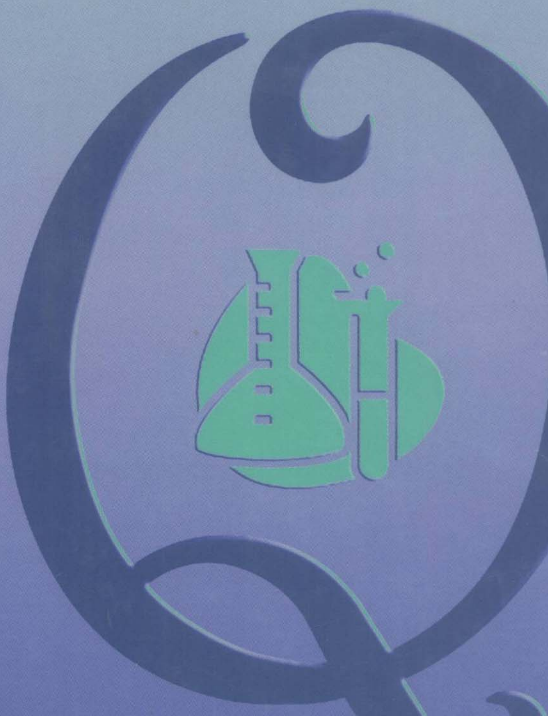
ANALYTICAL CHEMISTRY SERIES

Quality Assurance and Quality Control in the Analytical Chemical Laboratory

A Practical Approach

Piotr Konieczka • Jacek Namieśnik



CRC Press
Taylor & Francis Group

ANALYTICAL CHEMISTRY SERIES

Quality Assurance and Quality Control in the Analytical Chemical Laboratory

A Practical Approach

Piotr Konieczka • Jacek Namieśnik

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

No claim to original U.S. Government works
Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-1-4200-8270-8 (Hardcover)

Library of Congress Cataloging-in-Publication Data

Konieczka, Piotr.
Quality assurance and quality control in the analytical chemical laboratory : a practical approach / Piotr Konieczka, Jacek Namiesnik.
p. cm. -- (Analytical chemistry)
Includes bibliographical references and index.
ISBN 978-1-4200-8270-8 (hardcover : alk. paper)
1. Chemical laboratories--Quality control. 2. Chemistry, Analytic--Qualitative. I. Namiesnik, Jacek. II. Title. III. Series.

QD75.4.Q34K66 2009
542'.1--dc22 2009000685

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

ANALYTICAL CHEMISTRY SERIES

Series Editor
Charles H. Lochmüller
Duke University

Quality and Reliability in Analytical Chemistry, *George E. Baiulescu, Raluca-Ioana Stefan, Hassan Y. Aboul-Enein*

HPLC: Practical and Industrial Applications, Second Edition, *Joel K. Swadesh*

Ionic Liquids in Chemical Analysis, *edited by Mihkel Koel*

Environmental Chemometrics: Principles and Modern Applications, *Grady Hanrahan*

Quality Assurance and Quality Control in the Analytical Chemical Laboratory: A Practical Approach, *Piotr Konieczka and Jacek Namieśnik*

Preface

The aim of this book is to provide practical information about quality assurance/quality control (QA/QC) systems, including definition of all tools, understanding of their uses, and an increase in knowledge about the practical application of statistical tools during analytical data treatment.

Although this book is primarily designed for students and academic teachers, it may also prove useful to the scientific community, particularly among those who are interested in QA/QC. With its comprehensive coverage, this book can be of particular interest to researchers in the industry and academia, as well as government agencies and legislative bodies.

The theoretical part of the book contains information on questions relating to quality control systems.

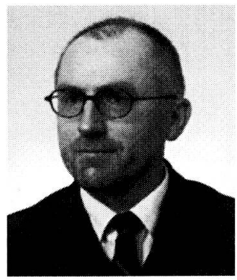
The practical part includes more than 60 examples relating to validation parameter measurements, using statistical tests, calculation of the margin of error, estimating uncertainty, etc. For all examples, a constructed calculation datasheet (Excel) is attached, which makes problem solving easier.

The accompanying CD contains more than 60 Excel datasheet files, each consisting of three main components: problem, data, and solution; in some cases, additional data such as graphs and conclusions are also included. After saving an Excel file on the hard disk, it is possible to use it on different data sets. It should be noted that in order to obtain correct calculations, it is necessary to use it appropriately. The user's own data should be copied only into yellow marked cells (be sure that your data set fits the appropriate datasheet). Solution data will be calculated and can be read from green marked cells.

We hope that with this book, we can contribute to a better understanding of all problems connected with QA/QC.

About the Authors

Piotr Konieczka (MSc 1989, PhD 1994, DSc 2008-GUT), born in 1965, has been employed at Gdańsk University of Technology since 1989 and is currently working as a tutor. His published scientific output includes 1 book, 6 book chapters, and more than 40 papers, as well as more than 70 lectures and communications. His research interests include metrology, environmental analytics and monitoring, and trace analysis.



Jacek Namieśnik (MSc 1972-GUT, PhD 1978-GUT, DSc 1985-GUT, Prof. 1998), born in 1949, has been employed at Gdańsk University of Technology since 1972. Currently a full professor, he has also served as vice dean of the Chemical Faculty (1990–1996) and dean of the Chemical Faculty (1996–2000 and 2005–present). He has been the head of the Department of Analytical Chemistry since 1995, as well as chairman of the Committee of Analytical Chemistry of the Polish Academy of Sciences since 2007, and Fellow of the International Union of Pure and Applied Chemistry (IUPAC) since 1996. He was director of the Centre of Excellence in Environmental Analysis and Monitoring in 2003–2005. Among his published scientific papers are 7 books, more than 300 papers, and more than 350 lectures and communications published in conference proceedings; he has 7 patents to his name. He is the receipient of various awards, including Professor *honoris causa* from the University of Bucharest (Romania) (2000), the Jan Hevelius Scientific Award of Gdańsk City (2001), and the Prime Minister of Republic of Poland Award (2007). His research interests include environmental analytics and monitoring and trace analysis.

List of Abbreviations

AAS	Atomic Absorption Spectrometry
ANOVA	ANalysis Of VAriance
BCR	Bureau Communautaire de Reference (Standards, Measurements, and Testing Programme–European Community)
CDF	Cumulative Distribution Function
CITAC	Cooperation on International Traceability in Analytical Chemistry
CL	Central Line
CRM	Certified Reference Material
CV	Coefficient of Variation
CVAAS	Cold Vapor Atomic Absorption Spectrometry
EN	European Norm
GC	Gas Chromatography
GLP	Good Laboratory Practice
GUM	Guide to the Expression of Uncertainty in Measurement
IAEA	International Atomic Energy Agency
ICH	International Conference on Harmonization
IDL	Instrumental Detection Limit
ILC	InterLaboratory Comparisons
IQR	InterQuartile Value
IRMM	Institute for Reference Materials and Measurements
ISO	International Organization for Standardization
IUPAC	International Union of Pure and Applied Chemistry
JCGM	Joint Committee for Guides in Metrology
LAL	Lower Action (control) Limit
LOD	Limit of Detection
LOQ	Limit of Quantification
L-PS	Laboratory-Performance Study
LRM	Laboratory Reference Material
LWL	Lower Warning Limit
MB	Method Blank
M-CS	Material-Certification Study
MDL	Method Detection Limit
M-PS	Method Performance Study
NIES	National Institute for Environmental Studies
NIST	National Institute of Standards and Technology
NRCC	National Research Council of Canada
PRM	Primary Reference Material
PT	Proficiency Test
$QA_{(1)}$	Quality Assessment
$QA_{(2)}$	Quality Assurance
QA/QC	Quality Assurance/Quality Control

QC	Quality Control
QCM	Quality Control Material
RH	Relative Humidity
RM	Reference Material
RSD	Relative Standard Deviation
S/N	Signal-to-Noise ratio
SD	Standard Deviation
SecRM	Secondary Reference Material
SI	Le Systeme Internationale d'Unités
SOP	Standard Operating Procedure
SRM	Standard Reference Material
UAL	Upper Action (control) Limit
USP	United States Pharmacopea
UWL	Upper Warning Limit
VIM	Vocabulaire International des Termes Fondamentaux et Généraux de Métrologie
VIRM	European Virtual Institute for Reference Materials

Contents

1 Basic Notions of Statistics

1.1 INTRODUCTION

Mathematical statistics is a branch of mathematics that applies the theory of probability to examining regularities in the occurrence of certain properties of material objects or phenomena that occur in unlimited quantities. Statistics presents these regularities by means of numbers.

Statistics is not only art for art's sake. It is a very useful tool that can help us find answers to many questions. Statistics is especially helpful for analysts, because it may clear many doubts and answer many questions associated with the nature of an analytic process, for example:

- How exact the result of determination is.
- How many determinations should be conducted to increase the precision of a measurement.
- Whether the investigated product fulfills the necessary requirements and/or norms.

Yet, it is important to remember that statistics should be applied in a reasonable way.

1.2 DISTRIBUTIONS OF RANDOM VARIABLES

1.2.1 Characterization of Distributions

The application of a certain analytical method unequivocally determines the distribution of measurement results (properties), here treated as independent random variables. A result is a consequence of a measurement. The set of obtained determination results creates a distribution (empirical).

Each defined distribution is characterized by the following parameters:

- A cumulative distribution function (CDF) X is determined by F_X and represents the probability that a random variable X takes on a value less than or equal to x; a CDF is (not necessarily strictly) right-continuous, with its limit equal to 1 for arguments approaching positive infinity, and equal to 0 for arguments approaching negative infinity; in practice, a CDF is described shortly by: $F_X(x) = P(X \leq x)$.
- A density function that is the derivative of the CDF: $f(x) = F'_X(x)$.

Below are the short characterizations of the most frequently used distributions:

- normal distribution
- uniform distribution (rectangular)
- triangular distribution

Normal distribution, also called the Gaussian distribution (particularly in physics and engineering), is a very important probability distribution used in many domains. It is an infinite family of many distributions, defined by two parameters: mean (location) and standard deviation (scale).

Normal distribution, $N(\mu_x, \mathrm{SD})$ is characterized by the following properties:

- an expected value μ_x
- a median $\mathrm{Me} = \mu_x$
- a variance SD^2

Uniform distribution (also called continuous or rectangular) is a continuous probability distribution for which the probability density function within the interval $\langle -a, +a \rangle$ is constant and not equal to zero, but outside the interval is equal to zero.

Because this distribution is continuous, it is not important whether the endpoints $-a$ and $+a$ are included in the interval. The distribution is determined by a pair of parameters $-a$ and $+a$.

Uniform distribution is characterized by:

- an expected value $\mu_x = 0$
- a median $\mathrm{Me} = 0$
- a variance $\mathrm{SD}^2 = a^2/3$

Triangular distribution over the interval $\langle -a, +a \rangle$ is characterized by:

- an expected value $\mu_x = 0$
- a median $\mathrm{Me} = 0$
- a variance $\mathrm{SD}^2 = a^2/6$

The distribution of a random variable provides complete information on an investigated characteristic (e.g., concentration, content, physiochemical property). Unfortunately, such complete information is seldom available. As a rule, characteristic inference is drawn using the analysis of a limited number of elements (samples) representing a fragment of the whole set that is described by the distribution. Then, one may infer a characteristic using an estimation of some of its parameters (statistical parameters) or its empirical distribution.

Statistical parameters are numerical quantities used in the systematic description of a statistical population structure.

These parameters can be divided into four basic groups:

- measures of location
- measures of statistical dispersion
- measures of asymmetry
- measures of concentration

1.3 MEASURES OF LOCATION

Measures of location use one value to characterize the general level of the value of the characteristic in a population [1].

The most popular measures of location are the following:

- arithmetic mean
- truncated mean
- mode
- quantiles:
  - quartiles
  - median
  - deciles

Arithmetic mean is the sum of all the values of a measurable characteristic divided by the number of units in a finite population:

$$x_m = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1.1}$$

Here are the selected properties of the arithmetic mean:

- The sum of the values is equal to the product of the arithmetic mean and the population size.
- The arithmetic mean fulfills the condition:

$$x_{\min} < x_m < x_{\max} \tag{1.2}$$

- The sum of deviations of individual values from the mean is equal to zero:

$$\sum_{i=1}^{n} (x_i - x_m) = 0 \tag{1.3}$$

- The sum of squares of deviations of each value from the mean is minimal:

$$\sum_{i=1}^{n} \left(x_i - x_m\right)^2 = \min \tag{1.4}$$

- The arithmetic mean is sensitive to extreme values of the characteristic.
- The arithmetic mean from a sample is a good approximation (estimation, estimator) of the expected value.

The *truncated mean* x_{wk} is a statistical measurement calculated for the series of results, among which the extrema (minima or maxima) have a high uncertainty concerning their actual value [2]. Its value is calculated according to the formula:

$$x_{wk} = \frac{1}{n}\left[(k+1)x_{(k+1)} + \sum_{i=k+2}^{n-k-1} x_{(i)} + (k+1)x_{(n-k)}\right] \quad (1.5)$$

where

x_{wk} = truncated mean
n = number of results in the series
k = number of extreme (discarded) results

Mode Mo is the value that occurs most frequently in a data set. In a set of results, there may be more than one value that can be a mode, because the same maximum frequency can be attained at different values.

Quantiles Q are values in an investigated population (a population presented in the form of a statistical series) that divide the population into a certain number of subsets. Quantiles are data values marking boundaries between consecutive subsets.

The 2-quantile is called the *median*, 4-quantiles are called *quartiles*, 10-quantiles are *deciles*, and 100-quantiles are *percentiles*.

A quartile is any of three values that divide a sorted data set into four equal parts, so that each part represents 1/4 of the sampled population.

The first quartile (designated Q_1) divides the population in a such a way that 25% of the population units have values lower than or equal to the first quartile Q_1, and 75% units have values higher than or equal to the first quartile. The second quartile Q_2 is the median. The third quartile (designated Q_3) divides the population in a such a way that 75% of the population units have values lower than or equal to the third quartile Q_3, and 25% units have values higher than or equal to the quartile.

The median measurement is the middle number in a population arranged in a nondecreasing order (for a population with an odd number of observations), or the mean of the two middle values (for those with an even number of observations).

A median separates the higher half of a population from the lower half; half of the units have values smaller than or equal to the median, and half of them have values higher than or equal to the median. Contrary to the arithmetic mean, the median is not sensitive to other units in a population. This is usually perceived as its advantage, but sometimes may also be regarded as a flaw; even immense differences between outliers and the arithmetic mean do not affect its value.

Hence, other means have been proposed, for example, the truncated mean. This mean, less sensitive to outliers than the standard mean (only a large number of outliers can significantly influence the truncated mean) and standard deviation, is calculated using all results, which transfers the extreme to an accepted deviation range — thanks to the application of appropriate iterative procedures.

The first decile represents 10% of the results that have values lower than or equal to the first decile, and 90% of the results have values greater than or equal to it.

1.4 MEASURES OF DISPERSION

Measures of dispersion (variability) are usually used to determine differences between individual observations and mean value [1].

The most popular measures of dispersion are:

- range
- variance
- standard deviation
- average deviation
- coefficient of variation (CV)

The range R is a difference between the maximum and minimum value of an examined characteristic:

$$R = x_{\max} - x_{\min} \tag{1.6}$$

It is a measure characterizing the empirical variability region of the examined characteristic, but does not give information on the variability of individual values of the characteristic in the population.

Variance SD^2 is an arithmetic mean of the squared distance of values from the arithmetic mean of the population. Its value is calculated according to the formula:

$$SD^2 = \frac{1}{n-1}\sum_{i-1}^{n}\left(x_i - x_m\right)^2 \tag{1.7}$$

Standard deviation SD, the square root of the variance, is the measure of dispersion of individual results around the mean. It is described by the equation:

$$SD = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - x_m\right)^2}{n-1}} \tag{1.8}$$

Standard deviation equals zero only when all results are identical. In all other cases it has positive values. Thus, the greater the dispersion of results, the greater the value of the standard deviation SD.

It must be remembered that dispersion of results occurs in each analytical process. Yet it is not always observed, for example, because of the resolution of a measuring instrument being too low.

Properties of the standard deviation:

- If a constant value is added to or subtracted from each value, the standard deviation does not change.
- If each measurement value is multiplied or divided by any constant value, the standard deviation is also multiplied/divided by that same constant.