Pharmaceutical Statistics

David S. Jones

BSc, PhD, CEng CChem, FIM, FRSS, MRSC, MPDSNI

Professor of Biomaterials Science
School of Pharmacy
Queen's University of Belfast
Belfast, UK

London • Chicago Pharmaceutical Press

Published by the Pharmaceutical Press
Publications division of the Royal Pharmaceutical Society of Great Britain

1 Lambeth High Street, London SE1 7JN, UK
100 South Atkinson Road, Suite 206, Grayslake, IL 60030-7820, USA

First published 2002

Text design by Barker/Hilsdon, Lyme Regis, Dorset
Typeset by MCS Ltd, Salisbury, Wiltshire
Printed in Great Britain by TJ International, Padstow, Cornwall

ISBN 0 85369 425 7

A catalogue record for this book is available from the British Library

and should be recalled whenever the reader is checking calculations. If such an approximate relationship is not observed, then it is strongly advised that all the calculations should be re-checked.

2.2.4.1 General comments on the standard deviation

The standard deviation is the most commonly used measure of the dispersion of data, because it may be related to the probability of a measurement occurring within certain regions on the frequency distribution. Thus, in normal (symmetrical), and indeed in moderately skewed (asymmetrical) distributions:

- 68.27% of all values are included within the numerical range described by $\bar{X} + s$ and $\bar{X} - s$, namely one standard deviation around the mean.
- 95.45% of all values are included within the numerical range described by $\bar{X} + 2s$ and $\bar{X} - 2s$, namely two standard deviations around the mean.
- 99.73% of all values are included within the numerical range described by $\bar{X} + 3s$ and $\bar{X} - 3s$, namely three standard deviations around the mean.

In the example described above concerning the time required for the release of 50% of the original loading of therapeutic agents, the mean and standard deviation were calculated to be 23.6 ± 2.3 h. Consequently

- 68.27% of all values are included within the numerical range described by 21.3 h (i.e. 23.6 – 2.3 h) to 25.9 h (i.e. 23.6 + 2.3 h).
  Therefore, in the current example, 10 out of 15 values were distributed within this range.
- 95.45% of all values are included within the numerical range described by 19.0 h (i.e. 23.6 – 4.6 h) to 28.2 h (i.e. 23.6 + 4.6 h).
  Therefore, in the current example, 14 out of 15 values were distributed within this range.
- 99.73% of all values are included within the numerical range described by 16.7 h (i.e. 23.6 – 6.9 h) to 30.5 h (i.e. 23.6 + 6.9 h).
  Therefore, in the current example, all values were distributed within this range.

The standard deviation (and indeed the variance) is dramatically affected by extreme values in a population, a point that should be considered whenever the variation of a set of data is under discussion. The effects of extreme values on the variance is illustrated in the following example.

EXAMPLE 2.9 *The concentrations (mg/5 mL) of a penicillin antibiotic in five separate bottles of a paediatric suspension have been examined using an iodometric technique. Calculate the mean and standard deviation and consider the contribution of each observation to the sample variance.*