_k$	$n_k$	$f_k = \frac{n_k}{n}$	$N_k = n$	$F_k = 1$

*Absolútna početnosť*  $n_j$  hodnoty štatistického znaku  $x_j$  predstavuje počet štatistických jednotiek s rovnakou hodnotou  $x_j$ ,  $j = 1, 2, 3, \dots, k$ .

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

$$f_j = \frac{n_j}{n}, \quad j = 1, 2, 3, \dots, k. \quad (1.1)$$

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

$$n_1 + n_2 + \dots + n_j = \sum_{i=1}^j n_i = N_j. \quad (1.2)$$

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

$$F_j = f_1 + f_2 + \dots + f_j = \frac{n_1 + n_2 + \dots + n_j}{n} = \frac{N_j}{n}. \quad (1.3)$$

Relatívne početnosti sa často vyjadrujú v percentách. Hodnota  $100 f_j\%$  udáva, koľko percent prvkov zo štatistického súboru má hodnotu sledovaného štatistického znaku  $x_j$ . Hodnota  $100 F_j\%$  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

$$\sum_{j=1}^k n_j = N_k = n, \quad \sum_{j=1}^k f_j = F_k = 1. \quad (1.4)$$

**Príklad 1.** Počas 15 pracovných dní sme zisťovali počet chybných výrobkov za smenu. Zistili sme nasledujúce hodnoty (v ks): 8; 10; 9; 7; 9; 8; 10; 7; 10; 9; 8; 8; 9; 8; 7. Zostavme kompletnú tabuľku početností.

#### Riešenie.

Rozsah štatistického súboru je  $n = 15$ . V štatistickom súbore sa opakujú štyri rôzne hodnoty (7, 8, 9, 10) a tak variačná tabuľka početností bude mať štyri riadky.

Trieda	Hodnoty znaku	Početnosť		Kumulatívna početnosť	
		absolútna	relatívna	absolútna	relatívna
$j$	$x_j$	$n_j$	$f_j$	$N_j$	$F_j$
1	7	3	$\frac{3}{15} = 0,200$	3	$\frac{3}{15} = 0,200$
2	8	5	$\frac{5}{15} = 0,333$	3+5=8	$\frac{8}{15} = 0,533$
3	9	4	$\frac{4}{15} = 0,267$	8+4=12	$\frac{12}{15} = 0,800$
4	10	3	$\frac{3}{15} = 0,200$	12+3=15	$\frac{15}{15} = 1,000$

Z variačnej tabuľky napríklad zistíme, že počas piatich pracovných dní bolo zaznamenaných 8 chybných výrobkov z celkového počtu 15 záznamov. Počet dní, v ktorých bolo zistených najviac 9 chybných výrobkov je 12, čo v relatívnom vyjadrení predstavuje 80,0% ( $100 F_3\%$ ).



### 1.1.2.2 Intervalové triedenie

Intervalové triedenie sa používa predovšetkým v prípade veľkého počtu rôznych hodnôt štatistického znaku, ktoré sa spravidla málo alebo vôbec neopakujú. Ide o 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 musíme 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  $I_j = (t_{j-1}; t_j)$ .

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

$$z_j = \frac{t_{j-1} + t_j}{2}, \quad j = 1, 2, \dots, k, \quad (1.5)$$

kde  $k$  je počet triednych intervalov.

Šírku triedneho intervalu  $h$  môžeme približne vypočítať podľa vzťahu

$$h \doteq \frac{R_V}{k}, \quad (1.6)$$

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

$$R_V = x_{(n)} - x_{(1)} = x_{\max} - x_{\min}. \quad (1.7)$$

Na určenie optimálneho počtu triednych intervalov  $k$  existuje niekoľko odporúčaných postupov a spôsobov. Uvedieme niekoľko vzťahov:

$$k \leq 5 \log n, \quad (1.8)$$

$$k \approx 1 + 3,322 \log n, \quad (1.9)$$

$$0,55 n^{0,4} \leq k \leq 1,25 n^{0,4}. \quad (1.10)$$

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

**Príklad 2.** Pri procese obrábania komponentov do automobilov vzniká materiál mikroskopických rozmerov, tzv. mikrotrieska. Namerané údaje sú (v  $\mu\text{m}$ ): 6,7; 3,9; 4,2; 9,1; 13,7; 9,6; 19,1; 21,9; 12,4; 14,5; 11,4; 17,2; 18,1; 23,2;

17,9; 11,8; 16,2; 10,5; 22,9; 9,8; 5,3; 10,2; 11,4; 12,3; 20,9; 22,1; 7,1; 6,4; 14,5; 16,2; 18,1; 17,3; 9,2; 8,9; 15,8. Zostavme kompletnú tabuľku početností.

### Riešenie.

Zo získaných 35 údajov vieme, že minimálny rozmer mikrotriesky je 3,9  $\mu\text{m}$  a maximálny 23,2  $\mu\text{m}$ . Počet *triednych intervalov*  $k$  nájdeme pomocou vzorca (1.10)

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

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  $h$  približne vypočítame podľa vzťahu (1.6)

$$h \doteq \frac{R_V}{k} = \frac{x_{\max} - x_{\min}}{k} = \frac{23,2 - 3,9}{4} = 4,825,$$

pričom budeme uvažovať hodnotu 4,9. Pri špecifikácii triednych intervalov volíme interval tak, aby horná hranica patrila do intervalu. Výsledok intervalového triedenia je v nasledujúcej tabuľke.

$j$	$(t_{j-1}; t_j)$	$z_j$	$n_j$	$f_j$	$N_j$	$F_j$
1	(3,8; 8,7)	6,25	6	$\frac{6}{35} = 0,171$	6	$\frac{6}{35} = 0,171$
2	(8,7; 13,6)	11,15	12	$\frac{12}{35} = 0,343$	18	$\frac{18}{35} = 0,514$
3	(13,6; 18,5)	16,05	11	$\frac{11}{35} = 0,314$	29	$\frac{29}{35} = 0,829$
4	(18,5; 23,4)	20,95	6	$\frac{6}{35} = 0,171$	35	$\frac{35}{35} = 1,000$

Každému triednemu intervalu určíme triedny znak ako stred intervalu  $z_j$ . Namerané údaje rozdelíme tak, aby sa každá hodnota nachádzala práve v jednom triednom intervale. Výsledkom triedenia je variačná tabuľka rozdelenia početností. Z tabuľky zistíme, že v súbore sa nachádza 12 vzoriek mikrotriesky, ktorých rozmer je v intervale (8,7; 13,6), čo predstavuje približne 34,30% (100  $f_2$ %) z celkového počtu vzoriek. Počet vzoriek, ktoré majú rozmer nanajvyš 18,5  $\mu\text{m}$  je 29, čo v relatívnom vyjadrení predstavuje 82,90% (100  $F_3$ %).

### 1.1.3 Číselné charakteristiky

Štatistické spracovanie údajov pomocou tabuliek a grafov nám môže uľahčiť analýzu údajov, ale nie je to vždy postačujúce. Na ďalšie popísanie štatistického

súboru zavádzame *číselné charakteristiky (miery)*. Patria sem charakteristiky polohy, variability, šikmosti a špicatosti.

### 1.1.3.1 Charakteristiky polohy

*Charakteristiky polohy* vyjadrujú určitú úroveň (polohu) znaku, okolo ktorej sú ostatné hodnoty viac či menej koncentrované. Môžeme ich rozdeliť na *stredné hodnoty* a *kvantily*. Ak sa stredné hodnoty počítajú pomocou všetkých štatistických jednotiek v súbore, tak hovoríme o *priemeroch* (napr. aritmetický, geometrický, harmonický, kvadratický). Do druhej skupiny stredných hodnôt (tzv. *ostatné stredné hodnoty*) zaraďujeme tie, ktoré sú založené len na vybraných hodnotách štatistického súboru. Patrí sem napríklad modus a medián.

K najčastejšie používaným priemerom patrí *aritmetický priemer* ( $\bar{x}$ ). Jeho nevýhodou je veľká citlivosť na odľahlé, extrémne hodnoty a možný fiktívny charakter vypočítanej hodnoty. Aritmetický priemer je definovaný ako súčet nameraných hodnôt  $x_i$ ,  $i = 1, 2, \dots, n$  štatistického znaku delený ich počtom  $n$ , t. j.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (1.11)$$

Ak sú hodnoty štatistického znaku roztriedené do  $k$  tried, tak aritmetický priemer je daný vzťahom

$$\bar{x} = \frac{1}{n} \sum_{j=1}^k x_j n_j. \quad (1.12)$$

Ak sú hodnoty štatistického znaku roztriedené do  $k$  triednych intervalov, tak aritmetický priemer je daný vzťahom

$$\bar{x} = \frac{1}{n} \sum_{j=1}^k z_j n_j, \quad (1.13)$$

kde  $z_j$  reprezentuje stred  $j$ -teho intervalu.

Jednou z vlastností aritmetického priemeru je, že súčet odchýlok jednotlivých hodnôt  $x_i$  štatistického znaku od aritmetického priemeru  $\bar{x}$  sa rovná nule, t. j.

$$\sum_{i=1}^n (x_i - \bar{x}) = 0. \quad (1.14)$$

*Geometrický priemer* ( $\bar{x}_G$ ) je definovaný ako  $n$ -tá odmocnina zo súčiny hodnôt štatistického znaku, t. j.

$$\bar{x}_G = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 x_2 x_3 \dots x_n}. \quad (1.15)$$

Ak sú hodnoty štatistického znaku roztriedené do  $k$  tried, tak geometrický priemer je daný vzťahom

$$\bar{x}_G = \sqrt[n]{\prod_{j=1}^k x_j^{n_j}}. \quad (1.16)$$

Ak sú hodnoty roztriedené do  $k$  triednych intervalov, tak geometrický priemer je daný vzťahom

$$\bar{x}_G = \sqrt[n]{\prod_{j=1}^k z_j^{n_j}}. \quad (1.17)$$

*Modus* ( $\hat{x}$ ,  $Mo$ ) je definovaný ako hodnota štatistického znaku, ktorá sa v štatistickom súbore vyskytuje najčastejšie. Nemusí byť určený jednoznačne. V prípade, že hodnoty sú roztriedené do intervalov, modus vypočítame podľa vzťahu

$$\hat{x} = a_0 + h \frac{d_1}{d_1 + d_2}, \quad (1.18)$$

kde  $a_0$  začiatok modálneho interval (t. j. intervalu, v ktorom sa nachádza modus),  $h$  je šírka intervalu,  $d_1$  rozdiel absolútnych početností modálneho intervalu a predchádzajúceho intervalu,  $d_2$  rozdiel absolútnych početností modálneho intervalu a nasledujúceho intervalu.

*Medián* ( $\tilde{x}$ ,  $Me$ ) je hodnota štatistického znaku z radu hodnôt zoradených podľa veľkosti, ktorá delí tento rad na dve rovnako početné časti. Inak povedané, je to prostredná hodnota štatistického súboru usporiadaného do variačného radu. Ak máme nepárny počet hodnôt, tak medián je prostredná hodnota, pre ktorý platí

$$\tilde{x} = x_{\left(\frac{n+1}{2}\right)}. \quad (1.19)$$

Ak rozsah súboru  $n$  je párne číslo, tak medián vypočítame podľa vzťahu

$$\tilde{x} = \frac{1}{2} \left( x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n+2}{2}\right)} \right). \quad (1.20)$$