

Martin Otava

Základní principy navrhování experimentů

*Pokroky matematiky, fyziky a astronomie*, Vol. 63 (2018), No. 3, 196–211

Persistent URL: <http://dml.cz/dmlcz/147444>

## Terms of use:

© Jednota českých matematiků a fyziků, 2018

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library*  
<http://dml.cz>

# Základní principy navrhování experimentů

*Martin Otava*

*Abstrakt.* Navrhování experimentů je disciplína umožňující zvolit optimální průběh experimentu v závislosti na zkoumané hypotéze. V tomto článku ukážeme, jak proces navrhování experimentu probíhá v praxi a na jakých metodách a principech je založen. Vzhledem k širokému záběru však zůstaneme jen u přehledových informací pro jednotlivé modely a pokročilejší postupy. Jako příklad poslouží problém z farmaceutické výroby tablet.

## 1. Úvod

Sousloví „navrhování experimentů“ může být interpretováno jednak jako aplikování sady nástrojů za účelem získání maxima informací s vynaložením minimálního úsilí prostřednictvím série vhodně navržených experimentů, jednak jako disciplína, která se vývojem právě takových nástrojů zabývá. V tomto článku se soustředíme na aplikovanou část tohoto problému: jak vybrat optimální typ experimentu pro ověření jisté hypotézy. Nejčastěji je optimalita vyjádřena minimalizací počtu nutných měření v rámci jednoho experimentu, tj. zjišťujeme, kolik měření musíme provést, aby kvalita odhadů zůstala na jisté „rozumné“ úrovni. Je-li předem dostupná informace o přibližném rozptylu, lze optimalitu vztáhnout i k síle testu, tj. schopnosti identifikovat efekt jisté velikosti jako statisticky významný.

Navrhováním experimentů máme obvykle na mysli situaci, kdy hodnotu veličin, jejichž vliv na odezvu plánujeme zkoumat, můžeme jistým způsobem nastavit. Nastavení pak často opravdu znamená, že máme jistotu o přesné hodnotě dané veličiny. Typickým příkladem jsou zemědělské či průmyslové experimenty, kdy zkoumáme vliv hnojiv, typu půdy, osiva, typu součástí či nastavení na výrobním stroji. Odezvou pak může být výnos či kvalita produktu. V jistých případech se může skutečná hodnota realizovaná během experimentu lišit od hodnoty nastavené. V případě typu součástky máme jistotu ohledně hodnoty nastavení, ale kvalita produktu může být měřena velice nepřesně, a tedy můžeme mít nejistotu nejen v měření odezvy, ale i nezávisle proměnných veličin. Tímto případem se však nadále nebudeme zabývat a budeme předpokládat, že jsme schopni provést nastavení velice přesně. Důležitou vlastností návrhu je pak vedle samotného počtu měření i nastavení zkoumaných veličin pro každé jednotlivé pozorování.

Optimální situace pro spolupráci se statistikem je ta, kdy je zahrnut do diskuze o experimentu od začátku, podílí se na jeho navrhování a následná analýza sebraných dat je pak provedena pomocí metod kompatibilních s návrhem tohoto experimentu. Bohužel velice často není takový postup z nejrůznějších příčin prakticky možný a spolupráce se statistikem začíná až v momentě, kdy jsou data sebrána a je potřeba je vyhodnotit. Z tohoto důvodu je velké množství literatury věnováno navrhování experimentů pro odborníky v nejrůznějších oborech, kteří nutně nemusí disponovat sta-

---

Mgr. MARTIN OTAVA, Ph.D., Janssen-Cilag s. r. o., Walterovo náměstí 329/1, 158 00 Praha 5, e-mail: [motava@its.jnj.com](mailto:motava@its.jnj.com)

tistickým vzděláním. Ti se pak soustředí pouze na některé aspekty problematiky dle specifík příslušných oborů.

V tomto článku bychom se chtěli pokusit žádný ze základních aspektů neignorovat. Vzhledem k širokému záběru však zůstaneme jen u přehledových informací pro jednotlivé modely a pokročilejší postupy.

## 2. Základní pojmy

### 2.1. Shrnutí nutných statistických základů

Obdobně jako v článku [8] budeme teorii ilustrovat na příkladu náhodné veličiny  $X$ , která reprezentuje množství účinné látky v hypotetické 550 mg tabletě. Obsah udávaný výrobcem budiž 500 mg a veličina  $X$  pak vyjadřuje skutečné množství pro jednotlivé tablety. Budeme uvažovat normální rozdělení se střední hodnotou  $\mu = \mathbb{E}(X)$  a rozptylem  $\sigma^2 = \text{var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2$ . Hustota veličiny  $X$  tedy bude mít tvar

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\sigma^2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right). \quad (1)$$

Známe-li parametry  $\mu$  a  $\sigma^2$ , můžeme pak spočítat pravděpodobnost, že hodnota účinné látky bude v intervalu  $(a, b)$ , ze vzorce

$$P(a < X < b) = \int_a^b f(x) dx. \quad (2)$$

Připomeňme si, že neznámé parametry  $\mu$  a  $\sigma^2$  je nutné odhadnout z datového souboru. V případě jednoduchého vzorku nezávislých pozorování  $x_1, \dots, x_N$  veličiny  $X$  bychom jako odhad  $\mu$  použili aritmetický průměr  $\hat{\mu} = \bar{X} = \frac{1}{N} \sum_{i=1}^N x_i$  a jako odhad  $\sigma^2$

$$\text{číslo } s^2 = \hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu})^2.$$

Velice často však nastane situace, kdy se populace veličiny  $X$ , v našem případě tablety, skládá z několika různých subpopulací, kdy každá má jinou střední hodnotu. Může jít například o tablety vyrobené na jiném přístroji či v jiné továrně, což může způsobit drobnou odchylku ve střední hodnotě a případně i v rozptylu (tuto možnost nebudeme nadále pro zjednodušení uvažovat). V takovém případě pak můžeme modelovat střední hodnotu veličiny  $X$  jako funkci několika nezávisle proměnných veličin, například pomocí lineární regrese (obdobně jako v článku [2]).

Přesné definice jednotlivých pojmů lze najít v základní učebnici matematické statistiky [1]. Pro podrobnější informace o regresních metodách doporučujeme publikaci [10].

### 2.2. Názvosloví

Názvosloví navrhování experimentů se vyvíjelo jistým způsobem paralelně k názvosloví matematické i aplikované statistiky. Snadno by tak mohlo dojít ke zmatení čtenáře, pokud by pojmy nebyly jasně vymezeny. Vzhledem k tomu, že autor je vzděláním

i praxi statistikem, použité pojmy vychází spíše ze statistického názvosloví, ale některé jsou přejaty i z názvosloví technického.

- *Odezva*: Závisle proměnná náhodná veličina, která je naším hlavním objektem zájmu. Cílem je určit, jak jiné nezávisle proměnné veličiny (faktory) ovlivňují odezvu.
- *Faktor*: Nezávisle proměnná veličina, jejíž vliv na odezvu chceme vyhodnotit. V některém statistickém softwaru je faktorem míněna pouze nominální veličina, ale v našem případě bude faktor reprezentovat jakoukoli veličinu včetně číselných měření.
- *Parametr*: V technické literatuře je někdy parametrem procesu míněn faktor. V našem případě budeme používat parametr ve smyslu ryze statistickém, tj. jako neznámou konstantu, jejíž hodnotu se snažíme odhadnout.
- *Pozorování*: Pozorováním budeme mít na mysli jednotlivý pokus (jeden krok experimentu), tj. jedno výstupní měření odezvy, které odpovídá jednomu nastavení faktorů. V praxi může výsledné číslo být složeno z více pozorování ve statistickém slova smyslu (například průměr tří měření pro zvýšení přesnosti). Klíčová je vlastnost, že jedno pozorování odpovídá jedné sadě nastavení faktorů a realizaci nějakého výrobního procesu (či osetí pole). Každé pozorování je tedy spojeno s náklady a časem, stejně jako přechod mezi pozorováními kvůli nutné změně nastavení faktorů.

### 2.3. Praktický příklad problému

Různé aspekty navrhování experimentů budeme demonstrovat na hypotetickém příkladu z výroby tablet. Předpokládejme, že během vývoje výrobního procesu chceme zjistit, jak může nastavení jednoho z přístrojů, mixéru, ovlivnit obsah účinné látky v tabletě. Uvažujme následující nastavitelné faktory:

- Existují různé typy mixéru od stejného výrobce o kapacitě 25, 50 a 75 litrů.
- Každý z mixérů může být vybaven lopatkami s různým počtem listů, od tří do pěti.
- Lopatky mixéru mohou být nainstalovány pod různým úhlem, od třiceti do šedesáti stupňů.

Soustředíme se na odhad vlivu těchto faktorů na střední hodnotu účinné látky (odezva) a budeme předpokládat normalitu reziduí.

V následující sekci popíšeme jednotlivé metody a poté demonstrováme jejich aplikaci na tomto příkladu.

## 3. Obecné principy

Ve stručnosti si představme několik teoretických konceptů, které pak v další sekci demonstrováme na konkrétním příkladu.

### 3.1. Modelování střední hodnoty

Jak bylo vysvětleno výše, cílem navrhovaného experimentu je charakterizace funkčního vztahu mezi střední hodnotou odezvy  $X$  a skupinou faktorů reprezentujících jistá nastavení systému, která můžeme ovlivnit. V našem případě půjde o nastavení přístrojů ve výrobním procesu. Otázkou, která nastavení zahrnout do uvažovaného experimentu,

většinou přenecháme odborníkům z příslušné vědní disciplíny. Statistik je obvykle velmi nápomocen v tom, jaký funkční vztah uvažovat a jakým způsobem nastavení z praktického přehledu převést na náhodné veličiny ve statistickém modelu

$$X = g(A_1, \dots, A_M) + \varepsilon, \quad (3)$$

kde  $g$  je lineární funkce (lineární v parametrech),  $A_1, \dots, A_M$  jsou náhodné veličiny reprezentující faktory a  $\varepsilon \sim N(0, \sigma^2)$  je reziduální variabilita.

### 3.2. Modelování rozptylu

Veškeré další faktory, které by mohly mít nějaký dopad na střední hodnotu odezvy  $X$ , ale nebyly zahrnuty do modelu (protože o nich nevíme nebo je nedokážeme měřit či nedokážeme nastavit), zvyšují reziduální variabilitu v modelu. Vyšší variabilita vede ke snížené přesnosti odhadů, přesnosti predikce budoucích pozorování a nižší síle testů. Proto je důležité možné zdroje variability identifikovat a do modelu je zahrnout.

Typickým příkladem je výrobní den. V různé dny mohou být ve výrobní hale různé klimatické podmínky, teplota a vlhkost vzduchu, může být přítomen jiný operátor (neboť některá nastavení mohou záviset na lidském faktoru), materiál může pocházet z jiného zdroje a konečně tablety musí být změřeny v laboratoři, kde může docházet k obdobným variacím v postupu a podmínkách. V ideálním případě bychom mohli všechny tyto faktory sledovat a nevysvětlená variabilita mezi dny by se blížila k nule, ale to není v mnoha praktických případech možné. Pokud pak není den zahrnut do modelu, mohlo by se stát, že skupina tablet v první den má vyšší střední hodnotu odezvy než v den druhý. Vliv na odhad střední hodnoty může být kompenzován randomizací, ale vyšší variability reziduí v modelu se tím nezabavíme. Jednou z možností je zahrnout den do modelu jako „náhodný efekt“ a od reziduální variability tak oddělit variabilitu kvůli různému výrobnímu dni:

$$X = g(A_1, \dots, A_M) + d + \varepsilon_E, \quad (4)$$

kde  $d \sim N(0, \sigma_{den}^2)$  reprezentuje variabilitu kvůli různým výrobním dnům a  $\varepsilon_E \sim N(0, \sigma_E^2)$ . Reziduální variabilita v tomto modelu je menší nežli původní reziduální variabilita při ignorování efektu dnů, tedy  $\sigma_E^2 \leq \sigma^2$ .

Dalším zajímavým příkladem využití náhodných efektů je situace, kdy důležitou roli hraje osoba analytika v laboratoři. Příprava vzorků může být velmi subjektivní záležitost, i když typicky jsou odchylky velice malé. Pokud máme několik konkrétních analytiků, kteří se budou na našem experimentu podílet a můžeme jim přiřadit různé experimenty dle našeho uvážení, můžeme zahrnout analytika jako běžnou veličinu do modelu střední hodnoty a navrhovacího procesu, neboť můžeme „nastavit“, který analytik bude analyzovat který vzorek. Zejména v případě dlouhodobých experimentů však není praktické předpokládat, že budeme mít úplnou kontrolu nad tím, kdo bude který vzorek měřit. Důvodem může být např. nemoc analytika či jeho absence z jiného důvodu. Často tedy máme pouze informaci, že například první tři pozorování budou měřena jedním analytikem, další tři jiným a tak dále. V tom případě je chování veličiny reprezentující analytika obdobné jako u různých výrobních dnů a použijeme náhodný efekt.

Dodejme, že máme-li jen málo úrovní příslušného faktoru (jen dva dni, jen dva analytici), je lepší zahrnout jeho efekt do střední hodnoty, neboť odhadování rozptylu

na základě nízkého počtu pozorování je velmi nespolehlivé. Při vyšším počtu úrovní (pět různých dní) je lepší modelovat vliv jako náhodný efekt. Nejenže to lépe odpovídá tomu, o co se zajímáme v našem experimentu, ale navíc přidání dalších dnů do modelů vyřešíme jen jedním parametrem (pro rozptyl) namísto několika pro každou úroveň.

### 3.3. Blokový design

Počet pozorování ovlivněných jedním náhodným efektem (tj. měření ve stejný den, stejným analytikem, pocházející ze stejného zdroje materiálu) může být kontrolovatelný z naší strany nebo může být do jisté míry náhodný (například při obtížně proveditelném měření nemusí být jisté, kolik pozorování denně zvládneme provést). Ve druhém případě nemůžeme přítomnost náhodných efektů zahrnout nijak systematicky do návrhu experimentu, neboť strukturu předem neznáme. Můžeme pouze provést navýšení minimálního počtu pozorování kvůli zvýšení počtu parametrů při následné analýze dat, která bude náhodné efekty brát v úvahu.

V případě, že počet pozorování ovlivněných jedním náhodným efektem je známý (či ho dokonce můžeme určit), je možné zahrnout přítomnost náhodných efektů do návrhu experimentu. Tato technika se nazývá blokový design a cílem je uspořádat různá nastavení našich plánovaných pozorování tak, abychom minimalizovali dopad případného vychýlení kvůli náhodným efektům na odhady vlivu faktorů na odezvu. V návrhu vezmeme v úvahu „náhodné bloky“ (tj. skupiny pozorování měřené stejným analytikem, ve stejný den atd.) o velikosti odpovídající počtu pozorování ovlivněných stejným náhodným efektem. Úplné bloky pak označují situaci, kdy prakticky zopakujeme celý původně plánovaný experiment uvnitř každého bloku. Tedy pokud jsme původně plánovali deset pozorování a máme tři analytiky, každého necháme provést těchto deset pozorování a tím jejich případný vliv na odhady parametrů zcela eliminujeme. Úplné bloky však vyžadují jednak zvýšení počtu pozorování, jednak je často nemožné je použít z praktických důvodů (například lze provést pouze tři pozorování denně). V takovém případě použijeme bloky o menší velikosti a uspořádáme do nich pozorování tak, aby se případný vliv bloků na odhadované parametry minimalizoval a experiment byl co nejvíce vyvážený. Z hlediska modelu pak k těmto blokům budeme přistupovat jako k náhodným efektům (viz předchozí odstavec).

### 3.4. Randomizace

V jistých případech nelze uvažovaný faktor do modelu žádným způsobem zahrnout, neboť ho nemůžeme měřit ani přesně stanovit, kdy se mění (na rozdíl od jednoznačného oddělení různých dní či analytiků). Například můžeme vědět, že jistá charakteristika vstupního materiálu může mít dopad na hodnotu veličiny  $X$ , ale nejsme schopni ji technicky změřit či se může měnit dynamicky během výroby. Podobně proudění vzduchu či lokální vlhkost mohou být obtížně měřitelné faktory. Jejich vliv se započte do reziduální variability, ale v jistých případech může dokonce vychýlit některý z našich odhadů, pokud by hodnota těchto veličin korelovala s hodnotami nastavení některého z faktorů. Například předpokládejme, že pokaždé, když bychom vybrali pro faktor  $A_1$  nastavení  $a_1$ , měli bychom materiál vysoké kvality a při nastavení  $a_2$  materiál nízké kvality. Vliv kvality by pak nebylo možné oddělit od vlivu faktoru  $A_1$  a náš odhad tohoto vlivu by byl vychýlen. Randomizace pořadí prováděných pozorování umožňuje

částečně bránit takovým vychýlením tím, že pozorování jsou prováděna v náhodném pořadí, místo toho, abychom nejdříve provedli všechna pozorování pro  $A_1 = a_1$  a poté všechna pozorování pro  $A_1 = a_2$ .

### 3.5. Split-plot

Split-plot design jsou metodologií používanou v případech, kdy je velice obtížné, drahé či časově náročné měnit nastavení jednoho z faktorů, o jehož vliv se zajímáme. Například můžeme mít čtyři faktory reprezentující různá nastavení přístroje ve výrobní lince a pátý reprezentující dva různé typy tohoto přístroje. Zatímco změnu nastavení lze provést pomocí ovladačů, změna pátého faktoru vyžaduje odpojení jednoho přístroje z linky a zapojení druhého. V takovém případě by klasická randomizace pořadí pozorování vedla k velkým časovým prodlevám v realizaci experimentu kvůli nutnosti časté výměny přístrojů. Řešením je vytvoření návrhu, ve kterém se obtížně měnitelná veličina vyskytuje v blocích o stejném nastavení (podobně jako při blokovém designu) a experiment je uspořádán tak, že můžeme odhadnout vliv tohoto faktoru na odezvu. Narozdíl od blokových designů v tomto případě bloky reprezentujeme jako fixní efekty v modelu střední hodnoty (ne jako náhodné efekty), neboť chceme odhadnout vliv blokového faktoru na odezvu.

### 3.6. Odhadnutelnost modelu

Jak bylo uvedeno výše, navrhování experimentů se primárně zabývá dvěma otázkami: Kolik pozorování je potřeba a v jakém nastavení zkoumaných faktorů by měla být provedena. První otázka úzce souvisí s modelem, který bude použit k vyhodnocení experimentu. V zásadě totiž platí, že pro každý parametr v modelu je třeba jedno pozorování, abychom mohli odhadnout všechny parametry. Samozřejmě v tomto případě mluvíme o minimálním počtu pozorování, což nutně neznamená, že jde o doporučený počet.

Uvažujme jednoduchý příklad, kdy chceme odhadnout pouze střední hodnotu z normálního rozdělení se známým rozptylem. Odhadujeme tedy jeden parametr a minimální počet pozorování k obdržení odhadu je tedy jedno jediné pozorování. Přirozeně však odhadování průměrné hodnoty populace na základě jednoho pozorování nebude příliš dobrý nápad, pokud není rozptyl prakticky nulový.

Rozhodnutí, kolik pozorování nad minimální hodnotu bychom měli použít, lze založit na síle testu nebo délce konfidenčního intervalu pro příslušný parametr. Je však třeba mít představu o rozptylu v experimentu. V případě, kdy tato informace není k dispozici, lze uvažovat jako minimální počet pozorování takový, který umožní odhad rozptylu s „rozumnou přesností“. Ta je často definována vzhledem k počtu stupňů volnosti, které zjednodušeně řečeno reprezentují rozdíl mezi informací obsaženou v datech a informací nutnou pro odhad parametrů. Typicky použijeme přinejmenším o pět pozorování více, než je minimální nutný počet pro odhadnutelnost všech parametrů v modelu, ale toto doporučení je silně zjednodušené a může se lišit dle konkrétní situace.

### 3.7. Síla testu

Navrhovaný experiment již nutně bere v úvahu, jaký statistický test bude použit k jeho vyhodnocení. Ať už bude test jakýkoli, bude testovat nějaký parametr vzhledem k fixní hodnotě (typicky nule) a rozhodovat, zda případná nenulová hodnota byla napozoro-

vána jen náhodou či se statisticky významnou jistotou. Statistická významnost je prakticky funkce velikosti odhadu parametru a nejistoty, kterou v tomto odhadu máme. Síla testu nám říká, jaká je pravděpodobnost, že v náhodně vygenerovaném vzorku o jistém počtu pozorování napozorujeme pro danou hypotetickou hodnotu parametru statisticky významný výsledek, při jisté hladině významnosti a nejistotě (reprezentované obvykle rozptylem). Při dané významnosti, velikosti souboru a rozptylu pak obvykle řešíme otázku, jaká má být velikost souboru, abychom byli schopni jistý „zajímavý“ efekt prohlásit za významný. „Zajímavým“ přitom rozumíme praktické, nikoli statistické hledisko. Čím menší absolutní velikost parametru chceme být schopni prohlásit za významnou, tím větší soubor potřebujeme, abychom dosáhli dané síly testu.

V praxi tedy odborník stanoví, jaký efekt je pro něj prakticky zajímavý do té míry, že kdyby byl napozorován, chtěl by mít jistotu (na jisté hladině významnosti), že experiment výsledek vyhodnotí jako statisticky významný. Na základě přibližného předpokladu o rozptylu a požadované síle testu pak statistik určí, kolik pozorování je potřeba.

Namísto síly testu může být jako kritérium použita délka konfidenčních intervalů. Obě kritéria spolu úzce souvisí, ovšem délka intervalu může být definována ve smyslu vyžadované přesnosti odhadu, takže není nutné určovat, co je „prakticky zajímavý“ efekt veličiny.

V obou případech ovšem potřebujeme mít alespoň přibližnou znalost rozptylu. Pokud není k dispozici jasná odborná znalost, menší pilotní experiment může napovědět, jaké hodnoty dávají smysl.

### 3.8. Shrnutí

Blokový design, randomizace, fixní veličiny v modelování střední hodnoty a náhodné efekty v modelu jsou jen různými způsoby, jak řešit stejnou situaci, kdy odezvu ovlivňuje několik potenciálních faktorů, o jejichž efekt se v zásadě nezajímáme a chceme jej pouze oddělit od efektu faktorů, které potřebujeme odhadnout. Randomizace cílí na redukci vlivu faktorů, jejichž hodnotu neznáme, zatímco náhodné efekty typicky použijeme v situaci, kdy hodnotu faktoru známe. Pokud ji nejsme schopni kontrolovat, použijeme náhodné efekty jen při analýze, zatímco v opačném případě využijeme principu blokových designů. Metoda split-plotu je velmi podobná, ale využijeme ji v případě, kdy se explicitně zajímáme o vliv faktoru, který tvoří bloky, na odezvu.

Odhadnutelnost modelu a posuzování síly testu jsou dva doplňující se přístupy k určení nutného počtu pozorování. Zatímco princip odhadnutelnosti je použitelný v každém případě, k určení síly testu je nutné vzít v úvahu informaci o rozptylu a také určit, jaký efekt je z praktického pohledu zajímavé detekovat pomocí statistických testů.

## 4. Navrhování experimentů v praxi

Připomeňme si náš příklad. Chceme zjistit, jak může nastavení mixéru ovlivnit obsah účinné látky v tabletě. Uvažujeme následující nastavitelné faktory:

- Existují různé typy mixéru od stejného výrobce o kapacitě 25, 50 a 75 litrů.
- Každý z mixérů může být vybaven lopatkami s různým počtem listů, od tří do pěti.
- Lopatky mixéru mohou být nainstalovány pod různým úhlem, od třiceti do šedesáti stupňů.



Soustředíme se na odhad vlivu těchto faktorů na střední hodnotu účinné látky (odezva) a budeme předpokládat normalitu reziduí.

V následujících odstavcích začneme s velmi jednoduchou verzí tohoto příkladu a budeme ji postupně rozšiřovat.

#### 4.1. Dvě populace

Začneme nejprve s výrazně jednodušším případem; uvažujme, že chceme studovat pouze rozdíl mezi dvacetipětilitrovým a padesátilitrovým mixérem za jistého nastavení ostatních faktorů. Označme si jako  $T^I$  indikátor situace, kdy používáme větší mixér. Indikátorem rozumíme takovou veličinu, která nabývá hodnoty jedna, pokud používáme větší mixér a hodnoty nula, pokud používáme menší mixér.

Namísto toho, abychom předpokládali, že všechny vyprodukované tablety se budou řídit jedním normálním rozdělením  $X \sim N(\mu, \sigma^2)$ , budeme uvažovat, že máme dvě rozdílné populace tablet. Tablety z menšího mixéru budou rozděleny dle  $X^1 \sim N(\mu_1, \sigma^2)$  a z většího dle  $X^2 \sim N(\mu_2, \sigma^2)$ . Odhadnout tato rozdělení můžeme jednoduše tak, že získáme vzorek z obou populací (tj. z obou mixérů) a odhadneme střední hodnotu a rozptyl zvlášť.

Všimněme si ovšem, že parametr  $\sigma^2$  je pro obě rozdělení stejný, ale byl by odhadován dvakrát, což nutně vede k snížení efektivity využívání dostupné informace a také k problému s interpretací, neboť bychom ve výsledku měli pro stejný parametr dva různé odhady.

Pokud chceme uvažovat stejný rozptyl pro oba mixéry, pak hodnotu odezvy  $X$ , která reprezentuje obsah účinné látky v celé populaci tablet, lze vyjádřit jako

$$X = \alpha + \beta T^I + \varepsilon, \quad (5)$$

kde  $\varepsilon \sim N(0, \sigma^2)$  a  $T^I$  označuje indikátor toho, že tableta byla vyrobena za použití většího mixéru. Zde  $\alpha$  reprezentuje střední hodnotu při použití menšího mixéru, a tedy  $\alpha = \mu_1$ , zatímco  $\beta$  reprezentuje rozdíl mezi střední hodnotou menšího a většího mixéru, tj.  $\beta = \mu_2 - \mu_1$ . Náhodná veličina  $\varepsilon$  se často nazývá náhodná chyba, ale ve skutečnosti zahrnuje nejruznější chyby měření a vliv faktorů, které jsme nevzali v úvahu (např. teplota ve výrobní hale).

Zdůrazněme, že jsme použili předpoklad stejných rozptylů. Není-li tento předpoklad splněn, pak může být nejlepším řešením odhadnout obě populace zvlášť. Podobně bychom se měli mít na pozoru ohledně předpokládané normality. Dodejme, že normální rozdělení  $\varepsilon$  vede k normálnímu rozdělení odezvy  $X$ . Tato vlastnost je specifická právě pro normální rozdělení a neplatí pro jiná rozdělení.

#### 4.2. Faktoriální experimenty

Prozkoumejme nyní složitější případ, kdy budeme uvažovat tři faktory namísto jednoho. Každý z nich bude faktorem s dvěma možnými, ne nutně číselnými, hodnotami:

- $T^I$  definujme jako výše: indikátor pro použití většího (padesátilitrového) místo menšího mixéru.
- $L^I$  definujme jako indikátor použití čtyř listů lopatky namísto tří listů lopatky.
- $U^I$  definujme jako indikátor nastavení šedesáti stupňů namísto třiceti stupňů.

Nejjednodušší model pro vztah odezvy  $X$  a tří faktorů reprezentujících nastavení bychom pak mohli napsat jako

$$X = \alpha + \beta_1 T^I + \beta_2 L^I + \beta_3 U^I + \varepsilon, \quad (6)$$

kde  $\varepsilon \sim N(0, \sigma^2)$ . Parametr  $\alpha$  reprezentuje střední hodnotu, jsou-li všechny indikátory nulové, a jednotlivé parametry  $\beta_k$  pak rozdíl středních hodnot, změníme-li některé nastavení. Všimněme si, že tento model vysvětluje působení jednotlivých faktorů nezávisle. Je-li například  $\beta_2 = 1,3$ , použití čtyř listů na lopatce zvýší obsah účinné látky o 1,3 neohledě na typ mixéru a nastavený úhel. To nezní jako reálný předpoklad; pokud nemáme takovou předchozí informaci, zdá se rozumnější připustit možnost, že vliv počtu listů a úhlu by se mohl lišit v závislosti na velikosti použitého mixéru. Tuto možnost připustíme přidáním takzvaných interakcí do modelu:

$$X = \alpha + \beta_1 T^I + \beta_2 L^I + \beta_3 U^I + \gamma_{12} T^I L^I + \gamma_{13} T^I U^I + \gamma_{23} L^I U^I + \delta_{123} T^I U^I L^I + \varepsilon, \quad (7)$$

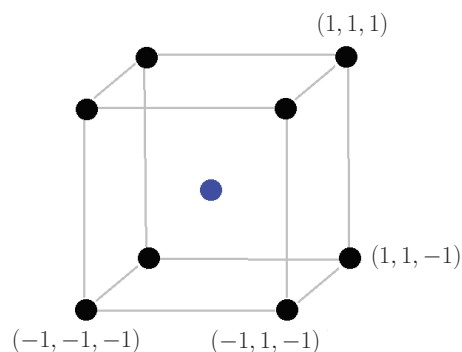
kde parametry  $\gamma$  jsou interakcemi druhého řádu a  $\delta$  interakcí třetího řádu. Připomeňme, že veličiny použité v modelu jsou indikátory druhého stupně jednotlivých nastavení, tedy nabývají hodnot nula či jedna, a parametr  $\gamma_{12}$  bude v modelu přítomen právě tehdy, budou-li oba indikátory rovny jedné, tj. když použijeme lopatku se čtyřmi listy ve větším (padesátilitrovém) mixéru.

Zatímco v modelu (6) by druhý typ lopatky zvýšil obsah účinné látky o  $\beta_2$ , v rozšířeném modelu (7) toto platí pouze pro menší mixér. Pro větší mixér bude velikost změny  $\beta_2 + \gamma_{12}$  pro úhel třicet stupňů a  $\beta_2 + \gamma_{12} + \delta_{123}$  pro šedesát stupňů. Interakce druhého řádu tedy umožňují, aby se vliv jednotlivých faktorů měnil v závislosti na hodnotě jiného faktoru. Interakce třetího řádu pak umožňují změnu interakce druhého řádu v závislosti na třetím faktoru. Další zobecnění je možné: s přidáním dalších faktorů vstupují do modelu stále vyšší a vyšší stupně interakcí.

### 4.3. Počet pozorování pro faktoriální experiment

Prosté spočítání parametrů modelu (7) vede k výsledku, že je třeba minimálně devět pozorování, abychom mohli odhadnout všechny parametry. Základní faktoriální experiment pro tři faktory s dvěma úrovněmi je však v literatuře často uváděn jako experiment s osmi pozorováními. To je důsledek toho, že cílem je odhad střední hodnoty a parametr  $\sigma^2$  je ignorován.

Osm pozorování je přirozeně rozděleno tak, že každá kombinace tří možných nastavení faktorů je pozorována právě jednou, jak můžeme vidět na obrázku 1. Ten ukazuje tři faktory, každý znázorněný na jedné ose s hodnotou  $-1$  označující jednu možnou hodnotu faktoru a hodnotou  $1$  označující druhou možnou hodnotu faktoru. V takovém experimentu s osmi pozorováními pak odhadujeme střední hodnotu odezvy v každé kombinaci nastavení na základě jediného pozorování, což nám zřejmě nedává žádnou informaci o nejistotě našeho odhadu. Zároveň ani odhad této střední hodnoty není nijak přesný. Pokud informaci o nejistotě chceme zahrnout, je nutné přidat další pozorování. Pokud přidáme jen jediné, budeme odhadovat rozptyl na základě dvou pozorování v jedné z kombinací nastavení, proto je obvykle doporučováno přinejmenším několik dalších pozorování v několika různých nastaveních.



Obr. 1. Příklad faktoriálního experimentu s třemi faktory (každý se dvěma možnými hodnotami) a osmi pozorováními (černé body). Modrý bod uprostřed značí centrální pozorování, které je často přidáno do experimentu, dává-li smysl (tedy existuje-li pro každý faktor logické nastavení mezi nižší a vyšší hodnotou)

Určení celkového počtu pozorování by pak mělo vzít v úvahu i sílu testu, pokud je k dispozici předchozí znalost o rozptylu. Analýza faktoriálních experimentů se obvykle provádí metodami analýzy rozptylu a můžeme tedy uvažovat odhady vyplývající z tohoto přístupu (v případě, že chceme použít jako kritérium délku konfidenčních intervalů) či příslušné statistické testy pro výpočet jejich síly. Je důležité si ujasnit, zda chceme definovat sílu vzhledem ke globálním testům (alespoň jeden parametr je významný), vzhledem ke konkrétnímu faktoru, o který se nejvíce zajímáme (například velikost mixéru je pro nás nejdůležitější faktor), nebo vůči všem testům na jednotlivé parametry. V posledním případě může být nutné uvažovat i o mnohonásobném porovnávání a případné úpravě hladiny významnosti.

#### 4.4. Redukce faktoriálních experimentů

Každá další úroveň nastavení znamená další parametry v modelu, jeden pro hlavní efekt a další pak pro interakce nejrůznějších stupňů. Pokud bychom například náš příklad rozšířili přidáním 75litrového mixéru, tj. jednou další úrovní pro typ mixéru, museli bychom zavést nový indikátor  $T_2^I$  a celkově by nám přibyly čtyři parametry: jeden hlavní efekt, dvě interakce nového indikátoru s  $L^I$  a  $U^I$  a trojná interakce. Další přidání nové úrovně počtu lopatek by zvýšilo počet parametrů ještě více, protože bychom museli uvažovat interakce s třemi úrovněmi typu mixéru namísto původních dvou.

Obdobně přidání dalších veličin by vedlo k rychlému zvýšení počtu parametrů, neboť by bylo třeba odhadnout interakce všech řádů se všemi dosavadními veličinami a taktéž zvýšit řád interakcí. Z tohoto důvodu jsou často faktoriální experimenty neproveditelné v praxi, máme-li uvažovaných veličin příliš mnoho. Typickým příkladem jsou experimenty na počátku vývoje výrobního procesu, tak zvané „screening experiments“, kdy máme jen málo znalostí o procesu a účelem experimentu je odhalit, které z mnoha veličin mají nějaký dopad na kvalitu výsledného produktu.

Řešením je pak redukce faktoriálních experimentů. Ve stručnosti to znamená, že rezignujeme na odhad efektu interakcí vyšších řádů a soustředíme se na hlavní

efekty a interakce nízkých řádů, přičemž se snažíme, aby odhadované parametry byly nezávislé. Tento přístup vede k dramatickému snížení nutných experimentů za cenu předpokladu, že důležité efekty se nebudou projevovat pouze v interakcích vyšších řádů. Tento předpoklad se zdá být v praxi poměrně rozumný.

Kromě redukováných faktoriálních experimentů existují další metody, jak navrhovat malé experimenty pro screening, například orthogonal arrays [5] nebo nedávno vyvinuté definitive screening designs [6]. Super-saturated designs [4] se vypořádávají se situací, kdy maximální počet možných experimentů je nižší než počet faktorů, které potřebujeme vyhodnotit. To má za následek, že musíme provést nějaký výběr veličin, které v modelu budou, neboť nemůžeme odhadnout ani základní model se všemi hlavními efekty.

#### 4.5. Kvantitativní veličiny

Uvažujme nyní náš původní příklad v plném znění. Pro typ mixéru a počet lopatek tak máme tři úrovně, zatímco pro úhel máme dokonce 31 možných úrovní, pokud předpokládáme, že jakýkoli celočíselný úhel mezi třiceti a šedesáti stupni může být použit jako nastavení. Pokud bychom uvažovali úhel jako faktor, potřebovali bychom třicet pozorování jen pro odhad hlavního efektu jednotlivých nastavení.

Namísto kvalitativních veličin však můžeme uvažovat kvantitativní veličiny. Označme jako  $T^K$ ,  $L^K$  a  $U^K$  veličiny, které budou nabývat hodnoty  $-1$  pro minimální nastavení a hodnoty  $1$  pro maximální a budeme je uvažovat jako spojité veličiny:

- $T^K = -1$  pro 25 litrů,  $T^K = 0$  pro 50 litrů a  $T^K = 1$  pro 75 litrů objemu mixéru.
- $L^K = -1$  pro tři listy,  $L^K = 0$  pro čtyři listy a  $L^K = 1$  pro pět listů lopatky.
- $U^K = -1$  pro třicet stupňů,  $U^K = 0$  pro čtyřicet pět stupňů a  $U^K = 1$  pro šedesát stupňů.

Pro veličinu  $L^K$  je tento koncept jasně hypotetický, neboť „tři a půl“ listu lopatky nedává v praxi smysl. Pro veličinu  $T^K$  pak můžeme spekulovat, do jaké míry by mohla být provedena interpolace pro například 37litrový mixér. Konečně, veličina  $U^K$  se nejvíce podobá spojité veličině, neboť máme definované všechny celočíselné hodnoty. Kvalitativně však mezi tím není velký rozdíl, neboť všechny veličiny mají konečnou množinu hodnot, zatímco my předpokládáme, že jsou spojité na reálném oboru. Jakkoli je tento model v tomto smyslu špatně, může se v praxi ukázat jako velice užitečný.

Model (7) tak můžeme přepsat ve tvaru

$$X = \alpha + \beta_1 T^K + \beta_2 L^K + \beta_3 U^K + \gamma_{12} T^K L^K + \gamma_{13} T^K U^K + \gamma_{23} L^K U^K + \delta_{123} T^K U^K L^K + \varepsilon. \quad (8)$$

Přestože model vypadá na první pohled stejně, interpretace parametrů se liší, neboť veličiny již nejsou pouze indikátory, ale nabývají interpretovatelných hodnot na intervalu  $-1$  až  $1$ . Proto jednotlivé parametry vyjadřují, jak by se změnila hodnota  $X$ , pokud se zvýší příslušná veličina o jedna (vzhledem k transformaci veličin na obor  $-1$  až  $1$  to znamená změnu veličiny o polovinu experimentálního oboru, tj. pro  $U^K$  změnu o patnáct stupňů). Dodejme, že tento model předpokládá, že zvýšení veličiny

o hodnotu jedna bude mít stejný efekt na odezvu, ať je počáteční hodnota jakákoli, tj. zvýšení z  $-1$  na  $0$  bude mít stejný efekt jako zvýšení z  $0$  na  $1$ .

Přestože počet parametrů je stejný jako pro model (7), všimněme si, že zde máme tři úrovně pro veličiny  $T$  a  $L$  a 31 možných úrovní pro  $U$ , zatímco předchozí model připouštěl pouze dvě úrovně nastavení pro každou veličinu. Faktoriální model s takovým množstvím úrovní by vedl k 279 parametrům. Toto je jedna z hlavních výhod používání kvantitativních veličin, kdykoli je možné hodnoty faktorů seřadit a přidělit jim číselnou hodnotu. Samozřejmě ne vždy je to možné, například je-li faktorem jméno operátora výrobní linky, jméno firmy dodávající výrobní materiál či typ přístroje využitého k měření. Na druhou stranu hlubší porozumění procesu často vede k odhalení měřitelných veličin v některých z těchto případů. Místo jména dodavatele můžeme měřit kvalitu dodávaného materiálu či místo typu přístroje můžeme měřit nějakou operační veličinu (např. teplotu v přístroji).

Důležité pro modelování je, že přidání dalších uvažovaných úrovní pro kvantitativní veličiny nezvyšuje počet parametrů v modelu. Důvodem je skutečnost, že neodhadujeme střední hodnotu zvláště na všech úrovních, ale pouze několik parametrů, které určí funkční závislost. Proto ani není nutné obdržet pozorování na všech možných úrovních všech veličin, ale stačí pozorování ve středu a v krajních bodech oborů těchto veličin. Z tohoto důvodu není problém zahrnout i veličinu jako je  $L^K$ , protože nebude třeba používat nastavení „tři a půl listu lopatky“, které by bylo v praxi nerealizovatelné. Druhým užitečným důsledkem je možnost interpolace, pokud dává pro danou veličinu smysl. Například i když pro veličinu  $U^K$  budeme mít pozorování pouze v 30, 45 a 60 stupních, můžeme na základě modelu předpovědět střední hodnotu pro nastavení 50 stupňů.

Zopakujme, že cenou za použití modelu (8) je předpoklad lineární závislosti mezi  $X$  a příslušnou veličinou. Pokud je tento předpoklad nesprávný, může být výsledný odhad zcela špatně. Smysluplnost tohoto předpokladu může být do jisté míry ověřena tím, jak dobře daný model vysvětluje napozorovaná data. V případě, že model příliš zjednodušuje funkční závislost a nevypadá přijatelně, můžeme uvažovat nejrůznější komplikovanější modely.

#### 4.6. Odezvové plochy

V praxi se často uvažují pouze interakce druhého řádu, neboť efekty interakcí vyšších řádů jsou obvykle zanedbatelné. Na druhou stranu nám reprezentace faktorů spojitými veličinami umožňuje modelovat nejen lineární závislost, ale i komplikovanější funkce. Kombinace interakcí druhého řádu a kvadratických efektů se nazývá „response surface design“, v češtině metoda odezvových ploch:

$$X = \alpha + \beta_1 T^K + \beta_2 L^K + \beta_3 U^K + \gamma_{12} T^K L^K + \gamma_{13} T^K U^K + \gamma_{23} L^K U^K + \omega_1 (T^K)^2 + \omega_2 (L^K)^2 + \omega_3 (U^K)^2 + \varepsilon. \quad (9)$$

V tomto případě  $\omega_1$ ,  $\omega_2$ ,  $\omega_3$  reprezentují kvadratické efekty. Ty jsou při některých typech experimentů velice důležité; pokud je skutečné optimální nastavení takové, kdy jsou všechny faktory nastavené blízko střední úrovně, tj. transformované spojitě veličiny blízko nule, pak je přirozené, že odezva zde nabývá svého minima/maxima a na celém experimentálním oboru pak pozorujeme kvadratickou funkci.

Právě určení optimálního nastavení uvažovaných faktorů je obvykle výstupem těchto experimentů. Pokud existuje nějaké jasné kritérium udávající, jaké hodnoty odezvy jsou vyhovující, pak může být ustanoven takzvaný „design space“, tedy obor hodnot, na kterém střední hodnota odezvy splňuje daná kritéria. Dodejme, že při použití bayesovského přístupu a zohlednění  $\varepsilon$  může být pro různá nastavení snadno odhadnuta pravděpodobnost, že individuální pozorování (namísto střední hodnoty) budou splňovat požadovaná kritéria.

Dodejme, že v praxi jsou často počty pozorování nutné pro experiment založený na odezvové ploše velmi vysoké, máme-li více než tři až pět faktorů. Z tohoto důvodu se pro větší počet faktorů často používají dvoustupňové přístupy. Nejdříve proběhne takzvaný screening založený například na redukovaných faktoriálních experimentech, který se soustředí především na odhad hlavních efektů a bere v úvahu jen limitovaný počet interakcí. Výstupem je zmenšená skupina faktorů, které jsou poté vyhodnoceny pomocí dalšího experimentu založeného na odezvoových plochách.

#### 4.7. Optimální experimenty

Jak bylo vysvětleno v předchozí sekci, odhadnutelnost parametrů, síla testů či požadovaná přesnost jsou obvyklými kritérii při určování nutného počtu pozorování. Následně je důležité rozhodnout, jaké nastavení hodnot faktorů by při daném počtu pozorování měla jednotlivá pozorování mít. Je to důležitá otázka, neboť toto rozhodnutí může velmi silně ovlivnit výsledky našeho experimentu například ve smyslu přesnosti předpovědi ve vztahu k nějakému kritériu v různých částech experimentálního oboru hodnot. V zásadě pracujeme s tím, že rozptyl je vyšší při nastavení (a v jeho okolí ve smyslu příslušného eukleidovského prostoru), kde je realizováno méně pozorování, nežli při nastavení, kde jich je více. Smyslem je tedy dosáhnout rovnoměrného rozložení pozorování na celém mnohorozměrném oboru hodnot, ovšem na druhou stranu mít dostatek pozorování blízko sebe za účelem dobrého odhadu reziduální variability.

Optimalita je obvykle definována ve smyslu D-optimality či I-optimality, které se soustředí na minimalizaci rozptylu odhadů (prostřednictvím minimalizace determinantu varianční-kovarianční matice) či průměrného rozptylu predikce (na celém uvažovaném experimentálním oboru). Optimalizace je prováděna numerickými algoritmy a pracuje s fixním počtem pozorování, přičemž se snaží určit optimální nastavení faktorů pro jednotlivá pozorování. Zároveň algoritmy berou v úvahu specifikovaný model, který chceme na závěr odhadnout.

V praxi tedy vezmeme například model (8), určíme kritérium optimality a počet pozorování a obdržíme výsledná nastavení pro celý experiment. Výsledek pak můžeme vidět v tabulce 1. Experiment je již randomizovaný, tj. nepozorujeme logickou posloupnost v pořadí nastavení. Celkový počet pozorování je šestnáct a každý řádek určuje nastavení hodnot faktorů pro dané pozorování. Na rozdíl od faktoriálního experimentu návrh neobsahuje všechny možné kombinace tří veličin. Všimněme si, že přestože jsme uvažovali faktory jako spojité, návrh obsahuje pouze hodnoty  $-1, 0, 1$ , takže nebude problém provést nastavení ani pro počet lopatek a typ mixéru.

Například první realizované pozorování bude nastaveno jako padesátilitrový mixér (nulová hodnota odpovídá střední úrovni) s pěti lopatkami pod úhlem 45 stupňů. Pozorování 6 a 7 reprezentují všechny faktory nastavené na nejnižší, respektive nejvyšší, úrovni. V experimentu není přítomno žádné pozorování, pro které by byly všechny

Pozorování	Mixér	Lopatky	Úhel
1	0	1	0
2	-1	1	-1
3	-1	-1	0
4	-1	0	1
5	0	-1	-1
6	-1	-1	-1
7	1	1	1
8	1	-1	-1
9	0	-1	1
10	-1	1	1
11	-1	0	-1
12	1	0	0
13	1	1	-1
14	0	0	-1
15	1	-1	1
16	-1	-1	1

Tab. 1. Příklad navrženého experimentu pro model založený na odezvové ploše

faktory nastavené na střední úrovni, ani není žádné pozorování opakováno vícekrát. V mnoha praktických případech stojí za zvážení přidat do návrhu jak centrální pozorování, tak opakování některých pozorování (typicky extrémních pozorování, tj. 6 a 7). Záleží ovšem na konkrétním kontextu, interpretaci jednotlivých nastavení a maximálního přijatelného počtu pozorování.

#### 4.8. Náhodné efekty

Předpokládejme nyní, že můžeme provést pouze tři pozorování denně a vliv dne nás nezajímá, ale chtěli bychom se vyhnout případnému vychýlení a zvýšené reziduální variabilitě. Den tedy můžeme zahrnout jako náhodný efekt a odhadneme rozptyl příslušný rozdílu mezi jednotlivými dny. Všimněme si, že náš základní experiment má šestnáct pozorování, proto potřebujeme přinejmenším šest dní a dává tedy smysl použít dny jako náhodný efekt. Vzhledem k tomu, že víme, kolik pozorování denně lze provést, můžeme pomocí bloků zahrnout den do návrhu již při přípravě experimentu. Dostaneme pak model

$$\begin{aligned}
X = & \alpha + \beta_1 T^K + \beta_2 L^K + \beta_3 U^K + \gamma_{12} T^K L^K + \gamma_{13} T^K U^K + \gamma_{23} L^K U^K + \\
& + \omega_1 (T^K)^2 + \omega_2 (L^K)^2 + \omega_3 (U^K)^2 + d + \varepsilon
\end{aligned} \tag{10}$$

s jedním dodatečným parametrem. Při aplikaci D-optimálního experimentu s náhodnými bloky se třemi pozorováními pak můžeme požadovat o dvě pozorování navíc (tj. šest plných dní) a získáme nastavení v tabulce 2. Poslední sloupec pak udává, ve které dny bude které pozorování realizováno.

#### 4.9. Poznámka ohledně randomizace

Kterýkoli návrh experimentu by měl být podroben randomizaci pořadí pozorování, aby se zajistilo, že je případný vliv neznámých faktorů na střední hodnotu odezvy mi-

Pozorování	Mixér	Lopatky	Úhel	Den
1	1	1	-1	1
2	0	-1	0	1
3	1	0	1	1
4	-1	1	1	2
5	1	0	0	2
6	0	-1	1	2
7	1	-1	-1	3
8	0	1	0	3
9	-1	0	0	3
10	-1	1	-1	4
11	0	0	0	4
12	-1	-1	1	4
13	-1	-1	-1	5
14	0	0	1	5
15	1	-1	0	5
16	-1	0	0	6
17	1	1	1	6
18	0	0	-1	6

Tab. 2. Příklad navrženého experimentu s náhodným efektem dne

nimalizován. Oba výše uvedené návrhy randomizovány byly, tj. nepozorujeme logickou posloupnost v pořadí nastavení.

Mějme však vždy na paměti praktická rizika úplné randomizace a konzultujme je pokud možno s osobami, které budou experiment provádět. Je třeba předem vyjasnit, zda randomizace pořadí nezvyšuje riziko nesprávných výsledků (tj. zda obdržíme správná data), a pokud ano, jaký typ randomizace je přijatelný.

Je taktéž dobré stanovit předem postup pro případ, že nějaké pozorování selže (například výpadek elektrického proudu): zda má být pozorování zopakováno, zda je lépe zopakovat je na konci, či zda by měl být zopakován celý blok. První možnost v našem případě není možná, neboť používáme náhodné bloky a nejspíš není možné provést jedno mimořádné pozorování ve stejný den. Dávalo by tedy smysl buď zopakovat celý blok (například pokud selže hned první pozorování v bloku) nebo realizovat chybějící pozorování na konci. Pokud to umožňují prostředky, je vhodné v takovém případě přidat ještě jedno či dvě další pozorování, aby bylo možné odhadnout efekt tohoto bloku. V některých speciálních případech si dokonce můžeme dovolit ztracené pozorování nenahrazovat, například pokud se jedná o pozorování, které je v návrhu několikrát opakováno, a jedno chybějící pozorování neohrozí odhadnutelnost a nemá velký dopad na sílu testu.

## 5. Diskuze

Navrhování experimentů je jedna z oblastí aplikace statistiky, která vedla k položení základů oboru. Zatímco první koncepty se objevily již v 18. století, současný přístup byl formován R. A. Fisherem ve třicátých letech 20. století [3]. Přestože postupy Fishera se



stále hojně využívají (faktoriální experimenty), téma rozhodně není vyčerpáno a vývoj metod pokračuje i v současné době. Ve smyslu minimalizace počtu pozorování v experimentu to mohou být uvedené supersaturované experimenty nebo tzv. definitive screening designs, ve smyslu optimalizace nastavení pak nejrůznější kritéria pro optimální experimenty. Obecně vývoj v matematické statistice jako takové umožňuje nové přístupy i v této aplikované doméně.

Z pohledu pronikání metod do praxe provozované nejen statistiky se ukázaly v minulosti velmi důležité přístupy zjednodušující metodologii na minimum. Příkladem mohou být Taguchiho metoda [9] či přístupy založené na Paretově principu (například Red X). Tyto statisticky spíše suboptimální metody pomohly rozšířit povědomí o užitečnosti metodického přístupu k navrhování experimentů mezi odbornou veřejnost. Jejich výhodami byly především snadná pochopitelnost a intuitivnost bez nutnosti pochopení obsáhlé teorie. V uplynulých desetiletích jsou postupně nahrazovány přístupy popsány v tomto článku i díky tomu, že pronikání nových metod do praxe je značně zjednodušeno dostupností specializovaných výpočetních programů se snadno ovladatelným uživatelským rozhraním.

Tento článek se pokusil nastínit, jak proces navrhování experimentu vypadá, jak probíhá v praxi, jaké aspekty je třeba vzít v úvahu a jaké metody lze využít. Látka byla ovšem podána jen velice přehledově vzhledem k tomu, že navrhování experimentů je téma velice široké. S ohledem na neustále pokračující výzkum a přesah tématu do mnoha aplikovaných oborů je obtížné doporučit vyčerpávající literaturu. Některé publikace se více soustředí na randomizaci, reprezentativnost pozorování, modelování rozptylu a sílu testů, zatímco jiné na odhad střední hodnoty, minimalizaci počtu nutných pozorování a odhadnutelnost parametrů.

V kontextu průmyslových experimentů autor doporučuje publikaci [7].

**Poděkování.** Autor děkuje RNDr. Jiřímu Dvořákovi, Ph.D., doc. RNDr. Antonínu Slavíkovi, Ph.D., a Ing. Ivanu Millerovi za užitečné komentáře, které přispěly k výraznému vylepšení článku.

#### L i t e r a t u r a

- [1] ANDĚL, J.: *Statistické metody*. MatfyzPress, Praha, 2007.
- [2] DVOŘÁK, J.: *O dětech, čápech a kauzalitě*. PMFA 62 (2017), 264–274.
- [3] FISHER, R. A.: *The design of experiments*. Macmillan, 1935.
- [4] GEORGIU, D. S.: *Supersaturated designs: A review of their construction and analysis*. J. Statist. Plann. Inference 144 (2014), 92–109.
- [5] HEDAYAT, A. S., SLOANE, N. J. A., STUFKEN, J.: *Orthogonal arrays: Theory and application*. Springer, 1999.
- [6] JONES, B., NACHTSHEIM, C. J.: *Effective design-based model selection for definitive screening design*. Technometrics 59 (2017), 319–329.
- [7] MONTGOMERY, D. C.: *Design and analysis of experiments*. 8th edition, Wiley, 2012.
- [8] OTAVA, M.: *Stručný průvodce statistickými intervaly*. PMFA 62 (2017), 7–16.
- [9] TAGUCHI, G., YU-IN, W.: *Introduction to off-line quality control*. Central Japan Quality Control Association, Magaya, 1979.
- [10] ZVÁRA, K.: *Regrese*. MatfyzPress, Praha, 2008.