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.

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, zhluková 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

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 republicae* vyjadruje *stav veci verejnej* alebo *štát*. Slovo *statistik* sa začalo používať až okolo roku 1750. Zaviedol ho nemec 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) štatistka 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 slovne 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). *Diskrétne 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 diskrétne, 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 diskrétne 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 slovne, 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 diskrétny 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, $x_2, x_3, \ldots, x_n$ 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 štatistického 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 najmenšej po najväčšiu hodnotu, tak získame *variačný rad*. Variačný rad môžeme zapísať v tvare

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

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	Početnosť		Kumulatívna početnosť	
	znaku	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, $\; j = 1, 2, 3, \ldots, k$.

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

$$f_j = \frac{n_j}{n}, \; j = 1, 2, 3, \ldots, k. \tag{1.1}$$

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

$$n_1 + n_2 + \cdots + n_j = \sum_{i=1}^{j} n_i = N_j. \tag{1.2}$$

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

$$F_j = f_1 + f_2 + \cdots + f_j = \frac{n_1 + n_2 + \cdots + n_j}{n} = \frac{N_j}{n}. \tag{1.3}$$

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

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

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	Početnosť		Kumulatívna početnosť	
	znaku	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 nanajvýš 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 $(t_{j-1}; t_j\rangle$, resp. $\langle t_{j-1}; t_j)$. V nasledujúcej časti budeme pracovať s typom intervalu $I_j = (t_{j-1}; t_j\rangle$.

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

$$z_j = \frac{t_{j-1} + t_j}{2}, \; j = 1, 2, \ldots, k, \tag{1.5}$$

kde k je počet triednych intervalov.

Šírku triedneho intervalu h môžeme približne vypočítať podľa vzťahu

$$h \doteq \frac{R_V}{k}, \tag{1.6}$$

kde R_V je *variačné rozpätie*, pre ktoré platí

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

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

$$k \leq 5 \log n, \tag{1.8}$$

$$k \approx 1 + 3{,}322 \log n, \tag{1.9}$$

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

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 µm): 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 µm a maximálny 23,2 µm. 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	$(t_{j-1};\, t_j\rangle$	z_j	n_j	f_j	N_j	F_j
1	$(3{,}8;\, 8{,}7\rangle$	6,25	6	$\frac{6}{35} = 0{,}171$	6	$\frac{6}{35} = 0{,}171$
2	$(8{,}7;\, 13{,}6\rangle$	11,15	12	$\frac{12}{35} = 0{,}343$	18	$\frac{18}{35} = 0{,}514$
3	$(13{,}6;\, 18{,}5\rangle$	16,05	11	$\frac{11}{35} = 0{,}314$	29	$\frac{29}{35} = 0{,}829$
4	$(18{,}5;\, 23{,}4\rangle$	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 $(8{,}7;\, 13{,}6\rangle$, čo predstavuje približne 34,30% ($100\, f_2$%) z celkového počtu vzoriek. Počet vzoriek, ktoré majú rozmer nanajvýš 18,5 µm 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* ($\overline{x}$). 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, $i = 1, 2, \ldots, n$ štatistického znaku delený ich počtom n, t. j.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{1.11}$$

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

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{k} x_j\, n_j. \tag{1.12}$$

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

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{k} z_j\, n_j, \tag{1.13}$$

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 $\overline{x}$ sa rovná nule, t. j.

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

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

$$\overline{x}_G = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1\, x_2\, x_3 \ldots x_n}. \tag{1.15}$$

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

$$\overline{x}_G = \sqrt[n]{\prod_{j=1}^{k} {x_j}^{n_j}}. \tag{1.16}$$

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

$$\overline{x}_G = \sqrt[n]{\prod_{j=1}^{k} {z_j}^{n_j}}. \tag{1.17}$$

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

$$\hat{x} = a_0 + h\,\frac{d_1}{d_1 + d_2}, \tag{1.18}$$

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 ($\widetilde{x}$, 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

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

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

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