Ovládání a základy statistiky v softwaru STATISTICA



Copyright © StatSoft CR s.r.o. 2013

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.

SIAINSINGA

Obsah:

Obsah		2
1	Spuštění programu STATISTICA	4
2	Načtení souboru	6
Přík	ad – import dat z Excelu	6
3	Zpracování chybějících dat	8
4	Vytvoření základní výpočtů1	5
4.1	Tabulka četností	5
4.2	Popisné statistiky1	7
4.2.1	Soubor Temperat CZ.sta - měření dílů jednotlivými operátory1	7
4.2.2	Editace tabulky (Anglické popisky apod.)1	9
4.2.3	Rozdělení spojité proměnné dle kategorie2	1
5	Vytvoření grafu	2
5.1	Histogram	2
5.2	Krabicový graf (Box Plot)	2
6	Uložení práce	3
6.1	Uložení celého sešitu výstupů	3
6.2	Uložení tabulky v softwaru	4
6.3	Uložení grafu	5
6.4	Přidání výstupů do Protokolu/Microsoft Wordu2	6
7	Další možnosti načtení souborů	1
7.1	Otevření textového souboru	1
8	Správce výstupů	3
8.1	Výstup do Microsoft Word / do protokolu STATISTICA	3
9	Ověření normality v softwaru STATISTICA	5
10	Jednovýběrový t test	2
11	Testy odlehlých hodnot	3
12	Připojení do databází pomocí STATISTICA Query4	4
Prá	ce v rozhraní STATISTICA Query 4	5
13	Úprava načtených dat4	6

SUALISIICA

Pı	roměnné a případy	
Tı	ransformace dat	
Pa	oužití filtru	
14	Automatizace rutinních analýz	
15	Analýza rozptylu	

1 Spuštění programu STATISTICA

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

	Domu	pravit	Zobrazit	Vložit	Formát	Statistik	y Data Mi	ining Grafy	Nástroje	Data	Enterprise Ná	pověda
Nový)tevřít Uložit Soubo	Tisk	Zavřit O	tevřít Uložit Projekt	Zaviit	Přidat do sešitu • p	Pñdat do rotokolu - do Výstup	Pñdat Wordu - Pa	nel Makro M lýz Vástroje	ožnosti 17	f Rezervovat f Vrátit se změnami) Zrušit SharePoint	日 (1) (2) (3) (3) (4) (4) (4) (4) (4) (4) (4) (4) (4) (4
	Soubor	t		Projekt			Výstup		Năstroje		SharePoint	
Klasick	é nabídky ((styl MS	Office 2	003)								
Klasicki A STATE	é nabídky (STICA Cz - Te	(styl MS	Office 2	003)								

Vybereme Pás karet - po potvrzení **OK** se obrazovce se objeví rychlá navigace, kterou zavřeme a

tá vás	STATISTICA	
Co by	ste chtěli udělat nejdříve?	1
	Otevřít dat. soubor ST	ATISTICA
1	🔘 Otevřít sešit Excelu	
1	🔘 Vytvořit dotaz k extern	í databázi
	Otevřít protokol	
E	🔘 Otevřít pracovní sešit	
₽	🔘 Otevřít makro	
R	🔘 Otevřít skript R	
DM	💮 Otevřít projekt Data M	lineru
Μ	Otevřít projekt STATIS	STICA
\Diamond	🔘 Otevřít elektronickou p	př <mark>íruč</mark> ku
9	🕐 Přehrát video:	
osledr	ně otevřené soubory	
vstupn	í data barven í xlsx	
Statista	ina data.sta	
uprave	na data.sta	
106 pro	měnné sta ka chuhy 2011 vlev	
		20
_ Příš	të tento dialog nezobrazov	/at
-		7 74



máme zde okno aplikace *STATISTICA*:

základní nabídk	a					
			STATISTICA Cz . (Data: Orien	acniMereni (2s krát 26ř))		
Soubor Domů Upravit Zobrazit Fon	mát Statistiky Data mining	Grafy Nástroje	Data Scorecard			R Feature Finder
	A 2D x Matica	Grafu bloku dat z	16 Mactaix			
	All 3D Selve T * Ikonv	Grafy vstuppich dat *	The Value of the state of the s			
Histogram Bodový Průměry Krabice Rozptyl Spojnice	₩ 3D XYZ * He Kategorizované *	Bill Dávk. (skupin) analýza	DB ranca nee grana			
Běžný	Vice	and the second	Nástroje			
1 2 vzdalnost poznatka 1 100.02 meren 1 2 100.02 meren 1 3 99.98 [Korekce 4 100.02 meren 1 6 100.01 meren 1 7 100.01 meren 1 8 99.98 meren 1 9 100.01 meren 1 10 100.03 korekce 11 100.01 meren 1 12 99.96 meren 1 13 100.02 meren 1 14 100.04 meren 1 15 100 meren 1 16 99.98 meren 1 17 99.99 meren 1 18 100.03 meren 1 20 100.01 meren 1 21 100.02 meren 1 23 100.01 meren 1 24 100.02 meren 1 25 100.02 meren 1						
					-	
panel analýz		stavov	nápovědu stiskněte F1 ný řádek	pracovi výstupy	ní plocha s ta	bulkou dat a

- 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*.

•			STATISTICA Cz - [Data: Tat	oulka1* (10s krát 10ř)]	
Zobrazit Vložit	Formát Statistiky Data Mining	Grafy Nástroj	e Data Scorecard	Nápověda	Možnosti 🛛 🙀 🗕
ich 🔹 📋 Záhlaví/zápatí	😑 Uložit vodorovně 🛛 📮 Zavřít vše				Skyl
🚽 Události 🕶	Uložit svisle				Zprava doleva
pce 🕶 🗹 Stavový řádek	🔁 Uspořádat na sebe 🔁 Přepnout okna	Klasické			✓ Ribbon bar
it	Okna	Rozhran			

2 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
- vložením 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 **Temperat CZ.xls**

Soubor	Domů	Upravit	Zobra	izit
🛃 Ulož	it it jako		Napos	ledy o
🍃 Otev	ŕít			Orier C:\U
🎽 Otev	rřít příklac	ly		Orier C:\U
🔄 Zavr	it tnosti			Trigo C:\Us
Posledn	ĩ	<	EE III	Freez C:\Us
Nový			North Street of	mere H:\St

Při otevírání "Excelovských" tabulek mámě několik možností, jak k tabulkám přistupovat:

Import	souboru: excel.xls	? 🗙
	Importovat všechny listy do sešitu	
	Importovat vybraný list do tabulky	
8	Otevřít jako pracovní sešit Excelu	
Nas	tavit jako výchozí	



• *Importovat vybraný list do tabulky* – nejčastější možnost – pokud máme více listů, tak upřesníme list, který chceme importovat – **vybereme:**

í	Vybrat list pro import.	? ×	
	Listy v souboru MS	OK Stomo	
tevřít so	ubor MS Excel		?
Soul	bor: excel.xls		ОК
 I.řádel 	< jako názvy proměnných		Storno
1.sloup	ec jako názvy případů		Proměnné
_ Importo	ovat formát buněk		
Sloupce	: od 1 🔷 do 4		

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

🖨 do 28

÷

Pokud mám v původním souboru textové popisky, ale formát proměnné je číselný, tak mě *STATISTICA* upozorní, převedeme tyto textové popisky v číselné proměnné na chybějící hodnotu, nebo je naimportujeme jako textové a následně se v načteném souboru podíváme a smažeme je.

Importovat hodnoty text	ových popisků	2 X
Importovaná data pro pr aby byla data importovár nebo chcete import zrušit Provést pro všechny č	oměnnou 7 jsou textová, ale na jako hodnoty textových p ? číselné proměnné, pokud jso	e typ proměnné je číselný. Chcete, popisků, převedena na chybějící data u data textová
Zrušit import	Převést na ChD	Importovat jako textové popisky

Řádky: od

1



Pozn.:	Přidejm	e tuto t	abulku o	do sešitu	výsledků	ì:		
	19 6	G	• 🗄 =					
Soubo	Do	mů	Upravit	Zobra	izit F	ormát	Statis	stiky
	1		,	E			THE	12
						>	C C	1-8
Novy	Otevrit •	Ulozit	Projekt	Pridat do sešitu -	protoko	do lu ▼ do	Wordu *	Worksp
	Sou	bor		Př	idat do n	ového s	ešitu	
		C:\U	sers\Mil	D: Fr	eezing SA	OS-2 B	ox plot	Ĩem
		Ope	rátor	Cast	Pokus	Mere	eni	

3 Zpracování chybějících dat

Načtení souboru

Postup si představíme na kompletním příkladu, jak postupovat. Pro zopakování začneme samotným datovým souborem a jeho načtením. Máme excelovský soubor, do kterého byly ručně zadány výsledky dotazníkového šetření. Soubor obsahuje řadu chybějících hodnot а překlepů. Přes Soubor -> Otevřít načteme tento datový soubor:

V posledním kroku mě software *STATISTICA* upozorňuje na to, že v proměnné, kterou vyhodnotil jako číselnou, se vyskytují textové popisky. Ve verzi *12* lze zaškrtnou "*Provést pro všechny*…" a nově kliknout na *Převést na ChD*, v tomto případě budou textové popisky v číselných proměnných (např. *N/A* apod.) převedeny na chybějící pozorování, tedy na prázdnou buňku.

Starší verze tuto možnost nemají, a proto si ukážeme případ, kde tyto textové popisky v číslených proměnných máme.

Editor textových hodnot

Dvojklikem na záhlaví každé proměnné můžeme vyvolat dialog konkrétní proměnné a v části *Textové hodnoty* se lze podívat, jestli se zde nějaký text (kterému by software přiřadil číselnou reprezentaci) nevyskytuje:



9 /	1	•	• B / U	<u>∧</u> -	3, 9g I
	Jméno	Тур	Kód ChD	Délka	DI. jmér propojer
3	1 číslo dotazníku	Double -	-999999998		100000
	2 Zména přístupu k	Double	-993999998		
	3 Způsob změry	Text	-999999998		8
	4 Vliv zaškolení	Integer	-999999998		
1 3	5 Nesrozumitelné č	Byte	-999999998		
	6 Odbornost lektora	Double -	-999999998		
	7 Byl/a jste spokoje	Double -	-999999998		
	8 Kurz bude užitečr	Double -	-999999998		
1	9 Úroveň studijních	Double -	-999999998		
1	Lektor se vyjadřov	Double -	-999999998		
1	1 Pozice	Double -	-999999998		
1	2 Věková kat.	Double +	-9999999998		
1	3 Dosavadní praxe	Double -	-999999998		
			a).		

Textový popisek má od softwaru přiřazenu číselnou reprezentaci, pokud je proměnná typu Double, lze se na tuto reprezentaci v Editoru textových hodnot podívat. Pokud je proměnná typu Text, přiřazení čísel proběhne automaticky až v případě využití proměnné k analýze. Máteli v softwaru kategorické proměnné, které budou vstupovat do analýz jako grupovací proměnné (faktory), doporučujeme mít všechny tyto proměnné jako číselný typ Double s právě zmíněnými textovými popisky. Číselnou reprezentaci si mohu libovolně překódovat (v Editoru textových hodnot) na vlastní hodnoty (vhodné a využitelné například u pořadí sloupcových grafů nebo při řazení případů

číselně, apod.). Změnu z *Text* na *Double* provedeme buď jednotlivě ve specifikaci jednotlivých proměnných nebo hromadně ve specifikaci všech proměnných, tedy po kliknutí na tlačítko *Všechny specif.* v dialogu kterékoli proměnné.





Vlastní překódování bychom potom provedli individuálně, například takto:

<u>Poznámka:</u> Textové popisky jsou vlastně přiřazení textu jakékoli číselné hodnotě, což je vhodné především pro přehlednost souboru, kde můžeme vidět buď textové popisky, nebo číselnou reprezentaci.

g contor	textoryennou	ine farmereral		A	
∑ jeřadí •	<u>P</u> řečísluje •	Kopíruje z •	Aglikuje na	Seřadí • Přečísluje	• Kopíruje z • Ar
T	ext. hodnota	Čiselně	Popis	Text. hodnota	a Číselně Po
V	aki. Š	101		zákl. VŠ	3
n	naturita	103		maturita	2
-					

U proměnných číselných jsou samozřejmě textové popisky nežádoucí, pojďme se nyní podívat na to, jak bychom je detekovali.

Detekce neexistujících kategorií

Jednou z možností, jak se podívat na jednotlivé proměnné je tabulka četností. V základních statistikách vybere *Tabulku četností*:



V případě našeho datového souboru (výsledky dotazníkového šetření) vybereme všechny proměnné a klikneme na *Výpočet*.

Postupně se proklikám jednotlivými tabulkami četností v sešitu výsledků a snadno identifikuji, jestli se v datech nevyskytují jiné kategorie, než mají, kolik je chybějících hodnot, atd.





	I	Tabulka čet	ností:Vliv zaškol	ení			
k ladní statistik <mark>y a t</mark> abul	Kategorie	Četnost	Kumulativní četnost	Rel.četnos	Kumulati rel.četno	vní ost	
Dialog frekvenčních t	1	18	18	29,032	26 29,0)323	
Tabulka četnosti:č	2	6	24	9 677	42 38	7097	
Tabulka četností:2	3	15		Tabulka četr	iostí:Zména přís	stupu k práci	
Tabulka četností:z	4	8	Kategorie	Cetnost	Kumulativní četnost	Rel.četnost	Kumulativní rel.četnost
Tabulka četnosti:	NUA	12	1	9	9	14,51613	14,516
Tabulka četností:ľ	N/A	1	2	35	44	56,45161	70,967
Tabulka četnosti:(CND	2	3	7	51	11,29032	82,258
Tabulka četností:E			4	5		Tabulka četni	osti:Uroveň studi
Tabulka četností:k			5	3		Četnost	Kumulativní
Tabulka četností:l		-	10	2	Kategorie		četnost
Tabulka četností:l		-	22	1	_1	46	46
Tabulka četností:F			ChD	0	2	4	50
Tabulka četností:\					10	11	61
Tabulka četností:[?	1	62
					ChD	0	62

V případě, že v datovém souboru máme i spojité proměnné, tak tyto proměnné načteme zvlášť v druhém kroku, v dialogu tabulky četností přepneme na *Detaily* a zvolíme například *Pěkné intervaly:*



Takto můžeme například identifikovat hodnoty, které jsou například mimo reálně možné meze.

Berme tuto metodu pouze jakousi základní hrubou detekci nevhodných dat, rozsahy intervalů bychom pro potřeby popisné statistiky optimalizovali samozřejmě pro každou proměnnou zvlášť.

V hlavní roli průzkumník

Nyní bychom chtěli identifikovat případy (řádky), ve kterých se "škodlivá" data vyskytují, to bude dalším krokem v naší analýze. Využijeme grafickou metodu, kterou je *Graf chybějících hodnot*. V záložce *Grafy -> 2D -> Grafy chybějících hodnot dat nebo dat mimo rozsah* otevřeme dialog tohoto grafu a vybereme proměnné.

Prom	ěnné:			
zad A V	vberte promě	nné pro graf Chl	D	? ×
Por	Fela datamatar	, ,		
102 1-	cisio dotazniku	k práci	<u>^</u>	OK

Přepneme na kartu *Detaily* a v roletce zvolíme *Oboje* (tedy detekci dat mimo rozsah i ChD).







V části Zadejte platný rozsah dat je možné zvolit rozmezí hodnot. které jsou platné. V našich datech máme dva možné typy rozsahů, rozdělíme tedy proměnné na dvě skupiny a určíme pro ně rozsahy. Zvolíme první a druhou sadu proměnných a upřesníme jejich rozsah (to ie výhodné především u dotazníků, kdy víme předem, jaké jsou možné výsledky otázky, které otázky jsou například na škále 1-10, atd.), po té klikneme Ok a získáme graf. Jedná se o graf, který vykresluje místa, kde v souboru chybí pozorování

nebo je zde pozorování mimo stanovený rozsah. Jsou tedy vyobrazeny jen problémové místa souboru.

Najedeme-li kurzorem na **konkrétní označené pozorování**, získáme informaci o čísle případu (*v obrázku jde o pozorování č. 18*). Naším cílem je identifikovat všechna tato pozorování v datovém souboru. Jednou z možností je využít interaktivního průzkumníka grafu. V záložce *Upravit* vyberme *Průzkumníka* (to platí pro nabídky typu *Pás karet* nebo klikneme do grafu pravým tlačítkem – například vedle nadpisu - a vybereme *Průzkumník*).





Poté obdélníkovým výběrem vyberte označte body grafu – při zapnutém Průzkumníku dáte kurzor do plochy grafu, následně stiskněme levé tlačítko myši a označme (roztáhněme čtverec) celou plochu grafu.

V dialogu *Průzkumníka* zvolme potom např. *Obarvit* a klikněme na *Použít* a následně na *Konec*:



Případy, které přísluší označeným bodům v grafu, byly obarveny přímo v datovém souboru. Klávesou *F1* v dialogu *Průzkumník* vyvoláme nápovědu, kde je popsán význam jednotlivých možností. Novinkou ve verzi *STATISTICA 12* je možnost (přes pravé tlačítko myši) označená data nechat vygenerovat jako podmnožinu do nové tabulky. Na takovéto podmnožině se poté přehledně podíváme na jednotlivá vadná pozorování.



Možnost Ověřit data

Velmi obdobně, jako jsme využili před chvílí graf hodnot mimo rozsah, můžeme najít data mimo rozsah i jinak. Stačí použít funkcionalitu **Ověřit data**, kterou najdeme v záložce **Data**-

Overu-O	veru
data	Zde si
můžeme	zadat
velký	počet
podmínel	k a
omezení,	které
mají data	splňovat

ulky	A Třídit Jména Případy ▼	Specifikace IIII Posun Všechny specifikace Svazky Textové popisky Proměnné *		Externí data * ++ Sloučit	² 0 ⁷ Vzorkování □ Přímý ♥ Vst ² 0 ⁷ Semínko ♥ Ověřit ▼	
	Případy	Proměnn	ié	Sprav	Ověřit dat	ta

(tyto podmínky lze pomocí tlačítka *Uložit jako* uložit pro pozdější použití, taktéž lze pomocí *Otevřít* podmínky nahrát). Data, která nejsou platná poté můžeme označit pomocí tlačítka *Označit neplatné* nebo jít jedno neplatné pozorování po druhém, podobně jako funguje vyhledávání textu v souborech (tlačítko *Nalézt první* a poté přejít na další pomocí klávesové zkratky *ctrl+F3*). Takto je možné neplatné pozorování v souboru postupně kontrolovat a případně přímo manuálně opravovat.

Případy jsou platné, poku	d jsou splněny:
Všechny podmínky	Alespoň jedna podmínka
Podmínka 1	Nalézt první
Platné pokud: 🔹	
v1>0.5	Označit neplatné
Dederfeler 2	🕅 Jako vyjmuté
Podminka 2	Stomo
Platné pokud: 🔹	Storio
	Zrušit vše
Podmínka 3	Otevřít
Platné pokud: 👻	📳 Uložit jako
	Rozsah
	Od případu:
Podmínka 4	1
Platné pokud: 🔹	Do případu:
	10 🛋



4 Vytvoření základní výpočtů

4.1 Tabulka četností

Soubor Domú Upra	▼ vit Zobraz	it Formát	Statistiky	Základní statistiky a tabulky: Temp	oerat CZ ? 💦
Základní Vicenásobná ANOV	A Neparametri	cké Prokládání	Rozdělení a	Základní výsledky M. Popisné statistiky Korelační matice	CK Stomo
Základní statistiky Spustí modul Základní stati	Základ stiky/tabulky	4 Meren		탄국 t+test, nezávislé, dle skupin 당편 t+test, nezávislé, dle proměn. 당첨 t+test, závislé vzorky 팀고 t+test, samost, vzorek	► Možnosti ◄
1 Novak	1	1 108,64	99	家最 Rozklad & jednofakt. ANOVA	
2 Novak	2	1 111,91	29	Rozklad	
3 Novak	3	1 96,230	04	Tabulky četnost í	

Volba proměnné:

Proměnné	é: Operator-Pokus	Výpočet
Možnosti	Popisné statistiky Normalita [Stomo
Základr	Vyberte proměnné na analýzu	8 🗙
	i omin	
Kategor. me	2 - Cast	OK
Všech	4 - Mereni	Storno
🕐 Přesn		[Svazky]
🔘 "Pěkr		·
🔘 Vel. k		
počá		
🔿 Celoč	Vybrativše Doztábnout Dřiblá	8+
O Speci		ac j
Uživa	1-3	
O OLITO		



Označili jsem 3 promění 🔼 📕 🖉 🗢 🕂 🗑 🔻	né, získáme	3 tabulky	četností:	-	
Soubor Domů Upravit	Zobrazit	Formát	Statistiky	Data	mining Gra
Základní Vícenásobná ANOVA N statistiky regrese	Jeparametrické F statistiky Základ	Prokládání Roz rozdělení sir	dělení a nulace	2 Pokrc Vícer. Analý	čilé modely ▼ /průzkumné ▼ (za síly testu [Pokročilé/Vícero
🗁 PS3*	<	Tabulka če	tností:Oper	rator (T	emperat CZ)
Základní statistiky a tabul	k al Kategorie	Četnost	Kumulat četnos	tivní st	Rel.četnost
Tabulka četnosti:	Novak	23		23	19,16667
Tabulka četnosti:	Novotny	24		47	20,00000
Tabulka četností:	Pokorny	24		71	20,00000
	Cech	24		95	20,00000
	Havlice	23		118	19,16667
	Novák	1		119	0,83333
	?	1		120	0,83333
	ChD	0		120	0,0000

Detekujeme chybně napsané/duplicitní kategorie....

Uložení výsledné tabulky – přes pravé tlačítko myši na vybrané tabulce v sešitu: Základ







CW.				
SIL	Název souboru:	Tabulka četností (SPC01)	•	Uložit
	Uložit jako typ:	Tabulky STATISTICA (*.sta)	•	Storne

4.2 Popisné statistiky

Na popisnou statistiku si vyzkoušejme 2 příklady:

4.2.1 Soubor Temperat CZ.sta - měření dílů jednotlivými operátory

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





Vybereme proměnné (pro více proměnných držím při výběru myši **Ctrl**) – vybrali jsme spojitou (měřenou) proměnnou "měření". V softwaru je několik možností, jak popisné charakteristiky získat:

Proměnné:	Mereni				Výpočet
Normalita	Pravd. &	bod.grafy	Kateg. grafy	Možnosti	Stomo
Základní vý	sledky	Detailní	výsledky	Robustnost	Možnosti
Souhm	: Statistiky	Grafy 1	Grafy 2	ו	

Na části Detailní výsledky si lze vybrat přesně to, co potřebuji (průměry, medián apod.):



Proměnné;	Mereni			Výpoče
Normalita Základní výsle	Pravd. & bod. grafy	Kateg grafy sledky	Možnosti Robustnost	Stomo
Míry polohy & N	stiky G1 G2 Variace, momenty ✓ Směrodat. odchylka	Výpočet statistik Kvantily, rozpět	:: í	Anal.skup
 %plat.pozorov. Průměr Souč. Medián 	Meze sp. směr. odch. Interval: 95,00 🖆 % Variační koeficient V Bozptví	Dolní & horr	ní kvartily nranice 10,00 🖆 %	

Tlačítkem *Anal.skupiny* si rozdělíme výpočet podle jednotlivých operátorů a podle jednotlivých dílů:

ké Prok	ládání Rozdé	🕅 🗰 Ví ělení a	cer./průzkumné 🔻 🛞 🖞 F	PLS, PCA, 🔛 Multiva	ariate 🔐 DOE	by Dávk. analza (d
rozo	dělení simu 1 Operátor	llace №	Popisné statistiky:	Temperat CZ v PS3 Mereni Pravd. & bod. grafy	Kateg. grafy Možn	osti ∫ Stomo
1	Novak	1	Zakladni vysled	Ky Detailní v	rýsledky Robustno	st 🛛 🔊 Možnosti 🔻
2	Novak	2	Souhm: Statis	stiky 🛛 🚟 성 G1 🚟 성 G	2 Výpočet statistik:	-
3	Novak	3	Míry polohy & N	Variace, momenty	Kvantily, rozpětí	Anal.skup
4	Novak	4	🔽 Počet platn.	Směrodat. odchylka	Analýza po skupinách	8 23
5	Novak	5	%plat.pozorov.	Meze sp. směr. odch.	1	
6	Novak	6	Průměr		Skup. proměnná(é)	Operator
7	Novak	7	Souč.		7-muta	
9 10 11 12 13 14 15 16	Nov Nov Nov Nov Nov Nov Nov Nov Nov Nov	perátor ast okus ereni		OK Storno [Svazky] Pro zobrazení odpovídajících proměnných	Výstup do jediné složky Pojmenovat výstupy Výsledky za všech. skupin Sloučit tabulkové výsledky	Neuspořádané Vzestupně Sestupně jedné tabulce OK Storno
17	Nov			zvolte "Ukázat pouze". Pro		
18	Nov Vvl	prat vše Ro	ztáhnout Přiblížit	stiskněte F1.		
20	Nov					
21	Nov	romennych		_		
22	Nov 1-2					
23		ouz <mark>e odpovídajíc</mark>	í proměnné			

Výsledkem je tabulka popisných statistik (**průměrně změřené hodnoty** rozdělené dle typu dílu a operátora, **variabilita (kolísavost)** měření dle operátorů a dílů):





⇒ PS3 < Image: Solution of the second se		Souhrnné výs Popisné stati	sledky istiky (Te	emperat CZ)					
🛓 👍 Základní statistiky a tak	Proměnná	Operátor	Cast	N platných	Průměr	Minimum	Maximum	Rozptyl	Sm.odcn
🛓 🔤 Dialog popisných s	Mereni	Pokorny	6	3	94,55128	91,61418	98,69031	13,60138	3,6880
Souhrnné výslev	Mereni	Pokorny	7	3	108,3405	104,7909	110,5487	9,637447	3,1044
	Mereni	Pokorny	8	3	109,5204	106,2551	111,5393	8,145827	2,8540
	Mereni	Cech	1	3	106,9716	104,0340	109,3760	7,347468	2,7106
	Mereni	Cech	2	3	110,6898	107,9589	113,6553	8,153521	2,8554
	Mereni	Cech	3	3	91,00978	88,49283	93,94303	7,556149	2,7488
	Mereni	Cech	4	3	96,00987	92,86737	101,0552	19,47571	4,4131
	Mereni	Cech	5	3	110,4953	109,3737	111,0870	0,944392	0,9717
	Mereni	Cech	6	3	92,84646	90,04922	94,49323	5,929981	2,4351
	Mereni	Cech	7	3	106,6603	104,4496	109,0504	5,315978	2,3056
	Mereni	Cech	8	3	109,2533	106,5765	110,7176	5,389916	2,3216
	Mereni	Havlicek	1	3	103,9523	103,1088	105,1161	1,084257	1,0412
	Mereni	Havlicek	2	3	106,8802	105,9350	108,1119	1,246285	1,1163
	Mereni	Havlicek	3	3	88,65133	87,57439	90,18211	1,854524	1,3618
	Mereni	Havlicek	4	3	95,25764	94,47401	96,49270	1,171619	1,0824
	Mereni	Havlicek	5	3	110,6907	108,5007	112,8080	4,642172	2,1545
	Mereni	Havlicek	6	3	90,62983	88,98065	91,63175	2,071292	1,4391
	Mereni	Havlicek	7	3	104,6683	101,7325	106,8967	7,042481	2,6537
	Mereni	Havlicek	8	3	104,9393	104,5638	105,3204	0,143134	0,3783

V zápětí se podíváme na možnosti editace tabulky. Více o průměrech a mírách kolísavosti se dočtete v:

17/09/2012 StatSoft ACADEMY – charakteristiky polohy

http://www.statsoft.cz/file1/PDF/newsletter/2012_09_17_StatSoft_popisna_statistika.pdf

15/10/2012 StatSoft ACADEMY - charakteristiky variability

http://www.statsoft.cz/file1/PDF/newsletter/2012_10_15_StatSoft_Popisne_statistiky_-_miry_variabily.pdf

4.2.2 Editace tabulky (Anglické popisky apod.)

Dvojklikem do např. záhlaví tabulky

5	tatistiky roz	dělení simula	ace N	🖁 Analýza síly tes	tu 📳 VEPA	C
Zák	lad			Pokročilé/	Vícerozměrné	
<		Souhrnné výs Popisné stati	ledky stiky			
ky a tak	Proměnná	Operátor	Cast	N platných	Průměr	Minimun
sných s	Mereni	Novak	1	3	106,4896	103,61
ié výslei	Mereni	Novak	2	3	112,7327	111,91
	Mereni	Novak	3	3	93,69352	90,071
	Mereni	Novak	4	3	96,26198	93,006
	Mereni	Novak	5	3	111,5526	110,73
	Mereni	Novak	6	3	96,28348	94,203
	Mereni	Novak	7	3	108,2578	107,43
	Mereni	Novak	8	3	109.5162	107.22

- CTRL+A označíme celý text, následně CTRL+C zkopírujeme a vložíme např. do překladače Google apod.
- ➢ Nový text zkratkou CTRL+V vložíme do záhlaví:





	Summary of	results				Proménná 4				
Proměnná	descriptive st Operator	atistics Cast	N platných	Průmer	Mini	A Arial	▼ 10	• B	ΙŪ×₂×	č <u>∆</u> •
Mereni	Novak	1	3	106,4896	10	a di Imaan		-		
Mereni	Novak	2	3	112,7327	11	Jmeno		Typ:	Double	- OK
Mereni	Novak	3	3	93,69352	90			- /II	Fa.	Storno
Mereni	Novak	4	3	96,26198	93	Typ dat:	utomaticky 🔻	Delka:	8	
Mereni	Novak	5	3	111,5526	11		Stav ořípadu	Kád ChD	-000000009	▲ << >>
Mereni	Novak	6	3	96,28348	94			Rou Cho.	-3333333333	
Mereni	Novak	7	3	108,2578	10	Formát zobrazení				
Mereni	Novak	8	3	109,5162	10	T OFINAL 2001 azerii				Všechny specif.
Mereni	Novotny	1	3	100,9402	97	Obecné				
Mereni	Novotny	2	3	104,2950	10	Cislo				Textové hodnoty
Mereni	Novotny	3	3	85,53490	83	Čas				Hade /Statist
Mereni	Novotny	4	3	92,01059	90	Vědecký				Houringstatisting
Mereni	Novotny	5	3	107,3354	10	Měna				Vlastnosti
Mereni	Novotny	6	3	89,67676	88	Zlomky				
Mereni	Novotny	7	3	101,5493	98	Vlastní				[Svazky]
Mereni	Novotny	8	3	105,1770	10					
Mereni	Pokorny	1	3	103,1411	10					
Mereni	Pokorny	2	3	111,9501	10	Dlouhé jméno (popis či	i výraz s Funkcemi):		V Prův. funkcemi
Mereni	Pokorny	3	3	92,89787	91		25 AC	-		
Mereni	Pokorny	4	3	96,58988	95					
Mereni	Pokorny	5	3	113,1069	11					

Dvojklikem na proměnnou *Průměr* vyvoláme dialog proměnné a změníme její název. Nyní se podíváme, jak si stojí jednotlivý operátoři v grafickém výstupu:



4.2.3 Rozdělení spojité proměnné dle kategorie

4		Proměnné:	syst		Výpočet
1	2	Manage Bag	Decid & Lead and . 1	Kata and 1 Mažardi	
syst	Nou				Storno
80 14	0 nekura	Zakladni vysled	JKY Detailni v	ysledky Hobustnost	Možnosti
81 15		Souhrn: Stati	istiku 🚟 G1 🚟 G	2 Vúpočet statistik:	
82 16	0 nekura	Míru polobu & N	Variace momentu	Kuantilu rozoětí	Anal.skup
83 14	0 nekura	Mily poloriy & N	Valiace, momenty		
64 16	10 nekura	Pocet plath.	Mara an amár adab	Minimum & maximum	
85 12	5 nekura	%plat.pozorov.	Interval OF 00 Class	📃 Dolní & horní kvartily	
00 14	U siiny	Průměr		🔄 🔄 Kvantilové hranice	SULET
8/ 15		🔲 Souč.	Variační koeticient	První: 10.00	CRSES 1
00 14	U silny	🔲 Medián	Rozptyl		Váž. momenty
89 12	Usilny	Modus	🛛 🗹 Sm. chyba průměru 🏑	190.00 kaz	9
90 20	0 silny	🔲 Geom. prům.	V Meze spolehl. prům.	Analýza po skupinách	8 4
91 12	0 silny	🔲 Harm. průměr	Interval: 95,00 🚔 %		
92 12	U sliny		Sikmost	Skup. proměnná(é) Kour	
A Výběr o	dle proměnný	ch	? X		
				✓ Zapnuto	Usporadani skupi
1 - syst			OK	📃 Výstup do jediné složky	Neuspořád
2 - Kour				Pojmenovat výstupy	🔘 Vzestupně
			Storno	Výsledky za všech, skupiny	🔘 Sestupně
			[Svazky]	Sloucit tabulkove vysledky v jed	ine tabulce
1				Of Of	Storno
1			Pro zobrazeni odpovidaji oloh		
1			odpovídajících proměnných		
			Pro zobrazeni odpovídajících proměnných zvolte "Ukázat		
			Pro zobrazeni odpovidajicich proměnných zvolte "Ukázat pouze", Pro více informaci		
1 1 1 1 Vybrat v	vše Roztáł	pout Přihlížit	Pro zobrazeni odpovidajicich proměnných zvolte "Ukázat pouze". Pro vice informaci stiskněte F1.		
1 1 1 1 1 Vybrat v	rše Roztáł	nout Přiblížit	Pro zobrazeni odpovidajicich proměných zvolte "Ukázat pouze". Pro vice informaci stiskněte F1.		
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	rše Roztáh	inout Přiblížit	Pro zobrazeni odpovidajicich proměnných zvolte "Ukázat pouze". Pro více informaci stiskněte F1.		
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	rše Roztáł nných	inout Přiblížit	Pro zobrazeni odpovidajicioh proměnných zvolte "Ukázat pouze". Pro více informaci stiskněte F1.		

Na kartě *Detailní výsledky* v dialogu *Popisné statistiky: Statistiky —> Základní statistiky —* > *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/



5 Vytvoření grafu

5.1 Histogram

Přes Grafy -> 2D grafy -> Histogramy

	260) 🏟 🛐 🗢					
Soubor	Domů	Upravit	Zobrazit	Formát	Statisti	iky Data mining	Grafy
Histogram	Bodový graf	Průměry Krabic	e Rozptyl S	pojnice	2D * 3D Sekv. * 3D XYZ *	Matice S Ikony M Kategorizované •	Grafy blo Grafy vst
		Běžný				Více	

Vybereme proměnné pro obě období:

kladní Detaily Vzhled	Kategorizovaný Možnosti 1 Možnosti 2	ОК
Proměnné: Prom.: 1-2	Vuberte proměnné pro histogram	Stomo
ýp grafu: adh Béžný Mil Vícenásobný	1 - KW2 Scrap (108) 2 - KW1 Scap (108) 3 - KW2 108 (%) Scrap + coating 4 - KW1 108 (%) Scrap + coating	OK Anal. skup. Storno [Svazky]
Typ proložení:		Galerie grafù lizace: Auto

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:

ikladní Detaily Vzhled Kategorizova	ný Možnosti 1 Možnosti 2			ок	
Typ grafu:	ásobný Proměnné:	1-2 nic		Stomo Možnosti v	
Grupovací intervaly © Celočís. mód V Auto	Max (neodlehlé)	Vyberte promění	stitet cases	Filtr případů ý graf	2
Vsechny hodnoty Nesetř. 123 321 Kátegorie: 10 Kódy: Žádné	Medián 25% Min (neodlehlé) Střední bod Hodn.: Medián ▼	1 - KW2 Scrap (108) 2 - KW1 Scap (108) 3 - KW2 I08 (%) Scr 4 - KW1 I08 (%) Scr	ap + coating ap + coating	1 - KW2 Scrap (108) 2 - KW1 Scap (108) 3 - KW2 I08 (%) Scrap + coating 4 - KW1 I08 (%) Scrap + coating	St [Sva
	Styl: Bod Společný rozptyl				



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



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

6 Uložení práce

6.1 Uložení celého sešitu výstupů

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

Přes Soubor -> Uložit jako...



Uložíme soubor, ve kterém je všechno, co jsme vygenerovali. 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

<u><u> </u></u>	obrazit V <u>I</u> ožit <u>F</u> ormát S <u>t</u> a	atistiky Data <u>m</u> ining	<u>G</u> rafy <u>N</u> ástr	oje <u>D</u> ata	<u>P</u> racovní se
🛛 🗅 🖨 🖬 🖉 🎒 🕻), 👗 🖻 🖻 🝼 🗠	🙀 Přidat do seši	tu 🐐 Přidat do	protokolu	 Přidat do
Arial	▼ 10 ▼ B I U	≡≡⊒⊠	<u>•</u> • <u>•</u> • <u>•</u> •	N	.00 100 14
PS1*	- hulles (Carrow 102)		Popisné stat	istiky (Scr	ap 108)
Zakladni statistiky a t	abulky (Scrap 108)	Proměnná	N platných	Průměr	Minimum
Dialog popisnych	statistik	KW2 Scrap (108)	24	25,54167	13,00000
2D histogramy (Histogram z Histogram z	<u>V</u> ložit Odstranit Přejmenovat Evtrahovat jako samostatné c	Insert 08) Delete F2	20	32,40000	16,00000
	Uložit položku(-y) jako				
G-Gy Kvalita (Scrap IO G-Gy Dialog výsle G-Gy X-diagra	vyjmout dokument ze sesitu <u>K</u> opírovat dokument sešitu Vložit jako dokument sešitu	Ctrl+C			
🛛 🕞 Kvalita (Scrap I0	Vlastnosti				

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:

oru:	X-diagram a diagram MR.stg
/p:	Soubory grafů STATISTICA (*.stg)
n í pro ti	Soubory grafů STATISTICA (*.stg) Soubory grafů STATISTICA 6.X (*.stg) Soubory grafů STATISTICA 7.X (*.stg) Bitové mapy (*.bmp) Soubory JPEG (*.jpeg, *.jpg) Soubory Portable Network Graphics (*.png) Metasoubory (*.wmf) Enhanced Metasoubory (*.emf) Soubory PDF (*.pdf) Soubory GIF (*.gif)
	Soubory TIFF (*.tif)

6.4 Přidání výstupů do Protokolu/Microsoft Wordu

1. Založme si nový protokol:



a přidejme popis protokolu:

					0	4		1000
oubor Domů	Zpráva St	<mark>ztistily Dete minin</mark>	g Grefy	Scorecard	Feature Fin	der	Mo	žnosti 🔻 🛛
Najít	b <mark>ac Nahradit</mark> bab Opakovat ∲ Přejít na /nahradit	Courier New B I U E = = A · Or Formát	• 10 • A	Rozložení	Dokument	•	÷	



-	*			
	Název souboru:	weekly_pistons	•	Ulož
	Uložit jako typ:	Protokoly STATISTICA (*.str)	•	Storm





	R Feature Finder	Možnosti 🕈 🙀 🧾
e I		Restory Window
lat ikony		
Jt okna ▼		
A 20	20 21 20 23	00 01 02 00
26. 1. 27. 1. 28. 1	·29···30···31···32···	5

2. Minimalizujme si v pravém horním rohu protokol:

3. V sešitě výsledků označíme (*při stisknutém Ctrl*) výstupy, které chceme přidat do reportu a klikneme na *Přidat do protokolu*:

ový Otevřít Uložit Soubor	Projekt	Přidat do sešitu *	Přidat do protokolu * Přidat	Přida do Word do nové	Přidat do u * Workspace * ho protokolu	Panel analýz	Makro Možnosti Nástroje	館 Reze 館 Vráti ゆ Zruši Sha	rvovat it se změnami it arePoint	😑 Uspořád]]] Uspořád 🔁 Uspořád	at vodorovně 🧐 at svisle 🛛 🛱 at na sebe 🦓 Okna
PS3* - Sirka Te PS3* 	sty seskup ta (Cover C2	ení, prom	week	y_piston C2.568 V <	estr Fooy	J			Sirka Test X-diagram Centrální p	y seskupení, vřímka: 100,0	proměnná
🖨 🦾 Dia	og výsled X & R-dia	ků diagrar gram: pro	nů řízení kva m.: Sirka	ity	Zóny A/B/C: 3,0 Testy pro spec.	00/2,000 případy)/1,000 * sigma (pravidla sledů)		od podsk.	do podsk.	
	X & R-dia	gram: pro	m.: Sirka		9 podskupin na	stejné st	traně centrální pří	mky	OK	OK	
	X & R-dia	gram: pro	m.: Sirka		6 podskupin za	sebou si	toupajících/klesaj	icích	OK	OK	
	Sirka Test	v seskuper	ní proměnna	: (Cov	14 podskupin ko	olísá nah	oru a dolů		OK	OK	
and a					2 z(e) 3 podsk.	v zóně A	nebo mimo ni		OK	OK	
					4 z(e) 5 podsku	ipin v zo	oně B nebo mim	o ni	OK	OK	
					47 mardalauria co	1 1 0			01/	OK	
					15 podskupin V.	zone C			UK	UN	



SINA TAST

Nyní si podíváme na náš protokol:

	1	C. 4	. נו או				J	мынс	A (2 - F 3)	- Lauva	LESTA SERV	upen, prom	enna, (Cover	CEISIO V FUUJ
Soubo	Don	nů 📃	Upravit	Zobrazi	t Formát	Statistiky	Data	mining	Grafy	Nástr	oje Da	ata Sešit	Scoreca	ard
	1			FI		INF IS	27	Λ_{a}		-	Rezer	vovat	📑 Uspořa	ádat vodorovr
						and the	- Sale	7 41	2	0	Vrátit	se změnami	Uspořa	ádat svisle
Nový	Otevřít	Uložit	Projekt	Přidat do	Přidat do	Přidat Přid	lat do	Panel	Makro N	Iožnosti	Zrušit		🔁 Uspořa	ádat na sebe
0.52	Sout	or		sesina - F	Vieto	wordu - work	space -	analyz -	Nástroje		Sha	reDoint		01
	- 5001	501		<u></u>	vysu	41			Ivasci uje		3110	repoint	1	
E	PS3 - Sir	ka Test	ty seskup	ení, proměn	ná: (Cover CZ.:	ita v PS3)							[
	PS3					<						Sirka Tes	ty seskupe	ní, proměnná
ė.	Cov	er CZ.s	ta									X-diagram		
		(valita	(Cover C	Z.sta v PS3)								Centrální	přímka: 100),077482 Sig
	ė-2	Dial	log výsle	dků diagram	ů řízení kvality	Zóny A/E	3/C: 3,00	00/2,000	/1,000 *	sigma		od	do	
			X & R-di	agram; prom	n:: Sirka	Testy pro	spec.	pripady (pravidla s	ledú)	1.1.1	podsk.	podsk.	
		-	X & R-di	agram; prom	h.: Sirka	9 podsku	ipin na s	steine st	rane cent	raini prii	nky	OK	OK	
			X & R-di	agram; prom	n.: Sirka	6 podsku	ipin za s	sebou st	oupalicici	n/kiesaji	cich	OK	OK	
			Sirka Tes	ty seskupení	, proměnná: (C	0\ 14 podsk	Rupin Ko	usa nah	pru a dolu	no ni		OK	OK	
						2 Z(e) 3	podsk. V	nin v zó	nebu min	ho mim	o ni		OK	
						15 podel	cupin v 3	zóně C	ne o nei	oo min	U III		OK	
						8 nodeki	ipin min	no zónu	С			OK	OK	
						o pousice	apart train	10 20110	0				UIL	
						-								
						11.								·
									_					F
1			III			🖡 🚰 X R-	diagram; p	orom.: Sirk	a 📗 Sirk	ca Testy s	eskupení, p	roměnná: (Cov	ver CZ.sta v P	S3)
			-											
Week	. @		23											
		النص		1					7					
AE	X/R: Si	irka: C	over CZ	Nasta	avení analýzy zp	ůsobi 📠 Nas	stavení ar	nalýzy způ	sobi		Připrave	na		Cov

Pokud mám všechna okna maximalizovaná (sešit výsledků i protokol), tak přepínáme pomocí zkratky <u>*CTRL* + *TAB*</u>



Výsledný protokol lze dále editovat, pokud se některé tabulky nezobrazili celé, tak je roztáhnu myší.



Výsledný protokol uložím jako PDF, nebo jako *RTF* (formát, který lze otevřít ve Wordu a přeuložit jako *.docx)





Protokol STATISTICA:	*		III
Theorem Stansmen.	Název souboru:	Cov	ver (nabeh) 🗸
	Uložit jako typ:	Prot	otokoly STATISTICA (*.str)
		Prot	otokoly STATISTICA (*.str)
		Prot	tokol STATISTICA 7 (*.str)
		We	ebová stránka (* htm. * html)
		Text XMI PDF	xtové soubory (*.bxt) 1L soubory jako text (*.xml) 1F Soubory (*.pdf)
Chi kvadrát	8.11.20	13 19:4	45 Složka souborů
Protokol1	Otevřít	8 13:	338 Rich Text Format 481
Protokoli	Upravit	8 13:	38 Protokol STATISTI 5
TRIAL12	Nový	17:5	58 Aplikace 539 517
5x260	Tisk	16:4	40 Microsoft Excel W 18 495
5x260	Otestovat systémem AVG	16:3	39 Tabulka STATISTI 10 196
📗 shlukova 🦯	Scan with Microsoft Security Essentials	11.5	59 Tabulka STATISTI 7
	Otevřít v programu	w I	Word (aplikace klasické pracovní plochy)
-	Softwareb02 Print2PDF		WordPad
	Sdílet s		Zvolit výchozí program
8	HP File Sanitizer - ničení	-	or recompleterom 2000

Nebo lze importovat přímo do MS Word (záleží na verzi Office, vždy lze uložit jako RTF a přeuložit). Postup shrnuje obrázek níže:







🤊 😁 (.) 👘 🗄 🔻 Domů Zpráva	Statistiky Data mi	ning Grafy Sco	recard			
rtevřít Uložit Projekt	Přidat do sešitu → protokolu → do Výst	Přidat Přidat do Wordu * Workspace *	Panel nalýz v Nástroje	諸語 Rezervovat 諸語 Vrátit se zm 夕 Zrušit SharePoir		
Uložit jako	PDF ² ky jako PS1\Cover CZ\Kvali PS1\Cover CZ\Kvali	Cover (ná 25.1.2013	`` <u>4</u> ``5 <u>`</u> `6`'? а́běh)			
		Histogram hodnot	X-diagram a diagram MR; promě X : .57950 (.57950)	ná: Prom1 sigma: ,27754 (27	7	
				Možnosti výstupu Výstupní tabulky jako Objekty (ve velikosti, jak j Tabulky normální velikost Použít současné nastaven	jsou v okně protokolu) ti (na jednotlivé stránky) í a příště již tento dialog nezobi	OK Stomo

7 Další možnosti načtení souborů

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 se budou ukládat výstupy. Ze základní nabídky vybereme *Nástroje* -*Možnosti*.... Otevře se dialog *Možnosti*, ve kterém přejdeme na záložku *Správce výstupů*:

Nástroje Pracovní sešit Okr
Panel analýz 🔹 🕨
dol
Sloučit g <u>r</u> af
Uzamknout graf
M <u>akro</u>
<u>V</u> lastní
M <u>o</u> žnosti

Noznosti → Obecné → Analýzy/grafy → Zobrazení → Omezení/meze → Správce výstupů	Umístit všechny výsledky (tabulky, grafy) do: Individuálních oken Délka fronty: 10
DokumentyTabulky datGrafyPracovní sešityMakraProtokolyData MinerPodraní In-Place Database	Více sešitů (pro každou analýzu/graf samostatný) Jednoho sešitu (společného pro všechny analýzy/grafy) Existujícího sešitu: Imístit výsledky do sešitu automaticky Nové výsledky nahoru
	Výstup protokolu: Do jednoho protokolu (společný pro všechny analýzy/grafy) Procházet Výstup Microsoft Word: Do jednoho dokumentu Word (společný pro všechny analýzy/grafy) Procházet Výstup Microsoft Word: Do jednoho dokumentu Word (společný pro všechny analýzy/grafy) Procházet Výstup Microsoft Word: Do jednoho dokumentu Word (společný pro všechny analýzy/grafy) Procházet Vožit tabulku dat do Wordu jako objekt Další detaily: Žádné Výchozí písmo: Courier New 9 Zamykání protokolů (GxP Protokoly) Uzamčeno Značka vytvoření: Protokol vytvořen &[Date] &[Time] uživatelem &[User] na &[Computer] Časová značka: &[Date] &[Time] uživatelem &[User] na &[Computer]
	Tisk. rozlišen i obrázků Metasouboru/schránky/protokolu: 600 🖨 DPI

Můžeme zvolit některé z těchto možností:

- *individuální okna* každá tabulka či graf se zobrazuje v samostatném oknu. 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 Ověření normality v softwaru STATISTICA

Jedním ze základních předpokladů mnoha statistických analýz je normalita. Pokud některý test či metoda normální 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:

Proměnné:	Prom.: Prom1	Intervaly	Stomo
		Prom.: Prom1	🔊 Možnosti
Typ grafu:	Typ proložení:	Všechny hodnoty	Anal, skup,
Vícenásobný	Normální	Nesetř. 123 321	SELECT Filtr případů
Dvojité-Y	Beta	🖉 🔘 Kategorie: 10 🚔	
Zobrazovaný typ	Exponenciální	Hranice: žádné Kódy: žádné	Galerie grafi
Standardní 🔹	Gama	Různé podskupiny	Aktualizace: Auto
Mezery mezi sloup	Geometrické	Time Změnit proměnnou	
Zobrazit procenta	Statistiky		

Zajímavý článek o tomto tématu naleznete zde:

http://www.statsoft.cz/file1/PDF/newsletter/2013_10_09_StatSoft_Jak_se_pozna_normalita_pomoci_grafu.pdf





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



Pozn.: Pokud použijeme K-S test, P hodnota se zobrazuje intervalem, pro přesnou P hodnotu využijte modul Rozdělení a simulace (viz níže modul Rozdělení a simulace).

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



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

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**

Základní výsledky C Normalita Pravd. & bod. gr	Detailní výsledky afy Kate) :g. grafy	Robustnost Možnosti	Storno
Normalita Pravd. & bod. gr	rafy Kate	eg. grafy	Možnosti	Možnos
U 3D bod. graf se jmény	<u>P</u> olo-normálr	pravděpod. (ní pravděpor	graf d. graf	Anal.sk
Kategoriz, bodové grafy	🝸 Normál, p-gra	af s odstr. tre	endem	
🤌 Plošný gra <u>f</u>				SELECT f

Nebo přes záložku Grafy:



trendem (odstraněn lineární trend).

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

Zakadam vznied Nategon.	ovaný Možnosti 1 Možnosti 2	ОК
Typ grafu:	Proměnné:	Stomo
Polo-nomální	žádné Vyberte proměnné pro	p-graf
Statistiky Shapiro-Wilksův test	- Rozlože □ Vice 4 - Π 2 - Hi 3 - Pi 4 - Π	OK Storn
	Vybrat vše Roztáhno	ut Přiblížit
	Proměnné:	
	1-4	
	Deurse a descuidación premi	inna.



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, Anderson – Darling atd.

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 F1 v políčku pro zaškrtnutí příslušného testu vyvolá nápovědu k tématu a doporučení k jednotlivým testů:



The Shapiro-Wilk W test is used in testing for normality. If the W statistic is significant, then the hypothesis that the respective distributi because of its good power properties as compared to a wide range of alternative tests (Shapiro, Wilk, & Chen, 1968). STAT/ST/CA impli to large samples (with up to 2,000 observations; see Basic Statistics). See also Kolmogorov-Smirnov Test and Lilliefors Test.

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

ſ	Rozdělení
	Iabulky četností 🔝 Histogramy
	Kategorizace

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.

A

Modul Rozdělení a simulace

Modul, který slouží přímo testování různých pro rozdělení ie Rozdělení a simulace:



Na kartě **Základ** vybereme proměnné a přepneme na kartu **Spojité proměnné**. Zde vybereme Normální rozdělení.

Copyright © 2014

Základ

LANK.

Spustit simulaci

Návrh simulace





áklad Spojité proměnné Disk	trétní pro <mark>měnné Možnosti</mark>		1	<u>0</u> K
<< Předch. Další >> ACC	CELERATION -			Storno
Rozdělení			2	Možnos
Vybrat vše Odstranit vš	ie Změnit			
🔽 Normáln í				Otevřít
Log-normální	Posun (práh/poloha):	0	SELECT CRSES	E 8
Folded normal	Posun (práh/poloha):	0	8	
📰 Polonormáln í	Posun (práh/poloha):	0	4	
Rayleighovo	Posun (práh/poloha):	0	8	
Weibullovo	Posun (práh/poloha):	0	8	
🔲 Gaussovské směsí	Počet rozdělní:	2	8	
🔲 Johnsonovo rozdělení (obe	ec. ne-normální, jednod. mód)	No.		
Rozdělení extrém. hodnot			ChE) vynechár
Zobecněné Paretovo	Posun (práh/poloha):	0	9	Párově
Trojúheln íkové				Cele pripa

Volba konkrétního výstupu pro dané proměnné. Tlačítkem Souhrnné statistiky rozdělení získáme výstupy z testů normality:

Výsledky proložení dat rozdě Základ Uložt proložení	leními: Cars.st	a v PS2	114							
Proměnné << >> HAN Rozdělení << >> BRA MAN	IDLING ELERATION KING DLING AGE									
Popisné statistiky		Souhm	Souhmé stat	istiky rozdělení]					
Empirická distr. funkce	• 📶	Histogram s proložen ím	Kumulativ	n í histogram						
Graf <u>P</u> -P		Graf <u>Q</u> -Q	Fabic Krabic	ový graf]					
PS2* <			Souhrn rozdě	lení for Pron	něnná: HAND	DLING (Cars.sta	v PS2)			
Gars.sta			K-S d	K-S p-hodn.	AD stat.	AD p-hodn.	Chí-kvadrát	Chí-kvadr. p-hodn.	Chí-kvadr. SV	(p
🛓 🔯 Dialog výsledků j	Normální	(poloha,měřítko)	0,371527	0,003112	3,300678	0,019628	21,28409	0,000004	1,000000	
- Souhrn: ACC			1							
Souhrn rozdě										

Následující příklad slouží k ověření normality vybraných veličin:

Příklad - Normalita a důležitost náhodného výběru

<u>Úkol</u>: 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. Poklepáním 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 zvolme Normální pravděpodobnostní grafy. Na něm se body vyskytují na přímce.
- 4. Na záložce *Detaily* dialogu *Základní statistiky a tabulky Popisné statistiky* kromě nabídnutých možností zaškrtněme 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 Setřídit*. 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*.
- 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.

10 Jednovýběrový t test

	1												
	vzdalenost				_								
1	100,02												
2	100,01								. 9	X			
3	99,98		And t-te	est pro sa	mo:	statne pru	imery: Orien	itachiivieren					
4	100,09												
5	100,01			roměnné:	vzd	lalenost			SUHH V	ýpočet 🛛			
6	100,01												
7	100,01		Zákl	adní výsled	lky	Detailní v	ýsledky Mo	žnosti	Sto	rno			
8	99,99		D.	faran žuć h		-A.,	<u> </u>						
9	100,05		nt	aerenichi h	oun	oty			Moz Moz	nosti 🔻			
10	100,03		•	Test všec	h pr	ůměrů vůči	i: 100		bull a				
11	100,01		0	Test prům	ěrů	vůči různýr	0 2.1	Coocif	Ana Ana	al.skup			
12	99,96			volitelným	kor	istantám	991	opeon.					
13	100,02			a beautic allo	اند خان		n žun úsla		CASES 1	@ ⊻			
14	100,04			UUIAZICUIU	una	illiena piui		- 121					
15	100		V	ýpočet me	zí s	polehl.;	Interval: 95,	0(🛢 %	Váž. m	omenty			
16	99,98			ícerozměri	hý te	est (Hotellin	iaovo T²)		SV =				
17	99,99						15		@ W-1	D N-1			
18	100.03		p-no	dnota pro a	zvyr	azneni: 🔍	JJ 🔽 🗖						
									I CLD				
1	ter ser ter												
PS 2*	••••••	-	-										
📥 🗁 2D	histogramy (Trigonor		Test prům	érů vůči refe	renč	ní konstante	é (hodnotě) (Or	rientacniMeren	li) Defense žeć		CV/	-	
Ĺ	Histogram z Yi	Proměnná	Prumer	Sm.odch.	IN	Sm.cnyba	-95.000%	+95.000%	konstanta	L	50	р	
- 📑 Zá	kladní statistiky a tabu	vzdalenost	100,0123	0,029300	26	0,005746	100,0005	100,0241	100,0000	2,141918	25	0,042130	
📄 🚔 Zá	kladní statistiky a tabu												
	Test průměrů vůč												

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

Test je signifikantní, zamítáme nulovou hypotézu: $H0: \mu = \mu_0 = 100m$

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

Test průměrů vůči referenční konstantě (hodnotě) (OrientacniMereni)

	Int. spolehl95,000%	Int. spolehl +95,000%
vzdalenost	100,0005	100,0241

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

Kompletní řešený 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



11 Testy odlehlých hodnot

Pro objektivní vylučování extrémních hodnot na základě vypočteného testovacího kritéria u souborů dat, které odpovídají Normálnímu rozdělení náhodné veličiny, je v softwaru implementován Grubbsův test

× 🖪 🔒 🖉	0 9 C	🔒 🖥 🔻						
Soubor	Domů	Upravit	Zobrazit	Formát	Statistiky			
Základny statistiky Základny Zá	rícenásobná regrese Cákladní sta kladní výslev M Popisné Korelačn	i ANOVA Ne zi tistiky a tab dky statistiky í matice	12 eparametrické statistiky śkład ułky: Pistons	Prokládání rozdělení CZ R CZ O Stomo	Rozdělení a simulace			
Proměnn	né: Velik	ost						
Normalita	Pr	avd. & bod. g	grafy	Kateg. graf	y Možnosti			
Základni	í výsledky		Detailní výsl	edky	Robustnost			
	Souhm_: statistiky							
Výpočet stati Umístění	stik:							
Useknut	ý průměr	Usek	nuté % příp.	5,00				
Winsoriz								
	ovany prům	ër Elsek	nuté % nřín. –	5 00				

Další možností je využití krabicového grafu v záložce Grafy:

Základní Detaily Vzhled Kategoriz	zovaný Možnosti 1 Možnosti 2]		Krabicový graf	z syst seskupenj	i Kour	-
Typ grafu:	žný 🗘 Eroměnn Žný Cávislé p	é: rom.: syst síprom.: Kour			*		
Grupovací intervaly	Max (neodlehlé)	Krabicový					
Prom.: Kour	75%	Hodn: Kvantily 🔻					1
Celočís. mód I Auto	• Medián	Koeficient: 25	T		0	T	
Všechny hodnoty	25%	- Swatca		-	0		-
Nesetř. 123 321	Min (neodienie)				T		
🔘 Kategorie: 10 🚔	Hada i Madián 🖉 🖛	Rozsan neodieni.					
Hranice: žádné	Hodn.: Wiedian	Koeficient: 1	B				1
Kódy: žádné	Styl: Bod 🔻	Odlehlé hodnoty		· · · · · · · · · · · · · · · · · · ·			
Vícenásobné podmnožiny	Společný rozptyl	Odl. hodn. & extrémy 🔹					-
Změnit proměnnou	Rozložení více krabic	Koeficient: 1,5		\perp			
Proložení	Překnývající	🖽 🔲 Propojit střední body					
Vypnuto	Ořez, evtrém rozlož · 0 🛋 🖤	- Zohr čistá data					Medián
Lineámí 📃	Statietiky	- Rozmíetit					T Rozsah neodleh
M Polynomiáln í	Kruskal-Wallisův test	1102million	nekuřák	bývalý	silný	slabý	o Odlehlé
C Logaritmické 👻	F test a p (ANOVA)	vyp. Vyp. Vyp. Vyp. Vyp. Vyp. Vyp. Vyp. V		Kour			# Extrémy



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*:



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.

Zprostředkovatel	Připojení	Upřesnění	Vše	
Chcete-li se připo 1. Vyberte neb	ojit k datům o zadejte na	serveru SQL ázev serveru:	Server, za	idejte tyto údaje
MILOS-H	P\SQLEXP	RESS		Aktualizovat
 Zadejte infoi Použít Použít 	mace pro p integrovan zadané už	řihlášení k se é zabezpeče ivatelské jmér	erveru: ní Windov no a heslo	vs NT
Uživatelské	jméno:			
Heslo:	Ē			
🗌 Prázdné	heslo	Povolit ulože	ní hesla	
3. (Vyberte	databázi na	serveru:		
Sample	DataSQL			•
O Pripojit d	atabázový	soubor jako n	ázev data	báze:
				A CONTRACTOR OF CONTRACTOR
Název s	ouboru:			
			Testo	wat nënojan (
			Tosco	var pripojern
		K C	Stomo	Nápověd
		AV.	500110	Napoveu

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

Vyberte	data, ke kt středkovat	erým se chcete elé OLE DB	připojit:	~
Crys Crys Micr Micr Micr Micr Micr Micr Micr Micr	al Decision al Decision soft Jet 4.0 psoft OLE D psoft OLE D	s Closed Integra s Closed XML A DLE DB Provid 12.0 Access Da B Provider for A B Provider for In B Provider for In B Provider for C B Provider for C B Provider for C B Provider for S B Provider for S B Simple Provider of	tor Base DO Provider der atabase Engine Ol unalysis Services S Data Mining Servic ndexing Service nternet Publishing DBC Drivers DLAP Services 8.0 DIAP Services 8.0 DIAP Services 4.0 DIAP Server der	LE DB Pro 3.0 ces
MSD SQL	Native Clie			
MSD SQL Stat	Native Clie Soft OLE DE	B Provider for S	TATISTICA Sprea	dsheets 💌





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 pust". 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*.

Spojení	Kritéria	Nápověda
!+ (!)	₩3 ►]
	Pi	renést data do programu STATISTICA Cz(F5)



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

méno: Dotaz 1		Spusti
am chcete vložit data?		Upravi
🔿 Do nové tabulky	i	Storno
💿 Do aktivní tabul	ky počínaje	
Proměnné	1	
Případu	1	
Vlastnosti		
Přizpůsobit tabulki	u	
Max. řádků:	10000	Vlastnosti

13 Úprava načtených dat

Proměnné a případy

	Vybrat pouze jména proměnných	Přidání a odebrání proměnných provedeme následujícím způsobem: V záhlaví tabulky klikneme pravým uchem myši
	Statistiky bloku dat	 a zobrazíme dialog, v němž můžeme vybrat možnost
<u></u>	Grafy blo <u>k</u> u dat	• Odebrat proměnné nebo Přidat proměnné.
r S	<u>G</u> rafy vstupních dat	 Při přidávání proměnných se zobrazí dialog, v němž
Ж	Vyjmout Ctrl+	Dřídat proměnné
B _B	Kopírovat Ctrl+	
	Kopírovat se z <u>á</u> hlavími	Kolik: Použijte 0 v poli "Za" pro vložení OK
a	V <u>l</u> ožit Ctrl+	Za: datum nebo stiskněte F2 pro výběr ze Storno
	Vloži <u>t</u> jinak	Jméno: NProm Typ: Double
	Přidat proměnné	- Kód ChD: -999999998 ➡ Délka: 8 ➡
	Odstranit proměnné	Formát zobrazení Pokud se mají
	Přesunout proměnné	Obecné Desetinná místa: 2 💭 nové proměnné a morčina datí je
	Kopírovat proměnné	Datum 1000000-100000 přilš velká, je Čas 1000.00; -1000.00 přilš velká, je Vědeckú 1000.00; -1000.00 vytelejší přidat
	Specifikace proměnné	Měna 1 000.00; (1 000.00) současné Procenta přepočítat jejich
	Správce skupin	∠lomky hodnotu pomoci Vlastní dávkové transformace dat
ŦŦ	Vyplnit/standardizovat <u>b</u> lok	(nabídka Data).
	Odstranit	Dlouhé jméno (popis či výraz s Funkcemi):
	<u>F</u> ormát	•
	Ozn <u>a</u> čení buněk	 L Vzorce: použite jména proměnných nebo v1, v2, V0 je č. případu. Příklady: (a) = mean(v1:v3, sqrt(v7), AGE) (b) = v1+v2; komentář (po ;)

uživatel specifikuje počet přidávaných proměnných, název proměnné, ze kterou 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 jejích 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 popisek. Velikost *4 byty*.

• Byte

Celá čísla v rozmezí 0 až 255, nelze vložit desetinná čísla, každé číselné hodnotě lze přiřadit textový popisek. Velikost *1 byte*.

• 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:

Přidat případy			? 🛛
Kolik:	10	•	ОК
Vložit za případ:	5	•	Storno
Před prvním připade	m vložit D	i.	

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.

Dlouhé iméno (popis či výraz s Funkcemi):	Prův. funkcemi
r Propopisy použijte libovolný text. Provýrazy použijte jm. Příkladu: (a) = mean(v1:v3. sort(v7). AGE) (b) = v1+v	prom. či v1, v2, v0 pro č. případu. 2: komentář (po :)

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:





Dlouhé jméno (popis či výraz s Funkcemi):	🗹 Prův. funkcemi
=Log2(v2)+v3	

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

STATIST	TICA Cz
2	Výraz je správně. Přepočítat proměnnou?
	Hodnoty, které nelze spočítat (např. =Sqrt(v2-1), =1/v2, kde v2=0), budou zkonvertovány na chybějící data (ChD). Pokud bude argument výrazu ChD, výsledkem celého výrazu bude ChD.
	Ano Ne

Použití filtru

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



Filtr případů pro a	nalýzu/graf			? 🛛
Filtr se užije jen pro tut	o analýzu/graf		Z <u>m</u> ěnit zdroj pr	o filtr případů
Zapnout <u>fi</u> ltr		묮 Přehled prom.	: O <u>d</u> stranit vše	OK
vcetne pripadu vsechny				Storno
 některé, vybran 	é:			🔁 Otevřít
výra <u>z</u> em: Funkce	V5>1			📳 Uložit jako
nebo čí <u>s</u> ly případů:				
- Mimo případů (z mno	ožiny případů definova	iné v sekci Včetně příp	adů'):	
určené výrazem: Funkce	V5>12			
ne <u>b</u> o čísly případů:	1-6			
Čísly případů: Zadejt Vybrané Použij výrazem: Použij Příkla V příp Textor	te čísla případů neborozs te stejné operátory, funkt te názvy proměnných net dy: (a) v1=0 OR vět>11 radě konfliktu budou mít p vé hodnoty označte pomo	rahy. Např.: 1; 3; 5-12 be a syntaxi jako ve vzoro bo v1, v2 v0 je číslo pi 3 (b) pohlaví="MUŽ" AND vřednost jména proměnnýc cói \$ ve tvaru "hodnota"\$.	ích v tabulce: ňpadu (v0<4 znamená v4<)(v5+v6) h před textovými hodno	pňpady 1-3). otami proměnných.

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 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, anebo 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*.

15 Analýza rozptylu

ANOVA

Analýza rozptylu je užitečná v situacích, kdy nás zajímá vliv jedné nebo více *nominálních* proměnných (též zvaných *faktory*) na proměnnou *kvantitativní*. Příkladem může být analýza velikosti tržeb v závislosti na ročním období, analýza účinků určitého léku u různých skupin pacientů, analýza mezd podle dle dosaženého vzdělání atd.

Zkoumáme-li závislost pouze na jednom faktoru, hovoříme o jednofaktorové analýze rozptylu. Celý soubor se rozčlení do příslušného počtu skupin (podle počtu úrovní faktoru) a předmětem zkoumání jsou potom střední hodnoty těchto skupin – jejich shoda či rozdílnost. Faktor může obecně nabývat libovolného počtu hodnot a testová hypotéza má pak tvar H_0 :

 $\mu_1 = \mu_2 = ... = \mu_k$, čímž v podstatě říká, že sledovaná proměnná není závislá na úrovni faktoru a že při všech jeho úrovních nabývá zhruba stejných hodnot, přičemž rozdíly jsou způsobeny pouze náhodným kolísáním. Alternativní hypotéza tvrdí, že alespoň jedna z uvedených rovností neplatí.

Podstatou je, jak už název napovídá, rozklad rozptylu zkoumané (závislé) proměnné, a to jednak na část, která vzniká v důsledku skutečné rozdílnosti jednotlivých skupin, tzv. *meziskupinový rozptyl*, a jednak na část zapříčiněnou náhodným kolísáním, tzv.

vnitroskupinový (reziduální) rozptyl. Testovým kritériem je pak podíl těchto složek. Pokud je meziskupinová variabilita dostatečně velká oproti reziduální, test vede k zamítnutí hypotézy o rovnosti středních hodnot.

Stejně jako regrese i analýza rozptylu je založena na obecném lineárním modelu. ANOVA je v podstatě součástí (speciálním případem) regrese.

Další návodné články k tomuto tématu naleznete v archivu newsletterů StatSoft Academy: <u>http://www.statsoft.cz/o-firme/archiv-newsletteru/</u>

✓ Anova dvojného třídění:

<u>http://www.statsoft.cz/file1/PDF/newsletter/2012_11_12_StatSoft_Analyza_rozptylu.pdf</u> V Neparametrická Anova:

http://www.statsoft.cz/file1/PDF/newsletter/2013_06_04_StatSoft_Neparametricka_anova.pdf



Příklad – jednofaktorová ANOVA: Patnáct pozemků bylo náhodně rozděleno do tří skupin. Na dvou z nich byla použita hnojiva A a B, třetí skupina byla kontrolní bez hnojení. Určete, zda použité hnojivo má vliv na výnos obilí. Stanovení hypotézy:

 H_0 : Použité hnojivo nemá vliv na výnos obilí.

 H_1 : Použité hnojivo má vliv na výnos obilí.

Test provedeme na 5% hladině významnosti.

 Otevřeme datový soubor *Hnojiva.sta*. V prvních dvou sloupcích jsou uvedeny výnosy při použití hnojiv, ve třetím jsou výnosy z pozemků nehnojených. Poněvadž takto uspořádaný soubor neobsahuje žádnou proměnnou, která by označovala úroveň faktoru (tyto úrovně jsou uvedeny pouze v záhlaví), je potřeba data převést do vhodnějšího tvaru, který program *STATISTICA* očekává. To lze provést dvěma způsoby. První způsob:

Vytvoříme nový soubor *Hnojiva (upraveny).sta* \rightarrow *Soubor* – *Nový* - *Tabulka dat*. Počet proměnných nastavíme na 2 a počet případů na 15. V nové tabulce vytvoříme proměnnou *Hnojivo* (faktor) a proměnnou *Výnos*. Kopírováním vložíme data. Původní

a upravenou tabulku ukazují následující obrázky.

Data: Hnojiva.sta (3s krát 5ř)							
Výnosy							
obilí		Hnojivo A	Hnojivo B	Bez hnojeni			
	1	71	69	65			
	2	68	74	62			
	3	73	72	57			
	4	73	71	63			
	5	69	74	60			

🎹 Data: Hnojiva_uprav (2s krát						
	1	2				
	Hnojivo	Výnos				
1	A	71				
2	A	68				
3	A	73				
4	A	73				
5	A	69				
6	В	69				
7	В	74				
8	В	72				
9	В	71				
10	В	74				
11	zadne	65				
12	zadne	62				
13	zadne	57				
14	zadne	63				
15	zadne	60				

Druhý způsob:

StejnéhoVhodnějšího tvaru dat lze dosáhnout rychleji a jednodušeji seskupením dat. *Data – Přeskupování… - z*áložka *Seskupování*. Proměnné vybere všechny. Jméno cílové proměnné bude *Výnos* a jméno kódové proměnné *Hnojivo*. Potvrdíme tlačítkem *OK*. Upravenou tabulku ukazuje následující obrázek.



Data: Tabulka3* (2s krát 15ř)							
Výnosy							
obilí	1	2					
	Výnos	Hnojivo					
1	71	Hnojivo A					
2	69	Hnojivo B					
3	65	Bez hnojeni					
4	68	Hnojivo A					
5	74	Hnojivo B					
6	62	Bez hnojeni					
7	73	Hnojivo A					
8	72	Hnojivo B					
9	57	Bez hnojeni					
10	73	Hnojivo A					
11	71	Hnojivo B					
12	63	Bez hnojeni					
13	69	Hnojivo A					
14	74	Hnojivo B					
15	60	Bez hnojeni					

2. Ověříme předpoklad normality dat.

K ověření normality zvolíme Shapirův-Wilkův test, který najdeme v záložce *Statistiky – Základní statistiky/tabulky - Popisné statistiky – Normalita*. Proměnná, kterou testujeme, je Výnos. Po stisknutí tlačítka *Tabulky četností* se nám spolu s tabulkou četností objeví i výsledky testu normality dat.

	Tabulka četností:Výnos (Hnojiva (upravena)) K-S d=,17211, p> .20; Lilliefors p> .20 Shapiro-Wilk W=,90249, p=,10388							
	Četnost	Četnost Kumulativní Rel.četn. Kumul. % Rel.četn. Kumul. %						
Kategorie		četnost	(platných)	(platných)	všech	všech		
55,00000 <x<=60,00000< td=""><td>2</td><td>2</td><td>13,33333</td><td>13,3333</td><td>13,33333</td><td>13,3333</td></x<=60,00000<>	2	2	13,33333	13,3333	13,33333	13,3333		
60,00000 <x<=65,00000< td=""><td>3</td><td>5</td><td>20,00000</td><td>33,3333</td><td>20,00000</td><td>33,3333</td></x<=65,00000<>	3	5	20,00000	33,3333	20,00000	33,3333		
65,00000 <x<=70,00000< td=""><td>3</td><td>8</td><td>20,00000</td><td>53,3333</td><td>20,00000</td><td>53,3333</td></x<=70,00000<>	3	8	20,00000	53,3333	20,00000	53,3333		
70,00000 <x<=75,00000< td=""><td>7</td><td>15</td><td>46,66667</td><td>100,0000</td><td>46,66667</td><td>100,0000</td></x<=75,00000<>	7	15	46,66667	100,0000	46,66667	100,0000		
ChD	0	15	0,00000		0,00000	100,0000		

3. Předpoklady homogenity rozptylů.

Analýzu spustíme volbou Statistiky - Základní statistiky/tabulky – Rozklad & Jednofakt. ANOVA. Zvolíme Proměnné: Závislá proměnná je Výnos, Grupovací proměnná je Hnojivo. Přepneme na záložku Skupiny tabulek a v oddělení Výstupní tabulky vybereme Celková tabulka průměrů, Analýza rozptylu, Leveneův test a Brown & Forsythe (HOV). Po stisknutí Výpočet je třeba znovu zadat grupovací proměnnou Hnojivo. Znovu klikneme na tlačítko Výpočet a dostaneme požadované výstupy. Výše zmiňované testy homogenity rozptylů ani zde neprokázaly rozdíl rozptylů mezi jednotlivými skupinami (jejich p hodnoty přesahují 0,05). Předpoklad homogenity je tedy splněn.



2 - Hnojivo	2 - Hnojivo	2 - Hnojivo	2 - Hnojivo	2 - Hnojivo	2 - Hnojivo	OK Storno
Podrobn.) Přiblížit	Podrobn. Přiblížit List2:	Podrobn. Přiblížit List3:	Podrobn. Přiblížit	Podrobn. Přiblížit	Podrobn, Přiblížit List6:	Esvazky J Pro zobrazeni odpovidajicich proměnných zvolte "Ukázat pouze", Pro vice informaci stiskněte F1.

4. *Tabulka průměrů* ukazuje průměry a rozptyly v jednotlivých skupinách. Výstup vlastní analýzy rozptylu ukazuje následující obrázek.

	Analýza rozptylu (Hnojiva_uprav) Označ. efekty jsou význ. na hlad. p < ,05000							
	SČ	SČ SV PČ SČ SV PČ F p						
Proměnná	efekt efekt efekt chyba chyba chyba							
Výnos	336,9333	2	168,4667	76,00000	12	6,333333	26,60000	0,000039

Na základě této tabulky můžeme tvrdit, že zamítáme hypotézu o stejných středních hodnotách. Mezi skupinami je statisticky významný rozdíl a výnos tedy závisí na použitém hnojivu.

Alternativním způsobem, jak spustit předešlé analýzy je i následující postup: *Statistiky - Základní statistiky/tabulky – Rozklad & Jednofakt. ANOVA.* Zvolíme *Proměnné: Závislá proměnná* je *Výnos, Grupovací proměnná* je *Hnojivo.* Místo zvolení záložky *Skupiny tabulek* potvrdíme výběr proměnných tlačítkem OK. Zobrazí se následující tabulka:

🕼 Statistiky dle skupin - výsled	ky: Tabulka1	? ×
ZÁVISLÉ : 1 proměnná:)	Výnos	
GRUPOV. : 1-Hnojivo (3): Hnojivo A Hnojivo B	<u>₿</u> ±
Základní výsledky Popisné stati	stiky ANOVA & testy Post-hoc	
Analýza rozptylu	🚝 Kategoriz, normál, pravd. grafy	Storno
📄 Provést Welchův F-test	🖉 Kategoriz, polo-normál, p-grafy	🔈 Možnosti 🔻
Testy homog. rozptylu	E Kateg, p-grafy s odstr. trendy	Anal.skup
Leveneovy testy	🥖 Gra <u>f</u> průměrů vs. směr. odch.	
Brown-Forsytheho testy		
	Graf interakcí	
p-hodnota pro zvýraznění: ,05	✓ Zobrazit intervaly spolehl. průměru:	





Touhle cestou se dostaneme nejen k výběru výpočtů analýzy rozptylu a testů homogenity rozptylů, ale také k různým možnostem vizualizace dat.

Příklad – dvojné třídění: Máme tři kmeny krys, jejichž obecná schopnost úspěšně se pohybovat v bludišti by se dala popsat jako dobrá, nestálá, nebo špatná. Čtyři krysy z každého kmenu byly vychovávány ve stimulujícím prostředí, čtyři v prostředí omezeném. Cílem je určit, zda kmen, prostředí nebo obojí má vliv na počet chyb, které krysa v bludišti udělá.

- 1. Otevřeme soubor *Rats CZ.sta*.
- 2. Z nabídky Statistiky vybereme položku ANOVA a zobrazí se úvodní panel Obecná ANOVA/MANOVA. Zvolíme položku Vícefaktorová ANOVA jako Typ analýzy, v poli Metoda specifikace ponecháme Rychlé nastavení. Klikneme na OK. Zobrazí se dialog ANOVA/MANOVA Vícefaktorová ANOVA. Klikneme na tlačítko Proměnné a zvolíme Chyby jako závislou proměnnou a Kmen a Prostředí jako kategoriální prediktory.
- 3. Klikneme dvakrát na OK a dostaneme se tak do dialogu ANOVA Výsledky. Tento dialog poskytuje spoustu možností pro volbu nejrůznějších výsledků. Jsou uspořádány na osmi záložkách. Pokud by nám tyto výsledky nestačily, je možné se přepnout do ještě obsáhlejšího výsledkového dialogu stiskem tlačítka Více výsledků více výsledků. Zpět do původního výsledkového dialogu se vrátíte stiskem tlačítka Méně. Nyní kliknutím na tlačítko Všechny efekty/grafy na záložce Základ zobrazíme dialog Tabulka všech efektů. Oba efekty (Prostředí i Kmen) jsou označeny jako významné (označeny hvězdičkou *), ale efekt interakce významný není.

igma-omezena para Iekompozice efektiv	imetrizace /ní hypotézy					
Efekt	SČ	Stupně volnosti	PČ	F	P	Storno
Prostredi Kmen Prostredi [*] Kmen	5551, 7940, 16,	1	5551, 3970, 8,	5,823 4,164 ,008	,027* ,033* ,992	Zavřít dialog při 0 Zobrazit Graf Tabulka Průměry: Nevážené Vážené Met. nejm. čtv.
						✓ Výpočet sm. chyt Zobr. +/- sm. ch.

4. Marginální průměry je možné vypočítat a zobrazit v grafu tak, že efekt interakce v tabulce vybereme (kliknutím) a klikneme na *OK*. Zobrazí se dialog *Uspořádání faktorů*, v němž určíme, jak bude vypadat vytvořený graf. Pro účely tohoto příkladu nastavíme Kmen v seznamu *Osa x, horní* a Prostředí v seznamu *Vzor čáry*. Kliknutím na *OK* vytvoříme příslušný graf.

Osa x, horní	Vzor čar	
Prostredi Kmen	Prostredi Kmen	
		Specif. uspořádání faktorů v grafu





Vidíme, že krysy vychované v omezeném prostředí, dělaly více chyb než krysy vychované ve stimulujícím prostředí nezávisle na kmeni. Současně krysy se špatnou schopností orientovat se v bludišti dělaly nejvíc chyb, nejméně jich dělaly chytré krysy.



5. Výsledky ANOVA lze zobrazit také ve formě tabulky kliknutím na tlačítko *Všechny efekty* na záložce *Základ*. Významné efekty jsou zvýrazněny červeně.

	Jednorozměrné testy významnosti pro Chyby (Rats Sigma-omezená parametrizace					
	Dekompozice efektivní hypotézy					
	SČ	Stupně	PČ	F	р	
Efekt		volnosti				
Abs. člen	100233,4	1	100233,4	105,1353	0,000000	
Prostredi	5551,0	1	5551,0	5,8225	0,026705	
Kmen	7939,8	2	3969,9	4,1640	0,032635	
Prostredi*Kmen	16,1	2	8,0	0,0084	0,991604	
Chyba	17160.8	18	953.4			

Jak již bylo řečeno, v této analýze se projevil významný efekt faktorů *Prostředí* a *Kmen*. Zdůrazněme ovšem, že test významnosti nám neříká nic o tom, která (nebo které) ze skupin krys se od ostatních v počtu chyb významně liší. Abychom to zjistili, můžeme provést *Post-hoc testy*.

6. Klikneme na tlačítko *Více výsledků* a následně na záložku *Post-hoc*. V poli *Efekt* zvolíme *Kmen*, abychom mohli provést porovnání marginálních průměrů pro tento efekt. Kliknutím na tlačítko *Schefféův* se v tabulce zobrazí výsledky Schefféova testu:

	Scheffeho test; proměnná Chyby (Rats CZ) Pravděpodobnosti pro post-hoc testy Chyba: meziskup. PČ = 953,38, sv = 18,000				
Ĵ.	Kmen	{1}	{2}	{3}	
C. buňky		39,375	73,000	81,500	
1	Dobra,		0,121814	0,044361	
2	Nestala	0,121814		0,860446	
3	Spatna	0,044361	0,860446		





Tato tabulka zobrazuje statistickou významnost rozdílů průměrů pro všechny páry skupin krys. Jak je vidět, pouze rozdíl mezi 1. a 3. skupinou, tj. mezi hloupými a chytrými krysami, je statisticky významný na hladině významnosti 0,05. Lze tedy utvořit závěr, že pouze hloupé krysy dělaly významně více chyb než krysy chytré, zatímco průměrné krysy se od zbývajících dvou skupin nijak významně neliší.

7. Samozřejmě je třeba otestovat předpoklady, za kterých lze metodu ANOVA uplatňovat. Přepneme se proto na záložku *Předpoklady*. Jedním z předpokladů je homogenita rozptylů. *STATISTICA* poskytuje několik testů tohoto předpokladu ve skupině *Homogenita rozptylů/kovariancí* na záložce *Předpoklady*. Pro účely tohoto příkladu klikneme na tlačítko *Leveneův test (ANOVA)*.

	(1110) 1
👺 ANOVA Výsledky 1: Rats CZ	?- 🛛
Profily Vlastní testy Rezidua 1 Rezidua 2 Matice Protokol Detaily Průměry Plánované porovnání Post-hoc Předpoklady Proměnné Chyby Efekt: "Prostredi"*Kmen Homogenita rozptylů/kovariancí IIII Cochran, Hartley, Bartlett IIII Eveneňy test (ANITVA)	± Méně Zavřít Zavřít Změnit Možnosti▼ Anal.Skup.
Rozdělení prom. uvnitř skupin Bozdělení prom. uvnitř skupin Histogramy Norm. p-graf Bozd graf Bod. graf Polo-norm. graf z-transf. vnitřních korel. Graf průměrů vs. sm. odchylky	

Níže uvedená tabulka s výsledky tohoto testu nevykazuje žádné údaje indikující, že by rozptyl v jednotlivých skupinách byl významně odlišný (tj. podmínka homogenity rozptylů je splněna).

	Leveneův test homogenity rozptylů (Rats CZ Efekt: "Prostredi"*Kmen Stupně volnosti pro všechna F: 5, 18				
	PČ	PČ	F	р	
	Efekt	Chyba			
Chyby	226,4604	186,4132	1,214830	0,342166	

ANOVA předpokládá, že rozdělení závislé proměnné v jednotlivých skupinách je normální. I přesto je ANOVA velice robustní vzhledem k porušení tohoto předpokladu. Pro posouzení typu rozdělení závislé proměnné je možné využít několika grafů, které jsou ve skupině *Rozdělení vnitřních odchylek* nebo *Rozdělení prom. uvnitř skupin* na záložce *Předpoklady*. Pro přesnější ověření je možné použít např. *Shapirův-Wilkův* test normality v modulu *Základní statistiky/tabulky*.







8. Jak se zdá z výsledků analýzy, můžeme s velkou pravděpodobností říci, že faktory genetických dispozic i prostředí výchovy mají významný efekt na schopnost krys pohybovat se v bludišti.

