

7.5 Statistika

Statistické údaje neboli **statistická data** jsou číselné údaje o **hromadných jevech**, tj. čísla o různých skutečnostech (přírodních, společenských či jiných jevech) sledovaných ne jednotlivě, ale ve velkém počtu případů (hromadě). Na základě nich jednak objevujeme zákonitosti, které se projevují teprve při studiu hromadných jevů, jednak nabýváme spolehlivějších podkladů pro svá rozhodnutí.

Příklady statistických údajů:

Číselné údaje o počtu obyvatelstva, o objemu výroby ve státě či v jednotlivých podnicích, o dovozu nebo vývozu určitého zboží, o příjmech a výdajích obyvatelstva, o kvantitativně určených technologických vlastnostech výrobků a mnohé další.

Statistika ve smyslu **statistické činnosti** znamená získávání statistických údajů (pozorováním, měřením, vážením apod.), jejich zpracování (třídění, výpočet statistických charakteristik, předkládání výsledků) a hodnocení.

Statistika jako **statistická teorie** je věda o metodách sběru, zpracování a vyhodnocování statistických údajů. **Matematická statistika** vychází ze shromážděných statistických údajů a zabývá se jejich matematickým (statistickým) zpracováním a rozбором získaných výsledků.

Statistický soubor, statistické jednotky

Množinu všech objektů statistického pozorování (osob, věcí, jevů apod.) shromážděných na základě toho, že mají jisté společné vlastnosti, nazýváme **statistickým souborem**. Jednotlivé prvky této množiny se nazývají **prvky (elementy) statistického souboru** nebo též **statistické jednotky**. Počet všech prvků statistického souboru se nazývá **rozsah souboru** n .

Příklady statistických souborů:

Při sčítání lidu je statistickým souborem množina všech obyvatel státu nebo množina všech domácností, při soupisu domů v určitém místě představuje statistický soubor množina všech domů v daném místě, při zkoumání průmyslové produkce v podnicích určitého resortu je statistickým souborem množina všech podniků daného resortu apod.

Věcné, časové a prostorové vymezení statistického pozorování (zjišťování) jednoznačně určuje soubor všech v dané situaci v úvahu přicházejících statistických jednotek, jenž se nazývá **základní statistický soubor**. Statistické zjišťování zaměřené na všechny jednotky základního souboru nazýváme **úplné (vyčerpávající) zjišťování**. Často však statistika provádí zjišťování zaměřené jen na část jednotek základního souboru, takové zjišťování se nazývá **výběrové** a odpovídající statistické soubory jsou **výběrové soubory**. Důvodem k výběrovému zjišťování je především okolnost, že základní soubory mohou být natolik rozsáhlé, že zkoumání všech jednotek by bylo příliš pracné, zdlouhavé, nákladné nebo dokonce neproveditelné. Dalším důvodem je i skutečnost, že při **náhodném výběru** můžeme využít teorie pravděpodobnosti k dostatečně spolehlivým a přesným úsudkům o charakteru základního souboru.

Statistické znaky

Společnou vlastnost statistických jednotek (prvků statistického souboru), jejíž proměnnost je předmětem statistického zkoumání (pozorování a zpracování), nazýváme **statistickým znakem**; zpravidla se značí x . Jednotlivé údaje znaku se nazývají **hodnoty znaku**; značí se $x_1, x_2, \dots, x_n$.

Příklady statistických znaků:

Při sčítání lidu je jimi věk, pohlaví, zaměstnání obyvatel apod., při zkoumání průmyslové produkce podniků výše plánované a skutečné produkce, počet pracovníků, velikost mzdových fondů apod.

Hodnoty znaků mohou být vyjádřeny buď čísly, nebo jiným způsobem (zpravidla slovním popisem). V prvním případě mluvíme o **znacích kvantitativních**, např. to může být tělesná výška, váha, množství produkce, počet pracovníků apod. Ve druhém případě mluvíme o **znacích kvalitativních**, které se mohou vyskytovat ve dvou druzích (znaky alternativní, např. muž — žena, voják — nevoják, prospěl — neprospěl) nebo ve více druzích (např. povolání, národnost, náboženství, rodinný stav).

Četnosti, jejich rozdělení a grafické znázornění

Předpokládejme, že při statistickém šetření (zjišťování) nás zajímá jistý statistický znak x , jenž nabývá hodnot $x_1, x_2, \dots, x_n$, kde n je rozsah uvažovaného statistického souboru. Nechť celkový počet různých hodnot znaku x je $k \leq n$. Rovnost platí jen v případě, že každá jednotka souboru nabývá jiné hodnoty sledovaného znaku. V praxi je tento případ výjimečný, zejména u rozsáhlých souborů dochází k opakovanému výskytu stejného znaku pro různé statistické jednotky.

Počet statistických jednotek, jimž přísluší stejná hodnota znaku x_j pro $j = 1, 2, \dots, k$, se nazývá **četnost (absolutní četnost) hodnoty znaku x_j** ; označíme ji n_j . Podíl četnosti n_j hodnoty znaku x_j a **rozsahu souboru n** se nazývá **relativní četnost hodnoty znaku x_j** ; budeme ji značit v_j , tedy

$$v_j = \frac{n_j}{n}.$$

Součet všech četností je rozsah souboru:

$$\sum_{j=1}^k n_j = n_1 + n_2 + \dots + n_k = n,$$

a proto součet všech relativních četností je roven jedné:

$$\begin{aligned} \sum_{j=1}^k v_j &= v_1 + v_2 + \dots + v_k = \frac{n_1}{n} + \frac{n_2}{n} + \dots + \frac{n_k}{n} = \\ &= \frac{1}{n}(n_1 + n_2 + \dots + n_k) = 1 \end{aligned}$$

V praxi se relativní četnosti vyjadřují v procentech; číselnou hodnotu v procentech získáváme vynásobením relativní četnosti stem.

Uspořádáme-li všechny různé hodnoty znaku a jim odpovídající četnosti do tabulky 7.7, nazýváme ji **tabulkou rozdělení četností**. Obdobně se sestavuje **tabulka rozdělení relativních četností**, resp. spojená **tabulka rozdělení četností a relativních četností**.

ny x_j pro $j = 1, 2, \dots, k$, kde k je celkový počet třídních intervalů uvažovaného statistického souboru.

Poznámka. K určení vhodného počtu k intervalů se užívají různé empirické vzorce, např. tzv. Sturgesův vzorec:

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

kde n je rozsah statistického souboru.

Četnost výskytu hodnot statistického znaku v určitém třídním intervalu charakterizovaném třídním znakem x_j budeme opět značit n_j a příslušnou relativní četnost v_j . Jejich určením pro $j = 1, 2, \dots, k$ získáváme **skupinové (intervalové) rozdělení četností**, které zapisujeme zpravidla ve formě tabulky.

Příklad sestavení tabulky skupinového rozdělení četností:

Byly naměřeny výšky 300 osob v mezích od 153 do 197 cm. Navrhněte (podle Sturgesova vzorce) jejich rozdělení do skupin (intervalů) a popište sestavení tabulky skupinového (intervalového) rozdělení četností.

Řešení. Zvolíme výškové intervaly těže délky, jejichž počet je podle Sturgesova vzorce $k \approx 1 + 3,3 \log 300 \approx 9$.

Pro každý j -tý interval ($j = 1, 2, \dots, 9$) určíme počet naměřených dat (výšek osob), které v něm leží, tj. jejich četnost n_j . Získáme tak např. následující tabulku 7.9.

Tabulka skupinového (intervalového) rozdělení četností řešené úlohy

Tab. 7.9

j	Intervaly výšky x (v cm)	Střed y intervalů x_j (v cm)	Četnosti n_j
1	153 — 157	155	7
2	158 — 162	160	20
3	163 — 167	165	35
4	168 — 172	170	49
5	173 — 177	175	84
6	178 — 182	180	60
7	183 — 187	185	27
8	188 — 192	190	14
9	193 — 197	195	4
Součet četností (rozsah souboru) n			300

Grafická znázornění rozdělení četností

Na základě četnostních tabulek lze provést **grafické znázornění rozdělení četností** v pravouhlé soustavě souřadnic, ve které na vodorovné ose x znázorňujeme hodnoty statistického znaku, resp. třídního intervalu nebo šířku třídních intervalů a na svislé ose y znázorňujeme jim odpovídající četnosti. Nejčastěji používaná grafická znázornění rozdělení četností jsou tzv. polygon a histogram četností.

Tabulka rozdělení četností

Tab. 7.7

j	x_j	n_j
1	x_1	n_1
2	x_2	n_2
...	...	...
k	x_k	n_k

Poznámka. Nejjednodušší způsob, jak sestavit tabulku rozdělení četností, je čárkovací metoda: Jednotlivé různé hodnoty sledovaného znaku zapíšeme pod sebe, pak probíráme postupně jednotky statistického souboru a zaznamenáváme čárkou výskyt hodnoty znaku do příslušného řádku. (Výhodné je pro přehlednost umístit každou pátou čárku vodorovně přes předchozí čtyři.) Sečtením čárek v jednotlivých řádcích obdržíme četnosti.

Příklad sestavení tabulky rozdělení četností:

Při zjišťování počtu nezletilých dětí ve třiceti vybraných rodinách byly získány tyto výsledky: 1, 1, 0, 2, 3, 4, 2, 2, 3, 0, 1, 2, 2, 4, 3, 3, 0, 1, 1, 1, 2, 2, 0, 2, 1, 1, 2, 3, 3, 2. Uspořádejte získané údaje do tabulky rozdělení četností, vypočítejte relativní četnosti a vyjádřete je v procentech.

Řešení. Označme x sledovaný znak (počet nezletilých dětí v domácnosti), x_j ($j = 1, 2, \dots, 5$) jeho hodnoty, n_j četnosti, v_j relativní četnosti jejich výskytu ve vyšetřovaných domácnostech. Jsou uvedeny v tabulce 7.8.

Tabulka rozdělení četností a relativních četností řešené úlohy

Tab. 7.8

j	x_j	n_j	v_j	v_j v %
1	0	4	0,13	13
2	1	8	0,27	27
3	2	10	0,33	33
4	3	6	0,20	20
5	4	2	0,07	7
součet		30	1,00	100

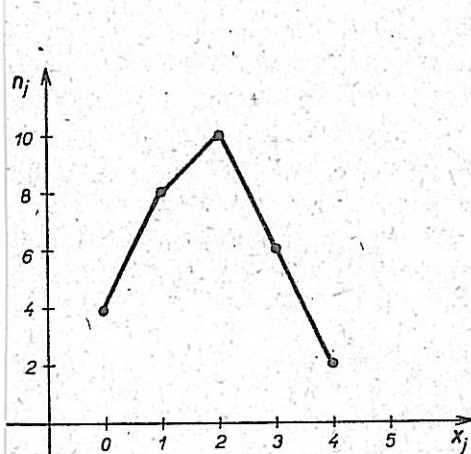
Skupinové rozdělení četností

V praxi se často stává, že rozsah statistického souboru je velký a také počet zjištěných hodnot kvantitativního statistického znaku je značný, takže z tohoto rozsáhlého a nepřehledného statistického materiálu se obtížně vyzvojují praktické závěry. Proto se blízké hodnoty znaku sdružují do **skupin (tříd)** tvořených obvykle **intervaly (třídními intervaly)**. Pokud je to možné, je nejjednodušší volit všechny intervaly téže délky (šířky). Hodnoty znaku, jež se dostaly do téhož intervalu, lze potom reprezentovat jedinou hodnotou — středem intervalu, tzv. **třídním znakem**. Třídní znaky budeme značit písmem-

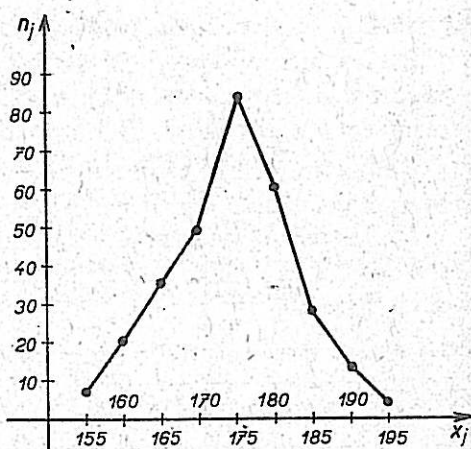
Polygon četnosti neboli **spojnicový diagram** získáme spojením bodů, jejichž první (x -ová) souřadnice je hodnota znaku, resp. třídního znaku a druhá (y -ová) souřadnice odpovídající četnost.

Histogram četnosti nebo **sloupkový diagram** tvoří množina obdélníků (se základnami na ose x), jejichž obsahy jsou přímo úměrné znázorňovaným četnostem. Je vhodný zejména pro znázornění skupinového (intervalového) rozdělení četností; základny obdélníků („sloupků“) na ose x mají délky rovné délkám (šířkám) třídních intervalů, pokud jsou vesměs stejné délky, pak výšky obdélníků udávají odpovídající četnosti.

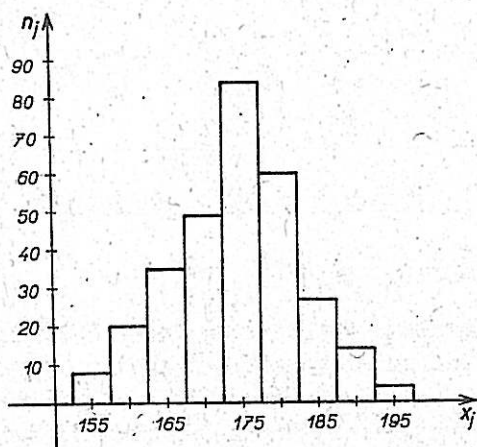
Poznámka. Analogicky lze graficky znázorňovat rozdělení relativních četností.



Obr. 7.5



Obr. 7.6



Obr. 7.7

Příklady sestrojení polygonu četnosti (spojnicového diagramu) a histogramu četnosti (sloupkového diagramu):

Znáznorníte graficky rozdělení četností podle tabulky 7.8 a skupinové (intervalové) rozdělení četností podle tabulky 7.9.

Řešení. V obr. 7.5 a 7.6 jsou sestrojeny polygony četnosti (spojnicové diagramy) k tab. 7.8 a k tab. 7.9. V obr. 7.7 je sestrojen histogram četnosti (sloupkový diagram) k tab. 7.9.

Charakteristiky statistického souboru

Statistickými charakteristikami nazýváme čísla, která podávají stručnou souhrnnou informaci o uvažovaném statistickém souboru z různých hledisek. Je-li předmětem našeho zájmu jediný kvantitativní znak, jde především o tzv. charakteristiky polohy (úrovně) a charakteristiky variability (proměnnosti, rozptýlení).

A. Charakteristiky polohy (úrovně) hodnot znaku neboli jeho střední hodnoty jsou čísla, která nějak charakterizují „průměrnou hodnotu“ sledovaného kvantitativního znaku. Patří mezi ně zejména aritmetický průměr, dále medián, modus, harmonický průměr a geometrický průměr.

Aritmetický průměr $\bar{x}$ hodnot $x_1, x_2, \dots, x_n$ kvantitativního znaku x je definován jako podíl součtu hodnot znaku a jejich počtu (rozsahu souboru) n , tj. je dán vzorcem

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i. \quad (1)$$

Zabýváme-li se statistickým souborem většího rozsahu, zjednodušíme si výpočet aritmetického průměru užitím tabulky rozdělení četností znaku x . Má-li hodnota znaku x_1 četnost n_1 , hodnota x_2 četnost n_2 atd. až hodnota x_k četnost n_k (kde $n_1 + n_2 + \dots + n_k = n$), pak aritmetický průměr je dán vzorcem

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n} = \frac{1}{n} \sum_{i=1}^k n_i x_i. \quad (2)$$

Poznámka. Aritmetický průměr počítaný podle vzorce (2) lze považovat za vážený aritmetický průměr s váhami danými četnostmi $n_1, n_2, \dots, n_k$.

Významné vlastnosti aritmetického průměru:

a) Součet všech rozdílů $x_i - \bar{x}$ jednotlivých hodnot znaku x_i a jejich aritmetického průměru $\bar{x}$ se rovná nule: $\sum_{i=1}^n (x_i - \bar{x}) = 0$.

b) Přičteme-li, resp. odečteme-li od každé hodnoty znaku konstantní číslo $a > 0$, pak aritmetický průměr nových hodnot znaku se rovná aritmetickému průměru původních hodnot zvětšenému, resp. zmenšenému o číslo a .

c) Násobíme-li, resp. dělíme-li každou hodnotu znaku určitou konstantou $b \neq 0$, pak aritmetický průměr nových hodnot se rovná aritmetickému průměru původních hodnot násobenému, resp. dělenému konstantou $b \neq 0$.

Vlastností b) a c) se užívá ke zjednodušenému výpočtu aritmetického průměru tzv. metodou vhodně zvoleného počátku, která spočívá v tom, že zvolíme novou proměnnou $u_i = x_i - a$, vypočteme $\bar{u}$ a pak

$$\bar{x} = \bar{u} + a, \quad (3)$$

anebo při skupinovém rozdělení četností se stejně dlouhými intervaly zvolíme novou proměnnou $u_i = \frac{x_i - a}{b}$, vypočteme $\bar{u}$ a pak

$$\bar{x} = b\bar{u} + a. \quad (4)$$

Příklady výpočtu aritmetického průměru hodnot kvantitativního znaku:

1. Vypočtete aritmetický průměr počtu nezletilých dětí ve třiceti vyšetřovaných rodinách s výsledky uvedenými v tabulce 7.8.

Řešení. Aritmetický průměr počtu dětí vypočteme podle vzorce (2):

$$\bar{x} = \frac{0 \cdot 4 + 1 \cdot 8 + 2 \cdot 10 + 3 \cdot 6 + 4 \cdot 2}{30} = \frac{8 + 20 + 18 + 8}{30} = 1,8$$

2. Vypočtete aritmetický průměr výšek 300 osob, jež jsou uvedeny v tabulce 7.9.

Řešení. Aritmetický průměr výšek osob vypočteme opět podle vzorce (2):

$$\bar{x} = \frac{155 \cdot 7 + 160 \cdot 20 + \dots + 190 \cdot 14 + 195 \cdot 4}{300} = 174,4 \text{ (cm)}$$

Užití aritmetického průměru hodnot kvantitativního znaku statistického souboru je velmi časté. Avšak ne vždy musí být aritmetický průměr vhodnou charakteristikou polohy (úrovně) hodnot statistického znaku. Méně vhodný je především v situacích, kdy hodnoty znaku nejsou rovnoměrně rozložené kolem aritmetického průměru a v případech, kdy v souboru jsou extrémně nízké nebo vysoké hodnoty. Použití aritmetického průměru je zcela nevhodné, jestliže součet hodnot sledovaného znaku nemá věcný smysl.

V některých situacích se uplatní harmonický průměr nebo geometrický průměr.

Harmonický průměr $\bar{x}_H$ hodnot znaku $x_1, x_2, \dots, x_n$ je definován jako podíl rozsahu souboru a součtu převrácených hodnot znaku, tj. je dán vzorcem

$$\bar{x}_H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = n : \sum_{i=1}^n \frac{1}{x_i}. \quad (5)$$

Máme-li sestavenou tabulku rozdělení četností, podle níž hodnota znaku x_1 má četnost n_1 , hodnota x_2 četnost n_2 atd. až hodnota x_k četnost n_k (kde $n_1 + n_2 + \dots + n_k = n$), pak harmonický průměr těchto hodnot znaku vypočteme podle vzorce

$$\bar{x}_H = \frac{n}{n_1 \cdot \frac{1}{x_1} + n_2 \cdot \frac{1}{x_2} + \dots + n_k \cdot \frac{1}{x_k}} = n : \sum_{i=1}^k \frac{n_i}{x_i}. \quad (6)$$

Poznámka. Harmonický průměr počítaný podle vzorce (6) se často nazývá **vážený harmonický průměr s váhami** danými četnostmi $n_1, n_2, \dots, n_k$.

Užití harmonického průměru přichází v úvahu poměrně zřídka, má-li věcný smysl součet převrácených hodnot sledovaného znaku.

Příklady užití harmonického průměru hodnot kvantitativního znaku:

1. V určité dílně, v níž se vyrábějí stejné výrobky, byly naměřeny šesti dělníkům tyto časy potřebné ke zhotovení jednoho výrobku: 3, 4, 5, 6, 10, 12 minut. Určete dobu, které je v průměru třeba ke zhotovení jednoho výrobku.

Řešení. Výkony jednotlivých dělníků jsou velmi rozsáhlé, např. první vyrobí za tutéž dobu čtyřikrát více výrobků než poslední, proto postrádá věcný smysl počítat aritmetický průměr naměřených časů. Avšak součet jejich převrácených hodnot udává celkovou část produkce všech dělníků za 1 minutu a tedy průměrná doba potřebná ke zhotovení jednoho výrobku je dána harmonickým průměrem podle vzorce (5):

$$\begin{aligned} \bar{x}_H &= 6 : \left(\frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{10} + \frac{1}{12} \right) = \\ &= 6 : \frac{20 + 15 + 12 + 10 + 6 + 5}{60} = 6 \cdot \frac{60}{68} \approx 5,3 \text{ (min)} \end{aligned}$$

Poznámka. K těmto výsledku lze dospět touto úvahou: Za 1 hodinu, tj. za 60 minut, vyrobí jednotliví dělníci 20, 15, 12, 10, 6, 5 výrobků. Celkem tedy vyrobí 68 výrobků za 6 · 60 minut čili jeden výrobek vyrobí průměrně za $\frac{6 \cdot 60 \text{ minut}}{68} \approx 5,3$ minuty.

2. Máme údaje o plánované hrubé produkci (objemu výroby), skutečné hrubé produkci (objemu výroby) a odtud vypočteném procentu splnění plánu 5 závodů průmyslového podniku uvedené v tabulce 7.10. (Procento splnění plánu = $\frac{\text{skutečná hrubá produkce}}{\text{plánovaná hrubá produkce}} \cdot 100$.) Určete průměrné procento splnění plánu za celý podnik.

Řešení. Průměrné procento splnění plánu lze určit jako vážený harmonický průměr procent splnění plánu jednotlivých závodů podle vzorce (6), přičemž jako váhy (četnosti) vystupují skutečné hrubé produkce (objemy výroby) jednotlivých závodů:

Závod	Plánovaná hrubá produkce (v tis. Kč)	Skutečná hrubá produkce (v tis. Kč)	Procento splnění plánu
1	2 500	2 630	105,2
2	1 700	1 654,1	97,3
3	2 900	3 279,9	113,1
4	2 400	2 820	117,5
5	1 450	1 444,2	99,6
Celkem	10 950	11 828,2	$\bar{x}_H$

$$\bar{x}_H = 11\,828,2 : \left(\frac{2\,630}{105,2} + \frac{1\,654,1}{97,3} + \frac{3\,279,9}{113,1} + \frac{2\,820}{117,5} + \frac{1\,444,2}{99,6} \right) =$$

$$= 11\,828,2 : (25 + 17 + 29 + 24 + 14,5) = 11\,828,2 : 109,5 \doteq 108.$$

Poznámka. K těmto výsledkům lze dospět takto: Průměrné procento splnění plánu = $\frac{\text{celková skutečná hrubá produkce}}{\text{celková plánovaná hrubá produkce}} \cdot 100$. Z plánované hrubé produkce jednotlivých závodů v tabulce 7.10 je určena plánovaná hrubá produkce celého podniku 10 950 (v tis. Kč). Po dosazení do uvedeného vztahu dostáváme hledané průměrné procento splnění plánu $\frac{11\,828,2}{10\,950} \cdot 100 \doteq 108$.

Geometrický průměr $\bar{x}_G$ hodnot kvantitativního znaku $x_1, x_2, \dots, x_n$ je definován vzorcem

$$\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} \quad (7)$$

Geometrického průměru se ve statistice užívá prakticky jen k výpočtu průměrného ročního tempa růstu průmyslové, stavební nebo zemědělské výroby.

Příklady užití geometrického průměru hodnot kvantitativního znaku:

Jsou dány v procentech tyto údaje o růstu určitého druhu výroby v devíti po sobě jdoucích letech: 103,5; 104,7; 107,6; 105,8; 112,7; 116,5; 115,3; 108,5; 110,6. Máme vypočítat průměrné roční tempo růstu této výroby za uvedené období.

Řešení. Průměrné roční tempo růstu je vyjádřeno geometrickým průměrem podle vzorce (7):

$$\bar{x}_G = \sqrt[9]{103,5 \cdot 104,7 \cdot 107,6 \cdot 105,8 \cdot 112,7 \cdot 116,5 \cdot 115,3 \cdot 108,5 \cdot 110,6} \doteq 109,4$$

Doplňující charakteristiky polohy (úrovně) kvantitativních znaků jsou modus a medián. Definujeme:

Modus různých hodnot $x_1, x_2, \dots, x_n$ znaku, resp. třídního znaku x je hodnota, která má největší četnost. Značí se $\text{Mod}(x)$.

Modus je užíván k odhadu střední hodnoty znaku souboru nejčastěji tehdy, máme-li sestavenou tabulku rozdělení četností pro statistický soubor s velkým rozsahem nebo když modus má v souboru specifický význam.

Příklad, kdy je prakticky používán modus jako charakteristika polohy:

Při sledování počtu úrazů na jednotlivých pracovištích v určitém podniku v průběhu dne nás zajímá především modus, tj. nejčetnější hodnota počtu úrazů, abychom tak zjistili, na kterém pracovišti a v kterou denní dobu dochází k úrazům nejčastěji. Podle toho po zjištění lze pak provést příslušná opatření.

Medián všech hodnot $x_1, x_2, \dots, x_n$ znaku, resp. třídního znaku x statistického souboru, v němž jsou prvky uspořádány podle velikosti hodnot sledovaného znaku, je prostřední hodnota znaku: a) U souborů, jejichž rozsah n je liché číslo, se medián rovná hodnotě znaku prostředního, tj. prvku s indexem $\frac{n+1}{2}$. b) U souborů, jejichž rozsah n je sudé číslo, se medián rovná aritmetickému průměru hodnot dvou prostředních prvků, tj. prvků s indexy $\frac{n}{2}$ a $\frac{n}{2} + 1$. Značí se $\text{Med}(x)$.

Medián je užíván jako střední hodnota znaku souboru zejména tehdy, když jsou v něm zastoupeny prvky s hodnotami znaku mimořádně odlišnými (malými nebo velkými) oproti ostatním hodnotám znaku. V těchto případech je medián lepší charakteristikou polohy hodnot znaku než aritmetický průměr, který závisí na všech, tedy i na extrémních hodnotách znaku.

Příklad určení $\text{Mod}(x)$ a $\text{Med}(x)$ hodnot kvantitativního znaku x statistického souboru:

Pro rozdělení četností podle tabulky 7.11 určete aritmetický průměr, modus a medián hodnot x_j znaku x .

Tabulka rozdělení četností hodnot znaku řešené úlohy

Tab. 7.11

Hodnoty znaku x_j	Četnosti n_j
1	8
2	9
3	49
4	50
5	14
6	7
7	2
8	1
Součet n	110

Řešení. $\bar{x} \doteq 3,79$ (podle vzorce (2)), $\text{Mod}(x) = 4$ (neboť největší četnost 50 má hodnota znaku 4), $\text{Med}(x) = 4$ (neboť $\frac{x_{55} + x_{56}}{2} = 4$).

Střední hodnoty jako charakteristiky polohy (úrovně) hodnot kvantitativního znaku statistického souboru představují jistou jeho hodnotu, kolem níž je v jistém smyslu nejvíce soustředěno rozdělení četností hodnot znaku. Kromě charakteristik polohy se zavádějí charakteristiky variability hodnot kvantitativního znaku, jimiž se budeme zabývat v následujících odstavcích.

B. Charakteristiky variability (měnlivosti, rozptýlení) hodnot znaku jsou čísla, která nějak charakterizují, jak se hodnoty znaku prvků souboru liší od zvolené charakteristiky polohy (střední hodnoty), resp. od sebe navzájem. Patří mezi ně variační rozpětí, průměrná absolutní odchylka, rozptyl, směrodatná odchylka a variační koeficient.

Variační rozpětí R je rozdíl mezi největší a nejmenší hodnotou znaku prvků daného souboru:

$$R = x_{\max} - x_{\min} \quad (8)$$

Je pouze orientační charakteristikou variability hodnot znaku.

Rozdíl mezi hodnotou znaku x_i a zvolenou střední hodnotou, např. aritmetickým průměrem $\bar{x}$, se nazývá **odchylka hodnoty znaku** x_i od střední hodnoty.

Dokonalejší charakteristiky variability hodnot znaku:

Průměrná absolutní odchylka $\bar{d}$ je aritmetický průměr absolutních hodnot odchylek hodnot znaku všech prvků souboru od aritmetického průměru hodnot znaku:

$$\bar{d} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|, \quad (9)$$

resp. pro skupinové rozdělení četností (má-li hodnota x_i četnost $n_i, \dots$, hodnota x_k četnost n_k , $\sum_{i=1}^k n_i = n$) ve váženém tvaru:

$$\bar{d} = \frac{1}{n} \sum_{i=1}^k n_i |x_i - \bar{x}|. \quad (10)$$

Poznámka. Absolutní hodnoty odchylek se zde berou vzhledem k tomu, že průměrná odchylka hodnot x_i od aritmetického průměru $\bar{x}$ je vždy nulová, neboť pro součet odchylek x_i od aritmetického průměru $\bar{x}$ platí $\sum_{i=1}^n (x_i - \bar{x}) = 0$.

Nejdůležitějšími charakteristikami variability hodnot znaku v statistickém souboru jsou rozptyl a směrodatná odchylka.

Rozptyl s_x^2 (značí se též s^2) je aritmetický průměr druhých mocnin odchylek hodnot znaku od aritmetického průměru:

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2; \quad (11)$$

resp. pro skupinové rozdělení četností (při obdobném označení jako u vzorce (10)) ve váženém tvaru:

$$s_x^2 = \frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^2. \quad (12)$$

Provedeme-li ve vzorcích (11), (12) naznačené umocnění, lze je upravit na tvary, z nichž se rozptyl zpravidla prakticky počítá:

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2, \quad (13)$$

tj. rozptyl se rovná rozdílu aritmetického průměru druhých mocnin hodnot znaku a druhé mocniny jejich aritmetického průměru; resp. pro skupinové rozdělení četností ve váženém tvaru:

$$s_x^2 = \frac{1}{n} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2, \quad (14)$$

Použijeme-li metody vhodně zvoleného počátku zavedením nové proměnné $u_i = \frac{x_i - a}{b}$, kde a, b jsou reálné konstanty (a lib., $b > 0$), vypočteme nejprve rozptyl nové proměnné

$$s_u^2 = \frac{1}{n} \sum_{i=1}^n u_i^2 - \bar{u}^2, \quad (15)$$

resp. pro skupinové rozdělení četností ve váženém tvaru:

$$s_u^2 = \frac{1}{n} \sum_{i=1}^k n_i u_i^2 - \bar{u}^2 \quad (16)$$

a pak rozptyl původní proměnné podle vzorce

$$s_x^2 = b^2 \cdot s_u^2. \quad (17)$$

Směrodatná odchylka s_x (značí se též s) je druhá odmocnina z rozptylu (daného některým ze vzorců (11) až (15)). Výhodou směrodatné odchylky je, že charakterizuje variabilitu hodnot znaku v měřících jednotkách znaku, zatímco rozptyl je vyjádřen ve druhých mocnínách těchto jednotek.

Poznámka. Směrodatná odchylka je citlivější charakteristikou variability hodnot znaku v porovnání s průměrnou absolutní odchylkou, neboť zesiluje váhu odchylek pro extrémní hodnoty.

Příklad výpočtu charakteristik variability ilustrující jejich význam v porovnání s charakteristikou polohy (aritmetickým průměrem):

Uřídíme průměrnou absolutní odchylku a směrodatnou odchylku pro tři soubory o témž rozsahu $n = 5$ (pro jednoduchost a názornost výpočtu volíme n malé), které mají též aritmetický průměr $\bar{x}$ hodnot sledovaného znaku x :

a) 10, 10, 10, 10, 10
b) 8, 9, 10, 11, 12
c) 0, 5, 10, 15, 20

Řešení	soubor a)	soubor b)	soubor c)
Aritmetický průměr:	$\bar{x} = 10$	$\bar{x} = 10$	$\bar{x} = 10$
Průměrná absolutní odchylka:	$\bar{d} = 0$	$\bar{d} = \frac{6}{5} = 1,2$	$\bar{d} = 6$
Rozptyl:	$s_x^2 = 0$	$s_x^2 = 2$	$s_x^2 = 50$
Směrodatná odchylka:	$s_x = 0$	$s_x = \sqrt{2} \doteq 1,41$	$s_x = 5\sqrt{2} \doteq 7,07$

Z řešeného příkladu je patrné, že aritmetický průměr charakterizuje průměrnou hodnotu znaku, ale nevyjadřuje nic bližšího o hodnotách znaku, z nichž byl vypočten. Lze říci, že čím větší je variabilita hodnot znaku, tím méně reprezentativní je aritmetický průměr či jiná charakteristika polohy. Informaci o rozptýlení hodnot znaku kolem aritmetického průměru podává průměrná absolutní odchylka nebo lépe rozptyl, resp. směrodatná odchylka.

Relativní (poměrné) charakteristiky variability

Pro dva nebo více statistických souborů ke vzájemnému porovnání variability hodnot sledovaných znaků výrazně odlišné úrovně nebo vyjádřených v různých měřicích jednotkách se zavádějí tyto charakteristiky:

Poměrná průměrná absolutní odchylka $\bar{p}$ je podíl průměrné absolutní odchylky a příslušného aritmetického průměru sledovaného znaku x :

$$\bar{p} = \frac{\bar{d}}{\bar{x}} \quad (18)$$

Variační koeficient v (značí se často též v_x) je definován jako podíl směrodatné odchylky a aritmetického průměru sledovaného znaku x :

$$v = \frac{s_x}{\bar{x}}, \quad (19)$$

resp. v procentech

$$v = \frac{s_x}{\bar{x}} \cdot 100 \% \quad (20)$$

Příklad užití variačního koeficientu:

Porovnejte diferenciaci (variabilitu) mezd pracovníků dvou podniků na základě údajů o jejich příjmech v tabulce 7.12.

Tabulka údajů o příjmech pracovníků podniků pro řešenou úlohu

Tab. 7.12

1. podnik		2. podnik	
Měsíční příjem x_i (v Kč)	Počet pracovníků n_i	Hodinová mzda x_i (v Kč)	Počet pracovníků n_i
2 000	20	10	30
2 500	30	15	30
3 000	30	20	15
3 500	15	25	10
4 000	5	30	5

Řešení. Sledovaný statistický znak x (příjem pracovníka) je vyjádřen v obou podnicích v různých jednotkách (měsíční a hodinová mzda). K porovnání variability mezd použijeme proto variační koeficienty. Podle vzorců (2), (14), (19) postupně dostáváme

pro 1. podnik $\bar{x} = 2775$, $s_x = 558,46$, $v = 0,201$,

pro 2. podnik $\bar{x} = 16,11$, $s_x = 5,9056$, $v = 0,367$.

Výsledek: Diferenciace (variabilita) mezd v 1. podniku je nižší než ve 2. podniku.