

Cvičení 2: Analýza rozptylu dvojného třídění

Příklad na analýzu rozptylu bez interakcí:

V rámci pedagogicko – psychologického výzkumu bylo sledováno, zda čas potřebný k vyřešení určité úlohy závisí na denní době a hlučnosti okolí. Bylo proto vybráno 12 studentů s přibližně stejnými studijními výsledky a rozděleno do tří skupin. První skupina řešila úlohu ráno, druhá v poledne a třetí večer. V každé skupině vždy jeden student pracoval v tichém prostředí, druhý poslouchal reprodukovanou hudbu, třetí rozhlasovou hru a čtvrtý silný pouliční hluk. Počet minut potřebných k vyřešení úlohy je uveden v tabulce:

	ticho	hudba	hra	hluk
ráno	6	7	8	6
v poledne	8	5	10	5
večer	7	6	12	7

Na hladině významnosti 0,05 testujte hypotézu, zda doba potřebná k vyřešení úlohy nezávisí na denní době a na hlučnosti okolí. V případě zamítnutí nulové hypotézy zjistěte, které dvojice řádků resp. sloupců se liší na hladině významnosti 0,05.

Návod:

Načtete datový soubor hlucnost_okoli.sta se třemi proměnnými X, A, B a 12 případy. Proměnná X obsahuje počet minut, A – denní doba (1 – ráno, 2 – v poledne), B – hlučnost okolí (1 – ticho, 2 – hudba, 3 – hra, 4 - hluk).

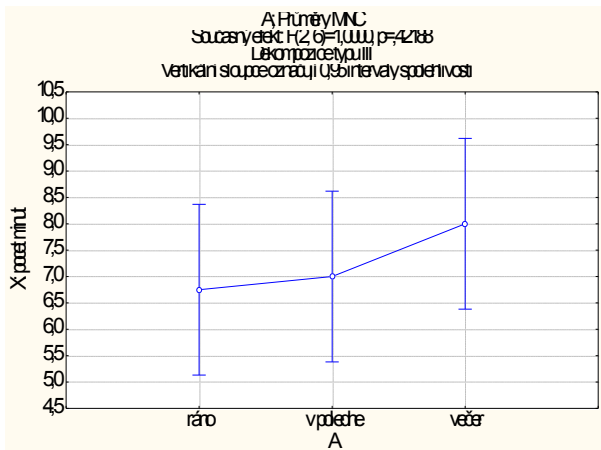
Nejprve spočítáme průměry pro denní doby a pro hlučnost okolí:

Statistiky – ANOVA – Typ analýzy ANOVA hlavních efektů, Metoda specifikace: Rychlé nastavení – OK, Proměnné – Seznam závislých proměnných X, Kategor. nezáv, prom. (faktory) A, B – OK – Možnosti – Parametrizace – odškrtneme Sigma-omezená, zaškrtneme Bez absolutního členu – OK – Průměry – vybereme Efekt A (resp. B) – Vš. Marginální tabulky.

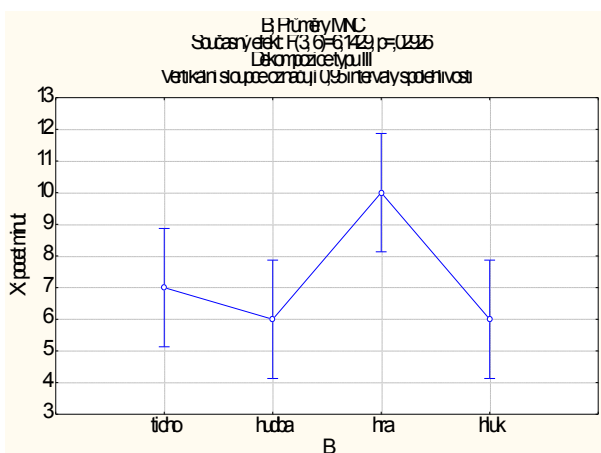
A; Vážené průměry (hlučnost okolí)						
Současný efekt: $F(2, 6)=1,0000$, p						
Dekompozice typu III						
Č. buř	A	X	X	X	X	N
		Prům	Sm.C	-95,0	+95,0	
1	ráno	6,750	0,478	5,226	8,273	4
2	v pole	7,000	1,224	3,102	10,89	4
3	več	8,000	1,354	3,690	12,30	4

B; Vážené průměry (hlučnost okolí)						
Současný efekt: $F(3, 6)=6,1429$						
Dekompozice typu III						
Č. buř	B	X	X	X	X	N
		Prům	Sm.C	-95,0	+95,0	
1	tich	7,000	0,577	4,515	9,484	3
2	hud	6,000	0,577	3,515	8,484	3
3	hr	10,00	1,154	5,031	14,96	3
4	hlk	6,000	0,577	3,515	8,484	3

Současně můžeme nechat vykreslit grafy závislostí počtu minut potřebných k vyřešení úlohy denní době a poté na hlučnosti prostředí.



Vidíme, že průměrná doba potřebná k vyřešení úlohy se zvyšuje s postupující denní dobou.



Nejvyšší průměrnou dobu potřebovali studenti, kteří při řešení poslouchali rozhlasovou hru, naopak nejkratší doba stačila těm, kteří poslouchali hudbu či byli vystaveni hluku z ulice.

Dále získáme tabulku analýzy rozptylu dvojného třídění bez interakcí:

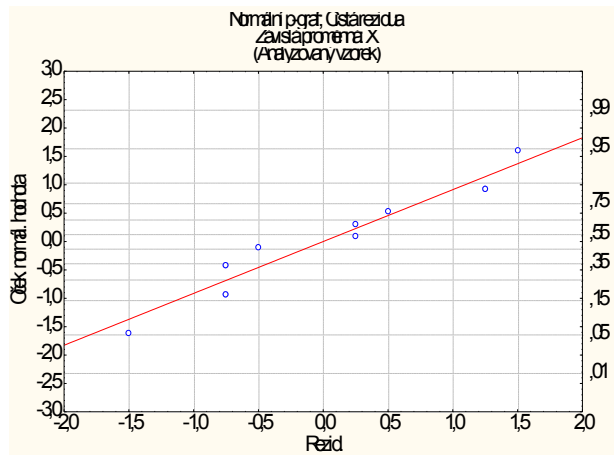
Návrat do ANOVA Výsledky – Všechny efekty.

Jednorozměrné testy významnosti pro Přeparametrizovaný model Dekompozice typu III					
Efek	SC	Stupň volno	PC	F	p
A	3,500	2	1,750	1,000	0,421
B	32,25	3	10,75	6,142	0,029
Chyť	10,50	6	1,750		

Vidíme, že na hladině významnosti 0,05 je významný faktor B, tj. hlučnost okolí. Vliv denní doby není prokazatelný na hladině významnosti 0,05.

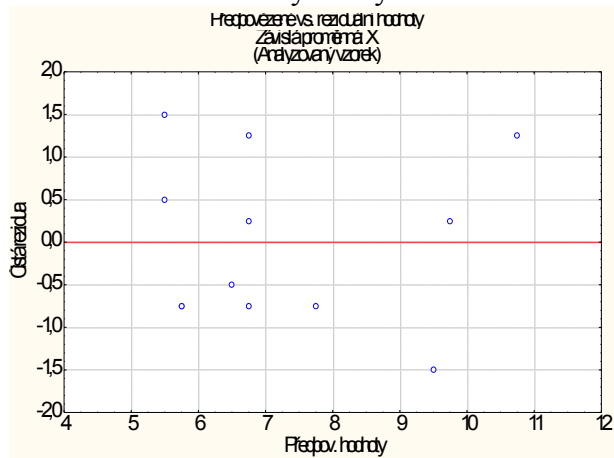
Než přistoupíme k mnohonásobnému porovnávání, budeme ještě analyzovat rezidua.

Návrat do ANOVA Výsledky – Rezidua – P-graf reziduí



Normální pravděpodobnostní graf reziduí svědčí o tom, že rezidua se řídí normálním rozložením.

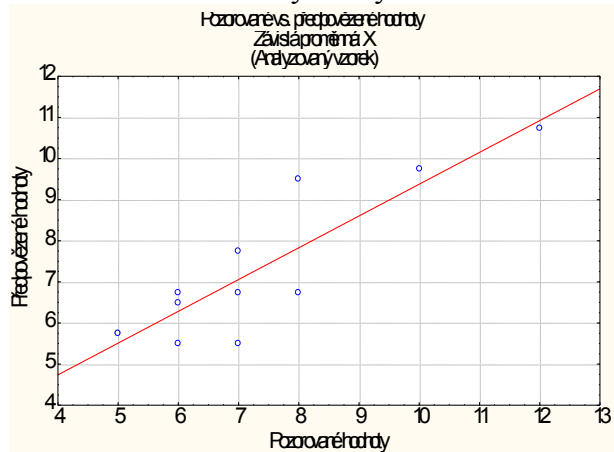
Návrat do ANOVA Výsledky – Rezidua – Před. & rezidua



Graf závislosti reziduí na predikovaných hodnotách vypadá jako náhodný mrak bodů, což je v pořádku.

Podíváme se ještě na graf závislosti predikovaných hodnot na pozorovaných hodnotách:

Návrat do ANOVA Výsledky – Rezidua – Poz. & před.



Provedeme mnohonásobné porovnávání: Návrat do ANOVA Výsledky – Více výsledků – Post-hoc – Efekt B – Tukeyův HSD.

Tukeyov HSD test; proměnná X					
Přibližné pravděpodobnosti pro					
Chyba: meziskup. PC = 1,7500,					
Č. bur	B	{1}	{2}	{3}	{4}
		7,000	6,000	10,00	6,000
1	tich		0,793	0,112	0,793
2	hud	0,793		0,038	1,000
3	hr	0,112	0,038		0,038
4	hlu	0,793	1,000	0,038	

Na hladině významnosti 0,05 se liší skupiny (2,3) a (3,4), tj. (hudba, hra) a (hra, hluk).

Příklad na analýzu rozptylu s interakcemi: Velké jezero na severu USA bylo rozděleno na pět oblastí a z každé oblasti byly odebrány tři vzorky vody. U každého vzorku byla provedena dvě opakovaná stanovení obsahu fosforu (v mg/l). Výsledky laboratorních analýz obsahu fosforu jsou uvedeny v tabulce:

	Vzorek 1	Vzorek 2	Vzorek 3
Oblast 1	0,010 0,008	0,009 0,012	0,011 0,006
Oblast 2	0,013 0,017	0,008 0,010	0,012 0,011
Oblast 3	0,009 0,015	0,010 0,014	0,017 0,011
Oblast 4	0,011 0,015	0,008 0,013	0,010 0,014
Oblast 5	0,014 0,006	0,018 0,010	0,005 0,013

Na hladině významnosti 0,05 vyšetřete, zda oblasti a odebrané vzorky mají vliv na koncentraci fosforu ve vodě. Dochází k vzájemnému ovlivňování těchto faktorů?

Návod:

Vytvořte nový datový soubor se třemi proměnnými X, A, B a 30 případy. Do X napište hodnoty obsahu fosforu, proměnná A reprezentuje oblasti 1 – 5 a proměnná B vzorky 1 – 3.

Statistiky – ANOVA – Typ analýzy Vícefaktorová ANOVA. Metoda specifikace: Rychlé nastavení – OK, Proměnné – Seznam závislých proměnných X, Kategor. nezáv, prom. (faktory(A, B – OK – Možnosti – Parametrizace – odškrtneme Sigma-omezená, zaškrtneme Bez absolutního členu – OK – Všechny efekty.
Dostaneme tabulku analýzy rozptylu dvojného třídění s interakcemi.

Jednorozměrné testy významnosti pro Přeparametrizovaný model Dekompozice typu III					
Efek	SC	Stupň volno	PC	F	p
A	0,000	4	0,000	0,726	0,587
B	0,000	2	0,000	0,131	0,877
A*B	0,000	8	0,000	0,727	0,666
Chyba	0,000	12	0,000		

Na hladině významnosti 0,05 se neprokázal vliv faktoru A, B ani interakcí.

Příklad k samostatnému řešení na analýzu rozptylu dvojného třídění s interakcemi (příklad je převzat z bakalářské práce Mariky Dienové)

Na spálení do cementárny se dodávají různé druhy odpadů, nás budou zajímat emulzní topné oleje. Zjišťuje se jejich výhřevnost (veličina X - v MJ/kg) v závislosti na době odebrání vzorku (faktor A – buď čerstvě po dodání nebo těsně před spálením) a na dodavateli odpadů (faktor B – buď dodavatel I, II nebo III).

	I	II	III
Po dodání	36,33	38,46	38,43
	36,8	37,65	38,56
	37,28	38,36	38,62
Před spálením	10,44	26	20,11
	18,66	25,18	35,82
	15,96	24,22	26,13

Proveďte analýzu rozptylu dvojného třídění s interakcemi.