Design and Analysis in Chemical Research

Design and Analysis in Chemical Research

Edited by

ROY L. TRANTER

Statistics and Data Evaluation Manager
QA Compliance Group
Glaxo Wellcome Operations
County Durham, UK

CRC Press

First published 2000

Published by
Sheffield Academic Press Ltd
Mansion House, 19 Kingfield Road
Sheffield S11 9AS, England

ISBN 1-85075-994-4

Published in the U.S.A. and Canada (only) by
CRC Press LLC
2000 Corporate Blvd., N.W.
Boca Raton, FL 33431, U.S.A.
Orders from the U.S.A. and Canada (only) to CRC Press LLC

U.S.A. and Canada only:
ISBN 0-8493-9746-4

Printed on acid-free paper in Great Britain by
Bookcraft Ltd, Midsomer Norton, Bath

British Library Cataloguing-in-Publication Data:
A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data:
Design and analysis in chemical research / edited by Roy Tranter.
p. cm. -- (Sheffield analytical chemistry : v. 3)
Includes bibliographical references and index.
ISBN 0-8493-9746-4 (alk. paper)
1. Chemistry, Analytic --Statistical methods. 2.Experimental design. I. Tranter, Roy. II. Series.
QD75.4.S8D48 1999
543'.07'2--dc21 99-28562
CIP

Design and Analysis in Chemical Research

Sheffield Analytical Chemistry

Series Editors: J.M. Chalmers and R.N. Ibbett

A series which presents the current state of the art of chosen sectors of analytical chemistry. Written at professional and reference level, it is directed at analytical chemists, environmental scientists, food scientists, pharmaceutical scientists, earth scientists, petrochemists and polymer chemists. Each volume in the series provides an accessible source of information on the essential principles, instrumentation, methodology and applications of a particular analytical technique.

Titles in the Series:

Inductively Coupled Plasma Spectrometry and its Applications
Edited by S.J. Hill

Extraction Methods in Organic Analysis
Edited by A.J. Handley

Design and Analysis in Chemical Research
Edited by R.L. Tranter

Spectroscopy in Process Analysis
Edited by J.M. Chalmers

Preface

Within the chemical sciences, statistics has the reputation of being hard and usable only by mathematicians or masochists. It also has the reputation of either telling us what we already know or making predictions that are wrong. So why should we bother with it?

Both of these reputations are undeserved and often stem from a dry theoretical statistics course or from experience of the statistics used publicly for essentially political aims. But 'real' statistics is quite different. It is the aim of this book to show that it is essentially an extension of the logical processes used by chemists every day, and that its use can, and does, bring greater understanding of problems more quickly and easily than the purely intuitive or "let's try and see" approaches.

For this we must be careful to distinguish between the tools we use to make the statistical calculations—the equations, algorithms and software—and the thought processes that allow us to decide which is the best tool to use and which is the best method for interpreting the results of its use. The latter is best described as *statistical thinking*. It encompasses the tools but extends the context to include awareness and appreciation of the sciences of measurement, experimentation and logic. It is the philosophy of rational data analysis and interpretation.

Statistics is a mathematical subject and it does involve equations—we cannot get away from this—but the equations are generally no more difficult than those routinely used in spectroscopy, kinetics, structure-activity relationships, molecular modelling or any other chemical system requiring computation. Many of the statistics equations (and concepts) are much simpler! What is different about statistics is that it is all about handling variability and uncertainty. Most chemists are more comfortable with certainty and, for many, the concept of error is associated with mistakes and poor work.

Simply accepting that all measurements have some uncertainty associated with them is the first, major step in coming to terms with statistics. The second is accepting that uncertainty can be measured and handled in a quantitative way. Once we can get a handle on uncertainty, we can control it. Although it is not possible, in any practical sense, to remove uncertainty, we can certainly reduce it and its effects, and so increase our confidence in the chemical truths that we discover.

Statistics is a broad subject and it has very wide application in the sciences, engineering, financial and behavioural arenas. It can be presented in many ways, all good in their own contexts. Here we have chosen to concen-

trate on principles and interpretation rather than on formal derivation and proof. References are given for those wanting to get into the latter. All we need to be aware of here is that all the methods we describe are well established, well documented and widely accepted. This does not mean that they give the "truth"—only that their properties are well understood and that they will give consistent and interpretable results when used appropriately.

You are not entering an equation-free zone. Far from it. But the equations needed to understand and/or to implement a particular tool are given with explanation and interpretation to help you come to terms with them.

One of the best ways of understanding how or where to use statistics is through applications. We have included many examples of actual or possible use from a wide area. The context of research chemistry is used throughout, but many of the good, easy-to-explain examples come from analytical and process chemistry, where quantitative measurement and interpretation are the norm. However, remember our objective of establishing principles. The principles in these examples apply to all branches of research chemistry, including organic and inorganic synthesis and molecular design, as well as the more obvious topics of physical chemistry and chemical physics.

In a book of this size, it is impossible to cover the whole of the vast subjects of statistics and chemometrics. So we have chosen to cover the basic statistical methods that underpin the *statistical thinking* approach. These are chapters 1-8. Chapters 9-13 describe the tools that are frequently used in chemical situations where a quantitative model is needed to describe or test a relationship between variables. Chapter 2 focuses on data quality, as we can have no confidence in statistical results if we have no confidence in the data, no matter how impressive the calculations might be.

These days, the computation of statistics is a trivial task. However, the easy availability of good hardware and software does mean that it is very much easier to do statistical computation without actually understanding what is being done. The corollary is that you have more time to develop that understanding and to consider the interpretation of the results that you have calculated. If you do get stuck or want advice on how to get away from the 'black box' approach, there are many quite friendly and helpful statisticians out there.

Finally to the authors. They are all people well known and well respected in their areas. Some are professionally trained statisticians and some are chemists with a deep understanding and appreciation of statistics. All are practical users of statistics and have wide experience of using statistics and the principles of *statistical thinking* in many areas of chemistry. I am greatly indebted to them for the time and effort they have put in to writing their chapters, and for the patience they have shown to me as editor of this volume. You will find much to appreciate in their work. Enjoy it and apply it!

Roy Tranter

Contributors

Dr M. Robert Alecio	Director, Positive Probability Limited, 9 Church Street, Isleham, Ely, Cambridgeshire CB7 5RX, UK Email: Robert@Alecio.freeserve.co.uk
Mr Anthony G. Ferrige	Director, Positive Probability Limited, 9 Church Street, Isleham, Ely, Cambridgeshire CB7 5RX, UK Email: gfv63@dial.pipex.com
Marion Gerson	Director, Centre for Quality Engineering, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK Email: m.e.gerson@ncl.ac.uk
Mrs Sonya Godbert	Senior Consultant Statistician, Statistical Services, Glaxo Wellcome Research and Development, Park Road, Ware SG12 0DP, UK
Professor Theodora Kourti	Department of Chemical Engineering, McMaster University, Hamilton, Ontario L8S 4L7, Canada Email: kourtit@mcmaster.ca
Professor Olav M. Kvalheim	Department of Chemistry, The University of Bergen, Allegate 41, N-5007 Bergen, Norway Email: olav.kvalheim@kj.uib.no
Dr Ivan Langhans	CQ Consultancy, Kapeldreef 60, B-3001 Heverlee, Belgium Email: CQ@CQConsultancy.be
Dr Willem Melssen	Department of Analytical Chemistry, Katholieke Universität Nijmegen, Postbus 9010, 6500 GL Nijmegen, The Netherlands Email: willem@sci.kun.nl

Dr Max A. Porter Statistics Manager, Glaxo Wellcome UK International Actives Supply, North Lonsdale Road, Ulverston, Cumbria LA12 9DR, UK
Email: map41247@GlaxoWellcome.co.uk

Dr John M. Thompson School of Mathematics and Statistics, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
Email: dr.jmthompson@cwcom.net
jmt@for.mat.bham.ac.uk

Dr Roy L. Tranter Statistics and Data Evaluation Manager, QA Compliance, Glaxo Wellcome UK International Product Supply, Harmire Road, Barnard Castle, County Durham DL12 8DT, UK
Email: rlt48033@GlaxoWellcome.co.uk

Contents