1.2	Triedenie podľa	kvantitatívnych	znakov
-----	-----------------	-----------------	--------

### POJMY

✓	Kvantitatívny štatistický znak	✓	Intervalové (skupinové) rozdelenie
$\checkmark$	Diskrétny znak		početnosti
$\checkmark$	Spojitý znak	✓	Počet tried (intervalov)
$\checkmark$	Štatistické triedenie	$\checkmark$	Rozpätie intervalu
$\checkmark$	Triediaci znak	$\checkmark$	Frekvenčná tabuľka
$\checkmark$	Trieda	$\checkmark$	Triedna početnosť
$\checkmark$	Triedna početnosť	$\checkmark$	Absolútne početnosti
$\checkmark$	Zásada jednoznačnosti	$\checkmark$	Relatívne početnosti
$\checkmark$	Zásada úplnosti	✓	Kumulatívne početnosti – absolútne
$\checkmark$	Rozdelenie početnosti	✓	Kumulatívne početnosti – relatívne

### VZORCE

relatívna početnosť

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

kumulatívna absolútna početnosť

 $N_1 = n_1 \qquad N_i = N_{i-1} + n_i$ 

počet tried

 $m = \sqrt{n}$ 

rozpätie intervalu

 $h = \frac{x_{max} - x_{min}}{m}$ 

# FUNKCIE V EXCELI

=COUNT(hodnota1,hodnota2,...) =SQRT(číslo) =MAX(číslo1, číslo2,...) =MIN(číslo1, číslo2,...)

#### RIEŠENÝ PRÍKLAD

V praktickej ukážke budeme prezentovať dva typy triedenia kvantitatívnych údajov a to:

- rad rozdelenia početnosti
- intervalové (skupinové) rozdelenie početnosti

#### Zadanie2:

Roztrieď te poľnohospodárske podniky podľa počtu stredísk, na ktoré sa členia podniky z územného hľadiska.

### Riešenie:

Štatistický znak počet stredísk predstavuje diskrétny kvantitatívny znak s malým počtom obmien, preto pri triedení roztriedime podniky do radu rozdelenia početnosti. V prvom kroku zistíme obmeny štatistického znaku pomocou funkcií **MAX** a **MIN**. Zistili sme, že minimálna hodnota v súbore je 0, t.j. podnik predstavuje z územného hľadiska jeden kompaktný celok (nie je členený na strediská). Maximálna hodnota je 4, čiže podniky majú najviac štyri strediská. Na základe týchto hodnôt v ďalšom kroku určíme triedy (vypíšeme hodnoty od 0 po 4). Vlastné triedenie budeme realizovať prostredníctvom Excelu. Jedná sa o kvantitatívne triedenie, t.j. využijeme analytické nástroje v Exceli. Postup je nasledovný:

- 1. Výber z menu Tools/Data Analysis...<sup>1</sup>
- 2. Z ponuky vyberieme Histogram, ktorý slúži na triedenie kvantitatívnych údajov.
- 3. Vyplníme vstupné okno.
  - *Input Range* predstavuje vstupnú oblasť, t.j. vysvietime údaje, ktoré sa majú triediť. V našom prípade vysvietime štatistický znak počet stredísk.
  - Do *Bin Range* vkladáme informácie, na základe ktorých sa má uskutočniť triedenie, t.j. vysvietime triedy, ktoré sme si predtým vypísali. Vysvecujeme o jednu triedu menej (hodnoty od 0 po 3), pretože Excel poslednú triedu robí automaticky.
  - Ak chceme, aby Excel vypočítal aj kumulatívne relatívne početnosti a výsledky triedenia zobrazil graficky, aktivujeme si voľby *Cumulative Percentage* a *Chart Output*.
  - Označíme výstupnú oblasť. Stačí označiť len jednu bunku, ktorá predstavuje ľavý horný roh výstupnej oblasti. Ak chceme mať výstup vedľa

<sup>&</sup>lt;sup>1</sup> Ak sa **Data Analysis** ... v ponuke nenachádza, je potrené z **Tools** vybrať voľbu **Add-Ins...** a aktivovať **Analysis ToolPak**, resp. **Analysis ToolPak VBA**. Po označení a potvrdení daných volieb sa v zozname **Tools** objaví aj ponuka **Data Analysis...** V prípade, že sa tak nestane, je potrebné skontrolovať, či je Excel nainštalovaný v kompletnej verzii.

vypísaných tried, ako výstupnú oblasť označíme bunku vedľa bunky s textom TRIEDA.

Nasledujúci obrázok znázorňuje ako má vyzerať vyplnené vstupné okno.

Histogram		×
Input Input Range: Bin Range: Labels	\$I\$2:\$I\$130 🗾 \$K\$7:\$K\$10 🔜	OK Cancel Help
Output options Qutput Range: New Worksheet Ply: New Workbook Pareto (sorted histogram) Cumulative Percentage Chart Output	\$L\$6	

Po potvrdení cez OK dostávame nasledovný výstup.

n x max	129	) 			
	xi	ין ni	Fi	fi	Ni
TRIEDA	Bin	Frequency	Cumulative %		
0	0	) 10	7,75%	7,75%	10
1	1	15	19,38%	11,63%	25
2	2	25	38,76%	19,38%	50
3	3	3 57	82,95%	44,19%	107
4	More	22	100,00%	17,05%	129
		129		100,00%	



Poznámka: Časť vyplnená zelenou farbou a graf predstavuje výstup Excelu. Relatívne a kumulatívne početnosti boli dopočítané podľa príslušného vzorca.

#### Interpretácia výsledkov:

Vo výstupnej tabuľke Excelu sa nachádzajú hodnoty: **Bin** (triedy<sup>2</sup>), **Frequency** (absolútne početnosti) a **Cumulative %** (kumulatívne relatívne početnosti). Jednotlivé triedy je vhodné charakterizovať aj prostredníctvom relatívnych a absolútnych početností, ktoré sú dopočítané v posledných dvoch stĺpcoch tabuľky.

Súčasťou výstupu je aj grafické zobrazenie. V grafe sa nachádzajú dva typy grafov a to histogram z absolútnych početností a kumulatívny ogiv<sup>3</sup> z kumulatívnych relatívnych početností.

Z výsledkov frekvenčnej tabuľky vyplýva (hodnoty n<sub>i</sub>), že bez stredísk je 10 podnikov. Naopak, najviac stredísk (štyri) má 22 poľnohospodárskych podnikov. V súbore sa najčastejšie vyskytuje hodnota 3, čo znamená, že najčastejšie majú analyzované podniky tri strediská, čo predstavuje 44,19%. Kumulatívna absolútna početnosť pri tejto triede je 107, t.j., 107 podnikov má do troch stredísk (vrátane), čo je 82,95%. Obdobným spôsobom by sme mohli interpretovať ostatné hodnoty vo výstupe.

## Zadanie3:

Roztrieď te poľnohospodárske podniky podľa priemerného mesačného zárobku<sup>4</sup>.

#### Riešenie:

Keďže priemerný mesačný zárobok predstavuje spojitý znak, pri triedení budeme vychádzať z intervalového rozdelenia početnosti.

Skôr ako pristúpime k triedenie je potrebné najskôr vypočítať:

- počet tried (intervalov), do ktorých budeme triediť
- rozpätie intervalu

V Exceli budeme postupovať nasledovne:

• do jedného stĺpca si vypíšeme označenia hodnôt, ktoré potrebujeme vypočítať

<sup>&</sup>lt;sup>2</sup> Posledná hodnota v stĺpci je označená ako More, ktorá v prípade radu rozdelenia početnosti reprezentuje len jedno číslo. V našom prípade More=4.

<sup>&</sup>lt;sup>3</sup> Ogivná krivka predstavuje neklesajúcu krivku, na základe priebehu ktorej vidíme ako sa menia početnosti v jednotlivých triedach. Čím je priebeh strmší, tým je nárast medzi triedami väčší.

<sup>&</sup>lt;sup>4</sup> Priemerný mesačný zárobok je vypočítaná hodnota v rámci jedného roku. Predstavuje priemer mesačných zárobkov.

 do vedľajšieho stĺpca pomocou príslušných funkcií a vzorcov počítame potrebné hodnoty, tak ako je to uvedené v nasledujúcej tabuľke:

n	129	rozsah štatistického súboru, je možné určiť cez funkciu COUNT
		počet tried vypočítame cez funkciu SQRT(n),
m	11,36	pri ďalšom výpočte použijeme zaokrúhlenú hodnotu <b>11</b>
x max	25 166,66	maximum určené cez funkciu MAX
x min	3 333,33	minimum určené cez funkciu MIN
		rozpätie intervalu vypočítame pomocou vzorca h = (max-min)/m
		hodnota je pre nás len orientačná, pri ďalších výpočtoch budeme
h	1 984,85	vychádzať zo zaokrúhlenej hodnoty 2 000.

- po určení hodnôt m a h je možné prejsť k zostrojeniu intervalov. Aby boli dodržané zásady triedenia, zostrojíme prvý a posledný interval zľava, resp. sprava otvorený.
- problematické môže byť <u>určenie prvého intervalu</u>, pretože nie je presne stanové pravidlo ako vypočítať hornú hranicu prvého intervalu. Orientačne vychádzame z minimálnej hodnoty, pretože táto hodnoty by sa v tomto intervale mala nachádzať. Všeobecne platí, že nemá význam zostrojiť taký prvý interval, v ktorom bude 0 hodnôt, ale ani taký interval, v ktorom by bolo príliš veľa hodnôt vzhľadom na ďalšie intervaly (napr. 6,2,0,1, atď.). Znamená to, že ak je v prvom intervale nulová početnosť, je potrebné jeho hornú hranicu zvýšiť a naopak, ak je v prvom intervale vysoká početnosť vzhľadom na nasledujúce, je potrebné hornú hranicu znížiť. V našom prípade je stanovená ako vhodná hranica hodnota 4 200.
- Ak máme určený prvý interval, ostané intervaly sa dopočítajú tak, aby počet intervalov sa rovnal 11 a rozpätie každého intervalu (interval 2-10, pretože 1. a 11. interval sú otvorené) bolo 2 000.

Intervaly je potrebné vypísať do samostatných buniek v Exceli, osobitne dolné hranice (DH) a horné hranice (HH). Prvý a posledný interval je otvorený, preto na príslušnom mieste v Exceli necháme prázdne bunky (viď obr. ....).

- Až teraz môžeme pristúpiť k triedeniu. Triedenie robíme podobne ako v predchádzajúcom prípade prostredníctvom voľby **Tools/DataAnalysis/Histogram**.
- Vyplníme vstupné okno ako je to zobrazené na nasledujúcom obrázku. Oblasti, ktoré je potrebné určiť, sú obdobné ako v predchádzajúcom príklade s tým, že ako oblasť *Bin Range* vysvietime len prvých desať horných hraníc stanovených intervalov, pretože posledný interval robí Excel automaticky.

listogram			? ×
Input			
Input Range:	\$C\$3:\$C\$131	<u>.</u>	UK
<u>B</u> in Range:	\$G\$9:\$G\$18		Cancel
🗖 Labels			Help
-Output options	\$H\$8	<u></u>	
C New Worksheet Ply:			
C New <u>W</u> orkbook			
Pareto (sorted histogram)			
Cumulative Percentage			
🗹 Chart Output			

• Po potvrdení cez **OK** dostávame nasledovný výstup:

n	129						
m	11,36	11					
x max	25 166,66						
x min	3 333,33						
h	1 984,85	2 000					
			_	ni	Fi	fi	Ni
Číslo int.	DH	HH	Bin	Frequency	Cumulative %		
1		4200	4200	2	1,55%	1,55%	2
2	4200	6200	6200	4	4,65%	3,10%	б
3	6200	8200	8200	9	11,63%	6,98%	15
4	8200	10200	10200	22	28,68%	17,05%	37
5	10200	12200	12200	29	51,16%	22,48%	66
6	12200	14200	14200	35	78,29%	27,13%	101
7	14200	16200	16200	16	90,70%	12,40%	117
8	16200	18200	18200	5	94,57%	3,88%	122
9	18200	20200	20200	3	96,90%	2,33%	125
10	20200	22200	22200	2	98,45%	1,55%	127
11	22200		More	2	100,00%	1,55%	129
				129		100%	



 Výstup voľby Histogram je tvorený vyfarbenými bunkami a grafom<sup>5</sup>. Posledné dva stĺpce (relatívne početnosti a kumulatívne absolútne početnosti) sú manuálne dopočítané.

#### Interpretácia výsledkov:

Z výsledkov triedenia vyplýva, že najpočetnejšiu triedu predstavuje interval č. 6, z ktorého môžeme vyčítať, že v 35 podnikoch mali zamestnanci priemerný mesačný zárobok od 12 200 Sk po 14 200 Sk, čo predstavuje 27,13%. Kumulatívna absolútna početnosť v tomto intervale je 101, čo znamená, že v 101 poľnohospodárskych podnikoch bol priemerný mesačný zárobok do 14 200 Sk, čo predstavuje 78,29%. Zastúpenie v ostatných intervaloch je nižšie, pričom najmenej, do 4 200 Sk bol priemerný mesačný zárobok v 2 podnikoch a viac ako 22 200 Sk mali zamestnanci mesačný zárobok takisto v 2 podnikoch.

<sup>&</sup>lt;sup>5</sup> Štandardný graf bol upravený, pretože sa jedná o spojitý znak, t.j. stĺpce histogramu sa musia nachádzať vedľa seba (bola zrušený medzera medzi jednotlivými stĺpcami).