

Téma 1.: Základní informace o systému STATISTICA, výpočet různých 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 8.0 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.

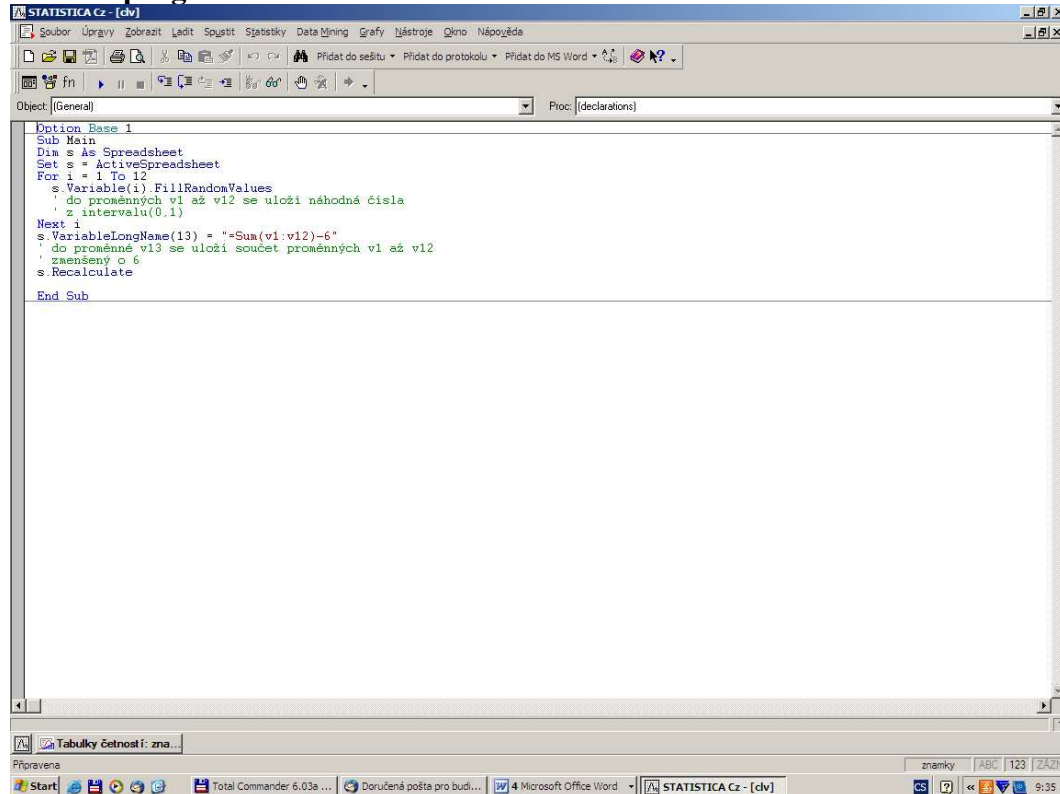
Ukázka pracovního sešitu:

Kategorie	Četnost	Kumulativní četnost	Rel. četnost	Kumulativní rel. četnost
vyborně	7	7	35.00000	35.0000
velmi dobře	3	10	15.00000	50.0000
dobře	2	12	10.00000	60.0000
neprospěl	8	20	40.00000	100.0000
ChD	0	20	0.00000	100.0000

- **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:

Ukázka programovacího okna:



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

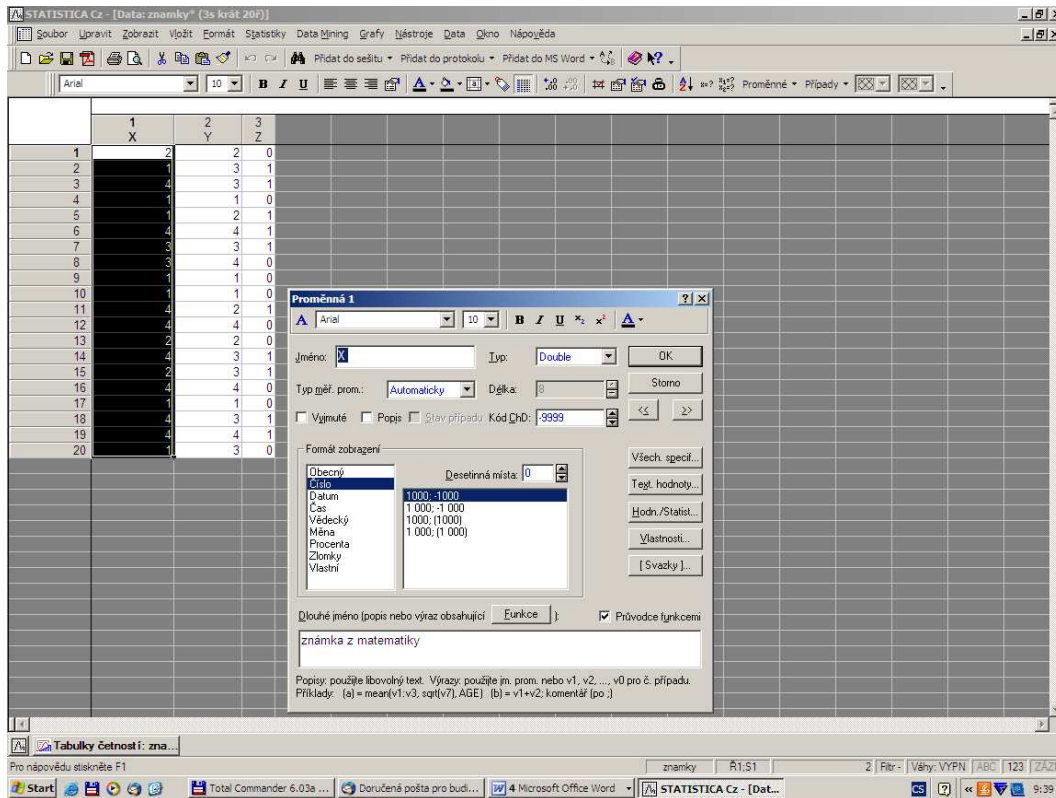
Ve verzi 8.0 je možno používat Projekty. Tento prvek umožňuje uložit soubor STATISTICA Projekt, a "zmrazit" tak doposud provedenou práci včetně umístění všech oken s objekty na obrazovce. Po otevření tohoto souboru se může přímo navázat na již provedenou část analýzy. Lze také uložit zaznamenané makro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

Vytvoření datového okna

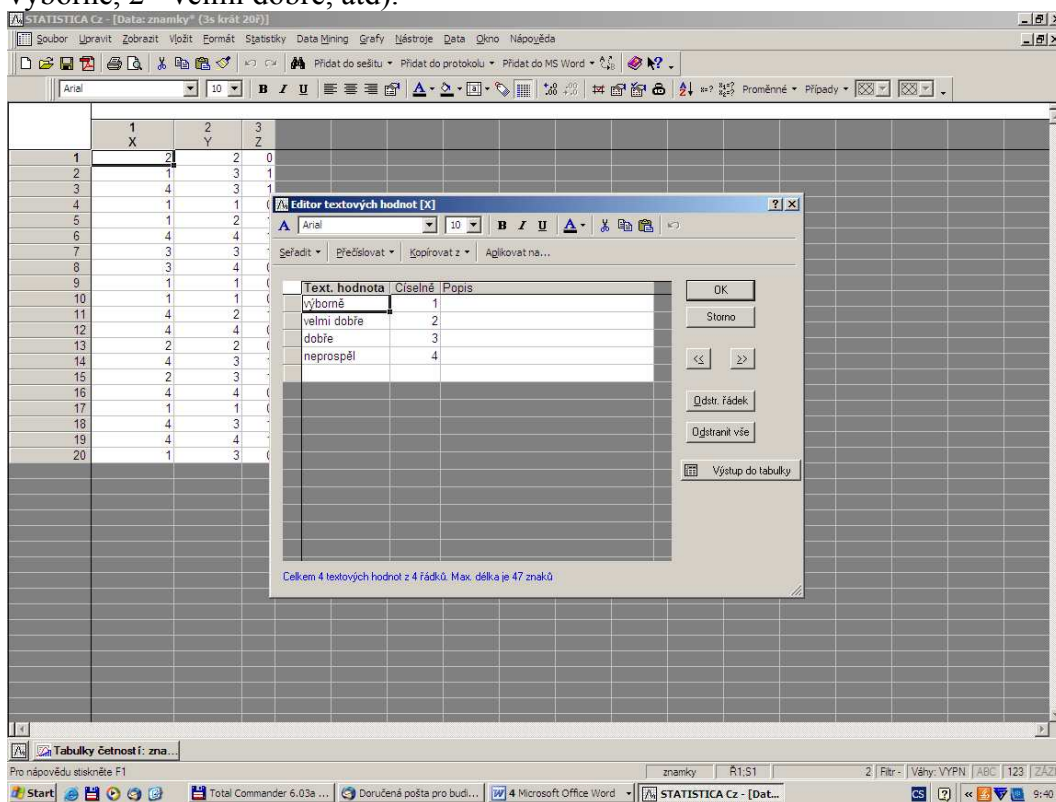
Po spuštění systému STATISTICA zavřeme všechna okna, která se objeví.

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: $=v3/v2*100$). Některé funkce jsou předem vloženy – položka „Funkce“.

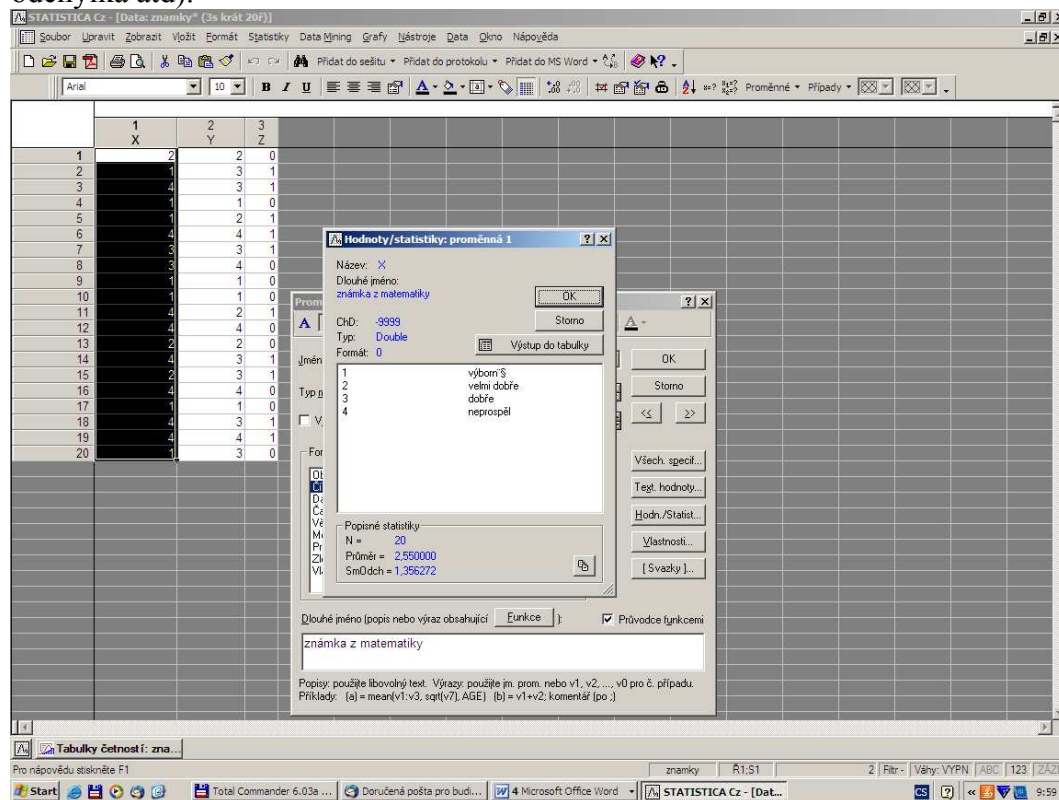


V položce „Textové hodnoty“ lze nastavit k variantám proměnné textový popis (např. 1 – výborně, 2 - velmi dobře, atd).



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.: Zapišeme do datového okna programu STATISTICA datový soubor, který bude obsahovat známky z matematiky, angličtiny a údaje o pohlaví dvaceti studentů.

Návod: Soubor – Nový – Počet proměnných 3, Počet případů 20, OK.

Úkol 2.: Proměnné nazveme X, Y, Z a vytvoříme jim návěští (X - známka z matematiky, Y - známka z angličtiny, Z - pohlaví studenta). Popíšeme, 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). Soubor uložíme pod názvem znamky.sta

Návod: 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 3.: Zjistíme absolutní a relativní četnosti úspěšných matematiků, angličtinářů a oboustranně úspěšných studentů.

Návod: Statistika – Základní statistiky a tabulky – Tabulky četností – OK – Proměnné X, Y – OK – Výpočet. Pro proměnnou X dostaneme tabulku:

Kategorie	Tabulka četností: X: známka z matematiky (znamky)			
	Četnost	Kumulativní četnost	Rel. četnost	Kumulativní rel. četnost
výborně	7	7	35,00000	35,0000
velmi dobře	3	10	15,00000	50,0000
dobře	2	12	10,00000	60,0000
neprospěl	8	20	40,00000	100,0000
ChD	0	20	0,00000	100,0000

Aby student uspěl, musí mít známku 1, 2 nebo 3. Ve sloupci Kumulativní četnost vidíme, že takových studentů bylo 12. Ve sloupci Kumulativní relativní četnost zjistíme, že úspěšných studentů z matematiky bylo 60%.

Analogicky zjistíme požadované údaje pro úspěšné angličtináře:

Kategorie	Tabulka četností: Y: známka z angličtiny (znamky)			
	Četnost	Kumulativní četnost	Rel. četnost	Kumulativní rel. četnost
výborně	4	4	20,00000	20,0000
velmi dobře	4	8	20,00000	40,0000
dobře	7	15	35,00000	75,0000
neprospěl	5	20	25,00000	100,0000
ChD	0	20	0,00000	100,0000

Úspěšných angličtinářů bylo 15, tj. 75%.

Výpočet absolutní četnosti oboustranně úspěšných studentů lze provést např. pomocí filtru. Statistika – Základní statistiky a tabulky – Tabulky četností – OK – Proměnné Z – OK, Select cases – Zapnout filtr – včetně případů – některé, vybrané pomocí výrazu $X \leq 3$ and $Y \leq 3$ – OK – Výpočet. Dostaneme tabulku

Kategorie	Tabulka četností: Z: pohlaví studenta (znamky) Zhrnout podmínku: $X \leq 3$ and $Y \leq 3$			
	Četnost	Kumulativní četnost	Rel. četnost	Kumulativní rel. četnost
žena	7	7	63,63636	63,6364
muž	4	11	36,36364	100,0000
ChD	0	11	0,00000	100,0000

Vidíme, že oboustranně úspěšných studentů je 11, tj. $11/20 = 55\%$.

Úkol 4.: Vypočteme podmíněnou relativní četnost úspěšných matematiků mezi úspěšnými angličtináři.

Návod: Statistika – Základní statistiky a tabulky – Tabulky četností – OK – Proměnné X - OK, Select cases – Zapnout filtr – včetně případů – některé, vybrané pomocí výrazu $Y \leq 3$ – OK – Výpočet. Dostaneme tabulku

Kategorie	Tabulka četností: X: známka z matematiky (znamky) Zhrnout podmínku: Y<=3			
	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
výborně	7	7	46,66667	46,6667
velmi dobře	3	10	20,00000	66,6667
dobře	1	11	6,66667	73,3333
neprospěl	4	15	26,66667	100,0000
ChD	0	15	0,00000	100,0000

Ve sloupci Kumulativní četnost vidíme, že úspěšných matematiků mezi úspěšnými angličtináři bylo 11. Ve sloupci Kumulativní relativní četnost zjistíme, že to odpovídá 73,3%.

Úkol 5.: Utvoříme uspořádaný datový soubor známek z matematiky.

Návod: Kurzor nastavíme na X – Data – Seřadit – Proměnná X – Směr Vzestupně, Seřadit Číselně, zaškrtneme Vytvořit novou tabulku – OK. Ve vytvořeném datovém okně odstraníme proměnné Y, Z (Proměnné – Odstranit – Od proměnné Y Do proměnné Z – OK). Upravíme formát proměnné X (Formát – Proměnné – Autom. přízp. velikost).

Úkol 6.: Najdeme vektor variant známek z matematiky.

Návod: Úkol splníme např. tak, že v nastavení vlastností proměnné vybereme Hodn/Statist a dostaneme tabulku, v níž budou varianty 1, 2, 3, 4.

Úkol 7.: Pro datový soubor s údaji o známkách najdeme relativní četnost

- matematických jedničkářů
- oboustranně neúspěšných studentů.

Návod:

ad a) V tabulce, kterou jsme vytvořili v úkolu 3, najdeme tento údaj na 1. řádku, ve sloupci Relativní četnost:

Kategorie	Tabulka četností: X: známka z matematiky (znamky)			
	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
výborně	7	7	35,00000	35,0000
velmi dobře	3	10	15,00000	50,0000
dobře	2	12	10,00000	60,0000
neprospěl	8	20	40,00000	100,0000
ChD	0	20	0,00000	100,0000

Matematických jedničkářů tedy bylo 35%.

ad b) Výpočet absolutní četnosti oboustranně neúspěšných studentů lze provést např. pomocí filtru.

Statistiky – Základní statistiky a tabulky – Tabulky četností – OK – Proměnné Z – OK, Select cases – Zapnout filtr – včetně případů – některé, vybrané pomocí výrazu X>3 and Y>3 – OK – Výpočet. Dostaneme tabulku

Tabulka četností: Z: pohlaví studenta (znamky)				
Zhrnout podmínku: $X > 3$ and $Y > 3$				
Kategorie	Četnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
žena	2	2	50,00000	50,0000
muž	2	4	50,00000	100,0000
ChD	0	4	0,00000	100,0000

Vidíme, že oboustranně neúspěšní studenti jsou 4, tj. $4/20 = 20\%$.