Statistics and Chemometrics for Analytical Chemistry

James N. Miller, Jane C. Miller
and Robert D. Miller

Pearson

Seventh edition

Statistics and Chemometrics for Analytical Chemistry

Seventh Edition

Pearson

At Pearson, we have a simple mission: to help people make more of their lives through learning.

We combine innovative learning technology with trusted content and educational expertise to provide engaging and effective learning experiences that serve people wherever and whenever they are learning.

From classroom to boardroom, our curriculum materials, digital learning tools and testing programmes help to educate millions of people worldwide – more than any other private enterprise.

Every day our work helps learning flourish, and wherever learning flourishes, so do people.

To learn more, please visit us at **www.pearson.com/uk**

James N. Miller
Jane C. Miller
Robert D. Miller

Statistics and Chemometrics for Analytical Chemistry

Seventh Edition

Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • São Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

Pearson Education Limited
KAO Two
KAO Park
Harlow CM17 9NA
United Kingdom
Tel: +44 (0)1279 623623
Web: www.pearson.com/uk

Third edition published under the Ellis Horwood imprint 1993 (print)
Fourth edition published 2000 (print)
Fifth edition published 2005 (print)
Sixth edition published 2010 (print)
Seventh edition published 2018 (print and electronic)

ISBN: 978-1-292-18671-9 (print)
978-1-292-18672-6 (PDF)
978-1-292-18674-0 (ePub)

British Library Cataloguing-in-Publication Data
A catalogue record for the print edition is available from the British Library

Library of Congress Cataloging-in-Publication Data
Names: Miller, J. N. (James N.), 1943- author. | Miller, J. C. (Jane Charlotte), author. | Miller, Robert D. (Chemist), author.
Title: Statistics and chemometrics for analytical chemistry.
Description: Seventh edition / James N. Miller, Jane C. Miller, Robert D. Miller. | Harlow, United Kingdom : Pearson Education Limited, 2018.
Identifiers: LCCN 2017043378| ISBN 9781292186719 (print) | ISBN 9781292186726 (pdf) | ISBN 9781292186740 (epub)
Subjects: LCSH: Chemometrics—Textbooks.
Classification: LCC QD75.4.C45 M54 2018 | DDC 543/.072—dc23
LC record available at https://lccn.loc.gov/2017043378

10 9 8 7 6 5 4 3 2 1
22 21 20 19 18

Print edition typeset in 9.25/12 Stone Serif ITC Pro by iEnergizer Aptara®, Ltd.
Printed in Slovakia by Neografia

NOTE THAT ANY PAGE CROSS REFERENCES REFER TO THE PRINT EDITION

Contents

Preface to the seventh edition ix

Preface to the first edition xi

Publisher's acknowledgements xiii

Glossary of symbols xvii

1 Introduction 1

1.1 Analytical problems 1
1.2 Errors in quantitative analysis 2
1.3 Types of error 3
1.4 Random and systematic errors in titrimetric analysis 6
1.5 Handling systematic errors 8
1.6 Planning and design of experiments 11
1.7 Statistical calculations 12
Bibliography and resources 14
Exercises 15

2 Statistics of repeated measurements 16

2.1 Mean and standard deviation 16
2.2 The distribution of repeated measurements 18
2.3 Log-normal distribution 23
2.4 Definition of a 'sample' 24
2.5 The sampling distribution of the mean 25
2.6 Confidence limits of the mean for large samples 26
2.7 Confidence limits of the mean for small samples 27
2.8 Presentation of results 29
2.9 Other uses of confidence limits 30
2.10 Confidence limits of the geometric mean for a log-normal distribution 30
2.11 Propagation of random errors 31
2.12 Propagation of systematic errors 34
Bibliography 35
Exercises 35

3 Significance tests 37

3.1 Introduction 37
3.2 Comparison of an experimental mean with a known value 38
3.3 Comparison of two experimental means 39
3.4 Paired *t*-test 43
3.5 One-sided and two-sided tests 45
3.6 *F*-test for the comparison of standard deviations 47
3.7 Outliers 49
3.8 Analysis of variance 52
3.9 Comparison of several means 53
3.10 The arithmetic of ANOVA calculations 56
3.11 The chi-squared test 59
3.12 Testing for normality of distribution 61
3.13 Conclusions from significance tests 65
3.14 Bayesian statistics 67
Bibliography 70
Exercises 71

4 The quality of analytical measurements 75

4.1 Introduction 75
4.2 Sampling 76
4.3 Separation and estimation of variances using ANOVA 77
4.4 Sampling strategy 78
4.5 Introduction to quality control methods 79
4.6 Shewhart charts for mean values 80
4.7 Shewhart charts for ranges 82
4.8 Establishing the process capability 84
4.9 Average run length: CUSUM charts 87
4.10 Zone control charts (J-charts) 91
4.11 Proficiency testing schemes 93
4.12 Method performance studies (collaborative trials) 96
4.13 Uncertainty 100
4.14 Acceptance sampling 109
4.15 Method validation 112
Bibliography 116
Exercises 117

5 Calibration methods in instrumental analysis: regression and correlation 120

5.1 Introduction: instrumental analysis 120
5.2 Calibration graphs in instrumental analysis 122
5.3 The product–moment correlation coefficient 124
5.4 The line of regression of y on x 128
5.5 Errors in the slope and intercept of the regression line 129
5.6 Calculation of a concentration and its random error 131
5.7 Limits of detection 134

5.8 The method of standard additions 138
5.9 Use of regression lines for comparing analytical methods 140
5.10 Weighted regression lines 145
5.11 Intersection of two straight lines 150
5.12 ANOVA and regression calculations 152
5.13 Introduction to curvilinear regression methods 153
5.14 Curve fitting 156
5.15 Outliers in regression 160
Bibliography 162
Exercises 162

6 Non-parametric and robust methods 165

6.1 Introduction 165
6.2 The median: initial data analysis 166
6.3 The sign test 171
6.4 The Wald–Wolfowitz runs test 173
6.5 The Wilcoxon signed rank test 174
6.6 Simple tests for two independent samples 177
6.7 Non-parametric tests for more than two samples 180
6.8 Rank correlation 182
6.9 Non-parametric regression methods 183
6.10 Introduction to robust methods 186
6.11 Simple robust methods: trimming and winsorisation 187
6.12 Further robust estimates of location and spread 188
6.13 Robust ANOVA 190
6.14 Robust regression methods 191
6.15 Re-sampling statistics 193
6.16 Conclusion 194
Bibliography and resources 195
Exercises 196

7 Experimental design and optimisation 198

7.1 Introduction 198
7.2 Randomisation and blocking 200
7.3 Two-way ANOVA 201
7.4 Latin squares and other designs 204
7.5 Interactions 205
7.6 Identifying the important factors: factorial designs 211
7.7 Fractional factorial designs 216
7.8 Optimal experimental designs 219
7.9 Optimisation: basic principles and univariate methods 220
7.10 Optimisation using the alternating variable search method 223
7.11 The method of steepest ascent 225
7.12 Simplex optimisation 227
7.13 Simulated annealing 231
Bibliography and resources 232
Exercises 232

8 Multivariate analysis 235

8.1 Introduction 235
8.2 Initial analysis 237
8.3 Principal component analysis 238
8.4 Cluster analysis 242
8.5 Discriminant analysis 246
8.6 *K*-nearest neighbour (KNN) method 250
8.7 Disjoint class modelling 251
8.8 Regression methods 251
8.9 Multiple linear regression (MLR) 253
8.10 Principal components regression (PCR) 255
8.11 Partial least-squares (PLS) regression 257
8.12 Natural computation methods: artificial neural networks 260
8.13 Conclusion 261
Bibliography and resources 262
Exercises 262

Solutions to exercises 265

Appendix 1: Commonly used statistical significance tests 275

Appendix 2: Statistical tables 278

Index 287

Supporting resources

Visit **www.pearsoned.co.uk/miller** to find valuable online resources

For students
- Further exercises

For instructors
- Further exercises
- Complete Instructor's Manual
- PowerPoint slides of figures from the book

For more information please contact your local Pearson Education sales representative or visit **www.pearsoned.co.uk/miller**

Preface to the seventh edition

Since the publication of the sixth edition of this text in 2010 the importance of the proper application of statistical methods to chemical and biochemical measurements has continued to grow, as has the number and sophistication of the methods themselves. Enormous amounts of information on such methods, along with critical comparative assessments and examples of their application, are freely available from numerous websites. In this new edition we have tried to reflect some of these developments, while retaining from previous editions a pragmatic and, as far as possible, non-mathematical approach to statistical calculations. Many of the changes and additions have been made in response to readers and reviewers of the previous edition, and we are very grateful to them for their guidance.

Most importantly, the addition of a third author has allowed us to expand significantly the parts of Chapter 4 that deal with uncertainty estimation, through the Guide to the Estimation of Uncertainty in Measurement (GUM), and the validation of analytical methods. These topics are especially relevant in areas such as food analysis, forensic sciences and medical and environmental analysis, and are critical to the maintenance of the quality of analytical measurements. Other areas where the text has been more modestly expanded include the application of Bayesian methods, which are now increasingly being used in many areas (Chapter 3); testing for normality of distribution (Chapter 3); the estimation of limits of detection in calibration experiments (Chapter 5); the use of robust statistical methods (Chapter 6); and some additional material on experimental designs (Chapter 7). Major developments have also occurred in the application of multivariate methods (Chapter 8) but the constraints of space and the complexity of the methods have persuaded us to leave those topics untouched.

As in previous editions we have provided both worked examples and exercises for the reader. The former have again been implemented using Excel® or Minitab®, these being the programs perhaps most widely used in education and by practising scientists. Both are repeatedly updated; and add-ins (some free, others not) for Excel now provide a great number of advanced calculation methods.

Finally, we thank the Royal Society of Chemistry for permission to use data from papers published in *The Analyst*, and also the patient and expert staff at Pearson Education, especially Janey Webb, Agnibesh Das and Payal Rana.

James N. Miller
Jane C. Miller
Robert D. Miller
August 2017

Preface to the first edition

To add yet another volume to the already numerous texts on statistics might seem to be an unwarranted exercise, yet the fact remains that many highly competent scientists are woefully ignorant of even the most elementary statistical methods. It is even more astonishing that analytical chemists, who practise one of the most quantitative of all sciences, are no more immune than others to this dangerous, but entirely curable, affliction. It is hoped, therefore, that this book will benefit analytical scientists who wish to design and conduct their experiments properly, and extract as much information from the results as they legitimately can. It is intended to be of value to the rapidly growing number of students specialising in analytical chemistry, and to those who use analytical methods routinely in everyday laboratory work.

There are two further and related reasons that have encouraged us to write this book. One is the enormous impact of microelectronics, in the form of microcomputers and handheld calculators, on statistics: these devices have brought lengthy or difficult statistical procedures within the reach of all practising scientists. The second is the rapid development of new 'chemometric' procedures, including pattern recognition, optimisation, numerical filter techniques, simulations and so on, all of them made practicable by improved computing facilities. The last chapter of this book attempts to give the reader at least a flavour of the potential of some of these newer statistical methods. We have not, however, included any computer programs in the book – partly because of the difficulties of presenting programs that would run on all the popular types of microcomputer, and partly because there is a substantial range of suitable and commercially available books and software.

The availability of this tremendous computing power naturally makes it all the more important that the scientist applies statistical methods rationally and correctly. To limit the length of the book, and to emphasise its practical bias, we have made no attempt to describe in detail the theoretical background of the statistical tests described. But we have tried to make it clear to the practising analyst which tests are appropriate to the types of problem likely to be encountered in the laboratory. There are worked examples in the text, and exercises for the reader at the end of each chapter. Many of these are based on the data provided by research papers published in The Analyst. We are deeply grateful to Mr. Phil Weston, the Editor, for allowing us thus to make use of his distinguished journal. We also thank our

colleagues, friends and family for their forbearance during the preparation of the book; the sources of the statistical tables, individually acknowledged in the appendices; the Series Editor, Dr. Bob Chalmers; and our publishers for their efficient cooperation and advice.

J. C. Miller
J. N. Miller
April 1984

Publisher's acknowledgements

We are grateful to the following for permission to reproduce copyright material:

Figures

Portions of information contained in this publication/book as Figures in Chapter 4,7,8 are printed with permission of Minitab Inc. All such material remains the exclusive property and copyright of Minitab Inc. All rights reserved. This information may have been translated for your convenience from the original and official English language version, which can be found at www.minitab.com, or as embedded in Minitab Statistical Software. Minitab retains all rights therein, and Minitab disclaims any and all responsibility for any reliance by you upon the translated version, which you use at your own risk. In the event of any discrepancy the English language version shall control.

Tables

Tables 03A.2, 03A.3, 03A.4, 03A.11 from *Elementary Statistics Tables for all users of statistical techniques*, London; Boston: Allen & Unwin (Neave H.R. 1981) Copyright © 1981 Routledge, reproduced by permission of Taylor & Francis Books UK; Table 03A.5 from *Practical Statistics for the Analytical Scientist*, 2 ed., Royal Society of Chemistry (Ellison S.L.R, Barwick V.J. and Farrant T.J.D. 2009) p.208, Reproduced by permission of The Royal Society of Chemistry; Table 03A.6 from Statistical treatment for rejection of deviant values: Critical values of Dixon's "Q" parameter and related subrange ratios at the 95% confidence level, *Analytical Chemistry* 63(2), 139–46 (Rorabacher, D.B. 1991), Copyright © American Chemical Society 1991. Reprinted with permission from American Chemical Society; Table 03A.7 from *Elementary Statistics Tables* London; Boston: Allen & Unwin (Neave H.R. 1981) Copyright © 1981 Routledge, reproduced by permission of Taylor & Francis Books UK; Table 03A.8 from *Elementary Statistics Tables*, London; Boston: Allen & Unwin (Neave H.R. 1981) Copyright © 1981 Routledge, reproduced by permission of Taylor & Francis Books UK; Table 03A.12 from *Elementary Statistics Tables* London; Boston: Allen & Unwin (Neave, H.R. 1981) Copyright © 1981 Routledge, reproduced by permission of Taylor & Francis Books UK; Tables 03A.13 from *Elementary Statistics Tables*, London; Boston: Allen & Unwin

(Neave, H.R. 1981) Taylor & Francis, Copyright © 1981 Routledge, reproduced by permission of Taylor & Francis Books UK; Tables 03A.14 from *Elementary Statistics Tables*, London; Boston: Allen & Unwin (Neave, H.R. 1981) Taylor & Francis, Copyright © 1981 Routledge, reproduced by permission of Taylor & Francis Books UK; Table A.15 from The distribution of the largest of a set of estimated variances as a fraction of their total, *Annals of Human Genetics (London)*, 11, 47–52 (Cochran W.G. 1941), 1941 Blackwell Publishing Ltd/University College London. Version of Record online: 13 APR2011 published by permission of John Wiley & Sons doi:10.1111/j.1469-1809.1941.tb02271, John Wiley & Sons; Table 3.1 from A rapid quantitative assay of intact paracetamol tablets by reflectance near-infrared spectroscopy *Analyst*, 124, 163–167 (Trafford, A.D., Jee, R.D., Moffat, A.C. and Graham, P. 1999), Royal Society of Chemistry

Text

General Displayed Text on page 13 from Microsoft Excel® Used with permission from Microsoft; General Displayed Text on page 13 from Minitab® Statistical Software, MINITAB® and all other trademarks and logos for the Company's products and services are the exclusive property of Minitab Inc. All other marks referenced remain the property of their respective owners. See minitab.com for more information; General Displayed Text on page 13 from MATLAB®, The MathWorks, Inc.; General Displayed Text on page 13 from The Unscrambler® ©2017 CAMO Software AS; General Displayed Text on page 13 from Design-Expert ®, Copyright © Stat-Ease, Inc.; Extract on page 13 from Design-Ease®, Copyright © Stat-Ease, Inc.; General Displayed Text on page 13 from SciFinder®, American Chemical Society; General Displayed Text on page 14 from Microsoft Office®, Used with permission from Microsoft; Exercise 2.1 adapted from Spectrofluorimetric determination and thin-layer chromatographic identification of selenium in foods *Analyst*, 108, 505 (Moreno-Dominguez, T., Garcia-Moreno, C. and Marine-Font, A. 1983), Royal Society of Chemistry; Exercise 2.3 adapted from Determination of total mercury in hydrocarbons and natural gas condensate by atomic fluorescence spectrometry *Analyst*, 124, 185–189 (Shafawi, A., Ebdon, L., Foulkes, M., Stockwell, P. and Corns, W. 1999), Royal Society of Chemistry; Exercise 2.5 adapted from Determination of fluorene in sea-water by room temperature phosphorescence in organised media, *Analyst*, 123, 2217–2221 (Gonsalez, M.A. and Lopez, M.H., 1998), Royal Society of Chemistry; Exercise 3.2 adapted from Spectrophotometric study of the reaction of titanium(IV) with chlorophosphonazoI *Analyst* 108, 641–644 (Xing-chu, Q, and Ying-quen, Z. 1983), Royal Society of Chemistry; Example 3.2.1 adapted from Speciation of selenomethionine and selenourea using living bacterial cells, *Analyst*, 123, 919–927 (Aller, A.J. and Robles, L.C, 1998), Royal Society of Chemistry; Example 3.3.1 adapted from Classical wet ashing versus microwave-assisted attacks for the determination of chromium in plants *Analyst*, 1–4 (Sahuquillo, A., Rubio, R., and Rauret, G. 1999), Royal Society of Chemistry; Example 3.3.2 adapted from Determination of tin in organic matter by atomic-absorption spectrometry *Analyst*, 108, 109–115 (Analytical Methods Committe 1983), Royal Society of Chemistry; Example 3.3.3 adapted from Determination of thiol concentrations in haemolysate by resonance Raman spectrometry *Analyst*, 107, 195–199 (Banford, J.C., Brown, D.H., McConnell, A.A., McNeil, C.J., Smith, W.E., Hazelton, R.A., and Sturrock, R.D. 1982), Royal Society of Chemistry; Exercise 3.4 adapted from Modified gas-liquid

chromatographic method for determining bromide/total bromine in foodstuffs and soils *Analyst*, 108, 742–747 (Roughan, J.A., Roughan, P.A. and Wilkins, J.P.G. 1983), Royal Society of Chemistry; Exercise 3.5 adapted from Extraction and analysis of interstitial water from sandstone *Analyst* 107, 731–736 (Wheatstone, K.G. and Gelsthorpe, D. 1982), Royal Society of Chemistry; Exercise 3.6 from Simultaneous determination of urinary catecholamines and 5-hydroxyindoleamines by high-performance liquid chromatography with fluorescence detection *Analyst* 123, 307–311 (Yamaguchi, M., Ishida, J. and Yoshimura, M. 1998), Royal Society of Chemistry; Example 3.6.1 adapted from Determination of chemical oxygen demand of wastewaters without the use of mercury salts *Analyst*, 107, 1047–1053 (Ballinger, D., Lloyd, A., and Morrish, A. 1982), Royal Society of Chemistry; Exercise 3.8 adapted from Classical wet ashing versus microwave-assisted attacks for the determination of chromium in plants *Analyst* 124, 1–4 (Sahuquillo, A., Rubio, R. and Rauret, G. 1999), Royal Society of Chemistry; Exercise 3.10 adapted from Flow injection spectrophotometric determination of hydrogen peroxide using a crude extract of zucchini (Cucurbita pepo) as a source of peroxidase *Analyst* 123, 1809–1812 (da Cruz Vieira, I. 1998), Royal Society of Chemistry; Exercise 3.11 adapted from A rapid quantitative assay of intact paracetamol tablets by reflectance near-infrared spectroscopy *Analyst* 124, 163–167 (Trafford, A.D., Jee, R.D., Moffat, A.C. and Graham, P. 1999), Royal Society of Chemistry; Exercise 3.12 adapted from Distribution of zinc amongst human serum proteins determined by affinity chromatography and atomic-absorption spectrophotometry *Analyst* 108, 492–504 (Foote, J.W. and Delves, H.T. 1983), Royal Society of Chemistry; Exercise 3.13 adapted from Flame atomic-absorption spectroscopic determination of antimony in atmospheric particulates by using direct atomisation of the covalent hydride *Analyst* 107, 1488–1492 (Castillo, J.R., Lanaja, J., Marinez, M.C. and Aznarez, J. 1982), Royal Society of Chemistry; Exercise 5.8 adapted from Sulphide ion-selective electrode studies concerning Desulfovibrio species of sulphate-reducing bacteria *Analyst* 108, 43–52 (Al-Hitti, I.K., Moody, G.J. and Thomas, J.D.R. 1983), Royal Society of Chemistry; Exercise 5.9 adapted from Determination of lead in whole blood by electrothermal atomic-absorption spectrometry using graphite probe atomisation *Analyst* 108, 244–253 (Giri, S.K., Shields, C.K., Littlejohn D. and Ottaway, J.M. 1983), Royal Society of Chemistry; Example 5.9.1 adapted from Fluorimetric determination of phytic acid based on the activation of the oxidation of 2,2'-dipyridyl ketone hydrazone catalysed by Cu(II) *Analyst*, 124, 897–900 (March, J.G., Simonet, B.M., and Grases, F. 1999); Exercise 5.10 adapted from Comparative study of the luminescent properties of europium and terbium coordinated with thenoyltrifluoroacetone or pyridine-2,6-dicarboxylic acid in aqueous solutions *Analyst*, 123, 261–265 (Arnaud, N., Vaquer, E. and Georges, J. 1998), Royal Society of Chemistry; Exercise 5.11 adapted from Improved method for measuring hydrolyzable tannins using potassium iodate *Analyst* 123, 435–439 (Willis, R.B. 1998), Royal Society of Chemistry; Exercise 5.12 adapted from Effect of non-ionic surfactants as mobile phase additives on the fluorescence intensity of dansyl derivatives of biogenic amines in high-performance thin-layer chromatography *Analyst*, 123, 725–729 (Linares, R.M., Ayala, J.H. 1998), Royal Society of Chemistry; General Displayed Text on page 196 from Mac® is a registered trademark of Apple Inc. registered in the U.S. and other countries; Exercise 7.2 adapted from Comparison of microwave-assisted extraction and Soxhlet extraction for phenols in soil samples using experimental designs *Analyst* 123, 1679–1684 (Egizabal, A., Zuloaga, O., Extebarria, N., Fernández, L.A. and Madariaga, J.M. 1998), Royal Society of Chemistry; Exercise 7.3 adapted from Fluorimetric flow-injection method for

anionic surfactants based on protein–surfactant interactions, *Analyst* 123, 2257–2261 (Recalde Ruiz, D.L., Carvalho Torres, A.L., Andrés Garcia, E. and Díaz García, M.E. 1998), Royal Society of Chemistry; Exercise 7.4 adapted from Apparent and real reducing ability of polypropylene in cold-vapour atomic-absorption spectrophotometric determinations of mercury *Analyst*, 107, 179–184 (Kuldvere, A. 1982), Royal Society of Chemistry; Exercise 8.2 adapted from Elemental content of Vietnamese rice Part 1. Sampling, analysis and comparison with previous studies *Analyst*, 124, 553–560 (T.D.Phuong T.D., Choung P.V., Khiem D.T.and Kokot S. 1999), Royal Society of Chemistry.

Glossary of symbols

a	–	intercept of regression line
b	–	gradient of regression line
c	–	number of columns in two-way ANOVA
C	–	correction term in two-way ANOVA
C	–	used in Cochran's text for homogeneity of variance
d	–	difference between paired vaues
F	–	the ratio of two variances
G	–	used in Grubbs' test for outliers
h	–	number of samples in one-way ANOVA
k	–	coverage factor in uncertainty estimates
μ	–	arithmetic mean of a population
M	–	number of minus signs in Wald–Wolfowitz runs test
n	–	sample size
N	–	number of plus signs in Wald–Wolfowitz runs test
N	–	total number of measurements in two-way ANOVA
v	–	number of degrees of freedom
$P(r)$	–	probability of r
Q	–	Dixon's Q, used to test for outliers
r	–	product–moment correlation coefficient
r	–	number of rows in two-way ANOVA
r	–	number of smallest and largest observations omitted in trimmed mean calculations
R^2	–	coefficient of determination
R'^2	–	adjusted coefficient of determination
r_s	–	Spearman rank correlation coefficient
s	–	standard deviation of a sample
s^2	–	variance of a sample
$s_{y/x}$	–	standard deviation of y-residuals
s_b	–	standard deviation of slope of regression line
s_a	–	standard deviation of intercept of regression line
$s_{(y/x)w}$	–	standard deviation of y-residuals of weighted regression line
s_{x_0}	–	standard deviation of x-value estimated using regression line
s_B	–	standard deviation of blank
s_{x_E}	–	standard deviation of extrapolated x-value

$s_{x_{0w}}$ – standard deviation of x-value estimated by using weighted regression line
σ – standard deviation of a population
σ_0^2 – measurement variance
σ_1^2 – sampling variance
t – quantity used in the calculation of confidence limits and in significance testing of mean (see Section 2.4)
T – grand total in ANOVA
T_1 and T_2 – test statistics used in the Wilcoxon rank sum test
u – standard uncertainty
U – expanded uncertainty
w – range
w_i – weight given to point on regression line
$\overline{x}$ – arithmetic mean of a sample
x_0 – x-value estimated by using regression line
x_0 – outlier value of x
$\widetilde{x}_i$ – pseudo-value in robust statistics
x_E – extrapolated x-value
$\overline{x}_w$ – arithmetic mean of weighted x-values
X^2 – quantity used to test for goodness-of-fit
$\hat{y}$ – y-values predicted by regression line
y_0 – signal from test material in calibration experiments
$\overline{y}_w$ – arithmetic mean of weighted y-values
y_B – signal from blank
z – standard normal variable

1 Introduction

Major topics covered in this chapter

- **Errors in analytical measurements**
- **Gross, random and systematic errors**
- **Precision, repeatability, reproducibility, bias, accuracy, trueness**
- **Planning experiments**
- **Statistical calculations**

1.1 Analytical problems

Analytical scientists face both qualitative and quantitative problems. For example, the presence of boron in distilled water is very damaging in the manufacture of electronic components, so we might be asked: 'Does this distilled water sample contain any boron?' The comparison of soil samples in forensic science provides another qualitative problem: 'Could these two soil samples have come from the same site?' Other problems are single- or multi-component quantitative ones: 'How much albumin is there in this sample of blood serum?' 'This steel sample contains small amounts of chromium, tungsten and manganese – how much of each?'

Modern analytical chemistry is almost always a **quantitative** science, as a numerical result will generally be much more valuable than a qualitative one. Only by finding *how much* boron is present in a water sample can we decide whether its level is worrying, or how it might be reduced. Sometimes *only* a quantitative result has any value: almost all samples of blood serum contain albumin, so the only question is – how much? Two soil samples might be compared using particle size analysis, in which the proportions of the soil particles falling within a number, say ten, of particle-size ranges are determined. Each sample would then be characterised by these ten pieces of data, which can be used (see Chapter 8) to provide a quantitative rather than just a qualitative assessment of their similarity.

1.2 Errors in quantitative analysis

Since quantitative methods will be the norm in an analytical laboratory, we must accept that the errors that occur in such methods are crucially important. Our guiding principle will be that no *quantitative results are of any value unless they are accompanied by some estimate of the errors inherent in them.* (This principle naturally applies to any field of study in which numerical experimental results are obtained.) Several examples illustrate this idea, and they also introduce some types of statistical problem that we shall meet and solve in later chapters.

Suppose we synthesise an analytical reagent which we believe to be entirely new. We study it using a spectrometric method and it gives a value of 104 (normally our results will be given in proper units, but in this hypothetical example we use purely arbitrary units). On checking the reference books, we find that no compound previously discovered has given a value above 100 when studied by the same method in the same experimental conditions. So have we really discovered a new compound? The answer clearly depends on the errors associated with that experimental value of 104. If further work suggests that the result is correct to within 2 units, i.e. the true value probably lies in the range 104 ± 2, then a new compound has probably been discovered. But if the error may amount to 10 units (i.e. 104 ± 10), then it is quite likely that the true value is actually less than 100, in which case a new discovery is far from certain. So our knowledge of the experimental errors is crucial (in this and every other case) to the correct interpretation of the results. Statistically this example involves the comparison of our experimental result with an assumed or reference value: this topic is studied in detail in Chapter 3.

Analysts commonly perform several replicate determinations in the course of a single experiment. (The value and significance of such replicates is discussed in detail in the next chapter.) Suppose we perform a titration four times and obtain values of 24.69, 24.73, 24.77 and 25.39 ml. All four values are different, because of the errors and variations inherent in the measurements, and the fourth value (25.39 ml) is substantially different from the other three. Can this fourth value be safely rejected, so that the mean result is reported as 24.73 ml, the mean of the other three readings? In statistical terms, is the value 25.39 ml an *outlier*? The major topic of outlier rejection is discussed in detail in Chapters 3 and 6.

Another frequent problem involves the comparison of two (or more) sets of results. Suppose we measure the vanadium content of a steel sample by two separate methods. With the first method the average value obtained is 1.04%, with an estimated error of 0.07%, while for the second method the average value is 0.95%, with an error of 0.04%. Several questions then arise. Are the two average values significantly different, or are they indistinguishable within the limits of the experimental errors? Is one method significantly less error-prone than the other? Which of the mean values is actually closer to the truth? Again, Chapter 3 discusses these and related questions.

Many instrumental analyses are based on graphical methods. Instead of making repeated measurements on the same sample, we perform a series of measurements on a small group of *standards* containing known analyte concentrations covering a considerable range. The results yield a calibration graph that is then used to estimate by interpolation the concentrations of *test samples* ('unknowns') studied by the same procedure. All the measurements on the standards and on the test samples will be

subject to errors, so we must assess the errors involved in drawing the calibration graph, and the error in the concentration of a single sample determined using the graph. We can also estimate the limit of detection of the method, i.e. the smallest quantity of analyte that can be detected with a given degree of confidence. These and related methods are described in Chapter 5.

These examples show just a few of the problems arising from experimental errors in quantitative analysis. We must next study the various types of error in more detail.

1.3 Types of error

Experimental scientists distinguish three types of error – **gross**, **random** and **systematic** errors. Gross errors are so serious that there is no alternative to abandoning the experiment and making a completely fresh start. Examples include a complete instrument breakdown, accidentally dropping or discarding a crucial sample, or discovering during the course of the experiment that a supposedly pure reagent was in fact badly contaminated. Such errors (which occur even in the best laboratories!) are normally easily recognised.

We can make the distinction between random and systematic errors by studying a real experimental situation. Four students (A–D) each perform an analysis in which *exactly* 10.00 ml of *exactly* 0.1 M sodium hydroxide is titrated with *exactly* 0.1 M hydrochloric acid. Each student performs five replicate titrations, with the results shown in Table 1.1.

Student A's results have two characteristics. First, they are all very close to each other; all the results lie between 10.08 and 10.12 ml. In everyday terms we would say that the results are highly *repeatable.* Their second feature is that they are *all too high:* in this experiment (rather unusually) we know that the correct result should be exactly 10.00 ml. Evidently two entirely separate types of error have occurred. First, there are **random errors** – *these cause replicate results to differ from one another, so that the individual results fall on both sides of the average value* (10.10 ml in this case). Random errors affect the **precision**, or **repeatability**, of an experiment. In the case of student A the random errors are evidently small, so we say that the results are **precise**. In addition, however, there are **systematic errors** – *these cause all the results to be in error in the same direction* (in this case they are all too high). The total systematic error (in a given experiment there may be several sources of systematic error, some positive and others negative; see Chapter 2) is called the **bias** of the measurement. (The opposite of bias, or

Table 1.1 Data demonstrating random and systematic errors

Student	Results (ml)					**Average (ml)**	Comment
A	10.08	10.11	10.09	10.10	10.12	**10.10**	Precise, biased
B	9.88	10.14	10.02	9.80	10.21	**10.01**	Imprecise, unbiased
C	10.19	9.79	9.69	10.05	9.78	**9.90**	Imprecise, biased
D	10.04	9.98	10.02	9.97	10.04	**10.01**	Precise, unbiased

Table 1.2 Random and systematic errors

Random errors	Systematic errors
Affect **precision** – repeatability or reproducibility	Produce **bias** – an overall deviation of a result from the true value even when random errors are very small
Cause replicate results to fall on either side of an average value	Cause all results to be affected in one sense only, all too high or all too low
Can be estimated using replicate measurements	Cannot be detected simply by using replicate measurements
Can be minimised by good technique but not eliminated	Can be corrected, e.g. by using standard methods and materials
Caused by both humans and equipment	Caused by both humans and equipment

lack of bias, is sometimes referred to as **trueness** of a method: see Section 4.15.) The random and systematic errors here are easily distinguished using the experimental results, and may also have quite distinct causes in terms of techniques and equipment (see Section 1.4). The data obtained by student B are in direct contrast to those of student A. The average of B's five results (10.01 ml) is very close to the true value, so there is no evidence of bias, but the spread of the results is very large, indicating poor precision, i.e. substantial random errors. Comparison of these results with those of student A shows clearly that random and systematic errors can occur independently of one another. This is also shown by the data of students C and D. Student C's work has poor precision (range 9.69–10.19 ml) and the average result (9.90 ml) is (negatively) biased. Student D has achieved both precise (range 9.97–10.04 ml) and unbiased (average 10.01 ml) results. The distinction between random and systematic errors is summarised in Table 1.2, and in Fig. 1.1 as a series of *dot-plots*. This simple graphical method of displaying data, in which individual results are plotted as dots on a linear scale, is often used in exploratory data analysis (EDA, also called initial data analysis, IDA: see Chapters 3 and 6).

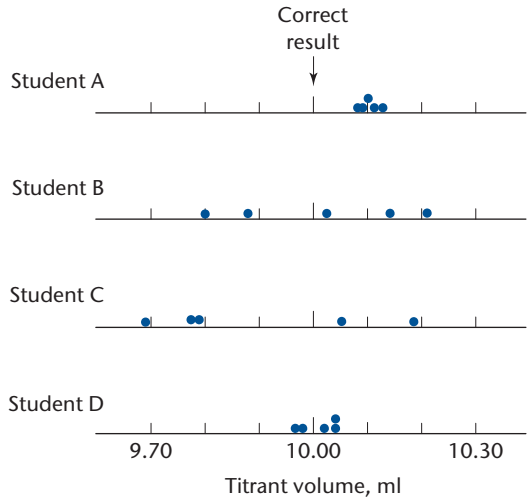

Figure 1.1 Bias and precision: dot-plots of the data in Table 1.1.

In many analytical experiments the most important question is – how far is our result from the true value of the concentration or amount that we are trying to measure? This is expressed as the **accuracy** of the experiment. Accuracy is defined by the International Organization for Standardization (ISO) as 'the closeness of agreement between a test result and the accepted reference value' of the analyte. The accuracy of a single result may thus be affected by *both* random and systematic errors. The accuracy of an average result also has contributions from both error sources: even if systematic errors are absent, the average result will probably not equal the reference value exactly, because of the occurrence of random errors (see Chapters 2 and 3). The results from student B demonstrate this. Four of B's five measurements show significant inaccuracy, i.e. are well removed from the true value of 10.00. But the average of the results (10.01) is very accurate, so it seems that the inaccuracy of the individual results is due largely to random errors and not to systematic ones. By contrast, all of student A's individual results, and the resulting average, are inaccurate: given the good precision of A's work, it seems that these inaccuracies are due to systematic errors. Note that, contrary to the implications of many dictionaries, accuracy and precision have entirely different meanings in the study of experimental errors.

In summary, precision describes random error, bias describes systematic error and the accuracy, i.e. closeness to the true value of a single measurement or a mean value, incorporates both types of error.

Another important area of terminology is the difference between **reproducibility** and **repeatability**. We can illustrate this using the students' results again. In the normal way each student would do the five replicate titrations in rapid succession, taking only an hour or so. The same set of solutions and the same glassware would be used throughout, the same preparation of indicator would be added to each titration flask, and the temperature, humidity and other laboratory conditions would remain much the same. In such cases the precision for each student would be the *within-run* precision: this is called the **repeatability**. Suppose, however, that for some reason the titrations were performed by each student on five different occasions in different laboratories, using different pieces of glassware and different batches of indicator. It would not be surprising to find a greater spread of the results in this case. The resulting data would reflect the *between-run* precision of the method, i.e. its **reproducibility**.

- **Repeatability** describes the precision of within-run replicates.
- **Reproducibility** describes the precision of between-run replicates.
- The reproducibility of a method is normally expected to be poorer (i.e. with larger random errors) than its repeatability.

One further lesson may be learned from the titration experiments. Obviously the data obtained by student C are unacceptable, and those of student D are the best. Sometimes, however, two methods may be available for a particular analysis, one of which is believed to be precise but biased, and the other imprecise but without bias. In other words we may have to choose between the types of results obtained by students A and B respectively. Which type of result is preferable? A dogmatic answer to this

question is impossible: in practice the choice between the methods will often be based on their costs, ease of automation, speed, and so on. But we must realise that a method substantially free from systematic errors may still, if it is very imprecise, give an average value that is (by chance) a long way from the correct value. On the other hand a precise but biased method (e.g. student A) can be converted into one that is both precise *and* unbiased (e.g. student D) *if the systematic errors can be discovered and hence removed*. Random errors can never be eliminated, though by careful technique we can minimise them, and by making repeated measurements we can measure them and evaluate their significance. Systematic errors can often be removed by careful checks on our experimental technique and equipment. This distinction between the two major types of error is further explored in the next section.

When a laboratory is supplied with a sample and requested to determine the concentrations of one of its constituents, it will estimate, or perhaps know from previous experience, the extent of the major random and systematic errors occurring. The customer supplying the sample may well want this information incorporated in a single statement, giving the *range within which the true concentration is reasonably likely to lie*. This range, which should be given with a probability (e.g. 'it is 95% probable that the concentration lies between . . . and . . .'), is called the **uncertainty** of the measurement. Uncertainty estimates are now very widely used in analytical chemistry and are discussed in more detail in Chapter 4.

1.4 Random and systematic errors in titrimetric analysis

The students' titration experiments showed that random and systematic errors can occur independently of one another, and thus presumably arise at different stages of an experiment. A complete titrimetric analysis can be summarised by the following steps:

1 Making up a standard solution of one of the reactants. This involves (a) weighing a weighing bottle or similar vessel containing some solid material, (b) transferring the solid material to a standard flask and weighing the bottle again to obtain by subtraction the weight of solid transferred (weighing *by difference*), and (c) filling the flask up to the mark with water (assuming that an aqueous titration is to be used).

2 Transferring an aliquot of the standard material to a titration flask by filling and draining a pipette properly.

3 Titrating the liquid in the flask with a solution of the other reactant, added from a burette. This involves (a) filling the burette and allowing the liquid in it to drain until the meniscus is at a constant level, (b) adding a few drops of indicator solution to the titration flask, (c) reading the initial burette volume, (d) adding liquid to the titration flask from the burette until the end point is adjudged to have been reached, and (e) measuring the final level of liquid in the burette.

So the titration involves some ten separate steps, the last seven of which are normally repeated several times, giving replicate results. In principle, we should examine each step to evaluate the random and systematic errors that might occur. In practice, it is simpler to examine separately those stages which utilise weighings (steps 1(a) and 1(b)), and the remaining stages involving the use of volumetric equipment. (It is not

intended to give detailed descriptions of the techniques used in the various stages. Similarly, methods for calibrating weights, glassware, etc. will not be given.) The tolerances of weights used in the gravimetric steps, and of the volumetric glassware, may contribute significantly to the experimental errors. Specifications for these tolerances are issued by such bodies as the British Standards Institute (BSI) and the American Society for Testing and Materials (ASTM). If a weight or a piece of glassware is within the tolerance limits, but not of exactly the correct weight or volume, a systematic error will arise. Thus, if the standard flask actually has a volume of 249.95 ml, this error will be reflected in the results of all the experiments based on the use of that flask. Repetition will not reveal the error: in each replicate the volume will be assumed to be 250.00 ml when it is actually less.

Weighing procedures are normally associated with very small *random* errors. In routine laboratory work a 'four-place' balance is commonly used, and the random error involved should not be greater than ca. 0.0002 g (the next chapter describes the terms used to express random errors). Since the quantity being weighed is normally about 1 g or more, the random error as a percentage of that weight is not more than 0.02%. A good standard material for volumetric analysis should (amongst other properties) have as high a formula weight as possible, to minimise these random weighing errors when a solution of a specified molarity is being made up.

Systematic errors in weighings can be significant, arising from adsorption of moisture on the surface of the weighing vessel; corroded or contaminated weights; and the buoyancy effect of the atmosphere, acting to different extents on objects of different density. Simple experimental precautions can be taken to minimise these systematic weighing errors. Weighing by difference (see above) cancels systematic errors arising from moisture and other contaminants on the surface of the bottle. (See also Section 2.12.) *If such precautions are taken*, the errors in the weighing steps will be small, and in most volumetric experiments weighing errors will probably be negligible compared with the volumetric ones. Hence gravimetric methods are usually used for the calibration of items of volumetric glassware, by weighing (in standard conditions) the water that they contain or deliver; and standards for top-quality calibration experiments (Chapter 5) are made up by weighing rather than volume measurements.

Most of the *random* errors in volumetric procedures arise in the use of volumetric glassware. In filling a 250 ml standard flask to the mark, the error (i.e. the distance between the meniscus and the mark) might be about ±0.03 cm in a flask neck of diameter ca. 1.5 cm. This corresponds to a volume error of about 0.05 ml – only 0.02% of the total volume of the flask. The error in reading a burette (the conventional type graduated in 0.1ml divisions) is perhaps 0.01–0.02 ml. Each titration involves two such readings (the errors of which are *not* simply additive – see Chapter 2); if the titration volume is ca. 25 ml, the percentage error is again very small. The experiment should be arranged so that the volume of titrant is not too small (say not less than 10 ml), otherwise such errors may become appreciable. (This precaution is analogous to choosing a standard compound of high formula weight to minimise the weighing error.) Even though a volumetric analysis involves several steps, each involving a piece of volumetric glassware, the random errors should evidently be small if the experiments are performed with care. In practice a good volumetric analysis should have a relative standard deviation (see Chapter 2) of not more than about 0.1%. Until fairly recently such precision was not normally attainable in instrumental analysis methods, and it is still not very common.

Volumetric procedures incorporate several important sources of systematic error. Perhaps the commonest error in routine volumetric analysis is to fail to allow enough time for a pipette to drain properly, or a meniscus level in a burette to stabilise. The temperature at which an experiment is performed has two effects. Volumetric equipment is conventionally calibrated at 20 °C, but the temperature in an analytical laboratory may differ from this by several degrees, and many experiments in biochemical analysis are carried out in 'cold rooms' at ca. 4 °C. The temperature affects both the volume of the glassware and the density of liquids.

Systematic indicator errors can be quite substantial, perhaps larger than the random errors in a typical titrimetric analysis. In the titration of 0.1 M hydrochloric acid with 0.1 M sodium hydroxide, we expect the end point to correspond to a pH of 7. In practice, however, we estimate this end point using an indicator such as methyl orange, which changes colour over the pH range ca. 3–4. So if the titration is performed by adding alkali to acid, the indicator will give an apparent end point when the pH is ca. 3.5, i.e. just before the true end point. The error can be evaluated and corrected by doing a *blank* experiment, i.e. by determining how much alkali is required to produce the indicator colour change in the *absence* of the acid.

It should be possible to consider and estimate the sources of random and systematic error arising at each distinct stage of any analytical experiment. It is very important to do this, so as to avoid major sources of error by careful experimental design (Sections 1.5 and 1.6). In many analyses (though not normally in titrimetry) the overall error is in practice dominated by the error in a single step: this point is further discussed in Chapter 2.

1.5 Handling systematic errors

Much of the rest of this text will deal with the handling of random errors, using a wide range of statistical methods. Usually we shall assume that systematic errors are absent (though methods which test for their occurrence will be described). But first we must discuss in more detail how systematic errors arise, and how they may be countered. The example of a titrimetric analysis given above showed that systematic errors cause the mean value of a set of replicate measurements to deviate from the true value. We also learned that systematic errors cannot be revealed merely by making repeated measurements, and that unless the true result of an analysis is known in advance – an unlikely situation! – very large systematic errors might go entirely undetected unless suitable precautions are taken. A few examples will clarify both the possible problems and their solutions.

The levels of transition metals in biological samples such as blood serum are important in biomedical studies. For many years determinations were made of the serum levels of chromium – with some startling results. Different workers, all studying pooled serum samples from healthy subjects, published concentrations varying from <1 to ca. 200 ng ml^{-1}. The lower results were mostly obtained later than the higher ones, and it gradually became apparent that the earlier values were due at least partly to contamination of the samples by chromium from stainless steel syringes, tube caps, and so on. The determination of traces of chromium, e.g. by atomic-absorption spectrometry, is relatively straightforward, and no doubt each study achieved results with

satisfactory precision; but in a number of cases the large systematic error introduced by the contamination was entirely overlooked. Similarly the normal levels of iron in seawater are now known to be in the parts per billion (ng ml^{-1}) range, but until fairly recently the concentration was thought to be much higher, perhaps tens of μg ml^{-1}. This misconception arose from the practice of sampling and analysing seawater in ship-borne environments containing high ambient iron levels. Methodological systematic errors of this kind are extremely common.

Systematic errors also occur widely when false assumptions are made about the accuracy of an analytical instrument. A monochromator in a spectrometer may gradually go out of adjustment, so that errors of several nanometres in wavelength settings arise, yet many photometric analyses are undertaken without appropriate checks being made. Commonplace devices such as volumetric glassware, stopwatches, pH meters and thermometers can all show substantial systematic errors, but many laboratory workers use them as though they are without bias. Many instrumental analysis systems are now computer controlled, minimising the number of steps and the skill levels required in their use. It is tempting to regard results from such instruments as beyond reproach, but (unless the devices are 'intelligent' enough to be self-calibrating – see Section 1.7) they are still subject to systematic errors.

Systematic errors arise not only from procedures or apparatus: they can also arise from human bias. Some chemists suffer from astigmatism or colour-blindness (the latter is more common among men than women) which might introduce errors in their readings of instruments and in other observations. A number of authors have reported various types of number bias, such as a tendency to favour even over odd numbers, or 0 and 5 over other digits, in the reporting of results. In short, systematic errors of several kinds are a constant, and often hidden, risk for the analyst, so very careful steps must be taken to minimise them.

There are several approaches to this problem, and any or all of them should be considered in each analytical procedure. The first precautions should precede any experimental work. The analyst should consider carefully each stage of the proposed procedure, the apparatus to be used and the sampling and analytical protocols to be adopted. The likely sources of systematic error, such as the instrument functions that need calibrating, the steps of the procedure where errors are most likely to occur, and the checks that can be made during the analysis, must be identified. Such foresight can be very valuable and is normally well worth the time invested. A little thinking of this kind should have revealed the possibility of contamination in the serum chromium determinations described above.

The second line of defence against systematic errors lies in the design of the experiment at every stage. We have seen (Section 1.4) that weighing by difference can remove some systematic gravimetric errors: these can be assumed to occur to the same extent in both weighings, so the subtraction process eliminates them. Another example of careful experimental planning is provided by the spectrometer wavelength error described above. If the concentration of a single substance is to be determined by absorption spectrometry, two approaches are possible. In the first, the sample is studied in a 1 cm pathlength spectrometer cell at a single wavelength, say 400 nm, and the concentration of the test component is determined from the well-known equation $A = \varepsilon bc$, where A, ε, b and c are respectively the measured absorbance, a reference value of the molar absorptivity (units l $mole^{-1}$ cm^{-1}) of the test substance, the pathlength (cm) of the spectrometer cell, and the molar concentration of this substance. Several systematic errors can arise here. The wavelength might actually be (say) 405 nm rather