

MILAN MELOUN, JIŘÍ MILITKÝ

# **Statistické zpracování experimentálních dat**

**v chemometrii, biometrii, ekonometrii  
a v dalších oborech přírodních, technických  
a společenských věd**

1993



## Obsah

1	CHYBY INSTRUMENTÁLNÍCH MĚŘENÍ .....	1
1.1	Klasifikace chyb měření .....	1
1.2	Charakteristika přesnosti přístrojů .....	3
1.2.1	Mezní hodnoty chyb a třída přesnosti přístroje .....	5
1.2.2	Zařazení přístroje do třídy přesnosti .....	6
1.2.3	Zaokrouhlování chyby výsledku .....	8
1.2.4	Složky chyby výsledku měření .....	9
1.3	Modely měření .....	10
1.4	Kvantilové odhady chyb .....	13
1.5	Sčítání kvantilových chyb .....	16
1.6	Momentové odhady chyb .....	18
1.6.1	Pravděpodobnostní interval chyby .....	18
1.6.2	Toleranční interval chyby .....	20
1.7	Chyba výsledků instrumentálních měření .....	21
1.7.1	Metoda Taylorova rozvoje .....	22
1.7.2	Metoda dvoubodové aproximace .....	27
1.7.3	Metoda simulací Monte Carlo .....	27
1.8	Souhrn: Postup určení chyby instrumentálních měření .....	30
1.9	Výpočetní programy .....	30
1.10	Ostatní řešené příklady .....	31
1.11	Úlohy k procvičování .....	40
1.12	Literatura .....	43
2	PRŮZKUMOVÁ ANALÝZA JEDNOROZMĚRNÝCH DAT .....	45
2.1	Metody průzkumové analýzy dat .....	46
2.1.1	Grafy identifikace statistických zvláštností dat .....	50
2.1.2	Konstrukce a identifikace rozdělení výběru .....	66
2.1.3	Identifikace rozdělení výběru pro diskrétní náhodné veličiny .....	81
2.1.4	Transformace dat .....	85
2.1.5	Zpětná transformace .....	90

2.2	Ověření předpokladů o datech	92
2.2.1	Určení minimální velikosti výběru	92
2.2.2	Ověření předpokladu nezávislosti prvků výběru	94
2.2.3	Ověření normality výběru	95
2.2.4	Ověření homogenity výběru	99
2.3	Souhrn: Postup průzkumové analýzy	101
2.4	Výpočetní programy	103
2.5	Ostatní řešené příklady	109
2.6	Úlohy k procvičování	119
2.7	Literatura	125
3	STATISTICKÁ ANALÝZA JEDNOROZMĚRNÝCH DAT	127
3.1	Bodové odhady parametrů polohy, rozptýlení a tvaru	128
3.1.1	Metoda maximální věrohodnosti	129
3.1.2	Výběrové charakteristiky	130
3.2	Intervalový odhad parametrů polohy a rozptýlení	138
3.2.1	Povaha intervalového odhadu	138
3.2.2	Konstrukce intervalových odhadů	139
3.3	Odhady parametrů vybraných rozdělení	142
3.3.1	Poissonovo rozdělení	142
3.3.2	Normální rozdělení	146
3.3.3	Laplaceovo rozdělení	149
3.3.4	Rovnoměrné rozdělení	151
3.3.5	Exponenciální rozdělení	153
3.3.5.1	Jednparametrové exponenciální rozdělení	153
3.3.5.2	Dvoupparametrové exponenciální rozdělení	154
3.3.6	Logaritmicko-normální rozdělení	157
3.3.6.1	Dvoupparametrové lognormální rozdělení	158
3.3.6.2	Tříparametrové lognormální rozdělení	162
3.4	Robustní odhady parametrů polohy a rozptýlení	164
3.4.1	Medián	165
3.4.2	Uřezaný průměr	167
3.4.3	Robustní M-odhady	171
3.4.4	Analýza malých výběrů	176
3.4.5	Neparametrické odhady rozptylů	178
3.5	Testování statistických hypotéz	181
3.5.1	Postup testování statistické hypotézy	182
3.5.2	Testy hypotéz o parametrech jednoho souboru	185
3.5.3	Testy hypotéz o parametrech dvou souborů	187
3.6	Souhrn: Postup vyhodnocení jednorozměrných výběrů	199
3.7	Výpočetní programy	201
3.8	Ostatní řešené příklady	203
3.9	Úlohy k procvičování	216

3.10 Literatura	219
4 STATISTICKÁ ANALÝZA VÍCEROZMĚRNÝCH DAT	221
4.1 Vícerozměrné náhodné veličiny	222
4.2 Charakteristiky vícerozměrných náhodných veličin	226
4.3 Odhady parametrů polohy, rozptýlení a tvaru	230
4.4 Zobrazení vícerozměrných dat	236
4.4.1 Zobecněné rozptylové grafy	237
4.4.2 Symbolové grafy	244
4.5 Ověření normality	251
4.5.1 Testy normality	251
4.5.2 Grafické ověření normality	253
4.6 Statistická analýza vektoru středních hodnot	255
4.7 Statistická analýza kovariančních matic	263
4.8 Souhrn: Postup při analýze vícerozměrných dat	267
4.9 Výpočetní programy	267
4.10 Ostatní řešené příklady	268
4.10 Úlohy k procvičování	273
4.11 Literatura	275
5 ANALÝZA ROZPTYLU (ANOVA)	277
5.1 Základní pojmy analýzy rozptylu	278
5.2 Jednofaktorová analýza rozptylu	280
5.2.1 Modely s pevnými efekty	281
5.2.1.1 Metodologie statistické analýzy	281
5.2.1.2 Technika vícenásobného porovnání	285
5.2.1.3 Lineární regresní model	286
5.2.1.4 Ověření normality chyb	288
5.2.1.5 Ověření konstantnosti rozptylu (homoskedasticity)	291
5.2.2 Modely s náhodnými efekty	291
5.3 Dvoufaktorová analýza rozptylu	294
5.3.1 Modely s pevnými efekty	296
5.3.1.1 Modely pro případ bez opakování měření	297
5.3.1.2 Vyvážené modely	303
5.3.1.3 Nevyvážené modely	309
5.3.2 Modely se smíšenými efekty	310
5.3.3 Modely s náhodnými efekty	311
5.4 Souhrn: Postup při analýze rozptylu	313
5.5 Výpočetní programy	314
5.6 Ostatní řešené příklady	316
5.7 Úlohy k procvičování	320
5.8 Literatura	326

6 LINEÁRNÍ REGRESNÍ MODELÝ	327
6.1 Formulace lineárního regresního modelu	327
6.2 Geometrie a předpoklady metody nejmenších čtverců	329
6.2.1 Geometrie metody nejmenších čtverců	329
6.2.2 Předpoklady metody nejmenších čtverců	338
6.3 Statistické vlastnosti metody nejmenších čtverců	340
6.3.1 Konstrukce intervalů spolehlivosti	350
6.3.2 Testování hypotéz	354
6.3.2.1 Test multikolinearity	358
6.3.2.2 Test významnosti absolutního členu	361
6.3.2.3 Testy složených hypotéz	365
6.3.2.4 Test shody dvou lineárních modelů	368
6.3.2.5 Testy vhodnosti lineárního modelu	372
6.3.3 Porovnání regresních přímek	380
6.3.3.1 Test homogenity úseků	381
6.3.3.2 Test homogenity směrnic	383
6.3.3.3 Test shody regresních přímek	384
6.4 Numerické problémy lineární regrese na počítači	387
6.4.1 Metoda ortogonálních funkcí	391
6.4.2 Metoda racionálních hodnot	394
6.5 Regresní diagnostika	399
6.5.1 Využití průzkumové analýzy dat	400
6.5.2 Posouzení kvality dat	401
6.5.2.1 Statistická analýza reziduí	402
6.5.2.2 Analýza prvků projekční (H) matice	409
6.5.2.3 Grafy identifikace vlivných bodů	411
6.5.2.4 Ostatní charakteristiky vlivných bodů	419
6.5.3 Posouzení kvality navrženého regresního modelu	428
6.5.3.1 Parciální regresní grafy	428
6.5.3.2 Parciální reziduální grafy	431
6.5.3.3 Znaménkový test vhodnosti modelu	435
6.5.4 Ověření předpokladů metody nejmenších čtverců	437
6.5.4.1 Heteroskedasticita (nekonstantnost rozptylu)	437
6.5.4.2 Autokorelace	439
6.5.4.3 Normalita chyb	440
6.6 Postupy při porušení předpokladů metody nejmenších čtverců	442
6.6.1 Omezení na parametry	442
6.6.2 Metoda zobecněných nejmenších čtverců (MZNČ)	447
6.6.2.1 Heteroskedasticita	450
6.6.2.2 Autokorelace	457
6.6.3 Multikolinearita	464
6.6.4 Všechny proměnné zatížené náhodnými chybami	471
6.6.5 Jiná rozdělení chyb	477

6.7 Kalibrace	493
6.7.1 Druhy kalibrace a kalibrační modely	494
6.7.2 Kalibrační přímka	497
6.7.3 Přesnost kalibrace	504
6.8 Souhrn: Postup při lineární regresní analýze	509
6.9 Výpočetní programy	511
6.10 Ostatní řešené příklady	514
6.11 Úlohy k procvičování	536
6.12 Literatura	547
<b>7 KORELACE</b>	<b>551</b>
7.1 Korelační modely	552
7.1.1 Korelační modely pro dvě náhodné veličiny	552
7.1.2 Korelační model pro více náhodných veličin	560
7.2 Korelační koeficienty	571
7.2.1 Párový korelační koeficient	571
7.2.2 Parciální korelační koeficient	578
7.2.3 Vícenásobný korelační koeficient	580
7.2.4 Pořadové korelace	583
7.3 Úlohy k procvičování	585
7.4 Literatura	588
<b>8 NELINEÁRNÍ REGRESNÍ MODEL Y</b>	<b>589</b>
8.1 Formulace nelineárního regresního modelu	592
8.2 Modely chyb měření	597
8.3 Formulace kritéria regrese	604
8.4 Geometrie nelineární regrese	611
8.5 Numerické postupy odhadování parametrů	618
8.5.1 Nederivační optimalizační postupy	620
8.5.1.1 Metody přímého hledání	620
8.5.1.2 Simplexové metody	622
8.5.1.3 Metody využívající náhodných čísel	629
8.5.1.4 Postupy speciálně pro metodu nejmenších čtverců	632
8.5.2 Derivační metody pro kritérium metody nejmenších čtverců	637
8.5.2.1 Gaussovy-Newtonovy metody	640
8.5.2.2 Metody Marquardtova typu	644
8.5.2.3 Postupy typu dog-leg	646
8.5.3 Komplikace procesu nelineární regrese	648
8.5.3.1 Neodhadnutelnost některých parametrů	648
8.5.3.2 Existence minima $U(\beta)$	650
8.5.3.3 Výskyt lokálních minim	651
8.5.3.4 Špatná podmíněnost parametrů v modelu	652
8.5.3.5 Malé rozmezí experimentálních dat	653

8.5.4 Testování spolehlivosti regresních algoritmů	656
8.6 Statistická analýza nelineární regrese	658
8.6.1 Nelinearita regresního modelu	660
8.6.1.1 Vychýlení odhadů parametrů	661
8.6.1.2 Asymetrie odhadů parametrů	665
8.6.2 Intervalové odhady parametrů	666
8.6.2.1 Oblasti spolehlivosti parametrů	666
8.6.2.2 Intervaly spolehlivosti parametrů	673
8.6.2.3 Intervaly spolehlivosti predikce	675
8.6.3 Testy hypotéz o odhadech parametrů	677
8.6.4 Těstnost proložení regresní křivky	679
8.6.4.1 Statistická analýza reziduí	680
8.6.4.2 Analýza vlivných bodů	682
8.7 Souhrn: Postup při testování navrženého modelu	685
8.8 Výpočetní programy	688
8.9 Ostatní řešené příklady	689
8.10 Úlohy k procvičování	698
8.11 Literatura	703
<b>9 INTERPOLACE A APROXIMACE</b>	<b>707</b>
9.1 Klasické interpolační postupy	708
9.1.1 Lagrangeova a Newtonova interpolační formule	710
9.1.2 Hermitovská interpolace	718
9.1.3 Racionální interpolace	719
9.2 Spline interpolace	722
9.2.1 Lokální Hermitovská interpolace	728
9.2.2 Kubické spline	735
9.3 Aproximace funkcí	744
9.4 Aproximace tabelárních závislostí	748
9.4.1 Polynomická aproximace	749
9.4.2 Úseková regrese	752
9.5 Numerické vyhlazování	760
9.5.1 Spline vyhlazování	760
9.5.2 Neparametrická regrese	772
9.5.3 Číslicová filtrace	775
9.6 Souhrn: Postup při interpolaci a aproximaci	787
9.7 Výpočetní programy	788
9.8 Ostatní řešené příklady	790
9.9 Úlohy k procvičování	791
9.10 Literatura	793



10 DERIVACE A INTEGRACE .....	795
10.1 Derivace .....	796
10.1.1 Analytická derivace .....	798
10.1.2 Numerické derivace .....	798
10.2 Integrace .....	802
10.2.1 Analytická integrace .....	802
10.2.2 Numerická integrace .....	802
10.3 Souhrn: Postup při numerické derivaci a integraci .....	810
10.4 Výpočetní programy .....	811
10.5 Ostatní řešené příklady .....	811
10.6 Úlohy k procvičování .....	813
10.7 Literatura .....	814
DODATEK	
SOFTWARE - Statistický systém ADSTAT .....	815
Rejstřík .....	825



## Předmluva autorů

Analýza a interpretace experimentálních dat v klasických i moderních technických, behavioristických a gnoseologických oborech patří mezi neustále se rozvíjející směry zkoumání, ležící na pomezí matematické statistiky a informatiky. V řadě oborů se vydělil samostatný směr zkoumání založený na analýze dat. Příkladem jsou chemometrie, biometrie, psychometrie, ekonometrie a řada dalších. Statistická analýza dat nabývá stále na větším významu a stává se často jedním ze základních přístupů v řadě přírodovědných, technických a sociálních věd.

### Zaměření knihy

Stále větší počet výkonných osobních počítačů třídy PC podporuje na pracovištích trend decentralizace a interaktivnosti při zpracování experimentálních dat a interpretaci výsledků. To klade větší nároky na pracovníky, kteří již těžko obhájí jednoduché postupy vyhodnocování dat, založené mnohdy na zjednodušených nebo i nesprávných předpokladech. Nabídka a možnosti počítačově orientovaného statistického zpracování dat nutí experimentátora k hlubší analýze, což vede většinou i k radikální změně pohledu na rutinně prováděnou výzkumnou práci. Existuje celé spektrum méně či více dokonalých a komplexních programů, programových systémů pro statistické vyhodnocování dat. Jiné jsou budovány jako univerzálně použitelné, i když zaměřené na specifické oblasti (chemometrie, biometrie, ekonometrie, medicínská statistika, obchodní statistika, statistika pro sociology, psychology, atd.).

Úlohy vyhodnocení experimentálních dat v technické praxi se vyznačují společnými vlastnostmi:

- (a) rozsahy zpracovávaných dat nejsou obvykle velké,
- (b) v datech se vyskytují výrazné nelinearity, neaditivita a vzájemné vazby, které je třeba identifikovat a popsat,
- (c) rozdělení dat jen zřídka odpovídá normálnímu běžně předpokládanému ve standardní statistické analýze,
- (d) v datech se vyskytují vybočující měření a různé heterogenity,

(e) statistické modely se často tvoří na základě předběžných informací z dat (datově orientované přístupy),

(f) parametry statistických modelů mají mnohdy definovaný fyzikální význam, a musí proto vyhovovat velikostí, znaménkem nebo vzájemným poměrem,

(g) existuje jistá neurčitost při výběru modelu popisujícího chování dat.

Z hlediska použití statistických metod je proto žádoucí mít možnost zkoumat statistické zvláštnosti dat (průzkumová analýza), ověřovat základní předpoklady o datech, hodnotit kvalitu výsledků s ohledem na základní schéma "data - model - statistická metoda", a konečně využívat i alternativních postupů statistické analýzy včetně robustních a adaptivních metod. Z uvedeného je patrné, že oblast zpracování experimentálních dat má řadu požadavků, které nelze v plné míře zajistit bez obecně koncipovaných statistických programů. V této knize se chceme zabývat metodami a postupy pro zpracování experimentálních dat převážně s využitím výpočetní techniky.

### **Okruh čtenářů**

Knihou představuje vysokoškolskou učebnici exaktního zpracování experimentálních dat a je určena především studentům vysokých škol technického, přírodovědného, ale i humanitního a ekonomického směru. Spolu se statistickým systémem ADSTAT je jako účinný prostředek statistické analýzy experimentálních dat vhodná také pro pracovníky výzkumných ústavů, podnikových a státních kontrolních laboratoří a zkušeben, ke kontrole kvality a jakosti. Je bezprostředně využitelná v chemické laboratoři a chemometrii, v ekonomice a ekonometrii, sociologii, medicíně, biologii, sledování kvality životního prostředí apod. Ukazuje se, že předešlá vydání knihy nacházejí uplatnění i v jiných technických a přírodovědných oborech.

### **Způsob výuky**

Způsob výkladu látky odpovídá praktickému uživatelskému zaměření učebnice. Jádrem postupu je rozbor problému na charakteristickém příkladu, a to od formulace problému přes zadání vstupních dat, volbu programu, výklad postupu řešení až k závěru. Pro porozumění výkladu stačí předběžné znalosti matematiky na úrovni vysoké školy technického směru, většinou však pouze znalosti matematiky středoškolské. Kniha obsahuje přes 400 řešených příkladů i řadu úloh k samostatnému procvičování látky. Data ke všem příkladům a úlohám jsou na zvláštní disketě, která je k dispozici s programem ADSTAT. Příklady a úlohy jsou převzaty z literatury nebo z praxe laboratoří technické kontroly kvality a dalších technických laboratoří.

Příklady pomáhají čtenáři procvičovat statistickou interaktivní analýzu dat, konfrontovat závěry více testů a diagnostických grafů a posoudit i jejich vypovídací schopnost. Právě na příkladech z laboratorní praxe, odlišných od školních

modelových úloh, se čtenář naučí reálně použitelné metody vyšetřování dat, a to i v případech, kdy data nevedou k jednoznačným závěrům.

### Předešlá vydání

Původní text 1. vydání byl vydán pod titulem M. Meloun, J. Militký: *Chemometrie*, VŠCHT Pardubice 1988. Po úpravě vyšel v 2. vydání pod titulem M. Meloun, J. Militký: *Chemometrie - Zpracování experimentálních dat na IBM PC*, SNTL Praha 1991. V této podobě získala kniha cenu Českého literárního fondu za rok 1991. Následující 3. vydání přineslo úpravy směrem k obecnějšímu použití vyhodnocování experimentálních dat vedle chemometrie i v biometrii a ekonometrii a mělo titul M. Meloun, J. Militký: *Statistické zpracování experimentálních dat na osobních počítačích (v chemometrii, biometrii, ekonometrii a v dalších oborech přírodních, technických a společenských věd)*, Finish Pardubice 1992. Toto vydání doznalo řadu kritických úprav a rozšíření o úlohy k samostatnému procvičování.

### Doporučený software

Při výběru vhodného software jsme sledovali kvalitu dle technických požadavků, kladených na software, dále úroveň obsluhy a konečně úroveň metod a algoritmů. **Technické požadavky** zahrnují minimální kapacitu operační paměti, velikost pevného disku, nutnost matematického koprocesoru, typ grafického zobrazení a typ tiskáren a ploterů. **Úroveň obsluhy** se týká způsobu manipulace s daty, ovládání systémů obvykle cestou hierarchických nabídek (menu) nebo ikon, interaktivní nebo dávkové zpracování dat a manipulace s výsledky. **Úroveň metod a algoritmů** jsme posuzovali dle numerického a statistického hlediska. Je třeba si uvědomit, že každá statistická metoda platí za jistých a často dosti striktních předpokladů, a pokud nejsou tyto předpoklady splněny, jsou výsledky buď velmi přibližné nebo zcela zavádějící a chybné. **Ověřování předpokladů statistických metod** je proto nezbytnou součástí kvalitních programů. Navíc je často důležité volit i alternativní postupy, které jsou užitečné při nesplnění některých předpokladů, resp. pracují adaptivně, přizpůsobují se chování dat.

V současné době začíná docházet postupně k unifikaci obsluhy programových balíků související s jejich aplikací pod operačními systémy typu WINDOWS. Rozdíl se projevují především v rozsahu nabídky různých metod a numerické stránce použitých algoritmů. Zde je kvalitativní posun velmi malý a řada nových verzí se značně zlepšenými možnostmi obsluhy je prakticky stejná s ohledem na úroveň metod a algoritmů.

Ve statistických programech se vyskytují různé techniky, vedoucí často i k rozporným závěrům. Některé statistické metody jsou použitelné jen pro některé typy modelů. V dalším jsou uvedeny programové balíky vhodné ke zpracování experimentálních dat a dále i univerzální balíky, které jsou v našich podmínkách běžně dostupné a patří ve světě k nejužívanějším:

**BMDP (BMDP Inc., USA)**

Ve verzi PC-90 jde o rozsáhlý programový systém, složený z izolovaných 41 programů ve 13 blocích, ovládaný přes speciální editor. Nevýhodou pro začátečníky je pevná organizace řídicích příkazů, které musí uživatel používat. Výhodou pro delší speciální výpočty je i možnost práce v dávkovém režimu. Programy obsahují množství velmi kvalitních metod jak klasických, tak moderních a celou řadu diagnostických charakteristik. Systém umožňuje zpracovávat i rozsáhlé datové soubory, a je proto oblíben zejména v přírodních oborech a medicíně. Ve verzi BMDP New System je umožněna práce pod WINDOWS, což sebou nese na jedné straně zjednodušení obsluhy a na druhé straně omezení flexibility

**SAS (SAS Institute Inc., USA)**

Velmi rozsáhlý programový soubor je pro technické aplikace převedený ze sálových počítačů. Je orientován na komplexní užití v technické praxi, a proto má i speciální moduly, např. pro řízení kvality, pro analýzu spolehlivosti, atd. Stojí za doporučení.

**SPSS (SPSS Inc., USA)**

Představitel rozsáhlých statistických systémů, převedených pro PC ze sálových počítačů. Je koncipován modulárně a umožňuje práci v příkazovém a interaktivním režimu, ale také v dávkovém režimu. Využívá pevně strukturovaných řídicích příkazů. Umožňuje zpracování rozsáhlých datových souborů, vyskytujících se např. v ekonomii a sociologii.

**STATGRAPHICS (STSC Inc., USA)**

Patří poněkud paradoxně mezi nejoblíbenější statistické programy, zejména mezi nestatistiky. Byl konstruován speciálně pro osobní počítače. Vyznačuje se jednoduchostí ovládní systémem nabídek a panelů. Umožňuje kreslit a zobrazovat i velmi náročné grafy.

**SOLO (BMDP Statistical Software Inc., USA)**

Zajímavý statistický systém v jazyce BASIC. Je řízený systémem nabídek a panelů bez předdefinovaných hodnot. V základním modulu obsahuje programy pro nejběžnější statistické úlohy a v pokročilém modulu i komplikovanější úlohy, např. pro analýzu časových řad, plánované experimenty, analýzu přežití, řízení kvality, kontrolní karty, atd. Patří mezi nejlepší při porovnání jeho možností s jeho cenou.

**NCSS (BMDP / SPSS, USA)**

Rozsahem nevelký, ale účinný statistický balík, který je pokračováním úspěšného balíku SOLO firmy BMDP Statistical Software. Obsahuje výkonný tabulkový procesor podobný Excelu s kapacitou 16384 řádků a 256 sloupců, v němž lze snadno upravovat a transformovat data. Po vyplnění formulářů pro analýzu nebo tvorbu grafu se spustí výpočet. Výstup je organizován jako editovatelný text doplněný grafy ve formátu \*.WMF. Metody jsou velmi kvalitní a umožňují i nestandardní postupy jako robustní metody, fuzzy metody shlukové analýzy. Menu nabízí tyto metody: ANOVA, MANOVA, GLM - ANOVA, shluková analýza, prokládání křivek, nelineární regrese, růstové křivky, polynomické křivky, popisná statistika, návrh experimentu, analýza časových řad, ARIMA, Box-Jenkinsovy

modely, spektrální analýza, faktorová a diskriminační analýza, hlavní komponenty, lineární regrese, analýza přežití, stepwise regrese, robustní regrese (L1), regulační diagramy. Dobře je řešen import a export dat a export grafiky.

#### **SYSTAT** (Systat / SPSS, USA)

Rozsahem a statistickou náplní podobný programu NCSS. Systat má za sebou třicetiletý vývoj až k interaktivní aplikaci pro Windows. Jedná se o spolehlivý software s poměrně širokým a vyváženým výběrem statistických metod, z nichž pro některé je však nutné znát některé syntaktické příkazy. Program se ovládá interaktivně pomocí menu, Pro uživatelské analýzy je možné využít jednoduchý jazyk Systat Basic. Poněkud obtížnější je práce s daty a jejich úprava. Systém nabízí dynamické 2D a 3D grafy, na nichž je možné interaktivně označovat jednotlivé body, či skupiny bodů s možností vazby na datový editor. Grafy je možné přenášet do jiných aplikací v objektovém formátu a částečně editovat. Ze statistických metod jsou k dispozici popisná statistika, neparametrické testy, ANOVA, návrh experimentu, korelační, diskriminační a faktorová analýza, hlavní komponenty, lineární, GLM a stepwise regrese, časové řady, Fourierova transformace, vyhlazování, nelineární regrese, shluková analýza.

#### **S-PLUS** (MathSoft, USA)

Jeden z nejrozsáhlejších statistických systémů pro PC. Software byl vyvinut v Bellových laboratořích AT&T až začátkem 80. let. Tím je dána moderní objektově-orientovaná filosofie a struktura systému, jehož základním prvkem je dynamický objekt. To umožňuje velmi efektivní práci s daty, výpočetními metodami, interaktivními grafy a výstupy atd. a snadnou komunikaci s jinými programy. Systém nabízí zhruba 1400 statistických, matematických a dalších funkcí prakticky ze všech oblastí statistické analýzy, které jsou přístupné pomocí redefinovatelného menu, tlačítek, nebo z příkazového řádku. Na zvláště vysoké úrovni jsou lineární regresní modely klasické, zobecněné lineární (GLIM), a aditivní (GAM), modely stabilizující varianci (AVAS, ACE), robustní metody (L1, uřezané nejmenší čtverce, nejmenší medián čtverců, M-odhady), analýza přežití, lineární a nelineární modely se smíšenými efekty, shluková analýza včetně robustních a fuzzy-metod, časové řady a zvláště modul Wavelets pro analýzu signálů, časových řad a obrazu, modul GARCH pro analýzu finančních řad. Přes značnou rozsáhlost systému je jeho dynamickou strukturou zajištěna neobyčejně snadná obsluha. Důležitým rysem systému je objektově orientovaný statistický jazyk S, pomocí něhož je možné sestavit libovolnou analýzu či výpočet, který může být snadno integrován do systému jako nová funkce. Systém je podporován anglickou dokumentací a řadou odborných publikací.

#### **ADSTAT** (TriloByte Statistical Software, Česká republika)

Nabízí 8 základních modulů obsahujících přes 30 statistických metod. Práci usnadňuje výkonný tabulkový editor, manažer diskových souborů a dynamický grafický výstup společně s přehledným tabulkovým výstupem. Data, která byla zadána v editoru lze zpracovávat libovolnými moduly. Grafické výstupy lze tisknout v prezentační kvalitě, nebo ukládat do souborů ve formátu TIF pro

snadný export do jiných programových balíků. ADSTAT je dodáván jako modulární systém zahrnující vždy základní uživatelské prostředí včetně datového editoru a tabelárního a grafického výstupu dat, které je doplněno o příslušný statistický modul zpracování dat. Moduly lze provozovat i zakoupit samostatně.

**QC-EXPERT** (TriloByte Statistical Software, Česká republika)

Český statistický software pro sledování jakosti a analýzu dat v technologii, laboratoři a obchodu. Program poskytuje praktický přehledný výstup v podobě protokolu a grafů, který je přímo použitelný jako dokumentace analýzy. Nabídka systému obsahuje následující témata: základní statistika a diagnostika dat, automatická mocninná transformace, jednofaktorová analýza rozptylu, z-skóre, klasická a robustní korelační analýza, modelování responsních povrchů a jejich optimalizace, metody pro statistickou přejímku, Shewhartovy regulační diagramy, regulační diagramy EWMA, CUSUM a Hotellingovy regulační diagramy pro více proměnných. Novinkou jsou dynamické regulační diagramy s predikovanými regulačními mezemi založené na klouzavých průměrech, které umožňují statistickou regulaci autokorelovaných procesů a procesů s nekonstantní střední hodnotou. Výstupy poskytují kompletní diagnostiku a interpretaci dat a analýz. Systém QC-Expert je doprovázen uživatelským manuálem a metodickou učebnicí Statistické řízení jakosti, která popisuje a vysvětluje použité metody a další souvislosti.

**Poděkování**

Není možné poděkovat všem spolupracovníkům, studentům a doktorantům, kteří nám pomáhali či přispěli praktickými úlohami, radami či konstruktivní kritikou.

Naše vřelé díky patří především ing. Dáše Militké za pečlivou korektorskou práci. V počítačové typografii při editování rukopisu v textovém editoru WordPerfect 6.0 nám byl významným pomocníkem ing. Václav Dušek. Paní Ludmile Vařekové patří naše díky za pomoc při sazbě původního rukopisu.

Díky patří také všem pedagogům a studentům řádného i licenčního studia, kteří nám poskytli cenné dotazy, podněty a připomínky k předešlým vydáním.

Pardubice, listopad 1997

Milan Meloun a Jiří Militký



## O autorech

**MILAN MELOUN** (\*1943), Prof. RNDr. DrSc., vystudoval přírodovědeckou fakultu University J. E. Purkyně (dnešní Masarykova) v Brně 1965. Je profesorem analytické chemie a chemometrie na Katedře analytické chemie Chemicko-technologické fakulty Univerzity v Pardubicích. Vyučoval statistické metody a analytickou chemii na Baghdádské univerzitě v Iráku a na Královské technice The Royal Institute of Technology ve Stockholmu. Je autorem a spoluautorem více než 70 originálních sdělení, 15 monografií a 7 vysokoškolských učebnic, 8 patentů a zlepšovacích návrhů a na konferencích přednesl více než 150 přednášek. Je členem redakčních rad zahraničních odborných časopisů *Talanta* a *Analytica Chimica Acta*, předsedou sekce Chemometrie při České společnosti chemické.

Většina jeho publikací se týká počítačově orientované analýzy instrumentálních dat při studiu rovnovah v roztocích a určování reakčního produktu. Knižně se uvedl dvoudílnou monografií *Computation of Solution Equilibria*, Folia UJEP Brno 1984 (spoluautor Josef Havel), která pak doplněná o extrakční rovnováhy (spoluautor Erik Högfeltdt) vyšla v roce 1988 v anglickém nakladatelství Ellis Horwood, Chichester. K této problematice se spolu s Josefem Havlem vrátil ve dvou obsáhlých kapitolách monografie *Computational Methods for the Determination of Formation Constants*, Plenum Press New York 1985, editované Davidem Leggettem.

V jeho vědecké práci představuje počítač spojovací článek mezi statistikou a analytickou chemií. Výsledkem je analytická chemometrie, předmět, který přednáší na Univerzitě (dříve VŠCHT) v Pardubicích od roku 1978. Dokladem je řada učebních textů, např. *Chemometrie*, VŠCHT Pardubice 1985 (spoluautor Milan Javůrek). V zahraničí přednášel dle textů svých učebnic, a to při dvouletém pobytu na Univerzitě v Bagdadu *Data Analysis by Statistical and Computing Technique*, University Baghdad Press, Baghdad 1980. Při hostování na Královské univerzitě ve Stockholmu používal svou sbírku příkladů *Introduction to Chemometrics*, která je postavena na interaktivní analýze statistického systému STATGRAPHICS. Vlastní přístup k analýze experimentálních dat se promítá i do kapitoly *Chemometrics in the Instrumental Laboratory* v monografii, editované Jaroslavem Churáčkem, *Advanced*

*Instrumental Methods of Chemical Analysis*, Academia Praha 1993, nebo *Hodnocení analytických výsledků* ve Vláčilově sbírce *Příklady z chemické a instrumentální analýzy*, SNTL Praha 1983.

Obsáhlá rešerše a zkušenosti spoluautora Jiřího Militkého přinesly řadu novějších postupů ze statistické analýzy dat, průzkumové analýzy a především interaktivní přístup k analýze dat na osobním počítači. Společně tak vzniklo první vydání učebnice *Chemometrie - Zpracování experimentálních dat na IBM PC*. Text byl přeložen do angličtiny a po doplnění o kapitoly vícerozměrné statistiky Michele Forinou vychází postupně jako dvoudílná učebnice u nakladatele Ellis Horwood, Chichester 1991 pod titulem *Chemometrics for Analytical Chemistry - Volume I. PC-Aided Statistical Data Analysis, Volume II. Regression Model Building and Testing*.

Na Univerzitě v Pardubicích přednáší v řádném studiu chemometrii, organizuje postgraduální a dvouleté licenční studium chemometrie a i krátkodobé intenzivní kurzy chemometrie pro aplikaci v průmyslu. V těchto formách studia chemometrie je užívána především učebnice Milan Meloun a Jiří Militký: *Statistické zpracování experimentálních dat na osobních počítačích*, Finish Pardubice 1992 nebo PLUS Praha 1994, jejíž další doplněné vydání je zde předkládáno. Ve výuce chemometrie nachází uplatnění i jeho učebnice rozšířeného a v laboratořích hojně užívaného software. Jedná se o prakticky zaměřenou učebnici textového editoru WordPerfect a učebnici české verze Quattro Pro, doplněnou o stovku vyřešených praktických příkladů.

**JIŘÍ MILITKÝ** (\*1949), Prof. Ing. CSc., ukončil fakultu textilní, specializace textilní chemie, na VŠST v Liberci roku 1973 s vyznamenáním. V letech 1974 až 1976 pracoval ve Státním výzkumném ústavu textilním, Liberec v oddělení matematického modelování textilních struktur. V letech 1976 až 1989 pracoval ve Výzkumném ústavu zušlechťovacím, Dvůr Králové n. L., kde se věnoval převážně zpracování experimentálních dat s využitím výpočetní techniky.

Od roku 1990 je vedoucím katedry textilních materiálů na Vysoké škole strojní a textilní v Liberci. V roce 1982 obhájil kandidátskou disertační práci z oblasti fyziky textilních vláken. V roce 1989 byl jmenován docentem a v roce 1992 se habilitoval. V prosinci 1993 byl jmenován řádným profesorem.

Je členem několika vědeckých společností (The Textile Institute, JČMF) a pracuje ve výboru sekce Chemometrie při České společnosti chemické.

Jeho publikační činnost zahrnuje oblasti textilního inženýrství, modelování kinetických procesů v pevné fázi a zpracování experimentálních dat. Je autorem nebo spoluautorem 606 vědeckých příspěvků (publikací, monografií, referátů a posterů). Jeho první kniha *Modifikovaná PES vlákna* (spoluautoři Jiří Kryštůfek, Jiří Vaníček a Oldřich Hartych) vyšla v SNTL v roce 1984. Zcela přepracované a rozšířené vydání bylo publikováno nakladatelstvím Elsevier v roce 1991. S Jiřím Kryštůfkem zpracoval knihu *Barvení akrylových vláken a směsí*, která vyšla

v SNTL Praha v roce 1987. Ve spolupráci s Milanem Melounem publikoval učebnice a monografie z oblasti využití interaktivních statistických metod v chemometrii.

Jiří Militký publikoval celkem 8 knih, z nichž tři jsou zaměřeny do oblasti zpracování experimentálních dat s využitím výpočetní techniky. Moderní metody interaktivní statistické analýzy dat zpracoval do rozsáhlého seriálu příruček *Statistické metody v textilní praxi I - IV*, vydaného v letech 1982 až 1985 v Domě techniky Pardubice. Přehled metod regrese a matematického modelování publikoval v seriálu skript *Tvorba matematických modelů I - VI*, vydaných v letech 1983 až 1989 v Domě techniky Ostrava.

Vytvořil systém programů pro zpracování experimentálních dat v jazyce HPL. Tyto programy jsou charakteristické tím, že kromě stránky *statistické*, vycházející vždy nejdříve z ověřování předpokladů o modelech, datech a použité metodě, využívají také progresivních *numerických* postupů (zejména v oblasti lineární a nelineární regrese). Tyto algoritmy se později staly jádrem originálního programového systému ADSTAT.

Prezentoval příspěvky na řadě konferencí o *počítačové statistice* (Edinburg, Řím, Kodaň, Dubrovník, Vídeň, Neuchatel, Tampere, atd.), *chemometrii* (Montreal, Boloňa, Taormina, atd.) a *souvisejících vědních disciplínách* (Nice, Perugia, Ithaca, Honolulu, Kyoto, Mt Fuji, Interlaken, Bukurešť, Lodž, Budapešť, Stockholm, Norimberk, Hakone, Bolton, Espoo, atd.).

