

Téma č.7: 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:

Vytvořte nový datový soubor se třemi proměnnými X, A, B a 12 případy. Do proměnné X napište hodnoty 6 7 8 6 8 5 10 5 7 6 12 7, do A napište 1 1 1 1 2 2 2 2 3 3 3 3 a do B 1 2 3 4 1 2 3 4 1 2 3 4. Proměnným X, A, B vytvořte návěští (X – počet minut, A – denní doba, B – hlučnost okolí) a popište, co znamenají jednotlivé varianty proměnných A (1 – ráno, 2 – v poledne) a B (1 – ticho, 2 – hudba, 3 – hra, 4 - hluk).

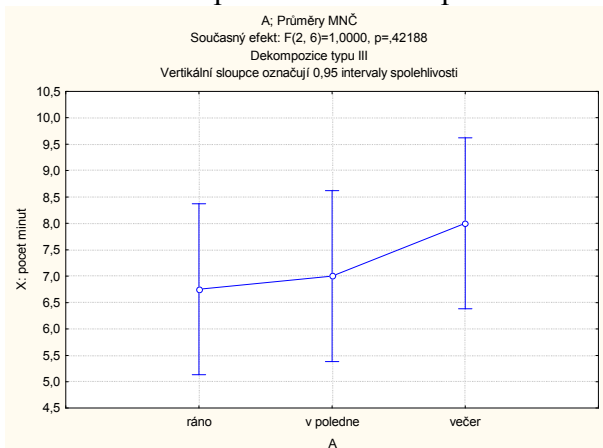
Nejprve spočítáme průměry pro denní doby a pro hlučnost okolí:

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.

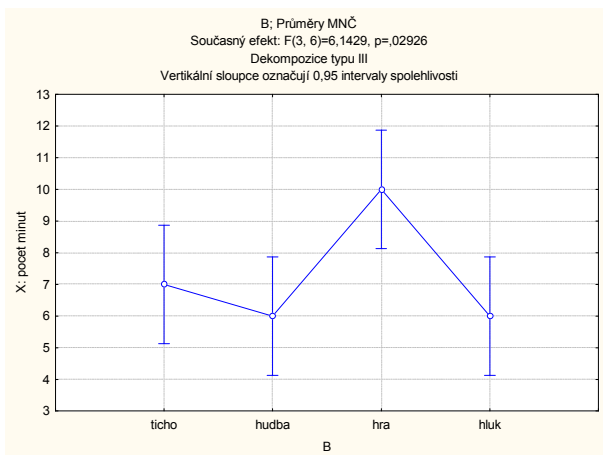
A; Vážené průměry (hlucnost_okoli) Současný efekt: $F(2, 6)=1,0000$, $p=,42188$ Dekompozice typu III						
Č. buňky	A	X Průměr	X Sm.Ch.	X -95,00%	X +95,00%	N
1	ráno	6,750000	0,478714	5,226520	8,27348	4
2	v poledne	7,000000	1,224745	3,102315	10,89768	4
3	večer	8,000000	1,354006	3,690947	12,30905	4

B; Vážené průměry (hlucnost_okoli) Současný efekt: $F(3, 6)=6,1429$, $p=,02926$ Dekompozice typu III						
Č. buňky	B	X Průměr	X Sm.Ch.	X -95,00%	X +95,00%	N
1	ticho	7,000000	0,577350	4,515862	9,48414	3
2	hudba	6,000000	0,577350	3,515862	8,48414	3
3	hra	10,000000	1,154701	5,031725	14,96828	3
4	hluk	6,000000	0,577350	3,515862	8,48414	3

Současně můžeme nechat vykreslit grafy závislosti počtu minut potřebných k vyřešení úlohy na 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:

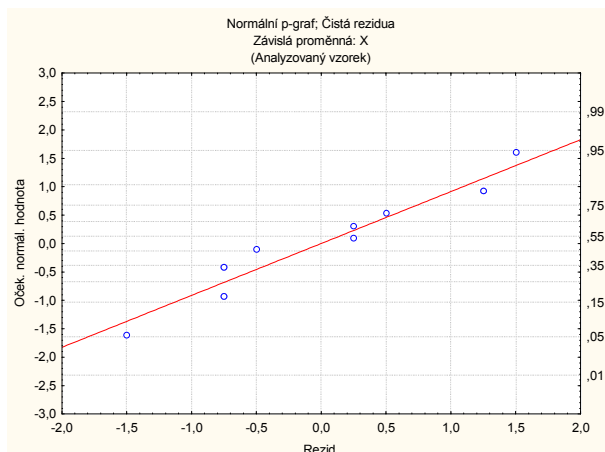
Návrat do ANOVA Výsledky – Všechny efekty.

Jednorozměrné testy významnosti pro X (hlucnost_okoli.sta)					
Přeparametrizovaný model					
Dekompozice typu III					
Efekt	SČ	Stupně volnosti	PČ	F	p
A	3,50000	2	1,75000	1,000000	0,421875
B	32,25000	3	10,75000	6,142857	0,029263
Chyba	10,50000	6	1,75000		

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

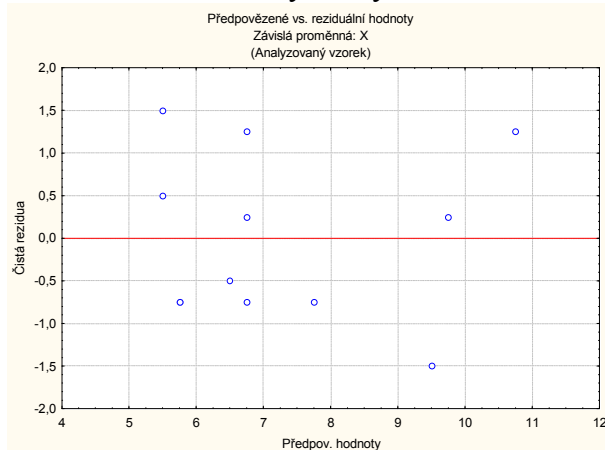
Než přistoupíme k mnohonásobnému porovnávání, budeme ještě analyzovat rezidua.

Návrat do ANOVA Výsledky – Rezidua – P-graf reziduí



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

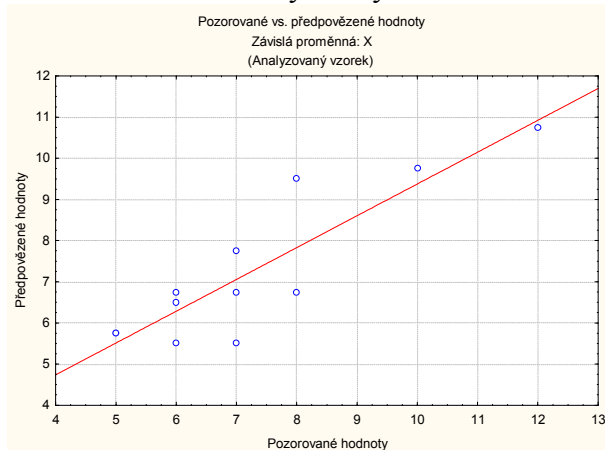
Návrat do ANOVA Výsledky – Rezidua – Před. & rezidua



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.

Tukeyův HSD test; proměnná X (hlucnost_okoli.sta) Přibližné pravděpodobnosti pro post hoc testy Chyba: meziskup. PČ = 1,7500, sv = 6,0000					
Č. buňky	B	{1} 7,0000	{2} 6,0000	{3} 10,000	{4} 6,0000
1	ticho		0,793146	0,112486	0,793146
2	hudba	0,793146		0,038069	1,000000
3	hra	0,112486	0,038069		0,038069
4	hluk	0,793146	1,000000	0,038069	

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:

Datový soubor se jmenuje fosfor_v_jezere. Proměnná X obsahuje 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 X (fosfor_v_jezere.sta) Přeparametrizovaný model Dekompozice typu III					
Efekt	SČ	Stupně volnosti	PČ	F	p
A	0,000038	4	0,000010	0,726010	0,587787
B	0,000003	2	0,000002	0,131313	0,877940
A*B	0,000077	8	0,000010	0,727904	0,666096
Chyba	0,000198	15	0,000013		

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.

Výsledek: Na hladině významnosti 0,05 je prokazatelný vliv faktorů A, B, vliv interakcí nikoliv.