

Analyza rozptylu, 2 faktory

Menu:	QCExpert	Anova	Dvoufaktorová
-------	----------	-------	---------------

Anova pro 2 faktory je rozšířením jednofaktorové analýzy rozptylu popsané v předchozím odstavci. Analyzuje se vliv dvou faktorů s pevnými efekty na výsledek pozorování. Výsledky pozorování Z_{ij} při n_X různých úrovních faktoru X a n_Y různých úrovních faktoru Y lze popsat modelem Anova pro 2 faktory X, Y :

$$Z_{ij} = Z_0 + \alpha_i X_i + \beta_j Y_j + \lambda_{ij} + \varepsilon_{ij}, \quad i = 1, \dots, n_X, \quad j = 1, \dots, n_Y,$$

kde Z_0 je absolutní člen (celková střední hodnota), α a β jsou vektory příspěvků jednotlivých úrovní, prvky matice λ_{ij} se nazývají interakce obou faktorů a ε_{ij} je náhodná chyba s normálním rozdělením a definičně nulovou střední hodnotou, $\varepsilon \sim N(0, \sigma^2)$. Definičně je dáno $\sum \alpha_i = 0$, $\sum \beta_j = 0$, $\sum \lambda_{i(j)} = 0$, $\sum \lambda_{j(i)} = 0$.

Data a parametry

Data jsou uspořádána ve 3 sloupcích. Dva sloupce obsahují jména úrovní prvního a druhého faktoru (na pořadí faktorů nezáleží). Úrovně jsou v tabulce zapsány jako textové hodnoty, například: „malý, střední, velký“. Třetí sloupec obsahuje číselné hodnoty pozorované pro odpovídající kombinace úrovní faktorů. Oba faktory mohou mít dvě nebo více úrovní. Pro každou kombinaci úrovní musí být zadáno aspoň jedno pozorování.

Příklad uspořádání dat v datové tabulce Faktor X „Odrůda“ má 3 úrovně: $X_1 = \text{Brazil}$, $X_2 = \text{Dlouholistý}$, $X_3 = \text{Cassablanca}$. Faktor Y „Hnojivo“ má 2 úrovně: $Y_1 = \text{NPK}$, $Y_2 = \text{Ledek Ch}$. Pro každou kombinaci úrovní máme k dispozici 1 pozorování Z_{ij} nazvané „Výtěžek“.

Odrůda	Hnojivo	Výtěžek
Brazil	NPK	14.6
Brazil	Ledek Ch	17.4
Dlouholistý	NPK	13.3
Dlouholistý	Ledek Ch	12.6
Cassablanca	NPK	17.5
Cassablanca	Ledek Ch	14.9

Jedna kombinace úrovně prvního a druhého faktoru se nazývá cela, nebo buňka. Každá cela musí obsahovat alespoň jedno pozorování. Je-li ve všech celách stejný počet pozorování, jedná se o vyvážený plán, resp. vyváženou analýzu rozptylu. V opačném případě se jedná o analýzu rozptylu nevyváženou. Podle počtu měření při jednotlivých kombinacích, tedy počtu měření v cele lze rozlišit 3 typy analýzy rozptylu:

1 pozorování v každé cele – *Vyvážená analýza rozptylu bez opakování*

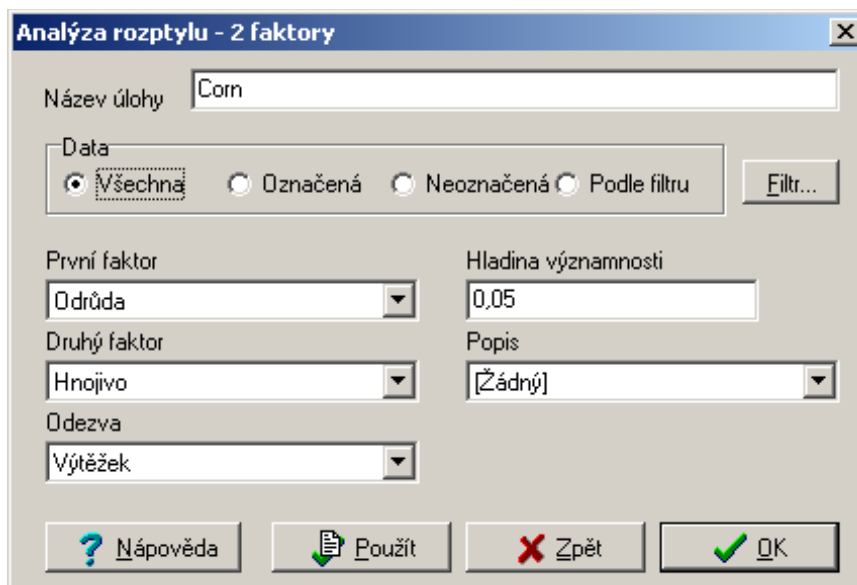
stejný počet $n > 1$ pozorování v každé cele – *Vyvážená analýza rozptylu s opakováním*

nestejný počet $n_{ij} > 0$ pozorování v každé cele – *Nevyvážená analýza rozptylu*

V dialogovém okně se vybere první a druhý faktor a pozorovaná odezva, zvolí se hladina významnosti, na níž se budou provádět testy (obvyklá hodnota je 0.05). Po stisku *OK* se provede analýza a výsledky se zapíší do okna Grafy a do Protokolu.

Omezení – maximální počet úrovní kteréhokoliv faktoru je 32, $n_X \leq 32$, $n_Y \leq 32$.

Při větším počtu úrovní, nebo více faktorech použijte modul *Zobecněná analýza rozptylu*.



Obrázek 1 ANOVA 2 faktory - Vstupní dialogový panel

Program automaticky rozezná o který typ se jedná a podle toho provede výpočty. Výstup je popsán v následujících odstavcích.

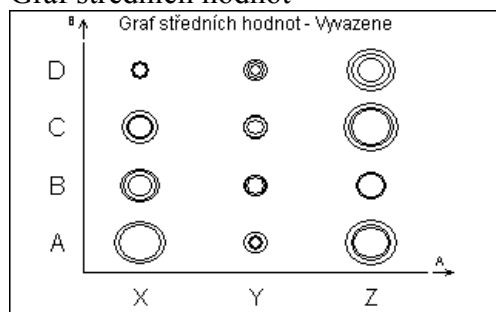
Protokol

Analýza rozptylu 2 faktory	
Typ modelu	Automaticky rozpoznáný typ analýzy: - vyvážený bez opakování, nebo - vyvážený s opakováním, nebo - nevyvážený
Faktory úrovně	Názvy obou faktorů a jejich úrovně
Počet opakování	V případě vyváženého modelu Anova: počet opakování v celách n_0 v případě nevyváženého modelu počty opakování pro každou celu, n_{ij}
Tabulka středních hodnot	Tabulka aritmetických průměrů v každé cele, tedy pro každou kombinaci úrovní
Průměry pro faktor	Celkové průměry při každé úrovni jednoho faktoru bez ohledu na úroveň druhého faktoru.
Celkový průměr	Celkový průměr všech hodnot. Tato hodnota by odpovídala střední hodnotě odezvy, kdyby na ni neměl vliv žádný faktor.
Parametry modelu	Odhady příspěvků jednotlivých úrovní obou faktorů
Úroveň faktoru	Názvy všech úrovní prvního a druhého faktoru
Příspěvek úrovně	Odhady příspěvků úrovní α , β
Tabulka ANOVA	Výsledná tabulka analýzy rozptylu

Zdroj variability	Název zdroje variability, jehož velikost významnost se analyzuje
1. faktor	Hodnocení variability 1. faktoru
2. faktor	Hodnocení variability 2. faktoru
Interakce	Hodnocení variability interakce faktorů
Rezidua	Hodnocení variability reziduí
Celkem	Hodnocení celkové variability
Součet čtverců	Součet čtverců odchylek příslušejících jednotlivým zdrojům
Průměrný čtverec	Průměrný čtverec odchylek příslušejících jednotlivým zdrojům, odpovídá rozptylu
Stupně volnosti	Počet stupňů volnosti příslušející jednotlivým zdrojům
Směr. odch.	Směrodatná odchylka příslušející jednotlivým zdrojům
F-kritérium	Hodnota testovacího F-kritéria vypočtená z dat pro první a druhý faktor a pro interakce
Kritický kvantil	Kritická hodnota F-rozdělení, je-li F-kritérium větší, než kritický kvantil, je příslušný parametr významným zdrojem variability.
Závěr	Slovní formulace výsledku F-testu: Významný, nebo Nevýznamný
p-hodnota	Odpovídající p-hodnota F-testu.

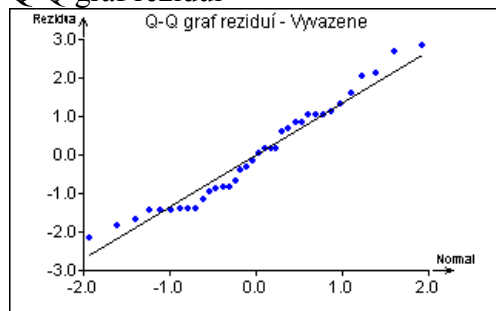
Graf

Graf středních hodnot



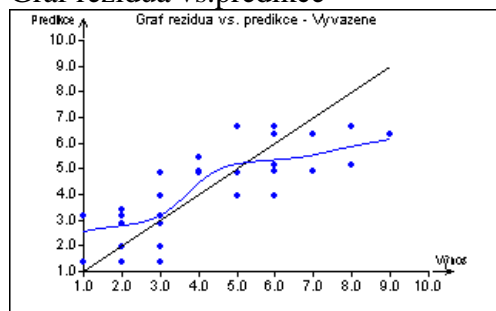
Graf středních hodnot vizuálně porovnává hodnoty odezvy pro jednotlivé úrovně faktorů. Velikost pozorované odezvy je vyjádřen velikostí kružnice. Délka kružnic je přibližně úměrná velikosti odezvy. Na první pohled jsou patrné rozdíly mezi hodnotami v jednotlivých celcích. Statistická významnost musí být ovšem potvrzena ve sloupci Závěr v protokolu.

Q-Q graf reziduí



Q-Q graf reziduí slouží k posouzení normality reziduí, případně k odhalení vybočujících měření. V případě normálního rozdělení reziduí bez vybočujících hodnot leží všechny body přibližně na černé přímce.

Graf rezidua vs. predikce

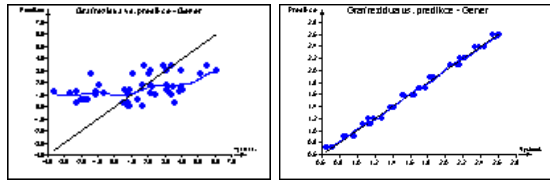


Graf rezidua vs. predikce vyjadřuje efektivitu (těsnost proložení, fit) modelu Anova. V případě nevýznamnosti modelu leží body spíše na vodorovné linii (ilustrace A), čím více se body přimykají k černé přímce $y = x$, tím je model Anova přesnější a významnější (ilustrace B). Přibližný průběh střední hodnoty predikce je znázorněna modrou křivkou.

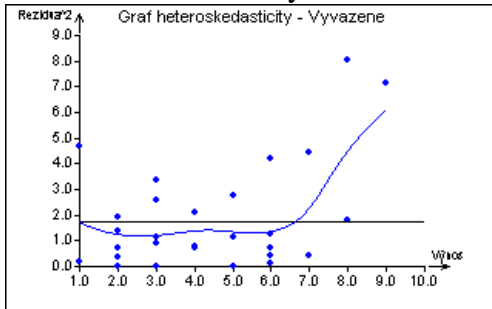
Ilustrace:

A

B

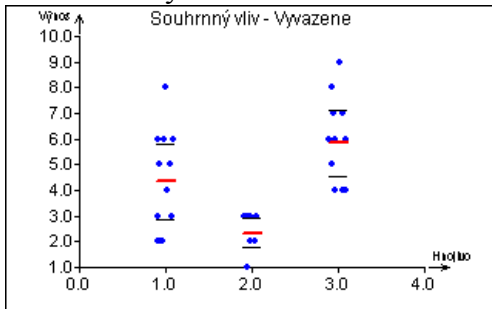


Graf heteroskedasticity



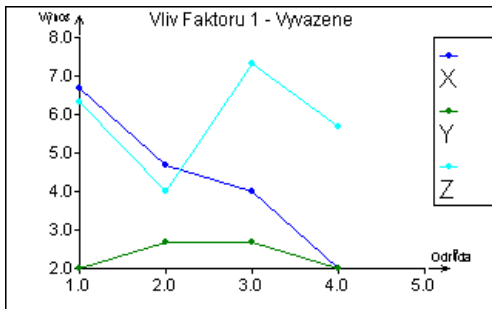
Graf heteroskedasticity graficky znázorňuje závislost velikosti chyby na pozorované hodnotě. Model Anova předpokládá konstantní chybu (konstantní σ^2). Modrá křivka neparametrického vyhlazení čtverce reziduí by měla ležet přibližně na černé vodorovné přímce, která představuje odhad σ^2 . Odchyluje-li se křivka od konstantního σ^2 , data vykazují heteroskedasticitu (nekonstantnost rozptylu) a výsledky analýzy jsou méně spolehlivé.

Graf Souhrnný vliv



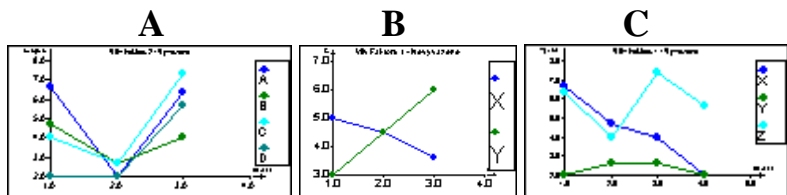
Tento graf vyjadřuje souhrnný vliv faktoru na odezvu. Pro jednotlivé úrovně faktoru znázorňuje jednotlivé hodnoty odezvy a střední hodnoty odezvy s přibližnými intervaly spolehlivosti. Je-li faktor významný (viz protokol), jsou vyznačené střední hodnoty v grafu vyznačeny červeně.

Graf vlivu faktoru

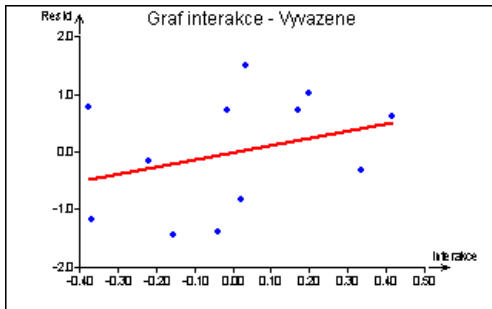


Graf vlivu faktoru vyjadřuje střední hodnoty faktoru při různých hodnotách druhého faktoru. Mají-li v obou grafech vlivu faktoru všechny křivky podobný průběh (ilustrace A), má faktor vliv na odezvu. Vykazují-li křivky výrazně opačný průběh (ilustrace B), svědčí to o významné interakci. Jsou-li křivky uspořádány náhodně, chaoticky (ilustrace C), nemá faktor výrazný vliv. Jedná se však pouze o kvalitativní posouzení, tento graf nenahrazuje testy významnosti v protokolu.

Ilustrace:



Graf vlivu interakce



Graf vlivu interakce znázorňuje graficky významnost interakce. Vykazují-li body rostoucí tendenci, projevuje se v modelu vliv interakce. Statistická významnost vlivu interakce je vyjádřena barvou přímky. Je-li interakce statisticky významná (viz protokol), je přímka vyznačena červeně.