

Popisné statistiky kategoriálních dat.

Četnosti



**SLEZSKÁ
UNIVERZITA**

FAKULTA VEŘEJNÝCH
POLITIK V OPAVĚ

doc. Ing. Petr Sed'a, Ph.D.

Co se dnes dozvíte?

- Jednostupňové třídění, rozdělení četností,
- Tabulky a grafy rozdělení četností,
- Třídění metrických znaků, intervalové rozdělení četností.

Jednostupňové třídění:

tabulka rozdělení četností

- řádky tabulky představují jednotlivé třídy (kategorie)
- sloupce tabulky vyjadřují četnosti (počty jednotek)

Třídící
znak

ZNÁMKA	POČET MATEMATIKA	POČET STATISTIKA
1	12	18
2	35	37
3	61	54
4	18	17
CELKEM	126	126

35 studentů dostalo z matematiky dvojku

Jak správně roztrždit data?

třidy mohou být:

- jednotlivé obměny třídícího znaku
- množiny hodnot třídícího znaku (například intervaly)

Každá hodnota třídícího znaku musí náležet **právě jedné** třídě

Příklad: Všechny muže ve městě rozdělíme na 2 skupiny:

- a) muži, kteří chodí k holiči
- b) muži, kteří se holí sami

Absolutní a relativní četnosti:

četnosti – vyjadřují zastoupení jednotlivých tříd

četnost (absolutní četnost) n_i
počet prvků třídy i

35 studentů dostalo dvojku

relativní četnost p_i
vyjadřuje část z celku (obvykle v %)

28% studentů dostalo dvojku

$$p_i = \frac{n_i}{n}$$

četnost

velikost souboru

Kumulativní četnosti:

ordinálních a metrických znaků

lze je uspořádat

kumulativní četnost kn_i

počet hodnot menších nebo rovných dané třídě

$$kn_i = n_1 + n_2 + \dots + n_i = \sum_{k=1}^i n_k$$

47 studentů dostalo **nejhůř** dvojku

relativní kumulativní četnost kp_i

$$kp_i = \frac{kn_i}{n}$$

37% studentů dostalo **nejhůř** dvojku

Obecná tabulka četností:

NÁZEV (HODNOTA) TŘÍDY	ABSOLUTNÍ ČETNOSTI	RELATIVNÍ ČETNOSTI	KUMULATIVNÍ ABSOLUTNÍ ČETNOSTI	KUMULATIVNÍ RELATIVNÍ ČETNOSTI
třída 1	n_1	p_1	$kn_1 = n_1$	$kp_1 = p_1$
třída 2	n_2	p_2	$kn_2 = n_1 + n_2$	$kp_2 = p_1 + p_2$
...
třída m	n_m	p_m	$kn_m = n$	$kp_m = 1$
CELKEM	n	1	x	x

Platí vztahy:

$$kn_i = kn_{i-1} + n_i \quad kp_i = kp_{i-1} + p_i$$

pro $i = 2, 3, \dots, m$

Příklad 1 - Tabulka četností

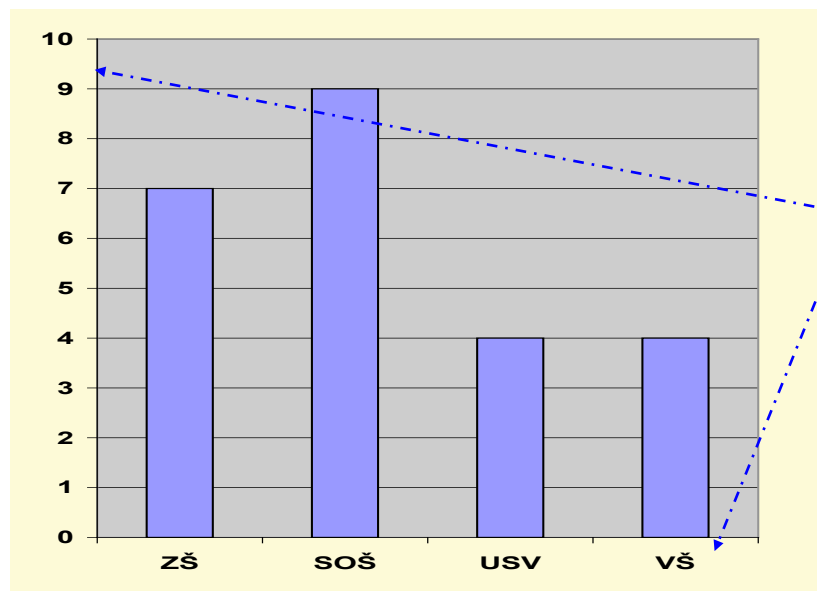


ZNÁMKA Z MATEMATIKY	ČETNOSTI		KUMULATIVNÍ ČETNOSTI	
	n_i	p_i	kn_i	kp_i
1	12	0,095	12	0,095
2	35	0,278	47	0,373
3	61	0,484	108	0,857
4	18	0,143	126	1,000
CELKEM	126	1,000	x	x

85,7% studentů udělalo zkoušku

Histogram rozdělení četnosti:

sloupcový graf četností (histogram)
vhodný pro nominální a ordinální proměnné



osa x – jednotlivé kategorie

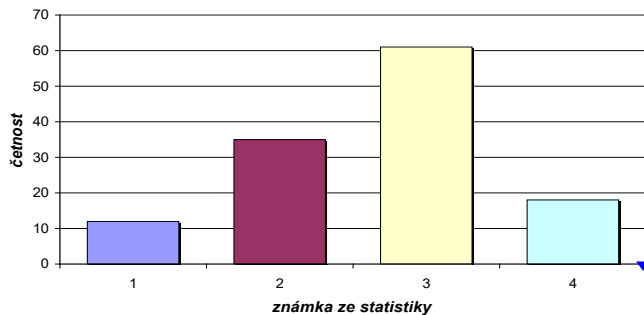
seřazené abecedně (nominální
znaky) nebo podle uspořádání

osa y – vynášené četnosti

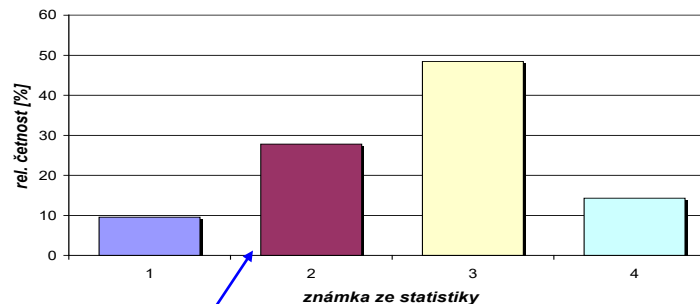
- absolutní četnosti
- relativní četnosti
- kumulativní četnosti

Histogram četností – příklad:

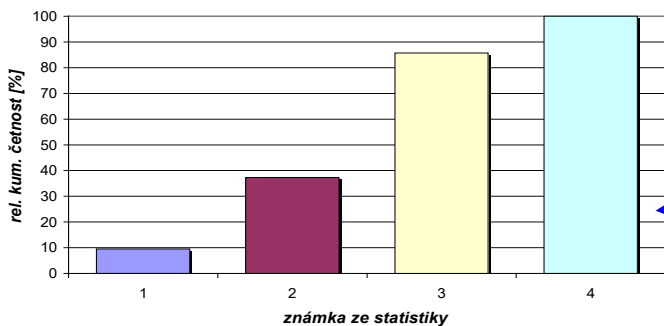
Histogram absolutních četností



Histogram relativních četností



Histogram relativních kumulativních četností



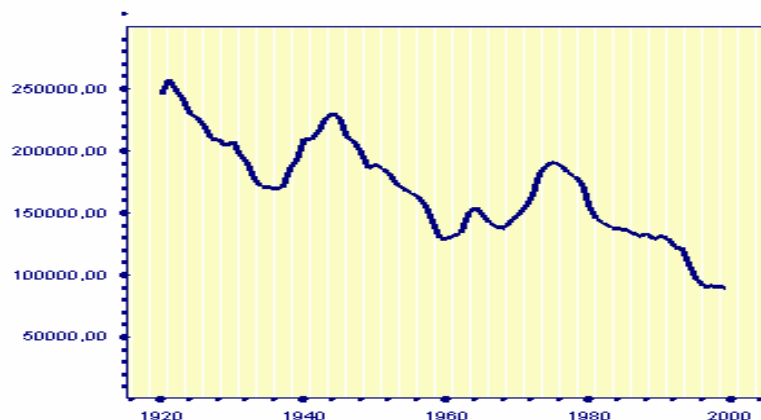
- absolutní a relativní četnosti mají stejný průběh, liší se pouze v ose y
- kumulativní četnosti mají vždy neklesající průběh

Polygon rozdění četností:

spojnicový graf četností (polygon)

vhodné pro ordinální a metrické proměnné

**Počet živě narozených dětí v České republice
1920 - 1999**



zvláštní případ:

graf časové řady

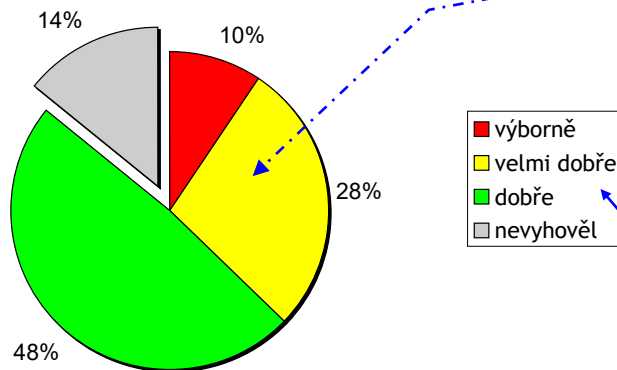
kategoriální proměnná
je čas

Výsečový graf (pie chart):

výsečový graf četností

vhodné zejména pro nominální proměnné s malým počtem obměň

Výsečový graf relativních četností



plochy výsečí a středové
úhly úměrné četnostem

popisky obvykle v %

*důležité obměny lze
zvýraznit vysunutím*

legenda grafu

Intervalové rozdělení četností:

vhodné pro spojité metrické znaky (s velkým počtem obměn)

rozdělení oboru hodnot třídícího znaku na jednotlivé intervaly

VĚK	četnosti		kumul. četnosti	
	abs.	rel.	abs.	rel.
26 - 30	6	25,0%	6	25,0%
31 - 35	5	20,8%	11	45,8%
36 - 40	3	12,5%	14	58,3%
41 - 45	5	20,8%	19	79,2%
46 - 50	2	8,3%	21	87,5%
51 - 55	2	8,3%	23	95,8%
56 - 60	1	4,2%	24	100,0%
Celkem	24	100,0%	x	x

intervals

Jak vytvořit intervaly?

- počet tříd (intervalů) v rozmezí 5 až 20
 - malý počet tříd – malá informační hodnota*
 - velký počet tříd – nepřehledná tabulka*
- hranice intervalů dobře zapamatovatelná čísla
 - dělitelná 5, 10, 20, ...*
- intervaly jednoznačně pokrývají celý obor hodnot
 - hraniční body intervalů patří pouze jednomu z nich*
- intervaly stejně široké
 - srovnatelnost intervalů mezi sebou*
- oba krajní intervaly mají nenulové četnosti

Jak vytvořit intervaly?

Vždy musí být splněny následující podmínky:

1. Všechny hodnoty znaku zařazené do intervalu musí být bez větší chyby zastupitelné středem intervalu x_j , přičemž $x_j = \frac{x_{j \max} - x_{j \min}}{2}$, který reprezentuje všechny hodnoty v jedné třídě se nazývá **třídní znak**. Hranice a středy tříd mají být zaokrouhlená čísla.
2. Interval je dostatečně velký z důvodu přehlednosti. Je vhodné zvolit interval stejné délky, tzv. **ekvidistantní**.
3. Každý prvek souboru je zařaditelný do právě jedné třídy. Třídy jsou tedy **disjunktní**. Pro splnění této podmínky je nutné přesně vymezit hranice intervalů.

Odhad počtu tříd:

Sturgesovo pravidlo pro odhad počtu tříd k :

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



n	k
10	4
20	5
50	7
100	8
500	10
1000	11

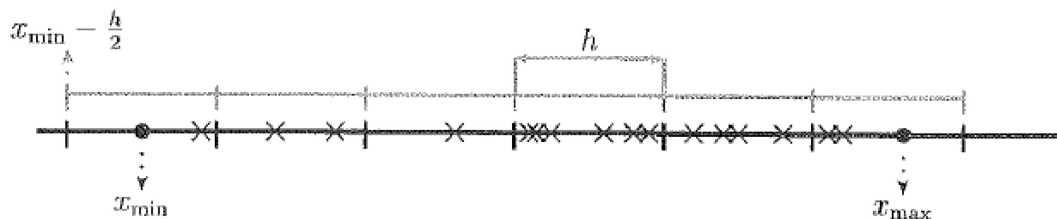
odhad šířky intervalu h :

$$h \approx \frac{x_{\max} - x_{\min}}{k}$$

vypočtené hodnoty jsou pouze doporučením
skutečné hodnoty → přehlednost tabulky
(zaokrouhlené hranice a šířky tříd)

Postup při vytvoření třídního rozdělení četností:

1. Určení variačního intervalu $\langle x_{min}, x_{max} \rangle$.
2. Volba počtu tříd k .
3. Určení šířky třídy h .
4. Volba dolní meze první třídy, obvykle: $x_{min} - \frac{h}{2}$.
5. Určení tříd.
6. Určení třídních znaků x_j .
7. Určení třídních četností n_j .

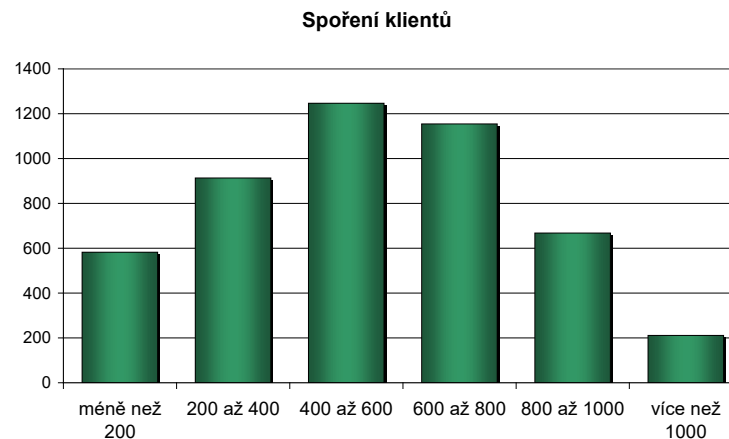


Příklad – intervalové rozdělení:

rozdělení pravidelných měsíčních úspor klientů spořitelny:

MĚSÍČNÍ SPOŘENÍ [KČ]	POČET KLIENTŮ
méně než 200	582
200 až 400	913
400 až 600	1247
600 až 800	1154
800 až 1000	668
více než 1000	211

tabulka četností



histogram

Příklad 1 – školní potřeby



U 18 studentů byly zjištěny výdaje na školní potřeby v měsíci září 2023. Sestavte **tabulku třídního rozdělení četností**. Zjištěné výdaje v Kč: 1510, 1580, 1600, 1620, 1620, 1700, 1730, 1750, 1810, 1860, 1900, 1930, 1950, 2050, 2100, 2200, 2380, 2460.

Variační interval $\langle x_{min}, x_{max} \rangle = \langle 1510, 2460 \rangle$

Počet tříd dle **Sturgesova pravidla**: $k \approx 1 + 3,3 \cdot \log n = 1 + 3,3 \cdot \log 18 \doteq 5$

Šířka třídy $h = \frac{950}{5} = 190 \doteq 200$

Sestavíme odpovídající tabulku četností. Hodnota třídícího znaku x_i , je nahrazena hodnotou středu intervalu. Při praktických výpočtech i tyto hodnoty **vhodně zaokrouhlíme**.

Příklad 1 – školní potřeby



U 18 studentů byly zjištěny výdaje na školní potřeby v měsíci září 2023. Sestavte tabulku třídního rozdělení četností. Zjištěné výdaje v Kč: 1510, 1580, 1600, 1620, 1620, 1700, 1730, 1750, 1810, 1860, 1900, 1930, 1950, 2050, 2100, 2200, 2380, 2460.

Třída v Kč	Třídní znak x_j	Absolutní čet. n_j	Relativní čet. p_j	Abs. kumul. čet. kn_j	Rel. kumul. čet. kp_j
1500 – 1700	1600	6	0,33	6	0,33
1701 – 1900	1800	5	0,28	11	0,61
1901 – 2100	2000	4	0,22	15	0,83
2101 – 2300	2200	1	0,06	16	0,89
2301 - 2500	2400	2	0,11	18	1,00
Celkem	x	18	1	x	x

V tomto případě jsme nemohli zvolit dolní mez první třídy rovnu $x_{min} - \frac{h}{2} = 1510 - 100 = 1410$, protože bychom pěti třídami o šířce 200 nepokryli interval.

Příklad 2 – benzínová pumpa



V čase mezi 0:00 hod a 06:00 hod. jistého dne bylo u benzínové čerpací stanice obslouženo 54 vozidel. V čase od 06:00 do 10:00 to bylo 96 vozidel. Mezi časy 10:00 a 15:00 bylo v průměru obslouženo 36 vozidel za hodinu. Další měření bylo prováděno v čase od 15:00 do 20:00, kdy bylo obslouženo celkem 20% ze všech automobilů toho dne. Konečně v čase mezi 20:00 a 24:00 bylo u čerpací stanice obslouženo 50 vozidel.

- a) Z údajů, které máte k dispozici, sestavte přehlednou tabulku četností.
- b) Doplněte tabulku ji o průměrné počty obsloužených vozidel a vypočtete průměrný počet vozidel obsloužených za jednu hodinu v každém měřeném úseku a celkem.
- c) Určete průměrný počet obsloužených vozidel za hodinu přes den (6:00 – 20:00) a v nočních hodinách.

Příklad 2 – benzínová pumpa



a) Tabulka četností:

ČASOVÝ ÚSEK	četnosti		kumulativní četnosti	
	absolutní	relativní	absolutní	relativní
0:00 – 6:00	54	11,4%	54	11,4%
6:00 – 10:00	96	20,2%	150	31,6%
10:00 – 15:00	180	37,9%	330	69,5%
15:00 – 20:00	95	20,0%	425	89,5%
20:00 – 24:00	50	10,5%	475	100,0%
CELKEM	475	100,0%	x	x

Příklad 2 – benzínová pumpa



b) Rozšíření tabulky o průměry:

ČASOVÝ ÚSEK	Délka úseku	Průměrný počet
0:00 – 6:00	6	9,0
6:00 – 10:00	4	24,0
10:00 – 15:00	5	36,0
15:00 – 20:00	5	19,0
20:00 – 24:00	4	12,5
CELKEM	24	19,8

c) Přes den se u benzínové pumpy obslouží v průměru 26,5 vozidel za hodinu, přes noc 10,4 vozidel za hodinu. Počítáme jako vážené aritmetické průměry.

1. Ramík J. a Čemerková Š. *Statistika A*. Opava, Karviná: SLU, 2000. **(kapitola 2)**.





Děkuji za pozornost.