STATISTIKA

1. Základní pojmy

Statistika – věda o metodách sběru, zpracování a vyhodnocování statistických údajů. Zkoumá společenské, přírodní, technické a jiné jevy vždy na dostatečně rozsáhlém souboru údajů.

Matematická statistika vychází ze shromážděných statistických údajů. Zabývá se jejich matematickým zpracováním a rozborem výsledků.

Statistické údaje (data) jsou číselné údaje o společenských, přírodních, technických a jiných skutečnostech, o tzv. **statistických jevech.**

Statistický soubor je soubor osob, věcí, událostí, jevů apod. shromážděných na základě určité společné vlastnosti.

Statistické jednotky jsou jednotlivé prvky statistického souboru.

Rozsah souboru (ozn. ***n***) je počet všech prvků statistického souboru.

Statistický znak je společná vlastnost statistického souboru, kterou vyšetřujeme. Ten může být buď **kvantitativní** (je vyjádřen číselným údajem, jako např. známka z testu, výška, váha, hrubý roční příjem, atd.) nebo **kvalitativní** (je vyjádřen slovním popisem – např. národnost, občanství, povolání, druh nemoci apod.).

2. Zpracování statistického souboru

Třídění souboru

Ve většině šetření je počet různých hodnot sledovaného znaku x menší, než počet jednotek n tohoto statistického souboru. Znamená to tedy, že několik různých statistických jednotek téhož souboru má stejné hodnoty. Hovoříme o tzv. **četnosti** nebo také **absolutní četnosti.**

Absolutní četnost (četnost) hodnoty znaku (ozn. n_j) udává počet statistických jednotek, kterým přísluší stejná hodnota znaku x_j, tzn. že pomocí této absolutní četnosti lze **roztřídit soubor do *r* tříd** ($r < n$). Tedy místo n hodnot $x_1, x_2, \cdots, x_n$ pracujeme pouze s r hodnotami $x_1^*, x_2^*, \cdots, x_r^*$, přičemž každá tato hodnota x_j^* má svou absolutní četnost n_j.

Součet četností všech možných hodnot znaku se rovná rozsahu soboru, tedy $\sum_{j=1}^{r} n_j = n$.

Relativní četnost hodnoty znaku ($v_j = \frac{n_j}{n}$) udává, jaká část souboru má hodnotu znaku x_j^*.

Součet relativních četností se rovná jedné, tedy $\sum_{j=1}^{r} v_j = 1$.

Relativní četnosti se velmi často vyjadřují v procentech a pak platí, že jejich součet je 100 %.

Všechny různé hodnoty znaku x_j^* (tedy třídy $x_1^*, x_2^*, \cdots, x_r^*$) a jim odpovídající četnosti n_j (tedy $n_1, n_2, \cdots, n_r$), resp. relativní četnosti v_j (tedy $v_1, v_2, \cdots, v_r$) lze zapsat do tabulky, kterou nazýváme **tabulka rozdělení četností a relativních četností.**
Rozdělení četností (absolutních či relativních) lze znázornit graficky, provádíme tzv. **grafické znázornění četností.** Nejčastěji užíváme polygon četností a histogram četností.

Příklad 2.1. **Ve třídě byly zaznamenány tyto známky z testu:**

Prvotní tabulka (netříděný soubor):

Jméno žáka	*Známka z testu (x_j)*	*Jméno žáka*	*Známka z testu (x_j)*	*Jméno žáka*	*Známka z testu (x_j)*	*Jméno žáka*	*Známka z testu (x_j)*	*Jméno žáka*	*Známka z testu (x_j)*
Adam	2	Frézová	2	Janeček	1	Opletalová	2	Šedivá	3
Béla	3	Galatek	3	Kadeřavý	4	Paleček	1	Ulrych	2
Čejek	1	Hanáček	5	Lakomá	2	Remunda	3	Válek	3
Dlouhá	3	Chromá	4	Mastný	1	Řečný	3	Wilczek	5
Elin	4	Indra	3	Nedoma	3	Sitek	2	Žežulka	1

Proveďte roztřídění tohoto souboru na vhodné třídy a stanovte absolutní a relativní četnost každé třídy. Proveďte grafické znázornění četností.

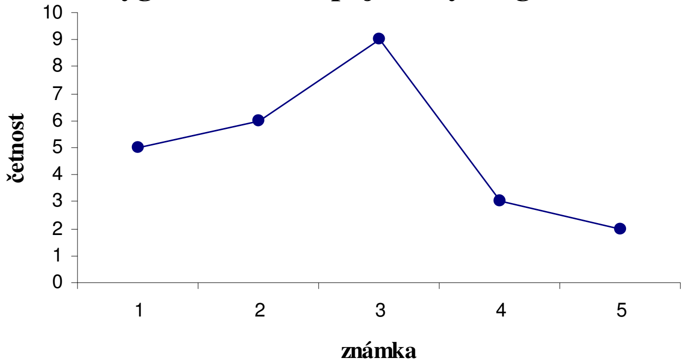
Řešení: V prvotní tabulce je zaznamenán tzv. **netříděný soubor**, který pomocí tříd a četností roztřídíme:

Tabulka rozdělení četností a relativních četností (tříděný soubor)*:*

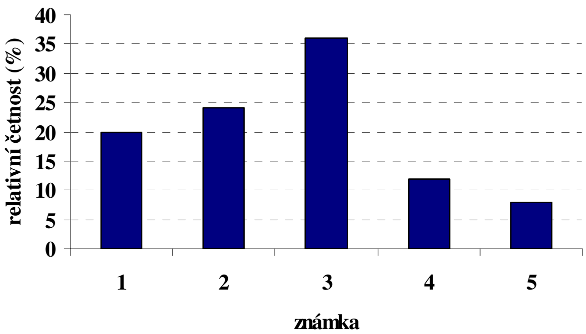
Třída x_j^*	*Četnost znaku* n_j	*Relativní četnost znaku* $\frac{n_j}{n}$
$x_1^* = 1$	$n_1 = 5$	$\frac{5}{25} = 0{,}2$
$x_2^* = 2$	$n_2 = 6$	$\frac{6}{25} = 0{,}24$
$x_3^* = 3$	$n_3 = 9$	$\frac{9}{25} = 0{,}36$
$x_4^* = 4$	$n_4 = 3$	$\frac{3}{25} = 0{,}12$
$x_5^* = 5$	$n_5 = 2$	$\frac{2}{25} = 0{,}08$
součet	$\sum_{j=1}^{5} n_j = 25$	$\sum_{j=1}^{5} \nu_j = 1$

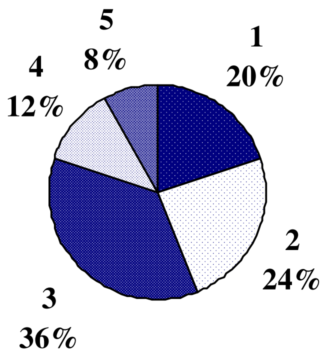
Grafické znázornění rozdělení četností:

Polygon četnosti (spojnicový diagram)



Histogram četnosti (sloupkový a kruhový diagram)





Poznámka: *Postupují-li hodnoty kvantitativního znaku po příliš malých krocích nebo je jich příliš mnoho, sdružujeme je v intervaly a hodnoty z téhož intervalu nahradíme středem tohoto intervalu.*

Příklad 2.2. **Statistickým soborem je 200 žáků školy, sledovaným znakem jejich výška. Byly zjištěny tyto údaje:**

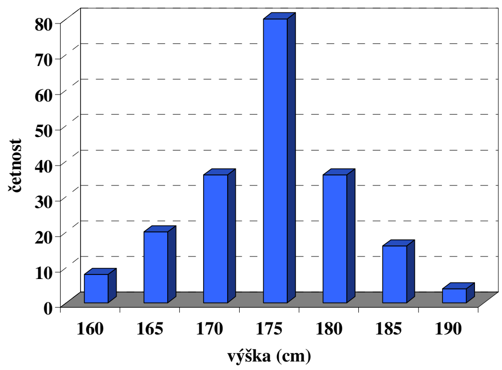
Výška (cm)	158-162	163-167	168-172	173-177	178-182	183-187	188-192
Počet žáků	8	20	36	80	36	16	4

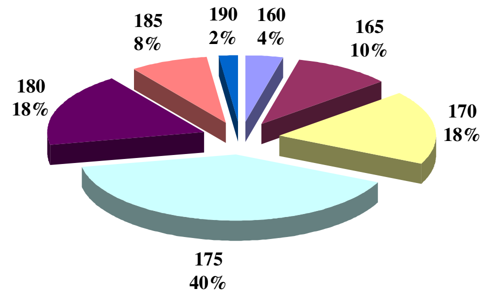
Najděte střed každého intervalu (třídy) a stanovte absolutní a relativní četnost každé třídy. Proveďte grafické znázornění četností histogramy četnosti.

Řešení: Místo celého intervalu zapíšeme jeho střed:

x_j^* ***(cm)***	160	165	170	175	180	185	190
n_j ***(počet žáků)***	8	20	36	80	36	16	4
v_j ***(%)***	4%	10%	18%	40%	18%	8%	2%

Histogram četnosti (sloupkový a kruhový diagram)





Cvičení 2:

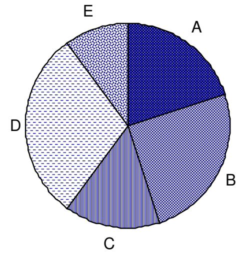
1) Při zjišťování věku posluchačů jedné studijní skupiny na vysoké škole byly zjištěny tyto hodnoty: 18, 19, 18, 18, 19, 18, 20, 21, 20, 21, 22, 22, 18, 18, 18, 19, 19, 18, 19, 20. Určete rozsah souboru, sestavte tabulku rozdělení četností jednotlivých hodnot znaku „věk“ a určete relativní četnosti. Sestrojte spojnicový diagram (tzv. polygon četnosti) rozdělení četností.

2) Ve třídě s 25 žáky prospělo s vyznamenáním 7 žáků, prospělo 14 žáků, neprospěli 3 žáci, nebyl klasifikován 1 žák. Vypočtěte relativní četnosti znaku „prospěch“ a ukažte, že jejich součet je roven 1. Sestrojte kruhový diagram (tzv. histogram) rozdělení četností.

3) V prodejně pánské obuvi zaznamenávali velikosti prodaných párů během dne s tímto výsledkem: 41, 41, 41, 42, 42, 41, 39, 41, 37, 41, 45, 41, 42, 38, 40, 39, 38, 41, 41, 38, 42, 39, 44, 43, 43, 44, 39, 39, 43, 43, 40, 42, 43, 41, 41, 43, 40, 40, 40, 42, 42, 42, 41, 40, 42. Určete rozsah souboru, vypočtěte absolutní a relativní četnosti znaku „velikost“. Relativní četnosti vyjádřete v procentech a ověřte, zda jejich součet je roven 100%. Sestrojte sloupkový diagram rozdělení četností.

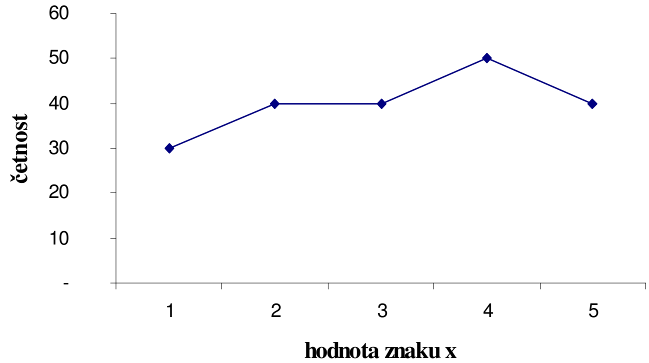
4) Ze „Školní statistické ročenky 2008“ vyjměte část tabulky **Počítačové dovednosti jednotlivců v ČR v roce 2007** týkající se věkových skupin (str.126). Tabulku zapište a sestrojte spojnicový diagram a histogram (sloupkový či kruhový).
Poznámka: „Školní statistickou ročenku 2008“ naleznete též na internetových stránkách Českého statistického úřadu (www.czso.cz) v sekci pro studenty.

5) Kruhový diagram vyjadřuje v procentech volební preference pěti politických stran. Jsou-li volební preference strany *A* znázorněny kruhovou výsečí se středovým úhlem 72°, jsou preference této strany:
a) 15 % b) 20 % c) 25 % d) 30 %

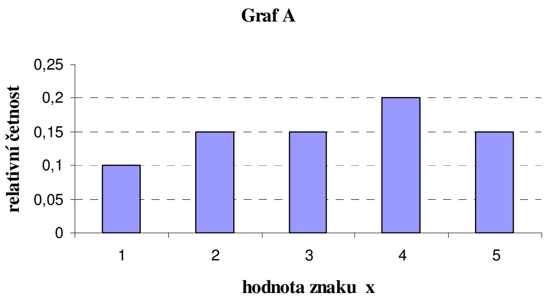


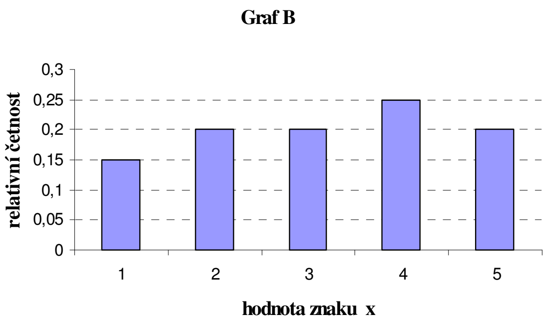
V: 20 %

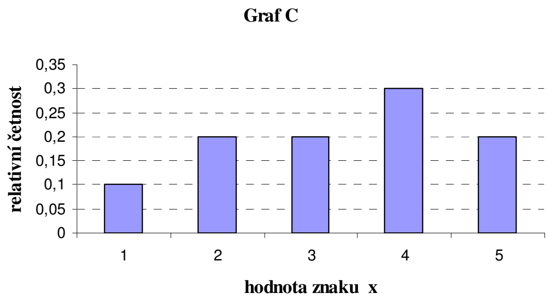
6) Na obrázku je spojnicový diagram, který znázorňuje četnosti všech pěti hodnot daného znaku *x*:

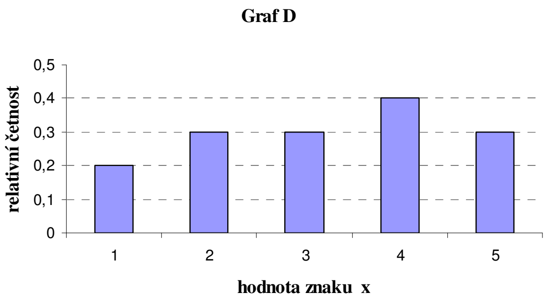


Z následujících sloupkových diagramů znázorňujících relativní četnosti hodnot tohoto znaku vyberte ten, který odpovídá uvedenému spojnicovému diagramu:









V: Graf B

3. Statistické charakteristiky

Statistickými charakteristikami nazýváme čísla, která podávají stručnou a souhrnnou informaci o souboru. Pokud se omezíme na podmínku, že vyšetřujeme statistický soubor na základě jediného kvantitativního znaku, jedná se o **charakteristiky polohy (úrovně)** a **variability (proměnlivosti)**. Základní statistické charakteristiky slouží pro vzájemné porovnávání různých statistických souborů.

Charakteristiky polohy hodnoty znaku (neboli **střední hodnoty znaku**) jsou čísla, která zastupují celý soubor. Jedno číslo nahrazuje dlouhou řadu hodnot zkoumaného znaku a je pak snadné porovnat dva i více statistických souborů. Charakterizují polohu znaku na číselné ose. Mezi ně patří **aritmetický průměr, geometrický průměr, harmonický průměr, medián, modus.**

Aritmetický průměr (ozn. $\bar{x}$) je nejužívanější statistická veličina, která v jistém smyslu vyjadřuje typickou hodnotu popisující daný soubor mnoha hodnot (např. průměrný plat v ČR). Má smysl jako velmi důležitá charakteristika daného souboru tehdy, pokud jsou odchylky naměřených hodnot nahodilé a v souboru se nevyskytují extrémně nízké nebo vysoké hodnoty.
Definice aritmetického průměru:

- aritmetický průměr **prostý**:

$$\bar{x} = \frac{1}{n}\left(x_1 + x_2 + \cdots + x_n\right) = \frac{1}{n}\sum_{i=1}^{n} x_i ,$$

- aritmetický průměr **vážený**:

$$\bar{x} = \frac{1}{n}\left(n_1x_1 + n_2x_2 + \cdots + n_rx_r\right) = \frac{1}{n}\sum_{j=1}^{r} n_jx_j ,$$

kde $n_1, n_2, \cdots, n_r$ jsou četnosti tříd $x_1^*, x_2^*, \cdots, x_r^*$. Užíváme tam, kde statistický soubor je složen z r podsouborů. Máme-li tedy sestavenu tabulku rozdělení četností, užijeme pro výpočet aritmetického průměru vážený aritmetický průměr.

Příklad 3.1. **V jednom ročníku školy jsou dvě paralelní třídy. Třída A má 20 studentů, třída B 32 studentů. Při jednom testu získali žáci těchto tříd tyto body:**

- **Třída *A*: 62, 67, 71, 74, 76, 77, 78, 79, 79, 80, 80, 81, 81, 82, 83, 84, 86, 89, 93, 98.**
- **Třída *B*: 80, 81, 82, 83, 84, 85, 86, 87, 87, 88, 88, 89, 89, 89, 90, 90, 90, 90, 91, 91, 91, 92, 92, 93, 93, 94, 95, 96, 97, 98, 99, 100.**

Vypočtěte aritmetický průměr bodů ve třídě *A*, ve třídě *B* a v celém ročníku.

Řešení: $\bar{x}_A = \dfrac{1600}{20} = 80$, $\bar{x}_B = \dfrac{2880}{32} = 90$

Chceme-li vypočítat průměr bodů za celý ročník, nesmíme ho počítat jako průměr ze získaných dvou průměrů, tj. $\dfrac{80+90}{2} = 85$!!! Každý z nich má totiž jinou „váhu“. První zastupuje 20 žáků, druhý 32.
Proto celkový průměr vyřešíme

- buď jako podíl součtu všech bodů v ročníku a celkového počtu žáků v ročníku

$$\bar{x} = \frac{1600+2880}{20+32} = \frac{4480}{52} \doteq 86{,}154$$

- nebo pomocí váženého průměru

$$\bar{x} = \frac{20\cdot 80 + 32\cdot 90}{20+32} = \frac{4480}{52} \doteq 86{,}154$$

V tomto případě stačilo k určení celkové průměru znát průměry a počty žáků ve třídách (nemusíme již znát body u jednotlivých žáků).

Závěr: Průměr bodů ve třídě *A* je 80 bodů, ve třídě *B* 90 bodů, celkový průměr v ročníku je 86,154.

Příklad 3.2. **Výsledky srovnávací písemné práce z matematiky ve dvou paralelních třídách 4.*A* a 4.*B* ukázaly následující:**
Třída 4.*A* (počet žáků 30) napsala písemnou práci s průměrem 2,50,
třída 4.*B* (počet žáků 25) napsala tuto práci s průměrem 2,60, ale na rozdíl od sousední třídy zde nikdo neměl nedostatečnou známku.
Marek, student třídy 4.*B*, si uvědomil, že kdyby napsal písemku lépe, mohla skončit jeho třída lépe než sousední. Jakou známku z písemky dostal?

Řešení: $\bar{x}_A = 2{,}50, \quad \bar{x}_B = 2{,}60$

Označíme-li celkový počet bodů získaných třídou 4.*B* jako *s*, pak platí, že

$$\bar{x}_B = \frac{s}{25} = 2{,}60 \quad \Rightarrow \quad s = 2{,}60 \cdot 25 \quad \Rightarrow \quad s = 65$$

Kdyby Marek dostal známku o *m* stupňů lepší než ve skutečnosti a jeho třída by tedy dopadla lépe než 4.*A*, muselo by platit

$$\frac{65 - m}{25} < 2{,}50 \quad \Rightarrow \quad m > 2{,}5$$

Závěr: Marek se tedy musel zlepšit o 3 stupně. Protože však žádný žák v jeho třídě neměl pětku, měl tedy čtyřku a musel by se zlepšit až na jedničku.

Příklad 3.3. **Průměrná denní teplota se v meteorologii stanovuje jako průměr z teploty vzduchu naměřené v 7 hodin ráno, teploty ve 14 hodin a teploty v 21 hodin, přičemž poslední údaj se započítává s dvojnásobnou vahou. Vypočtěte průměrnou denní teplotu, jestliže ranní teplota byla -4°C, odpolední 2°C a večerní -3°C.**

Řešení: Platí tedy $\bar{t} = \dfrac{t_7 + t_{14} + 2 \cdot t_{21}}{4} \quad \Rightarrow \quad \bar{t} = \dfrac{-4 + 2 + 2 \cdot (-3)}{4} = -2\,°$

Závěr: Průměrná denní teplota byla -2°C.

Příklad 3.4. **Průměrný měsíční plat zaměstnanců firmy v prvním až třetím čtvrtletí byl 13 600 Kč. Ve čtvrtém čtvrtletí vyplatila firma zaměstnancům mimořádné odměny, a tak jejich průměrný měsíční plat za celý loňský rok činil 14 800 Kč. O kolik korun byla průměrná měsíční mzda zaměstnanců firmy ve čtvrtém čtvrtletí větší ve srovnání s jejich průměrnou měsíční mzdou v prvním až třetím čtvrtletí?**

Řešení: I.- III.čtvrtletí, tj. 9 měsíců ……………………13 600 Kč za 1 měsíc
IV.čtvrtletí, tj. 3 měsíce ……………………… x Kč za 1 měsíc
celý rok, tj. 12 měsíců ……………………… 14 800 Kč za 1 měsíc

Odtud plyne rovnice:

$$13600 \cdot 9 + 3x = 14800 \cdot 12$$

$$x = 18400$$

Rozdíl činí $18400 - 13600 = 4800\,\text{Kč}$.

Závěr: Průměrná měsíční mzda zaměstnanců firmy ve čtvrtém čtvrtletí byla o 4 800 Kč větší ve srovnání s jejich průměrnou měsíční mzdou v prvním až třetím čtvrtletí.

Vlastnosti aritmetického průměru:

- Změní-li se každá z hodnot $x_1, x_2, \cdots, x_n$ o stejnou konstantu *c*, pak se o tuto konstantu změní i aritmetický průměr:

$$\frac{(x_1+c)+(x_2+c)+\cdots+(x_n+c)}{n}=\frac{(x_1+x_2+\cdots+x_n+nc)}{n}=\frac{x_1+x_2+\cdots+x_n}{n}+\frac{nc}{n}=\overline{x}+c$$

- Změní-li se každá z hodnot $x_1, x_2, \cdots, x_n$ *k*-násobně, pak se *k*-násobně změní i aritmetický průměr:

$$\frac{(kx_1)+(kx_2)+\cdots+(kx_n)}{n}=\frac{k(x_1+x_2+\cdots+x_n)}{n}=k\overline{x}$$

Příklad 3.5. **Každý z *n* stromků vyrostl za 5 let do dvojnásobné výše. Kolikrát se zvětšila průměrná výška všech těchto stromků?**

Řešení: Jedná se o přímé užití výše uvedené vlastnosti:

$$\frac{(2x_1)+(2x_2)+\cdots+(2x_n)}{n}=\frac{2(x_1+x_2+\cdots+x_n)}{n}=2\overline{x}$$

Závěr: Průměrná výška stromků se zvětšila dvakrát.

Příklad 3.6. **Majitel stavební firmy vyplatil za řádné plnění zakázky každému ze šesti pomocníků stejnou částku, každému ze čtyř zedníků také stejnou částku, ale vyšší než pomocníkům, nejvyšší částku vyplatil mistrovi. Průměrně vyplacená částka byla 6 730 Kč. Později se majitel firmy rozhodl vyplatit ještě každému z uvedených pracovníků 10 % částky, kterou mu vyplatil původně. Jak vysoká byla průměrná dodatečně vyplacená částka?**

Řešení: Původně bylo v průměru zaměstnancům vyplaceno:

$$\overline{x}=\frac{6x+4y+z}{11}=6730 \text{ Kč}$$

Po dodatečném zvýšení bylo celkem průměrně vyplaceno:

$$\frac{6\cdot 1{,}1x+4\cdot 1{,}1y+1{,}1z}{11}=\frac{1{,}1\cdot(6x+4y+z)}{11}=1{,}1\cdot 6730=1\cdot 6730+0{,}1\cdot 6730=\overline{x}+673$$

Závěr: Průměrná dodatečně vyplacená částka byla 673 Kč.

Příklad 3.7. **Na sportovní utkání chce obec prodávat vstupenky za jednotnou cenu, a to za 50 Kč, 100 Kč, 150 Kč nebo 200 Kč. Podle předběžného průzkumu je 218 zájemců ochotno koupit si vstupenku nejvýše za 50 Kč, dalších 205 zájemců nejvýše za 100 Kč, dalších 148 zájemců vstupenku nejvýše za 150 Kč a pouze dalších 76 zájemců vstupenku až do ceny 200 Kč. Při jaké ceně vstupenky bude mít obec nejvyšší zisk?**

Řešení: Vstupenku za 50 Kč by koupilo (218 + 205 + 148 + 76) zájemců, zisk by byl

$$(218+205+148+76)\cdot 50=32350 \text{ Kč}.$$

Zisk za vstupenky po 100 Kč ………….. $(205+148+76)\cdot 100=42900$ Kč.

Zisk za vstupenky po 150 Kč ………….. $(148+76)\cdot 150=33600$ Kč.

Zisk za vstupenky po 200 Kč ………..… $76\cdot 200=15200$ Kč.

Závěr: Zisk obce bude největší, bude-li vstupenka stát 100 Kč.

Geometrický průměr z n kladných čísel $a_1, a_2, ..., a_n$ je definován jako n-tá odmocnina z jejich součinu:

$$\sqrt[n]{a_1 \cdot a_2 \cdot \ldots \cdot a_n} = \left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}}$$

Geometrický průměr $\bar{x}_G$ se ve statistice využívá k výpočtu průměrného tempa růstu v národohospodářských časových řadách, tedy tempa růstu průmyslové či jiné výroby.
Jsou-li hodnoty růstu $x_1, x_2, \cdots, x_n$ uvedeny v procentech, pak průměrné roční tempo růstu

$$\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \cdots \cdot x_n}$$

získáme rovněž v procentech.

Příklad 3.8. **Vypočtěte průměrný růst cen za poslední čtyři roky, jestliže byl zaznamenán postupný nárůst o 20 % z předchozího roku a pak 10 %, poté 15 % pokles cen, ale pak zase 10 % růst cen.**

Řešení: Průměrné roční tempo růstu cen vyjádříme geometrickým průměrem:

$$\bar{x}_G = \sqrt[4]{120 \cdot 110 \cdot 85 \cdot 110} = 105{,}40\ \%$$

Závěr: Průměrné roční tempo růstu cen bylo 5,4 %.

Poznámka: Do vzorce pro geometrický průměr lze samozřejmě dosazovat i desetinná čísla vystihující růst či pokles:

$$\bar{x}_G = \sqrt[4]{1{,}20 \cdot 1{,}10 \cdot 0{,}85 \cdot 1{,}10} = 1{,}054$$, což odpovídá vzrůstu o 5,4 %.

Příklad 3.9. **Geometrický průměr čtyř kladných čísel 4, 16, *x*, *x* je 2. Určete neznámá čísla, vypočtěte pak jejich aritmetický průměr a porovnejte s průměrem geometrickým.**

Řešení: $\sqrt[4]{4 \cdot 16 \cdot x \cdot x} = 2 \quad \Rightarrow \quad x = 0{,}5$

Aritmetický průměr $\bar{x} = \dfrac{4 + 16 + 0{,}5 + 0{,}5}{4} = 5{,}25$

Závěr: Hledaná čísla jsou rovna číslu 0,5. Aritmetický průměr všech čtyř čísel je 5,25, což je číslo větší než jejich průměr geometrický.

Nerovnost aritmetického a geometrického průměru

V matematice říká **nerovnost aritmetického a geometrického průměru** (krátce **AG-nerovnost**), že aritmetický průměr skupiny nezáporných čísel je vždy větší nebo roven geometrickému průměru těchto čísel. Navíc, rovnost nastává tehdy a jen tehdy, pokud jsou všechna čísla ve skupině stejná.
Formálně se nerovnost zapíše

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \cdot x_2 \cdot \cdots \cdot x_n} .$$

Příklad 3.10. **Dokažte AG-nerovnost pro dvě nezáporná čísla *x*, *y*.**

Řešení: $\dfrac{x+y}{2} \geq \sqrt{xy} \quad \Rightarrow \quad (x+y)^2 \geq 4xy \quad \Rightarrow \quad x^2 - 2xy + y^2 \geq 0 \quad \Rightarrow \quad (x-y)^2 \geq 0$

Tatro nerovnost je splněna pro všechna nezáporná čísla x, y (druhá mocnina je vždy nezáporná). Rovnost nastává pouze pro $x = y$.

Harmonický průměr z nenulových hodnot statistického souboru je definován jako podíl rozsahu souboru (počtu členů) a součtu převrácených hodnot znaků. Jinými slovy je to převrácená hodnota aritmetického průměru převrácených hodnot zadaných členů.

$$\bar{x}_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

Harmonický průměr užíváme nejčastěji při výpočtu průměrné rychlosti či průměrného času.

Příklad 3.11. **Řidič zkušebního automobilu jel do cílového místa rychlostí 70 km/h a zpět rychlostí 90 km/h. Jakou průměrnou rychlost dosáhl na celé trase?**

Řešení: $$\bar{x}_h = \frac{2}{\frac{1}{70} + \frac{1}{90}} = 78{,}75$$

Závěr: Průměrná rychlost řidiče byla 78,75 km/h.

Příklad 3.12. **Tři pracovníci opakovaně provádějí stejnou operaci. Prvnímu trvá operace 2 minuty, druhému 3 minuty, třetímu 4 minuty. Jak dlouho trvá průměrně jedna operace?**

Řešení: $$\bar{x}_h = \frac{3}{\frac{1}{2} + \frac{1}{3} + \frac{1}{4}} \doteq 2{,}8$$

Závěr: Jedná operace trvá průměrně asi 2 minuty 48 sekund.

●●●●●●

Přejděme nyní od průměrů k doplňujícím charakteristikám polohy, jako jsou **modus a medián**.

Modus znaku *x* je hodnota, která má v souboru největší četnost.
Značíme Mod (x) nebo taky $\hat{x}$. Představuje jakousi typickou hodnotu sledovaného souboru a jeho určení předpokládá roztřídění souboru podle obměn znaku. Modus nemusí být určen jednoznačně (tzn., že se stejnou nejvyšší frekvencí se může vyskytovat více hodnot).
Jeho výhodou je, že ho lze použít i pro nečíselná data, kde např. aritmetický průměr použít nelze. Např. modus souboru {*jablko, pomeranč, hruška, pomeranč, jablko, jablko, hruška*} je jablko.

Medián znaku *x* je prostřední hodnota znaku, jsou-li hodnoty $x_1, x_2, \cdots, x_n$ uspořádány vzestupně.
Značíme Med (x) nebo taky $\tilde{x}$. Medián dělí řadu podle velikosti seřazených výsledků na dvě stejně početné poloviny. Platí, že nejméně 50 % hodnot je menších nebo rovných a nejméně 50 % hodnot je větších nebo rovných mediánu.
Pro nalezení mediánu daného souboru stačí hodnoty seřadit podle velikosti a vzít hodnotu, která se nalézá uprostřed seznamu. Pokud má soubor sudý počet prvků, obvykle se za medián označuje aritmetický průměr prostředních dvou hodnot.
Základní výhodou mediánu jako statistického ukazatele je fakt, že není ovlivněný extrémními hodnotami. Proto se často používá v takových souborech, u kterých aritmetický průměr dává obvykle nevhodné výsledky. Např. u souboru { 1, 2, 2, 3, 9 } je medián (stejně jako modus) roven dvěma, což je zřetelně vhodnější ukazatel převažující tendence než aritmetický průměr, který je zde roven 3,4.
Medián je nejpoužívanější **kvantil** (konkrétně kvantil dělící soubor na dvě části). Kromě mediánu se velmi často používají **kvartily** (soubor se dělí na čtyři části), **decily** (na deset částí) a **percentily** (na sto částí).

Příklad 3.13. **Ve třídě s 31 studenty byla zjišťována výše jejich kapesného na týden. Výsledky šetření jsou zpracovány v následující tabulce rozdělení četností:**

Výše kapesného	50 Kč	100 Kč	200 Kč	500 Kč
Počet studentů	15	12	3	1

Určete průměrnou hodnotu, modus a medián kapesného ve třídě. Porovnejte tyto charakteristiky polohy.

Řešení: Průměr: $\bar{x} = \dfrac{15 \cdot 50 + 12 \cdot 100 + 3 \cdot 200 + 1 \cdot 500}{31} \doteq 98 \text{ Kč}$

Modus: Nejčastěji vyskytující se hodnota je 50,- Kč $\Rightarrow$ $\hat{x} = 50,-$ Kč.

Medián: Seřadíme-li 31 hodnot vedle sebe ve vzrůstajícím pořadí, pak prostřední hodnota je na 16. místě, což je 100,- Kč $\Rightarrow$ $\tilde{x} = 100,-$ Kč

Závěr: Protože jedna hodnota znaku v tomto souboru je výrazně odlišná, je jako doplňková charakteristika polohy vhodnější medián. To dokládá i porovnání aritmetického průměru s mediánem. Modus je výrazně nižší než aritmetický průměr a medián.

Cvičení 3(1 část):

1) Při statistickém průzkumu v jedné obci byly zjištěny následující počty členů v domácnostech této obce. Vypočtěte průměrný počet členův jedné domácnosti.

Počet členů domácnosti	1	2	3	4	5	6	7	8
Počet domácností	52	123	119	85	22	63	19	17

V: $\bar{x} \doteq 3{,}504$

2) Průměrný obsah stříbra v náhodně vybraných vzorcích rudy udává následující tabulka. Vypočtěte průměrný obsah stříbra v této rudě.

Obsah stříbra v %	0-4	4-8	8-12	12-16	16-20	20-24
Počet vzorků	188	75	44	24	13	8

V: $\bar{x} \doteq 5{,}72$ %

3) Určete aritmetický průměr, modus a medián čísel $x_1, x_2, \cdots, x_{15}$, jestliže se číslo 2 vyskytuje mezi nimi pětkrát, číslo 7 osmkrát a čísla 10 a 12 jednou.

V: $\bar{x} \doteq 5{,}87$; $\hat{x} = 7$; $\tilde{x} = 7$

4) Ve třídě je 8 žáků zařazeno do volitelného předmětu informatika, 10 do cvičení z biologie a 14 do anglické konverzace. Průměrný prospěch v informatice byl 1,60, ve cvičení z biologie 1,40 a v anglické konverzaci 1,20. Jaký je průměrný prospěch třídy ve volitelných předmětech?

V: $\bar{x} \doteq 1{,}3625$

5) Aritmetický průměr tří čísel je 38,4. Součet dvou z nich 77,4. Určete třetí číslo.

V: 37,8

6) Několik jablek má průměrnou hmotnost 180 g. Kdybychom k nim přidali jedno jablko o hmotnosti 210 g, zvětšila by se průměrná hmotnost jablek o 3g. Určete původní počet jablek.

V: 9 jablek

7) Deset hráčů soutěžilo v hodu na koš. První hráč získal 11 bodů, druhý 8 bodů, třetí také 8 bodů, čtvrtý dosáhl aritmetického průměru počtu bodů prvních tří hráčů. Podobně pátý a každý další hráč získal počet bodů, který se rovná aritmetickému průměru počtu bodů všech hráčů, kteří házeli na koš před ním. Kolik bodů získal desátý hráč?

V: 9 bodů

8) Sportovci čtyř družstev jednoho oddílu byli testováni na fyzickou zdatnost. Každý obdržel známku od 1 (nejlepší) do 5 (nejhorší). Výsledky jsou uvedeny v následující tabulce:

	1	**2**	**3**	**4**	**5**
1. družstvo	4	3	13	5	2
2. družstvo	4	6	10	4	5
3. družstvo	5	9	3	3	3
4. družstvo	4	7	11	4	4

a) Jaká byla průměrná známka v celém oddílu? Počítejte s přesností na dvě desetinná místa.
b) Které družstvo bylo v průměru nejlepší a které nejhorší?
c) Určete četnosti jednotlivých známek v celém oddílu a sestrojte příslušný polygon četnosti.
d) Určete relativní četnosti (v %) jednotlivých známek v celém oddílu s přesností na dvě desetinná místa.

(Samostatná práce! Vypracovaný úkol odevzdejte na volném listu papíru.)

9) Určete medián čísel $x,\ 3,\ 4x-3,\ x+4,\ -16,\ 9,\ x-4$, jejichž aritmetický průměr je 4.

V: $\tilde{x} = 5$

10) Statistický soubor o rozsahu 1 000 statistických jednotek vyšetřujeme z hlediska jistého kvantitativního znaku x. Vzestupně uspořádané hodnoty znaku x pro jednotlivé statistické jednotky jsou $x_1, x_2, \cdots, x_{1000}$. Předpokládejme, že se hodnoty znaku x u všech jednotek vyšetřovaného souboru zvětší o pět (pětkrát). Jak se změní aritmetický průměr, medián a modus tohoto souboru?

V: vše se zvětší o pět (pětkrát)

11) Číslo n je z naměřených hodnot 3, n, 5, 11, 7, 8, 10, 11, 11 největší. Určete toto číslo n, jestliže medián tohoto souboru je roven aritmetickému průměru.

V: $n = 24$

12) V akciové společnosti je průměrný plat 13 500 Kč, přitom 30 % pracovníků s nejnižším platem má průměrně 9 000 Kč. Této skupině pracovníků byla počátkem roku zvýšena mzda jednotně o 500 Kč. O kolik procent vzrostl průměrný plat v celé společnosti následkem tohoto zvýšení nejnižších platů?

V: o 1,11 %

13) U sta studentů byla zjišťována výše kapesného. Určete chybějící údaje a, b v následující tabulce, jestliže bylo zjištěno, že průměrná výše kapesného studentů byla 873,50 Kč. Určete medián a modus v tomto souboru.

Výše kapesného (Kč)	800	a	1500	580	1210	1000	700
Počet studentů	30	12	8	10	b	20	15

V: a = 750 Kč, b = 5, $\tilde{x} = 800$, $\hat{x} = 800$

14) Vypočtěte průměrnou rychlost automobilu, který jede z místa A do místa B stálou rychlostí 80 km/h a zpět tutéž trasu rychlostí 120 km/h.

V: 96 km/h

15) Tři popelky přebírají hromadu hrachu. První Popelka by ho přebrala za 2 hodiny, druhá za 3 hodiny, třetí za 6 hodin. Za jak dlouho by hromadu přebrala „průměrná Popelka“?

V: 3 hodiny

16) Jeden stroj vyrobí výrobek za 20 sekund, druhý za 30 sekund, třetí za 6 sekund. Určete průměrnou rychlost výroby jednoho výrobku.

V: 12 sekund

17) Geometrický průměr pěti kladných čísel 2, 4, 8, x, x je 4. Vypočtěte jejich harmonický průměr.

V: $x = 4$, $\bar{x}_h = \frac{40}{11}$

18) Průmyslový podnik vykázal v posledních letech tyto hodnoty růstu produkce: 105,5 %, 100,6 %, 104,2 %, 101,2 %, v pátém roce pak pokles o 2,4 %. Jaké bylo jeho průměrné roční tempo růstu?

V: 101,78 %

19) V tabulce jsou uvedeny koeficienty růstu prodeje automobilů značky Škoda a značek dovozových zahraničních automobilů v ČR v letech 2002 – 2006. Určete, zda byl v uvedeném období vyšší přírůstek u značky Škoda nebo u značek zahraničních automobilů.

Rok	***2002***	***2003***	***2004***	***2005***	***2006***
Škoda	1,131	1,212	1,149	0,814	0,929
Zahraniční značky	1,416	1,563	1,042	0,820	1,141

V: $\bar{x}_G$ (Škoda) = 1,036 < $\bar{x}_G$ (zahr.) = 1,166

20) Ze „Školní statistické ročenky 2008“ vyjměte část tabulky **Výdaje na 1 žáka/studenta podle druhu/typu školy** týkající středního vzdělávání v uvedených letech (str.110). Tabulku zapište, sestrojte sloupkový diagram výdajů v těchto letech na gymnáziích, SOŠ a SOU a pak určete průměrné výdaje na jednoho žáka na gymnáziích, SOŠ a SOU v těchto letech.

Poznámka: „Školní statistickou ročenku 2008“ naleznete též na internetových stránkách Českého statistického úřadu (www.czso.cz) v sekci pro studenty.

●●●●●●

Jak již bylo vyloženo, každou charakteristiku polohy chápeme jako číslo, kolem něhož jednotlivé hodnoty znaku kolísají. Velikost tohoto kolísání vyjadřují **charakteristiky variability (proměnlivosti)** znaku. V souboru, ve kterém jako charakteristiku polohy zvolíme aritmetický průměr, je vhodnou charakteristikou variability **rozptyl a směrodatná a odchylka**.

Průměrná absolutní odchylka (ozn. $\bar{d}$) je aritmetický průměr absolutních hodnot odchylek jednotlivých hodnot znaku všech prvků daného souboru od jejich aritmetického průměru:

v neváženém tvaru: $\bar{d} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$ nebo *ve váženém tvaru*: $\bar{d} = \frac{1}{n}\sum_{j=1}^{r}n_j|x_j - \bar{x}|$

Rozptyl znaku x (ozn. s_x^2) je aritmetický průměr druhých mocnin odchylek hodnot znaku od aritmetického průměru:

v neváženém tvaru: $s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ nebo *ve váženém tvaru*: $s_x^2 = \frac{1}{n}\sum_{j=1}^{r}n_j(x_j - \bar{x})^2$

Jednotka (fyzikální jednotka) rozptylu je druhou mocninou jednotky hodnot znaku (např. jsou-li hodnoty znaku uvedeny v cm, pak jednotkou rozptylu je cm^2).

Směrodatná odchylka (ozn. s_x) je druhá odmocnina z rozptylu:

v neváženém tvaru: $s_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ nebo *ve váženém tvaru*: $s_x = \sqrt{\frac{1}{n}\sum_{j=1}^{r}n_j(x_j - \bar{x})^2}$

Jednotka (fyzikální jednotka) směrodatné odchylky je stejná jako jednotka hodnoty znaku.
Směrodatná odchylka ukazuje, jak se rozprostírají hodnoty v souboru dat. Přesněji, je to míra průměrné vzdálenosti hodnot dat od jejich průměru. Pokud budou všechny hodnoty dat stejné, pak směrodatná odchylka bude nulová.

Variační koeficient (ozn. v_x) je mírou relativní variability a užívá se při porovnávání variability různých souborů. Je definován jako podíl směrodatné odchylky a aritmetického průměru. Vyjadřuje se obvykle v procentech:

$$v_x = \frac{s_x}{\bar{x}} \cdot 100\%$$

Má smysl jen tehdy, nabývá-li znak x jen nezáporných hodnot.

Příklad 3.14. **V příkladu 3.13.** (str. 10) **jsme vyšetřovali průměrné týdenní kapesné 31 studentů. Vypočtěte průměrnou absolutní odchylku, rozptyl, směrodatnou odchylku a variační koeficient průměrného týdenního kapesného studentů.**

Řešení: Již jsme vypočetli průměr kapesného $\bar{x} = 98$ Kč.
Pro další potřebné výpočty zpracujeme tabulku:

Výše kapesného x_j	***50 Kč***	***100 Kč***	***200 Kč***	***500 Kč***
$\lvert x_j - \bar{x}\rvert$	48	2	102	402
Počet studentů n_j	15	12	3	1

Průměrnou absolutní odchylka: $\bar{d} = \frac{15 \cdot 48 + 12 \cdot 2 + 3 \cdot 102 + 1 \cdot 402}{31} \doteq 47$ Kč

Rozptyl: $s_x^2 = \frac{15 \cdot 48^2 + 12 \cdot 2^2 + 3 \cdot 102^2 + 1 \cdot 402^2}{31} \doteq 7336\,(\text{Kč})^2$

Směrodatná odchylka: $s_x = \sqrt{7336} \doteq 85{,}65$ Kč

Variační koeficient: $v_x = \frac{85{,}65}{98} \cdot 100 = 87{,}4\,\%$

Tato hodnota vypovídá o tom, že v daném souboru je velký rozptyl kolem střední hodnoty.

Cvičení 3 (2.část):

21) Ve které ze dvou pětičlenných skupin pracovníků (viz tabulka) se stejným průměrným platem je lépe uplatňována mzdová diferenciace (mzdy nejsou tzv. nivelizovány) na základě vypočteného variačního koeficientu:

	1.	*2.*	*3.*	*4.*	*5.*	*celkem*
Mzdy 1. skupiny	7 224	8 312	6 368	8 630	9 246	39 780
Mzdy 2. skupiny	8 468	4 950	9 464	10 472	6 426	39 780

V: $v_x(1) = 12{,}94\,\%,\ \ v_x(2) = 25{,}29\,\%$

22) Při opakovaném měření délky součástky mikrometrem byly naměřeny následující hodnoty v mm: 2,11; 2,01; 2,09; 2,11; 2,02; 2,03; 2,03; 2,10; 2,05; 2,05. Vypočtěte průměr, směrodatnou odchylku a variační koeficient (charakterizuje přesnost měření).

V: $\bar{x} = 2{,}06\ mm;\ \ s_x = 0{,}037\ mm;\ \ v_x = 1{,}8\,\%$

23) Skupina 20 studentů odpracovala o prázdninách v lese 3 160 hodin (viz tabulka). Vypočtěte aritmetický průměr, směrodatnou odchylku a variační koeficient:

Počet odpracovaných hodin	100	140	160	180	190
Počet studentů	3	4	4	5	4

V: $\bar{x} = 158\ h;\ \ s_x = 29{,}9\ h;\ \ v_x = 18{,}92\,\%$

24) Následující tabulka uvádí denní dojivost krav v litrech. Vypočtěte průměrnou dojivost, směrodatnou odchylku a variační koeficient:

Dojivost za 1 den	0-2	2-4	4-6	6-8	8-10	10-12
Počet krav	5	8	15	30	25	17

V: $\bar{x} = 7{,}26\ l;\ \ s_x = 2{,}71\ l;\ \ v_x = 37{,}33\ \%$

25) Určete v procentech průměrný obsah bromidu stříbrného ve fotografických roztocích (viz tabulka) a vypočtěte směrodatnou odchylku a variační koeficient:

Číslo roztoku	1	2	3	4	5	6	7	8
Obsah AgBr v %	38,25	40,42	40,35	38,62	37,10	40,55	37,23	39,84

V: $\bar{x} = 39{,}045\,\%;\ \ s_x = 1{,}34\,\%;\ \ v_x = 3{,}43\ \%$

26) Sir Weldon opakoval 4 096krát hod 12 kostkami a v každém hodu zaznamenal počet šestek. Rozdělení četností tohoto znaku udává následující tabulka. Určete aritmetický průměr, modus, medián, směrodatnou odchylku a variační koeficient:

Počet šestek	0	1	2	3	4	5	6	7 a více
Četnost	447	1145	1181	796	380	115	24	8

V: $\bar{x} \doteq 2{,}00;\ \ \hat{x} = 2;\ \ \tilde{x} = 2;\ \ s_x \doteq 1{,}29;\ \ v_x = 64{,}5\ \%$

27) Ze „Školní statistické ročenky 2008“ vyjměte část tabulky **Sňatky podle vzájemného věku snoubenců v roce 2006** týkající věku nevěsty (první tučně vytištěný řádek, str.47). Vypočtěte aritmetický průměr věku nevěsty a určete medián a modus tohoto souboru.

Poznámka: „Školní statistickou ročenku 2008“ naleznete též na internetových stránkách Českého statistického úřadu (www.czso.cz) v sekci pro studenty.

4. Statistická závislost znaků

Statistická jednotka je nositelkou nějakého statistického znaku. Často je třeba zkoumat případy, kdy je statistická jednotka nositelkou dvou znaků. Nás přitom zajímá, jaká je ***souvislost mezi těmito znaky*** (např. výnos a vodní srážky, prospěch žáka a jeho absence apod.)

Označme tyto znaky x a y, rozsah souboru jako obvykle n a vytvořme **prvotní tabulku:**

Statistická jednotka (číslo i)	*Hodnota znaku x*	*Hodnota znaku y*
1	x_1	y_1
2	x_2	y_2
$\vdots$	$\vdots$	$\vdots$
n	x_n	y_n

Vzájemná souvislost mezi znaky x a y je vyjádřena tzv. **koeficientem korelace *r*** :

$$r = \frac{k_{xy}}{s_x \cdot s_y},$$

kde k_{xy} je tzv. **kovariance** : $k_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$,

s_x, s_y jsou směrodatné odchylky statistických znaků x a y.

Vlastnosti koeficientu korelace *r*:

- $r \in \langle -1, 1 \rangle$
- Je-li $r > 0$, pak hodnota znaku y roste s rostoucím znakem x,
  je-li $r < 0$, pak hodnota znaku y klesá s rostoucím znakem x.

Podle velikosti koeficientu korelace r hovoříme o různých **stupních vazby mezi znaky *x* a *y***:
Jestliže

- $0 \le |r| < 0{,}3$, pak se jedná o **nulový** stupeň vazby,
- $0{,}3 \le |r| < 0{,}5$, pak se jedná o **mírný** stupeň vazby,
- $0{,}5 \le |r| < 0{,}7$, pak se jedná o **význačný** stupeň vazby,
- $0{,}7 \le |r| < 0{,}9$, pak se jedná o **vysoký** stupeň vazby,
- $0{,}9 \le |r| \le 1$, pak se jedná o **těsný** stupeň vazby.

Poznámka: Je-li mezi znaky lineární závislost, je r = ± 1.

Závislost mezi znaky lze často vystihnout **lineární závislostí** $y = ax + b$, jejímž grafem je tzv. **regresní přímka**.
Její rovnice je

$$y = \frac{k_{xy}}{s_x^2} \cdot (x - \bar{x}) + \bar{y}$$

Příklad 4.1. **Na 10 pokusných polích byl pozorován hektarový výnos a množství srážek (viz tabulka). Jaká je vzájemná vazba množství srážek a hektarových výnosů? Určete rovnici regresní přímky a sestrojte graf.**

Pole (i)	*Závlaha (mm)* x_i	*Výnos (q/1ha)* y_i
1	150	33
2	200	42
3	125	30
4	150	35
5	185	45
6	145	30
7	160	38
8	170	40
9	210	45
10	190	42

Řešení: **průměry:** $\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 168{,}5$; $\bar{y} = \frac{1}{10}\sum_{i=1}^{10} y_i = 38$

směrodatné odchylky:

$$s_x = \sqrt{\frac{1}{10}\sum_{i=1}^{10}(x_i - 168{,}5)^2} = \sqrt{665{,}25} \quad ; \quad s_y = \sqrt{\frac{1}{10}\sum_{i=1}^{10}(y_i - 38)^2} = \sqrt{29{,}6}$$

kovariance: $k_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \Rightarrow k_{xy} = \frac{1}{10}\sum_{i=1}^{10}(x_i - 168{,}5)(y_i - 38) = 130{,}5$

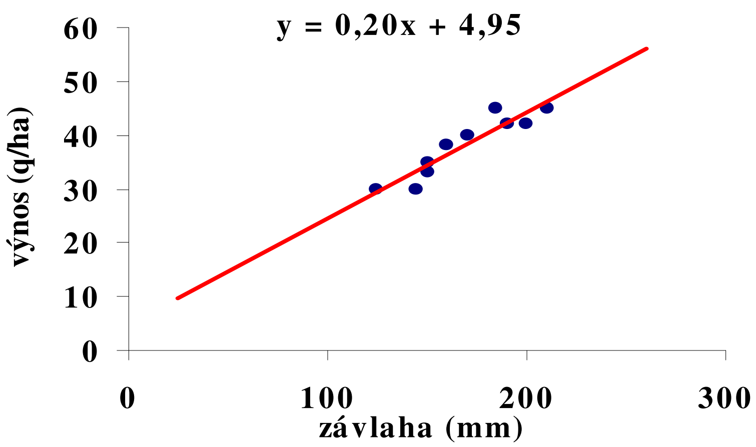
koeficient korelace: $r = \frac{k_{xy}}{s_x \cdot s_y} \Rightarrow r = \frac{130{,}5}{\sqrt{665{,}25}\cdot\sqrt{29{,}6}} \doteq 0{,}93$

regresní přímka: $y = \frac{k_{xy}}{s_x^2}\cdot(x - \bar{x}) + \bar{y} \Rightarrow y = \frac{130{,}5}{665{,}25}\cdot(x - 168{,}5) + 38$

$$y \doteq 0{,}20x + 4{,}95$$

Závěr: Mezi srážkami a výnosy je těsný stupeň vazby.

Graf:



Příklad 4.2. **V následující tabulce jsou uvedeny známky z matematiky na konci 1. a 2. ročníku. Na základě hodnoty koeficientu korelace určete stupeň vazby mezi těmito známkami. Určete regresní přímku a sestrojte graf.**

Známka na konci 1. ročníku	*Známka na konci 2.ročníku*			
	1	2	3	4
1	3	1		
2	2	6	5	
3		7	8	
4			2	1

Řešení: Tabulku nutno přepsat do tvaru uvedeného na začátku této kapitoly, abychom si ujasnili hodnoty x_i a y_i:

1. ročník (x_i)	*2. ročník* (y_i)	*Četnost* (x_i, y_i)
1	1	3
1	2	1
2	1	2
2	2	6
2	3	5
3	2	7
3	3	8
4	3	2
4	4	1
Celkem žáků		***35***

průměry: $\bar{x} = \frac{1}{35}\sum_{i=1}^{35} x_i \doteq 2{,}49$; $\bar{y} = \frac{1}{35}\sum_{i=1}^{10} y_i \doteq 2{,}34$

směrodatné odchylky:

$$s_x = \sqrt{\frac{1}{35}\sum_{i=1}^{35}(x_i - 2{,}49)^2} = \sqrt{0{,}6498} \quad ; \quad s_y = \sqrt{\frac{1}{35}\sum_{i=1}^{35}(y_i - 2{,}34)^2} = \sqrt{0{,}5682}$$

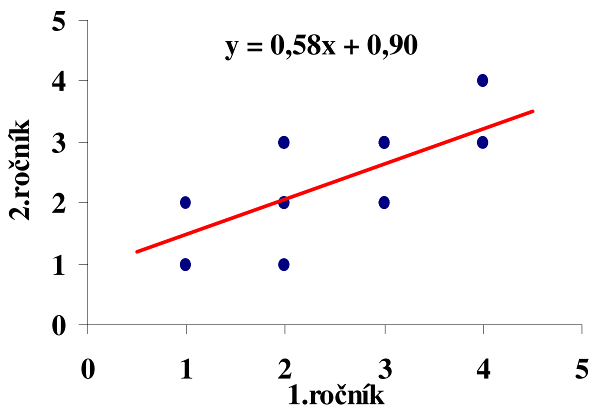
kovariance: $k_{xy} = \frac{1}{35}\sum_{i=1}^{35}(x_i - 2{,}49)(y_i - 2{,}34) = 0{,}3763$

koeficient korelace: $r = \frac{0{,}3763}{\sqrt{0{,}6498}\cdot\sqrt{0{,}5682}} \doteq 0{,}619$

regresní přímka: $y = 0{,}58x + 0{,}90$

Závěr: Mezi známkami z 1. a 2. ročníku je význačný stupeň vazby.

Graf:



Cvičení 4:

1) Vypočítejte koeficient korelace mezi věkem matky a pořadím narozeného dítěte (Návod: Uvažujte vždy střed intervalu věku matky. Tabulku přepište podobně jako v příkladě 4.2.)

Věk matky	***Pořadí narozených***			
	1	2	3	4
20-24	66 649	50 606	9 372	1 649
25-29	18 336	40 105	17 879	4 028
30-34	3 799	11 426	9 633	3 404

V: $r = 0{,}391$ (mírný stupeň vazby)

2) Ze „Školní statistické ročenky 2008“ vyjměte část tabulky **Jednotlivci v ČR využívající internet k nakupování** týkající se věkových skupin v roce 2007 (str.124). Vypočtěte koeficient korelace a určete typ vazby mezi věkem a užíváním internetu k této činnosti. Určete regresní přímku a sestrojte graf.
Poznámka: „Školní statistickou ročenku 2008“ naleznete též na internetových stránkách Českého statistického úřadu (www.czso.cz) v sekci pro studenty.

3) V následující prvotní tabulce jsou uvedeny známky 12 studentů z matematiky (ozn. x) a fyziky (ozn. y). Srovnejte variabilitu známek z matematiky a fyziky. Vypočtěte koeficient korelace mezi těmito známkami a zjistěte, jaká je vzájemná vazba mezi známkami z matematiky a fyziky. Určete regresní přímku závislosti známky z fyziky (y) na známce z matematiky (x) a sestrojte graf.

Žák č. i	***Známka z M x_i***	***Známka z F y_i***	***Žák č. i***	***Známka z M x_i***	***Známka z F y_i***
1	3	2	7	2	2
2	3	2	8	3	1
3	1	2	9	2	2
4	4	4	10	1	1
5	1	1	11	2	1
6	2	1	12	2	2

V: $v_M = 41{,}42\,\%$, $v_F = 47{,}37\,\%$, $r = 0{,}595$, $y = 0{,}5x + 0{,}7$

4) Dopravní podnik zjišťuje závislost délky dráhy ujeté vagónem po generální opravě a náklady na údržbu. Na základě uvedených údajů vyjádřete závislost nákladů na délce ujeté dráhy.

Vagón č. i	***Délka dráhy (km) x_i***	***Náklady (hal/km) y_i***
1	30 000	30
2	60 000	40
3	70 000	42
4	80 000	47
5	90 000	50
6	95 000	52

(Samostatná práce! Vypracovaný úkol odevzdejte na volném listu papíru.)

●●●●●●