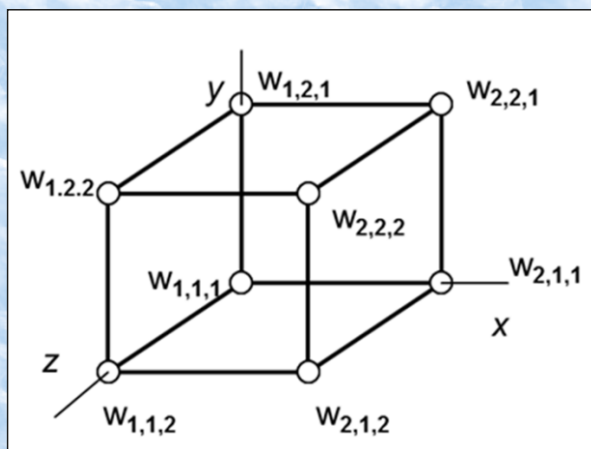
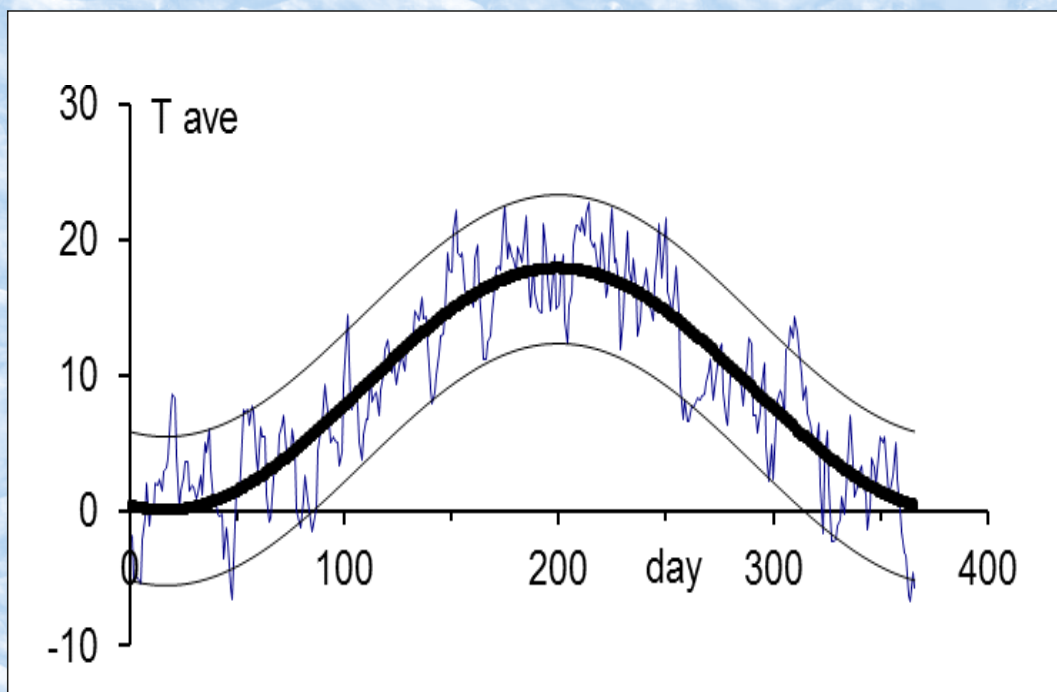


Jaroslav Menčík



# Úvod do experimentální analýzy



Tato publikace neprošla jazykovou korekturou.

Recenzenti:

Prof. Dr.Rer.Nat. Dietrich Munz

Doc. Ing. Jiří Němeček, PhD.

Prof. Michael V. Swain

© prof. Ing. Jaroslav Menčík, CSc., 2017

ISBN 978-80-7560-066-0 (tisk)

ISBN 978-80-7560-067-7 (pdf)

Knih je volně přístupná na <http://hdl.handle.net/10195/66960>.

## Úvod do experimentální analýzy

Kniha podává přehled základních metod pro navrhování a vyhodnocování experimentů. Jsou ukázány hlavní kroky experimentálního zkoumání, vysvětleny příčiny chyb a kolísání měřených veličin, a shrnuty základy statistických metod pro vyhodnocování. Je ukázáno, jak popisovat výsledky měření, od parametrů až po rozdělení pravděpodobnosti. Následuje charakterizování vztahů mezi vyšetřovanými veličinami a prokládání empirických dat regresními funkcemi. Je vysvětlena tvorba intervalů spolehlivosti pro zjišťované parametry a stanovení rozsahu měření pro zajištění požadované přesnosti výsledku nebo pro ověření hypotézy. Jedna kapitola je věnována teorii podobnosti a rozměrové analýze, které pomáhají zmenšit rozsah experimentů a činí jejich výsledky obecnějšími. Je vysvětlena analýza rozptylu, plánování pokusů a experimentální nalezení maxima nebo minima. Tyto postupy umožňují najít významné faktory a optimální parametry určitého procesu nebo konstrukce. Kapitola o citlivostní analýze ukazuje, jak kolísání vstupních veličin přispívá k odchylkám vyšetřované veličiny od optimální nebo jmenovité hodnoty. To se využívá ke stanovení tolerancí. Poslední kapitola je věnována nástrojům pro zkoumání chování náhodných veličin, jako je simulační metoda Monte Carlo. Použití jednotlivých metod je ilustrováno na příkladech a kapitoly jsou doplněny odkazy na literaturu.

Kniha je volně přístupná na <http://hdl.handle.net/10195/66960>.

## Poděkování

Během přípravy této knihy jsem využil zkušeností získaných v práci a přednášení v různých institucích. Jsem vděčný svým kolegům za jejich rady, jakož i studentům za jejich otázky. Obzvláštění dík si zaslouží Prof. Milan Lánský a Prof. Miroslav Steiner z Univerzity Pardubice, Prof. Michael Swain z University of Sydney, Prof. Dietrich Munz z Karlsruhe Institute of Technology (KIT), Prof. Eckhard Quandt z Univerzity Kiel, Dr. Andy Bushby z University of London, Prof. Hitoshi Tokura z Tokyo Institute of Technology, Dr. Gaston Rauchs z Luxembourg Institute of Science and Technology a docent Jiří Němeček z ČVUT v Praze. Poděkování náleží i mým bývalým kolegům z VÚSABu Jablonec nad Nisou, Ing. Radovanu Drahorádovi a Ing. Miroslavu Matuškoví, CSc., s nimiž jsem ve výzkumné činnosti začínal.

## **Introduction to Experimental Analysis**

This book brings a review of basic methods for design and evaluation of experiments. It shows the principal steps and explains the causes of dispersion and errors of the measured values, as well as statistical methods for their evaluation. It is shown how the measured values can be described by simple characteristics or probability distributions. Then, characterisation of relations between investigated quantities is described, including fitting of the data by regression functions. It is explained how to create confidence intervals, which contain the true values of the parameters, and how many tests are necessary for obtaining the results with the demanded accuracy or for the verification of a certain hypothesis. One chapter is devoted to the theory of similarity and dimensional analysis, which help one to reduce the extent of experiments and make the results more general. Other chapters explain the analysis of variance, design of experiments, and experimental finding of a maximum or minimum. These procedures facilitate finding of the most important factors and optimum parameters. Sensitivity analysis shows how variations of the input quantities cause deviations of the investigated quantity from the optimum or nominal value. The last chapter is devoted to the efficient tools for the study of random influences, such as the Monte Carlo simulation technique. The use of the described methods is illustrated on examples and the individual chapters are supplemented by references.

The book can be downloaded on <http://hdl.handle.net/10195/66961>.

# OBSAH

1.	Úvod	7
2.	Etapy experimentálního výzkumu	9
3.	Chyby a kolísání měřených hodnot	15
4.	Základy pravděpodobnosti a statistiky pro experimentální analýzu	19
5.	Určování charakteristik zkoumaných veličin	33
6.	Vztahy mezi dvěma a více veličinami	43
7.	Vyrovňování empirických dat regresními funkcemi	51
8.	Intervaly spolehlivosti, testování hypotéz a nezbytné množství dat	65
9.	Rozměrová analýza a teorie podobnosti	81
10.	Analýza rozptylu	97
11.	Plánování experimentů (DOE)	103
12.	Experimentální nalezení extrémů	113
13.	Citlivostní analýza	121
14.	Simulační metody pro zkoumání náhodných veličin a vlivů	127
	Rejstřík	137

***Není na světě nic praktičtějšího, než dobrá teorie.***

Ludwig Boltzmann

***Měřte vše měřitelné, a neměřitelné učíte měřitelným.***

Galileo Galilei

# 1. Úvod

Účelem výzkumu je získávat informace o předmětu našeho vyšetřování. Často je třeba najít optimální složení vyvíjeného materiálu nebo parametry určitého stroje, součástky nebo procesu. Požadované informace lze získat teoretickou činností, pozorováním a experimentem. **Teoretická činnost** znamená práci s abstraktními modely. Při **pozorování** je náš postoj pasivní, pouze pozorujeme, co se děje a zaznamenáváme charakteristické jevy a hodnoty, aniž bychom zasahovali do vyšetřovaného objektu nebo měnili podmínky během pozorování. V některých případech je pozorování jedinou možností, jak získat informace. Příkladem je studium hvězd nebo lidské společnosti, anebo vývoj nového léku, který by mohl léčit, ale také zabíjet. **Experiment** je cílevědomá činnost, která má přinést více informací díky řízenému působení na vyšetřovaný reálný objekt nebo jeho model. **Model** může být hmotný nebo počítačový. Pozorování a experiment umožňují získání vstupních hodnot a verifikaci hypotéz o zkoumaném objektu.

Experimenty jsou velmi důležité pro získání informací. Během dlouhé doby byly vyvinuty různé metody, které dělají experimentování a vyhodnocování výsledků více systematickým a účinným. Každý rok mnoho studentů a dalších lidí přichází do styku s experimentálním výzkumem poprvé a musí se učit, jak jej provádět. Stále je ale platný poznatek, který kdysi učinil Pascal: “Teprve na konci práce víme, jak jsme měli začít”. Tato knížka chce zmírnit tento problém tím, že podává stručný přehled základních metod a přístupů.

Autor pracoval řadu roků v aplikovaném výzkumu. Později, na Univerzitě Pardubice, přednášel o výzkumných metodách studentům z různých zemí a oborů. S touto rozsáhlou zkušeností se rozhodl připravit stručnou knížku pro všechny, kteří chtějí získat úvodní informace o tom, jak účinně organizovat a vyhodnocovat experimenty, bez ohledu na oblast zkoumání. Aby si popsané postupy mohl procvičit i čtenář bez zvláštních nástrojů, lze ukázkové příklady řešit s pomocí „všudypřítomného“ Excelu.

Tato knížka o experimentální analýze je rozdělena do čtrnácti kapitol. Následující kapitola ukazuje základní kroky experimentálního zkoumání. Kapitola 3 popisuje, jaké chyby se mohou vyskytnout při experimentech nebo měření a ukazuje cesty k jejich odstranění nebo zmenšení. Následující kapitola vysvětluje základy pravděpodobnosti a statistických metod užívaných při experimentální analýze. Kapitola 5 ukazuje určování důležitých charakteristik vyšetřovaných veličin, od průměru přes histogram až k rozdělení pravděpodobnosti. Kapitola 6 je věnována vztahům mezi dvěma nebo více veličinami. Kapitola 7 vysvětluje prokládání empirických dat regresními funkcemi a určování jejich parametrů. Kapitola 8 je věnována stanovení rozsahu experimentů pro získání výsledků s požadovanou přesností. Vysvětluje i konstrukci intervalů spolehlivosti a statistické testy pro ověření, zda existuje významný rozdíl mezi různými soubory. Kapitola 9 uvádí základy teorie podobnosti a rozměrové analýzy. To jsou účinné nástroje, které zmenšují rozsah potřebných experimentů a činí jejich výsledky obecnějšími. Kapitola 10 vysvětluje základy analýzy rozptylu, která ukazuje významnost různých faktorů. Kapitola 11 je věnována metodám plánování pokusů. Ty umožňují rychle najít nejvýznamnější faktory a nejlepší parametry konstrukce nebo určitého procesu. Kapitola 12 ukazuje postupy pro experimentální nalezení maxima nebo minima určité funkce. Kapitola 13 (Analýza citlivosti) ukazuje, jak proměnlivost vstupních veličin přispívá k odchylkám vyšetřované veličiny od jmenovité nebo optimální hodnoty. Poslední kapitola je věnována nástrojům pro zkoumání chování náhodných veličin, jako je metoda Monte Carlo a metoda LHS.



## 2. Etapy experimentálního výzkumu

Každý experimentální výzkum má následující etapy: formulování účelu a cílů, příprava experimentů a měření, jejich provedení a vyhodnocení, formulace závěrů a publikování výsledků. Jednotlivé etapy jsou dále podrobně popsány.

### 1) Příprava experimentů

Příprava experimentů sestává z následujících kroků:

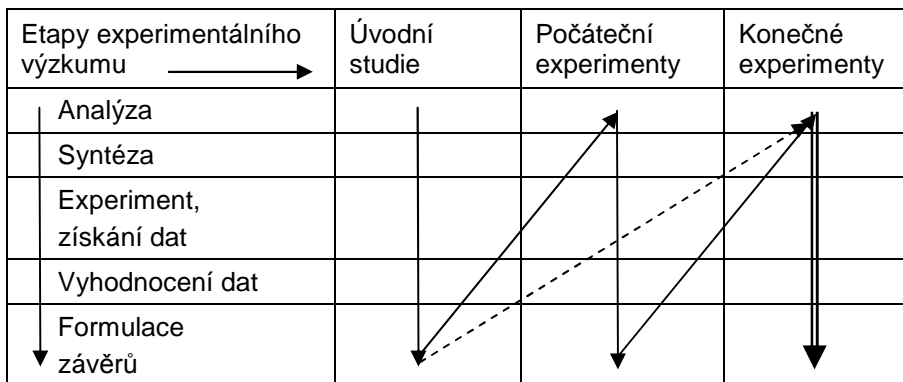
- Seznámení s problémem (analýza), formulace předmětu a cíle zkoumání.
- Volba charakteristických fyzikálních a dalších veličin popisujících chování a důležité vlastnosti vyšetřovaného objektu nebo jevu. Někdy je od začátku jasné, jaké veličiny použít, ale při zkoumání něčeho zcela nového musíme někdy dokonce příslušnou veličinu vytvořit. Nebojte se! Připomeňme si slavný Galileův výrok: „Změřte vše měřitelné, a neměřitelné učíte měřitelným!“ Koho by například před sto lety napadlo měřit množství informací!
- Výběr vhodných druhů experimentů a měřicích metod s ohledem na naše možnosti: vybavení, finanční prostředky a čas.
- Příprava plánu experimentů a měření, včetně jejich rozsahu (volba úrovní sledovaných veličin a jejich rozmezí, předběžné stanovení počtu experimentů a testů).
- Příprava přístrojů a potřebného vybavení, získání vzorků.

Plán všech experimentů, včetně časového rozvrhu, musí být zaznamenán předem.

### 2) Provedení experimentů, pozorování a měření

Experimenty jsou prováděny podle předem připraveného plánu. Výzkum je ale vždy doprovázen určitou mírou nejistoty, takže počáteční plán a časový rozvrh musí být často upraven s ohledem na dosažené výsledky. Někdy se proto experimenty provádějí ve dvou nebo i více etapách, jak je znázorněno na Obr. 2.1.

Experimenty v první etapě mají omezený rozsah a slouží pro lepší stanovení rozsahu a podmínek zbývajících experimentů.



*Obr. 2.1. Etapy experimentálního výzkumu, podle [2.1].*

Všechny experimenty musí být popsány ve výzkumném nebo provozním deníku. Tyto záznamy obsahují informaci o datu a době testů, soupis použitých přístrojů (včetně jejich typů, výrobních čísel a uspořádání nebo zapojení), seznam osob, které se pokusů zúčastnily (včetně jejich role), a popis experimentů a pozorování. Je-li to možné, tyto záznamy se dělají s použitím počítače nebo laptopu, ale mohou být dělány i ručně. Práci mohou někdy ušetřit předtištěné formuláře. Užitečné jsou též fotografie a skici. Experimentátor může také využít videokameru nebo magnetofon, a popřípadě zde komentovat svoji činnost nebo pozorování. Toto vše je nutno dělat pečlivě, protože někdy již není možno měření nebo pozorování provést znovu, jestliže z nějakého důvodu došlo ke ztrátě původních dat. Popis experimentů by měl být dělán tak pečlivě, aby je kdokoliv mohl v budoucnu zopakovat.

### 3) Vyhodnocení výsledků

Zpracování získaných hodnot zahrnuje jejich třídění, vytvoření tabulek a diagramů, návrh a ověření vztahů mezi sledovanými veličinami, včetně stanovení konstant v regresních funkcích, sestavení konfidenčních intervalů a testování různých hypotéz o vyšetřovaných jevech.

#### **4) Formulování závěrů**

Získané výsledky slouží pro formulaci závěrů, případně pro přípravu plánu dalších prací. Potom je možné napsat zprávu. Výsledky se také často publikují ve formě prezentace na konferenci nebo článku v časopise.

#### **5) Publikování výsledků**

Publikování je velmi významné zejména ve vědecké komunitě (univerzity, výzkumné ústavy). Výzkumné zprávy (v čitelné formě!) jsou ale důležité i pro výzkumná nebo vývojová oddělení. Připravujeme-li informaci o našem výzkumu k publikování, např. na konferenci nebo v časopise, je užitečné dodržovat některá osvědčená pravidla. Vědecký článek je obvykle uspořádán následujícím způsobem:

1. Úvod, se stručnou formulací účelu práce.
2. Přehled stavu znalostí o zkoumané problematice, např. přehled relevantních publikací (knihy, články v časopisech a sbornících konferencí, výzkumné zprávy).
3. Podrobný popis použitých metod a přístrojů.
4. Popis experimentů a měření.
5. Zpracování měření a analýza výsledků.
6. Závěry.
7. Soupis pramenů.

Někdy se na konec článku před soupis literatury zařazuje poděkování (Acknowledgment). Zde může autor vyjádřit poděkování za pomoc, za podporu z grantového projektu nebo z jiných zdrojů, a také za svolení publikovat některé výsledky nebo části převzaté z jiných prací; v takových případech je vždy nutno uvést pramen. Chceme-li převzít obrázek nebo data z jiného článku nebo knihy, musíme také získat svolení od držitele copyrightu a zmínit to v poděkování. Naše práce musí být také vždy prostá jakéhokoliv plagiátorství.

Pokud se týká podrobností úpravy článku pro uveřejnění ve vědeckém časopise, autor by se měl vždy seznámit se stylem obvyklým v daném časopise, včetně způsobu uvádění citací; každý časopis má svůj styl. Pokyny pro autory bývají na

webových stránkách. Pořadatelé konferencí na svých stránkách obvykle také uvádějí požadovaný vzor uspořádání příspěvku.

### Příklad – Návrh experimentů

Při konkrétním výzkumu zpravidla existuje více možností, jak získat požadované informace. Výzkumník musí rozhodnout, kterou metodu použije s ohledem na své zkušenosti, finance a čas, který má pro zkoumání k dispozici, vybavení které je dostupné “doma”, anebo které bude nutno koupit či pronajmout, a také s ohledem na požadavky na výsledky z hlediska jejich důležitosti a přesnosti. Rozmanitost možností experimentálního zkoumání můžeme ilustrovat na jednoduché úloze – zjištění technického stavu spalovacího motoru. Odpovídající informace lze získat například: 1) z výkonové charakteristiky (výkon, kroučící moment a měrná spotřeba paliva, zjištěné motorovou brzdou), nebo 2) z kompresního poměru a těsnosti spalovacího prostoru, nebo 3) z hluku a vibrací motoru nebo jeho částí, 4) ze spotřeby paliva a maziv, 5) ze stavu maziv (chemické složení, obsah kovových částic), 6) ze složení výfukových plynů (CO, NO<sub>x</sub>...), 7) z výkonu nutného na protáčení motoru naprázdno (charakterizuje mechanické ztráty), 8) z opotřebení válců (měřeného přímo nebo prostřednictvím kovových částic v oleji) aj. Jakákoliv z těchto metod může poskytnout více nebo méně relevantní informaci, a volba je na výzkumníkovi. Čtenář sám může zkusit navrhnout další metody.

### **Modely a simulace**

Častým úkolem při výzkumu je najít vhodný model studovaného procesu nebo objektu, nebo vyjádřit vliv důležitých faktorů na určitou vlastnost nebo jev. Na druhé straně, vlastnosti nebo jevy se často zkoumají právě prostřednictvím vhodného modelu nebo simulace. Pro lepší pochopení zde vysvětlíme některé pojmy z této oblasti.

**Model** je zjednodušený objekt nebo systém, který může pomoci při analýze problému, obvykle s nižšími náklady a v kratší době. Reprezentuje celý systém nebo jeho část a může být vytvořen buď v hmotné podobě (např. z kovu nebo dřeva) nebo v matematické podobě vhodné pro demonstrování jeho chování.

**Simulace** zahrnuje vystavení modelu různým vstupním podmínkám a pozorování jeho chování. Může pracovat s hmotnými modely vystavenými působení

skutečného prostředí, nebo s matematickými modely vystavenými matematickým rušivým vstupům, jež simulují očekávané podmínky.

Model může být popisný nebo prediktivní. **Popisný model** pomáhá pochopit objekt reálného světa, systém nebo jev (např. rozřezaný model spalovacího motoru). **Prediktivní model** pomáhá pochopit a předpovídat jeho chování.

Modely mohou být klasifikovány jako statické nebo dynamické, deterministické nebo pravděpodobnostní, a ikonické nebo analogové nebo symbolické. Vlastnosti **statických modelů** se s časem nemění, zatímco **dynamické modely** uvažují jevy měnící se s časem. **Deterministické modely** se užívají, jestliže výsledek vyšetřovaného jevu je vždy stejný. **Pravděpodobnostní modely** jsou nezbytné, pokud výsledek nebo hodnoty nastávají s určitou pravděpodobností. **Ikonický model** vypadá jako skutečná věc (například model letadla v měřítku pro zkoušky ve větrném tunelu). **Analogové modely** jsou takové, které se chovají jako reálné objekty nebo systémy; nemusí však stejně vypadat. Existuje mnoho analogií mezi fyzikálními jevy; známá je například membránová analogie pro zkoumání kroucení různých profilů na základě pozorování tvaru nafukované membrány podobného profilu. **Symbolické modely** jsou abstrakcemi důležitých kvantifikovatelných částí určitého systému. Matematická rovnice, popisující závislost výstupních parametrů na vstupních parametrech, je symbolický model. Rozlišujeme **teoretické modely**, které vycházejí ze všeobecně uznávaných přírodních zákonů, a **empirické modely**, které jsou přibližným matematickým popisem experimentálních údajů. Oba druhy modelů se souhrnně označují jako **matematické modely**.

Při matematickém modelování jsou části systému představovány idealizovanými prvky, které mají stejné podstatné charakteristiky jako daný systém, a jejichž chování lze popsat matematickými rovnicemi. Pouze nejjednodušší matematické modely mohou být studovány klasickými analytickými metodami. Počítače výrazně rozšířily použití matematického modelování. Numerické metody a snadnost, s jakou umožňují testování různých stavů vyšetřovaného objektu nebo systému učinily z **počítačového modelování** a simulací velice výkonné nástroje výzkumu a vývoje. Simulace provozu určitého systému prostřednictvím matematického modelu umožňuje získat informace s nižšími náklady a v kratší době, než kdyby se muselo experimentovat s reálnými objekty. Kromě toho existují situace, při kterých by experimentování nebylo možné, ať již kvůli nákladům,

bezpečnosti nebo času. Například piloti letadel trénují na letových simulátorech a operátoři atomových reaktorů se učí na reaktorových simulátorech.

Více o přípravě experimentů lze najít v literatuře, např. [1 – 4].

Než začneme vysvětlovat jednotlivé postupy experimentální analýzy, pokusíme se toto vážné téma trochu odlehčit. Tibor Dévényi v knížce [5] přirovnal vědeckou činnost k práci čtyřdobého motoru: 1. Sání = (stadium literatury), 2. Komprese (= provádění experimentů, vyhodnocování měření), 3. Jiskra a exploze (získání nápadu, potvrzení hypotézy, formulování závěrů), 4. Výfuk (= publikování výsledků). Podobnost je zřejmá. To však neznamená, že naše publikace by mohly být tak škodlivé, jako výfukové plyny.

## **Literatura ke kapitole 2**

1. Bernard, J.: Technický experiment. ČVUT, Praha, 1999. 74 s.
2. Montgomery, D. C.: Design and analysis of experiments. Wiley, New York, 2012 (8th edition). 730 s.
3. Kropáč, O.: Metody experimentálního výzkumu. ČVUT, Praha, 1979. 139 s.
4. Dieter, G. E.: Engineering design. 2nd Edition. McGraw-Hill, New York, 1991. 721 s.
5. Dévényi, T.: Kariéra Dr. Gézy Támhletoho aneb vědci a hlodavci. Mladá fronta, Praha, 1981. 184 s.

### 3. Chyby a kolísání měřených hodnot

Přesné hodnoty vyšetřovaných veličin jsou známy jen málokdy. Všeobecně lze rozlišit tři druhy hodnot [1, 2]:

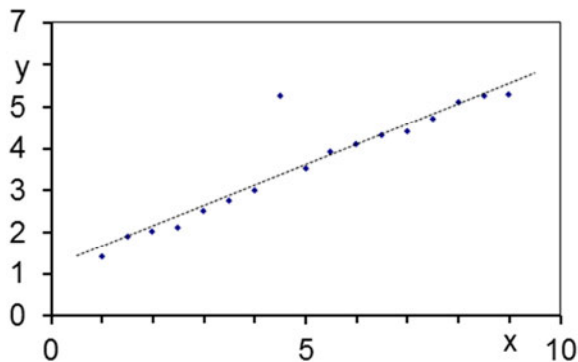
- skutečné,
- naměřené,
- použité v následujících výpočtech.

Naměřené hodnoty se často liší od hodnot přesných v důsledku různých chyb vznikajících při měření. Příčiny těchto chyb budou diskutovány v následujících odstavcích. Avšak i když je měření přesné, mohou naměřené hodnoty méně nebo více kolísat i když se zkoušky nebo měření opakují za prakticky stejných podmínek. Jednou příčinou je vnitřní proměnlivost měřené veličiny nebo jevu. Například pevnost křehkého materiálu se mění od vzorku ke vzorku v důsledku rozdílné velikosti materiálových defektů, které jsou rozhodující pro pevnost a lom jednotlivých vzorků. Také podmínky opakovaných zkoušek se mohou mírně měnit, například teplota nebo vlhkost okolí. Někdy je měření ovlivněno vibracemi nebo dalšími činiteli. A konečně, hodnoty použité v navazujících výpočtech se mohou lišit od naměřených, protože se v nich často užije průměrná hodnota nebo kvantil místo jednotlivých naměřených hodnot; příkladem je součinitel teplotní roztažnosti nebo jmenovitá pevnost materiálu.

Podívejme se nyní na chyby při měření. Můžeme rozlišit tři druhy chyb: hrubé, systematické a náhodné [3 – 5].

**Hrubá chyba** se projevuje jako hodnota, která je zřetelně mimo obvyklé rozmezí dalších hodnot (Obr. 3.1). Hrubé chyby vznikají v důsledku nepozornosti při odečítání naměřených hodnot, užitím nesprávného rozsahu měřicího přístroje, nebo technickou chybou. Mohou být odhaleny opakováním měření, vizuální kontrolou nakreslené řady hodnot, nebo statistickými testy tzv. odlehlých hodnot [6].

**Systematické chyby** vznikají trvalým vlivem skrytých faktorů (např. vyšší teplota v místě měření nebo nepřesnost měřicího přístroje). Tyto vlivy způsobují trvalé posunutí měřených hodnot, kladné nebo záporné. Nemohou být odhaleny pouhým opakováním měření, ale použitím jiné metody nebo jiných podmínek experimentu.



**Obr. 3.1.** Příklad odlehlé hodnoty (tzv. outlier).

Hlavní příčiny systematických chyb jsou:

- 1) nepřesnost našich smyslů (zrak, sluch), špatné psychické rozpoložení (práce ve stresu nebo spěchu, únava, vyčerpání),
- 2) nepřesnost měřicích přístrojů a metod,
- 3) nemožnost docílit vhodné podmínky experimentu (teplota, tlak, vlhkost, žádné parazitní vibrace),
- 4) měření samotné může ovlivňovat měřenou veličinu (příklady: relativně těžký snímač připevněný k lehké součásti změní její dynamickou charakteristiku, elektrický proud může zvýšit teplotu tenzometrického snímače, a tedy i jeho odpor),
- 5) Nevhodná metoda nebo aproximace použitá při zpracování naměřených hodnot (například regresní funkce užitá pro širší interval než byl interval, ve kterém se prováděla měření pro stanovení jejích konstant). Nízké počty číslic ve výpočtech, protože chyby v důsledku zaokrouhlování se při zřetězených výpočtech sečítají (blíže viz konec kapitoly).

Systematické chyby vyloučit následujícími způsoby:

- Duševní pohoda personálu, bez stresu během experimentů a měření.
- Použití dostatečně přesných přístrojů. Orientační pravidlo říká, že chyba měřicího přístroje by měla být alespoň desetkrát menší než přípustná chyba ve stanovení měřené veličiny. Na příklad, má-li tloušťka určité součásti být zjištěna s přesností 0,01 mm, je nutné použít měřidlo s přesností ne horší než 0,001 mm. Důležité přístroje je nutno čas od času kalibrovat.



- Jednotlivé prvky v měřicím řetězci „snímač - spojovací kabely - zesilovač - měřicí přístroj...“ jsou uspořádány v sérii a jejich chyby a nepřesnosti se sečítají. Nejúčinnější cesta ke zvýšení přesnosti je nahradit člen s nejnižší přesností přesnějším. Výzkumný pracovník by proto měl znát přesnosti použitých přístrojů. U dynamických problémů je nutno užívat přístroje s vhodnými dynamickými charakteristikami.
- Vyloučení nežádoucích vlivů vhodným uspořádáním experimentu (například prováděním všech měření při stejné teplotě).
- Eliminace nežádoucích vlivů přepočtem naměřených hodnot pomocí korekčních faktorů (např. pro teplotu).
- Stálé vyvažování experimentu (např. užití Wheatstoneova můstku).
- Znáhodnění experimentů, tj. použití náhodných kombinací hodnot jednotlivých vstupních veličin při posloupnosti sérií testů.
- Použití dostatečně vysokého počtu číslic, zvláště pokud se naměřené hodnoty dále zpracovávají (viz níže).

**Náhodné chyby.** Tyto chyby jsou způsobeny náhodnými vlivy, které jsou mimo naši kontrolu. Jejich velikost od jednoho pokusu ke druhému kolísá. Lze je odhalit opakováním experimentu, a opakování se také užívá ke snížení jejich vlivu. Zde se zlepšení dosáhne metodami matematické statistiky, například stanovením intervalu spolehlivosti, který obsahuje skutečnou hodnotu s vysokou pravděpodobností. Více k této otázce lze najít v následující kapitole a v kapitolách 7 a 8.

Na tomto místě je vhodné říci několik slov na téma **optimální počet číslic při zpracování naměřených hodnot**. Říkáme, že číslo má  $n$  **významných číslic**, jestliže jeho absolutní chyba není větší než polovina řádu  $n$ -té číslice. Jestliže jakákoliv vstupní veličina má  $n$  významných číslic, pak nanejvýš  $n$  číslic bude významných v konečném výsledku. Více číslic výsledku již nezvýší jeho přesnost. Má-li výsledek mít  $n$  významných číslic, je nutno provádět všechny mezisoučty s minimálně  $n + 1$  číslicemi; výsledek je pak zaokrouhlen na  $n$  číslic. Při násobení a dělení se jednotlivé činitele zaokrouhlují tak, aby měly alespoň o jednu číslici více než činitel s nejmenším počtem významných číslic (tj. s největší relativní chybou). Více k této problematice lze najít např. v [7, 8].

### Literatura ke kapitole 3

1. Pechoč, V.: Vyhodnocování měření a početní metody v chemickém inženýrství. SNTL, Praha, 1981. 226 s.
2. Bernard, J.: Technický experiment. ČVUT, Praha, 1999. 74 s.
3. Příručka měřicí techniky pro strojírenství a energetiku. SNTL, Praha, 1965. 928 s.
4. Jenčík, J., Kuhn, L.: Technická měření ve strojnictví. SNTL, Praha, 1982. 584 s.
5. Taylor, J. R.: An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements. University Science Books, Herndon, 1997. 327 s.
6. Kupka, K.: Statistické řízení jakosti. Trilobyte, Pardubice, 1997. 191 s.
7. Nektivinda, M., Šrubař, J., Vild, J.: Úvod do numerické matematiky. SNTL, Praha, 1976. 288 s.
8. Děmidovič, B. P., Maron, I. A.: Základy numerické matematiky. SNTL, Praha, 1966. 721 s.

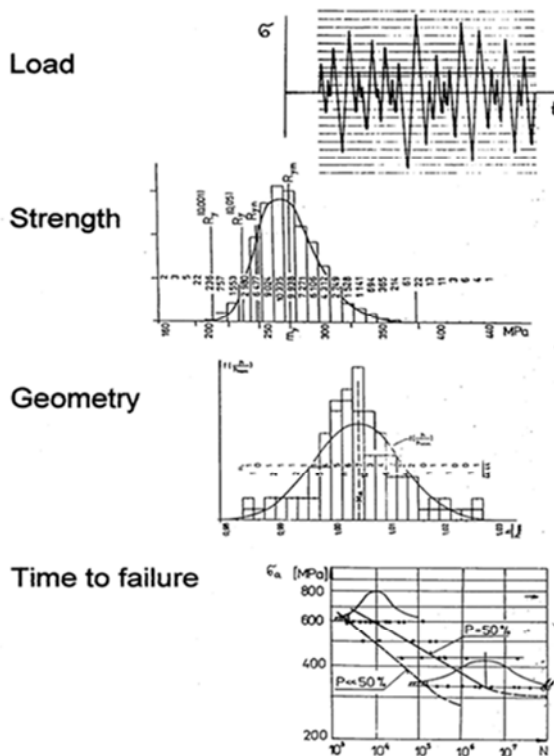
## 4. Základy pravděpodobnosti a statistiky pro experimentální analýzu

Výskyt různých jevů i hodnoty vyšetřovaných veličin jsou často doprovázeny nejistotou. To je v důsledku činitelů, které nemůžeme ovlivnit, a proto je označujeme jako náhodné. Pro lepší práci s nimi užíváme pojem pravděpodobnost a statistické metody. Odpovídající postupy pomáhají při řešení mnoha problémů. Protože počítače mohou udělat všechnu namáhavou práci, jediná věc, kterou uživatel pravděpodobnostních metod potřebuje, je pochopení základních pojmů a procedur. V této kapitole jsou vysvětleny hlavní pojmy, jako je náhodná veličina, pravděpodobnost, základní a výběrový soubor, průměr, střední hodnota, rozptyl, směrodatná odchylka, variační koeficient, rozdělení pravděpodobnosti, distribuční funkce, hustota pravděpodobnosti, kvantil, kritická hodnota, konfidenční interval a testování hypotéz. Jsou ukázána i důležitá rozdělení pravděpodobnosti. Další podrobnosti je možno najít ve statistické literatuře, například [1 – 5].

**Pravděpodobnost** je kvantitativní míra možnosti, že nastane náhodný jev. Klasická definice pravděpodobnosti  $P$  vychází z četnosti výskytu určité události při mnohonásobném opakování pokusu nebo pozorování:

$$P \approx n / N \quad (4.1)$$

kde  $N$  je celkový počet pokusů (předpokládaný velmi vysoký,  $N \rightarrow \infty$ ) a  $n$  je počet výskytů určitého jevu, například viditelná určitá strana mince po dopadu, počet dnů s maximální teplotou vyšší než  $20^\circ\text{C}$ , nebo počet vadných součástí v dávce. Pravděpodobnost je bezrozměrová veličina, která může nabývat hodnot mezi 0 a 1; nula značí nemožný jev a 1 značí jev jistý. **Náhodná veličina** je veličina, která může nabývat různých hodnot s určitými pravděpodobnostmi. Náhodné veličiny jsou **nespojité** (diskrétní) nebo **spojité**. Příkladem nespojité náhodné veličiny je počet mrtvých při dopravní nehodě nebo počet zatěžovacích cyklů stroje do poruchy. Spojitá náhodná veličina může nabýt jakékoliv hodnoty (v určitém intervalu), například pevnost materiálu, rychlost větru, délka, hmotnost, doba do poruchy, trvání opravy nebo pravděpodobnost poruchy. Některé příklady jsou ukázány na Obr. 4.1.



**Obr 4.1.** Příklady náhodných veličin [6]. Load – zatížení, Strength – pevnost, Geometry – geometrie, Time to failure – čas do poruchy.

Náhodné veličiny mohou být popsány pomocí **rozdělení pravděpodobnosti** nebo pomocí jednoduchých čísel, která se nazývají **parametry**, pokud se vztahují k **základnímu souboru** (= množina všech hodnot nebo prvků vyšetřované veličiny), anebo **charakteristiky**, jestliže jsou počítány pouze ze vzorku omezeného rozsahu, tzv. **výběru**. Parametry se obvykle označují řeckými písmeny a charakteristiky latinkou.

### Popis pomocí parametrů

Hlavní parametry (resp. charakteristiky) náhodných veličin jsou uvedeny níže, spolu se vzorci pro výpočet ze vzorků omezené velikosti.

**Střední hodnota  $\mu$**  (resp. **průměr  $\bar{x}$** ) charakterizuje polohu náhodné veličiny na číselné ose; odpovídá tak jejímu těžišti,

$$\mu = \int_{-\infty}^{+\infty} x f(x) dx, \quad \bar{x} = \frac{\sum x_j}{n} \quad (4.2)$$

$x_j$  je  $j$ -tá hodnota;  $n$  – velikost vzorku.

**Rozptyl**  $\sigma^2$  (nebo  $s^2$ ) – charakterizuje rozptýlení dané veličiny a počítá se jako

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx; \quad s^2 = \frac{\sum (x_j - \bar{x})^2}{n - 1} \quad (4.3)$$

**Směrodatná odchylka**  $\sigma$  (nebo  $s$ ) je definována jako druhá odmocnina z rozptylu,

$$\sigma = \sqrt{\sigma^2}, \quad s = \sqrt{\frac{\sum (x_j - \bar{x})^2}{n - 1}} \quad (4.4)$$

Má stejný rozměr jako vyšetřovaná veličina  $x$  a je proto užívána pro charakterizování rozptýlení častěji než rozptyl podle vztahu (4.3).

**Variační koeficient**  $v$  charakterizuje rozptýlení ve srovnání se střední hodnotou,

$$v = \frac{s}{\bar{x}} \quad (4.5)$$

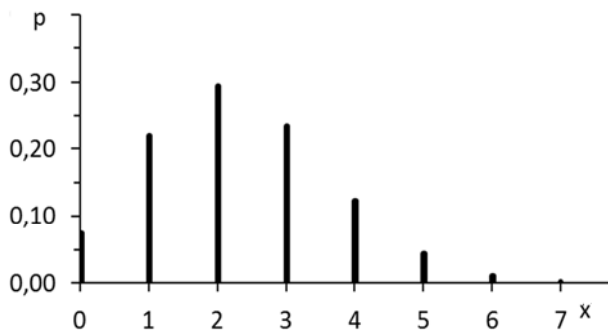
Může se používat např. pro porovnávání náhodné proměnlivosti různých veličin.

Nevýhodou průměrné hodnoty  $\bar{x}$  je její citlivost na extrémní hodnoty ve vyšetřovaném souboru; přidání velmi vysoké nebo velmi nízké hodnoty může způsobit významnou změnu průměru. Méně citlivou (tj. robustní) charakteristikou středu řady hodnot je **medián**  $m$ . To je hodnota uprostřed řady hodnot seřazených od nejmenší do největší; například  $m = 4$  pro řadu 2, 6, 1, 8, 10, 4, 3.

### Popis pomocí rozdělení pravděpodobnosti

Podrobnější informace se získá z **rozdělení pravděpodobnosti**, které ukazuje, jak je náhodná proměnná rozložena podél číselné osy. Pro diskrétní veličiny se užívá **pravděpodobnostní funkce**  $p(x)$ , Obr. 4.2, která udává pravděpodobnosti, se kterými náhodná proměnná  $x$  nabývá jednotlivých hodnot  $x^*$ ,

$$p(x^*) = P(x = x^*) \quad (4.6)$$



**Obr. 4.2.** Pravděpodobnostní funkce. Příklad: Binomické rozdělení, parametr  $p = 0.23$ ;  $n = 10$  [6].

**Hustota pravděpodobnosti  $f(x)$**  se užívá pro spojité veličiny a ukazuje, kde se tato veličina vyskytuje více nebo méně často (Obr. 4.3 nahoře). Matematicky vyjadřuje pravděpodobnost, že proměnná  $x$  bude ležet uvnitř nekonečně úzkého intervalu mezi  $x^*$  a  $x^* + dx$ .

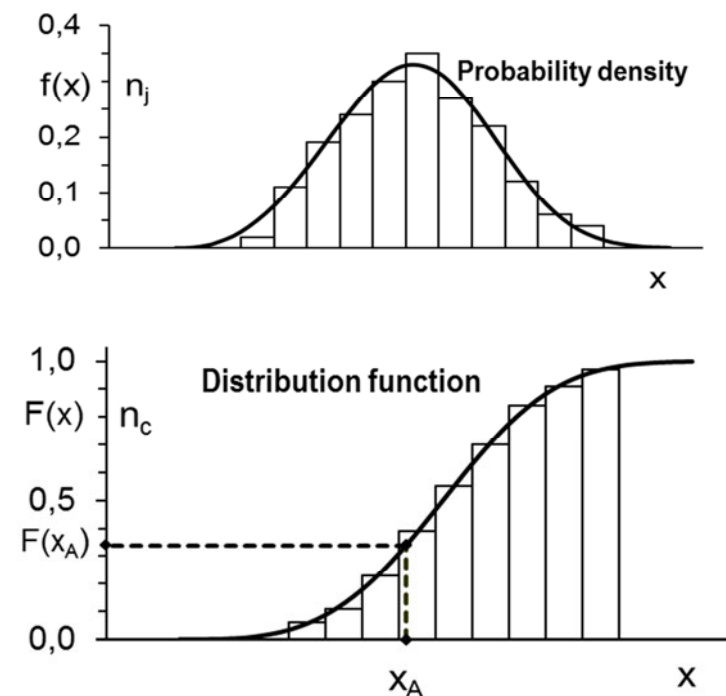
**Distribuční funkce  $F(x)$**  se používá pro diskrétní i spojité veličiny (Obr. 4.3 dole) a vyjadřuje pravděpodobnost, s jakou náhodná veličina  $x$  nabude hodnoty menší nebo rovné  $x^*$ ,

$$F(x^*) = P(x \leq x^*) \quad (4.7)$$

Obě veličiny jsou navzájem spojeny relacemi

$$f(x) = dF/dx, \quad F(x) = \int_{-\infty}^x f(x) dx, \quad \text{popř.} \quad F(x) = \sum_{i=1}^n p(x_i). \quad (4.8)$$

Obrázek 4.3 ukazuje dvě možnosti zobrazování těchto funkcí: prostřednictvím histogramů nebo analytických výrazů. **Histogram** (sloupkový diagram) se získá tak, že se rozmezí všech možných hodnot rozdělí do několika intervalů stejné šířky, stanoví se počty hodnot v každém intervalu a nakreslí se obdélníky o výškách úměrných těmto počtům. Aby výsledky byly obecnější, četnosti výskytů v jednotlivých intervalech se vydělí celkovým počtem všech hodnot. Tak se získají **relativní četnosti** a **relativní kumulativní četnosti**, které přibližně odpovídají hustotě pravděpodobnosti a distribuční funkci. Stanovení těchto charakteristik z empirických hodnot bude podrobně vysvětleno v Kapitole 5.



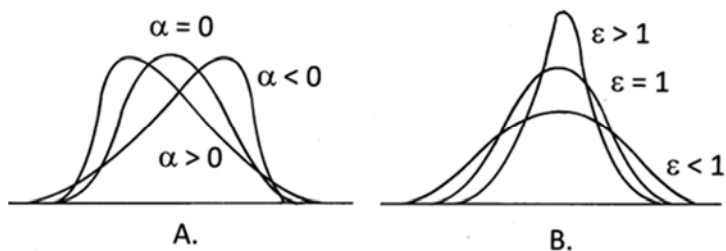
**Obr. 4.3.** Hustota pravděpodobnosti  $f(x)$  a distribuční funkce  $F(x)$  spojité veličiny [6]. Histogramy ukazují relativní četnost ( $n_j$ ) a relativní kumulativní četnost ( $n_{c,j}$ ).

Pravděpodobnost určitého jevu (např. výška  $x$  sněhové pokrývky menší než  $x_\alpha$ ) je možno určit jako odpovídající plochu pod křivkou  $f(x)$  od  $-\infty$  do  $x_\alpha$ , nebo přímo jako hodnotu  $F(x_\alpha)$  distribuční funkce.

**POZNÁMKA:** Pravděpodobnost je bezrozměrová, ale hustota pravděpodobnosti má rozměr, rovný převrácené hodnotě rozměru uvažované veličiny, např.  $\text{m}^{-1}$ ,  $\text{MPa}^{-1}$  nebo  $\text{K}^{-1}$ . Například délka  $L$  určité součásti je rozložena tak, že  $F(L \leq 2.00 \text{ m}) = 0.45$  a  $f(L = 3.00 \text{ m}) = 0.026 \text{ m}^{-1}$ .

**Tvar** rozdělení pravděpodobnosti je rychle charakterizován následujícími dvěma veličinami: šikmostí a špičatostí. **Šikmost**  $\alpha$  (koeficient asymetrie) informuje, jestli je rozdělení symetrické ( $\alpha = 0$ ) nebo prodloužené směrem doprava ( $\alpha > 0$ ) nebo doleva ( $\alpha < 0$ ). **Špičatost**  $\varepsilon$  informuje, zdali je rozdělení ostřejší ( $\varepsilon > 0$ ) nebo

tupější ( $\varepsilon < 0$ ) než normální rozdělení ( $\varepsilon = 0$ ). Obě veličiny jsou znázorněny schematicky na Obr. 4.4; jejich definice lze najít v učebnicích statistiky [1 – 5].



**Obr. 4.4.** Šikmost  $\alpha$  a špičatost  $\varepsilon$  rozdělení pravděpodobnosti.

Velmi důležité jsou také následující dvě veličiny. **Kvantil**  $x_\alpha$  je taková hodnota náhodné veličiny, že pravděpodobnost, že  $x$  bude menší nebo rovna  $x_\alpha$ , je pouze  $\alpha$ ,

$$P(x \leq x_\alpha) = \alpha \quad (4.9)$$

Kvantily jsou inverzní k hodnotám distribuční funkce. Na Obr. 4.3 je vyobrazen  $\alpha$ -kvantil, který odpovídá pravděpodobnosti  $F(x_\alpha)$ ,

$$x_\alpha = F^{-1}(\alpha) \quad (4.10)$$

Kvantily se užívají pro stanovení „zaručené“ nebo „bezpečné“ hodnoty určité veličiny, jako je například minimální očekávatelná pevnost nebo doba do poruchy.

**Kritická hodnota**  $x^\beta$  je taková hodnota náhodné veličiny  $x$ , že pravděpodobnost jejího překročení je pouze  $\beta$ ,

$$P(x > x^\beta) = \beta \quad (4.11)$$

(V tomto případě  $\beta$  neznamena exponent, ale hořejší index !)

Kritické hodnoty se užívají pro stanovení maximální očekávatelné hodnoty určité veličiny, jako je např. rychlost větru nebo výška sněhu v určité oblasti. Užívají se také při testování hypotéz, například jestli dva vzorky pocházejí se stejného základního souboru. Pravděpodobnost  $\beta$  je doplňková k  $\alpha$ , to znamená  $\beta = 1 - \alpha$ , a

$$x^\beta = x_{1-\alpha}, \quad x_\alpha = x^{1-\beta} \quad (4.12)$$

Více o základních definicích a pravidlech pro počet pravděpodobnosti lze najít v [1 – 5].



## Důležitá rozdělení pravděpodobnosti

Existují některá rozdělení, která jsou velmi důležitá. Pro nespojité veličiny to je binomické a Poissonovo rozdělení. Hlavní rozdělení pro spojité veličiny jsou normální, logaritmicko-normální, Weibullovo a exponenciální. K některým účelům se užívá rovnoměrné rozdělení, chí-square ( $\chi^2$ ) rozdělení, Studentovo  $t$ -rozdělení a  $F$ -rozdělení. Jejich stručné popisy následují; podrobnosti lze najít v literatuře, např. [1 – 5].

**Binomické rozdělení** (Obr. 4.2) vyjadřuje pravděpodobnost výskytu  $x$  příznivých událostí v  $n$  pokusech, jestliže tato pravděpodobnost v každém pokusu je rovna  $p$ . Příkladem je počet  $x$  vadných kusů ve výběru o velikosti  $n$ , jestliže jejich podíl v celém základním souboru je  $p$ . Pravděpodobnostní funkce je

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad (4.13)$$

a střední hodnota je  $\mu = np$ . Toto rozdělení je diskrétní a má pouze jeden parametr  $p$ , který lze určit z celkového počtu  $m$  příznivých výsledků v  $n$  pokusech jako  $p = m/n$ ;  $n$  by mělo být hodně vysoké.

**Poissonovo rozdělení** je podobné binomickému, ale je vhodnější pro vzácné jevy s nízkými pravděpodobnostmi výskytu  $p$ . Pravděpodobnostní funkce, udávající pravděpodobnost výskytu  $x$  příznivých výsledků v  $n$  pokusech je

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (4.14)$$

kde  $\lambda$  parametr rozdělení. ( $\lambda$  odpovídá průměrnému výskytu  $x$  a také součinu  $np$  u binomického rozdělení.)

**Normální rozdělení**, nazývané také Gaussovo rozdělení, má symetrický zvonovitý tvar (Obr. 4.3 a 4.5). Užívá se velmi často pro spojité veličiny, obzvláště je-li kolísání způsobeno mnoha náhodnými vlivy a rozptyl není příliš velký. Hustota pravděpodobnosti je

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad (4.15)$$

se střední hodnotou  $\mu$  a směrodatnou odchylkou  $\sigma$  jako parametry. Pro distribuční funkci  $F(x)$  neexistuje explicitní vzorec; musí se počítat jako integrál z hustoty pravděpodobnosti; viz vzorec (4.8). V praxi se pro výpočet  $F$  užívají různé přibližné vztahy; viz např. [7].

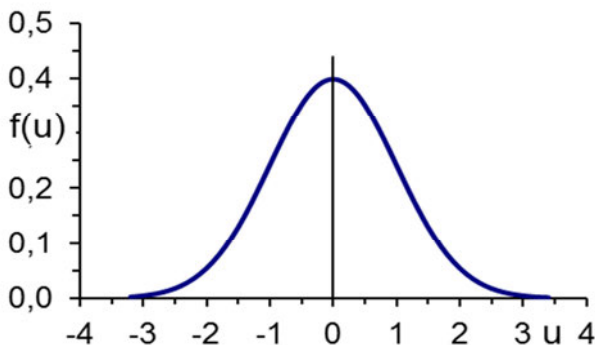
**Normované normální rozdělení** odpovídá normálnímu rozdělení s parametry  $\mu = 0$  a  $\sigma = 1$  (Obr. 4.5). Vztah pro hustotu pravděpodobnosti se obvykle uvádí ve tvaru

$$f(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) \quad (4.16)$$

$u$  je normovaná proměnná, která souvisí s proměnnou  $x$  normálního rozdělení vztahem

$$u = (x - \mu)/\sigma \quad (4.17)$$

Vyjadřuje vzdálenost  $x$  od střední hodnoty  $\mu$  v násobcích směrodatné odchylky. Je užitečné si pamatovat, že 68.27% všech hodnot veličiny s normálním rozdělením leží uvnitř intervalu  $\mu \pm \sigma$ , 95.45% leží uvnitř  $\mu \pm 2\sigma$ , a 99.73% (tj. „téměř“ všechny) uvnitř intervalu  $\mu \pm 3\sigma$ .



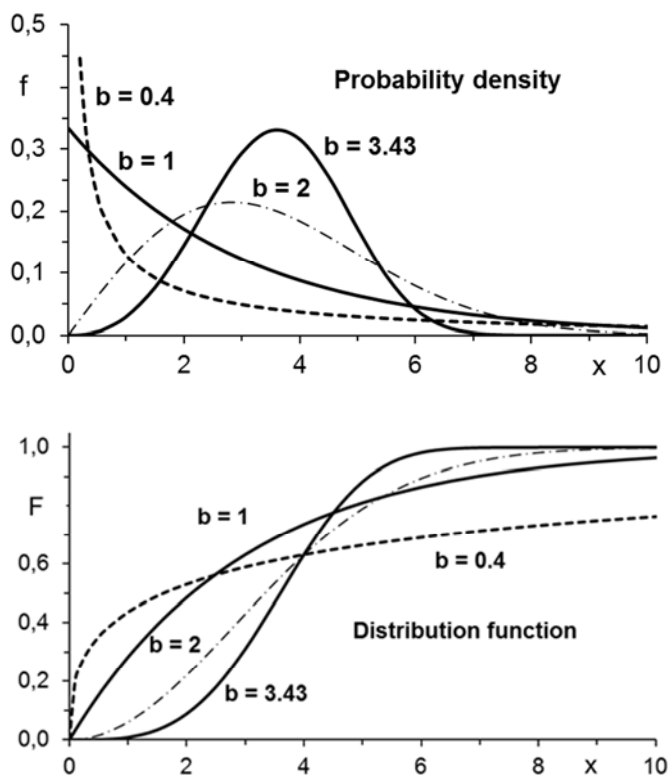
**Obr. 4.5.** Normované normální rozdělení ( $\mu = 0$ ,  $\sigma = 1$ ).

**Logaritmicko-normální rozdělení** (označované též lognormální) je nesymetrické (protažené směrem vpravo, podobné Weibullovo rozdělení s  $b = 2$  na Obr. 4.6) a odpovídá případu, jestliže logaritmus náhodné veličiny má normální rozdělení.

**Weibullovo rozdělení** (Obr. 4.6) má distribuční funkci

$$F(x) = 1 - \exp\{-[(x - x_0)/a]^b\} \quad (4.18)$$

se třemi parametry: parametr měřítka  $a$ , parametr tvaru  $b$ , and prahová hodnota  $x_0$ , která odpovídá minimální možné hodnotě  $x$ . Vztah pro hustotu pravděpodobnosti  $f(x)$  lze získat snadno jako derivaci distribuční funkce. Weibullovo rozdělení je velmi pružné díky parametru tvaru  $b$  (Obr. 4.6). Často se užívá pro aproximaci pevnosti nebo doby do poruchy, ale i třeba rozdělení průměru sazí. Patří mezi tzv. **rozdělení extrémních hodnot** [8], a vyskytuje se například, je-li porucha objektu způsobena jeho nejslabším místem. Stanovení parametrů tohoto velmi důležitého rozdělení z empirických dat bude vysvětleno v kapitole 5.



**Obr. 4.6.** Weibullovo rozdělení pro různé hodnoty parametru tvaru  $b$  [6]. Probability density - hustota pravděpodobnosti, Distribution function - distribuční funkce

Kromě flexibility má Weibullovo rozdělení zvláštní výhodu při posuzování spolehlivosti. Parametr tvaru  $b$  ve výrazu (4.18) souvisí s charakterem poruch a všeobecně informuje o etapě života vyšetřovaného objektu. Hodnoty  $b < 1$  jsou

typické pro období časných poruch, zatímco  $b > 1$  odpovídá období stárnutí a dožívání. Hodnota  $b \approx 1$  charakterizuje období užitečného života, s poruchami z mnoha různých příčin.

**Exponenciální rozdělení** je zvláštním případem Weibullova rozdělení (4.18) pro parametr tvaru  $b = 1$  (viz Obr. 4.6), s distribuční funkcí

$$F(t) = 1 - \exp[-(t/T_0)] \quad (4.19)$$

Používá se například pro doby mezi poruchami  $t$ , vznikajícími z mnoha různých příčin, např. ve složitých systémech sestávajících z velkého počtu součástí. Toto rozdělení má jenom jeden parametr, zde  $T_0$ , který odpovídá střední hodnotě  $\mu$  a má stejnou hodnotu jako směrodatná odchylka  $\sigma$ . (Poznámka: v případě času byl zde užít symbol  $t$  místo  $x$ ; rozdíl je pouze formální. Kromě toho, minimální hodnota  $x_0$  se často předpokládá rovna nule.)

Následující čtyři rozdělení jsou důležitá zejména pro vytváření konfidenčních intervalů, pro statistické testy a pro simulaci metodou Monte Carlo, která bude vysvětlena v kapitole 14.

**Rovnoměrné rozdělení** má konstantní hustotu pravděpodobnosti ( $f = konst$ ) v definičním intervalu  $\langle a, b \rangle$ , takže vypadá jako obdélník. Střední hodnota odpovídá průměru obou mezí,  $\mu = (a + b)/2$ , a rozptyl je  $\sigma^2 = (b - a)^2/12$ .

**Rozdělení chí-kvadrát ( $\chi^2$ )** je rozdělením součtu  $n$  náhodných veličin, kde každá je definována jako čtverec normované normální proměnné. Důležitým parametrem je počet stupňů volnosti, který je v tomto případě roven  $n$ . Blíže viz např. [1 – 5].

**Rozdělení  $t$** , nazývané také Studentovo, má veličina vytvořená jako poměr čísel s normovaným normálním rozdělením a čísel s rozdělením chí-kvadrát. Vypadá podobně jako normované normální rozdělení (Obr. 4.5), ale je nižší a širší, obzvláště pro menší počty stupňů volnosti; viz [1 – 5]. Užívá se např. pro tvorbu konfidenčních intervalů (viz dále).

**$F$ -rozdělení** odpovídá poměru dvou veličin s rozdělením chí-kvadrát a užívá se pro porovnávání dvou rozptylů. Toto rozdělení závisí na počtu stupňů volnosti každé proměnné [1 – 5].

Rychlou zhrůnou informaci lze získat z Wikipedie a zdrojů citovaných v ní. Hodnoty distribučních funkcí a kvantilů zmíněných rozdělení je možné najít ve

speciálních tabulkách nebo ve statistických či univerzálních programech, včetně Excelu.

Nyní vysvětlíme dva důležité pojmy matematické statistiky.

**Konfidenční interval**, neboli interval spolehlivosti. Důsledkem náhodné proměnlivosti vyšetřované veličiny je, že každé měření a následující výpočty dávají rozdílné hodnoty v závislosti na užitém vzorku. V takovém případě se zpravidla z naměřených hodnot stanovuje aritmetický průměr  $\bar{x} = \Sigma x_i/n$ . Ten ale neříká nic o tom, jak daleko je od skutečné střední hodnoty  $\mu$ . Z tohoto důvodu se často určuje konfidenční interval, který s vysokou pravděpodobností obsahuje skutečnou hodnotu  $\mu$ . Na příklad,  $\alpha$ -konfidenční interval pro střední hodnotu je

$$\bar{x} - t_{\alpha, n-1} \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha, n-1} \frac{s}{\sqrt{n}} \quad (4.20)$$

$\bar{x}$  a  $s$  jsou průměr a směrodatná odchylka vzorku o  $n$  hodnotách, a  $t_{\alpha, n-1}$  je tzv.  $\alpha$ -kritická hodnota oboustranného  $t$ -rozdělení pro  $n - 1$  stupňů volnosti. Pravděpodobnost, že skutečná střední hodnota  $\mu$  bude ležet uvnitř intervalu (4.20), je  $1 - \alpha$ , a  $\alpha$ , že se bude nacházet vně. (Praktická aplikace bude ukázána později.) Intervaly spolehlivosti lze také sestřojovat pro jiná rozdělení; blíže viz [1 – 5].

**POZNÁMKA:** Existují také jednostranné kritické hodnoty. Takováto hodnota ( $\alpha'$ ) odpovídá pravděpodobnosti, že  $\mu$  bude buď (jenom) vyšší nebo (jenom) nižší než příslušná kritická hodnota. Vztah mezi  $\alpha'$  a veličinou  $\alpha$  je  $\alpha' = \alpha/2$ . Při použití statistických tabulek nebo počítačových nástrojů musí člověk proto vědět, jak byla příslušná veličina definována.

**Testování hypotéz.** Často musíme rozhodnout, který ze dvou druhů výrobků je lepší, anebo jestli existuje skutečný rozdíl mezi dvěma skupinami naměřených hodnot. Takové rozhodnutí může být založeno na porovnání hodnot charakteristického parametru, například průměru. Obvykle ale jednotlivé naměřené hodnoty kolísají, takže existují i rozdíly mezi vypočtenými charakteristikami. Je-li tento rozdíl velký, je rozhodnutí snadné. V opačném případě musíme také přihlídnout k tomu, že je částečně způsoben náhodnými příčinami. Ke spolehlivějšímu rozhodování se užívají statistické testy, které mohou odhalit, jestli rozdíly mezi příslušnými charakteristikami porovnávaných souborů jsou pouze náhodné, anebo jestli odrážejí skutečné rozdíly mezi oběma základními soubory

(např. mezi oběma typy výrobků). Takovéto testy sestávají z několika kroků. V prvním kroku je formulována takzvaná **nulová hypotéza**: „**mezi oběma soubory neexistuje žádný rozdíl**“. (**Alternativní hypotéza** je „**mezi oběma soubory existuje významný rozdíl**“.) Ve druhém kroku se ze statistických charakteristik obou vzorků vypočítá hodnota testového kritéria; forma tohoto kritéria záleží na druhu problému; příslušné vzorce lze najít ve statistické literatuře [1 – 5]. Je-li nulová hypotéza platná, mají hodnoty testového kritéria určité rozdělení. Ve třetím kroku porovnáme vypočtenou hodnotu kritéria s *kritickou hodnotou* tohoto rozdělení. Je-li vypočtená hodnota vyšší než kritická hodnota odpovídající nízké pravděpodobnosti  $\alpha$ , nastal jev, který byl očekáván pouze s touto velmi malou pravděpodobností (např. 5% nebo 1%). Učiníme závěr, že rozdíl mezi oběma soubory není náhodný; nulovou hypotézu odmítneme na hladině významnosti  $\alpha$ . Jestliže vypočítaná hodnota kritéria je menší než kritická, přijímáme nulovou hypotézu a činíme závěr, že není rozdíl mezi oběma soubory. Také říkáme, že rozdíl mezi nimi není statisticky významný. Z tohoto hlediska je důležité, jakou pravděpodobnost  $\alpha$  považujeme za významnou; to je záležitostí naší volby.

**POZNÁMKA.** Při tomto testu existuje pravděpodobnost  $\alpha$ , že náš závěr „oba soubory se liší“, založený na zamítnutí nulové hypotézy, je chybný, a ve skutečnosti mezi nimi neexistuje rozdíl. Toto je takzvaná chyba prvního druhu. Jestliže nulová hypotéza nebyla zamítnuta, existuje opačné riziko, že se oba soubory ve skutečnosti liší, což by byla chyba druhého druhu. Pravděpodobnost tohoto chybného závěru je  $\beta$ . Vyšší důvěra ve správné zamítnutí nulové hypotézy (tedy nižší  $\alpha$ ) také znamená vyšší riziko přijmutí alternativní hypotézy. Obvykle je nutno přijmout určitý kompromis.

Testy hypotéz jsou podrobně vysvětleny v literatuře, např. [1 – 5], a jsou také k dispozici v různých statistických nebo univerzálních počítačových programech. Také Excel nabízí některé testy, např. pro rozdíl mezi dvěma středními hodnotami nebo mezi rozptyly dvou souborů. Aplikace budou ukázány v Kapitole 8.

## Pořadová statistika

Častým úkolem při experimentální analýze je najít pro skupinu naměřených hodnot (např. pevnost nebo doba do poruchy) parametry rozdělení pravděpodobnosti nebo určitý kvantil či hodnotu distribuční funkce. V některých případech to je prosté; například parametry normálního rozdělení jsou průměr a směrodatná odchylka.

Někdy je to ale složitější, např. u Weibullova tříparametrického rozdělení, anebo jestliže histogram naměřených hodnot má složitější tvar. Naštěstí je v takových případech poměrně snadné naměřeným hodnotám přiřadit hodnoty distribuční funkce, a to následujícím jednoduchým způsobem. Nejprve seřadíme naměřené hodnoty od nejmenší ( $j = 1$ ) do největší ( $j = n$ );  $j$  je pořadové číslo a  $n$  je celkový počet hodnot. Odpovídající hodnoty distribuční funkce se obvykle vypočítají jako

$$F_j = j / (n + 1) \quad (4.21)$$

Vysvětlení vzorce (4.21) je prosté. Máme-li např. 100 hodnot dob do poruchy  $t$  a uspořádáme je od nejkratší do nejdelší, potom pravděpodobnost  $F$ , že  $t$  bude menší nebo rovno nejnižší z oněch 100 hodnot,  $t_1$ , je přibližně 1 : 100. Pravděpodobnost, že  $t \leq t_2$  je 2/100, atd.; všeobecně  $F_j = j/n$ . V rovnici (4.21) byla ke jmenovateli přičtena jednotka kvůli matematické korektnosti: pravděpodobnost  $F$ , že  $t$  bude kratší nebo rovna  $t_n$ , musí být menší než 1, protože při provedení dalších měření by se mohly objevit i hodnoty vyšší než  $t_n$ .

POZNÁMKA: Existují i další empirické vzorce pro výpočet hodnot  $F_j$ , například  $F_j = (j - 1/2)/n$ . Žádný vzorec ale nelze doporučit jednoznačně, zvláště přihlédneme-li k tomu, že při určování parametrů rozdělení mohou vzniknout větší chyby v důsledku toho, že hodnoty ve výběrovém souboru, ze kterých se parametry počítají, mohou značně kolísat od výběru k výběru (zejména u malých rozsahů), než v důsledku vzorce užitého pro výpočet  $F_j$ .

### Neparametrické metody

Statistické testy a postupy obvykle předpokládají, že vyšetřovaná veličina má určité rozdělení pravděpodobnosti, a pracují s jeho parametry. Existují ale i tzv. neparametrické metody [9, 10], které nevyžadují žádný předpoklad o rozdělení, ani znalost jeho parametrů. Tyto metody lze užívat, jestliže schází jakákoliv informace o rozdělení pravděpodobnosti. Na druhou stranu obvykle potřebují větší soubor o větším rozsahu  $n$ , aby poskytly stejnou sílu informace jako metody parametrické.

Nejdůležitější neparametrické metody, aplikovatelné při experimentálním výzkumu jsou: (1) určování kvantilů, (2) testy dobré shody, používané pro ověření, má-li soubor určité rozdělení, (3) testy pro ověření, zda dva výběry pocházejí ze stejného základního souboru, a (4) testy korelace mezi dvěma veličinami. Tyto testy obvykle pracují s uspořádanými (seřazenými) hodnotami. V následujícím odstavci

bude vysvětlena první z metod, která se užívá velmi často. Další metody budou vysvětleny později.

## Kvantily

V předchozím odstavci bylo vysvětleno přiřazování hodnot  $F_j$  distribuční funkce jednotlivým hodnotám  $y_j$ , seřazeným podle velikosti. Nalezení  $\alpha$ -kvantilu  $y$  je opačný úkol; je to taková hodnota  $y_\alpha$  řady čísel  $y$ , která odpovídá hodnotě  $\alpha$  distribuční funkce  $F$ . Není-li přesná hodnota  $F = \alpha$  známa, lze ji nalézt ze sousední menší a větší hodnoty  $F$  interpolací. Kvantil je také možno najít interpolací ze sousedních hodnot  $y$ .

**Poděkování.** Části této kapitoly byly dříve publikovány v Kapitole 2 knihy [6].

## Literatura ke kapitole 4

1. Freund, J. E., Perles, B. E.: Modern elementary statistics. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 2006 (12th edition). 576 s.
2. Felix, M., Bláha, K.: Matematickostatistické metody v chemickém průmyslu. SNTL, Praha, 1962. 336 s.
3. ČSN 010253. Statistické metody v průmyslové praxi. Vydavatelství ÚNM, Praha, 1972, 114 s.
4. Suhir, E.: Applied Probability for Engineers and Scientists. McGraw-Hill, New York, 1997. 593 s.
5. Montgomery, D. C., and Runger, G. C.: Applied Statistics and Probability for Engineers. John Wiley, New York, 2006 (4th edition). 784 s.
6. Menčík, J.: Concise reliability for engineers. InTech, Rijeka, 2016, 204 s., ISBN <http://www.intechopen.com/books/concise-reliability-for-engineers>.
7. Abramowitz, M. and Stegun, I.: Handbook of mathematical functions. National Bureau of Standards, Washington. 1972 (tenth printing). 1046 s.
8. Gumbel, J. E.: Statistics of Extremes. Columbia University Press, New York, 1958. 375 s.
9. Conover, W. J.: Practical nonparametric statistics. 3<sup>rd</sup> ed., 1999. Wiley, 584 s.
10. ČSN 010253. Statistické metody v průmyslové praxi III. Základní neparametrické metody. Vydavatelství ÚNM, Praha, 1974.



## 5. Určování charakteristik zkoumaných veličin

Zkoumané veličiny často kolísají z náhodných příčin. V takových případech bývají popisovány prostřednictvím parametrů a charakteristických hodnot nebo pomocí rozdělení pravděpodobnosti. Tato kapitola ukazuje, jak sestřít histogramy a určovat parametry rozdělení. Zvláštní pozornost je věnována Weibullovu rozdělení. Je též ukázáno posouzení vhodnosti navrženého rozdělení pomocí diagramů Q-Q.

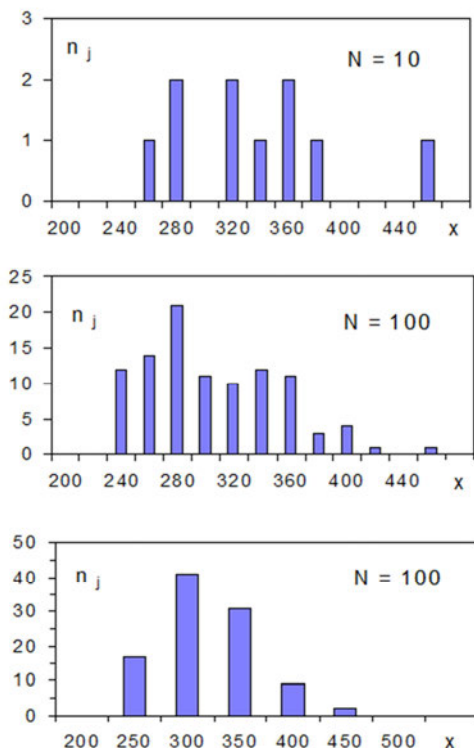
### Charakteristické hodnoty

Nejdůležitější parametry jsou střední hodnota  $\mu$  a směrodatná odchylka  $\sigma$ . Při vyhodnocování empirických dat jsou tyto parametry nahrazeny výběrovým průměrem  $\bar{x}$  a výběrovou směrodatnou odchylkou  $s$ , definovanými vztahy (4.2) a (4.4) v Kapitole 4. Další charakteristiky jsou koeficient asymetrie (šikmost) a špičatost, zmíněné tamtéž. Všechny tyto charakteristiky jsou jedinými zdroji informací o poloze náhodné veličiny na číselné ose a jejím rozptylu, máme-li naměřených hodnot velmi málo, méně než 15. Univerzální programy na jeden pokyn vypočtou všechny charakteristiky a vytisknou je ve formě tabulky. Například v Excelu stačí zadat pokyn Analýza dat a potom Popisná statistika.

POZNÁMKA. Zmíněné programy pro analýzu dat, a také velice užitečný program Řešitel, který umožňuje hledat extrémní hodnoty nebo řešit rovnice, jsou součástí každého Excelu, nejsou však běžně přístupné. Pokud se po stisknutí tlačítka Data neobjeví na hořejší liště vpravo pokyny Analýza dat a Řešitel, musíme je aktivovat. Po stisknutí tlačítka Soubor zvolíme Možnosti a v nich Doplňky. Poté označíme v seznamu Analytické nástroje a stiskneme tlačítko Přejít. V následujícím malém menu zaškrtneme Analytické nástroje a také Řešitel a stiskneme OK. To je vše.

### Histogram

První představu o rozdělení pravděpodobnosti vyšetřované veličiny můžeme získat z **histogramu** (Obr. 4.3 v kapitole 4 a Obr. 5.1 na následující straně). Histogram se



**Obr. 5.1.** Histogramy skupiny experimentálních hodnot pro rozdílné počty  $N$ .

sestrojuje ze všech zaznamenaných hodnot tak, že se rozmezí všech možných hodnot rozdělí do několika intervalů (tříd) stejné šířky, zjistí se počty hodnot v jednotlivých třídách, a nakreslí se obdélníky úměrné těmto četnostem [1, 2]. Histogramy se sestavují snadno při použití univerzálních programů jako Matlab, Mathcad, SPSS nebo Excel. U posledně jmenovaného k tomu slouží příkaz Histogram z menu Analýza dat (viz poznámku uvedenou výše); je nutno znát předem také počet tříd a jejich hranice. Bohužel neexistuje žádný univerzální vzorec pro stanovení počtu tříd  $m$ . V literatuře se nejčastěji uvádějí dva vzorce:

$$m = \text{INT}(2 \ln N) \quad , \quad m = \text{INT}(2 \sqrt{N}) \quad (5.1)$$

$N$  je celkový počet všech hodnot a INT znamená celočíselnou část výrazu (Integer). Tyto vzorce jsou ale vhodné pouze při několika desítkách hodnot. Protože univerzální počítačové programy vytvářejí histogramy okamžitě, je vhodné

nakreslit více variant histogramů, s různými počty tříd a s různými souřadnicemi jejich hranic, a vybrat takový, který vypadá nejlépe. Příklady histogramů s hezkým a nehezkým vzhledem jsou ukázány na Obr. 5.1. Máme-li důvod očekávat složitější rozdělení naměřených hodnot (např. se dvěma vrcholy), je pro vhodné zobrazení zapotřebí alespoň několik set hodnot.

Kromě histogramů, které udávají četnosti  $n_i$  v jednotlivých třídách ( $i$ ), je také možné vytvářet histogramy s kumulativními četnostmi; každá třída pak udává součet všech hodnot od levého konce až po vyšetřovanou  $j$ -tou třídu včetně:

$$n_{j,cum} = \sum_{i=1}^j n_i \quad (5.2)$$

Vydělíme-li počty hodnot v jednotlivých třídách celkovým počtem hodnot  $N$ , dostaneme relativní četnosti  $f_j$  a relativní kumulativní četnosti  $F_{j,cum}$ . Tyto dvě veličiny odpovídají přibližně hustotě pravděpodobnosti  $f$  a distribuční funkci  $F$ .

### Rozdělení pravděpodobnosti

Je výhodné, mohou-li empirická data být aproximována nějakým běžným rozdělením pravděpodobnosti. Velmi jednoduchá situace je s normálním, logaritmicko-normálním nebo exponenciálním rozdělením. Normální (Gaussovo) rozdělení (obr. 4.3 nebo 4.5 v kapitole 4) je plně popsáno střední hodnotou a směrodatnou odchylkou, definovanou v kapitole 3. Jeho použití je proto velmi snadné. V praxi se ale užívá výběrový průměr a směrodatná odchylka. Logaritmicko-normální rozdělení pracuje podobně s logaritmy naměřených hodnot. Exponenciální rozdělení má stejnou hodnotu směrodatné odchylky jako průměr, takže stačí vypočítat aritmetický průměr z naměřených hodnot. (Jejich odhady, počítané z empirických dat, se od sebe navzájem vždy poněkud liší. Je-li však rozdíl mezi  $\bar{x}$  a  $s$  malý, lze předpoklad o exponenciálním rozdělení přijmout.)

Velkou flexibilitu při aproximování různých tvarů empirických rozdělení vykazuje rozdělení Weibullovo, kterého si zde všimneme podrobněji.

### Weibullovo rozdělení

Všeobecný tvar jeho distribuční funkce (Obr. 4.6 v Kapitole 4) je

$$F(t) = 1 - \exp\{-[(t - t_0)/a]^b\} \quad (5.3)$$

s parametry  $a$ ,  $b$ ,  $t_0$ . **Parametr měřítka  $a$**  se vztahuje k hodnotám  $t$  a zajišťuje, že rozdělení nezávisí na jednotkách  $t$  (např. minuty nebo hodiny). Konstanta  $b$  je **parametr tvaru**. V závislosti na hodnotě  $b$  může Weibullovo rozdělení mít různé, dokonce velmi rozdílné tvary (Obr. 4.6); pro  $b = 1$  se změní v exponenciální rozdělení. Weibullovo rozdělení je vhodné pro charakterizování dob do poruchy, pevnosti nebo zatížení; proto se stalo oblíbeným při posuzování spolehlivosti. Užívá se však i v mnoha delších případech. Konstanta  $t_0$  je **prahová hodnota**, která odpovídá nejmenší možné hodnotě a charakterizuje polohu rozdělení na ose  $t$ . (Poznámka:  $t$  je obvyklý symbol pro čas; v případě jiných veličin lze užit jiné symboly.) V této části vysvětlíme dvě metody pro stanovení parametrů. První je založena na proložení transformovaných naměřených hodnot přímkou, druhá počítá parametry přímo pomocí vhodného optimalizačního programu.

### Dvouparametrické Weibullovo rozdělení

Pevnost nebo doba do poruchy nemohou mít záporné hodnoty, takže prahový parametr se často předpokládá roven nule,  $t_0 = 0$ . Distribuční funkce (5.3) má pouze dva parametry:

$$F(t) = 1 - \exp[-(t/a)^b] \quad (5.4)$$

Parametry  $a$  a  $b$  lze najít snadno, protože transformovaná data lze proložit přímkou [3]. Dvojitá logaritmická transformace a úprava změní rovnici (5.4) na tvar

$$\ln t = \ln a + (1/b) \ln\{\ln[1/(1 - F)]\} \quad (5.5)$$

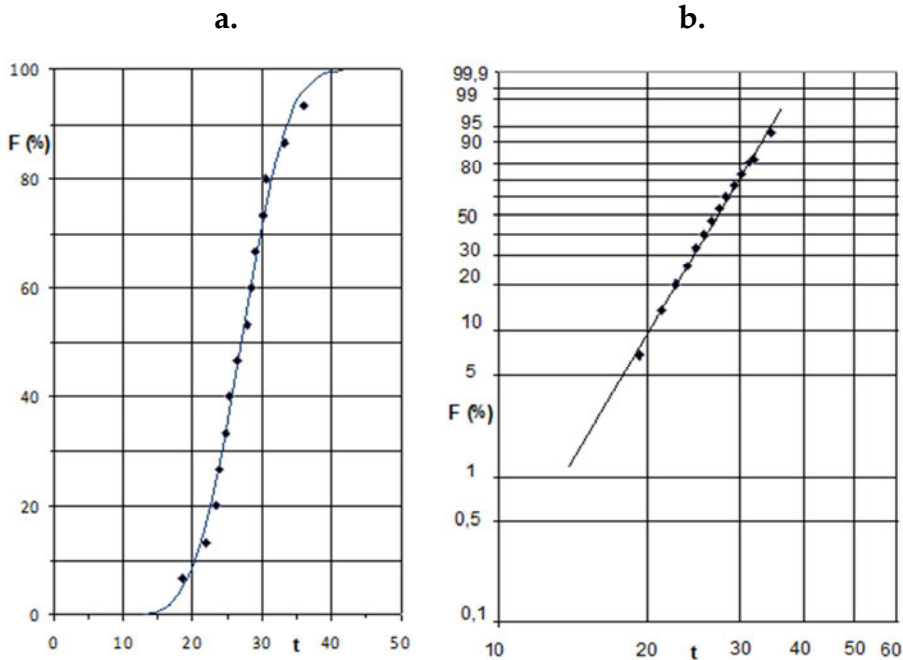
který odpovídá rovnici přímky (Obr. 5.2b)

$$Y = A + BX \quad (5.6)$$

kde  $Y = \ln t$ ,  $X = \ln\{\ln[1/(1 - F)]\}$ ,  $A = \ln a$ ,  $B = 1/b$ .

Tato metoda linearizace byla velmi oblíbená v minulosti, a dosud se někdy používá pro určování parametrů z provozních údajů pomocí speciálního diagramu, označovaného jako Weibullův papír (Obr. 5.2b). V tomto případě jsou třeba jednotlivé měřené hodnoty  $t_j$  a odpovídající hodnoty  $F_j$  empirické distribuční funkce. Hodnoty  $t_j$  se získávají uspořádáním  $n$  hodnot z provozu (např. doby do poruchy od nejmenší ( $j = 1$ ) do největší ( $j = n$ )). Odpovídající hodnoty distribuční funkce lze vypočítat podle kapitoly 4 jako

$$F_j = j / (n + 1) \quad (5.7)$$



**Obr. 5.2.** Weibullova distribuční funkce  $F(t)$ : vlevo je původní souřadný systém, vpravo Weibullův pravděpodobnostní papír s transformovanými hodnotami [3].

$j$  je pořadové číslo a  $n$  je celkový počet hodnot. Regresní konstanty  $A$ ,  $B$  se získají proložení empirických dat přímkou (použitím Weibullova papíru nebo pomocí vhodného programu pro prokládání křivek, např. „Vložit spojnicí trendu“ v Excelu). Konstanty v původní distribuční funkci (5.4) se z  $A$  a  $B$  získají inverzní transformací:

$$b = 1/B, a = \exp(A) \quad (5.8)$$

Zakreslení naměřených hodnot v souřadném systému  $X = \ln\{\ln[1/(1 - F)]\}$ ,  $Y = \ln t$ , umožňuje dobrou vizuální kontrolu. V ideálním případě leží jednotlivé hodnoty na přímce.

### Tříparametrické Weibullovo rozdělení

Dvuparametrické rozdělení není vždy vhodné. Někdy transformovaná data neleží na přímce, nebo je zřejmé, že rozdělení by mělo mít prahovou hodnotu zřetelně vyšší než 0. V takovém případě bude tříparametrická funkce (5.3) lepší. Parametry

tohoto rozdělení lze najít popsáním postupem pro dvouparametrickou funkci, jestliže  $t$  v rovnici (5.4) nahradíme výrazem  $t - t_0$ ; konstanta  $t_0$  musí být definována předem. Rozložení transformovaných bodů se mění v závislosti na zvolené hodnotě  $t_0$ . Nejlepší hodnota  $t_0$  je taková, pro kterou se transformovaná data nejvíce blíží k přímce. Tento postup vyžaduje určitý cvik. Naštěstí existuje i jednodušší přímočarý postup [3], který vysvětlíme dále.

### Přímé stanovení parametrů

Konstanty  $a$ ,  $b$ , a  $t_0$  lze snadno získat bez jakékoliv transformace. Řešení rovnice (5.3) pro  $t$  dá následující výraz pro kvantily:

$$t = t_0 + a\{\ln[1/(1 - F)]^{1/b}\} \quad (5.9)$$

Tento výraz a **metoda nejmenších čtverců** se pak použijí pro nalezení takových hodnot  $a$ ,  $b$ , a  $t_0$ , které minimalizují součet čtverců rozdílů mezi naměřenými a vypočtenými hodnotami  $t$ ,

$$\Sigma(t_{j,\text{meas}} - t_{j,\text{calc}})^2 = \min ! \quad (5.10)$$

Je-li k dispozici vhodný řešitel pro takovouto minimalizaci (i Excel má takový), postačí připravit řadu naměřených hodnot  $t_{j,\text{meas}}$ , a druhou řadu hodnot  $t_{j,\text{calc}}$ , vypočítaných pro stejné hodnoty  $F_j$  pomocí rovnice (5.9) s použitím parametrů  $a$ ,  $b$ , a  $t_0$ . Řešitel na pokyn, aby minimalizoval výraz (5.10) změnami  $a$ ,  $b$ , a  $t_0$ , udělá všechnu potřebnou práci.

**POZNÁMKA.** Vzorec (5.9) je také vhodný pro nalezení „minimální zaručené hodnoty“ (např. pevnosti nebo doby do poruchy), tj. takové, že pravděpodobnost, že by se vyskytl ještě slabší kus, je velmi nízká (odpovídající zvolené hodnotě  $F$ ).

### Příklad 1.

Pevnost nové slitiny byla měřena na deseti vzorcích, s následujícími výsledky: 250 – 201 – 232 – 281 – 297 – 211 – 276 – 302 – 315 – 265 MPa.

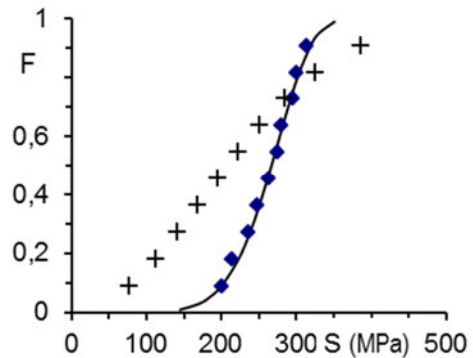
Nalezněte parametry tříparametrického Weibullova rozdělení a určete „zaručenou“ 1% pevnost.

**Řešení.** Všechny údaje jsou v tabulce 1. Naměřené pevnosti byly seřazeny od nejnižší k nejvyšší ( $S_{j,\text{meas}}$ ), a odpovídající hodnoty distribuční funkce byly vypočteny jako  $F_j = j/(N + 1)$ . Potom byly pro stejné hodnoty  $F_j$  vypočteny pomocí rovnice (9) hodnoty  $S_{j,\text{calc}}$  pro konstanty  $a$ ,  $b$ ,  $S_0$ , definované předem. Rovněž byl

vypočten součet čtverců rozdílů mezi naměřenými a vypočtenými pevnostmi. Jednotlivé hodnoty jsou nakresleny na obr. 5.3; kosočtverce představují naměřené hodnoty, zatímco křížky odpovídají výpočtům pro zvolené počáteční hodnoty  $a = 250$  MPa,  $b = 2$  a  $S_0 = 0$  MPa. Je zřejmé, že křížky nesouhlasí s naměřenými pevnostmi. Použití Řešitele dalo následující hodnoty:  $a = 280.6$  MPa,  $b = 6.659$  a  $S_0 = 0$  MPa. Tyto konstanty aproximují měřená data velmi dobře; viz též plnou čáru extrapolovanou k nižším a vyšším pravděpodobnostem. Zaručená 1%-pevnost, vypočtená jako  $S(F = 0.01)$ , je 140.6 MPa. (Čtenáři doporučujeme vyřešit znovu tento příklad pro získání praxe.)

**Tabulka 1.**

$j$	$S_{j,meas}$	$F_j$	$S_{j,calc}$
1	201.0	0.0909	197.2
2	211.0	0.1818	216.2
3	233.0	0.2727	232.9
4	250.0	0.3636	246.5
5	265.0	0.4545	258.4
6	276.0	0.5455	269.6
7	281.0	0.6364	280.7
8	297.0	0.7273	292.3
9	302.0	0.8182	305.4
10	315.0	0.9091	322.7

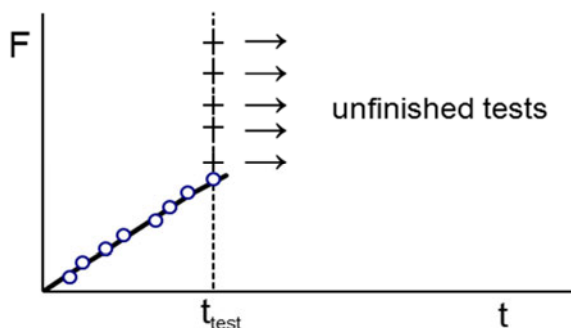


**Obr. 5.3.** Pevnosti naměřené ( $S_{j,meas}$ ) a vypočtené ( $S_{j,calc}$ ) a hodnoty  $F_j$  distribuční funkce.

V tomto příkladu bylo nutné omezit prahovou pevnost hodnotou  $S_0 \geq 0$ , protože první výpočet bez jakéhokoliv omezení dal zápornou hodnotu pevnosti  $S_0$ , která není možná. S tímto omezením Řešitel okamžitě „doporučil“ prahovou hodnotu  $S_0 = 0$ , takže rozdělení pravděpodobnosti má zde ve skutečnosti jen dva parametry  $a$ ,  $b$ . V některých případech může existovat velký rozdíl mezi hodnotami, předpověděnými pro malé pravděpodobnosti podle dvou- nebo tříparametrické Weibullovovy funkce. Dvoupametrické rozdělení s předpokládanou prahovou hodnotou  $S_0 = 0$  dá nižší dovolené napětí, což je bezpečnější. Na druhé straně, velikost průřezu takové součásti musí být větší, a tedy nákladnější. Někdy je nutno udělat kompromis mezi bezpečností a ekonomikou.

## Stanovení parametrů rozdělení z cenzorovaných dat

V některých případech je počet experimentálních hodnot omezen a pouze část výsledků je známa. Například doba do poruchy únavou složitých součástí kolísá, a je-li testována skupina takovýchto součástí pro zjištění rozdělení jejich životností, mohou být doby do porušení u některých součástí neprakticky dlouhé. Zkoušky životnosti se proto někdy ukončují po uplynutí určité doby  $t_{\text{test}}$  anebo po porušení určitého podílu zkoušených kusů. V tomto případě známe přesně doby do poruchy porušených dílů a také víme, že životnost zbývajících součástí bude delší (ale nevíme, jak bude dlouhá). Jiný případ nastane, jestli měřená veličina má hodnoty mimo použitelný rozsah měřicího přístroje; říkáme, že měřená data jsou cenzorována. Situace je znázorněna na Obr. 5.4. Jestliže je znám druh rozdělení pravděpodobnosti, můžeme stanovit parametry tohoto rozdělení z těch dat, pro která jsou doby do poruchy známy, jestliže každé z těchto hodnot ( $t_j$ ), uspořádaných podle velikosti, přiřadíme odpovídající hodnotu distribuční funkce  $F_j = j/(N+1)$ . Tyto otázky jsou velmi důležité při zkouškách životnosti, takže existují různá schémata a postupy pro zpracování výsledků; viz např. [4].



Obr. 5.4. Cenzorovaná data ze zkoušek životnosti. Unfinished tests - nedokončené zkoušky

## Q - Q graf

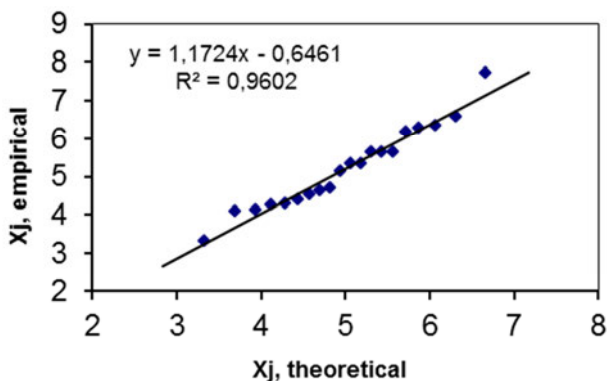
Jednoduchým nástrojem pro porovnání dvou rozdělení pravděpodobnosti je tzv. kvantil-kvantilový graf, krátce Q-Q graf [5]. Do tohoto diagramu se zakreslují odpovídající kvantily proti sobě. Souřadnice  $x_j$  bodu  $P_j$  zde odpovídá  $j$ -kvantilu jednoho rozdělení a souřadnice  $y_j$  odpovídá stejnému kvantilu druhého rozdělení. Jsou-li obě porovnávaná rozdělení podobná, leží body v Q-Q grafu přibližně na přímce  $y = x$  (Obr. 5.5). Kvantil-kvantilový graf názorně informuje, jsou-li poloha,



tvár a šikmost porovnávaných rozdělení podobné nebo rozdílné. Q–Q grafy se používají pro srovnání dvou souborů dat nebo pro porovnání empirického rozdělení s teoretickým.

Kvantil-kvantilový graf se vytvoří následujícím způsobem. Porovnávané veličiny (první označíme  $x$  a druhou  $y$ ) se uspořádají od nejmenší do největší, odpovídající hodnoty distribuční funkce  $F$  se vypočítají z rovnice (5.7), a dvojice hodnot kvantilů  $x_j, y_j$  pro stejné hodnoty  $F_j$  se zakreslí v souřadném systému  $x, y$  (Obr. 5.5). To je nejsnadnější, mají-li oba soubory stejný počet hodnot. V ostatních případech je nutno přepočítat kvantily pro stejné pravděpodobnosti interpolací ze sousedních bodů.

Obrázek 5.5 ukazuje Q–Q graf pro 20 teoretických hodnot normálního rozdělení (s průměrem  $\mu = 5.0$  a směrodatnou odchylkou  $\sigma = 1.0$ ), porovnaných s dvaceti “empirickými” hodnotami, vygenerovanými v tomto případě pro stejné parametry metodou Monte Carlo. Vidíme, že jednotlivé body leží přibližně na přímce.



**Obr. 5.5.** Kvantil-kvantilový (Q-Q) diagram pro normální rozdělení (parametry:  $\mu = 5$ ,  $\sigma = 1$ ,  $n = 20$ ). Vodorovná osa – teoretické hodnoty, svislá osa – naměřené hodnoty.

**Poděkování.** Část této kapitoly byla dříve publikována v kapitole 11 knihy [3].

## Literatura ke kapitole 5

1. Pechoč, V.: Vyhodnocování měření a početní metody v chemickém inženýrství. SNTL, Praha, 1981. 226 s.

2. Felix, M., Bláha, K.: Matematickostatistické metody v chemickém průmyslu. SNTL, Praha, 1962. 336 s.
3. Menčík, J.: Concise reliability for engineers (Kapitola 11). InTech, Rijeka, 2016, *Open Access*: <http://www.intechopen.com/books/concise-reliability-for-engineers>, 204 p. ISBN 978-953-51-2278-4.
4. Bednařík, J. a kol.: Technika spolehlivosti v elektronické praxi. Praha, SNTL, 1990. 336 s.
5. Kupka, K.: Statistické řízení jakosti. Trilobyte, Pardubice, 1997, 2001. 191 s.

## 6. Vztahy mezi dvěma a více veličinami

V této kapitole jsou definovány různé veličiny pro charakterizování vztahů mezi dvěma i více proměnnými, jako je kovariance, korelace a koeficient determinace. Je také ukázáno jejich použití při stanovení konstant v regresní funkci a hodnocení jakosti aproximace. Autokorelace ukazuje, jsou-li mezi sebou korelovány hodnoty v určité řadě. Závěrem je vysvětleno získávání informací prostřednictvím takzvaného data-miningu.

### Kovariance and korelace

Jedním z úkolů experimentálního výzkumu je zjistit zda existuje vztah mezi dvěma nebo více veličinami a jak je těsný, a popsat jej vhodným matematickým výrazem. Takovýto vztah buď neexistuje, nebo je méně či více těsný, až popř. deterministický. Tuto těsnost lze charakterizovat pomocí koeficientu kovariance nebo korelace. Pro dvě proměnné,  $x$  a  $y$ , je **kovariance**  $s_{xy}$  definována jako

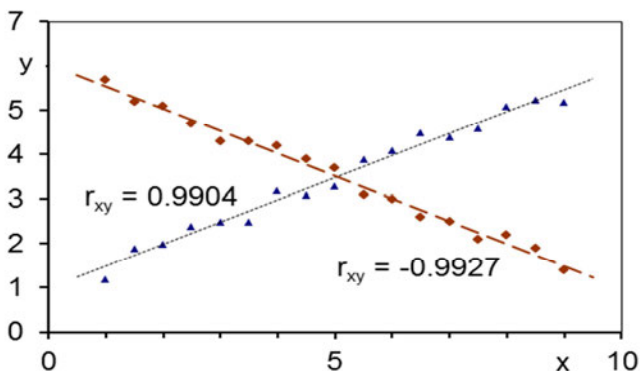
$$s_{xy} = \frac{\sum (x_j - \bar{x})(y_j - \bar{y})}{n-1} \quad (6.1)$$

$\bar{x}$  a  $\bar{y}$  jsou průměrné hodnoty obou veličin a  $n$  je počet dvojic  $x_j, y_j$ . Kovariance může být pozitivní, jestliže obě veličiny spolu vzrůstají (popřípadě klesají), nebo negativní, jestliže jedna veličina vzrůstá a druhá klesá.

Nedostatkem kovariance je, že tento koeficient může nabývat hodnoty od  $-\infty$  do  $+\infty$ , v závislosti na hodnotách  $x, y$ . Náznornější mírou těsnosti vztahu je **koeficient korelace**  $r_{xy}$ , definovaný jako kovariance dělená směrodatnými odchylkami obou proměnných:

$$r_{xy} = \frac{s_{xy}}{s_x s_y} \quad (6.2)$$

Korelační koeficient je vlastně kovariance normalizovaná vzhledem k rozptylu obou veličin. Hodnoty  $r_{xy}$  se pohybují mezi 0 (žádná korelace) a +1 (deterministický nebo funkční vztah) při pozitivní korelaci. Při negativní korelaci se  $r_{xy}$  pohybuje mezi 0 a -1 (Obr. 6.1).



**Obr 6.1.** Korelace kladná (trojúhelníky) a záporná (kosočtverce).

POZNÁMKA: Vyjadřování vztahu dvou veličin pomocí regresní funkce bude probráno v Kapitole 7.

Analogickou charakteristikou nezávislou na parametrech je **Spearmanův koeficient pořadové korelace**:

$$r_s = 1 - \frac{6 \sum d_j^2}{n(n^2 - 1)} \quad (6.3)$$

$n$  je počet dvojic hodnot  $x_j, y_j$  veličin  $x, y$ . Každé hodnotě  $x_j$  je přiřazeno pořadové číslo  $p_j$ , a hodnotám  $y$  jsou přiřazena pořadová čísla  $q_j$  ( $j = 1$  odpovídá nejnižším hodnotám). Potom se vytvoří rozdíly pořadových čísel jednotlivých dvojic  $x_j, y_j$  jako  $d_j = p_j - q_j$  a užity v rovnici (6.3).

Relativně vysoké hodnoty korelačního koeficientu,  $r = 0.8$  a více, naznačují funkční vztah. Nicméně při pochybnostech je vhodné provést statistický test významnosti (viz Kapitolu 8).

**Pozor!** Vysoký stupeň korelace neznamená nutně příčinný vztah. Může existovat jiný činitel, který ovlivňuje obě veličiny ( $x, y$ ) podobným způsobem. Dobře známý je následující humorný příklad. Výzkum, provedený v několika vesnicích, ukázal, že existuje vysoká korelace mezi počtem narozených dětí a počtem čápů ve vesnici. Dokazuje to snad, že čápi nosí děti? Nikoliv; vysvětlení souvisí s velikostí vesnice: ve větších vesnicích se rodí více dětí, ale obvykle větší vesnice mají více rybníků, a tedy i čápů.

Velmi užitečnou veličinou je **koeficient determinace**  $r^2$ , který je vhodný pro charakterizování kvality lineárních i nelineárních regresních funkcí. Obrázek 6.2 ukazuje regresní přímku, jednu hodnotu  $y_{\text{meas}}$  naměřenou pro jistou hodnotu  $x$ , dále odpovídající vypočtenou hodnotu  $y_{\text{calc}}(x)$  na regresní přímce, a průměrné hodnoty  $\bar{x}$ ,  $\bar{y}$  (na obrázku  $x_{\text{ave}}$ ,  $y_{\text{ave}}$ ) skupiny  $x_j$  a  $y_j$ . Vzdálenost  $y_j$  od průměru  $\bar{y}$  je možno vyjádřit jako

$$y_j - \bar{y} = (y_j - y_{j,\text{reg}}) + (y_{j,\text{reg}} - \bar{y}), \text{ or } \Delta_{\text{tot}} = \Delta_{\text{res}} + \Delta_{\text{reg}} \quad (6.4)$$

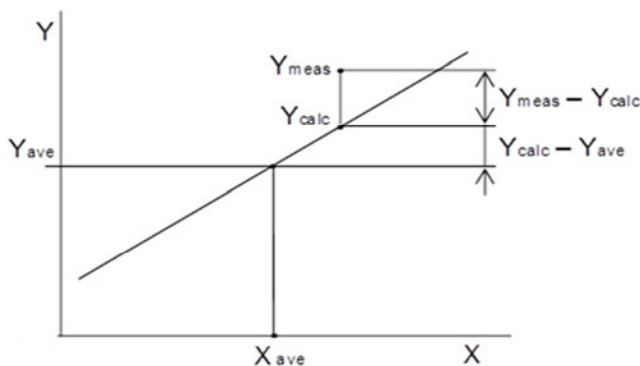
$y_{j,\text{reg}}$  je odpovídající hodnota na regresní přímce;  $\Delta_{\text{tot}}$  znamená celkový rozdíl,  $\Delta_{j,\text{reg}}$  znamená rozdíl  $j$ -té hodnoty na regresní přímce a průměru, a  $\Delta_{\text{res}}$  je zbytkový (residuální) rozdíl, což je vzdálenost  $j$ -té měřené hodnoty a odpovídající hodnoty na regresní čáře, Lze dokázat, že platí i následující vztah:

$$\sum_j (y_j - \bar{y})^2 = \sum_j (y_j - y_{j,\text{reg}})^2 + \sum_j (y_{j,\text{reg}} - \bar{y})^2 \quad (6.5)$$

Součet je proveden pro všech  $n$  hodnot. Rovnici (6.5) lze přepsat jako

$$SS_{\text{tot}} = SS_{\text{reg}} + SS_{\text{res}} \quad (6.6)$$

$SS_{\text{tot}}$  je součet čtverců celkových rozdílů,  $SS_{\text{reg}}$  je součet čtverců rozdílů mezi body na regresní křivce a celkovým průměrem  $\bar{y}$ , a  $SS_{\text{res}}$  součet čtverců vzdáleností jednotlivých naměřených hodnot od odpovídajících bodů na regresní křivce (tj. reziduálních vzdáleností).



**Obr. 6.2.** Koeficient determinace  $r^2$ ; reziduální složka ( $y_{\text{meas}} - y_{\text{calc}}$ ) a regresní složka ( $y_{\text{calc}} - y_{\text{ave}}$ ) celkového rozdílu  $y_{\text{meas}} - y_{\text{ave}}$ . Indexy znamenají: *meas* – měřený, *calc* – vypočtený (regresní), *ave* – průměrný.

Jestliže rovnici (6.6) vydělíme výrazem  $n - 1$ , obdržíme

$$s_{\text{tot}}^2 = s_{\text{reg}}^2 + s_{\text{res}}^2 \quad (6.7)$$

$SS$  znamená součet kvadrátů rozdílů a  $s^2$  znamená rozptyl; Indexy mají stejný význam jako v předchozím případě; *res* odpovídá zbytkovému (reziduálnímu) rozptylu jednotlivých naměřených hodnot kolem regresní funkce. Koeficient determinace je definován jako

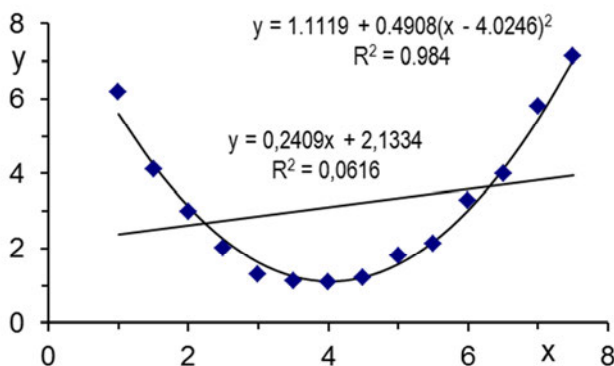
$$r^2 = SS_{\text{reg}}/SS_{\text{tot}} = s_{\text{reg}}^2/s_{\text{tot}}^2 \quad (6.8)$$

a vyjadřuje, jaká část celkového rozptylu je tvořena regresní funkcí. Pro deterministický vztah je  $r^2 = 1$ . Rovnici (6.8) lze přepsat na tvar

$$r^2 = (SS_{\text{tot}} - SS_{\text{res}})/SS_{\text{tot}} = 1 - (SS_{\text{res}}/SS_{\text{tot}}), \text{ or } r^2 = 1 - (s_{\text{res}}^2/s_{\text{tot}}^2) \quad (6.9)$$

Koeficient determinace (6.8) byl odvozen pro lineární vztah  $y(x)$ . Je-li však počítán podle rovnice (6.9) prostřednictvím zbytkového rozptylu, může být také používán pro charakterizování kvality nelineárních regresních funkcí. Z toho vyplývá význam zbytkového rozptylu při určování parametrů regresních funkcí (viz následující kapitola).

Zkoumáme-li vztah mezi dvěma veličinami, vždy bychom měli nejprve udělat graf naměřených hodnot, a teprve potom posuzovat těsnost vztahu. Obrázek 6.3 ukazuje skupinu naměřených hodnot, která zřetelně vykazuje nelineární vztah. Kdybychom bez tohoto poznatku stanovili koeficient korelace pro lineární vztah, obdrželi bychom velmi nízkou hodnotu, říkající, že neexistuje žádná (lineární) korelace ( $r =$



**Obr. 6.3.** Dvě aproximace naměřených hodnot: dobrá aproximace (kvadratická funkce) a špatná (lineární funkce).  $R^2 =$  koeficient determinace.

0,286;  $r^2 = 0,0616$ ). Naproti tomu koeficient determinace pro kvadratickou regresní funkci je  $r^2 = 0,984$  and  $r = 0,992$ , což znamená velmi dobrou aproximaci.

### Vícenásobná korelace

Zjištění korelací je první krok při zkoumání vztahů mezi vyšetřovanými veličinami. Korelace mohou existovat mezi dvěma, ale i více proměnnými, tzv. vícenásobná korelace. Korelace v experimentálních datech se obvykle najdou pomocí vhodného statistického programu. Například v Excelu příkaz CORREL, aplikovaný na skupinu párovaných hodnot  $x_j$  a  $y_j$ , poskytne hodnotu (lineárního) korelačního koeficientu  $r_{xy}$ . V našem případě je ale účinnější vytvořit graf naměřených hodnot  $y(x)$  a nakreslit tam regresní funkci pomocí příkazu *Vložit spojnicí trendu*, čímž dostaneme velmi názorný obrázek. Je také užitečné si vyžádat (z menu) koeficient determinace  $r^2$ , který charakterizuje těsnost vztahu.

Vícenásobnou korelaci budeme nyní ilustrovat na příkladu tepelného zpracování oceli.

#### Příklad.

Bylo zapotřebí vyšetřit, jak je tvrdost ( $H$ ) a pevnost ( $S$ ) kalené oceli určité značky ovlivněna teplotou zpracování ( $T$ ) a časovou prodlevou ( $t$ ) při vysoké teplotě, a zda jsou spolu korelovány. Osm vzorků (A - H) bylo zpracováno za různých podmínek, jak je naznačeno v hořejší části Tabulky 1, uvedené na následující straně. Tabulka vícenásobných korelací ve spodní části byla vytvořena v Excelu. K tomu byly použity pokyny Analýza dat a Korelace; poté bylo do příslušného okénka v menu Korelace vloženo pole obsahující vstupní hodnoty a byla označena buňka pro umístění korelační tabulky. Po stisknutí OK se objeví korelační tabulka uvedená dále.

Korelační tabulka ukazuje názorně, které veličiny jsou silně korelovány, a které nikoliv. Například koeficient korelace mezi pevností a tvrdostí je v našem případě velmi vysoký (0,980), zatímco korelace mezi časovou prodlevou na vysoké teplotě a tvrdostí je nízká (-0,545). Zanedbání nepodstatných veličin umožní zjednodušení vzorců a pozdějších výpočtů. Vysoká korelace dvou veličin umožňuje použít jakoukoliv z nich, což může někdy zjednodušit práci. Například stanovení pevnosti je kvůli výrobě vzorků mnohem náročnější než měření tvrdosti. Pokud testy se současným měřením pevnosti a tvrdosti odhalí vysokou korelaci, jako v našem případě, je možné v budoucnu měřit pouze tvrdost a počítat pevnost z ní pomocí

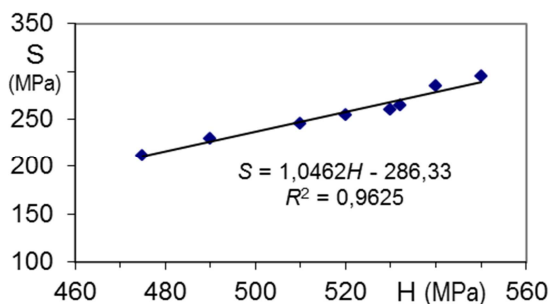
**Tabulka 1.** Vstupní data (nahore) a tabulka vícenásobných korelací (dole).

Vzorek	Teplota	Čas	Tvrdost	Pevnost
A	800	10,3	475	212
B	840	9,8	510	246
C	920	9,6	540	285
D	820	10	490	230
E	870	9,7	520	255
F	850	10,5	530	260
G	900	9,4	550	295
H	860	9,1	532	265

	Teplota	Čas	Tvrdost	Pevnost
Teplota	1			
Čas	-0,581	1		
Tvrdost	0,907	-0,545	1	
Pevnost	0,951	-0,573	0,980	1

vhodného vzorce, jak je ukázáno na Obr. 6.4.

**Obr. 6.4.** Pevnost  $S$  jako funkce tvrdosti  $H$  (obojí v MPa):  $S = 1.046 H - 286.33$ .

### Autokorelace

Až dosud jsme zkoumali, jestli dvě veličiny jsou navzájem korelovány. Je ale také možné zkoumat, jestli hodnoty v jedné řadě dat jsou korelovány navzájem mezi sebou. Příkladem mohou být denní průměrné teploty. Je pravděpodobné, že rozdíl mezi dvěma teplotami bude nižší, jestliže odpovídá dvěma následujícím dnům, než kdyby činil několik týdnů nebo měsíců. Tímto způsobem je možno charakterizovat řadu hodnot pomocí tzv. autokorelace. Autokorelační koeficient se dostane podobně jako korelační koeficient dvou proměnných  $x$  a  $y$  (např. v Excelu příkazem CORREL). Jediný rozdíl nyní je, že místo veličiny  $y$  použijeme opět

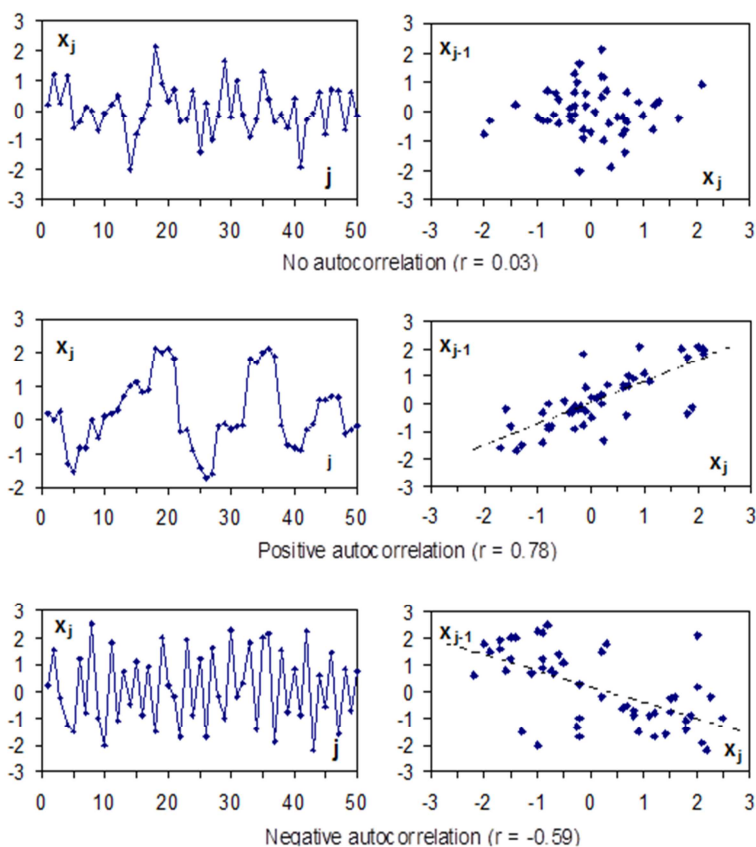


hodnoty  $x$ , ale posunuté o 1, 2, ... nebo  $n$  míst. Takto vytvořená nová řada je označena  $y'$ , a hovoříme o autokorelaci prvního, druhého... nebo  $n$ -tého řádu. Například tabulka dole odpovídá autokorelaci prvního řádu.

$y :$	$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$	$y_7$	$y_8$	$y_9$	$\dots$
$y' :$		$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$	$y_7$	$y_8$	$\dots$

Autokorelace může být kladná nebo záporná. Oba případy jsou na obr. 6.5.

POZNÁMKA: Autokorelace se užívá zejména při analýze dynamických procesů a časových řad a při zpracování signálů.



**Obr. 6.5.** Autokorelace: a) žádná ( $r = 0.03$ ), b) pozitivní ( $r = 0.78$ ), c) negativní ( $r = -0.59$ ). Diagramy vlevo znázorňují časové řady; diagramy vpravo ukazují autokorelace prvního řádu;  $x_{j-1} = f(x_j)$ .

Více informací o kovarianci a korelaci lze najít např. v knihách [1 – 5].

**Data-mining.** V některých oborech, např. v chemii, biologii, astronomii či medicíně, ale i v bankovníctví nebo obchodě, existují veliká množství dat. A výkonné počítače jsou dnes schopny je zpracovávat. To vedlo ke vzniku oboru, kterému se říká data-mining (dolování z dat). Na rozdíl od tradiční analýzy dat, kdy vyslovíme určitou hypotézu a snažíme se ji potvrdit na základě údajů, které pak získáváme pozorováním nebo experimenty, vycházíme při data-miningu naopak z obrovských množství dat, která již jsou k dispozici, a snažíme se v nich najít nějaké charakteristické znaky nebo souvislosti, a tedy skryté a potenciálně užitečné informace. Kvůli obrovským množstvím analyzovaných dat (TB), je pro data mining nezbytné použití vhodného software. Existují specializované programy pro tento účel, např. STATISTICA Data Miner, SAS Enterprise Miner a [SPSS Clementine](#). Mezi známé nekomerční SW patří [Weka](#) a [Orange](#). Bližší informace o data miningu lze najít v [6, 7].

## Literatura ke kapitole 6

1. Freund, J. E.: Modern elementary statistics. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1981 (6th edition). 561 s.
2. Freund, J. E., Perles, B. E.: Modern elementary statistics. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 2006 (12th edition). 576 s.
3. Suhir, E.: Applied Probability for Engineers and Scientists. McGraw-Hill, New York, 1997. 593 s.
4. Montgomery, D. C., and Runger, G. C.: Applied Statistics and Probability for Engineers. John Wiley, New York, 2006 (4th edition). 784 s.
5. Kupka, K.: Statistické řízení jakosti. Trilobyte, Pardubice, 1997. 191 s.
6. Berka, P.: Dobývání znalostí z databází. Academia, Praha, 2003. 366 s.
7. Piatetsky-Shapiro G., Parker, G.: "Lesson: Data Mining, and Knowledge Discovery: An Introduction". Introduction to Data Mining. KD Nuggets. 2011, Retrieved 30 August 2012. [http://www.kdnuggets.com/data\\_mining\\_course/x1-intro-to-data-mining-notes.html](http://www.kdnuggets.com/data_mining_course/x1-intro-to-data-mining-notes.html)

## 7. Vyrovnávání empirických dat regresními funkcemi

Tato kapitola se zabývá vyjádřováním vztahů mezi různými veličinami pomocí regresních funkcí. Jsou ukázány typické tvary těchto funkcí a je vysvětleno stanovení jejich konstant. Protože vstupní data zpravidla vykazují proměnlivost z náhodných příčin, regresní funkce nedává zcela přesné hodnoty. Spolehlivost předpovědi lze zvýšit vytvořením konfidenčního pásu kolem regresní funkce.

Je výhodné, máme-li empirická data popsána analytickým výrazem, například

$$y = f(x), \text{ nebo } w = f(x, y, z, \dots) \quad (7.1)$$

podle toho, závisí-li vyšetřovaná proměnná na jedné nebo více veličinách. Takovýto výraz, **regresní funkce**, poskytuje zhuštěnou informaci a umožňuje další zpracování získaných dat. Kroky při hledání regresní funkce jsou následující: 1) návrh vhodného tvaru regresní funkce, a 2) stanovení nejvhodnějších hodnot konstant v této funkci. V některých případech je ještě nutné: 3) posoudit kvalitu navržené aproximace, popřípadě rozhodnout, která z několika navržených aproximací je nejlepší.

### Návrh regresní funkce

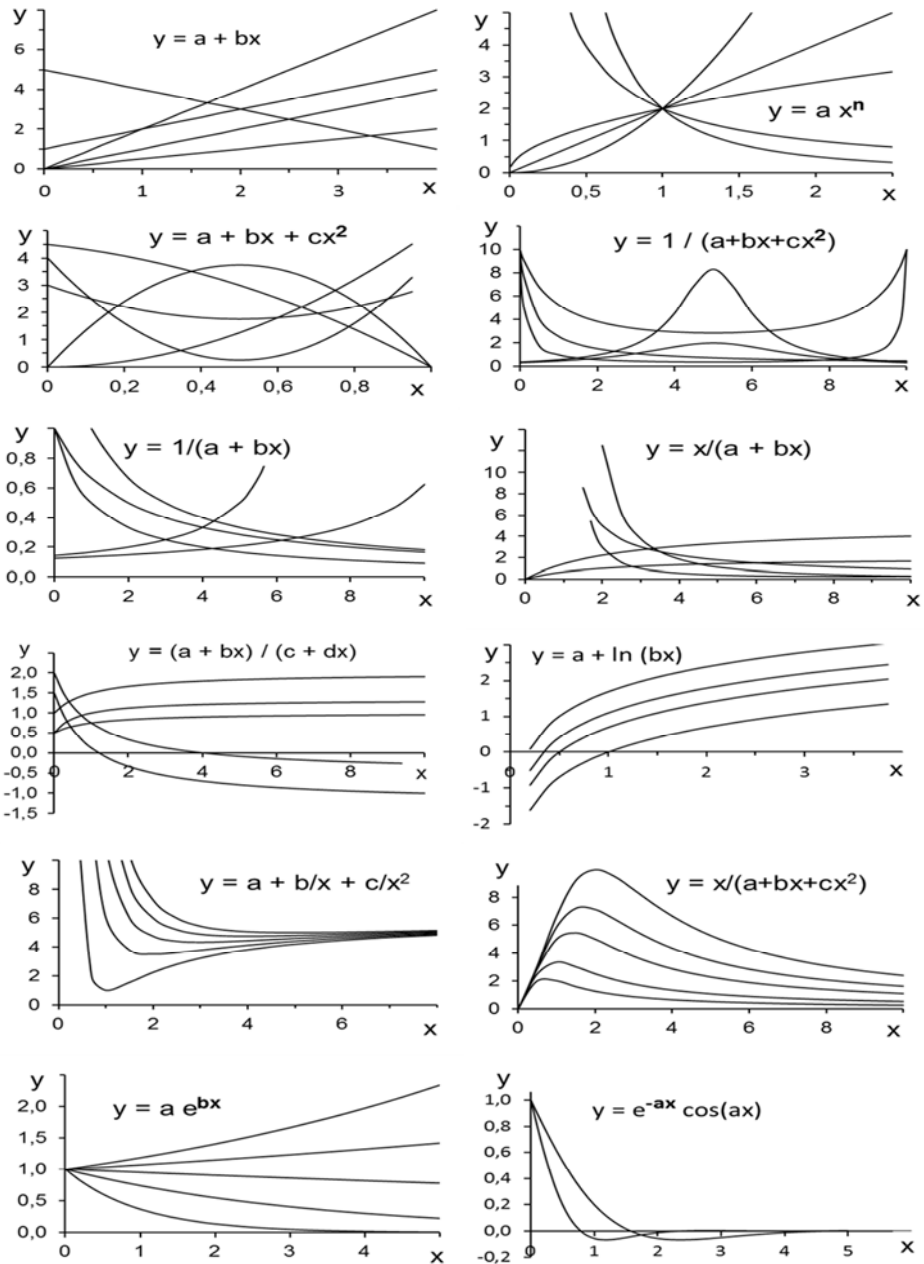
První představu lze získat z grafu se všemi naměřenými hodnotami  $y(x)$ . **Obraz řekne více než tisíc slov!** V případě více nezávisle proměnných,  $w = f(x, y, z, \dots)$ , lze tvar regresní funkce odhadnout z grafů  $w = f(x; y = konst, z = konst)$ ,  $w = f(y; x = konst, z = konst)$ ..., odpovídajících řezům odezvovou plochou. S volbou regresní funkce může pomoci Obr. 7.1, který ukazuje tvary různých funkcí. Dříve, než proložíme naměřená data regresní funkcí, je vhodné přemýšlet o charakteru vyšetřovaného jevu nebo procesu. Obecně, regresní funkce může monotónně růst nebo klesat, může mít mocninný charakter, anebo její růst může být exponenciální či logaritmický, může klesat jako hyperbola nebo exponenciální funkce se záporným exponentem, může mít určité maximum či minimum. Může se blížit asymptoticky k jisté hodnotě pro velmi vysoké hodnoty nezávisle proměnné.

Řešení některých úloh vede k periodickým funkcím (sin, cos...). Někdy se aproximace hledá ve tvaru řady (např. polynom, Fourierova řada s trigonometrickými členy nebo Pronyho řada s členy exponenciálními). Při návrhu regresní funkce může pomoci znalost analytického řešení podobných úloh. Tvary několika jednoduchých funkcí jsou znázorněny na Obr. 7.1; viz též [1, 2]. V některých případech mohou být užitečné funkce typické pro rozdělení pravděpodobnosti, například normální nebo Weibullovo rozdělení; viz Obr. 4.7 a 4.8 v Kapitole 4). V těchto případech se jedná o tvary hustot pravděpodobnosti a distribučních funkcí. Dva příklady, harmonická (kosinová) funkce a funkce s několika exponenciálními členy, jsou uvedeny později v této kapitole.

### Určování regresních konstant

Některé univerzální programy (včetně Excelu) nabízejí několik regresních funkcí pro aproximování empirických dat a mohou v nich určovat parametry s užitím vlastních algoritmů. V takových případech stačí vytvořit graf naměřených hodnot  $x$  a  $y$ . Potom, po kliknutí pravým tlačítkem myši na řadu bodů, se objeví menu, ze kterého lze zvolit vhodnou funkci, např. lineární, polynomickou, mocninnou, exponenciální nebo logaritmickou. Postup byl ukázán v Kapitole 5. Je důležité požadovat (z menu), aby byl v grafu ukázán také tvar regresní rovnice a koeficient determinace  $r^2$ , který charakterizuje kvalitu aproximace. (Podrobné vysvětlení  $r^2$  bylo dáno v Kapitole 6). To je užitečné zejména, když je třeba porovnávat různé regresní funkce.

Při návrhu regresní funkce musíme brát v úvahu více hledisek. Jedním z nich je, aby výpočty s touto funkcí byly poměrně jednoduché. Snadno se pracuje s polynomy, např.  $y = a + bx + cx^2 + dx^3 + \dots$ . Také univerzální programy pro prokládání dat křivkami nabízejí takové funkce. Polynomy lze naštěstí užívat i pro funkce složitějších tvarů, jestliže původní data vhodným způsobem přetransformujeme. Tabulka 1 za Obr. 7.1 ukazuje několik funkcí, které mohou být snadno převedeny na polynomický tvar. Při takovýchto transformacích se ale mění i charakter rozptýlení jednotlivých hodnot kolem regresní funkce. Klasické stanovení regresních konstant je založeno na metodě nejmenších čtverců (viz dále), která dává nejpřesnější výsledky, pokud je uvedený rozptyl konstantní, nezávislý na hodnotách nezávisle proměnné  $x$  (tzv. homoskedasticita). Jestliže se transformací výrazně změnil charakter rozptylu, je třeba transformované hodnoty  $y$  přenásobit vhodnými váhami. Blíže k těmto otázkám viz např. [3].



**Obr. 7.1.** Tvary různých funkcí pro aproximování naměřených hodnot. Různé křivky u jednotlivých funkcí odpovídají různým hodnotám regresních konstant.

**Tabulka 1.** Transformace různých funkcí na polynomický tvar [1].

Původní funkce	Transformace	Transformovaná funkce
$w = a + \frac{b}{x} + \frac{c}{x^2} + \dots + \frac{k}{x^q}$	$t = \frac{1}{x}$	$y = a + bt + ct^2 + \dots + kt^q$
$w = ae^{bx}$	$y = \ln w$	$y = \ln a + bx$
$w = \exp(a + bx + cx^2)$	$y = \ln w$	$y = a + bx + cx^2$
$w = ax^b$	$y = \ln w$ $t = \ln x$	$y = \ln a + bt$
$w = \frac{1}{a + bx + cx^2 + \dots + kx^q}$	$y = \frac{1}{w}$	$y = a + bx + cx^2 + \dots + kx^q$
$w^n = \frac{x^m}{a + bx + cx^2 + \dots + kx^q}$	$y = \frac{x^m}{w^n}$	$y = a + bx + cx^2 + \dots + kx^q$
$w = \ln(a + bx + cx^2)$	$y = \exp(w)$	$y = a + bx + cx^2$

Rozložení naměřených hodnot někdy neodpovídá žádné z předdefinovaných funkcí a je vhodnější navrhnout svůj vlastní výraz  $y = f(x)$ . Obrázek 1 může pomoci při jeho hledání. Regresní konstanty by měly být takové, že vzdálenosti mezi jednotlivými naměřenými hodnotami a odpovídajícími vypočtenými hodnotami budou minimální. Pro tento účel se zpravidla používá **metoda nejmenších čtverců** [4 – 6], která minimalizuje součet čtverců rozdílů mezi naměřenými a vypočtenými hodnotami  $y$ :

$$SS_{\text{res}} = \sum (y_{j,\text{meas}} - y_{j,\text{calc}})^2 = \min ! \quad (7.2)$$

Index *res* znamená residuální (zbytkový), *meas* – měřený, a *calc* – vypočtený. Součet se dělá přes všech  $n$  hodnot  $y(x_j)$ . Řešiče, instalované v univerzálních programech, tuto minimalizaci usnadňují. Jedna ukázka byla v Kapitole 5 (příklad 1); dvě další aplikace budou ukázány později v této kapitole. Použití těchto řešičů má velkou výhodu: není nutná žádná transformace regresní funkce ani přiřazování vah jednotlivým hodnotám, neboť se nemění charakter jejich rozptýlení kolem regresní funkce.

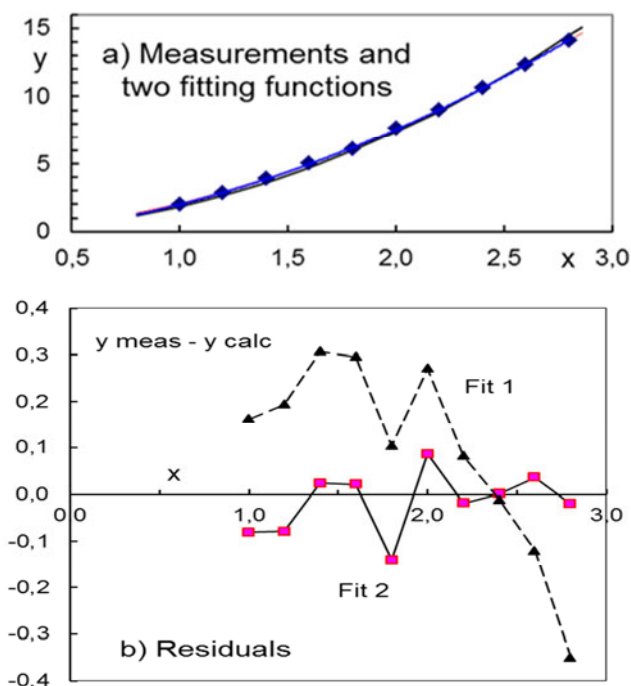
## Hodnocení jakosti aproximace

Někdy dá jednoznačnou odpověď o kvalitě aproximace vizuální posouzení. Velmi jednoduchou veličinou pro hodnocení je **koeficient determinace**  $r^2$ , vysvětlený v předchozí kapitole. Čím je  $r^2$  blíže k 1, tím je aproximace lepší. V případě pochybností je vhodné udělat test významnosti  $r$ , popsany v Kapitole 8.

Pro podrobnější charakterizování jsou vhodná takzvaná **rezidua**. Ta jsou definována jako rozdíly mezi naměřenými a vypočtenými hodnotami,

$$\Delta_j = y_{j,\text{meas}} - y_{j,\text{calc}} \quad (7.3)$$

a kreslí se jako funkce nezávisle proměnné  $x$ . (Platí  $SS_{\text{res}} = \sum (\Delta_j)^2$ ). Rozdíly mezi dvěma křivkami bývají velmi dobře viditelné, což je užitečné, jestliže se obě křivky, kreslené v původních souřadnicích, překrývají. (Obr. 7.2a ukazuje dvě aproximační funkce, obr. 7.2b ukazuje rezidua  $\Delta(x)$ .) Další informace o kvalitě aproximace vyplývá z rozložení reziduí. Fit 2 s náhodně rozptýlenými kladnými a



**Obr. 7.2.** a) Naměřená data ( $y$ ) vyrovnaná dvěma funkcemi (Fit 1 a Fit 2), b) Průběh reziduí  $\Delta_j = y_{j,\text{meas}} - y_{j,\text{calc}}$ ; Fit 2 je lepší než Fit 1.

zápornými hodnotami  $\Delta$  je vhodnější, než Fit 1 se systematickou postupnou změnou  $\Delta$  z kladných hodnot do záporných s rostoucím  $x$ .

Výhodná jsou **relativní** (neboli **normalizovaná**) **rezidua**,

$$A_{j,\text{rel}} = (y_{j,\text{meas}} - y_{j,\text{calc}}) / y_{j,\text{calc}} \quad (7.4)$$

která nezávisí na měřítku  $y$ .

Rozdíly mezi naměřenými hodnotami a odpovídajícími hodnotami na regresní křivce mohou sloužit ke třem účelům: 1) jejich kvadráty se používají v kritériu (7.2) pro optimalizaci, 2) mohou být použity pro ověření, je-li rozložení jednotlivých bodů kolem regresní křivky normální (což je podmínka použitelnosti metody nejmenších čtverců), a 3) slouží pro vytvoření konfidenčního pásu kolem regresní křivky. Tento pás může být použit pro spolehlivější předpovědi, než jsou předpovědi založené pouze na regresní funkci.

Nyní vysvětlíme dvě metody pro stanovení regresních konstant: analytickou a numerickou.

### **Analytické stanovení regresních konstant**

Konstanty v regresní funkci (7.1) se určují z podmínky minima součtu čtverců odchylek hodnot vypočtených od naměřených,  $SS_{\text{res}}$ . Minimum analytické funkce se nachází tam, kde je její derivace rovna nule. Vypočtené hodnoty závisí na regresních konstantách, kterých je zpravidla několik. Je tedy nutno vytvořit parciální derivace součtu (7.2) podle jednotlivých konstant, a každou derivaci položit rovnou nule [3 – 5]. Tím dostaneme soustavu rovnic a jejím vyřešením pak získáme hledané konstanty. Postup ukážeme na lineární regresi

$$y = a + bx \quad (7.5)$$

Rovnice (7.2) vypadá (součet čtverců  $SS_{\text{res}}$  je zde zapsán jako  $S$ ):

$$S = \sum (a + bx_j - y_{j,\text{meas}})^2 \quad (7.6)$$

Derivace  $\partial S / \partial a$  a  $\partial S / \partial b$  jsou

$$\partial S / \partial a = 2 \sum (a + bx_j - y_{j,\text{meas}}) \quad , \quad \partial S / \partial b = 2 \sum (a + bx_j - y_{j,\text{meas}}) x_j \quad (7.7)$$



Položíme-li oba výrazy rovny nule, dostaneme soustavu dvou lineárních rovnic o dvou neznámých  $a$  a  $b$ . Jejím řešením obdržíme

$$b = \frac{n \sum x_j y_j - \sum x_j \sum y_j}{n \sum x_j^2 - (\sum x_j)^2}, \quad a = \frac{\sum y_j - b \sum x_j}{n} \quad (7.8)$$

Podobný postup lze užít i pro další regresní funkce a vede obecně na soustavu  $n$  lineárních rovnic pro  $n$  neznámých regresních constant. Jednodušší ale je použít vhodný počítačový řešič, vysvětlený v Kapitole 5 a v následujícím odstavci.

### Počítačem podporované stanovení regresních konstant

Univerzální počítačové programy jako Excel, Matlab nebo Mathcad obsahují řešitele, které mohou najít minimum zvolené závislosti. To činí stanovení regresních konstant velmi snadným. Je pouze nutné připravit řadu naměřených hodnot ( $y_{j,\text{meas}}$ ) a řadu hodnot  $y_{j,\text{calc}}$ , vypočtených prostřednictvím rovnice (7.1) pro stejné hodnoty  $x_j$  s použitím příslušných parametrů, např.  $a$ ,  $b$  v rovnici (7.5). Řešitel na příkaz minimalizovat výraz (7.2) nebo (7.5) najde měněním  $a$ ,  $b$  jejich nejlepší hodnoty s použitím vlastního algoritmu. Pro tyto výpočty musí být buňky pro regresní konstanty ( $a$ ,  $b$ ) připraveny předem, jakož i buňka obsahující výraz (7.1). Hledání nejlepších hodnot regresních konstant začíná tím, že těmto konstantám přiřadíme počáteční hodnoty. Potom „požádáme“ řešič (Solver), aby našel takové hodnoty konstant, pro které bude obsah buňky se vzorcem (7.6), tj. součet čtverců rozdílů, minimální. Tato hodnota ( $SS_{\text{res}}$ ) také charakterizuje kvalitu aproximace a je užitečná, jestliže je nutno minimalizaci opakovat.

Určování regresních konstant pomocí Řešitele vyžaduje určitou praxi. Kvalita vypočtených „nejlepších“ hodnot regresních konstant někdy závisí na jejich počátečních hodnotách. V nejhorším případě optimalizační proces nekonverguje a je nutno zvolit jiné počáteční hodnoty. Kromě toho, Řešitel zpravidla hledá konstanty, zajišťující minimum z matematického hlediska, a může navrhnout i hodnoty, které nemají reálný smysl. Navíc, běžný optimalizační algoritmus obvykle hledá jenom nejbližší extrém a „netuší“, že může existovat extrémů více. (K této otázce viz též Kapitulu 12.)

Někdy je nutno provést optimalizační výpočty ve dvou nebo více krocích, protože první optimalizace zřejmě nedala přesné hodnoty. Konstanty získané v prvním procesu se pak užijí jako vstupní hodnoty při opakovaném procesu. Je obecně

vhodné nechat vždy proběhnout optimalizační proces alespoň ještě jednou; porovnání součtu čtverců rozdílů (7.2) informuje, jestli hledání minima je u konce.

Stanovení regresních konstant bude nyní ilustrováno na dvou příkladech.

### Příklad 1. Odezva na zatížení viskoelastického materiálu

Časový průběh deformace součásti z viskoelastického materiálu při konstantním zatížení připomíná exponenciální funkci. Jednoduchá exponenciální funkce ale často není vyhovující. Lepší aproximací může být tzv. Pronyho řada, což je součet několika exponenciálních funkcí,

$$y = a_0 + a_1 \exp(-t/\tau_1) + a_2 \exp(-t/\tau_2) + \dots \quad (7.9)$$

$a_1, a_2, \dots$  jsou konstanty a  $\tau_1, \tau_2, \dots$  jsou takzvané relaxační časy, které jsou také neznámé. Obrázek 7.3 ukazuje časový průběh vnikání indentoru do PMMA (polymetylmetakrylátu). Pro aproximaci byla užita následující regresní funkce [6]:

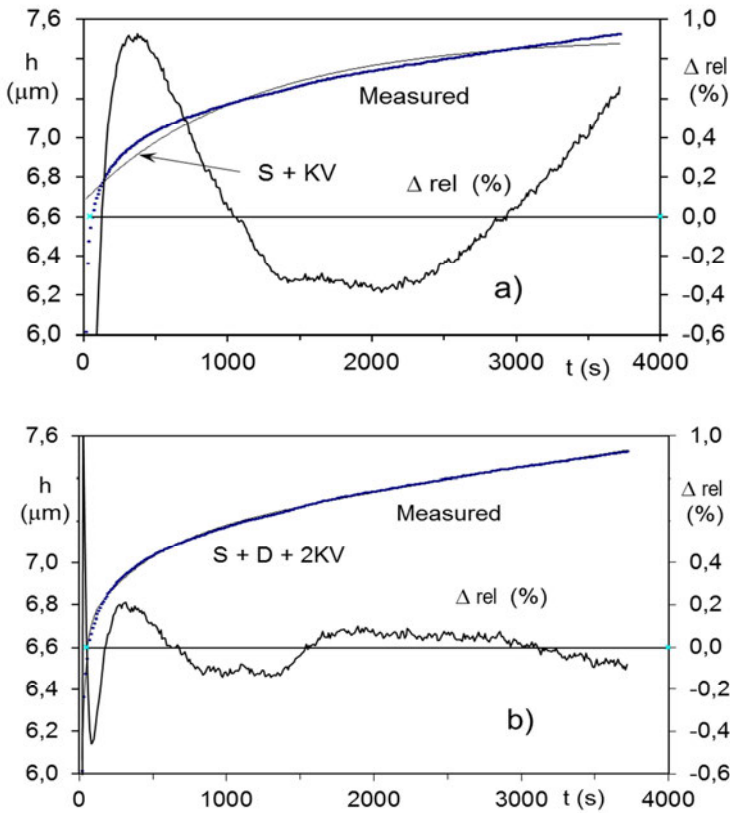
$$y(t) = F K [A_0 + c_v t - \sum B_j \exp(-t/\tau_j)] \quad (7.10)$$

$F$  je zatížení indentoru,  $K$  je konstanta odpovídající jeho geometrii, a  $A_0, c_v, B_j$  a  $\tau_j$  ( $j = 1, 2, 3, \dots$ ) jsou regresní konstanty, nalezené metodou nejmenších čtverců. Obrázek 7.3 na následující straně ukazuje dvě aproximace, se třemi a šesti regresními konstantami. Pro srovnání jsou u obou případů v grafech také ukázána relativní residua  $\Delta_{rel}$ . Ta jsou definována rovnicí (7.4) a nezávisí na měřítku pro  $y$ . Residua mohou pomoci rozlišit různé aproximace, obzvláště jestliže ve společných souřadnicích  $y(x)$  vypadají téměř stejně; viz též Obr. 7.2. Ve vyšetřovaném případě je aproximace se šesti konstantami zřetelně lepší.

**POZNÁMKA:** Komerční programy pro analýzu konstrukcí metodou konečných prvků umožňují práci s Pronyho řadami.

### **Konfidenční pás pro předpověděné hodnoty**

Regresní funkce obvykle slouží pro předpovědi sledované veličiny. Jestliže ale byla vytvořena z hodnot, vykazujících velký rozptyl, předpověď nebude velmi spolehlivá. To lze zlepšit vytvořením konfidenčního pásu kolem regresní funkce. Můžeme-li předpokládat, že rozložení naměřených bodů  $y_j$  kolem regresní funkce  $y_{reg}(x)$  je normální a odpovídající reziduální rozptyl (stanovený z mnoha bodů) je



**Obr. 7.3.** Vnikání indentoru do PMMA při konstantním ztížení [6]. Měření (tečkovaně) a dvě aproximace (plné tenké čáry).  $h$  – hloubka,  $t$  – čas. a) Model S+KV (3 konstanty), b) Model S+D+2KV (6 konstant). Model S+D+2KV aproximuje měřené hodnoty velmi dobře; naměřené a vypočtené hodnoty splývají a rozdíly jsou viditelné pouze pomocí reziduí  $\Delta_{\text{rel}}$  (tenké cik-cak čáry). S - pružina, D - tlumič, KV - Kelvinovo-Voigtovo těleso (pružina a tlumič paralelně).

konstantní, nezávislý na  $x$ , je možné sestrojít hranice konfidenčního pásu (viz Obr. 7.4 pro následující Příklad 2) jako

$$y_{U,L}(x) = y_{\text{reg}}(x) \pm |u_{\alpha}| s_{\text{res}} \quad (7.11)$$

U a L označují horní a dolní mez tohoto pásu,  $u_\alpha$  je  $\alpha$ -kvantil standardního normálního rozdělení, a  $s_{\text{res}}$  je residuální směrodatná odchylka, definovaná jako

$$s_{\text{res}} = \sqrt{\{SS_{\text{res}}/\nu\}} \quad (7.12)$$

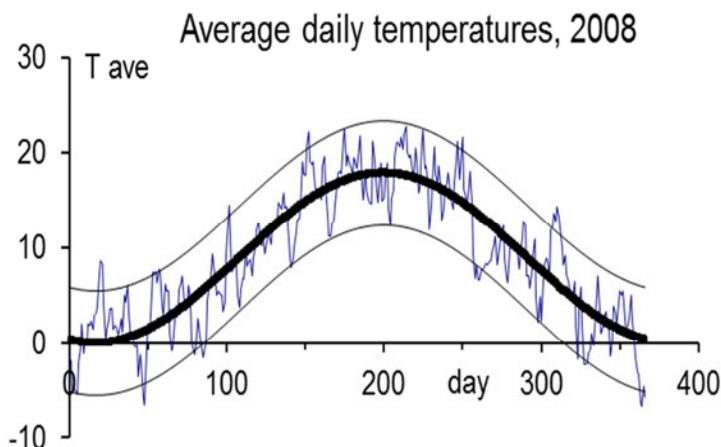
$SS_{\text{res}}$  je součet čtverců rozdílů experimentálních hodnot a odpovídajících bodů na regresní křivce, a  $\nu$  je počet stupňů volnosti, rovný počtu naměřených hodnot mínus počet regresních konstant (např. dvě u lineárního vztahu). Pravděpodobnost, že předpověděná hodnota bude vně mezí ( $y_{U,L}$ ), je  $2\alpha$ .

### Příklad 2. Vyjádření průměrných denních teplot během roku

Venkovní teploty kolísají během dne i během roku. Nicméně tato kolísání vykazují určitou pravidelnost (den, noc, léto, zima...). Přihlédneme-li k takovéto pravidelnosti, mohou předpovědi teploty v určitém okamžiku být přesnější. Například v Ústí nad Orlicí byly měřeny teploty během roku 2008. Pro průměrné denní teploty byla použita kosinová funkce:

$$T = T_0 + A \cos [\pi(x_j - x_0)/186] \quad (7.13)$$

$x_j$  je pořadové číslo dne (1. leden má číslo 1) a  $T_0$ ,  $A$  a  $x_0$  jsou konstanty (186 dní je poloviční délka sledovaného přestupného roku 2008). Obrázek 7.4 ukazuje jednotlivé průměrné teploty a graf regresní funkce (7.13) s konstantami nalezenými



**Obr. 4.** Průměrné denní teploty  $T_{\text{ave}}$  v Ústí během roku 2008. Naměřené hodnoty (spojené tenkými čarami), regresní křivka (tlustá čára) a konfidenční pás pro spolehlivost 90%.

řešitelem v Excelu. Jejich hodnoty jsou  $T_0 = 10,86^\circ\text{C}$ ,  $A = -9,41^\circ\text{C}$ ,  $x_0 = 14,17$ . Residuální směrodatná odchylka byla  $3,236^\circ\text{C}$ . Obrázek 7.4 také ukazuje interval spolehlivosti pro teploty (viz dále).

Nyní předpovíme průměrnou teplotu např. pro 23. říjen a určíme také interval spolehlivosti, který by měl obsahovat skutečnou průměrnou teplotu s pravděpodobností 90%.

V rovnici (7.13) měl 23. říjen pořadové číslo  $x = 297$ , a předpověděná teplota je

$$T = 10,86 - 9,41 \cos[\pi(297 - 14,17)/186] = 9,52^\circ\text{C}.$$

Konstrukce pásu spolehlivosti. Konfidenční pás pro teploty je možno sestavit pomocí rovnice (7.11). Pro residuální směrodatnou odchylku  $s_{\text{res}} = 3,236^\circ\text{C}$  a 5%–kvantil normovaného normálního rozdělení ( $=1,645$ ) je poloviční šířka intervalu spolehlivosti rovna  $\Delta = s_{\text{res}} \times u_{0,05} = 3,236 \times 1,645 = 5,32$ , a dolní a horní mez 90% konfidenčního intervalu jsou:  $T_L = 9,52 - 5,32 = 4,20^\circ\text{C}$ ,  $T_U = 9,52 + 5,32 = 14,84^\circ\text{C}$ . Průměrná teplota, naměřená pro tento 297. den, je  $7,43^\circ\text{C}$ , a leží uvnitř konfidenčních mezí; viz Obr. 7.4.

Z grafu je také vidět, že několik teplot během roku je vně konfidenčních mezí. To je pochopitelné, protože tyto meze byly sestaveny pro konfidenci 90%; deset procent naměřených hodnot tedy může ležet mimo pás spolehlivosti.

### Vícenásobná regrese

Často je nutno vyjádřit, jak proměnná  $y$  závisí na několika vstupních veličinách:

$$y = f(x_1, x_2, \dots) \quad (7.14)$$

Nejjednodušší je lineární vztah

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \dots \quad (7.15)$$

Konstanty  $a_0, a_1, \dots$  lze získat jednoduše vícenásobnou lineární regresi, dostupnou v univerzálních počítačových programech. Zde bude na příkladu ukázáno použití funkce LINREGRESE v Excelu.

#### Příklad 3. Vícenásobná lineární regrese

Máme  $n = 10$  hodnot pevnosti  $y$  hliníkové slitiny měřených pro různý podíl hořčíku ( $x_1$ ) a různé teploty tepelného zpracování ( $x_2$ ); viz tabulku níže. Předpokládáme, že

pevnost na nich závisí jako  $y = a_0 + a_1x_1 + a_2x_2$ .

<u>Vstupní data</u>		
$y$	$x_1(\%)$	$x_2(^{\circ}\text{C})$
234	10,3	800
256	9,8	840
290	9,6	920
248	10,0	820
255	9,7	870
260	10,5	850
285	9,4	900
250	9,1	860
244	9,9	830
270	10,2	890

Řešení. Nejprve vytvoříme (“kliknutím a tažením”) prázdné pole s pěti řádky a s počtem sloupců rovným počtu regresních konstant; v našem případě 3. V následujícím kroku otevřeme menu „Vložit funkci“, najdeme LINREGRESE, vložíme pole hodnot  $y$  do hořejšího okénka a pole všech vstupních hodnot  $x_1$ ,  $x_2$  do okénka pod ním, a zapíšeme slovo PRAVDA (popř. TRUE) do dvou spodních políček, a současně stiskneme klávesy CTRL, SHIFT, ENTER. To je vše. Tabulka s výsledky je dole. Čtenář si může tuto proceduru zopakovat.

#### Tabulka výsledků

a2	a1	a0
0,46907	5,489513	-197,333
0,065626	5,815328	96,55583
0,891138	6,640678	#N/A
28,65067	7	#N/A
2526,91	308,6903	#N/A

Jednotlivé řádky tabulky obsahují: řádek 1: regresní konstanty zleva doprava:  $a_2$ ,  $a_1$ ,  $a_0$ ; řádek 2: směrodatné odchylky jednotlivých regresních konstant; řádek 3: koeficient determinace  $r^2$  a směrodatná odchylka  $y$ ; řádek 4: F-statistika a počty stupňů volnosti, potřebné pro testování, jestli vztah mezi  $y$  a nezávislými veličinami není pouze náhodný; řádek 5: regresní součet čtverců  $SS_{\text{reg}}$  a residuální

součet čtverců  $SS_{\text{res}}$ , definovaný v kapitole 5. Tři buňky s #N/A neobsahují žádné hodnoty. Podrobné vysvětlivky lze najít v menu u příkazu *LINREGRESE*. Koefficient determinace  $r^2 = 0,891$  (viz řádek 3), což je přijatelné. Čtenář může vypočítat  $y$  pro zvolené hodnoty  $x_1, x_2$  a porovnat s tabulkou vstupních hodnot.

Hledaná regresní funkce (viz konstanty v prvním řádku hořejší tabulky) je tedy

$$y = -197,333 + 5,48951 x_1 + 0,46907 x_2$$

Funkce pro vícenásobnou lineární regresi může být někdy použita i pro nelineární regresi. Například, součin

$$y = a x_1 x_2 x_3 \quad (7.16)$$

může být změněn logaritmickou transformací na součet

$$z = b + u_1 + u_2 + u_3 \quad (7.17)$$

kde  $z = \log y$ ,  $b = \log a$ ,  $u_1 = \log x_1$ ,  $u_2 = \log x_2$ , a  $u_3 = \log x_3$ . Podobně je možné transformovat i jednotlivé proměnné. Například ve výrazu  $y = a_1 x_1 + a_2 x^3 + a_3 \sin(cx)$  lze definovat nové proměnné jako  $w = x^3$  a  $v = \sin(cx)$ . Poznámka. Regresní konstanta  $b$  musí být potom transformována zpět do původního systému jako  $a = 10^b$ . Neumožňuje-li předpokládaný výraz vícenásobnou lineární regresi, mohou být regresní konstanty nalezeny užitím vhodného řešiče, jak bylo popsáno dříve.

**POZNÁMKA.** Přímé nalezení parametrů ve Weibullově rozdělení z naměřených hodnot, popsané v Kap. 5, není nic jiného, než určení konstant v regresní funkci.

### Pohyblivé průměry

Někdy nemáme žádnou představu, jakou funkci použít pro aproximaci určité veličiny, která představuje časovou řadu. Vidíme pouze spoustu datových bodů. Někdy lze odhalit trend lépe, jestliže původní data vyhladíme užitím pohyblivých průměrů. Takovýto průměr se vypočítá z  $p$  sousedních bodů; počet  $p$  volíme. Například původní řadu  $x_1, x_2, x_3, x_4, \dots, x_n$  nahradíme (pro zvolený počet  $p = 3$ ) řadou  $y_1 = (x_1+x_2+x_3)/3$ ,  $y_2 = (x_2+x_3+x_4)/3$ ,  $y_3 = (x_3+x_4+x_5)/3, \dots$ ,  $y_{n-2} = (x_{n-2}+x_{n-1}+x_n)/3$ . Tato nová řada má jenom  $n - 2$  členů, obecně  $n - (p - 1)$ , a je hladší než řada původní. Je možné použít několik aproximací, pro různé hodnoty  $p$ , a zvolit tu, která vypadá nejlépe. Univerzální počítačové programy (včetně Excelu) umožňují snadnou aplikaci pohyblivých průměrů na empirická data.

## Literatura ke kapitole 7

1. Kropáč, O.: Metody experimentálního výzkumu. ČVUT, Praha, 1979. 139 s.
2. Pechoč, V.: Vyhodnocování měření a početní metody v chemickém inženýrství. SNTL, Praha, 1981. 226 s.
3. Meloun, M., Militký, J.: Statistické zpracování experimentálních dat. PLUS spol. s r. o., Praha, 1994. 840 s.
4. Felix, M., Bláha, K.: Matematickostatistické metody v chemickém průmyslu. SNTL, Praha, 1962. 336 s.
5. [https://en.wikipedia.org/wiki/Regression\\_analysis](https://en.wikipedia.org/wiki/Regression_analysis). Leden, 14, 2017.
6. Freund, J. E., Perles, B. E.: Modern elementary statistics. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 2006 (12th edition). 576 p.
7. Menčík, J., He, L.H., Němeček, J.: Characterization of viscoelastic-plastic properties of solid polymers by instrumented indentation. Polymer Testing, 30, 2010, 101 - 109.



## 8. Intervaly spolehlivosti, testování hypotéz, a nezbytné množství dat

Důležitým úkolem v počátečním stadiu jakéhokoliv výzkumu je stanovení nezbytného rozsahu experimentů. Experimenty stojí peníze (vzorky a související materiál, přístroje a další zařízení, která je nutno koupit nebo najmout), a také vyžadují čas a pracovní kapacity (plus odpovídající osobní náklady). Jejich rozsah by proto neměl být nepřiměřeně velký, ale odpovídající: 1) účelu zkoumání, 2) důležitosti očekávaných výsledků, a 3) požadované přesnosti. Toto vše by mělo být vyjasněno předem.

Informaci o přesnosti zjištěného parametru lze získat z intervalu spolehlivosti pro něj. Šířka tohoto intervalu závisí na počtu naměřených hodnot. Naopak ze vztahu pro konfidenční interval můžeme dostat vzorec pro potřebný počet měření. Podobně rozhodnutí o testované hypotéze je založeno na hodnotě příslušného testového kritéria, která také závisí na počtu měření. Tento počet, důležitý pro posouzení nebo zvýšení síly testu, je možno zpětně odvodit z tohoto kritéria.

V této kapitole budou ukázány vzorce pro konfidenční intervaly a statistické testy, a také pro počty měření potřebné pro zajištění požadované přesnosti některých charakteristik (průměr, směrodatná odchylka a jiné parametry), počet hodnot pro konfidenční intervaly pro body na regresní přímce, a množství hodnot pro některé statistické testy (rozdíl mezi dvěma soubory, statistická významnost korelace dvou veličin, test dobré shody). Každý oddíl začneme příslušnými vztahy; aplikace budou ukázány na příkladech.

### Konfidenční intervaly a nezbytný počet naměřených hodnot

#### Střední hodnota

Konfidenční interval pro střední hodnotu  $\mu$  je [1 – 5]:

$$\bar{x} - t_{\alpha, \nu} s / \sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha, \nu} s / \sqrt{n} \quad (8.1)$$

$\bar{x}$  je průměrná hodnota ( $= \sum x_j/n$ ),  $n$  je počet naměřených hodnot,  $s$  je výběrová směrodatná odchylka, a  $t_{\alpha, \nu}$  je  $\alpha$ -kritická hodnota  $t$ -rozdělení pro  $\nu = n - 1$  stupňů volnosti. Poloviční šířka intervalu spolehlivosti

$$\Delta = t_{\alpha, \nu} s / \sqrt{n} \quad (8.2)$$

vyjadřuje nejistotu při určení střední hodnoty a odpovídá možné nepřesnosti. Tento výraz je možné přepsat tak, abychom získali počet hodnot (nebo testů) nezbytných pro zajištění, že skutečná střední hodnota  $\mu$  se nebude od průměrné hodnoty  $\bar{x}$  lišit více než o  $\Delta$ :

$$n = (t_{\alpha, \nu} s / \Delta)^2 \quad (8.3)$$

Pravděpodobnost, že rozdíl může být větší, je  $\alpha$ . Vidíme, že potřebný počet měření stoupá výrazně s rostoucím rozptylem jednotlivých hodnot ( $s$ ) a s rostoucím požadavkem na přesnost (tj. s menší přípustnou chybou  $\Delta$ ). Zhruba lze říci, že pro zmenšení nepřesnosti na polovinu je nutno uskutečnit čtyřikrát více zkoušek. Určité zmírnění plyne z faktu, že  $t_{\alpha, \nu}$  klesá s rostoucím  $n$ , zvláště pro nízké počty  $n$ .

Podle rovnice (8.3) by bylo v podstatě možné docílit jakoukoliv přesnost, ale za vysokou cenu. Proto je často nutný určitý kompromis. Někdy je rozumné udělat zpočátku jenom několik zkoušek, abychom získali odhad směrodatné odchylky  $s$ , a vypočítat předběžný počet zkoušek  $n$  pomocí rovnice (8.3). Je-li  $n$  vysoké, je rozumné udělat asi polovinu testů, vypočítat zpřesněné odhady  $s$  a  $n$  (s použitím zpřesněné hodnoty  $t$ ) a pak vykonat zbývající zkoušky.

### Příklad 1.

Tloušťky frézovaných součástí, naměřené na 10 kusech, byly:  $D = 16,02 - 15,99 - 16,03 - 16,00 - 15,98 - 16,04 - 16,00 - 16,01 - 16,01 - 15,99$  mm. Vypočítejte: a) průměrnou hodnotu a směrodatnou odchylku. Předpokládejte, že tloušťky mají normální rozdělení, a vypočítejte b) 95% konfidenční interval pro střední hodnotu, a také c) interval, který bude obsahovat 95% všech tloušťek.

### Řešení.

a) Průměrná hodnota je  $\bar{D} = (\sum D_i)/n = 16,007$  mm a směrodatná odchylka je  $s = 0,01889$  mm.

b) 5%–konfidenční interval pro střední hodnotu, vypočtený podle rovnice (8.1), je (pro oboustrannou kritickou hodnotu  $t_{0,05; 10-1} = 2,2622$ ):

$$16,007 - 2,2622 \frac{0,01889}{\sqrt{10}} < \mu_D < 16,007 + 2,2622 \frac{0,01889}{\sqrt{10}}$$

$$15,993 < \mu_D < 16,020 \text{ mm, nebo } \mu_D \in (\bar{D} \pm \Delta) = 16,007 \pm 0,0135 \text{ mm}$$

Chceme-li zvýšit přesnost stanovení střední hodnoty tloušťky tak, aby se nelišila od vypočtené průměrné hodnoty  $\bar{D}$  o více než  $\Delta = 0,005$  mm, rovnice (8.3) dává

$$n = (t_{\alpha, \nu} s / \Delta)^2 = (2,002465 \times 0,01889 / 0,005)^2 = 57.2$$

2,0025 je kritická hodnota  $t$ -rozdělení  $t_{\alpha, \nu}$  pro hladinu významnosti  $\alpha = 5\%$  a počet stupňů volnosti  $\nu = n - 1 = 57$ . Při výpočtu byla použita směrodatná odchylka  $s = 0,01889$  mm jako dříve, protože lepší odhad nebyl k dispozici. Zpřesněnou střední hodnotu tloušťky tedy obdržíme jako průměr ne méně než 58 hodnot.

c) Jednotlivé hodnoty lze očekávat (za předpokladu normálního rozdělení) uvnitř intervalu  $\bar{D} - u_{\alpha/2} \times s < D < \bar{D} + u_{\alpha/2} \times s$ , kde  $u_{\alpha/2}$  je  $\alpha/2$  – kritická hodnota normovaného normálního rozdělení (odpovídající pravděpodobnosti  $\alpha/2$ , že tloušťka bude větší než hořejší mez konfidenčního intervalu, a  $\alpha/2$ , že bude menší než dolní mez). V našem případě je  $u_{0,025} \approx 1,96$ , takže  $16,007 - 1,96 \times 0,01889 < D < 16,007 + 1,96 \times 0,01889$ ; tedy  $D \in (15,970; 16,044)$ . Spolehlivost předpovědi lze zvýšit, užijeme-li místo konfidenčního interval takzvaný **toleranční interval**, který vysvětlíme později v této kapitole.

## Rozptyl

Interval spolehlivosti pro rozptyl  $\sigma^2$  normálního rozdělení je dán vztahem [1 – 5]

$$(n - 1)s^2 / \chi^2_{\alpha/2, \nu} \leq \sigma^2 \leq (n - 1)s^2 / \chi^2_{1 - \alpha/2, \nu} \quad (8.4)$$

$\chi^2_{1 - \alpha/2, \nu}$  je  $(1 - \alpha/2)$ -kritická hodnota rozdělení chí-kvadrát pro  $\nu = n - 1$  stupňů volnosti;  $\chi^2_{\alpha/2, \nu}$  je  $\alpha/2$ -kritická hodnota. (Poznámka:  $\alpha$ -kritická hodnota je totožná s  $(1 - \alpha)$ -kvantilem.) Šířka interval spolehlivosti závisí na počtu hodnot  $n$ . Počet měření, potřebný pro požadovanou šířku intervalu, je možno získat ze vztahu mezi počtem hodnot a odpovídajícími kritickými hodnotami rozdělení  $\chi^2$ . K tomu mohou posloužit univerzální programy obsahující statistické funkce (např. Excel).

**Příklad 2.**

Stanovte 90%-ní konfidenční interval pro směrodatnou odchylku z Příkladu 1.

Řešení. Rovnice (8.4) udává konfidenční meze pro rozptyl. Pro směrodatnou odchylku  $s = 0,01889$  mm je rozptyl  $s^2 = 0,0003658$  mm<sup>2</sup>. Další nutné hodnoty jsou:  $\nu = n - 1 = 10 - 1 = 9$ ,  $\chi^2_{0,05}(\nu = 9) = 19,9190$ ,  $\chi^2_{0,95}(\nu = 9) = 3,3251$ . Dolní (L) a horní (U) konfidenční meze pro rozptyl jsou

$$s^2(L) = 9 \times 0,0003658 / 19,9190 = 0,0001612 \text{ mm}^2, \quad s_L = 0,0126975 \text{ mm}$$

$$s^2(U) = 9 \times 0,0003658 / 3,3251 = 0,0009901 \text{ mm}^2, \quad s_U = 0,0314659 \text{ mm}$$

a konfidenční meze pro směrodatnou odchylku (počítanou jako odmocnina z rozptylu) jsou  $s_L = 0,0127$  mm,  $s_U = 0,0315$  mm. (Všimněte si velké šířky konfidenčního interval pro  $s$ , ve srovnání s odhadnutou hodnotou  $s = 0,0189$  mm !)

**Parametr exponenciálního rozdělení**

Exponenciální rozdělení hraje velmi důležitou roli např. ve spolehlivosti. Obvykle se jím aproximují doby mezi poruchami z mnoha příčin ve složitých elektrických, mechanických a dalších objektech nebo systémech sestávajících z mnoha částí. Pravděpodobnost poruchy během intervalu  $(0; t)$  je

$$R(t) = e^{-t/T_{\text{mean}}} \quad (8.5)$$

$t$  je čas a  $T_{\text{mean}}$  je střední doba do poruchy nebo mezi poruchami. V praxi se střední doba určuje jako aritmetický průměr naměřených dob do poruchy,

$$T_{\text{mean}} \approx T_{\text{ave}} = \Sigma t_{ij} / n \quad (8.6)$$

sečítá se všech  $n$  hodnot  $t_{ij}$ . Doby do poruchy jednotlivých prvků ale kolísají, a střední doba, vypočtená podle rovnice (8.6), je pouze odhadem. Je proto nutná znalost konfidenčních mezí pro  $T_0$ . Dolní (L) a horní (U) mez jsou dány vztahem

$$T_L = \frac{2r}{\chi^2_{\alpha/2}(\nu)} \bar{t} \leq T_{\text{mean}} \leq \frac{2r}{\chi^2_{1-\alpha/2}(\nu)} \bar{t} = T_U \quad (8.7)$$

$\chi^2_{\alpha/2}(\nu)$  je  $\alpha/2$ -kritická hodnota a  $\chi^2_{1-\alpha/2}(\nu)$  je  $(1-\alpha/2)$ -kritická hodnota  $\chi^2$ -rozdělení pro  $\nu$  stupňů volnosti [1 – 5]. Pravděpodobnost, že skutečná doba do poruchy bude

kratší než  $t_L$  nebo delší než  $t_U$  je  $\alpha$ . Počet stupňů volnosti závisí na uspořádání zkoušek. Jsou-li ukončeny po poruše  $r$  kusů, platí  $\nu = 2r$  [6]. S rostoucím počtem  $r$  se rozdíl mezi dolní a horní kritickou hodnotou zmenšuje. Konfidenční interval pro střední dobu se také zužuje a předpověď  $T_{\text{mean}}$  je přesnější. Tímto způsobem je možné předem stanovit počet porouchaných kusů, při kterém bude možno test ukončit, abychom získali požadovaný interval spolehlivosti pro střední dobu do poruchy.

### Příklad 3.

Deset elektrických součástek bylo testováno pro zjištění dob do poruchy. Takovéto testy mohou u některých zkoušených kusů trvat velmi dlouho. Zkoušky se proto často ukončují po uplynutí určitého času (viz Obr. 5.4 v Kapitole 5). V našem případě bylo trvání zkoušky stanoveno jako  $t_T = 500$  hodin. Během této doby selhalo pouze 6 součástí ( $r = 6$ ), a to v dobách 65 – 75 – 90 – 120 – 250 – 410 hodin. Čtyři součásti test přečkaly. Je zapotřebí odhadnout střední dobu do poruchy a sestrojít oboustranný konfidenční interval (pro stupeň konfidence 90%).

### Řešení.

Střední hodnota a směrodatná odchylka dob do poruchy šesti porušených součástí byly blízké: 168.33 a 136.33 hodin. Je proto možné předpokládat exponenciální rozdělení.

Kumulovaná doba testů, počítaná podle [6], byla:

$$t_{\text{tot}} = \sum_{i=1}^6 t_i + 4 \times t_T = 60 + 75 + 90 + 120 + 250 + 410 + 4 \times 500 = 3010 \text{ hodin}$$

Průměrná doba do poruchy je  $t_{\text{ave}} = t_{\text{tot}}/r = 3010 / 6 = 501,67$  h

Dolní a horní konfidenční mez pro skutečnou střední dobu  $T_{\text{mean}}$ , s ohledem na to, že zkoušky byly ukončeny před porušením všech vzorků, jsou [6]:

$$T_L = \frac{2r}{\chi^2_{\alpha/2}(2r+2)} t_{\text{ave}} \leq t_{\text{mean}} \leq \frac{2r}{\chi^2_{1-\alpha/2}(2r)} t_{\text{ave}} = T_U \quad (8.8)$$

$\chi^2_{\alpha/2}(2r+2)$  je  $\alpha/2$ -kritická hodnota  $\chi^2$ -rozdělení pro  $2r+2$  stupně volnosti, a  $\chi^2_{\alpha/2}(2r)$  is  $\alpha/2$  – kritická hodnota  $\chi^2$ -rozdělení pro  $2r$  stupňů volnosti. Ve vyšetřovaném případě ( $r = 6$ ,  $\alpha = 10\%$ ), byly kritické hodnoty  $\chi^2_{0.05; 14} = 23,685$  a

$\chi^2_{0.95; 12} = 5,226$ . Vložíme-li je, spolu s  $t_{\text{ave}} = 501,67$  h do (8), dostaneme  $T_L = 254,4$  h a  $T_U = 1152,1$  h. Střední dobu do poruchy tedy můžeme očekávat v intervalu  $t_{\text{mean}} \in (254 \text{ h}; 1152 \text{ h})$ .

Tento konfidenční interval, stanovený ze šesti poruch, je velmi široký. Pokud by měl být užší (abychom získali přesnější odhad), bylo by nutné test prodloužit, aby došlo k poruše více součástí, anebo zvýšit počet testovaných kusů; viz [6] nebo kapitolu 20 v [7].

### Hodnoty předpověděné pomocí regresní čáry

Regresní přímka

$$y(x) = a + bx \quad (8.9)$$

se často užívá pro předpověď hodnot  $y$ , odpovídajících určitým hodnotám  $x$ . Konstanty  $a$ ,  $b$  ale byly stanoveny z měřených hodnot, které vykazují určitý rozptyl. Kdybychom použili jinou skupinu měření, dostali bychom poněkud odlišnou regresní přímku. Interval spolehlivosti pro bod na regresní přímce (Obr. 8.1) je [1–4, 8]:

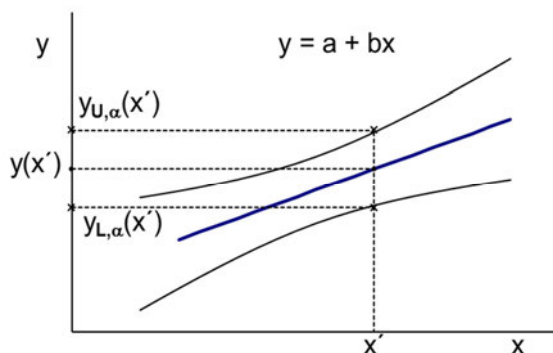
$$y = a + bx \pm t_{\alpha, \nu} s_{\text{res}} \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{(n-1)s_x^2}} \quad (8.10)$$

kde  $s_{\text{res}}(x)$  je zbytková směrodatná odchylka, definovaná jako

$$s_{\text{res}} = \sqrt{\frac{\sum (y_j - a - bx_j)^2}{n - 2}} \quad (8.11)$$

Všechny body konfidenčních mezí tvoří pás spolehlivosti, který je nejužší pro  $x = \bar{x}$ . Vzorec pro  $x = \bar{x}$  je identický se vztahem (1) pro konfidenční interval pro střední hodnotu. Podobně šířka konfidenčního intervalu v místě  $\bar{x}$  klesá s odmocninou z  $1/n$ . Konfidenční pás pro regresní přímku tedy může být zúžen, použijeme-li vyšší počet hodnot pro získání regresních konstant. Potřebný počet hodnot pro zajištění požadované přesnosti lze najít podobně jako u střední hodnoty. Úprava rovnice (8.3) dá

$$n = (\tau_{\alpha, \nu} s_{\text{res}} / \Delta)^2 \quad (8.12)$$



**Obr. 8.1.** Konfidenční interval pro regresní přímku.

### Toleranční meze

Někdy potřebujeme znát interval, který bude obsahovat  $P$  procent z celého souboru. Velmi důležitý případ je, má-li soubor (přibližně) normální rozdělení. Jsou-li známy parametry  $\mu$ ,  $\sigma$  základního souboru, potom  $P\%$  všech kusů bude ležet uvnitř mezí  $\mu \pm u_{1-P/2}\sigma$ , kde  $u_{1-P/2}$  je  $(1-P/2)$ -kritická hodnota normovaného normálního rozdělení.

Často jsou ale známy pouze výběrové charakteristiky  $\bar{x}$  a  $s$  místo  $\mu$  a  $\sigma$ . V takovém případě je nemožné stanovit odpovídající meze s určitostí. Můžeme pouze stanovit tzv. toleranční meze, které budou obsahovat podíl  $P$  základního souboru se zvolenou pravděpodobností  $\gamma$ . Dvoustranné toleranční meze (dolní a horní) lze vypočítat podle vztahu [5, 9, 10]

$$x_L, x_U = \bar{x} \pm ks \quad (8.13)$$

$\bar{x}$  a  $s$  jsou průměr a směrodatná odchylka vzorku o velikosti  $n$ , a  $k$  je konstanta, závislá na  $P$ ,  $n$ ,  $\gamma$ . Koeficienty  $k$  pro vybrané hodnoty  $P$ ,  $n$ ,  $\gamma$  lze najít ve statistických tabulkách [9, 10].

#### Příklad 4.

V Příkladu 1 byl počítán interval, který by měl obsahovat 95% všech frézovaných součástek: (15,970 mm; 16,044 mm). Tento interval byl vypočten za předpokladu, že  $\bar{D} = 16,007$  mm a  $s = 0,01889$  mm jsou parametry celkového základního

souboru. Tyto hodnoty ale byly vypočteny pouze ze vzorku  $n = 10$  kusů. Spolehlivější informace bude získána pomocí vzorce pro *toleranční interval*. Pro podíl  $P = 0.95$ , spolehlivost předpovědi  $\gamma = 0.90$ , a  $n = 10$  je  $k = 3,18$ ; viz [9, 10]. Interval, obsahující 95% všech kusů, je

$$\bar{D} \pm k \times s = 16,007 \pm 3,18 \times 0,01889 = 16,007 \pm 0,060 = (15,947; 16,067) \text{ mm}$$

Je širší než původní interval. Rozdíl mezi oběma interval se zvětšuje s rostoucí směrodatnou odchylkou a klesajícím množstvím naměřených hodnot, obzvláště je-li  $n < 10$ .

## Testování hypotéz

### Rozdíl dvou průměrů

Testové kritérium závisí na tom, je-li rozptyl obou vzorků (přibližně) stejný nebo ne. Zde si všimneme pouze případu s různými rozptyly, který je univerzálnější. V tomto případě je testová charakteristika

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (8.14)$$

a bude porovnána s kritickou hodnotou  $t$ -rozdělení pro hladinu významnosti  $\alpha$  a  $\nu$  stupňů volnosti, definovanou jako [8, 11]

$$\nu = \frac{[(s_1^2/n_1) + (s_2^2/n_2)]^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}} \quad (8.15)$$

### Příklad 5.

Byla navržena modifikovaná procedura pro přípravu určitého druhu plastické hmoty. Náklady byly nižší, ale také naměřená pevnost byla o něco nižší. Je nutno zjistit, jestli pokles pevnosti byl náhodný, nebo je-li statisticky významný. Charakteristiky obou vzorků byly:

Vzorek 1:  $n_1 = 31$ ,  $\bar{x}_1 = 31,03$ ,  $s_1^2 = 1,41$ ,  $s_1 = 1,19$

Vzorek 2:  $n_2 = 32$ ,  $\bar{x}_2 = 29,87$ ,  $s_2^2 = 1,84$ ,  $s_2 = 1,36$



Vložení těchto hodnot do rovnic (8.14) a (8.15) dává  $\nu = 60,4$  a  $t = 3,615$ . Tato hodnota testové charakteristiky  $t$  je větší než kritické hodnoty  $t$ -rozdělení pro hladinu významnosti  $\alpha = 0,05$  a počet stupňů volnosti  $\nu = 60$ , která je  $t_{0,05; 60} = 2,0003$ . Rozdíl mezi oběma průměrnými hodnotami je významný (na hladině 5%), což potvrzuje, že způsob přípravy má významný vliv na pevnost. Poznámka: Rozdíl je statisticky významný i na hladině významnosti 0,001.

### Porovnání přesnosti dvou měřících metod

Tento test je založen na porovnání rozptylů obou metod. Jejich poměr má  $F$ -rozdělení. Testové kritérium

$$F = s_1^2 / s_2^2 \quad (8.16)$$

bude porovnáno s kritickou hodnotou  $F$ -rozdělení pro  $\nu_1 = n_1 - 1$  a  $\nu_2 = n_2 - 1$  stupňů volnosti,  $F_\alpha(n_1 - 1, \nu_2 = n_2 - 1)$ . Bude-li  $F > F_\alpha$ , nulovou hypotézu (neexistuje rozdíl mezi oběma rozptyly) zamítneme. V opačném případě učiníme závěr, že rozdíl není významný.

### Test významnosti koeficientu korelace dvou veličin

Korelační koeficient  $r (= \sqrt{r^2})$  je někdy velmi vysoký, například 0,9 nebo více, takže můžeme předpokládat, že navržený funkční vztah mezi oběma veličinami je oprávněný. Někdy je však tento koeficient nižší, a my nevíme, jestli hodnoty jedné veličiny skutečně závisí na hodnotách druhé veličiny, nebo jestli s ní jsou korelovány jenom náhodně. V takových případech je užitečný statistický test. Korelační koeficient  $r$  je statisticky významný na hladině významnosti  $\alpha$ , platí-li

$$\frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} \geq t_{\alpha,\nu} \quad (8.17)$$

kde  $t_{\alpha,\nu}$  jednostranná  $\alpha$ -kritická hodnota  $t$ -rozdělení pro  $\nu = n - 2$  stupně volnosti;  $n$  je počet párů hodnot. Chceme-li mít jistotu, že korelační koeficient  $r$  je statisticky významný (na hladině  $\alpha$ ), vyplývá z Rovnice (17), že počet dvojic musí být

$$n \geq 2 + (t_{\alpha,\nu})^2 (1 - r^2)/r^2 \quad (8.18)$$

Protože samotná kritická hodnota  $t_{\alpha,\nu}$  závisí na  $n$ , jsou někdy zapotřebí dva nebo více iteračních kroků pro získání správného počtu hodnot  $n$ .

### Příklad 6.

Korelační koeficient mezi pevností a tvrdostí vyšetřované slitiny, stanovený ze 30 párů hodnot, byl  $r = 0,7$ . Je tato hodnota statisticky významná?

Testové kritérium (8.17) je

$$\frac{|0,7|\sqrt{30-2}}{\sqrt{1-0,7^2}} = 5,1867$$

Jednostranná kritická hodnota  $t$ -rozdělení pro  $\nu = n - 2 = 30 - 2 = 28$  stupňů volnosti je  $t_{0,05;28} = 1,701$  pro hladinu spolehlivosti 5%, a 2,763 pro hladinu spolehlivosti 0.5%. Vypočtená hodnota 5,1867 je mnohem vyšší než obě kritické hodnoty. Činíme tedy závěr, že mezi pevností a tvrdostí tohoto materiálu existuje silná korelace.

**Testy dobré shody.** Tyto testy se užívají pro ověření, jestli experimentální data mají určité rozdělení pravděpodobnosti. Nejčastěji se užívají dva druhy testů: Kolmogorovův-Smirnovův a  $\chi^2$ . **Kolmogorovův-Smirnovův test** vyšetřuje rozdíly mezi empirickou distribuční funkcí a referenční funkcí. Tyto rozdíly se vypočítají pro všechny hodnoty empirického rozdělení a maximální rozdíl (který má v případě platnosti nulové hypotézy také určité rozdělení) se porovná s kritickou hodnotou, kterou je možno najít ve speciálních tabulkách [9, 10]. Je-li tento rozdíl větší, zamítneme nulovou hypotézu a řekneme, že rozdělení naměřených hodnot neodpovídá předpokládanému rozdělení. V opačném případě hypotézu přijmeme. Následující příklad ilustruje použití tohoto testu.

### Příklad 7.

Je nutné ověřit, jestli dodávka NaOH pochází z výroby se střední hodnotou koncentrace  $\mu = 42.3$  a směrodatnou odchylkou  $\sigma = 1.5$ . Výsledky  $N = 50$  analýz byly následující:

44.0	45.0	42.5	41.9	41.9	41.2	41.0	43.3	42.3	40.0
42.2	43.6	44.4	40.2	41.5	42.7	45.0	43.0	43.2	41.1
42.4	42.4	40.8	39.2	43.2	44.2	42.3	43.3	44.7	41.1
41.9	42.7	41.4	44.4	43.0	40.1	42.0	39.7	42.9	42.7
41.5	38.8	43.4	43.9	40.8	40.5	42.8	41.6	43.0	42.8

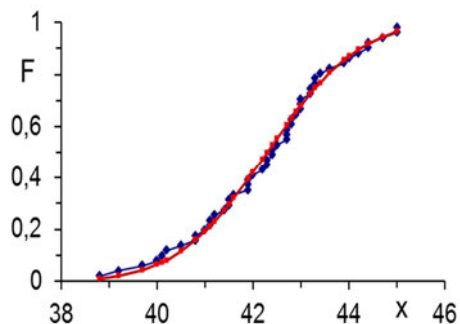
Výběrové charakteristiky byly: průměrná koncentrace  $\bar{x} = 42.27$ , směrodatná odchylka  $s = 1.48$ , a minimální a maximální hodnoty:  $x_{\min} = 38.8$ ,  $x_{\max} = 45.0$ .

**Řešení.** Pro Kolmogorovův-Smirnovův test (zkráceně K-S) potřebujeme hodnoty distribuční funkce pro rozdělení empirické (tj. naměřené) a pro předpokládané teoretické rozdělení. Nejprve seřadíme naměřené koncentrace ( $x$ ) od nejnižší ( $x_1$ ) do nejvyšší ( $x_N$ ). Odpovídající hodnoty empirické distribuční funkce byly počítány jako  $F_{j,\text{emp}} = j/(N+1)$ . Potom byly vypočteny hodnoty teoretické distribuční funkce pro stejné kvantily  $x_j$ , ale za předpokladu, že přísluší normálnímu rozdělení s parametry  $\mu = 42,3$  a  $\sigma = 1,5$ . (V Excelu lze pro výpočet užít příkaz `NORMDIST( $x_j, \mu, \sigma, \text{TRUE}$ )`.) Nakonec byly stanoveny rozdíly obou distribučních funkcí jako  $\Delta_j = |F_{j,\text{emp}} - F_{j,\text{theor}}|$ . Tabulka s částí výsledků je na následující straně a Obr. 8.2 ukazuje obě distribuční funkce.

Maximální rozdíl mezi empirickou a teoretickou distribuční funkcí byl  $\Delta_{\max} = 0.05612$  (pro  $j = 28$ ; viz tabulku). To je mnohem méně než kritická hodnota K-S rozdělení [10], která je  $D_{0.05}(50) = 0.188$  pro hladinu spolehlivosti  $\alpha = 0.05$  a počet hodnot  $N = 50$ . Proto můžeme přijmout hypotézu, že parametry dodávky jsou  $\mu = 42.3$  a  $\sigma = 1.5$ .

Tabulka k příkladu 7.

$j$	$x_j$	$F_{j,\text{emp}}$	$F_{j,\text{theor}}$	$\Delta_j$
1	38.8	0.01961	0.00982	0.00979
2	39.2	0.03922	0.01938	0.01983
.....				
<u>28</u>	<u>42.7</u>	<u>0.54902</u>	<u>0.60514</u>	<u>0.05612</u>
.....				
50	45.0	0.98039	0.96407	0.01632



**Obr. 8.2.** Kolmogorovův-Smirnovův test – srovnání empirického a teoretického rozdělení pro 50 hodnot.

Jiný test dobré shody je  $\chi^2$ -test (chí-kvadrát test). Ten je založen na myšlence, že pokud má výběrový soubor předpokládané rozdělení, potom rozdíly mezi empirickými a teoretickými (předpokládanými) hodnotami mají normované

normální rozdělení, a součet jejich kvadrátů má tedy rozdělení  $\chi^2$ . Použití tohoto testu je následující. Data jsou rozdělena do  $m$  tříd, a zjistí se četnosti jejich výskytu v jednotlivých třídách. Potom se zjistí četnosti pro předpokládané rozdělení a vypočítají se rozdíly obou četností v jednotlivých třídách. Hodnotící kritérium je [11]:

$$\chi^2 = \sum_{j=1}^m \frac{(n_j - np_j)^2}{np_j} \quad (8.19)$$

Je-li hodnota, vypočtená podle vzorce (8.19), větší než  $\alpha$ -kritická hodnota  $\chi^2$ -rozdělení pro  $(m-k-1)$  stupňů volnosti, hypotéza “vzorek má předpokládané rozdělení” se zamítne na hladině významnosti  $1-\alpha$ . V opačném případě hypotézu přijmeme. POZNÁMKA:  $k$  je počet parametrů distribuční funkce teoretického rozdělení. Podmínkou pro užití  $\chi^2$ -testu a počtu tříd  $m$  je, že  $np_j$  musí být rovno nebo větší než 5 (tj.  $np_j \geq 5$ ) pro každé  $j$ .

### Příklad 8.

Hypotézu z Příkladu 7 (50 vzorků pochází ze základního souboru s parametry  $\mu = 42.3$  a  $\sigma = 1.50$ ) nyní otestujeme pomocí  $\chi^2$ -testu. Vstupní hodnoty jsou stejné jako v předchozím příkladu.

Řešení. Rozmezí možných koncentrací (38,0 – 45,0) bylo rozděleno do 13 intervalů o šířce 0,5, a byly vypočteny četnosti jim odpovídající, podobně jako v předchozím příkladu.  $\chi^2$ -test ale může být použit jenom tehdy, je-li počet hodnot v každém subintervalu rovný nebo vyšší než 5. Proto byly vytvořeny nové intervaly (celkem 7) tak, že se některé původní intervaly sloučily, takže šířky nových nejsou všude stejné. Tabulka na následující straně obsahuje všechny hodnoty. Sloupec 1 ukazuje koncentrace, sloupec 2 udává odpovídající počty  $n_j$  vzorků v nových třídách, sloupec 3 ukazuje teoretické pravděpodobnosti  $p_j$  koncentrací, vypočítaných pomocí hodnot distribuční funkce, sloupec 4 ukazuje teoretické počty hodnot  $np_j$  pro  $n = 50$ , sloupec 5 udává rozdíly  $n_j - np_j$ , a sloupec 6 ukazuje dílčí hodnoty  $Z_j = (n_j - np_j)^2 / np_j$  pro předpokládané kritérium chí-kvadrát. Výsledná hodnota kritéria podle vzorce (8.19), je  $\chi^2 = 1,720$ . To je mnohem méně, než kritická hodnota  $\chi^2_{0.05}(4) = 9,488$ , odpovídající konfidenční úrovni  $\alpha = 0,05$  a počtu stupňů volnosti  $\nu = m - p - 1 = 4$ ;  $m$  je počet subintervalů (= 7) a  $p$  (=2) je

počet parametrů zkoumaného rozdělení [10, 11]. Můžeme proto připustit, že testovaných 50 odběrů pochází z dodávky s parametry  $\mu = 42.3$  a  $\sigma = 1.50$ ; podobně jako vyplynulo z Kolmogorovova-Smirnovova testu.

### Tabulka k příkladu 8

$x_j$	$n_j$	$p_j$	$np_j$	$n_j - np_j$	$Z_j$
38.51-40.50	7	0.1151	5.755	1.245	0.2693
40.51-41.50	9	0.1818	9.090	-0.090	0.0009
41.51-42.00	5	0.1238	6.190	-1.190	0.2288
42.01-42.50	6	0.1323	6.615	-0.615	0.0572
42.51-43.00	9	0.1266	6.330	2.670	1.1261
43.01-43.50	5	0.1085	5.425	-0.425	0.0333
43.51-45.00	9	0.1760	8.800	0.200	0.0045

### **Bayesovské metody**

Tento výraz označuje pravděpodobnostní metody, které umožňují kombinovat informace o nějakém jevu nebo veličině s předchozími informacemi z měření nebo pozorování. Použití doplňujících informací může zvýšit spolehlivost našich závěrů nebo zmenšit rozsah měření k tomu potřebných.

Bayesovské metody vycházejí z tzv. **Bayesovy věty** [7, 12, 13]. Předpokládejme, že jev  $B$  může nastat, jestliže nastal jiný jev  $A$ . Ten ale může nastat několika způsoby ( $A_1, A_2, \dots, A_n$ ), jež se navzájem vylučují. Pravděpodobnost současného výskytu obou jevů,  $B$  a  $A_j$ , je:

$$P(BA_j) = P(A_j) \times P(B|A_j) \quad (8.20)$$

$P(A_j)$  je pravděpodobnost jevu  $A_j$  a  $P(B|A_j)$  je podmíněná pravděpodobnost se kterou může nastat jev  $B$ , jestliže nastal jev  $A_j$ . Celková pravděpodobnost jevu  $B$  je

$$P(B) = \Sigma P(BA_j) \quad (8.21)$$

sečítá se pro všechny možné případy  $j = 1, 2, \dots, n$ . Bayesova věta se na celou záležitost dívá z opačné strany: „**Jestliže nastal jev  $B$ , jaká je pravděpodobnost, že to bylo v důsledku (nebo po) jevu  $A_j$ ?**“ S využitím rovnic (8.20) a (8.21) a toho, že  $P(BA_j) = P(A_jB)$ , je možno vyjádřit tuto pravděpodobnost jako [7, 12, 13]:

$$P(A_j|B) = P(A_j) \times P(B|A_j) / P(B) \quad (8.22)$$

Celková pravděpodobnost  $P(B)$  ve jmenovateli se počítá z jednotlivých pravděpodobností pomocí rovnic (8.21) a (8.20). Rovnice (8.22) je nejjednodušším tvarem **Bayesovy věty**. Její použití ukážeme na následujícím příkladu.

### Zvýšení spolehlivosti nedestruktivního testování.

Svařované součásti jsou testovány na přítomnost defektů (trhlin). Přístroj používaný pro nedestruktivní zkoušení není dokonalý. Klasifikuje vadu správně (jako vadu) pouze s pravděpodobností 98%, zatímco ve 2% nerozezná trhlinu a součást klasifikuje jako dobrou. Na druhou stranu, přístroj označí 96% dobrých součástí jako dobré, ale 4% klasifikuje chybně jako s trhlinou. Podle dlouhodobých záznamů o inspekcích obsahují 3% všech testovaných součástí trhliny. Otázky jsou: jestliže testovaná součást byla označena jako “vadná” (tj. s defektem), jaká existuje pravděpodobnost, že je skutečně: a) vadná, b) dobrá? A co když je součást klasifikována jako “dobrá”?

Řešení. Jev  $A_1$ : součást obsahuje vadu,  $A_2$ : součást je dobrá.  $P(A_1) = 0,03$ ;  $P(A_2) = 0,97$ . Jev  $B$ : součást je klasifikována jako vadná.  $P(B|A_1) = 0,98$ ;  $P(B|A_2) = 0,04$ . Podíl testovaných součástí, označených jako špatné, je:  $P(B) = 0,03 \times 0,98 + 0,97 \times 0,04 = 0,0682$ .

Případ 1a. Pravděpodobnost, že součást, označená jako špatná, je špatná, je  $P(A_1|B) = P(A_1) \times P(B|A_1) / P(B) = 0,03 \times 0,98 / 0,0682 = 0,431 = 43,1\%$ . Případ 1b. Pravděpodobnost, že součást, označená jako špatná, je ve skutečnosti dobrá, je  $P(A_2|B) = 0,97 \times 0,04 / 0,0682 = 0,569 = 56,9\%$ . [V důsledku vysokého podílu dobrých součástí (98%) je také podíl dobrých, ale zamítnutých dílů, vysoký. Může se proto vyplatit znovu přezkoušet zamítnuté součásti a snížit tak celkové ztráty.]

Jev  $B'$ : Součást je klasifikována jako dobrá.  $P(B'|A_1) = 0,02$ ;  $P(B'|A_2) = 0,96$ . Celkový podíl součástí, označených jako dobré, je:  $P(B') = 0,03 \times 0,02 + 0,97 \times 0,96 = 0,9318$ . Případ 2a. Pravděpodobnost, že součást, označená jako dobrá, je ve skutečnosti špatná, je  $P(A_1|B') = 0,03 \times 0,02 / 0,9318 = 0,00064 \approx 0,06\%$ . Případ 2b. Pravděpodobnost, že součást, označená jako dobrá, je ve skutečnosti dobrá, je  $P(A_2|B') = 0,99936 \approx 99,94\%$ . podobný přístup se užívá i v lékařství, například při hromadném snímkování na rakovinu.

Kombinace různých informací se uplatňuje i v následující úloze.

## Zlepšení odhadu parametrů normálního rozdělení

Střední hodnota  $\mu$  a směrodatná odchylka  $\sigma$  základního souboru s normálním rozdělením nejsou obvykle známy, takže se místo nich pracuje s jejich odhady, výběrovým průměrem a směrodatnou odchylkou  $m$  a  $s$  vypočtenými ze vzorku o velikosti  $n$ . Odhad střední hodnoty je možné zlepšit prostřednictvím intervalu spolehlivosti (8.1). Odhad může být dále zpřesněn, jsou-li k dispozici další informace, například odhady  $m_0$  a  $s_0$  z předchozích měření nebo záznamů. Jestliže je znám počet  $n_0$  těchto hodnot, a můžeme-li předpokládat, že všechny vzorky (nové i staré) pocházejí ze stejného souboru, je možno vypočítat zlepšený průměr jako vážený průměr z obou souborů,

$$m_u = (nm + n_0m_0) / n_u \quad ; \quad n_u = n + n_0 \quad (8.23)$$

$n_u$  je aktualizovaný počet hodnot. Aktualizovaná směrodatná odchylka je

$$s_u = \sqrt{\frac{(n-1)s^2 + (n_0-1)s_0^2 + nm^2 + n_0m_0^2 - n_um_u^2}{n_u - 1}} \quad (8.24)$$

Potom je možno stanovit zpřesněný konfidenční interval pro  $\mu$ , přičemž  $m$ ,  $s$ ,  $n$  ve vzorci (8.1) nahradíme aktualizovanými hodnotami  $m_u$ ,  $s_u$  a  $n_u$ . Není-li  $n_0$  známo, literatura [12 - 14] doporučuje vzorec:

$$n_0 = s^2 / s_0^2 \quad (8.25)$$

Tento vztah je založen na myšlence, že  $m_0$  a  $s_0$  nesou informace odpovídající fiktivnímu vzorku o velikosti  $n_0$ . Čím menší je rozptyl  $s_0^2$  ve srovnání s  $s^2$ , tím důležitější jsou původní výsledky, a tedy tím větší je velikost fiktivního vzorku.

Více o bayesovských metodách se lze dočíst v [7, 12 - 14] a v pramenech tam citovaných.

**Poděkování.** Část této kapitoly byla dříve publikována v kapitole 22 knihy [7].

## Literatura ke kapitole 8

1. Freund, J. E.: Modern elementary statistics. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1981 (6th edition). 561 s.
2. Freund, J. E., Perles, B. E.: Modern elementary statistics. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 2006 (12th edition). 576 s.

3. Suhir, E.: Applied Probability for Engineers and Scientists. McGraw-Hill, New York, 1997. 593 s.
4. Montgomery, D. C., and Runger, G. C.: Applied Statistics and Probability for Engineers. John Wiley, New York, 2006 (4th edition). 784 s.
5. ČSN 010250. Statistické metody v průmyslové praxi.) ÚNM, Praha, 1972.
6. Bednařík, J. a kol.: Technika spolehlivosti v elektronické praxi. Praha, SNTL, 1990. 336 s.
7. Menčík, J.: Concise reliability for engineers. InTech, Rijeka, 2016, 204 s. *Open Access*, viz: <http://www.intechopen.com/books/concise-reliability-for-engineers>, ISBN 978-953-51-2278-4.
8. Felix, M., Bláha, K.: Matematickostatistické metody v chemickém průmyslu. SNTL, Praha, 1962, 336 s.
9. Jílek, M.: Statistické toleranční meze. SNTL, Praha, 1988. 280 s.
10. Likeš, J., Laga, J.: Základní statistické tabulky. SNTL, Praha, 1978. 564 s.
11. ČSN 010253. Statistické metody v průmyslové praxi. Základní neparametrické metody. Vydavatelství Úřadu pro normalizaci a měření, Praha, 1974. 114 s.
12. Martz, H. F., Waller, R. A.: Bayesian Reliability Analysis. John Wiley, New York, 1982. 745 s.
13. Press, S. J.: Bayesian statistics. Principles, models and applications. John Wiley & Sons, New York, 1989. 256 s.
14. Holický, M., Marková, J.: Základy teorie spolehlivosti a hodnocení rizik. ČVUT, Praha, 2005. 115 s.
15. Ang, A. H. S., Tang, W. H.: Probability Concepts in Engineering Planning and Design. Vol. 1, Basic Principles. John Wiley, New York, 1975. 574 s.



## 9. Rozměrová analýza a teorie podobnosti

Rozměrová analýza a teorie podobnosti jsou silné nástroje, které významně zvyšují účinnost experimentálního zkoumání. Teorie podobnosti je velmi užitečná pro vytváření modelů a práci s nimi. To je obzvláště důležité, mají-li být zkoumány velké objekty, anebo jestliže experimentování s reálnými objekty nebo systémy je nemožné nebo velice obtížné. Rozměrová analýza a použití bezrozměrových proměnných zjednodušují experimenty, mohou ušetřit mnoho experimentální práce, a dělají výsledky obecnějšími. V této kapitole jsou definovány různé druhy podobnosti, ukázány příklady bezrozměrových veličin a je vysvětleno, jak mohou být systematicky vytvářeny. To je ilustrováno na praktických příkladech. Jsou též ukázány meze použitelnosti pravidel podobnosti.

### Rozměrová analýza

Každá fyzikální veličina je popsána číselnou hodnotou doprovázenou jednotkou. Číselná hodnota říká, kolikrát je vyšetřovaná veličina větší než její jednotka. Příklad délky je 5,3 m, příklad síly je 25 N, příklad času je 15,6 ms. Kromě základních jednotek (metr, kilogram, sekunda...), definovaných v systému SI (Système International), se užívají také různé odvozené jednotky a také předpony ( $\mu$ , m, k, M...), které označují řád.

Každá rovnice, která popisuje fyzikální jev, musí být rozměrově homogenní: její levá strana musí mít stejný rozměr jako pravá strana. Takováto kontrola homogenity rozměrů by se měla dělat vždy před použitím nově odvozeného vztahu. Vyšetřujme například vzorec pro výpočet průhybu  $y$  elastického nosníku zatíženého silou  $F$ . Z nauky o pružnosti je známo, že  $y$  bude přímo úměrné  $F$  a nepřímo úměrné ohybové tuhosti nosníku, definované jako  $E \times J$ , kde  $E$  je Youngův modul materiálu a  $J$  je moment setrvačnosti průřezu. Průhyb bude také úměrný určité mocnině  $S$  délky nosníku  $L$ . Představme si nyní, že neznáme exponent  $S$ . V takovém případě můžeme napsat základní tvar vzorce:

$$y = C \times F \times L^S / (E \times J) \quad (9.1)$$

$C$  je bezrozměrová konstanta. Nahradíme-li jednotlivé veličiny ve vztahu (9.1) jejich jednotkami, dostaneme

$$m = 1 \times N \times m^S / (Nm^{-2} \times m^4)$$

Rozměr pravé strany musí být stejný jako rozměr levé strany, tj. metr, obecně  $m^1$ . Součin všech členů, obsahujících  $m$ , je  $m^S \times m^2 \times m^{-4} = m^{S+2-4} = m^{S-2}$ . Porovnání exponentů na levé a pravé straně rovnice dá  $1 = S - 2$ . Z toho vyplývá  $S = 3$ , takže  $y = C \times F \times L^3 / (EJ)$ , což je vztah velmi dobře známý z mechaniky. Podmínka homogenity rozměrů (rozměr obou stran rovnice musí být stejný) tedy pomohla najít správný tvar vzorce.

Je-li jedna strana rovnice tvořena součtem několika členů, pak všechny musí mít stejný rozměr. Například svislá dráha  $y$  tělesa padajícího v gravitačním poli je obecně

$$y = y_0 + v_0 t + \frac{1}{2} g t^2 \quad (9.2)$$

$t$  je čas,  $y_0$  a  $v_0$  jsou poloha a rychlost tělesa v čase  $t = 0$ , a  $g$  je tíhové zrychlení. Podmínka rozměrové homogenity říká, že jednotlivé veličiny nemohou existovat ve fyzikální rovnici nezávisle, ale pouze ve skupinách se stejným rozměrem. Vydělíme-li rovnici (9.2) jedním z členů, např.  $y_0$ , změní se na bezrozměrový tvar

$$y/y_0 = 1 + v_0 t/y_0 + \frac{1}{2} g t^2/y_0 \quad (9.3)$$

s normalizovanými veličinami  $y/y_0$ ,  $v_0 t/y_0$  and  $g t^2/y_0$ .

Téměř každou fyzikální rovnici je možno přetransformovat na bezrozměrový tvar. Použití normalizovaných veličin má řadu výhod. Fyzikální rovnice, vyjádřené prostřednictvím bezrozměrových veličin jsou obecnější, než když jsou vyjádřeny pomocí rozměrových veličin. Relativní posunutí (dráha),  $y/y_0$ , nezávisí jednotlivě na  $v_0$ ,  $t$  a  $y_0$ , ale na jejich určitých kombinacích, uvedených v rovnici (9.3). Bezrozměrové veličiny tedy umožňují kombinovat výsledky experimentů vykonaných se vzorky o různé počáteční rychlosti a poloze, kdy jedinou podmínkou je jejich správná kombinace. (V hořejším příkladu nosníku hraje roli kombinace jeho rozměrů a materiálu.) Pro formulování určitého zákona je tedy možno využít více dat a širší rozmezí parametrů (viz Obr. 9.2 dále). Výsledky vyjádřené v bezrozměrové formě jsou také univerzálnější, použitelné pro celou třídu podobných objektů, s podobnou geometrií nebo fyzikálními vlastnostmi.

Navíc – a to je velmi důležité – použití bezrozměrových veličin může ušetřit experimentální práci, protože

**Vztah  $N$  veličin, jejichž rozměry lze vyjádřit pomocí  $D$  základních rozměrů, je (obvykle) možno nahradit vztahem**

$$P = N - D \quad (9.4)$$

### **bezrozměrových parametrů $\Pi$ .**

Podle tohoto, tzv. **Buckinghamova pravidla** [1 – 4], stanovení menšího počtu regresních konstant vyžaduje méně experimentů. Redukce experimentální práce je významná zejména pokud vyšetřovaný vztah obsahuje mnoho veličin a počet proměnných  $N$  je blízký počtu základních rozměrů,  $D$ . To je možno ilustrovat na předchozím příkladu padajícího tělesa. Rovnice (9.2) představuje vztah 5 veličin:  $y$ ,  $y_0$ ,  $v_0$ ,  $g$ , a  $t$ ; tj.  $N = 5$ . Tyto veličiny lze vyjádřit pomocí dvou základních rozměrů: metr a sekunda; tedy  $D = 2$ . Podle rovnice (4), počet bezrozměrových parametrů by měl být  $P = N - D = 5 - 2 = 3$ . A skutečně, rovnice (9.3) je vztahem tří bezrozměrových parametrů:  $y/y_0$ ,  $v_0t/y_0$  and  $gt^2/y_0$ . Stanovení potřebného počtu experimentů bude diskutováno v kapitole 11. Nicméně určitou představu získáme z jednoduchého příkladu. Jestliže je třeba vyšetřit vliv šesti faktorů, a každý bude na dvou úrovních (nízká a vysoká), počet nezbytných pokusů bude  $2^6 = 64$ . Pokud by počet bezrozměrových parametrů byl pouze 4, počet nezbytných pokusů klesne na  $2^4 = 16$ , tj. na jednu čtvrtinu!

### **Podobnost**

Použití bezrozměrových veličin je také velice důležité při studiu chování reálných objektů prostřednictvím modelů. Například stavba nové velké lodi, mostu nebo chemického reaktoru je doprovázena mnoha nejistotami, a možné ztráty v důsledku chybného designu by byly velmi vysoké. Proto se nejprve zpravidla postaví a testuje menší model. Pokud však má adekvátně odrážet chování skutečného objektu, musí mezi nimi existovat podobnost. Existují různé druhy podobnosti, například:

Geometrická podobnost, tj. shodnost tvaru včetně odpovídajících úhlů a úměrnost odpovídajících rozměrů (tzv. měřítko). Platí následující relace:

$$\text{Rozměr modelu} = \text{Měřítko} \times \text{Rozměr skutečného objektu}$$

Například model budovy v měřítku 1:20 má dvacetkrát menší rozměry než skutečný dům.

Statická podobnost: deformace modelu při konstantním zatížení jsou ve stejném poměru jako odpovídající deformace objektu (díla).

Kinematická podobnost: existuje poměr časové úměrnosti mezi odpovídajícími jevy pro model a pro objekt.

Dynamická podobnost: síly, působící v odpovídajících časech a místech na modelu a objektu jsou ve stejném poměru.

Teplotní podobnost: teplotní profily v modelu a prototypu jsou v odpovídajících časech podobné.

Chemická podobnost: rychlost chemických reakcí v modelu je úměrná rychlosti stejných reakcí v odpovídajících časech a místech objektu.

**Teorie podobnosti** pracuje s takzvanými **podobnostními čísly**. Čtenáři, kteří absolvovali středoškolskou výuku fyziky, znají např. Reynoldsovo číslo (Re), které pomáhá posoudit, jestli proudění kapaliny je laminární nebo turbulentní. Více příkladů bude uvedeno na konci kapitoly. Podobnostní čísla jsou bezrozměrová; vlastně každá bezrozměrová veličina může sloužit jako podobnostní číslo.

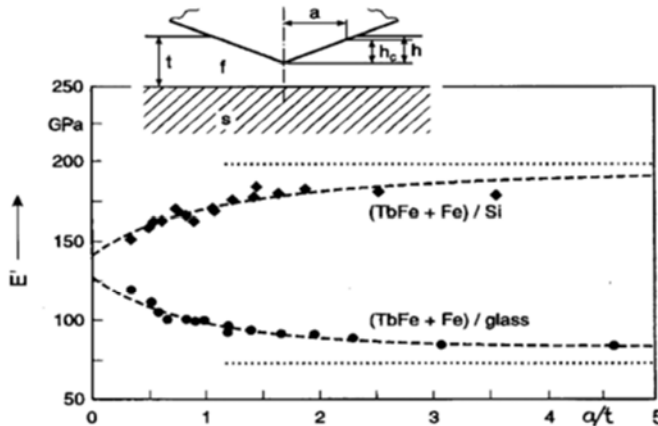
**Bezrozměrové proměnné** lze vytvářet různými způsoby. Nejjednodušší případ je poměr určité veličiny a její charakteristické hodnoty, např.  $x/x_0$  nebo  $\Delta x/x_0$  pro vzdálenost nebo posunutí. V mechanice je dobře známé poměrné prodloužení ( $\varepsilon = \Delta L/L$ ), Poissonovo číslo  $\mu$  (poměr relativního zkrácení v příčném směru k relativnímu prodloužení ve směru zatížení), nebo koeficient tření  $f$ , definovaný jako poměr síly, potřebné k posouvání tělesa po jiném tělese, a normálové síly, která obě tělesa přitlačuje k sobě. Jiným příkladem je poměrná poloha bodu v tělese, například

$$\xi = (x - x_{\min}) / (x_{\max} - x_{\min}) \quad (9.5)$$

$x$ ,  $x_{\max}$  a  $x_{\min}$  představují souřadnice. Podobně je možné vyjádřit čas. Bezrozměrová teplota,  $\theta = (T - T_{\infty}) / (T_0 - T_{\infty})$ , se užívá pro univerzální popis přenosu tepla ( $T_0$  je počáteční teplota a  $T_{\infty}$  je konečná teplota). V tomto případě také poloha vyšetřovaného místa může být v bezrozměrové podobě. Tento přístup bude ilustrován na příkladu z podobné oblasti.

Příklad. Stanovení modulu pružnosti tenkých povlaků instrumentovanou indentací.

Modul pružnosti  $E$  různých materiálů lze určovat (kromě jiných metod) také instrumentovanou indentací. Indentor je vtačován do vzorku a jeho zaboření je během zatěžování a odlehčování měřeno jako funkce zatížení. Modul pružnosti se pak vypočítá speciálním postupem z naměřených hodnot [5, 6]. Stanovení modulu pružnosti povlaku, vytvořeného na substrátu, je složitější. Odezva vzorku s povlakem na vnikání indentoru, a tedy i hodnota  $E$ , získaná v testu, závisí na modulu povlaku ( $E_c$ ) a modulu substrátu ( $E_s$ ), na tloušťce povlaku ( $t$ ) a na hloubce  $h$  zaboření indentoru. Naměřená hodnota  $E$  se postupně mění od hodnoty modulu povlaku (pro „nulové“ vniknutí indentoru) k hodnotě modulu substrátu pro značně velké hloubky zaboření (Obr. 9.1). POZNÁMKA. Křemík (Si) má vyšší modul pružnosti než povlak, zatímco sklo má nižší modul.

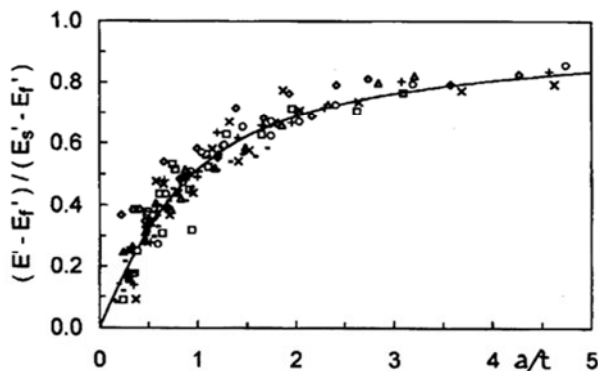


**Obr. 9.1.** Měření modulu pružnosti vzorků opatřených povlakem [7]. Povlak z TbTe/Fe byl nanesen na (a) křemíkový (Si) a (b) skleněný substrát (glass). Tečkované vodorovné přímky odpovídají modulům pružnosti substrátů.  $E'$  - redukovaný modul  $[= E/(1 - \mu^2)]$ ,  $a$  - poloměr kontaktu,  $t$  - tloušťka povlaku.

Pravou hodnotu modulu povlaku  $E_c$  je možno získat tak, že se hodnoty  $E$ , naměřené pro různé hloubky, proloží určitou funkcí  $\Phi$  a extrapolují k nulové hloubce zaboření. Tato bezrozměrová funkce  $\Phi$  může být podle [8] definována jako

$$\Phi(h/h_c) = [E(h/t) - E_s] / [E_c - E_s] \quad (9.6)$$

Význam bezrozměrového vyjádření všech veličin je demonstrován na Obr. 9.2. Ten ukazuje hodnoty naměřené na 25 vzorcích s povlaky z různých materiálů, o různých tloušťkách, na různých substrátech a s různými hloubkami zaboření indentoru [8]. Je vidět, že všechny hodnoty, zakreslené v normalizovaných souřadnicích, leží přibližně na křivce, získané teoretickým řešením kontaktu s polonekonečným tělesem opatřeným povlakem [9].



**Obr. 9.2.** Měření modulu pružnosti vzorků s povlakem; podle [8]. Hodnoty, naměřené pro 25 různých povlaků, substrátů a hloubek, jsou vyneseny v normalizovaných souřadnicích.  $a$  – poloměr kontaktu,  $t$  – tloušťka povlaku,  $s$  – substrát,  $f$  – povlak (film),  $E'$  – redukovaný modul =  $E/(1 - \mu^2)$ .

Bezrozměrové musí být také argumenty v matematických funkcích typu sin, cos, log nebo exp. V opačném případě by jakákoliv změna jednotek (např. metry místo mm) by změnila numerickou hodnotu výsledku. Bezrozměrové jsou také argumenty ve spojitéch rozděleních pravděpodobnosti. Na příklad normální rozdělení užívá argument  $\{1/2[(x - \mu)/\sigma]^2\}$ , kde  $\mu$  a  $\sigma$  jsou střední hodnota a směrodatná odchylka. Výraz v hranatých závorkách ale není nic jiného, než standardizovaná proměnná, která vyjadřuje vzdálenost  $x$  od střední hodnoty  $\mu$  jako násobek směrodatné odchylky  $\sigma$ . Podobně je bezrozměrový i argument ve Weibullově nebo exponenciálním rozdělení. (POZNÁMKA: Bezrozměrové veličiny se užívají častěji, než si myslíme.)

Tabulka na konci této kapitoly ukazuje příklady bezrozměrových veličin z různých oblastí.

Následující odstavce, vycházející z publikací [1, 2], vysvětlují formální postup pro vytváření bezrozměrových parametrů a poskytují i další rady.

### Vytváření bezrozměrových parametrů

Jednotlivé kroky ukážeme na předchozím případě pohybu tělesa v gravitačním poli.

1. Sepíšeme všechny veličiny a jejich rozměry:

$$y(\text{m}), y_0(\text{m}), v_0 (\text{m} \times \text{s}^{-1}), t(\text{s}), g(\text{m} \times \text{s}^{-2})$$

Veličin je v tomto případě  $N = 5$  a základní rozměry jsou dva: metr a sekunda ( $D = 2$ ).

2. Bezrozměrové parametry ( $\Pi$ ) budeme předpokládat v obecném tvaru:

$$\Pi = y^{x_1} \times y_0^{x_2} \times v_0^{x_3} \times t^{x_4} \times g^{x_5}$$

3. Levou a pravou stranu vyjádříme prostřednictvím rozměrů zúčastněných veličin:

$$1 = [\text{m}]^0 \times [\text{s}]^0 = [\text{m}]^{x_1} \times [\text{m}]^{x_2} \times [\text{m} \times \text{s}^{-1}]^{x_3} \times [\text{s}]^{x_4} \times [\text{m} \times \text{s}^{-2}]^{x_5}$$

4. Rovnost obou stran rovnice vyžaduje také rovnost exponentů u stejných základů.

Zde použijeme zápis obvyklý u soustav rovnic:

$$\text{metr:} \quad x_1 + x_2 + x_3 \quad x_5 = 0 \quad (\text{a})$$

$$\text{sekunda:} \quad -x_3 + x_4 - 2x_5 = 0 \quad (\text{b})$$

-----

Toto jsou dvě lineární rovnice s 5 neznámými. Kdybychom měli k dispozici 5 rovnic místo dvou, dostali bychom neznámé hodnoty  $x_1, \dots, x_5$  přímo řešením soustavy pěti rovnic. V našem případě ale máme o tři neznámé více než rovnic pro jejich stanovení:  $N - D = 5 - 2 = 3$ . Můžeme proto navrhnout tři exponenty a vypočítat zbývající dva. Tuto volbu můžeme udělat třikrát. V tomto případě navrhneme hodnoty  $x_1, x_3$  and  $x_5$ , přičemž budeme chtít, aby byly co nejjednodušší. Proto jedna ze zvolených konstant bude vždy rovna 1, a zbývající budou rovny 0. Je rozumné, jestliže jeden z těchto exponentů přísluší veličině, kterou zkoumáme.

Volba 1.  $x_1 = 1; x_3 = 0; x_5 = 0$ .

Dosazení těchto konstant do rovnice (b) dá  $x_4 = 0$ . Vložení  $x_1, x_3$  a  $x_5$  do rovnice (a) dá  $x_2 = -1$ . První bezrozměrový parametr tedy bude  $\Pi_1 = y^1 \times y_0^{-1} \times v_0^0 \times t^0 \times g^0 = y/y_0$ .

Volba 2.  $x_1 = 0; x_3 = 1; x_5 = 0.$

Dosadíme-li tyto konstanty do rovnic (a, b) a vyřešíme získanou soustavu podobným způsobem, dostaneme  $x_4 = 1$  and  $x_2 = -1$ . Druhý parametr je  $\Pi_2 = y^0 \times y_0^{-1} \times v_0^1 \times t^1 \times g^0 = v_0 t / y_0.$

Volba 3.  $x_1 = 0; x_3 = 0; x_5 = 1.$

Jejich dosazení do soustavy rovnic (a, b) a vyřešení dá  $x_4 = 2$  and  $x_2 = -1$ . Třetí parametr je  $\Pi_3 = y^0 \times y_0^{-1} \times v_0^0 \times t^2 \times g^1 = g t^2 / y_0.$  Čtenáři doporučujeme, aby si řešení zopakoval.

Pohyb padajícího tělesa je tedy možno vyjádřit jako

$$\Phi(\Pi_1, \Pi_2, \Pi_3) = \Phi(y/y_0, v_0 t/y_0, g t^2/y_0) = 0, \text{ nebo } y/y_0 = f(v_0 t/y_0, g t^2/y_0).$$

Naznačené parametry by byly vhodné, pokud bychom mohli snadno měnit  $t$  a měřit  $y$ . Je ale možné zvolit i jiné parametry. Například kdybychom mohli měnit  $y$  a měřit trvání pádu  $t$ , definovali bychom nejprve  $x_4, x_1$  a  $x_5$  (podobně jako výše), našli  $x_2, x_3$  a dostali  $\Pi_1 = v_0 t/y_0, \Pi_2 = y/y_0, \Pi_3 = y_0 g/v_0^2.$

### Další doporučení

1) Nalezený tvar bezrozměrového parametru někdy neodpovídá našim záměrům nebo experimentálním možnostem. Všeobecně lze vytvářet nové parametry nebo podobnostní čísla jako součin nebo poměr původních bezrozměrových parametrů, anebo použít jejich převrácenou hodnotu nebo mocninu. Protože původní parametry byly bezrozměrové, jsou i takto vzniklé nové parametry bezrozměrové.

2) Vyskytuje-li se v úloze několik veličin se stejným rozměrem, můžeme vytvářet bezrozměrové parametry přímo jako jejich poměry. Tak lze snížit počet argumentů, které se musí získat řešením systému rovnic, jak bylo naznačeno dříve. Ilustrovat to budeme na příkladu průhybu nosníku  $y$  s obdélníkovým průřezem ( $b \times h$ ) a délkou  $L$ , zatíženého osamělou silou  $P$ . Modul pružnosti materiálu je  $E$ . Proměnné a jejich rozměry jsou:  $y(m), w(m), h(m), L(m), P(N), E(Nm^{-2})$ ; to je 6 proměnných se dvěma rozměry. Počet bezrozměrových parametrů, potřebných pro popis problému, je  $P = N - D = 6 - 2 = 4$ . Okamžitě můžeme vytvořit tři parametry  $\Pi_1 = y/h, \Pi_2 = b/h$  a  $\Pi_3 = L/h$ . Zbývají dvě veličiny ( $P$  a  $E$ ), které musí být obsaženy ve čtvrtém parametru. S přihlédnutím k jejich rozměrům a k podmínce bezrozměrnosti musí být do  $\Pi_4$ , zahrnuta také jedna geometrická veličina, například  $h$  nebo jeho



mocnina. Tak dostaneme tento parametr jako  $\Pi_4 = P/(Eh^2)$ . Vyšetřovaný vztah lze tedy zapsat v následujícím bezrozměrovém tvaru:

$$y/h = f[P/(Eh^2), L/h, b/h] \quad (9.7)$$

Zdůrazněme, že pro zkoumání relativního průhybu  $y/h$  jsou důležité nikoliv jednotlivé veličiny  $L$  nebo  $P$ , atd., ale jejich poměry.

3) U některých úloh se vždy vyskytují bezrozměrové veličiny. Příklady jsou součinitel tření, Poissonovo číslo  $\mu$  pro příčnou kontrakci, nebo úhel  $\varphi$  (rad). Tyto veličiny se automaticky stávají argumenty v bezrozměrových vztazích.

4) Vytváříme-li bezrozměrové parametry, snažíme se využít existující znalosti o vyšetřovaném problému nebo podobných úlohách. Například průhyb pružného nosníku je přímo úměrný zatížení a nepřímo modulu pružnosti. Někdy známe analytické řešení úlohy pro velmi malé nebo velmi vysoké hodnoty určité proměnné. To může pomoci při hledání tvaru argumentů. Někdy je známo, že některé veličiny musí vystupovat v určité kombinaci. Tuto kombinaci pak můžeme uvažovat jako novou proměnnou, což může snížit celkový počet proměnných. Vyšetřujme na příklad sílu působící v kontaktu dvou těles. Jestliže se má vyšetřovat vliv tření, je možno sílu  $F$  (N) a kontaktní plochu  $A$  ( $\text{m}^2$ ) nahradit kontaktním tlakem  $p = F/A$  ( $\text{N}/\text{m}^2$ ).

5) Připravujeme-li určitý experiment, je nutné uvažovat všechny veličiny, které by mohly hrát roli. Jinak bychom mohli dostat nesprávné a zavadějící výsledky. Je méně nebezpečné zahrnout veličinu, jejíž důležitost je nejistá, a později se může ukázat, že ji lze pominout, než vynechat veličinu, která se později ukáže jako důležitá. Použití rozměrové analýzy někdy odhalí závažné nedostatky. Například objeví-li se určitý rozměr jenom u jedné veličiny, potom tato veličina z bezrozměrové rovnice vypadne a nebude zahrnuta do žádného bezrozměrového parametru. Jestliže ale tato veličina je zřejmě nezbytná pro popis vyšetřovaného jevu (např. jako závislá proměnná), je nutné přidat ještě další veličinu, která bude mít stejný rozměr. Budeme to ilustrovat na příkladu vyšetřování intenzity opotřebení obráběcího nástroje. Veličiny, které hrají roli, jsou: intenzita opotřebení  $w$  (m/s), rychlost vzájemného pohybu  $v$  (m/s) a tlak v kontaktní ploše  $p$  ( $\text{N}/\text{m}^2$ ). Bezrozměrový parametr lze hledat v obecném tvaru  $\Pi = w^{x_1} v^{x_2} p^{x_3}$ . Tento výraz můžeme přepsat prostřednictvím rozměrů jednotlivých veličin m, s, N:

$$[\text{m}]^0 [\text{s}]^0 [\text{N}]^0 = [\text{m} \times \text{s}^{-1}]^{x_1} \times [\text{m} \times \text{s}^{-1}]^{x_2} \times [\text{N} \times \text{m}^{-2}]^{x_3} \quad (9.8)$$

Levá strana odpovídá bezrozměrovému vyjádření. Z podmínky rovnosti exponentů u stejného základu vyplývá  $\text{N}^0 = \text{N}^{x_3}$ , takže  $x_3 = 0$ . Z experimentů je ale velmi dobře známo, že rychlost opotřebování závisí na kontaktním tlaku, takže  $x_3$  nemůže být rovno 0. Je tedy nezbytné přidat ještě nějakou další veličinu, která bude mít také rozměr  $\text{Nm}^{-2}$ . Může to být například tvrdost  $H$  ( $\text{Nm}^{-2}$ ), která charakterizuje odpor materiálu. Obecný tvar bezrozměrového parametru pak bude

$$\Pi = w^{x_1} v^{x_2} p^{x_3} H^{x_4} \quad (9.9)$$

Z tohoto výrazu můžeme snadno formulovat vhodný vztah bezrozměrových parametrů jako  $w/v = f(p/H)$ , a provést řadu experimentů pro nalezení vhodného tvaru funkce  $f$ . Ze cvičných důvodů zde ale stanovíme bezrozměrové argumenty také formálním postupem popsaným dříve.

Vyjádříme levou a pravou stranu (9.9) pomocí rozměrů zúčastněných veličin:

$$[\text{m}]^0 [\text{s}]^0 [\text{N}]^0 = [\text{m} \times \text{s}^{-1}]^{x_1} \times [\text{m} \times \text{s}^{-1}]^{x_2} \times [\text{N} \times \text{m}^{-2}]^{x_3} \times [\text{N} \times \text{m}^{-2}]^{x_4}$$

Podmínka rovnosti exponentů na levé a pravé straně vede k následující soustavě rovnic:

$$\begin{array}{lcl} m: & x_1 + x_2 - 2x_3 - 2x_4 & = 0 \\ s: & -x_1 - x_2 & = 0 \\ N: & x_3 + x_4 & = 0 \end{array}$$

Tyto tři rovnice nejsou dostačující pro určení čtyř exponentů. Obecně bychom mohli zvolit jeden exponent a potom dostat zbývající tři řešením soustavy 3 rovnic. Naneštěstí to v našem případě není možné, protože mezi rovnicemi existuje lineární závislost. (První rovnici můžeme získat ze součtu druhé a dvojnásobku třetí rovnice.) Musí tedy být dva bezrozměrové parametry, takže je nutno zvolit dva exponenty, například  $x_1$  and  $x_3$ .

Volba 1.  $x_1 = 1$ ;  $x_3 = 0$ .

Tato volba dává  $x_2 = -1$  a  $x_4 = 0$ , takže  $\Pi_1 = w/v$ .

Volba 2.  $x_1 = 0$ ;  $x_3 = 1$ .

Tato volba dává  $x_2 = 0$  a  $x_4 = -1$ , takže  $\Pi_2 = p/H$ .

Můžeme tedy vyšetřovat vztah  $I_1 = f(I_2)$ , tedy  $w/v = f(p/H)$ , podobně jako dříve.

### Omezení principu podobnosti

Princip podobnosti platí jenom za určitých podmínek, a mimo ně ztrácí svoji platnost [10]. Příkladem je přechod od pružných deformací k plastickým u součástí z kovových materiálů. Jsou-li napětí nižší než mez kluzu, existuje přímá úměrnost mezi napětími a přetvořeními, a je možno užívat princip superpozice. Zákony v pružnoplastické oblasti jsou ale nelineární, takže situace musí být řešena pro různá zatížení individuálně. Jiným případem je pružný kontakt dvou těles. Jsou-li napětí nízká a zatížená oblast velká, je možno použít vztahy pro homogenní izotropní elastická tělesa. Jestliže je ale zatížená oblast mnohem menší, srovnatelná s rozměry krystalických zrn nebo jiných složek mikrostruktury, je nutno zohlednit heterogenitu, popřípadě i anizotropii materiálu. Příkladem je beton a další kompozitní materiály při zkouškách nanoindentací, ale také krystal perlitu, jestliže je velikost vtisku srovnatelná s tloušťkou feritových a cementitových lamel. O nehomogenitě v takovýchto případech svědčí vysoký rozptyl jednotlivých naměřených hodnot. Naštěstí statistická analýza velkého množství (stovek) vpichů umožní odhalit vlastnosti jednotlivých fází [11]. V takových případech se dalším parametrem pro charakterizování materiálu stává rozložení mikrostrukturních jednotek. Tvrdost kovů, rostoucí s klesající velikostí vpichu, známá jako „indentation size effect“, je částečně způsobena klesajícím množstvím dislokací, které usnadňují plastické tečení. Pro velmi malé hloubky indentace se také stává důležitou drsnost povrchu. Měření při velmi nízkých zatíženích mohou zase být výrazně ovlivněna adhezními silami, zejména při testování velmi poddajných materiálů, jako jsou např. gely.

Existují i další případy, kdy princip podobnosti neplatí. Známa je závislost pevnosti součástí z křehkých materiálů na velikosti zatížené plochy nebo objemu. Křehký lom obvykle začne na přítomném slabém místě, jako je prasklé krystalické zrno v keramice nebo nepatrné škrábnutí na povrchu skla. Menší velikost zatížené oblasti znamená menší pravděpodobnost výskytu větší vady. Menší defekt může sloužit jako východisko poruchy pouze při vyšším napětí. Velmi malé předměty jsou proto pevnější. Podobně závisí na velikosti i mez únavy kovových součástí.

Je nutno mít také na paměti, že někdy se vyšetřovaná veličina mění se změnami určitého parametru poměrně pomalu, ale při dosažení jeho jisté úrovně se může

změnit velmi rychle. Vztah, popisující jisté chování nebo proces, často platí jenom v určitém rozmezí parametrů. Pokud je příslušný proces popsán prostřednictvím bezrozměrových veličin, jsou podmínky pro přechod z jednoho režimu do druhého charakterizovány tzv. **kritickou hodnotou** těchto veličin. Znáмым případem je přechod od laminárního proudění k turbulentnímu při dosažení kritické hodnoty Reynoldsova čísla. Je tedy nutno vždy nejprve uvažovat všechny možné vlivy, a počet faktorů snížit až po důkladné analýze.

## **Příklady bezrozměrových veličin**

### Materiálové vlastnosti

$E_1/E_2, H_1/H_2$  poměr modulů pružnosti nebo tvrdostí; indexy označují součásti nebo fáze,

$E(x)/E_0, H(x)/H_0$  poměry jako výše; index 0 označuje charakteristickou hodnotu,

$H/Y, E/Y, E/H$  poměr tvrdosti a meze kluzu nebo modulu pružnosti,

$\sigma/Y, \sigma/\sigma_u, Y/\sigma_u$  poměr napětí k mezi kluzu  $Y$ , pevnosti  $\sigma_u$ , povrchovému napětí

### Geometrie

$x/d$   $x$  – vzdálenost, hloubka zaboření indentoru,  $d$  – charakteristický rozměr vzorku nebo materiálu (kontaktní poloměr, průměr krystalického zrna, póru nebo vlákna, délka vzorku, šířka, výška nebo průměr, tloušťka povlaku, rozměr plastické zóny, vzdálenost dislokací nebo jiných materiálových defektů, vzdálenost od okraje vzorku ...),

$\Delta l/L$  poměrné posunutí, poměrné prodloužení ( $\epsilon$ ),  $\Delta l$  – prodloužení,  $L$  – základní délka,

$h/R, h/t_c$  poměr zaboření indentoru  $h$  k poloměru  $R$  jeho špičky nebo k tloušťce povlaku  $t_c$ ,

$h_c/h$  poměr hloubky kontaktu  $h_c$  k zaboření indentoru  $h$ .

### Síly a napětí

$F/F_0$  poměr zatížení  $F$  a síly adheze ( $F_0 = F_{ad}$ ) nebo jiné charakteristické síly,

$\sigma/\sigma_m$  poměr napětí  $\sigma$  k jmenovitému či střednímu napětí (tlaku)  $\sigma_m$ .

### Čas

$t/t_0$   $t_0$  – charakteristický čas (nárůstu zatížení, relaxační čas...).

Čtenář sám může najít další příklady.

## Podobnostní čísla užívaná často ve fyzice a technice

Podobnostní čísla byla pojmenována po významných vědcích a jsou označena prvními dvěma písmeny příslušného jména. Zde je uveden výběr.

Archimedes  $Ar = gd^3\rho'(\rho - \rho')/\eta^2$  ;  $\rho, \rho'$  – hustota kapaliny a tělesa,  $g$  – gravitační zrychlení,  $d$  – charakteristický rozměr,  $\eta$  – dynamická viskozita

Biot  $Bi = \alpha d/\lambda$  ;  $\alpha$  – koeficient přestupu tepla,  $d$  – charakteristický rozměr,  $\lambda$  – tepelná vodivost tělesa

Deborah  $De = t_r/t$  ;  $t_r$  – relaxační doba,  $t$  – čas

Euler  $Eu = \Delta p / \rho u^2$  ;  $\Delta p$  – rozdíl tlaku,  $\rho$  – hustota,  $u$  – charakteristická rychlost

Fourier  $Fo = a\tau/d^2$  ;  $a$  – teplotní vodivost,  $\tau$  – čas,  $d$  – charakteristický rozměr

Froude  $Fr = u^2/gd$  ;  $u$  – charakteristická rychlost,  $g$  – gravitační zrychlení,  $d$  – charakteristický rozměr

Galilei  $Ga = gd^3/\nu^2$  ;  $g$  – gravitační zrychlení,  $d$  – charakteristický rozměr,  $\nu$  – kinematická viskozita

Grashoff  $Gr = \beta\Delta Tgl^3/\nu$  ;  $\beta$  – teplotní roztažnost kapaliny,  $\Delta T$  – teplotní rozdíl,  $g$  – gravitační zrychlení,  $d$  – charakteristický rozměr,  $\nu = \eta/\rho$  = kinematická viskozita

Nusselt  $Nu = \alpha d/\lambda$  ;  $\alpha$  – součinitel přestupu tepla,  $d$  – charakteristický rozměr,  $\lambda$  – tepelná vodivost okolního prostředí

Péclet	$Pe = ud/a$ ; $u$ – rychlost, $d$ – charakteristický rozměr, $a$ – teplotní vodivost
Prandtl	$Pr = \nu/a$ ; $\nu$ – kinematická viskozita, $a$ – teplotní vodivost
Reynolds	$Re = ud\rho/\eta = ud/\nu$ ; $u$ – charakteristická rychlost, $d$ – charakteristický rozměr, $\rho$ – hustota kapaliny, $\eta$ – dynamická viskozita, $\nu = \eta/\rho$ = kinematická viskozita
Stanton	$St = \alpha/(\lambda u) = Nu/(Re.Pr)$ ; $\alpha$ – součinitel přestupu tepla, $\lambda$ – tepelná vodivost tekutiny, $u$ – rychlost tekutiny
Stokes	$Stk = ut/d$ ; $u$ – rychlost, $t$ – relaxační doba, $d$ – charakteristický rozměr
Weber	$We = \rho u^2 d/\sigma$ ; $\rho$ – hustota, $u$ – rychlost, $d$ – charakteristický rozměr, $\sigma$ – povrchové napětí

### Literatura ke kapitole 9

1. Kožešník, J.: Teorie podobnosti a modelování. Academia, Praha, 1983. 216 s.
2. Zlokarnik, M.: Scale-up in Chemical Engineering. 2nd Edition, Wiley, 2006, 296 s.
3. Szirtes, T.: Applied Dimensional Analysis and Modeling, McGraw-Hill, New York, 1997, 2nd Ed. 2007. 856 s.
4. Y.T. Cheng, Y.T., Cheng, C.M.: Scaling, dimensional analysis, and indentation measurements, Mat. Sci. Eng. R44 (2004) 91 – 149.
5. Oliver, W.C., Pharr, G.M.: An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. J. Mater. Res. 7 (1992), No. 6, 1564 – 1583.
6. Fischer-Cripps, A.C.: Nanoindentation. Springer, 2004, 282 s.
7. Menčík, J., Munz, D., Quandt, E., Ludwig, A.: Determination of elastic modulus of thin layers. Z. Metallkd. 90 (1999), No. 10, 766 – 773.
8. Menčík et al: Determination of elastic modulus of thin layers using nanoindentation. J. Mater. Res., Vol. 12 (1997), No. 9, 2475 – 2484.
9. Gao, H., Chiu, C.H., Lee, J.: Elastic contact versus indentation modelling of multilayered materials. Int. J. Solids Structures, 29 (1992), 2471 – 2492.

10. Menčík, J.: Limitations of Similarity Principle in Indentation Testing of Small Samples. *Key Engineering Materials* (Trans Tech Publications), Vol. 586 (Local Mechanical Properties IX), 2014. 47 – 50.
11. Němeček, J.: Nanoindentation based analysis of heterogeneous structural materials. Chapter 4 (p. 89 – 108) in *Nanoindentation in materials science* (J. Němeček, editor). Intech, Rijeka, 2012. Open access, available at <http://www.intechopen.com/books>

***S pomocí fyzikálních teorií se snažíme nalézt cestu bludištěm  
pozorovaných fakt a pochopit svět.***

Albert Einstein

***Fyzikální zákony by měly mít matematickou krásu.***

Paul Dirac



## 10. Analýza rozptylu

Častým úkolem ve výzkumu je posoudit vliv různých činitelů (faktorů), najít, který z nich má největší vliv, který má zanedbatelný vliv, atd. Má-li se porovnat vliv pouze dvou faktorů, postačí *t*-test pro rozdíl mezi středními hodnotami (viz Kapitulu 8). Někdy je ale nutno posoudit účinek tří nebo více činitelů. Například při výrobě určité sloučeniny lze použít různé suroviny, různé technologické postupy a různé aparáty, a my chceme vědět, které z těchto faktorů mají silnější vliv na výrobek. Bylo by možné testovat oddělené rozdíly mezi jednotlivými dvojicemi faktorů. Účinnější je ale takzvaná **analýza rozptylu** (analysis of variance, ANOVA), kterou dále stručně vysvětlíme.

Je-li možné třídít skupinu testů podle jednoho nebo více hledisek, potom je možné třídít také celkovou proměnlivost podle těchto hledisek. Základní myšlenka analýzy rozptylu je rozložit celkový rozptyl  $\sigma_{\text{tot}}^2$  vyšetřované veličiny na části způsobené jednotlivými faktory ( $\sigma_{ij}^2$ ) a zbytkovou část  $\sigma_{\text{res}}^2$ , způsobenou neidentifikovanými (náhodnými) vlivy:

$$\sigma_{\text{tot}}^2 = \sigma_{f_1}^2 + \sigma_{f_2}^2 + \dots + \sigma_{\text{res}}^2 \quad (10.1)$$

Porovnání rozptylů, odpovídajících jednotlivým faktorům, se zbytkovým rozptylem způsobeným náhodnými vlivy může odhalit, jestli vznikají skutečně účinkem příslušných faktorů, anebo jestli jsou pouze náhodné.

Celkovou proměnlivost výsledků experimentů je možno vyjádřit prostřednictvím součtu čtverců rozdílů mezi jednotlivými pozorováními a celkovým průměrem všech hodnot. Vliv jednotlivých faktorů lze vyjádřit pomocí čtverců rozdílů mezi průměrným efektem příslušného faktoru a celkovým průměrem. Zůstává zbytkový rozptyl, který je založen na čtvercích rozdílů mezi jednotlivými pozorováními a průměry pro jednotlivé činitele. (POZNÁMKA. Rozptyl se počítá jako součet čtverců odchylek dělený počtem stupňů volnosti.)

**Analýza rozptylu** je založena na testování hypotéz. Testuje nulovou hypotézu „neexistuje významný rozdíl mezi vlivem jednotlivých činitelů“, což také znamená

„všechny naměřené hodnoty pocházejí ze stejného základního souboru“, a provádí to porovnáním různých rozptylů. Výběrové rozptyly mají  $\chi^2$ -rozdělení a podíl dvou veličin s  $\chi^2$ -rozdělením má  $F$ -rozdělení. Pro posouzení vlivu jednotlivých faktorů se tedy užívá  $F$ -test.

Příslušný postup bude ukázán na jednofaktorové analýze rozptylu, tj. s jedním faktorem [1]. Vztahy pro potřebné výpočty jsou shrnuty v Tabulce 1.

TABULKA 1.

Zdroj rozptylu	Součet čtverců	Stupně volnosti	Průměrný rozptyl	$F$
Faktor	$S_{fj} = n \sum (y_{j\cdot} - y_{\cdot\cdot})^2$	$(p - 1)$	$s_{fj}^2 = S_{fj} / (p - 1)$	$s_{fj}^2 / s_{res}^2$
Reziduální	$S_{res} = \sum (y_{jk} - y_{i\cdot})^2$	$(N - p)$	$s_{res}^2 = S_{res} / (p - 1)$	
Celkový	$S_{tot} = \sum (y_{jk} - y_{\cdot\cdot})^2$	$(N - 1)$		

Jednotlivé symboly mají následující význam:

$S_{tot}$  – celkový součet čtverců rozdílů mezi jednotlivými hodnotami a celkovým průměrem.

$S_{fj}$  – součet čtverců rozdílů mezi jednotlivými hodnotami a průměrem odpovídajícími  $j$ -tému vyšetřovanému faktoru.

$S_{res}$  – zbytkový součet čtverců rozdílů mezi jednotlivými hodnotami a průměrnými hodnotami skupin, odpovídajících jednotlivým faktorům.

$N$  – počet všech testů (nebo pozorování)

$n$  – počet pozorování (nebo testů) pro jednotlivé faktory

$p$  – počet faktorů

$s_{fj}^2$  – průměrný rozptyl  $j$ -tého faktoru

$s_{res}^2$  – průměrný reziduální rozptyl

$y_{jk}$  –  $k$ -tá hodnota  $j$ -tého faktoru

$y_{j\cdot}$  – průměr hodnot  $j$ -tého faktoru

$y_{\cdot\cdot}$  – celkový průměr všech hodnot

$\nu$  – počet stupňů volnosti (rovný počtu hodnot pro stanovení charakteristiky minus počet regresních konstant použitých při jejím určení).

Jestliže neexistuje významný rozdíl mezi rozptyly, má testová charakteristika  $F$ -rozdělení. Je-li hodnota  $F$ , vypočtená z naměřených hodnot, větší než  $\alpha$ -kritická hodnota rozdělení  $F$ , nastal jev, který byl očekáván jenom s velmi nízkou pravděpodobností  $\alpha$ , a činíme závěr, že tento rozdíl není náhodný – nulovou hypotézu zamítáme na hladině významnosti  $\alpha$ . V opačném případě řekneme, že rozdíl mezi jednotlivými faktory není významný. Aplikaci ukážeme na následujícím příkladu.

Příklad. Je třeba zjistit, má-li druh motorového oleje vliv na spotřebu paliva. Byly porovnávány tři druhy olejů (A, B, C) a každý byl testován v pěti vozidlech. Jednotlivé spotřeby paliva a průměrné (všechny v l/100 km) byly následující:

Značka	Průměrná spotřeba					Průměr
A:	7,7	8,1	7,1	7,6	8,0	$y_{jA} = 7,7$
B:	7,0	5,8	7,4	6,6	7,0	$y_{jB} = 6,8$
C:	7,6	8,5	8,2	8,0	7,7	$y_{jC} = 8,0$

Celkový průměr byl  $y_{..} = 7,5$  l/100 km. Tyto hodnoty byly vloženy do vzorců v Tabulce 1, spolu s  $p = 3$ ,  $n = 5$ ,  $N = pn = 3 \times 5 = 15$ . Výsledky jsou v Tabulce 2.

TABULKA 2.

Zdroj rozptylu	Součet čtverců	Stupně volnosti	Průměrný rozptyl	$F$
Faktor	$S_{f1} = 4,1853$	$3 - 1 = 2$	$s_{f1}^2 = 2,0927$	9,5410
Zbytkový	$S_{res} = 2,6320$	$15 - 3 = 12$	$s_{res}^2 = 0,2193$	
Celkový	$S_{tot} = 6,8173$	$15 - 1 = 14$		

Hodnota testového kritéria je  $F = s_{f1}^2/s_{res}^2 = 9,54$ . To je mnohem více než kritická hodnota  $F$ -rozdělení pro hladinu spolehlivosti  $\alpha = 5\%$  a stupně volnosti  $v_1 = 3 - 1 = 2$ ,  $v_2 = 15 - 3 = 12$ , která je  $F(0,05; 2; 12) = 3,885$ . Vypočtená hodnota  $F = 9,54$  je dokonce vyšší než kritická hodnota pro významnost  $\alpha = 1\%$ :  $F(0,01; 2; 12) = 6,927$ . Můžeme si tedy být prakticky jisti, že druh oleje má vliv na spotřebu paliva. (Čtenář si může pro pocvičení celý test zopakovat.)

Porovnání vypočtené hodnoty s kritickou hodnotou, odpovídající jisté hladině významnosti, je klasický přístup, vyvinutý v době, kdy byly k dispozici jenom tabulky kritických hodnot určitých rozdělení (např.  $F$  nebo  $t$ ) odpovídající vybraným pravděpodobnostem (např. 5%). V dnešní době mohou univerzální programy (včetně Excelu) počítat hodnoty distribučních funkcí mnoha rozdělení. A distribuční funkce udává pravděpodobnost nepřekročení. Můžeme tedy přímo určovat pravděpodobnost, že rozdíly mezi jednotlivými faktory jsou významné. Pro  $F = 9,54$  tato pravděpodobnost činí  $0,9967 \approx 99,7\%$  a pravděpodobnost, že naměřené rozdíly byly pouze náhodné, je  $\alpha = 1 - 0,9967 = 0,0033 \approx 0,3\%$  (viz funkci v Excelu  $F.DIST(F;v_1;v_2;TRUE)$  pro  $F = 9,54$ ,  $v_1 = 2$ ,  $v_2 = 12$ ).

POZNÁMKA:  $\alpha$ -kvantil odpovídá  $(1 - \alpha)$ -kritické hodnotě.

Navíc, univerzální statistické programy umožňují přímé provádění analýzy rozptylu. Je pouze nutno znát vstupní data (např. naměřené spotřeby paliva pro oleje A, B, C). Příslušné programy vykonají všechny potřebné výpočty (včetně určení stupňů volnosti) a uvedou výslednou hodnotu  $F$  a kritickou  $F$ -hodnotu pro zvolenou hladinu významnosti  $\alpha$ , spolu s pravděpodobností  $P$ , že vliv daného faktoru je nevýznamný. Na příklad, Excel nabízí několik druhů analýzy rozptylu; jsou přístupné z menu Analýza dat, které vyhledáme v menu Data. Hořejší problém tří olejů patří do kategorie “Anova: Jeden faktor”. Postačí zapsat naměřené spotřeby oleje do pole (matice) se třemi řádky a pěti sloupci (3 značky oleje a 5 testů pro každou), vložit toto pole do příslušného vstupního okénka v menu, označit pokyn „sdužit řádky”, a zadat požadovanou hladinu spolehlivosti testu. Po stisknutí ENTER se výsledky ukáží ve dvou tabulkách. První tabulka udává počet hodnot v jednotlivých porovnávaných příkladech (tj. v řádcích), jejich součty, průměry a rozptyly. Druhá tabulka ukazuje všechny důležité hodnoty zmíněné nahore; součet čtverců je zde označen SS a průměrné rozptyly MS. Doporučujeme čtenáři, aby sám vyřešil tuto úlohu se stejnými vstupními hodnotami a porovnal svoje výsledky s výsledky uvedenými v této kapitole – lepšímu pochopení a praxi.

Toto byl jednoduchý problém s jedním faktorem – pro ilustraci. Analýza rozptylu se užívá pro různé problémy, s tříděním podle dvou, tří, ale i více faktorů. Více lze najít podrobných příručkách matematické statistiky, např. [1 – 5]. Stručné vysvětlivky bývají též dostupné prostřednictvím příkazu Help v počítačových programech pro statistickou analýzu.

## **Literatura ke kapitole 10**

1. Felix, M., Bláha, K.: Matematickostatistické metody v chemickém průmyslu. SNTL, Praha, 1962. 336 s.
2. Freund, J. E.: Modern elementary statistics. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1981 (6th edition). 561 s.
3. Freund, J. E., Perles, B. E.: Modern elementary statistics. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 2006 (12th edition). 576 s.
4. Suhr, E.: Applied Probability for Engineers and Scientists. McGraw-Hill, New York, 1997. 593 s.
5. Montgomery, D. C., and Runger, G. C.: Applied Statistics and Probability for Engineers. John Wiley, New York, 2006 (4th edition). 784 s.

***Nejvyšším soudcem každé fyzikální teorie je experiment.***

Lev Landau

***Náhoda slouží připraveným.***

Louis Pasteur

# 11. Plánování experimentů (DOE)

Každé experimentální zkoumání začíná s předběžnými pokusy pro získání určité představy. Práce v tomto stadiu jsou zčásti intuitivní, s cílem lépe úlohu definovat. Jakmile jsou cíle zkoumání definovány, je následujícím úkolem snížit velký počet možných proměnných na několik nejdůležitějších. Užitečná je statistická analýza, protože může pomoci při výběru ze všech možných modelů. Když později experimentování přechází do stadia optimalizace (např. chemického složení, procesu nebo konstrukčního řešení, je při hledání optimálních parametrů opět efektivní statistický design experimentů. Tato kapitola ukazuje, jak je možno plánovat experimenty účinně tak, abychom získali požadované informace s minimálním úsilím. Jsou vysvětleny důležité pojmy, jako bloky, znáhodňování a latinské čtverce, jakož i hlavní pravidla a tabulky pro plánování pokusů. Jejich použití je ukázáno na příkladech.

**Proměnné**, které hrají roli v experimentech, mohou být klasifikovány jako kvantitativní, kvalitativní, nebo binární. **Kvantitativní odezva**, která se měří na spojitě stupnici, je nejobvyklejší a při statistické analýze se s ní pracuje nejsnáze. **Kvalitativní odezva**, jako lesk nebo zápach, lze řadit na stupnici například od 0 pro nejhorší možnost, až po 10 pro nejlepší. **Binární odezva** má jednu hodnotu ze dvou, např. dobrý nebo špatný, shodný nebo neshodný, muž nebo žena.

**Faktory** jsou experimentální proměnné kontrolované výzkumníkem. Důležitou částí plánování experimentálního programu je identifikace důležitých proměnných, které mají vliv na odezvu, a rozhodnutí, jak budou v pokusech využity. Vědecký model experimentu se zkoumá z hlediska důležitých proměnných. Velmi užitečné jsou předchozí zkušenosti. Často také při zjišťování významných faktorů využíváme výhody rozměrové analýzy (viz Kapitulu 9).

Faktory mohou být navzájem nezávislé, kdy určitý faktor ovlivňuje odezvu bez ohledu na ostatní faktory. Někdy ale mezi dvěma nebo více faktory existuje interakce, což znamená, že vliv jedné proměnné na odezvu závisí na úrovni dalších proměnných.

## Náhodný výběr

Při každém experimentálním zkoumání s velkým množstvím zkoušek nebo měření na různých vzorcích je důležité, aby každý vzorek před zahrnutím do experimentu měl stejnou šanci, že bude vybrán. Takovýto, tzv. náhodný výběr lze zajistit znáhodněním. Jeden způsob, jak znáhodnit dávku vzorků, je přiřadit každému z nich číslo, vložit očíslované lístky do nádoby, zamíchat a vytahovat podobně jako v loterii. (Dnes lze užívat náhodná čísla generovaná počítačem.) Tímto způsobem lze předejít tomu, aby se projevovaly systematické chyby např. v důsledku postupných malých změn testovacího zařízení nebo výkonu operátora. Odebírají-li se vzorky z velkých výkovků, je nutno přihlížet k rozdílným vlastnostem v různých místech výkovku. Je-li třeba určit průměrné vlastnosti celého výkovku, budeme znáhodněním míst odběru ve výkovku minimalizovat ovlivnění výsledků zkoušek místem odběru. Kromě veličin, které jsou pod kontrolou experimentátora, jsou i jiné veličiny, které nemusí být pod kontrolou. Příkladem je způsob, jakým pracují různí operátoři, nebo rozdílná vlhkost vzduchu nebo další vlivy okolí. Tyto vlivy lze také snížit použitím návrhu s bloky, jejichž popis následuje.

## Znáhodněné bloky, latinské čtverce

Pro zvýšení spolehlivosti závěrů se experiment často vícekrát opakuje. Častým problémem při těchto experimentech je ale dodržení stejných podmínek. Například dávka homogenní suroviny je dostačující jenom pro jednu sérii testů, a pro další sérii musí být použita jiná dávka. Vlastnosti jednotlivých dávek často kolísají. Jestliže k této skutečnosti nepřihlížíme a hodnotíme všechny naměřené hodnoty (z více dávek) společně, budou výsledky vykazovat větší rozptyl v důsledku kombinace přirozeného rozptylu a rozdílů vlastností mezi dávkami. Tento nedostatek lze zmenšit nebo eliminovat, jsou-li experimenty připraveny tak, že rozdíly mezi opakováními budou odděleny od rozdílů v důsledku rozdílných dávek (viz kapitolu Analýza rozptylu a knihy [1 – 5].) Experimenty se rozdělí do skupin s přibližně stejnými podmínkami, tzv. **bloky**. Blokem může být surovina z jedné dávky, ve srovnání s materiály z jiných dávek. Blokem také může být skupina pokusů vykonaných ve stejném chemickém reaktoru při stejných tepelných podmínkách, analýzy prováděné stejným technikem nebo vzorky odebrané z kontinuálního procesu během poměrně krátké doby.



Chceme-li předejít systematickým chybám způsobeným stejným pořadím operací v každém bloku, doporučuje se provádět jednotlivé experimenty v náhodném pořadí, podobně, jak bylo popsáno v předchozím odstavci. Hovoříme potom o **znáhodněných blocích**. Jestliže chceme na příklad zkoumat vliv teploty na výsledek chemické reakce, a v tomto zkoumání budou použity suroviny ze čtyř dodávek ( $B_1, B_2, B_3, B_4$ ) a čtyř teplot ( $T_1, T_2, T_3, T_4$ ) pro každou dodávku, můžeme použít následující náhodné uspořádání:

Dávka	Pořadí teplot v každé dávce			
$B_1$	$T_1$	$T_3$	$T_4$	$T_2$
$B_2$	$T_3$	$T_1$	$T_4$	$T_2$
$B_3$	$T_4$	$T_2$	$T_3$	$T_1$
$B_4$	$T_1$	$T_4$	$T_2$	$T_3$

Toto uspořádání připomíná takzvané **latinské čtverce** [1 – 5]. Při těchto pokusech se používá stejný počet variant pro každý faktor a jsou uspořádány systematictěji. Příkladem je latinský čtverec “ $4 \times 4$ ” pro tři faktory, naznačený níže. Řádky odpovídají úrovním prvního faktoru, sloupce odpovídají druhému faktoru, a písmena  $A, B, C, D$  odpovídají třetímu faktoru. Vytváření takového systému (rotace písmen pro třetí faktor) je zřejmé z tabulky.

Řádky	S l o u p c e			
	1	2	3	4
1	$A$	$B$	$C$	$D$
2	$B$	$C$	$D$	$A$
3	$C$	$D$	$A$	$B$
4	$D$	$A$	$B$	$C$

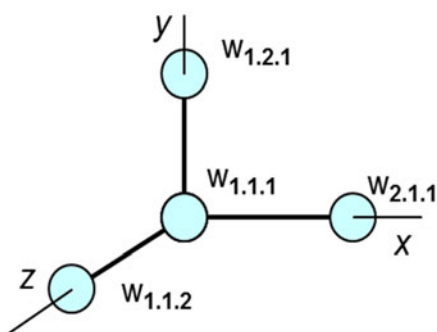
### Faktorové pokusy

Potřebný rozsah experimentů narůstá s počtem faktorů, které hrají roli. Chceme-li vědět, jestli faktor  $x$  má vliv na veličinu  $w$ , a jak silný je tento vliv, musíme udělat (alespoň) dva experimenty, tj. pro dvě hodnoty  $x$ . Je-li třeba vyšetřit vliv dvou faktorů, musíme udělat alespoň tři nebo čtyři pokusy. Tento počet silně narůstá s počtem faktorů. Naše práce bude efektivnější, použijeme-li pravidla pro **plánování experimentů** (tzv. design of experiments, DOE).

**DOE** znamená vytváření takových kombinací vstupních veličin, které poskytnou požadované informace s minimem experimentálního úsilí. Tento přístup budeme ilustrovat na případu tří nezávislých proměnných  $x$ ,  $y$ ,  $z$ , a závislé proměnné  $w$ . Nezávislé veličiny se obvykle nazývají **faktory**, a jejich jednotlivé hodnoty (**úrovně**) se označují indexy, např. 1 pro nižší úroveň a 2 pro vyšší úroveň, jestliže se užívají dvě úrovně. Například závisle proměnná odpovídající experimentu s faktorem  $x$  na nižší úrovni, s  $y$  na vyšší úrovni a se  $z$  na vyšší úrovni, se označí  $w_{1,2,2}$ . Užívají se i jiná označení pro uspořádání vstupních veličin, například  $-1$  pro nižší úroveň a  $+1$  pro vyšší úroveň. Používají-li se tři úrovně, použije se pro střední úroveň 0. Také se užívají symboly  $+$  a  $-$ , uspořádané do tabulek pro různé počty faktorů (viz konec této kapitoly).

**POZNÁMKA:** Případ se třemi faktory je zde užít pro ilustraci, protože si jej můžeme snadno představit v našem trojrozměrném prostoru. Popsané postupy lze rozšířit i na více faktorů.

Nejjednodušší uspořádání pro tři faktory je na Obr. 11.1. Udělají se čtyři experimenty, pro následující kombinace:  $w_{111}$ ,  $w_{211}$ ,  $w_{121}$ ,  $w_{112}$ . Vliv  $x$  na  $w$  se získá jako rozdíl hodnot  $w_{211} - w_{111}$ , vliv  $y$  se získá z rozdílu  $w_{121} - w_{111}$ , a vliv  $z$  je  $w_{112} - w_{111}$ . V těchto případech se vyšetřuje vždy vliv jedné veličiny, zatímco ostatní veličiny si ponechávají své původní úrovně. Chceme-li získat informaci o rozptylu, musíme experimenty nejméně dvakrát opakovat; to znamená, že bychom udělali nejméně osm pokusů.



**Obr. 11.1.** Jednoduchý experiment se třemi nezávislými faktory ( $x$ ,  $y$ ,  $z$ ), každý na dvou úrovních.

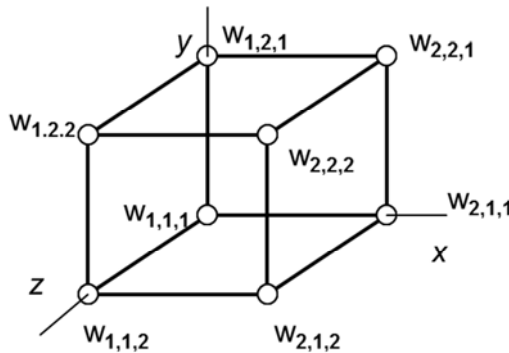
Více informací se získá z **úplného faktorového pokusu**. Veličina našeho zájmu,  $w$ , se určuje pro všechny možné kombinace všech faktorů a úrovní (Obr. 11.2). Potřebný počet experimentů je obecně

$$n_{\text{exp}} = (n_{\text{úroveň}})^{N_{\text{faktorů}}} \quad (11.1)$$

jestliže počet úrovní  $n_{\text{úroveň}}$  je stejný pro každý faktor. Pro tři faktory ( $N_{\text{faktorů}} = 3$ ), každý se dvěma úrovněmi, je počet experimentů  $n_{\text{exp}} = 2^3 = 8$ , s kombinacemi  $w_{111}$ ,  $w_{211}$ ,  $w_{121}$ ,  $w_{112}$ ,  $w_{221}$ ,  $w_{212}$ ,  $w_{122}$ ,  $w_{222}$  (viz Obr. 11.2 a tabulku na konci této kapitoly). Použijí-li se pro jednotlivé faktory různé počty úrovní, bude počet experimentů

$$n_{\text{exp}} = n_{\text{úroveň } 1} \times n_{\text{úroveň } 2} \times n_{\text{úroveň } 3} \times \dots \quad (11.2)$$

kde  $n_{\text{úroveň } j}$  označuje počet úrovní  $j$ -tého faktoru.



**Obr. 11.2.** Úplný faktorový experiment se třemi nezávislými faktory.

Výhodou úplného faktorového pokusu je, že pro posouzení vlivu jakéhokoliv faktoru se užijí výsledky všech experimentů. Tím se zvyšuje přesnost výsledků. Například, průměrný vliv faktoru  $x$  v našem případě se třemi faktory se stanoví sečtením jeho účinků pro různé hodnoty faktorů  $y$  a  $z$  a dělením čtyřmi (Obr. 11.2):

$$U_x = [(w_{221} - w_{121}) + (w_{211} - w_{111}) + (w_{222} - w_{122}) + (w_{212} - w_{112})] / 4 \quad (11.3)$$

Výsledek je možno přepsat:

$$U_x = [(w_{221} + w_{211} + w_{222} + w_{212}) - (w_{121} + w_{111} + w_{122} + w_{112})] / 4 \quad (11.4)$$

Výraz v první kulaté závorce je součet všech výsledků při  $x$  na vyšší úrovni, a výraz ve druhé závorce je součet všech výsledků při  $x$  na nižší úrovni. Rovnici (11.4) lze také zapsat jako

$$U_x = (w_{221} + w_{211} + w_{222} + w_{212} - w_{121} - w_{111} - w_{122} - w_{112}) / 4 \quad (11.5)$$

Podobně se získají vlivy  $y$  a  $z$ :

$$U_y = (w_{121} + w_{221} + w_{122} + w_{222} - w_{111} - w_{211} - w_{112} - w_{212}) / 4 \quad (11.6)$$

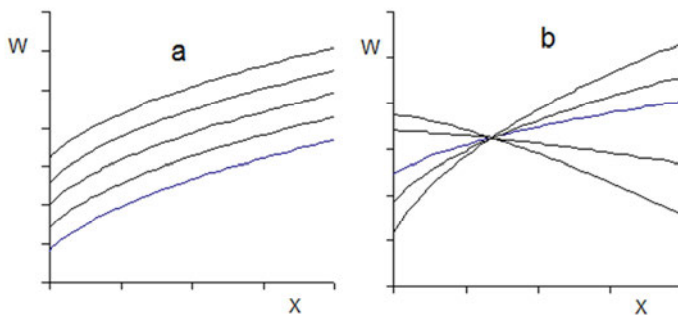
$$U_z = (w_{122} + w_{222} + w_{112} + w_{212} - w_{121} - w_{221} - w_{111} - w_{211}) / 4 \quad (11.7)$$

Jinou výhodou faktorových pokusů je možnost odhalit interakce ze stejných experimentů. **Interakce** znamená, že vliv jistého faktoru, například  $x$ , závisí také na hodnotách faktoru  $y$  nebo  $z$ , nebo na obou. Situace je schematicky znázorněna na Obr. 11.3 s křivkami pro různé hodnoty  $z$ ; levý obrázek je bez interakcí, a pravý obrázek odpovídá interakci  $x$ - $z$ . Pokud náš experiment se třemi faktory má odhalit interakci mezi faktory  $x$  a  $z$ , je řešení následující:

1) účinek  $x$  při jedné úrovni  $z$  se odečte od účinku  $x$  při druhé úrovni  $z$ :  $[(w_{212} - w_{112}) \text{ mínus } (w_{211} - w_{111})]$ .

2) Protože vliv  $y$  se neuvažuje, je nutno přidat podobné účinky také pro druhou úroveň  $y$ :  $[(w_{222} - w_{122}) \text{ mínus } (w_{221} - w_{121})]$ . Výsledek musí být opět dělen čtyřmi:

$$U_{xz} = [(w_{212} - w_{112}) - (w_{211} - w_{111}) + (w_{222} - w_{122}) - (w_{221} - w_{121})] / 4 \quad (11.8)$$



**Obr. 11.3.** Experimenty bez interakce (a) a s interakcí (b) některých faktorů. Jednotlivé křivky odpovídají různým hodnotám faktoru  $z$ .

Výraz (11.8) lze přepsat do tvaru:

$$U_{xz} = (w_{212} + w_{222} + w_{111} + w_{121} - w_{211} - w_{221} - w_{112} - w_{122}) / 4 \quad (11.9)$$

Účinky jiných interakcí lze získat podobným způsobem.

Následuje praktická ilustrace návrhu experimentů.

**Příklad.** Je třeba odhalit příčinu tvorby trhlinek na povrchu ocelových pružin při kalení. Jako tři nejvýznamnější faktory se jeví: teplota oceli před kalením ( $T_s$ ), teplota olejové lázně ( $T_o$ ), a obsah uhlíku v oceli ( $C$ ).

Pro kvantitativní charakterizování jejich vlivu byl navržen úplný faktorový experiment, s každým faktorem na dvou úrovních:

Úroveň	$T_s$ (°C)	$T_o$ (°C)	$C$ (%)
Nízká (–)	830	70	0,5
Vysoká (+)	910	120	0,7

Počet experimentů byl  $2^3 = 8$ . Kombinace úrovní a odpovídající počty trhlinek  $N$  nalezených na kalených pružinách jsou v následující tabulce (viz též Obr. 2;  $x$  odpovídá  $T_s$ ,  $y$  odpovídá  $T_o$ , a  $z$  odpovídá  $C$ ; + odpovídá vyšší úrovni a – odpovídá nižší úrovni):

Test č.	$T_s$	$T_o$	$C$	$T_s$ (°C)	$T_o$ (°C)	$C$ (%)	$N$
1	–	–	–	830	70	0,5	67
2	+	–	–	910	70	0,5	79
3	–	+	–	830	120	0,5	59
4	+	+	–	910	120	0,5	90
5	–	–	+	830	70	0,7	61
6	+	–	+	910	70	0,7	75
7	–	+	+	830	120	0,7	52
8	+	+	+	910	120	0,7	87

Průměrné vlivy teploty oceli ( $U_s$ ), teploty oleje ( $U_o$ ) a obsahu uhlíku ( $U_c$ ), vypočtené z rovnic (5) – (7), jsou:

$$U_s = (79+90+75+87-67-59-61-52)/4 = 23,0$$

$$U_o = (59+90+52+87-67-79-61-75)/4 = 1,5$$

$$U_c = (52+87+61+75-59-90-67-79)/4 = -5,0$$

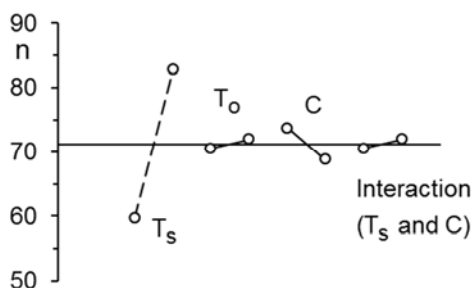
Největší vliv má teplota oceli; vliv obsahu uhlíku (v rozmezí 0,5 ÷ 0,7 %) je malý a vliv teploty oleje (v rozmezí 70°C ÷ 120°C) je zanedbatelný.

Interakce teploty oceli a obsahu uhlíku podle rovnice (9) je

$$U_{sc} = [(75+87+67+59)-(79+90+61+52)]/4 = 1,5$$

tedy také zanedbatelný. Další interakce by bylo možno najít podobným způsobem.

Velmi informativní je grafické znázornění (Obr. 4). Vodorovná osa představuje průměr všech hodnot, tj.  $(67+79+59+90+61+75+52+87)/8 = 71,25$ . Vliv teploty oceli je zobrazen dvěma body vlevo: jeden udává průměrný počet trhlin v případech, kdy teplota oceli byla nižší, tj.  $(59+67+52+61)/4 = 59,75$ , a druhý odpovídá vyšší teplotě,  $(90+79+87+75)/4 = 82,75$ . Pro lepší viditelnost jsou spojeny přímkou. POZNÁMKA:  $82,75 - 59,75 = 23,0 = U_s$ . Obrázek 4 také zobrazuje vliv teploty oleje a obsahu uhlíku.



**Obr. 11.4.** Vliv jednotlivých faktorů na jakost pružin.

Někdy je nutné vyšetřovat vliv většího počtu činitelů. Tabulka 1 na následující straně uvádí úrovně jednotlivých faktorů pro úplné faktorové pokusy se dvěma, třemi a čtyřmi faktory, každý na dvou úrovních. Čtyři testy jsou nutné pro dva faktory (sloupce A,B), osm testů pro tři faktory (A,B,C), a 16 testů pro čtyři faktory (A,B,C,D). Tabulku můžeme snadno rozšířit pro více faktorů, všimneme-li si, jak se střídají plusy a mínusy u jednotlivých faktorů, počínaje se sloupcem A pro jednofaktorový pokus a se sloupci A,B pro dvoufaktorový pokus.

**Tabulka 1.** Kombinace úrovní pro pokusy se dvěma, třemi a čtyřmi faktory.

Test	A	B	C	D
1	-	-	-	-
2	+	-	-	-
3	-	+	-	-
4	+	+	-	-
5	-	-	+	-
6	+	-	+	-
7	-	+	+	-
8	+	+	+	-
9	-	-	-	+
10	+	-	-	+
11	-	+	-	+
12	+	+	-	+
13	-	-	+	+
14	+	-	+	+
15	-	+	+	+
16	+	+	+	+

V průmyslovém výzkumu, například při optimalizaci výrobních podmínek, hraje často roli více faktorů, a počet pokusů pro úplný faktorový experiment by byl velmi vysoký (např. 256 pokusů pro 8 faktorů, každý na dvou úrovních). V takových případech se užívají **krácené faktorové experimenty**, kde jsou některé kombinace úrovní vynechány. Pro tento účel byla vyvinuta speciální schémata (tzv. ortogonální matice). Toto téma přesahuje rámec této knížky a čtenáře odkazujeme na knihy o navrhování experimentů, robustním designu a odpovídajících metodách, které vyvinul Genichi Taguchi a další autoři [6 – 9]. Pro navrhování experimentů obecně lze doporučit inženýrskou příručku [10] a podrobnou monografii [11].

### Literatura ke kapitole 11

1. Freund, J. E.: Modern elementary statistics. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1981 (6th edition). 561 s.

2. Freund, J. E., Perles, B. E.: Modern elementary statistics. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 2006 (12th edition). 576 s.
3. Suhir, E.: Applied Probability for Engineers and Scientists. McGraw-Hill, New York, 1997. 593 s.
4. Montgomery, D. C., Runger, G. C.: Applied Statistics and Probability for Engineers. John Wiley, New York, 2006 (4th edition). 784 s.
5. Felix, M., Bláha, K.: Matematickostatistické metody v chemickém průmyslu.) SNTL, Praha, 1962. 336 s.
6. Taguchi, G.: Introduction to Quality Engineering: Designing Quality into Products and Processes. Asian Productivity Organization, 1986. 191 s.
7. Ross, P. J.: Taguchi Techniques for Quality Engineering. McGraw-Hill, New York, 1996. 329 s.
8. Taguchi, G., Chowdhury, S., Wu, Y.: Taguchi's Quality Engineering Handbook. John Wiley & Sons, Hoboken, New Jersey, 2005. 1804 s.
9. Fowlkes, W. Y., Creveling, C. M.: Engineering methods for robust product design. (Using Taguchi methods in technology and product development.) Addison-Wesley, New York, 1995. 432 s.
10. Dieter, G. E.: Engineering design. 2nd edition. McGraw-Hill, New York, 1991. 721 s.
11. Montgomery, D. C.: Design and analysis of experiments. 8th edition, Wiley, 2012. 730 s.

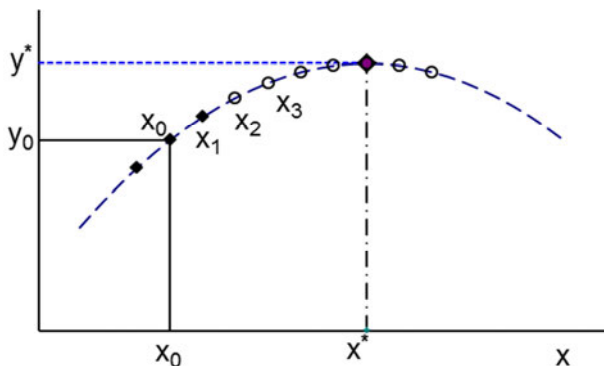


## 12. Experimentální nalezení extrémů

Obvyklou úlohou ve výzkumu je nalezení maxima nebo minima určité veličiny, například jistého parametru stroje nebo sloučeniny, účinnosti nebo nákladů. Tato kapitola ukazuje postupy pro nalezení extrému funkce (maximum nebo minimum) v případech, kdy její analytický tvar není znám a jednotlivé hodnoty je nutno zjišťovat experimenty nebo počítačovým modelováním pro určité kombinace hodnot vstupních veličin. Nejprve je vysvětlena intuitivní metoda postupných změn vstupních hodnot, dále metoda největšího spádu, a simplexová metoda, která je velmi účinná v případech s několika nezávislými proměnnými. Zmíněna je i novější metoda simulovaného žhání a metoda genetických algoritmů. Postupy budeme ilustrovat na příkladu hledání maxima.

### A. Postupné změny jednotlivých proměnných

Experimentální hledání maxima funkce jedné proměnné,  $y = f(x)$ , je velmi snadné (Obr. 12.1). Začneme se dvěma experimenty – pro hodnoty  $x_0$  a  $x_1$ . Jestliže hodnota  $y(x_1)$  byla vyšší než  $y(x_0)$ , uděláme další experiment s  $x$  změněným ve stejném směru,  $x_2 = x_1 + \Delta x$ , kde  $\Delta x$  je vhodně zvolený přírůstek. Tímto způsobem pokračujeme tak dlouho, až  $y$  začne klesat; maximum leží přibližně na hodnotě



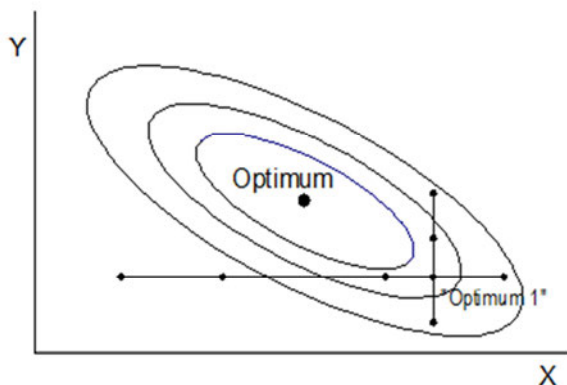
**Obr. 12.1.** Experimentální nalezení maxima funkce jedné nezávisle proměnné.

předchozího kroku nebo blízko ní (Obr. 12.1). Přesnější polohu maxima lze najít tak, že zde provedeme více pokusů.

Hledáme-li minimum, pohybujeme se opačným směrem.

Při **dvou nezávisle proměnných** předpokládáme, že funkci  $z(x, y)$  lze aproximovat polynomem, přinejmenším v blízkosti zvoleného výchozího bodu  $x_0, y_0$ . Provedeme zde několik experimentů, při kterých měníme pouze jednu proměnnou, např.  $x$ , přičemž  $y$  si ponechává svou původní hodnotu. Tímto způsobem pokračujeme ve směru rostoucího  $z$ , dokud  $z$  nezačne klesat. Předchozí krok odpovídá lokálnímu maximu  $z$ . Nyní se budeme z tohoto bodu pohybovat v kolmém směru a měníme pouze  $y$  a pokračujeme až do dosažení lokálního maxima  $z$ , a tak dále. Situace je vyobrazena na Obr. 12.2.

Je-li přibližně známa poloha maxima (nebo minima), lze také udělat několik pokusů kolem tohoto bodu, například 4 až 8 pro dvě nezávisle proměnné, a proložit získané hodnoty odezovou plochou (Obr. 13.1 v kapitole 13); polynom druhého řádu často stačí. Přesnou polohu extrému potom najdeme s použitím standardních matematických metod nebo vhodného řešitele. Více o tomto tématu lze najít v Kapitole 13 a literatuře tam doporučené.



*Obr. 12.2. Experimentální hledání maxima funkce dvou nezávislých proměnných  $x, y$ .*

## B. Metoda největšího spádu

Tato metoda se snaží blížit k maximu (nebo minimu) nejrychlejším způsobem, tj. ve směru **gradientu** k odezvové ploše [1– 3]. Nejprve je nutno najít tento gradient,

jak bude ukázáno dále v případě dvou nezávislých proměnných. Několik pokusů se provede kolem vhodně vybraného bodu  $x_0, y_0$ . Hodnoty  $z(x, y)$  mohou být proloženy polynomem, například

$$z = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots + b_1y + b_2y^2 + \dots \quad (12.1)$$

Vektor gradientu se obvykle získá prostřednictvím parciálních derivací,

$$\mathbf{grad} z = \partial z / \partial x \mathbf{i} + \partial z / \partial y \mathbf{j} + \dots \quad (12.2)$$

$\mathbf{i}$  a  $\mathbf{j}$  jsou jednotkové vektory ve směrech  $x$  a  $y$ . Pro polynom (1) je gradient

$$\mathbf{grad} z = (a_1 + 2a_2x + 3a_3x^2 + \dots) \mathbf{i} + (b_1 + 2b_2y + \dots) \mathbf{j} \quad (12.3)$$

V malém okolí bodu  $x_0, y_0$  obvykle postačí polynom prvního stupně (tečná rovina),

$$z = a_0 + a_1x + b_1y \quad (12.4)$$

s gradientem

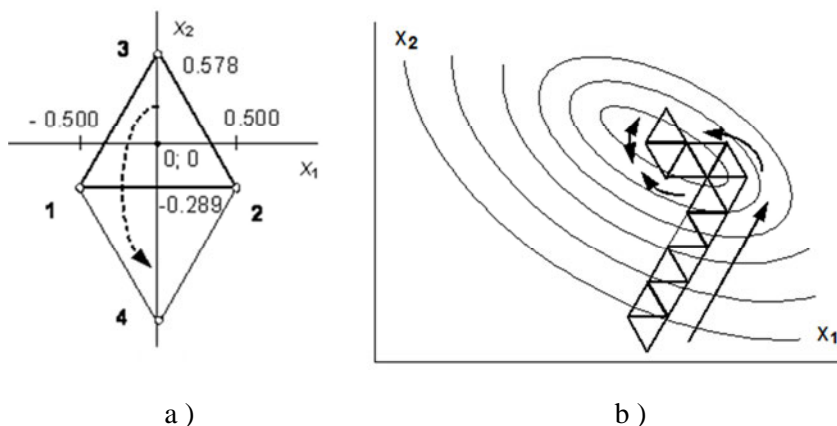
$$\mathbf{grad} z = a_1 \mathbf{i} + b_1 \mathbf{j} \quad (12.5)$$

Nyní pokračujeme tímto směrem k lokálnímu maximu, s kroky úměrnými  $a_1$  ve směru  $x$  a  $b_1$  ve směru  $y$ , až hodnoty  $z$  začnou klesat. Kolem tohoto bodu učiníme opět několik pokusů a nalezneme směr nejrychlejšího nárůstu, atd. Použití tohoto postupu při třech a více proměnných je podobné.

### C. Simplexová metoda

Toto je jednoduchá metoda, při které se vstupní veličiny blíží k optimu postupnými kroky podle algoritmu, který navrhl Spendley aj. [4]. Znalost gradientu není nutná. V prvním kroku se vytvoří tzv. *simplex*. To je jednoduché fiktivní konvexní těleso s  $n + 1$  vrcholy; jejich počet je o jeden vyšší, než počet  $n$  vstupních proměnných (například trojúhelník pro dvě nezávisle proměnné, a čtyřstěn pro tři vstupní proměnné). Souřadnice vrcholů odpovídají hodnotám vstupních parametrů. Pro všechny tyto body se vypočtou hodnoty výstupní veličiny  $z$ . V následujícím kroku se vytvoří nový simplex tak, že se vrchol s nejhorší hodnotou  $z$  nahradí novým, jehož hodnoty jsou mu zrcadlově symetrické. (V dvourozměrném prostoru se nový trojúhelníkový simplex získá překlopením původního trojúhelníku kolem hrany protilehlé nejhoršímu vrcholu, Obr. 12.3a.) Pro tento nový vrchol se opět vypočte hodnota závisle proměnné. Nyní se porovnají hodnoty závisle proměnné pro

všechny vrcholy tohoto nového simplexu, a zase se odvrhne nejhorší vrchol a nahradí novým. Takto se pokračuje tak dlouho, až optimalizovaná veličina dosáhne extrémní nebo přijatelné hodnoty. Dosažení extrému je obvykle indikováno oscilací simplexu mezi dvěma polohami, nebo pohybem simplexových těles kolem uzavřené křivky (Obr. 12.3b).



**Obr. 12.3.** Simplexová metoda. a) Bezrozměrový simplex 1-2-3 pro dvě nezávislé proměnné ( $x_1, x_2$ ) a jeden optimalizační krok (vytvoření vrcholu 4); b) pohyb směrem k optimu.

POZNÁMKA: Pozorný čtenář si jistě všimnul podobnosti mezi touto metodou a obrázkem 1 pro jednu nezávislou proměnnou.

Praktický postup je následující [4, 5]. Nejprve se zvolí souřadnice  $x_{i,0}$  výchozího bodu (= těžiště simplexu) a přírůstky jednotlivých proměnných  $\Delta x_{i,j}$ . (První index značí proměnnou;  $i = 1, 2, \dots, n$ , kde  $n$  je počet nezávislých proměnných; druhý index označuje krok.) Potom se souřadnice jednotlivých  $n+1$  vrcholů prvního simplexu vypočítají prostřednictvím přidruženého bezrozměrového regulárního simplexu jako [4, 5]:

$$x_{i,l} = x_{i,0} + z_i \Delta x_i \quad (12.6)$$

$z_i$  označuje poloměr koule vepsané ( $r_i$ ) nebo opsané ( $R_i$ ) přidruženému simplexu.

Těžiště tohoto simplexu je v počátku souřadného systému, a příslušné poloměry se počítají pomocí následující matice a vzorců:

$$\begin{bmatrix} -r_1 & -r_2 & -r_3 & \dots & -r_{n-1} & -r_n \\ R_1 & -r_2 & -r_3 & \dots & -r_{n-1} & -r_n \\ 0 & R_2 & -r_3 & \dots & -r_{n-1} & -r_n \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & R_{n-1} & R_n \\ 0 & 0 & 0 & \dots & 0 & R_n \end{bmatrix}, \quad r_i = \frac{1}{\sqrt{2i(i+1)}}; \quad R_i = \sqrt{\frac{i}{2(i+1)}} \quad (12.7)$$

Sloupec označuje proměnnou a řádek číslo vrcholu; matice má  $n$  sloupců a  $n + 1$  řádků. Odezva se potom počítá pro všech  $n + 1$  kombinací vstupních hodnot. Souřadnice  $x_i$  nového, obecně  $(j + 1)$ -vrcholu, se pak určí jako

$$x_{i,j+1} = \frac{2}{n} \sum_{i=1}^n x_{i,j} - x_{i,j}^* \quad (12.8)$$

První index značí proměnnou ( $x_1, x_2, \dots$ ), druhý index udává číslo vrcholu.  $x_{i,j}^*$  je souřadnice bodu s nejhorší hodnotou optimalizačního kritéria ( $y$ ), a  $(\sum x_{i,j})/n$  je průměr souřadnic všech vrcholů ( $j$ -tého simplexu) kromě nejhoršího. Tímto způsobem se získá všech  $n$  souřadnic nového vrcholu.

Postup objasníme na příkladu. Mějme  $n = 2$  nezávislé proměnné  $x_1$  and  $x_2$ . (Např.  $x_1$  je šířka obdélníkového průřezu součásti zatížené tahem, a  $x_2$  je jeho výška). Necht' souřadnice výchozího bodu jsou  $x_{1,0} = 200$  mm,  $x_{2,0} = 500$  mm, a jejich přírůstky při optimalizačních krocích budou  $\Delta x_1 = 20$  mm,  $\Delta x_2 = 30$  mm. (Jednotlivé proměnné mohou obecně mít různé rozměry.) Bezrozměrová matice (12.7) má nyní  $n = 2$  sloupce a  $n + 1 = 3$  řádky:

$$\begin{bmatrix} -r_1 & -r_2 \\ R_1 & -r_2 \\ 0 & R_2 \end{bmatrix} = \begin{bmatrix} -0.500 & -0.289 \\ 0.500 & -0.289 \\ 0 & 0.577 \end{bmatrix}$$

Levý (pravý) sloupec udává bezrozměrové souřadnice pro  $x_1$  ( $x_2$ ), a řádky přísluší prvnímu, druhému a třetímu vrcholu. Hodnoty  $r_j$  a  $R_j$  byly vypočítány podle vzorců na pravé straně výrazů (12.7).

Souřadnice tří vrcholů prvního simplexu (viz Obr. 12.3a) jsou následující: hodnota  $x_1$  prvního vrcholu je  $x_{1,1} = x_{1,0} - 0,500 \times \Delta x_1 = 200 - 0,500 \times 20 = 190$  mm, hodnota  $x_1$  druhého vrcholu je  $x_{1,2} = x_{1,0} + 0,500 \times \Delta x_1 = 200 + 0,500 \times 20 = 210$  mm, a třetí vrchol je  $x_{1,3} = x_{1,0} + 0 \times \Delta x_1 = 200 + 0 \times 20 = 200$  mm. Odpovídající hodnoty  $x_2$  jsou:  $x_{2,1} = x_{2,0} - 0,289 \times \Delta x_2 = 500 - 0,289 \times 30 = 491,33$  mm,  $x_{2,2} = x_{2,0} - 0,289 \times \Delta x_2 = 500 - 0,289 \times 30 = 491,33$  mm, a  $x_{2,3} = x_{2,0} + 0,578 \times \Delta x_2 = 500 + 0,578 \times 30 = 517,34$  mm.

Všechny hodnoty jsou uspořádány v následující tabulce:

Vrchol	$x_1$	$x_2$
1	190	491,3
2	210	491,3
3	200	517,3

Jestliže nejhorší hodnota optimalizačního kritéria patřila např. vrcholu č. 3, budou souřadnice nového vrcholu (č. 4) nyní  $x_{1,4} = 2(190 + 210)/2 - 200 = 200$  mm, a  $x_{2,4} = 2(491,3 + 491,3)/2 - 517,3 = 465,3$  mm. Nový simplex je definován vrcholy 1, 3 a 4 [viz také Obr. 12.3a]; bezrozměrové souřadnice vrcholu 4 jsou 0 a  $-0,289 - (0,289 + 0,578) = -1,156$ . Následuje výpočet optimalizačního kritéria ( $y$ ) v tomto vrcholu, porovnání hodnot  $y$  pro body 1, 2 a 4 a výpočet souřadnic vrcholu 5, a tak dále, až hodnota optimalizačního kritéria přestane růst.

Popsaná simplexová metoda má několik předností. Algoritmus je velmi jednoduchý a souřadnice nového vrcholu se počítají přímo ze souřadnic předchozího simplexu. Není třeba určovat žádný gradient. Každým krokem se přibližujeme k optimu. Metoda je vhodná i pro vyšší počty proměnných. Protože konstrukce nového vrcholu není založena na přesných hodnotách závisle proměnné, ale na jejich porovnání, nejsou zde nadměrné požadavky na přesnost výsledků v jednotlivých etapách optimalizace. Existují-li omezení pro některé proměnné, a vypočtené souřadnice nového vrcholu se octly mimo povolené meze, je možné vynechat vrchol s nejhorší hodnotou a snažit se zlepšit druhý nejhorší vrchol atd.

#### D. Další přístupy

Existují i další metody pro nalezení extrému určité funkce. Jsou ale náročnější a

vyžadují samočinný počítač a vhodný program. Zde zmíníme stručně dvě z nich: simulované žíhání a genetické algoritmy.

**Simulované žíhání** (simulated annealing). Až dosud byly popisovány metody, které se snažily blížít k maximu tak, že se v každém kroku změnila hodnota nezávisle proměnné (resp. proměnných) ve stejném směru, v jakém došlo při předcházejícím kroku ke zvýšení hodnoty sledované veličiny, resp. účelové funkce; viz např. Obr. 12.1 a 12.2. Nedostatkem je, že tento postup může najít pouze lokální maximum (nebo minimum). Někdy může ve větší oblasti existovat více lokálních maxim, a úkolem pak je najít globální maximum s nejvyšší hodnotou ze všech. (Hledání minima je analogické.) Název simulované žíhání byl vytvořen v analogii s tepelným zpracováním (žíháním), kdy se pomalým řízeným snižováním původní vysoké teploty snažíme docílit stavu tělesa s minimální vnitřní energií a vadami.

Metoda simulovaného žíhání postupuje po krocích. Ve srovnání s tzv. gradientními metodami optimalizace, popisovanými dříve, umožňuje tato metoda náhodné hledání v různých směrech, a s určitou pravděpodobností přijímá i řešení horší než řešení stávající. Díky tomu se může v následujících krocích dostat i z lokálního maxima a je schopna vyšetřovat více možných řešení. Postupně se pak blíží k celkovému maximu (popřípadě minimu) ve zkoumané oblasti.

Pro použití metody simulovaného žíhání je nezbytný vhodný počítačový program. Tím je například Simulated Annealing Solver, který je součástí souboru Global Optimization Toolbox v rámci výpočetního systému Matlab. Podrobnosti viz <http://www.mathworks.com> [6]. Obecné informace o této metodě lze také najít ve Wikipedii a pramenech v ní uvedených.

**Genetické algoritmy** řeší optimalizační úlohy tak, že napodobují principy biologického vývoje. Postupně generují různá řešení daného problému. Při řešení se pracuje s tzv. populací, jejíž každý jedinec představuje jedno řešení problému. To je obvykle reprezentováno binárními čísly, tj. řetězci nul a jedniček; používají se ale i jiná vyjádření, např. s maticemi. Na začátku optimalizačního procesu (v první generaci) je populace složena z naprosto náhodných členů. Z ní jsou stochasticky vybráni jedinci, kteří jsou pak modifikováni pomocí mutací a křížení, čímž vznikne nová populace. (Mutace zde znamená náhodnou změnu části jedince, zatímco křížení znamená prohození části několika jedinců mezi sebou.) Po přechodu do nové generace je pro každého jedince spočtena tzv. *fitness* funkce,

kteřá vyjadřuje jeho zdatnost, resp. kvalitu řešení reprezentovaného tímto jedincem. Podle ní pokračuje výběr a další modifikace. Tento postup se opakuje, čímž se kvalita řešení v populaci postupně vylepšuje (tj. „šlechtění“). Algoritmus se obvykle zastaví při dosažení postačující kvality řešení, případně po předem dané době.

Pro praktickou aplikaci genetických algoritmů je vhodný například program Genetic Algorithm Solver, který je součástí souboru Global Optimization Toolbox v rámci systému Matlab. Detaily lze opět najít na <http://www.mathworks.com>, některé obecné informace jsou ve Wikipedii a pramenech tam uvedených.

### Literatura ke kapitole 12

1. Felix, M., Bláha, K.: Matematickostatistické metody v chemickém průmyslu. SNTL, Praha, 1962. 336 s.
2. Arora, J.: Introduction to optimum design (third edition), Elsevier, 2011, 896 s.
3. Myers, R. H., and Montgomery, D. C.: Response surface methodology. Wiley, 2016 (4th Edition). 856 s.
4. Spendley W., Hext, G. R., Himsworth, F. R.: Sequential Application of Simplex Designs in Optimisation and Evolutionary Operation (1962) *Technometrics*, Vol. 4, No. 4, 441 – 461.
5. Tichomirov, V. B.: Planirovanije i analiz eksperimenta pri provedenii issledovanij v legkoj i tekstilnoj promyšlennosti. Legkaya industria, Moskva, 1974. 262 s.
6. Matlab – Simulated Annealing Solver; <http://www.mathworks.com>
7. Matlab – Genetic Algorithm Solver; <http://www.mathworks.com>



## 13. Citlivostní analýza

Vstupní veličiny často kolísají nebo se odchyľují od jmenovitých hodnot. Účelem citlivostní analýzy je ukázat, jak toto kolísání přispívá k odchylkám vyšetřované veličiny od její jmenovité nebo předpokládané hodnoty. To je důležité pro předpovídání odezvy za reálných podmínek. Analýzu citlivosti lze dělat s použitím analytických výrazů nebo simulačních pravděpodobnostních metod. V prvním případě je nutno znát vztah mezi výstupní veličinou  $z$  a vstupními proměnnými  $x_1, x_2, \dots, x_n$ , tzv. **odezovou funkci**. Přesný analytický výraz

$$y = f(x_1, x_2, \dots, x_n) \quad (13.1)$$

bývá k dispozici jenom u jednoduchých úloh. Často je nutno odezvu určovat pomocí experimentů nebo s použitím časově náročných numerických výpočtů. V takových případech se někdy užívají přibližné analytické výrazy, získané tak, že odezvu, určenou pouze pro několik kombinací vstupních parametrů, proložíme regresní funkcí.

Nejjednodušší tvar odezvové funkce je polynom, na příklad

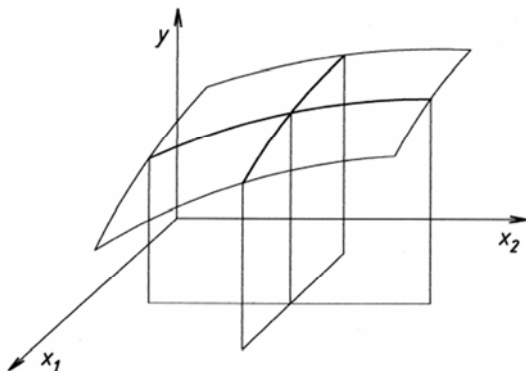
$$y_i = a_0 + a_{11}x_i + a_{12}x_i^2 + a_{13}x_i^3 + \dots \quad (13.2)$$

nebo

$$y_i = y_0 + a_i(x_i - x_{i,0}) + b_i(x_i - x_{i,0})^2 + \dots \quad (13.3)$$

Rovnice (13.3) vyjadřuje změny  $y$  jako funkci odchylek vstupní proměnné  $x_i$  od jmenovité hodnoty  $x_{i,0}$ . Index  $i$  označuje  $i$ -tou proměnnou, a  $y_i$  odpovídá této proměnné. Tyto regresní funkce odpovídají řezům odezvou plochou (Obr. 13.1). V blízkosti návrhového bodu ( $x_{i,0}$ ) jsou polynomy druhého stupně často dostačující.

Polynom, nebo dokonce lineární funkce mohou být užity i pro aproximování jiných vztahů (např.  $1/x$  nebo  $\sqrt{x}$ ), jestliže je provedena vhodná transformace. Řešiče v univerzálních programech umožňují snadné stanovení regresních koeficientů v složitých funkcích přímým užitím metody nejmenších čtverců, bez transformací, jak bylo popsáno v Kapitole 7. Další poznámky k odezovým funkcím lze ještě najít v Kapitole 14.



**Obr. 13.1.** Odezvová plocha pro 2 nezávislé proměnné, s řezy  $x_1 = konst$ ,  $x_2 = konst$  [1].

Citlivostní analýza závisí na tom, jestli odchylky jednotlivých veličin od jejich jmenovitých hodnot můžeme považovat za deterministické nebo náhodné [2, 3].

### Deterministické odchylky vstupních veličin

Citlivost odezvy na změny jednotlivých proměnných se zjistí pomocí parciálních derivací v příslušném bodě,

$$c_i = (\partial y / \partial x_i) \quad (13.4)$$

Citlivostní koeficienty  $c_i$  odpovídají konstantám  $a_{i,1}$  v rovnici (13.2) a  $a_i$  v rovnici (13.3). Další informaci poskytují **relativní citlivosti**,

$$c_{ri} = \frac{\partial y}{\partial x_i} \frac{x_{i,0}}{y_0} \approx \frac{\Delta y}{y_0} / \frac{\Delta x_i}{x_{i,0}} \quad (13.5)$$

$y_0$  a  $x_{i,0}$  jsou hodnoty odpovídající návrhovému bodu. Koeficient  $c_{ri}$  vyjadřuje relativní změnu  $y$  (například v procentech), způsobenou 1% odchylkou  $x_i$  od jmenovité hodnoty  $x_{i,0}$ . Pro lineární aproximaci,  $c_{ri} = a_i \times (x_{i,0}/y_0)$ .

Citlivostní analýza také ukáže, které vstupní veličiny mají zanedbatelný nebo malý vliv na proměnlivost výstupní veličiny  $y$ , a mohou být (přibližně) považovány za konstanty při složitější analýze s více vstupními veličinami. (Nezapomeňme, že rozptyl výstupu závisí na rozptylech vstupních veličin  $x_i$ , ale i na citlivostech  $c_i$  !)

Jsou-li přírůstky  $\Delta x_i$  malé, může být odezвовá plocha aproximována lineárním výrazem

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (13.6)$$

který představuje rovinu pro dvě nezávislé proměnné ( $n = 2$ ) a nadrovinu pro více vstupních proměnných. Konstanty odpovídají citlivostním koeficientům (kromě  $a_0$ ), a získají se proložením  $n + 1$  hodnot odezvy vícenásobnou lineární regresní funkcí. (Také Excel může být použit pro tento účel). Přírůstky  $y$  se počítají prostřednictvím prvých derivací. Pro  $y = f(x_1, x_2, \dots, x_n)$  je infinitesimální přírůstek  $y$  obecně

$$dy = (\partial y/\partial x_1)dx_1 + (\partial y/\partial x_2)dx_2 + \dots + (\partial y/\partial x_n)dx_n \quad (13.7)$$

kde  $\partial y/\partial x_i$  vyjadřuje parciální derivace. Při praktické analýze se diferenciály nahrazují malými konečnými přírůstky  $\Delta$ ,

$$\Delta y = (\partial y/\partial x_1)\Delta x_1 + (\partial y/\partial x_2)\Delta x_2 + \dots + (\partial y/\partial x_n)\Delta x_n \quad (13.8)$$

Použití citlivostní analýzy budeme ilustrovat na příkladu [1, 3] malé ploché pružiny pro měřicí přístroj (Obr. 13.2). Chceme znát citlivost poddajnosti této pružiny na změny jejích rozměrů a modulu pružnosti materiálu. Tato poddajnost je

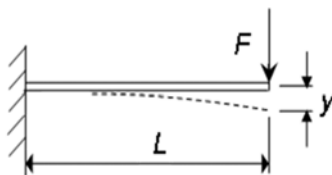
$$C = y/F = 4L^3/(Ewt^3) \quad (13.9)$$

$y$  je průhyb,  $F$  – zatížení,  $L$  – délka,  $E$  – modul pružnosti,  $w$  – šířka pružiny,  $t$  – tloušťka pružiny. Parciální derivace rovnice (9) podle první proměnné ( $x_1 = L$ ) je

$$\partial C/\partial L = 3L^2 \times 4/(Ewt^3) = [4L^3/(Ewt^3)] \times 3/L = (3/L) \times C \quad (13.10)$$

a přírůstek poddajnosti v důsledku malého přírůstku délky pružiny  $\Delta L$  je tedy

$$\Delta C = 3C (\Delta L/L) \quad (13.11)$$



**Obr. 13.2.** Pružina pro měřicí přístroj [1].

Vzorce pro jiné proměnné se získají podobným způsobem. Výsledný vztah, zahrnující změny všech proměnných, je

$$\Delta C = C (3\Delta L/L - \Delta E/E - \Delta w/w - 3\Delta t/t) \quad (13.12)$$

Relativní citlivost

$$\Delta C/C = 3\Delta L/L - \Delta E/E - \Delta w/w - 3\Delta t/t \quad (13.13)$$

ukazuje názorně vliv odchylek jednotlivých veličin. Bude-li pružina delší o 1% než je jmenovitá hodnota, bude poddajnost vyšší o 3%; jestliže modul pružnosti  $E$  bude vyšší o 1%, bude poddajnost nižší o 1%, atd. Konstanty u jednotlivých členů odpovídají jejich exponentům v rovnici (13.9), a znaménka závisí na tom, jestli příslušná veličina byla v čitateli nebo jmenovateli.

### Vliv náhodného kolísání vstupních veličin

Vliv náhodného kolísání vstupních veličin je možno posoudit prostřednictvím výrazu pro rozptyl funkce několika náhodných proměnných. Pro malý rozptyl,

$$s_y^2 = \left(\frac{\partial y}{\partial x_1}\right)^2 s_{x_1}^2 + \left(\frac{\partial y}{\partial x_2}\right)^2 s_{x_2}^2 + \dots + 2\left(\frac{\partial y}{\partial x_1}\right)\left(\frac{\partial y}{\partial x_2}\right) \text{cov}(x_1, x_2) + \dots \quad (13.14)$$

$s_{xi}$  je směrodatná odchylka  $x_i$ . Člen úplně vpravo je nenulový, jsou-li proměnné korelovány.

Odezvová plocha, zahrnující všechny faktory, může být přibližně vyjádřena jako polynom:

$$y = y_0 + \sum a_i (x_i - x_{i,0}) + \sum b_i (x_i - x_{i,0})^2 + \sum c_i (x_i - x_{i,0})(x_j - x_{j,0}) + \dots \quad (13.15)$$

Součet je dělaný pro všechny nezávislé proměnné. Konstanty se získají proložením hodnot kolem návrhového bodu regresní funkcí. Pro nekorelované proměnné a lineární aproximaci  $y$  podle vztahu (13.6) bude mít rovnice (13.14) tvar

$$s_y^2 = a_1^2 s_{x_1}^2 + a_2^2 s_{x_2}^2 + \dots + a_n^2 s_{x_n}^2 + \dots \quad (13.16)$$

Jednotlivé složky  $s_{yi}^2 = a_i^2 s_{xi}^2$  udávají rozptyly  $y$ , způsobené náhodným rozptylem  $i$ -té proměnné. Příspěvek  $s_{xi}^2$  k celkovému rozptylu  $s_y^2$  je větší pro větší rozptyl

proměnné  $x_i$  a pro větší citlivost ( $a_i$ ) výstupu  $y$  na změny  $x_i$ . Jestliže rovnici (13.16) podělíme  $s_y^2$ , dostaneme relativní podíly jednotlivých faktorů na celkovém rozptylu,

$$1 = a_1^2 \frac{s_{x1}^2}{s_y^2} + a_2^2 \frac{s_{x2}^2}{s_y^2} + \dots + a_n^2 \frac{s_{xn}^2}{s_y^2} + \dots \quad (13.17)$$

Vliv rozptylu jednotlivých vstupních veličin je možno posoudit prostřednictvím poměru variačního koeficientu  $i$ -té proměnné a variačního koeficientu výstupu, odpovídajícího rozptylu pouze této proměnné,

$$\omega_i = \frac{v_y}{v_{xi}} = \frac{s_y}{y_0} \bigg/ \frac{s_{xi}}{x_{i,0}} \quad (13.18)$$

### Analýza citlivosti pomocí simulačních metod

Vliv náhodné proměnlivosti vstupních veličin lze posoudit dokonce bez analytického výrazu pro odezovou funkci – prostřednictvím pravděpodobnostní simulační techniky Monte Carlo, která bude popsána v následující kapitole. V tomto případě citlivostní analýza spočívá v provedení  $m$  simulačních pokusů Monte Carlo, kdy jedinou náhodnou proměnnou je  $x_i$ , a v následujícím výpočtu dílčího rozptylu  $s_{yi}^2$  získaných hodnot  $y$ . Potom, s použitím charakteristik  $s_{xi}$ ,  $x_{i,0}$ ,  $y_0$ , a rovnic (13.16) a (13.18), lze vypočítat poměry variačních koeficientů  $v_i$  nebo citlivostní koeficienty  $a_i (= s_y/s_{xi})$  a koeficienty relativní citlivosti [2].

Přibližná hodnota celkového rozptylu se získá sečtením dílčích rozptylů,

$$s_y^2 = s_{y1}^2 + s_{y2}^2 + \dots + s_{yn}^2 + \dots \quad (13.19)$$

Přesnější hodnotu dostaneme, jestliže všechny vstupní veličiny ( $x_1$ ,  $x_2$ , ...  $x_n$ ) budeme uvažovat současně jako náhodné při simulacích Monte Carlo. Vydělíme-li rovnici (13.19) celkovým rozptylem  $s_y^2$ , dostaneme relativní vliv jednotlivých faktorů, podobně jako v rovnici (13.17).

Příklady použití analýzy nejistot při předpovědi životnosti je možno najít např. v [4, 5] a v kapitole 19 knihy [1].

**Poděkování.** Části této kapitoly byly již dříve publikovány v kapitole 19 knihy [1].

### Literatura ke kapitole 13

1. Menčík, J.: Concise reliability for engineers. InTech, Rijeka, 2016, ISBN 978-953-51-2278-4. *Publikace Open Access, dostupná na:* <http://www.intechopen.com/books/concise-reliability-for-engineers>. 204 s.
2. Novák, D., Teplý, B. and Shiraishi, N. Sensitivity Analysis of Structures: A Review. *Int. Conf. CIVIL COMP'93, August 1993*: p. 201-207. Scotland, Edinburgh.
3. Menčík, J.: Reliability-based parameter optimisation and tolerancing in structural design. *8th Int. Conf. ICOSSAR 2001*. Newport Beach, USA. *Structural Safety and Reliability* (R.B. Corotis, G.I. Schuëller and M. Shinozuka, eds), A.A. Balkema Publishers, Lisse, 188 s.
4. Jacobs, D. F., Ritter, and J. E. jr.: Uncertainty in minimum lifetime predictions. *J. Am. Ceram. Soc.*, 59 (1976) No. 11/12, str. 481 – 487.
5. Wiederhorn, S. M., Fuller, E. R. jr., Mandel, J., and Evans, A. G.: An error analysis of failure prediction techniques derived from fracture mechanics. *J. Am. Ceram. Soc.*, 59 (1976) No. 9/10, str. 403 – 411.

## 14. Simulační metody pro zkoumání náhodných veličin a vlivů

V dnešní době se velká část výzkumu provádí počítačovými simulacemi. Pravděpodobnostní simulační metody lze používat ke zkoumání vlivu náhodného kolísání různých veličin na vlastnosti strojů nebo syntetických látek, na únosnost konstrukcí, na fyzikální, chemické nebo biologické procesy, na spolehlivost, i v mnoha dalších případech. Velmi účinným nástrojem ke zkoumání náhodných jevů nebo procesů je technika Monte Carlo. Někdy se vyhovující výsledky získají s menším úsilím při použití metody LHS (Latin Hypercube Sampling).

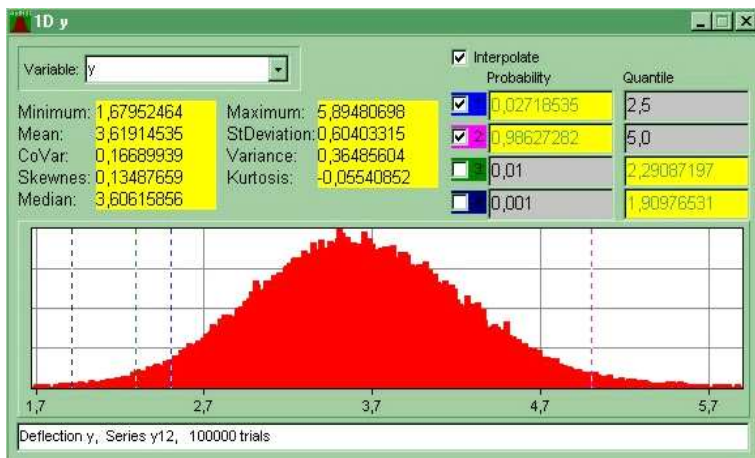
### Simulační metoda Monte Carlo

Metoda Monte Carlo je jednoduchá počítačová technika, založená na provádění četných fiktivních experimentů s použitím náhodných čísel [1 - 3]. Její použití je univerzální a nevyžaduje speciální znalosti teorie pravděpodobnosti. Jediná informace, kterou uživatel potřebuje, je vztah mezi vstupními veličinami a výstupem,

$$y = f(x), \text{ nebo } y = f(x_1, x_2, x_3, \dots) \quad (14.1)$$

a znalost rozdělení pravděpodobnosti vstupních veličin. Metoda MC opakuje pokusy s náhodnými čísly generovanými počítačem. V každém „pokusu“ se vstupním veličinám  $x_1, x_2, \dots, x_n$  přiřadí náhodná čísla, ale taková, že jejich rozdělení vždy odpovídá rozdělení pravděpodobnosti příslušné proměnné. S těmito čísly se pak provedou operace odpovídající analyzovanému systému a vypočítá se výstupní veličina  $y$  podle rovnice (14.1). Z výsledků lze vytvořit histogram, který odpovídá rozdělení  $y$ ; příklad je uveden na Obr. 14.1.

Vygenerované hodnoty lze užít pro výpočet průměrné hodnoty nebo pravděpodobnosti, že  $y$  bude nižší nebo vyšší než určitá zvolená hodnota  $y^*$ , anebo pro určení hodnot, které budou překročeny (nebo nedosaženy) jenom s určitou pravděpodobností (např. únosnost, doba do poruchy nebo maximální očekávatelné deformace nebo zatížení).



**Obr. 14.1.** Histogram vytvořený metodou Monte Carlo programem Ant-Hill [4, 5].

Existují různé komerční počítačové programy pro simulace Monte Carlo, např. [5 – 7], ale je možno si též vytvořit vlastní. Základem těchto programů je **generátor náhodných čísel**. Ve skutečnosti tato čísla nejsou zcela náhodná, protože se tvoří podle určitého (deterministického) algoritmu. Princip těchto generátorů je jednoduchý. Například tzv. kongruenční generátor vytváří náhodná čísla s rovnoměrným rozdělením v intervalu (0; 1) následujícím způsobem. Je zvoleno jedno číslo jako základ řady náhodných čísel  $u$  (například  $u_0 = 0.5284163$ ). Nyní, v prvním kroku, se toto číslo vynásobí jistým vhodným číslem  $Q$ , např. 997. Součin je  $997 \times 0.5284163 = 526.8310511$ . Prvé náhodné číslo  $u_1$  je pak vytvořeno jako část výsledku ležící za desetinnou čárkou; v našem případě  $u_1 = 0.8310511$ . Ve druhém kroku je  $u_1$  opět vynásobeno stejným číslem  $Q$ ,  $997 \times 0.8310511 = 828.5579467$ , a druhé náhodné číslo je vytvořeno jako desetinná část výsledku (tj.  $u_2 = 0.5579467$ ). Doporučujeme, aby čtenář vykonal několik kroků; pro kontrolu,  $u_3 = 0.2728599$ . Dlouhá řada těchto čísel má přibližně rovnoměrné rozdělení. Existuje mnoho dalších algoritmů. Například algoritmus pro normální rozdělení vychází z centrální limitní věty. Generátory náhodných čísel jsou také součástí univerzálních programů, jako je Matlab. Lze silně doporučit použití komerčních generátorů, neboť byly prověřeny podrobnými statistickými testy, že se chovají téměř jako skutečně náhodné. Také Excel má svůj generátor, avšak s omezenými možnostmi.

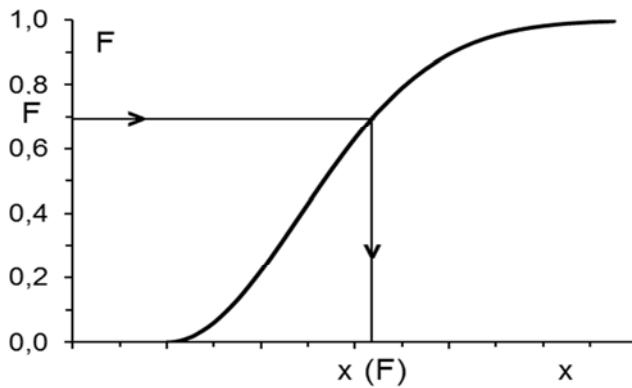


### Vytváření náhodných čísel se specifickými rozděleními

Komerční programy nabízejí často užívaná rozdělení, například rovnoměrné nebo normální. Náhodná čísla s jiným rozdělením, definovaným analyticky, lze vytvářet prostřednictvím rovnoměrného rozdělení. Základní myšlenkou je, že distribuční funkce  $F$  jakékoliv spojitě náhodné proměnné je také náhodná veličina, rozložená rovnoměrně v intervalu  $(0; 1)$ . Jestliže tedy distribuční funkce náhodné veličiny  $x$  je  $z = F(x)$ , potom náhodná čísla  $x$  lze získat z náhodných čísel  $z$  s rovnoměrným rozdělením v intervalu  $(0; 1)$  použitím inverzní transformace (Obr. 14.2):

$$x = F^{-1}(z) \quad (14.2)$$

Například distribuční funkce exponenciálního rozdělení je  $z = F(x) = 1 - \exp(-x/x_0)$



**Obr. 14.2.** Generování náhodných čísel  $x$  pomocí inverzní pravděpodobnostní transformace (14.2); podle [4].

s parametrem  $x_0$ . Inverzní transformace pro toto rozdělení je  $x = -x_0 \ln(1 - z)$ .

V některých případech má rozdělení náhodné veličiny složitý tvar, který lze popsat pouze histogramem, který se získá měřením.

Tento histogram se potom užije při konstrukci distribuční funkce  $F(x)$ . Tato funkce může být aproximována buď konstantními hodnotami  $F$  v jednotlivých sub-intervalech pro  $x$ , nebo interpolací uvnitř každé třídy,

$$F(x) = F_i + \frac{F_{i+1} - F_i}{x_{i+1} - x_i} (x - x_i), \quad x = x_i + \frac{F(x) - F_i}{F_{i+1} - F_i} (x_{i+1} - x_i) \quad (14.3)$$

$i = 1, 2, \dots, n$  označuje interval. Vzorec vpravo dává  $x$  odpovídající pravděpodobnosti  $F$ . Hodnoty  $F$  se generují jako náhodná čísla s rovnoměrným rozdělením.

Typickým rysem metody Monte Carlo je, že charakteristické hodnoty (průměr, kvantily, pravděpodobnosti odpovídající určitým hodnotám  $y$ , atd.), získané jako výsledek  $n$  simulačních pokusů, nejsou ve dvou sériích simulací nikdy stejné. Výsledky jsou tedy jen přibližné. Jsou ale blíže ke skutečným hodnotám při vyšším počtu pokusů. Počet těchto pokusů  $n$ , potřebný pro dosažení určité přesnosti výsledků, je přibližně dán vztahem

$$n = u_{\alpha/2}^2 (1 - P) / (P\delta^2) \quad (14.4)$$

$P$  je očekávaná (odhadnutá) pravděpodobnost vyšetřovaného jevu,  $\delta$  je přípustná relativní chyba při určení  $P$ , a  $u_{\alpha/2}$  je  $\alpha/2$ -kritická hodnota normovaného normálního rozdělení pro pravděpodobnost  $\alpha$ , že skutečná hodnota  $P$  bude vně intervalu  $P \pm \delta$ . Nezbytný počet simulací výrazně narůstá s klesající pravděpodobnostmi. Je-li například předpokládaná pravděpodobnost  $P = 0,01$ , dovolená relativní chyba  $\delta = 10\%$  a hladina spolehlivosti  $\alpha = 5\%$  (s hodnotou  $u_{\alpha/2} = 1,96 \approx 2$ ), bude zapotřebí  $\approx 40\,000$  simulačních pokusů. Pro  $P = 0,0001$  to bude již  $4\,000\,000$  pokusů atd. [Poznámka: Rovnice (14.4) vychází z toho, že počet příznivých výskytů určitého jevu o pravděpodobnosti  $P$  při  $n$  opakováních má binomické rozdělení, a toto rozdělení lze pro vysoké počty  $n$  aproximovat normálním rozdělením.]

### Složitější případy, metoda odezvové plochy

Přímé použití metody Monte Carlo je vhodné při jednoduchých vztazích  $y = f(x_1, x_2, \dots)$ . Často ale musí být odezva  $y$  nalezena numerickým řešením. Jestliže ale jeden takovýto pokus trvá minuty nebo více, tisíce simulačních pokusů by trvaly příliš dlouho. V takových případech je efektivnější kombinace techniky MC s metodou odezvové plochy, popsanou v kapitole 13. Podstata spočívá v tom, že „přesná“ odezva se vypočítá pouze pro vybrané hodnoty vstupních proměnných, výsledky se proloží jednoduchou regresní funkcí (odezvou plochou, Obr. 13.1 v kapitole 13), a simulace Monte Carlo se dělají s touto funkcí.

Vztah mezi výstupní veličinou  $y$  (deformace, nosnost konstrukce, amplituda vibrací) a vstupními proměnnými lze často aproximovat polynomem:

$$y = a_0 + \sum a_i x_i + \sum b_i x_i^2 + \dots + \sum c_{ij} x_i x_j + \dots \quad (14.5)$$

Tato aproximace je možná, jestliže vztah mezi vstupem a výstupem má podobný charakter (např.  $y \sim x^3$ ) nebo jestliže se výstupní veličina mění ve vyšetřovaném intervalu jenom málo. Pokud se liší od polynomu podstatně (např.  $y \sim 1/x^3$  nebo  $y \sim x^{1/2}$ ), rovnice (14.5) nemůže dát dobrou aproximaci v širším intervalu. Existuje několik cest ke zlepšení. Lineární nebo polynomickou funkci lze použít k aproximaci jiných výrazů, pokud ji vhodným způsobem transformujeme. Například vztah  $y = a/x^3$  lze vyjádřit jako  $y = az$ , jestliže zavedeme novou proměnnou  $z = 1/x^3$ ; vztah  $y = ax_1/x_2^2$  je možno použitím logaritmické transformace změnit na vícenásobnou lineární regresi  $Y = A_0 + A_1 X_1 + A_2 X_2$  atd. Někdy pomůže rozdělení definičního intervalu vstupních veličin do několika dílčích intervalů a použití různých regresních funkcí v jednotlivých subintervalech.

Jakost aproximace lze posoudit prostřednictvím residuální směrodatné odchylky  $s_{\text{res}}$ . Také rozdíly mezi „přesnými“ hodnotami a odpovídajícími hodnotami na odezvové ploše mohou sloužit jako kritérium. U dobré odezvové plochy jsou tyto rozdíly náhodně kladné a záporné. (Viz také pojem rezidua a Obr. 7.2 v kapitole 7.)

### **Použití metody Monte Carlo pro korelované veličiny**

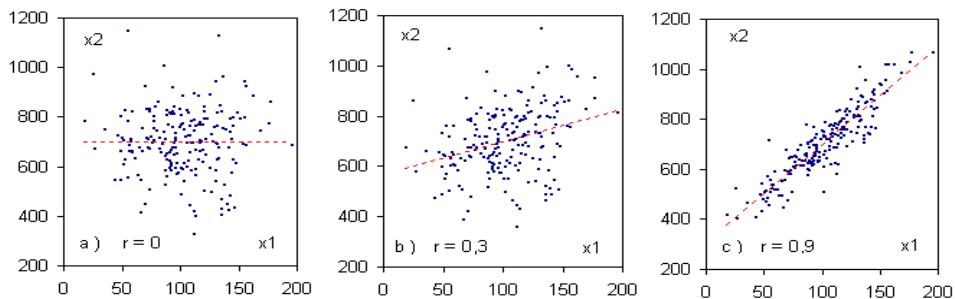
Použití techniky Monte Carlo v úlohách s více proměnnými je jednoduché, jestliže jednotlivé vstupní veličiny jsou vzájemně nezávislé (např. materiálové vlastnosti a geometrie součástí). Někdy ale mezi nimi existuje korelace (například mezi hustotou betonu a jeho modulem pružnosti). Zvláštním případem je autokorelace, kdy hodnota náhodné veličiny v určitém bodě souvisí částečně s hodnotami v okolí nebo v předcházejícím čase. Příkladem jsou vlastnosti základové půdy nebo teploty stavební konstrukce; ta se mění den ode dne, ale částečně závisí na roční sezóně.

Opomenutí korelací může vést k chybám. Například velmi nízká hodnota modulu pružnosti betonu by mohla být generována současně s velmi vysokou hodnotou pevnosti, což ale neodpovídá skutečnosti. Jestliže se korelace zohlední, odrážejí výpočty realitu lépe a předpovědi jsou přesnější. Někdy také není veličina potřebná pro analýzu k dispozici, ale může být nahrazena korelovanou veličinou. Například není-li možné provádět tahové zkoušky existující masivní ocelové konstrukce, lze někdy užít zkoušky tvrdosti.

Těsnost vztahu dvou veličin je charakterizována korelačním koeficientem  $r$ , definovaným

$$r = \text{cov}(xy) / (s_x s_y) \quad (14.6)$$

$\text{cov}(xy)$  je kovariance  $x$  a  $y$ , a  $s_x$  a  $s_y$  jsou směrodatné odchylky. Korelační koeficient  $r$  nabývá hodnot od  $-1$  do  $+1$ . Pro  $r = 0$  neexistuje vzájemný vztah, zatímco  $r = +1$  nebo  $-1$  odpovídá deterministickému (funkčnímu) vztahu. Pro  $r > 0$  rostou hodnoty  $x$  s rostoucím  $y$ , a klesají při  $r < 0$ . (POZNÁMKA: Korelační koeficient je roven druhé odmocnině koeficientu determinace  $r^2$ , vysvětleného v kapitole 6.) Na obr. 14.3 jsou ukázány tři příklady korelovaných veličin se stejnými středními hodnotami a směrodatnými odchylkami, ale s rozdílnými hodnotami korelačního koeficientu  $r$ .



**Obr. 14.3.** Veličiny  $x_1$  a  $x_2$  se stejnými středními hodnotami a směrodatnými odchylkami ( $\mu_1 = 100$ ,  $\mu_2 = 700$  a  $\sigma_1 = 30$ ,  $\sigma_2 = 150$ ) a různými korelačními koeficienty  $r$  [8, 9].

Je-li zapotřebí generovat korelované veličiny  $x_1$  a  $x_2$  a regresní funkce  $x_{2,\text{reg}} = f(x_1)$  je známa, jakož i koeficient determinace  $r^2$  této aproximace, lze užít následující postup [8, 9]. Nejprve se vygeneruje náhodná hodnota  $x_1$ . Potom se vygeneruje odpovídající hodnota  $x_2$  jako

$$x_2 = f(x_1) + \Delta x_2 = f(x_1) + u s_{2,\text{res}} = f(x_1) + u s_2 (1 - r^2) \quad (14.7)$$

$s_{2,\text{res}}$  je výběrová zbytková směrodatná odchylka  $x_2$  kolem regresní funkce  $f$ , a  $u$  je náhodný kvantil standardního normálního rozdělení (za předpokladu, že rozdělení jednotlivých hodnot  $x_2$  kolem  $f$  je normální). Výraz v pravé části rovnice (14.7) využívá skutečnost, že zbytkovou odchylku  $s_{2,\text{res}}$  proměnné  $x_2$  lze vyjádřit

prostřednictvím směrodatné odchylky  $s_2$  a koeficientu determinace  $r^2$ , příslušejícího regresní funkci  $x_2 = f(x_1)$ .

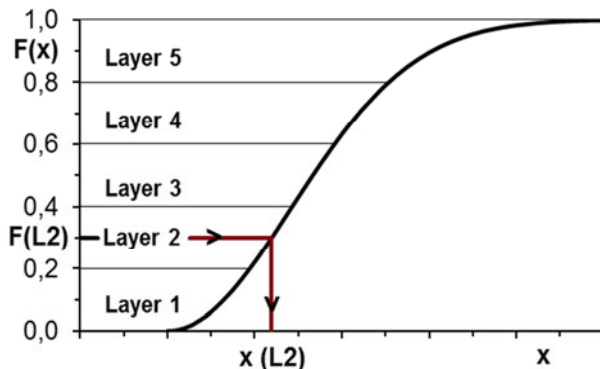
Pro více informací o metodě Monte Carlo je možno doporučit knihy [1, 2]. Mnoho příkladů praktických aplikací lze najít v [3]. Existují i různé komerční programy Monte Carlo pro inženýrské aplikace, např. [5 – 7]; některé další jsou zmíněny v kapitole 26 knihy [4].

### **Metoda LHS (Latin Hypercube Sampling)**

Metoda Monte Carlo má dva nedostatky. Zaprvé, obvykle potřebuje velmi vysoký počet simulací. Jestliže hodnoty vyšetřované veličiny se získávají časově náročnými numerickými výpočty, mohou simulace trvat velice dlouho. Také není

vždycky možné použít metodu odezvové plochy. Za druhé, může se stát, že náhodná čísla distribuční funkce  $F$  s rovnoměrným rozdělením (která slouží pro vytváření náhodných čísel s nestandardními rozděleními) nejsou v definičním intervalu  $(0; 1)$  rozdělena dostatečně rovnoměrně. Někdy jsou generována častěji v jedné části intervalu než druhé, a generovaná veličina má tak poněkud odlišné rozdělení, než bylo požadováno. Tento problém se může objevit zvláště, závisí-li výstupní veličina na mnoha proměnných.

Metoda, zvaná LHS (Latin Hypercube Sampling), tento nedostatek odstraňuje [10, 11]. Základní myšlenka LHS je podobná jako u generování náhodných čísel s nestandardním rozdělením prostřednictvím inverzní pravděpodobnostní transformace (14.2), ukázané na Obr. 14.2. Rozdíl je v tom, že LHS vytváří hodnoty  $F$  nikoliv generováním náhodných čísel rozptýlených chaoticky v intervalu  $(0; 1)$ , ale tak, že jim přiřazuje určité pevné hodnoty. Interval  $(0; 1)$  je rozdělen do několika vrstev stejné šířky, a hodnoty  $x$  se počítají inverzní transformací ( $F^{-1}$ ) z hodnot  $F$  odpovídajících středu každé vrstvy (Obr. 14.4). S dostatečně vysokým počtem vrstev (desítky až stovky) bude generovaná veličina  $x$  přibližně mít požadované rozdělení pravděpodobnosti. Tento přístup se označuje jako Stratified Sampling. Závisí-li výstupní veličina  $y$  na několika vstupních proměnných,  $x_1, x_2, \dots, x_n$ , je nutné, aby každé proměnné byly přiřazeny hodnoty odpovídající všem vrstvám, a aby veličiny a vrstvy jednotlivých proměnných byly náhodně zkombinovány. To se docílí tak, že se pořadová čísla vrstev přiřazují jednotlivým vstupním veličinám náhodně.



**Obr. 14.4.** Metoda LHS – princip [4, 10].

Postup je následující. Definiční interval distribuční funkce  $F$  každé z  $m$  proměnných se rozdělí na  $N$  vrstev.  $N$ , stejné pro všechny proměnné, také odpovídá počtu „pokusů“, tj. simulačních výpočtů. V každém pokusu se pořadová čísla vrstev náhodně přiřadí jednotlivým proměnným ( $X_1, X_2, \dots, X_m$ ). Tímto způsobem jsou vždy náhodně zkombinovány různé vrstvy jednotlivých proměnných. V praxi se toho docílí pomocí náhodných čísel a jejich seřazováním. Potom se každé vstupní proměnné přiřadí hodnota, odpovídající středu příslušné vrstvy její distribuční funkce.

Aplikace je ilustrována na případu se čtyřmi náhodnými veličinami ( $X_1, X_2, X_3, X_4$ ) a definičním intervalem pro  $F$  rozděleným do 5 vrstev (Obr. 14.4, 14.5). Pět vrstev je zde užito pro jednoduchost; obvykle se používá několik desítek vrstev. V našem případě bude pět pokusů, tj.  $Y$  bude počítáno pro pět kombinací čtyř vstupních veličin. To znamená, že se vygeneruje  $5 \times 4 = 20$  náhodných čísel s rovnoměrným rozdělením v intervalu  $(0; 1)$ ; viz tabulku v levé části Obr. 14.5. Potom se přiřadí čísla vrstev pro (např.) proměnnou  $X_1$  v jednotlivých pokusech s ohledem na pořadí náhodných hodnot pro  $X_1$  seřazených podle velikosti od největší do nejmenší. V našem případě vrstva č. 3 (s číslem 0,382) pro první pokus, vrstva č. 1 (s číslem 0,885) pro druhý pokus, vrstva č. 3 (s 0,863) pro třetí pokus, atd., v souladu s čísly 0,382 – 0,885 – 0,863 – 0,032 – 0,285 ve sloupci pro  $X_1$ . Podobné operace se provedou pro každou proměnnou. To znamená, že v prvním pokusu se proměnným  $X_1, X_2, X_3$  a  $X_4$  přiřadí hodnoty odpovídající 3., 5., 1. a 1. vrstvě jejich distribučních

## Náhodná čísla (RN)

Proměnná ----- Vrstva/pokus	$X_1$ <i>RN</i>	$X_2$ <i>RN</i>	$X_3$ <i>RN</i>	$X_4$ <i>RN</i>
1	0.382	0.101	0.596	0.899
2	0.885	0.958	0.014	0.407
3	0.863	0.139	0.245	0.045
4	0.032	0.164	0.220	0.017
5	0.285	0.343	0.554	0.357

## Čísla vrstev pro jednotlivé vrstvy (LN)

Proměnná ----- Vrstva/pokus	$X_1$ <i>LN</i>	$X_2$ <i>LN</i>	$X_3$ <i>LN</i>	$X_4$ <i>LN</i>
1	3	5	1	1
2	1	1	5	2
3	2	4	3	4
4	5	3	4	5
5	4	2	2	3

**Obr. 14.5.** Metoda LHS – přiřazování vrstev jednotlivým proměnným a pokusům.

funkcí atd. Potom se užije inverzní pravděpodobnostní transformace  $F^{-1}$  pro stanovení  $X_1$  z  $F_{1,1}$ , atd.; viz tabulku vpravo. Nyní se pětkrát vypočítá vyšetřovaná veličina  $Y = Y(X_1, X_2, X_3, X_4)$ . Získané hodnoty  $Y_1, Y_2, Y_3, Y_4, Y_5$  pak lze užít pro určení statistických charakteristik (průměrná hodnota, směrodatná odchylka...).

Obvykle se provede několik desítek nebo set pokusů, které umožní vytvoření distribuční funkce  $F(Y)$  a stanovení průměru, směrodatné odchylky, různých kvantilů a dalších charakteristik.

Více o metodě LHS lze najít v [10, 11].

**Poděkování.** Části této kapitoly byly již dříve publikovány v kapitolách 15 a 16 knihy [4].

### Literatura ke kapitole 14

1. Hammersley, J. M., Handscomb, D.: Monte Carlo Methods. John Wiley, New York, 1964. 184 s.
2. Marek, P., Guštar, M., Anagnos, T. Simulation-based reliability assessment for structural engineers. CRC Press, Boca Raton, 1996. 384 s.
3. Marek, P., Brozetti, J., Guštar, M., Tikalsky, P., editors. Probabilistic Assessment of Structures using Monte Carlo Simulation. ITAM CAS CR, Prague, 2003. ISBN 80-86246-19-1. 471 s.

4. Menčík, J.: Concise reliability for engineers. InTech, Rijeka, 2016, *An Open Access publication*,: <http://www.intechopen.com/books/concise-reliability-for-engineers>, ISBN 978-953-51-2278-4. 204 s.
5. AntHill program for Monte Carlo simulations: <http://www.sbra-anthill.com>
6. Novák, D., Rusina, R., Vořechovský, M.: FReET. A multipurpose probabilistic software for analysis of Engineering problems. <http://www.freet.cz>. 2015.
7. Petschacher, M.: VaP (Variables Processor). <http://www.petschacher.at>. 2016.
8. Menčík, J.: Simulační posuzování spolehlivosti při korelovaných veličinách. In: Spolehlivost konstrukcí, 23. - 24. 4. 2003, Dům techniky Ostrava, s. 151 – 156.
9. Čačko, J., Bílý, M., Bukovec, J.: Meranie, vyhodnocovanie a simulácia náhodných procesov. Bratislava, VEDA, 1984. 210 s.
10. Florian, A.: An efficient sampling scheme: Updated Latin Hypercube Sampling. Probabilistic Engineering Mechanics, 7 (1992), issue 2, s. 123 – 130.
11. Olsson, A., Sandberg, G., Dahlblom, G.: On Latin hypercube sampling for structural reliability analysis. Probabilistic Engineering Mechanics. 2002; 25; issue 1, s. 47 – 68.



## Rejstřík (číslo v závorce značí kapitolu)

algoritmy genetické	119 (12)
analýza citlivosti	125 (13)
analýza rozptylu	97 (10)
analýza	9, 10 (2)
ANOVA	97 (10)
autokorelace	48, 49 (6)
Bayesovské metody	77 (8)
bezrozměrné proměnné	84, 92 (9)
bezrozměrový parametr	87 (9)
bezrozměrový tvar	82 (9)
blok	104 (11)
blok znáhodněný	105 (11)
Buckinghamovo pravidlo	83 (9)
cenzenovaná data	40 (5)
citlivost odezvy	122 (13)
citlivost relativní	122 (13)
citlivostní analýza	125 (13)
četnost, č. relativní, kumulativní	22 (4)
data-mining	49 (6)
distribuční funkce	22, 23 (4), 129 (13)
DOE (Design of Experiments)	105 (11)
experiment	7 (1)
faktor	97 (10), 103 (11)
F-rozdělení	28 (4)
funkce odezvy	121 (13)
funkce pravděpodobnostní	21 (4)
Gaussovo rozdělení	25 (4)
gradient	114 (12)
graf kvantil-kvantilový (Q–Q)	40 (5)
histogram	22 (4), 33 (5)
hustota pravděpodobnosti	22, 23 (4)
charakteristické hodnoty	33 (5)
charakteristika	20 (4)

chí-kvadrát ( $\chi^2$ ) test	75 (8)
chyby hrubé	15 (3)
chyby náhodné	17 (3)
chyby systematické	15 (3)
interakce	108 (11)
interval spolehlivosti	29 (4)
koeficient determinace $r^2$	45 (6), 55 (7)
koeficient korelace	43, 44 (6)
Kolmogorovův-Smirnovův test	74 (8)
konfidenční interval	29 (4), 65, 67 (8)
konfidenční meze, exponenciální rozdělení	68 (8)
konfidenční pás	58 (7)
korelace	43, 47 (6)
korelované veličiny	131 (14)
kovariance	43 (6)
kritická hodnota	24 (4)
kvantil	24, 32 (4)
latinské čtverce	104, 105 (11)
maximum, minimum, hledání	113 (12)
medián	21 (4)
měření	9 (2)
metoda LHS (Latin Hypercube Sampling)	132 (14)
metoda Monte Carlo	127 (14)
metoda odezvové plochy	130 (14)
metoda nejmenších čtverců	38 (5), 54, 56 (7)
metoda největšího spádu	114 (12)
metody neparametrické	31 (4)
metody simulační	125 (13)
model	12 (2)
model empirický	13 (2)
model matematický	13 (2)
model teoretický	13 (2)
modelování počítačové	13 (2)
Monte Carlo (metoda)	127 (14)
náhodná čísla, generování	128 (14)
náhodná veličina	19 (4)

náhodné kolísání	124 (13)
odezvozá funkce	121 (13)
odezvozá plocha	122 (13), 130 (14)
odchylky deterministické	122 (13)
odchylky náhodné	124 (13)
parametr	20 (4), 35 (5), 87 (9)
parametr bezrozměrový	87 (9)
plánování pokusů	103, 105 (11)
počet pokusů Monte Carlo	129 (14)
podobnost	83 (9), 91 (9)
podobnostní čísla	84 (9)
pohyblivé průměry	63 (7)
pokus faktorový	105, 107, 110 (11)
pořadová statistika	30 (4)
pravděpodobnost	19 (4)
průměr	21 (4)
publikování	11 (2)
Q – Q graf (kvantil-kvantil)	40 (5)
regrese lineární, vícenásobná	61 (7)
regresní funkce	51 (7)
regresní konstanty, stanovení	52 (7)
regresní přímka	70 (8)
reziduum, relativní	55, 56 (7)
rozdělení binomické	22, 25 (4)
rozdělení binomické	22, 25 (4)
rozdělení exponenciální	28 (4)
rozdělení chí-kvadrát ( $\chi^2$ )	28 (4)
rozdělení logaritmicko-normální	26 (4)
rozdělení normální	25 (4)
rozdělení normované normální	26 (4)
rozdělení Poissonovo	25 (4)
rozdělení pravděpodobnosti	20 (4), 35 (5)
rozdělení rovnoměrné	28 (4), 128 (14)
rozdělení t (Studentovo)	28 (4)
rozdělení Weibullovo	25 - 27 (4), 36, 37 (5)
rozměr	81, 83 (9)

rozměrová analýza	81 (9)
rozptyl	21 (4), 67 (8)
rozptyl funkce více proměnných	124 (13)
Řešitel	57 (7)
simplex, simplexová metoda	115 (12)
simulace	12 (2)
simulované žíhání	119 (12)
směrodatná odchylka	21 (4)
soubor základní	20 (4)
Spearmanův koeficient pořadové korelace	44 (6)
střední hodnota	21 (4), 65 (8)
šikmost (koeficient asymetrie)	23 (4)
špičatost	24 (4)
teorie podobnosti	84 (9)
testy dobré shody	74 (8)
testování hypotéz	29 (4), 72 (8)
toleranční meze	71 (8)
transformace funkcí	54 (7)
transformace pravděpodobnostní	128, 129, 133 (14)
úroveň	106 (11)
variační koeficient	21 (4)
věta Bayesova	78 (8)
vícenásobná korelace	47 (6)
vrstvy	133, 134 (14)
výběr, náhodný	20 (4), 104 (11)
znáhodňování	103 (11)

Název: Úvod do experimentální analýzy  
Autor: prof. Ing. Jaroslav Menčík, CSc.  
Vydavatel: Univerzita Pardubice  
Do tisku: květen 2017  
Stran: 142  
Náklad: 20  
Vydání: první  
Tisk: Polygrafické středisko Univerzity Pardubice

Tato publikace neprošla jazykovou korekturou.

ISBN 978-80-7560-066-0 (tisk)

ISBN 978-80-7560-067-7 (pdf)

Kniha je volně přístupná na <http://hdl.handle.net/10195/66960>.

ISBN 978-80-7560-067-7 (pdf)