1. POPISNÁ ŠTATISTIKA

1.1 Základné pojmy

Cieľom popisnej štatistiky je výsledky štatistického skúmania vyjadriť v prehľadnej, koncentrovanej forme, a to:

- pomocou tabuliek,
- pomocou grafického znázornenia,
- pomocou popisných charakteristík.

Základným pojmom v popisnej štatistike je *štatistický súbor.* Je to konečná (aj keď niekedy veľmi rozsiahla) neprázdna množina prvkov, ktoré sú nositeľmi určitého hromadného javu. Počet prvkov tejto množiny nazývame *rozsah štatistického súboru.* Štatistický súbor môžu tvoriť napríklad ľudia žijúci v určitom meste, žiaci jednej triedy, deti narodené v jednom roku, atď.

Prvky štatistického súboru nazývame *štatistické jednotky.* Štatistická jednotka je teda základný prvok, na ktorom možno skúmať konkrétny prejav určitého hromadného javu. Štatistická jednotka musí byť vymedzená vecne, časovo a priestorovo. Pri *vecnom vymedzení* je potrebné presne definovať, čo budeme rozumieť pod štatistickými jednotkami (napr. deti určitého veku). *Časové vymedzenie* znamená určenie časového obdobia alebo okamihu, v ktorom je štatistická jednotka skúmaná (napr. deti narodené v r. 1997). *Priestorové vymedzenie* štatistických jednotiek znamená určenie miesta alebo územia, na ktorom sa koná štatistické skúmanie (napr. deti narodené v r. 1997 v okrese Nitra).

Hromadný jav, ktorý pozorujeme na štatistických jednotkách, nazývame *štatistický znak*. Ak je štatistickým súborom množina žiakov v určitej triede, tak pozorovaným štatistickým znakom môže byť hmotnosť žiakov, výška a vek žiakov, národnosť žiakov, farba ich očí, zamestnanie ich rodičov a pod.

Štatistické znaky možno deliť z rôznych hľadísk. Základné delenie znakov je delenie na kvantitatívne a kvalitatívne znaky.

Kvantitatívne znaky sú také znaky, ktorých hodnoty sú reálne čísla (napr. výška, váha a vek žiakov, počet predaných automobilov na Slovensku a pod.). *Kvalitatívne znaky* sú znaky, ktoré túto vlastnosť nemajú (napríklad národnosť, farba očí, vzdelanie, pohlavie, atď.). Nadobúdajú úrovne, ktoré popisujeme slovne. Špeciálnym prípadom kvalitatívneho znaku je tzv. *alternatívny – dichotomický znak*. Je to taký znak, ktorý nadobúda len dve úrovne (napr. pohlavie alebo výskyt

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

choroby). Kvalitatívne znaky sa snažíme vyjadriť kvantitatívne. Pri alternatívnom znaku je to jednoduché. Ak je pozorovaným znakom napríklad výskyt choroby, tak v prípade, že sa choroba vyskytuje, zapíšeme 1, v opačnom prípade zapíšeme 0. V pedagogických, psychologických a sociologických výskumoch sa pracuje väčšinou s kvalitatívnymi znakmi, ktoré sa snažíme kvantifikovať pomocou stupníc alebo pomocou bodového hodnotenia. Prospechová stupnica známok v škole je kvantifikovaním vedomostnej úrovne žiakov, bodovanie výkonov v športe je kvantifikovaním výkonnostnej úrovne športovca a pod.

Ak štatistický znak nadobúda konečne veľa alebo spočítateľne veľa hodnôt, nazýva sa *diskrétny znak* (napr. počet vyrobených výrobkov, počet bodov, ktorí získali žiaci v teste). *Spojité znaky* sú také znaky, ktoré môžu nadobúdať ľubovoľnú hodnotu z nejakého intervalu (napr. hmotnosť alebo výška desaťročných detí, výška rastliny, rýchlosť vetra, ale aj vedomostná úroveň žiakov, kvalita vody).

V tomto texte sa budeme zaoberať prevažne kvantitatívnymi a kvantifikovateľnými kvalitatívnymi znakmi. Budeme ich označovať veľkými písmenami X, Y, ... a hodnoty, ktoré môžu nadobúdať, malými písmenami $x_1, x_2, ...$ resp. $y_1, y_2, ...$.

1.2 Triedenie štatistických údajov

Prvým krokom pri spracovávaní údajov je ich triedenie. Triedením sa nazýva usporiadanie údajov do skupín (tried). Ak sa pri triedení používa iba jeden triediaci znak (napr. výška žiakov), hovoríme o *jednostupňovom* triedení. Ak sa štatistický súbor triedi súčasne podľa dvoch alebo viacerých štatistických znakov (napr. výška a hmotnosť žiakov), hovoríme o *dvojstupňovom* resp. *viacstupňovom* triedení.

Pri triedení musíme dodržať zásadu

- *úplnosti*, t.j. musíme zatriediť každú nameranú hodnotu,
- *jednoznačnosti*, t.j. musíme triedy (ak sa jedná o diskrétny znak) resp. intervaly (ak sa jedná spojitý znak) zvoliť tak, aby sme každú štatistickú jednotku jednoznačne zatriedili práve do jednej triedy resp. intervalu.

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

1.2.1 Jednoduché triedenie

Ak na prvkoch štatistického súboru zisťujeme hodnoty určitého kvantitatívneho znaku, získame množstvo číselných údajov. Ak je rozsah štatistického súboru n, dostaneme postupnosť hodnôt $x_1, x_2, ..., x_n$. Je to tzv. prvotná tabuľka. Býva často neprehľadná, najmä vtedy, keď sme uskutočnili rozsiahlejší počet meraní. Prvotnú tabuľku môžeme upraviť tak, že hodnoty usporiadame podľa veľkosti do neklesajúcej postupnosti $x_1' \le x_2' \le ... \le x_n'$. Tejto postupnosti hovoríme usporiadaná tabuľka. Hodnota x_1' je minimálna hodnota znaku a x_n' je maximálna hodnota znaku. Rozdiel $R = x_{\max} - x_{\min}$ (medzi maximálnou a minimálnou hodnotou znaku) je tzv. rozpätie (resp. variačná šírka) znaku..

Z usporiadanej tabuľky môžeme zároveň vidieť, či sa jednotlivé hodnoty znaku vyskytujú viackrát. Ak je tomu tak, potom môžeme zostaviť tzv. *tabuľku rozdelenia početností*, ktorá má vo všeobecnosti nasledujúci tvar.

Tabuľka 1. 1. Tabuľka rozdelenia početností

x_i	f_i
x_1	f_1
x_2	f_2
.	.
.	.
.	.
x_k	f_k

V prvom stĺpci sú rôzne hodnoty znaku a v druhom stĺpci sú ich (absolútne) početnosti. Symbol f_i označuje početnosť výskytu hodnoty x_i ($i = 1,2,...,k$). Pritom platí: $f_1 + f_2 + ... + f_k = n$.

Hovoríme, že sme uskutočnili tzv. jednoduché triedenie. Rovnako postupujeme aj v prípade kvalitatívneho znaku, iba s tým rozdielom, že do prvého stĺpca píšeme úrovne pozorovaného znaku.

Príklad 1.1 Žiaci 6. A dostali z písomnej práce z matematiky nasledovné známky: 3, 2, 1, 1, 2, 2, 3, 4, 4, 1, 2, 3, 3, 3, 5, 4, 2, 2, 2, 3, 3, 3, 3, 1, 3. Zostrojte tabuľku rozdelenia početností.

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

Riešenie. Štatistický súbor tvoria žiaci 6. A, ktorí písali písomku z matematiky. Každý žiak je štatistickou jednotkou. Pozorovaným znakom X je známka z matematiky. Nadobúda päť hodnôt 1, 2, 3, 4, 5. Rozsah súboru je n = 25. Hodnoty z prvotnej tabuľky usporiadame do neklesajúcej postupnosti. Dostaneme postupnosť: 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 5. Tabuľka rozdelenia početností známok z matematiky má nasledujúci tvar:

Tabuľka 1.2.

x_i	f_i
1	4
2	7
3	10
4	3
5	1
Spolu	25

Tabuľku rozdelenia početností môžeme doplniť o ďalšie stĺpce, o stĺpec relatívnych početností, o stĺpec kumulatívnych početností a o stĺpec kumulatívnych relatívnych početností. Kumulatívna početnosť hodnoty x_i je súčet $f_1 + ... + f_i$ (pre i = 1, 2,..., k). Stĺpec relatívnych početností je daný podielom jednotlivých početností a rozsahu súboru. Relatívne početnosti je možné vyjadriť aj v percentách.

Tabuľka 1.3. Tabuľka rozdelenia početností

Úrovne znaku x_i	Absolútna početnosť f_i	Kumulatívna početnosť	Relatívna početnosť f_i/n	Kumulatívna relatívna početnosť	Relatívna početnosť v %
x_1	f_1	f_1	f_1/n	f_1/n	$\frac{f_1}{n}\cdot 100$
x_2	f_2	f_1+f_2	f_2/n	$(f_1+f_2)/n$	$\frac{f_2}{n}\cdot 100$
.	.	.	.	.	.
x_k	f_k	$f_1+\ldots+f_k$	f_k/n	$(f_1+\ldots+f_k)/n$	$\frac{f_k}{n}\cdot 100$
Spolu	n	-	1	-	100

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

Príklad. 1.2 Tabuľku rozdelenia početností známok z matematiky z príkladu 1.1 doplníme o ďalšie stĺpce.

Tabuľka 1.4. Rozdelenie početností známok z matematiky

Známky x_i	Absolútna početnosť f_i	Kumulatívna početnosť	Relatívna početnosť f_i/n	Kumulatívna relatívna početnosť	Relatívna početnosť v %
1	4	4	0,16	0,16	16
2	7	11	0,28	0,44	28
3	10	21	0,40	0,84	40
4	3	24	0,12	0,96	12
5	1	25	0,04	1	4
Spolu	25	-	1	-	100

Tabuľka rozdelenia početností umožňuje získať niektoré informácie rýchlo a v prehľadnej forme. Napríklad zo stĺpca „absolútna početnosť" môžeme zistiť, koľko žiakov z triedy napísalo písomku z matematiky na 3, je ich 10. Koľko žiakov napísalo písomku lepšie ako na štvorku zistíme zo stĺpca „kumulatívna početnosť", vidíme, že ich je 21. V prípade, že by sme chceli vyjadriť početnosť žiakov, ktorí napísali písomku z matematiky lepšie ako na štvorku v percentách, príslušnú hodnotu 0,84 v stĺpci „kumulatívna relatívna početnosť" vynásobíme 100. Dostávame, že 84 % žiakov dostalo z písomky z matematiky lepšiu známku ako 4.

Príklad 1.3 Pri záverečných skúškach žiaci písali previerku z matematiky, ktorá sa skladala z 12 príkladov. Skúšky sa zúčastnilo 28 žiakov. Počet správne vyriešených príkladov u jednotlivých žiakov bol takýto: 10, 6, 4, 3, 10, 11, 9, 8, 8, 6, 9, 8, 7, 8, 6, 7, 8, 4, 8, 5, 10, 7, 6, 7, 8, 7, 6, 3. Zostavte tabuľku rozdelenia početností počtu správne vyriešených príkladov.

Riešenie. Štatistický súbor tvoria žiaci, ktorí písali písomku z matematiky. Každý žiak je štatistickou jednotkou. Pozorovaným štatistickým znakom je počet správne vyriešených príkladov. Rozsah súboru je $n = 28$.

Hodnoty z prvotnej tabuľky usporiadame do neklesajúcej postupnosti: 3, 3, 4, 4, 5, 6, 6, 6, 6, 6, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 8, 9, 9,

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

10, 10, 10, 11. Tabuľka rozdelenia početností počtu správne vyriešených príkladov má nasledujúci tvar.

Tabuľka 1.5. Rozdelenie početností počtu správne vyriešených príkladov

Hodnoty znaku x_i	Absolútna početnosť f_i	Kumulatívna početnosť	Relatívna početnosť f_i/n	Kumulatívna relatívna početnosť	Relatívna početnosť v %
3	2	2	0,07	0,07	7
4	2	4	0,07	0,14	7
5	1	5	0,04	0,18	4
6	5	10	0,18	0,36	18
7	5	15	0,18	0,54	18
8	7	22	0,25	0,79	25
9	2	24	0,07	0,86	7
10	3	27	0,10	0,96	10
11	1	28	0,04	1	4
Spolu	28	-	1	-	100

1.2.2 Intervalové a skupinové triedenie

Jednoduché triedenie je vhodné použiť iba pri diskrétnom znaku, ktorý vykazuje iba menší počet hodnôt. Ak sme namerali veľké množstvo rôznych hodnôt diskrétneho znaku a v prípade spojitého znaku, vytvoríme tzv. *intervalové rozdelenie početností* (v prípade spojitého znaku) resp. *skupinové rozdelenie početností* (v prípade diskrétneho znaku). Niekoľko po sebe idúcich hodnôt spájame do jednej triedy (intervalu). Každému triednemu intervalu priradíme jeho početnosť (počet tých hodnôt, ktoré do danej triedy padli) a triedny znak, čo je stred triedneho intervalu. Triedny znak označujeme x_i. Postupujeme tak, že najskôr určíme šírku triednych intervalov. Ak chceme všetky hodnoty pozorovaného znaku zaradiť do k triednych intervalov, šírku intervalov (označíme ju h) určíme približne podľa vzorca

$$h \doteq \frac{x_{\max} - x_{\min}}{k},$$

kde $x_{\max}$ je maximálna hodnota a $x_{\min}$ je minimálna hodnota pozorovaného znaku. Dostaneme intervaly

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

$$\left(x_i - \frac{h}{2},\ x_i + \frac{h}{2}\right\rangle, \quad i = 1,2,\ldots,k.$$

V niektorých prípadoch môže mať prvý interval dolnú hranicu $-\infty$ a podobne posledný interval môže mať hornú hranicu ∞. Pre určenie počtu intervalov existujú viaceré pravidlá ([14]). Môžeme použiť vzťah $k \doteq \sqrt{n}$, kde n je rozsah súboru, alebo počet intervalov stanovíme sami. Počet intervalov nemá byť príliš malý, aby sa nestratila podstatná časť informácie o rozdelení početností hodnôt znaku, ale taktiež nemá byť príliš veľký, pretože sa zníži prehľadnosť triedenia.

Príklad 1.4 Pri meraní výšky 14 žiakov sme dostali takéto hodnoty (v cm): 158, 161, 168, 163, 174, 179, 176, 172, 178, 168, 179, 174, 173, 179. Zostavte tabuľku rozdelenia početností pre namerané hodnoty výšky žiakov.

Riešenie. Sledovaným štatistickým znakom je výška žiakov. Je to spojitý znak, a preto by nebolo vhodné robiť jednoduché triedenie. Vytvoríme intervalové rozdelenie početností.

Platí: $x_{\min}$ = 158, $x_{\max}$ = 179. Počet intervalov k určíme podľa vzťahu $k \doteq \sqrt{n}$. Keďže n = 14, $k \doteq \sqrt{14} \doteq 4$. Šírku intervalov vypočítame podľa vzťahu:

$$h \doteq \frac{x_{\max} - x_{\min}}{k} = \frac{179 - 158}{4} \doteq 5.$$

Na základe vypočítaných hodnôt by sme mali hodnoty pozorovaného znaku rozdeliť do 4 intervalov so šírkou 5, pričom by sme mali začať najmenšou nameranou hodnotou, ktorou je hodnota 158 cm. Tabuľka rozdelenia početností nameraných výšok žiakov bude však prehľadnejšia, ak zvolíme nasledujúce intervaly:

$$(155, 160\rangle, (160, 165\rangle, (165, 170\rangle, (170, 175\rangle, (175, 180\rangle.$$

Vidíme, že intervalov je 5 - viac ako sme určili výpočtom, a to z toho dôvodu, že sme posunuli dolnú hranicu. Zostrojíme tabuľku intervalového rozdelenia početností nameraných hodnôt výšky žiakov.

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

Tabuľka 1.6. Rozdelenie početností nameraných hodnôt výšky žiakov

Interval	Stred intervalu x_i	Absolútna početnosť f_i	Kumulatívna početnosť	Relatívna početnosť f_i/n	Relatívna početnosť v %	Kumulatívna relatívna početnosť
$\langle 155, 160)$	157,5	1	1	0,07	7	0,07
$\langle 160, 165)$	162,5	2	3	0,14	14	0,21
$\langle 165, 170)$	167,5	2	5	0,14	14	0,35
$\langle 170, 175)$	172,5	4	9	0,29	29	0,64
$\langle 175, 180)$	177,5	5	14	0,36	36	1
Spolu	-	14	-	1	100	-

Nasledujúci príklad je ukážkou skupinového triedenia hodnôt diskrétneho znaku.

Príklad 1.5 Študenti gymnázia písali test z anglického jazyka, v ktorom získali nasledujúce počty bodov: 19, 21, 22, 23, 29, 31, 30, 29, 32, 33, 24, 34, 27, 35, 34, 38, 35, 36, 37, 38, 34, 35, 39, 40, 43, 34, 35, 38, 42, 41, 42, 45, 48, 36, 37, 43, 40, 37, 40, 40. Zostrojte tabuľku rozdelenia početností počtu získaných bodov.

Riešenie. Sledovaným znakom je počet bodov za test. Je to diskrétny znak, ktorý vykazuje väčší počet rôznych hodnôt, a preto zostavíme tabuľku skupinového rozdelenia početností. Platí:

$$x_{\min} = 19,\ x_{\max} = 48,\ n = 40.$$

Počet skupín k určíme pomocou nasledujúceho vzťahu:

$$k \doteq \sqrt{n} = \sqrt{40} \doteq 6.$$

Keďže $h \doteq \dfrac{x_{\max} - x_{\min}}{k} = \dfrac{48-19}{6} \doteq 5$, hodnoty pozorovaného znaku rozdelíme do nasledujúcich skupín:

19 - 23, 24 - 28, 29 - 33, 34 - 38, 39 - 43, 44 - 48.

Zostrojíme tabuľku rozdelenia početností.

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

Tabuľka 1.7. Rozdelenie početností počtu bodov za test

Skupiny	Stred skupiny x_i	Absolútna početnosť f_i	Kumulatívna početnosť	Relatívna početnosť f_i /n	Relatívna početnosť v %	Kumulat. relatívna početnosť
19 - 23	21	4	4	0,10	10	0,10
24 - 28	26	2	6	0,05	5	0,15
29 - 33	31	6	12	0,15	15	0,30
34 - 38	36	16	28	0,40	40	0,50
39 - 43	41	10	38	0,25	25	0,75
44 - 48	46	2	40	0,05	5	1
Spolu	-	40	-	1	100	-

1. 3 Grafické znázornenie rozdelenia početností

Grafické znázorňovanie slúži pre rýchlu a názornú prezentáciu štatistických výsledkov, najmä vtedy, keď sa jedná o obsiahlejšie údaje a vzájomné porovnávanie viacerých súborov. Grafickým znázornením rozdelenia početností získame názornú predstavu o charaktere rozdelenia početností pozorovaného znaku. Existuje viacero možností grafického znázornenia rozdelenia početností. Spôsob znázornenia závisí od toho, či treba znázorniť rozdelenie početností kvantitatívneho alebo kvalitatívneho znak. Pri grafickom znázornení rozdelenia početností kvantitatívneho znaku spravidla používame pravouhlú súradnicovú sústavu, pričom na vodorovnú os nanášame hodnoty znaku resp. triedne intervaly, prípadne ich stredy a na zvislú os príslušné početnosti. Najčastejšie používané grafy rozdelenia početností sú *histogram* a *polygón.*

Pri intervalovom triedení najčastejšie používame *histogram*. Je to stĺpcový diagram so stĺpcami, ktorých základňa sa rovná šírke intervalu a výška i - teho stĺpca sa rovná početnosti f_i, $i = 1,2,...,k$.

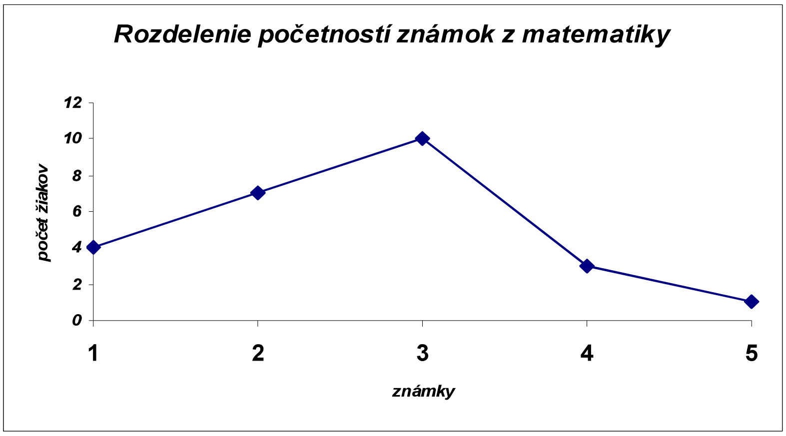
Polygón je spojnicový graf. Vznikne spojením bodov (x_i, f_i), $i = 1,2,...,k$. Polygón môžeme dostať aj z histogramu, a to tak, že spojíme lomenou čiarou stredy horných strán všetkých stĺpcov histogramu.

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

Ak body polygónu spojíme hladkou krivkou, dostaneme graf frekvenčnej krivky.

Ak namiesto hodnôt absolútnych početností na os y nanesieme hodnoty kumulatívnych početností, dostaneme polygón (resp. histogram) kumulatívnych početností.

Na nasledujúcich obrázkoch sú zostrojené grafy pre údaje z príkladov 1.1 a 1.4.



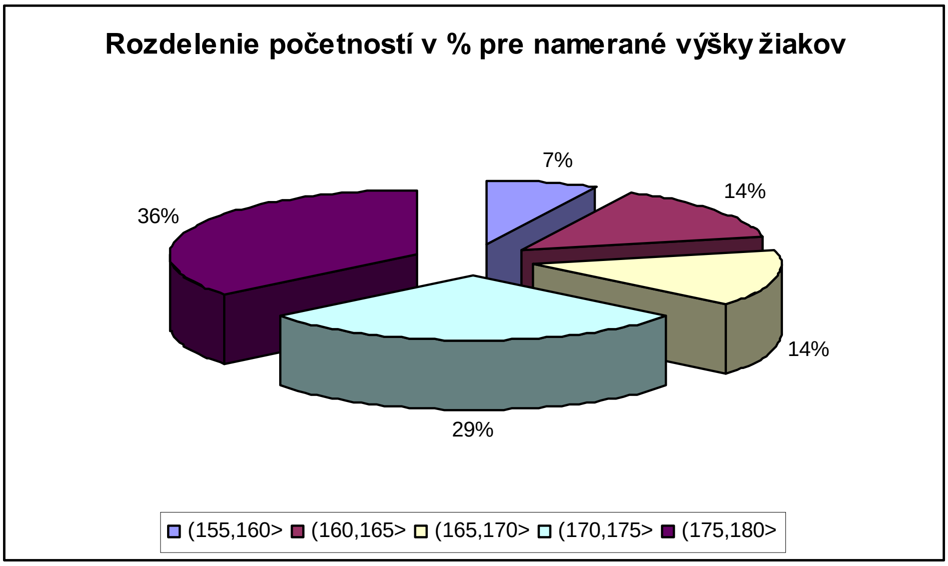
Obr. 1.1. Polygón rozdelenia početností známok z matematiky

Obr. 1.2. Histogram rozdelenia početností známok z matematiky

This document was created using SOLID CONVERTER PDF
To remove this message, purchase the product at www.SolidDocuments.com

Okrem polygónu a histogramu sa často používa aj *výsekový graf*, hlavne pri znázorňovaní rozdelenia početností kvalitatívneho znaku. S výsekovým grafom sa môžeme často stretnúť aj v dennej tlači. Jeho základom je kruh. Rozdelenie početností v percentách je vyjadrené v stupňoch, kde 1% = 3,60° (100 % = 360°).

Na obrázku 1.3 je zobrazený výsekový graf, ktorý vyjadruje štruktúru rozdelenia početností z príkladu 1.4.



Obr. 1.3. Výsekový graf pre namerané hodnoty výšky žiakov

1.4 Štatistické charakteristiky

Štatistický súbor môžeme charakterizovať aj pomocou číselných hodnôt, ktoré reprezentujú celý súbor. Poskytujú predstavu o celom sledovanom štatistickom súbore vo forme jednej alebo viacerých číselných hodnôt. Tieto číselné hodnoty, tzv. *štatistické charakteristiky*, môžeme rozdeliť do dvoch základných skupín:

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

***A)* Charakteristiky polohy** – charakterizujú úroveň hodnôt znaku v štatistickom súbore. Medzi najpoužívanejšie charakteristiky polohy patria *priemery* (aritmetický, harmonický, geometrický), *medián* a *modus.*

***B)* Charakteristiky variability** – charakterizujú mieru rozptýlenia hodnôt znaku. Na meranie variability sa používa *variačné rozpätie*, *priemerná odchýlka*, *smerodajná odchýlka*, *rozptyl* a ďalšie.

Štatistické charakteristiky sme prehľadne zapísali do nasledujúcej tabuľky. Keďže názvy štatistických charakteristík sa vo výstupnej zostave počítača veľmi často uvádzajú v anglickom jazyku, do tabuľky sme zaradili aj ich anglické názvy. Zároveň uvádzame aj ich ekvivalentné názvy.

***A)* Charakteristiky polohy** (means, averages, measures of central tendency)	***B)* Charakteristiky variability** (measures of variability, dispersions)
a) Aritmetický priemer, ozn. $\bar{x}$ (arithmetic mean, average)	*a) Variačné rozpätie*, ozn. R (range, variačná šírka)
b) Geometrický priemer, ozn. $\bar{x}_g$ (geometric mean)	*b) Rozptyl*, ozn. σ^2 (variance, disperzia)
c) Harmonický priemer, ozn. $\bar{x}_h$ (harmonic mean)	*c) Smerodajná odchýlka*, ozn. σ (standart deviation)
d) Medián, ozn. $\tilde{x}$ (median, prostredná hodnota)	*d) Priemerná odchýlka*, ozn. e (mean deviation, average deviation)
e) Modus, ozn. $\hat{x}$ (mode, najčastejšia hodnota)	*e) Medzikvartilová odchýlka*, ozn. Q (interquartill range)

Poznámka. Vo výberovej teórii sa počíta tzv. *výberový rozptyl*. Výberový rozptyl budeme označovať s^2. Je definovaný vzťahom

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 .$$

This document was created using SOLID CONVERTER PDF. To remove this message, purchase the product at www.SolidDocuments.com

1.4.1 Charakteristiky polohy

Charakteristiky polohy sú čísla, ktoré charakterizujú úroveň hodnôt znaku v štatistickom súbore. Postupne budeme definovať *priemery* (aritmetický, harmonický, geometrický), *medián* a *modus* a výpočty ich hodnôt budeme ilustrovať na príkladoch.

a) Aritmetický priemer (ozn. $\bar{x}$)

Aritmetický priemer je najčastejšie používaná charakteristika polohy. Priemery (aritmetický, harmonický a geometrický) sa počítajú zo všetkých hodnôt kvantitatívneho znaku. *Aritmetický priemer* definujeme ako súčet všetkých hodnôt znaku delený ich počtom, t.j. rozsahom súboru. Ak štatistický súbor má rozsah n a štatistický znak X nadobúda hodnoty $x_1, x_2, ..., x_n$, potom aritmetický priemer je daný vzťahom

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

Ak je daná tabuľka rozdelenia početností, potom aritmetický priemer počítame podľa vzorca

(1.2) $$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + ... + x_k f_k}{n} = \frac{1}{n}\sum_{i=1}^{k} x_i f_i , \quad \text{kde}$$

$$f_1 + f_2 + ... + f_k = n .$$

Absolútne početnosti f_i jednotlivých hodnôt x_i vo vzorci (1.2) vystupujú ako váhy príslušných hodnôt x_i, a preto uvedený aritmetický priemer nazývame aj vážený aritmetický priemer. Pri výpočte váženého aritmetického priemeru do tabuľky rozdelenia početností pridáme ďalší stĺpec s hodnotami $x_i f_i$ ($i = 1,2,..., k$). Výpočet aritmetického priemeru budeme ilustrovať na údajoch z príkladov 1.1 a 1.3.

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

Tabuľka 1.8.

Hodnoty znaku x_i	Absolútna početnosť f_i	Kumulatívna početnosť	Relatívna početnosť f_i/n	Kumulatívna relatívna početnosť	Relatívna početnosť v %	$x_i f_i$
1	4	4	0,16	0,16	16	4
2	7	11	0,28	0,44	28	14
3	10	21	0,40	0,84	40	30
4	3	24	0,12	0,96	12	12
5	1	25	0,04	1	4	5
Spolu	25	-	1	-	100	65

$$\bar{x} = \frac{1}{25}\sum_{i=1}^{5} x_i f_i = \frac{1}{25}\cdot 65 = 2{,}6 \doteq 3$$

Priemerná známka z písomnej práce z matematiky bola 3.

Tabuľka 1.9.

Hodnoty znaku x_i	Absolútna početnosť f_i	Kumulatívna početnosť	Relatívna početnosť f_i/n	Kumulatívna relatívna početnosť	Relatívna početnosť v %	$x_i f_i$
3	2	2	0,07	0,07	7	6
4	2	4	0,07	0,14	7	8
5	1	5	0,04	0,18	4	5
6	5	10	0,18	0,36	18	30
7	5	15	0,18	0,54	18	35
8	7	22	0,25	0,79	25	56
9	2	24	0,07	0,86	7	18
10	3	27	0,10	0,96	10	30
11	1	28	0,04	1	4	11
Spolu	28	-	1	-	100	199

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i f_i = \frac{1}{28}\cdot 199 \doteq 7,$$

t.j. priemerný počet správne vyriešených príkladov je 7.

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

Podobne budeme postupovať aj v prípade intervalového rozdelenia početností. Výpočet aritmetického priemeru hodnôt spojitého kvantitatívneho znaku ukážeme na príklade 1.4. Do tabuľky 1.6 doplníme stĺpec s hodnotami $x_i f_i$. Dostaneme tabuľku 1.10.

Tabuľka 1.10.

Interval	Stred intervalu x_i	Absolútna početnosť f_i	Kumulatívna početnosť	Relatívna početnosť f_i/n	Relatívna početnosť v %	$x_i f_i$
$\langle 155, 160 \rangle$	157,5	1	1	0,07	7	157,5
$\langle 160, 165 \rangle$	162,5	2	3	0,14	14	325
$\langle 165, 170 \rangle$	167,5	2	5	0,14	14	335
$\langle 170, 175 \rangle$	172,5	4	9	0,29	29	690
$\langle 175, 180 \rangle$	177,5	5	14	0,36	36	887,5
Spolu	-	14	-	1	100	2 395

Na základe výsledkov, ktoré sú uvedené v tabuľke, vypočítame aritmetický priemer:

$$\bar{x} = \frac{1}{14}\sum_{i=1}^{5} x_i f_i = \frac{1}{14} \cdot 2\,395 = 171{,}07.$$

Priemerná výška žiakov vybranej skupiny je približne 171,1 cm.

Analogicky vypočítame aritmetický priemer hodnôt z príkladu 1.5. Do tabuľky 1.7 doplníme stĺpec s hodnotami $x_i f_i$. Dostaneme tabuľku 1.11.

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

Tabuľka 1.11.

Trieda	x_i	Absolútna početnosť f_i	Kumulatívna početnosť	Relatívna početnosť f_i/n	Relatívna početnosť v %	Kumulat. relatívna početnosť	$x_i f_i$
19-23	21	4	4	0,10	10	0,10	84
24-28	26	2	6	0,05	5	0,15	52
29-33	31	6	12	0,15	15	0,30	186
34-38	36	16	28	0,40	40	0,50	576
39-43	41	10	38	0,25	25	0,75	410
44-48	46	2	40	0,05	5	1	92
Spolu	-	40	-	1	100	-	1400

$$\bar{x} = \frac{1}{40}\sum_{i=1}^{6} x_i f_i = \frac{1}{40} \cdot 1400 = 35$$

Žiaci dosiahli v teste z anglického jazyka priemerne 35 bodov.

Vážený aritmetický priemer z viacerých súborov (ozn. $\bar{x}_v$)

Často je potrebné zhrnúť informácie z viacerých štatistických súborov do jedného ukazovateľa. Napríklad v školskej praxi je potrebné vypočítať priemernú známku z dejepisu v troch paralelných triedach 6.A, 6.B a v 6.C šiesteho ročníka, kde máme nasledovné priemerné známky: $\bar{x}_{6A} = 2{,}4$; $\bar{x}_{6.B} = 2{,}0$; $\bar{x}_{6.C} = 1{,}2$. Ak je vo všetkých troch triedach rovnaký počet žiakov, môžeme priemernú známku z dejepisu v šiestom ročníku vypočítať podľa vzťahu: $\bar{x} = \frac{1}{3}\left(\bar{x}_{6.A} + \bar{x}_{6.B} + \bar{x}_{6.C}\right)$.

Ak v triedach nie je rovnaký počet žiakov, potom výpočet priemernej známky z dejepisu podľa uvedeného vzťahu je nesprávny. Aritmetické priemery súborov s rozdielnymi rozsahmi je potrebné „vážiť“ rozsahom príslušného súboru. Počítame *vážený aritmetický priemer* $\bar{x}_v$ z priemerov viacerých súborov podľa vzorca

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

$$\bar{x}_v = \frac{\bar{x}_1 \cdot n_1 + \bar{x}_2 \cdot n_2 + \ldots + \bar{x}_k \cdot n_k}{n_1 + n_2 + \ldots + n_k} = \frac{\sum_{i=1}^{k} \bar{x}_i n_i}{n}, \quad \text{kde}$$

$\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ sú aritmetické priemery súborov,

$n_1, n_2, \ldots, n_k$ je rozsah prvého, druhého,..., k - teho súboru,

$n = n_1 + n_2 + \ldots + n_k$,

k je počet súborov.

Vplyv rozsahov súborov na hodnotu váženého aritmetického priemeru $\bar{x}_v$ ukážeme na nasledujúcom príklade.

Príklad 1.6 V dvoch skupinách sa zisťovalo, ako často vynechajú žiaci vyučovanie počas jedného školského roku pre chorobu. V skupine A sa zistilo priemerne 20 hodín absencie, v skupine B 40 dní. Chceme zistiť priemerný počet dní absencie v oboch skupinách spolu.

Riešenie.

1. prípad: Skupina A a skupina B sú rovnako veľké.

$$n_1 = n_2 = 10; \quad \bar{x}_v = \frac{20 \cdot 10 + 40 \cdot 10}{20} = 30 \text{ dní absencie}$$

2. prípad: Skupina A je relatívne veľká, skupina B malá.

$$n_1 = 100, \; n_2 = 10; \quad \bar{x}_v = \frac{20 \cdot 100 + 40 \cdot 10}{110} = 21{,}82 \text{ dní absencie}$$

3. prípad: Skupina A je relatívne malá, skupina B veľká.

$$n_1 = 10, \; n_2 = 100; \quad \bar{x}_v = \frac{20 \cdot 10 + 40 \cdot 100}{110} = 38{,}18 \text{ dní absencie.}$$

b) Geometrický priemer (ozn. $\bar{x}_g$)

Geometrický priemer je definovaný ako *n*- tá odmocnina súčinu hodnôt znaku. Je definovaný pomocou nasledujúceho vzťahu:

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}, \qquad (x_i > 0, \; i = 1, 2, \ldots, n)$$

Ak sú údaje zapísané v tabuľke rozdelenia početností, geometrický priemer počítame podľa vzorca:

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

$$\overline{x}_g = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot \ldots \cdot x_k^{f_k}}\ , \quad \text{pričom} \sum_{i=1}^{k} f_i = n.$$

Používa sa ako charakteristika polohy znaku, ktorého hodnoty narastajú geometricky. Vyjadruje priemernú veľkosť zmeny.

Príklad 1.7 Zistite priemerný ročný koeficient rastu výroby v drevárenskom podniku, ak jeho výroba je zaznamenaná v nasledujúcej tabuľke:

Tabuľka 1.12.

Rok	1973	1974	1975	1976
Výroba v m^3	62 000	65 100	66 402	99 603

Riešenie. Ročné koeficienty rastu výroby sú

$$k_1 = \frac{65100}{62000} = 1{,}05; \ k_2 = \frac{66402}{65100} = 1{,}02; \ k_3 = \frac{99603}{66402} = 1{,}5.$$

Platí

$$99\,603 = 66\,402 \cdot k_3 = 65\,100 \cdot k_2 \cdot k_3 = 62\,000 \cdot k_1 \cdot k_2 \cdot k_3 .$$

Predpokladajme, že rast výroby je rovnomerný s ročným koeficientom rastu $\overline{k}$. Potom platí

$$99\,603 = 66\,402 \cdot \overline{k} = 65\,100 \cdot \overline{k} \cdot \overline{k} = 62\,000 \cdot \overline{k} \cdot \overline{k} \cdot \overline{k} .$$

Dostávame

$$62\,000 \cdot k_1 \cdot k_2 \cdot k_3 = 62\,000 \cdot \overline{k} \cdot \overline{k} \cdot \overline{k}, \quad \text{t.j.} \ (\overline{k})^3 = k_1 \cdot k_2 \cdot k_3 ,$$

odkiaľ vyplýva, že

$$\overline{k} = \sqrt[3]{k_1 \cdot k_2 \cdot k_3} = \sqrt[3]{1{,}05 \cdot 1{,}02 \cdot 1{,}5} \doteq 1{,}17.$$

Priemerný ročný koeficient rastu výroby v drevárenskom podniku je 1,17. Všimnime si, že sme počítali geometrický priemer. Aritmetický priemer nie je priemerným ročným koeficientom rastu výroby.

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

c) Harmonický priemer (ozn. $\bar{x}_h$)

Motiváciou pre zavedenie harmonického priemeru môže byť nasledujúci príklad.

Príklad 1.8 Je potrebné určiť priemerný čas, ktorý potrebuje jeden robotník na výrobu jedného výrobku z údajov v nasledujúcej tabuľke.

Tabuľka 1.13.

Robotník č. i	1	2	3	4	5	6	7	8	9	10
x_i	4	5	5	6	6	8	8	8	16	16

x_i označuje čas (v minútach) potrebný na výrobu jedného výrobku.

Riešenie. Pozorovaným znakom X je čas potrebný na výrobu jedného výrobku. Údaje zapíšeme do tabuľky rozdelenia početností.

Tabuľka 1.14.

x_i	4	5	6	8	16
f_i	1	2	2	3	2

Rozsah výberového súboru je $n = \sum_{i=1}^{5} f_i = 10$.

Uvažujme o počte výrobkov, ktorý vyrobia títo desiati robotníci za jednu hodinu (t.j. za 60 minút). Tento počet je:

$$\frac{60}{4} + \frac{60}{5}\cdot 2 + \frac{60}{6}\cdot 2 + \frac{60}{8}\cdot 3 + \frac{60}{16}\cdot 2 \quad \text{výrobkov.}$$

Označme $\bar{x}_h$ priemerný čas jedného robotníka potrebný na výrobu jedného výrobku. Platí

$$\frac{60}{4} + \frac{60}{5}\cdot 2 + \frac{60}{6}\cdot 2 + \frac{60}{8}\cdot 3 + \frac{60}{16}\cdot 2 = \frac{60}{\bar{x}_h}\cdot 10.$$

Po úprave dostávame rovnosť

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

$$\bar{x}_h \cdot \left(\frac{1}{4}+\frac{2}{5}+\frac{2}{6}+\frac{3}{8}+\frac{2}{16}\right) = 10,$$

odkiaľ vyplýva, že

$$\bar{x}_h = \frac{10}{\frac{1}{4}+\frac{2}{5}+\frac{2}{6}+\frac{3}{8}+\frac{2}{16}} \doteq 6{,}743.$$

Priemerný čas potrebný k zhotoveniu jedného výrobku je približne 6,7 minúty.

V uvedenom príklade aritmetický priemer nepredstavoval priemerný čas. Počítali sme tzv. harmonický priemer. Harmonický priemer budeme označovať $\bar{x}_h$. Je definovaný ako podiel rozsahu súboru a súčtu prevrátených hodnôt znaku:

$$\bar{x}_h = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}} \qquad (x_i > 0,\ i = 1, 2, \ldots, n)$$

Ak vychádzame z tabuľky rozdelenia početností, harmonický priemer počítame podľa vzorca:

$$\bar{x}_h = \frac{n}{\frac{f_1}{x_1}+\frac{f_2}{x_2}+\ldots+\frac{f_k}{x_k}} = \frac{n}{\sum_{i=1}^{k}\frac{f_i}{x_i}}, \quad \text{pričom } \sum_{i=1}^{k} f_i = n.$$

Harmonický priemer charakterizuje charakteristiku polohy súboru hodnôt, ak tieto hodnoty predstavujú výkonové limity.

Príklad 1.9 Auto prešlo 50 km. Prvých 10 km išlo rýchlosťou 50 km/h, ďalších 10 km rýchlosťou 80 km/h, potom 10 km rýchlosťou 90 km/h, ďalších 10 km rýchlosťou 60 km/h a posledných 10 km rýchlosťou 80 km/h. Akou priemernou rýchlosťou išlo auto?

Riešenie. Priemernú rýchlosť počítame podľa vzťahu

$$v = \frac{s}{t},$$

kde v označuje priemernú rýchlosť, s dráhu a t označuje čas. Platí

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

$$t = t_1 + t_2 + t_3 + t_4 + t_5 =$$

$$= \frac{s_1}{v_1} + \frac{s_2}{v_2} + \frac{s_3}{v_3} + \frac{s_4}{v_4} + \frac{s_5}{v_5} = \frac{10}{50} + \frac{10}{80} + \frac{10}{90} + \frac{10}{60} + \frac{10}{80},$$

$$v = \frac{s}{t} = \frac{50}{\frac{10}{50}+\frac{10}{80}+\frac{10}{90}+\frac{10}{60}+\frac{10}{80}} = \frac{50}{10\cdot\left(\frac{1}{50}+\frac{1}{80}+\frac{1}{90}+\frac{1}{60}+\frac{1}{80}\right)} = 68{,}7.$$

Všimnime si, že sme počítali harmonický priemer. Aritmetický priemer rýchlostí $\bar{x} = \frac{50+80+90+60+80}{5} = 72$ km/h nie je priemernou rýchlosťou.

d) Medián (ozn. $\tilde{x}$)

V štatistickej praxi sa najčastejšie ako charakteristika polohy používa aritmetický priemer, lebo závisí od všetkých pozorovaných hodnôt. Je však veľmi citlivý na to, ak sú niektoré hodnoty znaku extrémne veľké alebo malé. Extrémne hodnoty môžu spôsobiť, že priemer nie je najlepšou charakteristikou polohy. V takýchto prípadoch treba použiť inú charakteristiku polohy.

Na nasledujúcom príklade ukážeme vplyv extrémnych hodnôt na hodnotu aritmetického priemeru.

Príklad 1.10 Učiteľ zaznamenal u 11 žiakov nasledujúci počet hodín vymeškaných z vyučovania: 5, 12 , 6, 8, 10, 7, 5, 110, 2, 5, 6. Priemerný počet vymeškaných vyučovacích hodín je:

$$\bar{x} = \frac{1}{11}(5 + 12 + 6 + 8 + 10 + 7 + 5 + 110 + 2 + 5 + 6) = \frac{176}{11} = 16 \text{ hodín.}$$

Vidíme, že v dôsledku jednej extrémnej hodnoty je priemerný počet vymeškaných hodín značne skreslený a nevystihuje počet absencií v sledovanej skupine 10 žiakov. V takýchto prípadoch použijeme niektorú z ďalších charakteristík polohy, napríklad medián. *Medián* je prostredná hodnota, ktorú označujeme $\tilde{x}$. Je to hodnota, ktorá sa nachádza v strede súboru hodnôt usporiadaného do neklesajúcej postupnosti.

Určíme medián v súbore hodnôt z predchádzajúceho príkladu. Najskôr zoradíme hodnoty počtu vymeškaných hodín do neklesajúcej

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

postupnosti: 2, 5, 5, 5, 6, $\underline{\underline{6}}$, 6, 7, 8, 10, 110. Medián je hodnota, ktorá sa nachádza v strede postupnosti, t.j. $\tilde{x} = 6$ vymeškaných vyučovacích hodín.

Príklad 1.11 Pri štatistickom zisťovaní boli namerané hodnoty: 1, 2, 2, 3, 4, 5, 6, 7, 8, 8, 9. Rozsah súboru $n = 11$; ľahko vypočítame aritmetický priemer $\bar{x} = 5{,}18$. Medián $\tilde{x} = 5$, pretože je to prostredná nameraná hodnota.

V prípade, že rozsah súboru je párne číslo, potom medián určíme ako aritmetický priemer dvoch prostredných hodnôt. Uvedieme príklad.

Príklad 1.12 Pri štatistickom zisťovaní bolo nameraných týchto 12 hodnôt: 1, 2, 2, 3, 4, 5, 6, 7, 8, 8, 9, 10. Keďže rozsah súboru je párne číslo, stredné hodnoty sú 5 a 6. Za medián budeme považovať ich aritmetický priemer: $\tilde{x} = \dfrac{5+6}{2} = 5{,}5$.

Príklad 1.13 V nasledujúcej tabuľke sú uvedené údaje o mesačných príjmoch (v eurách) 20 členov družstva. Vypočítajte ich priemerný mesačný príjem.

Tabuľka 1.15.

mesačný príjem	300	400	500	600	700	8400
počet členov družstva	1	6	6	5	1	1

Riešenie. Vypočítajme aritmetický priemer: $\bar{x} = \dfrac{1}{20} \cdot 17800 =$ $= 890$ (eur). Ako je však vidno z tabuľky, majú všetci členovia, s výnimkou jedného, podstatne nižší príjem. To znamená, že aritmetický priemer v tomto prípade skresľuje skutočnosť. Vhodnejšou charakteristikou polohy príjmov členov družstva bude medián:

$$\tilde{x} = \frac{x_{10} + x_{11}}{2} = \frac{500 + 500}{2} = 500 \text{ (eur).}$$

Ak máme údaje roztriedené v tabuľke rozdelenia početností, medián určíme nasledovne: Na základe kumulatívnych početností určíme triedu

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

(tzv. mediánový interval), do ktorej patrí medián. Medián vypočítame podľa vzťahu:

$$\tilde{x} = x_m + h \cdot \frac{b+c-a}{2b}, \quad \text{kde}$$

x_m je počiatok mediánového intervalu;

a je kumulatívna početnosť intervalu, ktorý je v tabuľke rozdelenia početností pred mediánovým intervalom;

b je početnosť mediánového intervalu;

c je súčet početností za mediánovým intervalom;

h je šírka intervalu.

Príklad 1.14 Výpočet mediánu pre intervalové rozdelenie početností ukážeme na údajoch z príkladu 1.4. Mediánový interval je interval $\left(170, 175\right\rangle$. Vypočítajme medián:

$$\tilde{x} = x_m + h \cdot \frac{b+c-a}{2b} = 170 + 5 \cdot \frac{4+5-5}{8} = 172{,}5 \text{ cm.}$$

Príklad 1.15 Vypočítajte medián súboru hodnôt z príkladu 1.5. Mediánový interval je trieda 34-38. Vypočítame medián:

$$\tilde{x} = x_m + h \cdot \frac{b+c-a}{2b} = 34 + 5 \cdot \frac{16+12-12}{32} = 36{,}5 \doteq 37 \text{ bodov za test.}$$

e) Modus (ozn. $\hat{x}$)

Modus je hodnota znaku X, ktorá má v štatistickom súbore najväčšiu početnosť. Rovnako ako medián aj modus je charakteristika polohy, pre výpočet ktorej nepotrebujeme všetky namerané hodnoty – na rozdiel od aritmetického priemeru. Modus môžeme určiť iba vtedy, keď sa početnosti f_i hodnôt x_i znaku X odlišujú. Ak sú početnosti rovnaké, nemôžeme určiť modus.

Príklad 1.16 Pri pozorovaní diskrétneho štatistického znaku boli zistené tieto hodnoty: 3, 3, 3, 4, 4, 4, 6, 6, 6, 8, 8, 8. V tomto prípade nie je možné určiť modus, pretože všetky pozorované hodnoty majú rovnakú početnosť.

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

Príklad 1.17 V súbore hodnôt 3, 4, 4, 5, 5, 5, 6, 7, 7, 7, 7, 8 má najväčšiu početnosť hodnota 7, preto $\hat{x} = 7$.

Ak v súbore vystupujú dve alebo viac navzájom susediacich hodnôt rovnako často a ich početnosť je väčšia ako početnosť ostatných hodnôt, tak modus vypočítame ako aritmetický priemer najfrekventovanejších hodnôt.

Príklad 1.18 Pri pozorovaní diskrétneho štatistického znaku boli zistené tieto hodnoty: 3, 4, 4, 5, 5, 5, 6, 6, 6, 7, 8. Hodnoty 5 a 6 sa v tomto súbore vyskytujú s rovnakou frekvenciou (po tri razy), teda modus $\hat{x} = \frac{5+6}{2} = 5{,}5$.

Ak existujú dve navzájom nesusediace hodnoty s relatívne najväčšími početnosťami, tak obe tieto hodnoty uvádzame ako modus. V takomto prípade hovoríme, že rozdelenie je *bimodálne* (dvojvrcholové). Uvedieme príklad.

Príklad 1.19 V súbore hodnôt 3, 4, 4, 5, 5, 5, 5, 6, 6, 7, 8, 8, 8, 8, 9 sa najčastejšie vyskytujú dve hodnoty 5 a 8 (rovnako často). Obe hodnoty sú modus. Rozdelenie je bimodálne.

V prípade intervalového rozdelenia početností budeme za modus považovať stred intervalu s najväčšou početnosťou. Napríklad z tabuľky rozdelenia početností (tabuľka 1.7) v príklade 1.5 vidíme, že najčastejšie sa vyskytuje hodnota 36, t.j. študenti najčastejšie dostali za test z angličtiny 36 bodov. Pri grafickom znázornení je modus hodnota na x-ovej súradnici, v ktorej polygón početností dosahuje maximum.

Poznámka. Vo väčšine prípadov je aritmetický priemer najlepšou charakteristikou polohy, pretože sa počíta zo všetkých nameraných hodnôt. Aritmetický priemer by sa však nemal počítať v niektorých prípadoch, napríklad vtedy, ak je rozdelenie viacvrcholové (napr. bimodálne). Vtedy treba uprednostniť modus. Rovnako pre kvalitatívne údaje sa ako charakteristika polohy používa modus. Pri asymetrickom rozdelení treba uprednostniť medián.

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

1.4.2 Charakteristiky variability

Štatistické súbory sa môžu líšiť nielen v úrovni hodnôt znaku charakterizovanej niektorou charakteristikou polohy, ale aj variabilitou hodnôt pozorovaného znaku. Charakteristiky variability charakterizujú mieru rozptýlenia hodnôt znaku okolo ich charakteristiky polohy.

a) Variačné rozpätie (ozn. R)

Variačné rozpätie je rozdiel medzi maximálnou a minimálnou hodnotou. Počítame ho podľa vzťahu: $R = x_{\max} - x_{\min}$.

Variačné rozpätie skúmaného znaku poskytuje základný pohľad na menlivosť hodnôt znaku v štatistickom súbore. Jeho veľkosť závisí iba od krajných hodnôt, pričom jedna z nich alebo obe môžu byť extrémne hodnoty – pre daný súbor netypické, a preto môžu predstavu o variabilite značne skresľovať.

b) Priemerná odchýlka (ozn. e)

Priemerná odchýlka je lepšou charakteristikou variability ako variačné rozpätie, nakoľko jej veľkosť závisí od každej nameranej hodnoty štatistického súboru. Počítame ju ako aritmetický priemer absolútnych hodnôt všetkých odchýlok od ich aritmetického priemeru:

$$e = \frac{1}{n}\sum_{i=1}^{n}\left|x_i - \bar{x}\right|$$

alebo ak je daná tabuľka rozdelenia početností, podľa vzťahu:

$$e = \frac{1}{n}\sum_{i=1}^{k}\left|x_i - \bar{x}\right| \cdot f_i, \quad \text{pričom} \quad \sum_{i=1}^{k} f_i = n.$$

Príklad 1.20 Dvaja strelci strieľajú do terča a zásahy si zapisujú. Prvý strelec zasiahol 7, 8 a 9, druhý strelec 1, 10 a 13. Vypočítajte priemerné odchýlky počtu nastrieľaných bodov oboch strelcov.

Riešenie. Označme X počet nastrieľaných bodov 1. strelca a Y počet nastrieľaných bodov 2. strelca. Vypočítame aritmetické priemery počtu nastrieľaných bodov u oboch strelcov:

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

$$\bar{x} = \frac{1}{3}\cdot(7+8+9) = 8; \qquad \bar{y} = \frac{1}{3}\cdot(1+10+13) = 8.$$

To znamená, že obaja strelci priemerne nastrieľali 8 bodov. Napriek tomu, že priemerné výkony oboch strelcov sú rovnaké, ich výkony sú rozdielne. Túto odlišnosť vyjadríme pomocou priemernej odchýlky. Jej hodnotu vypočítame pomocou nasledujúcich tabuliek:

Tabuľka 1.16.

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x}\rvert$
7 8 9	-1 0 1	1 0 1
Σ	0	2

Tabuľka 1.17.

y_i	$y_i - \bar{y}$	$\lvert y_i - \bar{y}\rvert$
1 10 13	-7 2 5	7 2 5
Σ	0	14

1. strelec: $e_1 = \frac{1}{n}\sum_{i=1}^{n}\lvert x_i - \bar{x}\rvert = \frac{1}{3}\cdot 2 = 0,6\overline{6}$ bodu

2. strelec: $e_2 = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \bar{y}\rvert = \frac{1}{3}\cdot 14 = 4,67$ bodu.

Poznámka. Pretože súčet odchýlok hodnôt znaku od ich aritmetického priemeru je rovný nule, pri výpočte priemernej odchýlky sa počítal súčet absolútnych hodnôt uvedených odchýlok. Ďalšou možnosťou ako odstrániť vplyv znamienka je umocniť odchýlky na druhú. Ak odchýlky umocníme na druhú, nielenže odstránime vplyv znamienka, ale zároveň zvýrazníme extrémne hodnoty v súbore - malé odchýlky (menšie v absolútnej hodnote ako 1) budú po umocnení ešte menšie a veľké

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

odchýlky sa umocnením ešte viac zväčšia. Tým je motivovaná definícia *rozptylu* a *smerodajnej odchýlky*.

c) Rozptyl (ozn. σ^2)

Rozptyl (variancia, disperzia) je najčastejšie používaná charakteristika variability. Počítame ho ako aritmetický priemer štvorcov odchýlok hodnôt znaku od ich aritmetického priemeru podľa vzťahu:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

alebo, ak je daná tabuľka rozdelenia početností, podľa vzťahu:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x})^2 \cdot f_i \,, \quad \text{pričom} \quad \sum_{i=1}^{k} f_i = n.$$

Pri výpočte rozptylu je výhodné použiť nasledujúci upravený tvar predchádzajúceho vzorca:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k} x_i^2 \cdot f_i - \overline{x}^2 .$$

d) Smerodajná odchýlka (ozn. σ)

Smerodajná odchýlka σ je druhá odmocnina rozptylu: $\sigma = \sqrt{\sigma^2}$. Variabilita znaku sa zvykne charakterizovať pomocou smerodajnej odchýlky, pretože smerodajná odchýlka má rovnaký rozmer ako pozorovaný znak.

Príklad 1.21 Vypočítajme rozptyl a smerodajnú odchýlku zo súborov hodnôt z príkladu 1.20. Výpočet sa zjednoduší, ak tabuľky 1.16 a 1.17 doplníme o ďalšie stĺpce.

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

Tabuľka 1.18.

x_i	$x_i - \bar{x}$	$\lvert x_i - \bar{x}\rvert$	$(x_i - \bar{x})^2$
7	-1	1	1
8	0	0	0
9	1	1	1
Σ	0	2	2

Tabuľka 1.19.

y_i	$y_i - \bar{y}$	$\lvert y_i - \bar{y}\rvert$	$(y_i - \bar{y})^2$
1	-7	7	49
10	2	2	4
13	5	5	25
Σ	0	14	78

Rozptyl a smerodajná odchýlka zásahov prvého strelca je

$$\sigma_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{3}\cdot 2 = 0{,}66; \quad \sigma_1 = \sqrt{\sigma_1^2} = \sqrt{0{,}66} = 0{,}82.$$

Rovnako vypočítame rozptyl a smerodajnú odchýlku zásahov druhého strelca:

$$\sigma_2^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{3}\cdot 78 = 26; \quad \sigma_2 = \sqrt{\sigma_2^2} = \sqrt{26} = 5{,}1.$$

Z vypočítaných hodnôt smerodajných odchýlok vidíme, že kým zásahy prvého strelca sa od aritmetického priemeru odchyľujú o menej ako 1 v oboch smeroch (ležia teda medzi 7 a 9), zásahy druhého strelca sa odlišujú od priemeru viac ako o 5 (ležia medzi 3 a 13).

Príklad 1.22 Vypočítajme rozptyl a smerodajnú odchýlku súboru z príkladu 1.1. Tabuľku 1.8 doplníme o ďalšie stĺpce.

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

Tabuľka 1.20.

x_i	f_i	$x_i \cdot f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
1	4	4	-1,6	2,56	10,24
2	7	14	-0,6	0,36	2,52
3	10	30	0,4	0,16	1,6
4	3	12	1,4	1,96	5,88
5	1	5	2,4	5,76	5,76
Spolu	25	65	-	-	26

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot f_i = \frac{1}{25}\cdot 26 = 1{,}04; \quad \sigma = \sqrt{\sigma^2} = \sqrt{1{,}04} = 1{,}019 \doteq 1$$

Rovnakým spôsobom počítame rozptyl a smerodajnú odchýlku v prípade intervalového triedenia iba s tým rozdielom, že namiesto hodnôt znaku dosadíme do vzorca pre výpočet rozptylu stredy triednych intervalov. Výpočet budeme ilustrovať na nasledujúcom príklade.

Príklad 1.23 Vypočítajme rozptyl a smerodajnú odchýlku pre hodnoty z príkladu 1.4. Podobne ako v predchádzajúcom príklade, doplníme tabuľku 1.10 o ďalšie stĺpce. Dostaneme tabuľku 1.21.

Tabuľka 1.21.

Intervaly	Stredy intervalov x_i	Absolútna početnosť f_i	$x_i \cdot f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
$\langle 155, 160 \rangle$	157,5	1	157,5	-13,5	182,25	182,25
$\langle 160, 165 \rangle$	162,5	2	325	-8,5	72,25	144,5
$\langle 165, 170 \rangle$	167,5	2	335	-3,5	12,25	7
$\langle 170, 175 \rangle$	172,5	4	690	-1,5	2,25	9
$\langle 175, 180 \rangle$	177,5	5	887,5	6,5	42,25	211,25
Spolu	-	14	2395	-	-	554

Pomocou výsledkov z tejto tabuľky vypočítame rozptyl a smerodajnú odchýlku:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot f_i = \frac{1}{14}\cdot 554 = 39{,}57;$$

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

$$\sigma = \sqrt{\sigma^2} = \sqrt{39{,}57} = 6{,}3 \doteq 6 \text{ cm.}$$

Príklad 1.24 Vypočítame rozptyl a smerodajnú odchýlku pre hodnoty z príkladu 1.5. Najskôr zostavíme nasledujúcu tabuľku.

Tabuľka 1.22.

Intervaly	Stredy intervalov x_i	Absolútna početnosť f_i	$x_i \cdot f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot f_i$
19 - 23	21	4	84	-14	196	784
24 - 28	26	2	52	-9	81	162
29 - 33	31	6	186	-4	16	96
34 - 38	36	16	576	1	1	16
39 - 43	41	10	410	6	36	360
44 - 48	46	2	92	11	121	242
Spolu	-	40	1400	-	-	1660

Dostávame:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot f_i = \frac{1}{40} \cdot 1660 = 41{,}5,$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{41{,}5} = 6{,}44 \doteq 6 \text{ bodov.}$$

Poznámka. Ak charakteristikou strednej hodnoty je medián, príslušnou charakteristikou variability bude priemerná odchýlka e, pričom odchýlky budeme počítať od mediánu: $e = \frac{1}{n}\sum_{i=1}^{n}|x_i - \tilde{x}|$. V tomto prípade je možné použiť aj tzv. *medzikvartilovú odchýlku* (ozn. Q). K jej zavedeniu potrebujeme pojem *prvého* a *tretieho kvartilu* znaku X. Kým medián je prostredná hodnota v usporiadanej tabuľke hodnôt pozorovaného znaku, prvý kvartil Q_1 je hodnota „štvrtinová" a tretí kvartil Q_3 hodnota „trojštvrtinová". Presnejšie

$$Q_1 = \frac{1}{2}(x_{\frac{n}{4}} + x_{\frac{n}{4}+1}), \text{ ak je } n \text{ deliteľné štyrmi,}$$

$$Q_1 = x_{n_1}, \text{ ak } n \text{ nie je deliteľné štyrmi,}$$

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

kde n_1 je číslo $\frac{n}{4}$ zaokrúhlené na najbližšie vyššie celé číslo.

Podobne:

$$Q_3 = \frac{1}{2}\left(x_{\frac{3n}{4}} + x_{\frac{3n}{4}+1}\right), \text{ ak je } n \text{ deliteľné štyrmi,}$$

$$Q_3 = x_{n_3}, \text{ ak } n \text{ nie je deliteľné štyrmi,}$$

kde n_3 je číslo $\frac{3n}{4}$ zaokrúhlené na najbližšie vyššie celé číslo.

Medzikvartilová odchýlka znaku X je definovaná vzťahom:

$$Q = \frac{1}{2}(Q_3 - Q_1).$$

Príklad 1.25 Vypočítajme medzikvartilovú odchýlku mesačného príjmu z príkladu 1.13. Keďže $n = 20$ je deliteľné štyrmi, dostávame

$$Q_1 = \frac{1}{2}(x_5 + x_6) = 400, \quad Q_3 = \frac{1}{2}(x_{15} + x_{16}) = 600,$$

$$Q = \frac{1}{2}(Q_3 - Q_1) = \frac{600 - 400}{2} = 100 \text{ (eur).}$$

Podobne ako aritmetický priemer aj smerodajná odchýlka by tu bola príliš skreslená extrémne veľkou hodnotou $x_{20} = 8400$ eur.

Porovnanie variability

Ak máme dva súbory, často nás zaujíma, či hodnoty skúmaného znaku sú v prvom súbore viac rozptýlené ako hodnoty tohto znaku v druhom súbore. V prípade, že oba súbory sú rovnako veľké, ak sa meranie v oboch súboroch urobilo podľa rovnakej metodiky a zistili sa približne rovnako veľké priemery, variabilitu znaku v oboch súboroch môžeme porovnať pomocou ich rozptylov a smerodajných odchýlok.

V opačnom prípade na porovnanie variability oboch súborov použijeme *Pearsonov variačný koeficient* (ozn. v). Vypočítame ho ako podiel smerodajnej odchýlky a aritmetického priemeru:

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

$$v = \frac{\sigma}{\bar{x}}.$$

Variačný koeficient zvyčajne počítame v %:

$$v = \frac{\sigma}{\bar{x}} \cdot 100.$$

Vyjadruje mieru variability v percentách aritmetického priemeru.

Príklad 1.26 Pomocou variačného koeficienta porovnajme variabilitu zásahov strelcov z príkladu 1.20.

$$v_1 = \frac{\sigma_1}{\bar{x}} \cdot 100 = \frac{0{,}82}{8} . 100 = 10{,}2\ \%;$$

$$v_2 = \frac{\sigma_2}{\bar{y}} \cdot 100 = \frac{5{,}1}{8} . 100 = 63{,}5\ \%.$$

Z vypočítaných hodnôt variačného koeficienta vidíme, že variabilita zásahov druhého strelca je výrazne vyššia ako variabilita zásahov prvého strelca.

Príklad 1.27 U $n = 100$ osôb sa zisťoval počet zásahov pri experimente na reakčnom prístroji. Počet zásahov je mierou kvality výkonu. V nasledujúcej tabuľke sú uvedené priemerné počty zásahov ako aj smerodajné odchýlky (σ) a variačné koeficienty (v %) z prvého, piateho a desiateho pokusu. Chceme zistiť, či sa mení variabilita individuálnych výkonov počas tréningu.

Tabuľka 1.23.

	Priemerný počet zásahov	σ	v (v %)
1. pokus	13,85	4,75	34,3
5. pokus	22,60	4,65	20,6
10. pokus	24,50	3,90	15,9

Riešenie. Ak by sme pri porovnávaní individuálnych výkonov v jednotlivých pokusoch porovnávali iba smerodajné odchýlky, na základe vypočítaných hodnôt by sme získali dojem, že sa jedná iba o veľmi nepatrný rozdiel medzi výkonmi v uvedených pokusoch. To však nezodpovedá skutočnosti, pretože môžeme vidieť, že priemerné počty

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

zásahov v jednotlivých pokusoch sú značne rozdielne. Ak porovnáme variačné koeficienty, vidíme, že variabilita v priebehu tréningu výrazne klesá. Kým v 1. pokuse dosahovala smerodajná odchýlka 34 % priemeru, v 10. pokuse klesla na 16 %.

Variabilita znaku pozorovaného na prvkoch viacerých súborov

Predpokladajme, že určitý znak bol pozorovaný na jednotkách k štatistických súborov. Niekedy je potrebné charakterizovať variabilitu pozorovaného znaku na súbore vytvorenom z daných k súborov. Nech rozptyly v týchto súboroch sú $\sigma_1^2, \sigma_2^2, ..., \sigma_k^2$. Budeme rozlišovať dva prípady.

1. Všetky súbory majú rovnaké rozsahy ($n_1 = n_2 = ... = n_k$). V tomto prípade vypočítame celkový rozptyl podľa vzorca:

$$\sigma^2 = \frac{\sigma_1^2 + \sigma_2^2 + ... + \sigma_k^2}{k}; \ \sigma = \sqrt{\sigma^2}.$$

2. Nie všetky rozsahy sú rovnaké. Celkový rozptyl počítame podľa vzorca

$$\sigma^2 = \frac{1}{n}(n_1 \cdot \sigma_1^2 + n_2 \cdot \sigma_2^2 + ... + n_k \cdot \sigma_k^2); \quad n = n_1 + n_2 + ... + n_k.$$

Poznámka. Uvedené vzorce platia len pre súbory s približne rovnakými priemermi.

Príklad 1.28 Žiaci troch tried tretieho ročníka dostali z písomnej práce z matematiky tieto známky.
3.A: 1,2,2,2,3,3,1,4,4,1,2,2,2,3,5,4,2,3,3,2,1,2,3,1,3,2,2,4,1,2
3.B: 2,3,2,3,4,1,1,2,2,5,5,4,4,3,2,2,1,3,3,2,1,1,4,2,3,2,2,3,2,2,3,4,2
3.C: 1,1,2,3,1,1,1,2,1,2,3,2,1,1,1,1,1,1,2,1,1.

Vypočítajte priemernú známku z matematiky žiakov tretieho ročníka, rozptyl známok z matematiky žiakov tretieho ročníka a porovnajte variabilitu známok v jednotlivých triedach.

Riešenie. Zostrojíme najskôr tabuľky rozdelenia početností doplnené o príslušné stĺpce.

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com

Tabuľka 1.24.

x_i	3.A			3.B			3.C		
	f_i	$x_i f_i$	$x_i^2 f_i$	f_i	$x_i f_i$	$x_i^2 f_i$	f_i	$x_i f_i$	$x_i^2 f_i$
1	6	6	6	5	5	5	14	14	14
2	12	24	48	13	26	52	5	10	20
3	7	21	63	8	24	72	2	6	18
4	4	16	64	5	20	80	0	0	0
5	1	5	25	2	10	50	0	0	0
Σ	30	72	206	33	85	259	21	30	52

3.A:

$$\bar{x}_1 = \frac{72}{30} = 2{,}4; \quad \sigma_1^{\ 2} = \frac{1}{n_1}\sum_{i=1}^{k_1} x_i^2 \cdot f_i - \bar{x}_1^2 = \frac{206}{30} - 2{,}4^2 = 1{,}11; \quad v_1 = 43{,}9\%$$

3.B:

$$\bar{x}_2 = \frac{85}{33} = 2{,}6; \quad \sigma_2^2 = \frac{1}{n_2}\sum_{i=1}^{k_2} x_i^2 \cdot f_i - \bar{x}_2^2 = \frac{259}{33} - 2{,}6^2 = 1{,}24; \quad v_2 = 42{,}6\%$$

3.C:

$$\bar{x}_3 = \frac{30}{21} = 1{,}4; \quad \sigma_3^{\ 2} = \frac{1}{n_3}\sum_{i=1}^{k_3} x_i^2 \cdot f_i - \bar{x}_3^2 = \frac{52}{21} - 1{,}4^2 = 0{,}46; \quad v_3 = 47{,}7\%$$

Celkový priemer :

$$\bar{x}_v = \frac{2{,}4 \cdot 30 + 2{,}57 \cdot 33 + 1{,}42 \cdot 21}{30 + 33 + 21} = 2{,}22.$$

Celkový rozptyl:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{k} x_i^2 \cdot f_i - \bar{x}_v^2 = \frac{1}{84} \cdot (206 + 259 + 52) - 2{,}22^2 = 1{,}82.$$

Na základe vypočítaných hodnôt variačných koeficientov vidíme, že najväčšia variabilita známok je v 3. C triede.

This document was created using SOLID CONVERTER PDF To remove this message, purchase the product at www.SolidDocuments.com