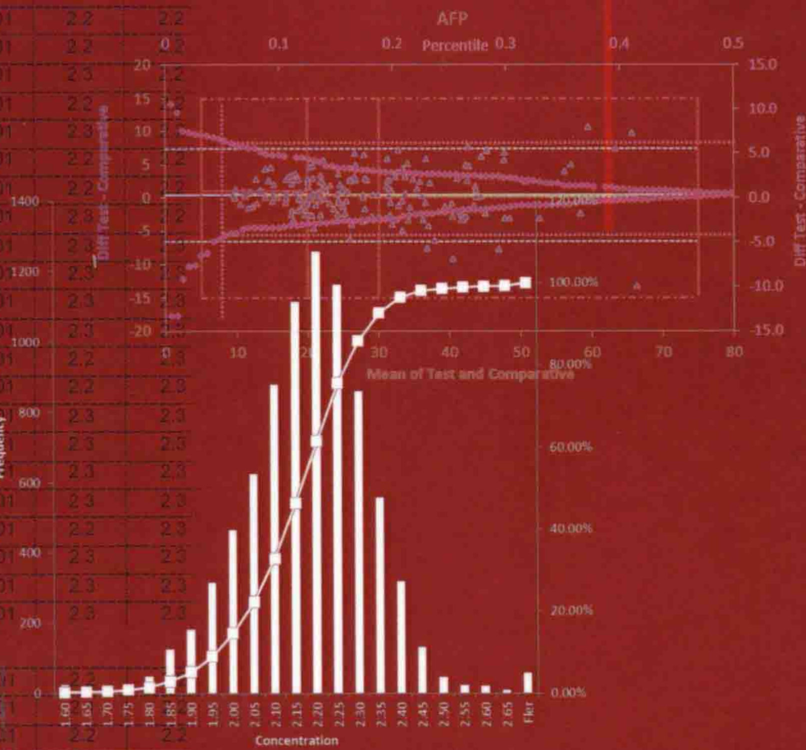
ELSEVIER

LABORATORY STATISTICS

Handbook of Formulas and Terms



Anders Kallner

LABORATORY STATISTICS

Handbook of Formulas and Terms

by

ANDERS KALLNER
Assoc. Professor (R)
Karolinska Univ. Laboratories
Stockholm SWEDEN

ELSEVIER
AMSTERDAM • BOSTON • HEIDELBERG • LONDON • NEW YORK • OXFORD
PARIS • SAN DIEGO • SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Elsevier
225 Wyman Street, Waltham, MA 02451, USA
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA

First edition

Library of Congress Cataloging-in-Publication Data
Kallner, Anders.
Laboratory statistics : handbook of formulas and terms / by Anders Kallner. – First edition.
pages cm
Includes bibliographical references and index.
ISBN 978-0-12-416971-5
1. Mathematical notation. 2. Mathematics–Terminology. I. Title.
QA41.K32 2014
519.503–dc23

2013019706

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

For information on all **Elsevier** publications
visit our web site at store.elsevier.com

Printed and bound in USA
14 15 16 17 18 10 9 8 7 6 5 4 3 2 1

ISBN: 978-0-12-416971-5

LABORATORY STATISTICS

Handbook of Formulas and Terms

Acknowledgment

Extracts from the *International Vocabulary of Metrology—Basic and General Concepts and Associated Terms* (VIM), 3rd edition, JCGM 200:2012 are published with permission of the Director of the International Bureau of Weights and Measures (BIPM), in his functions as Chairman of the JCGM. The member organizations of the JCGM also retain full internationally protected right on their titles, slogans, and logos included in the JCGM's publications. The JCGM does not accept any liability for the relevance, accuracy, completeness, or quality of reproduced information and materials. The only official version is the original version of the document published by the JCGM.

Introduction

When we have to add, or multiply, even big numbers everything goes almost mechanically. This is a routine work, . . ., the true mathematical thinking begins when one has to solve a real problem, that is to say, to identify a mathematical structure that would match the conditions of the problem, to understand principles of its functioning, to grasp connections with other mathematical structures, and to deduce the consequences implied by the logic of the problem. Such manipulations of structures are always immersed into various calculations since calculations form a natural language of mathematical structures. **Michel Heller (2008)**

This present "compendium" is for those who like me are engaged in practical laboratory work and do not have a major in statistical analysis and feel somewhat uncomfortable with the statistical jargon. We frequently face the need to analyze large amounts of data of various origins, collected for various purposes in routine or research work, and have discovered the power of spreadsheet programs in calculations and general data analysis.

Commercial statistical "packages" provide many of the analysis used in the laboratory. By necessity, the organization of the data in these packages has to accommodate many different requirements and is perhaps not optimal for a particular practical purpose. Laboratorians often desire to visualize their results graphically and interactively. The availability of spreadsheet programs has eliminated much of problems and hassle with calculations in statistics, provided simple understandable formulas are available. Indeed, simple spreadsheet programming can satisfy most of the necessary calculations and offer simple, efficient, and customized solutions.

This present compendium is not meant to be a "short course" in statistics but a source of a quick reference, repetition or explanation of formulas and concepts, and encourage development of statistical tools and routines in the research and routine laboratories.

Special attention has been given to expressions that can take different formats but, of course, give the same results. Exposing formulas in different formats may to some extent explain their origin, relation to other procedures, and their usage. We have tried to align formulas regarding style and terminology and group them in a logical order. Some formulas in the collection have been edited to facilitate applying in spreadsheet programs.

The selection of formulas in the compendium has developed during several courses in applied statistics for laboratorians and scientists with experimental projects. The number of worked examples is extensive and regularly enhanced by tables and figures. Whenever feasible the text makes reference to functions and routines in Microsoft EXCEL®.

Formulas have been collected and compared from many different sources, scientific literature, common textbooks, and the Internet. It is all out there, cast in different forms and shapes but may be difficult to find. An idea with this compendium is to have most of the statistical procedures used in the laboratory collected in one source and described in a standardized but not compressed format.

References to individual sources are not given but a list of contemporary literature.

A threat with preprogrammed routines is that, unless simple rules are violated and thus prevented from use, they will always produce an answer. The process of programming and calculating statistical routines has proved to deepen the understanding of the procedures and hopefully diminish erroneous use of established procedures. However, the author takes no responsibility for any erroneous decisions based on calculations using formulas in this compendium.

A comprehensive list of contents and an index facilitate the access of the desired concept or procedure.

VOCABULARY AND CONCEPTS IN METROLOGY

Many organizations have invested heavily in formulating internationally acceptable, clear, comprehensive, and understandable definitions of terms in metrology. Superficially, this may not seem to have any bearing on statistics. Basically, statistics is one way of formulating and expressing mathematical relationships, but we also need to agree on and use definitions of common concepts. The most extensive and internationally recognized list of concepts and their definitions is that created by the joint BIPM, ISO, IEC, IFCC, IUPAC, IUPAP, OIML, and ILAC document *International Vocabulary of Metrology—Basic and General Concepts and Associated Terms(VIM)*,downloadable at http://www.bipm.org/ (accessed 2013-06-30).

The definitions are reproduced *in extenso* from the VIM, but some notes have been deleted when pertaining to pure metrological problems.

The author is grateful for the interest and many excellent suggestions from students and other users of previous editions of the compendium. In particular, Professor Elvar Theodorsson, Department of Clinical Chemistry, University of Linköping, Sweden, has provided healthy critics.

Anders Kallner (anders.kallner@ki.se)

Some Notes on Nomenclature

Mathematical formulas may be difficult to decipher but are in fact unambiguous and comprehensive.

In this compendium, the formulas are not as compressed as they may be and therefore easier to understand. A few rules may help:

The number of items is abbreviated n or N.

$>$ is read "larger than," $<$ "smaller than," $\geq$ "larger than or equal to," $\leq$ "smaller than or equal to."

$\gg$ is read "much larger than," $\ll$ "much smaller than."

Fractions (division), a/b; multiplication $a \times b$.

Multiplications in the body of the text are written $\times$, i.e., $a \times b$.

Square root: $\sqrt{a}$ or explicit $\sqrt[2]{a}$, which allows for higher order roots.

Sum: $a_1 + a_2 + a_3 + \cdots + a_n$ is abbreviated: $\sum_{i=1}^{n} a_i$.

Sum of squares: $(a_1)^2 + (a_2)^2 + (a_3)^2 + \cdots + (a_n)^2$ is abbreviated $\sum_{i=1}^{n} a_i^2$.

A squared sum $(a_1 + a_2 + a_3 + \cdots + a_n)^2$ is $\left(\sum_{i=1}^{n} a_i\right)^2$.

Absolute value: $|a|$, i.e., disregarding any sign.

Standard deviation of a sample x is $s(x)$, $s(X)$, or s if there is no risk for misunderstanding.

Consequently, the standard error of the mean (SEM) is $s(\bar{x})$ or $s(\bar{X})$, as appropriate. The abbreviation SEM is also used.

The period (full stop) "." is used as the decimal sign and a comma "," as the 1000 separator.

Additional abbreviations as appropriate are explained in the text.

Some Greek letters used for certain purposes (small and capital):

Alpha: α and A, *Beta*: β and B, *Gamma*: γ and Γ, *Delta*: δ and Δ, *Epsilon*: ε and E, *Zeta*: ζ and Z, *Eta*: η and H, *Kappa*: κ and K, *Lambda*: λ and Λ, *My*: μ and M, *Xi*; ξ and Ξ, *Pi*: π and Π, *Rho*: ρ and P, *Sigma*: σ and Σ, *Tau*: τ and T, *Chi*: χ and X.

Contents

Formulas

BASICS

Logarithms and Exponents

The logarithm of a given number and a given base is the power to which the base must be raised to get the number.

If b is the base and a the given number, the logarithm is x. In many applications, the notation "log" refers to 10-logarithms (Briggs), i.e., the base 10 and ln refers to e-logarithms or "natural" logarithms with $e=2.7183$ as the base:

$$\text{If } {}^{b}\log(a) = x, \quad \text{then anti} \log(x) = a = b^{x} \tag{1}$$

$$\text{thus if } {}^{e}\log(a) = x; \quad \ln(a) = x; \quad \text{then anti} \ln(x) = a = e^{x} \tag{2}$$

and

$$\text{if } {}^{10}\log(a) = x, \quad \text{then anti} \log(x) = a = 10^{x} \tag{3}$$

$$\begin{aligned} a \times b = c; \quad & \log(a) + \log(b) = \log(c); \\ \frac{a}{b} = c; \quad & \log(a) - \log(b) = \log(c) \end{aligned} \tag{4}$$

$$\log(a^{b}) = b \times \log(a) \tag{5}$$

$$\frac{{}^{c}\log(a)}{{}^{c}\log(b)} = {}^{b}\log(a) \tag{6}$$

$$a^{-n} = \frac{1}{a^n} \tag{7}$$

$$a^{\frac{1}{n}} = \sqrt[n]{a} = \frac{1}{n} \times \log(a) \tag{8}$$

Microsoft EXCEL® commands: Natural logarithm: LN(a); antilog: EXP(LN(a)) (cf. 2).

10-logarithms (Briggs) LOG(a); antilog: $10^{\mathrm{LOG}(a)}$ (cf. 3).

Value of $e = e^1$: EXP(1) = 2.7183; e^b: EXP(b).

Examples

Let $a = 5$, $b = 10$, $c = 3$, and $n = 2$, then

$$^{10}\log(5) = 0.6990; \quad \text{anti}\log(0.6990) = 5 = 10^{0.6990}$$

Since $e = 2.7183$ and $^{e}\log(5) = \ln(5) = 1.61$; anti $\ln(1.61) = 5 = e^{1.61} = 2.7183^{1.61}$

$$5 \times 10 = 50; \quad \log(5) + \log(10) = \log(50);$$

$$0.6990 + 1 = 1.6990; \quad \text{anti}\log(1.6990) = 50$$

$$\frac{5}{10} = 0.5; \quad \log(5) - \log(10) = \log(0.5);$$

$$0.6990 - 1 = -0.3010; \quad \text{anti}\log(0.6990 - 1) = 0.5$$

$$\log(5^{10}) = 10 \times \log(5) = 6.990; \quad \frac{^{3}\log(5)}{^{3}\log(10)} = {}^{10}\log(5) = 0.6990$$

$$5^{-2} = \frac{1}{5^2} = \log(1) - 2 \times \log(5) = \text{anti}\log(0 - 2 \times 0.699)$$
$$= \text{anti}\log(-1.398) = \text{anti}\log(0.902 - 2) = 0.04$$

$$8^{1/3} = \sqrt[3]{8} = \frac{1}{3}\log(8) = \text{anti}\log\left(\frac{1}{3} \times 0.9031\right)$$
$$= \text{anti}\log(0.3010) = 2$$

Calculation of the logarithms, natural or 10-logaritms is directly available in spreadsheet programs. If mathematical tables or calculators are used, logarithms are conventionally expressed with four decimals to achieve sufficient precision for everyday use. Table values can be interpolated.