

Cvičení 1.: Základní informace o systému STATISTICA, bodové rozložení četností

Základní informace o programovém systému STATISTICA

Systém má modulární stavbu. V multilicenci pro Masarykovu univerzitu jsou k dispozici moduly: Základní statistiky/tabulky, Vícerozměrná regrese, ANOVA, Neparametrická statistika, Prokládání rozdělení, Pokročilé lineární/nelineární modely, Vícerozměrné průzkumné techniky, Průmyslová statistika & Six Sigma, Analýza síla testu, Automatizované neuronové sítě.

Velké množství informací o systému STATISTICA lze najít na webové stránce společnosti StatSoft, která je jejím distributorem v České republice (www.statsoft.cz). Z této stránky vede rovněž odkaz na elektronickou učebnici statistiky.

Instalace systému STATISTICA 9.1 je dostupná na <https://inet.muni.cz/app/soft/licence>

STATISTICA má několik typů oken:

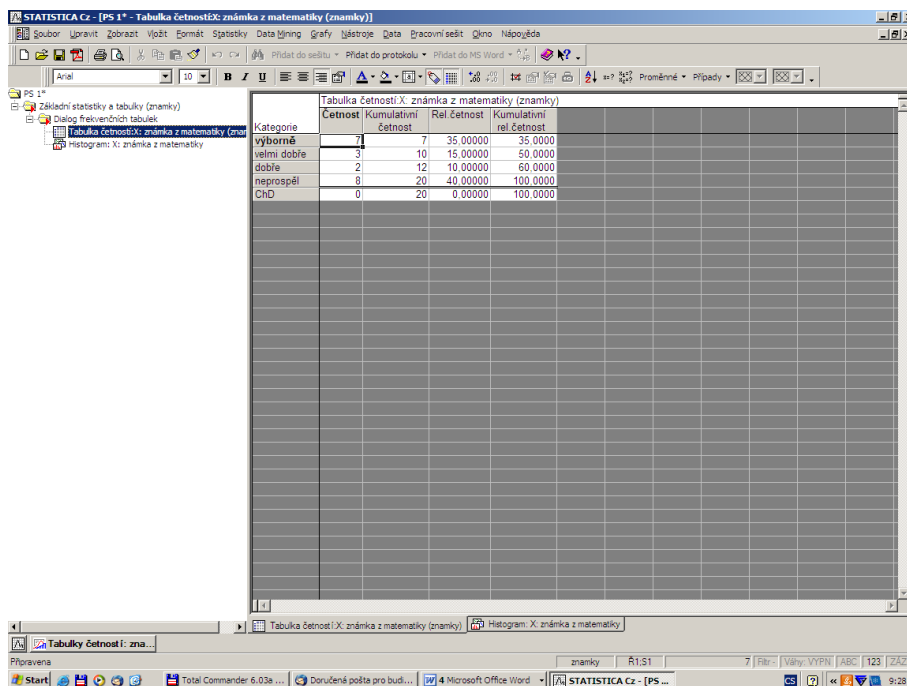
- **Datové okno (spreadsheet)** (má příponu sta, jeho obsah však lze exportovat i v jiných formátech). Do datového okna lze načítat datové soubory nejrůznějších typů (např. z tabulkových procesorů, databázové soubory, ASCII soubory).

Ukázka datového okna:

	1 X	2 Y	3 Z
1	2	2	0
2	1	3	1
3	4	3	1
4	1	1	0
5	1	2	1
6	4	4	1
7	3	3	1
8	3	4	0
9	1	1	0
10	1	1	0
11	4	2	1
12	4	4	0
13	2	2	0
14	4	3	1
15	2	3	1
16	4	4	0
17	1	1	0
18	4	3	1
19	4	4	1
20	1	3	0

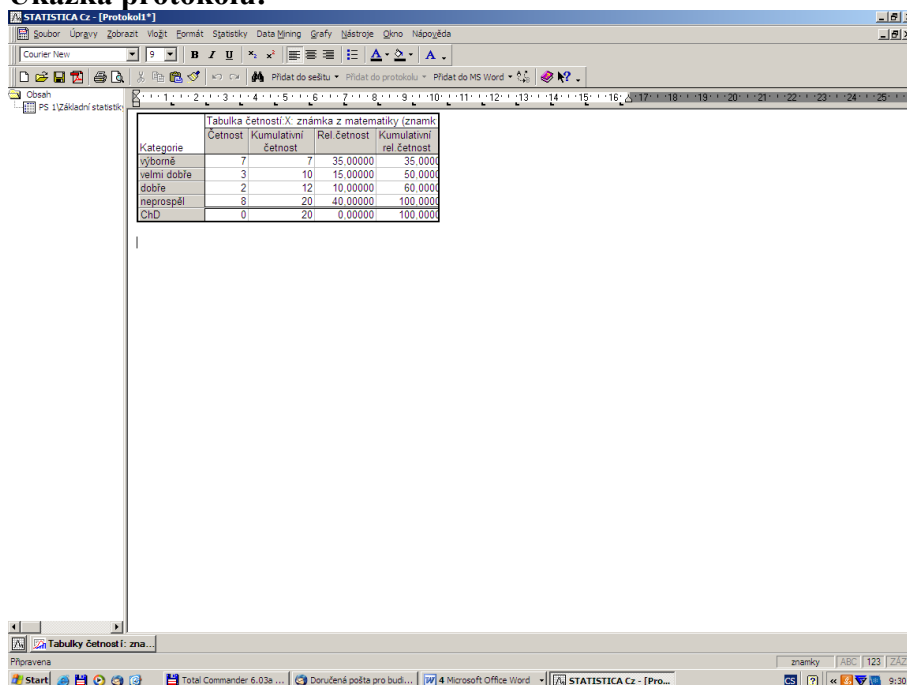
- **Pracovní sešit (workbook)** (má příponu stw). Do pracovního sešitu se ukládají výstupy, tj. tabulky a grafy. Skládá se ze dvou oken, v levém okně je znázorněna stromová struktura výstupů, v pravém jsou samotné výstupy. V levém okně se lze pohybovat myší nebo kurzorem, mazat, přesouvat, editovat apod. Výstupy mohou sloužit jako vstupy pro další analýzy a grafy.

Ukázka pracovního sešitu:

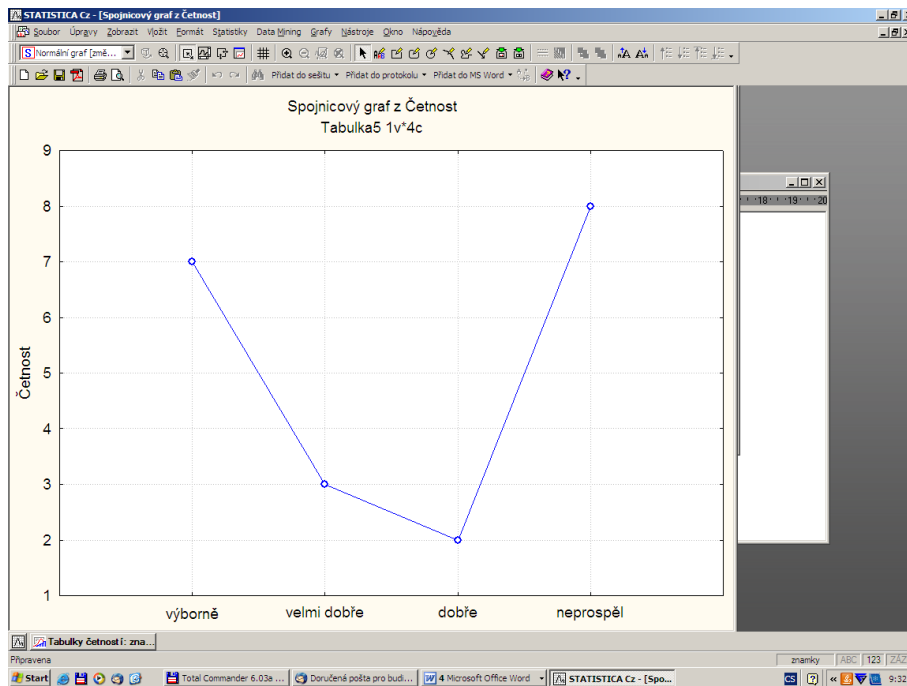


- **Protokol (report)** (má příponu str, lze ho uložit i ve formátu rtf, txt, htm či pdf). Pokud požadujeme, aby se výstupy ukládaly nejen do Pracovního sešitu, ale i do Protokolu či Wordu, postupujeme takto: Soubor – Správce výstupů – vybereme Výstup protokolu či Výstup MS Word – OK. Protokol se podobně jako Pracovní sešit skládá ze dvou oken. Do Protokolu můžeme vkládat vlastní text, vysvětlující komentáře, poznámky apod. Tabulky a grafy lze v Protokolu i Pracovním sešitu dále upravovat.

Ukázka protokolu:



- **Okno grafů** (přípona stg, lze ho uložit i jako bmp, jpg, png, tif a wmf či pdf). Získá se tak, že v Pracovním sešitu klikneme pravým tlačítkem na graf a vybereme Klonovat graf. Ukázka okna grafů:



- **Programovací okno** (přípona svb). Slouží pro zápis programů v jazyku STATISTICA Visual Basic. Vyvolá se z menu Nástroje – Makro.

Ukázka programovacího okna:

```

Option Base 1
Sub Main
Dim s As Spreadsheet
Set s = ActiveSpreadsheet
For i = 1 To 12
    s.Variable(i).FillRandomValues
    ' do proměnných v1 až v12 se uloží náhodná čísla
    ' z intervalu(0,1)
Next i
s.VariableLongName(13) = "Sua(v1-v12)-6"
' do proměnné v13 se uloží součet proměnných v1 až v12
' zapsány o 6
s.Recalculate
End Sub

```

Mezi jednotlivými typy oken se přepínáme pomocí položky Okno v hlavním menu.

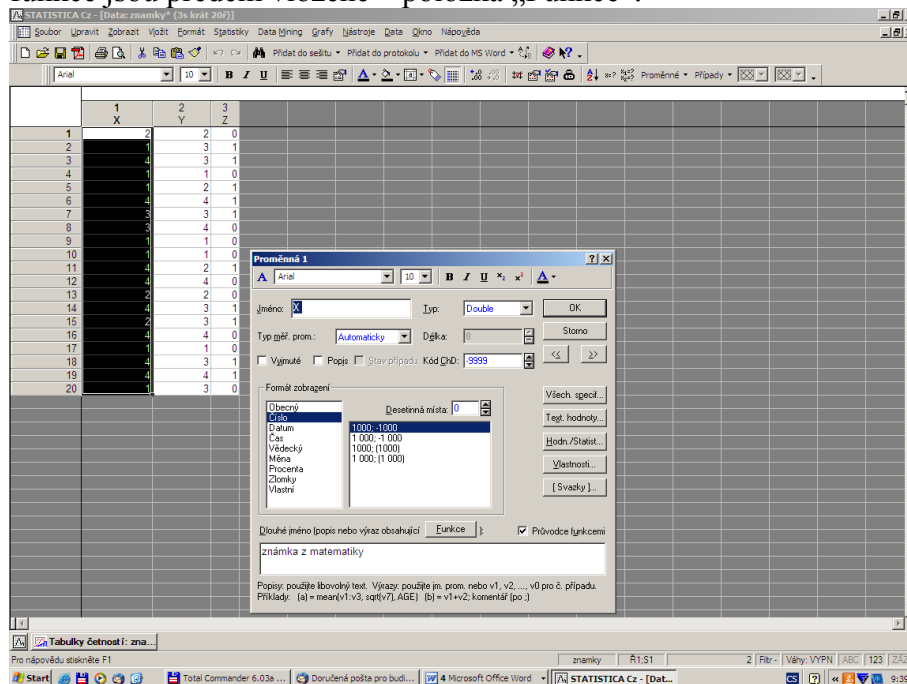
Od verze 8.0 je možno používat Projekty. Tento prvek umožňuje uložit soubor STATISTICA Projekt, a "zmrazit" tak doposud provedenou práci včetně umístění všech oken s objekty na obrazovce. Po otevření tohoto souboru se může přímo navázat na již provedenou část analýzy. Lze také uložit zaznamenané makro analýzy.

Vytvoření datového okna

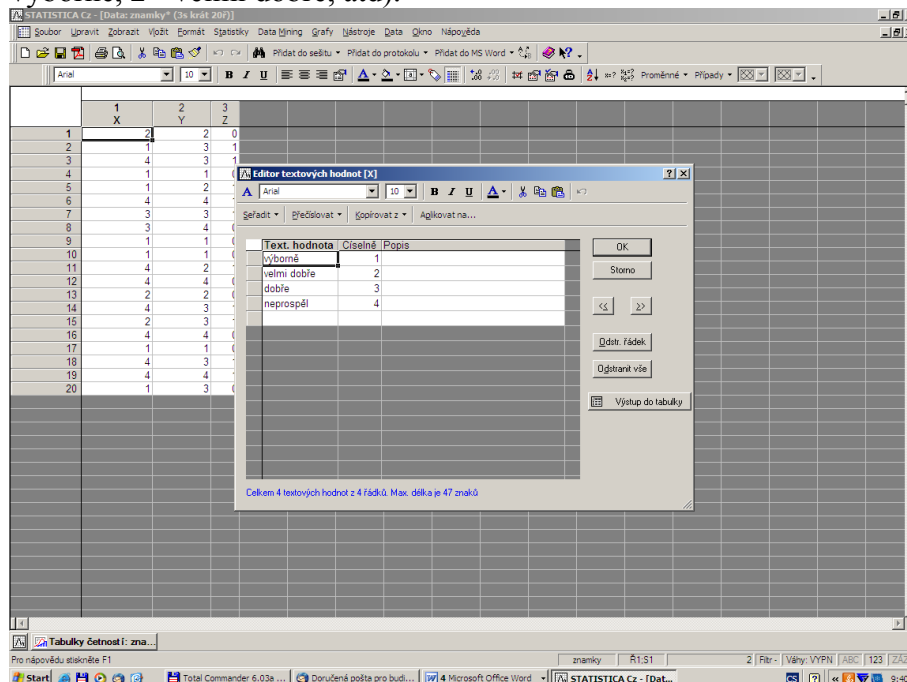
Po spuštění systému STATISTICA zavěme všechna okna, která se objeví.

Vytvoření nového souboru: Soubor – Nový – na liště Tabulka dat zvolíme potřebný počet proměnných (sloupce) a případů (řádky). Zde lze také zvolit formát dat, délku jmen případů atd. (tyto vlastnosti můžeme upravit i později).

Nastavení vlastností proměnné: 2x klikneme na název proměnné. Objeví se nové okno. Zde je možnost změny jména proměnné, formátu dat, nastavení kódu pro chybějící data apod. V položce „Dlouhé jméno“ lze nastavit dlouhý popis (tzv. návěští proměnné, např. X – známka z matematiky) a informace vztahující se k proměnné, ale spíše se „Dlouhé jméno“ používá pro vytvoření vzorců a transformaci proměnné (například: $=\sqrt{3}/\sqrt{2} * 100$). Některé funkce jsou předem vložené – položka „Funkce“.

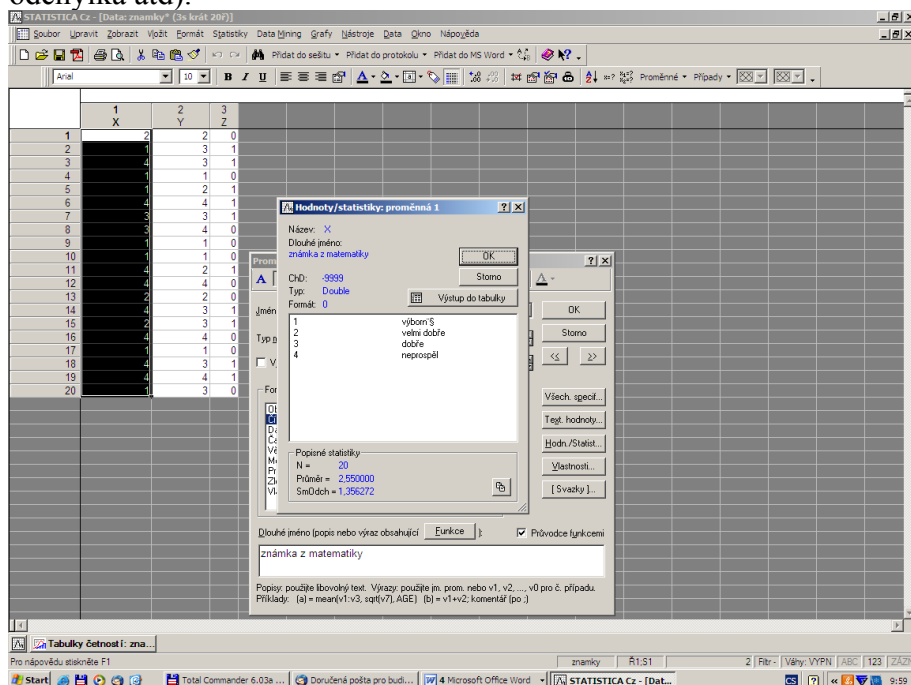


V položce „Textové hodnoty“ lze nastavit k variantám proměnné textový popis (např. 1 – výborně, 2 - velmi dobře, atd).



Poté můžeme přímo v datovém okně (ikona štítku v horní liště) měnit způsob zobrazení hodnot proměnné (buď číselný nebo slovní).

Položka Hodn/Statist udává celkové informace o proměnné (počet údajů, průměr, směrodatná odchylka atd).



Úkol 1.: Načtete soubor znamky.sta. Proměnným X, Y, Z vytvořte návěští (X - známka z matematiky, Y - známka z angličtiny, Z - pohlaví studenta). Popište, co znamenají jednotlivé varianty (u znaků X a Y: 1 - výborně, 2 - velmi dobře, 3 - dobře, 4 - neprospěl, u znaku Z: 0 - žena, 1 - muž).

Návod: Soubor – Otevřít – vybereme příslušný adresář se souborem znamky.sta – Otevřít. Kurzor nastavíme na Prom1 – 2x klikneme myší – Jméno X – Dlouhé jméno známka z matematiky, Text. hodnoty – 1 výborně, 2 velmi dobře, 3 dobře, 4 neprospěl, OK. U proměnné Y lze textové hodnoty okopírovat z proměnné X – v Editoru textových hodnot zvolíme Kopírovat z proměnné X. Přepínání mezi číselnými hodnotami a jejich textovým popisem se děje pomocí tlačítka s ikonou štítku.

Úkol 2.: Vytvořte

- tabulku rozložení četností známek z matematiky a známek z angličtiny,
- sloupkový diagram absolutních četností znaků X a Y,
- polygon absolutních četností znaků X a Y

Návod:

ad a) Statistika – Základní statistiky a tabulky – Tabulky četností – OK – Proměnné X, Y – OK - Výpočet.

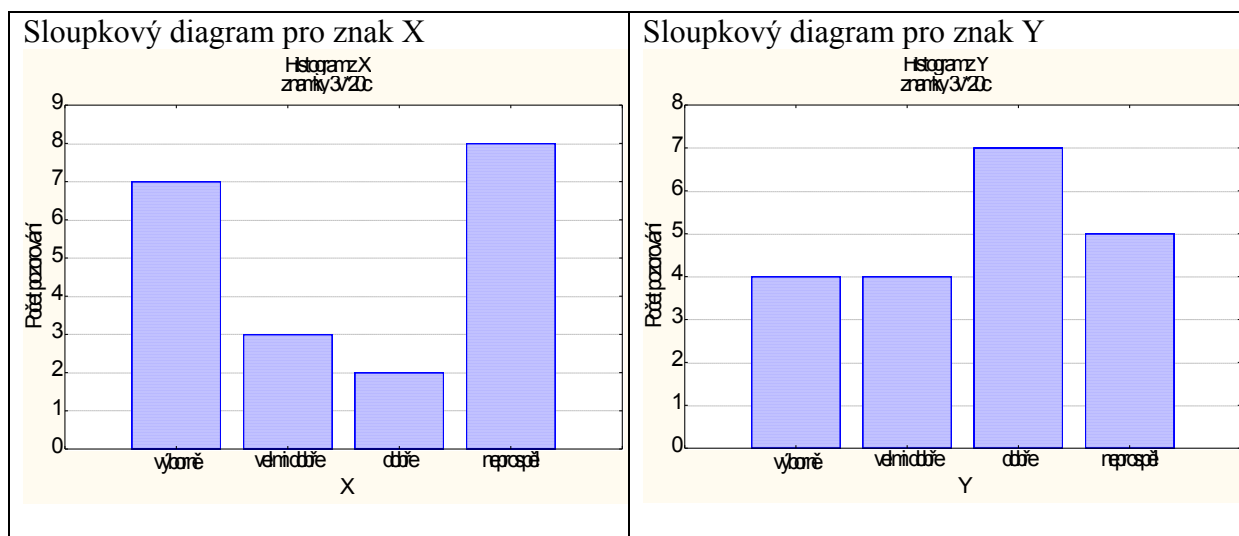
Tabulka rozložení četností známek z matematiky

Kategorie	Tabulka četností: X: známka z M			
	Četn.	Kumul. četnos	Rel.četn	Kumul. rel.četn
vyborne	7	7	35,00	35,00
velmi do	3	10	15,00	50,00
dobře	2	12	10,00	60,00
neprosp	8	20	40,00	100,00
ChD	0	20	0,00	100,00

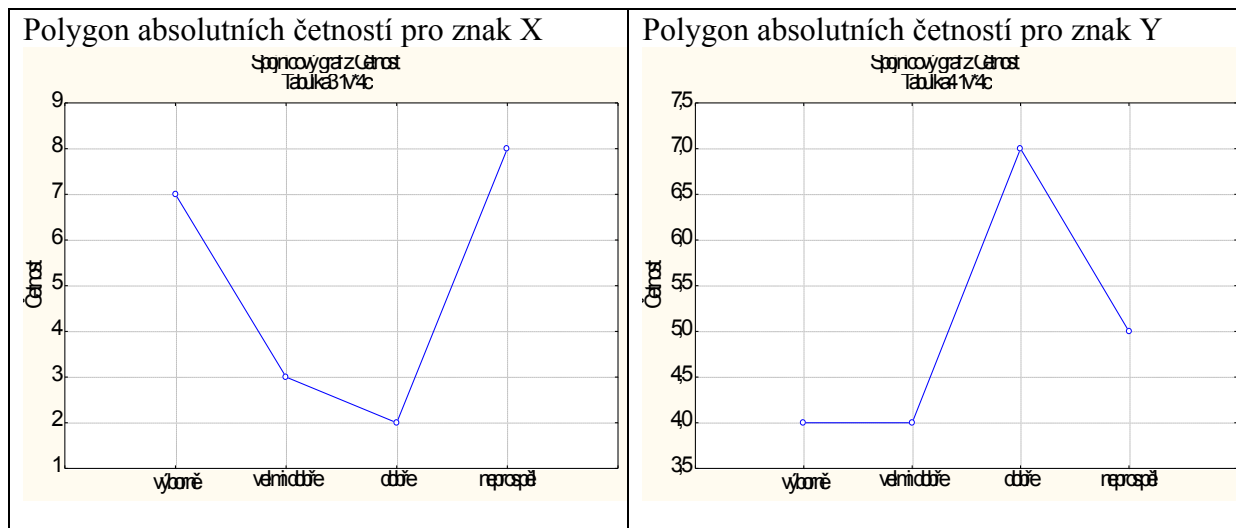
Tabulka rozložení četností známek z angličtiny

Kategorie	Tabulka četností: Y: známka z A			
	Četn.	Kumul. četnos	Rel.četn	Kumul. rel.četn
vyborne	4	4	20,00	20,00
velmi do	4	8	20,00	40,00
dobře	7	15	35,00	75,00
neprosp	5	20	25,00	100,00
ChD	0	20	0,00	100,00

ad b) Grafy – Histogramy – Proměnné X, Y – OK- vypneme Normální proložení – Detaily – zaškrtneme Mezery mezi sloupci - OK.



ad c) V pracovním sešitu vstoupíme do tabulky rozložení četností proměnné X resp. Y. Nastavíme se na řádek označený ChD. Pomocí Případy – Odstranit vymažeme tento řádek. Nastavíme se kurzorem na Četnost - klikneme pravým tlačítkem – Grafy bloku dat – Spojnicový graf: celé sloupce. Vykreslí se polygon absolutních četností.



Úkol 3.: Vytvořte tabulky rozložení četností známek z matematiky a angličtiny pouze

- a) pro ženy,
- b) pro muže.

Návod:

ad a) Statistika – Základní statistiky a tabulky – Tabulky četností – OK – Proměnné X, Y – OK – vybereme Select Cases - zaškrtneme Zapnout filtr – do okénka některé, vybrané pomocí výrazu napíšeme $Z = 0$, OK, Výpočet.

Tabulka rozložení četností známek z matematiky pro ženy:

Tabulka četností: X: známka z m				
Zhrnout podmínku: Z=0				
Kategorie	Četn	Kumulat četnos	Rel.četn	Kumulat rel.četn
výborn	5	5	50,00	50,00
velmi do	2	7	20,00	70,00
dobře	1	8	10,00	80,00
neprosp	2	10	20,00	100,00
ChD	10	10	0,00	100,00

Tabulka rozložení četností známek z angličtiny pro ženy:

Tabulka četností: Y: známka z a				
Zhrnout podmínku: Z=0				
Kategorie	Četn	Kumulat četnos	Rel.četn	Kumulat rel.četn
výborn	4	4	40,00	40,00
velmi do	2	6	20,00	60,00
dobře	1	7	10,00	70,00
neprosp	3	10	30,00	100,00
ChD	10	10	0,00	100,00

ad b) Statistika – Základní statistiky a tabulky – Tabulky četností – OK – Proměnné X, Y – OK – vybereme Select Cases - zaškrtneme Zapnout filtr – do okénka některé, vybrané pomocí výrazu napíšeme $Z = 1$, OK, Výpočet.

Tabulka rozložení četností známek z matematiky pro muže:

Tabulka četností: X: známka z m				
Zhrnout podmínku: Z=1				
Kategorie	Četn	Kumulat četnos	Rel.četn	Kumulat rel.četn
výborn	2	2	20,00	20,00
velmi do	1	3	10,00	30,00
dobře	1	4	10,00	40,00
neprosp	6	10	60,00	100,00
ChD	10	10	100,00	100,00

Tabulka rozložení četností známek z angličtiny pro muže:

Tabulka četností: Y: známka z a				
Zhrnout podmínku: Z=1				
Kategorie	Četn	Kumulat četnos	Rel.četn	Kumulat rel.četn
velmi do	2	2	20,00	20,00
dobře	6	8	60,00	80,00
neprosp	2	10	20,00	100,00
ChD	10	10	100,00	100,00

Úkol 4.: Nadále budeme pracovat s celým datovým souborem. Vytvoříme kontingenční tabulku simultánních absolutních četností znaků X a Y a graf simultánní četnostní funkce.

Návod: Statistika – Základní statistiky/tabulky – odškrtneme Zapnout filtr – OK - Kontingenční tabulky – OK – Select cases — Specif. tabulky - List 1 X, List 2 Y, OK, Výpočet.

Kontingenční tabulka (znamky)					
Četnost označených buněk > 10					
(Marginální součty nejsou označeny)					
X	Y	Y	Y	Y	Rad.
	výboř	velmi do	dobř	nepros	souč
výboř	4	1	2	0	7
velmi do	0	2	1	0	3
dobře	0	0	1	1	2
nepros	0	1	3	4	8
vs.skup	4	4	7	5	20

Vidíme, že ve výběrovém souboru byli 4 studenti, kteří měli z obou předmětů „výborně“, jeden student, který měl z matematiky „výborně“ a z angličtiny „velmi dobře“ atd. až 4 studenti, kteří z obou předmětů neprospěli.

Úkol 5.: Vytvořte kontingenční tabulku sloupcově a řádkově podmíněných relativních četností znaků X a Y.

Návod: Aktivujeme na liště Výsledky: kontingenční tabulky – Možnosti - zaškrtneme ve sloupci Výpočet tabulek volbu Procenta z počtu ve sloupci (resp. Procenta z počtu v řádku) – Výpočet.

Kontingenční tabulka sloupcově podmíněných relativních četností :

Kontingenční tabulka (znamky) Četnost označených buněk > 10 (Marginální součty nejsou označeny)						
	X	Y	Y	Y	Y	Rad.
		výboř	velmi do	dobř	nepros	souč
Cetnost	vyběrn	4	1	2	0	7
Sloupc. c		100,0	25,0	28,5	0,00	
Cetnost	velmi do	0	2	1	0	3
Sloupc. c		0,00	50,0	14,2	0,00	
Cetnost	dobře	0	0	1	1	2
Sloupc. c		0,00	0,00	14,2	20,0	
Cetnost	nepros	0	1	3	4	8
Sloupc. c		0,00	25,0	42,8	80,0	
Cetnost	vs. skup	4	4	7	5	21

Interpretace např. 4. řádku ve 2. sloupci: V souboru byli 4 studenti, kteří měli velmi dobře z angličtiny. Mezi nimi byl jeden, který neprospěl z matematiky, což představuje $1/4 = 25\%$.

Kontingenční tabulka řádkově podmíněných relativních četností:

Kontingenční tabulka (znamky) Četnost označených buněk > 10 (Marginální součty nejsou označeny)						
	X	Y	Y	Y	Y	Rad.
		výboř	velmi do	dobř	nepros	souč
Cetnost	vyběrn	4	1	2	0	7
Rádk. c		57,1	14,2	28,5	0,00	
Cetnost	velmi do	0	2	1	0	3
Rádk. c		0,00	66,6	33,3	0,00	
Cetnost	dobře	0	0	1	1	2
Rádk. c		0,00	0,00	50,0	50,0	
Cetnost	nepros	0	1	3	4	8
Rádk. c		0,00	12,5	37,5	50,0	
Cetnost	vs. skup	4	4	7	5	21

Interpretace např. 2. sloupce ve 4. řádku: V souboru bylo 8 studentů, kteří neprospěli z matematiky. Mezi nimi byl jeden, který měl velmi dobře z angličtiny, což představuje $1/8 = 12,5\%$.

Úkol 6.: Zápis dat zadaných pomocí absolutních četností

U 30 domácností byl zjišťován počet členů. 2 domácnosti měly 1 člen, 6 domácností 2 členy, 4 domácnosti 3 členy, 10 domácností 4 členy, 5 domácností 5 členů a 3 domácnosti měly 6 členů. Zapište tato data do systému STATISTICA, vytvořte variační řadu a zodpovězte následující otázky:

- Kolik procent domácností má právě 5 členů?
- Kolik procent domácností má nejvýše 5 členů?
- Kolik procent domácností má aspoň 5 členů?
- Kolik procent domácností má aspoň 2 a nejvýše 5 členů?

Návod: Vytvoříme nový datový soubor o 6 případech a dvou proměnných, jednu nazveme X a druhou četnost. Do proměnné X napíšeme čísla 1, ..., 6 (lze též vyplnit tak, že do Dlouhého jména této proměnné napíšeme =v0) a do proměnné četnost napíšeme čísla 2, 6, 4, 10, 5, 3.

Statistika – Základní statistiky/tabulky – Tabulky četností – Proměnné X – OK – klikneme na ikonu závaží – Proměnná vah četnost – OK – Stav Zapnuto – OK – Výpočet.

Dostaneme tabulku:

Kategorie	Četno	Kumula četnos	Rel.četn	Kumula rel.četn
1	2	2	6,66	6,66
2	6	8	20,00	26,66
3	4	12	13,33	40,00
4	11	23	33,33	73,33
5	5	28	16,66	90,00
6	3	31	10,00	100,00
ChD	0	31	0,00	100,00

Z této tabulky lze vyčíst odpovědi na všechny výše položené otázky.

Ad a) 16,7%, ad b) 90%, ad c) 26,7%, ad d) 83,3%

Úkol 7.: Zápis dat zadaných pomocí kontingenční tabulky

221 žáků 2. stupně základní školy bylo dotázáno, zda v současné době drží nějakou dietu.

Výsledky průzkumu jsou uvedeny v kontingenční tabulce. Znak X udává pohlaví žáka (1 – hoch, 2 – dívka), znak Y nabývá varianty 1, když respondent drží dietu a 2, když dietu nedrží.

X	Y		n _j
	drží dietu	nedrží dietu	
hoch	9	100	109
dívka	57	55	112
n _k	66	155	221

Zapište tato data do systému STATISTICA a zodpovězte následující otázky:

- Kolik procent hochů drží dietu?
- Kolik procent dívek drží dietu?
- Kolik procent těch žáků, kteří drží dietu, jsou hoši?
- Kolik procent těch žáků, kteří nedrží dietu, jsou dívky?

Návod: Vytvoříme nový datový soubor o třech proměnných X, Y, četnost a čtyřech případech. Do proměnné X napíšeme dvakrát pod sebe 1 (hoch) a dvakrát pod sebe 2 (dívka). Do proměnné Y napíšeme jedničku (1 – drží dietu) a dvojku (2 – nedrží dietu) a znovu jedničku a dvojku. D proměnné četnost napíšeme uvedené četnosti, tj. 9, 100, 57, 55. Popíšeme, co znamenají proměnné X, Y a co znamenají jejich varianty.

Výsledný soubor:

	1 X	2 Y	3 četno
1	hoch	drží die	9
2	hoch	nedrží c	100
3	dívka	drží die	57
4	dívka	nedrží c	55

Statistiky – Základní statistiky/tabulky – Kontingenční tabulky - Specif. tabulky – List 1 X, List 2 Y, OK, Váhy - četnost, Stav zapnuto – OK. Na záložce Možnosti zaškrtneme Procenta

z počtu v řádku a Procenta z počtu ve sloupci – Výpočet. Tím získáme kontingenční tabulku řádkově a sloupcově podmíněných relativních četností.

	X	Y drží di	Y nedrží d	Řad. souč
Četnost	hoch	9	10	19
Sloupc. c		13,6%	64,5%	
Řadk. če		8,2%	91,7%	
Četnost	divka	5	6	11
Sloupc. c		86,3%	35,4%	
Řadk. če		50,8%	49,1%	
Četnost	Vs.skl	14	8	22

Ad a) 8,3 % hochů drží dietu, ad b) 50,9 % dívek drží dietu, ad c) z těch žáků, kteří drží dietu, je 86,4 % dívek, ad d) z těch žáků, kteří nedrží dietu, je 35,5 % dívek.