PLÁN EXPERIMENTU A STATISTICKÁ KOREKTNOST MODELU

BOHUMIL MAROŠ

ABSTRAKT. Kvalitní příprava, realizace a analýza vědecko-výzkumných experimentů z laboratorních či (polo)provozních pokusů předpokládá osvojení si určitého souboru poznatků a systematicky vybudované metodiky známé pod zkratkou DOE – Design of Experiments. V případě použití „klasického“ regresního přístupu k analýze dat experimentů pro původní (netransformovaná) data z experimentu sestaveného a realizovaného podle metodiky DOE může dojít k různým nesrovnalostem a numerickým, statistickým i principiálně závažným problémům s dopadem na znehodnocení celé experimentální práce. V příspěvku jsou rozebírány numerické a následně statistické aspekty zmíněného nesouladu.

1. Úvod

Správný návrh, provedení a následná analýza (vyhodnocení) technologicko-technických experimentů nezbytně patří (nebo rozhodně by měla patřit) k základnímu vybavení studentů i absolventů (nejen) technických univerzit, výzkumných i vědeckých pracovníků. Zanedbání určitých principů plánování a provádění experimentů může vést ke znehodnocení experimentátorské práce. Cílem příspěvku je proto stručné seznámení s pojmy, principy a metodikou *plánování experimentů*, jako i upozornění a dokumentace možných problémů na srozumitelném příkladu.

2. Základní pojmy a principy plánování experimentů

Častým úkolem (nejen) technologicko-technické praxe je zjištění vazeb a vztahů mezi určitými veličinami zkoumaného procesu. Jde zvláště o případy, kdy proces je velice složitý a neexistuje pro jeho popis (dostatečně) vhodný matematický fyzikálně-technický model.

2.1. Základní pojmy DOE

Základním a nejčastějším cílem experimentu je určení, zda určité *faktory* (ovlivňující, vstupní, vysvětlující veličiny) mají vliv na sledovanou (ovlivňovanou, výstupní, vysvětlovanou) veličinu, často nazývanou *odezva* (response). Další, či následující úlohou může být nalezení takové úrovně faktorů, aby bylo dosaženo *optima* (maxima, minima) sledované veličiny. Potřebná data pro sestavení modelu lze získat

---

2010 *MSC.* Primární 34K06, 34K25.

Klíčová slova. Regresní analýza, DOE, ortogonalita.

Práce byla podporována projektem A-Math-Net – Síť pro transfer znalostí v aplikované matematice (CZ.1.07/2.4.00/17.0100).

buď *pozorováním* veličin procesu bez cíleného zásahu do něj (tzv. *pasivní* či neplánovaný experiment) anebo uskutečněním *experimentu* s cílenými zásahy do procesu (tzv. *aktivní* či *plánovaný* experiment). Avšak ani v případě aktivních experimentů nemají vždy experimentátoři dostatek znalostí o principech efektivního provedení experimentů a vůle k jejich uskutečnění. Je třeba s politováním konstatovat, že *žádnou analýzou (žádnou metodou) experimentálních dat nelze obejít špatně či nedostatečně připravený experiment* [3].
Termínem *experiment* se označuje soustava *pokusů* (také měření, pozorování), která je v případě plánovaného experimentu *vhodným způsobem uspořádána.* V podstatě lze říci, že v plánovaném experimentu jde o vytvoření takových podmínek, aby *rozsah* experimentu byl co nejmenší, ale objem i forma *informací* co nejkvalitnější. V důsledku to také znamená, že u faktorů budeme vybírat jejich vhodné *úrovně* v rámci zvolených (technologicky vhodných) intervalů.
Požadavek na efektivitu se uplatňuje ještě mnohem výrazněji u experimentu s mnoha *faktory* (nezávislými proměnnými). Jeho důsledné respektování vedlo k vytvoření samostatného odvětví aplikované statistiky, tzv. plánovaného (řízeného) experimentu. Různé návrhy uspořádání *měření* a metody jejich *vyhodnocení* se souhrnně označují jako plánování (navrhování) experimentů. V literatuře se často označují jako *Design of Experiment* (*DOE*).

2.2. Základní principy a přístupy metody DOE

Mezi základní principy a přístupy tvorby plánů experimentů podle metody DOE patří následující:

- *replikace*: je opakování měření při stejné úrovni nebo kombinaci úrovní faktorů. Tímto způsobem lze odhadnout nepřesnost měření (náhodnou složku) a zvýšit spolehlivost závěru tím,že můžeme použít test adekvátnosti (přiměřenosti, vhodnosti, použitelnosti) modelu,
- *rozdělení do bloků*: pro poměrně stejné podmínky experimentu za účelem odlišení dalšího zdroje variability,
- *znáhodnění*: aby nevznikla systematická chyba (např. stejným pořadím úrovní faktoru v každém bloku), provedeme jednotlivé pokusy (měření) experimentu v náhodném pořadí. Celé schéma komplexního experimentu pak nazýváme *znáhodněné (replikované) bloky.* Ty dovolují rozložit celkovou variabilitu na složku odpovídající efektům úrovní zkoumaného faktoru, složku nepřesnosti měření, složku odpovídající blokům a reziduální (zbytkovou, nevysvětlenou) složku, jež zahrnuje vliv ostatních (nezkoumaných) činitelů.

Z důvodu minimalizace počtu pokusů v experimentu se používají (pouze) *dvou* a *tříúrovňové* experimenty, které mohou být *úplné* (faktorové) a *zkrácené.*
Plánovaný experiment (*DOE*) poskytuje návod jak *současně měnit všechny faktory* na několika *málo* jejich *úrovních* najednou (z toho vyplývá *minimalizace* počtu pokusů v experimentu) při získání *maxima* informací pro výstavbu vhodného a použitelného modelu zkoumaného procesu.

2.3. Možné problémy při nepoužití metodiky DOE

Pokud budeme spoléhat jen na tzv. „selský rozum“, či „expertní posouzení“ (což zvláště „láká“ u malého počtu, tj. u 2 a 3 faktorů i zkušené experimentátory), pak může (téměř určitě) dojít k několika nepříjemným skutečnostem, např.:

- špatně provedeme *výběr bodů* (hodnot pokusů, úrovní faktorů) s důsledkem *nesprávně indikované statistické nevýznamnosti* některých faktorů (přičemž tyto jsou významné a vlivné), a tím i ke stanovení nesprávného („okleštěného“) modelu,
- sestavíme nevhodný *návrh s přebytečnými údaji* pro model za „dobrým“ účelem získání maxima (ale ne všech potřebných) informací,
- *metoda „pokus-omyl“* jenom málokdy (když se „náhodou strefíme“) poskytne užitečné informace o závislostech,
- *metoda změny pouze jednoho faktoru* při dalších faktorech konstantních (tzv. jednofaktorový experiment, plán) vede jak k příliš (zbytečně) vysokému počtu pokusů (a s tím souvisejícího dlouhého času a vynaložení mnoha peněz na takovýto experiment), tak i k zásadní skutečnosti nemožnosti podchycení (často existujících) interakcí faktorů,
- použití *nevhodné metody analýzy* (při správné metodice návrhu a provedení experimentu metodou DOE) např. lineární regresní analýzy místo analýzy přístupem DOE na *původní* data může také (neočekávaně a překvapivě) vést výše uvedené *nesprávné indikaci statistické nevýznamnosti* některých (ve skutečnosti významných) faktorů s důsledkem na stanovení nesprávného (příliš zjednodušeného) modelu.

3. Plán experimentu podle DOE

S ohledem na *minimalizaci* počtu pokusů (a tím i času a peněz na jejich provedení), lze uvažovat, že každý faktor z má pouze dvě (krajní) úrovně. Ty lze pomocí lineární transformace převést na hodnoty s úrovněmi -1 a $+1$. Tyto kódované (transformované) bezrozměrné faktory budou dále označovány symboly $x_0, x_1, \ldots, x_k$ pro k faktorů.
Lineární transformace původního faktoru z_j na její kódovanou hodnotu x_j, $j = 1, \ldots, k$, se provede pomocí vztahu

$$x_j = \frac{z_j - \dfrac{z_{j,max} + z_{j,min}}{2}}{\dfrac{z_{j,max} - z_{j,min}}{2}} \tag{3.1}$$

Plán experimentu je matice $\mathbf{X}$ tvořená ze sloupců kódovaných faktorů, přičemž první sloupec je složen ze samých jedniček a odpovídá fiktivní proměnné x_0. Příslušný regresní parametr je b_0. Řádky matice plánu představují (kódované) hodnoty úrovní jednotlivých faktorů. Počet řádků matice plánu je roven počtu všech pokusů (měření), včetně opakovaných.
Z důvodu podchycení *náhodných chyb měření* a umožnění *testování adekvátnosti modelu* mohou být přidány další *opakované pokusy* (v *krajních* bodech nebo ve středu plánu čili ve *středovém, centrálním* bodu, ve kterém mají všechny kódované faktory hodnotu 0).

Ortogonální plán je takový plán experimentu $\mathbf{X}$, ve kterém jsou všechny sloupcové vektory matice na sebe *kolmé*.
Ze vztahu (3.1) je zřejmé, že kódováním dochází také k převodu z původních fyzikálních jednotek faktorů na *bezrozměrový* tvar.

3.1. Plán experimentu podle DOE v konkrétním případě

Při tváření oceli za studena nebo za ohřevu klade materiál odpor, jenž se mění jak s velikostí deformace, tak s nastavením tvářecí teploty. Úkolem je nalézt vhodný model pro ocel 14240.3, jenž bude popisovat chování tohoto *přetvárného odporu* v závislosti na *stupni deformace* a *teplotě*. Jeden faktor je tedy logaritmický stupeň deformace φ [-], druhým faktorem je teplota t [°C]. Sledovanou veličinou bude přetvárný odpor σ [MPa]. Oblast zkoumání bude v oblasti, jež je vymezena oběma faktory: $\varphi \in \langle 0,0; 1,4\rangle$, $t \in \langle 20; 700\rangle$.
Chceme nalézt co nejjednodušší vztah $\sigma = f(\varphi, t)$, ale takový, aby statistická analýza potvrdila významnost tohoto vztahu. V literatuře [3] je ukázán *postup* výstavby modelu (v součinnosti s upravovanou či doplňovanou maticí plánu) od jednoduchého *lineárního bez interakcí*, přes *lineární model s interakcemi* až po *úplný kvadratický model*. V našem případě uvažovaný hledaný model bude mít jednodušší tvar, a to:

$$\sigma = \beta_0 + \beta_1 \cdot \varphi + \beta_2 \cdot t + \beta_3 \cdot \varphi \cdot t.$$

Odhad tohoto modelu můžeme zapsat ve tvaru

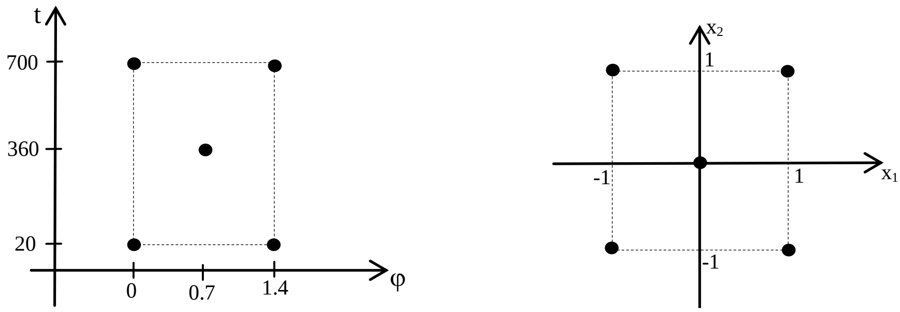
$$\hat{\sigma} = b_0 + b_1 \cdot \varphi + b_2 \cdot t + b_3 \cdot \varphi \cdot t, \tag{3.2}$$

kde b_j jsou odhady regresních parametrů β_j, $j = 0, 1, 2, 3$. Koeficient b_0 je tzv. absolutní (konstantní) člen. Obecně je vhodné s ním vždy počítat (zařazovat jej do modelu), protože podchycuje vliv měřítka (fyzikálních hodnot) odezvy, současně i vliv případné nezařazené vysvětlující proměnné, či existující nelinearitu a umožňuje korektní stanovení statistických ukazatelů (jakými jsou např. koeficient determinace modelu, Durbinova-Watsonova statistika apod.).
Pro daný rozsah uvažovaných veličin φ a t se přepočtou ze vztahu (3.1) transformované (kódované) veličiny takto:

$$x_1 = \frac{\varphi - 0,7}{0,7}, \quad x_2 = \frac{t - 360}{340}. \tag{3.3}$$

Abychom mohli *testovat adekvátnost modelu*, byla přidána ještě dvě *opakovaná* měření, a to ve *středovém* bodu $[0,7; 360]$, což po transformaci (3.3) přejde na bod $[0; 0]$. Původní a transformované hodnoty proměnných jsou názorně viditelné na obr. 1:

Měření uskutečníme v náhodném pořadí, např. v bodech $[1; -1]$, $[-1; -1]$, $[0; 0]$, $[-1; 1]$, $[1; 1]$, $[0; 0]$. Matice plánu $\mathbf{X}$ s transformovanými hodnotami úrovní faktorů v pokusech experimentu pro testovaný model tvaru (3.2) při daném pořadí měření



Obrázek 1. Původní a transformované hodnoty plánu.

pak bude:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{3.4}$$

Druhý sloupec pro x_1 odpovídá původní proměnné φ, třetí sloupec pro x_2 odpovídá původní proměnné t a poslední sloupec pro x_3 odpovídá interakci obou proměnných a získá se vynásobením hodnot z druhého a třetího sloupce.

Pomocí skalárních součinů lze jednoduše ukázat, že tento *plán* je *ortogonální*, tzn., že všechny sloupcové vektory jsou na sebe kolmé.

Po provedeném měření v pořadí, jež určuje matice plánu, jsme získali postupně tyto hodnoty přetvárného odporu: 816, 380, 510, 330, 390 a 497. To znamená, že např. hodnota přetvárného odporu $\sigma = 816\,\mathrm{MPa}$ se obdržela při logaritmickém stupni přetvoření $\varphi = 1,4$ a tvářecí teplotě $t = 20\,^\circ\mathrm{C}$, hodnota $\sigma = 510\,\mathrm{MPa}$ se obdržela při logaritmickém stupni přetvoření $\varphi = 0,7$ a tvářecí teplotě $t = 360\,^\circ\mathrm{C}$, atd.

3.2. Analýza experimentu

Pro úplný faktoriální (faktorový) dvou-úrovňový plán (ale i pro další typy experimentů) je podrobný *postup* (algoritmus, popis) návrhu i *analýzy* uskutečněného experimentu *metodou DOE* uveden např. v literatuře [5]. V této literatuře definovaným názorným postupem je možné „ručně“, „na papíře“ za pomocí kalkulačky, či elegantněji a obecněji prostřednictvím např. tabulkového procesoru Excel, uskutečnit analýzu DOE a získat tvar i koeficienty modelu. Je logické, že v dnešní době je lepší (komplexnější, spolehlivější, rychlejší) použít nějaký vhodný SW obsahující tuto metodu. Existuje speciální SW pro metodiku DOE známý pod anglickou zkratkou (akronymem) CADEX/DOE (Computer Aided Design and Analysis of Experiments / Design of Experiments), mezi něž patří např. programy *Design Expert* či *MODDE.* Avšak asi nejčastěji mají uživatelé (studenti, výzkumníci

a vědci) k dispozici obecné statistické (např. *Minitab, Statgraphics, JMP, Statistica, NCSS*), či matematické (*Matlab*) balíky programů, které lze pro dané účely použít.
Další alternativou, kterou lze použít, je klasická *lineární regrese*, jež je běžně dostupná v různých programech.
Odhad modelu (lineární 1. řádu s interakcí 2. řádu) s transformovanými proměnnými budeme potom uvažovat v analogickém tvaru jako model (3.2), tj.:

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_1 \cdot x_2,$$

kde $\hat{y}$ je vypočtená (predikovaná) hodnota σ z modelu, x_1, resp. x_2, je transformovaná proměnná φ, resp. t.
Neznámé odhadované parametry b_j, $j = 0, \ldots, 3$, získáme ze vztahu (metoda nejmenších čtverců)

$$\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} = \mathbf{M}^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} = \mathbf{V}\mathbf{X}^{\mathrm{T}}\mathbf{y}, \tag{3.5}$$

kde $\mathbf{b} = (b_0, b_1, b_2, b_3)^{\mathrm{T}}$ je vektor odhadovaných parametrů, $\mathbf{X}$ je matice plánu (3.4) a vektor $\mathbf{y} = (816, 380, 510, 330, 390, 497)^{\mathrm{T}}$ obsahuje naměřené hodnoty přetvárného odporu. Matice $\mathbf{M}$ je momentová matice a matice $\mathbf{V}$ se nazývá varianční.
Prakticky lze uvedený vektor regresních koeficientů $\mathbf{b}$ velice elegantně a názorně určit (podle vztahu (3.5) v systému *Matlab, Mathcad, Minitab* či *Statistica*, odkud dostaneme následující vektor odhadů regresních koeficientů:

$$\mathbf{b} = (487,167 \quad 124,000 \quad -119,000 \quad -94,000)^{\mathrm{T}}$$

pro transformované proměnné. Tyto hodnoty můžeme podle (3.3) zpětně přepočítat, abychom obdrželi hodnoty vektoru $\mathbf{b}$ pro původní proměnné φ a t:

$$\mathbf{b} = (389,637 \quad 319,238 \quad -0,073529 \quad -0,394958)^{\mathrm{T}}.$$

Nicméně, pro ověření statistické významnosti modelu, jako i významnosti jeho regresních koeficientů, je nutné stanovit další hodnoty různých statistik a uskutečnit potřebné testy.

4. Regresní analýza příkladu

Pokud uskutečníme statistickou vícenásobnou lineární regresní analýzu pro *kódovaná* a *původní* (netransformovaná) *data* uvedeného příkladu v systému *Matlab* (ale obecně i v libovolném jiném matematickém, statistickém, či ekonometrickém programu), pak sice dostaneme (po přepočtu, po zpětné transformaci) *numericky* stejné hodnoty regresních koeficientů, avšak jejich *statistická* významnost (statistické vlastnosti) se budou u obou modelů (pro oba druhy dat) lišit.

4.1. Statistické „pozadí“ hodnocení regrese

Před hodnocením statistických výstupů regrese je vhodné definovat vztahy vedoucí k jejich určení. Skutečnost, že vypočtený odhad regresního parametru modelu má nenulovou hodnotu ještě neznamená, že je statisticky významně odlišný od nuly (větší nebo menší než nula). To lze posoudit až po určení jeho nepřesnosti charakterizované směrodatnou odchylkou.

Nestranným odhadem rozptylu σ_2 náhodných veličin je náhodná veličina

$$s^2 = \frac{RSC}{n-k-1}, \quad RSC = (y-\hat{y})^{\mathrm{T}}(y-\hat{y}),$$

kde n je počet měření, k je počet faktorů v modelu a RSC je reziduální součet čtverců. Rozptyl jednotlivých náhodných veličin b_j je dán vztahem

$$s^2(b_j) = s^2 \cdot v_{jj}, \tag{4.1}$$

kde v_{jj} je diagonální prvek matice $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$. Jestliže platí, že

$$T = \frac{|b_j|}{\sqrt{s^2(b_j)}} \geq t\left(1-\frac{\alpha}{2}, n-k-1\right), \tag{4.2}$$

kde $t(\dots)$ je kvantil Studentova rozdělení, pak zamítáme hypotézu o statistické nevýznamnosti („nulovosti“) regresního koeficientu na hladině α. Jinými slovy, j-tý sloupec matice $\mathbf{X}$ má významný vliv na hodnoty vektoru $\mathbf{y}$.
Další možností hodnocení statistické významnosti regresního koeficientu je určit tzv. *phodnotu* (p-value), čili dosaženou hladinu pravděpodobnosti (významnosti) zamítnutí hypotézy o nevýznamnosti tohoto regresního koeficientu a porovnat ji se zvolenou „kritickou“ hladinou významnosti testování α (většinou se volí hodnota $0,05$).

4.2. Transformovaná data

S využitím výše definovaných vztahů byla v programu *Minitab* uskutečněna lineární regresní analýza s diagnostikou statistické významnosti regresních koeficientů nejprve pro transformovaná (kódovaná) data podle (3.4) s tímto výsledkem:

Regression Analysis: sigma versus fi; t; fi*t

```
The regression equation is
sigma = 487 + 124 fi - 119 t - 94,0 fi*t


Predictor      Coef  SE Coef       T      P    VIF
Constant    487,167    8,587   56,73  0,000
fi           124,00    10,52   11,79  0,007  1,000
t           -119,00    10,52  -11,32  0,008  1,000
fi*t         -94,00    10,52   -8,94  0,012  1,000

S = 21,0337   R-Sq = 99,4%   R-Sq(adj) = 98,6%
```

(4.3)

V tomto výstupu Minitabu ve sloupci *Coef* jsou hodnoty jednotlivých koeficientů b_j, ve sloupci *SE Coef* hodnoty směrodatných odchylek b_j, ve sloupci T jsou hodnoty T ze vztahu (4.2), ve sloupci P jsou p-hodnoty a ve sloupci *VIF* jsou hodnoty Variance Inflation Factor definované jako

$$VIF_j = v_{jj}\sum_{i=1}^{n}(X_{ij}-\overline{X}_j)^2.$$

Ve výstupu Minitabu se ještě vypíše analýza rozptylu:
V řádku *Residual Error* a sloupci *SS* je hodnota RSC a sloupci *MS* je odhad rozptylu s^2. Důležitá je zde p-hodnota v řádku *Lack of Fit*, poněvadž zde se testuje

```
Analysis of Variance

Source           DF      SS     MS       F      P
Regression        3  153492  51164  115,65  0,009
Residual Error    2     885    442
  Lack of Fit     1     800    800    9,47  0,200
  Pure Error      1      85     85
Total             5  154377
```
(4.4)

adekvátnost modelu. Je-li $P < \alpha$, pak vypočtený model není adekvátní (neodpovídá naměřeným hodnotám). Tento řádek se ve výstupu neobjeví, pokud nejsou žádná opakovaná měření.

4.3. Původní data

V kapitole 4.2 byla prováděna regresní analýza z transformovaných dat na interval $\langle -1; 1\rangle$. Nyní provedeme regresní analýzu na původní data pro proměnné φ a t. To znamená, že matice plánu nebude podle (3.4), ale bude mít tvar:

$$\mathbf{X} = \begin{pmatrix} 1 & 1,4 & 20 & 28 \\ 1 & 0 & 20 & 0 \\ 1 & 0,7 & 360 & 252 \\ 1 & 0 & 700 & 0 \\ 1 & 1,4 & 700 & 980 \\ 1 & 0,7 & 360 & 252 \end{pmatrix}.$$

Opět pomocí Minitabu obdržíme:

Regression Analysis: sigma versus fi; t; fi*t

```
The regression equation is
sigma = 390 + 319 fi - 0,0735 t - 0,395 fi*t


Predictor       Coef  SE Coef      T      P    VIF
Constant      389,64    20,79  18,74  0,003
fi            319,33    21,88  14,59  0,005  2,121
t           -0,07353  0,04374  -1,68  0,235  2,000
fi*t        -0,39496  0,04419  -8,94  0,012  3,121

S = 21,0337   R-Sq = 99,4%   R-Sq(adj) = 98,6%
```
(4.5)

```
Analysis of Variance

Source           DF      SS     MS       F      P
Regression        3  153492  51164  115,65  0,009
Residual Error    2     885    442
  Lack of Fit     1     800    800    9,47  0,200
  Pure Error      1      85     85
Total             5  154377
```
(4.6)

Na první pohled je zřejmé, že analýza rozptylu v (4.4) a (4.6) je naprosto shodná. Jiné je to však ve sloupcích P pro p-hodnoty v (4.3) a (4.5). Zatímco všechny p-hodnoty v (4.3) jsou menší než $0,05$, tak p-hodnota v (4.5) pro teplotu je $0,235$, což je mnohem větší hodnota než $0,05$. Z toho plyne, že proměnná t neovlivňuje přetvárný odpor. Každý technik okamžitě ví, že něco není v pořádku, protože tvářecí teplota má významný vliv na velikost přetvárného odporu. Ovšem podle (4.6) je tento model adekvátní, takže by měl být správný.
Na druhé straně je jednoduše ověřitelné, že samotné hodnoty regresních koeficientů pro původní data i transformovaná data jsou číselně shodné po provedení zpětné transformace.

4.4. Shrnutí

Porovnáním výsledků obou přístupů (pro oba druhy dat) je jasné, že i když oba modely *stejně kvalitně* vysvětlují rozptyl odezvy (vykazují stejné hodnoty koeficientu determinace), *liší se* v hodnocení statistické významnosti vlivu (efektu, koeficientu) teploty tváření na přetvárný odpor oceli. Otázka tedy zní, čím (jakou skutečností) je tato disproporce způsobena, což je obsahem následující kapitoly.

5. Rozbor statistického problému

Pro zjištění příčin vzniklého problému byla definována některá statistická i numerická indikační kritéria.

5.1. Příčiny problému

Na základě informací z literatury, např. [2], [4], lze zjistit, že *nesprávná indikace statistické nevýznamnosti regresoru* (vysvětlující proměnné) je důsledkem tzv. *multikolinearity* (téměř souběžnosti) několika *regresorů* (jejich vzájemné korelovanosti, souvislosti), či jinak řečeno, jejich *neortogonality* (nekolmosti). Tato skutečnost způsobuje (kromě jiného) *velké rozptyly* jednotlivých odhadů regresních koeficientů, díky kterým výsledky *t-testů* indikují nesprávně statistickou významnost těchto koeficientů.
Je překvapivé, že i v tomto případě vychází *koeficient determinace* vysoký a regresní *model* může dobře *popisovat* (avšak už ne dobře a správně *vysvětlovat*) experimentální data. Nesprávně určena *statistická* významnost vychází z *numerického* hlediska tzv. *špatné podmíněnosti* tzv. momentové matice $\mathbf{M} = \mathbf{X}^{\mathrm{T}}\mathbf{X}$ (kde $\mathbf{X}$ je matice plánu).

5.2. Indikační kritéria neortogonality

Pro testování uvedeného jevu existuje a je doporučeno několik kritérií, a to jak numerických, tak i statistických, přičemž *kritéria*:

- *numerická* vycházející ze spektrálního rozkladu momentové a (její inverzní) varianční matice se stanovením tzv. *vlastních čísel, stopy* a *determinantu matic*,
- *statistická* vycházejí z *korelační* matice, používá se u nich statistických pojmů, či pro hodnocení se využívá kvantilů pravděpodobnostních rozdělení.

Další obdobné členění těchto kritérií může být podle skutečnosti, z jaké *matice* regresorů se vychází, tj. zda se vychází z matice:

- plánu $\mathbf{X}$,
- momentové $\mathbf{M} = \mathbf{X}^{\mathrm{T}}\mathbf{X}$,
- variační $\mathbf{V}$, která je inverzní maticí k matici momentové: $\mathbf{V} = \mathbf{M}^{-1}$,
- anebo z normované verze varianční matice $\mathbf{R}$, která je formálně shodná s *korelační maticí* vysvětlujících proměnných (regresorů).

Seznam vybraných a na datech příkladu použitých kritérií vychází z literatury [1], [2], [4]. U níže uvedeného přehledu jsou u kritérií definované *kritické hodnoty*, při jejichž překročení lze uvažovat o existenci *multikolinearity* (*neortogonality*) regresorů:

A. Kritéria numerická

- *skalární součiny* dvojic vysvětlujících proměnných (z matice plánu $\mathbf{X}$): pokud nejsou všechny *nulové*, pak plán není ortogonální,
- *vlastní čísla* momentové matice $\mathbf{M}$: podíl největšího a nejmenšího vlastního čísla matice ($cond(\mathbf{M}) = \frac{\lambda_{max}}{\lambda_{min}}$) větší než 10^6,
- *vlastní čísla*: nejmenší vlastní číslo je blízké nule,
- *determinant* matice $\mathbf{M}$: hodnota větší než 10^6, nebo hodnota menší než 10^{-6},
- *stopa matice* $\mathbf{M}$: $\mathrm{tr}(\mathbf{M})$, tzn. součet diagonálních prvků, je větší než 10^6,
- *stopa matice* $\mathbf{V}$: hodnota $\mathrm{tr}(\mathbf{V})$ větší než 1.0,
- *korelační matice* $\mathbf{R}$: pokud nejsou všechny mimodiagonální prvky (Pearsonovy párové korelační koeficienty) rovné nule, pak plán není nekorelovaný (není nezávislý),
- *číslo podmíněnosti* K matice $\mathbf{R}$: $\mathrm{cond}(\mathbf{R})$ větší než 10^3,
- *determinant matice* $\mathbf{R}$: blízký nule.

B. Kritéria statistická

- *VIF-faktory* (Variance Inflation Factor - inflační faktory rozptylu regresorů [2]): rovné jedné (prediktory nejsou korelované, plán je nekorelovaný a je *ortogonální*), větší než 1, ale menší než 5 (indikace mírné korelovanosti a neortogonality plánu), větší jak 5, ale menší než 10 (významná korelovanost a neortogonalita), větší než 10 (multikolinearita, neortogonalita). V případě $VIF > 1$ jsou odhady regresních koeficientů (číselně) správné, ale nejsou správné jejich *p-hodnoty*. $VIF_j = v_{jj} \sum_{i=1}^{n} (X_{ij} - \overline{X}_j)^2$, $j = 1, \ldots, k$, v_{jj} je diagonální prvek variační matice $\mathbf{V}$,
- *Testy diagonality* výběrové korelační matice $\mathbf{R}$, tzn. testuje se nulová hypotéza, že tato matice je jednotková [2]: testovací veličinou je hodnota $W = -n \cdot \ln(|\mathbf{R}|)$, kde $|\mathbf{R}|$ značí hodnotu determinantu matice $\mathbf{R}$. Pokud $W > \chi^2_\alpha(\frac{k(k-1)}{2})$, pak se jedná o multikolinearitu.

- Častěji je však používaná *korigovaná* (zpřesněná) hodnota *testu diagonality* na počet faktorů (regresorů) podle *Farrara a Glaubera* [1]:

$$FG = -n \cdot \left(1 - \frac{2k+11}{6n}\right) \cdot \ln(|\mathbf{R}|) = \chi^2_\alpha \left(\frac{k(k-1)}{2}\right),$$

- *F-test* regresorů způsobujících multikolinearitu: Pokud platí

$$FD_j = \frac{n-k}{k-1}(VIF_j - 1) > F_\alpha(k-1, n-k),$$

kde $F_\alpha(k-1, n-k)$ je hodnota kvantilu F-rozdělení.

6. Způsoby řešení problému

Podle informací z literatury byly dále sestaveny návrhy způsobů řešení problému. Pro ověření navrhovaných opatření byla uskutečněna aplikace kritérií a způsobů řešení na data příkladu s hodnocením jejich detekční schopnosti a účinnosti.
V literatuře [2], [4], [5], je doporučeno v případě výskytu multikolinearity použít *lineární transformaci* dat proměnných z na proměnné x několika následujícími způsoby ($i = 1, 2, \ldots, n$, kde n je celkový počet pokusů):

- *centrování hodnot proměnných*:
  $x_i = z_i - \overline{z}$, kde $\overline{z}$ je aritmetický průměr proměnné z,
- *normování hodnot proměnných*:
  $x_i = \dfrac{z_i - \overline{z}}{s_z}$, kde s_z je směrodatná odchylka proměnné z,
- *DOE transformace* (kódování) - viz předchozí vztah (3.1).

Všechny uvedené způsoby lineární transformace dat vedou k *nulové střední hodnotě* a *symetrii* transformovaných dat s příznivým dopadem na zajištění *ortogonality* a *nekorelovanosti* plánu s takovýmito proměnnými. Nicméně, zde je třeba obecně konstatovat, že multikolinearitu regresorů (faktorů) nelze vždy uvedenými způsoby odstranit. Je to tehdy, když jsou faktory mezi sebou z principu závislé (např. délka a objem tělesa), při používání vyšších mocnin (které jsou navzájem závislé) proměnných, apod.
Pro analyzovaná data se rozmezí jejich hodnot pro původní i transformované proměnné pohybovalo v následujících intervalech:

data	φ	t	$\varphi \cdot t$
původní	$\langle 0,0; 1,4 \rangle$	$\langle 20; 700 \rangle$	$\langle 0; 980 \rangle$
centrovaná	$\langle -0,7; 0,7 \rangle$	$\langle -340; 340 \rangle$	$\langle -238; 238 \rangle$
normovaná	$\langle -1,225; 1,225 \rangle$	$\langle -1,225; 1,225 \rangle$	$\langle -1,5; 1,5 \rangle$
DOE transformovaná	$\langle -1; 1 \rangle$	$\langle -1; 1 \rangle$	$\langle -1; 1 \rangle$

Tabulka 1. Intervaly hodnot.

Prakticky to znamená, že jak *normování* tak i *DOE transformace* „stlačují" hodnoty do intervalu $\langle -2; 2\rangle$, což je doporučené rozmezí s ohledem na zaručení numerické stability výpočtů a omezení kumulačních chyb. Uvedené rozmezí poskytuje nejmenší odchylky od správného řešení, které bychom obdrželi, kdyby počítač uměl zapsat každé číslo na nekonečně dlouhou mantisu.

6.1. Ověření kritérií a způsobů řešení

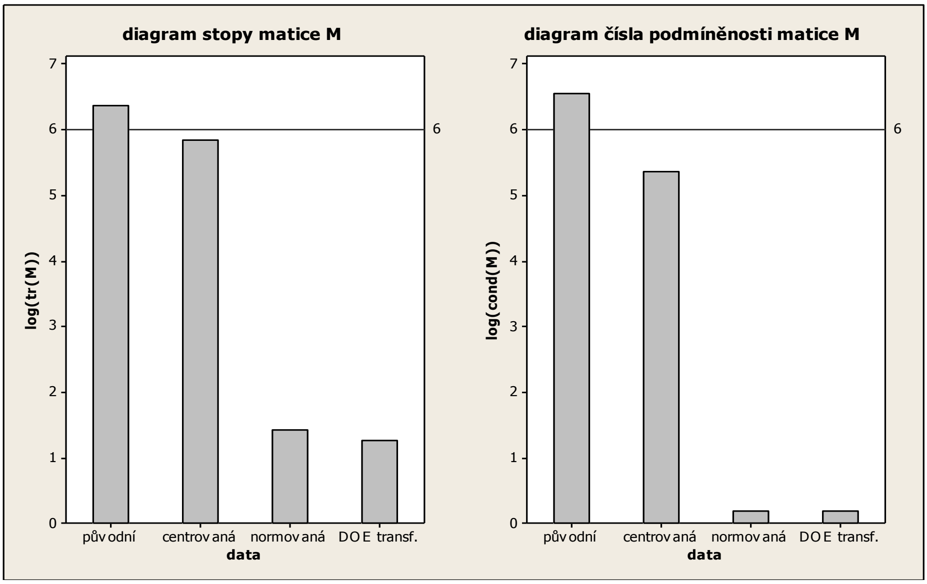
Po transformaci původních dat a výpočtu kritérií v prostředí *Minitab, Statistica* a *Matlab* obdržíme naprosto stejné výsledky, které jsou přehledně uspořádány v tab. 2 (kde jsou tučným písmem označeny hodnoty kritérií, které překročily kritické hodnoty a signalizují narušení *ortogonality* a *nekorelovanosti* plánu). Pro test Farrara a Glaubera je kritická hodnota $\chi^2_{0,05}(3) = 7,815$ a pro F-test je $F_{0,05}(2,3) = 9,552$.

kritéria		původní	centrovaná	mormovaná	DOE transf.
$\det(\mathbf{M})$		$\mathbf{1,232 \cdot 10^{12}}$	$\mathbf{1,232 \cdot 10^{12}}$	$1,944 \cdot 10^3$	384
$\mathrm{tr}(\mathbf{M})$		$\mathbf{2,328 \cdot 10^6}$	$6,89 \cdot 10^5$	27	18
$\mathrm{tr}(\mathbf{V})$		$\mathbf{2,059}$	$0,677$	$0,611$	$0,917$
$\mathrm{cond}(\mathbf{M})$		$\mathbf{3,641 \cdot 10^6}$	$2,359 \cdot 10^5$	$1,5$	$1,5$
$\mathrm{cond}(\mathbf{R})$		$10,371$	$10,371$	$10,371$	$10,371$
$\det(\mathbf{R})$		$0,321$	$0,321$	$0,321$	$0,321$
FG		$3,6$	$3,6$	$3,6$	$3,6$
VIF	φ	$\mathbf{2,121}$	1	1	1
	t	$\mathbf{2,000}$	1	1	1
	$\varphi \cdot t$	$\mathbf{3,121}$	1	1	1
FD	φ	$1,682$	0	0	0
	t	$1,500$	0	0	0
	$\varphi \cdot t$	$3,182$	0	0	0

Tabulka 2. Tabulka kritérií multikolinearity a korelovanosti dat.

Hodnoty a porovnání ukazatelů *stopa matice* **M**, jakož i číslo podmíněnosti matice **M** jsou pro původní a transformovaná data graficky (v logaritmickém měřítku) znázorněny na obr. 2:

Z obr. 2 je zřejmé, že *centrování* částečně (a snad i dostatečně) snižuje hodnoty stopy matice **M** a číslo podmíněnosti. Výrazné (téměř o 5 řádů) je však toto snížení u dalších způsobů transformace, tj. u *normování* a *DOE transformace*.



Obrázek 2. Sloupcový graf stopy a čísla podmíněnosti matice **M** pro analyzovaná data.

V tab. 3 můžeme sledovat, jak se výrazně mění velikost vlastních čísel, jestliže použijeme na původní data různé transformace. A nakonec je vidět to nejdůležitější - jak jednotlivé transformace ovlivňují významnost odhadů jednotlivých regresních parametrů.

		původní	centrovaná	normovaná	DOE transf.
vlastní čísla matice **M**	λ_1	$0,599$	6	6	6
	λ_2	$3,699$	$1,96$	6	4
	λ_3	$2.928 \cdot 10^5$	$4.624 \cdot 10^5$	6	4
	λ_4	$2.035 \cdot 10^6$	$2.266 \cdot 10^5$	9	4
p-hodnoty regresních koef.	b_0	$0,0028$	$0,0031$	$0,0000$	$0,0003$
	b_1	$0,0047$	$0,0071$	$0,0022$	$0,0016$
	b_2	$\mathbf{0,2348}$	$0,0077$	$0,0204$	$0,0156$
	b_3	0.0123	0.0123	0.0123	0.0123

Tabulka 3. Vlastní čísla matice **M** a p-hodnoty regresních koeficientů modelu (3.2).

6.2. Hodnocení kritérií a způsobů řešení

Z předchozích výsledků je jasné, že *neortogonalitu* (nekolmost) a *korelovanost* (*multikolinearitu* a *špatnou podmíněnost*) matice plánu s *původními* proměnnými:

- signalizovala (pouze) tři kritéria: *inflační faktory VIF* (větší než 1), *stopa matice* **M**, tj. tr(**M**) a *číslo podmíněnosti matic* **M**, které překročily kritickou mez 10^6. Je však třeba říci, že *malé* číslo podmíněnosti sice zaručuje správný numerický výsledek, avšak *velké* číslo pouze *může* (ale nemusí) způsobit numerické problémy tzn. malé číslo podmíněnosti je pouze *postačující*, ale nikoliv nutná podmínka k numerické stabilitě a statistické správnosti výsledku výpočtu,
- nejvýrazněji (pro techniky a technology) signalizovala *nesprávná* a nevěrohodná *statistická nevýznamnost* vlivu *teploty tváření* na přetvárný odpor. Jako *způsoby řešení* uvedeného problému se rýsují následující:
- použití metody *DOE*, která zajišťuje *ortogonalitu* a dobrou *numerickou podmíněnost* (malá čísla podmíněnosti díky transformovaným, normovaným, tzv. kódovaným datům) plánu experimentu s dopadem na *statistickou korektnost* a *spolehlivost* závěrů,
- při použití *regresní analýzy* (např. díky nemožnosti použít metodu DOE s příslušným SW) pro objektivní a spolehlivé rozpoznání uvedených nectností, a tím i možnosti výskytu numericko-statistické nekorektnosti indikace nevýznamnosti (některých) regresních koeficientů, je nutné a vhodné *porovnat* tuto významnost (pomocí t-statistik či p-hodnot) pro *modely* s *původními* a *transformovanými* daty. Výraznější nesouhlas (významnost versus nevýznamnost) pak signalizuje špatnou podmíněnost plánu. Hodnocení statistické významnosti koeficientů modelu je korektní pouze u modelu s *transformovanými* proměnnými, přičemž číselné hodnoty samotných regresních koeficientů jsou správné i pro model s původními proměnnými.

Obecně je zde viditelné, že doporučené způsoby *lineární transformace* původních dat vedou k odstranění multikolinearity díky ortogonalitě a nekorelovanosti plánů s *transformovanými* proměnnými a také k numericky i statisticky věrohodným výsledkům.

7. Statistický důsledek neortogonality

Při operacích se špatně podmíněnými maticemi rostou numerické chyby díky nutnému zaokrouhlení („ořezání“) čísel při omezené délce slova počítače. To se zvláště projeví při *inverzi momentové matice* **M**, ze které následně vychází výpočet rozptylů i směrodatných odchylek regresních koeficientů a příslušné t-testy jejich významnosti.
Z výše uvedeného vztahu (4.1) pro výpočet rozptylu regresních koeficientů je vidět *přímý dopad nepřesně určené inverzní matice* (díky špatné podmíněnosti momentové matice vycházející z matice plánu) na následně nepřesně stanovený (většinou nadhodnocený) *rozptyl regresního koeficientu.*
Špatná podmíněnost momentové matice *může* být způsobena až o $2,5$ řádu *rozdílnými hodnotami* (úrovněmi) základních faktorů v matici plánu (průměrná hodnota

$\varphi = 0,7$, střední hodnota $t = 360\,^\circ\mathrm{C}$). Právě díky *nadhodnocení* odhadu směrodatné odchylky u regresního koeficientu teploty tváření t došlo k *podhodnocení* příslušné t-statistiky a nepřekročení kritické hodnoty s dopadem na nesprávnou signalizaci statistické nevýznamnosti tohoto vlivu (faktoru).

8. Závěr

Na základě matematicko-statistického rozboru jednoduchého „technologického“ příkladu s lineárním modelem a interakcí faktorů, kde se pro *původní* data nečekaně vyskytl numericko-statistický problém nesprávné indikace *statistické nevýznamnosti* jednoho regresního koeficientu technologicky i fyzikálně vlivného faktoru, a to *teploty tváření*, lze vyslovit *závěry* a *doporučení*:

- Díky ortogonalitě (a malému číslu podmíněnosti) plánu je možné *v jednom kroku* (jednorázově) správně (korektně) stanovit statisticky významné a nevýznamné faktory. Užití *regresní analýzy* pro *původní* (netransformované) proměnné může (díky *neortogonalitě* plánu) poskytnout statisticky i fyzikálně, či technologicky *nesprávné* hodnocení nevýznamnosti (některých) regresních koeficientů (v analyzovaném příkladu to byl vliv, tj. regresní koeficient *teploty tváření*), přičemž však jejich hodnoty jsou numericky správné. Proto pouze regresní analýza s *transformovanými* proměnnými poskytne i statisticky *správné* výsledky,
- pokud je experiment sice *aktivní*, plánovaný, ale *ne podle metodiky DOE*, pak jeho vyhodnocení pomocí *regresní analýzy* poskytne správné hodnoty regresních koeficientů faktorů, čísla podmíněnosti jsou (většinou) malá, ale díky neortogonalitě plánu a určité multikolinearitě faktorů dochází ke statistické nekorektnosti jejich hodnocení, tj. k nesprávné indikaci statistické nevýznamnosti fyzikálně či technologicky vlivných faktorů, a to jak pro jejich *původní*, tak i pro *transformované* hodnoty. Tato skutečnost znamená, že není k dispozici žádná statisticky korektní metoda vyhodnocení experimentu, ale pouze metoda numericky korektní, kterou je *regresní analýza* (analýzu metodou DOE nelze v tomto případě použít),
- tzv. *pasivní* experiment je nejhorší variantou pro analýzu: plán je (zásadně) neortogonální, faktory jsou kolineární, podmíněnost je velká, a tak lze (asi výrazněji a častěji pro větší rozsah dat a větší počet faktorů) očekávat při analýze pomocí *regresní analýzy* jak numerickou nestabilitu (díky numerickým chybám a jejich kumulaci), tj. *numericky* nepřesné a až *nesprávné* hodnoty (některých) regresních koeficientů, tak i jejich nesprávně indikovanou *statistickou nevýznamnost*, což znamená statistickou nekorektnost výsledků,
- metoda *DOE* poskytuje *spolehlivé* a *efektivní* (optimální) přístupy k návrhu a analýze přínosných experimentů, jako i *numericky* a *statisticky* korektní výsledky. Proto je velice potřebné, aby vědci, výzkumníci, technici i technologové, ale i studenti tuto metodiku znali a zásadně ji používali,
- protože všechny tři nejpoužívanější SW *Matlab, Minitab, Statistica* poskytly shodně špatné výsledky při stanovení spolehlivosti odhadu regresních parametrů pro původní hodnoty, doporučuje se opatrnost při aplikaci výsledků regrese bez transformace.

Reference

[1] V. Garaj, I. Šujan: *Ekonometria*, ALFA/SNTL, Bratislava, 1980.
[2] P. Hebák, J. Hustopecký: *Vícerozměrné statistické metody s aplikacemi*, SNTL/ALFA, Praha, 1987.
[3] B. Maroš, T. Trávníček: *Plánování experimentu*, Sborník 5th International Conference *Aplimat* 2006, Bratislava, Ústav matematiky FSI STU Bratislava, 2006.
[4] M. Meloun, J. Militký: *Statistické zpracování experimentálních dat*, PLUS, Praha, 1994.
[5] J. Tošenovský, D. Noskievičová: *Statistické metody pro zlepšování jakosti*, Montanex, Ostrava, 2000.
[6] K. Zvára: *Regrese*, 1. vyd., Matfyzpress MFF UK Praha, Praha, 2008.

Bohumil Maroš, Ústav matematiky, Fakulta strojního inženýrství, Vysoké učení technické v Brně, Technická 2, 616 69 Brno, Česká republika,
e-mail: maros@fme.vutbr.cz