

Cvičení 3: Vícerozměrná jednofaktorová analýza rozptylu (MANOVA)

Na 45 vzorcích rudy pocházejících ze tří ložisek byly zjištěny hodnoty těchto čtyř proměnných:

X1 ... obsah vanadu v popelu (v promile)

X2 ... obsah železa v popelu (v promile)

X3 ... obsah nasycených uhlovodíků (v setinách procenta)

X4 ... obsah aromatických uhlovodíků (v setinách procenta)

Data jsou uložena v souboru ropa.sta.

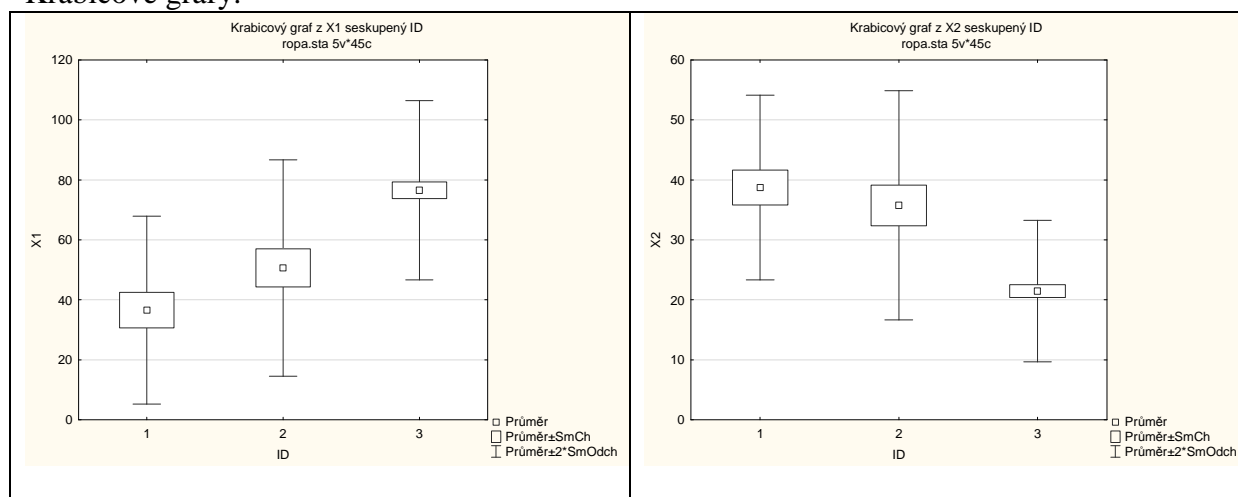
Úkol 1.: Ve všech třech skupinách vypočtete průměry a směrodatné odchylky proměnných X1, X2, X3, X4. Zjistěte rovněž rozsahy skupin. Vytvořte krabicové grafy proměnné X_i ve všech třech skupinách, $i = 1, 2, 3, 4$.

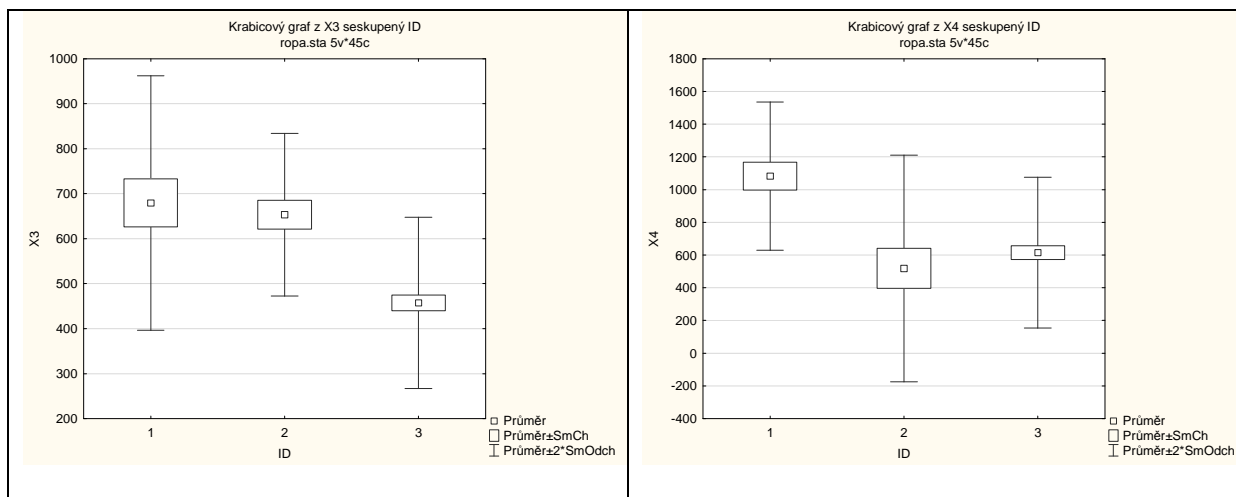
Výsledek:

Tabulka číselných charakteristik

Proměnná	Souhrnné výsledky Popisné statistiky (ropa.sta)			
	ID	N platných	Průměr	Sm.odch.
X1	1	7	36,571	15,6403
X2	1	7	38,714	7,6966
X3	1	7	679,571	141,4318
X4	1	7	1082,571	226,1260
X1	2	8	50,6250	18,0471
X2	2	8	35,7500	9,5581
X3	2	8	653,2500	90,2754
X4	2	8	518,1250	346,3580
X1	3	30	76,5333	14,9406
X2	3	30	21,4667	5,8882
X3	3	30	457,4667	95,2430
X4	3	30	614,8667	230,5085

Krabicové grafy:





Úkol 2.: Na hladině významnosti 0,05 testujte hypotézu, že proměnné X1, X2, X3, X4 se ve všech třech skupinách řídí normálním rozložením.

Výsledek:

Testy normality

Proměnná	Souhrnné výsledky Testy normality (ropa.sta)					
	ID	N	max D	Lilliefors p	W	p
X1: vanad (v promile)	1	7	0,279595	p < ,10	0,837889	0,094950
X2: zezezo (v promile)	1	7	0,256734	p < ,20	0,894783	0,300555
X3: nasyc. uhlovodiky (v des. promile)	1	7	0,164851	p > .20	0,944596	0,680404
X4: arom uhlovodiky (v des. promile)	1	7	0,218850	p > .20	0,886488	0,256840
X1: vanad (v promile)	2	8	0,268226	p < ,10	0,812765	0,039143
X2: zezezo (v promile)	2	8	0,222404	p > .20	0,916727	0,403873
X3: nasyc. uhlovodiky (v des. promile)	2	8	0,222340	p > .20	0,891501	0,241660
X4: arom uhlovodiky (v des. promile)	2	8	0,270404	p < ,10	0,798241	0,027410
X1: vanad (v promile)	3	30	0,114117	p > .20	0,955701	0,239602
X2: zezezo (v promile)	3	30	0,165019	p < ,05	0,939091	0,085977
X3: nasyc. uhlovodiky (v des. promile)	3	30	0,189553	p < ,01	0,884710	0,003623
X4: arom uhlovodiky (v des. promile)	3	30	0,115612	p > .20	0,954858	0,227664

Úkol 3.: Na hladině významnosti 0,05 testujte hypotézu, že varianční matice proměnných X1, X2, X3, X4 jsou ve všech třech skupinách shodné.

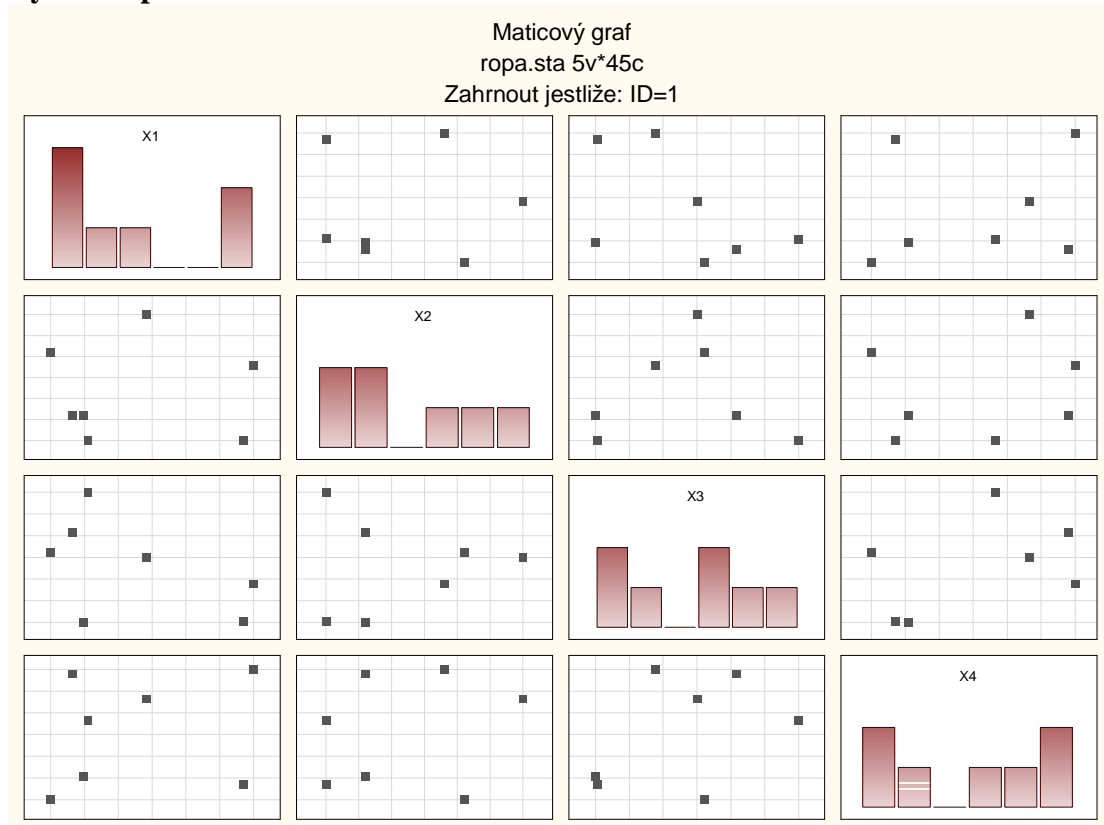
Výsledek Boxova M-testu:

Boxův M test (ropa.sta)				
Efekt: ID (Vypočteno pro všechny proměnné)				
	Boxovo M	Chí-kv.	SV	p
Boxovo M	35,34766	27,23627	20	0,128747

Lze konstatovat, že důležité předpoklady vícerozměrné analýzy rozptylu jsou splněny.

Úkol 4.: Pomocí maticových grafů prověřte, že vztahy mezi proměnnými X1, X2, X3, X4 jsou ve všech třech skupinách přibližně lineární.

Výsledek pro 1. naleziště:



Úkol 5.: Na hladině významnosti 0,05 testujte hypotézu, že střední hodnoty proměnných X1, X2, X3, X4 jsou ve všech třech skupinách shodné. Použijte Wilksův, Pillaiův, Hotellingův – Lawleův a Royův test.

Výsledek:

Efekt	Vícerozměrné testy významnosti. (ropa.sta) Sigma-omezená parametrizace Dekompozice efektivní hypotézy					
	Test	Hodnota	F	Efekt SV	Chyba SV	p
Abs. člen	Wilksův	0,01616	593,4657	4	39	0,000000
	Pillaiův	0,98384	593,4657	4	39	0,000000
	Hotelling	60,86828	593,4657	4	39	0,000000
	Royův	60,86828	593,4657	4	39	0,000000
ID	Wilksův	0,17959	13,2570	8	78	0,000000
	Pillaiův	1,08176	11,7808	8	80	0,000000
	Hotelling	3,11290	14,7863	8	76	0,000000
	Royův	2,53997	25,3997	4	40	0,000000

Úkol 6.: Pomocí simultánního testu založeného na Wilksově statistice testujte na hladině významnosti 0,05 hypotézu, že proměnné X1, X2, X3, X4 nezpůsobují rozdíly mezi skupinami.

Výsledek:

Vypočtené testové statistiky pro simultánní testy a kritická hodnota:

	1 K1	2 K2	3 K3	4 K4	5 kvantil
1	30,114241	32,723182	27,859025	17,643474	15,507313

Úkol 7.: Na hladině významnosti 0,05 proveďte vícerozměrnou obdobu mnohonásobného porovnávání, tj. zjistěte, které dvojice skupin se liší.

Výsledek:

Výsledek pro 1. a 2. skupinu:

t-testy; grupováno: ID: naleziste (ropa.sta) Skup. 1: 1; Skup. 2: 2 Hotellingovo 45,6734 F(4,10)=8,7833 p<,00261											
Proměnná	Průměr 1	Průměr 2	t	sv	p	Poč.plat 1	Poč.plat. 2	Sm.odch. 1	Sm.odch. 2	F-poměr Rozptyly	p Rozptyly
X1	36,571	50,6250	-1,59930	13	0,133764	7	8	15,6403	18,0471	1,331443	0,743087
X2	38,714	35,7500	0,65470	13	0,524074	7	8	7,6966	9,5581	1,542203	0,613888
X3	679,571	653,2500	0,43578	13	0,670148	7	8	141,4318	90,2754	2,454458	0,265396
X4	1082,571	518,1250	3,67238	13	0,002814	7	8	226,1260	346,3580	2,346116	0,318519

Vypočtenou p-hodnotu (tj. 0,00261) porovnáme s $\frac{\alpha}{\binom{3}{2}} = \frac{0,05}{3} = 0,01\bar{6}$. Vidíme, že 1. a 2. skupina se liší.

Výsledek pro 1. a 3. skupinu

t-testy; grupováno: ID: naleziste (ropa.sta) Skup. 1: 1; Skup. 2: 3 Hotellingovo 125,397 F(4,32)=28,662 p<,00000											
Proměnná	Průměr 1	Průměr 3	t	sv	p	Poč.plat 1	Poč.plat. 3	Sm.odch. 1	Sm.odch. 3	F-poměr Rozptyly	p Rozptyly
X1	36,571	76,5333	-6,32043	35	0,000000	7	30	15,6403	14,9406	1,095851	0,776819
X2	38,714	21,4667	6,58961	35	0,000000	7	30	7,6966	5,8882	1,708565	0,309137
X3	679,571	457,4667	5,05771	35	0,000013	7	30	141,4318	95,2430	2,205100	0,142430
X4	1082,571	614,8667	4,84954	35	0,000025	7	30	226,1260	230,5085	1,039138	1,000000

I v tomto případě nulovou hypotézu zamítáme na hladině významnosti 0,05.

Výsledek pro 2. a 3. skupinu:

t-testy; grupováno: ID: naleziste (ropa.sta) Skup. 1: 2; Skup. 2: 3 Hotellingovo 44,5444 F(4,33)=10,208 p<,00002											
Proměnná	Průměr 2	Průměr 3	t	sv	p	Poč.plat 2	Poč.plat. 3	Sm.odch. 2	Sm.odch. 3	F-poměr Rozptyly	p Rozptyly
X1	50,6250	76,5333	-4,17559	36	0,000180	8	30	18,0471	14,9406	1,459063	0,441637
X2	35,7500	21,4667	5,31026	36	0,000006	8	30	9,5581	5,8882	2,634953	0,061803
X3	653,2500	457,4667	5,21782	36	0,000008	8	30	90,2754	95,2430	1,113082	0,958255
X4	518,1250	614,8667	-0,94544	36	0,350739	8	30	346,3580	230,5085	2,257752	0,116036

Vidíme, že i 2. a 3. skupina se liší na hladině významnosti 0,05

Úkol 8.: Na hladině významnosti 0,05 zjistěte, které proměnné způsobují rozdíly mezi jednotlivými dvojicemi skupin.

Výsledek:

Podíváme-li se na tabulky v úkolu 7, můžeme konstatovat, že:

- naleziště 1 a 2 se liší pouze v obsahu aromatických uhlovodíků
- naleziště 1 a 3 se liší v obsahu všech čtyř látek
- naleziště 2 a 3 se neliší pouze v obsahu aromatických uhlovodíků.