TECHNICKÁ UNIVERZITA V LIBERCI

Nové možnosti rozvoje vzdělávání na Technické univerzitě v Liberci

Specifický cíl A2: Rozvoj v oblasti distanční výuky, online výuky a blended learning

NPO_TUL_MSMT-16598/2022

Plánování průmyslových experimentů

doc. Ing. Vladimír Bajzík, Ph.D.

Financováno Evropskou unií NextGenerationEU

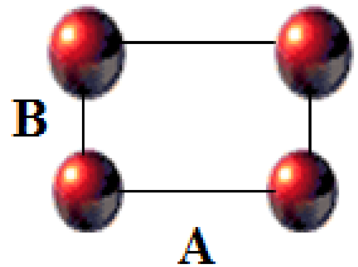
Národní plán obnovy

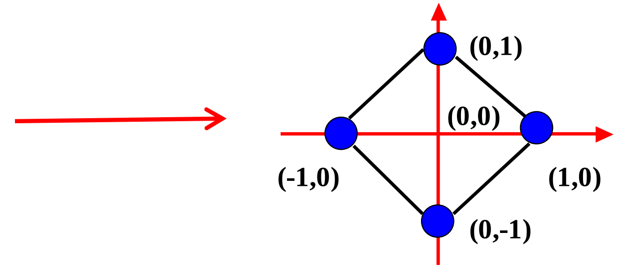
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

Plánování průmyslových experimentů II

Vladimír Bajzík

Faktorové plány

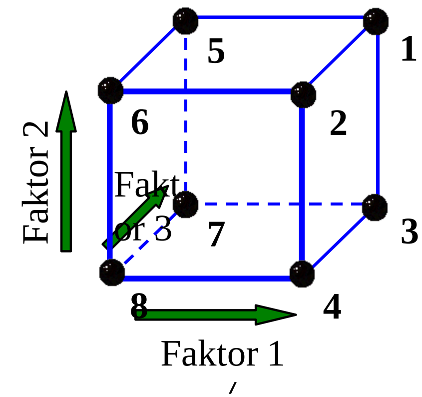




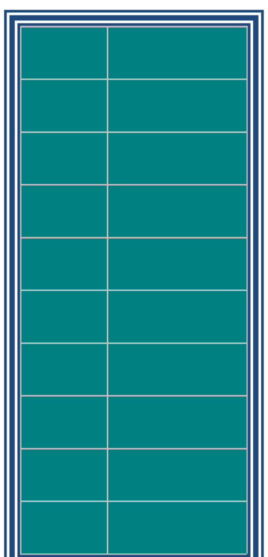
- Standardní pro složitější modely (interakce) a pro regresní analýzu

N^M^ M faktorů - nejjednodušší **2²**

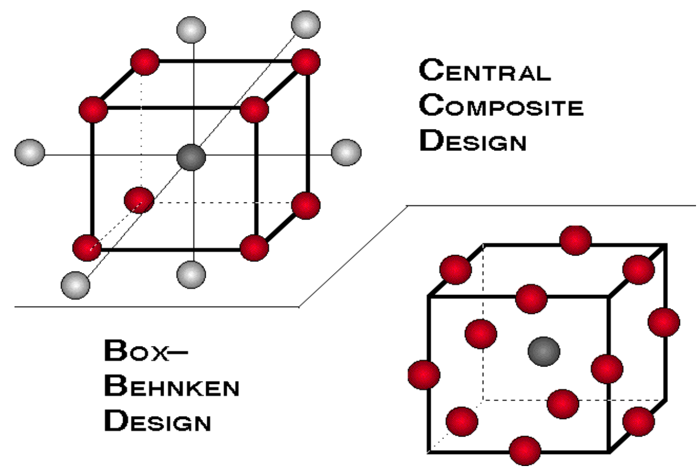
na N úrovních



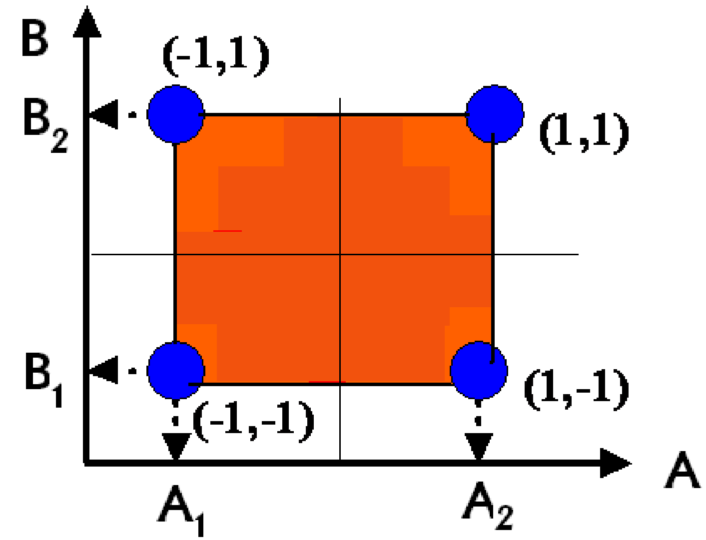
Charakteristika:
a) Vždy všechny kombinace úrovní všech faktorů.
b) Systematické mapování s minimálním počtem pokusů.
c) Výhodou je ortogonalita matice plánu



m	efekty
$\binom{m}{2}$	párové interakce
$\binom{m}{2}$	interakce 2 hého řádu
.	.
$\binom{m}{2}$	interakce m-1 ního řádu



Faktorové plány 2^2



Efekt faktoru = průměrná změna odezvy při změně úrovně faktoru

$$\hat{A} = \frac{y_{21} + y_{22}}{2} - \frac{y_{11} + y_{12}}{2}$$

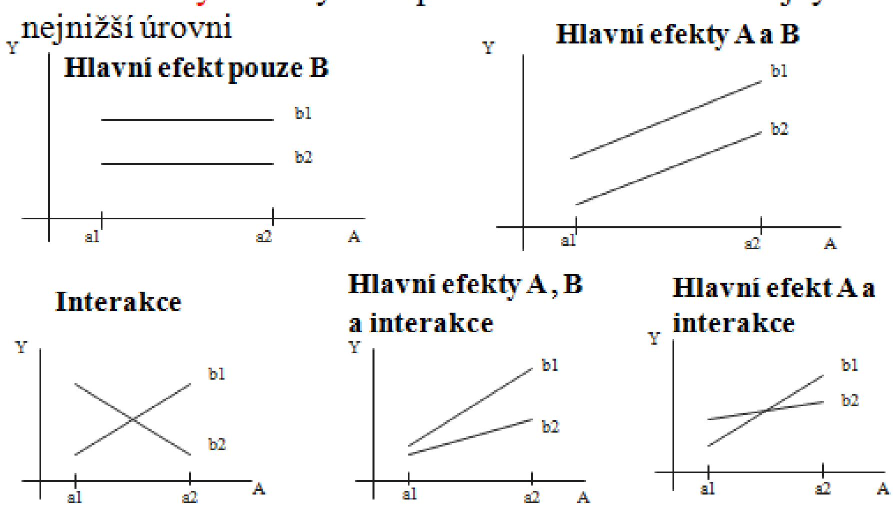
$$\hat{B} = \frac{y_{12} + y_{22}}{2} - \frac{y_{11} + y_{21}}{2}$$

Interakce mezi faktory = společné působení obou faktorů na velikost odezvy

A	B	y
-1	-1	y_{11}
-1	1	y_{12}
1	-1	y_{21}
1	1	y_{22}

Nastává, je-li rozdíl v odezvě mezi dvěma úrovněmi jednoho faktoru výrazně jiný při různých úrovních druhého faktoru, např. y_{21} - y_{11} a y_{22} - y_{12}.

Hlavní efekty: rozdíly mezi průměrnou odezvou na nejvyšší a nejnižší úrovni



	B1	B2
A1	y11	y12
A2	y21	y22

Faktorové plány 2^2

Statistický model pro dva faktory A, B:

$$y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij}$$

- y_{ij}: odezva
- μ: společná úroveň
- α_i: vliv i-té úrovně faktoru A, i=1,2,..., k; $\sum_{i=1}^{k} \tau_i = 0$
- β_j: vliv j-té úrovně faktoru B, j=1,2,... r; $\sum_{j=1}^{r} \beta_j = 0$
- $(\alpha\beta)_{ij}$: vliv interakce i-té a j-té úrovně faktorů A a B
- ϵ_{ij}: náhodná chyba při i-té a j-té úrovni faktorů A a B

Obecně: na každé kombinaci úrovní lze provést q opakování měření – s opakováním

Faktorové plány 2^2

Matice modelu = návrh kombinací úrovní faktorů

test	A	B	AB
1	-	-	+
2	+	-	-
3	-	+	-
4	+	+	+

Faktorové plány 2^2

Matice modelu = návrh kombinací úrovní faktorů

test	A	B	AB	y
1	-	-	+	y11
2	+	-	-	y21
3	-	+	-	y12
4	+	+	+	y22

	B1	B2
A1	y11	y12
A2	y21	y22

	B-		B+		
A-	y11 y12	$\bar{y}_{A-B-}$	y31 y32	$\bar{y}_{A-B+}$	$\bar{y}_{A-}$
A+	y21 y22	$\bar{y}_{A+B-}$	y41 y42	$\bar{y}_{A+B+}$	$\bar{y}_{A+}$
	$\bar{y}_{B-}$		$\bar{y}_{B+}$		

$$\hat{A} = \bar{y}_{A+} - \bar{y}_{A-}$$

$$\bar{y}_{A+} = \frac{y_{21} + y_{22}}{2}$$

$$\bar{y}_{A-} = \frac{y_{11} + y_{12}}{2}$$

$$\hat{B} = \bar{y}_{B+} - \bar{y}_{B-}$$

$$\widehat{AB} = \bar{y}_{AB+} - \bar{y}_{AB-}$$

$$\bar{y}_{AB-} = \frac{y_{21} + y_{12}}{2}$$

$$\bar{y}_{AB+} = \frac{y_{11} + y_{22}}{2}$$

Faktorové plány 2^2

Matice modelu = návrh kombinací úrovní faktorů

test	A	B	AB	y
1	-	-	+	y11, y12
2	+	-	-	y21, y22
3	-	+	-	y31, y32
4	+	+	+	y41, y42

	B1	B2
A1	y11, y12	y31, y32
A2	y21, y22	y41, y42

$$\hat{A} = \bar{y}_{A+} - \bar{y}_{A-}$$

$$\bar{y}_{A-} = \frac{y_{11} + y_{12} + y_{31} + y_{32}}{4}$$

$$\bar{y}_{A+} = \frac{y_{21} + y_{22} + y_{41} + y_{42}}{4}$$

$$\hat{B} = \bar{y}_{B+} - \bar{y}_{B-}$$

$$\widehat{AB} = \bar{y}_{AB+} - \bar{y}_{AB-}$$

$$\bar{y}_{AB-} = \frac{y_{21} + y_{22} + y_{31} + y_{32}}{4}$$

$$\bar{y}_{AB+} = \frac{y_{11} + y_{12} + y_{41} + y_{42}}{4}$$

Faktorové plány 2^2

Účinek koncentrace a teploty na viskozitu výsledného produktu

Odezva: požadované viskozity je dosahováno při době reakce mezi 7 až 9 hodinami => budeme sledovat dobu reakce v závislosti na úrovních faktorů

Faktor: koncentrace (A), teplota (B), 2 úrovně

Počet replikací: 2

Počet měření: $2^2 \times 2 = 8$

Experiment: **2^2 (2 faktory, 2 úrovně)**

Faktor A: koncentrace 42% a 48%

Faktor B: teplota 175°C a 195°C

test	A	B	AB
1	-	-	+
2	+	-	-
3	-	+	-
4	+	+	+

Náhodné uspořádání experimentů

	A	B	y
1	+	-	9.3
2	-	+	5.5
3	-	+	6.5
4	-	-	9.0
5	+	+	1.3
6	-	-	9.0
7	+	+	1.8
8	+	-	8.0

Faktorové plány 2^2

A	B	AB	y	y
-	-	+	9.0	9.0
+	-	-	9.3	8.0
-	+	-	5.5	6.5
+	+	+	1.8	1.3

$$\hat{A} = \frac{9.3 + 8 + 1.8 + 1.3}{4} - \frac{9 + 9 + 5.5 + 6.5}{4} = \frac{-9.6}{4} = -2.4$$

$$\hat{A} = \frac{-9 - 9 + 9.3 + 8 - 5.5 - 6.5 + 1.8 + 1.3}{4} = -2.4$$

$$\hat{B} = \frac{-9 - 9 - 9.3 - 8 + 5.5 + 6.5 + 1.8 + 1.3}{4} = -5.05$$

$$\widehat{AB} = \frac{+9 + 9 - 9.3 - 8 - 5.5 - 6.5 + 1.8 + 1.3}{4} = -2.05$$

A: změna koncentrace vede ke zkrácení reakce o 2,4 hod
B: změna teploty vede ke zkrácení reakce o 5,05 hod
Interakce: hodnota efektu interakce nemá zvláštní význam

Faktorové plány 2^2

A	B	AB	y	y	$\bar{y}$	s^2
-	-	+	9.0	9.0	9.0	0
+	-	-	9.3	8.0	8.65	0,845
-	+	-	5.5	6.5	6.0	0,500
+	+	+	1.8	1.3	1.55	0,125
-2.4	-5.05	-2.05				0,3675

$\alpha = 0,05,\ N = 2,\ r = 2 \Rightarrow \nu = 4,\ t_{0,975}(4) = 2,77,$

kde s^2 je průměr z rozptylů : (0+0,845+0,5+0,125)/4=0,3675

X – odhad efektu A,B,AB

2-výběrový t-test

r=počet replikací

N=počet faktorů

$$\left|\hat{X}\right| \approx t_{1-\alpha/2}(\nu)\sqrt{\frac{s^2}{r2^{N-2}}}\ ,\quad \nu = (r-1)2^N$$

$$\Rightarrow \boxed{X = 1.19}$$

Faktorové plány 2^3

Statistický model pro tři faktory A, B, C:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijk}$$

odezva

společná úroveň

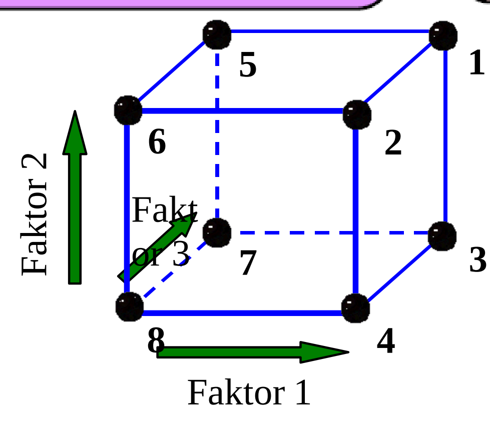
vliv i-té úrovně faktoru A, i=1,2,..., r

vliv j-té úrovně faktoru B, j=1,2,... s

vliv j-té úrovně faktoru C, k=1,2,...t

vlivy smíšených interakcí i-té, j-té a k-té úrovní faktorů A, B a C

náhodná chyba při i-té,j-té a k-té úrovni faktorů A, B a C



Faktorové plány 2^3

Matice modelu = návrh kombinací úrovní faktorů

test	A	B	C	AB	AC	BC	ABC
1	-	-	-	+	+	+	-
2	+	-	-	-	-	+	+
3	-	+	-	-	+	-	+
4	-	-	+	+	-	-	+
5	+	+	-	+	-	-	-
6	+	-	+	-	+	-	-
7	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+

Faktorové plány 2^3 - příklad

Účinek koncentrace a teploty a času na účinnost pracího prášku

Odezva: účinnost, měřená opticky na prané látce
Faktor: koncentrace (A), teplota (B), čas (C), 2 úrovně

Počet replikací: 2

Počet měření: $2^3 x2 = 16$

Experiment: **2^2 (3 faktory, 2 úrovně)**

Faktor A: koncentrace 10% a 40%
Faktor B: teplota 40°C a 60°C
Faktor C: čas 5min. a 15min.

test	A	B	C	AB	AC	BC	ABC
1	-	-	-	+	+	+	-
2	+	-	-	-	-	+	+
3	-	+	-	-	+	-	+
4	-	-	+	+	-	-	+
5	+	+	-	+	-	-	-
6	+	-	+	-	+	-	-
7	-	+	+	-	-	+	-
8	+	+	+	+	+	+	+

	A	B	C	y
1	-	-	-	37
2	-	-	-	45
3	+	-	-	48
4	+	-	-	56
5	-	+	-	59
6	-	+	-	68
7	+	+	-	102
8	+	+	-	90
9	-	-	+	43
10	-	-	+	35
11	+	-	+	63
12	+	-	+	54
13	-	+	+	71
14	-	+	+	77
15	+	+	+	122
16	+	+	+	107

Faktorové plány 2^3 - příklad

	A	B	C	AB	AC	BC	ABC	y	y	$\bar{y}$	s^2
1	-	-	-	+	+	+	-	37	45	41	32
2	+	-	-	-	-	+	+	48	56	52	32
3	-	+	-	-	+	-	+	59	68	63.5	40.5
4	-	-	+	+	-	-	+	102	90	96	72
5	+	+	-	+	-	-	-	43	35	39	32
6	+	-	+	-	+	-	-	63	54	58.5	40.5
7	-	+	+	-	-	+	-	71	77	74	18
8	+	+	+	+	+	+	+	122	107	114.5	112.5
průměr +	80,250	87,000	71,500	72,625	69,375	70,375	67,250			průměr rozptylů:	47.4375
průměr -	54,375	47,625	63,125	62,000	65,250	64,250	67,375				
efekt	25,875	39,375	8,375	10,625	4,125	6,125	-0,125				

$$\alpha = 0,05,\ N = 3,\ r = 2 \Rightarrow \nu = 8,\ t_{0,975}(8) = 2,306, \Rightarrow X = 7.941$$

$$\alpha = 0,01,\ N = 3,\ r = 2 \Rightarrow \nu = 8,\ t_{0,995}(8) = 3,355, \Rightarrow X = 11.555$$

$$\alpha = 0,001, N = 3,\ r = 2 \Rightarrow \nu = 8, t_{0.9995}(8) = 5,041, \Rightarrow X = 17.367$$

Faktorové plány

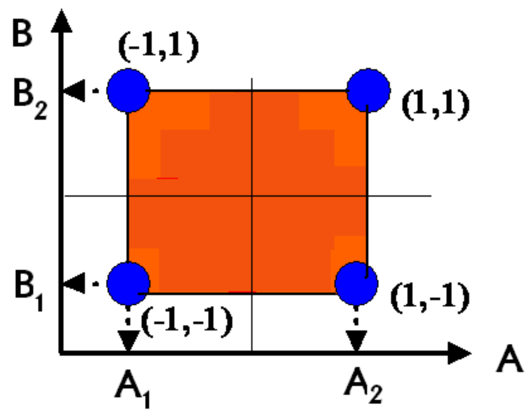
Matice plánu: kódované hodnoty

všech kombinací úrovní

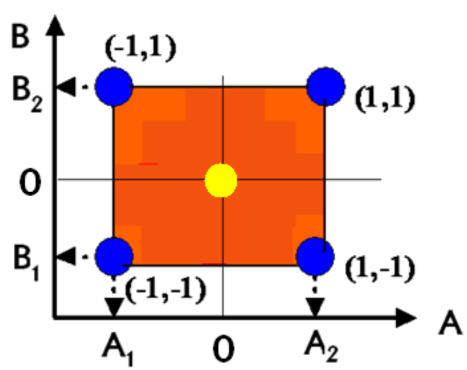
Výsledková plocha: funkce y = f(A,B..)

Plán: **M** (n x m) matice, jejíž sloupce jsou faktory a interakce a řádky jsou jednotlivé experimenty

Odezva: hodnoty *y* tj. výsledky měření pro danou kombinaci úrovní faktorů



$$\begin{pmatrix} A_1 & B_1 \\ A_1 & B_2 \\ A_2 & B_1 \\ A_2 & B_2 \end{pmatrix} \rightarrow \begin{pmatrix} -1 & -1 \\ -1 & 1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix}$$



$$\begin{pmatrix} A_1 & B_1 \\ A_1 & B_2 \\ A_2 & B_1 \\ A_2 & B_2 \\ A_0 & B_0 \\ A_0 & B_0 \end{pmatrix} \rightarrow \begin{pmatrix} -1 & -1 \\ -1 & 1 \\ 1 & -1 \\ 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Faktorové plány

x_1	x_2	y
1	1	y_1
1	-1	y_2
-1	1	y_3
-1	-1	y_4

Regresní model $E(y) = a_0 + a_1 x_1 + a_2 x_2$

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \mathbf{e} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix} \mathbf{a} + \mathbf{e}$$

Vzhledem k ortogonalitě sloupců matice **X**

$$\mathbf{X}^T\mathbf{X} = N\mathbf{E} \quad \hat{\mathbf{a}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \frac{1}{\mathrm{N}}\mathbf{X}^T\mathbf{y} \qquad \hat{a}_0 = \sum_{i=1}^{4} y_i / 4$$

$$\hat{a}_1 = (y_1 + y_2 - y_3 - y_4)/4 \qquad \hat{a}_2 = (y_1 - y_2 + y_3 - y_4)/4$$

Faktorové plány

$$E(y) = a_0 + a_1 x_1 + a_2 x_2 + a_{12} x_1 x_2$$

$pokus$	x_0	x_1	x_2	$x_1 x_2$	y
1	1	1	1	1	y_1
2	1	1	-1	-1	y_2
3	1	-1	1	-1	y_3
4	1	-1	-1	1	y_4

$$\hat{a}_{12} = (y1 - y2 - y3 + y4)/4$$

Místo $y_1, .. y_4$ se při opakování berou průměrné hodnoty $\overline{y}_1, .. \overline{y}_4$

Nezávislý odhad chyb.
Několik pokusů v bodě (0,0).
Z rozptylu

$$\sigma_0^2 \Rightarrow \sigma^2$$

$$\hat{a}_0 = \overline{y}_0$$

Faktorové plány 2^M

- Přidávání faktorů

- Pro 2^k faktorový pokus potřebujeme nejméně 2^k měření
- Případ 4 faktorů vyžaduje 2^4 =16 tj. 2x tolik pokus; než pro 3 faktory (2^3 =8).
- Je možno v 8 pokusech sledovat více faktorů?
- Nahrazení málo významných interakcí (faktorů) jinými faktory

Polo faktorový experiment

- Při 8 experimentech se testují 4 faktory při zanedbání interakce ABC.
- Polohy faktoru D se řídí znaménkovým schématem pro ABC.

Faktorové plány 2^M - krácení

Pokus	A	B	C	AB	AC	BC	D
1	-	-	-	+	+	+	-
2	-	-	+	+	-	-	+
3	-	+	-	-	+	-	+
4	-	+	+	-	-	+	-
5	+	-	-	-	-	+	+
6	+	-	+	-	+	-	-
7	+	+	-	+	-	-	-
8	+	+	+	+	+	+	+

Kombinace D a ABC

Mohou být přidány další faktory na vrub interakcí)

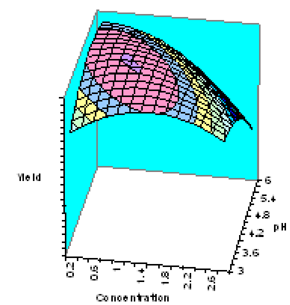
Faktorové plány 3^2

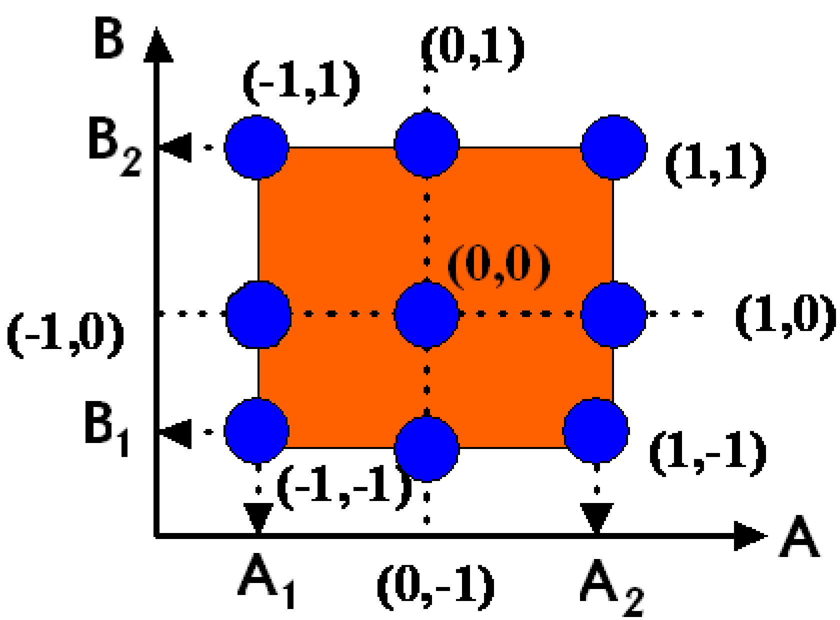
- Dva faktory A, B na třech úrovních

$$E(y) = a_0 + a_1 x_1 + a_2 x_2 + a_{12} x_1 x_2 + a_{11} x_1^2 + a_{22} x_2^2$$

Ortogonalita plánu vzhledem k x_1^2, x_2^2

x_1	x_2
-1	-1
0	-1
1	-1
-1	0
0	0
1	0
-1	1
0	1
1	1



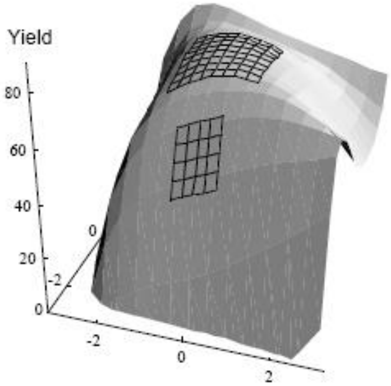


Odchylky od průměru 2/3 pro x_1^2, x_2^2

$$E(y) = a_0 + a_1 x_1 + a_2 x_2 + a_{12} x_1 x_2 +$$
$$+ a_{11}(x_1^2 - 2/3) + a_{22}(x_2^2 - 2/3)$$

Již je ortogonální pro matici plánu

Hodnocení křivosti odezvy



- Významné interakční členy
- Využití centrálních bodů (opakování)

  Průměrná hodnota všech výsledků (kromě centrálních bodů) y_F (n_F hodnot) a průměrná hodnota centrálních bodů y_C (n_C hodnot a s_C směrodatná odchylka). Párový t test resp. interval spolehlivosti :

$$\bar{y}_F - \bar{y}_C \pm t_{\alpha/2}(n_C - 1) s_C \sqrt{\frac{1}{n_C} + \frac{1}{n_F}}$$

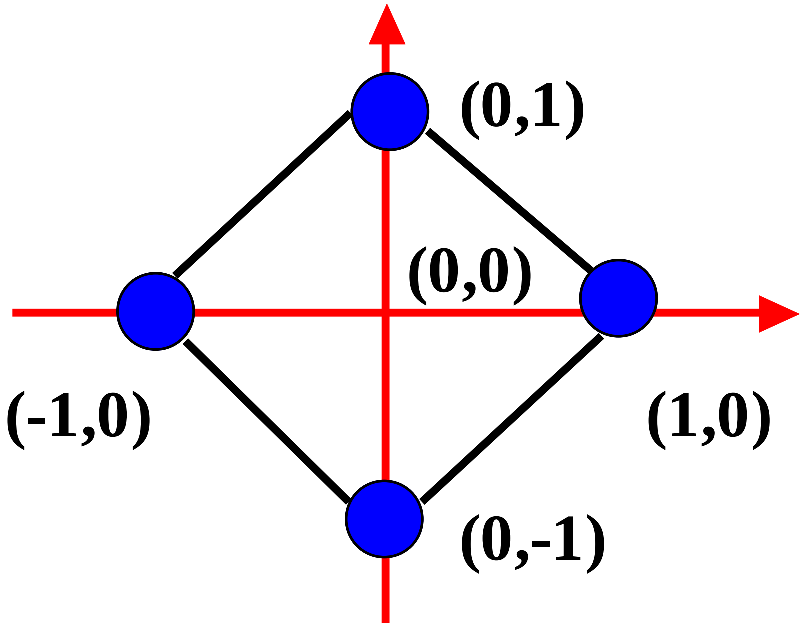
Pokud tento interval obsahuje nulu je odezva lineární (není výrazná křivost odezvy)

Pootočené plány

x_1	x_2
1	0
0	1
-1	0
0	-1

- model

$$E(y) = a_0 + a_1 x_1 + a_2 x_2$$



$$\hat{a}_0 = \sum_{i=1}^{4} y_i / 4$$

$$\hat{a}_1 = (y_1 - y_3) / 4$$

$$\hat{a}_2 = (y_2 - y_4) / 4$$

Vícefaktroiální návrhy experimentů

Příklad

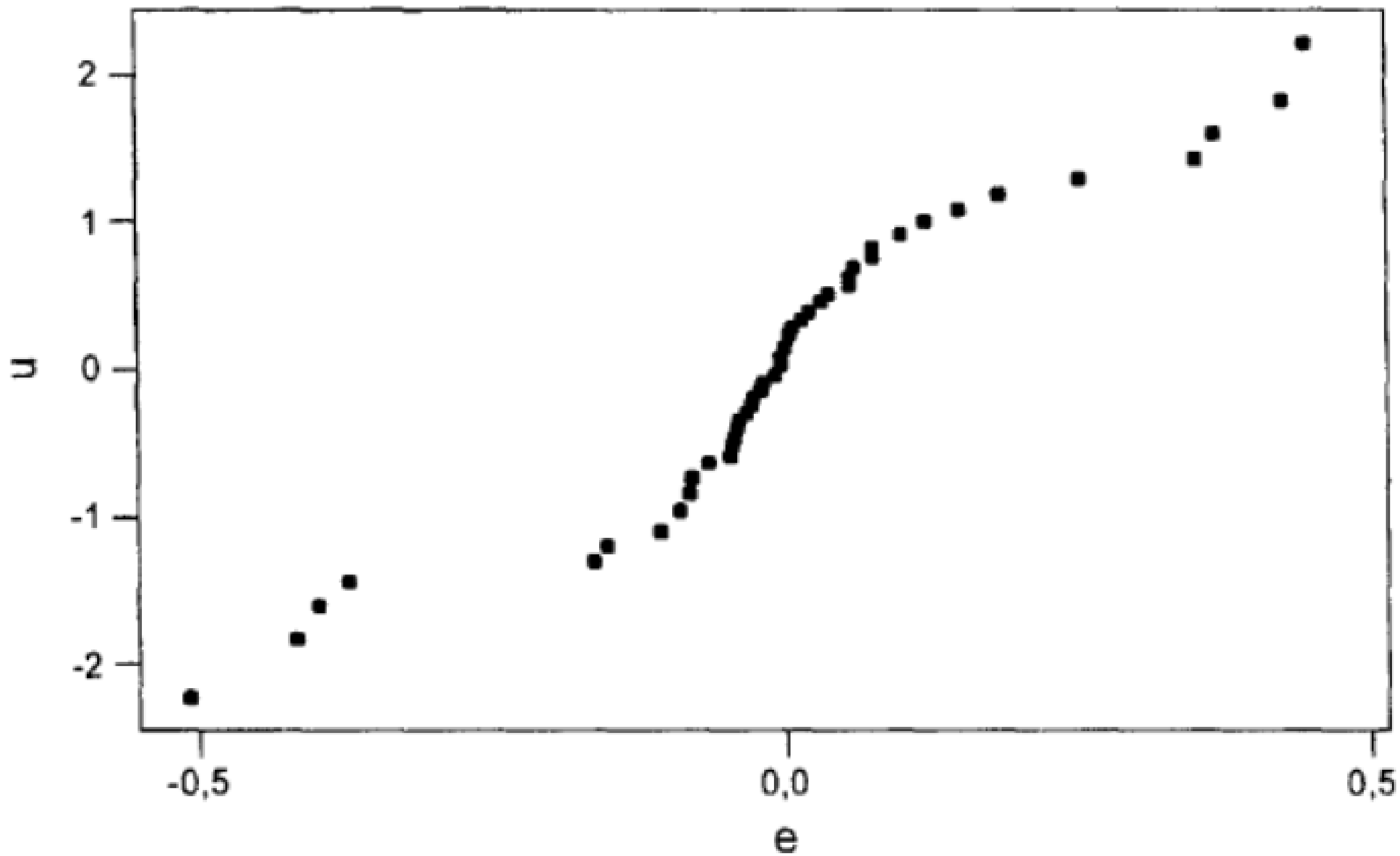
Matice návrhu:

A	B	C	D	y			$\overline{y}$	s^2
−	−	−	−	13,896	13,932	13,914	13,914	0,0003
+	−	−	−	13,588	13,964	14,328	13,960	0,1369
−	+	−	−	14,274	14,154	14,082	14,170	0,0094
+	+	−	−	13,970	13,738	13,738	13,815	0,0179
−	−	+	−	13,846	13,896	13,870	13,871	0,0006
+	−	+	−	14,264	14,432	14,228	14,308	0,0119
−	+	+	−	14,028	14,108	14,060	14,065	0,0016
+	+	+	−	14,000	13,640	13,592	13,744	0,0497
−	−	−	+	14,794	14,860	14,914	14,856	0,0036
+	−	−	+	14,718	15,198	15,490	15,135	0,1519
−	+	−	+	14,876	14,958	14,932	14,922	0,0018
+	+	−	+	15,034	15,384	15,170	15,196	0,0311
−	−	+	+	14,778	14,682	14,850	14,770	0,0071
+	−	+	+	14,962	14,504	14,136	14,534	0,1712
−	+	+	+	15,058	14,938	14,936	14,977	0,0049
+	+	+	+	15,424	15,036	14,470	14,977	0,2302

Vícefaktroiální návrhy experimentů

Příklad

Pravděpodobnostní graf reziduí:



Vícefaktroiální návrhy experimentů

Stejné podmínky měření

- Lze je zajistit prostým uspořádáním experimentu (časovým, organizačním, materiálovým)
- Často používáme zavedení dalších, tzv. blokových faktorů (blokováním)

Rozdělení měření do bloků nelze provádět libovolně.

test	A	B	C	AB	AC	BC	ABC
1	-	-	-	+	+	+	-
a	+	-	-	-	-	+	+
b	-	+	-	-	+	-	+
c	-	-	+	+	-	-	+
ab	+	+	-	+	-	-	-
ac	+	-	+	-	+	-	-
bc	-	+	+	-	-	+	-
abc	+	+	+	+	+	+	+

Vícefaktroiální návrhy experimentů

Stejné podmínky měření

- Lze je zajistit prostým uspořádáním experimentu (časovým, organizačním, materiálovým)
- Často používáme zavedení dalších, tzv. blokových faktorů (blokováním)

Rozdělení měření do bloků nelze provádět libovolně.

test	A	B	C	AB	AC	BC	ABC
1	-	-	-	+	+	+	-
a	+	-	-	-	-	+	+
b	-	+	-	-	+	-	+
c	-	-	+	+	-	-	+
ab	+	+	-	+	-	-	-
ac	+	-	+	-	+	-	-
bc	-	+	+	-	-	+	-
abc	+	+	+	+	+	+	+

Rozdělení měření do 2 bloků podle interakce ABC - vliv interakce ABC potom nelze odlišit od vlivu rozdělení na bloky

- hlavní blok
- alternativní blok

Vícefaktroiální návrhy experimentů

Stejné podmínky měření

- Lze je zajistit prostým uspořádáním experimentu (časovým, organizačním, materiálovým)
- Často používáme zavedení dalších, tzv. blokových faktorů (blokováním)

Rozdělení měření do 4 bloků metodou sudá-lichá podle interakcí AC a BD:

test	A	B	C	AB	AC	BC	ABC
1	-	-	-	+	0	0	-
a	+	-	-	-	1	0	+
b	-	+	-	-	0	1	+
c	-	-	+	+	1	1	+
ab	+	+	-	+	1	1	-
ac	+	-	+	-	2	1	-
bc	-	+	+	-	1	2	-
abc	+	+	+	+	2	2	+

	BC sudá	BC lichá
AC sudá	1 abc	b ac
AC lichá	a bc	c ab

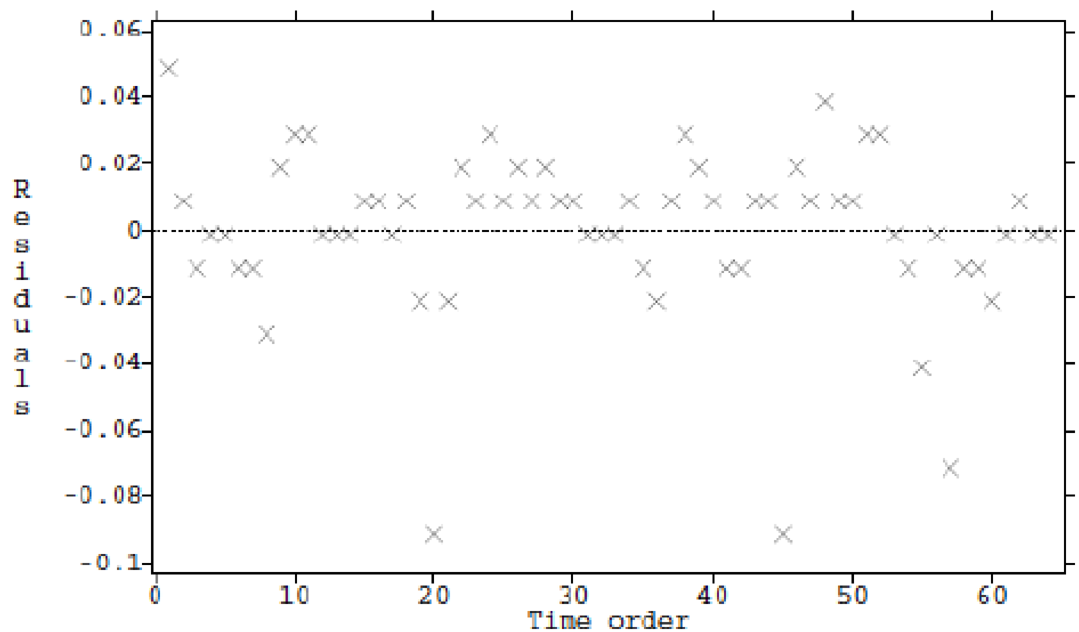
Vícefaktroiální návrhy experimentů

Nezávislost měření

Závislost v posloupnosti měření se zjišťuje prostřednictvím reziduí $r_{ij} = y_{ij} - \bar{y}_{i.}$

Tato "pořadová závislost" může vznikat posunem v měřícím zařízení, únavou operátora, změnou podmínek měření apod.

Grafická analýza - graf reziduí v čase (indexový graf)

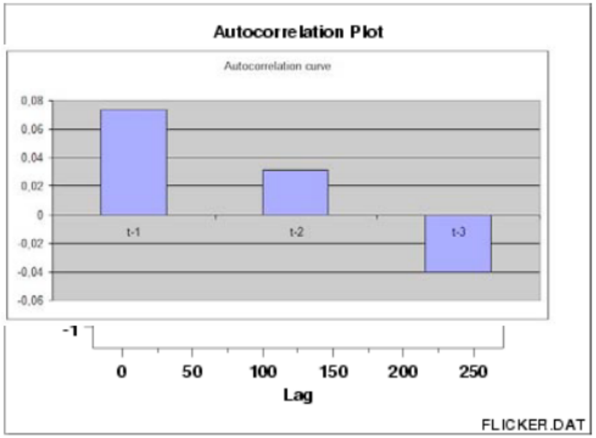


Vícefaktroiální návrhy experimentů

Nezávislost měření

Závislost v posloupnosti měření se zjišťuje prostřednictvím reziduí $r_{ij} = y_{ij} - \bar{y}_{i.}$

- Tato "pořadová závislost" může vznikat posunem v měřícím zařízení, únavou operátora, změnou podmínek měření apod.



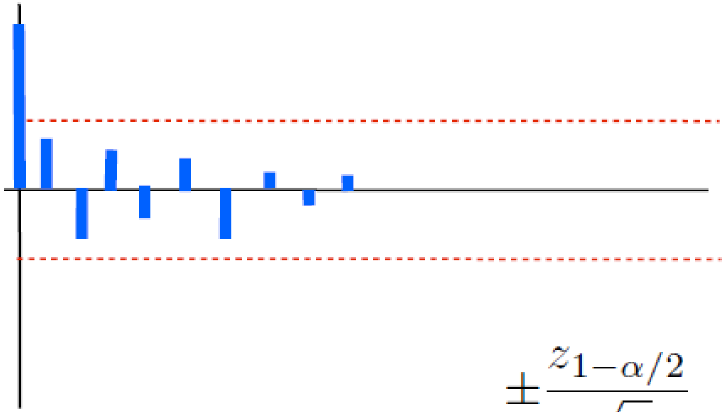
Grafická analýza - graf reziduí v čase (indexový graf)

Autokorelační funkce reziduí

$$c_{i,k} = Cov(r_{i,j}, r_{i,j-k})$$

$$\hat{c}_{i,k} = \frac{1}{n}\sum_{j=1}^{n-k} r_{i,j} r_{i,j+k}$$

$$\hat{r}_{i,k} = \frac{\sum_{j=1}^{n-k} r_{i,j} r_{i,j+k}}{\sum_{j=1}^{n} r_{i,j}^2}$$



$$\pm \frac{z_{1-\alpha/2}}{\sqrt{n}}$$

Vícefaktroiální návrhy experimentů

Nezávislost měření

Co se stane, když jsou data závislá?

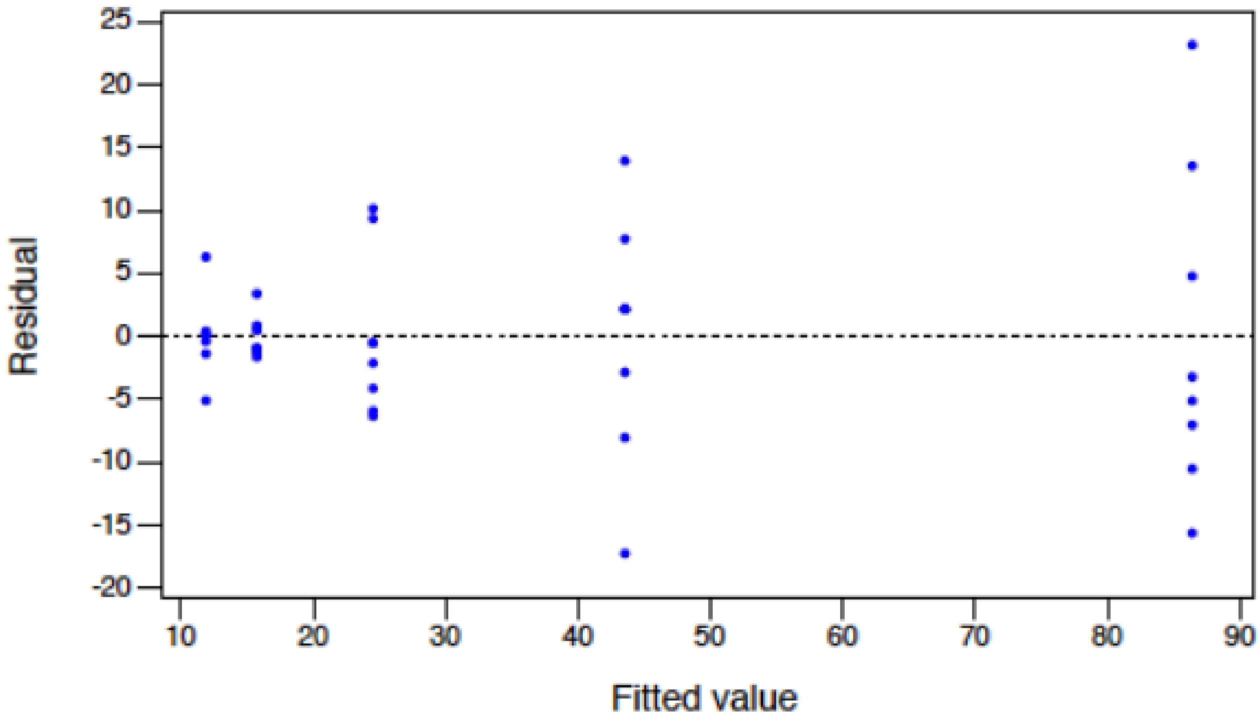
- Největší změnou je to, že rozptyly průměrů se stávají vychýlené (velikost vychýlení závisí na typu závislosti). To způsobí vychýlenost odhadů pro jednotlivé skupiny a pro efekty faktorů a jejich interakcí.
- Vliv randomizace může být na závislých datech různý účinek. Na stejných datech může jedno znáhodnění pořadí zvýšit pravděpodobnost chyby I. druhu, jiné naopak snížit.
- Náhodné seřazení kombinací v blocích také nepomůže, pokud není znáhodněno pořadí měření.

Jak se vypořádat se závislostí dat?

- Použít analýzu časových řad ke zjištění trendu a případné periodicity časové řady měření
- Pomocí autokorelační (parciální autokorelační) funkce zjistit závislosti v čase a určit ARMA model pro časovou řadu měření
- Na základě předchozích kroků provést "očištění" časové řady měření a dále pracovat pouze s rezidui, která by už měla být nezávislá

Vícefaktroiální návrhy experimentů

Konstantní rozptyl



- O'Brienův test
- Brown-Forsythův tet
- Levenův test
- Bartlettův test

první tři jsou založeny na transformaci do nové odezvy z_{ij} a použití ANOVA

vyžaduje předpoklad normality!

Vícefaktroiální návrhy experimentů

Konstantní rozptyl

- O'Brien
$$z_{ij} = \frac{(n_j + w - 2)n_j\left(y_{ij} - \bar{y}_{j.}\right)^2 - w s_j^2\left(n_j - 1\right)}{\left(n_j - 1\right)\left(n_j - 2\right)}$$

- Brown-Forsythe
$$z_{ij} = \left| y_{ij} - \tilde{y}_j \right|$$

y_{ij} - odezva pro i-té měření v j-té skupině
$\bar{y}_j$ - průměr odezvy ve skupině j
s^2_j - rozptyl odezvy ve skupině j
n_j - je počet měření ve skupině j
w - konstanta (0,5)
$\tilde{y}_j$ - medián odezvy ve skupině j
p - počet skupin

- Levene
$$z_{ij} = \left| y_{ij} - \bar{y}_j \right|$$

testová statistika (ANOVA) má F-rozdělení o (p-1) a (N-p) stupních volnosti

$$F = \frac{(N-p)\sum_{j=1}^{p} n_j\left(\bar{z}_{j.} - \bar{\bar{z}}_{..}\right)^2}{(p-1)\sum_{j=1}^{p}\sum_{i=1}^{N}\left(z_{ij} - \bar{z}_{j.}\right)^2}$$

Vícefaktroiální návrhy experimentů

Konstantní rozptyl **Jak lze nehomogenitu rozptylů odstranit?**

V podstatě jediný způsob "napravení" nehomogenity rozptylů je transformace.
V následující tabulce jsou některé uvedeny:

Rozdělení	Transformace	Nový rozptyl
Binomické rozdělení $X \approx Binom(n, p)$ $Var(\hat{p}) = p(1-p)/n$	$\arcsin(\sqrt{\hat{p}})$	$\frac{1}{4n}$
Poissonovo rozdělení $X \approx Poisson(\lambda)$ $Var(X) = \lambda$	$\sqrt{X}$	$\frac{1}{4}$
Pokud lze vyjádřit $Var(X) = g(\mu)$ pomocí $E(X) = \mu$	$y = \int_a^x \frac{1}{\sqrt{g(v)}} dv$	1

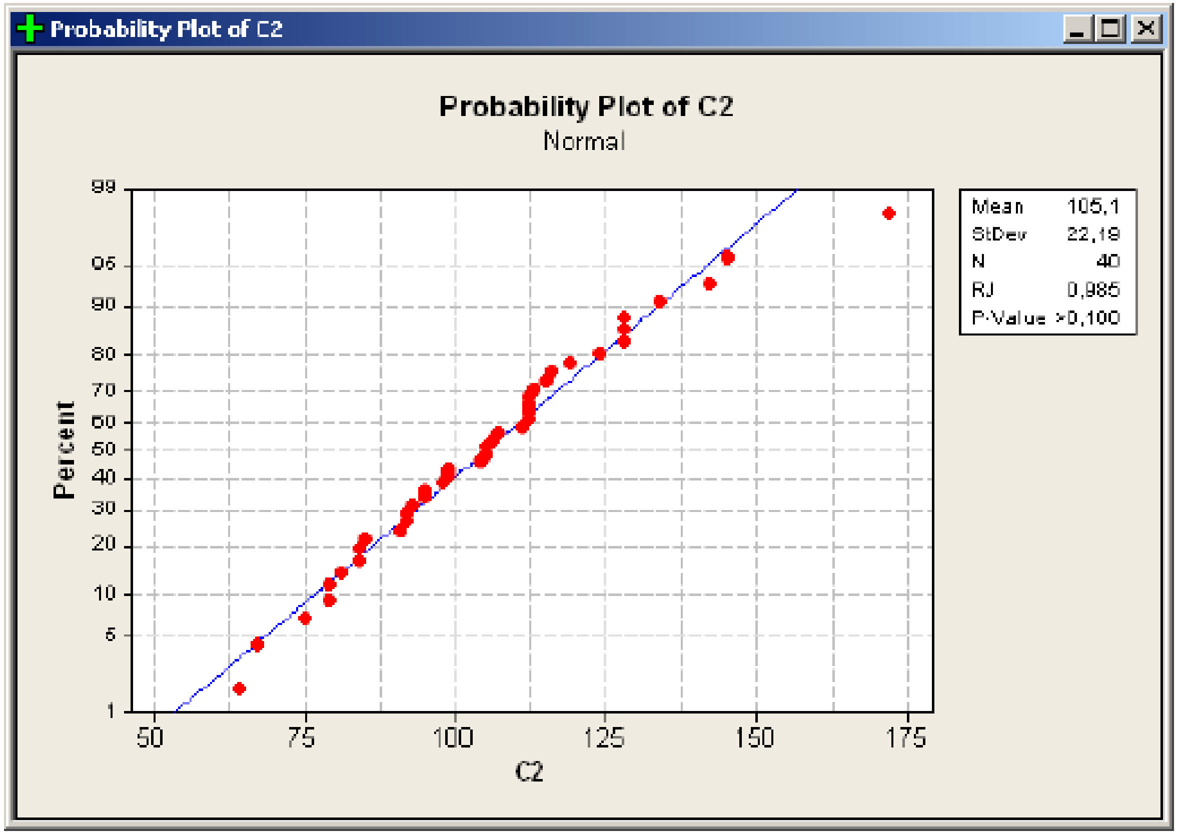
Alternativou je použití neparametrického testu (Kruskal-Wallis) namísto F-testu v ANOVA

Vícefaktroiální návrhy experimentů

Normalita reziduí

Testy normality:

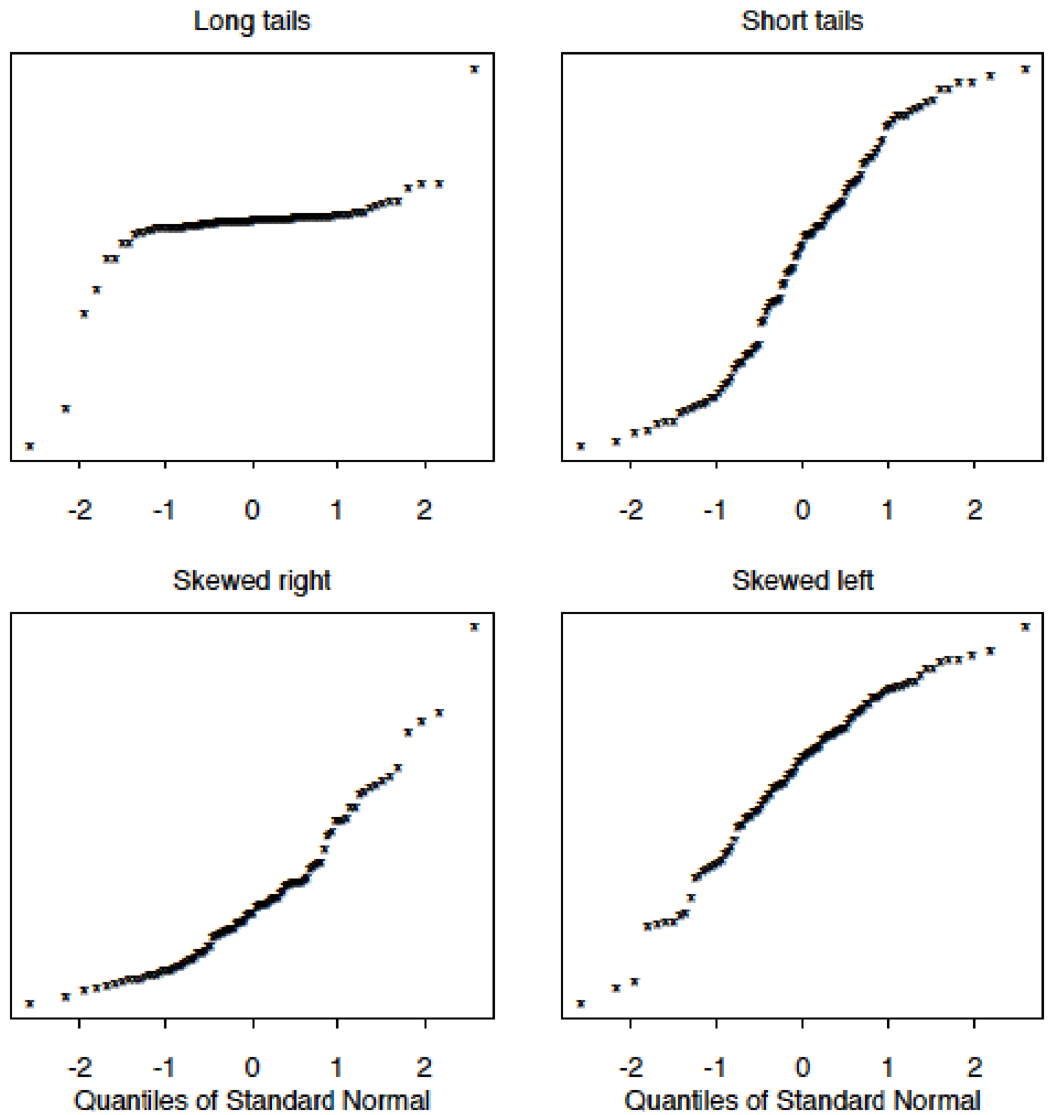
- Chí-kvadrát test
- Anderson-Darling
- Shapiro-Wilk (Ryan-Joiner)
- Test normality založený na šikmosti a špičatosti



Vícefaktroiální návrhy experimentů

Normalita reziduí

Normální pravděpodobnostní graf (rankitový graf, Q-Q graf)



rankit je aproximace normálního rozdělení: $$\frac{i-3/8}{n+1/4}$$

Vícefaktroiální návrhy experimentů

Normalita reziduí

Test normality založený na šikmosti a špičatosti

výběrový koeficient šikmosti:

$$S_k = \frac{M_3}{s^3} = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{x})^3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{x})^2}\right)^3}, \qquad ES_k = 0, \qquad DS_k = \frac{6(n-2)}{(n+1)(n+3)}$$

$$\boxed{\frac{|S_k|}{\sqrt{DS_k}} \geq z_\alpha}$$

výběrový koeficient špičatosti:

$$E_k = \frac{M_4}{s^4} - 3 = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{x})^2\right)^2} - 3, \qquad EE_k = \frac{-6}{n+1}, \qquad DE_k = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}$$

$$\boxed{\frac{|E_k - EE_k|}{\sqrt{DE_k}} \geq z_\alpha}$$

kde $z_\alpha = \Phi^{-1}(1 - \alpha/2)$

Vícefaktroiální návrhy experimentů

Neparametrická alternativa k ANOVA

Kruskal – Wallisův test

je neparametrickou obdobou analýzy rozptylu jednoduchého třídění.
Slouží k ověření nulové hypotézy H_0, že ***k > 2*** nezávislých náhodných výběrů o rozsazích $n_1, n_2, \ldots, n_k$ pochází z jednoho základního souboru. Předpokládáme, že tyto náhodné výběry byly pořízeny ze základních souborů se spojitými distribučními funkcemi $F_1(x), F_2(x), \ldots, F_k(x)$.
Nulovou hypotézu H_0 můžeme zapsat takto:

$$H_0: F_1(x) = F_2(x) = \ldots = F_k(x) \text{ pro všechna x.}$$

Postup při stanovení testového kritéria:

1) Máme k dispozici ***k*** výběrových souborů o četnostech $n_1, n_2, \ldots, n_k$.
2) Všechny výběrové soubory sloučíme do jediného souboru.
3) Každé hodnotě souboru přiřadíme vzestupně pořadové číslo, stejným hodnotám pak pořadí průměrné.
4) Následně sečteme pořadová čísla jednotlivých pozorování pro každý původní výběrový soubor zvlášť a získáme součty $T_1, T_2, \ldots, T_k$ (T_i ; $i = 1, \ldots, k$, je tedy součet pořadových čísel pro i-tý výběr).

Vícefaktroiální návrhy experimentů

Neparametrická alternativa k ANOVA

Testové kritérium má tvar
$$KW = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{T_i^2}{n_i} - 3 \cdot (n+1),$$
kde $n = n_1 + n_2 + \ldots + n_k$.
Statistika KW má za platnosti H_0 při $n_i \to \infty$ asymptoticky χ^2 – rozdělení o *k-1* stupních volnosti.
Pokud $KW > \chi^2_{\alpha(k-1)}$, přijímáme hypotézu alternativní, podle které se hodnoty nejméně dvou porovnávaných výběrových souborů od sebe průkazně liší.

Jestliže se v posloupnosti zjištěných údajů vyskytnou shodné hodnoty, kterým se přiřazuje průměrné pořadí, je nutno hodnotu KW dělit korekčním faktorem
$$K = 1 - \frac{\sum_{i=1}^{p} \left(t_i^3 - t_i\right)}{n^3 - n},$$
kde p je počet tříd se stejným pořadím a t_i počet pořadí v i-té třídě. Opravené testové kritérium se stanoví jako
$$KW_{opr.} = \frac{KW}{K}.$$

Vícefaktroiální návrhy experimentů

Neparametrická alternativa k ANOVA

V případě, že zamítneme nulovou hypotézu, tvrdíme, že všechny výběry nepocházejí z téhož rozdělení.
Obvykle pak ve druhé etapě statistického zpracování ř ešíme otázku, které výběry se od sebe statisticky významně liší.
K tomuto lze použít postupy, které se souhrnně nazývají neparametrické metody mnohonásobného porovnávání.
Pracujeme-li s vyváženým pokusným plánem, tzn. má-li všech ***k*** výběrů stejný rozsah (platí-li $n_1 = n_2 = \ldots = n_k = N$), můžeme Kruskal – Wallisův test doplnit **Neményiho metodou mnohonásobného srovnávání**.

Je-li diference $|T_i - T_j|$ větší nebo rovna kritické hodnotě D_α pro Neményiho metodu, zamítá se hypotéza o neprůkaznosti diference, tzn. že i-tý a j-tý výběr pocházejí z téhož rozdělení.
Tabulkové hodnoty se hledají pro hladinu významnosti α, pro k počet porovnávaných tříd a N opakování v každé třídě ($n_1 = n_2 = \ldots = n_k = N$).
Tímto postupem zhodnotíme $\frac{k(k-1)}{2}$ diferencí $|T_i - T_j|$.

Vícefaktroiální návrhy experimentů

Neparametrická alternativa k ANOVA

Pokud rozsahy jednotlivých výběrových souborů nejsou stejné (nevyvážený plán), můžeme zjistit, které výběry se od sebe statisticky významně liší, pomocí přibližné Dunnovy metody mnohonásobného srovnávání.
Jestliže

$$\left|T_i - T_j\right| > u_{\frac{2\alpha}{k(k-1)}} \cdot \sqrt{\frac{n(n-1)}{12}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)},$$

kde $u_{\frac{2\alpha}{k(k-1)}}$ je kritická hodnota rozdělení N(0; 1), zamítáme na hladině významnosti α hypotézu, že i-tý výběr (s rozsahem n_i) a j-tý výběr (s rozsahem n_j) pocházejí z téhož rozdělení.

Vícefaktroiální návrhy experimentů

Neparametrická alternativa k ANOVA

Příklad

V polním pokusu byly ověřovány čtyři varianty hnojení silážní kukuřice rozdílnou dávkou NPK v hnojivech, označené jako H_1 až H_4. Každá varianta byla ověřována na 8 parcelách s výnosem sklizené hmoty v tunách z parcely, uvedeným v tabulce dat. Ověřte, zda se výnosy u jednotlivých variant hnojení průkazně liší.

H_0: výnosy jednotlivých variant jsou shodné
H_1: výnosy jednotlivých variant jsou rozdílné

Vícefaktroiální návrhy experimentů

Neparametrická alternativa k ANOVA

$$KW = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{T_i^2}{n_i} - 3 \cdot (n+1)$$

Varianta	Výnos kukuřice v t z parcely číslo								Součet
	1	2	3	4	5	6	7	8	T_i
H_1	1,29	1,19	1,23	1,33	1,27	1,29	1,31	1,20	
pořadí	23,5	8,5	15	28,5	20	23,5	27	11,5	157,5
H_2	1,30	1,33	1,29	1,37	1,35	1,25	1,38	1,29	
pořadí	26	28,5	23,5	31	30	18,5	32	23,5	213
H_3	1,20	1,24	1,25	1,24	1,20	1,21	1,28	1,17	
pořadí	11,5	16,5	18,5	16,5	11,5	14	21	7	116,5
H_4	1,03	1,14	1,09	1,20	1,07	1,19	1,01	1,05	
pořadí	2	6	5	11,5	4	8,5	1	3	41

$$KW = \frac{12}{32 \cdot 33}\left(\frac{157{,}5^2}{8} + \frac{213^2}{8} + \frac{116{,}5^2}{8} + \frac{41^2}{8}\right) - 3 \cdot 33 = 22{,}3473$$

Vzhledem k výskytu stejných údajů (bylo použito průměrné pořadí) je vhodné opravit testové kritérium korekčním faktorem:

$$K = 1 - \frac{\boxed{(2^3 - 2)} + \boxed{(4^3 - 4)} + \boxed{(2^3 - 2)} + \boxed{(2^3 - 2)} + \boxed{(4^3 - 4)} + \boxed{(2^3 - 2)}}{32^3 - 32} = 0{,}9956$$

Vícefaktroiální návrhy experimentů

Neparametrická alternativa k ANOVA

Součet T_i
157,5
213
116,5
41

$$KW_{opr.} = \frac{22{,}3473}{0{,}9956} = 22{,}446 \qquad \chi^2_{0{,}05\,(3)} = 7{,}815 \qquad \chi^2_{0{,}01\,(3)} = 11{,}34$$

Opravené testové kritérium KW = 22,446 > χ^2 => na hladině významnosti $\alpha = 0{,}05$ i 0,01 přijímáme alternativní hypotézu, podle které se průkazně liší hodnoty výnosu nejméně ve 2 třídách.

Podrobnější vyhodnocení Neményiho metodou:

Tabulka diferencí mezi součty pořadí pro jednotlivé varianty hnojení (výnosy kukuřice) $T_i - T_j$

Třída T_j / Třída T_i	Diference $T_i - T_j$		
	H_2	H_3	H_4
H_1	55,5	41	116,5 ^{x}
H_2		96,5 ^{x}	172 ^{xx}
H_3			75,5

$D_{0{,}05} = 96{,}4$ (N = 8, K = 4)
$D_{0{,}01} = 116{,}8$ (N = 8, K = 4)

Statisticky významně se liší hodnoty výnosů kukuřice mezi variantami hnojení $H_2 - H_4$ a to na hladině významnosti $\alpha = 0{,}01$ (diference $T_i - T_j$ jsou označeny ^{xx}). Na hladině významnosti $\alpha = 0{,}05$ se dále statisticky významně liší hodnoty výnosů kukuřice mezi variantami hnojení $H_1 - H_4$ a $H_2 - H_3$.

Vícefaktroiální návrhy experimentů

Test extrémních odchylek hodnot znaku

V řadě pozorovaných hodnot se někdy objeví hodnota extrémně se lišící od ostatních, tzn. výrazně vybočuje z rozpětí ostatních naměřených hodnot.
Je třeba posoudit, zda je tato odchylka pouze náhodná nebo zda je uvedená hodnota zatížena „hrubou chybou“.
Pro objektivní posouzení této otázky existuje skupina testů, které se nazývají testy extrémních odchylek.

Dixonův test

Pozorovaná hodnota, která se extrémně liší od ostatních, je zřejmě buď nejmenší hodnotou (x_1) nebo největší hodnotou (x_n).

Nulová hypotéza H_0 tvrdí, že (x_1), resp. (x_n), je vybrána ze stejného normálně rozděleného základního souboru jako ostatní hodnoty.
Pro posouzení, zda hodnota (x_1) nebo hodnota (x_n) je zatížena hrubou chybou, užíváme testovacího kritéria

Vícefaktroiální návrhy experimentů

Test extrémních odchylek hodnot znaku

$$Q_1 = \frac{X_2 - X_1}{X_n - X_1} \quad \text{nebo} \quad Q_n = \frac{X_n - X_{n-1}}{X_n - X_1},$$

kde x_2 je druhá nejmenší pozorovaná hodnota a x_{n-1} je předposlední hodnota v řadě pozorování, uspořádaných vzestupně podle velikosti.

Jestliže vypočtená hodnota Q_1, resp. Q_n, překročí kritickou hodnotu $Q_{1\alpha} = Q_{n\alpha}$ (nalezenou v tabulkách pro Dixonův test, hladinu významnosti α a n rozsah souboru), zamítáme nulovou hypotézu na hladině významnosti α a hodnotu (x_1), resp. (x_n), jako údaj zkreslený hrubou chybou, ze souboru vyloučíme.

Tzn. že nulová hypotéza se zamítá, pokud platí Q_1 $(Q_n) > Q_{\alpha(n)}$.

Vícefaktroiální návrhy experimentů

Test extrémních odchylek hodnot znaku

Příklad

Bylo provedeno 5 měření vlhkosti zrna jarního ječmene s těmito výsledky (v %): 18,0 18,2 19,6 18,3 18,4.

Hodnota 19,6 vyvolává v této ř adě údajů podezření, že je ovlivněna nějakou hrubou chybou. Dixonovým testem budeme testovat hypotézu, že hodnota 19,6 není zatížena hrubou chybou.

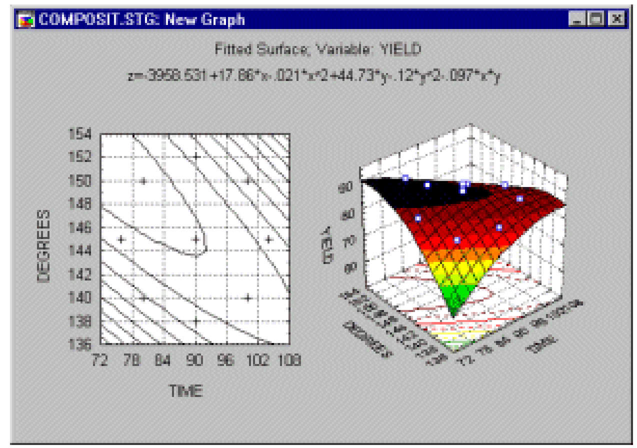
$$Q_5 = \frac{19{,}6 - 18{,}4}{19{,}6 - 18{,}0} = 0{,}75$$

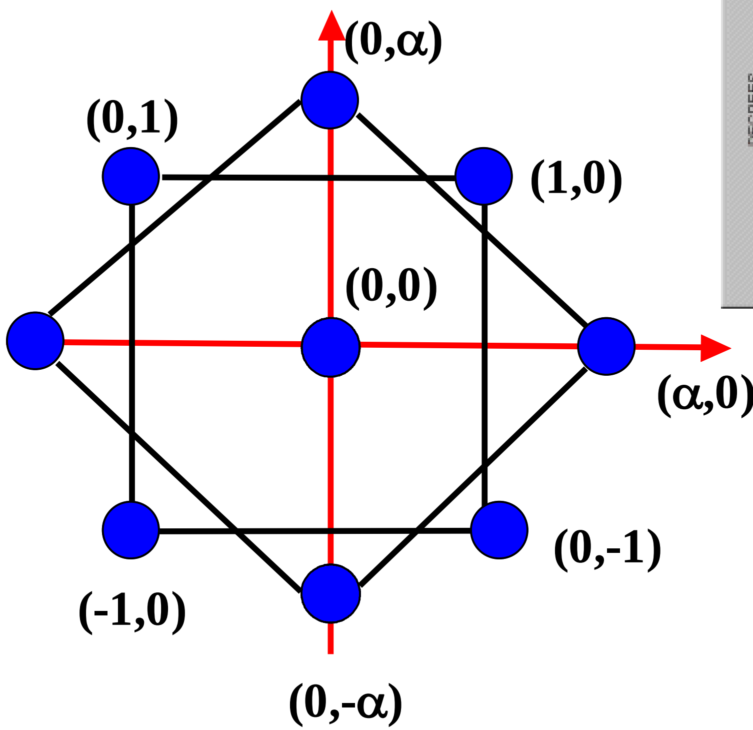
Protože $Q_5 = 0{,}75 > Q_{\alpha(5)} = 0{,}642$, zamítáme nulovou hypotézu na hladině významnosti 0,05 a hodnotu 19,6 ze souboru vyloučíme.

Centrální kompozitní plán

x_1	x_2
1	1
1	-1
-1	1
-1	-1
α	0
$-\alpha$	0
0	α
0	$-\alpha$
0	0

- Pro zjednodušení výpočtu odhadů je zase vhodné ortogonalizovat (použít odchylky od průměrů)





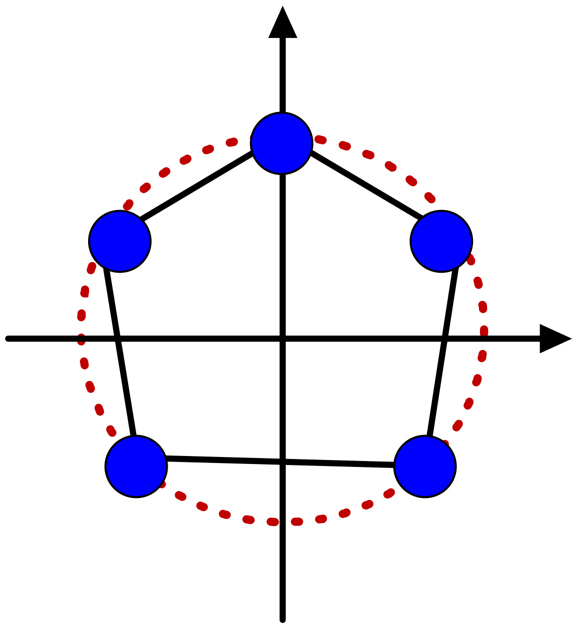
$$\beta = \frac{1}{N}\left(2^m + 2\alpha^2\right)$$

$$E(y) = \bar{y} + a_1x_1 + a_2x_2 + a_{12}x_1x_2 + $$
$$+a_{11}(x_1^2 - \beta) + a_{22}(x_2^2 - \beta)$$

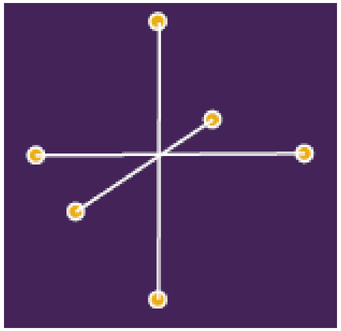
$$a_0 \;=\; \bar{y} - \beta(a_{11} + a_{22})$$

Rotabelové plány

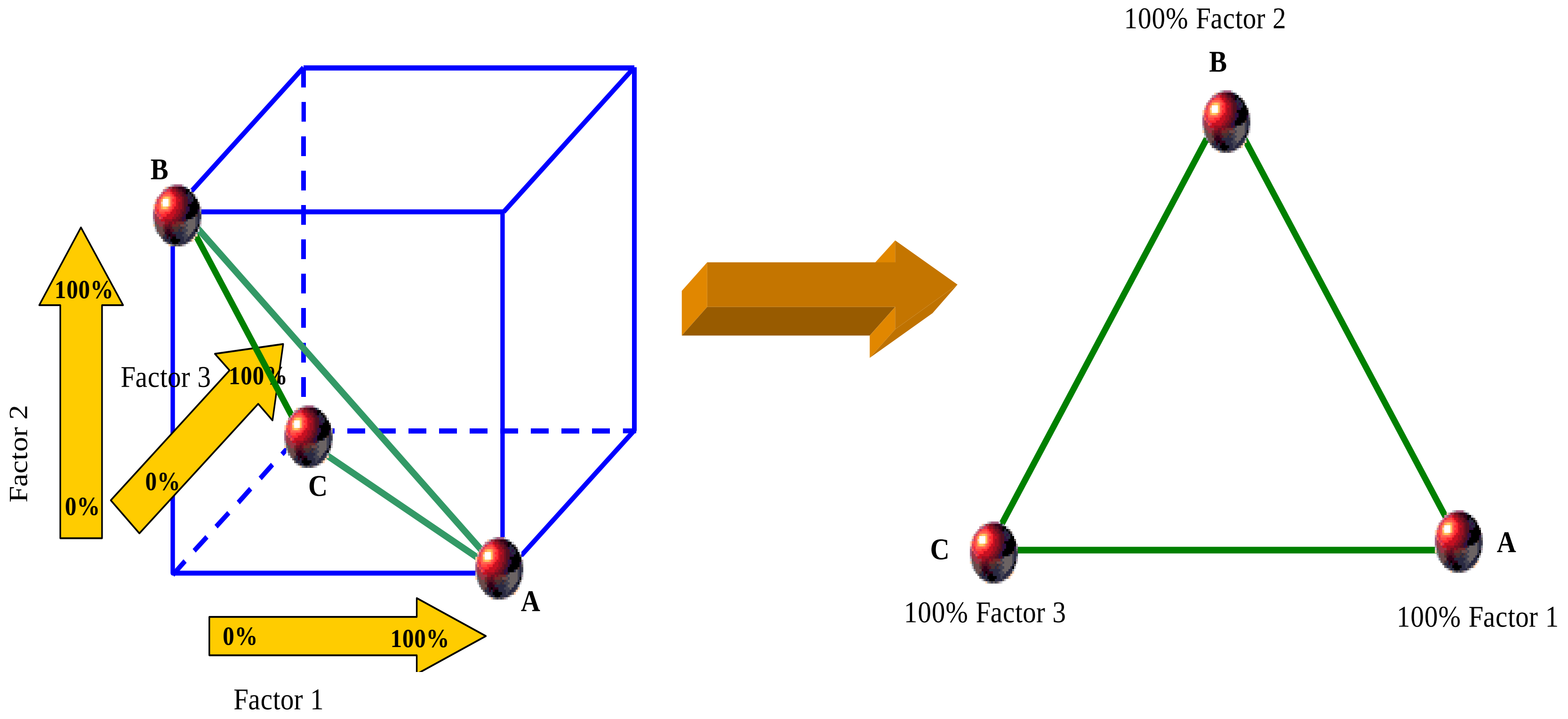
- Rotabelové plány pro m=2 jsou dány vrcholy *n*-úhelníka vepsaného do kružnice. Nemusí být obecně kompozitní.
- Výhoda- méně experimentů



Speciální případ kompozitních plánů. Speciální volba α tak, aby ve všech periferních bodech bylo D(y predikované) stejné. Optimální je $\alpha = 2^{m/4}$



Komponentní plány



Komponentní plány