

## Téma 1. Příklady

1. U 44 studentů desátého patra vysokoškolské koleje byla zjišťována vzdálenost místa bydliště do Českých Budějovic. Roztříděné údaje jsou uvedeny v tabulce č. 1.

**Tab. 1.**

Vzdálenost (km)	Počet dojíždějících studentů
(140;240>	23
(240;340>	14
(340;440>	4
(440;540>	2
(540;640>	1

- a) Proveďte rozbor příkladu z hlediska základní statistické terminologie.  
b) Propočítejte všechny druhy četností a komentujte hodnoty ve třetím řádku.

*Řešení:*

Jde o intervalově tříděná data. Třídícím znakem je vzdálenost místa bydliště do Českých Budějovic. Modálním intervalem je interval obsahující vzdálenost od 140 – 240 Km.

Vzdálenost (km)	Absolutní četnost	Relativní četnost	Kumulativní četnost	Kumulativní relativní četnost
(140;240>	23	23/44	23	23/44
(240;340>	14	7/22	37	37/44
(340;440>	<b>4</b>	<b>1/11</b>	<b>41</b>	<b>41/44</b>
(440;540>	2	1/22	43	43/44
(540;640>	1	1/44	44	1

Ze 44 studentů desátého patra pouze 4 dojíždějí z místa bydliště, které je vzdáleno (340;440> km. Tito čtyři studenti představují 1/11 z celkového počtu dotazovaných studentů. Dále můžeme říci, že 41 studentů dojíždí na kolej ze vzdálenosti (140;440> km. Tato část studentů činí 41/44, tj. cca 93,18 % dotazovaných studentů.

2. V tabulce 2 jsou uvedeny hodnoty sklizňových ztrát ( $t \cdot ha^{-1}$ ) získaných ze 60 odběrných míst. Intervalově roztříděte data – využijte všechny možnosti, které balík Statistica nabízí – a vyberte nejvhodnější třídění.

**Tab. 2.**

0,29	0,10	1,20	0,24	0,32	0,41
0,43	0,48	0,51	2,12	3,40	0,11
0,44	0,37	0,50	0,64	0,09	1,13
0,50	0,33	0,41	0,43	0,51	0,63
0,92	0,34	0,52	0,31	1,41	0,03
0,71	0,54	0,52	0,31	0,43	0,53
0,63	0,25	0,23	1,74	0,10	0,72

1,13	0,54	0,41	0,27	0,63	0,01
0,31	0,23	0,56	0,37	0,42	1,92
0,41	0,38	0,16	0,52	0,71	0,52

*Řešení:*

Soubor má 60 dat, tj.  $n = 60$ . Budeme třídit intervalově. Počet intervalů zjistíme pomocí Sturgesova vzorce:  $k = 1 + 3,3 \log n$ . Vypočítáme, že  $k = 6,867899$ . Roztřídíme tedy soubor do 7 skupin ( můžeme zkusit i do 6 skupin ).

Použijeme-li nabídku *Přesný poč. intervalů* a nastavíme-li počet na 7, -viz obr. 1, získáme tabulku četností 1 .

Obr. 1

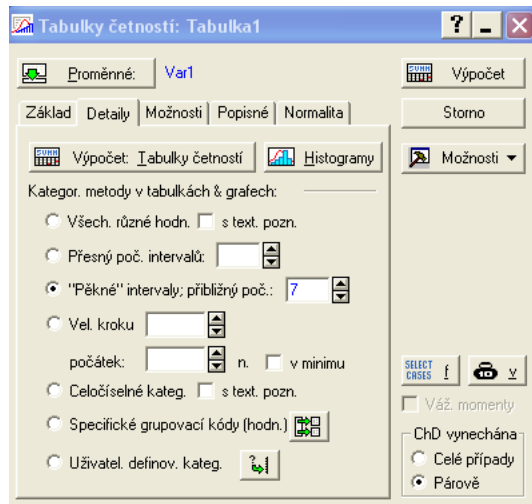


Tab. četností 1

		Tabulka četností:Var1 (Tabulka1)			
OD	DO	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
-,272500	<x<=,292500	13	13	21,66667	21,66667
,292500	<x<=,857500	38	51	63,33333	85,00000
,857500	<x<=1,422500	5	56	8,33333	93,33333
1,422500	<x<=1,987500	2	58	3,33333	96,66667
1,987500	<x<=2,552500	1	59	1,66667	98,33333
2,552500	<x<=3,117500	0	59	0,00000	98,33333
3,117500	<x<=3,682500	1	60	1,66667	100,00000
ChD		0	60	0,00000	100,00000

Vybereme-li nabídku „*Pěkné intervaly*“; *přibližný poč.*, viz obr. 2, získáme výstup uvedený v tab. četností 2. Všimněte si, že software roztřídil soubor pouze do 5 intervalů. Navíc interval ( 4,000000 ; 5,000000 > vykazuje nulovou absolutní četnost a je tedy zbytečný. Intervaly jsou ale skutečně „pěkné“.

Obr. 2.



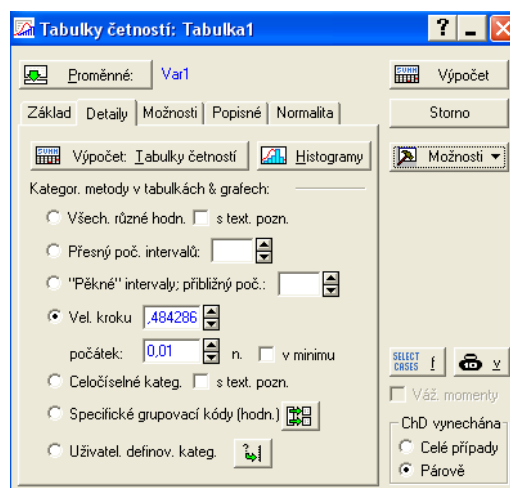
Tab. četností 2

		Tabulka četností:Var1 (Tabulka1)			
OD	DO	Četnost	Kumulativní četnost	Rel. četnost	Kumulativní rel. četnost
0,000000	<x<=1,000000	52	52	86,66667	86,66667
1,000000	<x<=2,000000	6	58	10,00000	96,66667
2,000000	<x<=3,000000	1	59	1,66667	98,33333
3,000000	<x<=4,000000	1	60	1,66667	100,00000
4,000000	<x<=5,000000	0	60	0,00000	100,00000
ChD		0	60	0,00000	100,00000

Toto třídění vede k dosti podstatné ztrátě informací. Je příliš hrubé, nicméně přesto vystihuje charakter posklizňových ztrát (tj. 52 případů mělo hodnotu menší nebo rovnu 1,00 t·ha<sup>-1</sup>).

Použijeme-li nabídku *Vel. Kroku*, musíme velikost kroku spočítat. V modulu *Popisná statistika* lze najít minimální a maximální hodnotu statist. znaku. (při tak malém souboru lze nalézt min. a max. přímo ze zadání úlohy). Min = 0,01 t·ha<sup>-1</sup>, max = 3,4 t·ha<sup>-1</sup>. Rozdíl mezi maximální a minimální hodnotou je 3,39 t·ha<sup>-1</sup>. Budeme-li třídit do 7 intervalů, pak délka jednoho intervalu - tj. velikost kroku - je  $3,39/7 = 0,484286$ . Nastavení modulu je uvedeno na obr.3 a výsledná sestava v tab. četností 3.

Obr. 3:



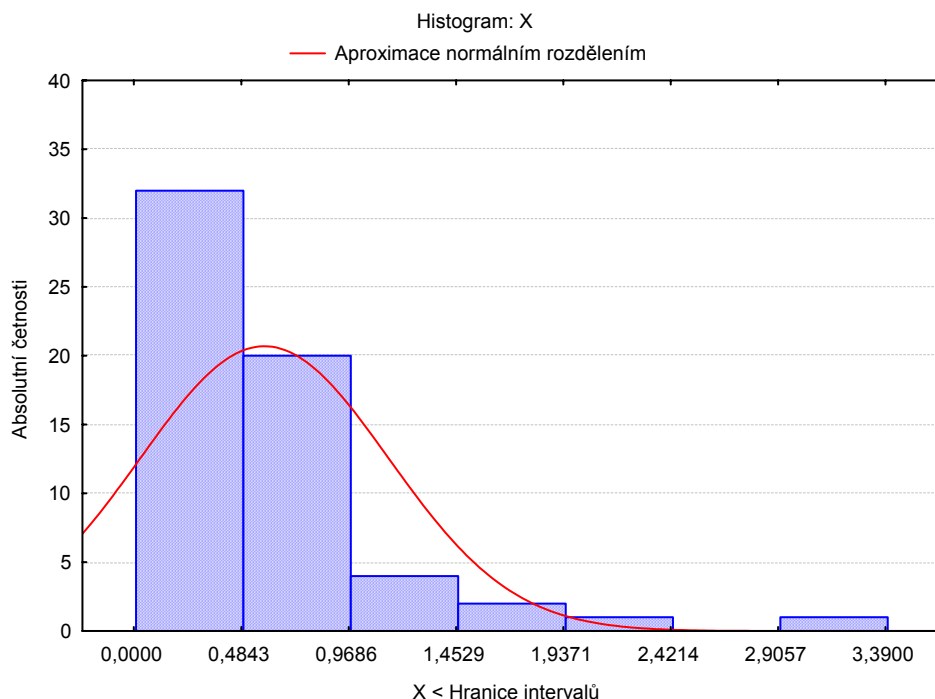
Počátek třídění je nastaven na hodnotu 0,01.

Tab. četností 3

		Tabulka četností:Var1 (Tabulka1)			
OD	DO	Četnost	Kumulativní četnost	Rel. četnost	Kumulativní rel. četnost
0,0100000	<=x<,4942860	32	32	53,33333	53,3333
,4942860	<=x<,9785720	20	52	33,33333	86,6667
,9785720	<=x<1,462858	4	56	6,66667	93,3333
1,462858	<=x<1,947144	2	58	3,33333	96,6667
1,947144	<=x<2,431430	1	59	1,66667	98,3333
2,431430	<=x<2,915716	0	59	0,00000	98,3333
2,915716	<=x<3,400002	1	60	1,66667	100,0000
3,400002	<=x<3,884288	0	60	0,00000	100,0000
ChD		0	60	0,00000	100,0000

Všimněte si, že program přidal navíc jeden interval, který vykazuje nulovou absolutní četnost! Pokud bychom chtěli znázornit provedené intervalové třídění prostřednictvím histogramu, stačí stisknout tlačítko s názvem *Tabulky četností: Tabulka1* v levém dolním rohu (dostaneme se tak o krok zpět) a poté stisknout tlačítko *Histogram*. Výsledný grafický výstup je zobrazen na grafu 1.

Graf 1: Histogram



3. V tabulce č. 3. je uveden počet lidí, přítomných na každé z 36 lekcí anglického jazyka. Soubor dat roztrídíte. Odůvodněte zvolený způsob třídění. Výsledek třídění graficky znázorníte.

**Tab. 3.**

30	23	25	30	30	18
23	20	25	23	32	23
17	30	25	27	27	18
22	17	31	24	25	28
32	25	26	24	21	17
27	35	25	34	28	25

*Řešení:*

Zvolíme intervalové třídění;  $n = 36$ . Pomocí Sturgesova vzorce spočítáme  $k = 6,1357$ . Soubor tedy roztrídíme do šesti intervalů. Minimální hodnota statistického znaku je 17, maximální hodnota statistického znaku je 35.

Třídíme diskretní statistický znak, je proto nutno stanovit správně hranice intervalů.

Hranice intervalu	Interval	Prvky intervalu
$(15,2 - 18,8>$	17 – 18	17, 17, 17, 18, 18
$(18,8 - 22,4>$	19 – 22	20, 21, 22
$(22,4 - 26,0>$	23 – 26	23, 23, 23, 23, 24, 24, 25, 25, 25, 25, 25, 25, 25, 25, 26
$(26,0 - 29,6>$	27 – 29	27, 27, 27, 28, 28
$(29,6 - 33,2>$	30 – 33	30, 30, 30, 30, 31, 32, 32
$(33,2 - 36,8>$	34 – 35	34, 35

