Analýza variance (ANOVA) dvoucestná, faktory s pevnými efekty; interakce; mnohonásobná srovnání

Pokud máme experiment, kdy sledujeme více faktorů, je již třeba dbát na to, který z nich má pevné a který náhodné efekty. Např. při sledování efektu zálivky a hnojení se bude jednat o ***dvoucestnou ANOVu s faktory s pevnými efekty*** (Typ I.). Tohoto typu je většina experimentů u nichž jsou úrovně dané experimentátorem. Úrovně jsou předem jasné a nejsou statistickým výběrem z více možností jako je tomu u faktorů s náhodným efektem (např. geneticky podmíněná variabilita organismů...).

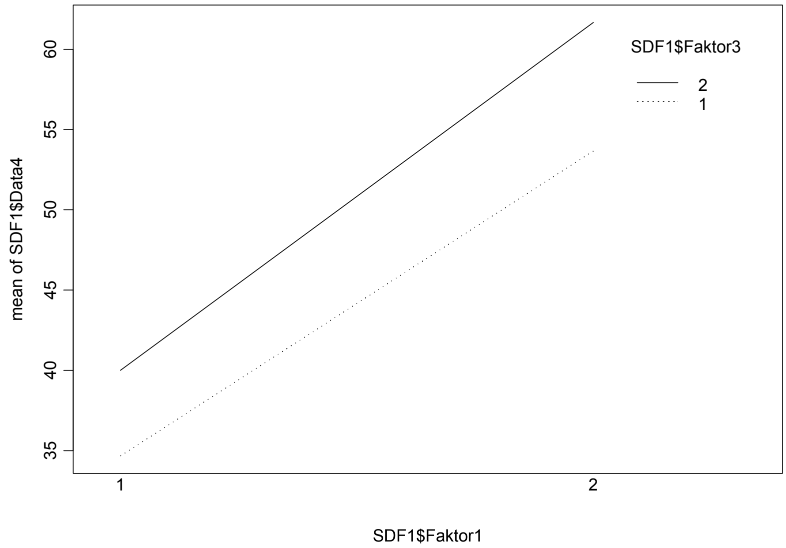
zálivka –zal	1	1	1	1	1	1	2	2	2	2	2	2
hnojení -hnoj	1	1	1	2	2	2	1	1	1	2	2	2
výška rostlin [cm] Data4	33	35	36	38	40	42	52	53	56	63	62	60

Analýzu budeme počítat z menu Statistcs>ANOVA>Fixed Effects...
Vzorec pro model, který zahrnuje i interakce, je v S+ `Data4~Faktor1*Faktor3`. To je zkrácený zápis pro `Data4~zal+hnoj+zal:hnoj`. Vzorec Faktor1*Faktor3 vyjadřuje veškeré kombinace, tedy i interakci. Samotná interakce mezi F1 a F2 se v S+ zapisuje `F1:F2`. Tedy obecný zápis modelu `Data4~zal*hnoj` zahrnuje jak efekty jednotlivých faktorů tak i jejich interakce.

```
              Df Sum of Sq  Mean Sq  F Value     Pr(F)
       zal     1  1240.333 1240.333 381.6410 0.0000000
       hnoj    1   133.333  133.333  41.0256 0.0002080
   zal:hnoj    1     5.333    5.333   1.6410 0.2360697
  Residuals    8    26.000    3.250
```

Výsledek ukazuje, že jednotlivé působení obou faktorů je vysoce průkazné, avšak jejich vzájemná interakce vychází neprůkazná. Tedy zálivka i hnojení jako samostatný faktor zvýší růst rostlin a jejich společné působení je aditivní (jejich vliv na růst se sčítá).
Dalším krokem by tedy mělo být zjednodušení modelu. Stejně jako v postupné regresi použijeme funkci update. Pokud jsme předchozí model uložili pod jménem pokus01 a zjednodušený bude pokus02, pak voláme: `pokus02<-update (pokus01, .~. –zal:hnoj)`.
Nyní porovnáme oba modely funkcí `anova (pokus01, pokus02)` a ověříme, zda se je nový model významně horší než původní. Poté si výsledek zobrazíme pomocí funkce `summary`.

Grafické zobrazení průměrů pro jednotlivé faktory i s interakcemi vyvoláme příkazem v "Commands" okně: `interaction.plot(SDF1$zal, SDF1$hnoj, SDF1$Data4)`, kde funkce `interaction.plot` má tři parametry (první faktor, druhý faktor, závislá proměnná).



Úsečky vymezené příslušnými průměry jsou přibližně rovnoběžné, což znamená, že přírůstek rostlin vlivem druhého faktoru je zhruba stejný v obou hladinách faktoru prvního.

Pokud je však působení dvou faktorů multiplikativní, což znamená, že jejich efekt na studovaný parametr se v podstatě násobí. V následujícím příkladu jsme v minulém datovém souboru zvětšili poslední tři hodnoty o 10.

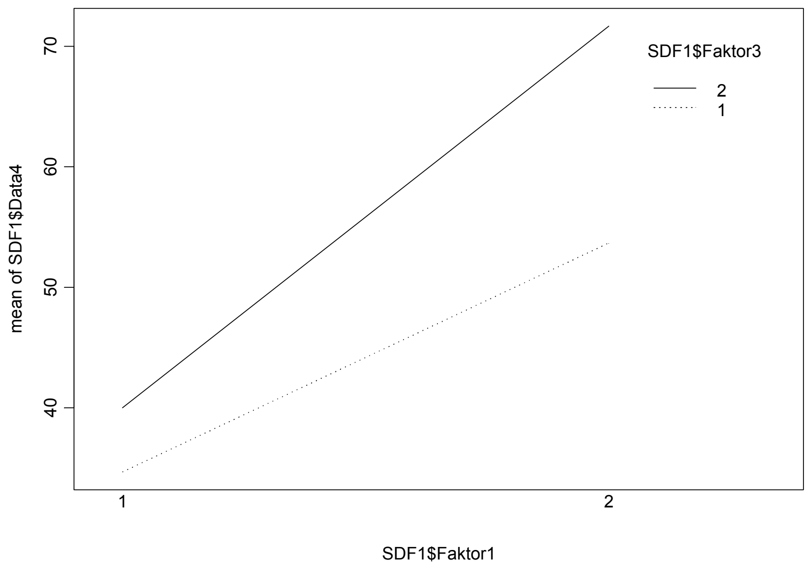
zálivka F1	1	1	1	1	1	1	2	2	2	2	2	2
hnojení F3	1	1	1	2	2	2	1	1	1	2	2	2
výška rostlin [cm] Data4	33	35	36	38	40	42	52	53	56	73	72	70

Opět voláme ANOVu s pevnými efekty a zadáme obecný model s interakcemi `Data4 ~ Faktor1 * Faktor3`.

```
                Df Sum of Sq  Mean Sq  F Value        Pr(F)
        Faktor1  1  1925.333 1925.333 592.4103 0.0000000087
        Faktor3  1   408.333  408.333 125.6410 0.0000035986
Faktor1:Faktor3  1   120.333  120.333  37.0256 0.0002943417
      Residuals  8    26.000    3.250
```

Výsledkem je opět vysoce průkazná ANOVA pro oba faktory. Oproti minulému příkladu je však průkazná i interakce mezi oběma faktory (Faktor1:Faktor3).

Grafické zobrazení interakcí ukazuje na nerovnoběžnost úseček a tedy na multiplikativní efekt obou faktorů. Pokud se úsečky rozbíhají jde o pozitivní interakci (synergismus). Opačně jde o negativní typ multiplikativní interakce (interference).



Mnohonásobná srovnání pokud je signifikantí interakce: zaměříme se na srovnání průměrů každého faktoru odděleně na každé úrovni druhého faktoru a vice versa. Poté analýza postupuje stejně jako pro jednocestnou ANOVu (viz příloha tohoto materiálu). V případě nesignifikantní interakce počítáme mnohonásobná srovnání pro každý signifikantní faktor. Více o pozadí ANOVy například v knize Experiments in Eology; A. J. Underwood 1997.

Chceme-li průměry pro jednotlivé kombinace faktorů, pak voláme stejně jako v minulém příkladě funkci `tapply`. Jedinou změnou bude použití více „shlukovacích“ proměnných. To se řeší použitím funkce `list`, viz příklad: `tapply (vaha, list(zal, hnoj), mean)`.

Výpočet stupňů volnosti v tabulce ANOVY:

	df
Faktor A	a – 1; (a – počet hladin faktoru A)
Faktor B	b – 1; (b – počet hladin faktoru B)
Interakce A:B	(a – 1)(b – 1)
Error	ab(n – 1); (n – počet opakování v rámci kombinace ošetření)
Total	abn – 1

Tento materiál byl vytvořen na základě příkladů a textů P. Sklenáře pro kurz biostatistiky v NCSS.

Mnohonásobná srovnání u ANOVY se dvěma faktory s pevnými efekty
Příklad s faktorem zal (např. zálivka se dvěma hladinami – sucho/mokro) a faktorem hnoj (hnojení se třema hladinami –bezN/stredN/hodneN). Pro každou kombinaci je k dispozici pět opakování. Závislou proměnnou je výška rostliny (vyska).

Interakce mezi faktory je signifikatní
srovnáváme průměry každého faktoru odděleně na každé úrovni druhého faktoru. To provedeme i pro prohozené faktory (viz tabulky 2 a 3). Počítáme vlastně několik jednoduchých (jednofaktorových) ANOV, kdy vybereme potřebné řádky (v S+ můžeme použít při tvorbě modelů jednoduchých podmínek umístěných v hranatých závorkách. Například pro vliv hnojení při malé zálivce voláme:
`model01<-aov(vyska~hnoj, subset=(zal=="sucho"))`. Pro tento model pak spočteme mnohonásobná srovnání. (Pozn. pokud jsou hladiny faktoru kódovány slovy či nečíselnými znaky je třeba je uzavřít do uvozovek.)

Tab. 1 - Tabulka průměrů je pro každou kombinaci ošetření:

		Hnojení		
		Málo	Středně	Hodně
Zálivka	Sucho	0,12	0,11	0,24
	Mokro	0,17	0,47	0,65

Tab. 2 – mnohonásobná porovnávání: Efekt hnojení při každé hladině zálivky

	Hnojení při malé zálivce			Hnojení při vysoké zálivce		
hnojení	málo	středně	hodně	málo	středně	hodně
	0,12	0,11	0,24	0,17	0,47	0,65
rozdíly:	a	a	b	a	b	c

Tab. 3 – mnohonásobná porovnávání: Efekt zálivky pro každou hladinu hnojení

	Zálivka při málo hnojení		středním hnojení		vysokém hnojení	
zálivka	sucho	mokro	sucho	mokro	sucho	mokro
	0,12	0,17	0,11	0,47	0,24	0,65
rozdíly:	a	a	a	b	a	b

Interakce mezi faktory je nesignifikantní
Pokud není interakce mezi hlavními faktory signifikantní, pak mnohonásobná srovnání provádíme pro každý faktor (samozřejmě jen pro ty, kde byla hodnota F statistiky dostatečně signifikantní) zvlášť. Vyloučená interakce vlastně potvrzuje, že jednotlivé faktory jsou na sobě nezávislé a tak můžeme hladiny pro srovnávání průměrovat. Což znamená, že spočteme opět jednoduché ANOVy, kde budeme srovnávat každý faktor zvlášť. Pokud by data v předchozím příkladu ukazovala na nesignifikantní interakci mezi zálivkou a hnojením, pak jednoduše spočteme mnohonásobná srovnání pro dvě ANOVy (první: `model01<-aov(vyska~zal)`) a druhou (`model02<-aov(vyska~hnoj)`). Pro každou spočteme mnohonásobná porovnání.