

STATISTICA

Téma 1. Práce s datovým souborem

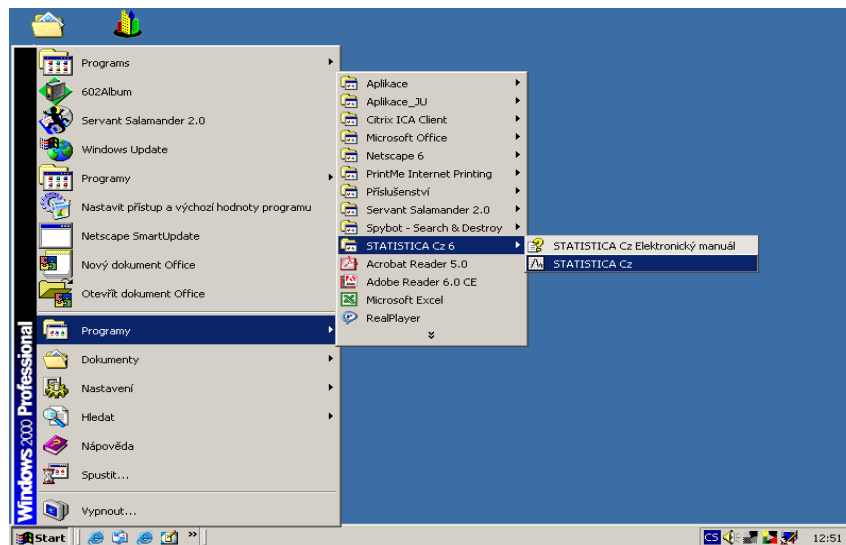
1) Otevření datového souboru

Program *Statistika.cz* otevíráme z ikony *Start*, nabídka *Programy*, podnabídka *Statistika Cz 6*.

Ze dvou nabídnutých možností vybereme *Statistika Cz* - viz. obr 1.

Po otevření souboru se objeví standardní tabulka, velmi podobná tabulce v *Excelu*. Má 10 řádků a 10 sloupců –viz obr.2. Řádky nazýváme *případy* a sloupce nazýváme *proměnné*. Všimněme si, že sloupce jsou pojmenovány Var1 až Var10. Řádky jsou číslovány od 1 do 10.

Obr. 1



Obr. 2

A screenshot of the 'Data: Tabulka2 (10s krát 10ř)' window. It shows a 10x10 grid with columns labeled 'Var1' through 'Var10' and rows labeled '1' through '10'. The cell at row 1, column 1 is selected.

	1 Var1	2 Var2	3 Var3	4 Var4	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

2) Editace datového souboru

V souboru dat se pohybujeme buď pomocí myši nebo šipek na klávesnici. Do sloupců zapisujeme „hodnoty“ (číselné i textové) statistických znaků (proměnných). Chceme-li zadat vícečíslnou řadu, nemusíme vypisovat každý řádek zvlášť, ale vyplníme dvě políčka, ta označíme a tahem myši za černý čtvereček v pravém dolním rohu označených políček kopírujeme řadu do dalších polí.

3) Přidávání a odstraňování proměnných a případů

V hlavním menu klikneme na nabídku **Vložit**, zvolíme možnost **Přidat případy**, popřípadě **Přidat proměnné** a vyplníme okno, které se otevře –viz obr. 3a, 3b.

Obr. 3a

The dialog box 'Přidat proměnné' contains the following fields and options:

- Kolik: 1
- Za: Var4
- Jméno: NProm
- Typ: Double
- Kód ChD: -9999
- Délka: 8
- Formát zobrazení: Obecný (selected), Číslo, Datum, Čas, Vědecký, Měna, Procenta, Zlomky, Vlastní
- Long name field: Dlouhé jméno (popis n. výraz obsahující Funkce)
- Buttons: OK, Storno
- Help text: Použijte 0 v poli "Za" pro vložení před první proměnnou. Poklepejte nebo stiskněte F2.
- Info text: Pokud se mají spočítat hodnoty nové proměnné a množina dat je příliš velká, je rychlejší přidat proměnné a současně přepočítat jejich hodnotu pomocí dávkové transformace dat (nabídka Data).

Obr. 3b

The dialog box 'Přidat případy' contains the following fields and options:

- Kolik: 36
- Vložit za případ: 0
- Buttons: OK, Storno
- Text: Před prvním případem vložit 0.

Při odstraňování proměnných (případů) musíme danou proměnnou (případ) označit, poté klikneme na pravé tlačítko myši a z nabídky, která se objeví viz. obr.4a, obr.4b, vybereme **Odstranit proměnné** resp. **Odstranit případy**. Sloupec nebo řádek označíme tak, že klikneme myší na jeho záhlaví.

Obr. 4a

The dialog box 'Odstranit proměnné' contains the following fields and options:

- Od proměnné: Bilkoviny
- Do proměnné: Bilkoviny
- Buttons: OK, Storno
- Text: Pro výběr ze seznamu poklepejte na poříčko proměnné (příp. F2).

Obr. 4b

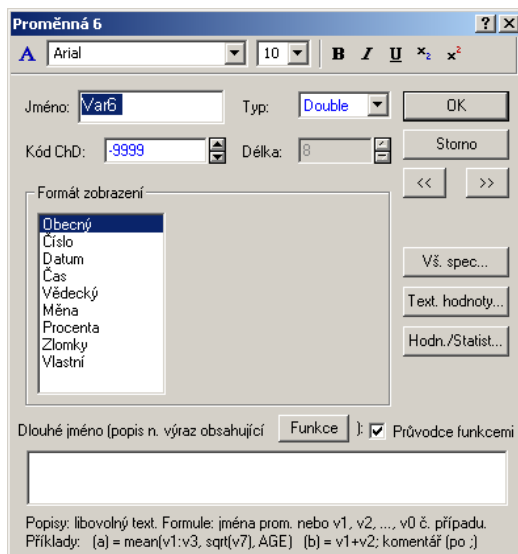
The dialog box 'Odstranit případy' contains the following fields and options:

- Od případu: 1
- Do případu: 36
- Buttons: OK, Storno

4) Přejmenování proměnných

Při přejmenovávání proměnné stačí dvakrát kliknout myší do záhlaví této proměnné. Do okna, které se objeví - viz. obr.5, zapíšeme nové jméno proměnné.

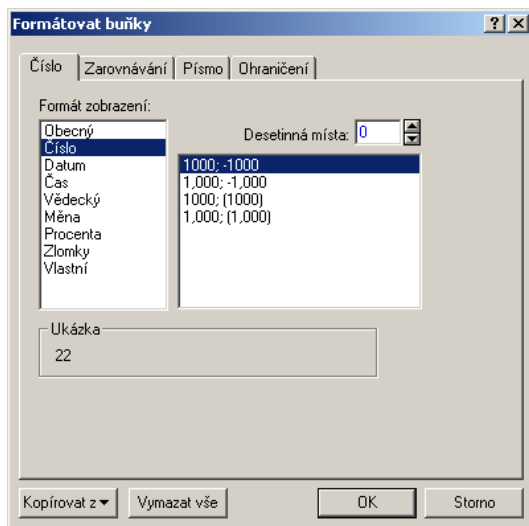
Obr. 5



5) Nastavení počtu desetinných míst

Pokud chceme u numerické proměnné nastavit určitý počet desetinných míst, dvakrát klikneme myší do jejího záhlaví; v okně, které se objeví, vybereme nabídku **číslo** a do příslušného políčka zapíšeme požadovaný počet desetinných míst. Stejného efektu dosáhneme prostřednictvím nabídky **Formát**, podnabídky **Formát buňky**. Objeví se okno -viz.obr. 6, které vyplníme.

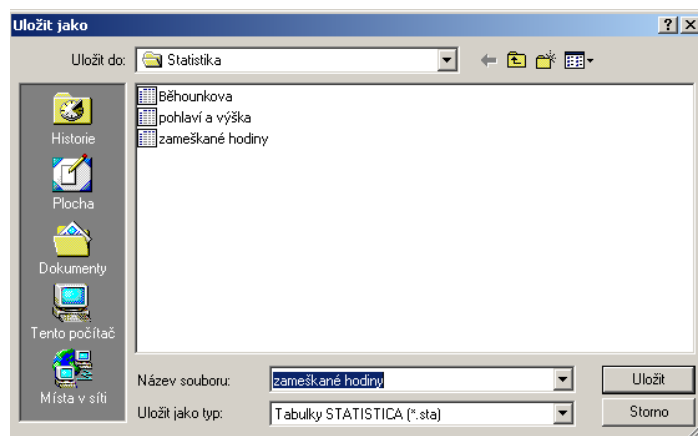
Obr. 6



6) Uložení souboru

Chceme-li uložit soubor, vybereme v hlavním menu nabídku **Soubor**, podnabídku **Uložit jako**. Objeví se okno- viz. obr.7, kam napíšeme název ukládaného souboru. Zásadně neukládáme pod názvem **Tabulka**. Tento název je vyhrazen pro novou tabulku. Ukládáme-li graf, má soubor příponu ***stw**, datový soubor má příponu ***sta**.

Obr. 7



7) Oprava chyb

Oprava chyb se provádí přepsáním příslušného chybně zadaného údaje.

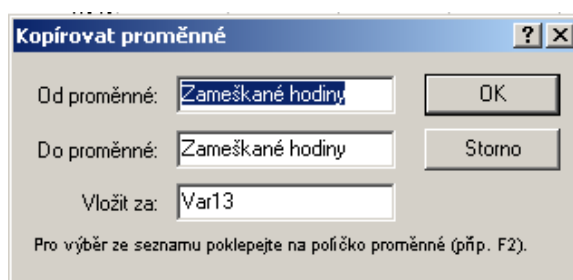
8) Kopírování dat z Excelu do Statisticy (ze Statisticy do Excelu)

Data, která chceme z Excelu zkopírovat označíme, poté stiskneme klávesy CTRL C, přepneme do programu *Statistica*, kurzor dáme na správné místo a data vložíme prostřednictvím kláves CTRL V. Kopírování ze *Statisticy* probíhá obdobně. Můžeme si však zvolit mezi možnostmi **Kopírovat** nebo **Kopírovat se záhlavím**. Tato alternativa se nabídne, kopírujeme-li prostřednictvím menu **Úpravy**. Pozor! Ve výpočetním středisku se při této manipulaci často počítač kousne.

9) Kopírování hodnot proměnných na jiné místo v tabulce

Potřebujeme-li zkopírovat proměnné na jiné místo v tabulce, vyplníme okno, které se otevře, jakmile v hlavní nabídce zvolíme ikonu **Vložit/Kopírovat proměnné** –viz obr.8.

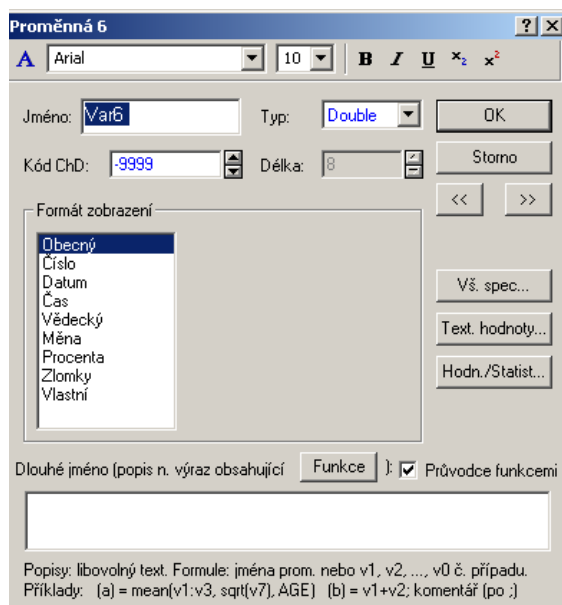
Obr. 8



V programu *Statistica* lze vytvářet proměnné, které jsou výsledkem aritmetických operací (popř. výpočtu funkcí) mezi již existujícími proměnnými. Postupujeme tak, že dvakrát klikneme na záhlaví proměnné, kterou chceme vytvářet. Zobrazí se okno -viz. obr.9; do jeho dolní části zapíšeme rovnítko a potřebnou aritmetické operaci, či vložíme potřebnou funkci.

Např. existují-li proměnné *dolnimez* a *hornimez*, pak proměnnou *středintervalu* vytvoříme předpisem $=(dolnimez+hornimez)/2$ Pozor! Názvy proměnných musí být jednoslovné. Klikneme na tlačítko **OK**. Objeví se okno se zprávou, že vzorec byl správně zadán a s dotazem, zda si přejeme provést výpočet hodnot nové proměnné. Výpočet se provede po stisknutí tlačítka **ano**.

Obr. 9

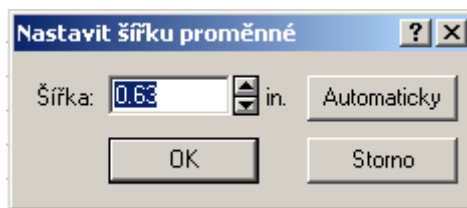


11) Úprava šířky sloupce

Šířku vybraného sloupce upravíme tak, že v hlavním menu vybereme nabídku **Formát**, podnabídku **Proměnné**, variantu **Šířka proměnné**. Objeví se okno – viz obr.10. Do něho zapíšeme požadovanou šířku proměnné.

Šířku sloupce upravit i tahem myši jako v *Excelu*.

Obr. 10



STATISTICA

Téma 1. Třídění

Cílem třídění je uspořádat, tj. rozřadit, zjištěné hodnoty statistického znaku (proměnné) do skupin.

Výsledkem třídění je tabulka, obsahující hodnoty znaku (popřípadě rozmezí hodnot znaku), absolutní četnosti, popřípadě další typy četností.

Dle charakteru znaků rozlišujeme třídění:

1. *prosté* – používáme je v případě, že třídícím znakem je kategoriální znak nebo numerický znak s malým počtem obměn hodnot.
2. *intervalové* – používáme je v případě, že třídícím znakem je numerický znak s velkým počtem hodnot. K odhadu přibližného počtu intervalů používáme Sturgesovo pravidlo.

Mezi nejčastěji užívané četnosti patří:

- 1) *absolutní četnost*; udává, kolik statistických jednotek má určitou hodnotu znaku,
- 2) *relativní četnost*; udává, kolik procent statistických jednotek má určitou hodnotu znaku.

Méně často užíváme *kumulativní četnost* a *kumulativní relativní četnost*.

Možnosti software (předpokládejme např., že máme naeditovány druhy nápojů a jejich ceny v různých restauracích-viz obr.1).

Obr. 1

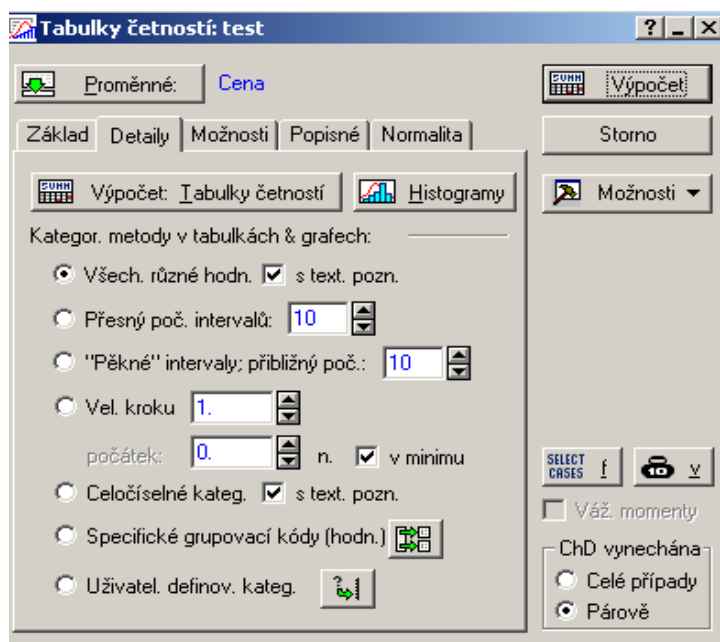
	1 Cena	2 Nápoje(káva, čaj, capuccino)	3 Var3	4 Var4	5 Var5	6 Var6	7 Var7	8 Var8	9 Var9	10 Var10
1	10 ě									
2	12 ě									
3	15 k									
4	20 c									
5	17 ě									
6	15 k									
7	30 c									
8	25 k									
9	35 c									
10	18 k									
11	12 ě									
12	15 k									
13	25 k									
14	28 c									
15	30 c									
16										

V menu *Statistika* vybereme submenu *Základní statistiky/tabulky* a v něm nabídku *Tabulky četností*. Výběr potvrdíme tlačítkem *OK*.

V horní části otevřeného okna zvolíme tlačítko *Proměnné*. Otevře se seznam proměnných, z něhož vybereme proměnné, které chceme rozřadit - např. proměnnou *Cena*.

Okno *Tabulky četností* má několik karet –viz obr.2. Nejprve obvykle volíme kartu *Možnosti*, a v ní zatrhneme typy četností které chceme počítat. Vlastní třídění provedeme tak, že klikneme na ikonu *Výpočet*.

Obr. 2



1) **Třídění prosté** (třídíme nápoje dle ceny):

V kartě **Detaily**, v části **Kategor.metody v tabulkách & grafech**- viz obr. 2, vybereme způsob třídění. V případě prostého třídění zvolíme nabídku **Všechny různé hodnoty s textovou poznámkou**

Přejeme-li si, aby výstupem byla tabulka, klikneme myší na lištu **Výpočet:Tabulky četností**. Požadujeme –li grafický výstup, klikneme na lištu **Histogramy**.

Výsledek prostého třídění proměnné **Cena** je uveden na obr.3, výsledek prostého třídění proměnné **Nápoje** je uveden na obr. 4.

Obr. 3

Kategorie	Tabulka četností:Cena			
	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
10	1	1	5.00000	5.0000
12	2	3	10.00000	15.0000
15	3	6	15.00000	30.0000
17	1	7	5.00000	35.0000
18	1	8	5.00000	40.0000
20	1	9	5.00000	45.0000
25	2	11	10.00000	55.0000
28	1	12	5.00000	60.0000
30	2	14	10.00000	70.0000
35	1	15	5.00000	75.0000
ChD	5	20	25.00000	100.0000

Obr. 4

Tabulka četností:Nápoje(káva, čaj, cappuccino)				
Kategorie	Četnost	Kumulativní četnost	Rel. četnost	Kumulativní rel. četnost
č:	4	4	20.00000	20.0000
k	6	10	30.00000	50.0000
c	5	15	25.00000	75.0000
ChD	5	20	25.00000	100.0000

2) Třídění intervalové

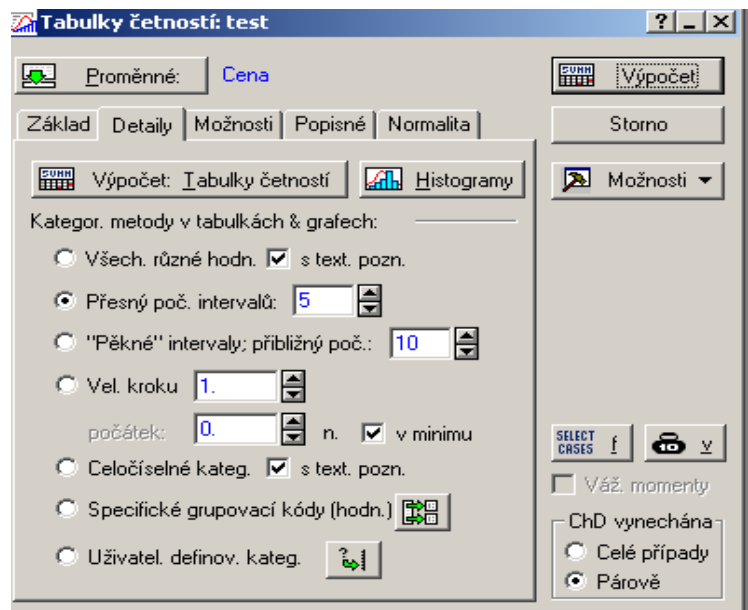
Při intervalovém třídění vybíráme v okně *Tabulky četností*, kartě *Details* některou ze tří následujících možností:

a) *Přesný poč. intervalů*

Hodnoty statistického znaku budou roztrženy do tolika intervalů, kolik zadáme. Počet intervalů zjistíme dle **Sturgesova pravidla**. V případě neceločíselného výsledku (např. 4,88) zkusíme obě možnosti, tj. 4 i 5 intervalů a na základě histogramu se pro některou z variant rozhodneme.

Nastavení okna i výsledná sestava jsou uvedeny na obr.5 a obr.6.

Obr. 5

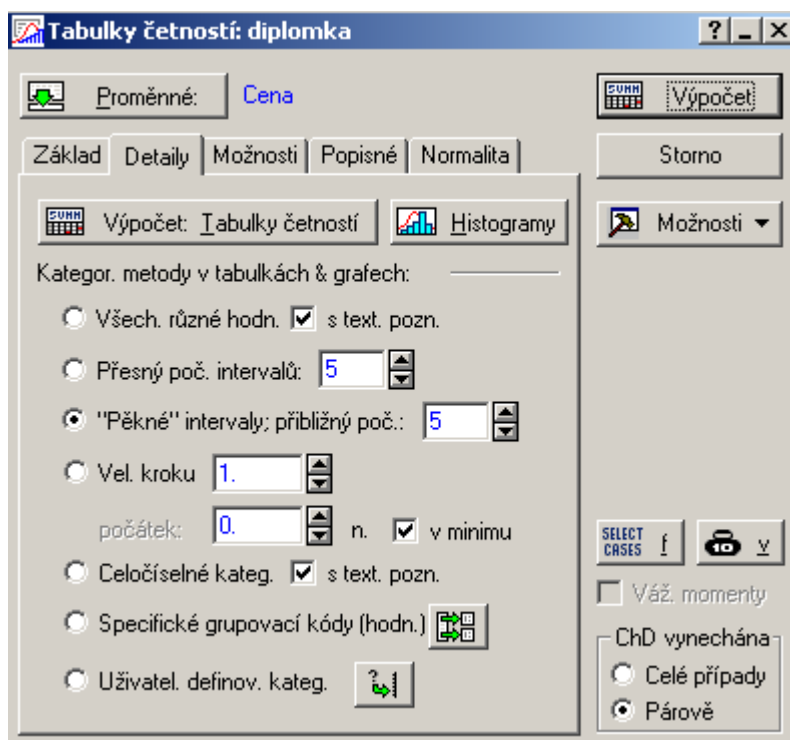


Obr. 6

		Tabulka četností:Cena			
OD	DO	Četnost	Kumulativní četnost	Rel. četnost	Kumulativní rel. četnost
6.875000	$x \leq 13.12500$	3	3	15.00000	15.0000
13.12500	$x \leq 19.37500$	5	8	25.00000	40.0000
19.37500	$x \leq 25.62500$	3	11	15.00000	55.0000
25.62500	$x \leq 31.87500$	3	14	15.00000	70.0000
31.87500	$x \leq 38.12500$	1	15	5.00000	75.0000
ChD		5	20	25.00000	100.0000

2) *Pěkné intervaly; přibližný poč.* – viz obr.7

Obr. 7



Pro porovnání výstupů s předchozím případem (obr. 6), rozřídíme nabídkou *Pěkné intervaly* proměnnou **Cena** také do pěti intervalů –viz obr. 8. V porovnání s obr 6 jsou intervaly v obr. 8 skutečně „pěkné“. Poznámka: zvolíme-li nabídku třídění do „pěkných intervalů“, respektuje program naše přání. Nemusí však respektovat počet intervalů, který jsme zadali. Tento údaj je pouze orientační.

Obr. 8

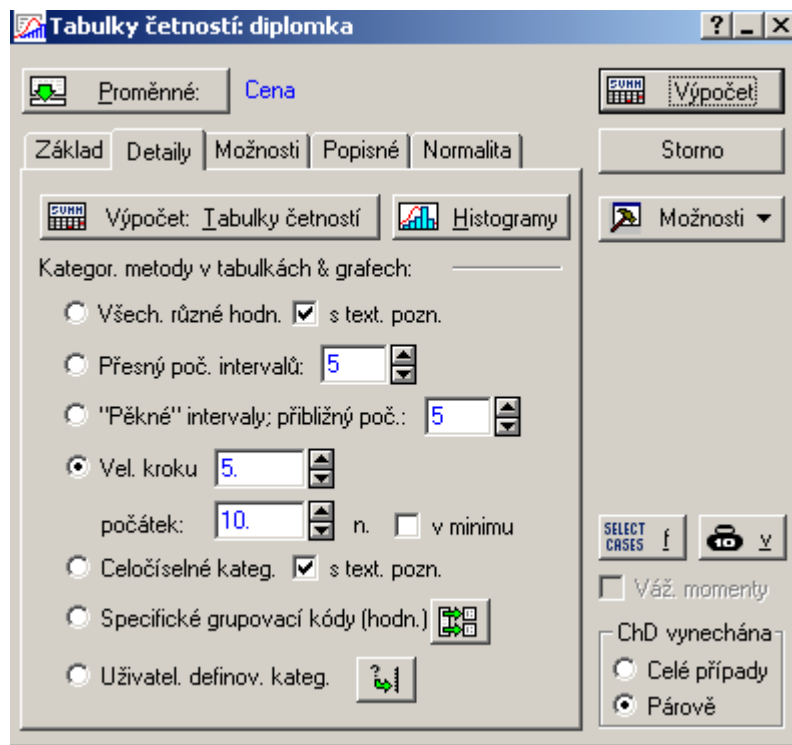
		Tabulka četností:Cena			
OD	DO	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
5.000000	$x \leq 10.000000$	1	1	5.00000	5.0000
10.000000	$x \leq 15.000000$	5	6	25.00000	30.0000
15.000000	$x \leq 20.000000$	3	9	15.00000	45.0000
20.000000	$x \leq 25.000000$	2	11	10.00000	55.0000
25.000000	$x \leq 30.000000$	3	14	15.00000	70.0000
30.000000	$x \leq 35.000000$	1	15	5.00000	75.0000
35.000000	$x \leq 40.000000$	0	15	0.00000	75.0000
40.000000	$x \leq 45.000000$	0	15	0.00000	75.0000
ChD		5	20	25.00000	100.0000

3) Vel. kroku

Zvolíme-li možnost *Velikost kroku*, musíme velikost kroku zadat. V našem případě jsme podle vzorečky $\frac{HORNÍMEZ - DOLNÍMEZ}{POČETINTERVALU(k)}$ spočítali $\frac{35-10}{4,88} = 5,12$. I když lze zadat neceločíselnou délku kroku, volíme raději celé číslo; tedy zadáme 5 nebo 6. Pro ukázkou jsme zvolili číslo 5.

Poté se rozhodneme, zda bude třídění začínat od určité konkrétní hodnoty, nebo zda třídění začne od minimální hodnoty statistického znaku. Svou volbu vyjádříme buď vyplněním políčka *Počátek* nebo zaškrtnutím políčka *v Minimu* – viz obr.9.

Obr. 9



Výsledek třídění je uveden na obr. 10.

Obr. 10

		Tabulka četností: Cena			
OD	DO	Četnost	Kumulativní četnost	Rel. četnost	Kumulativní rel. četnost
10.00000	<=x<15.00000	3	3	15.00000	15.0000
15.00000	<=x<20.00000	5	8	25.00000	40.0000
20.00000	<=x<25.00000	1	9	5.00000	45.0000
25.00000	<=x<30.00000	3	12	15.00000	60.0000
30.00000	<=x<35.00000	2	14	10.00000	70.0000
35.00000	<=x<40.00000	1	15	5.00000	75.0000
ChD		5	20	25.00000	100.0000