Yuri A.W. Shardt

Statistics for Chemical and Process Engineers

A Modern Approach

EXTRAS ONLINE

Springer

Statistics for Chemical and Process Engineers

Yuri A.W. Shardt

Statistics for Chemical and Process Engineers

A Modern Approach

Springer

Yuri A.W. Shardt
Institute of Automation and Complex Systems (AKS)
University of Duisburg-Essen
Duisberg, North Rhine-Westphalia
Germany

ISBN 978-3-319-21508-2 ISBN 978-3-319-21509-9 (eBook)
DOI 10.1007/978-3-319-21509-9

Library of Congress Control Number: 2015950483

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Foreword

The need for the development and understanding of large, complex data sets in a wide range of different fields, including economics, chemistry, chemical engineering, and control engineering is very important. In all these fields, the common thread is using these data sets for the development of models to forecast or predict future behaviour. Furthermore, the availability of fast computers has meant that many of the techniques can now be used and tested even on one's own computer. Although there exist a wealth of textbooks available on statistics, they are often lacking in two key respects: application to the chemical and process industry and their emphasis on computationally relevant methods. Many textbooks still contain detailed explanations of how to manually solve a problem. Therefore, the goal of this textbook is to provide a thorough mathematical and statistical background the regression analysis through the use of examples drawn from the chemical and process industries. The majority of the textbook presents the required information using matrices without linking to any particular software. In fact, the goal here is to allow the reader to implement the methods on any appropriate computational device irrespective of their specific availability. Thus, detailed examples, that is, base cases, and solution steps are provided to ease this task. Nevertheless, the textbook contains two chapters devoted to using MATLAB® and Excel®, as these are the most commonly used tools both in industry and in academics. Finally, the textbook contains at the end of each chapter a series of questions divided into three parts: conceptual questions to test the reader's understanding of the material; simple exercise problems that can be solved using pen, paper, and a simple, handheld calculator to provide straightforward examples to test the mechanics and understanding of the material; and computational questions that require modern computational software that challenge and advance the reader's understanding of the material.

This textbook assumes that the reader has completed a basic first-year university course, including univariate calculus and linear algebra. Multivariate calculus, set theory, and numerical methods are useful for understanding some of the concepts,

but knowledge is not required. Basic chemical engineering, including mass and energy balances, may be required to solve some of the examples.

The textbook is written so that the chapters flow from the basic to the most advanced material with minimal assumptions about the background of the reader. Nevertheless, multiple different courses can be organised based on the material presented here depending on the time and focus of the course. Assuming a single semester course of 39 h, the following would be some options:

1. *Introductory Course to Statistics and Data Analysis*: The foundations of statistics and regression are introduced and examined. The main focus would be on Chap. 1: Introduction to Statistics and Data Visualisation, Chap. 2: Theoretical Foundation for Statistical Analysis, and parts of Chap. 3: Regression, including all of linear regression. This course would prepare the student to take the Fundamentals of Engineering Exam in the United States of America, a prerequisite for becoming an engineer there.
2. *Deterministic Modelling and Design of Experiments*: In-depth analysis and interpretation of deterministic models, including design of experiments, is introduced. The main focus would be on Chap. 3: Regression and Chap. 4: Design of Experiments. Parts of Chap. 2: Theoretical Foundation for Statistical Analysis may be included if there is a need to refresh the student's knowledge of background information.
3. *Stochastic Modelling of Dynamic Processes*: In-depth analysis and interpretation of stochastic models, including both time series and prediction error methods, is examined. The main focus would be on Chap. 5: Modelling Stochastic Processes with Time Series Analysis and Chap. 6: Modelling Dynamic Processes. As necessary, information from Chap. 2: Theoretical Foundation for Statistical Analysis and Chap. 3: Regression could be used. The depth in which these concepts would be considered would depend on the orientation of the course: either a theoretical emphasis can be made, by focusing on the theory and proofs, or an application emphasis can be made, by focusing on the practical use of the different results.

As appropriate, material from Chap. 7: Using MATLAB® for Statistical Analysis and Chap. 8: Using Excel® to do Statistical Analysis could be introduced to show and explain how the students can implement the proposed methods. It should be emphasised that this material should not overwhelm the students nor should it become the main emphasis and hence avoid thoughtful and insightful analysis of the resulting data.

The author would like to thank all those who read and commented on previous versions of this textbook, especially the members of the process control group at the University of Alberta, the students who attended the author's course on process data analysis in the Spring/Summer 2012 semester, and members of the Institute of Automation and Complex Systems (Institute für Automatisierungstechnik und komplexe Systeme) at the University of Duisburg-Essen. The author would specifically wish to thank Profs. Steven X. Ding and Biao Huang for their support,

Oliver Jackson from Springer for his assistance and support, and the Alexander von Humboldt Foundation for the monetary support.

Downloading the data: The data sets, MATLAB® files, and Excel® templates can be downloaded from http://extras.springer.com/. Enter the ISBN of the book, ISBN 978-3-319-21508-2, and you will get the requested information.

Contents