

易于使用

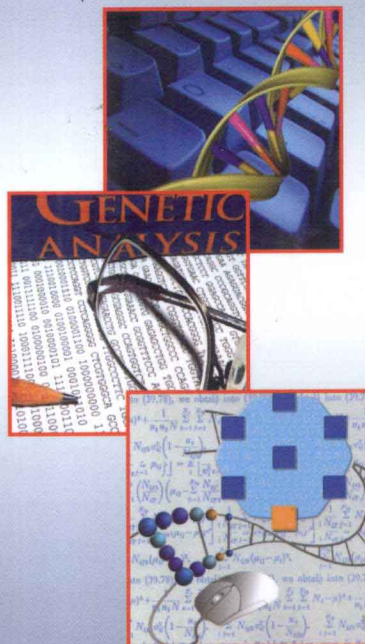
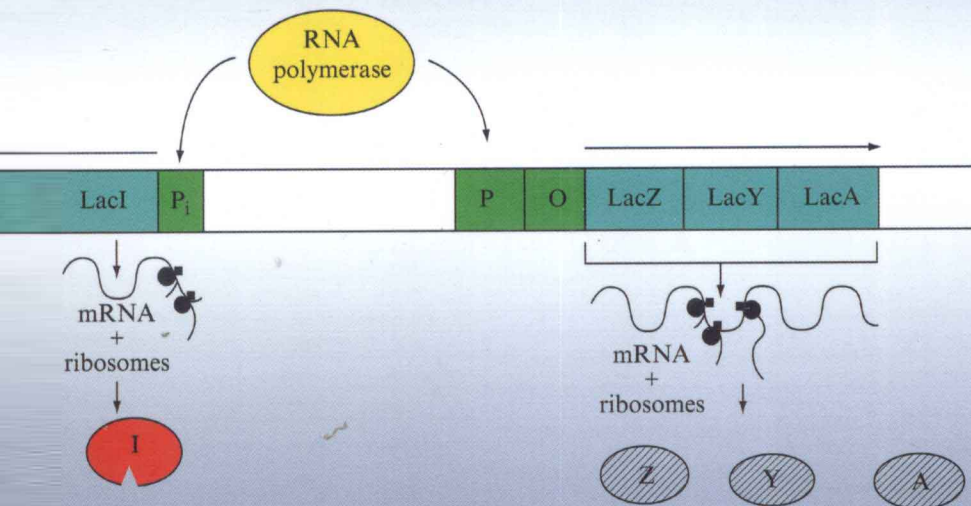
值得信赖

专业权威

数值计算方法精要

Essential Numerical Computer Methods

Michael L. Johnson



实验室解决方案

Essential Numerical Computer Methods

数值计算方法精要

Edited by
Michael L. Johnson



科学出版社

北京

图字:01-2011-5493 号

This is an annotated version of
Essential Numerical Computer Methods
Edited by Michael Johnson.

Copyright © 2011 ELSEVIER INC.
ISBN: 978-0-12-384997-7

All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publisher.

AUTHORIZED EDITION FOR SALE IN P. R. CHINA ONLY
本版本只限于在中华人民共和国境内销售

图书在版编目(CIP)数据

数值计算方法精要=Essential Numerical Computer Methods:英文/(美)约翰逊
(Johnson, M. L.)主编. —北京:科学出版社, 2012
(实验室解决方案)
ISBN 978-7-03-032912-7

I. ①数… II. ①约… III. ①数值计算-计算方法-英文 IV. ①0241

中国版本图书馆 CIP 数据核字(2011)第 246258 号

责任编辑:李小汀/责任印制:钱玉芬
封面设计:耕者设计工作室

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

北京佳信达欣艺术印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2012 年 1 月第 一 版 开本:787×1092 1/16

2012 年 1 月第一次印刷 印张:41

字数:970 000

定价:168.00 元

(如有印装质量问题,我社负责调换)

导 读

近几十年来，微电子、计算机和自动化等科学技术的飞速发展直接带动了高通量生物技术的迅猛发展，生物医学工作者面临着种类越来越丰富、数据量越来越大的生物医学数据，包括生理信号、生化信号、基因组数据、转录组数据、蛋白质组数据和代谢组数据等。如何更好地存储、分析和解释这些数据在整个生物医学研究中起着举足轻重的作用，已经成为生物医学研究中不可或缺的一环。

很多生物医学数据具有其特殊性，比如非线性、高维度、高噪声等。即使不同生物医学数据，其特性也具有很大的差别。从微观的 DNA、基因、蛋白质或其他小分子，到宏观的电生理信号，再到临床信息，这些数据既具有一定的共性，在更大程度上，更具有不同之处。不同的数据，不同的科学问题，需要不同的合适的算法，这样才能保证最大程度上的反映客观世界，得出科学的结论，指导进一步的生物医学研究。否则，恐怕会适得其反，会造成偏差甚至酿成错误。

因此，生物医学数据处理算法的研究则显得尤其重要。由 Michael Johnson 主编的 *Essential Numerical Computer Methods* 一书汇集了生物医学数据分析不同领域专家所撰写的章节，图文并茂，结合生物医学实际问题，深入浅出地阐述相应领域的计算方法。该书非常适合生物医学数据分析领域的研究人员阅读。该书对一些经典信号处理算法也有结合实例的诠释，这对于开拓不同领域的研究人员视野也大有益处。同时，对于在计算方法学领域已经具有一定造诣并正在寻求在生物医学领域应用的科研人员来说，一定也会发现该书的价值。另外，对于传统的生物医学研究人员，在面对各种各样的生物医学数据而苦于找不到合适的计算方法时，浏览一下本书，也会大有益处的。随着学科交叉的继续深入下去，我相信，从事生物医学研究的不同背景的人员都会从本书中从他们自己的角度吸收到所需营养。

崔庆华
于北京大学医学部

前 言

本书中的多章内容在近二十年前就首次出版了。在此期间，那些基本算法没有改变。然而，发生变化的是计算机速度和易用性获得了巨大提升，以及相应的计算机价格的快速下降。计算机速度的提升使得二十年前几乎不能用于生化实验室的算法现在得到普遍应用。在过去的二十年中，在技术和医学研究部门接受良好培训的博士生们并没有跟上数学、数值分析、统计学或计算机科学发展的脚步。

然而，在生物和生物医学研究中，计算机和计算方法的使用已经变得无处不在。一个主要原因就是美国国家卫生研究院（NIH）工作路线图对计算机和计算方法的重视。另一个因素是研究人员，特别是初级科学家、学生、杂志审稿人，以及 NIH 研究部成员所面临的日益增加的数学和计算复杂性。同时，计算机硬件和软件的快速发展迅速拉近了这些计算方法离科研领域普通工作人员的距离。

对于计算机和计算方法在生物和生物医学研究中的应用，人们的一个普遍认识是这些应用要么是基本的统计分析，要不就是 DNA 序列数据的检索。这些应用确实重要，但是它们仅仅揭开了当前或今后计算机和计算方法在生物医学研究领域的序幕。本书各章包含了丰富多彩的计算机和计算方法在生物医学研究领域的应用，这大大扩展了我们以前对该领域的认识。本书各章基本上是按照最初发表在 *Methods in Enzymology* 第 210、240、321、383、384、454 和 467 卷上的时间顺序排列的。这种按时间排列的顺序也为读者提供了一种从基本数值方法到其在专业生物和生物医学应用的渐进。

（崔庆华 译）

CONTRIBUTORS

Numbers in parentheses indicate the pages on which the authors' contributions begin.

- Robert D. Abbott** (67), Division of Biostatistics, Department of Internal Medicine, University of Virginia School of Medicine, Charlottesville, Virginia, USA
- G. Bard Ermentrout** (293), Department of Mathematics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
- James C. Boyd** (561), Department of Pathology, University of Virginia Health System, Charlottesville, Virginia, USA
- Marc Breton** (461), University of Virginia Health System, Charlottesville, Virginia, USA
- Emery N. Brown** (187, 199), Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital, Harvard Medical School/MIT Division of Health Sciences and Technology, Boston, Massachusetts, USA
- David E. Bruns** (561), Department of Pathology, University of Virginia Health System, Charlottesville, Virginia, USA
- William Clarke** (461), University of Virginia Health System, Charlottesville, Virginia, USA
- David F. Dinges** (225), Unit for Experimental Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA
- Harold B. Dowse** (479), School of Biology and Ecology, University of Maine, Orono, Maine, USA; Department of Mathematics and Statistics, University of Maine, Orono, Maine, USA
- Leon S. Farhy** (309, 585), Department of Medicine, Center for Biomathematical Technology, University of Virginia, Charlottesville, Virginia, USA
- Howard P. Gutgesell** (67), Division of Pediatric Cardiology, Department of Pediatrics, University of Virginia School of Medicine, Charlottesville, Virginia, USA
- E. R. Henry** (81), Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda MD, USA
- J. Hofrichter** (81), Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda MD, USA
- Abdul Salam Jarrah** (529), Virginia Bioinformatics Institute at Virginia Tech, Blacksburg, Virginia, USA; Department of Mathematics and Statistics, American University of Sharjah, Sharjah, United Arab Emirates
- Michael L. Johnson** (1, 37, 55, 173), University of Virginia Health System, Charlottesville, VA, USA
- Boris Kovatchev** (461), University of Virginia Health System, Charlottesville, Virginia, USA

- Michelle Lampl** (173), Emory University, Atlanta, GA, USA
- Reinhard Laubenbacher** (529), Virginia Bioinformatics Institute at Virginia Tech, Blacksburg, Virginia, USA
- Leslie M. Loew** (399), R. D. Berlin Center for Cell Analysis and Modeling, University of Connecticut Health Center, Farmington, CT, USA
- Greg Maislin** (225), Biomedical Statistical Consulting, Wynnewood, Pennsylvania, USA
- Anthony L. McCall** (585), Department of Medicine, Center for Biomathematical Technology, University of Virginia, Charlottesville, Virginia, USA
- Jay I. Myung** (511), Department of Psychology, Ohio State University, Columbus, Ohio, USA
- Erik Olofsen** (225), Department of Anesthesiology, P5Q, Leiden University Medical Center, 2300 RC, Leiden, The Netherlands
- Steven M. Pincus** (141), Independent Mathematician, Guilford, CT, USA
- Mark A. Pitt** (511), Department of Psychology, Ohio State University, Columbus, Ohio, USA
- Douglas Poland** (257), Department of Chemistry, The Johns Hopkins University, Baltimore, Maryland, USA
- James C. Schaff** (399), R. D. Berlin Center for Cell Analysis and Modeling, University of Connecticut Health Center, Farmington, CT, USA
- Christopher H. Schmid** (187, 199), Biostatistics Research Center, Division of Clinical Care Research, New England Medical Center, Tufts University, Boston, Massachusetts, USA
- Boris M. Slepchenko** (399), R. D. Berlin Center for Cell Analysis and Modeling, University of Connecticut Health Center, Farmington, CT, USA
- Walter F. Stafford III** (337), Analytical Centrifugation Research Laboratory, Boston Biomedical Research Institute, Boston, MA, USA; Department of Neurology, Harvard Medical School, Boston, MA, USA
- Martin Straume** (37, 55, 173), COBRA, Inc.
- Yun Tang** (511), Department of Psychology, Ohio State University, Columbus, Ohio, USA
- Joel Tellinghuisen** (361), Department of Chemistry, Vanderbilt University, Nashville, Tennessee, USA
- Hans P. A. Van Dongen** (225), Unit for Experimental Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA
- Donald G. Watts** (23), Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada K7L 3N6
- F. Eugene Yates** (425), The John Douglas French Alzheimer's Foundation, 11620 Wilshire Blvd, Suite 270, Los Angeles, CA, USA

PREFACE

Many of the chapters within this volume were first published almost two decades ago. Since then, these basic algorithms have not changed. However, what has changed is the huge increase in computer speed and ease of use along with the corresponding decrease in the costs. The increase in computer speed has made the use of some algorithms common that were almost never used in biochemistry laboratories two decades ago. During the past two decades, the training of the majority of senior M.D.s and Ph.D.s in clinical or basic disciplines at academic research and medical centers has not kept pace with advanced coursework in mathematics, numerical analysis, statistics, or computer science.

Nevertheless, the use of computers and computational methods has become ubiquitous in biological and biomedical research. One primary reason is the emphasis being placed on computers and computational methods within the National Institutes of Health (NIH) Roadmap. Another factor is the increased level of mathematical and computational sophistication among researchers, particularly among junior scientists, students, journal reviewers and NIH Study Section members. Yet another is the rapid advances in computer hardware and software that make these methods far more accessible to the rank-and-file members of the research community.

There exists a general perception that the applications of computers and computer methods in biological and biomedical research are either basic statistical analysis or the searching of DNA sequence data bases. While these are important applications, they only scratch the surface of the current and potential applications of computers and computer methods in biomedical research. The various chapters within this volume include a wide variety of applications that extend far beyond this limited perception. The chapters within this volume are basically in chronological order of original publication in *Methods in Enzymology* volumes 210, 240, 321, 383, 384, 454, and 467. This chronological order also provides a general progression from basic numerical methods to more specific biochemical and biomedical applications.

目 录

撰稿人	xiii
前言	xv
1. 最小二乘技术在生物化学中的应用	1
<i>Michael L. Johnson</i>	
I. 相对上一版的更新	1
II. 简介	2
III. 非线性最小二乘	4
IV. 为什么使用非线性最小二乘	6
V. 什么情况下使用非线性最小二乘	7
VI. 非线性最小二乘分析的结果有怎样的可信度	14
VII. 结论	20
参考文献	22
2. 非线性模型的参数估计	23
<i>Donald G. Watts</i>	
I. 简介	24
II. 讨论	35
参考文献	36
3. 残差分析：拟合优度检验	37
<i>Martin Straume and Michael L. Johnson</i>	
I. 相对上一版的更新	37
II. 简介	38
III. 残差散点图	40
IV. 残差的累积概率分布	41
V. 卡方统计量：对观察残差值频率与期望残差值频率之比定量化	43
VI. Kolmogorov-Smirnov 检验：除卡方检验之外的另一选择	44
VII. Runs 检验：对残差的趋势定量化	45
VIII. 序列滞后图：确定序列相关	47
IX. Durbin-Watson 检验：序列相关的定量检验方法	47
X. 自相关：检测时间序列实验中的序列相关	49
XI. 卡方检验：拟合优度的定量方法	50
XII. 野点：确定非正常样本点	51
XIII. 确定强影响点	52

XIV. 结论	53
参考文献	54
4 确定估计出的模型参数的完全置信概率的蒙特卡罗方法	55
<i>Martin Straume and Michael L. Johnson</i>	
I. 相对上一版的更新	55
II. 简介	56
III. 蒙特卡罗方法	58
IV. 为估计出的参数生成置信概率分布	58
V. 实现和解释	60
VI. 结论	65
参考文献	66
5. 异方差性和偏度对回归中预测的影响：人类心脏的生长建模	67
<i>Robert D. Abbott and Howard P. Gutgesell</i>	
I. 简介	67
II. 以人类心脏的生长建模为例	68
III. 估计方法	70
IV. 讨论	78
参考文献	79
6. 奇异值分解：实验数据分析中的应用	81
<i>E. R. Henry and J. Hofrichter</i>	
I. 相对上一版的更新	82
II. 简介	83
III. 定义与性质	91
IV. 含有噪声矩阵的奇异值分解	93
V. 奇异值分解在实验数据分析中的应用	107
VI. 一个简单例子的模拟：反应 $A \rightarrow B \rightarrow C$	120
VII. 小结	133
参考文献	138
7. 生物网络信号中的非正则性和异步性	141
<i>Steven M. Pincus</i>	
I. 相对上一版的更新	142
II. 简介	142
III. 正则性的定量化	146
IV. 实现和解释	148
V. 有代表性的生物学应用	152
VI. 与其他方法的关系	154

VII. 变正则性的力学假设	158
VIII. 互近似熵	158
IX. 面向更准确的网络建模型	166
X. 空间（向量）近似熵	167
XI. 小结	169
参考文献	170
8. 用近似熵辨别生长模型	173
<i>Michael L. Johnson, Michelle Lampl, and Martin Straume</i>	
I. 相对上一版的更新	173
II. 简介	174
III. 近似熵的定义和计算	175
IV. 为本章应用对近似熵计算的改进	176
V. 生长模型	177
VI. 期望模型依赖的近似熵分布	178
VII. 近似熵在本模型的应用实例	179
VIII. 小结	182
参考文献	184
9. 卡尔曼滤波在统计学计算问题中的应用	187
<i>Emery N. Brown and Christopher H. Schmid</i>	
I. 简介	187
II. 用卡尔曼滤波估计高斯似然值	189
III. 用卡尔曼滤波为贝叶斯推断计算后验概率密度	192
IV. 缺失值问题和卡尔曼滤波	193
V. 卡尔曼滤波算法的扩展	196
参考文献	196
10. 贝叶斯层次模型	199
<i>Christopher H. Schmid and Emery N. Brown</i>	
I. 简介	199
II. 高斯模型	203
III. 计算	208
IV. 实例：荟萃回归	211
V. 实例：婴儿生长的 Saltatory 模型	216
VI. 期望模型依赖的近似熵分布	219
VII. 模型检查	220
VIII. 小结	221
参考文献	222

11. 混合模型回归分析及个体差异的处理	225
<i>Hans P. A. Van Dongen, Erik Olofsen, David F. Dinges, and Greg Maislin</i>	
I. 简介	226
II. 实验和数据	227
III. 重复测量 ANOVA	228
IV. 混合模型回归分析	235
V. 一个备选线性混合效应模型	241
VI. 非线性混合模型回归分析	245
VII. 混合模型回归分析中的相关性结构	253
VIII. 小结	254
参考文献	255
12. 矩分布函数和最大熵方法	257
<i>Douglas Poland</i>	
I. 简介	257
II. 配体结合: 矩	259
III. 最大熵分布	265
IV. 配体结合: 分布函数	273
V. 热函分布	279
VI. 自关联分布	287
参考文献	291
13. 生物振荡中的数学	293
<i>G. Bard Ermentrout</i>	
I. 简介	293
II. 振荡子与兴奋性	294
III. 振荡子的干扰	299
IV. 耦合振荡子	304
参考文献	308
14. 带反馈的内分泌激素网络中振荡子的建模	309
<i>Leon S. Farhy</i>	
I. 简介	309
II. 内分泌激素网络建模中的一般原则	310
III. 单个激素分子丰度动力学的模拟	312
IV. 单个系统反馈环路驱动的振荡	315
V. 带有多个反馈环的网络	330
VI. 小结与讨论	333
参考文献	336

15. 沉淀速度实验中的边界分析	337
<i>Walter F. Stafford</i>	
I. 相对上一版的更新	338
II. 简介	338
III. 数据获取方法	340
IV. 分析用超速离心机中传输的测量	340
V. 传统分析方法	341
VI. 传输方法	342
VII. 平滑与差分	344
VIII. 表观沉淀系数分布函数的计算	349
IX. 沉淀系数的加权平均	354
X. 因散射效应校正分布函数的方法	357
XI. 讨论	358
参考文献	358
16. 等温滴定量热法中的统计误差	361
<i>Joel Tellinghuisen</i>	
I. 相对上一版的更新	361
II. 简介	364
III. 最小二乘中的方差-协方差矩阵	366
IV. 蒙特卡罗计算方法	372
V. Van't Hoff 分析: 最小二乘演示	373
VI. 等温滴定测热法	377
VII. 等温滴定测热法测热比较	390
VIII. 结论	394
参考文献	397
17. 基于虚拟细胞框架生理模拟	399
<i>James C. Schaff, Boris M. Slepchenko, and Leslie M. Loew</i>	
I. 相对上一版的更新	400
II. 简介	401
III. 细胞生理模拟的抽象过程	402
IV. 在现有模型上的应用	412
V. 结论	418
参考文献	424
18. 分形在生物学中的应用: 生化网络中的时间缩放	425
<i>F. Eugene Yates</i>	
I. 相对上一版的更新	426
II. 简介	427

III. 哺乳动物的分形形态学：一些分支	430
IV. 酶反应中的混沌	437
V. 确定生化反应中混沌和分形的实用指南	438
VI. 小结，或这些都意味着什么？	440
VII. 术语表	442
参考文献	458
19. 糖尿病连续葡萄糖检测数据获取和解释的分析方法	461
<i>Boris Kovatchev, Marc Breton, and William Clarke</i>	
I. 2010 年本领域的发展更新	462
II. 简介	463
III. 传感器误差的分解	466
IV. 平均血糖和偏差的测量	467
V. 危险和变异性的评估	469
VI. 系统稳定性的测量和绘图	472
VII. 基于时间序列预测将来的血糖值	473
VIII. 结论	476
参考文献	476
20. 生理和行为周期性的分析	479
<i>Harold B. Dowse</i>	
I. 简介	479
II. 生物学数据的类型及其获取方法	480
III. 时域分析	482
IV. 频域分析	488
V. 时/频分析及小波变换	496
VI. 信号调节	502
VII. 信号的强度和正则性	505
VIII. 对分析结果进行统计比较时一些实用的需要考虑的事	506
IX. 结论	507
参考文献	508
21. 计算模型的评估和比较	511
<i>Jay I. Myung, Yun Tang, and Mark A. Pitt</i>	
I. 相对上一版的更新	512
II. 简介	513
III. 模型评估和比较中的概念概览	513
IV. 模型比较方法	516
V. 模型比较实例：蛋白质折叠模型的选择	521
VI. 结论	525

参考文献	526
22. 生化网络中的代数模型	529
<i>Reinhard Laubenbacher and Abdul Salam Jarrah</i>	
I. 简介	530
II. 计算系统生物学	531
III. 网络推断	541
IV. 离散模型的逆向工程：一个实例	545
V. 讨论	553
参考文献	556
23. 蒙特卡洛模拟在为临床实验室检验建立分析质量需求中的应用：满足临床需求	561
<i>James C. Boyd and David E. Bruns</i>	
I. 相对上一版的更新	562
II. 简介	563
III. 建模方法	566
IV. 模拟研究中的方法	567
V. 结果	568
VI. 讨论	580
参考文献	582
24. 高血糖因子分泌和反调控的胰腺网络控制	585
<i>Leon S. Farhy and Anthony L. McCall</i>	
I. 相对上一版的更新	586
II. 简介	588
III. 糖尿病中高血糖因子反调控失常的机制	589
IV. 研究高血糖因子反调控缺陷的多学科交叉方法	591
V. 高血糖因子反调控控制的最初的定性分析	592
VI. STZ 处理大鼠中高血糖因子反调控控制机制的数学模型	595
VII. 用最小控制网络对正常胰腺的近似模拟及胰岛素缺陷胰腺中高血糖因子反调控异常的分析	598
VIII. 多学科交叉方法的优点和缺点	609
IX. 结论	612
参考文献	612
索引	619

(崔庆华 译)

封面图片：乳糖操纵子。编码区为蓝色表示，调控区为绿色表示，可扩散调控蛋白为红色表示。

CONTENTS

Contributors	xiii
Preface	xv
1. Use of Least-Squares Techniques in Biochemistry	1
<i>Michael L. Johnson</i>	
I. Update	1
II. Introduction	2
III. Nonlinear Least-Squares	4
IV. Why Use NLLS Analysis Procedures?	6
V. When to Use NLLS Analysis Procedures	7
VI. What Confidence Can Be Assigned to Results of NLLS Analysis?	14
VII. Conclusions	20
References	22
2. Parameter Estimates from Nonlinear Models	23
<i>Donald G. Watts</i>	
I. Introduction	24
II. Discussion	35
References	36
3. Analysis of Residuals: Criteria for Determining Goodness-of-Fit	37
<i>Martin Straume and Michael L. Johnson</i>	
I. Update	37
II. Introduction	38
III. Scatter Diagram Residual Plots	40
IV. Cumulative Probability Distributions of Residuals	41
V. χ^2 Statistic: Quantifying Observed Versus Expected Frequencies of Residual Values	43
VI. Kolmogorov-Smirnov Test: An Alternative to the χ^2 Statistic	44
VII. Runs Test: Quantifying Trends in Residuals	45
VIII. Serial Lag _n Plots: Identifying Serial Correlation	47
IX. Durbin-Watson Test: Quantitative Testing for Serial Correlation	47

X. Autocorrelation: Detecting Serial Correlation in Time Series Experiments	49
XI. χ^2 Test: Quantitation of Goodness-of-Fit	50
XII. Outliers: Identifying Bad Points	51
XIII. Identifying Influential Observations	52
XIV. Conclusions	53
References	54
 4. Monte Carlo Method for Determining Complete Confidence Probability Distributions of Estimated Model Parameters	 55
<i>Martin Straume and Michael L. Johnson</i>	
I. Update	55
II. Introduction	56
III. Monte Carlo Method	58
IV. Generating Confidence Probability Distributions for Estimated Parameters	58
V. Implementation and Interpretation	60
VI. Conclusion	65
References	66
 5. Effects of Heteroscedasticity and Skewness on Prediction in Regression: Modeling Growth of the Human Heart	 67
<i>Robert D. Abbott and Howard P. Gutgesell</i>	
I. Introduction	67
II. Example from Modeling Growth of the Human Heart	68
III. Methods of Estimation	70
IV. Discussion	78
References	79
 6. Singular Value Decomposition: Application to Analysis of Experimental Data	 81
<i>E. R. Henry and J. Hofrichter</i>	
I. Update	82
II. Introduction	83
III. Definition and Properties	91
IV. Singular Value Decomposition of Matrices Which Contain Noise	93
V. Application of Singular Value Decomposition to Analysis of Experimental Data	107
VI. Simulations for a Simple Example: The Reaction $A \rightarrow B \rightarrow C$	120
VII. Summary	133
References	138