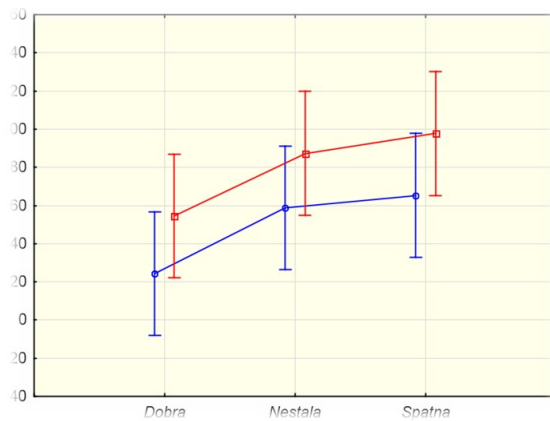
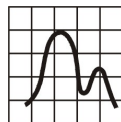


Manuál k ovládání programu *STATISTICA*



Mgr. Petra Beranová
Mgr. Lenka Blažková
Ing. Miloš Uldrich



StatSoft®

Copyright © StatSoft CR s.r.o. 2012

StatSoft CR s.r.o.
Ringhofferova 115/1
155 21 Praha 5 – Zličín
tel.: +420 233 325 006 • fax: +420 233 324 005 • e-mail: info@statsoft.cz • www.statsoft.cz

Všechna práva vyhrazena.

Kopírování, rozmnožování, publikování nebo přenos jakékoli části této publikace elektronickou, mechanickou, magnetickou, optickou, fotografickou nebo jakoukoli jinou cestou je zakázán bez písemné dohody se StatSoft CR s.r.o.

StatSoft, StatSoft logo, *STATISTICA*, *Data Miner*, SEPATH a GTrees jsou ochranné známky společnosti StatSoft, Inc. a jsou použity se souhlasem této společnosti. Další použité materiály mohou být chráněny právy k duševnímu vlastnictví jiných subjektů.

Obsah:

<i>Obsah:</i>	1
1 <i>Úvod</i>	3
2 <i>Spuštění programu STATISTICA</i>	3
3 <i>Načtení souboru</i>	6
<i>Příklad – import dat z Excelu</i>	6
4 <i>Vytvoření základní výpočtů – popisné statistiky</i>	8
<i>Rozdělení spojitě proměnné dle kategorie</i>	10
5 <i>Vytvoření grafu</i>	13
5.1 <i>Histogram</i>	13
5.2 <i>Krabicový graf (Box Plot)</i>	13
5.3 <i>Regulační diagram</i>	15
6 <i>Uložení práce</i>	16
6.1 <i>Uložení celé rozdělané práce</i>	16
6.2 <i>Uložení tabulky v softwaru</i>	17
6.3 <i>Uložení grafu</i>	19
6.4 <i>Přidání výstupů do Microsoft Word</i>	20
7 <i>Další možnosti načtení souborů</i>	21
7.1 <i>Otevření textového souboru</i>	24
8 <i>Správce výstupů</i>	27
8.1 <i>Výstup do Microsoft Word / do protokolu STATISTICA</i>	27
9 <i>Další příklady analýzy dat</i>	29
9.1 <i>Příklad – výpočet popisných statistik</i>	29
9.2 <i>Analýza způsobilosti procesu</i>	32
<i>Indexy způsobilosti procesu</i>	37
10 <i>Ověření normality v softwaru STATISTICA</i>	43
11 <i>Jednovýběrový t test</i>	50
12 <i>Připojení do databází pomocí STATISTICA Query</i>	51
13 <i>Úprava načtených dat</i>	54
<i>Proměnné a případy</i>	54

<i>Transformace dat</i>	55
<i>Použití filtru</i>	56
14 <i>Další grafické možnosti softwaru STATISTICA</i>	57
<i>Styly grafů</i>	59
<i>Přidání stylů pro grafické objekty</i>	61
<i>Automatické nastavení vzhledu grafu</i>	62
15 <i>Automatizace rutinních analýz</i>	64
16 <i>Vlastní menu v STATISTICA</i>	65
17 <i>Závěr</i>	69

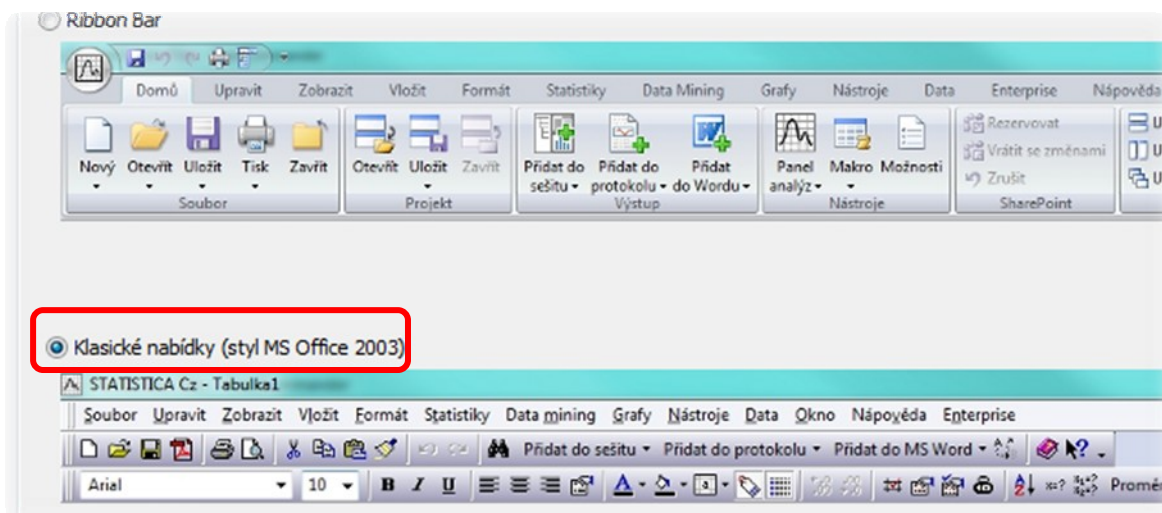
1 Úvod

Cílem manuálu je seznámení se základními možnostmi ovládní programu *STATISTICA* a rychle se zorientovat.

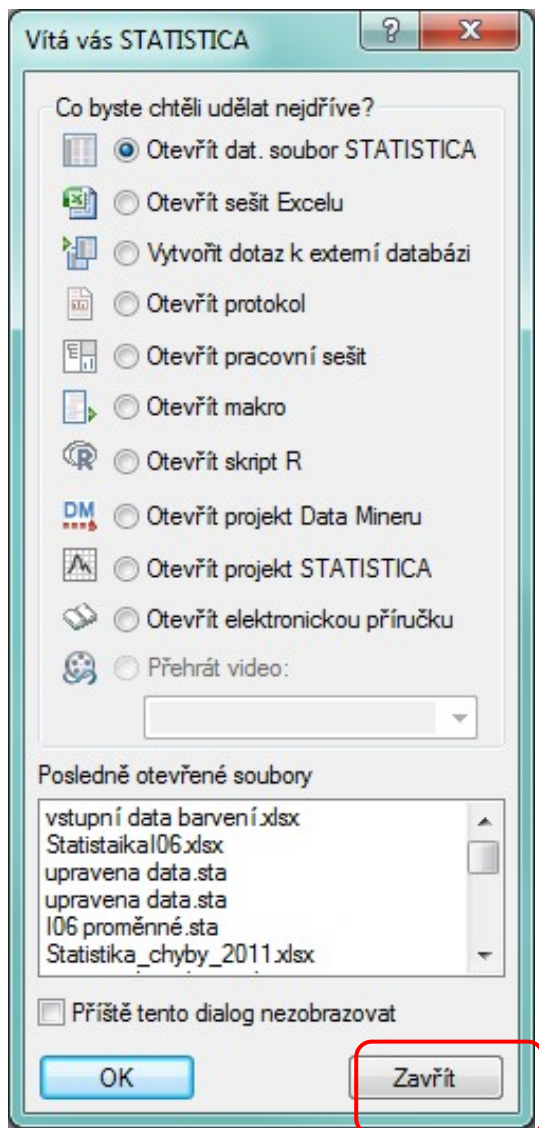


2 Spuštění programu *STATISTICA*

Při prvním spuštění nám dá program vybrat mezi 2 typy menu:



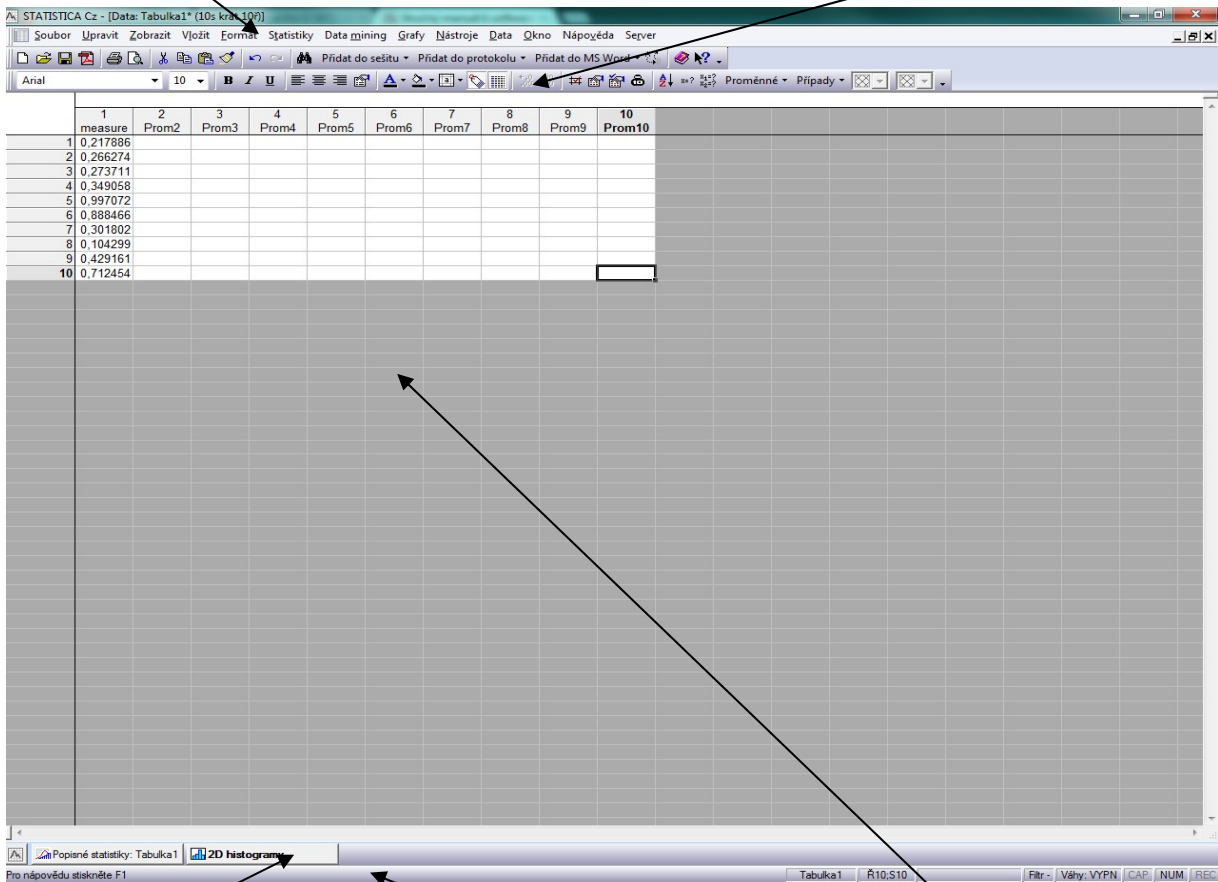
Po potvrzení **OK Klasické nabídky** se obrazovce se objeví rychlá navigace, kterou zavřeme a



máme zde okno aplikace *STATISTICA*:

základní nabídka

panely nástrojů s tlačítky



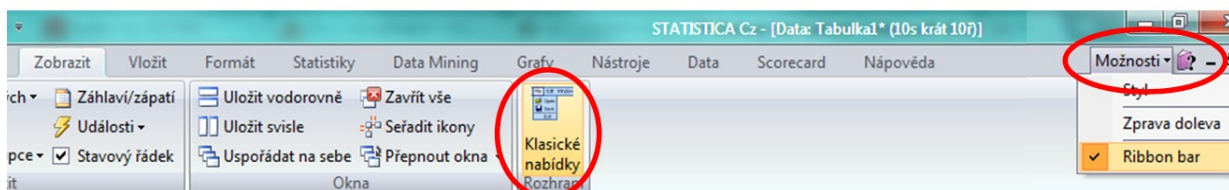
panel analýz

stavový řádek

pracovní plocha s dokumentem

- *základní nabídka* - slouží k ovládání systému, zpřístupňuje všechny nástroje programu
- *panely nástrojů s tlačítky* - jednodušší přístup k různým příkazům
- *panel analýz* - zde minimalizována okna všech spuštěných analýz, mezi kterými se lze přepínat
- *stavový řádek* - podává zkrácenou nápovědu a základní informace o aktivním dokumentu. Můžeme odtud např. ovládat filtry či váhy.

Software *STATISTICA* umožňuje práci v zobrazení *Ribbon bar*, přepnutí do klasického zobrazení provedete přes záložku *Možnosti* v pravém horním rohu, nebo přes záložku *Zobrazit*.



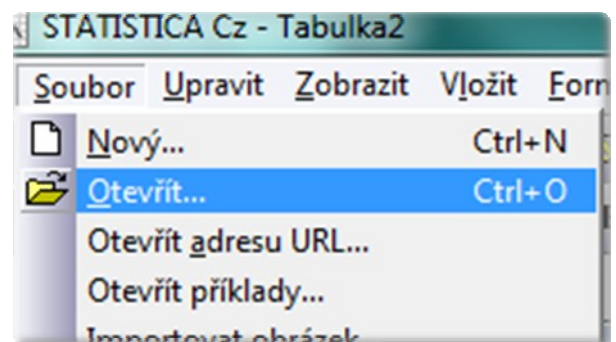
3 Načtení souboru

Data pro vlastní analýzu můžeme získat několika způsoby:

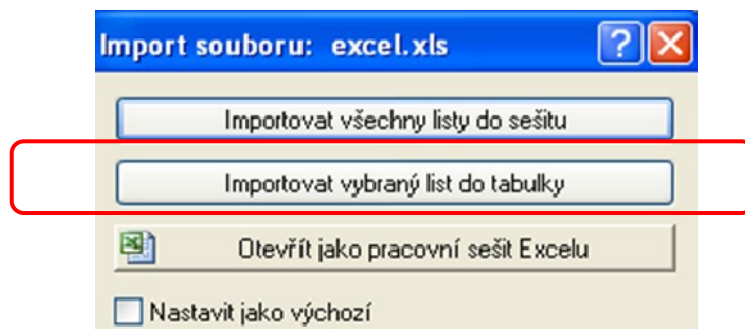
- *importem již uložených souborů různých formátů*
- *připojením k databázi* – pomocí SQL dotazů lze pracovat s daty uloženými například v databázi Oracle, MS SQL Server, Sybase atd.
- *otevřením tabulky Microsoft Excel v programu STATISTICA bez importu*
- *vložení dat do nové tabulky v programu STATISTICA*
- *sběrem dat on-line* - pokud je systém napojen na měřicí zařízení, naměřené hodnoty se dají ihned zpracovávat.

Příklad – import dat z Excelu

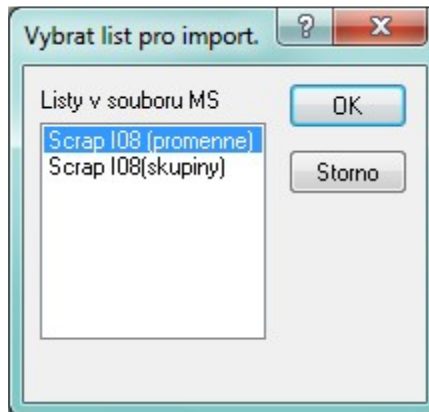
V menu **Soubor** a možnost **Otevřít** vybereme soubor **ScapI08.xlsx**



Při otvírání „Excelovských“ tabulek mámě několik možností, jak k tabulkám přistupovat:

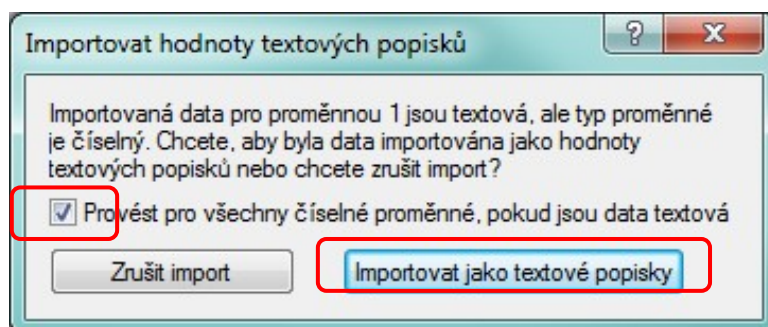


- **Importovat vybraný list do tabulky** – nejčastější možnost – pokud máme více listů, tak upřesníme list, který chceme importovat – **vybereme Scrap I08 (proměnné)**



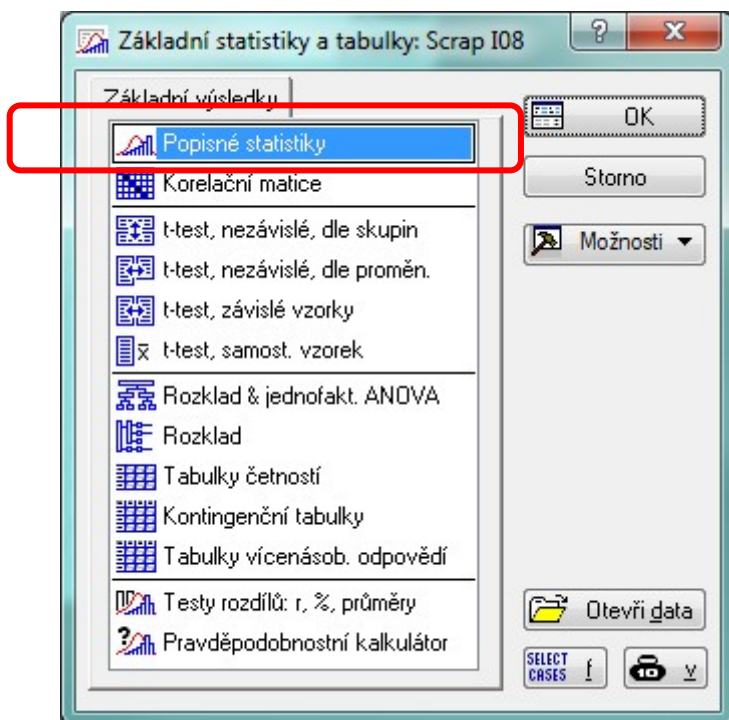
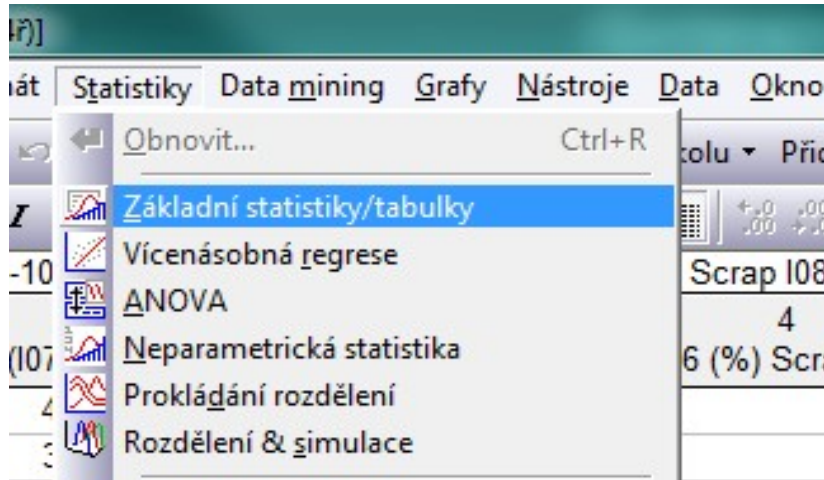
Data jsou načtena do tabulky softwaru *STATISTICA* (*.sta) stejně jako v případě načítání dat z textových souborů.

Pokud mám v původním souboru textové popisky, tak mě *STATISTICA* upozorní, zaškrtneme a pokračujeme:

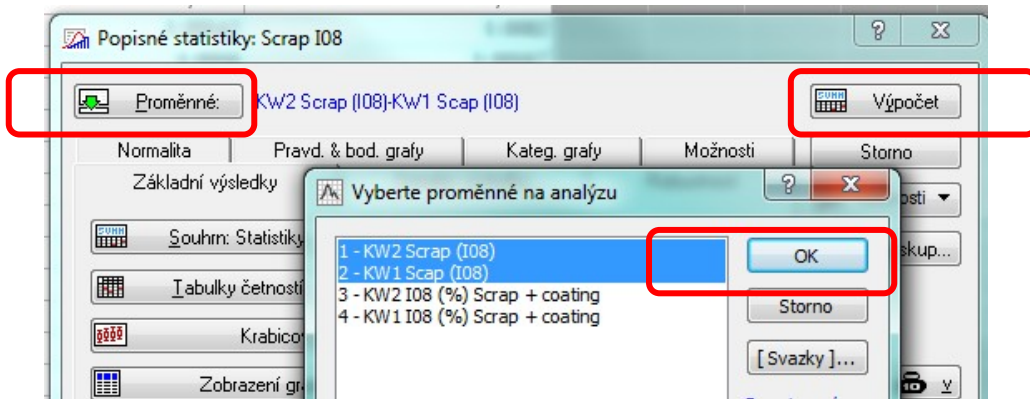


4 Vytvoření základní výpočtů – popisné statistiky

Přes záložku *Statistiky* -> *Základní statistiky/tabulky* -> *Popisné statistiky*



Vybereme proměnné (pro více proměnných držím při pokliknutí **Ctrl**) – vybrali jsme proměnné, které reprezentují počet vad na součástce (Scrap).

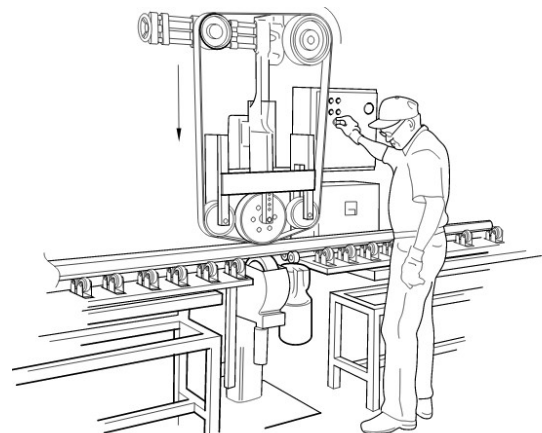


Máme zde tabulku základních statistických charakteristik.

Proměnná	N platných	Průměr	Minimum	Maximum	Sm.odch.
KW2 Scrap (I08)	24	25,54167	13,00000	40,00000	9,11272
KW1 Scrap (I08)	20	32,40000	16,00000	60,00000	10,72822

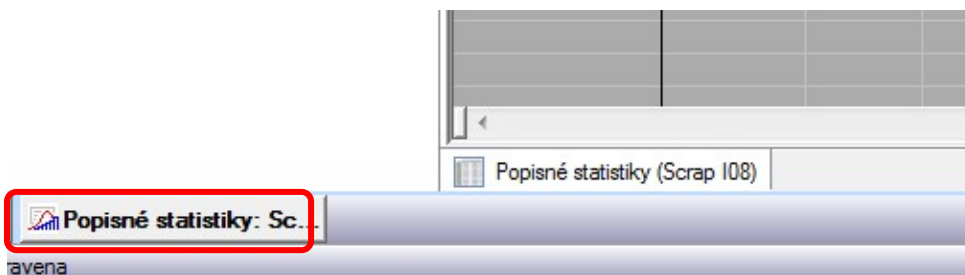
V souboru jsou naměřeny chyby na konkrétních vyrobených dílech (sklíčka), díly vyrobil jeden stroj.

- KW1 reprezentuje období před vyčištěním stroje
- období KW2 reprezentuje počet chyb na dílu, který byl vyroben po vyčištění stroje.

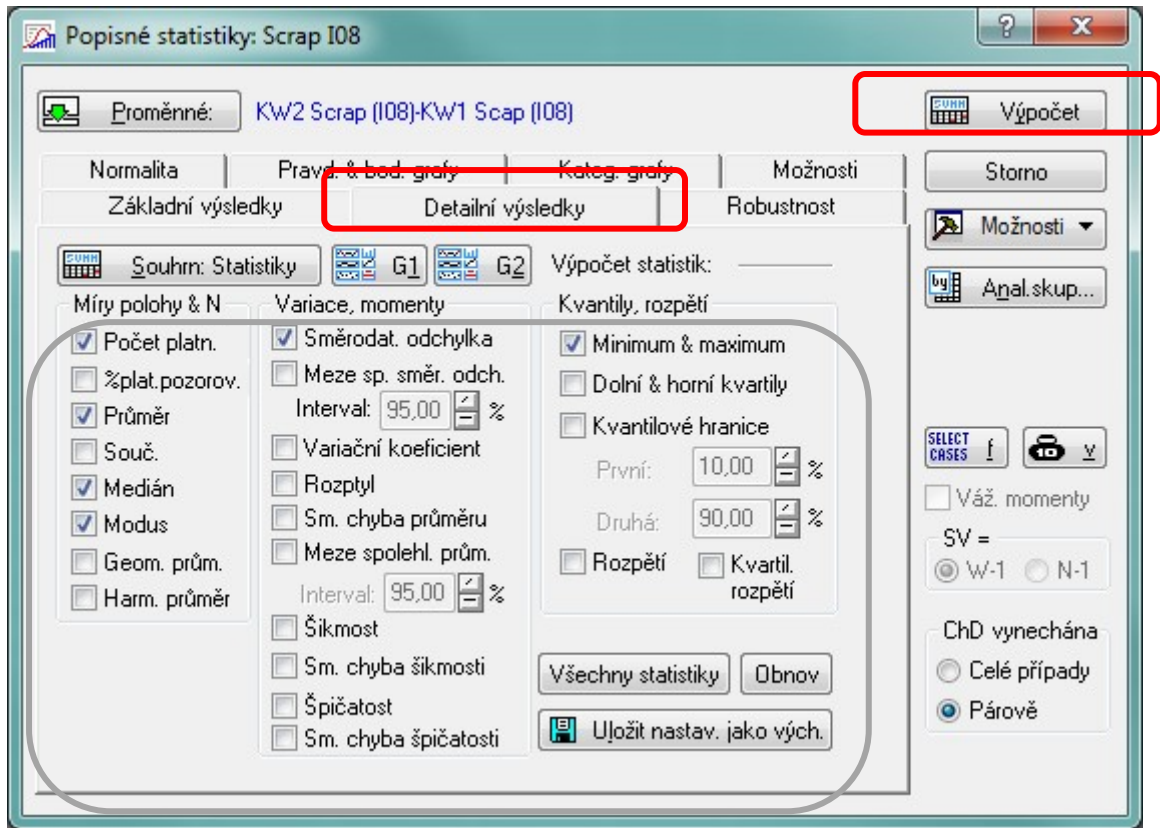


- Průměrný počet chyb na součástce se po vyčištění stroje snížil z **32** na **25**.
- Počet chyb také o něco méně kolísá, což reprezentuje menší směrodatná odchylka **9,1** oproti **10,78**.

Vyvoláme dole panel analýz a přepneme na kartu details:



Kde si lze v případě potřeby zaškrtnout další popisné statistiky:



Rozdělení spojitě proměnné dle kategorie

Vstupy pro analýzy máme často v této formě:

Měřená proměnná v jednom sloupci, příslušnost ke konkrétní skupině (dle pohlaví, vzdělání apod.) je potom ve sloupci druhém jako kategorie (faktor). Tento vstup je často vyžadován samotnou analýzou (dle skupin):

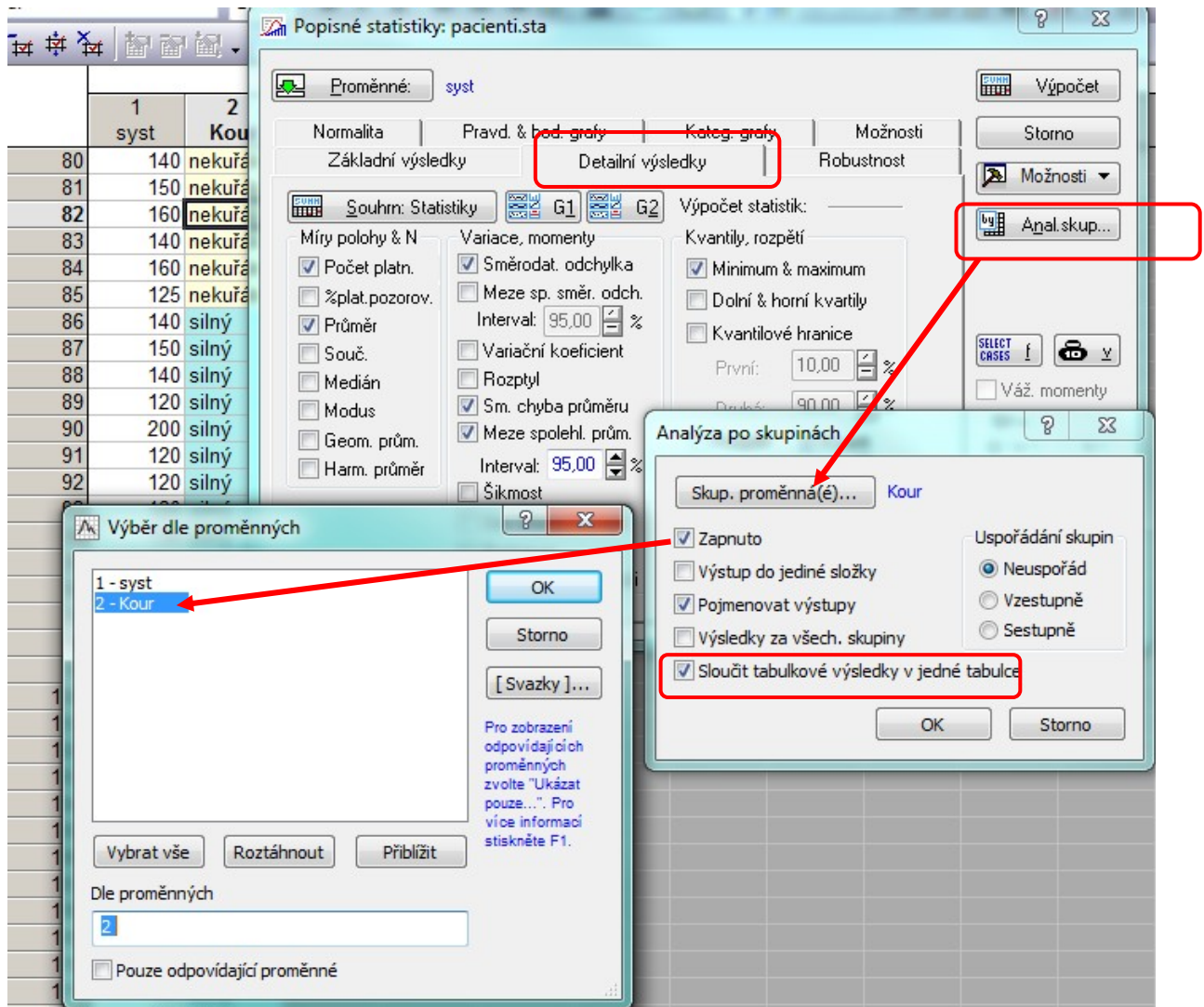
	1 Q _{Ab} (collagen)	2 Group
1	2,09496	1
2	2,02952	1
3	2,02856	1
4	2,06664	1
5	2,04344	1
6	2,03312	1
7	2,039264	1
8	2,08584	1
9	2,09984	1
10	2,07552	1
11	2,1592	1
12	2,12704	1
13	2,11264	1
14	2,10432	1
15	2,09984	2
16	2,07256	2
17	2,08584	2

 t-test, nezávislé, dle skupin

 t-test, nezávislé, dle proměn.

Dialog níže ukazuje, jak rozdělit výpočty popisných statistik podle kategoričké proměnné.

Na kartě **Detailní výsledky** v dialogu **Popisné statistiky: Statistika** → **Základní statistiky/tabulky** → **Popisné statistiky** → **anal. Skupiny**:



Kde vybereme proměnnou stupeň kouření (kouř) a charakteristiky polohy a variability tak vypočteme zvlášť pro jednotlivé kategorie.

Kompletní řešené příklady na char. variability a polohy, které ukážou další možnosti softwaru *STATISTICA* v této oblasti lze najít v našich newsletterech:

- Newsletter 20/08/2012
- Newsletter 17/09/2012
- Newsletter 15/10/2012

<http://www.statsoft.cz/o-firme/archiv-newsletteru/>

The screenshot shows the StatSoft website interface. At the top left is the StatSoft logo. In the top right, there is a navigation menu with three items: 'O firmě' (highlighted with a red box), 'Řešení', and 'Produkty'. Below the navigation is a banner with the text 'Komplexní analytické a reportovací řešení' and an illustration of three figures pushing a large yellow arrow upwards. A search bar and a 'DOMŮ' button are also visible.

On the left side, there is a vertical menu with several items, including 'Archiv newsletterů' (highlighted with a red box). The main content area is titled 'Archiv newsletterů' and contains the following text:

Vážení obchodní přátelé,
 v sekci, ve které se právě nacházíte je umístěn archiv newsletterů s
 médium, jehož aktuální podobu vám zasiláme pravidelně každý měsíc prost
 připravujeme ve dvou variantách.

Novinky ze StatSoftu - v tomto newsletteru najdete konkrétní příklady z
 studií, články obecnějšího charakteru, které však dávají konkrétní návod n
 podnikání. Také zde naleznete informace o současných i připravovaných k
 StatSoft CR.

StatSoft ACADEMY - tento newsletter se snažíme dělat tak, aby byl svým
 slova newsletter co nejvíce vzdálen. Je zaměřen zejména na vzdělávání
 informace z různých oblastí statistiky, praktické návody, vysvětlení slo
 formou atd. V tomto pravidelném měsíčním okénku zajímavostí pro vás
 zajímavé ceny.

K odběru newsletterů se můžete přihlásit zasláním mailu na newsletter@statsoft.com

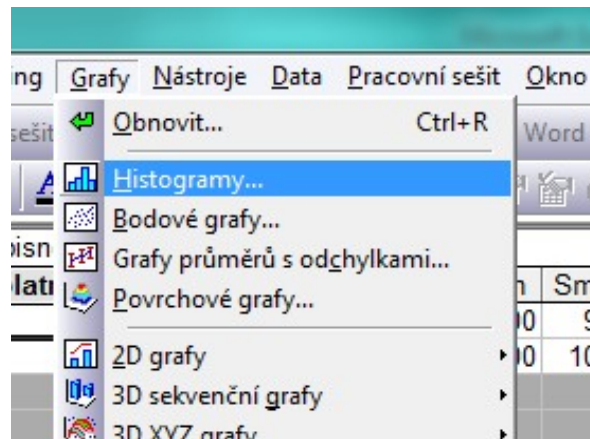
Aktuální i starší newslettery jsou k dispozici na následujících odkazech

- [05/02/2013 StatSoft ACADEMY](#)
- [08/01/2013 StatSoft ACADEMY](#)
- [10/12/2012 StatSoft ACADEMY](#)
- [26/11/2012 Novinky ze StatSoftu](#)
- [12/11/2012 StatSoft ACADEMY](#)

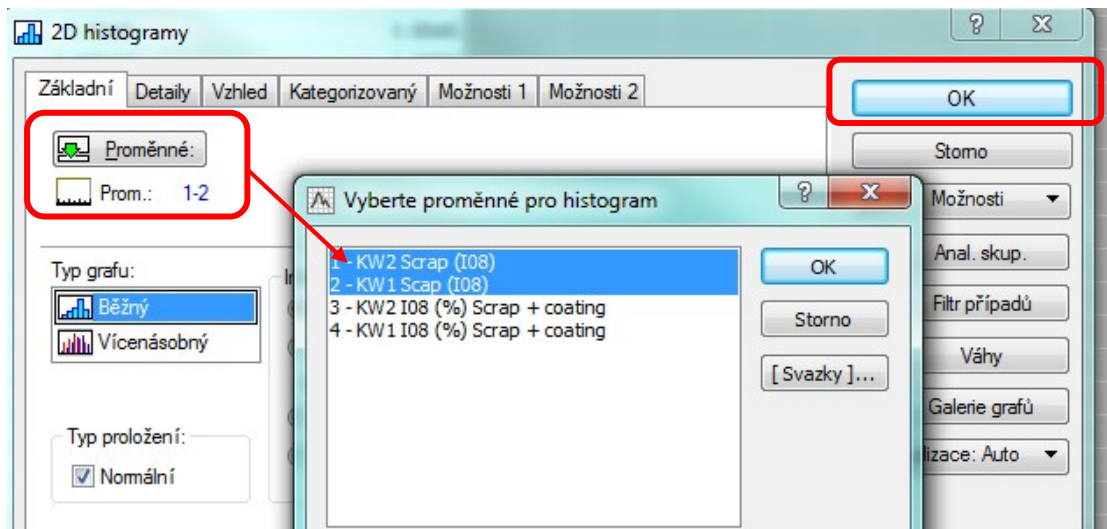
5 Vytvoření grafu

5.1 Histogram

Přes **Grafy** -> **2D grafy** -> **Histogramy**



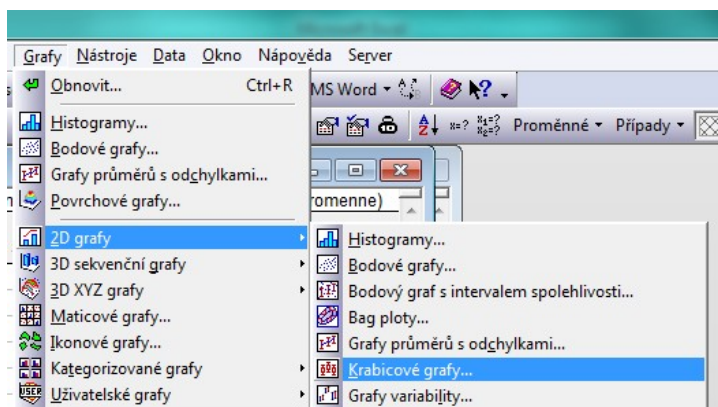
Vybereme proměnné pro obě období:



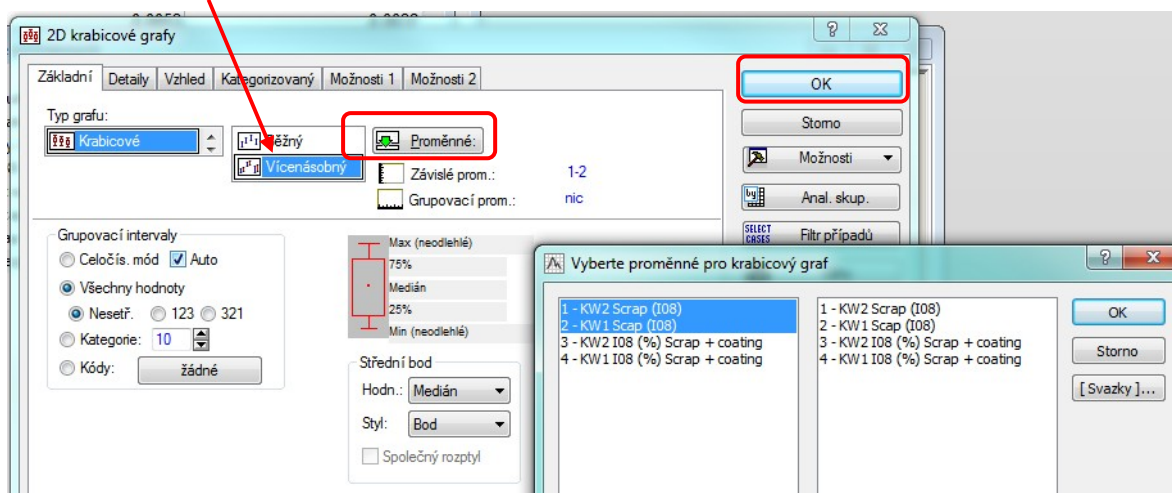
5.2 Krabicový graf (Box Plot)

Tímto grafem si vizuálně porovnáme oba naše vzorky, tedy před vyčištěním a po vyčištění stroje:

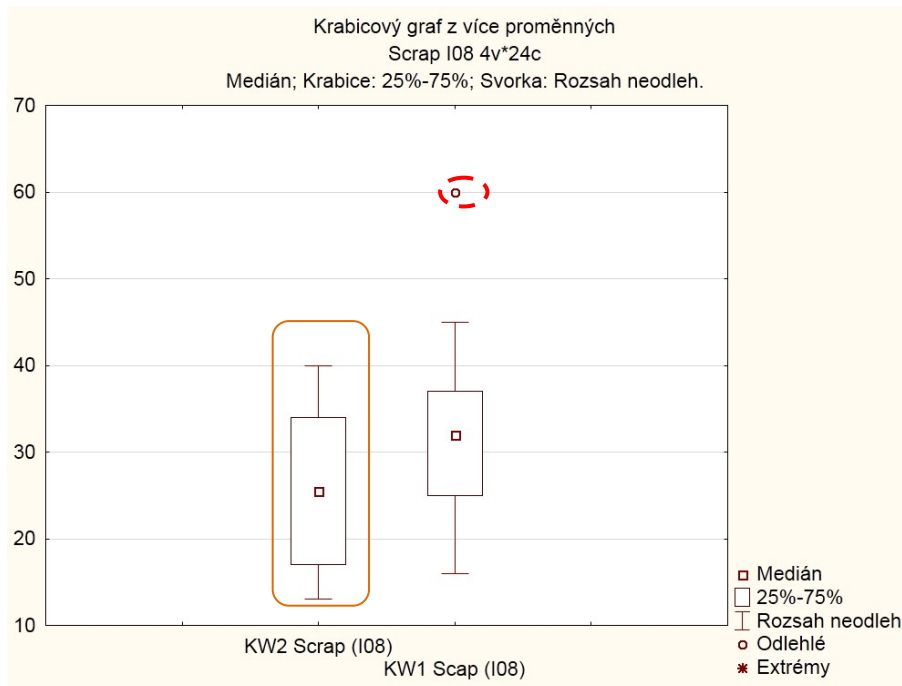
Přes **Grafy** -> **2D grafy** -> **Krabicové grafy**



Zvolíme **Vícenásobný** a opět vybereme proměnnou:



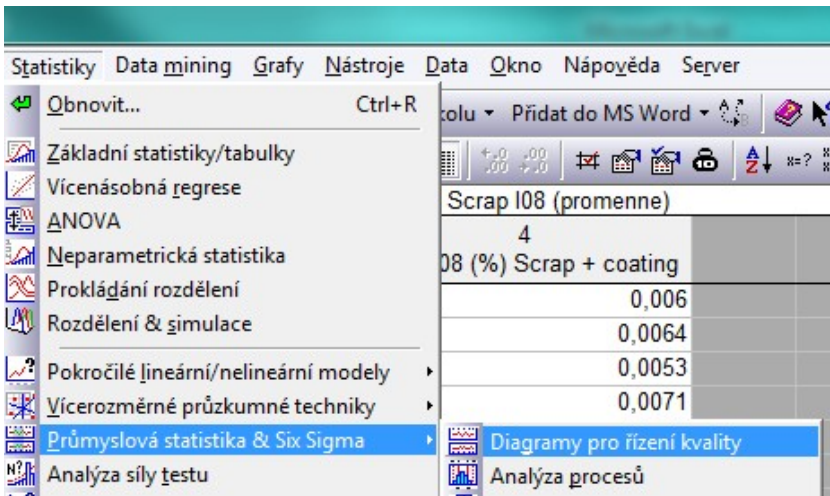
Z grafu je vidět, že v období po vyčištění stroje (**KW2**) došlo k celkovému poklesu variability souboru (krabice je niž):



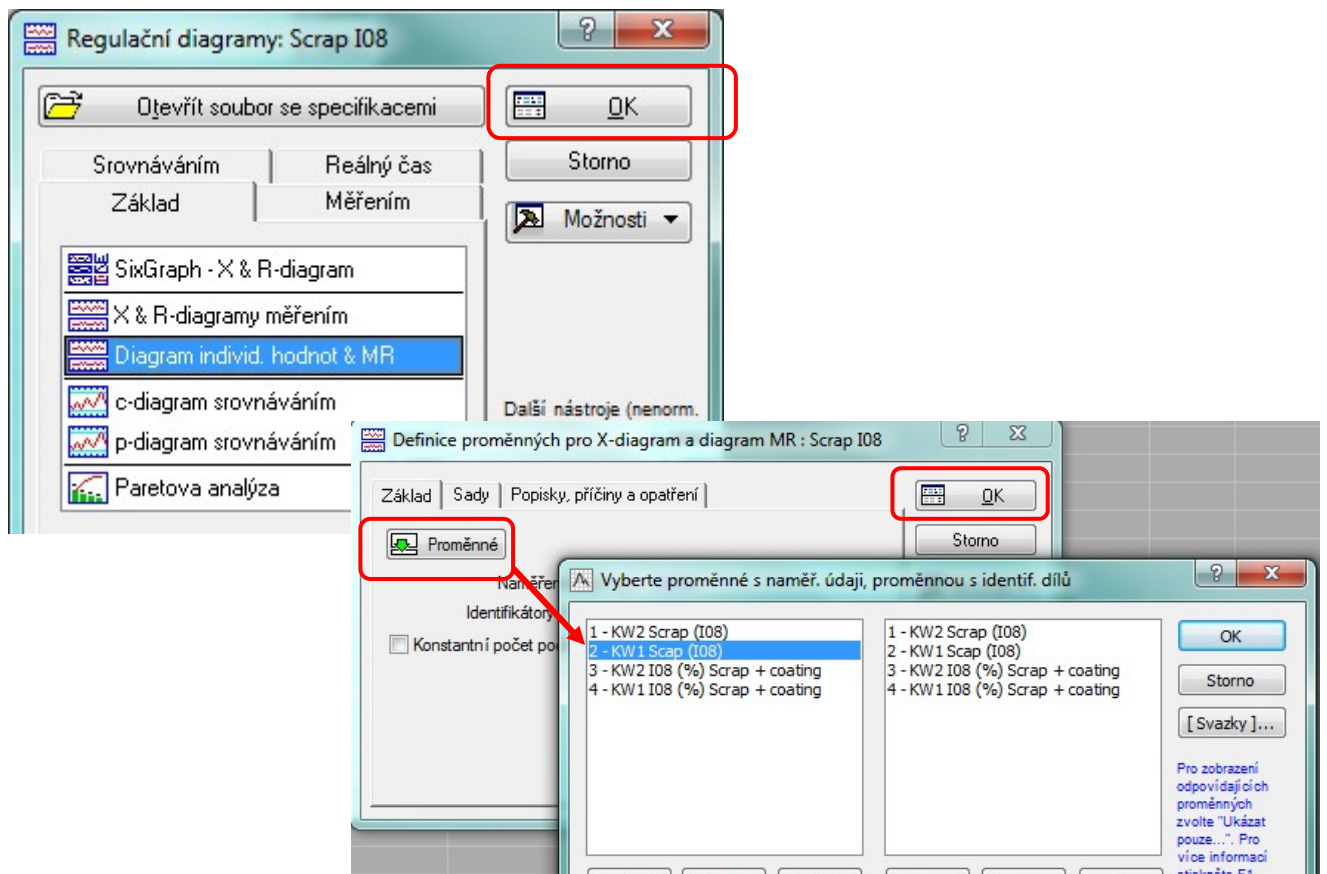
V souboru je jedno **odlehle pozorování**, které bylo naměřené v období KW1, je třeba zkontrolovat, jestli nejde o chybnou hodnotu operátora.

5.3 Regulační diagram

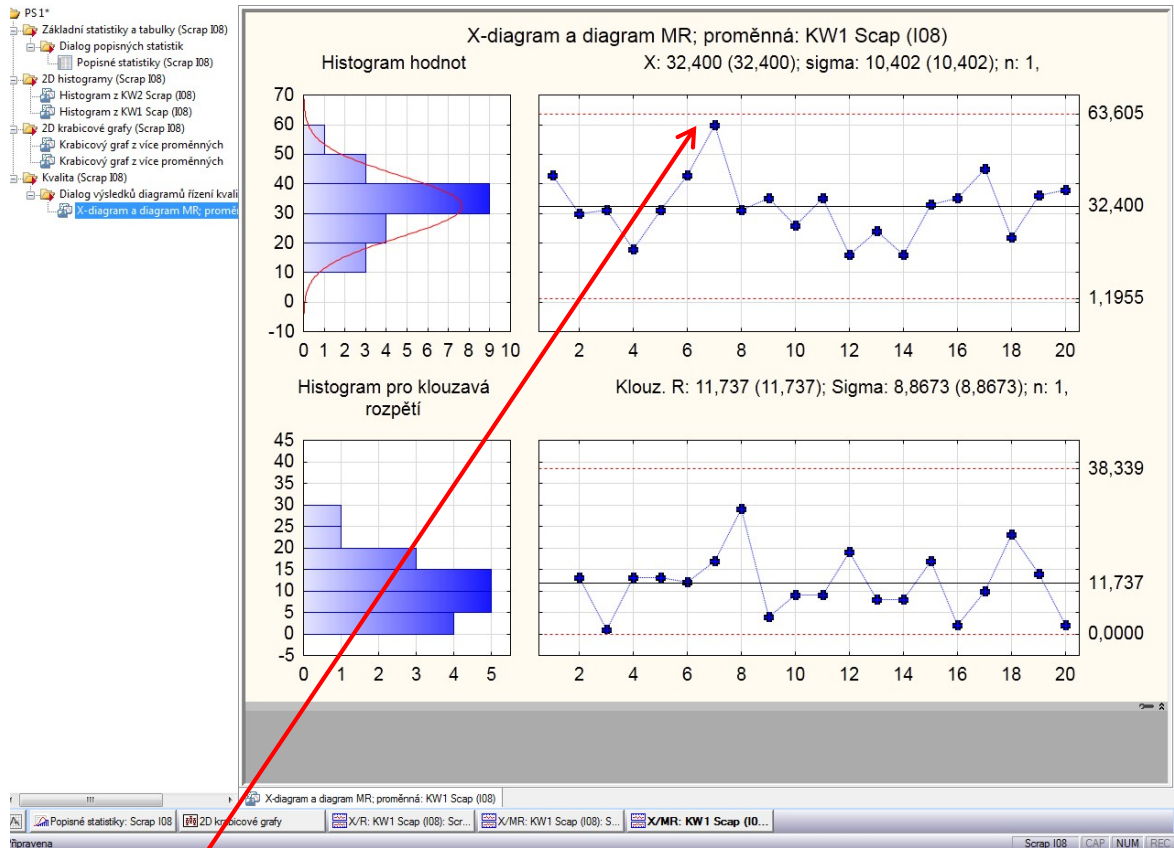
Pojďme se na proměnnou KW1 Scap (I08) podívat na regulačním diagramu, přes *Statistiky -> Průmyslová statistika -> Diagramy pro řízení kvality*



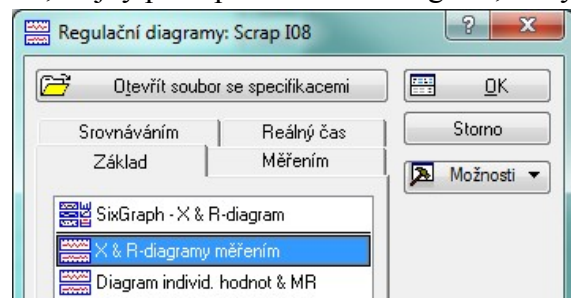
Vybereme diagram individuálních hodnot:



A získáme následující diagram:



Vidíme, že hodnota 60 nepřekročila meze diagramu, stejným postupem zvolíme diagram, který sleduje meze průměrných hodnot:

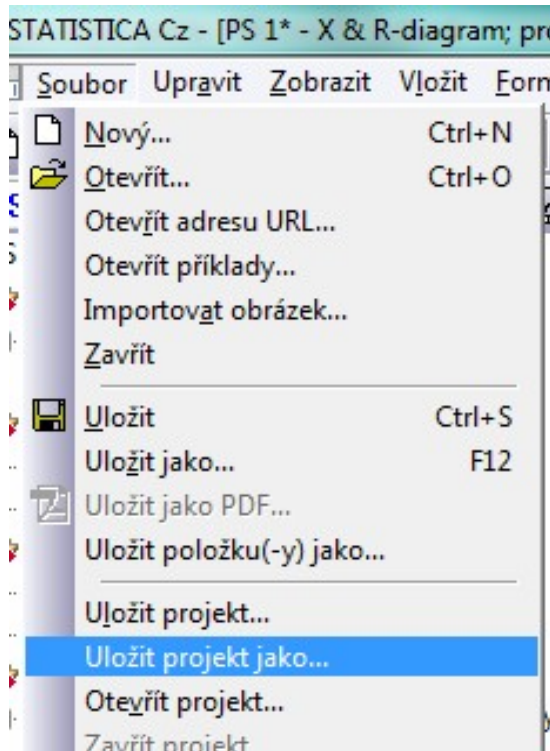


6 Uložení práce

6.1 Uložení celé rozdělané práce

Výstupy v souboru lze ukládat několika způsoby, začneme projektem, který je dobré použít, pokud chci uložit kompletní rozdělanou práci v softwaru *STATISTICA*:

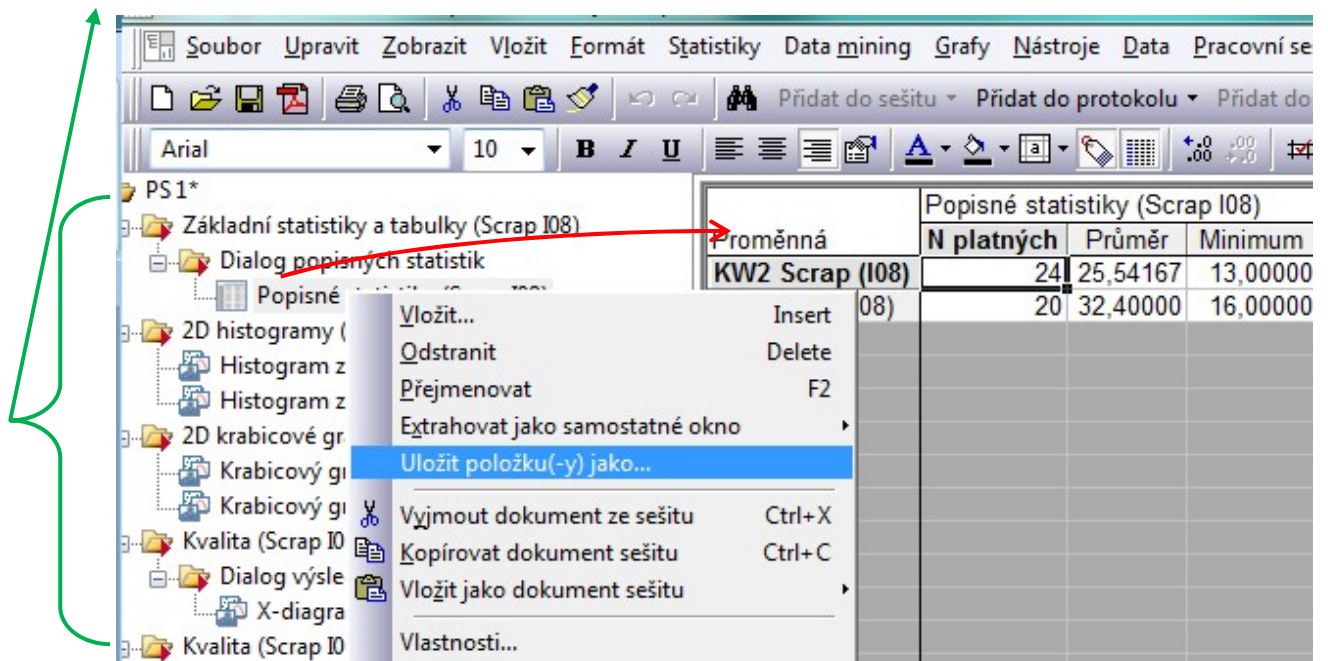
Přes *Soubor* -> *Uložit projekt jako...*



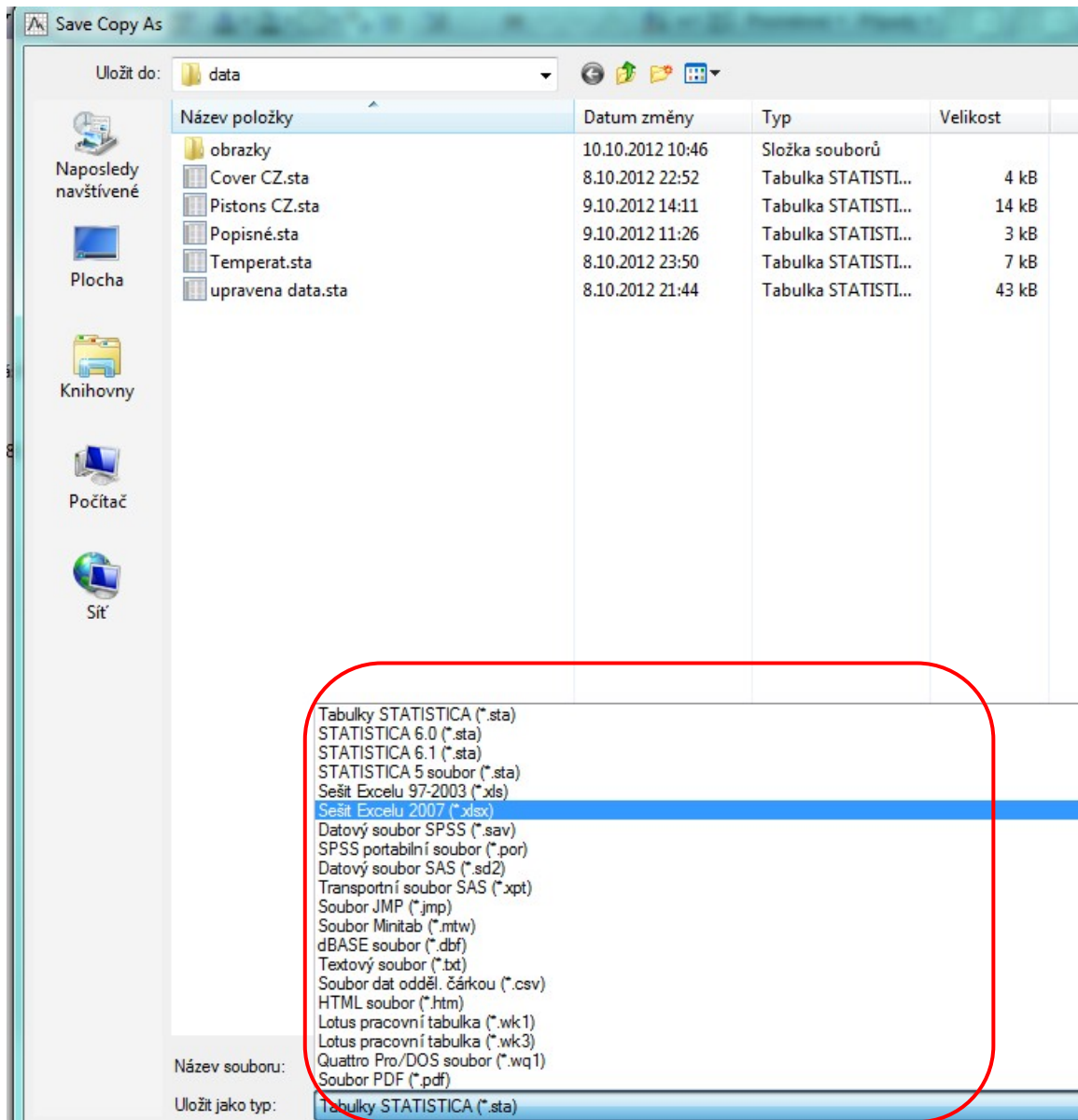
Uložíme soubor, ve kterém je všechno, co máme právě v softwaru otevřené. Tento soubor následně otevřeme přes **Soubor** -> **Otevřít** nebo dvojklik přímo na soubor.

6.2 Uložení tabulky v softwaru

– ve **stromu sešitu STATISTICA** klikneme přes pravé tlačítko na tabulku Popisných statistik



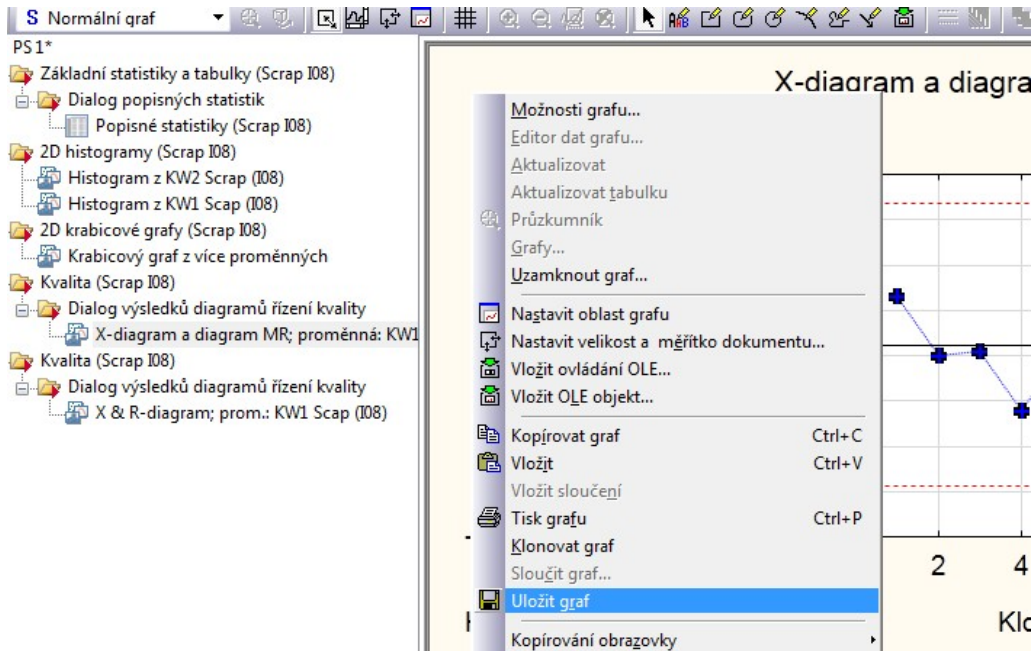
a zvolíme **Ulož položku(-y) jako...**



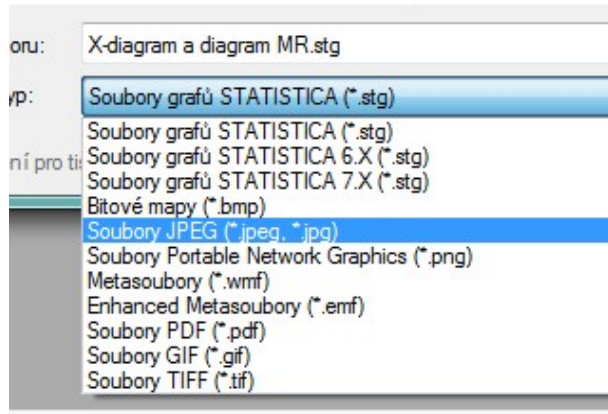
A tabulku si uložíme třeba ve formátu Excelu.

6.3 Uložení grafu

V příslušném grafu kliknu pravím tlačítkem

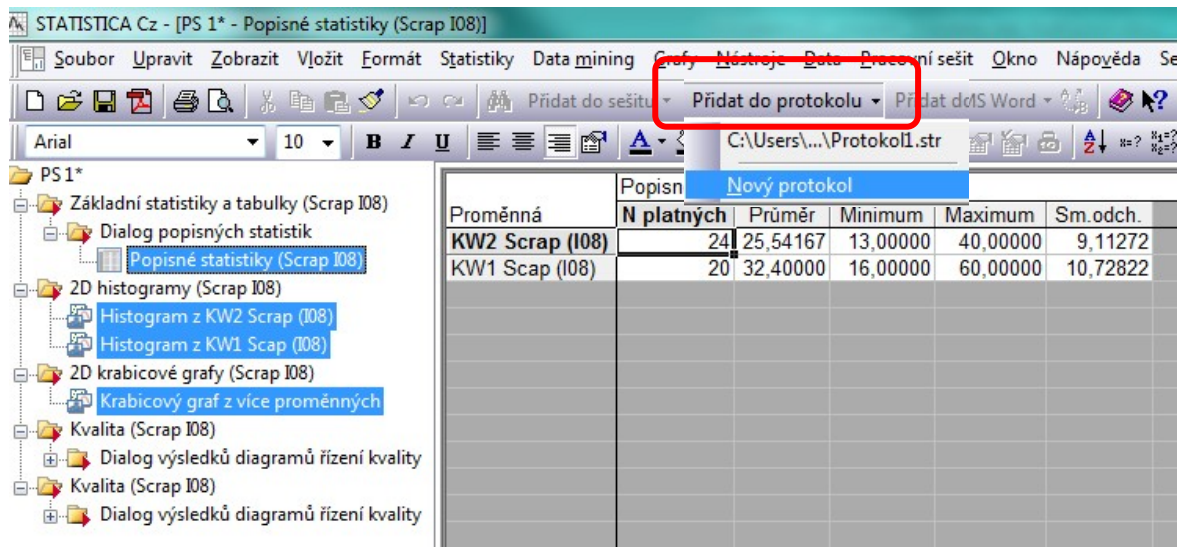


A opět vyberu formát pro uložení:

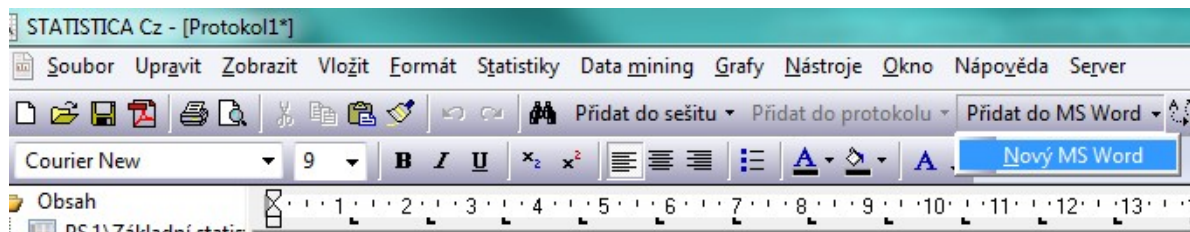


6.4 Přidání výstupů do Microsoft Word

Nejprve označíme (při stisknutém **Ctrl**) výstupy, které chceme přidat do MS Word a klikneme na Přidat do protokolu:



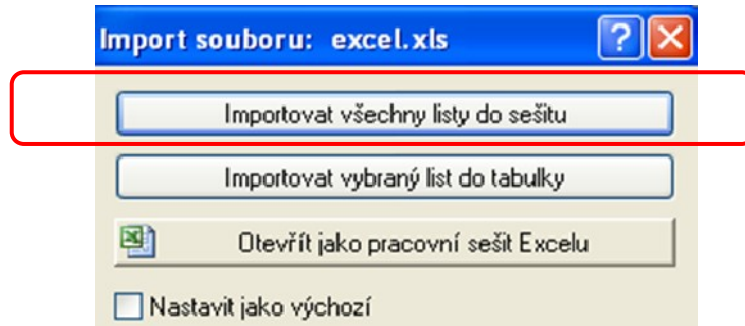
A po té klikneme na **Přidat do MS Word**:



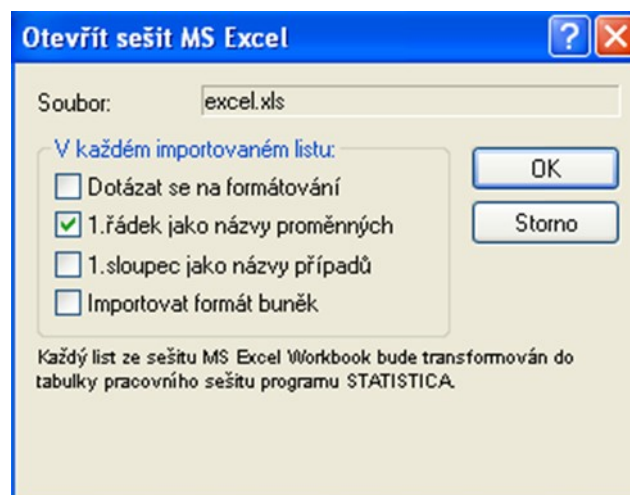
Celý proces se dá automatizovat viz **kapitola 8.1**.

7 Další možnosti načtení souborů

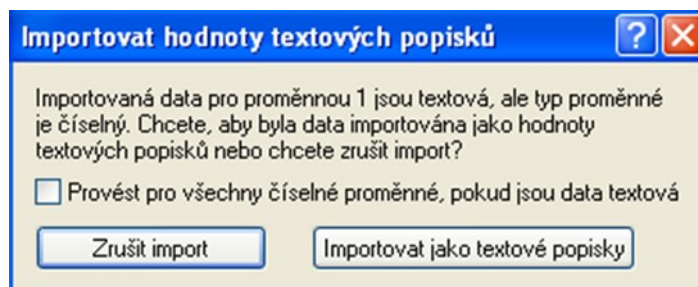
Při načítání z souboru typu Excel jsme si zatím ukázali pouze jednu možnost, nyní se podíváme na možnost *Importovat všechny listy do sešitu*



- **Importovat všechny listy do sešitu**
Při této možnosti můžeme podrobně nastavit pouze následující parametry (nikoli už velikost oblasti dat každého listu – tu *STATISTICA* určí automaticky):

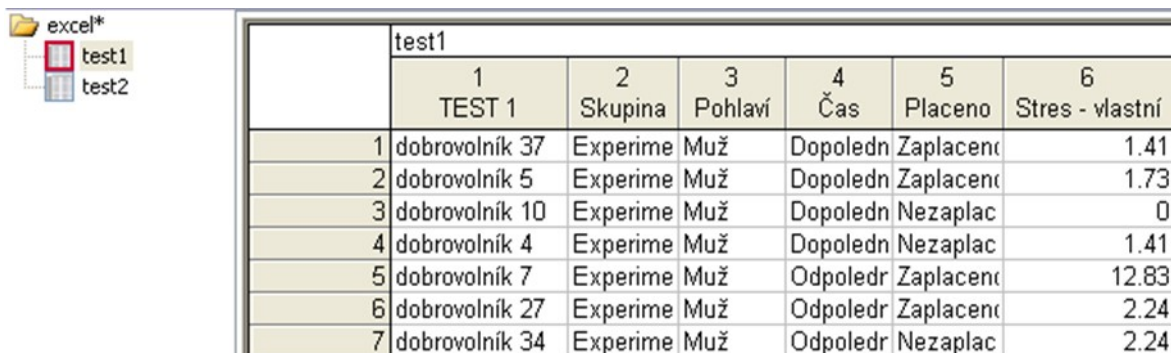


V případě, že listy obsahují i textové proměnné, je uživatel upozorněn na výskyt sloupce s textovými popisky:



Tuto informaci je třeba vzít na vědomí (tlačítko **Importovat jako textové popisky**) a to buď u každé proměnné obsahující textové hodnoty zvlášť, anebo zaškrtnutím možnosti **Provést pro všechny číselné proměnné, pokud jsou data textová** odsouhlasíme možné textové popisky u všech proměnných.

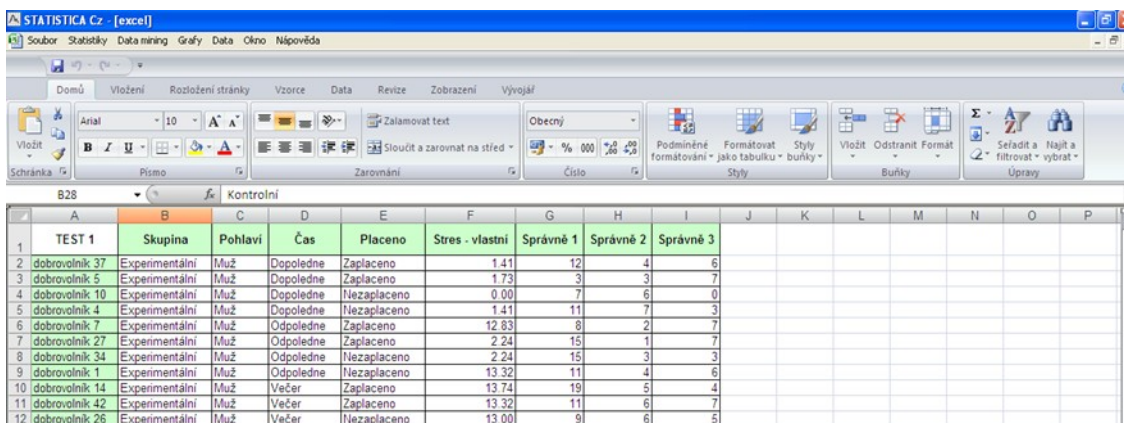
Možnost načíst všechny listy do sešitu načte ty listy excelovské tabulky, ve kterých jsou uložena nějaká data, do sešitu *STATISTICA* (.stw), aktivní list Excelu je vybrán jako *aktivní vstup* i ve *STATISTICA*, tj. případné grafy a výpočty probíhají nad tímto listem, pokud uživatel nezvolí jinou tabulku jako aktivní vstup – aktuální aktivní vstup je odlišen červeně orámovanou ikonou:



	test1					
	1	2	3	4	5	6
	TEST 1	Skupina	Pohlaví	Čas	Placeno	Stres - vlastní
1	dobrovolník 37	Experime	Muž	Dopoledn	Zaplaceno	1.41
2	dobrovolník 5	Experime	Muž	Dopoledn	Zaplaceno	1.73
3	dobrovolník 10	Experime	Muž	Dopoledn	Nezaplac	0
4	dobrovolník 4	Experime	Muž	Dopoledn	Nezaplac	1.41
5	dobrovolník 7	Experime	Muž	Odpoledr	Zaplaceno	12.83
6	dobrovolník 27	Experime	Muž	Odpoledr	Zaplaceno	2.24
7	dobrovolník 34	Experime	Muž	Odpoledr	Nezaplac	2.24

- **Otevřít jako pracovní sešit Excelu**

Pokud budeme chtít využívat funkce dostupné v Excelu a další nástroje aplikace Excel, je vhodné otevřít soubor jako pracovní sešit Excelu. V horní části menu je přístup k menu aplikace *STATISTICA* a v případě novější verze aplikace Excel je pod ní zobrazen i pás



	TEST 1	Skupina	Pohlaví	Čas	Placeno	Stres - vlastní	Správně 1	Správně 2	Správně 3
2	dobrovolník 37	Experimentální	Muž	Dopoledne	Zaplaceno	1.41	12	4	6
3	dobrovolník 5	Experimentální	Muž	Dopoledne	Zaplaceno	1.73	3	3	7
4	dobrovolník 10	Experimentální	Muž	Dopoledne	Nezaplaceno	0.00	7	6	0
5	dobrovolník 4	Experimentální	Muž	Dopoledne	Nezaplaceno	1.41	11	7	3
6	dobrovolník 7	Experimentální	Muž	Odpoledne	Zaplaceno	12.83	8	2	7
7	dobrovolník 27	Experimentální	Muž	Odpoledne	Zaplaceno	2.24	15	1	7
8	dobrovolník 34	Experimentální	Muž	Odpoledne	Nezaplaceno	2.24	15	3	3
9	dobrovolník 1	Experimentální	Muž	Odpoledne	Nezaplaceno	13.32	11	4	6
10	dobrovolník 14	Experimentální	Muž	Večer	Zaplaceno	13.74	19	5	4
11	dobrovolník 42	Experimentální	Muž	Večer	Zaplaceno	13.32	11	6	7
12	dobrovolník 26	Experimentální	Muž	Večer	Nezaplaceno	13.00	9	6	5

karet s nástroji Excelu. U starších verzí Excelu tento pás zobrazen není, nicméně do buněk sešitu lze vkládat vzorce využívající funkce Excelu v obou případech.

Oblast dat, se kterou chceme pracovat, definujeme při prvním spuštění grafu nebo analýzy. Je dobré nadefinovat maximální velikost oblasti, se kterou chceme pracovat, neboť je následně použita jako default i pro další spuštěné grafy a analýzy ve *STATISTICA*.

Vyberte rozsah Excelu pro analýzu: excel ? X

První řádek dat: 2 První sloupec dat: B OK

Poslední řádek dat: 49 Poslední sloupec: I Storno

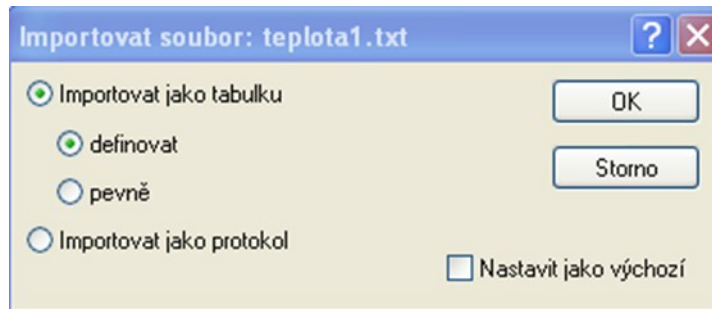
Vzít jména proměnných z řádku: 1 Použít výběr

Vzít jména případů ze sloupce: A

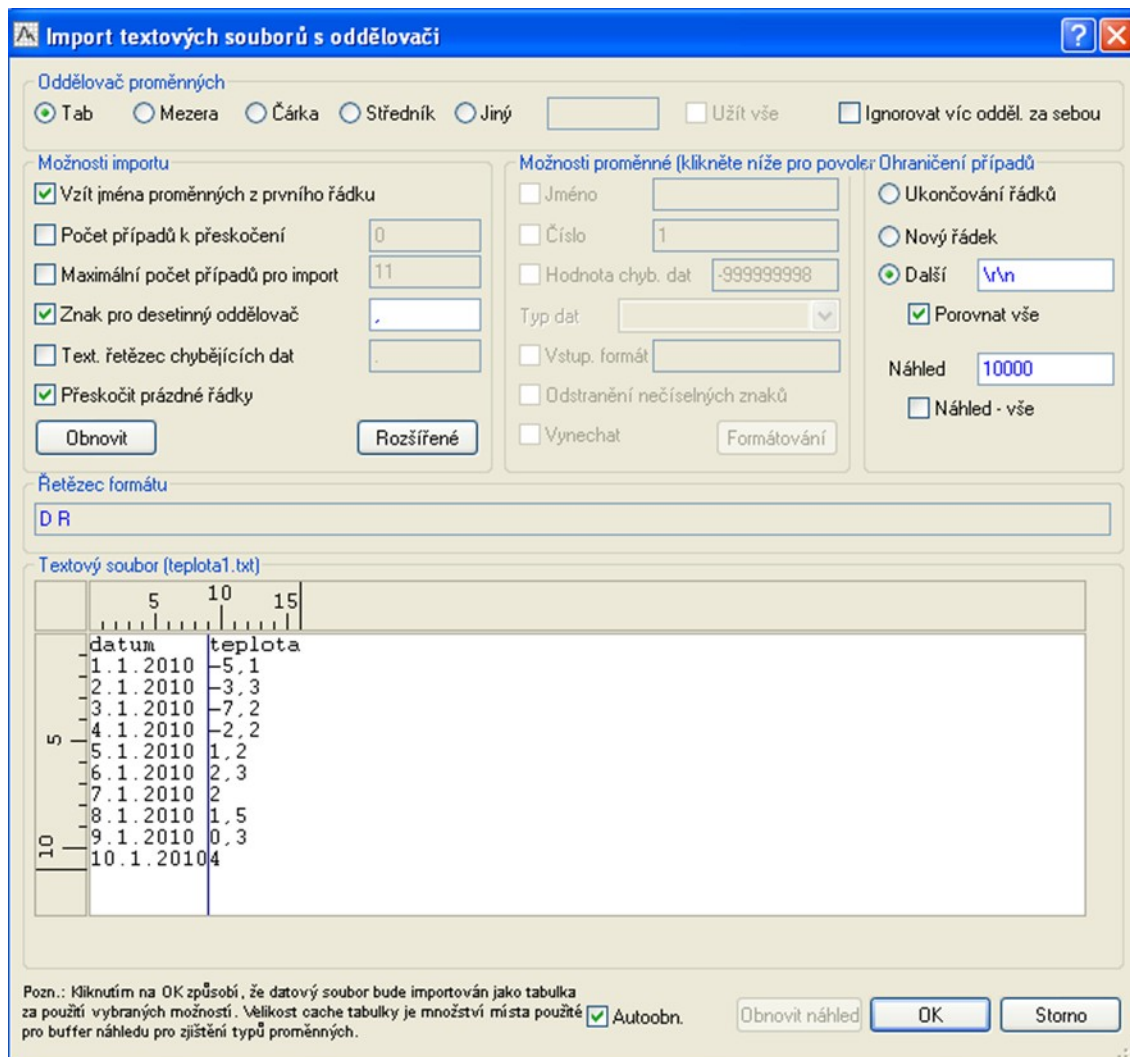
Ignorovat text ve sloupcích s čísly
 Vytvořit textovou proměnnou pro text
 Přehled/změna typů sloupců před importem

7.1 Otevření textového souboru

V menu **Soubor** zvolíme možnost **Otevřít...** a pomocí procházení úložišť osobního počítače nadefinujeme cestu k textovému souboru (např. s koncovkou .txt nebo .csv). Potvrdíme **OK** a zobrazí se následující dialog:



Ten necháme beze změny a opět potvrdíme **OK**. Definici, jak přesně chceme k obsahu textového souboru přistupovat, upřesníme prostřednictvím následujícího dialogu:



V horní části dialogu nastavíme oddělovač proměnných (defaultní nastavení je tabelátor nebo středník, podle typu dokumentu). Máme možnost nadefinovat i vlastní oddělovač – volba **Jiný** umožňuje vepsat vlastní typ oddělovače. Pokud je oddělovač tvořen celou skupinou znaků, je nutné zaškrtnout možnost **Užít vše**.

V dolním okně dialogu se automaticky zobrazuje náhled souboru tak, jak bude vypadat po načtení do *STATISTICA*, jednotlivé proměnné (sloupce) jsou odděleny svislými čarami. Pokud je textový soubor tvořen automaticky – jde například o výstup z nějakého programu – a na úvod dokumentu se zobrazuje hlavička identifikačních údajů a potom teprve samostatná data, máme možnost nastavit přeskočení prvních *n* řádků souboru (volba **Počet případů k přeskočení**). Dále je důležité si uvědomit, zda proměnné mají nějaký název – většinou chceme načíst tyto názvy jako záhlaví tabulky, proto i defaultní volba pro načtení souboru je **Vzít jména proměnných z prvního řádku**.

Zkontrolujeme také oddělovač desetinných míst, *STATISTICA* používá nastavení oddělovače pro Windows, tj. pokud otevíraný soubor vznikl například ve skriptu pro Linux systém, může být kódování desetinných míst tohoto souboru odlišné.

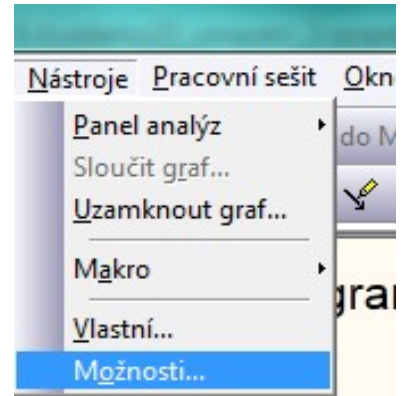
V tabulce náhledu můžeme myší vybrat konkrétní sloupec – proměnnou. Tím aktivujeme střední část menu **Možnosti proměnné**. Nyní lze nastavit jméno proměnné, nastavit datový typ anebo zvolený sloupec vyloučit z načítání.

Po nastavení všech parametrů potvrdíme **OK**. Výsledkem je otevření tabulky formátu *.sta* ve *STATISTICA*:

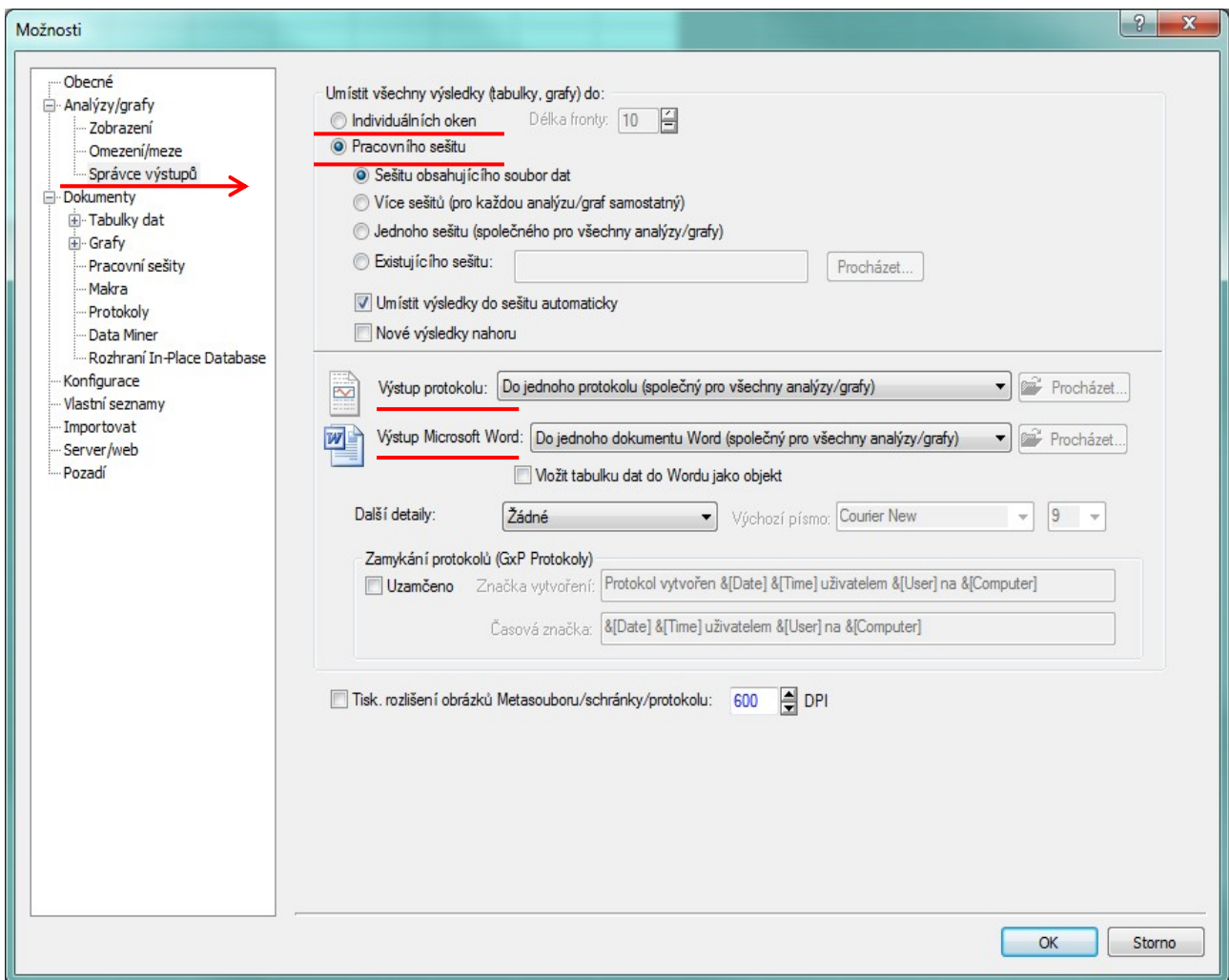
	1	2
	datum	teplota
1	1.1.2010	-5.10
2	2.1.2010	-3.30
3	3.1.2010	-7.20
4	4.1.2010	-2.20
5	5.1.2010	1.20
6	6.1.2010	2.30
7	7.1.2010	2.00
8	8.1.2010	1.50
9	9.1.2010	0.30
10	10.1.201	4.00

8 Správce výstupů

8.1 Výstup do Microsoft Word / do protokolu STATISTICA

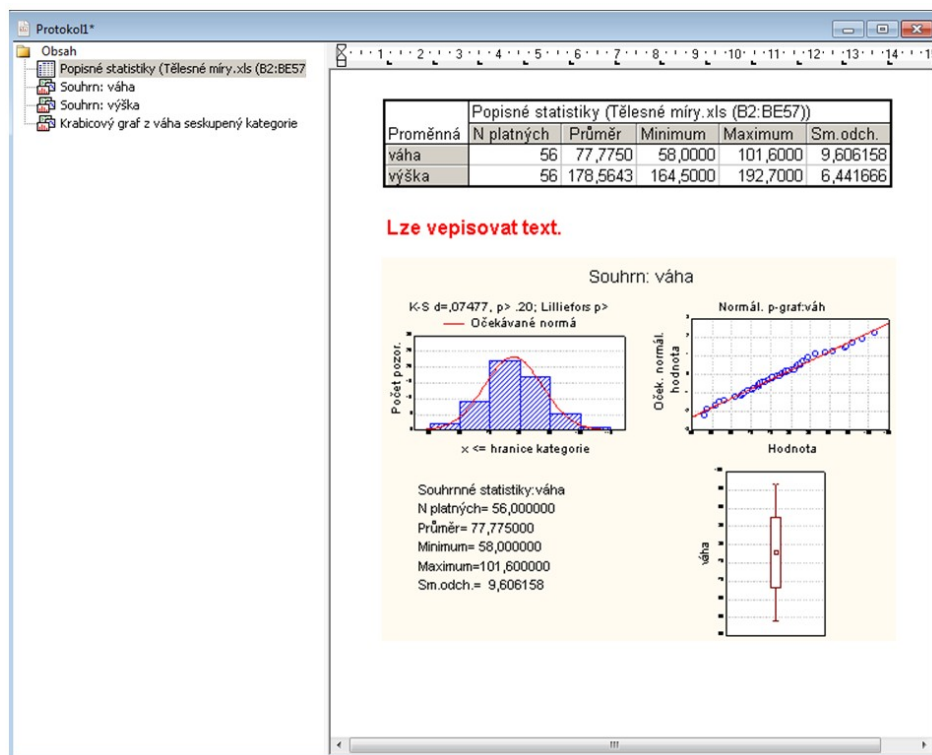


V programu *STATISTICA* můžeme nastavit, v jakém formátu budou ukládat výstupy. Ze základní nabídky vybereme *Nástroje - Možnosti...*. Otevře se dialog *Možnosti*, ve kterém přejdeme na záložku *Správce výstupů*:



Můžeme zvolit některé z těchto možností:

- *individuální okna* - každá tabulka či graf se zobrazuje v samostatném okně. Jednotlivá okna pak lze uložit ve formátu programu *STATISTICA* nebo v jiném formátu podle toho, zda se jedná o tabulku nebo graf. Pomocí nabídky **Soubor – Uložit** můžeme vybrat formáty **.xls*, **.txt*, **.htm*, **.pdf*, **.wmf*, **.jpg*, **.gif* atd.
- *pracovní sešit* - standardní formát výstupů v programu *STATISTICA* s příponou **.stw*. Právě v tomto formátu máme nyní výstupy z výše uvedených příkladů (pokud jsme neměnili výchozí nastavení). Okno pracovního sešitu je rozděleno na dvě části. Levá část zobrazuje stromovou strukturu (obdoba Průzkumníka). Pravá část je editorem vybraných dokumentů.
- *protokol* - má podobný vzhled jako pracovní sešit. V jeho levé části se zobrazuje seznam objektů protokolu. Pravá část je obdobou textového editoru. Na rozdíl od pracovního sešitu lze do protokolu mezi jednotlivé výstupy vepisovat text (viz následující ilustrační obrázek).

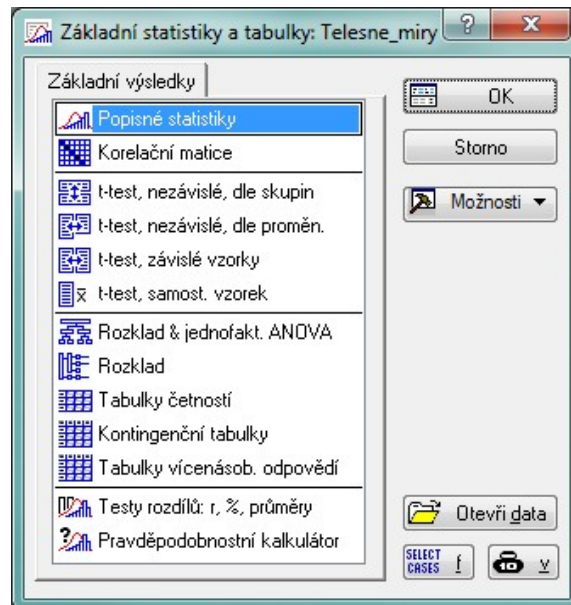


- *výstup do Microsoft Word* – výstupy se vkládají do dokumentu Microsoft Word, a mohou tak být jednoduše sdíleny s dalšími spolupracovníky.

9 Další příklady analýzy dat

9.1 Příklad – výpočet popisných statistik

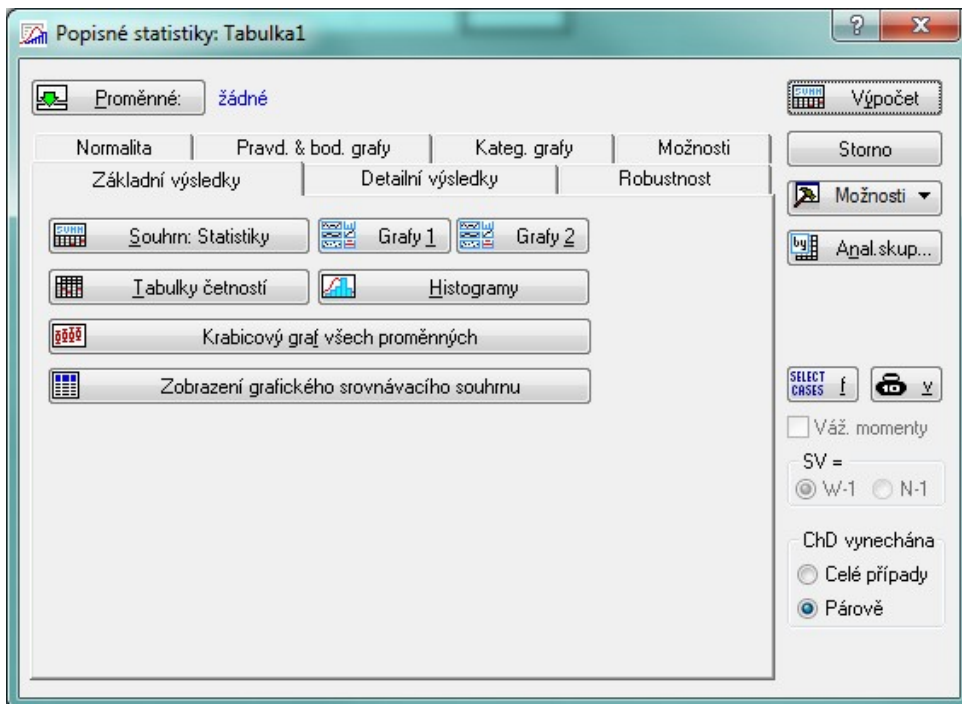
Na otevřené tabulce *Tělesné míry.xls* si ukážeme výpočet popisných statistik pro proměnné *váha* a *výška* u mužů. Přejdeme tedy na list *muži* a ze základní nabídky vybereme *Statistiky – Základní statistiky/tabulky*.



Poznámka:

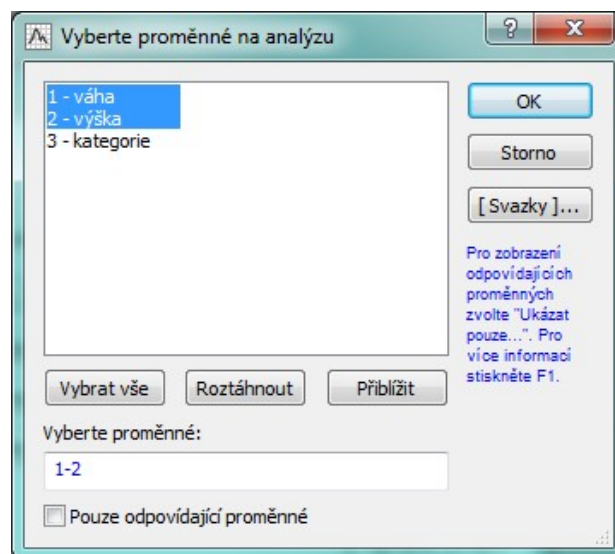
Pokud bychom pracovali se vstupními daty v tabulce formátu **.sta*, dostali bychom výše uvedenou nabídku přímo ze základní nabídky *Statistiky – Základní statistiky/tabulky*.

Nechme vybrané **Popisné statistiky** a kliknutím na **OK** přejdeme do okna vlastní analýzy:



Dialog obsahuje řadu záložek s nástroji pro výpočet různých popisných statistik a pro tvorbu grafů. Pokud se však nechceme zabývat podrobným nastavením, stačí kliknout na tlačítko **Výpočet** a program **STATISTICA** nás provede kroky potřebnými k výpočtu základních popisných statistik.

Nejprve vybereme proměnné *váha* a *výška* (stačí přes tyto dvě proměnné přejet myší při stisknutém levém tlačítku):



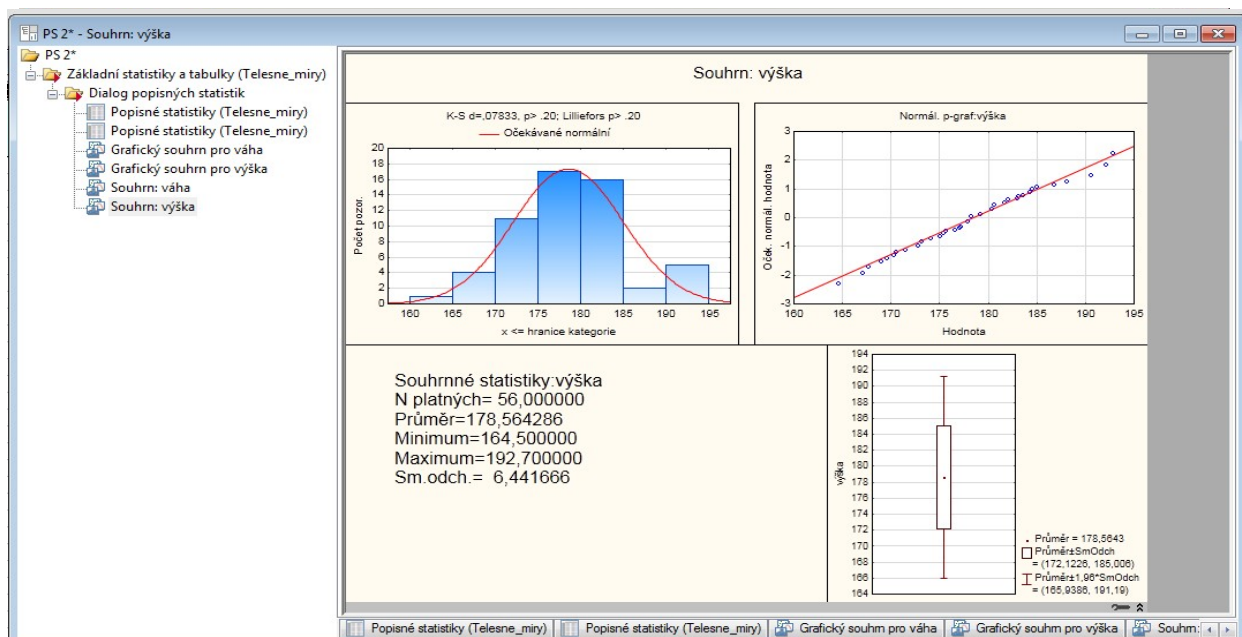
Poznámka:

STATISTICA také podporuje konvenci výběru pomocí kláves CTRL a SHIFT.

Kliknutím na **OK** dostaneme tabulku s výsledky:

Proměnná	N platných	Průměr	Minimum	Maximum	Sm.odch.
váha	56	77,7750	58,0000	101,6000	9,606158
výška	56	178,5643	164,5000	192,7000	6,441666

Výsledky můžeme snadno doplnit také o grafické výstupy. Obnovíme dialog *Popisných statistik*, který se minimalizoval na panel analýz a klikneme na tlačítko *Souhrn:Grafy*. Výsledek je na následujícím obrázku:



Rychle tak získáme pohled na data pomocí histogramu, normálního pravděpodobnostního grafu a krabicového grafu. Širokou nabídku dalších popisných statistik a grafických pohledů na data nalezneme na jednotlivých záložkách dialogu *Popisných statistik*.

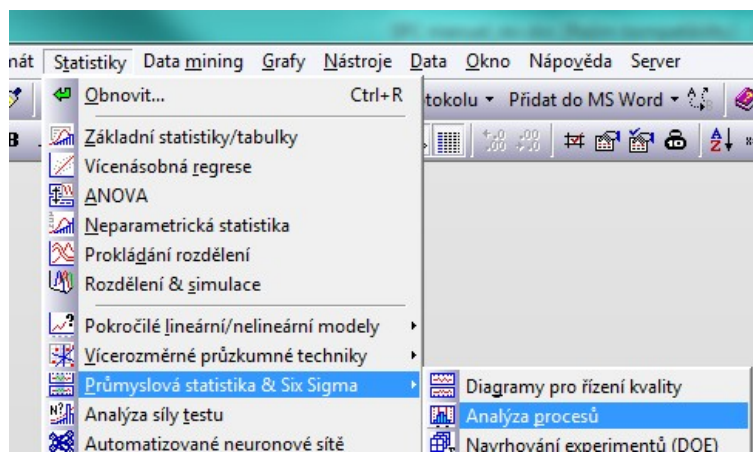
Pro další příklad si výstup s výsledky dosavadních analýz nechme otevřený. Ostatní okna na ploše v programu *STATISTICA* můžeme zavřít.

9.2 Analýza způsobilosti procesu

Úvod

Otevřeme si datový soubor *Pistons CZ.sta*.

Analýza způsobilosti procesu a výpočet indexů způsobilosti procesu lze provádět v modulu *Analýza procesů*, který je dostupný z nabídky *Statistika – Průmyslová statistika & Six Sigma*. Po otevření tohoto modulu se zobrazí úvodní panel.



Předpokládejme, že proces je statisticky zvládnutý (což jsme ověřili pomocí regulačních diagramů). Nyní však vyvstává otázka, do jaké míry náš proces v dlouhodobém měřítku odpovídá technickým specifikacím nebo obchodním cílům. Pokud např. vyrábíme pístní kroužky, chtěli bychom znát, kolik z nich má rozměr spadající dovnitř specifikačních mezí. Obecněji řečeno, otázkou je, jak způsobilý je proces (nebo dodavatel) produkovat výrobky vyhovující specifikacím.

Odpověď na tyto otázky je možné s využitím statistických technik. Vrátime-li se k našemu příkladu s pístními kroužky, ze vzorku určité velikosti můžeme odhadnout směrodatnou odchylku procesu, tj. směrodatnou odchylku průměru kroužků. Pokud je rozdělení hodnot normální, je možné poměrně snadno odvodit podíl kroužků, které budou specifikačním mezím vyhovovat. Podívejme se nyní na některé nejdůležitější indexy, které se běžně používají k popisu způsobilosti procesu.

Rozsah procesu

Nejdříve ovšem přijmeme úmluvu, podle níž se meze stanovují ve vzdálenosti $\pm 3\sigma$ od nominální hodnoty (σ je sm. odchylka procesu). Tyto meze by měly být shodné s těmi, které jsme předtím použili v regulačním diagramu, když jsme se snažili proces přivést do statisticky zvládnutého stavu. Tyto meze označují **rozsah procesu**. Pokud použijeme meze ve vzdálenosti $\pm 3\text{Sigma}$ ($\pm 3\sigma$), pak přibližně 99% všech pístních kroužků padne do těchto mezí.

Specifikační meze, LSL, USL

Bývá obvyklé, že technické požadavky stanovují také rozsah přijatelných hodnot. V našem příkladě je řečneme akceptovatelné, pokud průměr pístního kroužku bude v rozmezí $74.0 \pm$

.02 milimetrů. Dolní specifikační mez, LSL, je tedy $74 - 0,02 = 73,98$; horní specifikační mez, USL, je $74 + 0,02 = 74,02$. Rozdíl mezi USL a LSL se nazývá **specifikační rozsah**.

Potenciální způsobilost (C_p)

Nejjednodušší a nejpřímější indikátor způsobilosti procesu.

Je definován jako poměr **specifikačního rozsahu** a **rozsahu procesu**. Tento index se dá vyjádřit jako:

$$C_p = \frac{USL - LSL}{6\sigma}$$

Volně řečeno, tento index vyjadřuje, „**jak velká část**“ křivky normálního rozdělení se vejde **do specifikačních mezí (za předpokladu, že je proces vycentrován)**.

Poměr způsobilosti (C_r)

Tento index je vlastně ekvivalentní indexu C_p; je roven převrácené hodnotě, tj. $C_r = 1/C_p$. Tento index se dá tedy slovně popsat jako podíl specifikačního rozsahu, který je využíván vlastním procesem.

Dolní/horní potenciální způsobilost (C_{pl}, C_{pu})

Hlavní nevýhodou indexu C_p (nebo C_r) je to, že poskytuje nesprávné informace, pokud proces není vycentrován kolem nominální hodnoty (TS). „Excentricitu“ procesu je možné vyjádřit pomocí průměru procesu. Nejdříve je možné vypočítat dolní a horní potenciální způsobilost procesu:

$$C_{pl} = \frac{\mu - LSL}{3\sigma} \text{ a } C_{pu} = \frac{USL - \mu}{3\sigma}.$$

Je zřejmé, že pokud obě hodnoty nejsou shodné, proces není vycentrován.

Korekce (k)

Je možné upravit index C_p tak, aby reflektoval fakt, že proces není vycentrován. Můžeme vypočítat hodnotu korekce k:

$$k = \left| \frac{TS - \mu}{\frac{1}{2}(USL - LSL)} \right|$$

Tato korekce vyjadřuje excentricitu (nominální hodnota minus průměr) vztaženou ke specifikačnímu rozsahu.

Index C_{pk}

Úpravou indexu C_p pomocí korekce k můžeme vypočítat index C_{pk}, který počítá s tím, že proces není vycentrován.

$$C_{pk} = (1 - k)C_p$$

Pokud je proces vycentrován, pak je C_{pk} rovno nule a C_{pk} je shodné s C_p . Pokud se ovšem proces vzdálí od nominální hodnoty, C_{pk} je menší než C_p .

Potenciální způsobilost II (Cpm)

Jedna z nedávných úprav indexu C_p je zaměřena na odstranění vlivu (náhodného) nevycentrování. Můžeme vypočítat alternativní hodnotu směrodatné odchylky (σ_2) z n jednotlivých pozorování x_i ve vzorku jako:

$$\sigma_2 = \frac{\sqrt{\sum_{i=1}^n (x_i - TS)^2}}{n - 1}$$

Poté můžeme tento alternativní odhad směrodatné odchylky použít k výpočtu indexu C_p stejně jako předtím; tento nový index ovšem budeme označovat jako C_{pm} .

Výkonnost procesu vs. způsobilost procesu

Monitorujeme-li proces prostřednictvím regulačního diagramu, bývá často užitečné vypočítat také indexy způsobilosti. V případě, že data sestávají z několika vzorků (jako je tomu u regulačních diagramů), je možné vypočítat dva různé odhady variability dat. Jedním z nich je klasická směrodatná odchylka všech pozorování – v tomto případě budeme ignorovat fakt, že data pocházejí z více vzorků. Druhou možností je odhadovat inherentní variabilitu procesu z variability uvnitř jednotlivých vzorků.

Pokud použijeme k výpočtu standardních indexů celkovou variabilitu, říká se výsledným indexům *indexy výkonnosti* (protože popisují vlastní výkon procesu). Indexy vypočtené pomocí inherentní variability (směrodatné odchylky uvnitř vzorků) se obvykle nazývají *indexy způsobilosti* (protože popisují inherentní způsobilost procesu). Pro datové soubory, které obsahují více vzorků, vypočte *STATISTICA* indexy způsobilosti (C_p , C_{pk} , ...) i výkonnosti (P_p , P_{pk} , ...).

Toleranční meze

Než se na počátku 80. let rozšířilo používání indexů způsobilosti, byl pro odhad vlastností výrobního procesu obvyklou metodou výpočet a zkoumání tolerančních mezí. Postup byl následující. Předpokládejme, že sledovaný znak jakosti je v celé populaci výrobků normálně rozdělený. Můžeme provést intervalový odhad, který nám s určitou spolehlivostí zajistí, že určitá část populace bude ležet v odhadnutém intervalu. Jinými slovy, pokud máme dáno:

- (1) velikost vzorku (n),
- (2) průměr procesu,
- (3) směrodatnou odchylku procesu (*sigma*),
- (4) hladinu spolehlivosti α
- (5) procento populace, které chceme do intervalu zahrnout,
- (6)

můžeme vypočítat toleranční meze, které vyhoví všem parametrům. Modul *Analýza procesů* poskytuje uživateli možnosti pro stanovení těchto parametrů a pro výpočet tolerančních mezí. Najdete zde i možnosti pro výpočet neparametrických tolerančních mezí, které nejsou založeny na předpokladu normality.

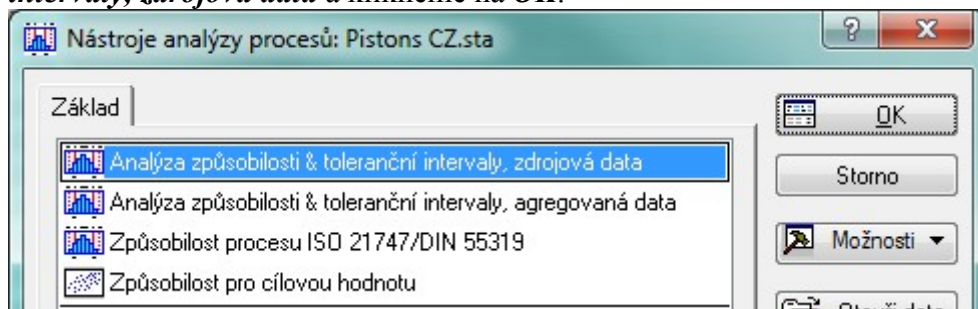
Ohodnocení normality a prokládání rozdělení

Představme si, že vyrábíme pístní kroužky. Specifikace říkájí, že jejich průměr by měl být **74.0 +/- 0.05 mm**.

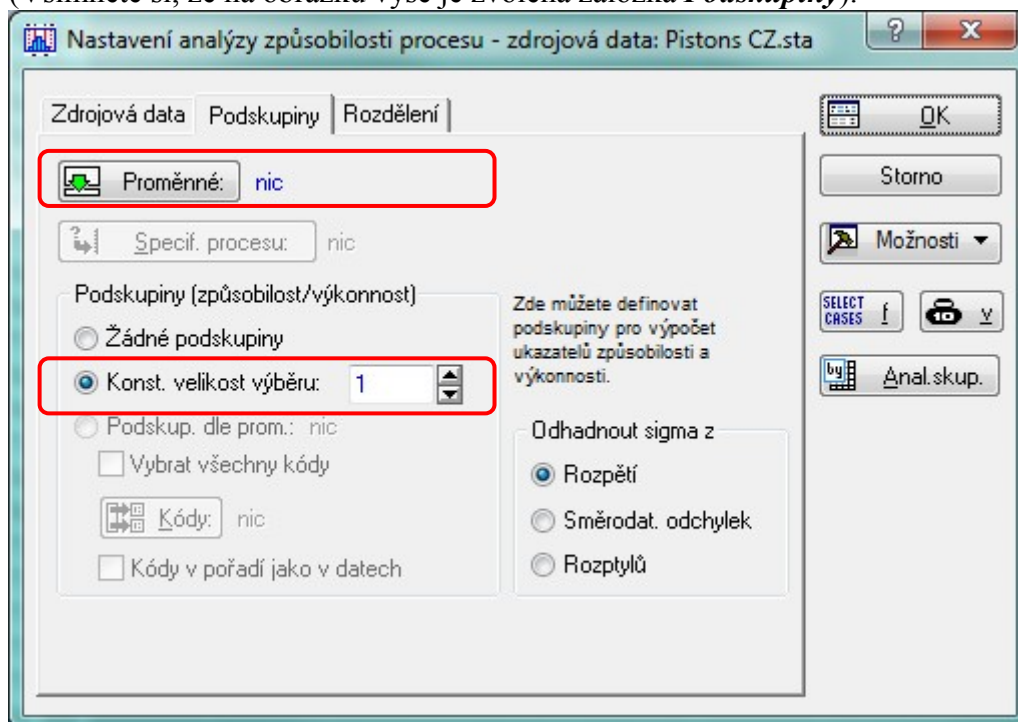
Z procesu jsme odebrali 25 vzorků po 5 pozorováních (hodnoty jsou uloženy v souboru *Pistons CZ.sta*). Chceme zjistit, jak způsobilý náš proces je.

Protože standardní indexy způsobilosti jsou založeny na předpokladu normality, měli bychom ověřit, zda data opravdu mají normální rozdělení. Poznamenejme ovšem, že data je možné proložit i dalšími spojitými rozděleními, včetně beta, exponenciálního, rozdělení extrémních hodnot, gama, log-normálního, Rayleighova, či Weibullova a výpočty indexů lze založit na směrodatné odchylce procesu odhadnuté z těchto distribucí. Máme možnost odhadnout parametry všech rozdělení a porovnat, jak dobře odpovídají datům.

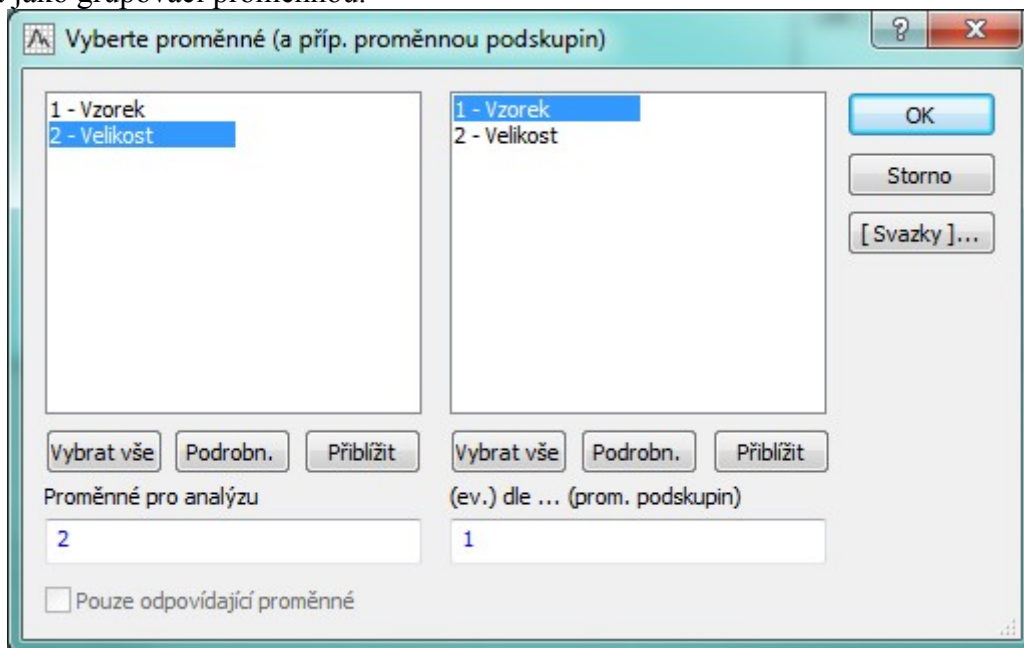
Např. prostřednictvím tlačítka **Otevři data** na úvodním panelu modulu *Analýza procesů* otevřeme soubor *Pistons CZ.sta*. Zvolíme možnost *Analýza způsobilosti & toleranční intervaly, zdrojová data* a klikneme na **OK**.



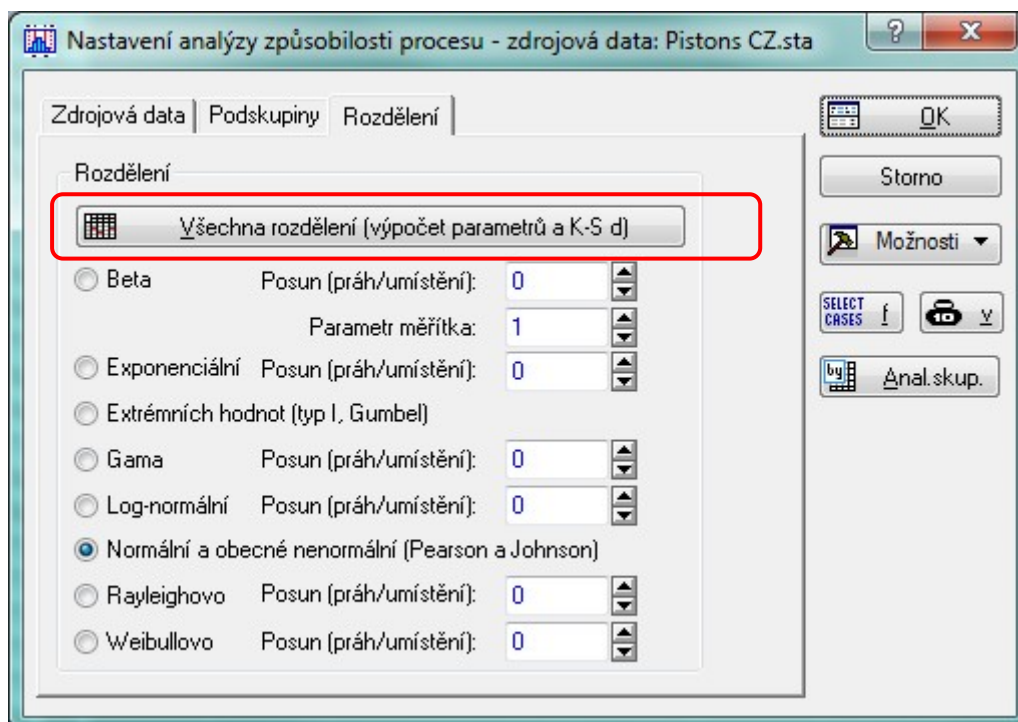
(Všimněte si, že na obrázku výše je zvolena záložka **Podskupiny**).



Klikněte na tlačítko **Proměnné** a zvolte proměnnou *Velikost* v poli **Proměnné pro analýzu** a *Vzorek* jako grupovací proměnnou.



Klikněte na **OK**. Všechna ostatní nastavení nechte na jejich implicitních hodnotách a přepněte se na záložku **Rozdělení**.



Klikněte na tlačítko **Všechna rozdělení (výpočet parametrů a K-S d)**.

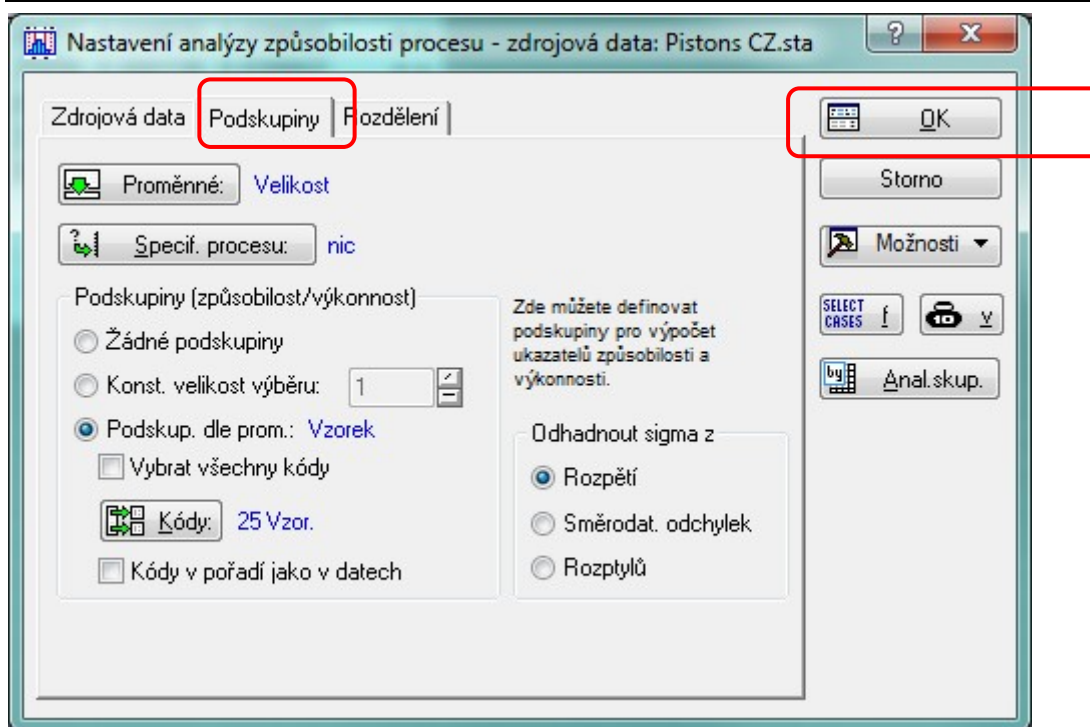
PS 4* Techniky analýzy p Dialog analýzy Odhady par		Odhady parametrů pro vš. rozdělení (Pistons CZ.sta) Proměnná: Velikost N = 125 (p-hodnoty předpokládají a-priorní znalost parametrů) Rozdělení jsou řazena (shora) z podle toho, jak dobře je lze proložit daty Na základě p-hodnoty je Kolmogorovův-Smirnovův test pro konkrétní rozd při $p > 0,20$ respektive $p > 0,10$ je rozdělení vhodným modelem pro data. Pozn.: Pokud je Normální (umístění, měřítko) vhodným modelem, použij				
Rozdělení	Vlastní Param 1	Vlastní Param 2	Param 1	Param 2	K-S d	K-S p-hodn.
Normální (umístění, měřítko)			74,00117	0,010070	0,039932	n.s.
Log-normální (práh, měřítko, tvar)	0,00		4,30408	0,000136	0,039952	n.s.
Nenormální (šikmost, špičatost)			-0,09795	0,446462	0,045565	n.s.
Rayleighovo (práh, měřítko)	0,00		52,32673		0,631781	$p < ,01$
Exponenciální (práh, měřítko)	0,00		74,00117		0,631951	$p < ,01$
Beta (práh, sigma, měřítko, tvar)	0,00	1,000000				--
Extrémních hodnot (umístění, měřítko)						--
Gama (práh, měřítko, tvar)	0,00					--
Weibullovo (práh, měřítko, tvar)	0,00					--

Tato tabulka obsahuje odhady parametrů pro různá pravděpodobnostní rozdělení proložená daty. Rozdělení jsou seřazena podle kvality proložení (od nejlepšího k nejhoršímu). Najdete zde i výsledky Kolmogorov-Smirnovova testu kvality proložení. Nulovou hypotézu tohoto testu představuje fakt, že data odpovídají rozdělení. Pokud je test významný (indikováno malými hodnotami úrovně p), je nepravděpodobné, že dané rozdělení odpovídá datům. **Označení „n.s.“ je zkratka pro „not significant“**, tj. „nevýznamné“, a tedy indikuje dobré proložení. Poznamenejme, že úrovně p jsou založeny na tabelovaných hodnotách. Jsou proto platné pouze v případě, že odhady parametrů jsou známy předem (což typicky nejsou), a měly by být proto interpretovány opatrně.

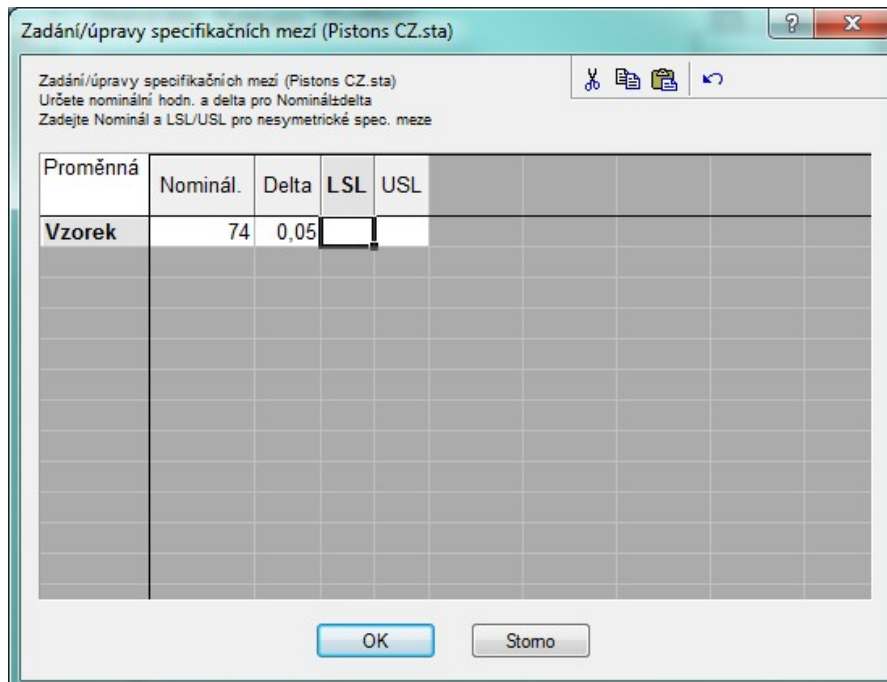
Jak je vidět, proložení dat normálním rozdělením se jeví jako nejlepší. Můžeme tedy aplikovat standardní indexy způsobilosti procesu. Někdy ovšem datům lépe odpovídají jiná rozdělení. V těch případech *STATISTICA* odhadne sigma procesu z tohoto rozdělení. Pak se ale musí dát pozor při interpretaci těch indexů, které zahrnují korekci na excentricitu procesu (protože některá z těchto rozdělení jsou velice šikmá).

Indexy způsobilosti procesu

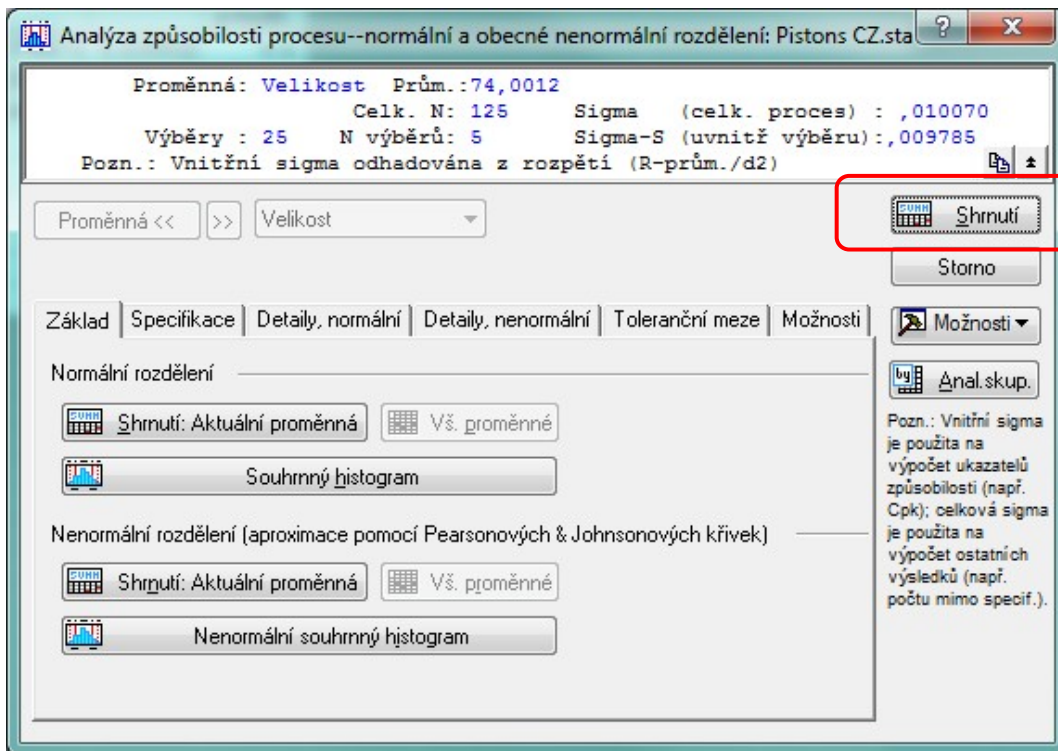
Vraťte se do dialogu specifikace analýzy



a klikněte na tlačítko **OK**. Objeví se dialogové okno **Zadání/úpravy specifikačních mezí**. Abychom mohli indexy způsobilosti počítat, je bezpodmínečně nutné nejprve zadat hodnoty technických specifikací. Do pole **Nominální** vložte 74 a do pole **Delta** hodnotu 0,05.



Klikněte na tlačítko **OK** a stiskněte **Shrnutí**.



STATISTICA vytvoří několik tabulek (pro **indexy způsobilosti** a pro **indexy výkonnosti**). Nyní se zaměříme jen na indexy způsobilosti:

The project tree on the left shows the hierarchy: 'Techniky analýzy procesů (Pistons CZ.sta)' > 'Dialog analýzy způsob. procesu, zdrojová data' > 'Dialog výsledků způsobilosti procesu, normální a obecné nenormální rozdělení' > 'Způsobilost procesu Z-bench, interv. spolehl. apod.' > 'Proměnná: Velikost (Pistons CZ.sta)'. The table on the right provides the following data:

Ukazatel způsobilosti	Hodnota
Dolní specifikační mez	73,95000
Nominální specifikace	74,00000
Horní specifikační mez	74,05000
Ukazatel způsobilosti CP	1,70323
CR (reciproké CP)	0,58712
Ukazatel způsobilosti CPK	1,66317
CPL (dolní ukaz. způsobilosti)	1,74329
CPU (horní ukaz. způsobilosti)	1,66317
K (necentrální korekce)	0,02352
CPM (potenc. způsobilost II)	1,64383

Tato tabulka obsahuje specifikace a různé indexy způsobilosti.

- Protože indexy způsobilosti jsou vyšší než 1, proces se zdá být způsobilý při splňování našich požadavků.

Ukazatel způsobilosti	Hodnota
Dolní specifikační mez	73,95000
Nominální specifikace	74,00000
Horní specifikační mez	74,05000
Ukazatel způsobilosti CP	1,70323
CR (reciproké CP)	0,58712
Ukazatel způsobilosti CPK	1,66317
CPL (dolní ukaz. způsobilosti)	1,74329
CPU (horní ukaz. způsobilosti)	1,66317
K (necentrální korekce)	0,02352
CPM (potenc. způsobilost II)	1,64383

Jinými slovy, produkujeme méně než 1% neshodných výrobků.

Ve skutečnosti proces využívá méně než 60% specifikačního rozsahu (74.0 +/- 0.05 mm), jak je vidět z hodnoty indexu CR. Využíváme zhruba **58%** stanoveného rozsahu tolerance.

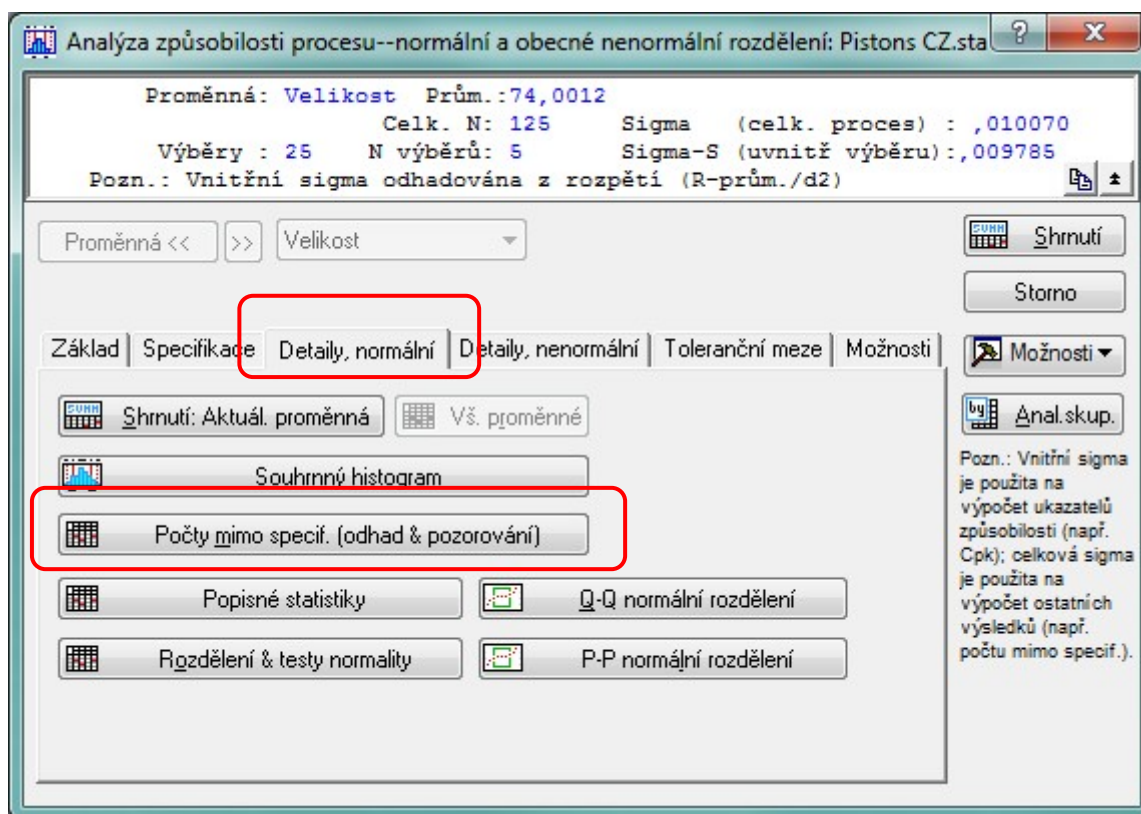
Proměnná: Velikost (Pistons CZ.sta)	
-3,000 *Sigma=73,9718	
+3,000 *Sigma=74,0305	
Ukazatel způsobilosti	Hodnota
Dolní specifikační mez	73,95000
Nominální specifikace	74,00000
Horní specifikační mez	74,05000
Ukazatel způsobilosti CP	1,70323
CR (reciproké CP)	0,58712
Ukazatel způsobilosti CPK	1,66317
CPL (dolní ukaz. způsobilosti)	1,74329
CPU (horní ukaz. způsobilosti)	1,66317
K (necentrální korekce)	0,02352
CPM (potenc. způsobilost II)	1,64383

Také se zdá, že proces je vycentrován kolem hodnoty 74, protože hodnota indexu C_{pk} je stále poměrně vysoká (1,66317) a necentrální korekce k je velice blízká nule:

Proměnná: Velikost (Pistons CZ.sta)	
-3,000 *Sigma=73,9718	
+3,000 *Sigma=74,0305	
Ukazatel způsobilosti	Hodnota
Dolní specifikační mez	73,95000
Nominální specifikace	74,00000
Horní specifikační mez	74,05000
Ukazatel způsobilosti CP	1,70323
CR (reciproké CP)	0,58712
Ukazatel způsobilosti CPK	1,66317
CPL (dolní ukaz. způsobilosti)	1,74329
CPU (horní ukaz. způsobilosti)	1,66317
K (necentrální korekce)	0,02352
CPM (potenc. způsobilost II)	1,64383

Je možné také provést analýzu očekávaného a pozorovaného počtu naměřených hodnot ležících mimo specifikační meze viz níže:

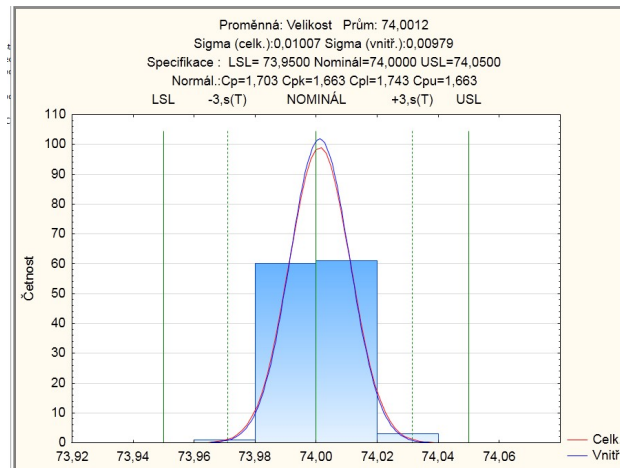
Na záložce **Detaily, normální** klikněte na tlačítko **Počet mimo specifikace**.



Tato tabulka obsahuje údaje o počtu a procentu pozorovaných jednotek ve vzorku, které padnou mimo specifikační meze (LSL, USL). Současně je uveden také počet a procento jednotek, u kterých se očekává, že padnou mimo specifikační meze, podle normálního rozdělení. Pro naše data ze specifikací nevypadlo ani jedno pozorování. Podle normálního rozdělení s daným průměrem procesu a směrodatnou odchylkou bychom mohli očekávat, **že pouze 0,0008% nebude odpovídat našim specifikacím.**

Techniky analýzy procesů (Pistons CZ.sta)		Proměnná: Velikost, rozdělení : Normál. (Pistons CZ.sta)				
Dialog analýzy způsob. procesu, zdrojová data		Specifikace : Dolní=73,9500 Nominál=74,0000 Horní=74,0500				
Odhady parametrů pro vš. rozdělení (Pistons CZ.sta)		Prům.74,001,Sm.od.: 0,1007				
Dialog výsledků způsobilosti procesu, normální a obecné		Pozorov.	Procenta Pozorov.	Očekáv.	Procenta Očekáv.	
Způsobilost procesu Z-bench, interv. spolehl. apoc	Proměnná: Velikost (Pistons CZ.sta)	Nad USL :	0	0,00	0,000078	0,000062
Způsobilost procesu Z-bench, interv. spolehl. apoc	Proměnná: Velikost (Pistons CZ.sta)	Pod LSL :	0	0,00	0,000023	0,000019
Způsobilost procesu Z-bench, interv. spolehl. apoc	Proměnná: Velikost, rozdělení : Normál. (Pistons CZ.sta)	Celk.	0	0,00	0,000101	0,000081

Souhrn analýzy způsobilosti je také možné nechat zobrazit graficky v podobě histogramu. Klikněte na tlačítko **Souhrnný histogram**.



(Poznámka: Tento histogram byl vytvořen pro deset kategorií. Na záložce **Možnosti** zrušte označení políčka **Pěkné intervaly** a zvolte počet kategorií.)

Výše uvedený histogram zobrazuje rozložení naměřených hodnot společně s proloženou křivkou normálního rozdělení (pro očekávané hodnoty). Současně jsou v histogramu vyznačeny specifikační meze (LSL a USL) a interval ± 3 sigma. Titulek grafu obsahuje hlavní indexy způsobilosti procesu.

Než byly do praxe uvedeny indexy způsobilosti, používaly se k ohodnocení charakteru výrobního procesu toleranční intervaly. Při dané velikosti vzorku (n), průměru procesu, směrodatné odchylce procesu (sigma), hladině spolehlivosti a podílu populace, který má spadat do tolerančního intervalu, je možné vypočítat toleranční meze. Stačí se přepnout na záložku **Toleranční meze**, určit výše zmíněné parametry (v našem případě změníme např. údaj v poli **Zahrnuto procent** na 90%) a kliknout na tlačítko **Toleranční meze, normální rozdělení**.

Data: Normální toleranční mez [Pistons C...	
Normální toleranční meze	
	Hodnota
Počet případů	125
Průměr	74,0012
Sigma	,010070
Dolní specifikační mez	73,9500
Nominální specifikace	74,0000
Horní specifikační mez	74,0500
Hladina spolehlivosti	95,0000%
% populace zahrnuto	90,0000%
Dolní toleranční mez	73,9826
Horní toleranční mez	74,0198

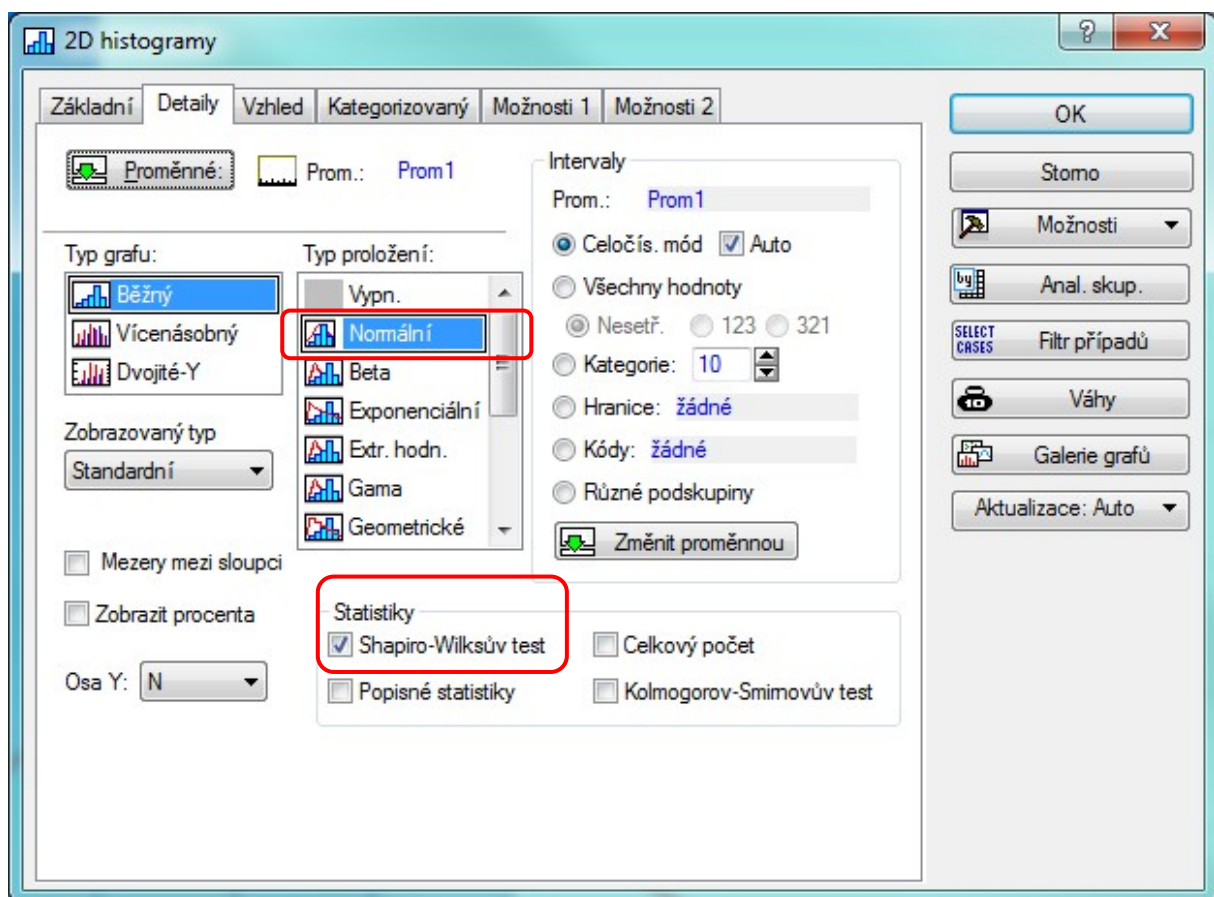
Kdybychom měli stručně shrnout údaje v tabulce, mohli bychom říci asi toto: jsme si na 95% jistí, že 90% populace bude ležet mezi hodnotami 73,9826 a 74,0198. Tyto hodnoty jsou popsány jako **Dolní a Horní toleranční mez**.

10 Ověření normality v softwaru STATISTICA

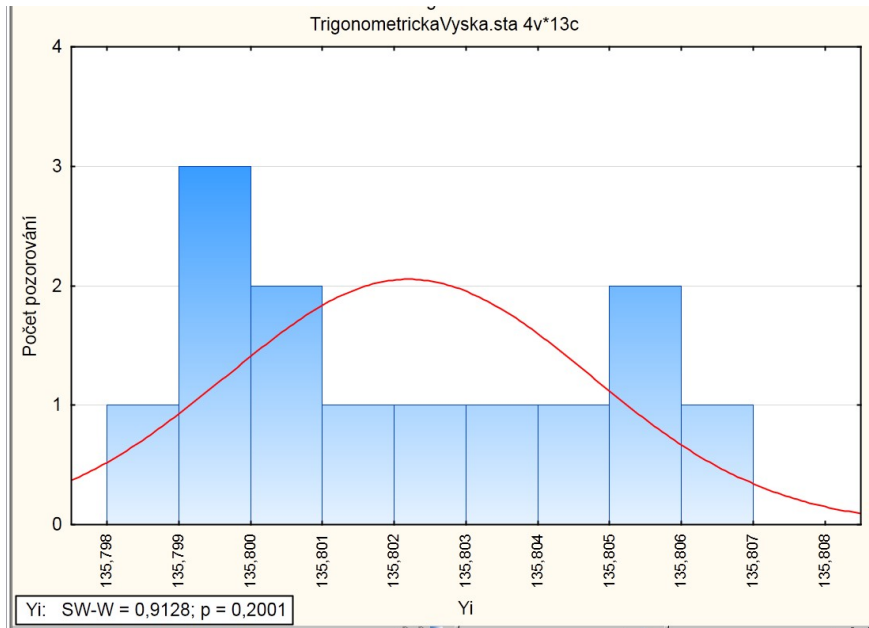
Jedním ze základních předpokladů mnoha statistických analýz je normalita. Pokud některý test či metoda normálního rozdělení předpokládá, je nutné to nejprve ověřit. K ověření lze použít mj. i statistické testy. Než však k testování normality přistoupíme, je dobré se zamyslet, zda se vůbec dá očekávat, že data jsou výběrem z normálního rozdělení. Pokud např. sledujeme platy obyvatelstva, víme, že nejsou omezené shora, zato jsou zdola omezené minimální mzdou, a rozhodně nejsou symetricky rozdělené kolem průměru. Takže prostou úvahou vyloučíme normalitu, aniž by bylo třeba provádět jakékoliv testy. Naopak u mnoha veličin, jako třeba byla v předchozím případě výška, je už z předchozích zkušeností známo, že se normálním rozdělením řídí. Potom testování také není nezbytné.

K ověřování normality systém STATISTICA poskytuje následující nástroje:

1. **Histogram** – vytvoříme histogram sledované proměnné a vizuálně ho porovnáme s normálním proložením:



Doplňkově si lze zaškrtnout Shapiro-Wilkův test pro otestování normality, v tomto konkrétním případě jsme nezamítli nulovou hypotézu o normalitě ($P(0,2) > 0,05$):



Dvojklikem do grafu vyvoláme dialog **Možnosti grafu** a graf si upravíme:

Histogram z Yi
TrigonometrickaVyska.sta 4v*13c

Zobraz. hodnoty měřítka

Auto - na hlavních značkách Použít textové hodn. z dat

Hodnoty dat: Zobr. jen hodnoty s text. popisy

Vlastní popisy:

Přeskakovat hodnoty

Vypnuto Auto Zobr. každý 2 popis

Formát hodnot

Obecné: ##0.)(##00)
Číslo: ##0
Datum: 0.)(0
Čas: 0
Vědecký: d/m/yy
Měna: dd/mm/yyyy
Procenta: d-mmm-yy
Zlomky: [E]S]0.###

Možnosti: Rozvrt: **Kolmo na osu**

Písmo...

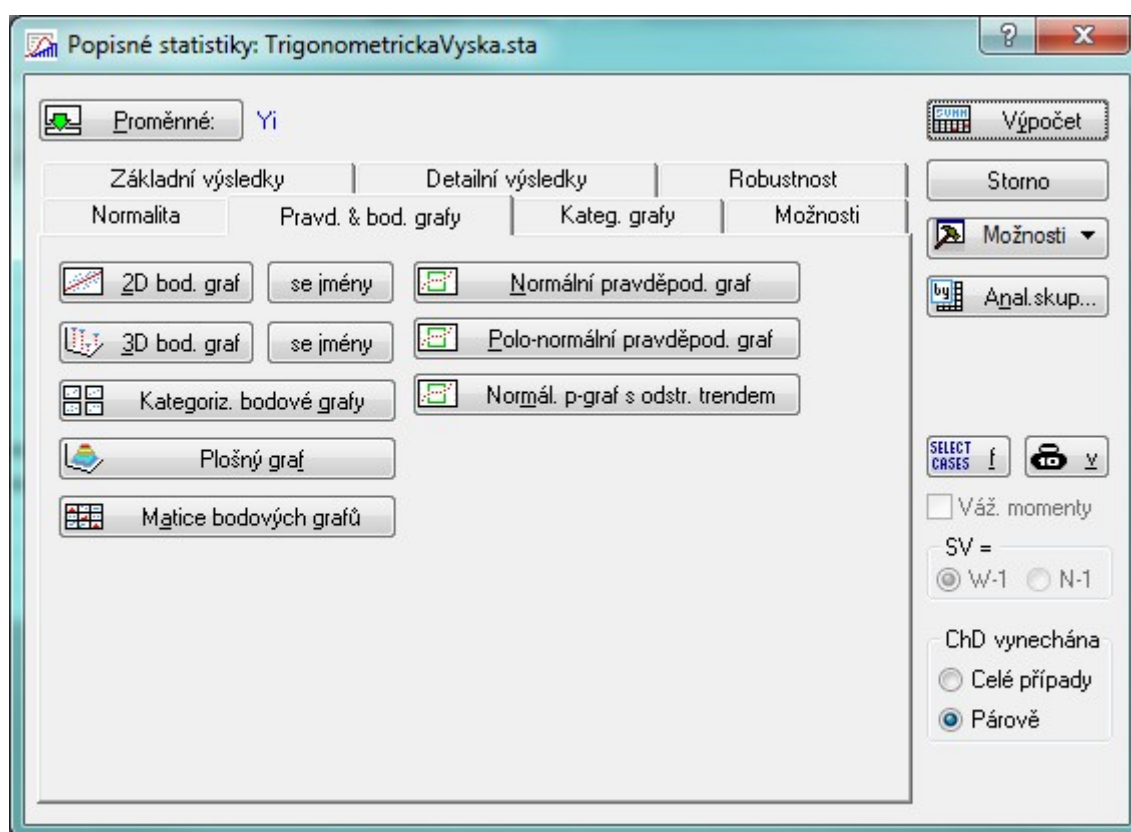
Písmo: Arial **Arial Rounded MT**
Řez písma: Normální **Polozhuštěné Tu**
Velikost: 8 9 10 11 12 14 16

Styl: Přeskřítnuté Podtržené
Barva: Černá
Ukázka: AaBbÁáÔô
Skript: Středoevropské

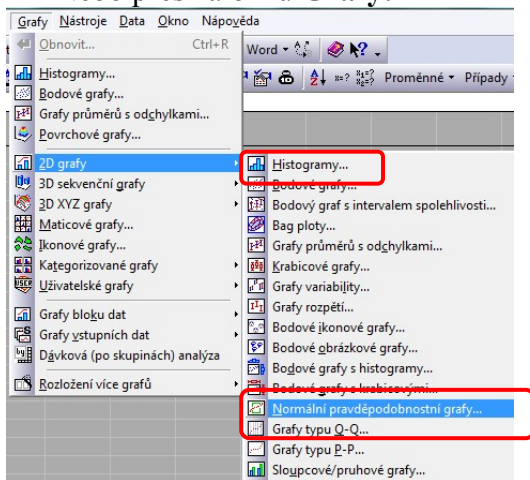
Toto písmo je typu Open Type. Bude použito pro tiskárnu i obrazovku.

Nevyvážený počet dat v jednotlivých intervalech nemusí nutně znamenat významné odchylky od normality, a proto je vhodnější použít **kvantilové grafy**:

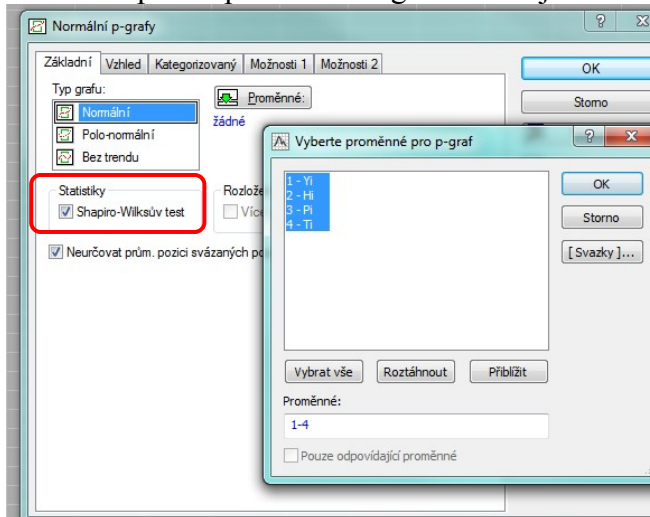
2. **Normální pravděpodobnostní graf** – jde o bodový graf, který porovnává kvantily spočtené z dat (osa x) s kvantily standardizovaného normálního rozdělení (osa y). Pokud veličina má normální rozdělení, leží body grafu na přímce. Tyto grafy lze vytvořit z nabídky *Statistika - Základní statistiky/tabulky - Popisné statistiky - Pravděpodobnostní & bodové grafy*. Kromě *Normálního pravděpodobnostního grafu* STATISTICA nabízí ještě *Polo-normální pravděpodobnostní graf* (obsahuje jen kladné hodnoty normálního rozdělení) a *Normální pravděpodobnostní graf s odstraněným trendem* (odstraněn lineární trend).



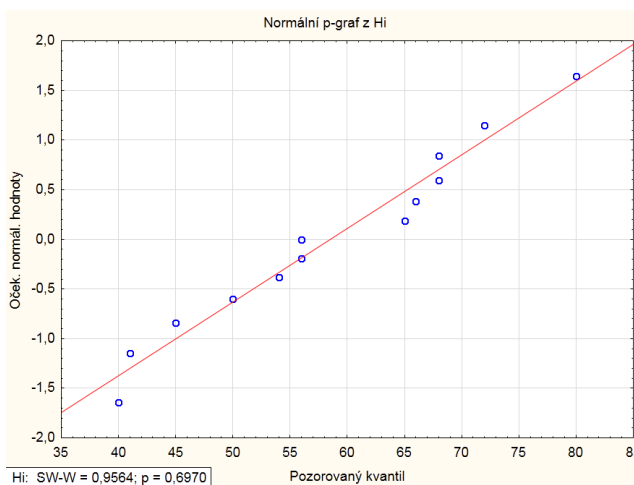
Nebo přes záložku Grafy:



Normální pravděpodobnostní graf obsahuje možnost zaškrtnout také Shapiro-Wilkův test:



Výsledný graf se statistikou SW testu:

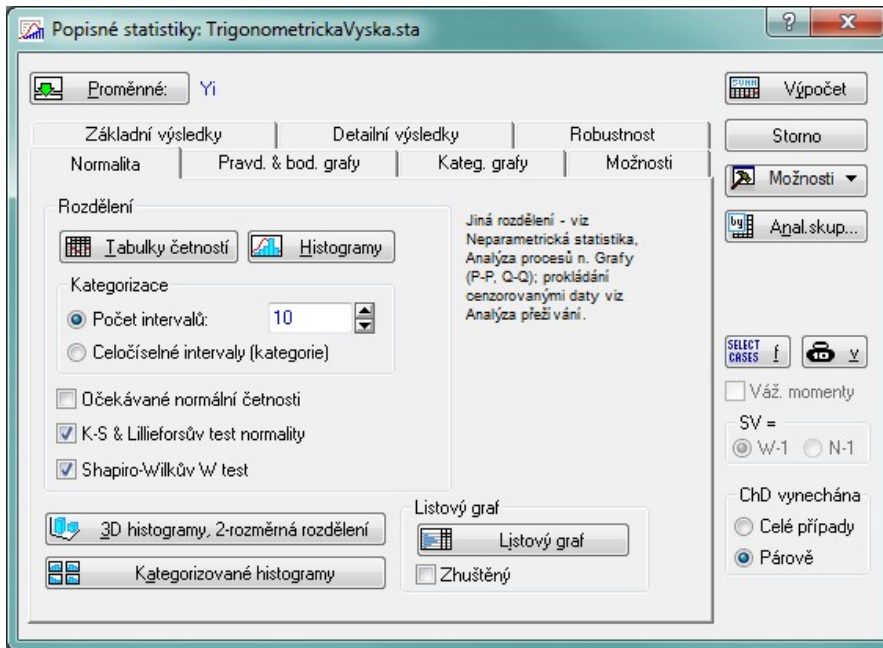


Zde nezamítáme nulovou hypotézu o normalitě $P(0,69) > 0,05$.

3. Testy

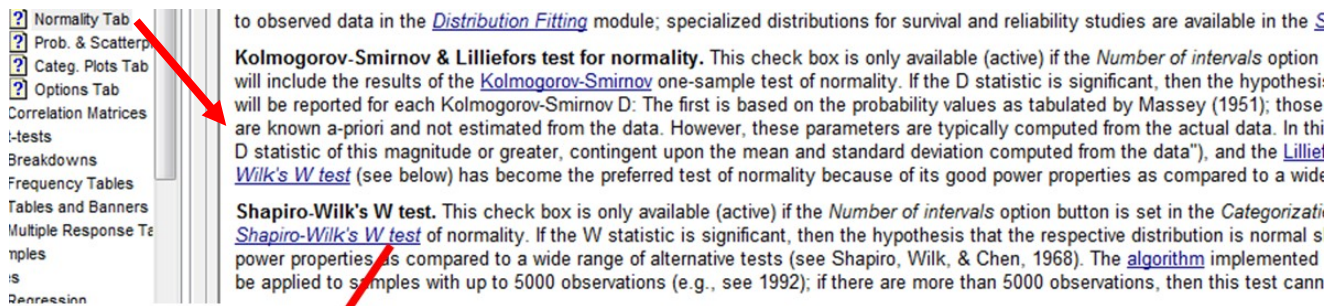
Kromě vizuálního ohodnocení jsou k dispozici také testy, které přímo s určitou pravděpodobností otestují, zda jsou data výběrem z normálního rozdělení, či nikoli. *STATISTICA* nabízí testy, např. Shapirův – Wilksův, Kolmogorovův – Smirnovův a Lillieforsův.

Přes *Statistiky* -> *Základní statistiky a tabulky* -> *Popisné statistiky* -> karta *Normalita*:



Jako nejjednodušší se doporučuje používat test Shapirův – Wilksův. Kolmogorovův – Smirnovův test se nedá použít přímo, protože předpokládá, že ověřujeme shodu našich dat s rozdělením, u kterého známe střední hodnotu a rozptyl. Ty se však většinou odhadují z dat samotných. Pro tento případ lze použít Lillieforsův test, který je modifikací Kolmogorovova – Smirnovova testu.

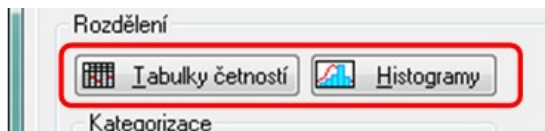
Klávesa **FI** v políčku pro zaškrtnutí příslušného testu vyvolá nápovědu k tématu a doporučení k jednotlivým testům:



Shapiro-Wilk W Test

The *Shapiro-Wilk W test* is used in testing for normality. If the *W statistic* is significant, then the hypothesis that the respective distribution is normal is rejected because of its good power properties as compared to a wide range of alternative tests (Shapiro, Wilk, & Chen, 1968). *STATISTICA* implements the *Shapiro-Wilk's W test* for normality. If the *W statistic* is significant, then the hypothesis that the respective distribution is normal is rejected because of its good power properties as compared to a wide range of alternative tests (see Shapiro, Wilk, & Chen, 1968). The algorithm implemented in *STATISTICA* can be applied to samples with up to 5000 observations (e.g., see 1992); if there are more than 5000 observations, then this test cannot be applied.

Shapiro-Wilkův test je zde upraven i pro relativně velké vzorky (5tis.). Po zaškrtnutí testu mám na výběr dvě možnosti reprezentace výsledku testu:



V modulu *Statistika - Prokládání rozdělení* se počítá test chí-kvadrát. Oboustranný či jednostranný T-test pro dva výběry pouze na základě statistik (průměry, směrodatné odchylky a rozsahy výběrů) je dostupný přes volbu **Základní statistiky a tabulky – Testy rozdílů: r, %, průměry**.

Následující příklad slouží k ověření normality vybraných veličin:

Příklad - Normalita a důležitost náhodného výběru

Úkol: Vytvoříme novou tabulku s proměnnou, která bude mít normální rozdělení. Ověříme její vlastnosti a otestujeme, zda jde skutečně o normální rozdělení. Vytvoříme náhodný a nenáhodný výběr a porovnáme výsledky. Poté v souboru *SpotřebaAut.sta* ověříme normalitu u proměnných *Zrychlení* a *Hmotnost*.

1. Vytvoříme novou tabulku o rozměrech *1s krát 1000 ř.* Zvolíme **Soubor - Nový - Tabulka. Počet proměnných 1 a Počet případů 1000**.
2. Poklepnáním na záhlaví se otevře dialog **Proměnná 1**, kam zadáme informace o proměnné: nazvěme ji Normální a do pole **Dlouhé jméno** vepíšeme funkci, která proměnnou vyplní. (Viz př. 2, bod 3.) *STATISTICA* disponuje funkcí **RndNormal** s parametrem *x*, který znamená směrodatnou odchylku. Pokud je zaškrtnut Průvodce funkcemi, po napsání = a počátečního písmene funkce program nabízí různé možnosti. Můžeme poklepat na zvolenou funkci a ta se sama vepíše do pole. Poté si můžeme zvolit směrodatnou odchylku a po kliknutí na **OK** se vygeneruje 1000 náhodných čísel z normálního rozdělení o střední hodnotě 0 a zadané směrodatné odchylce.

3. Nyní můžeme provést příslušné testy: Spustíme **Základní statistiky a tabulky - Tabulky četností**. Nejprve se podíváme na **Histogramy** na záložce **Detaily**, kde zadáme, že chceme **Přesný počet intervalů**, a to **10**. Vidíme, že rozdělení v histogramu odpovídá očekávanému normálnímu. Na záložce **Normalita** zadáme, že chceme **Shapirův-Wilksův W test**. Ve výsledné tabulce máme vysokou hodnotu p , takže nemůžeme zamítnout, že by data nepocházela z normálního rozdělení. Na záložce **Popisné** zvolíme **Normální pravděpodobnostní grafy**. Na něm se body vyskytují na přímce.
4. Na záložce **Detaily** dialogu **Základní statistiky a tabulky - Popisné statistiky** kromě nabídnutých možností zaškrtneme ještě **Šikmost** a **Špičatost**. a volme **Výpočet: Popisné statistiky**. V tabulce vidíme, že rozdělení je *symetrické* (šikmost je přibližně 0) a *normálně špičaté* (špičatost také přibližně 0).
5. Soubor vygenerovaných náhodných čísel z normálního rozdělení budeme považovat za celou populaci. Známe její průměr a směrodatnou odchylku. Nyní vytvoříme podsoubor čítající přibližně 50 hodnot z této populace. Volíme **Data - Náhodné vzorkování**. V záložce **Možnosti** vybereme **Výpočet pomocí přibližného počtu**. Na kartě **Jednoduché vzorkování** zvolíme **50** jako **Přibližný počet případů**. Tím se vytvoří nová tabulka s výběrem. Pokud porovnáme popisné statistiky u populace a výběru, shledáváme, že náš výběr slouží jako dobrý odhad pro celou populaci.
6. Nyní původní data setřídíme podle velikosti. Volíme **Data - Setřídít**. Tím se data po **OK** setřídí. Pomocí funkce **Data – Podmnožina** vytvoříme filtr, který vybere prvních 50 případů (klikneme na **Případy**, povolíme filtr a v části **Zahrnout** zadáme čísla případů 1-50). Tím jsme provedli nenáhodný výběr z dat. Pokud nyní porovnáme popisné statistiky u výběru i populace, vidíme, že by naše závěry byly silně zkreslené. Při zkoumání normality výběru se totiž ukáže, že výběr není výběrem z normálního rozdělení.
7. Otevřeme soubor **SpotřebaAut.sta**.
8. Spustíme **Statistika - Prokládání rozdělení**. Zvolíme **Normální**. Nastavíme **Proměnnou Zrychlení**. Pak už jen dáme **Graf pozorovaného a normálního rozdělení**. Na histogramu vidíme shodu s normálním rozdělením, stejně tak chí-kvadrát test ji nezamítá. Ještě by nás zajímal pravděpodobnostní graf. Ten je např. v modulu **Základní statistiky a tabulky – Tabulky četností - Popisné**. I na něm je vidět jasná shoda.
9. V případě **Hmotnosti** vidíme, že histogram neodpovídá normálnímu rozdělení. chí-kvadrát test ji také zamítá. Podíváme-li se na pravděpodobnostní graf, vidíme esovité zakřivení, stejně tak šikmost (0,53) naznačuje pravostranné zešikmení. Tato data nemůžeme považovat za výběr z normálního rozdělení.

11 Jednovýběrový t test

Pres *Statistiky* -> *Základní statistiky/tabulky* -> *t-test, samost. vzorek* se pak dostaneme k jednovýběrovému t-testu, kde definujeme referenční konstantu a klikneme na **Výpočet**:

The dialog box 't-test pro samostatné průměry: OrientacniMereni' is shown with the 'Možnosti' tab selected. The 'Referenční hodnoty' section has 'Test všech průměrů vůči:' set to 100. The 'Výpočet mezí spolehl.' checkbox is checked, with an interval of 95,00%. The 'Výpočet' button is highlighted with a red box.

The results table is displayed in the 'Test průměrů vůči referenční konstantě (hodnotě) (OrientacniMereni)' window. The table shows the mean, standard deviation, sample size, and confidence intervals for the variable 'vzdalenost'.

Proměnná	Průměr	Sm.odch.	N	Sm.chyba	Int. spolehl. -95,000%	Int. spolehl. +95,000%	Referenční konstanta	t	SV	p
vzdalenost	100,0123	0,029300	26	0,005746	100,0005	100,0241	100,0000	2,141918	25	0,042130

Test je signifikantní, zamítáme nulovou hypotézu: $H_0: \mu = \mu_0 = 100m$

Skutečná průměrná naměřená vzdálenost přístroje se s 95% P nachází v intervalu:

Test průměrů vůči referenční konstantě (hodnotě) (OrientacniMereni)		
	Int. spolehl. - -95,000%	Int. spolehl. - +95,000%
vzdalenost	100,0005	100,0241

Je rozdíl také prakticky významný? Má přístroj sys. chybu?

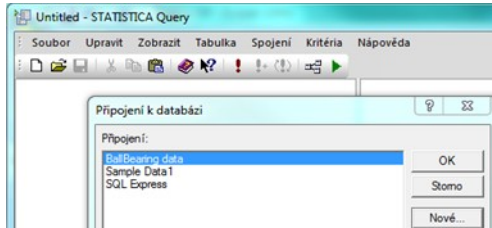
Kompletní řešení příklad na tento test lze najít v našem newsletteru z 08/01/2013 *StatSoft ACADEMY*:

http://www.statsoft.cz/file1/PDF/newsletter/2013_01_08_StatSoft_Test.pdf

12 Připojení do databází pomocí *STATISTICA Query*

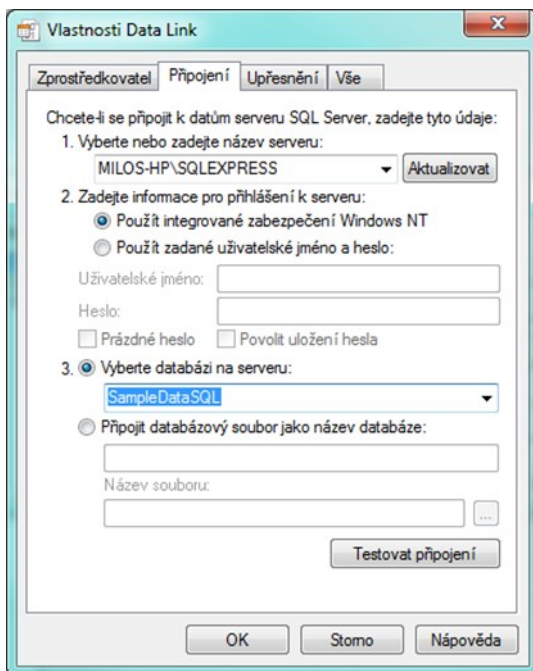
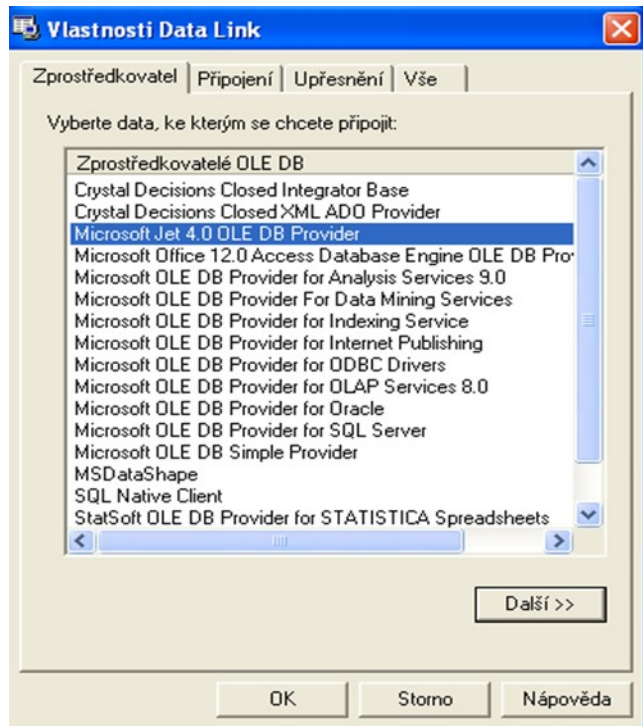
STATISTICA umožňuje přímé připojení do všech standardních databází přes konvence OLE DB a ODBC. Připojení probíhá v několika fázích:

Přes *Soubor - Získat externí data - Vytvořit dotaz* se dostaneme do okna rozhraní *STATISTICA Query*:



Zde tlačítkem *Nové* zvolíme možnost definovat nové připojení. V okně *Vlastnosti Data Link* vybereme vhodnou možnost z dostupných ovladačů pro připojovanou databázi: V dalším kroku vybereme server, u některých

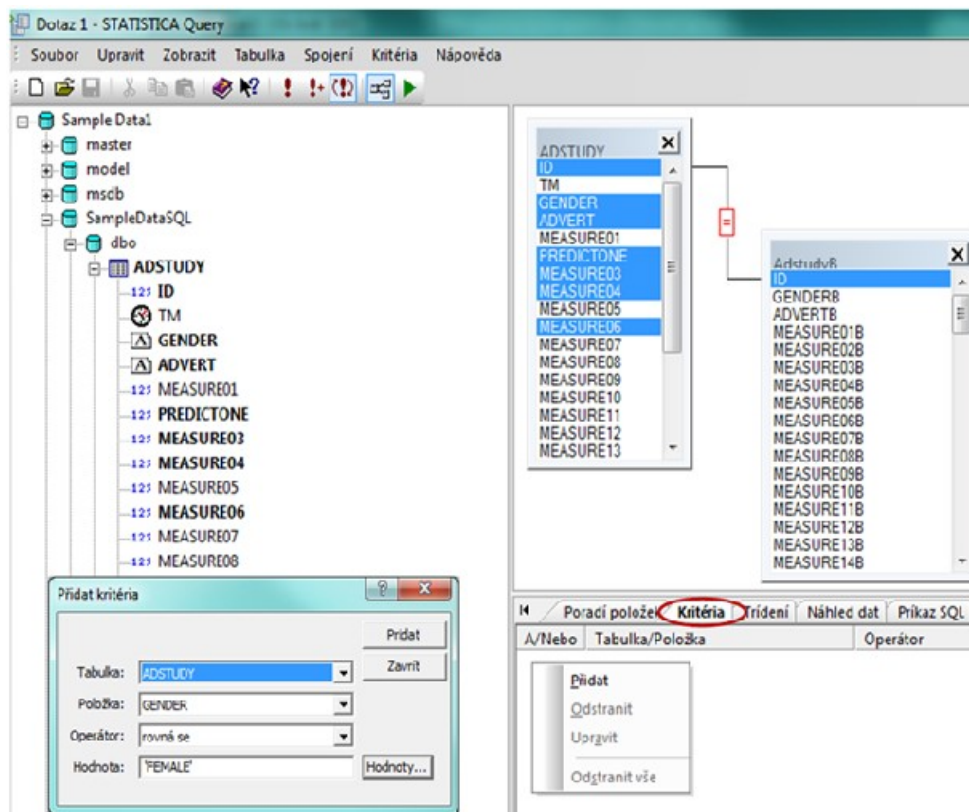
typů ovladačů, resp. databází, musíme cestu zadat ručně (např. *Access - Jet.OLEDB.4.0*). Dále zvolíme typ zabezpečení pro přístup do databáze a v rolovacím menu vybereme konkrétní databázi na serveru, který jsme definovali předchozím kroku. Vhodné je také otestovat připojení a v dalším kroku zvolíme název pro nové připojení, máme možnost zobrazit náhled připojovacího řetězce.



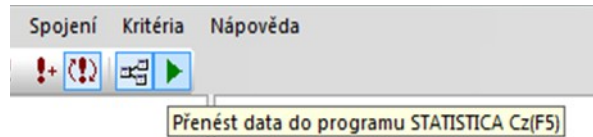
Práce v rozhraní *STATISTICA Query*

V rozhraní *STATISTICA Query* lze pracovat dvěma způsoby. První způsob využívá **grafický režim** a umožňuje práci i těm, kteří potřebují z databáze získávat konkrétní data, ale nemají potřebné znalosti dotazovacího jazyka SQL. Grafický režim funguje na principu „Táhni a pusť“. V levé části hlavního okna vidíme jednotlivé tabulky v databázi (na obrázku je to např. *ADSTUDY*), které lze přetáhnout do hlavního okna v pravé části menu. Kliknutím na jednotlivé názvy polí tabulky v hlavním okně (*ID*, *GENDER*...) vybereme, která pole z databáze chceme nahrát a automaticky tak již vytváříme SQL dotaz, který můžeme ve spodní části okna také nechat zobrazit (**Příkaz SQL**). Tlačítko **Náhled dat** umožňuje sledovat vybraná data.

Spojení tabulek je převzato z databáze, anebo jej lze nadefinovat přímo v prostředí *STATISTICA Query*, a to přetažením kurzoru z jedné tabulky na druhou (na konkrétním parametru, který slouží jako primární klíč), nebo přes záložku **Spojení – Přidat**. Možnost přidat spojení vyvoláme také kliknutím pravého tlačítka myši ve volném prostoru hlavního okna. Kliknutí ve spodní části rozhraní *STATISTICA Query* (viz následující obrázek) vyvoláme možnost přidání doplňkových **omezení** pro jednotlivé parametry.



Chceme-li upřesnit již vygenerovaný SQL dotaz či napsat nový bez využití grafického módu, přes záložku **Zobrazit** přepneme **grafický režim** na skriptovací. Přes záložku **Soubor – Uložit jako/Otevřít** lze hotové dotazy ukládat a načítat. Samotné spuštění dotazu probíhá přes zelenou ikonu v horní liště, nebo přes klávesu **F5**.

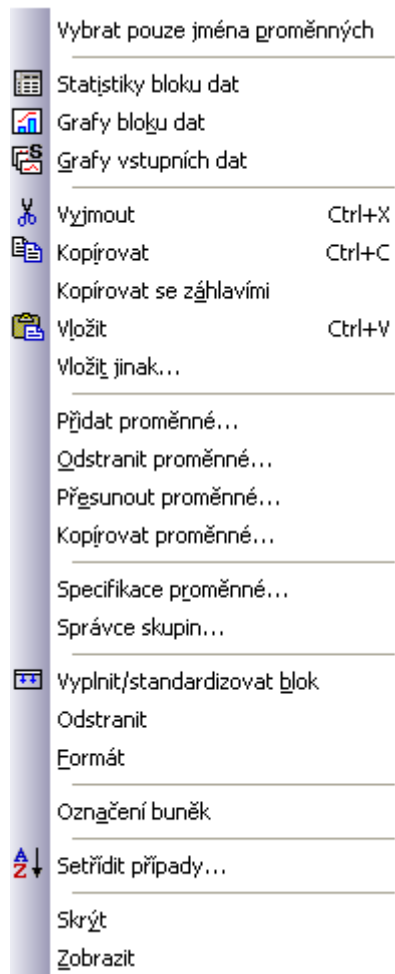


Defaultní nastavení *STATISTICA* je načítat data do aktivní tabulky dat, pokud chcete načíst data do nové prázdné tabulky, vyberte tuto možnost v následujícím dialogu:



13 Úprava načtených dat

Proměnné a případy



Přidání a odebrání proměnných provedeme následujícím způsobem: V záhlaví tabulky klikneme pravým uchem myši a zobrazíme dialog, v němž můžeme vybrat možnost **Odebrat proměnné...** nebo **Přidat proměnné**. Při přidávání proměnných se zobrazí dialog, v němž

uživatel specifikuje počet přidávaných proměnných, název proměnné, ze které se mají nové proměnné vložit, jméno proměnné (Pokud přidáváme více než jednu proměnnou,

bude zadaný název použit u všech těchto proměnných – pro odlišení bude ukončen pořadovým číslem přidávané proměnné. Přejmenování proměnných můžeme nicméně provést následně.), typ hodnot proměnné a způsob zobrazení jejich hodnot. Rozlišujeme čtyři typy hodnot proměnných, a sice:

- **Double**
Defaultní typ. Využívá se pro numerické hodnoty a umožňuje ukládat 64 bitová reálná čísla s přesností na 15 desetinných míst. Rozsah přibližně od $-1,7 * 10^{308}$ do $1,7 * 10^{308}$. Kód *chybějících dat* je -999999998.
- **Integer**
Celá čísla v rozmezí -2 147 483 648 a 2 147 483 647. Každé číselné hodnotě lze přiřadit textový popis. Velikost 4 byty.
- **Byte**
Celá čísla v rozmezí 0 až 255, nelze vložit desetinná čísla, každé číselné hodnotě lze přiřadit textový popis. Velikost 1 byty.
- **Text**

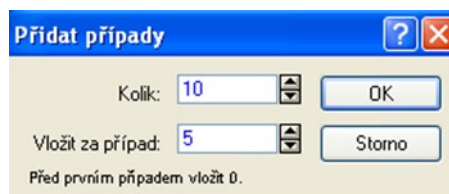
Textové řetězce s neomezenou délkou bez číselné reprezentace. Pro účely numerických výpočtů jsou různým řetězcům přiřazeny *ad-hoc* různé číselné hodnoty.

Kód *chybějících dat* je *prázdný řetězec*.

Přiřazením vhodného typu můžeme šetřit místo nutné pro uložení datové tabulky v paměti počítače.

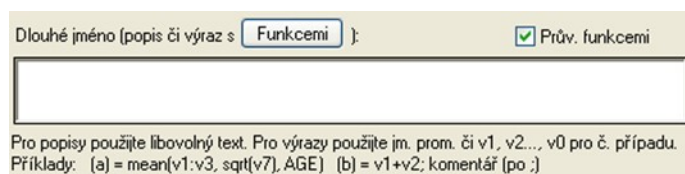
Do okna dialogu pro přidání proměnných s názvem **Dlouhé jméno** je možné vkládat matematické, statistické, logické, textové ale i jiné funkce, jejichž vstupem jsou ostatní proměnné tabulky, nicméně vkládání těchto funkcí doporučujeme provádět až po přidání proměnných. Pokud se funkce odkazují na proměnné, které se v tabulce vyskytují až za přidávanými proměnnými, nejsou odkazy pomocí písmene *v* a čísla sloupce proměnné jednoznačné.

Při přidávání případů je potřeba zadat, jen kolik řádků chceme do tabulky přidat a za který řádek se mají vložit:

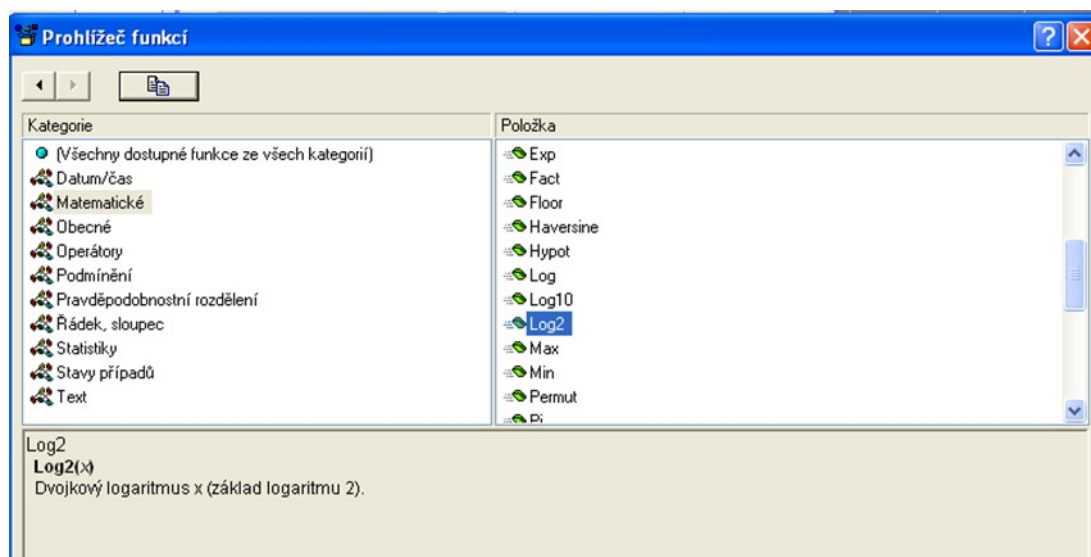


Transformace dat

Pro transformaci dat je ideální nadefinovat novou proměnnou, která bude funkcí proměnných původních. V záhlaví tabulky klikneme dvakrát na název nové proměnné a v dialogu podobném dialogu pro přidávání proměnných klikneme v dolní části na tlačítko Funkcemi.



Zobrazí se **Prohlížeč funkcí**, kde jsou dostupné všechny funkce, které jsou ve *STATISTICA* definovány. Můžeme je vybírat v levé části okna prohlížeče podle jejich typu, v pravé části



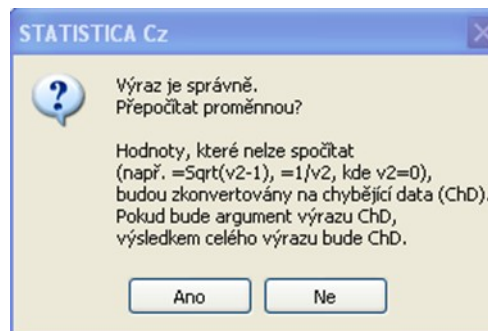
okna potom vybereme konkrétní funkci a v dolní části okna se zobrazí nápověda k vybrané funkci (popis toho, co funkce dělá a jaké má vstupní parametry).

Odkaz na jiné proměnné tabulky se tvoří buď použitím názvu proměnné (pokud název obsahuje mezery, je třeba ho uvádět v uvozovkách) anebo užitím písmena *v* a čísla sloupce

proměnné (například $v8$ odkazuje na proměnnou v osmém sloupci tabulky). Výraz $v0$ označuje pořadová čísla řádků (případů).

Zápis transformace pro novou proměnnou může vypadat například takto:

Potvrdíme volbu tlačítkem *OK*, *STATISTICA* zobrazí ještě dialog, kde odsouhlasíme přepočítání hodnot nové proměnné:



Použití filtru

Nejpohodlnější je nejspíš použití filtru při samotném volání analýzy nebo tvorbě grafu. V pravé části některého z úvodních dialogů je

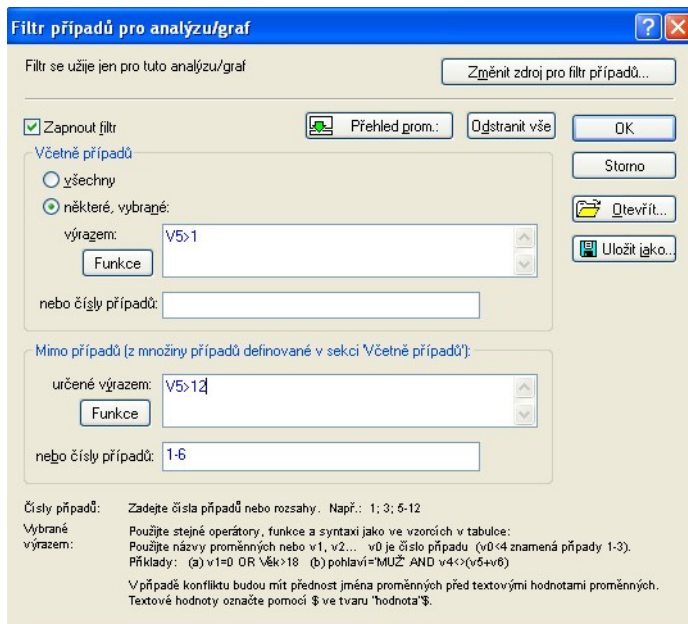
umístěno



tlačítko

SELECT CASES.

Pomocí něj zobrazíme dialog, v němž je třeba zatrhnout možnost **Zapnout filtr**. Tím se zpřístupní pole pro zadání podmínek pro zahrnutí nebo vyloučení některých řádků tabulky. Pro názornost uvádíme následující příklad zadání podmínek filtru.



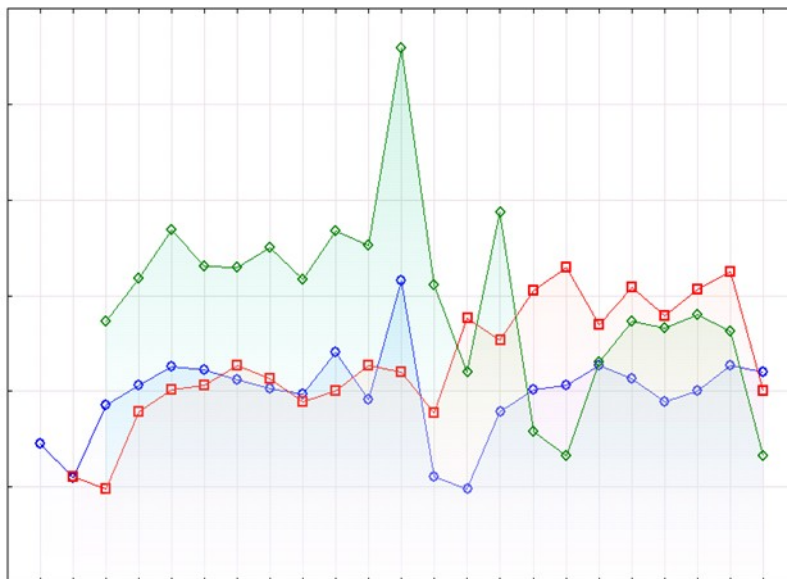
Do analýzy budou zahrnuty případy, pro které je splněna podmínka, že hodnota proměnné v pátém sloupci je větší než 1, a vyloučeny budou řádky 1 až 6 a dále ty případy, které sice splňují podmínku $v5 > 1$, ale u nichž je hodnota páté

proměnné větší než 12.

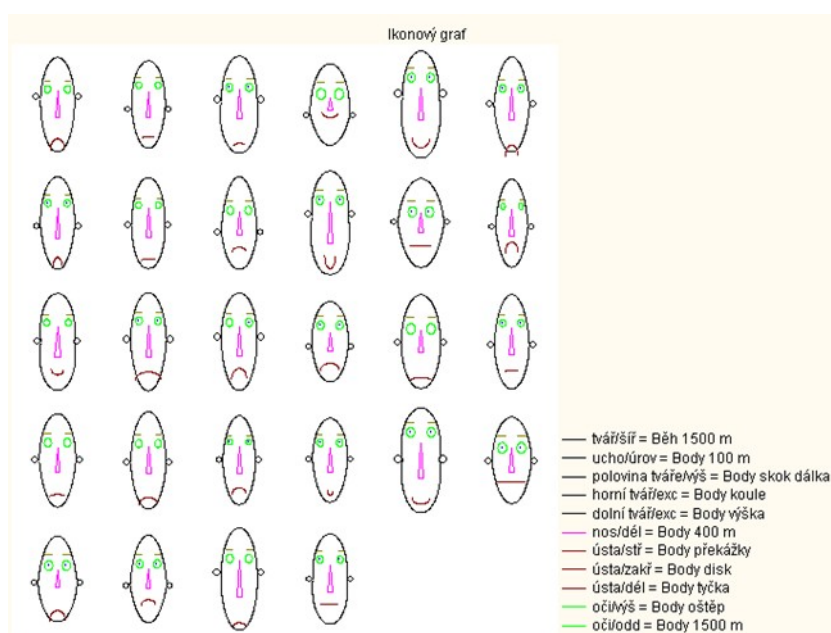
14 Další grafické možnosti softwaru STATISTICA

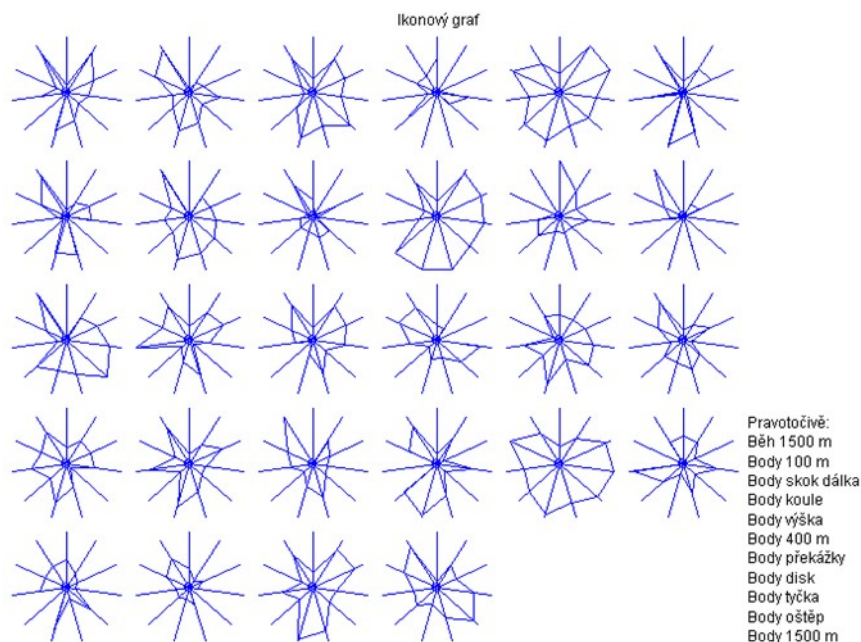
Kompletní nabídka grafů je k dispozici ze základní nabídky **Grafy**.

Software *STATISTICA* obecně poskytuje bohaté možnosti, co se týče grafů. Ve verzi *STATISTICA 10* a vyšší lze v porovnání se staršími verzemi programu nově upravovat i *průhlednost oblastí* nebo *značek* v grafech a tím získat například následující kvalitní výstup pro prezentaci:

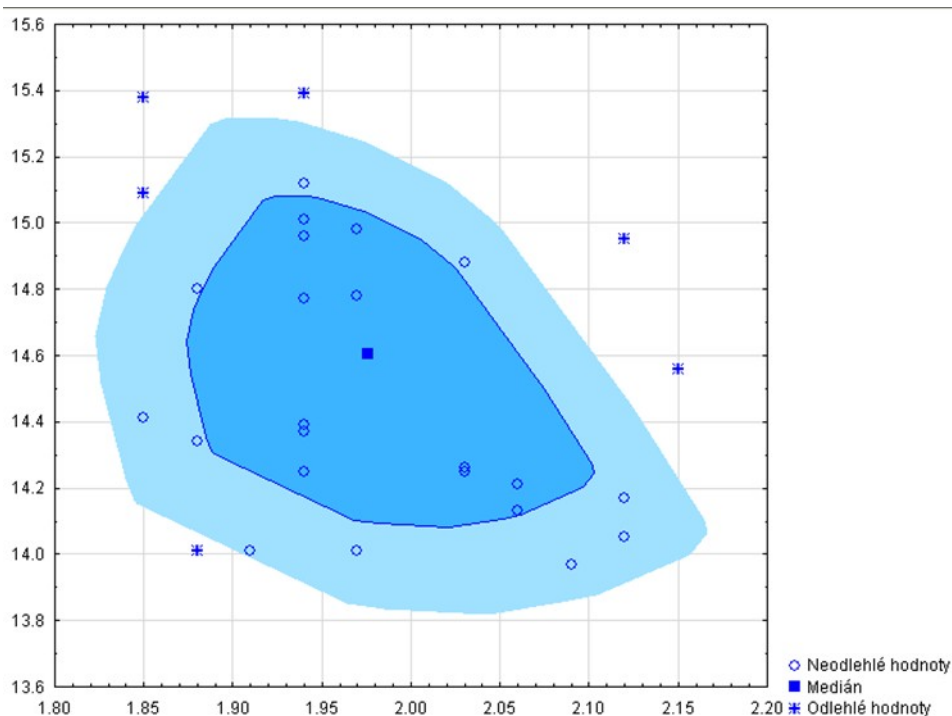


Ze zajímavých grafů, kterými *STATISTICA* disponuje, uvedme **ikonové grafy**, které slouží pro zobrazení a porovnání jednotlivých případů (řádků tabulky) podle měřených veličin (proměnných, sloupců tabulky),

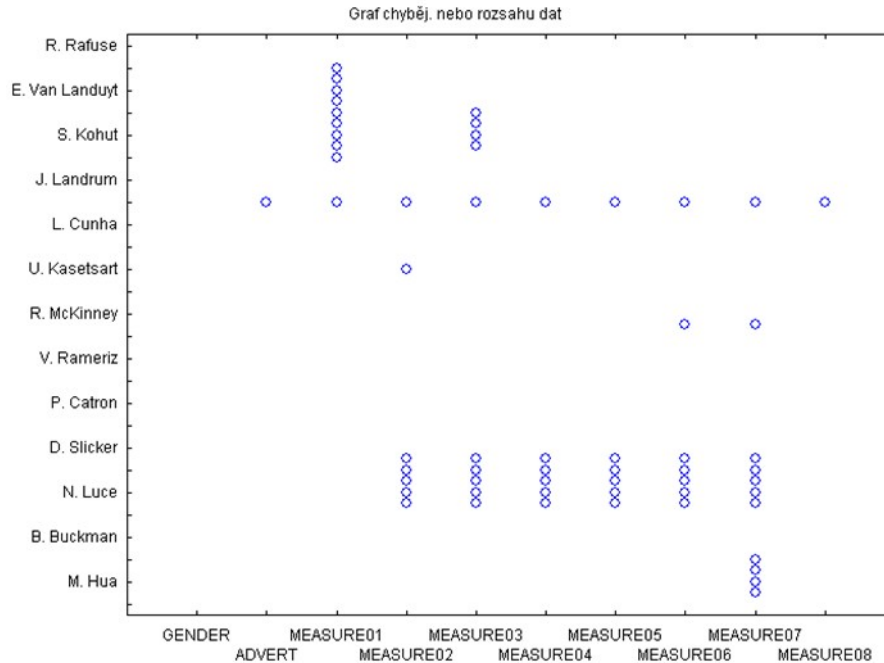




dále tzv. **bag ploty**, které jsou dvourozměrnou variantou krabicových grafů. Umožňují identifikaci odlehlých porozování vzhledem ke dvěma dimenzím současně,



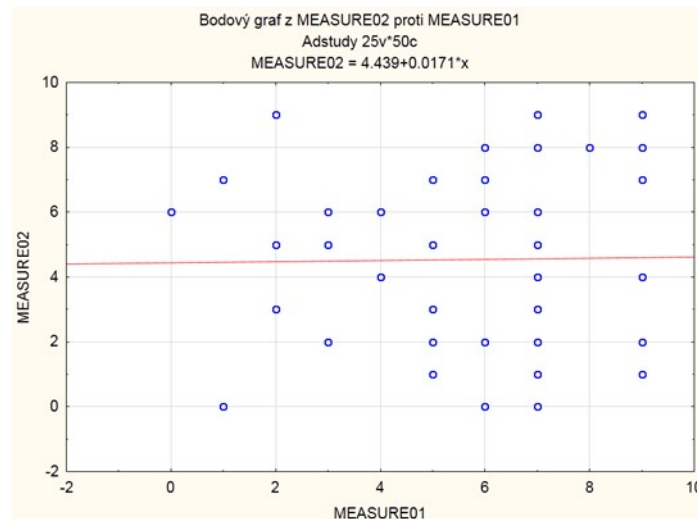
a grafy chybějících hodnot, které uživatele rychle upozorní na ty řádky nebo proměnné, které obsahují větší množství chybějících dat (body vynesené do grafu představují právě chybějící údaje pro proměnnou danou osou x a případ uvedený na ose y):



Použití grafů je zcela intuitivní, proto se v dalším textu budeme věnovat pouze pokročilé práci s grafickými styly.

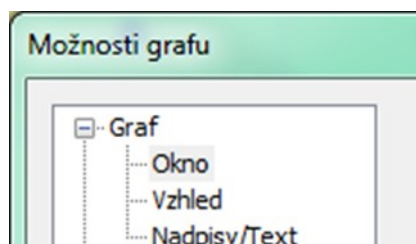
Styly grafů

V záložce **Soubor - Otevřít příklady** ve složce **Datasets** vybereme soubor *Adstudy.sta*
 V záložce **Grafy - 2D Grafy - Bodové grafy** zvolíme *MEASURE1* jako proměnnou X a *MEASURE2* jako proměnnou Y.



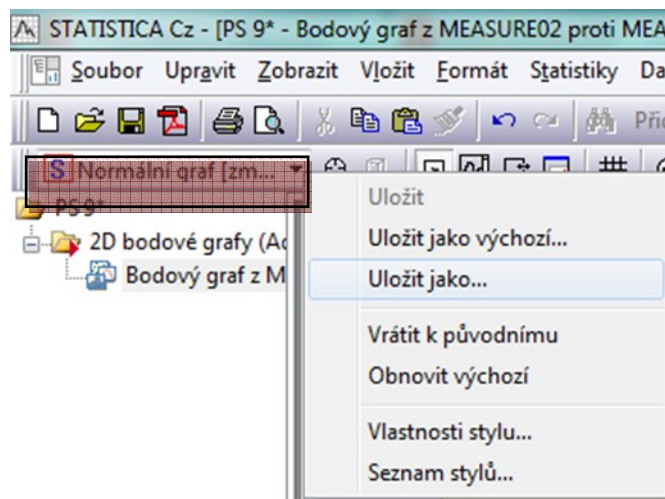
Z grafu je vidět, že obě měření jsou vzájemně nezávislá. Nyní graf přizpůsobíme podle potřeby uživatele, například:

Dvojklikem do prázdného okna grafu vyvoláme nabídku **Možnosti grafu**, kde zvolíme možnost **Okno** a změníme barvu vnitřního pozadí na světle žlutou a vnějšího pozadí na světle modrou.

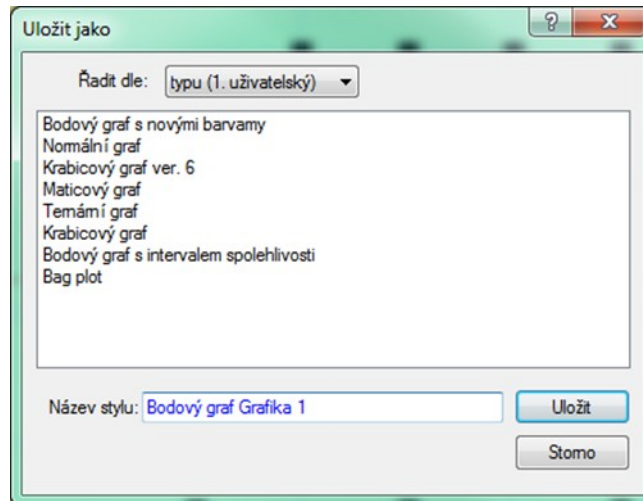


Dále v záložce **Spojnice - Obecné** změníme značku z prázdného modrého kroužku na plný černý čtverec. Nyní máme bodový graf, který odpovídá našim představám, a jehož styl chceme opakovaně využívat i pro další bodové grafy.

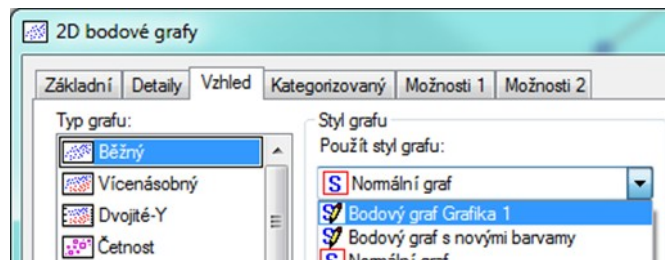
V levém horním rohu v záložce **Grafické styly** přes pravé tlačítko myši volíme možnost **Uložit jako...**



Styl grafu pojmenujeme, název by měl indikovat i typ grafu, protože použití stylu bodového grafu například na histogram může mít nežádoucí výsledek. Zvolíme tedy název *Bodový graf Grafika 1*.



Při tvorbě dalšího bodového grafu potom v záložce *Vzhled – Styl grafu* vybereme právě uložený styl.

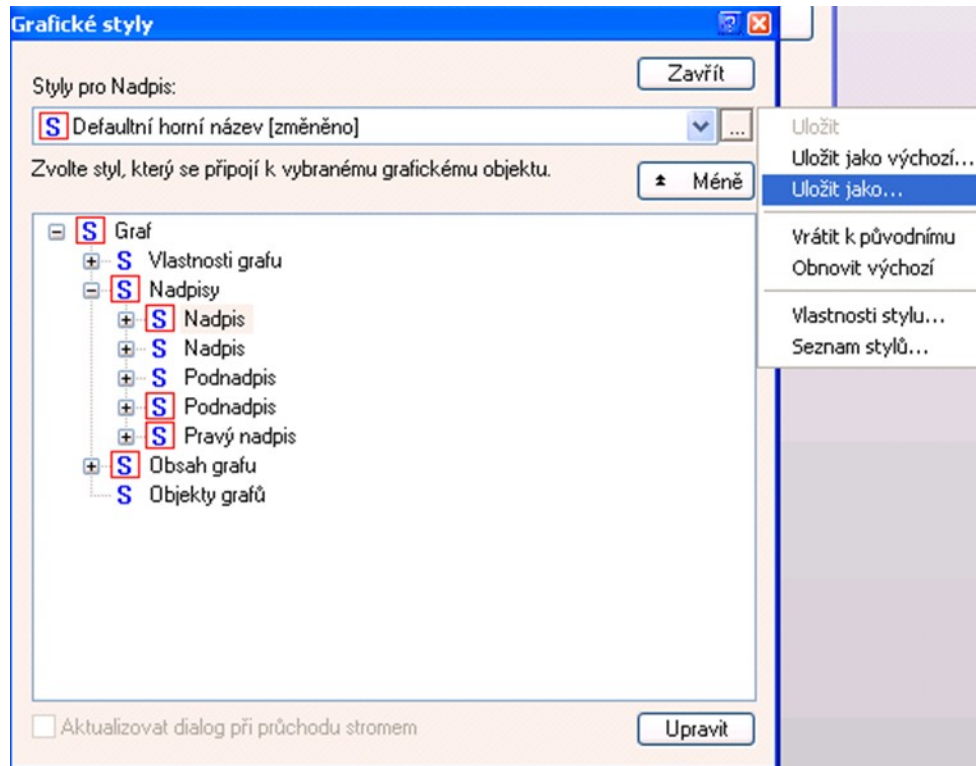


Pro již vytvořený graf lze také nastavit konkrétní styl, a to výběrem z rolovacího seznamu *Grafické styly* umístěném v levém horním rohu okna sešitu s grafem.

Přidání stylů pro grafické objekty

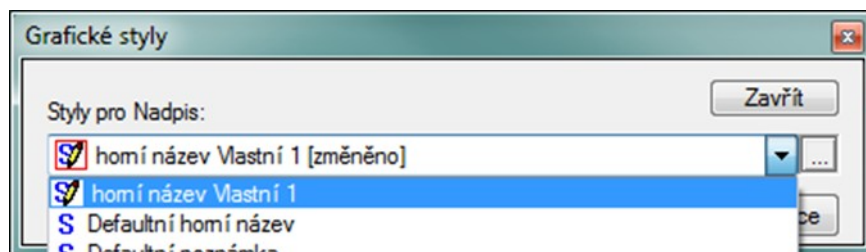
Ve *STATISTICA* lze ukládat i nastavení objektů, jako jsou šipky, obdélníky, ovály a jiném uživatelem dokreslené automatické tvary, které lze upravit skrze nastavení dostupná v panelu *Nástroje grafu*.

Jednotlivé atributy jsou označeny jako (A), skupiny atributů jako (AA) a skupiny vlastností symbolem (S). V grafu upravme například nadpis – dvojklikem na příslušný nadpis zobrazíme okno pro jeho editaci (zde provedeme změnu barvy, nastavení velikosti a fontu písma, atd.). Vzhled objektu nadpis budeme chtít dále opakovaně využívat. Klikneme na tlačítko *Styly...* dále klikneme na tlačítko *Více...* Klikneme myší na tlačítko se třemi tečkami vedle názvu stylu v okně *Styly pro nadpis* a vybereme možnost *Uložit jako...*



Případně volba ***Uložit jako výchozí*** způsobí, že námi upravený styl bude použit pro tento objekt u všech nově vytvářených grafů.

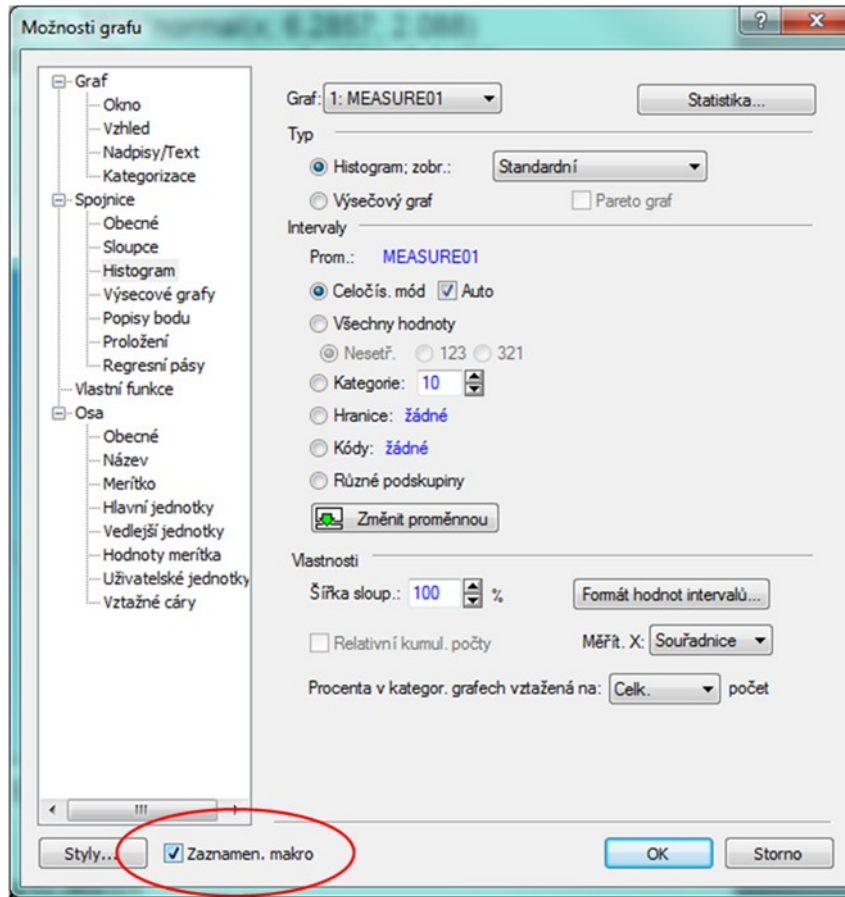
Pro aplikaci uživatelem definovaného stylu na konkrétní objekt v novém grafu vybereme ze stejné roletky jako výše dříve uložený styl.



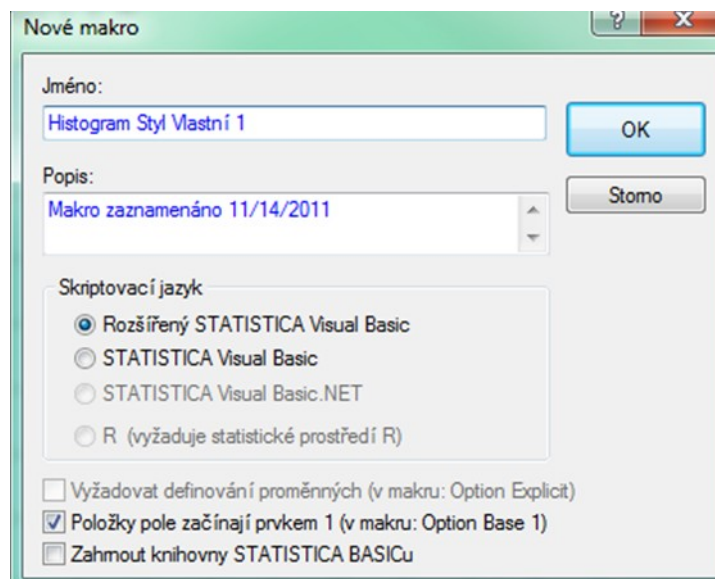
Pokročilejší uživatelé mohou definovat a používat grafické styly pro změnu jednotlivých grafických vlastností nebo hierarchicky organizovaných skupin vlastností, a to pomocí dialogu úprav ***Všechny možnosti***.

Automatické nastavení vzhledu grafu

V záložce ***Grafy - 2D Grafy - Histogram*** vybereme ze souboru *Adstudy.sta* jako závisle proměnnou *MEASURE01* a pro kategorizaci vybereme proměnnou *GENDER*. Dvojklikem do prostoru grafu vyvoláme dialog ***Možnosti grafu*** a v dolní části dialogu zaškrtneme ***Zaznamenat makro***



Dále provedeme úpravy histogramu podle libosti, to lze provádět buď přímo v dialogu **Možnosti grafu** anebo poklepaním na objekt grafu, který chceme upravit, což má za následek otevření dialogu **Možnosti grafu** přímo s volbami pro vybraný objekt, kde lze měnit jeho parametry. Změníme například barvu pozadí na světle žlutou, v kolonce **Nadpisy/Text** upravíme nadpis grafu, počet kategorií v kolonce **Kategorizace** upravíme na 8, v kolonce **Spojnice - Sloupce** upravíme barvu oblasti sloupce, atd. Klikneme na **OK** v pravém dolním rohu dialogu **Možnosti grafu**.



Zadáme název pro makro *Histogram Styl Vlastní 1* a potvrdíme **OK**, čímž zaznamenané makro uložíme. Nyní vytvoříme úplně jiný graf typu histogram. Analogické úpravy grafu nyní již není třeba provádět ručně – pro jejich aplikaci otevřeme dříve uložené makro *Histogram Styl Vlastní 1* a klávesou **F5** makro spustíme. V pracovním sešitě je poté k dispozici graf se vzhledem odpovídajícím vytvořené šabloně.

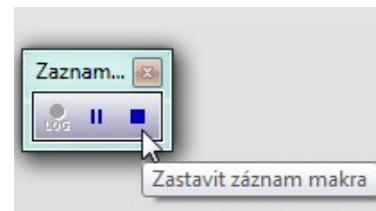
15 Automatizace rutinních analýz

Následující postup ukazuje tvorbu jednoduchého makra pro automatizaci rutinních činností. Software *STATISTICA* umožňuje vytvářet různé dávkové analýzy pomocí integrovaného jazyka *STATISTICA Visual Basic* (SVB), který lze využít ke zjednodušení prováděných úloh různé obtížnosti, od jednoduchých maker až po pokročilé projekty. Pomocí jazyka SVB může uživatel přistupovat prakticky ke každému funkčnímu prvku systémů a tedy i využívat vlastní rozšíření systému.

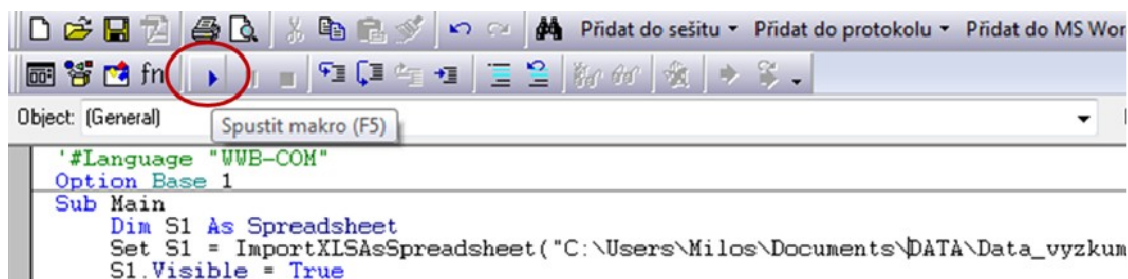
Všechny postupně prováděné analýzy lze snadno automaticky zaznamenávat pomocí záznamu makra. Tímto jednoduchým záznamem potom zcela automatizujeme často se opakující analýzy, a to i bez znalosti programování. **Postup tvorby záznamu makra je následující:**

Před vlastním spuštěním záznamu je třeba zvážit, zdali chceme provádět automatizovanou analýzu vždy nad již načtenou aktivní tabulkou *STATISTICA*, anebo bude načtení aktuálních dat také součástí kódu. V druhém ze zmíněných případů začneme nahrávat nejprve samotné otevírání příslušné tabulky.

Dále vybereme menu **Nástroje - Makro - Spustit záznam průběhu analýzy (hlavní makro)**. Nyní provedeme požadovanou posloupnost analýz nebo vytvoříme grafy, které dále upravujeme a podobně. Záznam ukončíme kliknutím na tlačítko **Zastavit záznam makra** na minipanelu, který se otevřel v okamžiku spuštění nahrávání makra,



nebo v menu **Nástroje - Makro - Zastavit záznam**. V následujícím dialogu si makro pojmenujeme a potvrdíme **OK**. Nyní máme k dispozici zaznamenaný kód, který můžeme upravit a následně uložit prostřednictvím nabídky **Soubor -> Uložit/Uložit jako...** Makro spustíme pomocí tlačítka **Spustit makro**, které je dostupné na hlavním panelu v okamžiku, kdy je aktivní okno s kódem makra, případně můžeme použít klávesu **F5**.



Všimněme si, že v příkladu zobrazeném na obrázku, je v kódu uložena cesta k souboru *Data_vyzkum*. Při spuštění makra proto bude vždy načtena aktuální verze tohoto souboru a analýzy se provedou nad aktuálními daty. Pokud bychom makro spustili již nad otevřenou tabulkou (Spreadsheet), v záznamu byl tento kód:

```
Dim S1 as Spreadsheet
Set S1 = ActiveDataSet
```

Makro by pak využívalo (a vyžadovalo) nějakou již otevřenou aktivní tabulku v aplikaci *STATISTICA*.

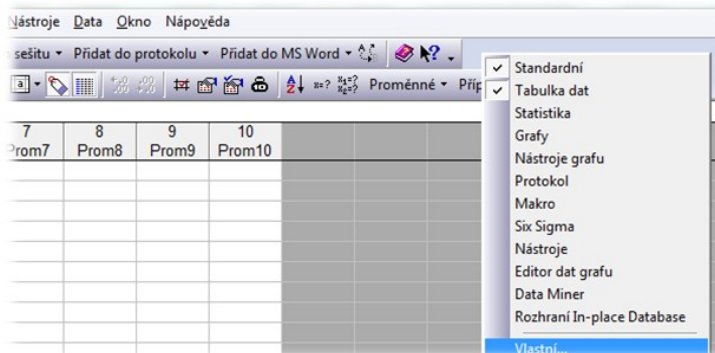
16 Vlastní menu v *STATISTICA*

Software *STATISTICA* je rozsáhlý modulární program, kde jednotlivé moduly nabízejí velkou řadu statistických metod, naši uživatelé obvykle využívají analýzy z různých modulů. Proto si ukážeme možnosti, jak si práci zjednodušit a přizpůsobit si záložky programu podle svých představ.

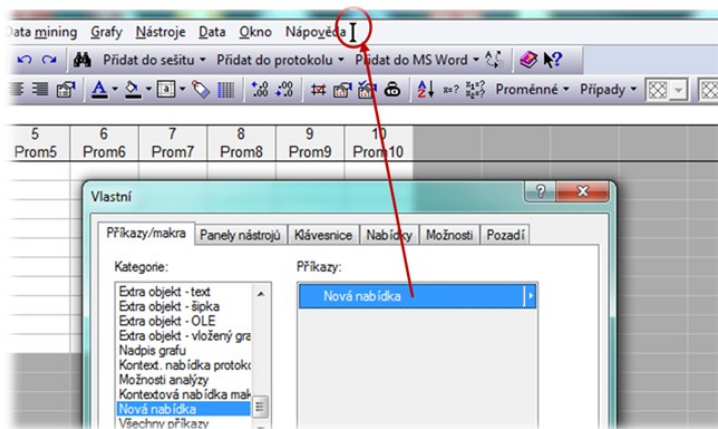
Často využívané analýzy a funkcionality softwaru, stejně jako vlastnoručně vytvořená makra si lze uspořádat do vlastního rozevíracího „menička“. Postup tvorby vlastního menu v softwaru je následující:

Postup

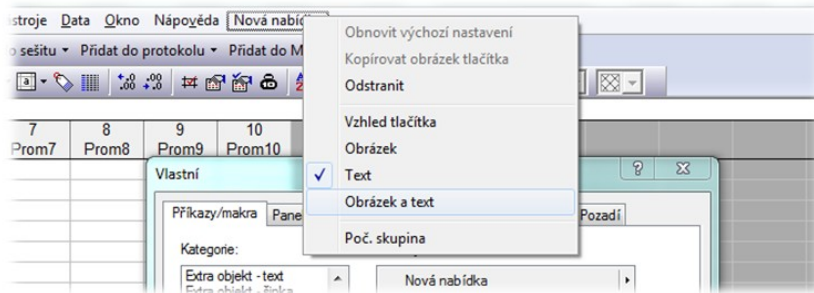
V rozhraní **Klasické nabídky** klikneme (pravým tlačítkem myši) na volnou část horní lišty.



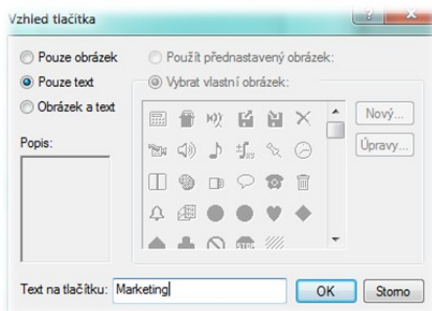
V dialogu *Vlastní* v oblasti **Kategorie** vyhledáme volbu **Nová nabídka** a z části **Příkazy**, na principu „táhni a pusť“, přetáhneme funkcionalitu **Nová nabídka** do horní lišty, tedy do základní nabídky analýz.



Dialog *Vlastní* necháme otevřený, pravým tlačítkem myši klikneme na nově vytvořenou záložku a otevřeme roletku:

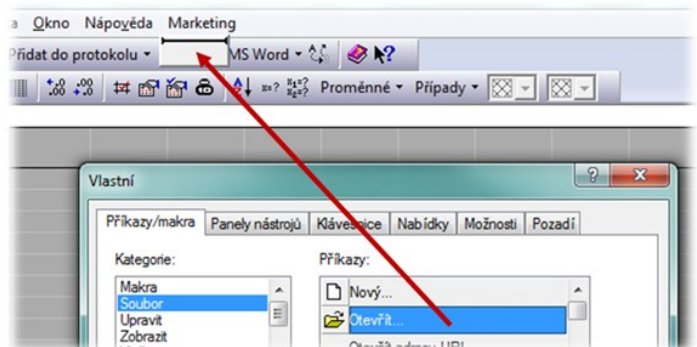


V roletce vybereme *Obrázek a text* a otevřeme dialog *Vzhled tlačítka*.

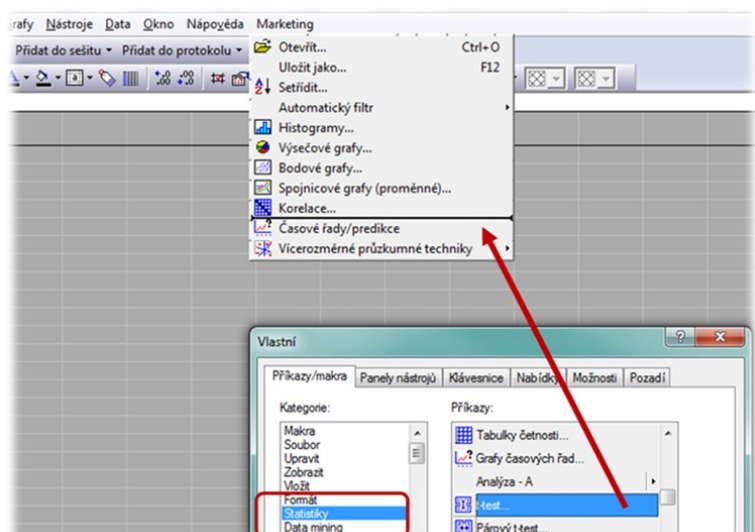


Záložka může být reprezentována textem, obrázkem anebo kombinací obou. Vybereme *Pouze text* a v okně *Text na tlačítku* napíšeme název budoucí roletky: Název může reprezentovat oblast, pro kterou se dané analýzy hodí, anebo, jak jsme si všimli u našich zákazníků, název konkrétního souboru.

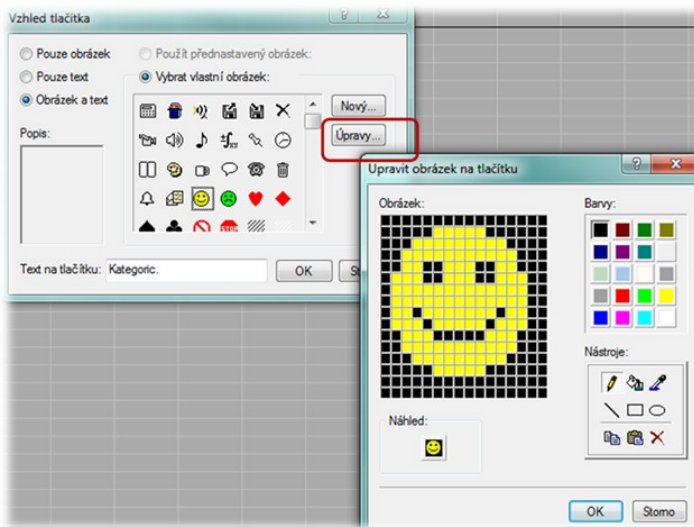
Nyní máme na liště nové tlačítko, v dialogu *Vlastní* v části *Kategorie* si vybereme funkcionality, které chceme do roletky přetáhnout. Funkcionality přetahujeme opět na principu „táhni a pusť“:



Pokračujeme například metodami. V části *Statistiky* si přetáhneme analýzy a oblasti analýz, které chceme ve své roletce mít:

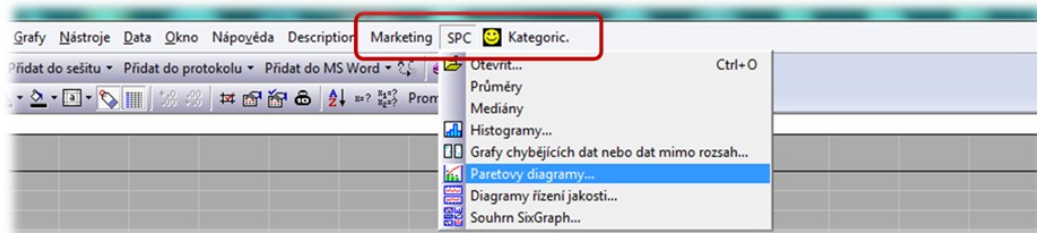


A nabídka je hotová.



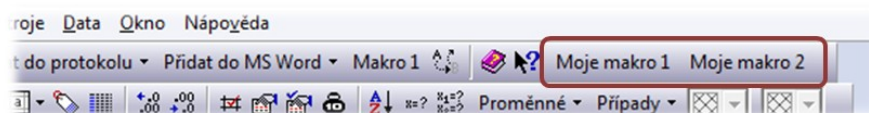
Pokud vybereme v nabídce **Obrázek a text**, tak si můžeme vybrat z celé řady přednastavených obrázků, tlačítkem **Úpravy** potom vyvoláme dialog pro úpravu obrázků.

Výsledkem potom může být nová sada záložek, která obsahuje funkcionality, jež uživatel nejčastěji používá na různé typy datových souborů.



Další možnosti

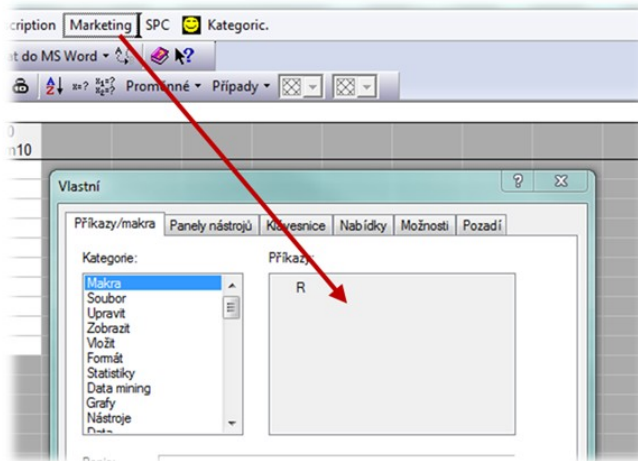
Pokud potřebujete přidat jen jednotlivé často používané funkcionality, pak nemusíte vytvářet přímo celé nové menu. Pokud Vám to více vyhovuje, můžete přidat do volných oblastí nabídek často používané analýzy ve formě tlačítka. Tato tlačítka umístíte jednotlivě přetažením do volné oblasti klasického menu. Následuje obrázek s dvěma vlastními makry (*Moje makro 1* a *Moje makro 2*), které se po kliknutí na tato tlačítka automaticky spustí:



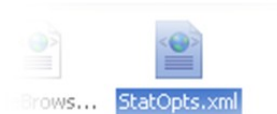
Smazání záložek

Smazání záložek lze dělat dvěma způsoby:

1. Při otevřeném dialogu *Vlastní* opět na principu „táhni a pusť“ přetáhneme konkrétní menu do oblasti *Příkazy* v dialogu *Vlastní*.

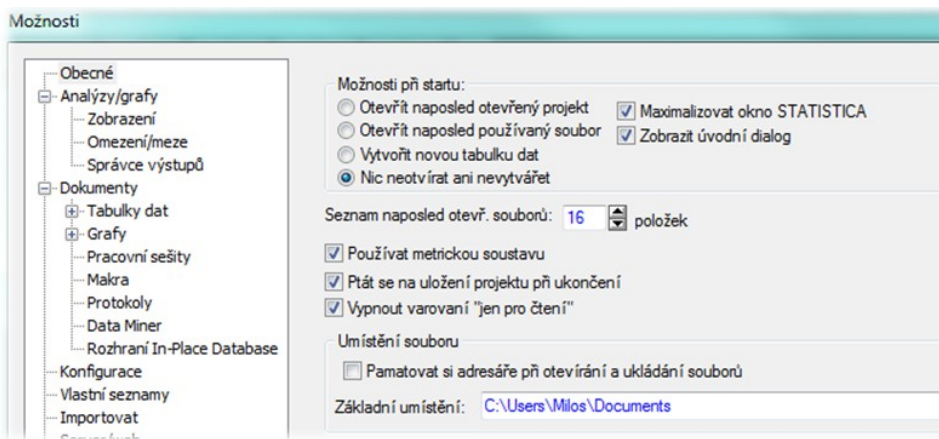


2. Druhou možností jak vrátit původní nastavení je smazání konfiguračního souboru **StatOpts.xml**, ve kterém jsou tato nastavení uložena. Dle Windows je uložen v:



C:\Documents and Settings\Uživatel (jméno účtu)\Data aplikací\StatSoft\STATISTICA\10.0.0.0.1029 pro Win XP nebo v: *C:\Users\jméno účtu\AppData\Roaming\StatSoft\STATISTICA\10.0.0.0.1029* pro Win 7/Vista.

Pozor! Smazání konfiguračního souboru odstraní veškerá nastavení, která byla v softwaru *STATISTICA* nakonfigurována, např. ve **Správci výstupů**:



Nastavení se netýkají připojení do databáze, které je uloženo v systémových registrech. Konfigurační soubory **StatOpts.xml** s konkrétními nastaveními lze zálohovat a např. dávkovými příkazy na úrovni Windows je potom na konkrétním PC měnit, je-li žádoucí, aby

se nastavení softwaru *STATISTICA* na konkrétním PC pod konkrétním uživatelem měnilo (např. v počítačové učebně). Konfigurační soubor je v rámci jedné verze přenosný mezi jednotlivými počítači, resp. mezi operačními systémy. Pro zobrazení uvedených složek je nutné v PC povolit **Zobrazení skrytých souborů a složek**.

17 Závěr

Tento manuál stručně popisuje základní ovládání programu *STATISTICA*. Jednotlivé kapitoly přinášejí „kuchařku“, která usnadní orientaci v prostředí programu *STATISTICA* a provede uživatele analýzou dat od jejich vstupu do programu až po finální výsledky. Zbývá tedy jen popřát, ať se vám práce s programem *STATISTICA* daří podle vašich představ!