Počítačová analýza vícerozměrných dat v oborech přírodních, technických a společenských věd

Prof. RNDr. Milan Meloun, DrSc. (Univerzita Pardubice, Pardubice)

20.-24. června 2011

Tato prezentace je spolufinancována Evropským sociálním fondem a státním rozpočtem České republiky.

4.1 NEPŘÍMÁ POZOROVÁNÍ A KORELACE

Příroda je vícerozměrná

Příroda je vícerozměrná:

znaky a *objekty* - *m*-tice znaků pro *n* objektů, kde *n* >> *m*
výhodné použít co nejmenší počet znaků *m*

U ***experimentálních dat*** lze řadu znaků zkonstantnit nebo znáhodnit

U ***neexperimentální data*** (pasivního pozorování) je vágnost v hledání skryté souvislosti.

a) Analýza experimentálních dat je zaměřena na redukci rozmčrnosti aby bylo možné zkoumat obecně *nelineární vztahy mezi znaky.*

b) U necxpcrimcntálních dat se redukce rozmčrnosti provádí až při statistické analýze a předpokládají se *lineární vztahy mezi znaky.*

Přímá a nepřímá pozorování:

Přímá měření poskytuje málo metod, např. měření délky měřítkem.

Nepřímé měření je např. měření teploty teploměrem, tj. měření délky rtuťového sloupce a přepočet na teplotu.

Kombinace přímých a nepřímých pozorování: je např. koncentrace jako podíl hmotnosti a objemu.

Zdrojová matice dat *X (n* x *m)*

Data: x_1 značí délku těla, x_2 značí šířku těla, x_3 je délka předního křídla, x_4 je délka zadního křídla, x_5 je počet průduchů, x_6 je délka tykadla I, x_7 je délka tykadla II, x_8 je délka tykadla III, x_9 je délka tykadla IV, x_{10} je délka tykadla V, x_{11} je počet tykadlových ostnů, x_{12} je délka posledního článku nohy, x_{13} je délka holeně, tibia, x_{14} je délka stehna, x_{15} je délka sosáku, x_{16} je délka kladélka, x_{17} je počet kladélkových trnů, x_{18} je řitní otvor, x_{19} je počet háčků zadních křídel.

Zdrojová matice dat *X (n* x *m)*

Znaky x_1 až x_{19} (sloupce, $m = 19$)

Objekty
(řádky zde indexované od $n = 1$ do 23)

i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}
1	21.2	11.0	7.5	4.8	5.0	2.0	2.0	2.8	2.8	3.3	3.0	4.4	4.5	3.6	7.0	4.0	8.0	0.0	3.0
2	20.2	10.0	7.5	5.0	5.0	2.3	2.1	3.0	3.0	3.2	5.0	4.2	4.5	3.5	7.6	4.2	8.0	0.0	3.0
3	20.2	10.0	7.0	4.6	5.0	1.9	2.1	3.0	2.5	3.3	1.0	4.2	4.4	3.3	7.0	4.0	6.0	0.0	3.0
4	22.5	8.8	7.4	4.7	5.0	2.4	2.1	3.0	2.7	3.5	5.0	4.2	4.4	3.6	6.8	4.1	6.0	0.0	3.0
5	20.6	11.0	8.0	4.8	5.0	2.4	2.0	2.9	2.7	3.0	4.0	4.2	4.7	3.5	6.7	4.0	6.0	0.0	3.0
6	19.1	9.2	7.0	4.5	5.0	1.8	1.9	2.8	3.0	3.2	5.0	4.1	4.3	3.3	5.7	3.8	8.0	0.0	3.5
7	20.8	11.4	7.7	4.9	5.0	2.5	2.1	3.1	3.1	3.2	4.0	4.2	4.7	3.6	6.6	4.0	8.0	0.0	3.0
8	15.5	8.2	6.3	4.9	5.0	2.0	2.0	2.9	2.4	3.0	3.0	3.7	3.8	2.9	6.7	3.5	6.0	0.0	3.5
9	16.7	8.8	6.4	4.5	5.0	2.1	1.9	2.8	2.7	3.1	3.0	3.7	3.8	2.8	6.1	3.7	8.0	0.0	3.0
10	19.7	9.9	8.2	4.7	5.0	2.2	2.0	3.0	3.0	3.1	0.0	4.1	4.3	3.3	6.0	3.8	8.0	0.0	3.0
11	10.6	5.2	3.9	2.3	4.0	1.2	1.0	2.0	2.0	2.2	6.0	2.5	2.5	2.0	4.5	2.7	4.0	1.0	2.0
12	9.2	4.5	3.7	2.2	4.0	1.3	1.2	2.0	1.6	2.1	5.0	2.4	2.3	1.8	4.1	2.4	4.0	1.0	2.0
13	9.6	4.5	3.6	2.3	4.0	1.3	1.0	1.9	1.7	2.2	4.0	2.4	2.3	1.7	4.0	2.3	4.0	1.0	2.0
14	8.5	4.0	3.8	2.2	4.0	1.3	1.1	1.9	2.0	2.1	5.0	2.4	2.4	1.9	4.4	2.3	4.0	1.0	2.0
15	11.0	4.7	4.2	2.3	4.0	1.2	1.0	1.9	2.0	2.2	4.0	2.5	2.5	2.0	4.5	2.6	4.0	1.0	2.0
16	18.1	8.2	5.9	3.5	5.0	1.9	1.9	1.9	2.7	2.8	4.0	3.5	3.8	2.9	6.0	4.5	9.0	1.0	2.0
17	17.6	8.3	6.0	3.8	5.0	2.0	1.9	2.0	2.2	2.9	3.0	3.5	3.6	2.8	5.7	4.3	10.0	1.0	2.0
18	19.2	6.6	6.2	3.4	5.0	2.0	1.8	2.2	2.3	2.8	4.0	3.5	3.4	2.5	5.3	3.8	10.0	1.0	2.0
19	15.4	7.6	7.1	3.4	5.0	2.0	1.9	2.5	2.5	2.9	4.0	3.3	3.6	2.7	6.0	4.2	8.0	1.0	3.0
20	15.1	7.3	6.2	3.8	5.0	2.0	1.8	2.1	2.4	2.5	4.0	3.7	3.7	2.8	6.4	4.3	10.0	1.0	2.5
21	16.1	7.9	5.8	3.7	5.0	2.1	1.9	2.3	2.6	2.9	5.0	3.6	3.6	2.7	6.0	4.5	0.0	1.0	2.0
22	19.1	8.8	6.4	3.9	5.0	2.2	2.0	2.3	2.4	2.9	4.0	3.8	4.0	3.0	6.5	4.5	0.0	1.0	2.5
23	15.3	6.4	5.3	3.3	5.0	1.7	1.6	2.0	2.2	2.5	5.0	3.4	3.4	2.6	5.4	4.0	0.0	1.0	2.0

PŘÍKLAD 9.4 *Vytvoření dendrogramu neuroleptik*

Neuroleptika redukují nežádoucí účinky přebytečného dopaminu a liší se ve svých účincích: potlačují nervozitu, záchvaty, třes, ospalost, parkinsonismus, vynechávání menstruace, vyrážky, zvýšené slinění atd. Cílem je provést klasifikaci neuroleptik do shluků podobných účinků.

Data: Data ***Neuroleptika*** (převrácená hodnota mediánové účinné dávky *1/ED50* [kg/mg]):

Lek název neuroleptika,

Nervoz potlačení nervozity,

Stereo potlačení stereotypního chování,

Tres potlačení záchvatu a třesu a

Usmr dávka smrtícího účinku.

Data

Lek	Nervoz	Stereo	Tres	Usmr
1 Chlorphromazine	3.846	3.333	1.111	1.923
2 Promazine	0.323	0.213	0.108	1.429
3 Trifluperazine	27.027	17.857	0.562	0.14
4 Fluphenazine	17.857	15.385	1.695	1.075
5 Perphenazine	27.027	27.027	1.961	2.083
6 Thioridazine	0.244	0.185	0.093	1.333
7 Pifluthixol	142.857	142.857	20.408	163.934
8 Thiothixene	4.348	4.348	0.047	0.345
9 Chorprothixene	5.882	2.941	4.545	4.167
10 Spiperone	62.5	47.619	11.765	0.847
11 Haloperidol	52.632	62.5	1.282	0.568
12 Azaperone	2.941	1.282	2.222	3.03
13 Pipamperone	0.327	0.187	1.724	0.397
14 Pimozide	20.408	20.408	0.107	0.025
15 Metitepine	15.385	10.204	10.204	27.027
16 Clozapine	0.161	0.093	0.327	0.323
17 Perlapine	0.323	0.323	0.37	0.067
18 Sulpiride	0.047	0.047	0.003	0.001
19 Butaclamol	10.204	9.091	1.471	0.025
20 Molindone	7.692	7.692	0.14	38138

PŘÍKLAD 1.1 *Popisné statistiky jednorozměrné analýzy zdrojové matice dat Hrách*

Hrách Zdrojová matice dat *Hrách obsahuje znaky smyslového posouzení charakteristik rozličných odrůd hrachu. Objekty zde představují vzorky pěti různých odrůd hrachu A až E, které byly sklízeny v pěti rozličných obdobích 1 až 5. Výsledná zdrojová matice o 12 znacích převážně smyslových charakteristik obsahuje 60 vzorků hrachu. Posouzení každého objektu hrachu bylo provedeno 10 porotci dvojím odhadem tak, že smyslové charakteristiky byly bodovány ve stupnici od 1 (nejhorší) do 9 (nejlepší). Tak bylo získáno 1200 řádků (objektů) postupem: 60 vzorků * 2 hodnocení * 10 porotců. V praxi se data obvykle průměrují, aby se kompenzovaly rozdíly v subjektivní škále přísnosti jednotlivých porotců. Výsledkem je pro každý z šedesáti objektů průměrná hodnota senzorického hodnocení. Cílem úlohy je: 1.* průměrovat data, 2. vynést původní data do grafu a 3. vypočítat popisné jednorozměrné statistiky.

Data

- ***Data:*** *Zdrojová matice dat n = 1200, m = 12 byla průměrována a výsledkem byla matice 60 * 12. Obsahovala průměrné hodnoty senzorického hodnocení pro znaky ve sloupcích: Aro je aroma, Slad je sladkost, Med je medovost, Bez je bezchuťovost, Klas je klasovost, Tvrd je tvrdost, Bel je bělost, Bar1 je barva 1, Bar2 je barva2, Bar3 je barva3, Slup je slupka, Ztr je ztráta.*

Data

Objekt	*Aro*	*Slad*	*Med*	*Bez*	*Klas*	*Tvrd*	*Bel*	*Bar1*	*Bar2*
	Bar3	*Slup*	*Ztr*						
B5	6.48	6.66	4.56	2.2	2.91	3.47	4.72	5.59	5.73
	5.99	4.26	3.25						
C4	5.75	6.09	3.81	2.32	4.03	3.77	4.17	5.73	5.75
	5.32	3.82	3.38						
B2	3.94	4.12	2.44	3.63	5.77	5.39	4.77	6.67	5.11
	4.60	3.5	3.03						

STATISTICA Cz - [Data: 11Hrach.sta (13s krát 82ř)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

Zdrojová matice dat

	1 Objekt	2 Aro	3 Slad	4 Med	5 Bez	6 Klas	7 Tvrd	8 Bel	9 Bar1	10 Bar2	11 Bar3	12 Slup	13 Ztr
1	B5	6,480	6,660	4,560	2,200	2,910	3,470	4,720	5,585	5,735	5,985	4,260	3,250
2	C4	5,750	6,090	3,810	2,320	4,030	3,770	4,170	5,730	5,745	5,325	3,820	3,380
3	B2	3,940	4,120	2,440	3,630	5,770	5,390	4,770	6,665	5,105	4,595	3,500	3,030
4	D5	6,600	6,120	4,440	1,930	3,310	4,460	4,860	5,160	5,740	6,565	2,120	3,940
5	D4	5,680	5,980	3,800	2,120	3,850	4,140	5,030	5,635	5,220	5,480	2,380	5,160
6	E2	4,740	4,660	2,880	2,940	5,650	5,770	5,310	5,940	5,270	5,890	1,750	3,640
7	B5	6,310	6,130	4,780	1,940	2,700	3,260	5,070	5,710	5,370	6,365	3,650	4,550
8	C5	6,200	6,020	4,650	1,780	3,120	3,740	5,250	5,655	5,475	5,960	2,510	3,800
9	C2	3,790	3,880	2,310	3,520	6,240	5,730	5,390	6,300	5,135	5,230	2,010	4,110
10	A4	5,680	6,340	3,750	2,790	4,170	3,870	4,520	4,920	5,760	4,570	2,970	4,410
11	D4	6,100	6,090	3,990	2,070	4,260	4,250	4,010	5,020	6,175	5,380	2,500	4,600
12	B1	3,410	3,180	1,820	4,640	6,240	7,430	4,260	4,835	5,955	4,550	1,850	4,270
13	D4	5,890	6,090	3,990	2,290	3,900	4,590	4,530	5,890	5,630	3,815	2,200	4,710
14	E4	5,770	5,320	3,880	2,260	4,220	4,990	5,050	5,335	5,590	5,540	2,160	3,650
15	B1	3,390	3,280	1,980	4,500	6,040	7,140	4,350	5,090	5,580	4,400	2,030	4,620
16	B5	6,570	6,880	4,830	1,970	2,920	3,390	3,860	4,615	6,665	6,660	2,220	3,580
17	D4	5,860	6,180	3,940	2,200	3,800	4,910	4,350	5,180	6,255	5,760	2,270	3,610
18	C2	3,960	4,480	2,300	3,940	6,230	6,410	4,470	5,105	5,720	4,970	2,050	4,530
19	A5	6,220	6,790	4,260	2,400	2,630	3,160	5,680	7,010	4,860	3,255	3,040	4,590
20	C3	5,110	5,250	3,090	3,270	5,280	5,240	5,610	6,595	5,110	3,950	2,740	4,230
21	B2	3,770	3,970	2,180	4,370	6,470	6,550	4,950	6,055	5,310	4,395	2,210	4,760
22	B5	7,090	6,090	5,180	1,740	2,570	3,180	5,230	5,920	5,515	4,115	2,090	3,100
23	D4	5,720	5,300	3,730	2,340	3,950	4,800	3,640	4,000	6,805	6,755	1,740	2,930
24	B1	3,220	3,210	1,950	4,420	6,240	7,270	4,600	6,030	5,600	4,165	1,680	3,580
25	A5	6,110	6,620	4,290	2,580	3,200	2,[illegible]	3,500	[illegible]	6,[illegible]	5,370	[illegible]	4,310
26	D4	6,070	6,270	3,980	2,190	3,890	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
27	A1	2,660	2,660	1,430	6,100	6,670	7,750	4,270	[illegible]	5,630	4,525	1,650	3,780
28	B3	5,260	5,490	3,460	3,030	4,850	4,170	5,220	5,410	5,415	6,120	3,080	3,950
29	C2	3,720	4,350	2,200	4,080	6,500	6,270	4,990	5,535	5,560	5,335	1,820	3,920
30	D3	5,430	5,190	3,470	2,400	4,430	5,260	4,460	4,780	5,720	5,895	1,610	3,770
31	B5	6,550	6,570	4,710	2,120	3,060	3,430	3,760	4,430	6,450	6,380	2,630	3,850
32	E3	5,530	5,410	3,680	2,470	4,720	5,780	3,880	4,335	6,470	6,790	1,800	2,950
33	C3	4,710	4,680	2,680	3,190	5,320	5,920	4,320	4,765	6,215	4,860	2,330	4,020
34	A5	6,280	7,030	4,910	2,380	2,190	2,600	4,560	5,905	5,585	4,950	3,630	3,380
35	D4	5,920	5,820	3,750	2,060	3,880	3,870	4,530	5,190	5,825	5,630	2,450	3,710
36	E4	6,090	5,720	3,800	1,940	4,440	4,450	3,940	4,625	6,510	7,175	2,180	2,990
37	B5	6,370	6,500	4,680	2,140	2,890	3,530	4,600	5,735	5,560	4,065	2,880	4,340
38	A4	5,710	5,680	3,970	2,650	4,390	3,720	5,320	6,280	5,120	6,075	2,390	2,550
39	C3	4,530	5,030	2,640	3,120	5,860	4,920	5,150	6,965	5,125	4,275	2,130	3,740
40	C4	5,950	6,280	4,040	2,190	3,930	3,610	4,120	5,395	5,815	4,505	3,090	5,340
41	D3	5,510	5,410	3,720	2,780	4,760	5,270	3,880	4,280	6,355	5,325	2,250	5,030
42	A1	3,100	3,430	1,800	4,860	6,220	7,070	4,140	5,275	5,580	3,695	2,050	4,730
43	A5	6,500	6,680	4,770	2,230	2,090	2,870	5,510	6,375	4,845	4,785	2,710	3,700
44	D3	5,460	5,410	3,270	2,970	5,150	4,980	3,610	4,305	6,600	5,435	2,370	4,440
45	A2	3,750	4,300	2,220	4,270	6,100	6,270	4,060	5,140	5,870	4,220	2,230	5,010
46	C4	5,860	5,270	3,730	2,500	3,860	4,300	4,150	4,495	6,230	6,140	2,110	3,290
47	A5	6,160	6,970	4,800	2,500	2,870	3,170	4,270	5,315	6,090	5,255	3,350	4,430
48	B2	3,870	3,880	2,230	4,060	5,990	6,310	4,450	5,530	5,785	4,980	1,930	3,620
49	B5	6,240	5,800	4,260	2,130	3,240	3,420	5,120	5,940	5,465	4,775	3,110	3,980

Pro nápovědu stiskněte F1

Ř1;S11 5,985 Filtr - Váhy: VYPN ABC 123 ZÁZN

Start Total Commander 6.0 - P... Doručená pošta - Micros... Prezentace2 STATISTICA Cz - [Dat... 10:28

PŘÍKLAD 2.2 Průzkumová analýza zdrojové matice dat demografického souboru Lidé

Vyšetřete grafickými diagnostikami průzkumové analýzy vícerozměrných dat, které ze 12 znaků demografického souboru dat *Lidé jsou nejvýhodnější k charakterizaci osob a které znaky mají největší míru rozptýlení. Matice obsahuje data pro n = 32 osob a m = 12 znaků, kde 16 osob bylo vybráno ze Skandinávie (kód A) a 16 osob ze Středomoří (kód B), 16 osob jsou muži (kód M) a 16 osob jsou ženy (kód F).*

Data: *Znaky obsahují u každé osoby výšku [cm], hmotnost [kg], délku vlasů [krátká: −1, dlouhá: +1], velikost boty [evropský standard], věk [roky], příjem [Euro], spotřeba piva [litry na rok], spotřeba vína [litry na rok], pohlaví [muž: −1, žena: +1], schopnost plavat [naměřený čas na uplavání 500 m], původ [A: −1 Skandinávie, B: +1 Středomoří], inteligenční kvocient IQ [evropský standardizovaný test IQ]. Mezi znaky jsou tři dichotomické, binární proměnné, a to pohlaví, délka vlasů a původ a ostatních 9 znaků nabývá kvantitativních hodnot.*

Osoba	Výška Plavání	Hmotnost Původ	Vlasy IQ	Boty	Věk	Příjem	Pivo	Víno	Sex
MA	198 98	92 −1	−1 100	48	48	45000	420	115	−1
MA	184 92	84 −1	−1 130	44	33	33000	350	102	−1
MA	183 91	83 −1	−1 127	44	37	34000	320	98	−1
FA	166 75	47 −1	−1 112	36	32	28000	270	78	1

PŘIKLAD 2.6 Sledování spotřeby proteinů v zemích Evropy

Sledována spotřeba proteinů v 25 zemích Evropy formou spotřeby 9 druhů potravin. Cílem je odhalit, zda existuje korelace mezi znaky, tj druhy potravin? Lze odhalit nějaké interakce mezi druhy potravin a zeměmi?

Data: v datech *Proteiny* jsou uvedeny znaky: *Cervene* značí spotřebu červeného masa, *Bile* značí spotřebu bílého masa, *Vejce* značí spotřebu vajec, *Mléko* se týká spotřeby mléka, *Ryby* značí spotřebu ryb, *Obiln* značí spotřebu obilnin, *Škrob* značí spotřebu škrobu, *Ořech* značí spotřebu ořechů, *Ovoce* značí spotřebu ovoce a zeleniny

Data

Země	*Cervene*	*Bile*	*Vejce*	*Mleko*	*Ryby*	*Obiln*	*Skrob*	*Orech*	*Ovoce*
Albania	10.1	1.4	0.5	8.9	0.2	42.3	0.6	5.5	1.7
Austria	8.9	14	4.3	19.9	2.1	28	3.6	1.3	4.3
Belgium	13.5	9.3	4.1	17.5	4.5	26.6	5.7	2.1	4
Bulgaria	7.8	6	1.6	8.3	1.2	56.7	1.1	3.7	4.2
Czechoslovakia	9.7	11.4	2.8	12.5	2	34.3	5	1.1	4
Denmark	10.6	10.8	3.7	25	9.9	21.9	4.8	0.7	2.4
East Germany	8.4	11.6	3.7	11.1	5.4	24.6	6.5	0.8	3.6
Finland	9.5	4.9	2.7	33.7	5.8	26.3	5.1	1	1.4
France	18	9.9	3.3	19.5	5.7	28.1	4.8	2.4	6.5
Greece	10.2	3	2.8	17.6	5.9	41.7	2.2	7.8	6.5
Hungary	5.3	12.4	2.9	9.7	0.3	40.1	4	5.4	4.2
Ireland	13.9	10	4.7	25.8	2.2	24	6.2	1.6	2.9
Italy	9	5.1	2.9	13.7	3.4	36.8	2.1	4.3	6.7
Netherlands	9.5	13.6	3.6	23.4	2.5	22.4	4.2	1.8	3.7
Norway	9.4	4.7	2.7	23.3	9.7	23	4.6	1.6	2.7
Poland	6.9	10.2	2.7	19.3	3	36.1	5.9	2	6.6
Portugal	6.2	3.7	1.1	4.9	14.2	27	5.9	4.7	7.9
Romania	6.2	6.3	1.5	11.1	1	49.6	3.1	5.3	2.8
Spain	7.1	3.4	3.1	8.6	7	29.2	5.7	5.9	7.2
Sweden	9.9	7.8	3.5	24.7	7.5	19.5	3.7	1.4	2
Switzerland	13.1	10.1	3.1	23.8	2.3	25.6	2.8	2.4	4.9
UK	17.4	5.7	4.7	20.6	4.3	24.3	4.7	3.4	3.3
USSR	9.3	4.6	2.1	16.6	3	43.6	6.4	3.4	2.9
West Germany	11.4	12.5	4.1	18.8	3.4	18.6	5.2	1.5	3.8
Yugoslavia	4.4	5	1.2	9.5	0.6	55.9	3	5.7	3.2

Úloha 4. Faktorová analýza při klasifikaci vzorků vín (Kompendium E408)

Pro 38 vzorků vín bylo nalezeno 24 analytických obsahů stopových prvků a charakteristických fyzikálně-chemických vlastností. Utvořte shluky podobných vlastností a dále shluky podobných vín.

Index	Cd	Mo	Mn	Ni	Cu	Al	Ba	Cr	Sr	Pb	B	Mg	Si	Na	Ca	P	K	Arom	Clar	Body	Flavor	Oakn	Quality	Reg
1	0.005	0.044	1.51	0.122	0.83	0.982	0.387	0.029	1.23	0.561	2.63	128	17.3	66.8	80.5	150	1130	3.3	1	2.8	3.1	4.1	9.8	1
2	0.055	0.16	1.16	0.149	0.066	1.02	0.312	0.038	0.975	0.697	6.21	193	19.7	53.3	75	118	1010	4.4	1	4.9	3.5	3.9	12.6	1
3	0.056	0.146	1.1	0.088	0.643	1.29	0.308	0.035	1.14	0.73	3.05	127	15.8	35.4	91	161	1160	3.9	1	5.3	4.8	4.7	11.9	1
4	0.063	0.191	0.959	0.38	0.133	1.05	0.165	0.036	0.927	0.796	2.57	112	13.4	27.5	93.6	120	924	3.9	1	2.6	3.1	3.6	11.1	1
5	0.011	0.363	1.38	0.16	0.051	1.32	0.38	0.059	1.13	1.73	3.07	138	16.7	76.6	84.6	164	1090	5.6	1	5.1	5.5	5.1	13.3	1
6	0.05	0.106	1.25	0.114	0.055	1.27	0.275	0.019	1.05	0.491	6.56	172	18.7	15.7	112	137	1290	4.6	1	4.7	5	4.1	12.8	1
7	0.025	0.479	1.07	0.168	0.753	0.715	0.164	0.062	0.823	2.06	4.57	179	17.8	98.5	122	184	1170	4.8	1	4.8	4.8	3.3	12.8	1
8	0.024	0.234	0.906	0.466	0.102	0.811	0.271	0.044	0.963	1.09	3.18	145	14.3	10.5	91.9	187	1020	5.3	1	4.5	4.3	5.2	12	1
9	0.009	0.058	1.84	0.042	0.17	1.8	0.225	0.022	1.13	0.048	6.13	113	13	54.4	70.2	158	1240	4.3	1	4.3	3.9	2.9	13.6	3
10	0.033	0.074	1.28	0.098	0.053	1.35	0.329	0.03	1.07	0.552	3.3	140	16.3	70.5	74.7	159	1100	4.3	1	3.9	4.7	3.9	13.9	1
11	0.039	0.071	1.19	0.043	0.163	0.971	0.105	0.028	0.491	0.31	6.56	103	9.5	45.3	67.9	133	1090	5.1	1	4.3	4.5	3.6	14.4	3
12	0.045	0.147	2.76	0.071	0.074	0.483	0.301	0.087	2.14	0.546	3.5	199	9.2	80.4	66.3	212	1470	3.3	0.5	5.4	4.3	3.6	12.3	2
13	0.06	0.116	1.15	0.055	0.18	0.912	0.166	0.041	0.578	0.518	6.43	111	11.1	59.7	83.8	139	1120	5.9	0.8	5.7	7	4.1	16.1	3
14	0.067	0.166	1.53	0.041	0.043	0.512	0.132	0.026	0.229	0.699	7.27	107	6	55.2	44.9	148	854	7.7	0.7	6.6	6.7	3.7	16.1	3
15	0.077	0.261	1.65	0.073	0.285	0.596	0.078	0.063	0.156	1.02	5.04	94.6	6.3	10.4	54.9	132	899	7.1	1	4.4	5.8	4.1	15.5	3
16	0.064	0.191	1.78	0.067	0.552	0.633	0.085	0.063	0.192	0.777	5.56	110	7	13.6	64.1	167	976	5.5	0.9	5.6	5.6	4.4	15.5	3
17	0.025	0.009	1.57	0.041	0.081	0.655	0.072	0.021	0.172	0.232	3.79	75.9	6.4	11.6	48.1	132	995	6.3	1	5.4	4.8	4.6	13.8	3
18	0.02	0.027	1.74	0.046	0.153	1.15	0.094	0.021	0.358	0.025	4.24	80.9	7.9	38.9	57.6	136	876	5	1	5.5	5.5	4.1	13.8	3
19	0.034	0.05	1.15	0.058	0.058	1.35	0.294	0.006	1.12	0.206	2.71	120	14.7	68.1	64.8	133	1050	4.6	1	4.1	4.3	3.1	11.3	1
20	0.013	0.03	2.82	0.058	0.05	0.623	0.349	0.082	2.91	0.171	3.54	208	9.3	79.2	66.4	266	1430	3.4	0.9	5	3.4	3.4	7.9	2
21	0.043	0.268	2.32	0.066	0.314	0.627	0.099	0.045	0.36	1.28	5.68	98.4	9.1	19.5	64.3	176	945	6.4	0.9	5.4	6.6	4.8	15.1	3
22	0.061	0.245	1.61	0.07	0.172	2.07	0.071	0.053	0.186	1.19	4.42	87.6	7.6	11.6	70.6	156	820	5.5	1	5.3	5.3	3.8	13.5	3
23	0.047	0.161	1.47	0.154	0.082	0.546	0.181	0.06	0.898	0.747	8.11	160	19.3	12.5	82.1	218	1220	4.7	0.7	4.1	5	3.7	10.8	2
24	0.048	0.146	1.85	0.092	0.09	0.889	0.328	0.1	1.32	0.604	6.42	134	19.3	125	83.2	173	1810	4.1	0.7	4	4.1	4	9.5	2
25	0.049	0.155	1.73	0.051	0.158	0.653	0.081	0.037	0.164	0.767	4.91	86.5	6.5	11.5	53.9	172	1020	6	1	5.4	5.7	4.7	12.7	3
26	0.042	0.126	1.7	0.112	0.21	0.508	0.299	0.054	0.995	0.686	6.94	129	43.6	45	85.9	165	1330	4.3	1	4.6	4.7	4.9	11.6	2
27	0.058	0.184	1.28	0.095	0.058	1.3	0.346	0.037	1.17	1.28	3.29	145	16.7	65.8	72.8	175	1140	3.9	1	4	5.1	5.1	11.7	1
28	0.065	0.211	1.65	0.102	0.055	0.308	0.206	0.028	0.72	1.02	6.12	99.3	27.1	20.5	95.2	194	1260	5.1	1	4.9	5	5.1	11.9	2
29	0.065	0.129	1.56	0.166	0.151	0.373	0.281	0.034	0.889	0.638	7.28	139	22.2	13.3	84.2	164	1200	3.9	1	4.4	5	4.4	10.8	2
30	0.068	0.166	3.14	0.104	0.053	0.368	0.292	0.039	1.11	0.831	4.71	125	17.6	13.9	59.5	141	1030	4.5	1	3.7	2.9	3.9	8.5	2
31	0.067	0.199	1.65	0.119	0.163	0.447	0.292	0.058	0.927	1.02	6.97	131	38.3	42.9	85.9	164	1390	5.2	1	4.3	5	6	10.7	2
32	0.084	0.266	1.28	0.087	0.071	1.14	0.158	0.049	0.794	1.3	3.77	143	19.7	39.1	128	146	1230	4.2	0.8	3.8	3	4.7	9.1	1
33	0.069	0.183	1.94	0.07	0.095	0.465	0.225	0.037	1.19	0.915	2	123	4.6	7.5	69.4	123	943	3.3	1	3.5	4.3	4.5	12.1	1
34	0.087	0.208	1.76	0.061	0.099	0.683	0.087	0.042	0.168	1.33	5.04	92.9	7	12	56.3	157	949	6.8	1	5	6	5.2	14.9	3
35	0.074	0.142	2.44	0.051	0.052	0.737	0.408	0.022	1.16	0.745	3.94	143	6.8	36.8	67.6	82	1170	5	0.8	5.7	5.5	4.8	13.5	1
36	0.084	0.171	1.85	0.088	0.038	1.21	0.263	0.072	1.35	0.899	2.38	130	6.2	101	64.4	99	1070	3.5	0.8	4.7	4.2	3.3	12.2	1
37	0.106	0.307	1.15	0.063	0.051	0.643	0.29	0.031	0.885	1.61	4.4	151	17.4	7.3	103	177	1100	4.3	0.8	5.5	3.5	5.8	10.3	1
38	0.102	0.342	4.08	0.065	0.077	0.752	0.366	0.048	1.08	1.77	3.37	145	5.3	33.1	58.3	117	1010	5.2	0.8	4.8	5.7	3.5	13.2	1

Úloha 6. Klasifikace vlastností rozličných druhů kávy (Kompendium E406)

U 43 vzorků kávy ze 30 zemí byly změřeny chemické a fyzikální vlastnosti. Nalezněte shluky podobných vlastností a shluky podobných prvků.

Data: 13 proměnných (sloupce): **i** index kávy, **j** je původ kávy, **x1** obsah vody, **x2** hmotnost zrn, **x3** extrakt, **x4** pH, **x5** volná acidita, **x6** obsah minerálů, **x7** tuky, **x8** kofein, **x9** trinonelin, **x10** kyselina chlorogeniková, **x11** kyselina neochlorogeniková, **x12** kyseliny isochlorogeniková, **x13** suma kyselin chlorogenikových.

Úloha 6. Klasifikace vlastností rozličných druhů kávy (Kompendium E406)

i	ii	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13
1	Mexico 1	8.9	156.6	33.5	5.8	32.7	3.8	15.2	1.1	1	5.4	0.4	0.8	6.6
2	Mexico 2	7.4	157.3	32.1	5.8	30.8	3.7	15	1.3	1	5.1	0.3	1	6.4
3	Guatemala	9.7	152.9	33.1	5.3	36.7	4.2	16.1	1.2	1	5.9	0.2	0.8	6.9
4	Honduras	10.4	174	31.5	5.6	34.2	3.9	15.8	1.1	0.9	5.9	0.4	0.6	6.8
5	Salvador 1	10.5	145.1	35.2	5.8	31.8	4.1	15.2	1.1	1	5.1	0.5	0.7	6.3
6	Salvador 2	10	156.4	34.5	5.8	32.6	3.9	15.4	1.2	0.8	5.3	0.4	0.7	6.4
7	Salvador 3	8.2	155.2	32.4	5.6	29.7	3.8	15.6	1.3	1.2	4.8	0.3	0.7	5.9
8	Nicaragua 1	9.2	167.8	30.6	5.9	28.9	3.8	15.1	1.3	1	5	0.3	0.7	5.9
9	Nicaragua 2	9.3	165.4	35.3	5.8	32.6	4.2	14.3	1.2	1	5.5	0.4	0.8	6.7
10	Costa Rica 1	7.1	180.3	33	5.8	29.3	4	15.1	1.3	1	5.1	0.3	0.7	6.1
11	Costa Rica 2	7.6	153.2	36	5.9	30.5	3.9	16.8	1.4	1.1	5.3	0.3	0.7	6.3
12	Costa Rica 3	7.3	159.6	35	5.8	29.9	3.7	16.5	1.2	1.2	5.5	0.3	0.7	6.5
13	Panama	9.3	161.8	32.4	5.8	31	3.7	15.5	1.3	1.2	5.6	0.3	0.6	6.6
14	Haiti	8.3	160.8	35.7	5.9	30	4.4	13	1.3	1	6.1	0.6	0.8	7.5
15	Dominica	11.6	174.8	32.5	5.4	35.2	3.7	14.5	1	1	5.7	0.3	0.5	6.5
16	Venezuela 1	9.7	169.1	34	5.8	31.6	4	15.7	1.3	1.3	5.1	0.3	0.3	6.2
17	Venezuela 2	10.6	163.7	35	5.8	35	3.8	15.8	1.2	1.1	6.1	0.3	0.9	7.3
18	Columbia 1	12	178.8	32.9	5.3	36.2	4.4	15.6	1.3	1	5.6	0.4	0.7	6.7
19	Columbia 2	10.6	169.1	33	5.3	37.5	4.4	15.1	1.2	1	6.1	0.1	0.6	6.9
20	Ecuador	11.6	148.5	34.6	5.3	39.4	4.2	14.6	1	1.1	5.7	0.5	0.4	6.6
21	Peru	10.1	153.7	34.5	6	28.4	3.7	15.9	1.3	1.1	6.1	0.4	0.8	7.3
22	Brasil 1	10.7	134.5	29.8	5.4	34.1	3.7	15.8	1.2	0.9	5.4	0.4	0.6	6.4
23	Brasil 2	9.7	160.7	33.8	5.3	37.2	4.2	15.2	1.1	0.9	5.4	0.3	0.5	6.2
24	Brasil 3	10.8	133.2	35	5.2	34.7	4.5	15.1	1.2	1.4	5	0.5	0.5	6
25	Brasil 4	11.1	131.7	29.8	5.4	33	4.1	15.8	1.1	1.2	5.1	0.5	0.5	6
26	Brasil 5	10.1	121.6	33.6	5.4	34.7	3.5	15.4	1.1	0.9	5.5	0.4	0.6	6.5
27	Cotedivoir	8	141.8	33.7	5.8	41.9	4.2	11	2	0.5	6.4	0.6	1.5	8.5
28	Togo	9	144.6	29.9	5.6	38	3.9	7.5	1.9	0.3	5.4	0.8	0.9	7.1
29	Cameroon	10.3	119.2	35.5	6.1	41.7	4.1	9.8	1.8	0.8	6	0.5	1.1	7.6
30	Congo	10	143.2	31.7	6.1	29.3	4.1	17	1.2	0.6	5.4	0.3	0.7	6.4
31	Angola 1	9.2	150.4	31.5	5.7	36.4	4.2	8.5	1.9	0.6	5.9	0.6	1.4	7.9
32	Angola 2	9.6	136.6	33.9	5.6	38.2	4	7.2	2.2	0.5	6.2	0.4	1.6	8.3
33	Angola 3	9.5	136.5	32	5.8	31.2	3.8	14.6	1.3	1	5.2	0.4	0.8	6.4
34	Ethiopie	9.3	124.2	35.6	5.8	31.8	3.8	15.7	0.9	0.9	5.5	0.2	0.8	6.5
35	Uganda 1	10.5	132.9	36.2	5.4	36.7	4	15.6	1	1	5.9	0.4	0.6	6.9
36	Uganda 2	10.7	181.2	33.1	5.8	30.7	3.9	15.8	1.3	1.1	5.3	0.3	0.6	6.2
37	Kenya	10.5	159.1	30.3	5.6	31.5	3.7	15.2	1.3	0.9	5.1	0.3	0.7	6
38	Tanganika	9.9	169.4	29	5.6	30.2	3.7	16.5	1.3	0.9	5	0.2	0.7	5.9
39	Madagascar	5	152	30.6	5.3	40.5	3.9	9.6	1.6	0.7	5.3	0.6	0.8	6.7
40	India	11.5	156.8	30.8	5.5	37.5	3.9	14.3	1.2	1	5.8	0.4	0.4	6.6
41	Sumatra	8.4	110.8	31.6	5.7	43.4	4.5	10.1	1.7	0.8	6.3	0.7	0.9	7.9
42	Java	5.6	163.1	34.5	5.5	33.3	4	16	1.2	1.1	5.1	0.3	0.8	6.3
43	Hawai	9.7	191.2	35.1	5.6	34.6	4.2	14.2	1.1	0.9	0.7	0.5	0.3	6.5

PŘÍKLAD 4.5 *Chromatografická analýza farmakologických sloučenin*

Byly měřeny hodnoty R_F pro 20 sloučenin s 18 eluenty. Žádné eluční činidlo však neprovedlo úplné rozdělení. Cílem je nalézt minimální výběr elučních činidel, které by daly dostatek informace pro kvalitativní analýzu.

Data

Datový soubor GIUSEPPE obsahuje 100 x R_F pro 20 sloučenin (v řádcích byla jména zkrácena na maximálně 8 písmen) a ve sloupcích je 18 elučních činidel představujících zde znaky: *i* vzorek, x_1 směs toluen : aceton : ethanol: 30 % amoniak = 45 : 45 : 7 : 3, x_2 směs ethylacetát: benzen : methanol : 30 % amoniak = 60 : 35 : 6.5 : 2.5, x_3 směs benzen : dioxan : ethanol : 30 % amoniak = 50 : 40 : 7.5 : 2.5, x_4 směs methanol : 30 % amoniak = 100 : 1.5, x_5 směs benzen : 2-propanol : methanol : 30 % amoniak = 70 : 30 : 20 : 5, x_6 směs ethylacetát: methanol : 30 % amoniak = 85 : 10 : 5, x_7 směs cyklohexan : toluen : diethylamin = 65 : 25 : 10, x_8 směs cyklohexan : toluen ; diethylamin = 75 : 15 : 10, x_9 směs cyklohexan : benzen : metanol : diethylamin = 70 : 20 : 10 : 5, x_{10} směs chloroform : aceton : diethylamin — 50 : 40 ; 10, x_{11} směs cyklohexan : chloroform : diethylamin = 50 : 40 : 10, x_{12} směs benzen : ethylacetát : diethylamin = 50 : 40 : 10, x_{13} směs xylen : methylethylketon : methanol : diethylamin = 40 : 40 : 6 : 2, x_{14} směs diethylether : diethylamin — 95 : 5, x_{15} směs ethylacetát : chloroform = 50 : 50, x_{16} směs ethylacetát : chloroform [A] = 50 : 50, x_{17} směs butanol : methanol = 40 : 60, x_{18} směs butanol: methanol [A] = 40 ; 60, kde [A] značí, že byl užit 0.1M methanolát draselný.

Data

i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}
Atropine	20	16	29	23	62	33	4	2	13	47	25	42	18	12	0	0	5	8
Biperide	91	90	87	68	92	87	73	72	64	85	81	86	68	94	11	40	40	65
Caffeine	55	42	52	68	77	54	8	5	13	60	30	51	41	20	13	12	54	57
Cocaine	81	82	81	71	87	82	46	41	38	81	72	80	52	72	6	24	30	57
Codeine	38	31	44	39	71	43	12	9	16	49	29	36	22	14	0	0	15	21
Cyclizin	71	72	80	64	85	80	49	47	40	75	71	73	39	59	2	9	34	54
Diazepam	76	79	80	78	85	80	28	21	29	80	61	75	72	54	54	50	85	87
Ketamine	77	79	80	76	86	79	71	33	32	81	66	76	67	66	27	37	66	79
Lignocaine	77	79	80	73	86	80	35	30	28	84	73	77	66	64	25	54	68	84
Lorazepam	47	34	53	77	73	46	2	0	12	52	7	22	47	8	28	15	85	79
Mebeveri	85	90	90	65	90	85	43	33	38	88	70	87	62	76	5	29	29	53
Methadon	85	84	88	48	89	83	63	64	48	85	73	86	38	86	1	10	13	29
Morphine	18	9	15	39	56	20	2	0	5	20	2	8	13	3	0	1	16	18
Naloxone	48	40	62	75	79	48	15	11	22	52	26	40	60	21	18	21	67	77
Papaverine	68	66	76	79	88	71	12	7	18	78	56	62	54	30	28	41	76	80
Pentazoc	72	66	81	65	87	76	22	18	26	69	41	54	39	44	2	11	32	59
Phenacet	64	58	62	79	87	66	4	1	13	68	18	41	58	24	41	40	86	84
Phenazon	66	53	70	83	86	65	30	22	24	77	60	66	45	54	15	21	68	74
Prazepam	81	83	86	83	88	81	41	31	35	83	66	82	75	74	65	67	86	88
Procaine	64	60	70	65	82	73	8	5	16	66	24	54	37	50	1	11	29	53

Dělení na strukturovaná a nestrukturovaná data

DRUHY DAT

Nestrukturovaná data

- matice ***X*** *(n* x *m*) nepředpokládá žádná speciální struktura mezi *znaky*
  - čili sloupci matice *X.*
- a) **Kvantitativní a semikvantitativní data:** vyšetřuje se
  - analýza parametrů polohy (vektoru průměrů),
  - rozptýlení (kovarianční respektive korelační matici),
  - přítomnost vybočujících bodů, předpoklady normality, standardní statistické testy: *PCA*

Analýza hlavních komponent (PCA):

lineární transformace původních os do souřadnicového systému hlavních komponent, které jsou vzájemné ortogonální (nekorelované).

V *PCA* osy postihují maximální množství informací vyjádřené variabilitou mezi objekty.

Relativní pozice objektů zůstává zachována.

Nový systém os je natočen do směrů, které postihují maximální *variabilitu minimalizují vzdálenosti objektů* od hlavních komponent.

Každý objekt má nové souřadnice, které se označují *skóre.*

b) Kvalitativní a semikvalitativní data:

Kvalitativní data bývají ve tvaru kontigenčníeh tabulek

- (lineární proměnné kódované 0 a 1).

Korespondenční analýza **(CA)** je PCA pro kontigenční tabulky.

Využívá ortogonálního rozkladu χ^2-statistiky, která vyjadřuje míru asociace.

Sloupce a řádky u CA jsou symetrické a lze je vyjádřit jedním grafem.

Korespondenční analýza seje duální, optimální, škálování nebo jako reciproké průměrování.

Vícenásobná korespondenční analýza (MCA): analyzuje několik binárních proměnných.

Vícerozměrněškálování(MDS)

vyjadřuje podobnosti či vzdálenosti mezi objekty.

Znázorňuje objekty na mapě tak, že eukleidovská vzdálenost zde odpovídá přibližně původním koeficientům podobnosti respektive vzdálenosti.

Klasická MDS je použita pro vzdálenosti a nemetrická MDS pro podobnosti.

Shluková analýza (CLU): se užívá když řádky a sloupce matice dat reprezentují stejný objekt.

Shluková analýza využívá znázornění ve stromové struktuře (dendrogramy).

Data musí obsahovat užitečnou informaci:

Předpokladem analýzy dat: data musí obsahovat požadovanou informaci.

Např. u stanovení koncentrace sloučeniny musí meření roztoku monitorovat tuto sloučeninu.

Žádná statistická metoda nemůže pomoci, když data neobsahují dostatečné množství informace o vlastnosti či jevu.

Objem informace v datech: závisí na způsobu formulování problému, dostatečná pozorování, měření, experimenty,

Relevantní data jsou data, která dostatečně vypovídají,

STRUKTUROVANÁ DATA

Pro jednu skupinu závisle proměnných

Matice závisle proměnných ***Y*** rozměru ***n*** x ***q*** matice nezávisle proměnných Z rozměru ***n*** x ***(m*** + ***q)***

a) Pro ***q*** = 1 jde o ***klasickou vícenásobnou regresi.***

b) Pro ***q =*** la ***Y*** je binární proměnná, jde o ***logistickou regresi.***

c) Pro ***q >*** 1, jde o ***vícerozměrnou lineární regresi*** (MLR).

d) Pro ortogonální sloupce matice F(čili znaky jsou nekorelované) užijeme ***standardní vícenásobnou regresi*** pro každý faktor zvlášť.

e) Při multikolinearitě (vysoké korelace mezi faktory v matici **Z**) užijeme řadu speciálních regresních metod:

Pro jednu skupinu závisle proměnných

1. ***Metoda parciálních nejmenších čtverců (PLS)*** kombinuje PCA a MLR, tj. využívá latentních vektoru k vyjádření jak závisle, tak i nezávisle proměnných.
2. ***Regrese na hlavních komponentách (PCR)*** využívá jako nezávisle proměnné jednotlivé hlavní komponenty.
3. ***Redundantní analýza (RA)*** je inverzní k PCR a určí se v ní hlavní komponenty pro matici ***Y*** příslušné skóry se pak užijí pro sérii vícenásobných regresí.
4. ***Vícenásobná analýza rozptylu (MANOVA),***
5. ***Diskriminační analýza (DA)*** provádí zařazení objektu do některé skupiny na základe znaku matice ***Z***.

Pro více skupin závisle proměnných

matice ***Y*** rozměru *n* x *q* dělena na dílčí matici $\boldsymbol{Y_{21}}$, rozměru n x q_1, na dílčí matici $\boldsymbol{Y_2}$ rozměru *n* x q_2 atd.

Kanonická korelační analýza (CCA) využívá kombinace vektoru $\boldsymbol{Y_1}$, $\boldsymbol{Y_2}$, ..., $\boldsymbol{Y_0}$ k hledání nových proměnných (kanonických proměnných), které mají nejvyšší korelace.

Analogií FA je ***vícerozměrná faktorová analýza*** (MFA), kam patří řada speciálních metod jako PARAFAC, TUCKER3, STATIS.

Prokrustova analýza (PA) je srovnání tabulek vzdáleností pro stejné objekty. V první fázi se vytvoří mapy MDS a pak se hledají transformace, které přiblíží body na obou mapách co nejblíže k sobě ve smyslu nejmenších čtverců.

Popisné charakteristiky vícerozměrných veličin
Intenzita vztahu mezi proměnnými:

Intenzita vztahu mezi proměnnými:

k charakterizaci *j*-tého znaku ξ_j čili sloupce zdrojové matice ***X*** se používá

střední hodnota $E(\xi_j) = \mu_j$ a

rozptyl $D(\xi_j) = \sigma_j^2$.

Míra intenzity vztahu mezi proměnnými ξ_i

a ξ_j , $j = I$.

***Druhý smíšený centrální moment*, kovariance**

$$cov(\xi_i, \xi_j) = E(\xi_i \xi_j) - E(\xi_i)E(\xi_j)$$

Popisné charakteristiky vícerozměrných veličin
Intenzita vztahu mezi proměnnými:

Kovariance $c(x_1, x_2)$ *mezi dvěma proměnnými* x_1 *a* x_2 je mírou jejich lineární závislosti:

a) Velká absolutní hodnota kovariance indikuje silnou lineární vazbu mezi dvěma proměnnými.

b) Malá hodnota kovariance znamená, že při změně x_1 se příliš nezmění x_2

c) Kovariance je mírou, která závisí na použitých jednotkách proměnných.

d) Limitní (maximální) hodnota kovariance je rovna odmocnině z rozptylů $s_{x_1}^2$ a $s_{x_2}^2$ tedy $c(x_1, x_2) = \sqrt{s_{x_1}^2 s_{x_1}^2}$

Popisné charakteristiky vícerozměrných veličin
Intenzita vztahu mezi proměnnými:

e) Pozitivní kovariance znamená přímou vazbu mezi x_1 a x_2, tj. při změně x_1, se změní x_2 ve stejném smyslu, růst x_1, je doprovázen růstem x_2.

f) Negativní kovariance znamená nepřímou vazbu mezi x_1 a x_2, tj. při změně x_1 se změní x_2 v opačném smyslu, růst x_1 je doprovázen poklesem x_2.

g) Nulová kovariance znamená nekorelovanost, tj. lineární nezávislost. Ještě stále však může být mezi x_1 a x_2 speciální typ nelineární závislosti.

Vlastnosti kovariance

a) Znaménko ukazuje na trend stochastické vazby mezi *j*-tým a *i*-tým sloupcem matice.

b) Je v absolutní hodnotě shora ohraničená součinem $\sigma_i \sigma_j$ tj. $\left|cov\left(\sigma_i, \sigma_j\right)\right| \leq \sigma_i \sigma_j$.

c) Je symetrickou funkcí svých argumentů.

d) Nemění se posunem počátku: pro čísla a_1, a_2, b_1, b_2 pak platí, že $cov(a_1\sigma_i + b_1, a_2\sigma_j + b_2) = a_1 a_2\, cov(\sigma_i, \sigma_j)$.

Vlastnosti kovariance

e) Pro nekorelované náhodné veličiny je $cov(\sigma_i, \sigma_j) = 0$:

1. $E(\sigma_i \sigma_j) = 0$ *a zároveň* $E(\sigma_i) = E(\sigma_j) = 0$, *což je případ centrovaných ortogonálních náhodných veličin, ne nutně nezávislých.*
2. $E(\sigma_i \sigma_j) = E(\sigma_i) = E(\sigma_j)$, *což je případ nezávislých náhodných veličin.*

f) Je *mírou intenzity lineární závislosti.*

Nevýhody kovariance:

hodnoty závisí na měřítku ξ_i a ξ_j. Velikost kovariance je omezena součinem $\sigma_i \sigma_j$.

Pearsonův párový korelační koeficient

$$\rho(\xi_i, \xi_j) = \rho_{ij} = \frac{cov(\xi_i, \xi_j)}{\sigma_i \sigma_j}$$

leží v rozmezí $-1 \leq \rho_{ij} \leq 1$:

pokud je $\rho_{ij} > 0$, jde o *pozitivně korelované* náhodné veličiny,

pokud je $\rho_{ij} <$ O, jde o *negativně korelované* náhodné veličiny.

Korelace

Korelace mezi dvěma proměnnými x_1 a x_2 je praktičtější mírou lineárního vztahu, jde o standardizovanou kovarianci a bezrozměrnou míru.

Standardizace se provádí podělením součinem směrodatných odchylek.

Nejužitečnější mírou vnitřního lineárního vztahu mezi dvěma proměnnými x_1 a x_2 je korelace, definovaná ***Personovým korelačním koeficientem r***

$$r = \sqrt{\frac{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sum_{i=1}^{n}(x_{1i} - \bar{x}_1)\sum_{i=1}^{n}(x_{2i} - \bar{x}_2)}}$$

Vlastnosti korelace:

a) $|\rho_{ij}| = 1$ ukazuje, mezi ξ_i a ξ_j existuje přesně lineární vztah.

b) Pokud jsou ξ_i a ξ_j vzájemně nekorelované, je $\rho_{ij} = 0$.

c) ξ_i a ξ_j pocházejí z vícerozměrného rozdělení a $\rho_{ij} = 0$ znamená, že proměnné jsou *vzájemně nezávislé*.

d) I pro nelineárně závislé náhodné veličiny může být $\rho_{ij} = 0$.

e) Korelační koeficient je invariantní vůči lineární transformaci ξ_i , ξ_j. Pro čísla a_1, a_2, b_1, b_2 platí vztah

$$\rho(a_1\xi_i + b_1, a_2\xi_j + b_2) = sign(a_1, a_2)\rho(\xi_i, \xi_j)$$

kde *sign(x)* je znaménková funkce, pro kterou platí

$$sign(x) = \begin{cases} -1 & pro\ x < 0 \\ 0 & pro\ x = 0 \\ 1 & pro\ x > 0 \end{cases}$$

Koeficient determinace

Koeficient determinace $D = r^2$ popisuje podíl celkového rozptylu, který lze objasnit tímto lineárním vztahem.

Korelace 0.0 značí, že mezi dvěma proměnnými není lineární vztah.

Korelace 1.0 značí, že mezi dvěma proměnnými je pozitivní lineární vztah.

Korelace -1.0 značí, že mezi dvěma proměnnými je negativní lineární vztah.

100D [%] vyjádřený v procentech je mírou k vystižení korelace, protože nezávisí na znaménku korelačního koeficientu.

Kauzalita versus korelace

Korelace je statistický pojem pro vyjádření míry lineárního vztahu a jde o čisté pojmovou míru.

Například: ročenky o demografii ukazují, že například počet narozených děti na vesnicích ve Skandinávii koreluje s počtem čápu vyskytujících se v tomto kraji s korelačním koeficientem $r \approx 0.75$. Přesto nelze přítomnost čápů v tomto kraji brát jako příčinu narozených dětí.

Ověření normality

Nejjednodušší metodou ověřování normality je test vícerozměrné šikmosti $g_{1,m}$ a vícerozměrné špičatosti $g_{2,m}$

$H_0: g_{1,m} = 0$ a $H_1: g_{2,m} = m(m+2)$.

Odhady parametrů polohy a rozptýlení:

Z vícerozměrného výběru definovaného n-ticí m-rozměrných objektů $\boldsymbol{x}_i^T = \left(x_{i,1}, x_{i,2}, \ldots, x_{i,m}\right)^T,\ i = 1, \ldots n,$ je možno stanovit *výběrový vektor středních hodnot* $\hat{\mu}$ určený vztahem

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i^T$$

- Pro odhad *kovarianční matice* S^0 platí

$$S^0 = \frac{1}{n}\sum_{i=1}^{n} (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})(\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})^T$$

Míra polohy náhodného vektoru se charakterizuje pomocí *vektoru středních hodnot* $\boldsymbol{\mu}_T = [E(\xi_1), \ldots, E(\xi_m)]$.

Míra rozptýlení pomocí *kovarianční matice* řádu *m* x *m*

$$\boldsymbol{C} = \begin{bmatrix} D(\xi_1) & cov(\xi_1,\xi_2) & \cdots & cov(\xi_1,\xi_i) & \cdots & cov(\xi_1,\xi_m) \\ cov(\xi_1,\xi_2) & D(\xi_2) & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & & & \vdots & \vdots \\ cov(\xi_1,\xi_m) & cov(\xi_2,\xi_m) & \cdots & cov(\xi_i,\xi_m) & \ldots & D(\xi_m) \end{bmatrix}$$

Místo kovarianční matice užijeme její normovanou verzi

korelační matici

$$\boldsymbol{R} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1i} & \cdots & \rho_{1m} \\ \rho_{12} & 1 & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & & & \vdots & \vdots \\ \rho_{1m} & \rho_{2m} & \cdots & \rho_{im} & \ldots & 1 \end{bmatrix}$$

má na diagonále samé jedničky a mimodiagonální prvky jsou

Perasonovy párové korelační koeficienty.

Pro vektor výběrových středních hodnot platí

$E(\hat{\boldsymbol{\mu}}) = \boldsymbol{\mu}$ a $D(\hat{\boldsymbol{\mu}}) = \frac{1}{n}\boldsymbol{C}$. Odhad $\hat{\boldsymbol{\mu}}$, je nevychýlený.

U odhadu kovarianční matice $E(\boldsymbol{S}^0) = \frac{n-1}{n}\boldsymbol{C}$ jde o vychýlený odhad. Používá se *výběrová korigovaná kovarianční matice*

$$\boldsymbol{S} = \frac{n-1}{n}\boldsymbol{S}^0$$

která je již nevychýleným odhadem kovarianční matice $\boldsymbol{C}$. Matice $\boldsymbol{S}^0$ je *výběrová kovarianční matice*.

Míry tvaru

Pokud máme dva vektory ξ_1 a ξ_2, které jsou nezávislé a rozdělené se střední hodnotou $\boldsymbol{\mu}$ a kovarianční maticí $\boldsymbol{C}$,

je *vícerozměrná šikmost* dána vztahem

$$g_{1,m} = E[(\boldsymbol{\xi}_1 - \boldsymbol{\mu})^T \boldsymbol{C}^{-1} (\boldsymbol{\xi}_2 - \boldsymbol{\mu})]^3$$

a pro *vícerozměrnou špičatost* platí

$$g_{2,m} = E[(\boldsymbol{\xi}_1 - \boldsymbol{\mu})^T \boldsymbol{C}^{-1} (\boldsymbol{\xi}_1 - \boldsymbol{\mu})]^2$$

Platí: $g_{1,m} = 0$ a $g_{2,m} = m(m+2)$.

Příklad 4.1 Popisné charakteristiky

Na úloze **B4.02** *Účinky neuroleptik při tlumení rozličných psychóz* si ukážeme odhady polohy, rozptýlení a tvaru vícerozměrné analýzy dat. K analýze znaků užijeme škálovaná data.

*Řešení:***1.**

Popisné statistiky: klasické odhady měr polohy a rozptýlení

Znak	n	$\bar{x}$	s
B402X1	20	20.10	33.90
B402X2	20	18.68	33.84
B402X3	20	3.01	5.21
B402X4	20	10.44	36.65

PŘÍKLAD 9.4 Vytvoření dendrogramu neuroleptik

Neuroleptika redukují nežádoucí účinky přebytečného dopaminu a liší se ve svých účincích: potlačují nervozitu, záchvaty, třes, ospalost, parkinsonismus, vynechávání menstruace, vyrážky, zvýšené slinění atd. Cílem je provést klasifikaci neuroleptik do shluků podobných účinků.

Data: Data ***Neuroleptika*** (převrácená hodnota mediánové účinné dávky *1/ED50* [kg/mg]):

- ***Lek*** název neuroleptika,
- ***Nervoz*** potlačení nervozity,
- ***Stereo*** potlačení stereotypního chování,
- ***Tres*** potlačení záchvatu a třesu a
- ***Usmr*** dávka smrtícího účinku.

Data

Lek	*Nervoz*	*Stereo*	*Tres*	*Usmr*
1 Chlorphromazine	3.846	3.333	1.111	1.923
2 Promazine	0.323	0.213	0.108	1.429
3 Trifluperazine	27.027	17.857	0.562	0.14
4 Fluphenazine	17.857	15.385	1.695	1.075
5 Perphenazine	27.027	27.027	1.961	2.083
6 Thioridazine	0.244	0.185	0.093	1.333
7 Pifluthixol	142.857	142.857	20.408	163.934
8 Thiothixene	4.348	4.348	0.047	0.345
9 Chorprothixene	5.882	2.941	4.545	4.167
10 Spiperone	62.5	47.619	11.765	0.847
11 Haloperidol	52.632	62.5	1.282	0.568
12 Azaperone	2.941	1.282	2.222	3.03
13 Pipamperone	0.327	0.187	1.724	0.397
14 Pimozide	20.408	20.408	0.107	0.025
15 Metitepine	15.385	10.204	10.204	27.027
16 Clozapine	0.161	0.093	0.327	0.323
17 Perlapine	0.323	0.323	0.37	0.067
18 Sulpiride	0.047	0.047	0.003	0.001
19 Butaclamol	10.204	9.091	1.471	0.025
20 Molindone	7.692	7.692	0.14	38138

Platí: je-li korelace mezi znaky malá, není třeba užít PCA a FA.

2. Kovarianční matice *C*:

Znak	*B402X1*	*B402X2*	*B402X3*	*B402X4*
B402X1	1140.90	1127.90	148.50	1044.50
B402X2	1127.90	1136.40	140.15	1051.20
B402X3	148.50	140.15	27.32	159.96
B402X4	1044.50	1051.20	159.96	1340.40

Vícerozměrná šikmost g_1: 32.14, vícerozměrná špičatost g_2: 46.708

3. Korelační matice *R*:

Znak	*B402X1*	*B402X2*	*B402X3*	*B402X4*
B402X1	1.0000	0.9905	0.8359	0.8445
B402X2	0.9905	1.0000	0.7864	0.8518
B402X3	0.8359	0.7864	1.0000	0.8238
B402X4	0.8445	0.8518	0.8238	1.0000

Korelace vystihuje míru lineární závislosti mezi dvěma znaky.

Struktury ukryté v datech:

Přirozeně se nalezne vždy nčjaká korelace mezi sloupci matice ***X***.

V řadě úloh jde o současný vliv ***několika*** rozličných znaků čili jeden znak je lineární kombinací ostatních znaků.

Pokud jde o strukturovaná data a výsledek *y* závisí na jediném znaku a kovariance $c(y, x_j)$ je dostatečné vysoká, jde o tzv. "selektivní znak".

případe vektoru vstupních veličin existuje více úrovní selektivity.

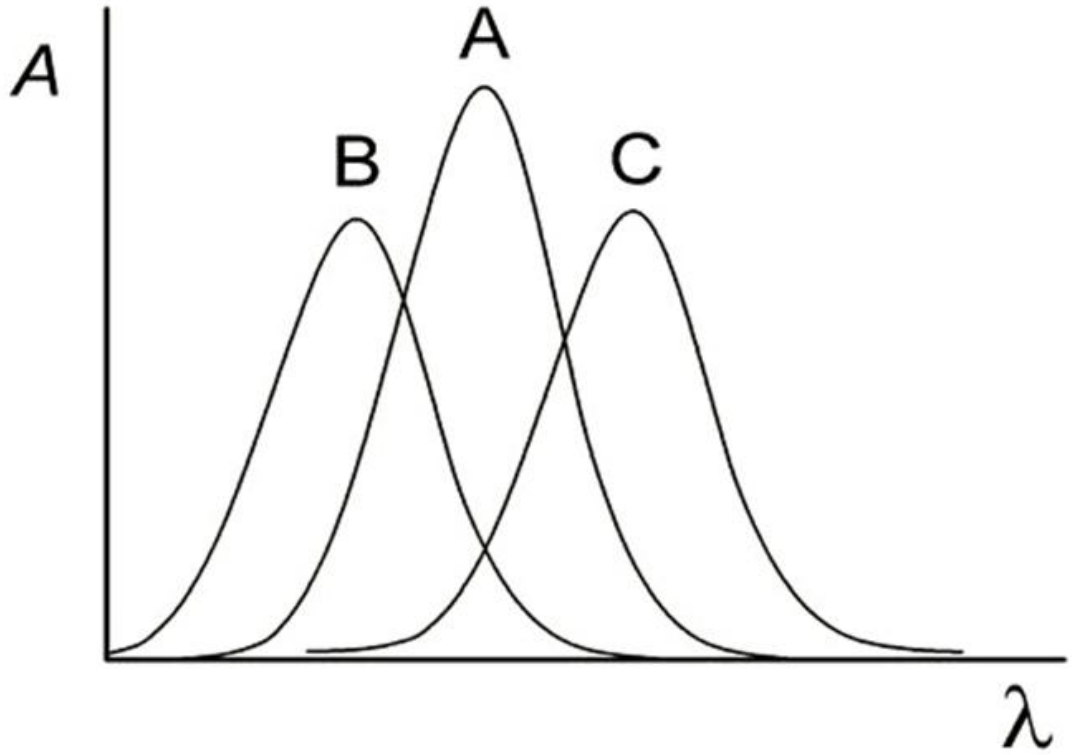
Data obsahují často znaky, které mohou být irrelevantní k výsledku y, které se pak zařazují mezi chyby.

Instrumentální sum a ostatní náhodné chyby budou vždy přítomny v datech.

Například, analýzou spekter koncentraci látky A, ve směsi s B a C.

Obr.1.1 Analýza signálu látky A při rušení signálem dvou látek, B a C

Signál látek B a C zde bude v roli šumu.



Co představuje rušivý šum?

Co je cílem stanovení, co je vytýčený model a co do modelu nepatří?

Vícerozměrná pozorování se proto modelují jako dvou součet složek: *struktura* a *šum.*

Struktura představuje část signálu, která objasňuje jak se ***X*** projeví při vysvětlování *y*, respektive ***Y.***

Šum představuje všechno ostatní, příspěvky od ostatních znaků a přístrojový šum. Šumová složka je vždy zkreslující a uživatel si ji obvykle přeje odstranit.

Vybočující body

Pro vybočující body platí **vlastnosti**

a) zkreslují odhady vektoru středních hodnot a kovarianční matice,

b) znehodnocují testy těchto parametrů,

c) ovlivňují výrazně výsledky vícerozměrných statistických metod, a

d) neumožňují tvorbu a selekci strukturních modelů.

Pro identifikaci odlehlých měření je obecně třeba:

1. definovat „čistá data",
2. určit pravděpodobnostní model dat a často i vybočujících bodů,
3. odhadnout parametry tohoto modelu.

Množina indexů $i = 1,2, ..., n$ odpovídá objektům, které rozkládá na podmnožinu potenciálně dobrých dat ***D*** a potenciálně vybočujících bodů ***V.***

Platí, že $\boldsymbol{I} = (\boldsymbol{D}, \boldsymbol{V})$.

Vybočující body

Počet potenciálně dobrých dat je n_D .

Počet potenciálně vybočujících bodů je n_V.

Podíl vybočujících bodů je pak $e = n_V/n$.

Hodnota výběrového průměru ze všech dat je pak

$$E(\bar{x}) = \boldsymbol{\mu}_0 + e\boldsymbol{\mu}$$

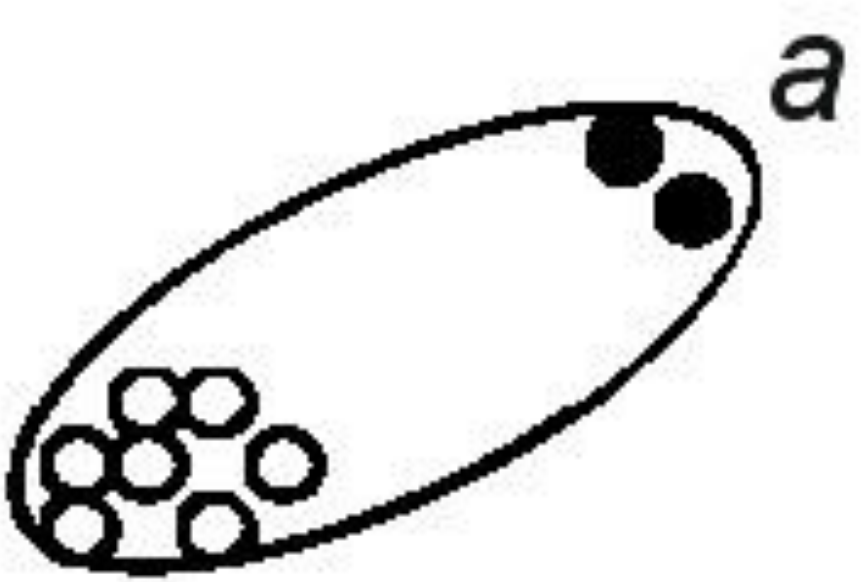
a očekávaná hodnota výběrové kovarianční matice $\boldsymbol{S}$ je

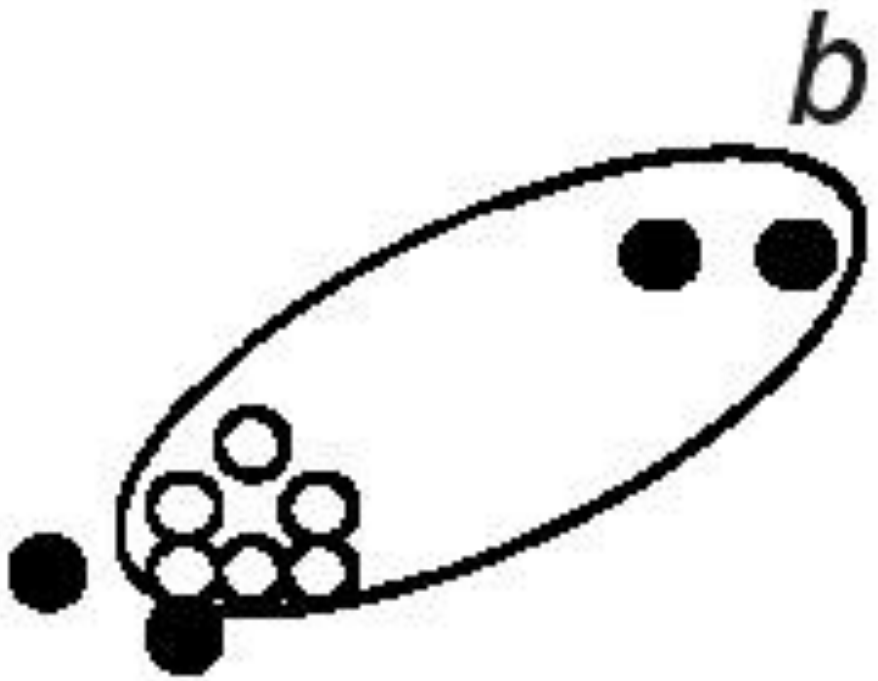
$$E(\boldsymbol{S}) = (1-e)\boldsymbol{C}_0 + e\boldsymbol{\Omega} + e(1-e)\boldsymbol{\mu}\boldsymbol{\mu}^T.$$

Výběrové průměry a kovarianční matice ze všech dat jsou závislé jak na podílu vybočujících bodů, tak i na jejich parameterech.

Běžný postup indikace vlivných bodů spočívá ve vypouštění skupin bodu (objektů), výpočtu korigovaných průměrů $\bar{\boldsymbol{x}}_k$ a kovarianční matice $\boldsymbol{S}_K$ a porovnání těchto parametrů s původními odhady $\bar{\boldsymbol{x}}$ a $\boldsymbol{S}$.

Vybočující body





K porovnání se používá Mahalanobisovy vzdálenosti

- $d_{ij} = \sqrt{(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{AD})^T [w(\boldsymbol{D}, \boldsymbol{p}) \boldsymbol{S}_{\boldsymbol{D}}]^{-1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{AD})}$
- kde $\overline{\boldsymbol{x}}_{AD}$ a $\boldsymbol{S}_{\boldsymbol{D}}$ jsou vektor aritmetických průměrů a kovarianční matice určené z potenciálně dobrých dat.
- **Korekční faktor** $w(\boldsymbol{D}, \boldsymbol{p})$ byl zaveden Hadim ve tvaru
- $w(\boldsymbol{D}, \boldsymbol{p}) = \left[1 + \frac{2}{n_D - 1 - 3m} + \frac{m+1}{n_D - m}\right]^2$
- Techniky indikace vybočujících bodů jsou citlivé na tzv. ***maskování***, kdy se vybočující body jeví jako korektní, vlivem zvětšení kovarianční matice.

Překryv

Může nastat také ***překryv***, kdy přítomnost vybočujících měření způsobí, že některá správná měření se dostanou mimo akceptovatelnou oblast, a to zkreslením kovarianční matice.

Obr. 1.2 Ukázka (a) maskování a (b) překryvu.

Vybočující body jsou na obrázku tmavé a znázornění vychází z faktu, že elipsa tvoří hraniční oblast oddělující dobrá (D) a vybočující (V) data.

PŘIKLAD 1.1 Analýza zdrojově matice dat Hrách

Zdrojová matice dat ***Hrách*** obsahuje znaky smyslového posouzení znaku odrůd hrachu.

Objekty jsou vzorky pěti odrůd hrachu A až E, sklízené v pěti rozličných obdobích 1 až 5.

Posouzení 10 porotci dvojmo, smyslové charakteristiky od 1 (nejhorší) do 9 (nejlepší), získáno 1200 řádků (objektů) tj. 60 vzorků x 2 krát opakováno 10 porotců. Původně ordinální dala se tak vlastně kardinalizovala.

Cílem je

1. průmčrovat data,
2. vynést původní data do grafu a
3. vypočítat popisné jednorozměrné statistiky.

Data

Data: matice dat *n* = 1200, *m* = 12 byla průměrována a výsledkem je matice 60 x 12 průměrných hodnot senzorického hodnocení pro znaky:

Aro je aroma, *Slad* je sladkost, *Med je* medovost, *Bez* je bezchuťovost, *Klas* je klasovost, *Tvrd* je tvrdost, *Bel* je bělost, *Bar1* je barva 1, *Bar2* je barva2, *Bar3* je barva3, *Slup* je slupka, *Ztr* je ztráta.

Objekt	*Aro*	*Slad*	*Med*	*Bez*	*Klas*	*Tvrd*	*Bel*	*Bar1*	*Bar2*	*Bar3*	*Slup*	*Ztr*
B5	6,48	6,66	4,56	2,2	2,91	3,47	4,72	5,59	5,73	5,99	4,26	3,25
...	...	...	...	...	...	...	...	...	...	...	...	...
C2	3.70	3.86	2,33	4.11	6.18	6.83	5,15	5,77	5.29	4.42	1,99	4,59

STATISTICA Cz - [Data: 11Hrach.sta (13s krát 82ř)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

Zdrojová matice dat

	1 Objekt	2 Aro	3 Slad	4 Med	5 Bez	6 Klas	7 Tvrd	8 Bel	9 Bar1	10 Bar2	11 Bar3	12 Slup	13 Ztr
1	B5	6,480	6,660	4,560	2,200	2,910	3,470	4,720	5,585	5,735	5,985	4,260	3,250
2	C4	5,750	6,090	3,810	2,320	4,030	3,770	4,170	5,730	5,745	5,325	3,820	3,380
3	B2	3,940	4,120	2,440	3,630	5,770	5,390	4,770	6,665	5,105	4,595	3,500	3,030
4	D5	6,600	6,120	4,440	1,930	3,310	4,460	4,860	5,160	5,740	6,565	2,120	3,940
5	D4	5,680	5,980	3,800	2,120	3,850	4,140	5,030	5,635	5,220	5,480	2,380	5,160
6	E2	4,740	4,660	2,880	2,940	5,650	5,770	5,310	5,940	5,270	5,890	1,750	3,640
7	B5	6,310	6,130	4,780	1,940	2,700	3,260	5,070	5,710	5,370	6,365	3,650	4,550
8	C5	6,200	6,020	4,650	1,780	3,120	3,740	5,250	5,655	5,475	5,960	2,510	3,800
9	C2	3,790	3,880	2,310	3,520	6,240	5,730	5,390	6,300	5,135	5,230	2,010	4,110
10	A4	5,680	6,340	3,750	2,790	4,170	3,870	4,520	4,920	5,760	4,570	2,970	4,410
11	D4	6,100	6,090	3,990	2,070	4,260	4,250	4,010	5,020	6,175	5,380	2,500	4,600
12	B1	3,410	3,180	1,820	4,640	6,240	7,430	4,260	4,835	5,955	4,550	1,850	4,270
13	D4	5,890	6,090	3,990	2,290	3,900	4,590	4,530	5,890	5,630	3,815	2,200	4,710
14	E4	5,770	5,320	3,880	2,260	4,220	4,990	5,050	5,335	5,590	5,540	2,160	3,650
15	B1	3,390	3,280	1,980	4,500	6,040	7,140	4,350	5,090	5,580	4,400	2,030	4,620
16	B5	6,570	6,880	4,830	1,970	2,920	3,390	3,860	4,615	6,665	6,660	2,220	3,580
17	D4	5,860	6,180	3,940	2,200	3,800	4,910	4,350	5,180	6,255	5,760	2,270	3,610
18	C2	3,960	4,480	2,300	3,940	6,230	6,410	4,470	5,105	5,720	4,970	2,050	4,530
19	A5	6,220	6,790	4,260	2,400	2,630	3,160	5,680	7,010	4,860	3,255	3,040	4,590
20	C3	5,110	5,250	3,090	3,270	5,280	5,240	5,610	6,595	5,110	3,950	2,740	4,230
21	B2	3,770	3,970	2,180	4,370	6,470	6,550	4,950	6,055	5,310	4,395	2,210	4,760
22	B5	7,090	6,090	5,180	1,740	2,570	3,180	5,230	5,920	5,515	4,115	2,090	3,100
23	D4	5,720	5,300	3,730	2,340	3,950	4,800	3,640	4,000	6,805	6,755	1,740	2,930
24	B1	3,220	3,210	1,950	4,420	6,240	7,270	4,600	6,030	5,600	4,165	1,680	3,580
25	A5	6,110	6,620	4,290	2,580	3,200	2,[illegible]	3,500	[illegible]	6,[illegible]	5,370	[illegible]	4,310
26	D4	6,070	6,270	3,980	2,190	3,890	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
27	A1	2,660	2,660	1,430	6,100	6,670	7,750	4,270	5,970	5,630	4,525	1,650	3,780
28	B3	5,260	5,490	3,460	3,030	4,850	4,170	5,220	5,410	5,415	6,120	3,080	3,950
29	C2	3,720	4,350	2,200	4,080	6,500	6,270	4,990	5,535	5,560	5,335	1,820	3,920
30	D3	5,430	5,190	3,470	2,400	4,430	5,260	4,460	4,780	5,720	5,895	1,610	3,770
31	B5	6,550	6,570	4,710	2,120	3,060	3,430	3,760	4,430	6,450	6,380	2,630	3,850
32	E3	5,530	5,410	3,680	2,470	4,720	5,780	3,880	4,335	6,470	6,790	1,800	2,950
33	C3	4,710	4,680	2,680	3,190	5,320	5,920	4,320	4,765	6,215	4,860	2,330	4,020
34	A5	6,280	7,030	4,910	2,380	2,190	2,600	4,560	5,905	5,585	4,950	3,630	3,380
35	D4	5,920	5,820	3,750	2,060	3,880	3,870	4,530	5,190	5,825	5,630	2,450	3,710
36	E4	6,090	5,720	3,800	1,940	4,440	4,450	3,940	4,625	6,510	7,175	2,180	2,990
37	B5	6,370	6,500	4,680	2,140	2,890	3,530	4,600	5,735	5,560	4,065	2,880	4,340
38	A4	5,710	5,680	3,970	2,650	4,390	3,720	5,320	6,280	5,120	6,075	2,390	2,550
39	C3	4,530	5,030	2,640	3,120	5,860	4,920	5,150	6,965	5,125	4,275	2,130	3,740
40	C4	5,950	6,280	4,040	2,190	3,930	3,610	4,120	5,395	5,815	4,505	3,090	5,340
41	D3	5,510	5,410	3,720	2,780	4,760	5,270	3,880	4,280	6,355	5,325	2,250	5,030
42	A1	3,100	3,430	1,800	4,860	6,220	7,070	4,140	5,275	5,580	3,695	2,050	4,730
43	A5	6,500	6,680	4,770	2,230	2,090	2,870	5,510	6,375	4,845	4,785	2,710	3,700
44	D3	5,460	5,410	3,270	2,970	5,150	4,980	3,610	4,305	6,600	5,435	2,370	4,440
45	A2	3,750	4,300	2,220	4,270	6,100	6,270	4,060	5,140	5,870	4,220	2,230	5,010
46	C4	5,860	5,270	3,730	2,500	3,860	4,300	4,150	4,495	6,230	6,140	2,110	3,290
47	A5	6,160	6,970	4,800	2,500	2,870	3,170	4,270	5,315	6,090	5,255	3,350	4,430
48	B2	3,870	3,880	2,230	4,060	5,990	6,310	4,450	5,530	5,785	4,980	1,930	3,620
49	B5	6,240	5,800	4,260	2,130	3,240	3,420	5,120	5,940	5,465	4,775	3,110	3,980

Pro nápovědu stiskněte F1

Ř1;S11 5,985 Filtr - Váhy: VYPN ABC 123 ZÁZN

Start Total Commander 6.0 - P... Doručená pošta - Micros... Prezentace2 STATISTICA Cz - [Dat... 10:28

Řešení

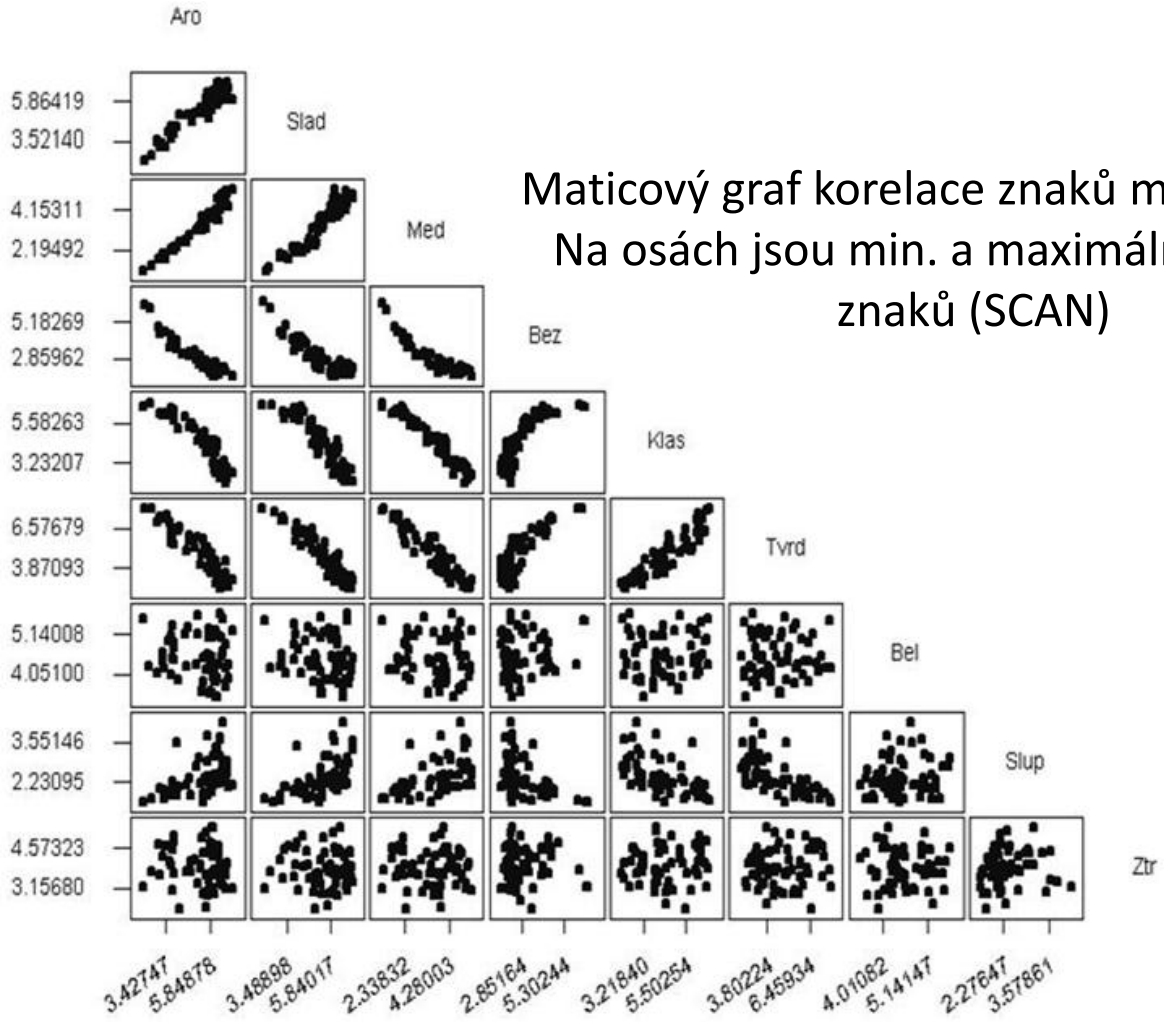
- ***Řešení:*** užity STATISTICA, SCAN, QC-Expert a MINITAB.
- **Průzkumová analýza dat: byly vypuštěny znaky týkající** se barvy hrachu *Bar1* až *Bar3* a k analýze byla použita matice rozměru 60 x 9.
- **Grafy původních dat: informaci v datech získáme z** *maticového grafu a korelačního koeficientu r* a *koeficientu determinace* $D = r^2$**. 100%.**

Matice korelačních koeficientů znaků

Tabulka 1.1 Matice korelačních koeficientů znaků.

	Aro	*Slad*	*Med*	*Bez*	*Klas*	*Tvrd*	*Bel*	*Slup*
Slad	0.951							
Med	0.975	0.949						
Bez	-0.952	-0.902	-0.904					
Klas	-0.933	-0.914	-0.969	0.836				
Tvrd	-0.924	-0.945	-0.944	0.855	0.927			
Bel	-0.128	-0.154	-0.09	0.108	0.086	0.032		
Slup	0.477	0.555	0.522	-0.417	-0.544	-0.627	0.11	
Ztr	-0.106	-0.019	-0.08	0.102	0.094	0.043	0.037	0.095

Maticový graf řekne více než matice korelačních koeficientů



Maticový graf korelace znaků matice *Hrách* . Na osách jsou min. a maximální hodnoty znaků (SCAN)

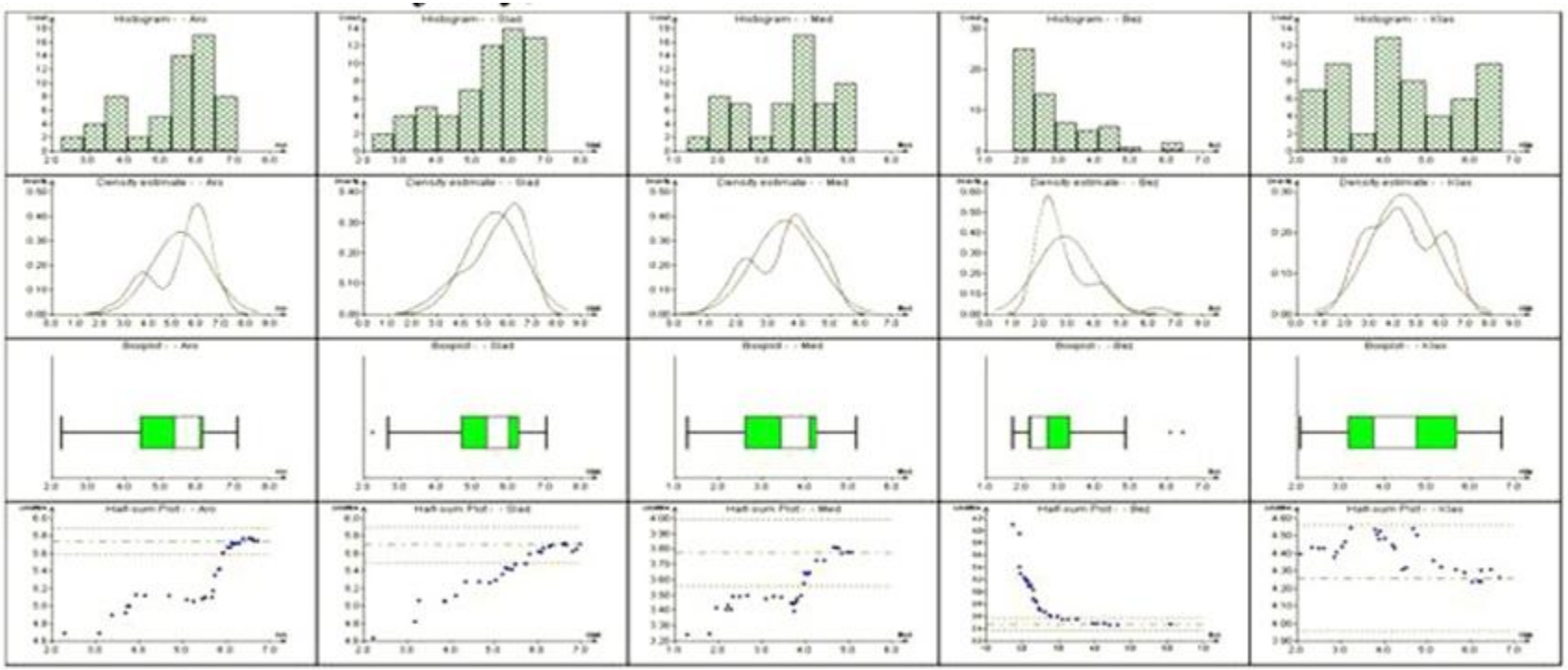
Exploratorní analýza dat

Znaky se vyšetří diagnostikami EDA:

histogram, jádrový odhad hustoty pravděpodobnosti, kvantilový graf, rankitový Q-Q graf, krabicový graf, graf polosum a symetrie, kruhový graf, atd

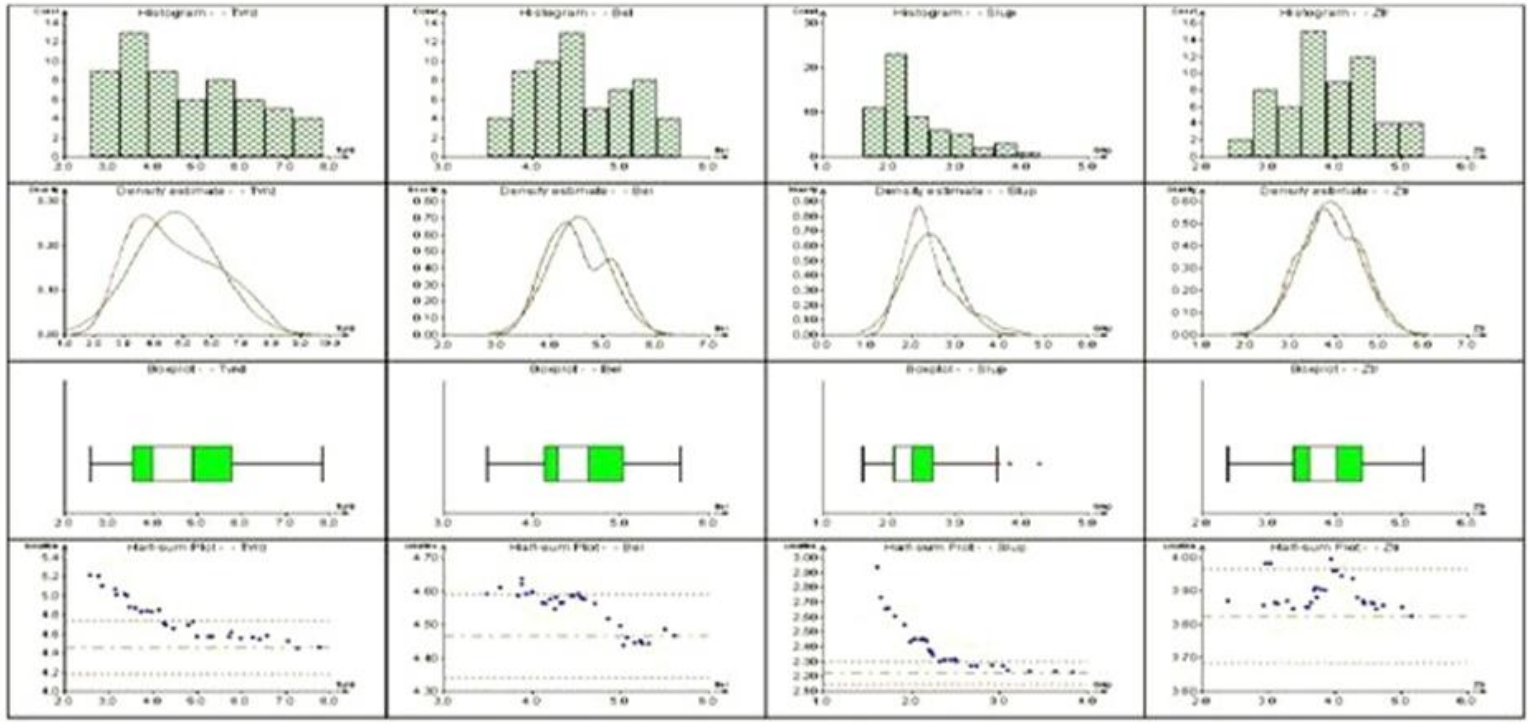
Zkoumá se vliv typu hrachu (A - E) a období sklizně (1-5) na znaky, rozdělení a odlehlé objekty,

Obr. 1.3a Histogram, graf hustoty pravděpodobnosti, krabicový graf a graf polosum pro znaky (zleva) Aro, Slad, Med, Ber a Klas, (QCEXPERT).



Obr. 1.3b Histogram, graf hustoty pravděpodobnosti, krabicový graf a graf polosum pro znaky (zleva) *Tvrd, Bel, Slup* a Ztr_9 (QCEXPERT).

EDA zde slouží především k odhalení velikosti proměnlivosti a odlehlých hodnot u všech sledovaných znaků, dále symetrie rozdělení a homogenity rozdělení.

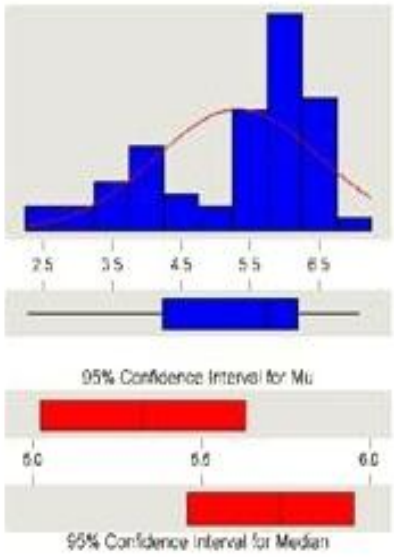


Vyšetříme, který znak dosahuje u objektů největší proměnlivosti a podle kterých znaků lze nejlépe rozlišovat mezi druhy hrachu.

- **Vyčíslení popisných statistik znaků o proměnlivosti objektů:**

který znak dosahuje u objektů hrachu největší proměnlivosti
a podle kterých znaků lze nejlépe rozlišovat mezi druhy hrachu?

Descriptive Statistics

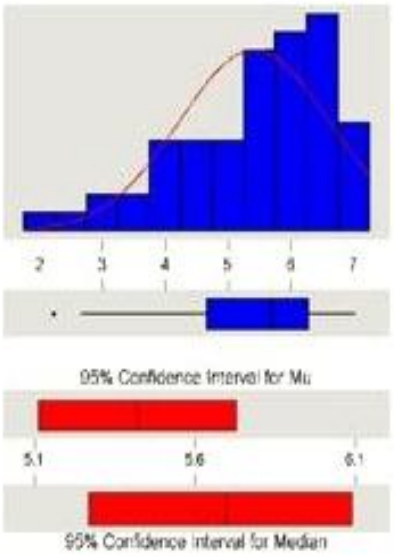


Variable: Aro

Anderson-Darling Normality Test	
A-Squared:	2.699
P-Value:	0.000
Mean	5.32617
StDev	1.18460
Variance	1.40328
Skewness	-8.6E-01
Kurtosis	-3.7E-01
N	90
Minimum	2.28000
1st Quartile	4.23750
Median	5.73500
3rd Quartile	6.19000
Maximum	7.09000
95% Confidence Interval for Mu	
5.02015	5.62218
95% Confidence Interval for Sigma	
1.00411	1.44481
95% Confidence Interval for Median	
5.45752	5.96416

Obr. 1.4.1 *Aro*, (MINITAB).

Descriptive Statistics

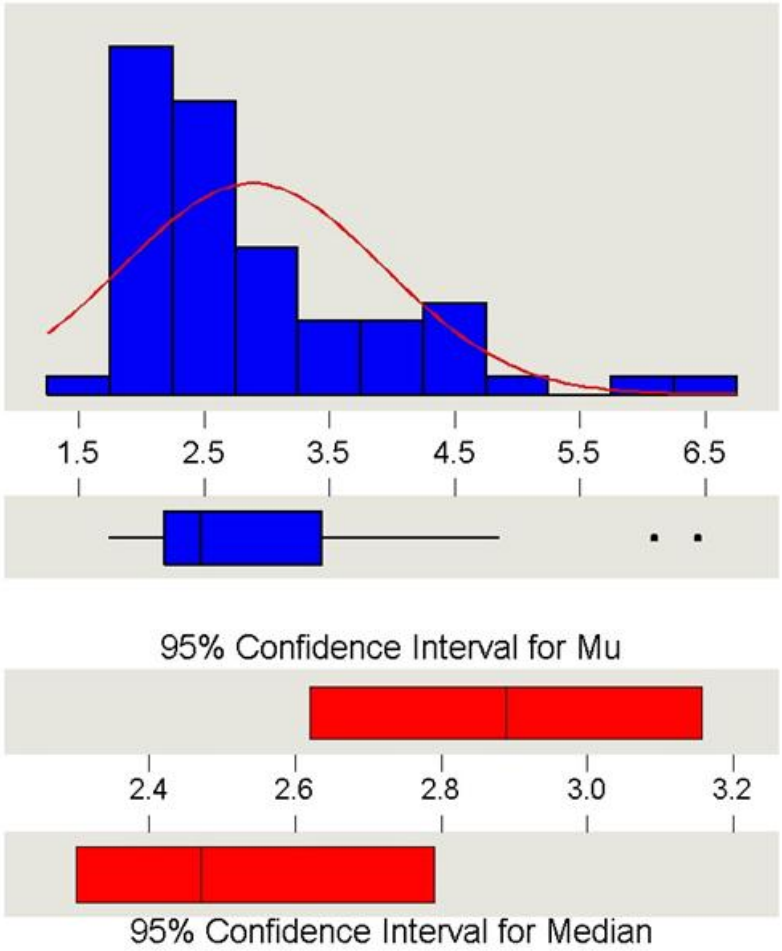


Variable: Slad

Anderson-Darling Normality Test	
A-Squared:	1.293
P-Value:	0.002
Mean	5.41933
StDev	1.19275
Variance	1.42264
Skewness	-7.9E-01
Kurtosis	-1.5E-01
N	90
Minimum	2.23000
1st Quartile	4.66500
Median	5.70000
3rd Quartile	6.28000
Maximum	7.03000
95% Confidence Interval for Mu	
5.11121	5.72745
95% Confidence Interval for Sigma	
1.01101	1.45475
95% Confidence Interval for Median	
5.26861	6.09000

Obr. 1.4.2 ... *Slad*, ...

Descriptive statistics



Variable: Bez

Anderson-Darling Normality Test

A-Squared:	3.475
P-Value:	0.000
Mean	2.88900
StDev	1.04022
Variance	1.08207
Skewness	1.52486
Kurtosis	2.16939
N	60
Minimum	1.74000
1st Quartile	2.18250
Median	2.47000
3rd Quartile	3.44250
Maximum	6.45000

95% Confidence Interval for Mu

2.62028 3.15772

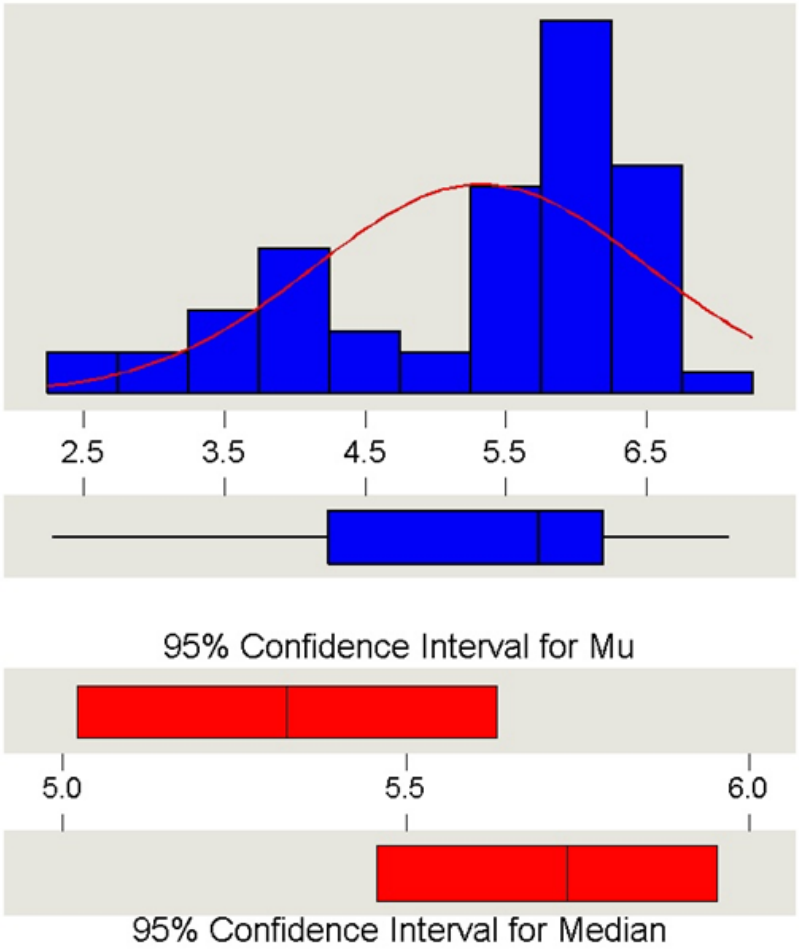
95% Confidence Interval for Sigma

0.88173 1.26872

95% Confidence Interval for Median

2.29931 2.79069

Descriptive statistics



Variable: Aro

Anderson-Darling Normality Test

A-Squared:	2.699
P-Value:	0.000

Mean	5.32617
StDev	1.18460
Variance	1.40328
Skewness	-8.6E-01
Kurtosis	-3.7E-01
N	60

Minimum	2.28000
1st Quartile	4.23750
Median	5.73500
3rd Quartile	6.19000
Maximum	7.09000

95% Confidence Interval for Mu

5.02015	5.63218

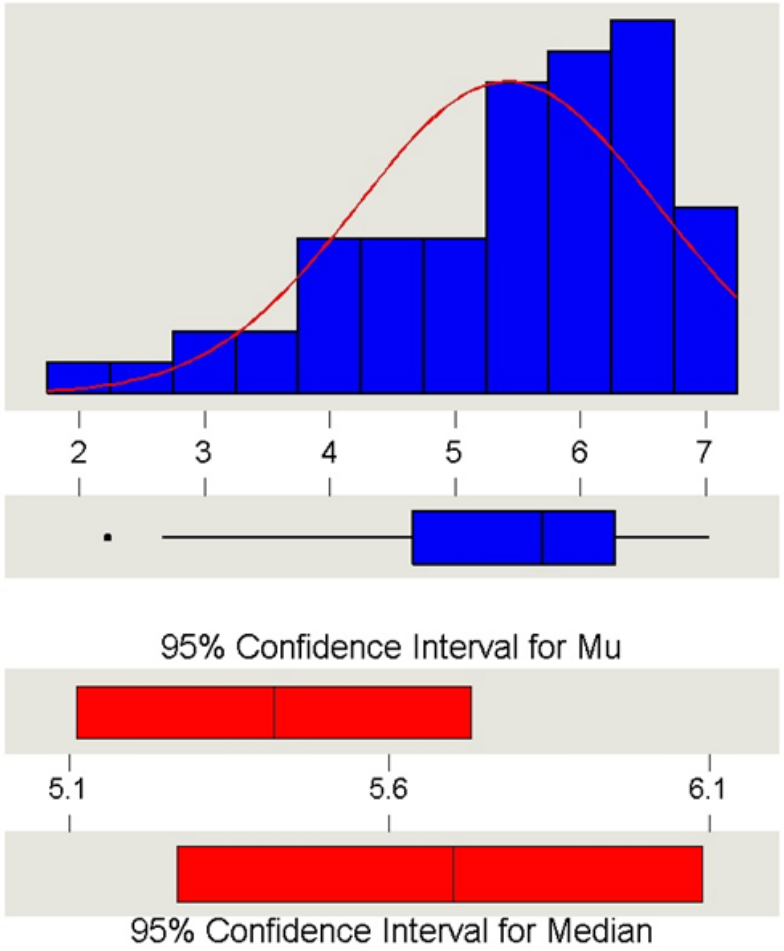
95% Confidence Interval for Sigma

1.00411	1.44481

95% Confidence Interval for Median

5.45792	5.95416

Descriptive statistics



Variable: Slad

Anderson-Darling Normality Test

A-Squared:	1.293
P-Value:	0.002

Mean	5.41933
StDev	1.19275
Variance	1.42264
Skewness	-7.9E-01
Kurtosis	-1.5E-01
N	60

Minimum	2.23000
1st Quartile	4.66500
Median	5.70000
3rd Quartile	6.28000
Maximum	7.03000

95% Confidence Interval for Mu

5.11121	5.72745

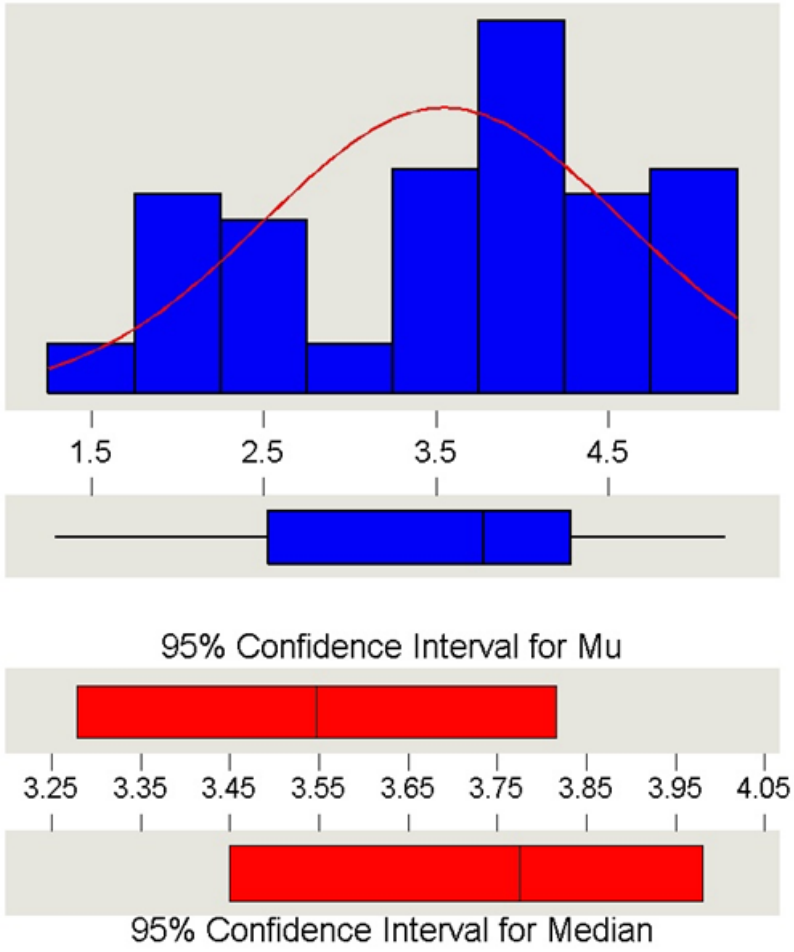
95% Confidence Interval for Sigma

1.01101	1.45475

95% Confidence Interval for Median

5.26861	6.09000

Descriptive statistics



Variable: Med

Anderson-Darling Normality Test

A-Squared:	1.180
P-Value:	0.004
Mean	3.54717
StDev	1.04217
Variance	1.08612
Skewness	-4.1E-01
Kurtosis	-8.8E-01
N	60
Minimum	1.29000
1st Quartile	2.52750
Median	3.77500
3rd Quartile	4.28250
Maximum	5.18000

95% Confidence Interval for Mu

3.27795 3.81639

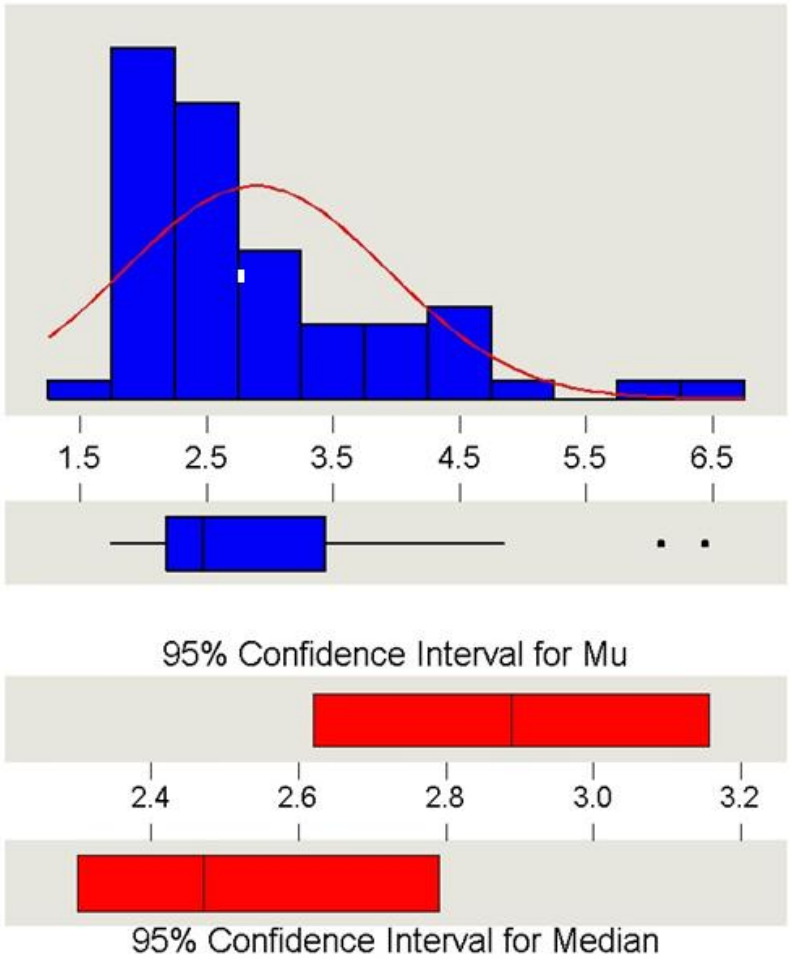
95% Confidence Interval for Sigma

0.88338 1.27110

95% Confidence Interval for Median

3.44890 3.98069

Descriptive statistics



Variable: Bez

Anderson-Darling Normality Test

A-Squared:	3.475
P-Value:	0.000

Mean	2.88900
StDev	1.04022
Variance	1.08207
Skewness	1.52486
Kurtosis	2.16939
N	60

Minimum	1.74000
1st Quartile	2.18250
Median	2.47000
3rd Quartile	3.44250
Maximum	6.45000

95% Confidence Interval for Mu

2.62028	3.15772

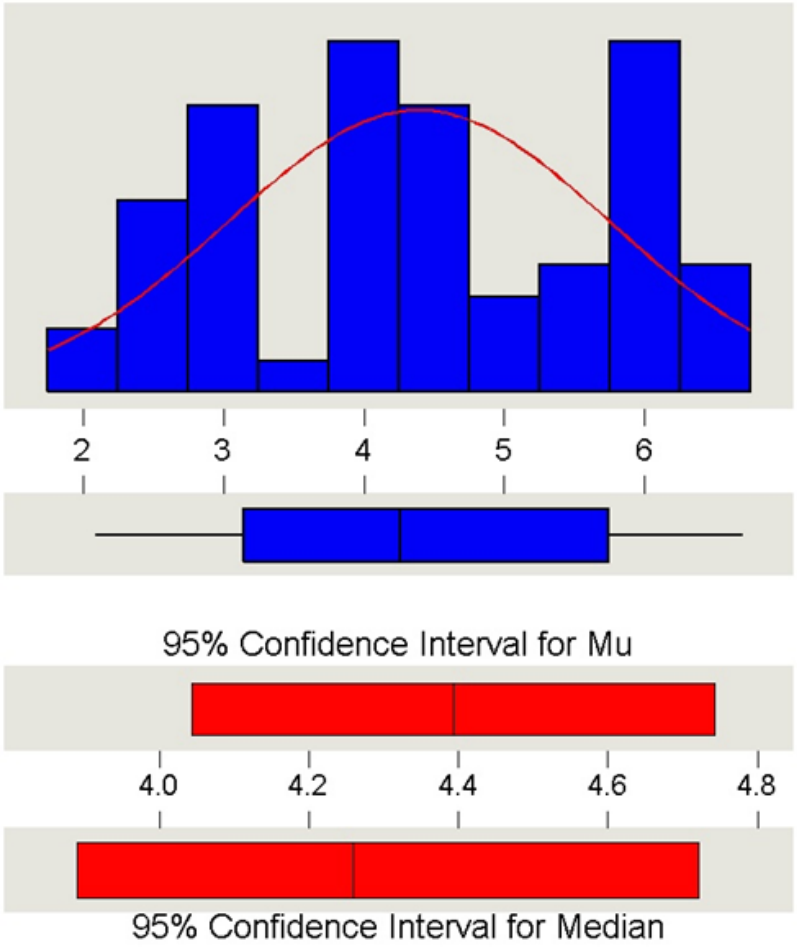
95% Confidence Interval for Sigma

0.88173	1.26872

95% Confidence Interval for Median

2.29931	2.79069

Descriptive statistics



Variable: Klas

Anderson-Darling Normality Test

A-Squared:	0.906
P-Value:	0.020

Mean	4.39333
StDev	1.35572
Variance	1.83798
Skewness	0.107576
Kurtosis	-1.16920
N	60

Minimum	2.09000
1st Quartile	3.14000
Median	4.26000
3rd Quartile	5.74000
Maximum	6.70000

95% Confidence Interval for Mu

4.04311	4.74355

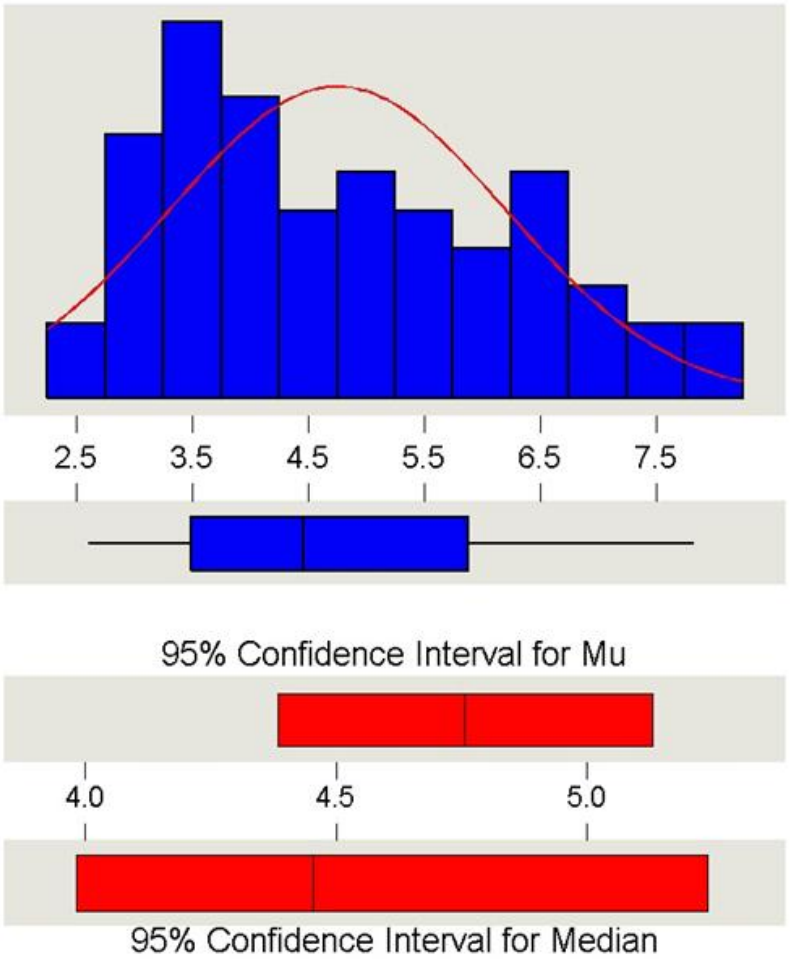
95% Confidence Interval for Sigma

1.14915	1.65352

95% Confidence Interval for Median

3.88931	4.72277

Descriptive statistics



Variable: Tvrd

Anderson-Darling Normality Test

A-Squared:	0.901
P-Value:	0.020

Mean	4.75683
StDev	1.44568
Variance	2.08999
Skewness	0.445675
Kurtosis	-8.6E-01
N	60

Minimum	2.60000
1st Quartile	3.48500
Median	4.45500
3rd Quartile	5.88500
Maximum	7.83000

95% Confidence Interval for Mu

4.38337	5.13029

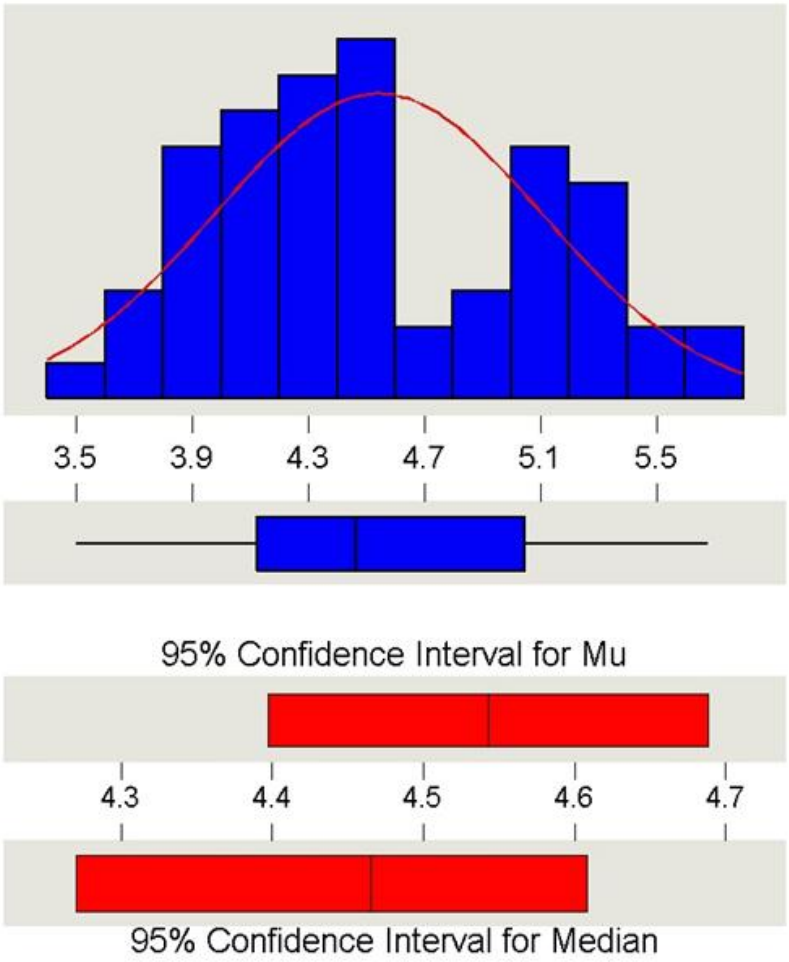
95% Confidence Interval for Sigma

1.22541	1.76324

95% Confidence Interval for Median

3.98168	5.24139

Descriptive statistics



Variable: Bel

Anderson-Darling Normality Test

A-Squared:	0.701
P-Value:	0.064
Mean	4.54283
StDev	0.56440
Variance	0.318543
Skewness	0.240026
Kurtosis	-9.3E-01
N	60
Minimum	3.50000
1st Quartile	4.12500
Median	4.46500
3rd Quartile	5.04500
Maximum	5.68000

95% Confidence Interval for Mu

4.39703 4.68863

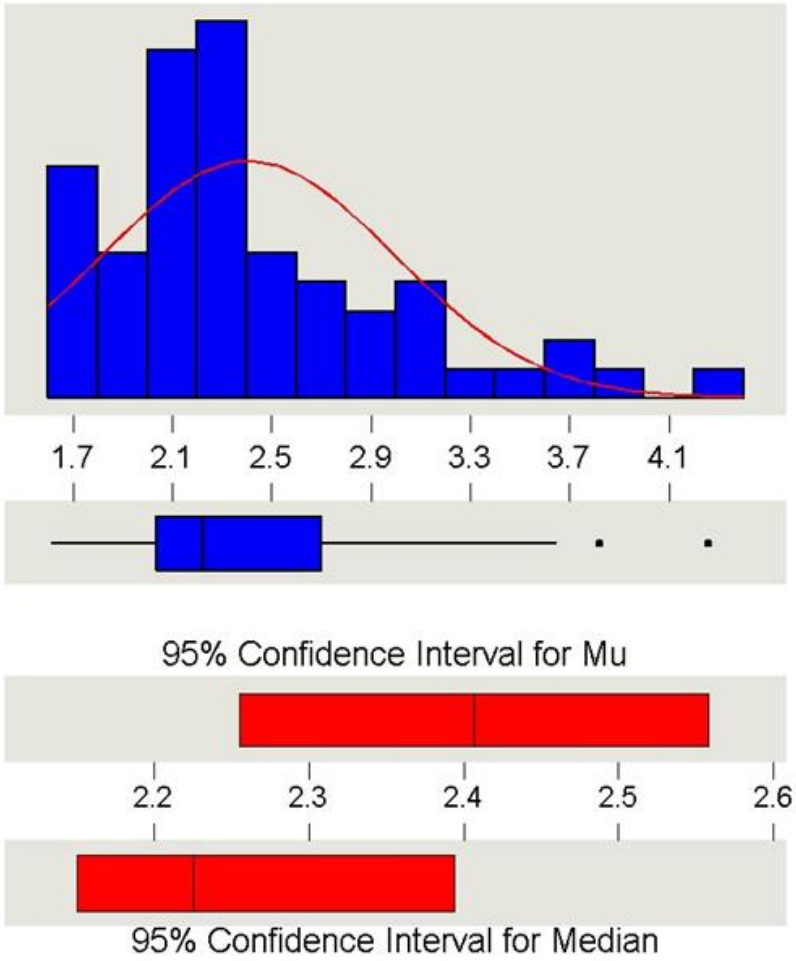
95% Confidence Interval for Sigma

0.47840 0.68837

95% Confidence Interval for Median

4.27000 4.60832

Descriptive statistics



Variable: Slup

Anderson-Darling Normality Test

A-Squared:	1.748
P-Value:	0.000

Mean	2.40683
StDev	0.58631
Variance	0.343761
Skewness	1.11795
Kurtosis	1.01405
N	60

Minimum	1.61000
1st Quartile	2.03500
Median	2.22500
3rd Quartile	2.70000
Maximum	4.26000

95% Confidence Interval for Mu

2.25537	2.55829

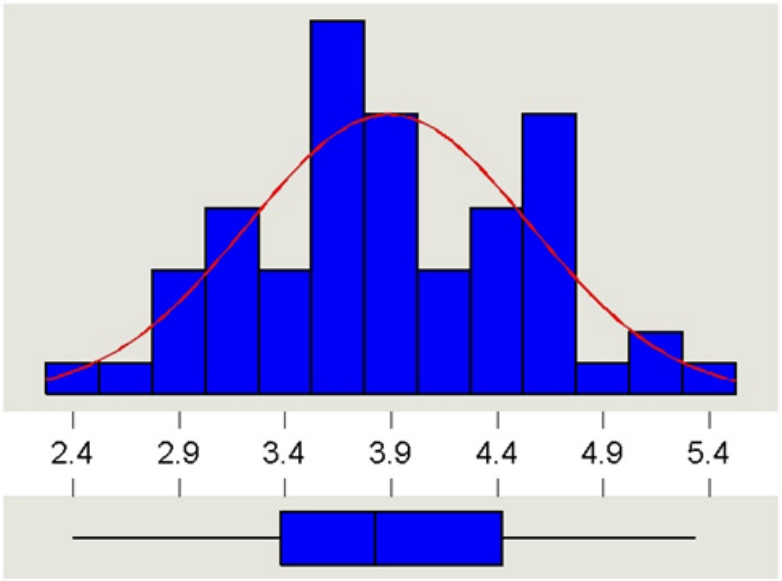
95% Confidence Interval for Sigma

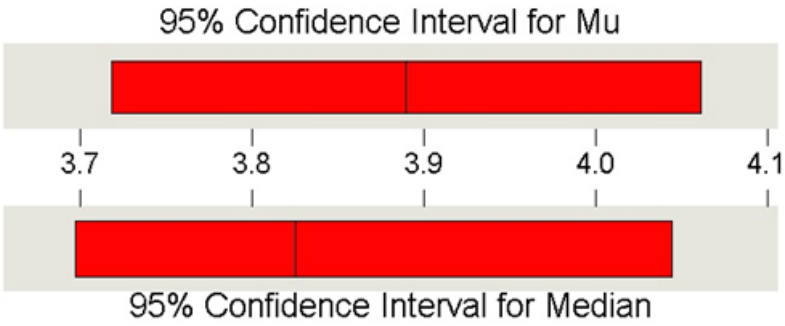
0.49698	0.71510

95% Confidence Interval for Median

2.15000	2.39416

Descriptive statistics





Variable: Ztr

Anderson-Darling Normality Test

A-Squared:	0.283
P-Value:	0.623
Mean	3.88967
StDev	0.66535
Variance	0.442695
Skewness	4.97E-03
Kurtosis	-5.3E-01
N	60
Minimum	2.40000
1st Quartile	3.38000
Median	3.82500
3rd Quartile	4.42750
Maximum	5.34000

95% Confidence Interval for Mu

3.71779 4.06155

95% Confidence Interval for Sigma

0.56398 0.81151

95% Confidence Interval for Median

3.69653 4.04485

STATISTICA Cz - [Data: 11Hrach.sta (13s krát 82ř)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

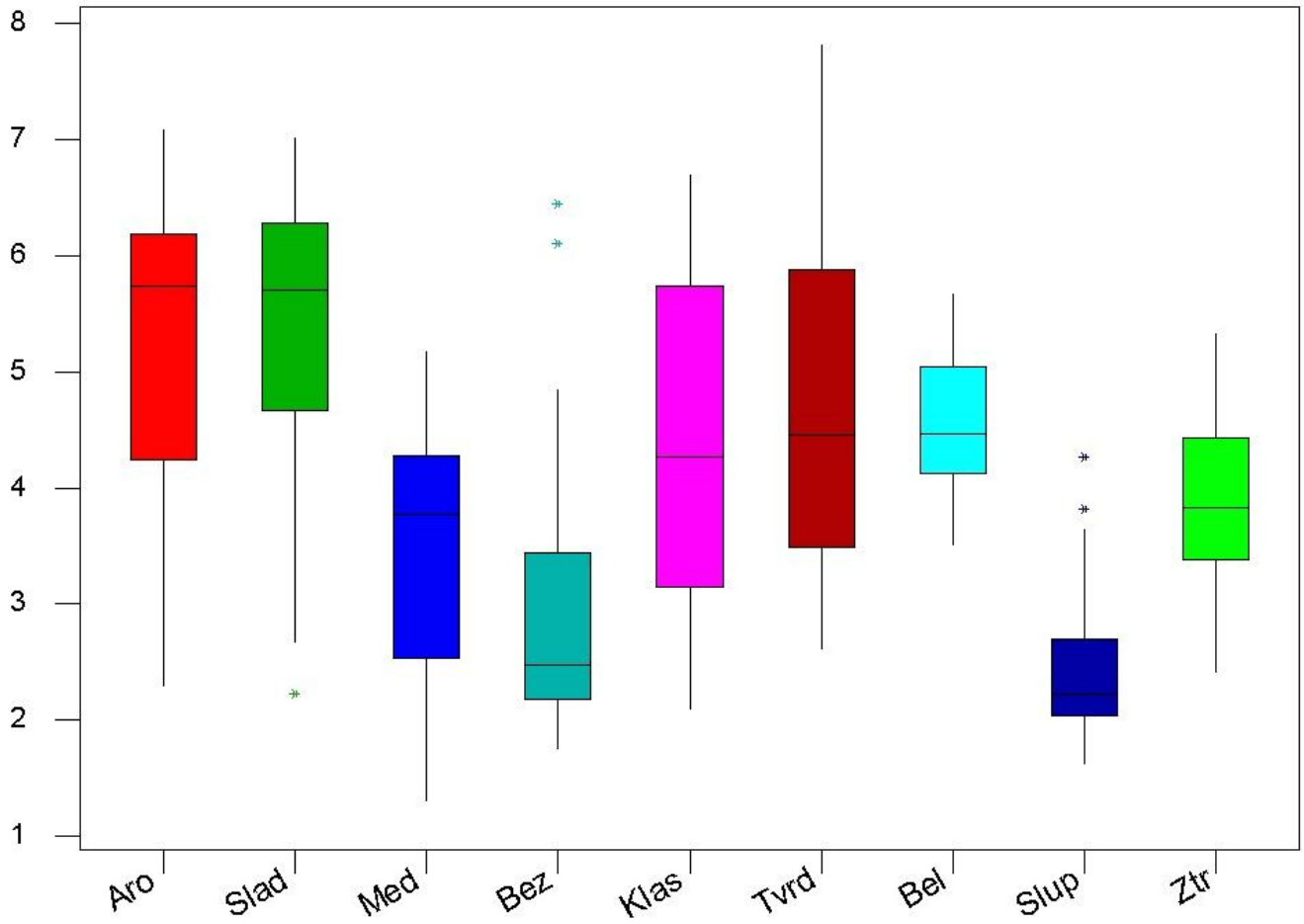
Zdrojová matice dat

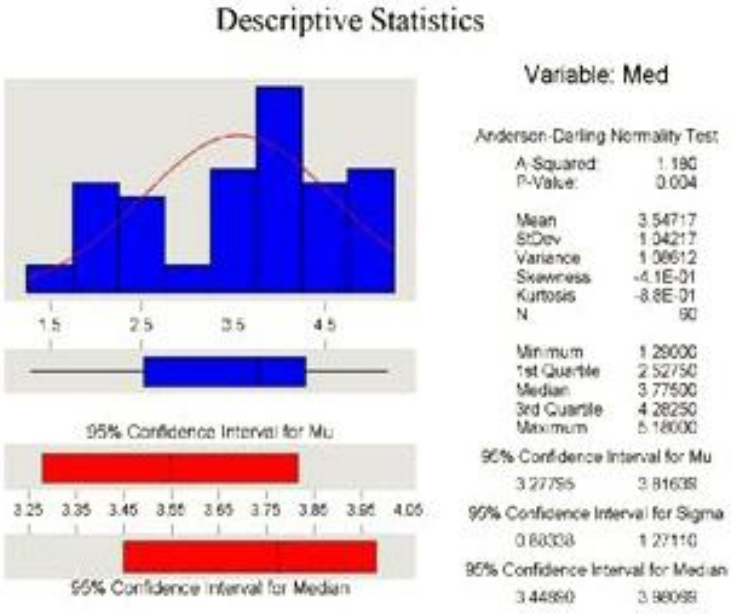
	1 Objekt	2 Aro	3 Slad	4 Med	5 Bez	6 Klas	7 Tvrd	8 Bel	9 Bar1	10 Bar2	11 Bar3	12 Slup	13 Ztr
1	B5	6,480	6,660	4,560	2,200	2,910	3,470	4,720	5,585	5,735	5,985	4,260	3,250
2	C4	5,750	6,090	3,810	2,320	4,030	3,770	4,170	5,730	5,745	5,325	3,820	3,380
3	B2	3,940	4,120	2,440	3,630	5,770	5,390	4,770	6,665	5,105	4,595	3,500	3,030
4	D5	6,600	6,120	4,440	1,930	3,310	4,460	4,860	5,160	5,740	6,565	2,120	3,940
5	D4	5,680	5,980	3,800	2,120	3,850	4,140	5,030	5,635	5,220	5,480	2,380	5,160
6	E2	4,740	4,660	2,880	2,940	5,650	5,770	5,310	5,940	5,270	5,890	1,750	3,640
7	B5	6,310	6,130	4,780	1,940	2,700	3,260	5,070	5,710	5,370	6,365	3,650	4,550
8	C5	6,200	6,020	4,650	1,780	3,120	3,740	5,250	5,655	5,475	5,960	2,510	3,800
9	C2	3,790	3,880	2,310	3,520	6,240	5,730	5,390	6,300	5,135	5,230	2,010	4,110
10	A4	5,680	6,340	3,750	2,790	4,170	3,870	4,520	4,920	5,760	4,570	2,970	4,410
11	D4	6,100	6,090	3,990	2,070	4,260	4,250	4,010	5,020	6,175	5,380	2,500	4,600
12	B1	3,410	3,180	1,820	4,640	6,240	7,430	4,260	4,835	5,955	4,550	1,850	4,270
13	D4	5,890	6,090	3,990	2,290	3,900	4,590	4,530	5,890	5,630	3,815	2,200	4,710
14	E4	5,770	5,320	3,880	2,260	4,220	4,990	5,050	5,335	5,590	5,540	2,160	3,650
15	B1	3,390	3,280	1,980	4,500	6,040	7,140	4,350	5,090	5,580	4,400	2,030	4,620
16	B5	6,570	6,880	4,830	1,970	2,920	3,390	3,860	4,615	6,665	6,660	2,220	3,580
17	D4	5,860	6,180	3,940	2,200	3,800	4,910	4,350	5,180	6,255	5,760	2,270	3,610
18	C2	3,960	4,480	2,300	3,940	6,230	6,410	4,470	5,105	5,720	4,970	2,050	4,530
19	A5	6,220	6,790	4,260	2,400	2,630	3,160	5,680	7,010	4,860	3,255	3,040	4,590
20	C3	5,110	5,250	3,090	3,270	5,280	5,240	5,610	6,595	5,110	3,950	2,740	4,230
21	B2	3,770	3,970	2,180	4,370	6,470	6,550	4,950	6,055	5,310	4,395	2,210	4,760
22	B5	7,090	6,090	5,180	1,740	2,570	3,180	5,230	5,920	5,515	4,115	2,090	3,100
23	D4	5,720	5,300	3,730	2,340	3,950	4,800	3,640	4,000	6,805	6,755	1,740	2,930
24	B1	3,220	3,210	1,950	4,420	6,240	7,270	4,600	6,030	5,600	4,165	1,680	3,580
25	A5	6,110	6,620	4,290	2,580	3,200	2,[illegible]	3,500	[illegible]	6,[illegible]	5,370	[illegible]	4,310
26	D4	6,070	6,270	3,980	2,190	3,890	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
27	A1	2,660	2,660	1,430	6,100	6,670	7,750	4,270	5,970	5,630	4,525	1,650	3,780
28	B3	5,260	5,490	3,460	3,030	4,850	4,170	5,220	5,410	5,415	6,120	3,080	3,950
29	C2	3,720	4,350	2,200	4,080	6,500	6,270	4,990	5,535	5,560	5,335	1,820	3,920
30	D3	5,430	5,190	3,470	2,400	4,430	5,260	4,460	4,780	5,720	5,895	1,610	3,770
31	B5	6,550	6,570	4,710	2,120	3,060	3,430	3,760	4,430	6,450	6,380	2,630	3,850
32	E3	5,530	5,410	3,680	2,470	4,720	5,780	3,880	4,335	6,470	6,790	1,800	2,950
33	C3	4,710	4,680	2,680	3,190	5,320	5,920	4,320	4,765	6,215	4,860	2,330	4,020
34	A5	6,280	7,030	4,910	2,380	2,190	2,600	4,560	5,905	5,585	4,950	3,630	3,380
35	D4	5,920	5,820	3,750	2,060	3,880	3,870	4,530	5,190	5,825	5,630	2,450	3,710
36	E4	6,090	5,720	3,800	1,940	4,440	4,450	3,940	4,625	6,510	7,175	2,180	2,990
37	B5	6,370	6,500	4,680	2,140	2,890	3,530	4,600	5,735	5,560	4,065	2,880	4,340
38	A4	5,710	5,680	3,970	2,650	4,390	3,720	5,320	6,280	5,120	6,075	2,390	2,550
39	C3	4,530	5,030	2,640	3,120	5,860	4,920	5,150	6,965	5,125	4,275	2,130	3,740
40	C4	5,950	6,280	4,040	2,190	3,930	3,610	4,120	5,395	5,815	4,505	3,090	5,340
41	D3	5,510	5,410	3,720	2,780	4,760	5,270	3,880	4,280	6,355	5,325	2,250	5,030
42	A1	3,100	3,430	1,800	4,860	6,220	7,070	4,140	5,275	5,580	3,695	2,050	4,730
43	A5	6,500	6,680	4,770	2,230	2,090	2,870	5,510	6,375	4,845	4,785	2,710	3,700
44	D3	5,460	5,410	3,270	2,970	5,150	4,980	3,610	4,305	6,600	5,435	2,370	4,440
45	A2	3,750	4,300	2,220	4,270	6,100	6,270	4,060	5,140	5,870	4,220	2,230	5,010
46	C4	5,860	5,270	3,730	2,500	3,860	4,300	4,150	4,495	6,230	6,140	2,110	3,290
47	A5	6,160	6,970	4,800	2,500	2,870	3,170	4,270	5,315	6,090	5,255	3,350	4,430
48	B2	3,870	3,880	2,230	4,060	5,990	6,310	4,450	5,530	5,785	4,980	1,930	3,620
49	B5	6,240	5,800	4,260	2,130	3,240	3,420	5,120	5,940	5,465	4,775	3,110	3,980

Pro nápovědu stiskněte F1

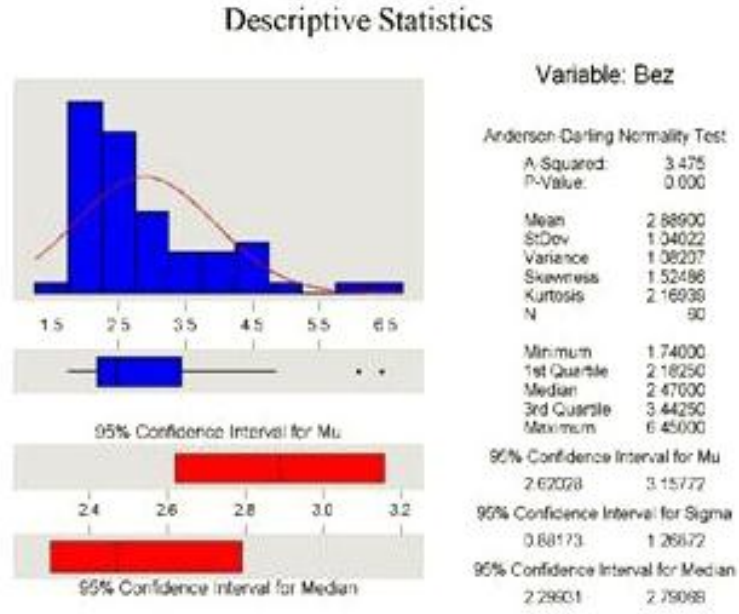
Ř1;S11 5,985 Filtr - Váhy: VYPN ABC 123 ZÁZN

Start Total Commander 6.0 - P... Doručená pošta - Micros... Prezentace2 STATISTICA Cz - [Dat... 10:28

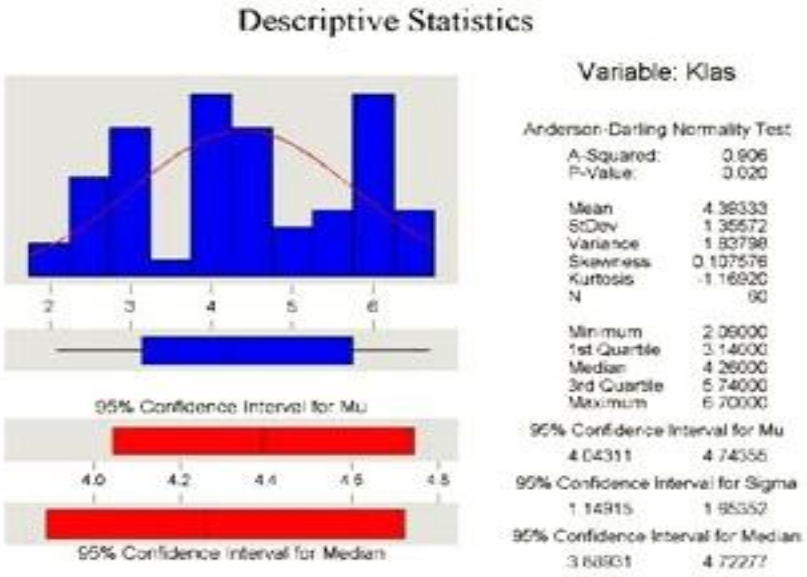

8
7
6
5
4
3
2
1
Aro
Slad
Med
Bez
Klas
Tvrd
Bel
Slup
Ztr



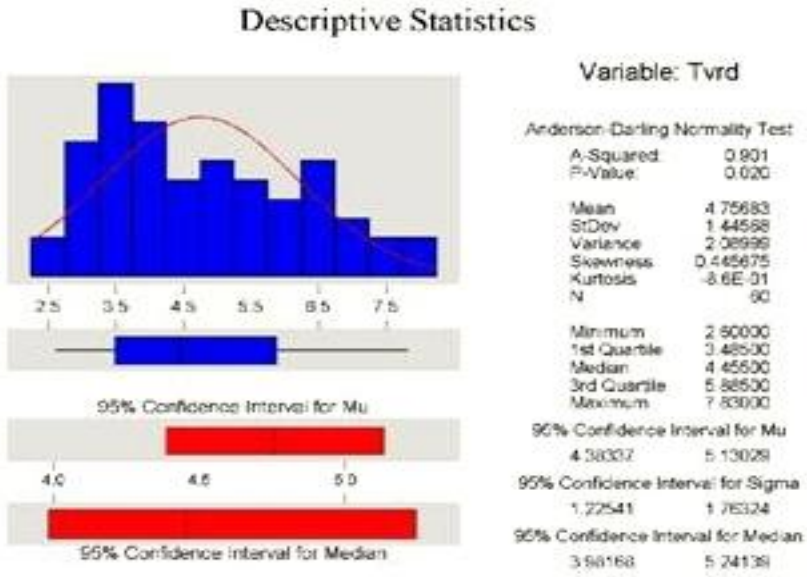
Obr. 1.4.3 ... *Med*



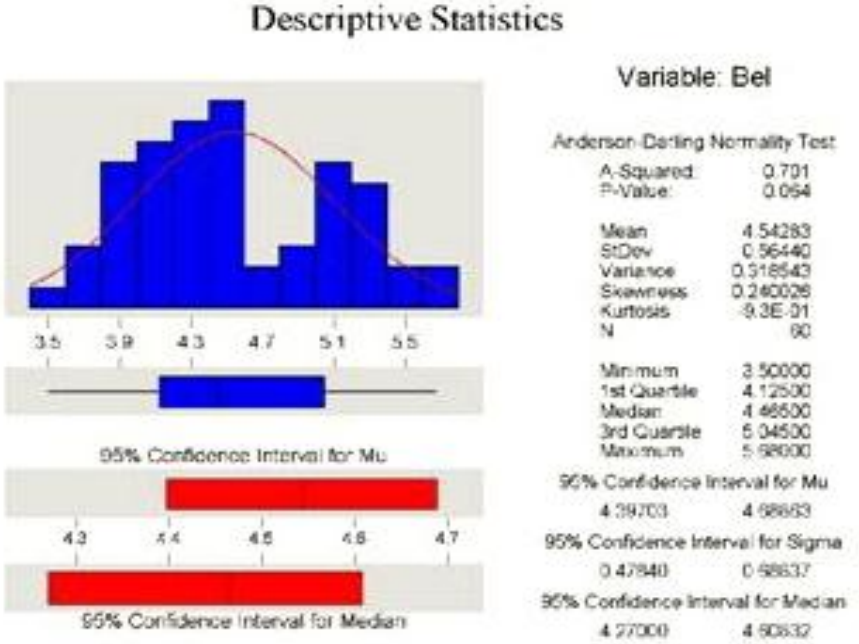
Obr. 1.4.4 ... *Bez*



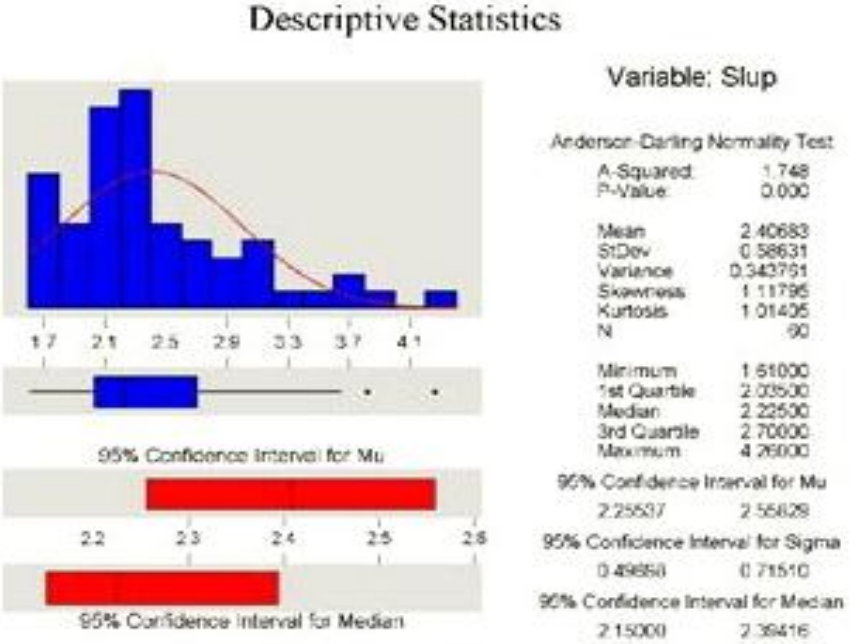
Obr. 1.4.5 ...*Klas*



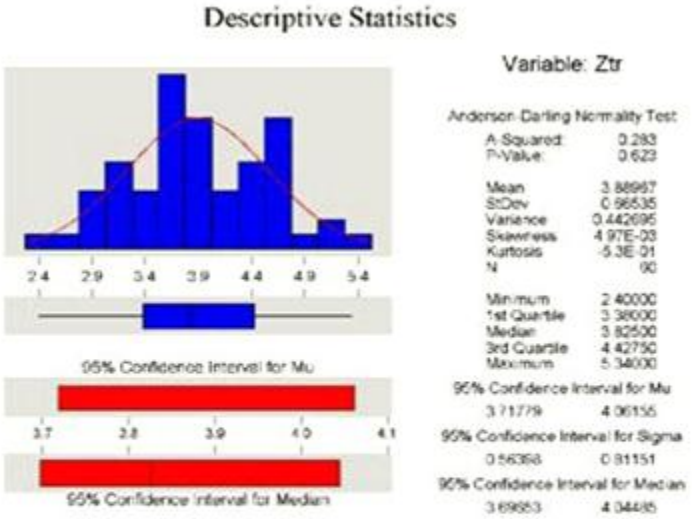
Obr. 1.4.6 *Tvrd*



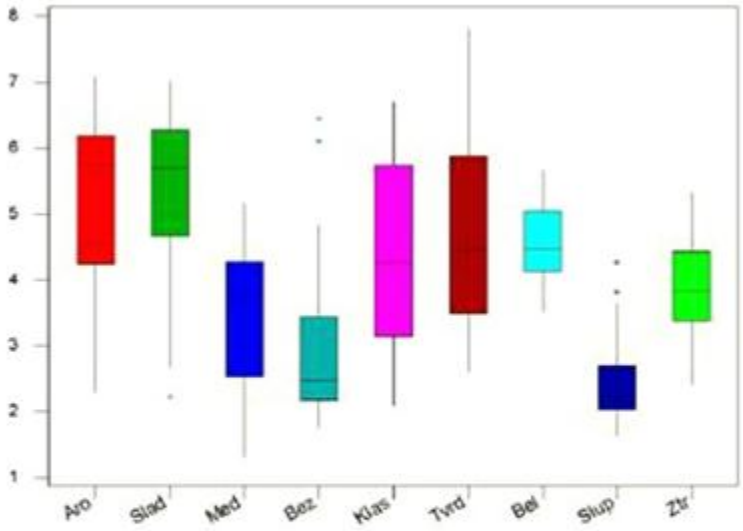
Obr. 1.4.7 ...*Bel*



Obr. 1.4.8 ... *Slup*



Obr. 1.4.9 *Ztr*,



Obr. 1.4.10 Krabicové grafy znaků vystihují proměnlivost znaku v objektech

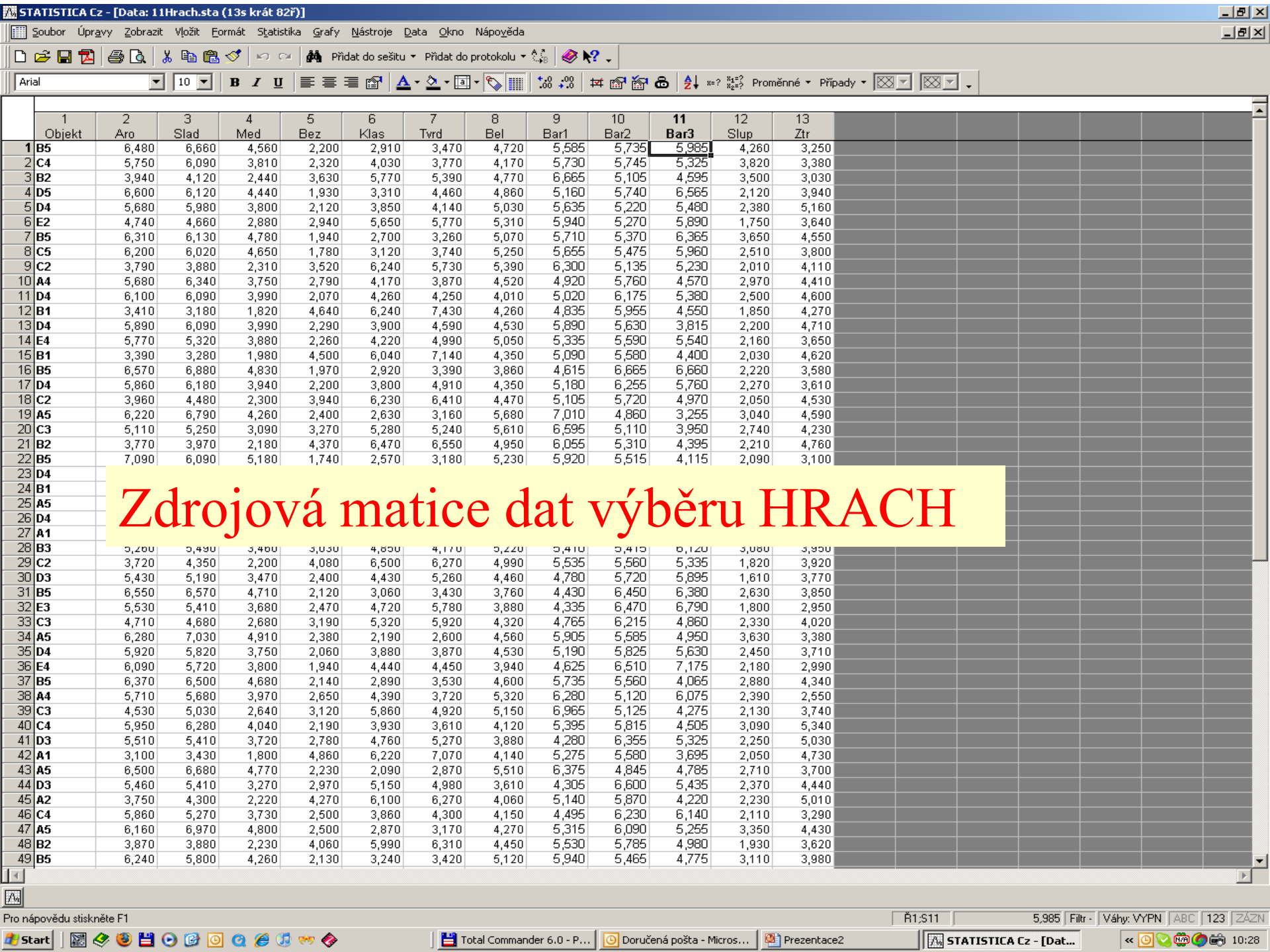
Závěr

Grafy zobrazují proměnlivost znaků při zobrazení všech objektů matice dat ***Hrách.***

Nejmenší proměnlivost mají poslední dva znaky.

Závěr: byl ukázán první pokus o zjištění proměnlivosti v datech, číselně i graficky.

Popisné statistiky ukázaly, že prvních šest znaků vykazuje největší rozptyl a lze je s výhodou využít ve vícerozměrné analýze dat.



	1 Objekt	2 Aro	3 Slad	4 Med	5 Bez	6 Klas	7 Tvrd	8 Bel	9 Bar1	10 Bar2	11 Bar3	12 Slup	13 Ztr
1	B5	6,480	6,660	4,560	2,200	2,910	3,470	4,720	5,585	5,735	5,985	4,260	3,250
2	C4	5,750	6,090	3,810	2,320	4,030	3,770	4,170	5,730	5,745	5,325	3,820	3,380
3	B2	3,940	4,120	2,440	3,630	5,770	5,390	4,770	6,665	5,105	4,595	3,500	3,030
4	D5	6,600	6,120	4,440	1,930	3,310	4,460	4,860	5,160	5,740	6,565	2,120	3,940
5	D4	5,680	5,980	3,800	2,120	3,850	4,140	5,030	5,635	5,220	5,480	2,380	5,160
6	E2	4,740	4,660	2,880	2,940	5,650	5,770	5,310	5,940	5,270	5,890	1,750	3,640
7	B5	6,310	6,130	4,780	1,940	2,700	3,260	5,070	5,710	5,370	6,365	3,650	4,550
8	C5	6,200	6,020	4,650	1,780	3,120	3,740	5,250	5,655	5,475	5,960	2,510	3,800
9	C2	3,790	3,880	2,310	3,520	6,240	5,730	5,390	6,300	5,135	5,230	2,010	4,110
10	A4	5,680	6,340	3,750	2,790	4,170	3,870	4,520	4,920	5,760	4,570	2,970	4,410
11	D4	6,100	6,090	3,990	2,070	4,260	4,250	4,010	5,020	6,175	5,380	2,500	4,600
12	B1	3,410	3,180	1,820	4,640	6,240	7,430	4,260	4,835	5,955	4,550	1,850	4,270
13	D4	5,890	6,090	3,990	2,290	3,900	4,590	4,530	5,890	5,630	3,815	2,200	4,710
14	E4	5,770	5,320	3,880	2,260	4,220	4,990	5,050	5,335	5,590	5,540	2,160	3,650
15	B1	3,390	3,280	1,980	4,500	6,040	7,140	4,350	5,090	5,580	4,400	2,030	4,620
16	B5	6,570	6,880	4,830	1,970	2,920	3,390	3,860	4,615	6,665	6,660	2,220	3,580
17	D4	5,860	6,180	3,940	2,200	3,800	4,910	4,350	5,180	6,255	5,760	2,270	3,610
18	C2	3,960	4,480	2,300	3,940	6,230	6,410	4,470	5,105	5,720	4,970	2,050	4,530
19	A5	6,220	6,790	4,260	2,400	2,630	3,160	5,680	7,010	4,860	3,255	3,040	4,590
20	C3	5,110	5,250	3,090	3,270	5,280	5,240	5,610	6,595	5,110	3,950	2,740	4,230
21	B2	3,770	3,970	2,180	4,370	6,470	6,550	4,950	6,055	5,310	4,395	2,210	4,760
22	B5	7,090	6,090	5,180	1,740	2,570	3,180	5,230	5,920	5,515	4,115	2,090	3,100
23	D4												
24	B1												
25	A5												
26	D4												
27	A1												
28	B3	5,260	5,490	3,460	3,030	4,850	4,170	5,220	5,410	5,415	6,120	3,080	3,950
29	C2	3,720	4,350	2,200	4,080	6,500	6,270	4,990	5,535	5,560	5,335	1,820	3,920
30	D3	5,430	5,190	3,470	2,400	4,430	5,260	4,460	4,780	5,720	5,895	1,610	3,770
31	B5	6,550	6,570	4,710	2,120	3,060	3,430	3,760	4,430	6,450	6,380	2,630	3,850
32	E3	5,530	5,410	3,680	2,470	4,720	5,780	3,880	4,335	6,470	6,790	1,800	2,950
33	C3	4,710	4,680	2,680	3,190	5,320	5,920	4,320	4,765	6,215	4,860	2,330	4,020
34	A5	6,280	7,030	4,910	2,380	2,190	2,600	4,560	5,905	5,585	4,950	3,630	3,380
35	D4	5,920	5,820	3,750	2,060	3,880	3,870	4,530	5,190	5,825	5,630	2,450	3,710
36	E4	6,090	5,720	3,800	1,940	4,440	4,450	3,940	4,625	6,510	7,175	2,180	2,990
37	B5	6,370	6,500	4,680	2,140	2,890	3,530	4,600	5,735	5,560	4,065	2,880	4,340
38	A4	5,710	5,680	3,970	2,650	4,390	3,720	5,320	6,280	5,120	6,075	2,390	2,550
39	C3	4,530	5,030	2,640	3,120	5,860	4,920	5,150	6,965	5,125	4,275	2,130	3,740
40	C4	5,950	6,280	4,040	2,190	3,930	3,610	4,120	5,395	5,815	4,505	3,090	5,340
41	D3	5,510	5,410	3,720	2,780	4,760	5,270	3,880	4,280	6,355	5,325	2,250	5,030
42	A1	3,100	3,430	1,800	4,860	6,220	7,070	4,140	5,275	5,580	3,695	2,050	4,730
43	A5	6,500	6,680	4,770	2,230	2,090	2,870	5,510	6,375	4,845	4,785	2,710	3,700
44	D3	5,460	5,410	3,270	2,970	5,150	4,980	3,610	4,305	6,600	5,435	2,370	4,440
45	A2	3,750	4,300	2,220	4,270	6,100	6,270	4,060	5,140	5,870	4,220	2,230	5,010
46	C4	5,860	5,270	3,730	2,500	3,860	4,300	4,150	4,495	6,230	6,140	2,110	3,290
47	A5	6,160	6,970	4,800	2,500	2,870	3,170	4,270	5,315	6,090	5,255	3,350	4,430
48	B2	3,870	3,880	2,230	4,060	5,990	6,310	4,450	5,530	5,785	4,980	1,930	3,620
49	B5	6,240	5,800	4,260	2,130	3,240	3,420	5,120	5,940	5,465	4,775	3,110	3,980

Zdrojová matice dat výběru HRACH

STATISTICA Cz - [Data: 11Hrach (13s krát 82ř)]

Soubor Upravit Zobrazit Vložit Formát Statistiky Data mining Grafy Nástroje Data Okno Nápověda

Přidat do sešitu Přidat do protokolu Přidat do MS Word

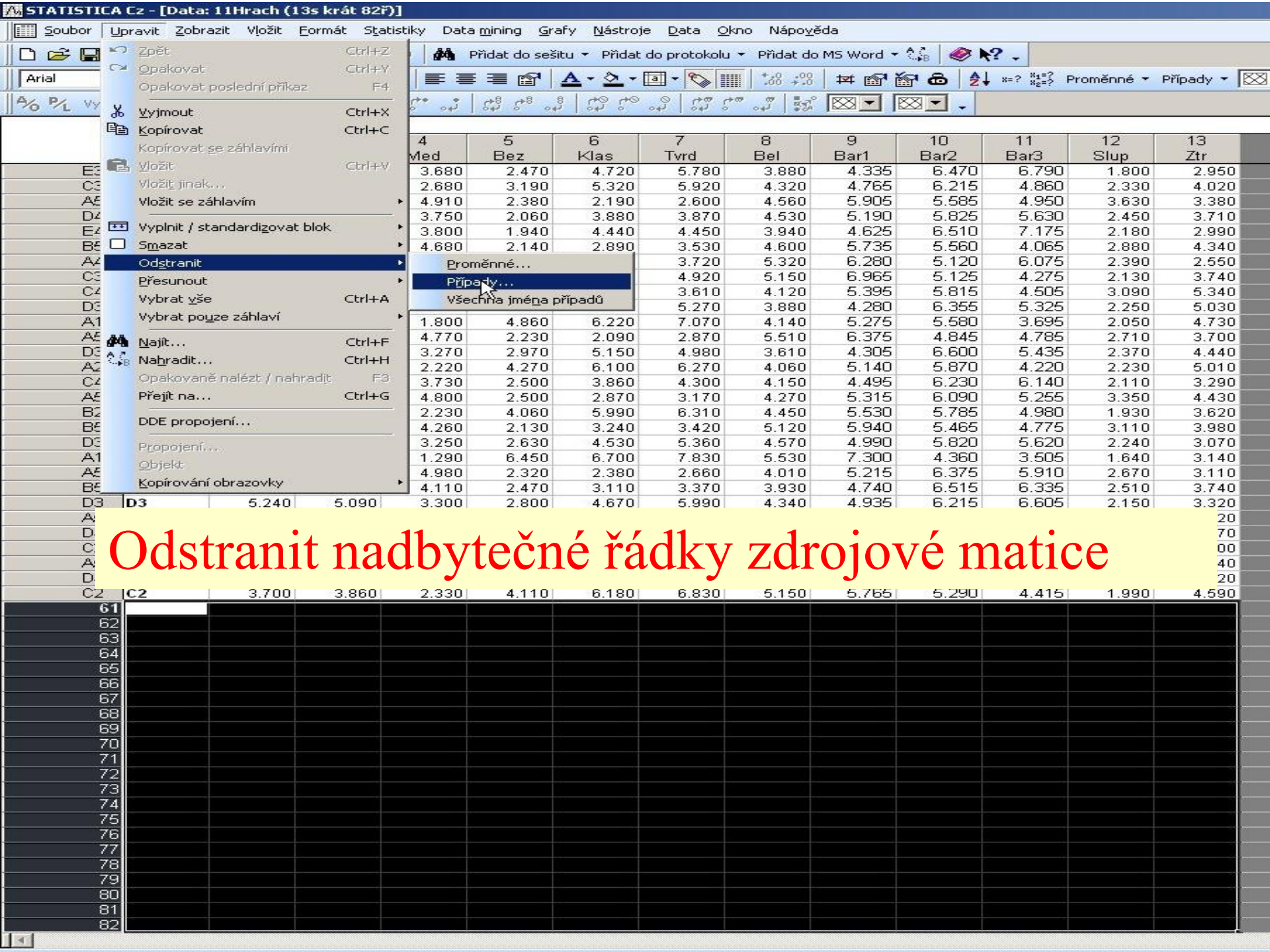
Arial 9 B I U

Proměnné Případy

Vypočtená data grafu... Přidat graf...

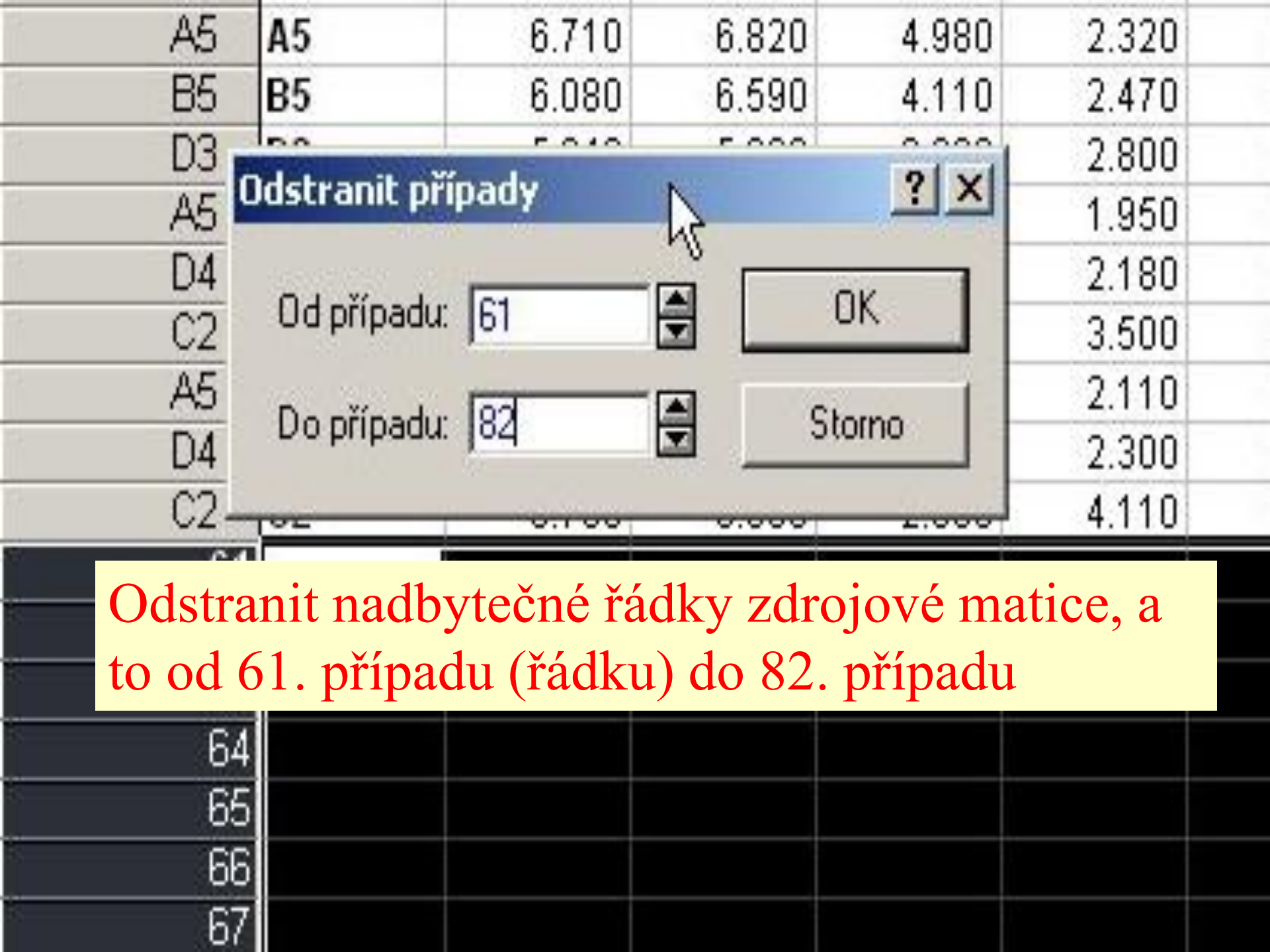
	1 Objekt	2 Aro	3 Slad	4 Med	5 Bez	6 Klas	7 Tvrd	8 Bel	9 Bar1	10 Bar2	11 Bar3	12 Slup	13 Ztr
C4	C4	5.750	6.090	3.810	2.320	4.030	3.770	4.170	5.730	5.745	5.325	3.820	3.380
B2	B2	3.940	4.120	2.440	3.630	5.770	5.390	4.770	6.665	5.105	4.595	3.500	3.030
D5	D5	6.600	6.120	4.440	1.930	3.310	4.460	4.860	5.160	5.740	6.565	2.120	3.940
D4	D4	5.680	5.980	3.800	2.120	3.850	4.140	5.030	5.635	5.220	5.480	2.380	5.160
E2	E2	4.740	4.660	2.880	2.940	5.650	5.770	5.310	5.940	5.270	5.890	1.750	3.640
B5	B5	6.310	6.130	4.780	1.940	2.700	3.260	5.070	5.710	5.370	6.365	3.650	4.550
C5	C5	6.200	6.020	4.650	1.780	3.120	3.740	5.250	5.655	5.475	5.960	2.510	3.800
C2	C2	3.790	3.880	2.310	3.520	6.240	5.730	5.390	6.300	5.135	5.230	2.010	4.110
A4	A4	5.680	6.340	3.750	2.790	4.170	3.870	4.520	4.920	5.760	4.570	2.970	4.410
D4	D4	6.100	6.090	3.990	2.070	4.260	4.250	4.010	5.020	6.175	5.380	2.500	4.600
B1	B1	3.410	3.180	1.820	4.640	6.240	7.430	4.260	4.835	5.955	4.550	1.850	4.270
D4	D4	5.890	6.090	3.990	2.290	3.900	4.590	4.530	5.890	5.630	3.815	2.200	4.710
E4	E4	5.770	5.320	3.880	2.260	4.220	4.990	5.050	5.335	5.590	5.540	2.160	3.650
B1	B1	3.390	3.280	1.980	4.500	6.040	7.140	4.350	5.090	5.580	4.400	2.030	4.620
B5	B5												
D4	D4												
C2	C2												
A5	A5												
C3	C3												
B2	B2												
B5	B5	7.090	6.090	5.180	1.740	2.570	3.180	5.230	5.920	5.515	4.115	2.090	3.100
D4	D4	5.720	5.300	3.730	2.340	3.950	4.800	3.640	4.000	6.805	6.755	1.740	2.930
B1	B1	3.220	3.210	1.950	4.420	6.240	7.270	4.600	6.030	5.600	4.165	1.680	3.580
A5	A5	6.110	6.620	4.290	2.580	3.200	2.860	3.500	4.950	6.220	5.370	2.150	4.310
D4	D4	6.070	6.270	3.980	2.190	3.890	4.240	3.850	4.460	6.675	6.205	2.200	3.700
A1	A1	2.660	2.660	1.430	6.100	6.670	7.750	4.270	4.970	5.630	4.525	1.650	3.780
B3	B3	5.260	5.490	3.460	3.030	4.850	4.170	5.220	5.410	5.415	6.120	3.080	3.950
C2	C2	3.720	4.350	2.200	4.080	6.500	6.270	4.990	5.535	5.560	5.335	1.820	3.920
D3	D3	5.430	5.190	3.470	2.400	4.430	5.260	4.460	4.780	5.720	5.895	1.610	3.770
B5	B5	6.550	6.570	4.710	2.120	3.060	3.430	3.760	4.430	6.450	6.380	2.630	3.850
E3	E3	5.530	5.410	3.680	2.470	4.720	5.780	3.880	4.335	6.470	6.790	1.800	2.950
C3	C3	4.710	4.680	2.680	3.190	5.320	5.920	4.320	4.765	6.215	4.860	2.330	4.020
A5	A5	6.280	7.030	4.910	2.380	2.190	2.600	4.560	5.905	5.585	4.950	3.630	3.380
D4	D4	5.920	5.820	3.750	2.060	3.880	3.870	4.530	5.190	5.825	5.630	2.450	3.710
E4	E4	6.090	5.720	3.800	1.940	4.440	4.450	3.940	4.625	6.510	7.175	2.180	2.990
B5	B5	6.370	6.500	4.680	2.140	2.890	3.530	4.600	5.735	5.560	4.065	2.880	4.340
A4	A4	5.710	5.680	3.970	2.650	4.390	3.720	5.320	6.280	5.120	6.075	2.390	2.550
C3	C3	4.530	5.030	2.640	3.120	5.860	4.920	5.150	6.965	5.125	4.275	2.130	3.740
C4	C4	5.950	6.280	4.040	2.190	3.930	3.610	4.120	5.395	5.815	4.505	3.090	5.340
D3	D3	5.510	5.410	3.720	2.780	4.760	5.270	3.880	4.280	6.355	5.325	2.250	5.030
A1	A1	3.100	3.430	1.800	4.860	6.220	7.070	4.140	5.275	5.580	3.695	2.050	4.730
A5	A5	6.500	6.680	4.770	2.230	2.090	2.870	5.510	6.375	4.845	4.785	2.710	3.700
D3	D3	5.460	5.410	3.270	2.970	5.150	4.980	3.610	4.305	6.600	5.435	2.370	4.440
A2	A2	3.750	4.300	2.220	4.270	6.100	6.270	4.060	5.140	5.870	4.220	2.230	5.010
C4	C4	5.860	5.270	3.730	2.500	3.860	4.300	4.150	4.495	6.230	6.140	2.110	3.290
A5	A5	6.160	6.970	4.800	2.500	2.870	3.170	4.270	5.315	6.090	5.255	3.350	4.430
B2	B2	3.870	3.880	2.230	4.060	5.990	6.310	4.450	5.530	5.785	4.980	1.930	3.620
B5	B5	6.240	5.800	4.260	2.130	3.240	3.420	5.120	5.940	5.465	4.775	3.110	3.980
D3	D3	5.690	4.970	3.250	2.630	4.530	5.360	4.570	4.990	5.820	5.620	2.240	3.070
A1	A1	2.280	2.230	1.290	6.450	6.700	7.830	5.530	7.300	4.360	3.505	1.640	3.140
A5	A5	6.710	6.820	4.980	2.320	2.380	2.660	4.010	5.215	6.375	5.910	2.670	3.110

Překopírovat 1. sloupec do nulového

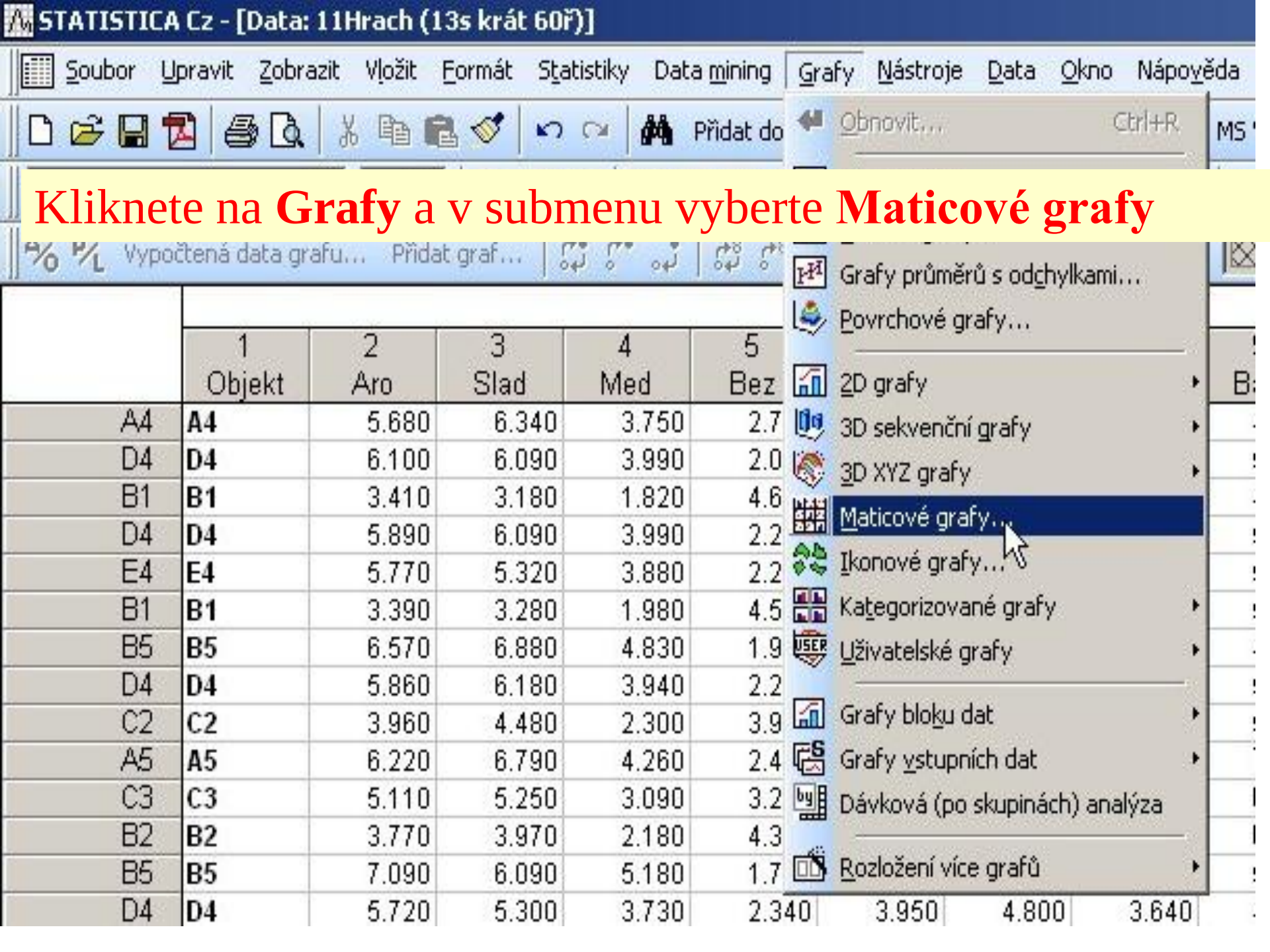


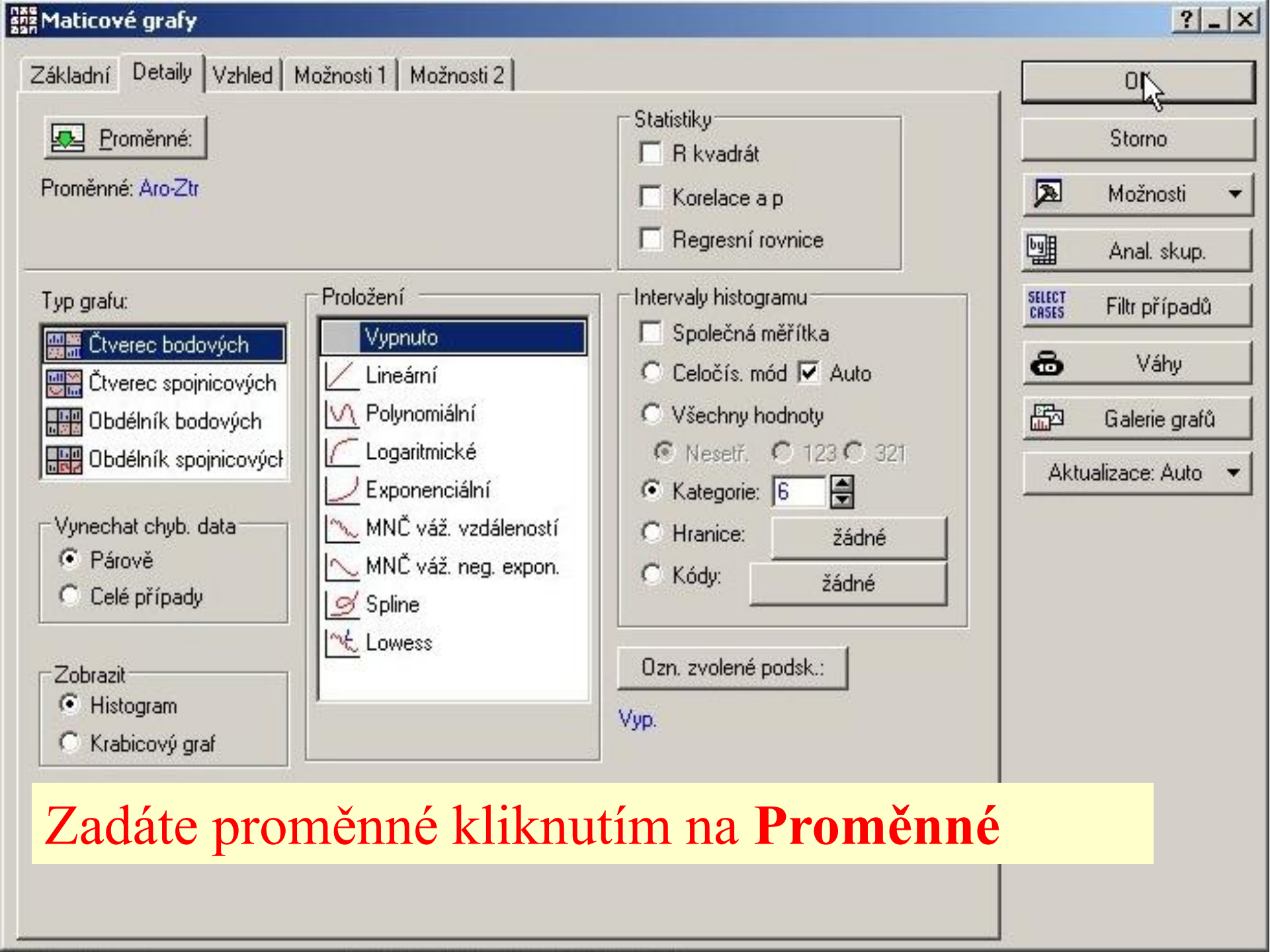
		4 Med	5 Bez	6 Klas	7 Tvrd	8 Bel	9 Bar1	10 Bar2	11 Bar3	12 Slup	13 Ztr
E3		3.680	2.470	4.720	5.780	3.880	4.335	6.470	6.790	1.800	2.950
C3		2.680	3.190	5.320	5.920	4.320	4.765	6.215	4.860	2.330	4.020
A5		4.910	2.380	2.190	2.600	4.560	5.905	5.585	4.950	3.630	3.380
D4		3.750	2.060	3.880	3.870	4.530	5.190	5.825	5.630	2.450	3.710
E4		3.800	1.940	4.440	4.450	3.940	4.625	6.510	7.175	2.180	2.990
B5		4.680	2.140	2.890	3.530	4.600	5.735	5.560	4.065	2.880	4.340
A4					3.720	5.320	6.280	5.120	6.075	2.390	2.550
C3					4.920	5.150	6.965	5.125	4.275	2.130	3.740
C4					3.610	4.120	5.395	5.815	4.505	3.090	5.340
D3					5.270	3.880	4.280	6.355	5.325	2.250	5.030
A1		1.800	4.860	6.220	7.070	4.140	5.275	5.580	3.695	2.050	4.730
A5		4.770	2.230	2.090	2.870	5.510	6.375	4.845	4.785	2.710	3.700
D3		3.270	2.970	5.150	4.980	3.610	4.305	6.600	5.435	2.370	4.440
A2		2.220	4.270	6.100	6.270	4.060	5.140	5.870	4.220	2.230	5.010
C4		3.730	2.500	3.860	4.300	4.150	4.495	6.230	6.140	2.110	3.290
A5		4.800	2.500	2.870	3.170	4.270	5.315	6.090	5.255	3.350	4.430
B2		2.230	4.060	5.990	6.310	4.450	5.530	5.785	4.980	1.930	3.620
B5		4.260	2.130	3.240	3.420	5.120	5.940	5.465	4.775	3.110	3.980
D3		3.250	2.630	4.530	5.360	4.570	4.990	5.820	5.620	2.240	3.070
A1		1.290	6.450	6.700	7.830	5.530	7.300	4.360	3.505	1.640	3.140
A5		4.980	2.320	2.380	2.660	4.010	5.215	6.375	5.910	2.670	3.110
B5		4.110	2.470	3.110	3.370	3.930	4.740	6.515	6.335	2.510	3.740
D3	D3	5.240	5.090	3.300	2.800	4.670	5.990	4.340	4.935	6.215	6.605	2.150	3.320
A													20
D													70
C													00
A													40
D													20
C2	C2	3.700	3.860	2.330	4.110	6.180	6.830	5.150	5.765	5.290	4.415	1.990	4.590

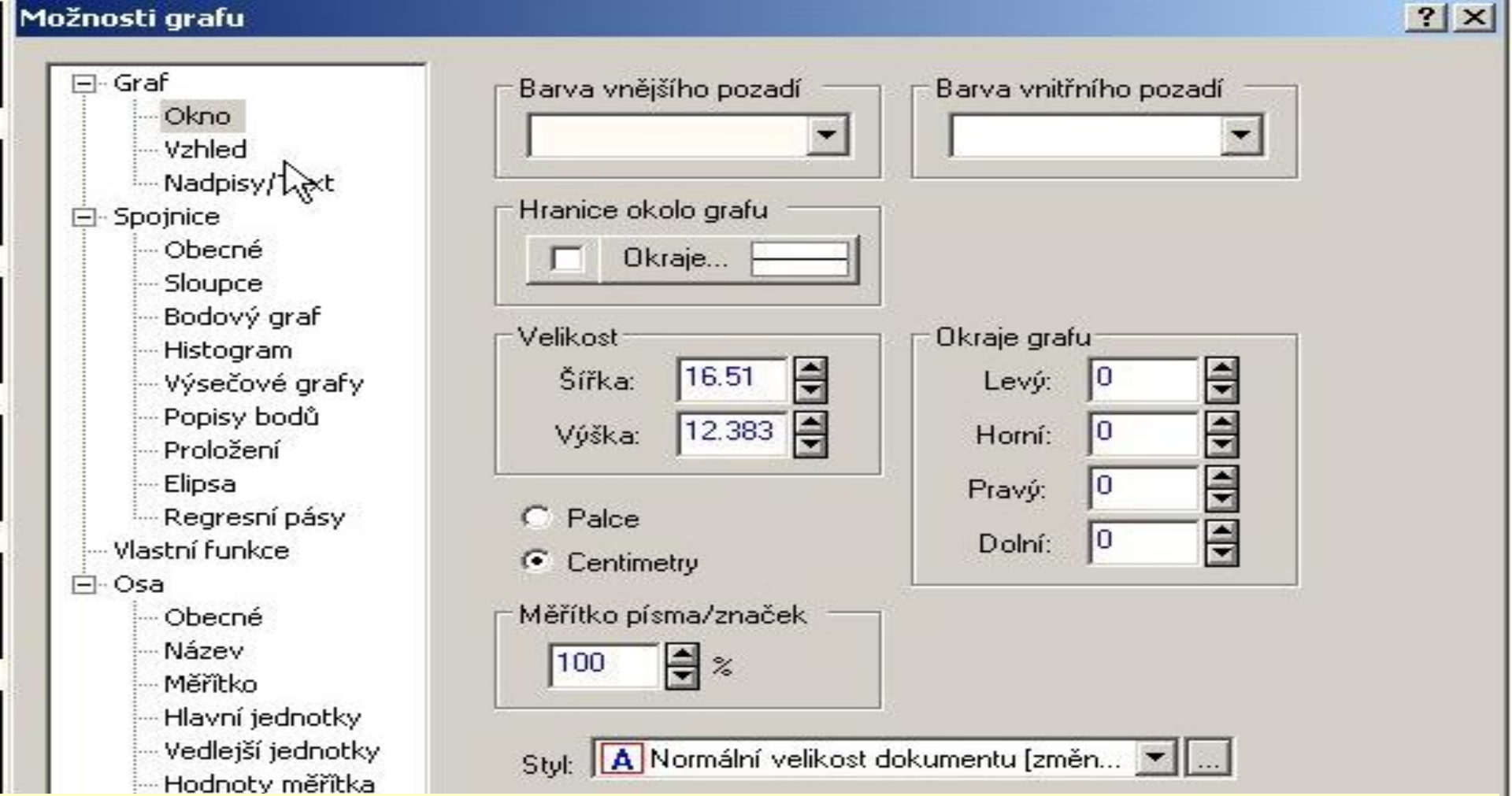
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82



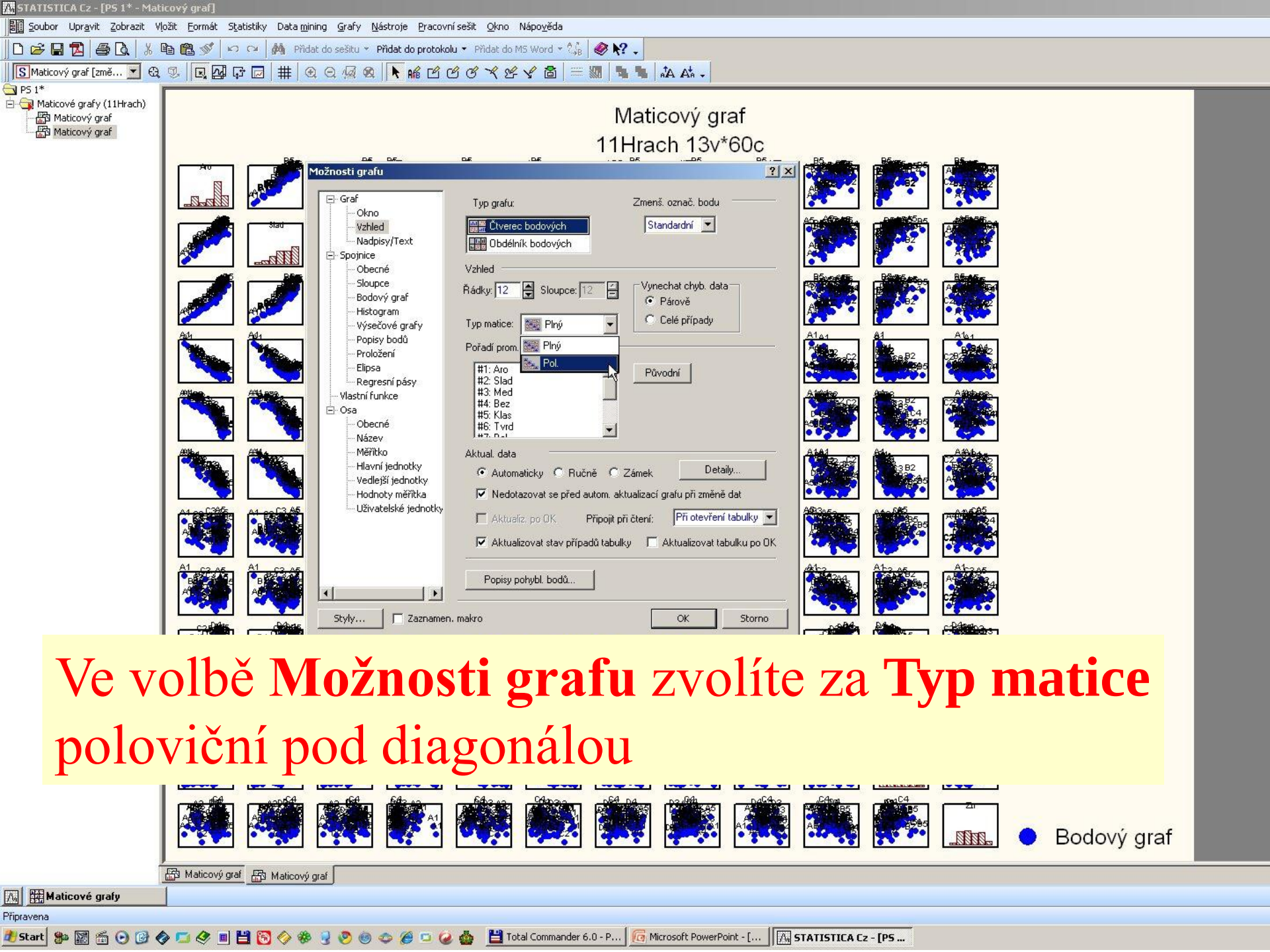
Odstranit nadbytečné řádky zdrojové matice, a to od 61. případu (řádku) do 82. případu


STATISTICA Cz - [Data: 11Hrach (13s krát 60ř)]
Soubor Upravit Zobrazit Vložit Formát Statistiky Data mining Grafy Nástroje Data Okno Nápověda
Přidat do
Obnovit... Ctrl+R
Kliknete na Grafy a v submenu vyberte Maticové grafy
Vypočtená data grafu... Přidat graf...
Grafy průměrů s odchylkami...
Povrchové grafy...
2D grafy
3D sekvenční grafy
3D XYZ grafy
Maticové grafy...
Ikonové grafy...
Kategorizované grafy
Uživatelské grafy
Grafy bloku dat
Grafy vstupních dat
Dávková (po skupinách) analýza
Rozložení více grafů
1 Objekt
2 Aro
3 Slad
4 Med
5 Bez
A4 A4 5.680 6.340 3.750 2.7
D4 D4 6.100 6.090 3.990 2.0
B1 B1 3.410 3.180 1.820 4.6
D4 D4 5.890 6.090 3.990 2.2
E4 E4 5.770 5.320 3.880 2.2
B1 B1 3.390 3.280 1.980 4.5
B5 B5 6.570 6.880 4.830 1.9
D4 D4 5.860 6.180 3.940 2.2
C2 C2 3.960 4.480 2.300 3.9
A5 A5 6.220 6.790 4.260 2.4
C3 C3 5.110 5.250 3.090 3.2
B2 B2 3.770 3.970 2.180 4.3
B5 B5 7.090 6.090 5.180 1.7
D4 D4 5.720 5.300 3.730 2.340 3.950 4.800 3.640

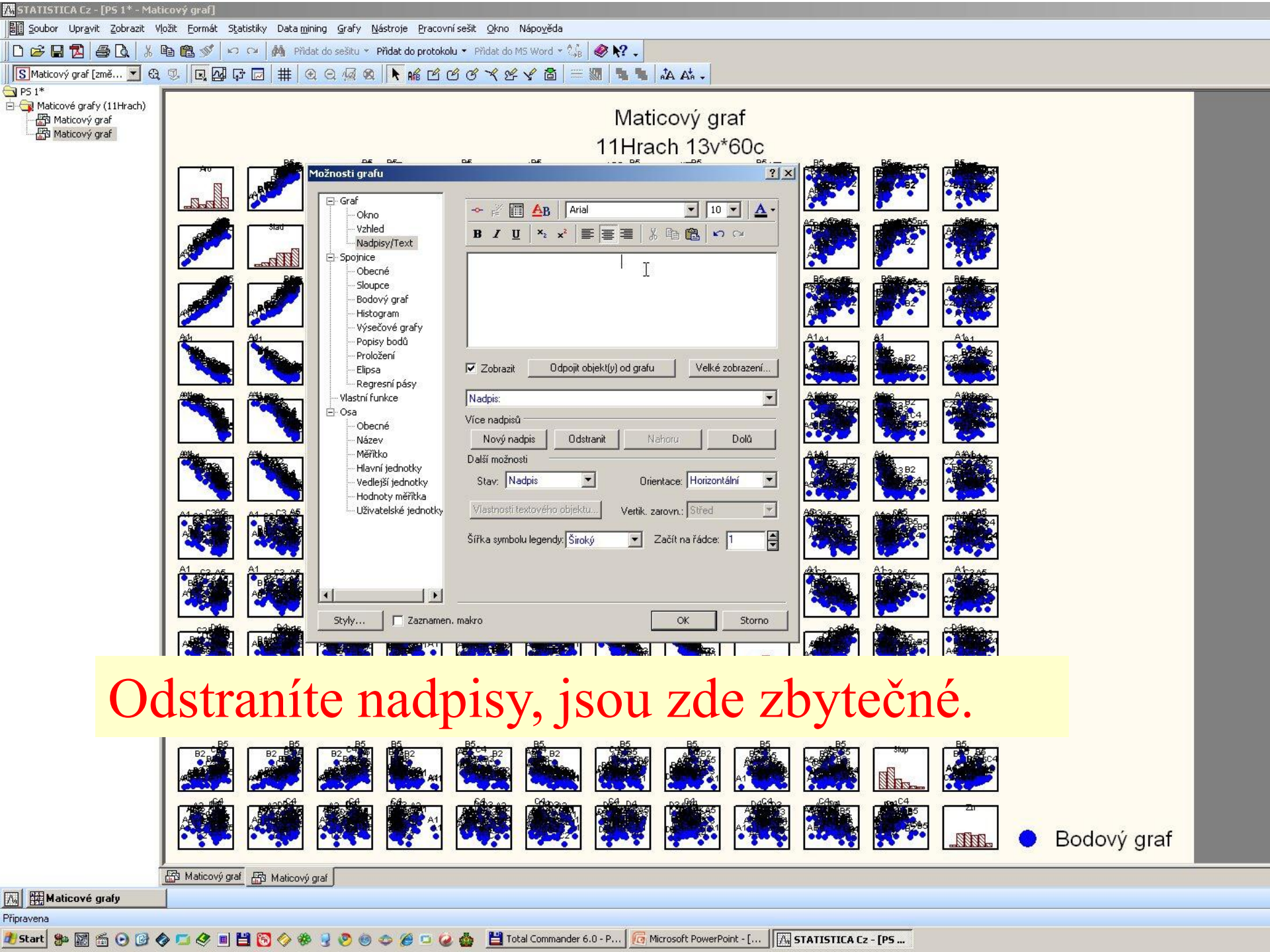

Maticové grafy
Základní
Detaily
Vzhled
Možnosti 1
Možnosti 2
OK
Storno
Možnosti
Anal. skup.
Filtr případů
Váhy
Galerie grafů
Aktualizace: Auto
Proměnné:
Proměnné: Aro-Ztr
Statistiky
R kvadrát
Korelace a p
Regresní rovnice
Typ grafu:
Čtverec bodových
Čtverec spojnicových
Obdélník bodových
Obdélník spojnicových
Proložení
Vypnuto
Lineární
Polynomiální
Logaritmické
Exponenciální
MNČ váž. vzdáleností
MNČ váž. neg. expon.
Spline
Lowess
Intervaly histogramu
Společná měřítka
Celočís. mód
Auto
Všechny hodnoty
Neseř.
123
321
Kategorie:
6
Hranice:
žádné
Kódy:
žádné
Vynechat chyb. data
Párově
Celé případy
Zobrazit
Histogram
Krabicový graf
Ozn. zvolené podsk.:
Vyp.
Zadáte proměnné kliknutím na **Proměnné**

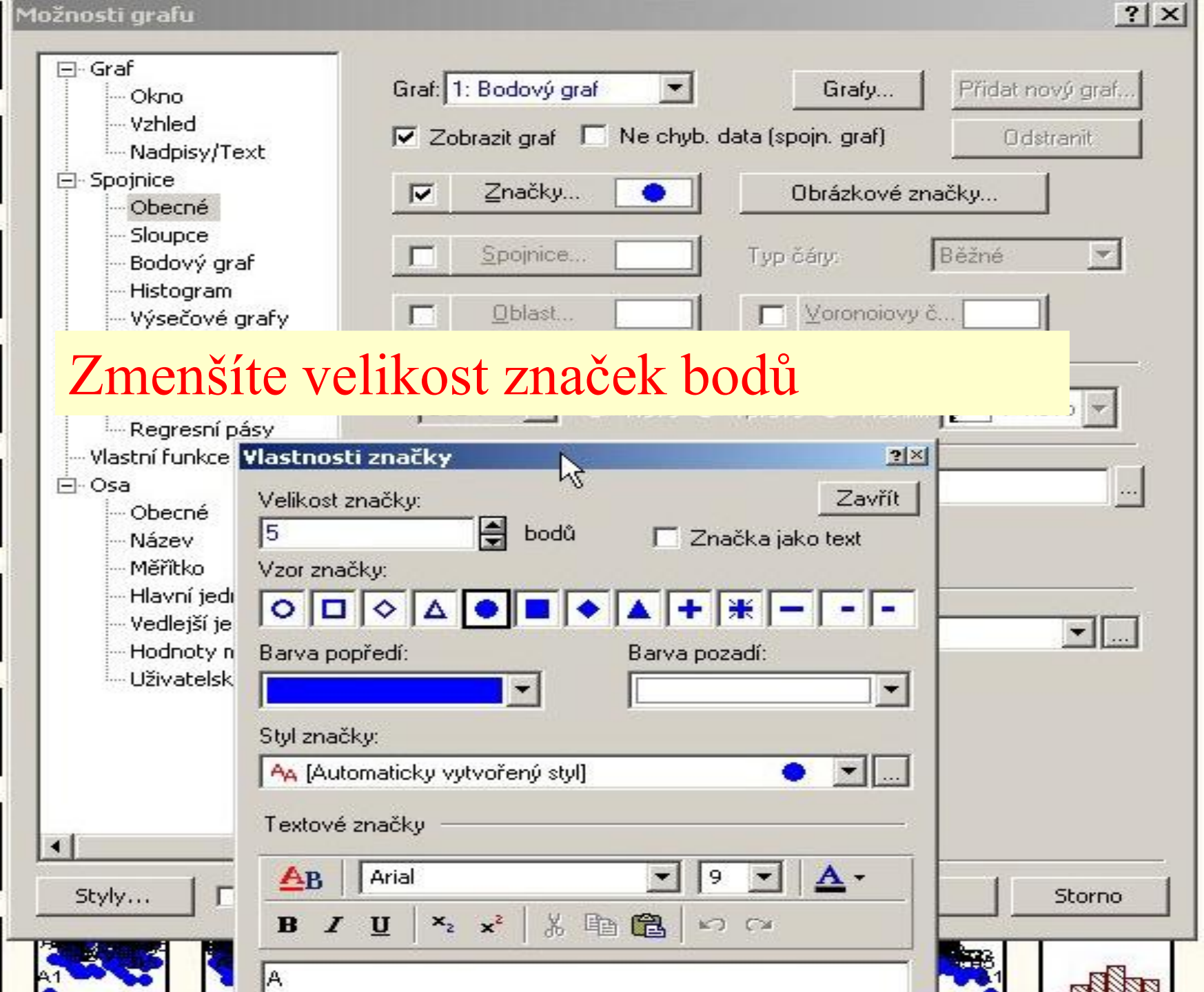


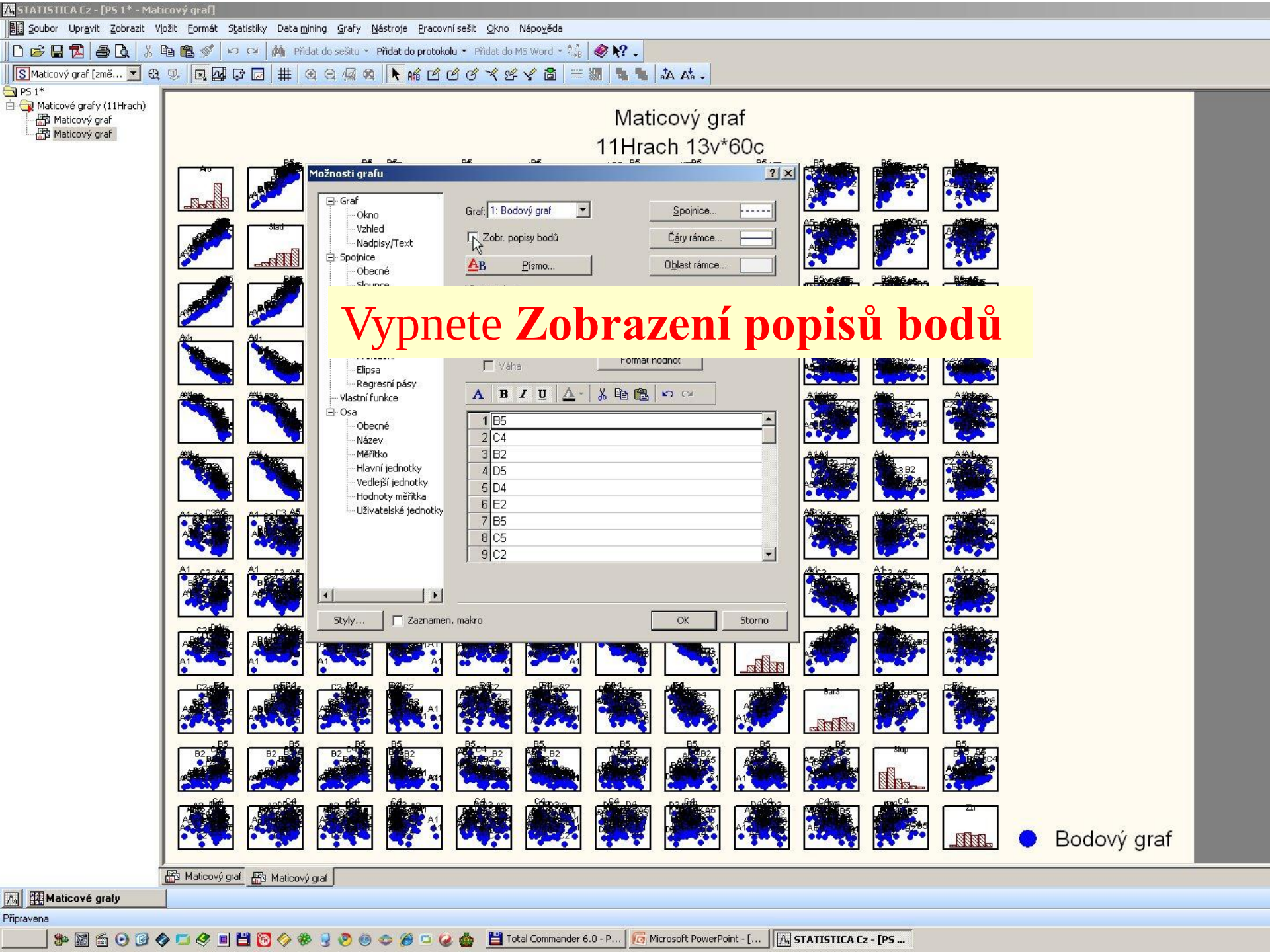
Zvolíte míru **Centimetry** a můžete nastavit velikost obrázku



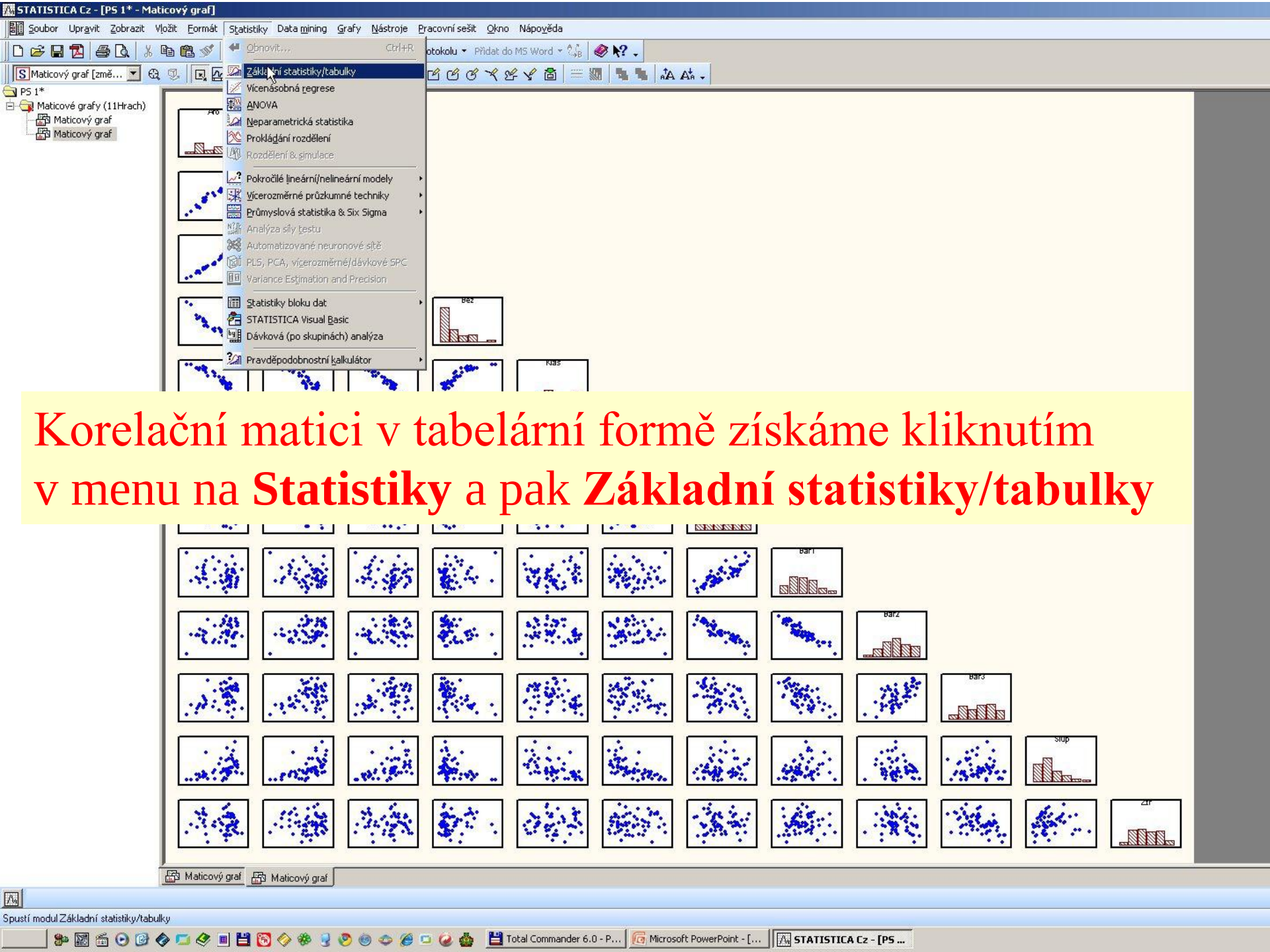
Ve volbě **Možnosti grafu** zvolíte za **Typ matice** poloviční pod diagonálou


STATISTICA Cz - [PS 1* - Maticový graf]
Soubor Upravit Zobrazit Vložit Formát Statistiky Data mining Grafy Nástroje Pracovní sešit Okno Nápověda
PS 1*
Maticové grafy (11Hrach)
Maticový graf
Maticový graf
Maticový graf
11Hrach 13v*60c
Možnosti grafu
Graf
Okno
Vzhled
Nadpisy/Text
Spojnice
Obecné
Sloupce
Bodový graf
Histogram
Výsečové grafy
Popisy bodů
Proložení
Elipsa
Regresní pásy
Vlastní funkce
Osa
Obecné
Název
Měřítko
Hlavní jednotky
Vedlejší jednotky
Hodnoty měřítka
Uživatelské jednotky
Arial
10
Zobrazit
Odpojit objekt(y) od grafu
Velké zobrazení...
Nadpis:
Více nadpisů
Nový nadpis
Odstranit
Nahoru
Dolů
Další možnosti
Stav: Nadpis
Orientace: Horizontální
Vlastnosti textového objektu...
Vertik. zarovn.: Střed
Šířka symbolu legendy: Široký
Začít na řádce: 1
Styly...
Zaznamen. makro
OK
Storno
Odstraníte nadpisy, jsou zde zbytečné.
Bodový graf
Maticový graf
Maticový graf
Maticové grafy
Připravena
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...


Možnosti grafu
Graf
Okno
Vzhled
Nadpisy/Text
Spojnice
Obecné
Sloupce
Bodový graf
Histogram
Výsečové grafy
Regresní pásy
Vlastní funkce
Osa
Obecné
Název
Měřítko
Hlavní jedn
Vedlejší je
Hodnoty n
Uživatelsk
Styly...
Graf: 1: Bodový graf
Grafy...
Přidat nový graf...
Zobrazit graf
Ne chyb. data (spojn. graf)
Odstranit
Značky...
Obrázkové značky...
Spojnice...
Typ čáry:
Běžné
Oblast...
Voronoiovy č...
Storno
Zmenšíte velikost značek bodů
Vlastnosti značky
Velikost značky:
5
bodů
Značka jako text
Zavřít
Vzor značky:
Barva popředí:
Barva pozadí:
Styl značky:
[Automaticky vytvořený styl]
Textové značky
Arial
9
A


STATISTICA Cz - [PS 1* - Maticový graf]
Soubor Upravit Zobrazit Vložit Formát Statistiky Data mining Grafy Nástroje Pracovní sešit Okno Nápověda
PS 1*
Maticové grafy (11Hrach)
Maticový graf
Maticový graf
Maticový graf
11Hrach 13v*60c
Možnosti grafu
Graf
Okno
Vzhled
Nadpisy/Text
Spojnice
Obecné
Elipsa
Regresní pásy
Vlastní funkce
Osa
Obecné
Název
Měřítko
Hlavní jednotky
Vedlejší jednotky
Hodnoty měřítka
Uživatelské jednotky
Graf: 1: Bodový graf
Zobr. popisy bodů
Písmo...
Spojnice...
Čáry rámce...
Oblast rámce...
Váha
Formát hodnot
1 B5
2 C4
3 B2
4 D5
5 D4
6 E2
7 B5
8 C5
9 C2
Styly...
Zaznamen. makro
OK
Storno
Vypnete Zobrazení popisů bodů
Bodový graf
Maticový graf
Maticový graf
Maticové grafy
Připravena
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...

Obdržíte takovýto maticový diagram korelace



Korelační matici v tabelární formě získáme kliknutím v menu na **Statistiky** a pak **Základní statistiky/tabulky**

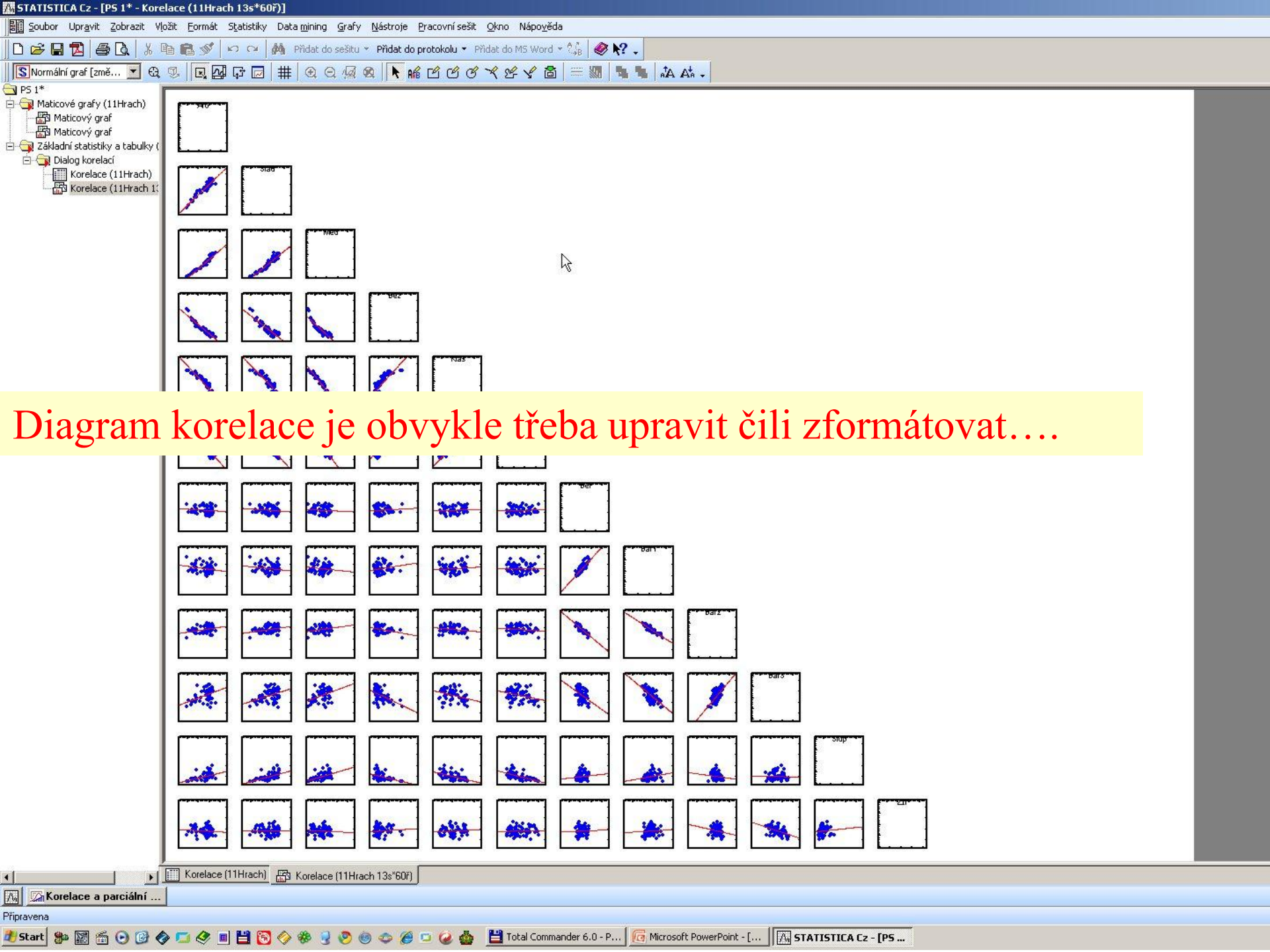
Pak zvolíte **Korelace a parciální korelace** a po zadání proměnných v záložce **Základní výsledky** kliknete na **Souhrn:Korelace**

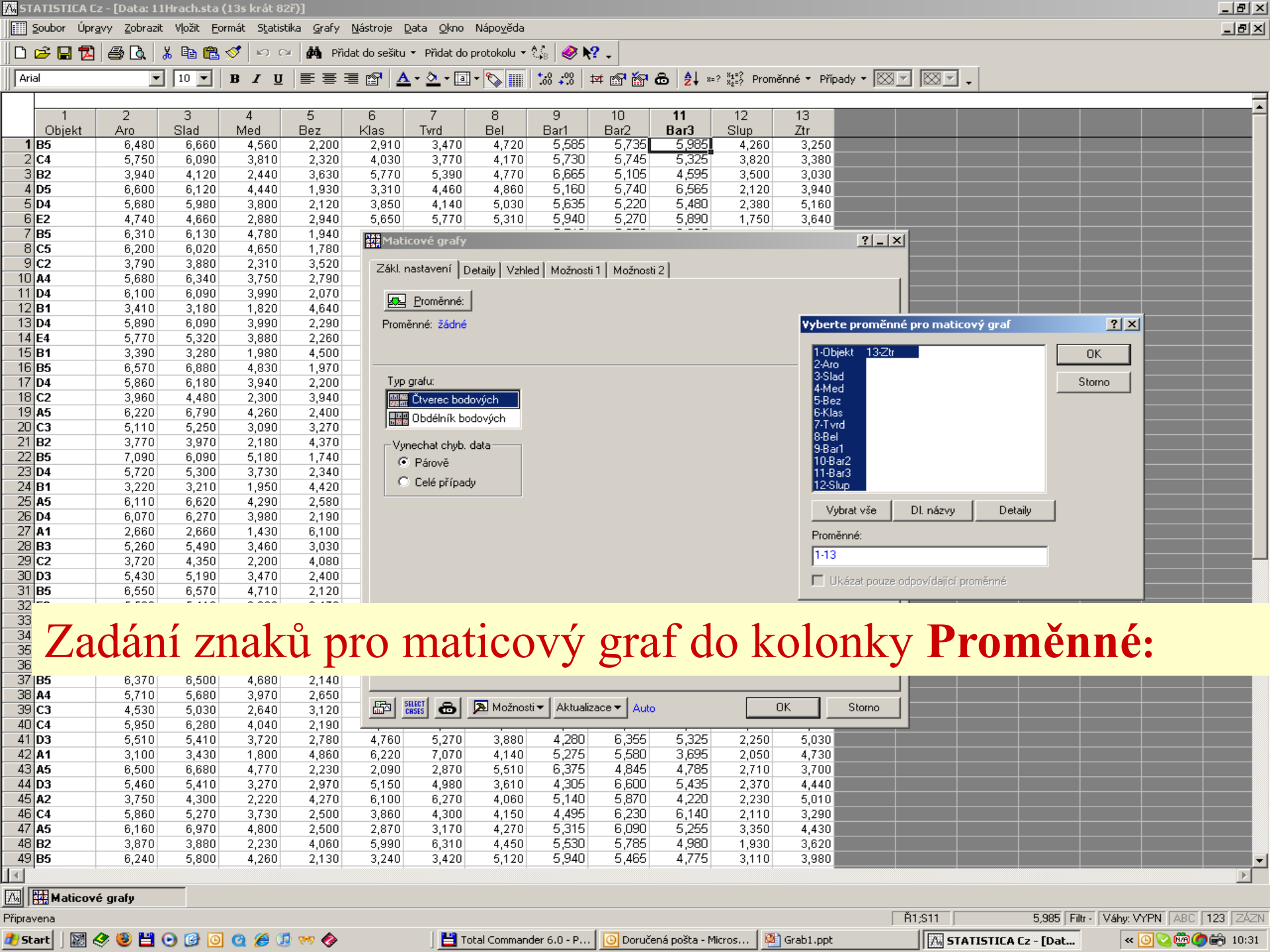
Korelace (11Hrach)
Označ. korelace jsou významné na hlad. p < .05000
N=60 (Celé případy vynechány u ChD)

Proměnná	Aro	Slad	Med	Bez	Klas	Tvrd	Bel	Bar1	Bar2	Bar3	Slup	Ztr
Aro	1.00	0.95	0.98	-0.95	-0.93	-0.92	-0.13	-0.22	0.34	0.45	0.48	-0.11
Slad	0.95	1.00	0.95	-0.90	-0.91	-0.95	-0.15	-0.17	0.33	0.39	0.55	-0.02
Med	0.98	0.95	1.00	-0.90	-0.97	-0.94	-0.09	-0.15	0.28	0.40	0.52	-0.08
Bez	-0.95	-0.90	-0.90	1.00	0.84	0.85	0.11	0.23	-0.35	-0.51	-0.42	0.10
Klas	-0.93	-0.91	-0.97	0.84	1.00	0.93	0.09	0.11	-0.23	-0.33	-0.54	0.09
Tvrd	-0.92	-0.95	-0.94	0.85	0.93	1.00	0.03	0.02	-0.17	-0.30	-0.63	0.04
Bel	-0.13	-0.15	-0.09	0.11	0.09	0.03	1.00	0.84	-0.89	-0.41	0.11	0.04
Bar1	-0.22	-0.17	-0.15	0.23	0.11	0.02	0.84	1.00	-0.91	-0.63	0.23	0.07
Bar2	0.34	0.33	0.28	-0.35	-0.23	-0.17	-0.89	-0.91	1.00	0.65	-0.13	-0.18
Bar3	0.45	0.39	0.40	-0.51	-0.33	-0.30	-0.41	-0.63	0.65	1.00	-0.02	-0.49
Slup	0.48	0.55	0.52	-0.42	-0.54	-0.63	0.11	0.23	-0.13	-0.02	1.00	0.10
Ztr	-0.11	-0.02	-0.08	0.10	0.09	0.04	0.04	0.07	-0.18	-0.49	0.10	1.00

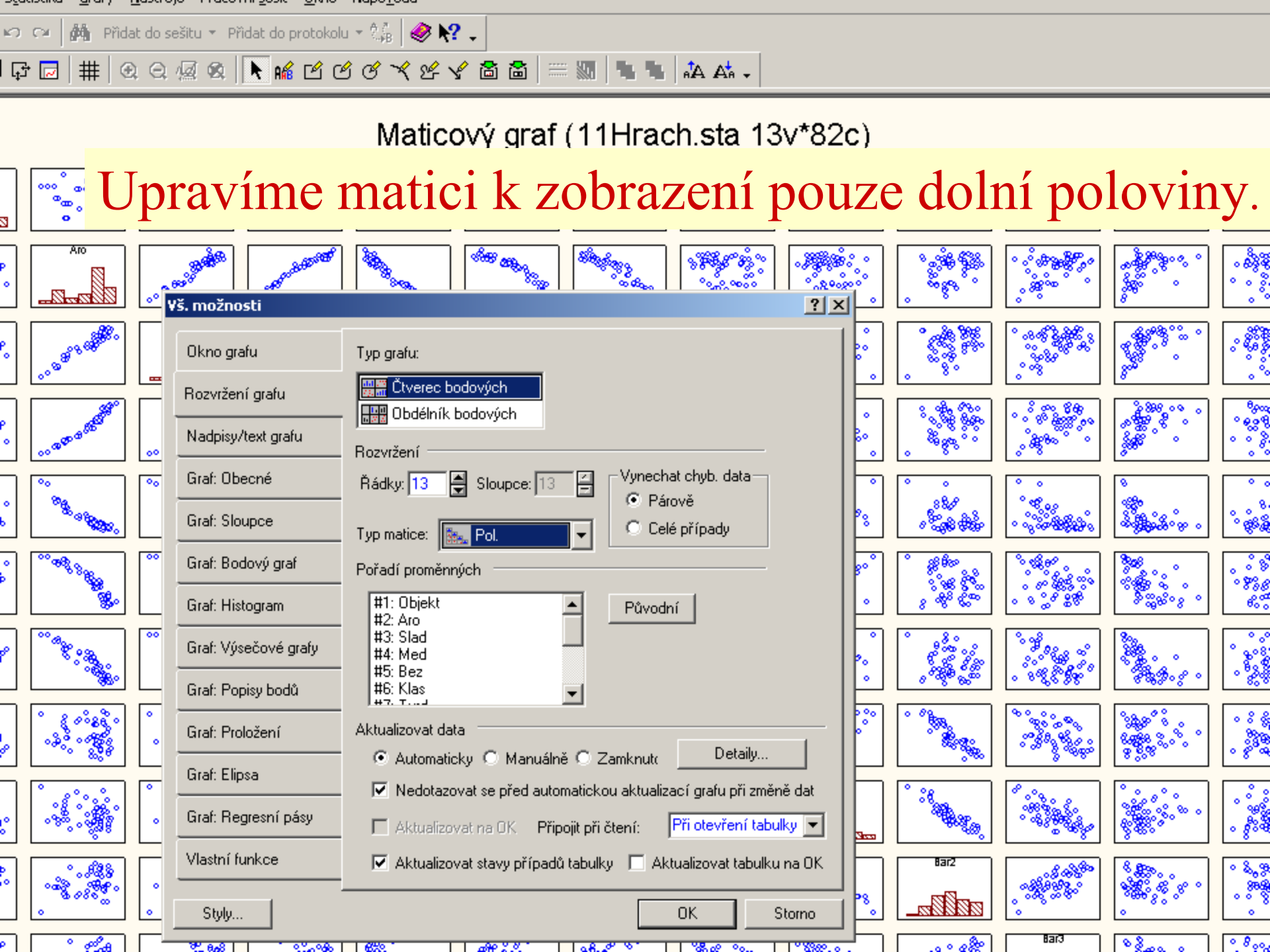
V korelační matici jsou **červeně** vyznačeny statisticky významné hodnoty Pearsonova korelačního koeficientu. Matice je symetrická dle diagonály.

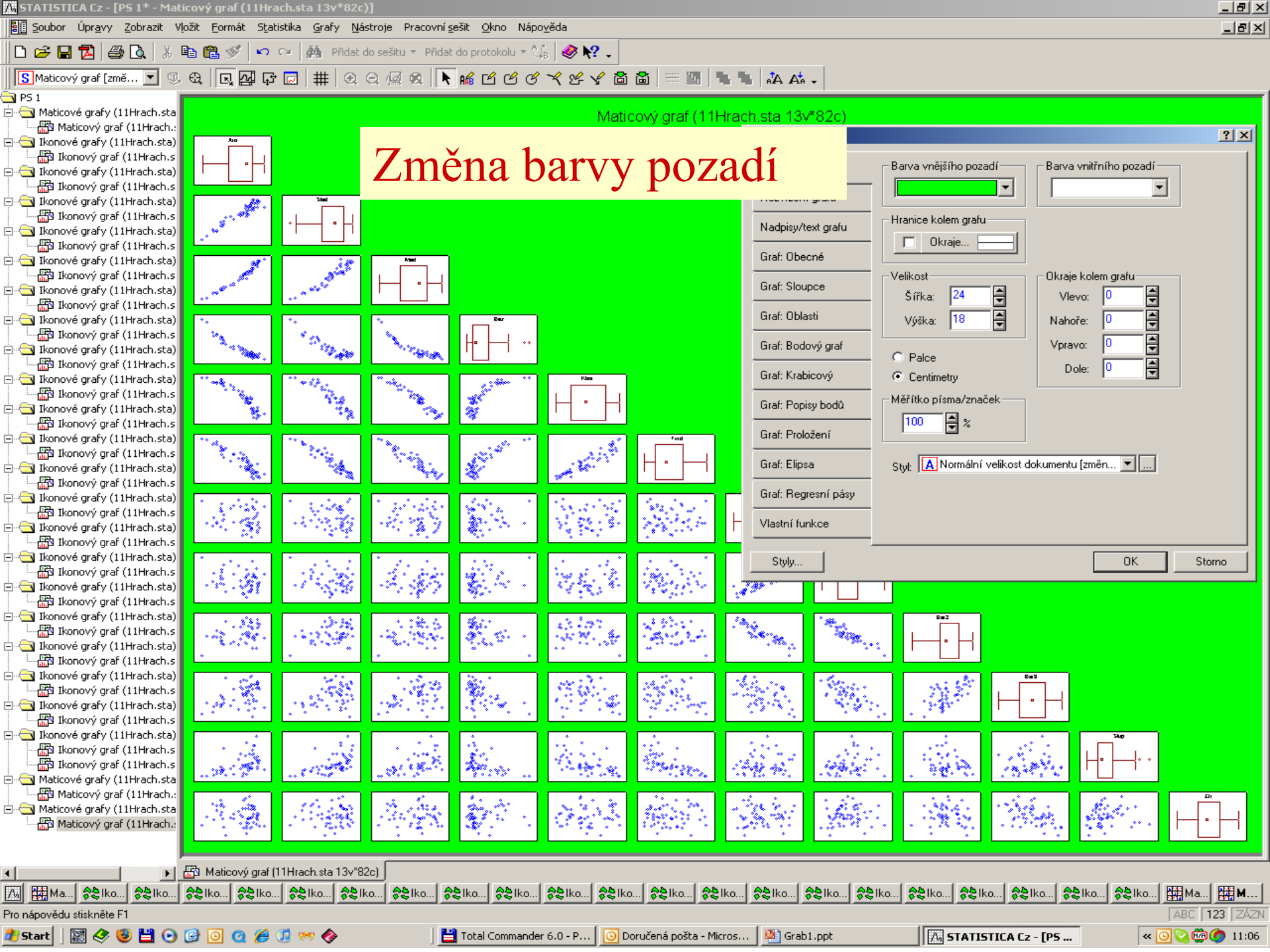
Po zadání zavolených (jenom některých) proměnných lze i zde zobrazit diagram korelace kliknutím na **Matice bod.grafů zvolených proměnných**

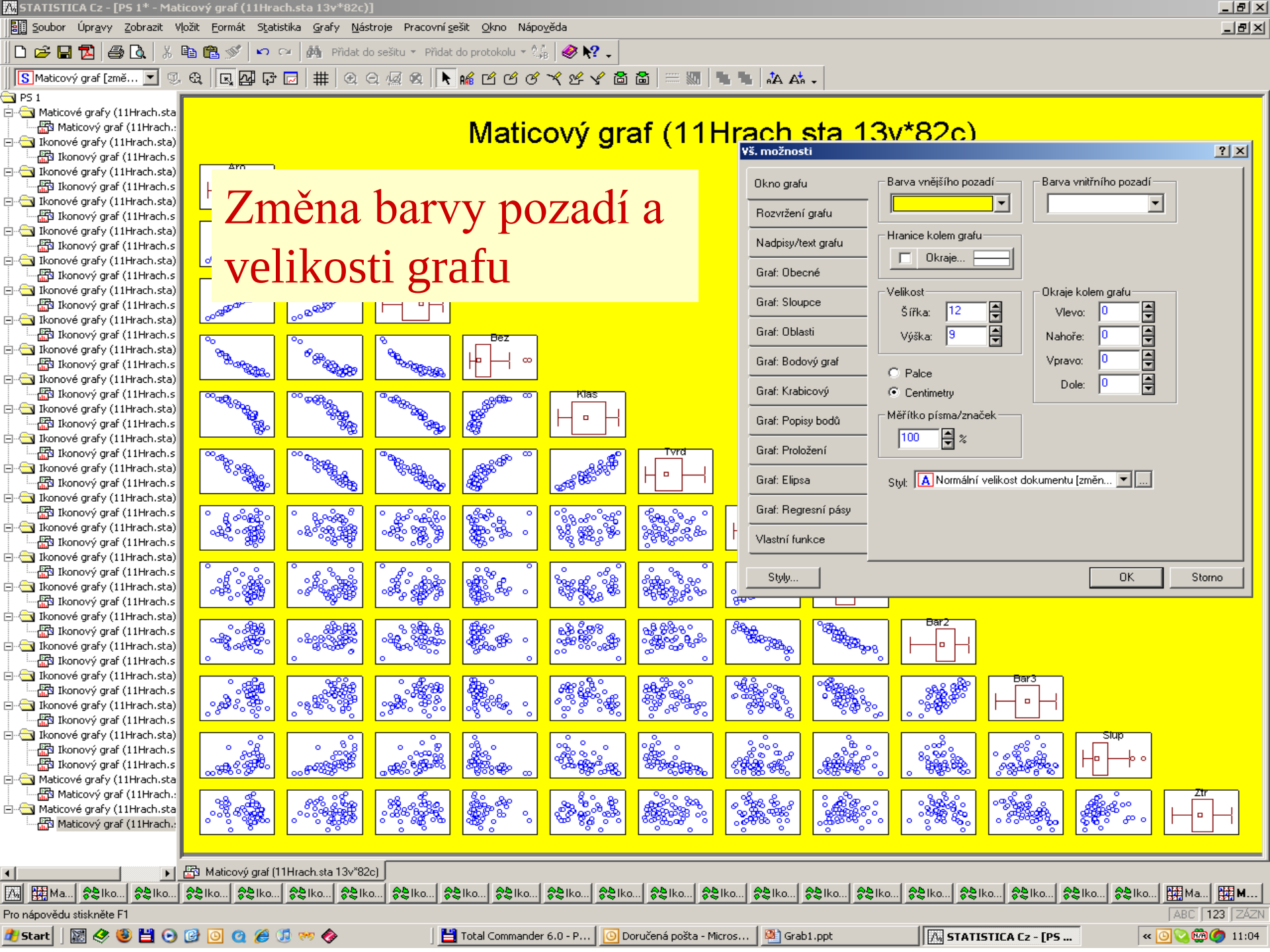

STATISTICA Cz - [PS 1* - Korelace (11Hrach 13s*60f)]
Soubor Upravit Zobrazit Vložit Formát Statistiky Data mining Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Přidat do MS Word
Normální graf [změ...
PS 1*
Maticové grafy (11Hrach)
Maticový graf
Maticový graf
Základní statistiky a tabulky (
Dialog korelací
Korelace (11Hrach)
Korelace (11Hrach 1
Diagram korelace je obvykle třeba upravit čili zformátovat….
Korelace (11Hrach)
Korelace (11Hrach 13s*60f)
Korelace a parciální ...
Připravena
Start
Total Commander 6.0 - P...
Microsoft PowerPoint - [...
STATISTICA Cz - [PS ...

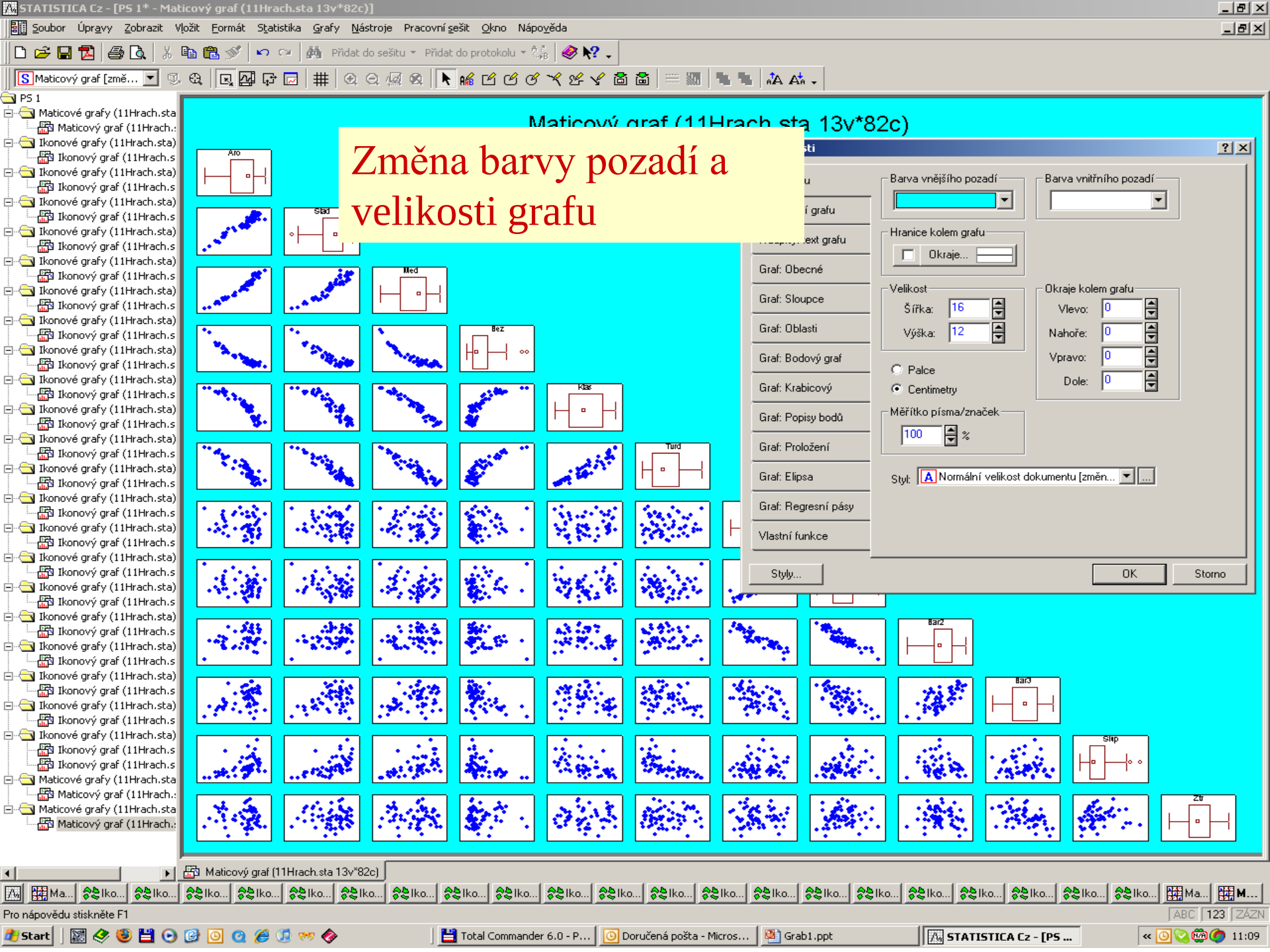


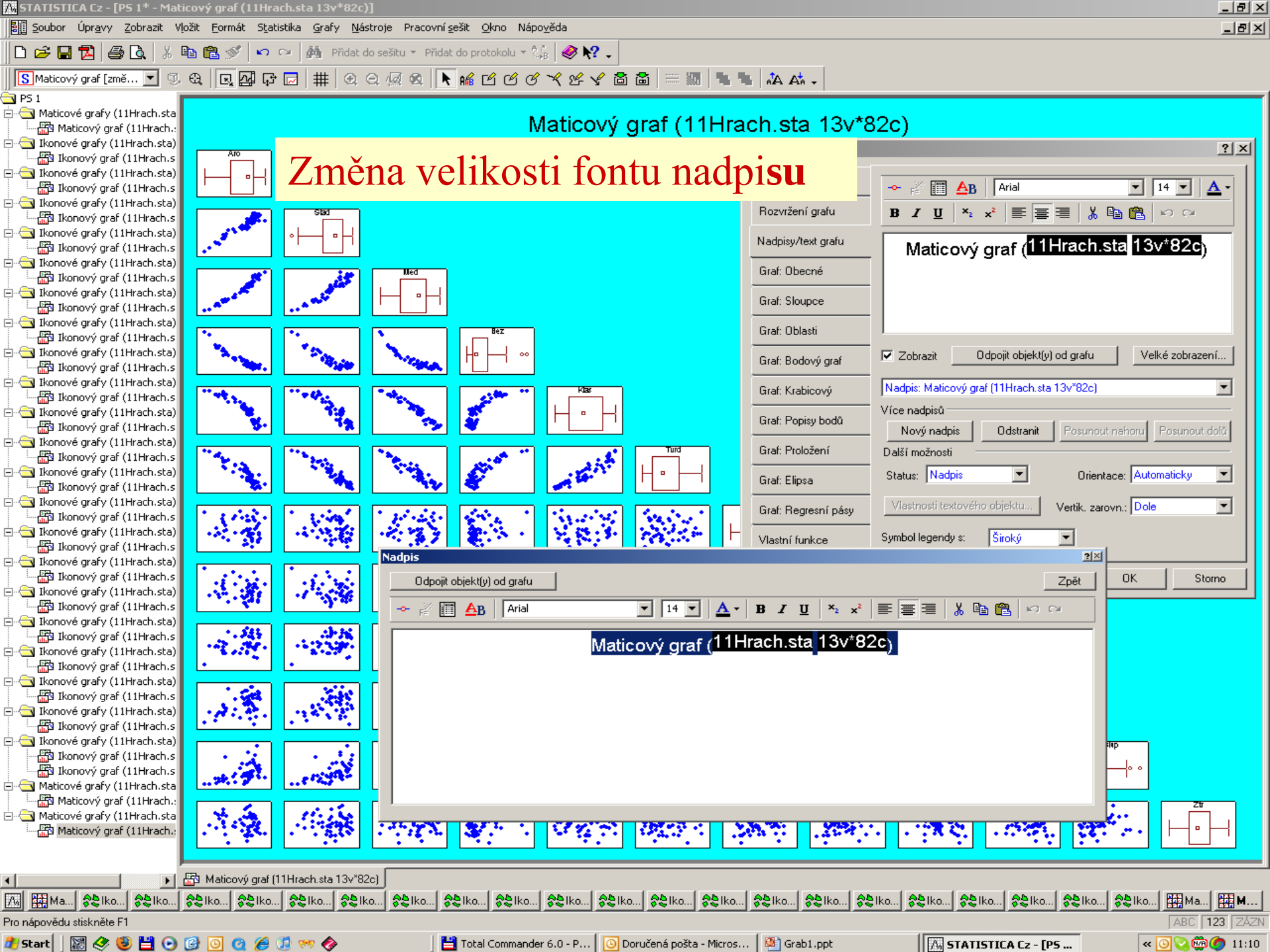
Zadání znaků pro maticový graf do kolonky **Proměnné:**

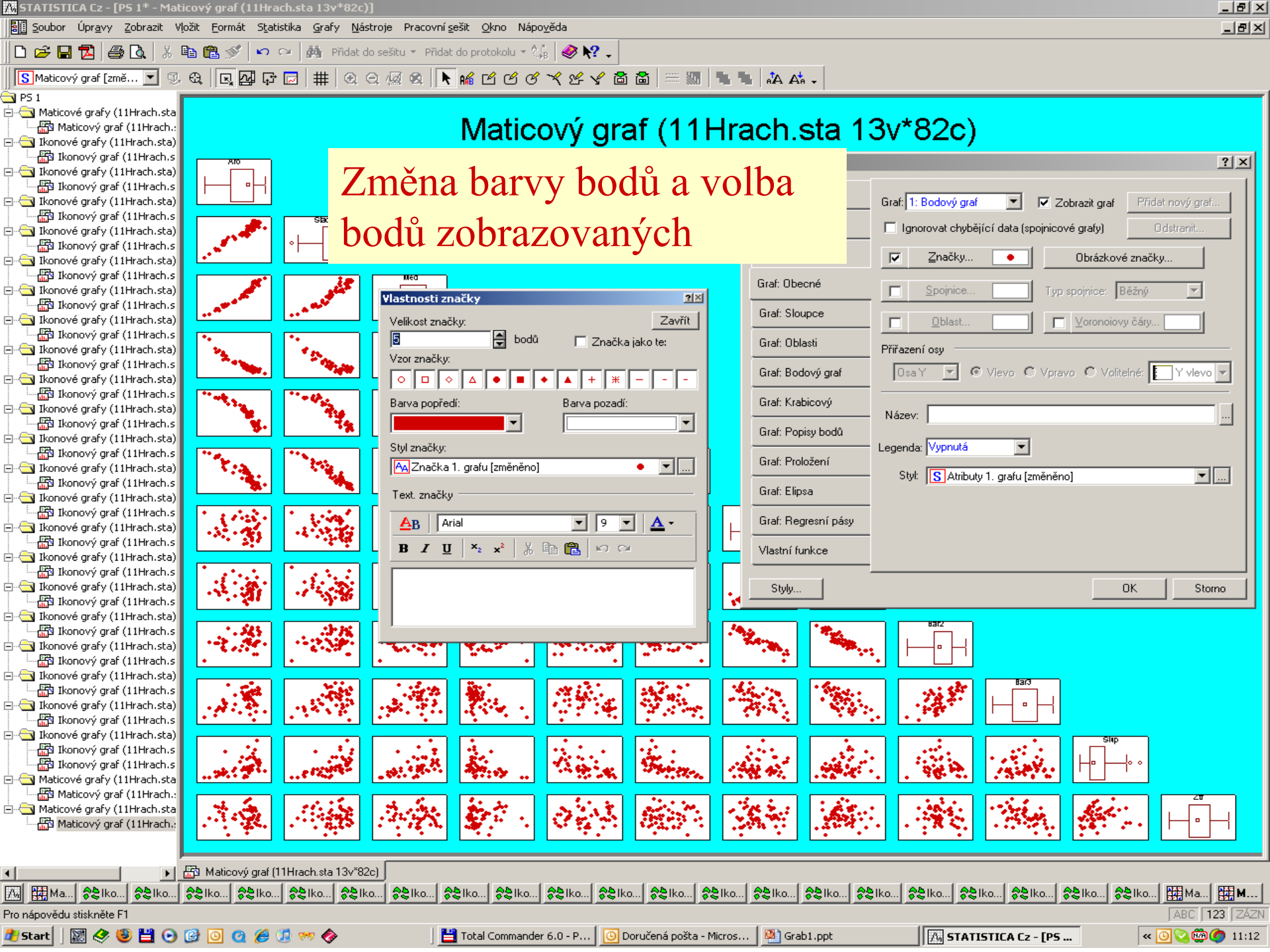

Přidat do sešitu
Přidat do protokolu
Maticový graf (11Hrach.sta 13v*82c)
Upravíme matici k zobrazení pouze dolní poloviny.
Vš. možnosti
Okno grafu
Rozvržení grafu
Nadpisy/text grafu
Graf: Obecné
Graf: Sloupce
Graf: Bodový graf
Graf: Histogram
Graf: Výsečové grafy
Graf: Popisy bodů
Graf: Proložení
Graf: Elipsa
Graf: Regresní pásy
Vlastní funkce
Typ grafu:
Čtverec bodových
Obdélník bodových
Rozvržení
Řádky: 13
Sloupce: 13
Vynechat chyb. data
Párově
Celé případy
Typ matice:
Pol.
Pořadí proměnných
#1: Objekt
#2: Aro
#3: Slad
#4: Med
#5: Bez
#6: Klas
Původní
Aktualizovat data
Automaticky
Manuálně
Zamknuto
Detaily...
Nedotazovat se před automatickou aktualizací grafu při změně dat
Aktualizovat na OK
Připojit při čtení:
Při otevření tabulky
Aktualizovat stavy případů tabulky
Aktualizovat tabulku na OK
Styly...
OK
Storno
Aro
Bar2
Bar3

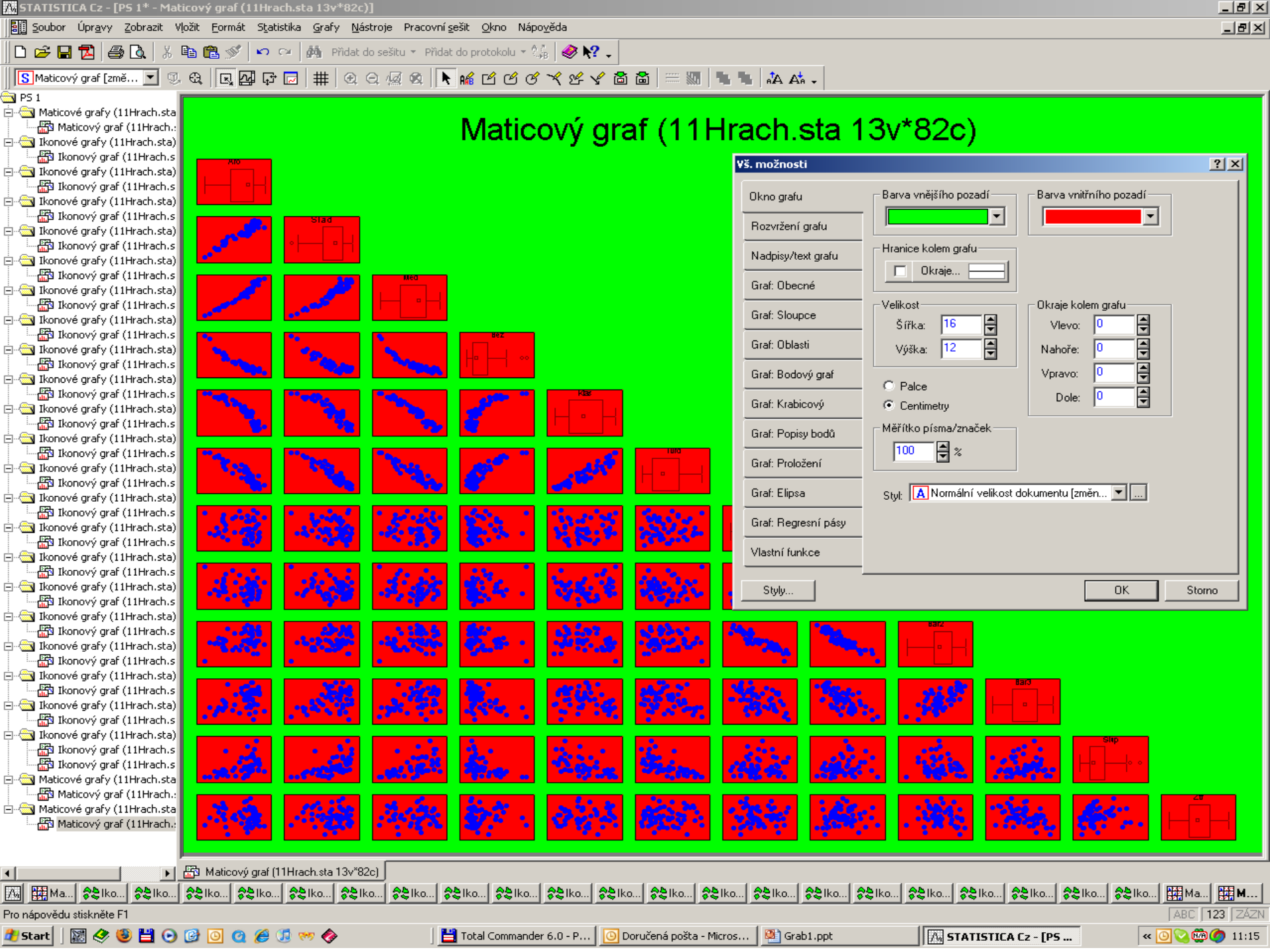

STATISTICA Cz - [PS 1* - Maticový graf (11Hrach.sta 13v*82c)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Maticový graf [změ...
PS 1
Maticové grafy (11Hrach.sta
Maticový graf (11Hrach.
Ikonové grafy (11Hrach.sta)
Ikonový graf (11Hrach.s
Maticový graf (11Hrach.sta 13v*82c)
Změna barvy pozadí
Nadpisy/text grafu
Graf: Obecné
Graf: Sloupce
Graf: Oblasti
Graf: Bodový graf
Graf: Krabicový
Graf: Popisy bodů
Graf: Proložení
Graf: Elipsa
Graf: Regresní pásy
Vlastní funkce
Styly...
Barva vnějšího pozadí
Barva vnitřního pozadí
Hranice kolem grafu
Okraje...
Velikost
Šířka: 24
Výška: 18
Palce
Centimetry
Okraje kolem grafu
Vlevo: 0
Nahoře: 0
Vpravo: 0
Dole: 0
Měřítko písma/značek
100 %
Styl: Normální velikost dokumentu [změn...
OK
Storno
Maticový graf (11Hrach.sta 13v*82c)
Pro nápovědu stiskněte F1
Start
Total Commander 6.0 - P...
Doručená pošta - Micros...
Grab1.ppt
STATISTICA Cz - [PS ...
11:06

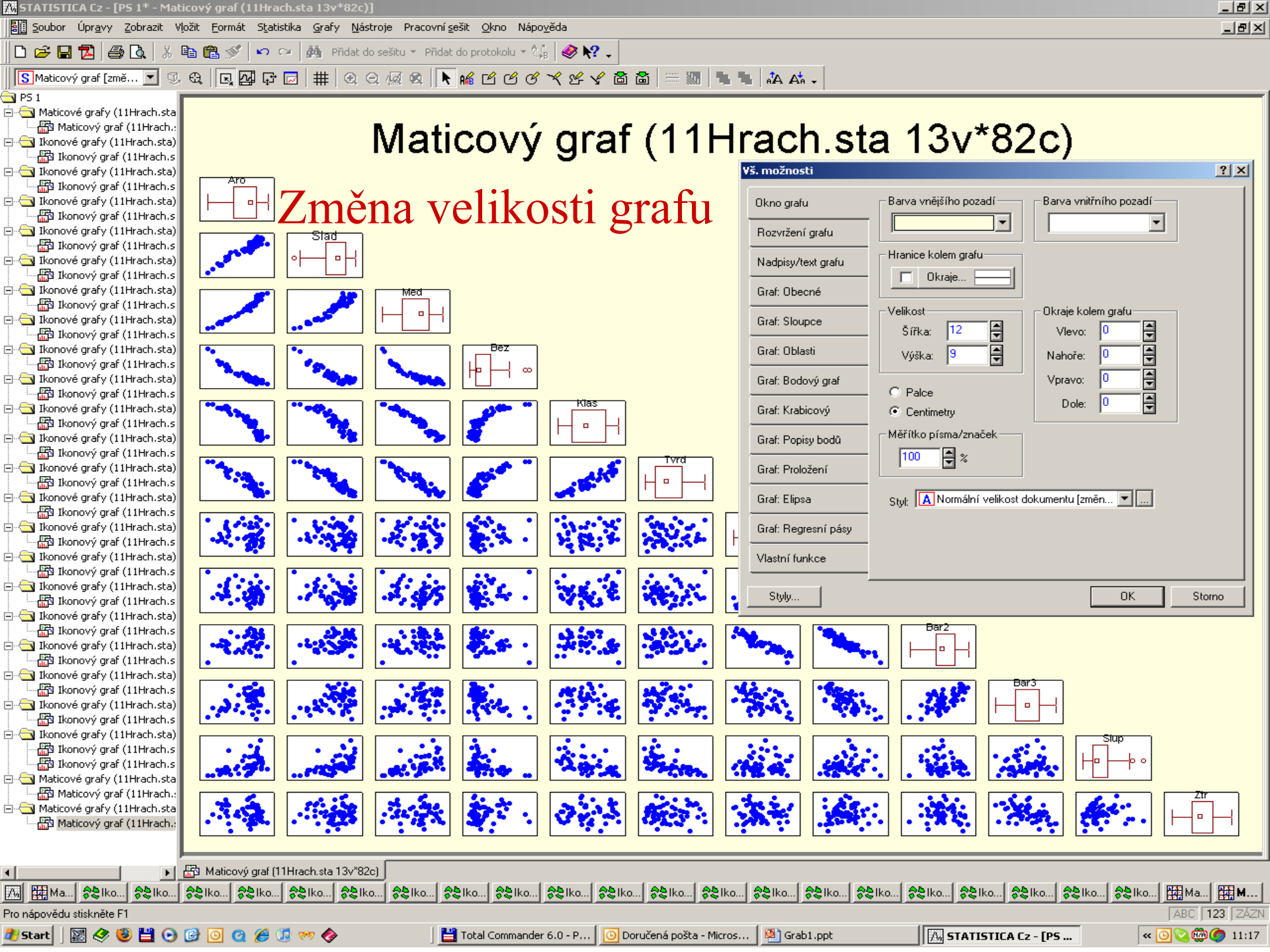

STATISTICA Cz - [PS 1* - Maticový graf (11Hrach.sta 13v*82c)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Maticový graf [změ...
PS 1
Maticové grafy (11Hrach.sta
Maticový graf (11Hrach.
Ikonové grafy (11Hrach.sta)
Ikonový graf (11Hrach.s
Maticový graf (11Hrach.sta 13v*82c)
Změna barvy pozadí a velikosti grafu
Vš. možnosti
Okno grafu
Rozvržení grafu
Nadpisy/text grafu
Graf: Obecné
Graf: Sloupce
Graf: Oblasti
Graf: Bodový graf
Graf: Krabicový
Graf: Popisy bodů
Graf: Proložení
Graf: Elipsa
Graf: Regresní pásy
Vlastní funkce
Barva vnějšího pozadí
Barva vnitřního pozadí
Hranice kolem grafu
Okraje...
Velikost
Šířka: 12
Výška: 9
Okraje kolem grafu
Vlevo: 0
Nahoře: 0
Vpravo: 0
Dole: 0
Palce
Centimetry
Měřítko písma/značek
100 %
Styl: Normální velikost dokumentu [změn...
Styly...
OK
Storno
Bez
Klas
Tvrd
Bar2
Bar3
Stup
Ztr
Maticový graf (11Hrach.sta 13v*82c)
Pro nápovědu stiskněte F1
Start
Total Commander 6.0 - P...
Doručená pošta - Micros...
Grab1.ppt
STATISTICA Cz - [PS ...
11:04

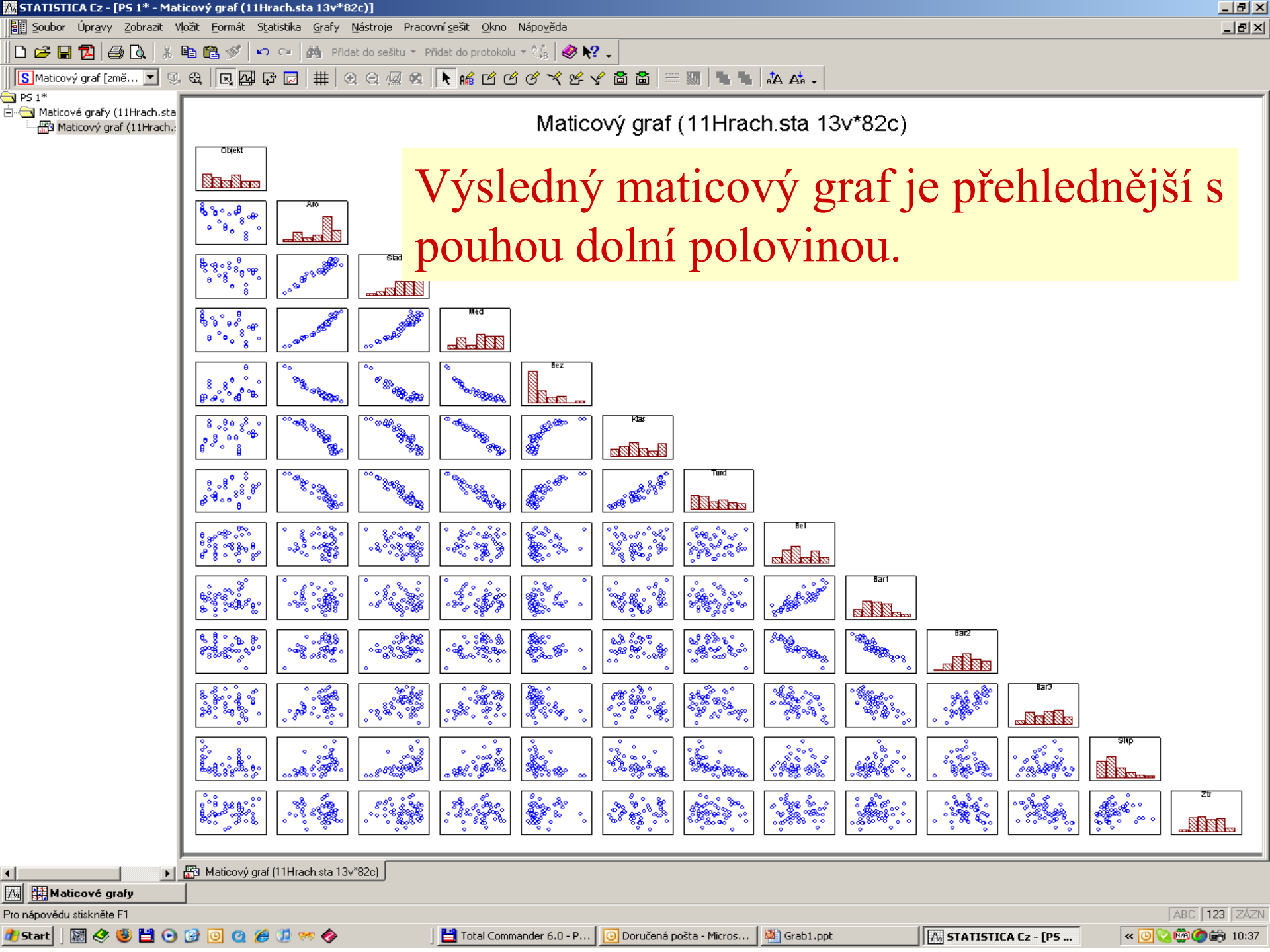

STATISTICA Cz - [PS 1* - Maticový graf (11Hrach.sta 13v*82c)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
PS 1
Maticové grafy (11Hrach.sta
Maticový graf (11Hrach.
Ikonové grafy (11Hrach.sta)
Ikonový graf (11Hrach.s
Maticový graf (11Hrach.sta 13v*82c)
Změna barvy pozadí a velikosti grafu
Barva vnějšího pozadí
Barva vnitřního pozadí
Hranice kolem grafu
Okraje...
Velikost
Šířka: 16
Výška: 12
Okraje kolem grafu
Vlevo: 0
Nahoře: 0
Vpravo: 0
Dole: 0
Palce
Centimetry
Měřítko písma/značek
100 %
Styl: Normální velikost dokumentu [změn...
Graf: Obecné
Graf: Sloupce
Graf: Oblasti
Graf: Bodový graf
Graf: Krabicový
Graf: Popisy bodů
Graf: Proložení
Graf: Elipsa
Graf: Regresní pásy
Vlastní funkce
Styly...
OK
Storno
Aro
Stad
Med
Bez
Klas
Turd
Bar2
Bar3
Slip
Zig
Maticový graf (11Hrach.sta 13v*82c)
Pro nápovědu stiskněte F1
Start
Total Commander 6.0 - P...
Doručená pošta - Micros...
Grab1.ppt
STATISTICA Cz - [PS ...
11:09

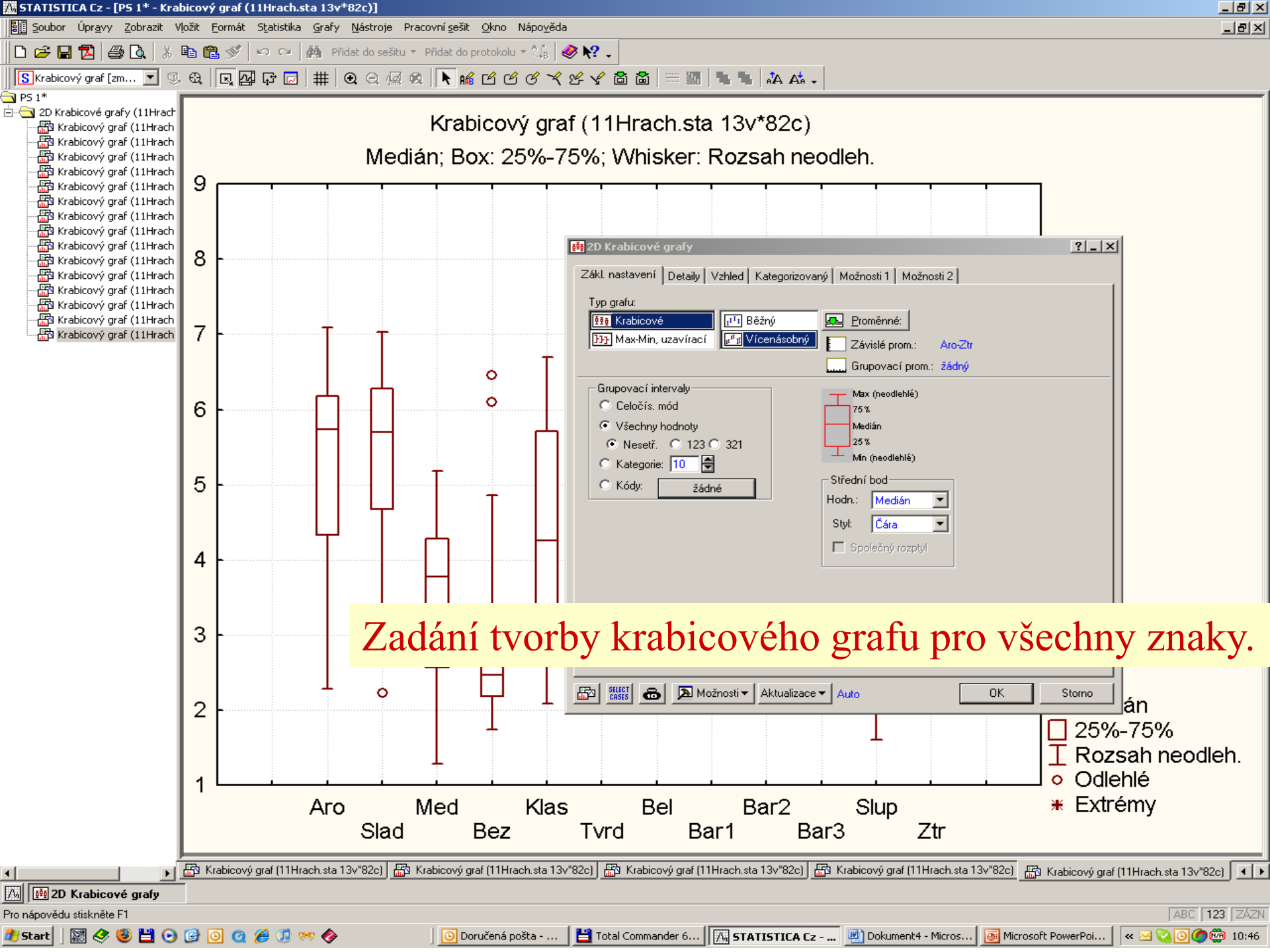

Změna velikosti fontu nadpisu
Maticový graf (11Hrach.sta 13v*82c)
Rozvržení grafu
Nadpisy/text grafu
Graf: Obecné
Graf: Sloupce
Graf: Oblasti
Graf: Bodový graf
Graf: Krabicový
Graf: Popisy bodů
Graf: Proložení
Graf: Elipsa
Graf: Regresní pásy
Vlastní funkce
Arial
14
Maticový graf (11Hrach.sta 13v*82c)
Zobrazit
Odpojit objekt(y) od grafu
Velké zobrazení...
Nadpis: Maticový graf (11Hrach.sta 13v*82c)
Více nadpisů
Nový nadpis
Odstranit
Posunout nahoru
Posunout dolů
Další možnosti
Status:
Nadpis
Orientace:
Automaticky
Vlastnosti textového objektu...
Vertik. zarovn.:
Dole
Symbol legendy s:
Široký
OK
Storno
Nadpis
Odpojit objekt(y) od grafu
Zpět
Maticový graf (11Hrach.sta 13v*82c)
Maticový graf (11Hrach.sta 13v*82c)
PS 1
Pro nápovědu stiskněte F1

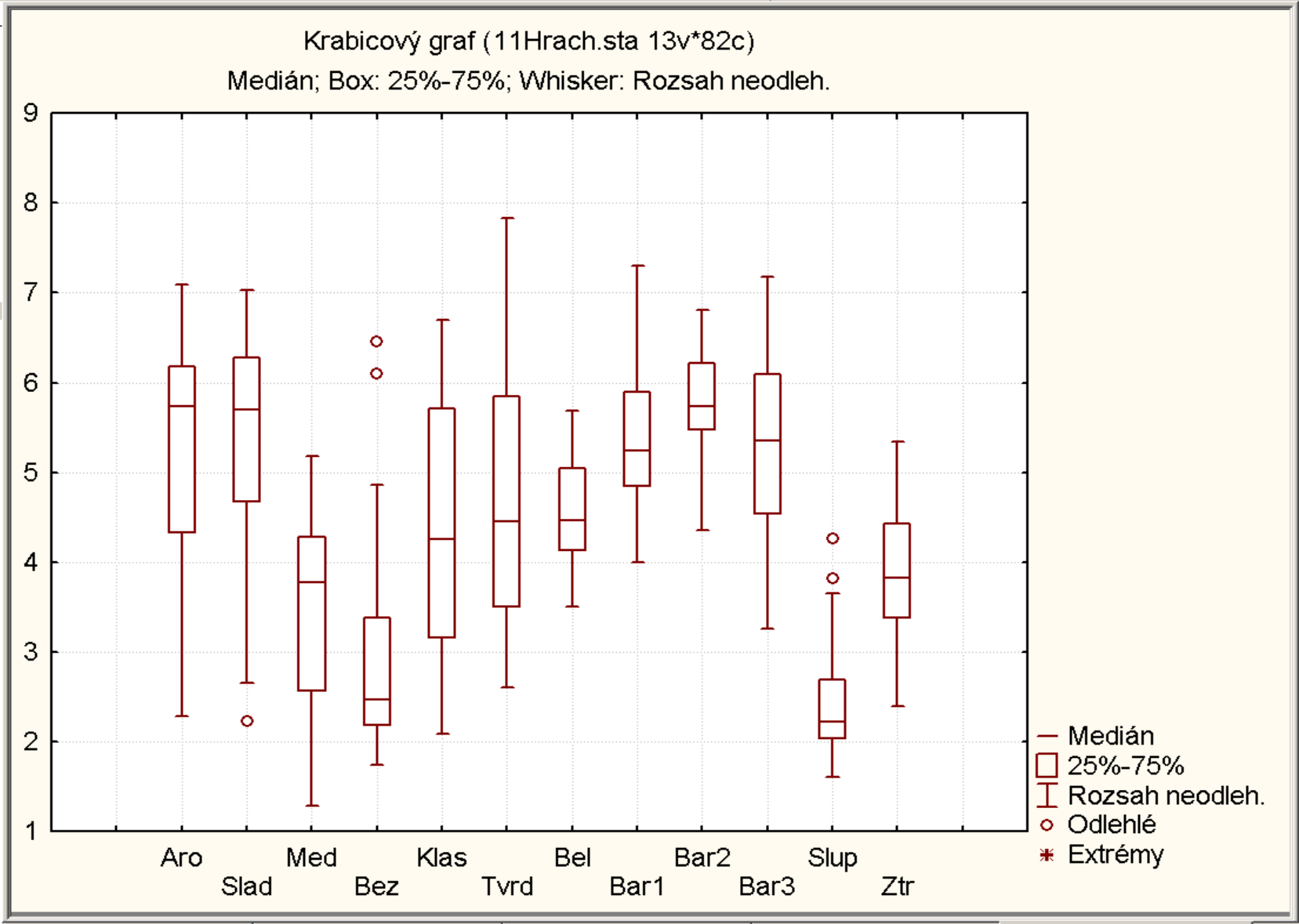

STATISTICA Cz - [PS 1* - Maticový graf (11Hrach.sta 13v*82c)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Maticový graf (11Hrach.sta 13v*82c)
Změna barvy bodů a volba bodů zobrazovaných
Vlastnosti značky
Velikost značky:
Zavřít
bodů
Značka jako te:
Vzor značky:
Barva popředí:
Barva pozadí:
Styl značky:
Značka 1. grafu [změněno]
Text. značky
Arial
Graf: 1: Bodový graf
Zobrazit graf
Přidat nový graf...
Ignorovat chybějící data (spojnicové grafy)
Odstranit...
Značky...
Obrázkové značky...
Spojnice...
Typ spojnice:
Běžný
Oblast...
Voronoiovy čáry...
Graf: Obecné
Graf: Sloupce
Graf: Oblasti
Graf: Bodový graf
Graf: Krabicový
Graf: Popisy bodů
Graf: Proložení
Graf: Elipsa
Graf: Regresní pásy
Vlastní funkce
Přiřazení osy
Osa Y
Vlevo
Vpravo
Volitelné:
Y vlevo
Název:
Legenda:
Vypnutá
Styl:
Atributy 1. grafu [změněno]
Styly...
OK
Storno
Maticové grafy (11Hrach.sta
Ikonové grafy (11Hrach.sta)
Ikonový graf (11Hrach.s
Maticový graf (11Hrach.sta 13v*82c)
Pro nápovědu stiskněte F1
Start
Total Commander 6.0 - P...
Doručená pošta - Micros...
Grab1.ppt
STATISTICA Cz - [PS ...
11:12

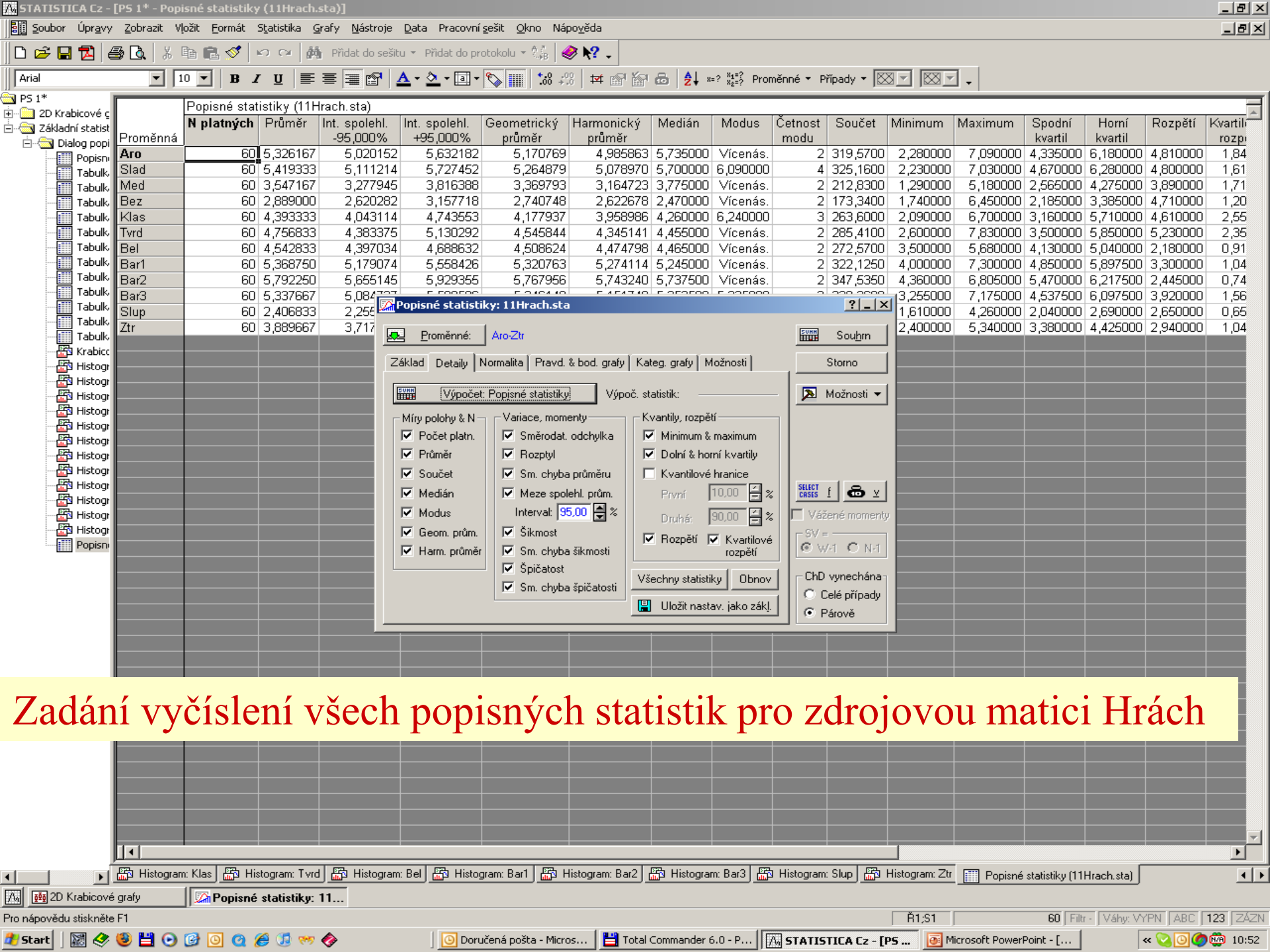

STATISTICA Cz - [PS 1* - Maticový graf (11Hrach.sta 13v*82c)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Maticový graf [změ...
PS 1
Maticové grafy (11Hrach.sta
Maticový graf (11Hrach.
Ikonové grafy (11Hrach.sta)
Ikonový graf (11Hrach.s
Maticový graf (11Hrach.sta 13v*82c)
Vš. možnosti
Okno grafu
Rozvržení grafu
Nadpisy/text grafu
Graf: Obecné
Graf: Sloupce
Graf: Oblasti
Graf: Bodový graf
Graf: Krabicový
Graf: Popisy bodů
Graf: Proložení
Graf: Elipsa
Graf: Regresní pásy
Vlastní funkce
Styly...
Barva vnějšího pozadí
Barva vnitřního pozadí
Hranice kolem grafu
Okraje...
Velikost
Šířka: 16
Výška: 12
Okraje kolem grafu
Vlevo: 0
Nahoře: 0
Vpravo: 0
Dole: 0
Palce
Centimetry
Měřítko písma/značek
100 %
Styl: Normální velikost dokumentu [změn...
OK
Storno
Pro nápovědu stiskněte F1
ABC 123 ZÁZN
Start
Total Commander 6.0 - P...
Doručená pošta - Micros...
Grab1.ppt
STATISTICA Cz - [PS ...
11:15


STATISTICA Cz - [PS 1* - Maticový graf (11Hrach.sta 13v*82c)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Maticový graf (11Hrach.sta 13v*82c)
Změna velikosti grafu
Vš. možnosti
Okno grafu
Rozvržení grafu
Nadpisy/text grafu
Graf: Obecné
Graf: Sloupce
Graf: Oblasti
Graf: Bodový graf
Graf: Krabicový
Graf: Popisy bodů
Graf: Proložení
Graf: Elipsa
Graf: Regresní pásy
Vlastní funkce
Barva vnějšího pozadí
Barva vnitřního pozadí
Hranice kolem grafu
Okraje...
Velikost
Šířka: 12
Výška: 9
Palce
Centimetry
Okraje kolem grafu
Vlevo: 0
Nahoře: 0
Vpravo: 0
Dole: 0
Měřítko písma/značek
100 %
Styl: Normální velikost dokumentu [změn...
Styly...
OK
Storno
Aro
Stad
Med
Bez
Klas
Tvrd
Bar2
Bar3
Stup
Ztr
PS 1
Maticový graf (11Hrach.sta 13v*82c)
Pro nápovědu stiskněte F1
Start
Total Commander 6.0 - P...
Doručená pošta - Micros...
Grab1.ppt
STATISTICA Cz - [PS ...
11:17


STATISTICA Cz - [PS 1* - Maticový graf (11Hrach.sta 13v*82c)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Maticový graf [změ...
PS 1*
Maticové grafy (11Hrach.sta
Maticový graf (11Hrach.:
Maticový graf (11Hrach.sta 13v*82c)
Výsledný maticový graf je přehlednější s pouhou dolní polovinou.
Objekt
Aro
Stab
Med
Bez
Kla
Turd
Bel
Bar1
Bar2
Bar3
Slup
Ztr
Maticový graf (11Hrach.sta 13v*82c)
Maticové grafy
Pro nápovědu stiskněte F1
ABC
123
ZÁZN
Start
Total Commander 6.0 - P...
Doručená pošta - Micros...
Grab1.ppt
STATISTICA Cz - [PS ...
10:37


STATISTICA Cz - [PS 1* - Krabicový graf (11Hrach.sta 13v*82c)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
PS 1*
2D Krabicové grafy (11Hrach
Krabicový graf (11Hrach
Krabicový graf (11Hrach.sta 13v*82c)
Medián; Box: 25%-75%; Whisker: Rozsah neodleh.
9
8
7
6
5
4
3
2
1
Aro
Slad
Med
Bez
Klas
Tvrd
Bel
Bar1
Bar2
Bar3
Slup
Ztr
2D Krabicové grafy
Zákl. nastavení
Detaily
Vzhled
Kategorizovaný
Možnosti 1
Možnosti 2
Typ grafu:
Krabicové
Běžný
Max-Min, uzavírací
Vícenásobný
Proměnné:
Závislé prom.: Aro-Ztr
Grupovací prom.: žádný
Grupovací intervaly
Celočís. mód
Všechny hodnoty
Nesetř. 123 321
Kategorie: 10
Kódy: žádné
Max (neodlehlé)
75 %
Medián
25 %
Min (neodlehlé)
Střední bod
Hodn.: Medián
Styl: Čára
Společný rozptyl
Možnosti
Aktualizace
Auto
OK
Storno
25%-75%
Rozsah neodleh.
Odlehlé
Extrémy
Zadání tvorby krabicového grafu pro všechny znaky.
Pro nápovědu stiskněte F1
Start
Doručená pošta - ...
Total Commander 6...
STATISTICA Cz - ...
Dokument4 - Micros...
Microsoft PowerPoi...
10:46




STATISTICA Cz - [PS 1* - Popisné statistiky (11Hrach.sta)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda
Popisné statistiky (11Hrach.sta)
Proměnná | N platných | Průměr | Int. spolehl. -95,000% | Int. spolehl. +95,000% | Geometrický průměr | Harmonický průměr | Medián | Modus | Četnost modu | Součet | Minimum | Maximum | Spodní kvartil | Horní kvartil | Rozpětí | Kvartil rozp
Aro | 60 | 5,326167 | 5,020152 | 5,632182 | 5,170769 | 4,985863 | 5,735000 | Vícenás. | 2 | 319,5700 | 2,280000 | 7,090000 | 4,335000 | 6,180000 | 4,810000 | 1,84
Slad | 60 | 5,419333 | 5,111214 | 5,727452 | 5,264879 | 5,078970 | 5,700000 | 6,090000 | 4 | 325,1600 | 2,230000 | 7,030000 | 4,670000 | 6,280000 | 4,800000 | 1,61
Med | 60 | 3,547167 | 3,277945 | 3,816388 | 3,369793 | 3,164723 | 3,775000 | Vícenás. | 2 | 212,8300 | 1,290000 | 5,180000 | 2,565000 | 4,275000 | 3,890000 | 1,71
Bez | 60 | 2,889000 | 2,620282 | 3,157718 | 2,740748 | 2,622678 | 2,470000 | Vícenás. | 2 | 173,3400 | 1,740000 | 6,450000 | 2,185000 | 3,385000 | 4,710000 | 1,20
Klas | 60 | 4,393333 | 4,043114 | 4,743553 | 4,177937 | 3,958986 | 4,260000 | 6,240000 | 3 | 263,6000 | 2,090000 | 6,700000 | 3,160000 | 5,710000 | 4,610000 | 2,55
Tvrd | 60 | 4,756833 | 4,383375 | 5,130292 | 4,545844 | 4,345141 | 4,455000 | Vícenás. | 2 | 285,4100 | 2,600000 | 7,830000 | 3,500000 | 5,850000 | 5,230000 | 2,35
Bel | 60 | 4,542833 | 4,397034 | 4,688632 | 4,508624 | 4,474798 | 4,465000 | Vícenás. | 2 | 272,5700 | 3,500000 | 5,680000 | 4,130000 | 5,040000 | 2,180000 | 0,91
Bar1 | 60 | 5,368750 | 5,179074 | 5,558426 | 5,320763 | 5,274114 | 5,245000 | Vícenás. | 2 | 322,1250 | 4,000000 | 7,300000 | 4,850000 | 5,897500 | 3,300000 | 1,04
Bar2 | 60 | 5,792250 | 5,655145 | 5,929355 | 5,767956 | 5,743240 | 5,737500 | Vícenás. | 2 | 347,5350 | 4,360000 | 6,805000 | 5,470000 | 6,217500 | 2,445000 | 0,74
Bar3 | 60 | 5,337667 | 5,084737 | | | | | | | | 3,255000 | 7,175000 | 4,537500 | 6,097500 | 3,920000 | 1,56
Slup | 60 | 2,406833 | 2,255 | | | | | | | | 1,610000 | 4,260000 | 2,040000 | 2,690000 | 2,650000 | 0,65
Ztr | 60 | 3,889667 | 3,717 | | | | | | | | 2,400000 | 5,340000 | 3,380000 | 4,425000 | 2,940000 | 1,04
Popisné statistiky: 11Hrach.sta
Proměnné: Aro-Ztr
Souhrn
Storno
Možnosti
Základ | Detaily | Normalita | Pravd. & bod. grafy | Kateg. grafy | Možnosti
Výpočet: Popisné statistiky
Výpoč. statistik:
Míry polohy & N
Počet platn.
Průměr
Součet
Medián
Modus
Geom. prům.
Harm. průměr
Variace, momenty
Směrodat. odchylka
Rozptyl
Sm. chyba průměru
Meze spolehl. prům.
Interval: 95,00 %
Šikmost
Sm. chyba šikmosti
Špičatost
Sm. chyba špičatosti
Kvantily, rozpětí
Minimum & maximum
Dolní & horní kvartily
Kvantilové hranice
První 10,00 %
Druhá: 90,00 %
Rozpětí
Kvartilové rozpětí
Všechny statistiky
Obnov
Uložit nastav. jako zákl.
Vážené momenty
SV = W-1 N-1
ChD vynechána
Celé případy
Párově
Zadání vyčíslení všech popisných statistik pro zdrojovou matici Hrách
Histogram: Klas
Histogram: Tvrd
Histogram: Bel
Histogram: Bar1
Histogram: Bar2
Histogram: Bar3
Histogram: Slup
Histogram: Ztr
Popisné statistiky (11Hrach.sta)
Pro nápovědu stiskněte F1
Start
Doručená pošta - Micros...
Total Commander 6.0 - P...
STATISTICA Cz - [PS ...
Microsoft PowerPoint - [...
10:52

STATISTICA Cz - [PS 1* - Popisné statistiky (11Hrach.sta)]

Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Data Pracovní sešit Okno Nápověda

Přidat do sešitu Přidat do protokolu

Arial 10 Proměnné Případy

PS 1*
- 2D Krabicové g
- Základní statist
  - Dialog popi
    - Popisn
    - Tabulk
    - Tabulk
    - Tabulk
    - Tabulk
    - Tabulk
    - Tabulk
    - Tabulk
    - Tabulk
    - Tabulk
    - Tabulk
    - Tabulk
    - Krabicc
    - Histogr
    - Histogr
    - Histogr
    - Histogr
    - Histogr
    - Histogr
    - Histogr
    - Histogr
    - Histogr
    - Histogr
    - Histogr
    - Popisn

Popisné statistiky (11Hrach.sta)

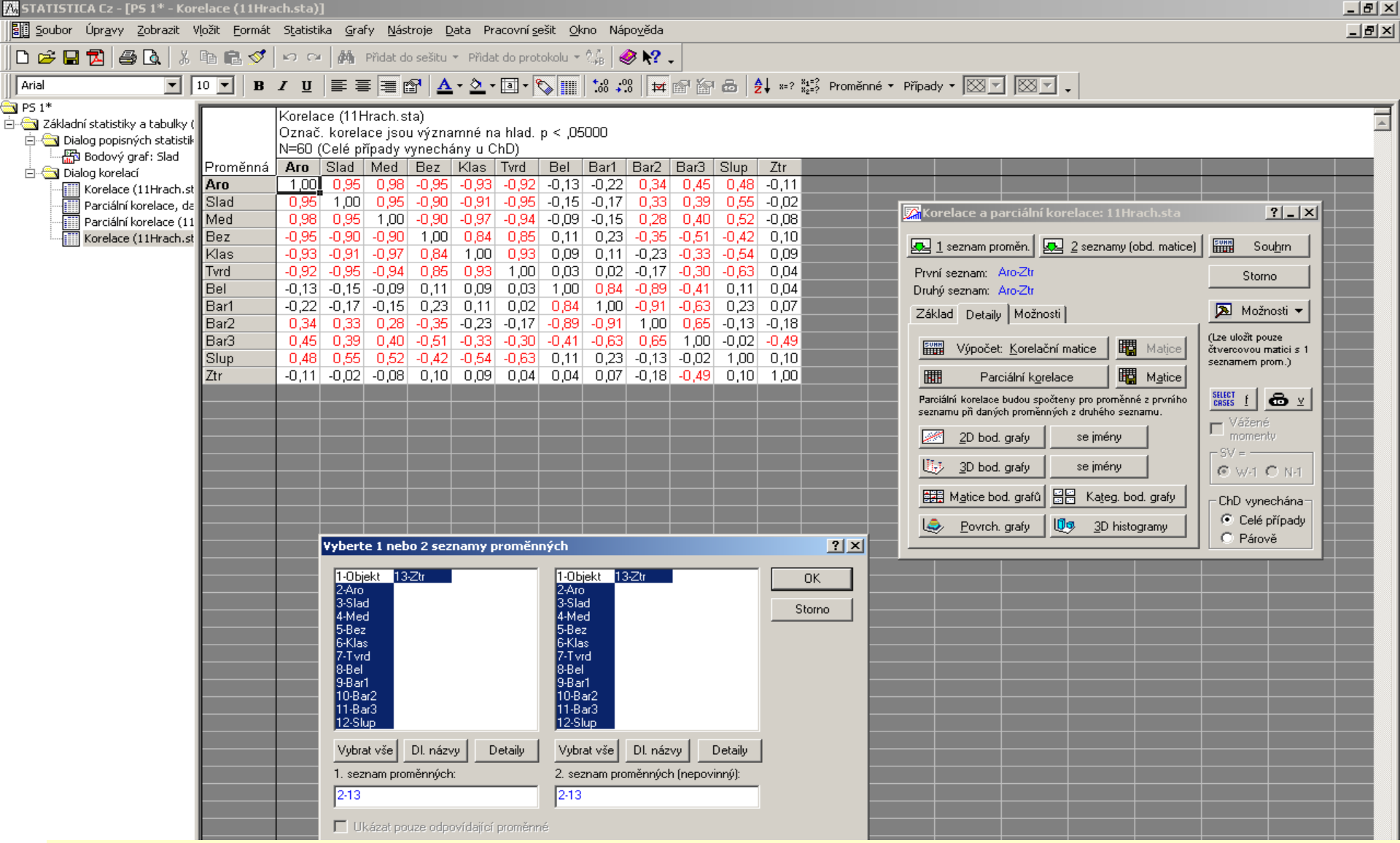
Proměnná	N platných	Průměr	Int. spolehl. -95,000%	Int. spolehl. +95,000%	Geometrický průměr	Harmonický průměr	Medián	Modus	Četnost modu	Součet	Minimum	Maximum	Spodní kvartil	Horní kvartil	Rozpětí	Kvartil rozp
Aro	60	5,326167	5,020152	5,632182	5,170769	4,985863	5,735000	Vícenás.	2	319,5700	2,280000	7,090000	4,335000	6,180000	4,810000	1,84
Slad	60	5,419333	5,111214	5,727452	5,264879	5,078970	5,700000	6,090000	4	325,1600	2,230000	7,030000	4,670000	6,280000	4,800000	1,61
Med	60	3,547167	3,277945	3,816388	3,369793	3,164723	3,775000	Vícenás.	2	212,8300	1,290000	5,180000	2,565000	4,275000	3,890000	1,71
Bez	60	2,889000	2,620282	3,157718	2,740748	2,622678	2,470000	Vícenás.	2	173,3400	1,740000	6,450000	2,185000	3,385000	4,710000	1,20
Klas	60	4,393333	4,043114	4,743553	4,177937	3,958986	4,260000	6,240000	3	263,6000	2,090000	6,700000	3,160000	5,710000	4,610000	2,55
Tvrd	60	4,756833	4,383375	5,130292	4,545844	4,345141	4,455000	Vícenás.	2	285,4100	2,600000	7,830000	3,500000	5,850000	5,230000	2,35
Bel	60	4,542833	4,397034	4,688632	4,508624	4,474798	4,465000	Vícenás.	2	272,5700	3,500000	5,680000	4,130000	5,040000	2,180000	0,91
Bar1	60	5,368750	5,179074	5,558426	5,320763	5,274114	5,245000	Vícenás.	2	322,1250	4,000000	7,300000	4,850000	5,897500	3,300000	1,04
Bar2	60	5,792250	5,655145	5,929355	5,767956	5,743240	5,737500	Vícenás.	2	347,5350	4,360000	6,805000	5,470000	6,217500	2,445000	0,74
Bar3	60	5,337667	5,084737	5,590596	5,246149	5,151749	5,352500	5,325000	2	320,2600	3,255000	7,175000	4,537500	6,097500	3,920000	1,56
Slup	60	2,406833	2,255373	2,558294	2,344012	2,287747	2,225000	Vícenás.	2	144,4100	1,610000	4,260000	2,040000	2,690000	2,650000	0,65
Ztr	60	3,889667	3,717788	4,061546	3,832079	3,772781	3,825000	Vícenás.	2	233,3800	2,400000	5,340000	3,380000	4,425000	2,940000	1,04

Histogram: Klas | Histogram: Tvrd | Histogram: Bel | Histogram: Bar1 | Histogram: Bar2 | Histogram: Bar3 | Histogram: Slup | Histogram: Ztr | Popisné statistiky (11Hrach.sta)

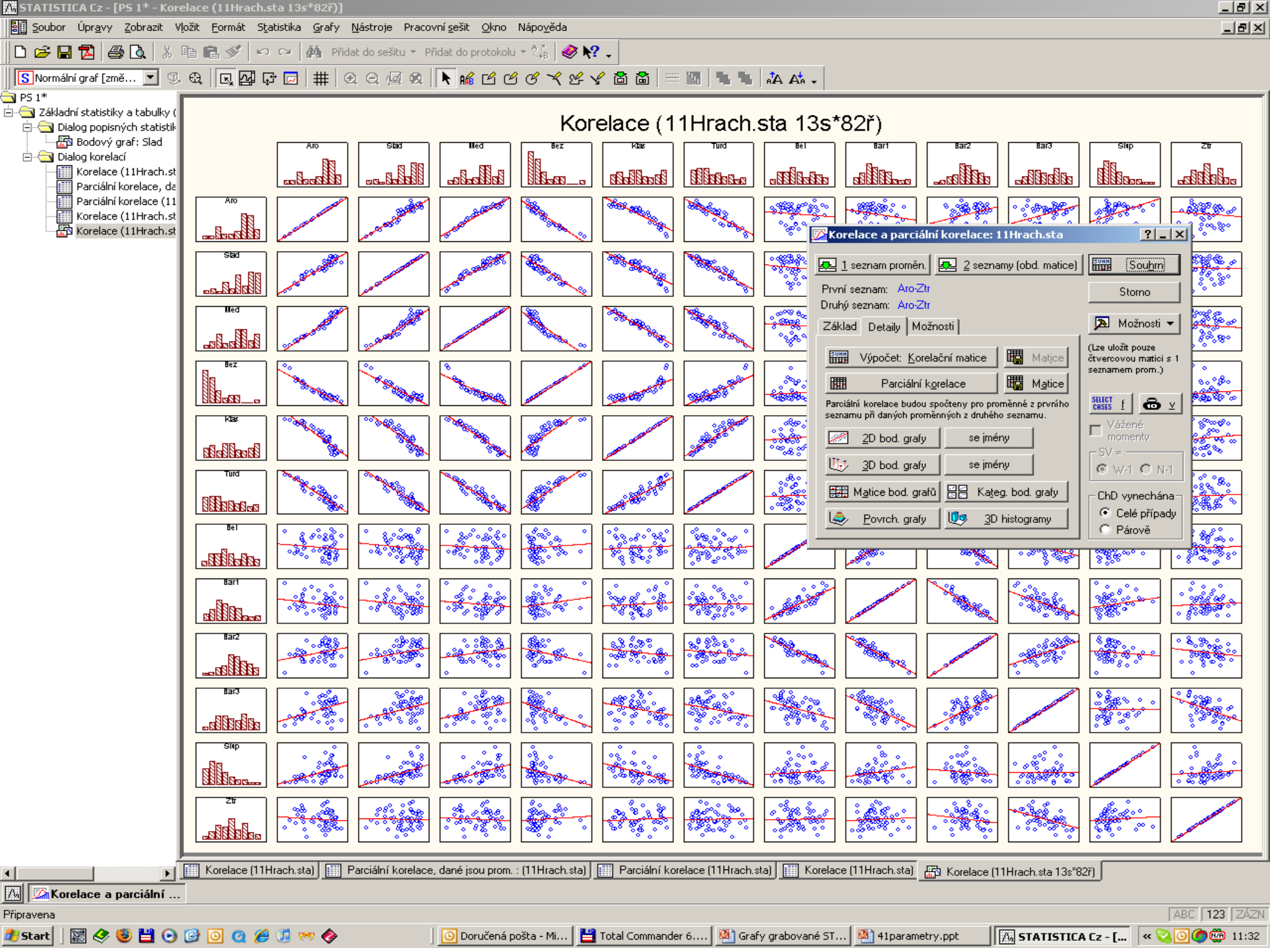
2D Krabicové grafy | Popisné statistiky: 11...

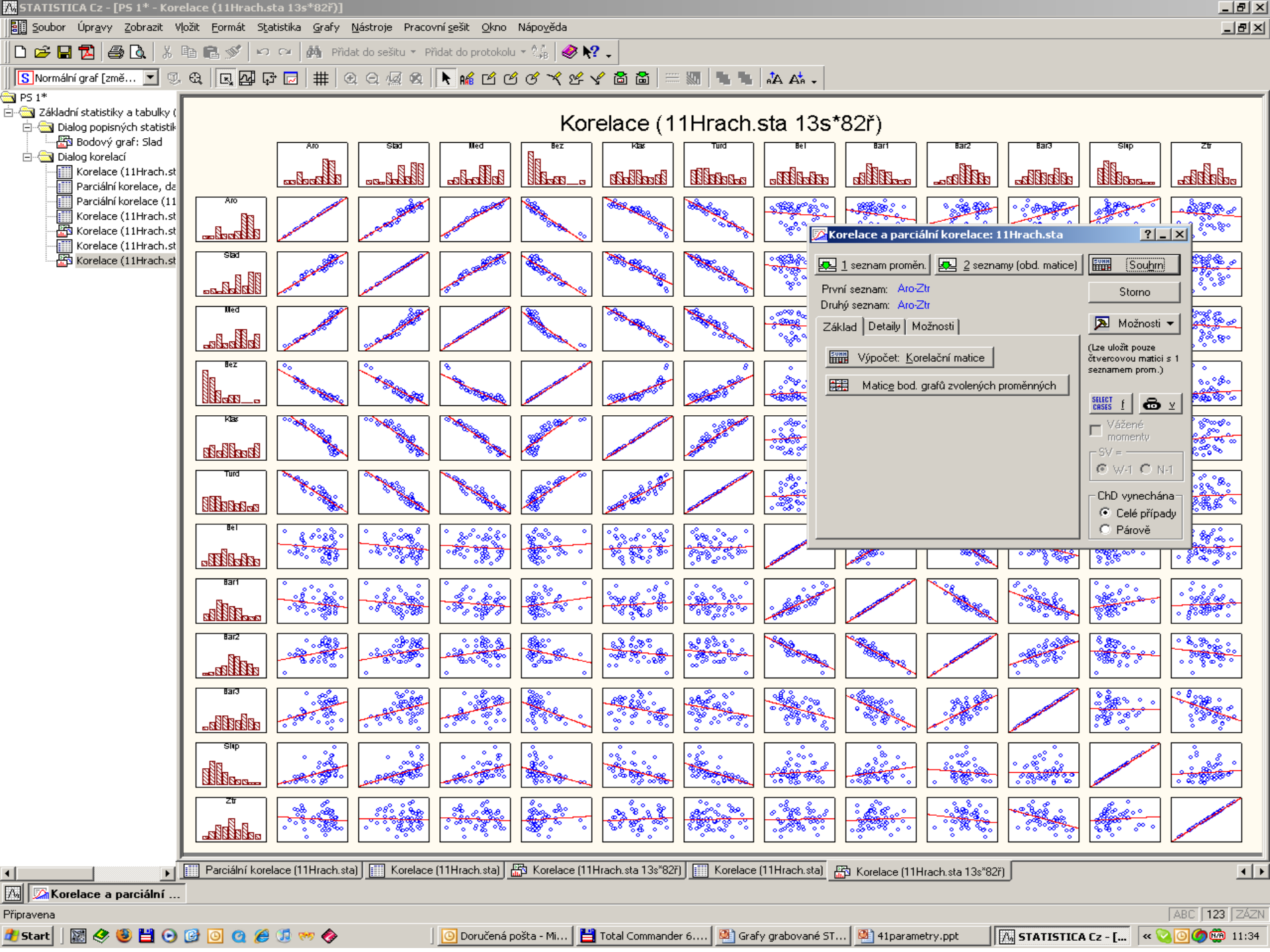
Pro nápovědu stiskněte F1 | R1;S1 | 60 | ABC | 123 | ZÁZN

Start | Doručená pošta - Micros... | Total Commander 6.0 - P... | STATISTICA Cz - [PS ... | Microsoft PowerPoint - [... | 10:53



Zadání testu významnosti jednotlivých korelačních koeficientů u vyčíslení korelační matice.


STATISTICA Cz - [PS 1* - Korelace (11Hrach.sta 13s*82ř)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Základní statistiky a tabulky (
Dialog popisných statistik
Bodový graf: Slad
Dialog korelací
Korelace (11Hrach.st
Parciální korelace, da
Parciální korelace (11
Korelace (11Hrach.st
Korelace (11Hrach.st
Korelace (11Hrach.sta 13s*82ř)
Aro
Slad
Med
Bez
Klas
Turd
Bel
Bar1
Bar2
Bar3
Slup
Ztr
Korelace a parciální korelace: 11Hrach.sta
1 seznam proměn.
2 seznamy (obd. matice)
Souhrn
První seznam: Aro-Ztr
Druhý seznam: Aro-Ztr
Storno
Základ
Detaily
Možnosti
Možnosti
(Lze uložit pouze čtvercovou matici s 1 seznamem prom.)
Výpočet: Korelační matice
Matice
Parciální korelace
Matice
Parciální korelace budou spočteny pro proměnné z prvního seznamu při daných proměnných z druhého seznamu.
2D bod. grafy
se jmény
3D bod. grafy
se jmény
Matice bod. grafů
Kateg. bod. grafy
Povrch. grafy
3D histogramy
Vážené momenty
SV =
W-1
N-1
ChD vynechána
Celé případy
Párově
Korelace (11Hrach.sta)
Parciální korelace, dané jsou prom. : (11Hrach.sta)
Parciální korelace (11Hrach.sta)
Korelace (11Hrach.sta)
Korelace (11Hrach.sta 13s*82ř)
Korelace a parciální ...
Připravena
ABC
123
ZÁZN
Start
Doručená pošta - Mi...
Total Commander 6....
Grafy grabované ST...
41parametry.ppt
STATISTICA Cz - [...
11:32


STATISTICA Cz - [PS 1* - Korelace (11Hrach.sta 13s*82ř)]
Soubor Úpravy Zobrazit Vložit Formát Statistika Grafy Nástroje Pracovní sešit Okno Nápověda
Přidat do sešitu
Přidat do protokolu
Normální graf [změ...
PS 1*
Základní statistiky a tabulky (
Dialog popisných statistik
Bodový graf: Slad
Dialog korelací
Korelace (11Hrach.st
Parciální korelace, da
Parciální korelace (11
Korelace (11Hrach.st
Korelace (11Hrach.st
Korelace (11Hrach.st
Korelace (11Hrach.st
Korelace (11Hrach.sta 13s*82ř)
Aro
Slad
Med
Bez
Klas
Turd
Bel
Bar1
Bar2
Bar3
Slup
Ztr
Korelace a parciální korelace: 11Hrach.sta
1 seznam proměn.
2 seznamy (obd. matice)
Souhrn
Storno
První seznam: Aro-Ztr
Druhý seznam: Aro-Ztr
Možnosti
Základ
Detaily
Možnosti
Výpočet: Korelační matice
Matice bod. grafů zvolených proměnných
(Lze uložit pouze čtvercovou matici s 1 seznamem prom.)
SELECT CASES
f
v
Vážené momenty
SV =
W-1
N-1
ChD vynechána
Celé případy
Párově
Parciální korelace (11Hrach.sta)
Korelace (11Hrach.sta)
Korelace (11Hrach.sta 13s*82ř)
Korelace (11Hrach.sta)
Korelace (11Hrach.sta 13s*82ř)
Korelace a parciální ...
Připravena
ABC
123
ZÁZN
Start
Doručená pošta - Mi...
Total Commander 6....
Grafy grabované ST...
41parametry.ppt
STATISTICA Cz - [...
11:34

Statistická analýza vektoru středních hodnot

A. Testování nulové hypotézy H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ v závislosti na alternativní H_1: $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$.

Data $\boldsymbol{X}$ jsou náhodným výběrem velikosti n z m-rozměrného normálního rozdělení $N(\boldsymbol{\mu}, \boldsymbol{C})$. Parametry $\boldsymbol{\mu}$ a $\boldsymbol{C}$ jsou neznámé a odhadují se pomocí výběrových charakteristik ji, $\boldsymbol{S}$.

K testování se používá Hotellingovy T^2 -statistiky

$$T^2 = n(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)^T \boldsymbol{S}^{-1} (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)$$

Testování

1. Je-li $T^2 \leq T^2(1-\frac{\alpha}{2})$, je na hladině významnosti α hypotéza $H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$ přijata.
2. Je-li $T^2 > T^2(1-\frac{\alpha}{2})$, hypotéza $H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$ se zamítá.

Při platnosti H_0 má veličina $C = (n - m)\, T^2 / (m\,(n - 1))$ *F*-rozdělení *s* m a $n - m$ stupni volnosti. Pokud je H_0 neplatná, má veličina *C* necentrální *F*-rozdělení. Pomocí veličiny *C* lze testovat hypotézu H_0 *F*-testem.

Testování

Využitím T^2-statistiky lze konstruovat konfidenční oblasti pro vektor $\boldsymbol{\mu}$. Platí totiž, že

$100(1-\alpha)\%$ oblast m-rozměrného vektoru je ohraničena povrchem elipsoidu ve tvaru

$$(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{S}^{-1}(\hat{\boldsymbol{\mu}}-\boldsymbol{\mu})=\frac{m(n-1)}{n(n-m)}F_{m,n-m}(1-\alpha),$$

kde $F_{m,n-m}(1-\alpha)$ je kvantil *F*-rozdělení s *m* a *n* - *m* stupni volnosti a rovnicí je definován *m*-rozměrný elipsoid se středem v místě $\hat{\boldsymbol{\mu}}$.

PŘIKLAD 4.5 Magnetizační vlastnosti ocelí (učebnice)

Pro 10 náhodně vybraných taveb železa (výběr V_1) byla zkoumána magnetická indukce x_1 [T] a koercitivní síly x_2 [A . m^{-1}] ocelí. Účelem je testovat nulovou hypotézu H_0: $\boldsymbol{\mu}$ = (1.75 70) a zkonstruovat 95% oblast spolehlivosti vektoru středních hodnot $\boldsymbol{\mu}$.

Data: Výběr V_1

i		1	2	3	4	5	6	7	8	9	10
x_1	[T]	1.788	1.710	1.843	1.725	1.740	1.731	1.780	1.746	1.828	1.796
x_2	[A . m^{-1}]	85.1	69.0	84.0	58.1	69.8	50.9	52.2	53.8	83.5	61.8

Řešení

Pro odhad vektoru středních hodnot platí $\boldsymbol{\mu} = (1.769, 66.82)^T$ a výběrová kovarianční matice a k ní matice inverzní mají tvar

$$\boldsymbol{S} = \begin{bmatrix} 0.002 & 0.376 \\ 0.376 & 184.3 \end{bmatrix}, \qquad \boldsymbol{S}^{-1} = \begin{bmatrix} 781.6 & -1.592 \\ -1.592 & 0.0087 \end{bmatrix}$$

Protože je $\boldsymbol{\mu}_0 = (1.75, 70)^T$, lze dosazením do rovnice vyčíslit Hotellingovu T^2-statistiku

$$T^2 = 10(0.019, -3.18) \begin{bmatrix} 781.6 & -1.592 \\ -1.592 & 0.0087 \end{bmatrix} \begin{bmatrix} 0.019 \\ -3.18 \end{bmatrix} = 5.503.$$

Odpovídající hodnota je C = 9 · 5.503/18 = 2.446. Kvantil *F*-rozdělení $F_{2,9}(0.95) = 4.459$. Protože je *C* menší než tento kvantil, hypotéza H_0 je přijata na hladině významnosti $\alpha = 0.05$.

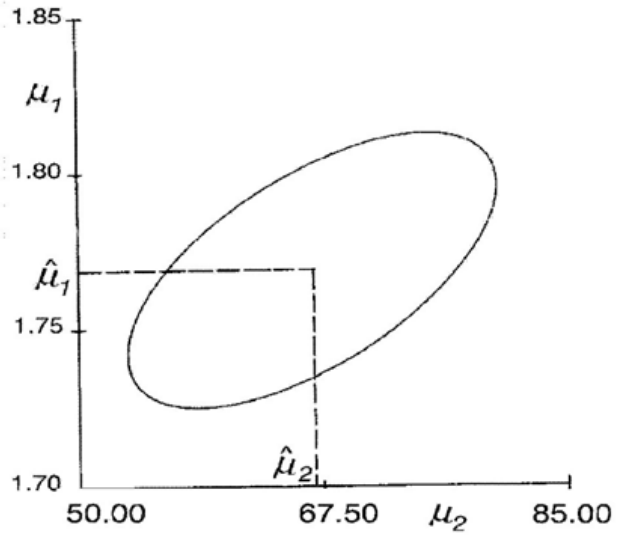
Závěr

Při konstrukci konfidenčního elipsoidu dostaneme

$$(1.769 - \mu_1,\quad 66.82 - \mu_2)\begin{bmatrix} 781.6 & -1.592 \\ -1.592 & 0.0087 \end{bmatrix}\begin{bmatrix} 1.769 - \mu_1 \\ 66.82 - \mu_2 \end{bmatrix} =$$

$$= (1.769 - \mu_1,\quad 66.82 - \mu_2)\begin{bmatrix} 781.6(1.769 - \mu_1) & -1.592(66.82 - \mu_2) \\ -1.592(1.769 - \mu_1) & 0.0087(66.82 - \mu_2) \end{bmatrix},$$

a po úpravách $2201.733 - 2665.56\,\mu_1 + 816.207\,\mu_1^2 + 4.66\,\mu_2 + 0.00901\,\mu_2^2 = 0$.



Vzhledem k proměnným μ_1, μ_2 jde o rovnici elipsy, která je znázorněna na obrázku.

- *Závěr:* Vektor středních hodnot se výrazně neliší od zadaných hodnot $\mu_1 = 1.75$ a $\mu_2 = 70$. Mezi znaky je výraznější kladná korelace, což je patrné i z tvaru konfidenčního elipsoidu.

B. Test shody dvou vícerozměrných středních hodnot $\boldsymbol{\mu_1}$ a $\boldsymbol{\mu_2}$ čili $H_0: \boldsymbol{\mu_1} = \boldsymbol{\mu_2}$ a $H_1: \boldsymbol{\mu_1} \neq \boldsymbol{\mu_2}$.

Vychází se ze dvou náhodných výběrů $\boldsymbol{X}_1$ a $\boldsymbol{X}_2$ vícerozměrných normálních rozdělení $N(\boldsymbol{\mu}_1, \boldsymbol{C}_1)$ a $N(\boldsymbol{\mu}_2, \boldsymbol{C}_2)$.

Testování závisí na tom, zda kovarianční matice $\boldsymbol{C}_1$ a $\boldsymbol{C}_2$ jsou shodné či nikoli:

1. Obě rozdělení mají shodnou kovarianční matici $\boldsymbol{C}_1 = \boldsymbol{C}_2 = \boldsymbol{C}$

Za odhad matice $\boldsymbol{C}$ se užije *společná* (pooled) *výběrová kovarianční matice* $\boldsymbol{S}_p$ dle $\boldsymbol{S}_p = \frac{(n_1-1)\boldsymbol{S}_1+(n_2-1)\boldsymbol{S}_2}{n_1+n_2-2}$, kde $\boldsymbol{S}_1$ a $\boldsymbol{S}_2$ jsou kovarianční matice výběru $\boldsymbol{X}_1$ a $\boldsymbol{X}_2$.

Hotellingova testační statistika má tvar $T^2 = \frac{n_1 n_2}{n_1+n_2}(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)^{\mathrm{T}} \boldsymbol{S}_p^{-1} (\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)$. Veličina $C = T^2 \frac{n_1+n_2-m-1}{m(n_1+n_2-2)}$ á za při platnosti hypotézy H_0 F-rozdělení s m a $n_1 + n_2 - m - 1$ stupni volnosti.

2. Kovarianční matice se významně liší, $C_1 \neq C_2$. (vícerozměrný Behrensův-Fisherův problém)

Vychází se ze dvou výběrů $\boldsymbol{X}_1$ velikosti n_1 a $\boldsymbol{X}_2$ velikosti n_2 pocházejících z m-rozměrných normálních rozdělení $N(\boldsymbol{\mu}_1, \boldsymbol{C}_1)$ a $N(\boldsymbol{\mu}_2, \boldsymbol{C}_2)$.

Nejdříve jsou určeny odhady středních hodnot $\hat{\boldsymbol{\mu}}_1$ a $\hat{\boldsymbol{\mu}}_2$, respektive kovariančních matic $\boldsymbol{S}_1$, $\boldsymbol{S}_2$. Protože $\boldsymbol{C}_1 \neq \boldsymbol{C}_2$, nelze pro určení společné kovarianční matice $\boldsymbol{S}_\mathrm{p}$ použít dřívější rovnici ale

lze sestavit statistiku $T_N^2 = (\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)^\mathrm{T} \left(\frac{\boldsymbol{S}_1}{n_1} + \frac{\boldsymbol{S}_2}{n_2}\right) (\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)$.

Tato veličina T_N^2 však již nemá ani χ^2-rozdělení ani Hotellingovo rozdělení. Použije se proto veličina $C_N = \frac{\hat{f}-m-1}{m\hat{f}} T_N^2$ má *F*-rozdělení s m a $\hat{f}$ stupni volnosti dle vztahu $\hat{f} = \frac{\mathrm{tr}(\boldsymbol{V}^2)+[\mathrm{tr}(\boldsymbol{V})]^2}{\sum_{i=1}^{2}[\mathrm{tr}(\boldsymbol{V}^2)+[\mathrm{tr}(\boldsymbol{V})]^2]/(n_i-1)}$, kde $\boldsymbol{V}_i = \boldsymbol{S}_i n_i$ a $\boldsymbol{V} = \sum_{i=1}^{2} \boldsymbol{V}_i$.

Johansen doporučuje použití statistiky

$$C_J = \left[m + 2A - \frac{6A}{m(m-1)+2}\right]^{-1} T_N^2,$$

která má opět *F*-rozdělení ale s $\hat{f}_1$ stupni volnosti, a veličina A se vyčíslí dle

$$A = \sum_{i=1}^{2} \left[\mathrm{tr}\left(E - V^{-1}V_i^{-1}\right)^2 + \mathrm{tr}^2\left(E - V^{-1}V_i^{-1}\right)^2\right] / 2(n_i - 1)$$

a $\hat{f}_1 = m(m+3)/A$.

Ke zjednodušení dojde v případě, kdy $n_1 = n_2$, tj. velikosti výběrů jsou stejné.

C. Hotellingova statistika se užívá i pro diferenční data $\boldsymbol{Z = X_1 - X_2}$

Pro dvourozměrnou normalitu obou výběrů mají data vektor středních hodnot $\boldsymbol{\mu_Z = \mu_1 - \mu_2}$ kovarianční matici $\boldsymbol{C_z = C_1 + C_2}$. Z dat $\boldsymbol{Z}$ se vypočtou popisné charakteristiky a matice $\boldsymbol{S_z}$ a pak statistika

$$T^2 = n\hat{\boldsymbol{\mu}}_Z^T \boldsymbol{S}_Z^{-1} \hat{\boldsymbol{\mu}}_Z.$$

Nulová hypotéza H_0: $\boldsymbol{\mu_1 = \mu_2}$ je tím převedena na ekvivalentní nulovou hypotézu H_0: $\boldsymbol{\mu_Z = 0}$

Došlo tím ke ztrátě *n*- 1 stupňů volnosti oproti případu, kdy $\boldsymbol{C_1 = C_2}$.

Veličina $C = T^2 \frac{n-m}{m(n-1)}$ má *F*-rozdělení s *m* a *n* - *m* stupni volnosti.

Příklad 4.6 Testování shody vektoru středních hodnot $\mu_1 = \mu_2$ (učebnice)

Za stejných podmínek jako u předešlého příkladu 4.5 byly u dalších deseti náhodně vybraných taveb železa (výběr V_2) zkoumány magnetické indukce x_1 [T] a koercitivní síly x_2 [A . m"[1]] ocelí. Testováním prověřte, zda má výběr V_1 a výběr V_2 shodný vektor středních hodnot, tj. $\boldsymbol{\mu_1 = \mu_2}$

Data: Výběr V_2

i		1	2	3	4	5	6	7	8	9	10
x_1	[T]	1.820	1.760	1.809	1.810	1.700	1.815	1.779	1.811	1.800	1.802
x_2	[A . m^{-1}]	95.9	66.1	80.3	62.8	59.4	74.6	58.0	78.9	63.0	65.0

Řešení

Úlohu budeme řešit

1. jednak pro shodné kovarianční matice $\boldsymbol{C_1 = C_2}$,
2. tak i pro různé kovarianční matice $\boldsymbol{C_1 \neq C_2}$.

Pro výběr V_1 platí $\widehat{\boldsymbol{\mu}}_1^T = (1.769, 66.82)$, $\boldsymbol{S}_1 = \begin{bmatrix} 0.00204 & 0.376 \\ 0.376 & 186.3 \end{bmatrix}$

a pro výběr V_2 platí $\widehat{\boldsymbol{\mu}}_1^T = (1.79, 70.4)$, $\boldsymbol{S}_2 = \begin{bmatrix} 0.00134 & 0.245 \\ 0.245 & 141.5 \end{bmatrix}$

ad 1) Řešení pro případ shodných kovariančních matic $\boldsymbol{C_1 = C_2}$.

Výběrová kovarianční matice j e rovna $\boldsymbol{S_p} = \begin{bmatrix} 0.00169 & 0.31 \\ 0.31 & 162.9 \end{bmatrix}$ a testovací T^2-statistika je pak $T^2 = 1.425$, což odpovídá veličině $C = 1.425 \cdot 17/36 = 0.673$ a kritická hodnota $F_{2,17}(0.95) = 3.5915$ je větší než C, a proto je přijata nulová hypotéza H_0 o shodě vektorů středních hodnot obou výběrů.

Řešení

ad 2) Řešení pro případ odlišných kovariančních matic $\boldsymbol{C_1 \neq C_2}$.

Protože jev obou výběrech stejný počet prvků *n*, bude po odečtení prvků obou výběrů $\hat{\boldsymbol{\mu}}_Z = ((-0.0219, -3.58), \boldsymbol{S}_Z = \begin{bmatrix} 0.0023 & 0.508 \\ 0.508 & 199.6 \end{bmatrix}$

Testovací T^2-statistika = 2.244 a $C = 8 \cdot 2.244/\ 18 = 0.997$ a kritická hodnota$F_{2,8}(0.95)$ = 4.459 je značně větší než *C,* přijímá se nulová hypotéza H_0 *o* shodě vektorů středních hodnot obou výběrů.

Závěr: Je patrné, že pro tento příklad nezávisle na shodě kovariančních matic vychází pro oba výběry shodné vektory středních hodnot $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$.

D. Test shody celkem r středních hodnot μ_i tzn. nulovou hypotézu H0: $\mu_1 = \mu_1 = \cdots = \mu_r$ proti alternativní H_A: $\mu_i \neq \mu_j$.

Vychází se z *k*-tice náhodných výběrů $\boldsymbol{X_k}$ velikosti n_k o kterých se předpokládá, že pocházejí z rozdělení $N(\boldsymbol{\mu_i}, \boldsymbol{C})$ lišících se pouze středními hodnotami. Z těchto výběrů jsou vypočteny odhady $\boldsymbol{\mu_j}$ a $\boldsymbol{C_j}$. Označí se $V_S = \sum_{j=1}^{k} V_j,\ n = \sum_{j=1}^{k} n_j$. a průměr $\tilde{\boldsymbol{\mu}} = \sum_{i=1}^{k} \frac{n_i}{n} \hat{\boldsymbol{\mu}}_i$.

Pak se vyčíslí matice $V_C = \sum_{i=1}^{k} (\hat{\boldsymbol{\mu}}_{\boldsymbol{i}} - \tilde{\boldsymbol{\mu}})(\hat{\boldsymbol{\mu}}_{\boldsymbol{i}} - \tilde{\boldsymbol{\mu}})^T$.

Z řady podobných testů k testování uvedené hypotézy uvedeme pouze Wilcoxovo λ kritérium $\lambda = \det \boldsymbol{V}_S \,/\, \det(\boldsymbol{V}_S + \boldsymbol{V}_C)$

Kvantily rozdělení veličiny λ, jsou publikovány v tabulkách.

E. Test hypotézy o shodnosti všech složek vektoru H_0: $\boldsymbol{\mu} = \mu_1 = \mu_2 = \cdots = \mu_m$, kde $\boldsymbol{A} = \boldsymbol{i} = (1,1,\ldots,1)^T$, $\boldsymbol{g} = \boldsymbol{\mu}$ a $r = 1$.

(Podobně lze formulovat hypotézy o shodě pouze některých složek, respektive jejich nulitě).

Při testování se vychází z výběru $\boldsymbol{X}$ velikosti n, na jehož základě se konstruuje výběrový průměr $\boldsymbol{\mu}$ a výběrová kovarianční matice $\boldsymbol{S}$. K testování nulové hypotézy H_0: $\boldsymbol{\mu} = \boldsymbol{Ag}$ se používá statistika

$$P_1 = \frac{n-m+r}{m-r} n\hat{\boldsymbol{\mu}}^T(\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{B}\boldsymbol{V}^{-1})\hat{\boldsymbol{\mu}},$$

která má v případě platnosti hypotézy H_0 F-rozdělení s $m-r$ a $n-m+r$ stupni volnosti, a matice $\boldsymbol{B}$ zde má tvar $\boldsymbol{B} = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{V}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^T$.

PŘIKLAD 4.7 Vliv přípravy vzorku na stanovení zinku v pšenici

Stryjewska určovala obsah kovů v obilovinách diferenciální pulzní voltametrií. Rozklad vzorku zrn byl prováděn mineralizací za mokra x_1, suchým zpopelněním x_2 a vysokotlakou mineralizací v autoklávu x_3. Na šesti vzorcích pšenice byl stanoven obsah zinku [ppm] pro všechny tři způsoby rozkladu. Testujte nulovou hypotézu H_0, že způsob rozkladu neovlivňuje významně stanovení zinku.

	Obsah zinku v pšenici [ppm]		
i	mineralizace za mokra	suché zpopelnění	tlaková mineralizace
	x_1	x_2	x_3
1	38.8	36.4	38136
2	42.7	33.6	32.5
3	41.1	35.7	34.8
4	42.8	37.1	37.9
5	55.3	36.0	37.0
6	41.6	31.8	37.9

Řešení

Řešení: Označíme-li střední hodnotu znaku x_1 jako μ_1 a podobně střední hodnotu x_2 jako μ_2 a u x_3 jako μ_3, můžeme přepsat uvedenou nulovou hypotézu na tvar H_0: $\mu_1 = \mu_2 = \mu_3 = \mu$. Předpokládejme, že hodnoty x_{ij} j= 1, 2, 3, pocházejí z třírozměrného normálního rozdělení.

Odhady středních hodnot jsou $\hat{\boldsymbol{\mu}}$ =(43.717, 35.1, 35.1) a odpovídající matice $\boldsymbol{S} = (n-1)^{-1}\boldsymbol{V}$, respektive $\boldsymbol{S}$ mají tvar

$$S = \begin{bmatrix} 34.31 & 1.83 & 7.91 \\ 1.83 & 4.00 & -0.84 \\ 7.91 & -0.84 & 9.46 \end{bmatrix}, \qquad S^{-1} = \begin{bmatrix} 0.038 & -0.025 & -0.034 \\ -0.025 & 0.271 & 0.045 \\ -0.034 & 0.045 & 0.138 \end{bmatrix}$$

Je zřejmé, že pro tento případ je *r= 1, m = 3, n = 6.* Matice $\boldsymbol{A}$ je $\boldsymbol{A}$ = (1, 1, 1)^T = $\boldsymbol{i}$. Matice $\boldsymbol{B}$ má pak tvar $\boldsymbol{B} = \frac{\boldsymbol{i}\boldsymbol{i}^T}{\boldsymbol{i}^T\boldsymbol{V}^{-1}\boldsymbol{i}} = \boldsymbol{K}\boldsymbol{i}\boldsymbol{i}^T$, v němž $\boldsymbol{i}\boldsymbol{i}^T$ je matice rozměru (3x3) obsahující samé jedničky. Pro koeficient K je $K = \sum_{i=1}^{3}\sum_{j=1}^{3}\tilde{V}_{ij} = 0.0837$, kde $\tilde{V}_{ij}$ jsou prvky matice $\boldsymbol{V}^{-1}$.

Závěr

Matice

$$\boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{B}\boldsymbol{V}^{-1} = \begin{bmatrix} 0.0017 & -0.0049 & -0.0069 \\ -0.0049 & 0.0539 & 0.0089 \\ -0.0069 & 0.0089 & 0.0376 \end{bmatrix}$$

a

$$P_1 = \frac{4}{2}\, 6\, (43.72 \;\; 35.1 \;\; 35.1) \begin{bmatrix} 0.0077 & -0.0049 & -0.0069 \\ -0.0049 & 0.0539 & 0.0089 \\ -0.0069 & 0.0089 & 0.0276 \end{bmatrix} \begin{bmatrix} 43.72 \\ 35.1 \\ 35.1 \end{bmatrix} =$$

$$= \frac{24}{2} 100.8059 = 1209.67$$

Kvantil *F*-rozdělení $F_{2,4}(0.95) = 6.9443$. Protože je P_1 značně vyšší, nelze přijmout hypotézu H_0 o shodě vektorů středních hodnot všech tří složek x_1, x_2, x_3.

Závěr: Způsob přípravy vzorku významně ovlivní určení obsahu zinku v pšenici.

Statistická analýza kovariančních matic

Z m-rozměrné normálně rozdělené náhodné veličiny $N(\boldsymbol{\mu}, \boldsymbol{C})$ se konstruuje náhodný výběr $\boldsymbol{X}$ velikosti n, který má prvky x_{ij}, $i = 1, \ldots, n, \quad j = 1, \ldots, m$. Při testování se využívá odhadů $\hat{\boldsymbol{\mu}}$ a $\boldsymbol{S}$ nebo výběrové korelační matice $\hat{\boldsymbol{R}}$.

A. Test sféricity: testuje se nulová hypotézaH$_0$: $\boldsymbol{C} = \sigma^2 \boldsymbol{E}$ proti alternativě H$_A$: $\boldsymbol{C} \neq \sigma^2 \boldsymbol{E}$, kde $\sigma^2 > 0$ je rozptyl a $\boldsymbol{E}$ je jednotková matice.

Testační statistika k testování sféricity je $T_S = \det \boldsymbol{S} \left(\frac{\operatorname{tr} \boldsymbol{S}}{m} \right)^m$, kde tr $\boldsymbol{S}$ je stopa matice $\boldsymbol{S}$. Tabulky kvantilů statistiky T_s jsou uvedeny v trabulkách.

B. Pro velké rozsahy výběru n.

Použije se statistika $S_T = -\left(n - 1 - \frac{2m^2+m+2}{6m} \right) \ln T_S$ která má χ^2-rozdělení s $\frac{(m-1)(m+2)}{2}$ stupni volnosti.

Při znalosti vlastních čísel $\lambda_1, \ldots, \lambda_m$ kovarianční matice $\boldsymbol{S}$ lze statistiku T_s vyjádřit ve tvaru

$$T_S = \prod_{i=1}^{n} \lambda_i \Big/ \left(\sum_{i=1}^{n} \frac{\lambda_i}{m} \right)^m .$$

Test sféricity je ekvivalentní testu rovnosti všech vlastních čísel kovarianční matice ***S,*** tj. H_0: $\lambda_1 = \lambda_2 = \cdots = \lambda_m$.

Použijeme-li místo matice ***S*** korelační matici ***R***, testuje se vlastně nulová hypotéza H_0: ***R = E*** proti alternativní H_A: ***R*** $\neq$ ***E.*** (čili hypotéza H_0 vyjadřuje nezávislost složek vícerozměrného normálního rozdělení).

Testovací statistika má jednoduchý tvar $T_R = -n \ln \det \boldsymbol{R}$ a má asymptotické χ^2-rozdělení s $m(m-l)/2$ stupni volnosti.

C. Bartelletův a Sugirův test:

Je možno použít také Bartelletovy statistiky $Q = -\left[n - \frac{2m+11}{6}\right] \ln \det \boldsymbol{R}$ přibližně χ^2-rozdělení s $m(m-l)/2$ stupni volnosti.

Sugirův test s testační statistikou $T_{SS} = \frac{(n-1)m}{2}\left(\frac{m \operatorname{tr} S^2}{\operatorname{tr} S^2} - 1\right)$.

Veličina T_{SS} má přibližně χ^2-rozdělení s $(m-1)(m+2)/2$ stupni volnosti.

Uvedené testy jsou speciálním případem testování nulové hypotézy H_0: $\boldsymbol{C = C_0}$ proti alternativní H_A: $\boldsymbol{C \neq C_0}$, kdy testační statistika má tvar

$$L_c = (n-1)\left(\ln \det \boldsymbol{C}_0 - \ m\ - \ln \det \boldsymbol{S} + \operatorname{tr}(\boldsymbol{S}\,\boldsymbol{C}_0^{-1})\right).$$

Platí, že $L = L_c(1 - D_1)$ má přibližně χ^2-rozdělení s $m(m+1)/2$ stupni volnosti a pro parametr D_1 lze psat $D_1 = \frac{2m^2+3m-1}{6(n-1)(m+1)}$.

Další aproximace spolu s kvantily lze nalézt v tabulkách.

D. Test správnosti korelační matice:

Test korelační matice ***R,*** která by se měla rovnat známé korelační matici $\boldsymbol{R_0}$, čili test nulové hypotézy H_0: $\boldsymbol{R = R_0}$ proti alternativní H_A: $\boldsymbol{R \neq R_0}$.

Testační statistika má tvar

$$L_R = (n - 1)[\ln \frac{\det \boldsymbol{R}_0}{\det \widehat{\boldsymbol{R}}} - m + \mathrm{tr}(\widehat{\boldsymbol{R}}\boldsymbol{R}_0^{-1})]$$

Asymptoticky má statistika L_R χ^2–rozdělení s $m(m-1)/2$ stupni volnosti.

PŘIKLAD 4.8 *Ověření nezávislosti stanovení zinku v pšenici na přípravě vzorků* (učebnice)

Pro data z příkladu 4.7 ověřte nulovou hypotézu H_0. $\boldsymbol{R} = \boldsymbol{E}$, tj. že jednotlivé způsoby rozkladu vzorku poskytují nezávislé výsledky.

f 1 0.156 0.439 0.156 1 -0.136 0.439 -0.136 1

Řešení: Korelační matice je $\widehat{\boldsymbol{R}} = \begin{bmatrix} 1 & 0.156 & 0.439 \\ 0.156 & 1 & -0.136 \\ 0.439 & -0.136 & 1 \end{bmatrix}$ a pro její determinant platí $\det \widehat{\boldsymbol{R}} = 0.745$. Použije se Bartelletovy statistiky $Q = -\left(6 - \frac{17}{6}\right) \ln 0.745 = 0.932.$

Jelikož je kvantil χ^2-rozdělení $\chi^2_{0.95}(3) = 2.353$ vyšší než statistika Q, hypotéza H_0 je přijata.

Závěr: Na základě uvedeného testu vychází, že způsoby rozkladu vzorku lze považovat za nezávislé.

E. Test shodnosti kovariančních matic:

Při testování shody vektorů středních hodnot je třeba testovat také shodu několika kovariančních matic, $\boldsymbol{C}_1 = \boldsymbol{C}_2 = \ldots = \boldsymbol{C}_k$.

Vychází se z výběrů X_i, $i = 1, \ldots, k$, velikosti n_i, pro které jsou určeny kovarianční matice $\boldsymbol{S}_i$.

Společná kovarianční matice je $\boldsymbol{S}_P = \frac{\sum_{i=1}^{k}(n_i-1)\boldsymbol{S}_i}{\sum_{i=1}^{k} n_i - k}$

K testování hypotézy H_0: $\boldsymbol{C}_1 = \boldsymbol{C}_2 = \ldots = \boldsymbol{C}_k$ lze použít testační statistiky

$$L_U = \left(\sum_{j=1}^{k} n_j - k\right) \ln\det \boldsymbol{S}_P - \sum_{j=1}^{k} (n_j - 1) \ln\det \boldsymbol{S}_j$$

Pro větší výběry je statistika bL_U přibližně s χ^2-rozdělením s $(m+1)m(k-1)/2$ stupni volnosti a koeficient b je roven

$$b = 1 - \left(\sum_{j=1}^{k} \frac{1}{n_j - 1} - \frac{1}{\sum_{j=1}^{k} n_j - k}\right) \frac{2m^2 + 3m - 1}{6(m+1)(k-1)}.$$

PŘIKLAD 4.9 Shoda kovariančních matic u dvou výběrů popisujících vlastnosti ocelí (učebnice)

Pro výběry taveb V_1 a V_2 z příkladů 4.5 a 4.6 je třeba před ověřením shody vektorů středních hodnot ověřit shodu kovariančních matic. Proveďte tento test s využitím statistiky L_U.

Řešení: Na základě údajů z příkladu 4.12 určíme, že L_U = 0.4625 a

$$b = 1 - \left[\frac{2}{9} - \frac{1}{18}\right]\frac{8+6-1}{6 \cdot 3} = 0.879.$$

Kvantil χ^2-rozdělení je roven $\chi^2_{0.95}(3)$= 2.353. Protože je hodnota bL_U = 0.407 výrazně nižší než kritická hodnota 2.353, je H_0 o shodnosti kovariančních matic přijata.

Závěr: Oba vzorky ocelí mají shodné kovarianční matice. Protože mají také shodné vektory středních hodnot, jde o dva homogenní výběry, pocházející z téhož rozdělení.