Ročník 2012 electroscope Číslo III

ANOVA – Základní metoda vyhodnocování experimentů

M. Motyčka, O. Tůmová
Katedra technologií a měření, Fakulta elektrotechnická, ZČU v Plzni,
Univerzitní 26, Plzeň
E-mail : mmotycka@ket.zcu.cz, tumova@ket.zcu.cz

Anotace:
Průmyslové experimenty jsou základním nástrojem pro zajištění kvality výroby. Nezastupitelnou roli mají především při prvotním vývoji daného výrobku a při následném zlepšování kvality. Je třeba použít systematický návrh experimentu, aby na jeho vyhodnocení mohyl být použity statistické metody. Základními statistickými metodami pro hodnocení experimentů jsou regresní analýza a analýza rozptylu ANOVA. Cílem tohoto článku je čtenáři poskytnout základní přehled o analýze rozptylu a jejího použití ve vyhodnocování experimentů.

Industrial experiments are a fundamental tool to ensure quality production. It has an essential role especially in the initial development of the product and the subsequent quality improvement. For statistical evaluation is necessary to use a systematic design of the experiment. Basic statistical methods for the evaluation experiments are regression analysis and analysis of variance ANOVA. The aim of this paper is to provide the reader an overview of the analysis of variance and its use in evaluation of experiments.

ÚVOD

V současnosti jsou velmi vysoké nároky na kvalitu výroby. Trend je takový, že se tyto nároky ještě zpřísňují. Proto je důležité, mít zajištěnou kvalitu výroby už od samotného prvopočátku. Je třeba tedy začít řešit kvalitu výroby už ve vývoji. Platí základní poučka v řízení kvality, že čím později se v životním cyklu u výrobku objeví případná zásadní chyba, tím jsou následky, především ekonomické, větší.

Jedním z nástrojů, které nám pomáhají se zajišťováním kvality výroby již ve vývoji, je průmyslový design experimentů. Pomocí experimentů je možné získat informace o chování znaku kvality v závislosti na různých faktorech, které působí na výrobní proces. Může se jednat o např. o vliv různých vstupních materiálů, výrobních nástrojů od různých dodavatelů, případně je možné zkoumat vlivy prostředí na kvalitu výroby. Důležitým faktorem působícím na výrobu je samozřejmě také nastavení samotných parametrů výrobního procesu.

V počátcích experimentální práce se provádějí často tzv. *screeningové experimenty*. Účelem tohoto typu experimentu je zmapovat všechny možné faktory, které mohou na výrobní proces působit a z nich vybrat takové, jejichž působení bude statisticky významné. Tyto experimenty mohou probíhat i se značným množstvím faktorů. Pokud zkoumáme vliv více faktorů, mohou do experimentu vstupovat navíc ještě interakce mezi faktory, a to teoreticky i několikerého řádu, tj. vliv tří a více faktorů současně. To znamená, že na odezvu nepůsobí jen nastavení daného faktoru, ale i současné nastavení faktorů ostatních, které do procesu vstupují. S jednotlivými interakcemi se proto v návrhu experimentu počítá jako se samostatnými faktory. Proto se u screeningových experimentů využívá metoda *d-optimálních plánů*. Zásadní výhodou této metody je, že umožňuje přímo zvolit počet experimentů, které chceme provádět. Při větším množství faktorů, a jejich úrovní, klasické experimenty bobtnají do obrovských rozměrů a stávají se tak především časově, ale tedy i ekonomicky, neúnosné.

D – OPTIMÁLNÍ PLÁNY

Metoda d – optimálních plánů vychází z úplného faktorového návrhu experimentů. Z toho faktorového experimentu se následně pomocí matematického modelu experimentu sestavuje tzv. matice kandidátních bodů. Počet řádků v této matici odpovídá počtu experimentů, které jsme si stanovili. U modelu procesu je důležité vědět, zda se jedná o model lineární, či zda do procesu vstupují některé interakce. V průmyslové praxi, pokud se nějaké interakce vyšetřují, tak se většinou řeší jen interakce prvého řádu, tj. interakce dvou faktorů. Interakce vyšších řádů se zanedbávají automaticky z předpokládaného důvodu jejich zanedbatelného vlivu na odezvu.

Minimální počet pokusů je dán počtem prvků v matematickém modelu. Za optimální se poté považuje takový experiment, který splňuje dané kritérium optimality. Těchto kritérií je několik. Nejčastěji používané je tzv. D-kritérium. Toto kritérium počítá s *informační maticí* $\mathbf{X}^T\mathbf{X}$ dané matice návrhu. Jako d-optimální se označí takový návrh, jehož informační matice má maximální determinant.

Nevýhodou této metody je značná výpočetní náročnost, která plyne z počtu možností, kterými je možné z daného plného návrhu vybrat kandidátní matici o daném počtu řádků. Z tohoto důvodu většina statistických softwarů, které s *d-plány* pracují, využívá některou z numerických metod, které byly pro tento účel vyvinuty. Většina těchto numerických metod, vychází z tzv. *výměnného algoritmu.*

Tento typ algoritmů vychází z matice návrhu $\mathbf{X_n}$, která obsahuje *n* počátečních experimentů. Cílem těchto algoritmů je záměna bodů v matici návrhu za body z kandidátní matice tak, aby změna velikosti determinantu informační matice byla co největší. K tomu se využívá tzv. *variační funkce d(x):*

$$d(x) = x'(X'_n X_n)^{-1} x\,. \qquad (1)$$

Po přidání experimentu x^j do matice návrhu vzniká nová matice $\mathbf{X_{(n+1)}}$. K popsání vztahu mezi maticí novou a původní může být použita matice informační a platí, že [10]:

$$(X'_{(n+1)} X_{(n+1)}) = (X'_n X_n) + (x_j \times x'_j). \qquad (2)$$

K výpočtu determinantu nové informační matice je poté možné použít právě variační funkci d(x). Platí následující vztah [1].

$$\left|X'_{(n+1)} X_{(n+1)}\right| = |X'_n X_n| \times (1 + d(x_j)). \qquad (3)$$

V případě, že z matice návrhu budeme odebírat prvek X_i tak, abychom získali experiment s původním počtem pokusů, budou vztahy pro výpočet odpovídat rovnicím (2) a (3) s tím rozdílem, že prvky se budou v rovnicích odečítat. Odebíráme takový experiment, jehož variační funkce je nejnižší.

Z tohoto základního principu vycházejí další numerické metody, např. metoda DETMAX nebo Fedorovův algoritmus. Podrobnější informace o těchto metodách je možné najít např. v [1] a [2].

ANALÝZA ROZPTYLU

Analýza rozptylu *ANOVA* je základním nástrojem, který se používá k vyhodnocování experimentů. Tato metoda umožňuje analyzovat rozptyl mezi jednotlivými experimenty a tedy i mezi nastavením jednotlivých faktorů. Tento rozptyl je ovlivněn jednak známými zdroji variability (tj. změnami úrovní jednotlivých faktorů), ale i dalšími vlivy. Mezi ty patří zejména chyby při nastavování jednotlivých úrovní. Jedná se nejen o přesnost samotného nastavení, ale například i o jeho stabilitu. Případným zdrojem variability mohou být například i faktory, které na proces působí, ale nebyly do samotného experimentu zařazeny (např. pokud na provedení celého experimentu je třeba více než jedna šarže materiálu). Dalším vlivem mohou být i tzv. *šumové faktory*, tj. faktory jejichž vliv na výrobní proces může být značný, ale nejsme schopni je jakýmkoli (ekonomicky přijatelným) způsobem regulovat. V takovémto případě se pak provádí tzv. *robustní návrh* experimentu, jehož úkolem je najít takové nastavení řiditelných faktorů, při kterém je možné vliv šumových faktorů zanedbat, resp. se stává statisticky nevýznamným. Všechny zdroje variability, mimo variabilitu způsobenou změnou nastavení úrovní jednotlivých faktorů, způsobují tzv. *reziduální rozptyl* (nebo též experimentální chybu). Pomocí analýzy rozptylu můžeme určit, které faktory jsou statisticky významné, a které faktory je možné z experimentu odstranit, protože jejich vliv není statisticky prokazatelný.

Obecně lze analýzu rozptylu chápat jako zobecnění t-testu. Základní nulovou hypotézou, která se pomocí anovy testuje, je možné zapsat takto:

$$H_0\colon \mu_1 = \mu_2 = \cdots = \mu_m, \qquad (4)$$

kde m > 2 a značí celkový počet faktorů se střední hodnotou μ_m a rozptylem σ^2.

Dalším důležitým krokem v analýze rozptylu je určit, zda budeme anovu provádět pro pevné nebo náhodné efekty. V případě pevných efektů provádíme analýzu pro předem stanovené efekty. Námi zjištěné výsledky jsou poté aplikovatelné jen na dané efekty a není možné tyto závěry žádným způsobem zobecnit na ostatní vlivy, které nebyly do analýzy rozptylu zahrnuty. Toto je případ, který nastává ve většině průmyslových experimentů. Druhým typem je analýza rozptylu pro efekty náhodné. V tomto případě provádíme analýzu pro efekty, které náhodně zvolíme z velkého množství. V tomto případě lze poté výsledky zobecnit i na efekty, které jsme přímo netestovali. Tento případ se ale v průmyslových experimentech vyskytuje jen zřídkakdy, a proto se jím toto článek zabývat nebude. Tuto problematiku lze dohledat ve velkém množství publikací, např. v [3].

Celkový pozorovaný rozptyl SS_T má tři složky:

- odchylka celkového průměru všech výsledků SS_m vzhledem k nule;
- rozptyl průměrů výsledků pro každou úroveň faktoru SS_A vztažený k průměru celkovému;
- rozptyl individuálních výsledků pro každou úroveň SS_E vztažený k průměru pro každou úroveň (tzv. reziduální rozptyl).

Odhad rozptylu mezi jednotlivými třídami je nestranný jen tehdy, pokud platí nulová hypotéza H_0. Pokud tato hypotéza neplatí, pak se budou statisticky významně lišit odhady rozptylu uvnitř tříd a mezi jednotlivými třídami. Příčinou tohoto rozdílu je to, že se alespoň jedna hodnota významně liší od ostatních. Testování hypotézy H_0 se prakticky provádí F – testem, kterým srovnáváme rozptyly mezi a uvnitř tříd:

$$F = \frac{rozptyl\ mezi\ průměry\ skupin\ {}^{SS_X}/_{f_X}}{rozptyl\ uvnitř\ skupin\ {}^{SS_E}/_{f_E}}, \quad (5)$$

kde f značí počty stupňů volnosti dané veličiny.

Testovaná veličina F má F – rozdělení s počtem stupňů volnosti daném typem experimentu a počtem faktorů. Hypotéza H_0 zamítáme, pokud platí:

$$F_{vypočtené} > F_{m-1,\, m(n-1)}(\alpha), \quad (6)$$

kde $m-1$ a $m(n-1)$ značí počty stupňů volnosti. Pokud nerovnost (6) neplatí, tak se nulová hypotéza H_0 o rovnosti středních hodnot nezamítá.

ANOVA pro jednofaktorový experiment

U tohoto typu experimentů se zkoumá jen vliv jednoho faktoru A, resp. vliv nastavení jednotlivých úrovní daného faktoru A na odezvu. Přičemž platí následující vztahy.

Součty čtverců jednotlivých zdrojů:

$$SS_A = \frac{\sum (X_{i.})^2}{n_i} - \frac{(X_{..})^2}{n}, \quad (7)$$

$$SS_T = \sum_{i=1}^{I} \sum_{p=1}^{n_i} X_{ip}^2 - \frac{(X_{..})^2}{n}. \quad (8)$$

Do celkového pozorovaného rozptylu ještě vstupuje vliv replikací (ve vzorcích s indexem p). Replikace je nezávislé opakování každé kombinace faktorů. Replikace, společně s randomizací a technikou bloků, patří mezi základní principy návrhu experimentů.

Reziduální součet čtverců se vypočte dle vztahu:

$$SS_E = SS_T - SS_A. \quad (9)$$

Jednotlivé stupně volnosti:

$$f_A = I - 1, \quad (10)$$

$$f_T = n - 1, \quad (11)$$

$$f_E = n - I. \quad (12)$$

Testovací statistika se poté vypočte dle:

$$F_A = \frac{SS_A}{f_A} \frac{f_E}{SS_E} = \frac{SS_A}{SS_E} \frac{n-I}{I-1}. \quad (13)$$

Aby platila nulová hypotéza $H_0 : \mu_0 = \mu_1$, musí platit $F_A \leq F_{krit}$, kde kritická hodnota odpovídá $F_{I-1,\, n-1}(\alpha)$. Hodnota α označuje tzv. *hladinu významnosti* (nebo také chybu I. druhu). Zjednodušeně můžeme tvrdit, že hladina významnosti určuje pravděpodobnost, že zamítneme správnou hypotézu. Nejčastěji se volí hodnota $\alpha = 5\%$.

Pro určení, které úrovně faktoru A způsobily zamítnutí hypotézy H_0, se používají tzv. metody mnohonásobného porovnávání. Mezi tyto metody patří například známá Schefféhо či Tukeyova metoda, případně metoda Duncanova či Fisherova LSD (Less significant difference).

ANOVA pro dvoufaktorový experiment s interakcemi

Z jednofaktorové analýzy rozptylu rozšířením o další faktory a případné interakce, získáme analýzu rozptylu pro mnohafaktorové experimenty. Princip fungování anovy u takto rozsáhlých experimentů, lze snadno předvést na dvoufaktorovém experimentu s interakcemi, které u tohoto typu experimentu mohou být jen prvého řádu. Samozřejmě pokud bude působících faktorů více, může do interakcí teoreticky vstupovat i 3 a více faktorů.

Nulová hypotéza je definována rovnicí (4). Vypočtená testovací statistika se porovnává s kritickou hodnotou F_x, která odpovídá počtům stupňů volnosti jednotlivých faktorů a zvolené hladině významnosti. Pokud chceme v experimentu vyšetřovat interakce, je nutné provést minimálně dvě replikace ($P \geq 2$). Celkový počet výstupů z experimentu se tedy rovná $n = I\ J\ P$, kde I a J jsou počty úrovní faktorů A a B. Model odezvy odpovídá rovnici:

$$X_{ijp} = \mu + \alpha_i + \beta_j + \tau_{ij} + \varepsilon_{ijp}, \quad (14)$$

kde μ je průměrná hodnota,
α_i, β_j je vliv faktorů A a B,
τ_{ij} je vliv interakcí mezi faktory A a B,
ε_{ijp} je náhodná chyba s rozdělením N (0; σ^2).

Součty jednotlivých čtverců mají tvar:

$$SS_A = \frac{1}{IP} \sum_{i=1}^{I} X_{i..}^2 - \frac{1}{n} X_{...}{}^2, \quad (15)$$

$$SS_B = \frac{1}{JP} \sum_{j=1}^{J} X_{.j.}^2 - \frac{1}{n} X_{...}{}^2, \quad (16)$$

$$SS_E = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{p=1}^{P} X_{ijp}^2 + \frac{1}{P} \sum_{i=1}^{I} \sum_{j=1}^{J} X_{ij}^2, \quad (17)$$

$$SS_T = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{p=1}^{P} X_{ijp}^2 + \frac{1}{P} X_{...}^2, \quad (18)$$

Tab. 1: Tabulka ANOVA pro dvoufaktorový experiment s interakcí

zdroj variability	součet čtverců	stupně volnosti (DF)	odhad rozptylu (MS)	testová veličina	kritická hodnota	P hodnota
faktor A	SS_A	f_A	SS_A / f_A	$F_A = \frac{SS_A/f_A}{SS_E/f_E}$	$F_{fA,fE}\ (\alpha)$	
faktor B	SS_B	f_B	SS_B / f_B	$F_B = \frac{SS_B/f_B}{SS_E/f_E}$	$F_{fB,fE}\ (\alpha)$	
interakce AB	SS_{AB}	f_{AB}	SS_{AB} / f_{AB}	$F_{AB} = \frac{SS_{AB}/f_{AB}}{SS_E/f_E}$	$F_{fAB,fE}\ (\alpha)$	
reziduální	SS_E	f_E	SS_E / f_E	---	---	
celkový	SS_T	f_T	---	---	---	

$$SS_{AB} = SS_T - SS_A - SS_B - SS_E. \quad (19)$$

A počty stupňů volnosti mají tvar:

$$f_A = I - 1, \quad (20)$$

$$f_B = J - 1, \quad (21)$$

$$f_E = n - IJ, \quad (22)$$

$$f_T = n - 1, \quad (23)$$

$$f_{AB} = (I - 1)(J - 1). \quad (24)$$

Vliv daného faktoru je prokázán na zvolené hladině významnosti α, pokud příslušná testovací statistika F_x je větší než hodnota kritická. Testovací statistika F_x se určí dle vztahu:

$$F_X = \frac{SS_X/_{f_x}}{SS_E/_{f_E}}, \quad (25)$$

kde SS_x a f_x odpovídá příslušným součtům jednotlivých rozptylů a stupňů volnosti příslušných danému faktoru.

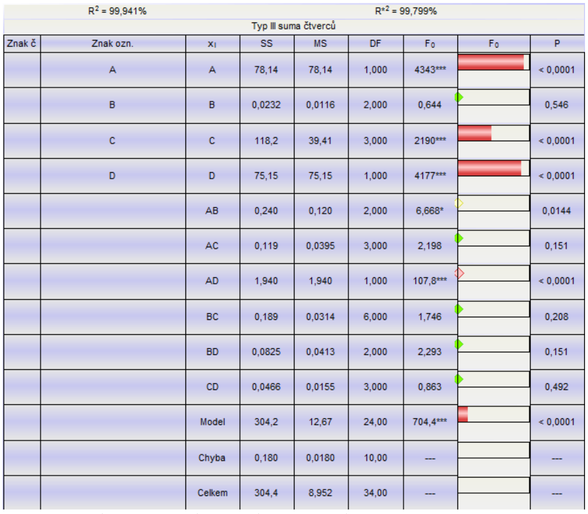
V případě experimentu bez interakcí se z výše uvedených rovnic vypustí ty členy, které odpovídají právě interakcím. Takovýto experiment lze provést i v případě, že počet replikací je roven P ≥ 1.

INTERPRETACE VÝSLEDKŮ ANOVA

Ve výstupech různých statistických SW se můžeme setkat s tzv. *P-hodnotou*. Tato hodnota odpovídá pravděpodobnosti F – rozdělení s danými stupni volnosti pro daný součet čtverců. Platí, že pokud je P > α, tak hypotézu H_0 nezamítáme. A naopak pokud platí, že P < α, zamítáme hypotézu H_0 o shodnosti průměrů a vliv daného faktoru můžeme považovat za prokázaný.

Interpretaci výsledků je možné demonstrovat na následujícím vzorovém příkladu. V tomto případě se bude jednat o čtyřfaktorový experiment, v němž jsou uvažovány interakce prvého řádu. K výpočtu analýzy rozptylu byl použitý statistický software Destra od Q-DAS GmbH. Je možné použít celou další řadu statistických SW a analýzu rozptylu zvládá i MS Excel.

Na obr. 1 vidíme výstup z multifaktorové analýzy rozptylu ze SW Destra.



R² = 99,941%						R*² = 99,799%			
Typ III suma čtverců									
Znak č	Znak ozn.	x_I	SS	MS	DF	F_0	F_0	P
	A	A	78,14	78,14	1,000	4343***		< 0,0001
	B	B	0,0232	0,0116	2,000	0,644		0,546
	C	C	118,2	39,41	3,000	2190***		< 0,0001
	D	D	75,15	75,15	1,000	4177***		< 0,0001
		AB	0,240	0,120	2,000	6,668*		0,0144
		AC	0,119	0,0395	3,000	2,198		0,151
		AD	1,940	1,940	1,000	107,8***		< 0,0001
		BC	0,189	0,0314	6,000	1,746		0,208
		BD	0,0825	0,0413	2,000	2,293		0,151
		CD	0,0466	0,0155	3,000	0,863		0,492
		Model	304,2	12,67	24,00	704,4***		< 0,0001
		Chyba	0,180	0,0180	10,00	---		---
		Celkem	304,4	8,952	34,00	---		---

Obr. 1 Analýza rozptylu pro daný experiment

Můžeme pozorovat, že statisticky významné jsou faktory A, C a D a interakce mezi faktory A a D. P – hodnoty u těchto faktorů jsou minimální. SW Destra také umožňuje eliminaci jednotlivých faktorů. V tomto případě jsou postupně z vyhodnocení odebírány faktory, jejichž P – hodnota je největší, tedy vliv na odezvu je nejmenší a celá analýza rozptylu je znovu přepočítána. Tím získáme přesnější výsledky, protože do experimentu již nevstupuje variabilita, o které jsme prokázali, že na odezvu nemá vliv. Výsledky po eliminaci vidíme na obr. 2.

R^2 = 99,705%					R^{*2} = 99,642%			
Typ III suma čtverců								
Znak č	Znak ozn.	x_i	SS	MS	DF	F_0	F_0	P
	A	A	82,02	82,02	1,000	2559***		< 0,0001
	C	C	132,5	44,15	3,000	1377***		< 0,0001
	D	D	77,58	77,58	1,000	2420***		< 0,0001
		AD	3,407	3,407	1,000	106,3***		< 0,0001
		Model	303,5	50,58	6,000	1578***		< 0,0001
		Chyba	0,898	0,0321	28,00	---		---
		Celkem	304,4	8,952	34,00	---		---

Obr. 2 Analýza rozptylu po eliminaci nevýznamných faktorů

Z obrázku 2 je patrné, že ve vyhodnocení zůstaly stejné faktory, jako v prvním případě a tedy můžeme považovat jejich vliv za prokázaný.

ZÁVĚR

Účelem tohoto článku bylo čtenáře seznámit s analýzou rozptylu a její aplikací pro vyhodnocování experimentů. Analýza rozptylu ANOVA je základní metoda, která nám umožňuje rozlišit, které faktory působící na výrobní proces jsou důležité a které nikoli. Po tomto zjednodušení je možné se podrobně zaměřit na faktory významné a provést komplexnější zkoumání vlivu těchto faktorů. K těmto účelům se nejčastěji využívá regresní analýzy.

Tato práce byla podpořena grantem Studentské grantové soutěže ZČU č. SGS-2012-026 „Materiálové a technologické systémy v elektrotechnice"

LITERATURA

[1] MITCHELL, T. J.
An Algorithm for the Construction of D-Optimal Experimental Designs. Technometrics. May 1974, Vol. 16, No. 2, s. 203-210. Dostupný také z WWW: http://www.jstor.org/

[2] DE AGUIAR, P. F., et al.¨
D-optimal designs. Chemometrics and Intelligent Laboratory Systems . 1995, vol. 30, Issue 2, p. 199-210. Dostupný také z WWW: http://www.sciencedirect.com

[3] MONTGOMERY, Douglas C.
Design and analysis of experiments. 7th ed. Hoboken: John Wiley , 2009, 656 s. ISBN 978-047-0398-821

[4] TŮMOVÁ, Olga.
Metrologie a hodnocení procesů. 1. vyd. Praha: BEN - technická literatura, 2009, 231 s. ISBN 978-807-3002-497

[5] MONTGOMERY, Douglas C.
Statistical Quality Control: A modern introduction. 6th ed. [s.l.]: John Wiley & Sons, 2009. 734 s. ISBN 978-0470-23397-9

[6] LIKEŠ, Jiří.
Navrhování průmyslových experimentů. 1. vyd. Praha: SNTL, 1968, 236 s