11. Analýza rozptylu dvojného třídění

Motivace

Zkoumáme vliv dvou faktorů A a B na závisle proměnnou veličinu X. Např. zjišťujeme, zda výnosy určité plodiny (náhodná veličina X) jsou ovlivněny typem půdy (faktor A) a způsobem hnojení (faktor B). Předpokládáme, že faktor A má a úrovní (tj. počet typů půdy) a faktor B má b úrovní (tj. počet způsobů hnojení). Přitom máme n_{ij} pokusů takových, že na i-tém typu půdy byl použit j-tý způsob hnojení. Výsledky (tzn. výnosy dané plodiny) těchto n_{ij} pokusů označíme $X_{ij1}, X_{ij2}, \ldots, X_{ijn_{ij}}$. Omezíme se na případy, kdy počet pozorování $n_{ij} = c \geq 1$ (jde o tzv. vyvážené třídění). Výsledky lze zapsat do tabulky:

		faktor B			
		1	2	…	b
faktor A	1	$X_{111}, \ldots, X_{11c}$	$X_{121}, \ldots, X_{12c}$	…	$X_{1b1}, \ldots, X_{1bc}$
	2	$X_{211}, \ldots, X_{21c}$	$X_{221}, \ldots, X_{22c}$	…	$X_{2b1}, \ldots, X_{2bc}$
	⋮	⋮	⋮	…	⋮
	a	$X_{a11}, \ldots, X_{a1c}$	$X_{a21}, \ldots, X_{a2c}$	…	$X_{ab1}, \ldots, X_{abc}$

Analogicky jako u analýzy rozptylu jednoduchého třídění předpokládáme, že data se řídí normálním rozložením, tj.

$X_{ij1}, X_{ij2}, \ldots, X_{ijc} \sim N(\mu_{ij}, \sigma^2)$, i = 1, …, a, j = 1, …, b

a jednotlivé náhodné výběry jsou stochasticky nezávislé, tedy

$X_{ijk} = \mu_{ij} + \varepsilon_{ijk}$,

kde ε_{ijk} jsou stochasticky nezávislé náhodné veličiny s rozložením $N(0, \sigma^2)$. Zajímá nás, zda všechny střední hodnoty jsou stejné. Přístup k problému se liší podle toho, zda faktory A, B jsou nezávislé (pak se jedná o analýzu rozptylu dvojného třídění bez interakcí) nebo se mohou nějakým způsobem ovlivňovat (jde o analýzu rozptylu dvojného třídění s interakcemi).

Označení

V analýze rozptylu dvojného třídění se používá následující označení.

$$n = abc, X_{ij.} = \sum_{k=1}^{c} X_{ijk}, M_{ij.} = \frac{1}{c} X_{ij.}, X_{i..} = \sum_{j=1}^{b}\sum_{k=1}^{c} X_{ijk}, M_{i..} = \frac{1}{bc} X_{i..}, X_{...} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} X_{ijk},$$

$$M_{...} = \frac{1}{n} X_{...}$$

Analogické označení zavedeme i pro jiné kombinace indexů.

Dvojné třídění bez interakcí

Předpokládáme, že řádkový faktor A a sloupcový faktor B se neovlivňují (např. to znamená, že každý ze čtyř způsobů hnojení působí stejně na každém ze tří druhů půdy).

Náhodné veličiny X_{ijk} se řídí modelem

M0: $X_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$ pro $i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, c$, přičemž ε_{ijk} jsou stochasticky nezávislé náhodné veličiny s rozložením $N(0, \sigma^2)$,

μ je společná část střední hodnoty závisle proměnné veličiny,

α_i je efekt faktoru A na úrovni i,

β_j je efekt faktoru B na úrovni j.

Parametry μ, α_i, β_j neznáme. Požadujeme, aby platily tzv. reparametrizační rovnice:

$$\sum_{i=1}^{a} \alpha_i = 0, \sum_{j=1}^{b} \beta_j = 0.$$

Podobně jako v analýze rozptylu jednoduchého třídění se počítají součty čtverců.

$S_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} (X_{ijk} - M_{...})^2$ … celkový součet čtverců,

počet stupňů volnosti $f_T = n - 1$,

$S_A = bc\sum_{i=1}^{a} (M_{i..} - M_{...})^2$ … součet čtverců pro řádkový faktor A,

počet stupňů volnosti $f_A = a - 1$,

$S_B = ac\sum_{j=1}^{b} (M_{.j.} - M_{...})^2$ … součet čtverců pro sloupcový faktor B,

počet stupňů volnosti $f_B = b - 1$,

$S_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} (X_{ijk} - M_{ij.})^2$ … reziduální součet čtverců,

počet stupňů volnosti $f_E = n - a - b + 1$.

Lze dokázat, že $S_T = S_A + S_B + S_E$.

Sčítanci $(M_{i..} - M_{...})$ a $(M_{.j.} - M_{...})$ představují bodové odhady efektů α_i a β_j.

Pokud by nezáleželo na sloupcovém faktoru B, platila by hypotéza $\beta_1 = \ldots = \beta_b = 0$ a dostali bychom model

M1: $X_{ijk} = \mu + \alpha_i + \varepsilon_{ijk}$

Platnost uvedené hypotézy ověřujeme pomocí testové statistiky

$F_B = \frac{S_B / f_B}{S_E / f_E}$, která se řídí rozložením F(b-1,n-a-b+1), je-li model M1 správný.

Hypotézu o nevýznamnosti sloupcového faktoru tedy zamítneme na hladině významnosti α, když platí:

$F_B \geq F_{1-\alpha}(b-1,n-a-b+1)$.

Kdyby nezáleželo ani na řádkovém faktoru, platila by hypotéza $\alpha_1 = \ldots = \alpha_a = 0$ a dostali bychom model

M2: $X_{ijk} = \mu + \varepsilon_{ijk}$

Rozdíl mezi modely M1 a M2 ověřujeme pomocí testové statistiky

$F_A = \dfrac{S_A / f_A}{S_E / f_E}$, která se řídí rozložením F(a-1,n-a-b+1), je-li model M2 správný.

Hypotézu o nevýznamnosti řádkového faktoru tedy zamítneme na hladině významnosti α, když platí:

$F_B \geq F_{1-\alpha}(a-1,n-a-b+1)$.

Při uvedeném postupu tedy zjišťujeme, zda záleží na sloupcovém efektu B. Pokud ne, platí model M1 a ptáme se, zda záleží na řádkovém efektu A, tj. zda platí model M2.

Postup lze samozřejmě provést i v jiném pořadí – nejdřív zkoumáme řádkový efekt A (tj. ověřujeme platnost modelu M1': $X_{ijk} = \mu + \beta_j + \varepsilon_{ijk}$) a poté sloupcový efekt B. Lze ukázat, že oba řetězce M0 → M1 → M2 a M0 → M1'→ M2 dají stejné výsledky. (To platí pouze za předpokladu, že $n_{ij} = c$ pro všechna i, j.)

Výsledky výpočtů zapisujeme do tabulky analýzy rozptylu dvojného třídění bez interakcí.

Zdroj variability	součet čtverců	st. vol.	podíl S/f	$F = \frac{S/f}{S_E/f_E}$
řádkový efekt A	S_A	$f_A = a-1$	S_A/f_A	$F = \frac{S_A/f_A}{S_E/f_E}$
sloupcový efekt B	S_B	$f_B = b-1$	S_B/f_B	$F = \frac{S_B/f_B}{S_E/f_E}$
reziduální	S_E	$f_E = n-a-b+1$	S_E/f_E	-
celkem	S_T	$f_T = n-1$	-	-

Schefféhoa Tukeyova metoda mnohonásobného porovnávání

Zjistíme-li, že existují významné rozdíly mezi řádky, můžeme pomocí Schefféhonebo Tukeyovy metody zjistit, které dvojice řádků se významně liší. Určíme tedy, které rozdíly $\alpha_i - \alpha_t$ jsou nenulové (na dané hladině významnosti).

Podle **Schefféhometody** zamítneme rovnost $\alpha_i = \alpha_t$, když

$$|M_{i..} - M_{t..}| > \sqrt{\frac{2(a-1)}{bc} \cdot \frac{S_E}{n-a-b+1} \cdot F_{1-\alpha}(a-1, n-a-b+1)}$$

a podle **Tukeyovy metody**, když

$|M_{i..} - M_{t..}| > \sqrt{\frac{1}{bc} \cdot \frac{S_E}{n-a-b+1}}\, q_{1-\alpha}(a, n-a-b+1)$, kde $q_{1-\alpha}(a,n-a-b+1)$ najdeme v tabulkách kvantilů studentizovaného rozpětí.

Jestliže zjistíme významný rozdíl mezi sloupci, určujeme podobně, které dvojice sloupců se mezi sebou liší, tj. které rozdíly $\beta_j - \beta_t$ jsou nenulové.
Podle **Scheffého metody** zamítneme rovnost $\beta_j = \beta_t$, když

$$\left|M_{.j.} - M_{.t.}\right| > \sqrt{\frac{2(b-1)}{ac} \cdot \frac{S_E}{n-a-b+1} \cdot F_{1-\alpha}(b-1, n-a-b+1)}$$

a podle **Tukeyovy metody**, když

$$\left|M_{.j.} - M_{.t.}\right| > \sqrt{\frac{1}{ac} \cdot \frac{S_E}{n-a-b+1}} q_{1-\alpha}(b, n-a-b+1).$$

Příklad:
Byl zaznamenán prodej určitého zboží (v kusech) během tří stejně dlouhých časových období. Přitom byl sledován jednak vliv balení zboží (řádkový faktor A, úroveň 1 – balení v sáčku, úroveň 2 – balení v krabičce) a jednak vliv druhu reklamy (sloupcový faktor B, úroveň 1 – bez reklamy, úroveň 2 – reklama v novinách, úroveň 3 – reklama v TV a novinách). Výsledky prodeje jsou zaznamenány v tabulce:

		B		
		1	2	3
A	1	1	1	6
	2	3	4	9

Na hladině významnosti 0,05 je třeba posoudit vliv reklamy a i vliv balení zboží na jeho prodej.

Řešení:
Data zpracujeme pomocí analýzy rozptylu dvojného třídění bez interakcí. Přitom $a = 2$, $b = 3$, $c = 1$, $n = 6$. Nejprve provedeme pomocné výpočty:
$X_{1..} = 8$, $X_{2..} = 16$, $M_{1..} = 8/3$, $M_{2..} = 16/3$, $X_{...} = 24$, $M_{...} = 24/6 = 4$, $X_{.1.} = 4$, $X_{.2.} = 5$, $X_{.3.} = 15$, $M_{.1.} = 4/2 = 2$, $M_{.2.} = 5/2$, $M_{.3.} = 15/2$

$$S_A = 3\left[\left(\frac{8}{3} - 4\right)^2 + \left(\frac{16}{3} - 4\right)^2\right] = \frac{32}{3} = 10,\overline{6},$$

$$S_B = 2\left[(2-4)^2 + \left(\frac{5}{2} - 4\right)^2 + \left(\frac{15}{2} - 4\right)^2\right] = 37,$$

$$S_T = (1-4)^2 + (1-4)^2 + (6-4)^2 + (3-4)^2 + (3-4)^2 + (9-4)^2 = 48,$$

$$S_E = S_T - S_A - S_B = 0,\overline{3}.$$

Výsledky zapíšeme do tabulky analýzy rozptylu dvojného třídění bez interakcí.

Zdroj variability	součet čtverců	st. vol.	podíl S/f	$F = \frac{S/f}{S_E/f_E}$

způsob balení	$10,\overline{6}$	1	$10,\overline{6}$	63,99
druh reklamy	37	2	18,5	110,98
reziduální	$0,\overline{3}$	2	$0,1\overline{6}$	-
celkem	48	5	-	-

Odpovídající kvantily: pro řádkový efekt $F_{0,95}(1,2) = 18,1$, pro sloupcový efekt $F_{0,95}(2,2) = 19$. Protože $F_A = 63,99 \geq 18,1$, zamítáme na hladině významnosti 0,05 hypotézu, že způsob balení nemá vliv na prodej zboží. Podobně $F_B = 110,98 \geq 19$, tedy na hladině významnosti 0,05 zamítáme hypotézu, že druh reklamy nemá vliv na prodej zboží. V tomto druhém případě lze pomocí Scheffého nebo Tukeyovy metody zjistit, které druhy reklamy se od sebe liší na hladině významnosti 0,05.
Nejprve vypočítáme absolutní hodnoty rozdílů sloupcových průměrů:

$$|M_{.1.} - M_{.2.}| = \left|2 - \frac{5}{2}\right| = 0,5, |M_{.1.} - M_{.3.}| = \left|2 - \frac{15}{2}\right| = 5,5, |M_{.2.} - M_{.3.}| = \left|\frac{5}{2} - \frac{15}{2}\right| = 5$$

Pravá strana Scheffého vzorce je $\sqrt{\frac{2}{2} \cdot 0,1\overline{6} \cdot 19} = 2,52$.

Vidíme, že podle Scheffého metody se na hladině významnosti 0,05 liší sloupce 1, 3 (tj. bez reklamy a s reklamou v TV a novinách) a sloupce 2, 3 (tj. s reklamou jen v novinách a reklamou v TV a novinách).

Pravá strana Tukeyova vzorce je $\sqrt{\frac{0,1\overline{6}}{2}} \cdot q_{0,95}(3,2) = \sqrt{\frac{0,1\overline{6}}{2}} \cdot 8,33 = 2,4$.

Podle Tukeyovy metody se na hladině významnosti 0,05 také liší sloupce 1, 3 a sloupce 2, 3. Výhodnější je hodnota získaná Tukeyovou metodou, protože je menší.

Výpočet pomocí systému STATISTICA:

Načteme datový soubor baleni_a_reklama.sta o třech proměnných X, A, B a 6 případech, kde X – prodej, A – typ balení (1 – sáček, 2 – krabička), B – druh reklamy (1 – bez reklamy, 2 – reklama v novinách, 3 – reklama v TV a novinách).
Statistiky – ANOVA – ANOVA hlavních efektů – Rychlé nastavení – OK – Závisle proměnná X, Kategor. nezáv. prom. A, B – OK – Možnosti – Parametrizace – odškrtneme Sigma-omezená, zaškrtneme Bez abs. členu – OK. Dostaneme tabulku analýzy rozptylu dvojného třídění bez interakcí.

	Jednorozměrné testy významnosti pro X (baleni_a_reklama.sta Přeparametrizovaný model Dekompozice typu III				
Efekt	SČ	Stupně volnosti	PČ	F	p
A	10,66667	1	10,66667	64,0000	0,015268
B	37,00000	2	18,50000	111,0000	0,008929
Chyba	0,33333	2	0,16667		

Vidíme, že p-hodnota pro testovou statistiku F_A je 0,015268, tedy na hladině významnosti 0,05 zamítáme hypotézu, že typ balení nemá vliv na prodej zboží. Podobně p-hodnota pro testovou statistiku F_B je 0,008929, což znamená, že na hladině významnosti 0,05 zamítáme hypotézu, že druh reklamy nemá vliv na prodej zboží.
Abychom zjistili, které dvojice druhů reklamy se liší na hladině významnosti 0,05, použijeme Schefféo (resp. Tukeyovu) metodu mnohonásobného porovnávání.
Návrat do ANOVA Výsledky – Více výsledků – Post-hoc – Efekt B –Tukeyův HSD.

	Tukeyův HSD test; proměnná X (baleni_a_reklama.sta) Přibližné pravděpodobnosti pro post hoc testy Chyba: meziskup. PČ = ,16667, sv = 2,0000			
Č. buňky	B	{1} 2,0000	{2} 2,5000	{3} 7,5000
1	bez reklamy		0,548301	0,010156
2	reklama v novinach	0,548301		0,012218
3	reklama v TV a novinach	0,010156	0,012218	

Vidíme, že na hladině významnosti 0,05 se liší dvojice (1,3) a (2,3).

Výpočet pomocí systému SPSS:

Analyze – General Linear Model – Univariate – Dependent Variable X, Fixed Factor(s) A, B – Model – zaškrtneme cystom – vybereme Main effects – Model A, B – odškrtneme Include intercept in model – Continue – Post Hoc – Post hoc Tests for B – zaškrtneme Tukey – Continue – OK

Tabulka analýzy rozptylu dvojného třídění bez interakcí:

Tests of Between-Subjects Effects

Dependent Variable: prodej

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Model	143,667[a]	4	35,917	215,500	,005
A	10,667	1	10,667	64,000	,015
B	37,000	2	18,500	111,000	,009
Error	,333	2	,167		
Total	144,000	6			

a. R Squared = ,998 (Adjusted R Squared = ,993)

Testová statistika F_A pro test významnosti způsobu balení se realizuje hodnotou 64, p-hodnota 0,015, tedy na hladině významnosti 0,05 zamítáme hypotézu, že způsob balení neovlivňuje prodej. Testová statistika F_B nabývá hodnoty 111, p-hodnota je 0,009, tedy na hladině významnosti 0,05 zamítáme hypotézu, že způsob reklamy neovlivňuje prodej.

Výsledky Tukeyovy metody mnohonásobného porovnávání pro faktor B (způsob reklamy):

Multiple Comparisons

Dependent Variable: prodej
Tukey HSD

(I) druh reklamy	(J) druh reklamy	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
bez reklamy	reklama v novinach	-,50	,408	,548	-2,90	1,90
	reklama v TV a novinach	-5,50*	,408	,010	-7,90	-3,10
reklama v novinach	bez reklamy	,50	,408	,548	-1,90	2,90
	reklama v TV a novinach	-5,00*	,408	,012	-7,40	-2,60
reklama v TV a novinach	bez reklamy	5,50*	,408	,010	3,10	7,90
	reklama v novinach	5,00*	,408	,012	2,60	7,40

Based on observed means.

*. The mean difference is significant at the ,05 level.

Podle p-hodnot uvedených ve sloupci Sig. vidíme, že na hladině významnosti 0,05 se liší dvojice (bez reklamy, reklama v TV a novinách) a (reklamy, reklama v TV a novinách).

Dvojné třídění s interakcemi

Nyní předpokládáme, že faktory A a B se mohou ovlivňovat (např. některý způsob hnojení má zcela specifický vliv na určitý typ půdy). Náhodné veličiny X_{ijk} se řídí modelem

M0: $X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$ pro $i = 1, \dots, a$, $j = 1, \dots, b$, $k = 1, \dots, c$, přičemž γ_{ij} je interakce mezi faktorem A na úrovni i a faktorem B na úrovni j. V této situaci předpokládáme, že $c \geq 2$. Parametry μ, α_i, β_j neznáme. Požadujeme, aby platily tzv. reparametrizační rovnice:

$$\sum_{i=1}^{a}\alpha_i = 0, \sum_{j=1}^{b}\beta_j = 0, \sum_{i=1}^{a}\gamma_{ij} = 0, \sum_{j=1}^{b}\gamma_{ij} = 0.$$

Nyní můžeme utvořit modely

M1: $X_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$

M2: $X_{ijk} = \mu + \alpha_i + \varepsilon_{ijk}$

M3: $X_{ijk} = \mu + \varepsilon_{ijk}$

(Lze samozřejmě použít i jiný řetězec modelů, kdy postupně klademe rovny nule parametry α_i, β_j, γ_{ij} v jiném pořadí.)

Vypočítáme součty čtverců S_T, S_A, S_B, S_{AB}, S_E, přičemž

$S_{AB} = c\sum_{i=1}^{a}\sum_{j=1}^{b}\left[\left(M_{ij.} - M_{i..}\right) - \left(M_{.j.} - M_{...}\right)\right]^2$ je součet čtverců pro interakce,

počet stupňů volnosti $f_{AB} = (a-1)(b-1)$.

Vliv interakcí je prokázán na hladině významnosti α, když

$$F_{AB} = \frac{S_{AB}/f_{AB}}{S_E/f_E} \geq F_{1-\alpha}\left((a-1)(b-1), n-ab\right).$$

Výsledky zapisujeme do tabulky analýzy rozptylu dvojného třídění s interakcemi:

Zdroj variability	součet čtverců	st. vol.	podíl S/f	$F = \dfrac{S/f}{S_E/f_E}$
řádkový faktor A	S_A	$f_A = a-1$	S_A/f_A	$F_A = \dfrac{S_A/f_A}{S_E/f_E}$
sloupcový faktor B	S_B	$f_B = b-1$	S_B/f_B	$F_B = \dfrac{S_B/f_B}{S_E/f_E}$
interakce A,B	S_{AB}	$f_{AB} = (a-1)(b-1)$	S_{AB}/f_{AB}	$F_{AB} = \dfrac{S_{AB}/f_{AB}}{S_E/f_E}$
reziduální	S_E	$f_E = n-a-b+1$	S_E/f_E	-
celkem	S_T	$f_T = n-1$	-	-

Je třeba si povšimnout, že součet $S_{AB} + S_E$ resp. $f_{AB} + f_E$ dá hodnotu S_E resp. f_E v tabulce bez interakcí.

Možné problémy v analýze rozptylu dvojného třídění s interakcemi

a) Ukáže-li se vliv interakcí významný, vzniká otázka, zda testovat vliv řádků resp. sloupců pomocí tabulky s interakcemi nebo provést novou analýzu rozptylu, ale tentokrát bez interakcí. Převládá názor, že je zapotřebí dokončit analýzu rozptylu s interakcemi.
b) Pokud interakce vyjdou významné a řádky a sloupce rovněž, zpravidla se nedoporučuje provádět mnohonásobné porovnávání, protože by se mohlo stát, že některá interakce by byla mnohem výraznější než příslušný řádkový resp. sloupcový efekt.
c) Nejsou-li interakce významné a řádky resp. sloupce ano, pak lze provést mnohonásobné porovnávání zcela analogicky jako v případě třídění bez interakcí, avšak je jiný počet stupňů volnosti f_E.

Tabulka odhadů různých parametrů a rozptylů těchto odhadů

parametr	odhad	rozptyl odhadu
μ	M	σ^2/n
$\mu + \alpha_i$	$M_{i..}$	σ^2/bc
$\mu + \beta_j$	$M_{.j.}$	σ^2/ac
$\mu + \alpha_i + \beta_j + \gamma_{ij}$	$M_{ij.}$	σ^2/c
α_i	$M_{i..} - M$	$\sigma^2(a-1)/n$
β_j	$M_{.j.} - M$	$\sigma^2(b-1)/n$
γ_{ij}	$(M_{ij.} - M_{i..}) - (M_{.j.} - M)$	$\sigma^2(a-1)(b-1)/n$

Neznámý rozptyl σ^2 nahradíme jeho odhadem $s^2 = \dfrac{S_e}{n-ab}$.

Příklad:
Byly zkoumány výnosy sena (v q/ha) v závislosti na typu půdy (řádkový faktor A, úroveň 1 – normální půda, úroveň 2 – kyselá půda) a na způsobu hnojení (sloupcový faktor B, úroveň 1 – bez hnojení, úroveň 2 – hnojení chlévskou mrvou, úroveň 3 – hnojení vápenatým hnojivem). Každá kombinace faktorů A a B byla realizována čtyřikrát nezávisle na sobě. Výnosy sena jsou uvedeny v tabulce:

		B		
		1	2	3
A	1	28 32 30 30	37 36 39 36	34 38 37 36
	2	31 27 30 29	34 34 30 38	42 40 41 39

Na hladině významnosti 0,05 máme posoudit vliv typu půdy a způsobu hnojení (včetně případných interakcí) na výnosy sena.

Řešení:
Data zpracujeme pomocí analýzy rozptylu dvojného třídění s interakcemi. Přitom a = 2, b = 3, c = 4, n = abc = 24. Nebudeme provádět pomocné výpočty, ale rovnou uvedeme tabulku výsledků.

Zdroj variability	součet čtverců	st. vol.	podíl S/f	$F = \frac{S/f}{S_E/f_E}$
typ půdy	0,166	1	0,166	0,04
způsob hnojení	318,25	2	159,125	41,81
interakce	55,084	2	27,542	7,24
reziduální	68,5	18	3,8056	-
celkem	442	23	-	-

Odpovídající kvantily: pro řádkový efekt $F_{0,95}(1,18) = 4,41$, pro sloupcový efekt $F_{0,95}(2,18) = 3,55$, pro interakce $F_{0,95}(1,18) = 4,41$.
Protože $F_A = 0,04 < 4,41$, nezamítáme na hladině významnosti 0,05 hypotézu, že typ půdy neovlivňuje výnos sena.
Dále $F_B = 41,81 \geq 3,55$, tedy na hladině významnosti 0,05 se prokázal rozdíl mezi použitými způsoby hnojení.
Jelikož $F_{AB} = 7,24 \geq 4,41$, zamítáme na hladině významnosti 0,05 hypotézu o nevýznamnosti interakcí (tj. aspoň jeden způsob hnojení působí jinak na půdu normální než kyselou).

Výpočet pomocí systému STATISTICA:
Načteme datový soubor seno.sta se třemi proměnnými A, B, X a 24 případy, kde X – výnos sena, A – typ půdy, B – způsob hnojení. Jednotlivé varianty

proměnných A a B mají tento význam: u proměnné A je 1 – normální půda, 2 – kyselá půda, u proměnné B je 1 – bez hnojení, 2 – chlévská mrva, 3 – vápenaté hnojivo.

Nejprve spočítáme průměry ve všech 6 skupinách:
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 – zaškrtneme Bez absolutního členu – OK – Průměry – Pozorované, nevážené.

Č. buňky	A*B; Nevážené průměry (seno.sta) Současný efekt: F(2, 18)=7,2372, p=,00494 Dekompozice typu III						
	A	B	X Průměr	X Sm.Ch.	X -95,00%	X +95,00%	N
1	normální	bez hnojení	30,00000	0,975392	27,95078	32,04922	4
2	normální	chlévská mrva	37,00000	0,975392	34,95078	39,04922	4
3	normální	vápenaté hnojivo	36,25000	0,975392	34,20078	38,29922	4
4	kyselá	bez hnojení	29,25000	0,975392	27,20078	31,29922	4
5	kyselá	chlévská mrva	34,00000	0,975392	31,95078	36,04922	4
6	kyselá	vápenaté hnojivo	40,50000	0,975392	38,45078	42,54922	4

Nyní provedeme testování hypotéz o vlivu faktorů:
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 – zaškrtneme Bez absolutního členu – OK – Všechny efekty.
Dostaneme tabulku analýzy rozptylu dvojného třídění s interakcemi.

Efekt	Jednorozměrné testy významnosti pro X (seno.sta) Přeparametrizovaný model Dekompozice typu III				
	SČ	Stupně volnosti	PČ	F	p
A	0,1667	1	0,1667	0,04380	0,836585
B	318,2500	2	159,1250	41,81387	0,000000
A*B	55,0833	2	27,5417	7,23723	0,004938
Chyba	68,5000	18	3,8056		

Vidíme, že p-hodnota pro testovou statistiku F_B je velmi blízká 0, tedy na hladině významnosti 0,05 zamítáme hypotézu, že způsob hnojení nemá vliv na výnosy sena. Podobně p-hodnota pro testovou statistiku F_{AB} je 0,004938, což znamená, že na hladině významnosti 0,05 zamítáme hypotézu, že způsob hnojení působí na oba typy půd stejně.

Vzhledem k tomu, že rozsahy výběrů v daných šesti skupinách jsou větší než 1 (c = 4), lze ověřit předpoklad o homogenitě rozptylů. Vrátíme se do ANOVA Výsledky a zvolíme Více výsledků – Předpoklady – Leveneův test.

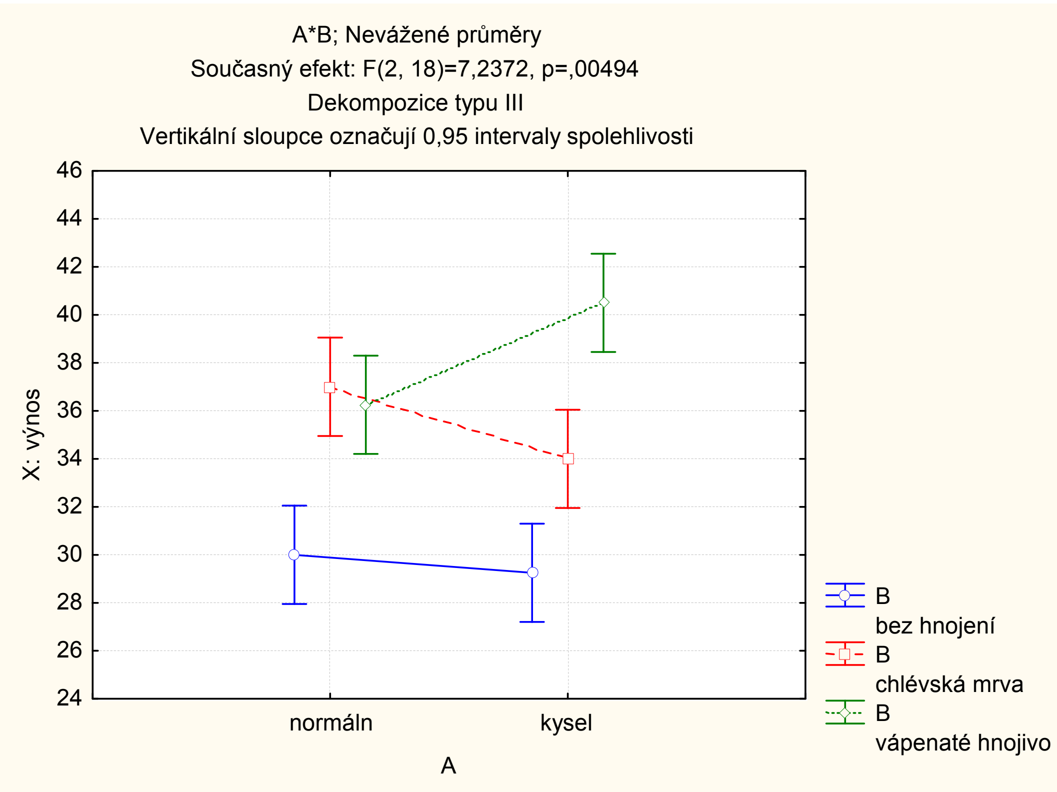
	Leveneův test homogenity rozptylů (seno.sta) Efekt: A*B Stupně volnosti pro všechna F: 5, 18			
	PČ Efekt	PČ Chyba	F	p
X	0,600000	1,555556	0,385714	0,852058

Zjistíme, že p-hodnota je 0,852058, tudíž tedy na hladině významnosti 0,05 nezamítáme hypotézu o homogenitě rozptylů.

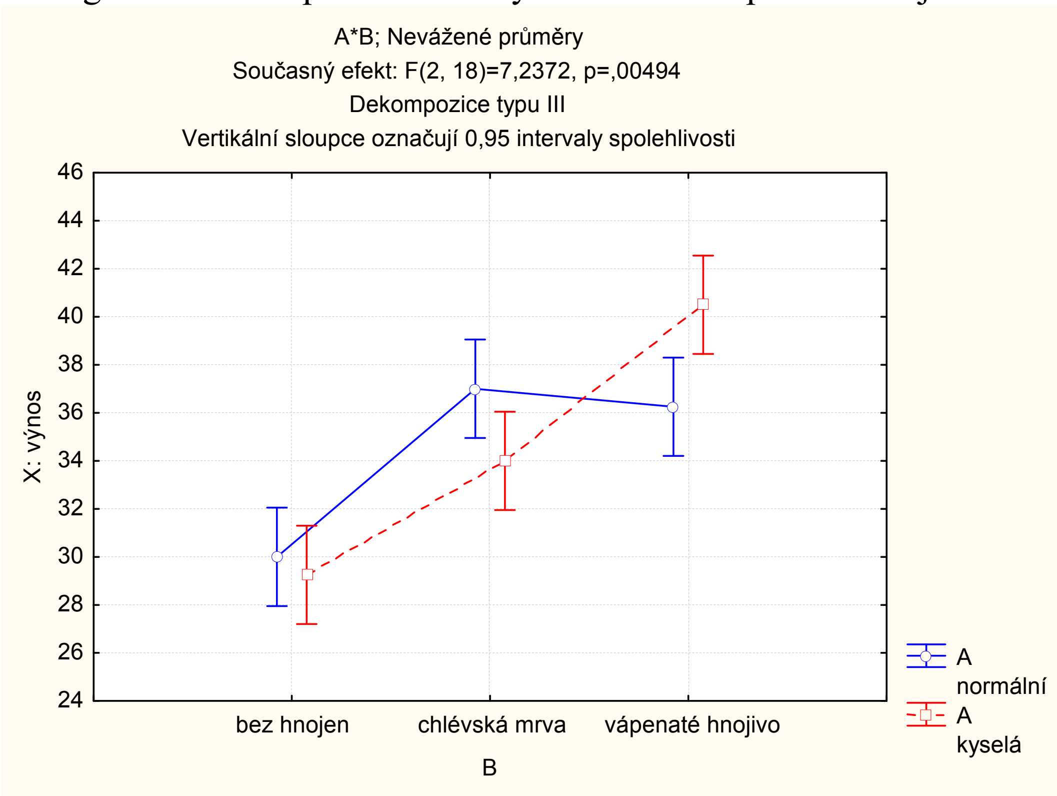
Normalitu všech šesti výběrů můžeme orientačně posoudit rovněž v Předpokladech pomocí N-P plotu.

Ve všech šesti případech lze konstatovat vcelku dobrou shodu s normálním rozložením.

Průměrné výnosy sena (spolu s 95% intervaly spolehlivosti) na normální a kyselé půdě při daných třech způsobech hnojení lze znázornit graficky. V ANOVA Výsledky zvolíme Průměry – Pozorované, nevážené – Graf. Lze vykreslit graf závislosti průměrného výnosu sena na typu půdy:



nebo graf závislosti průměrného výnosu sena na způsobu hnojení:



V grafu se objevuje křížení, které je typické pro případ, kdy působí interakce mezi faktory A, B.

Nyní pomocí Tukeyovy metody mnohonásobného porovnávání zjistíme, které dvojice výběrů se liší na hladině významnosti 0,05. Vrátíme se do Anova: Výsledky – klikneme na Více výsledků – Post hoc – Efekt B - Tukeyův HSD.

	Tukeyův HSD test; proměnná X (seno.sta) Přibližné pravděpodobnosti pro post hoc testy Chyba: meziskup. PČ = 3,8056, sv = 18,000			
Č. buňky	B	{1} 29,625	{2} 35,500	{3} 38,375
1	bez hnojení		0,000171	0,000149
2	chlévská mrva	0,000171		0,022412
3	vápenaté hnojivo	0,000149	0,022412	

Vidíme, že na hladině významnosti 0,05 se liší všechny tři dvojice skupin:

Výpočet pomocí systému SPSS:

Načteme datový soubor seno.sav.
Analyze – General Linear Model – Univariate – Dependent variable X, Fixed Factor(s) A, B – Model – odškrtneme Include intercept in model – Continue – Options – zaškrtneme Descriptive Statistics a Homogeneity tests – Continue – Plots – Horizontal Axis A, Separate Lines B – Add, Horizontal Axis B, Separate

Lines A – Add – Post Hoc – Post Hoc Tests for A, B – zaškrtneme Tukey – Continue – OK.

Tabulka průměrů a směrodatných odchylek ve zkoumaných šesti skupinách:

Descriptive Statistics

Dependent Variable: výnos

typ půdy	způsob hnojení	Mean	Std. Deviation	N
normalni	bez hnojeni	30,00	1,633	4
	chlevska mrva	37,00	1,414	4
	vapenate hnojivo	36,25	1,708	4
	Total	34,42	3,579	12
kysela	bez hnojeni	29,25	1,708	4
	chlevska mrva	34,00	3,266	4
	vapenate hnojivo	40,50	1,291	4
	Total	34,58	5,230	12
Total	bez hnojeni	29,63	1,598	8
	chlevska mrva	35,50	2,828	8
	vapenate hnojivo	38,38	2,669	8
	Total	34,50	4,384	24

Výsledky Levenova testu homogenity rozptylů:

Levene's Test of Equality of Error Variances[a]

Dependent Variable: výnos

F	df1	df2	Sig.
,386	5	18	,852

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: A+B+A * B

Na hladině významnosti 0,05 nelze zamítnout hypotézu o shodě rozptylů v daných šesti skupinách, protože p-hodnota je 0,852, což je větší než 0,05.

Tabulka analýzy rozptylu dvojného třídění s interakcemi:

Tests of Between-Subjects Effects

Dependent Variable: výnos

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Model	28939,500[a]	6	4823,250	1267,423	,000
A	,167	1	,167	,044	,837
B	318,250	2	159,125	41,814	,000
A * B	55,083	2	27,542	7,237	,005
Error	68,500	18	3,806		
Total	29008,000	24			

a. R Squared = ,998 (Adjusted R Squared = ,997)

Na hladině významnosti 0,05 se neprokázal vliv typu půdy (p-hodnota = 0,837), avšak prokázal se vliv způsobu hnojení (p-hodnota = 0,000) a vliv interakcí (p-hodnota = 0,005).

Tukeyova metoda mnohonásobného porovnávání pro způsob hnojení:

Multiple Comparisons

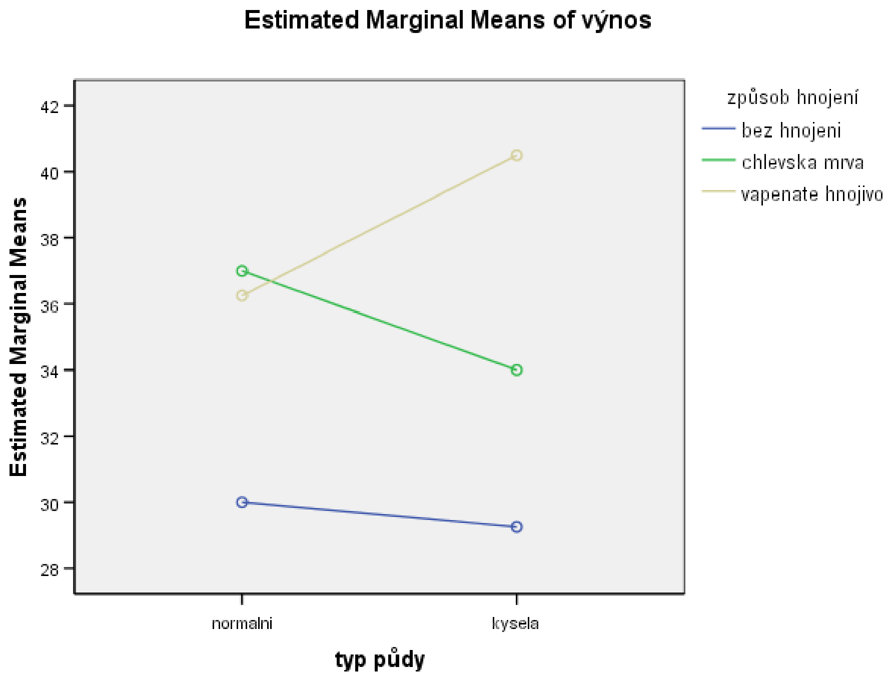
Dependent Variable: výnos
Tukey HSD

(I) způsob hnojení	(J) způsob hnojení	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
bez hnojeni	chlevska mrva	-5,88*	,975	,000	-8,36	-3,39
	vapenate hnojivo	-8,75*	,975	,000	-11,24	-6,26
chlevska mrva	bez hnojeni	5,88*	,975	,000	3,39	8,36
	vapenate hnojivo	-2,88*	,975	,022	-5,36	-,39
vapenate hnojivo	bez hnojeni	8,75*	,975	,000	6,26	11,24
	chlevska mrva	2,88*	,975	,022	,39	5,36

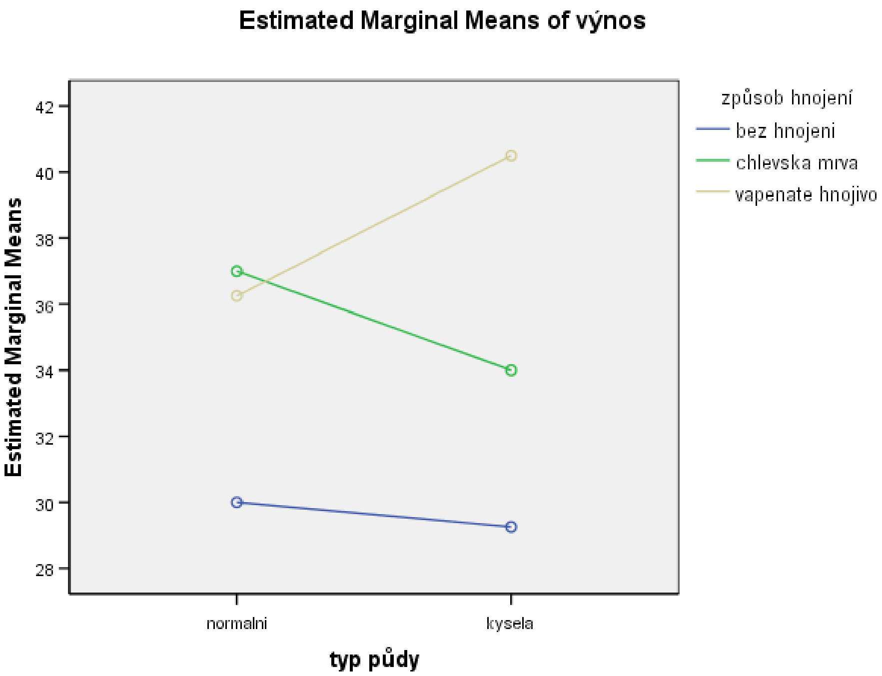
Based on observed means.
*. The mean difference is significant at the ,05 level.

Na hladině významnosti 0,05 se liší všechny tři dvojice způsobu hnojení.

Graf závislosti průměrného výnosu sena na typu půdy:



Graf závislosti průměrného výnosu sena na způsobu hnojení:



Estimated Marginal Means of výnos
Estimated Marginal Means
42
40
38
36
34
32
30
28
normalni
kysela
typ půdy
způsob hnojení
bez hnojeni
chlevska mrva
vapenate hnojivo