
Health, Safety, *and* Environmental Data Analysis

A Business Approach

Anthony J. Joseph

Health, Safety, *and* Environmental Data Analysis

A Business Approach

Anthony J. Joseph



LEWIS PUBLISHERS

Boca Raton

New York

Acquiring Editor: Ken McCombs
Editorial Assistant: Susan Alfieri
Project Editor: Helen Linna
Marketing Manager: Arline Massey
Direct Marketing Manager: Becky McEldowney
Cover design: Dawn Boyd
PrePress: Kevin Luong

Library of Congress Cataloging-in-Publication Data

Anthony Joseph

p. cm.

Includes bibliographical references and index.

ISBN 1-56670-233-X1.

1. Environmental science—statistical methods. 2. Industrial hygiene—statistical methods.

3. Occupational health—statistical methods I. Joseph, Anthony. II. Title

TD426.M62 1997

628.5'5—dc20

97-73212

CIP

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage or retrieval system, without prior permission in writing from the publisher.

The consent of CRC Press LLC does not extend to copying for general distribution, for promotion, for creating new works, or for resale. Specific permission must be obtained in writing from CRC Press LLC for such copying.

Direct all inquiries to CRC Press LLC, 2000 Corporate Blvd., N.W., Boca Raton, Florida 33431.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation, without intent to infringe.

© 1998 by CRC Press LLC

Lewis Publishers is an imprint of CRC Press LLC

No claim to original U.S. Government works

International Standard Book Number 1-56670-233-X

Library of Congress Card Number 97-73212

Printed in the United States of America 1 2 3 4 5 6 7 8 9 0

Printed on acid-free paper

Contents

Preface.....	vii
Overview.....	xi
List of Figures	xviii
List of Tables	xiv

Section A Basic Statistical Concepts

1	Presentation of Data.....	3
1.1	Introduction.....	3
1.2	Presentation	3
1.3	Accuracy and Errors.....	4
1.4	Arrays	5
1.5	Tally Marks	5
1.6	Frequency Distribution	5
1.7	Cumulative Frequency Table	8
1.8	Tabulation of Attributes	9
1.9	Graphical Displays.....	10
1.10	Graphs.....	15
1.11	Scatter Diagrams	16
1.12	Time Value in Presentation of Data	16
1.13	Key Points in Data Presentation.....	18
2	Measures of Location and Dispersion.....	21
2.1	Introduction.....	21
2.2	Mean.....	22
2.3	Median	22
2.4	Mode	22
2.5	Percentile	22
2.6	Selecting an Appropriate Measure of Central Tendency	22
2.7	Measures of Dispersion of Discrete Data	23
2.8	Measures of Dispersion of Grouped Data.....	24
2.9	Other Measures	26
2.10	Application of Mean and Standard Deviation.....	27

- 2.11 Rates 28
- 2.12 Limitations of Rates 30
- 2.13 Errors in Measures 30
- 3 Probability and Probability Distributions 33
 - 3.1 Introduction 33
 - 3.2 Simple Events 33
 - 3.3 Compound Events..... 34
 - 3.4 Expectations..... 36
 - 3.5 Probability Distributions 37
 - 3.6 Continuous Probability Distributions..... 40
 - 3.7 Guidelines in Selecting a Probability Distribution..... 41
 - 3.8 Examples of the Use of Probability Theory 41

Section B Data Gathering and Analysis

- 4 Sampling Methods 49
 - 4.1 Introduction 49
 - 4.2 Requirements of a Good Sample 49
 - 4.3 Sample Size 51
 - 4.4 Sampling Systems 51
 - 4.5 Sampling Considerations..... 53
 - 4.6 Developing a Questionnaire 55
 - 4.7 Execution of a Survey 55
 - 4.8 Survey Errors..... 58
 - 4.9 Recommended Attributes of Measurements..... 58
 - 4.10 Design of Experimental Studies 58
 - 4.11 Quality Control Sampling..... 61
 - 4.12 Considerations in Statistical Analysis of
Occupational Environmental Exposures..... 61
 - 4.13 Presentation of Report 62
- 5 Sampling Theory and Estimation 65
 - 5.1 Introduction 65
 - 5.2 Estimates..... 65
 - 5.3 Point Estimators 66
 - 5.4 Relationship between Mean Point Estimate and
Sample Size..... 67
 - 5.5 Interval Estimates..... 69
- 6 Testing and Inference..... 71
 - 6.1 Introduction..... 71
 - 6.2 Hypothesis Testing..... 71

6.3	Hypothesis Tests about a Population Mean :	
	Large Sample	72
6.4	Suggested Steps of Hypothesis Testing.....	73
6.5	Hypothesis Tests about a Population Mean:	
	Small Sample	74
6.6	Hypothesis Tests about a Population Proportion.....	75
6.7	Hypothesis Tests a Population Variance	76
6.8	Hypothesis Tests of Two Populations: Difference	
	Between Two Means.....	76
6.9	Hypothesis Tests of Two Populations: Differences	
	Between Two Proportions	77
6.10	Hypothesis Tests of two Populations: Comparing	
	Variances	77
6.11	Hypothesis Tests of more than two Population	
	Means	77
6.12	Hypothesis Testing and Decision Making	77
6.13	Goodness of Fit Test	80
6.14	An Illustrative Example	81
6.15	Non-parametric Methods.....	83

Section C Information Generation

7	Regression and Correlation	89
7.1	Introduction.....	89
7.2	Regression Equation.....	89
7.3	Coefficient of Correlation.....	91
7.4	Testing of Significance.....	92
7.5	Estimation and Prediction	93
7.6	Residual Analysis	94
7.7	Correlation Analysis	95
7.8	Multiple Regression	95
7.9	An Example	96
8	Time Series	99
8.1	Introduction.....	99
8.2	Smoothing Method: Moving Average	100
8.3	Forecasting.....	103
8.4	Exponential Smoothing.....	106
9	Linear Programming	107
9.1	Introduction.....	107
9.2	Formulating a Linear Programming Problem.....	107
9.3	Solving the Problem	108

10	Network and Gantt Chart	111
10.1	Introduction	111
10.2	Network Analysis Techniques	111
10.3	Allocation of Resources	116
11	Decision Analysis	119
11.1	Introduction	119
11.2	Stages in a Decision-Making Process	120
11.3	Definition of the Problem	120
11.4	Analysis of the System	120
11.5	Constructing the Model	121
11.6	Manipulation and Testing of Model	122
11.7	Implementation	123
11.8	Application of Regression Modeling	123
11.9	Conclusion	124
12	Project Examples	125
12.1	Introduction	125
12.2	Example 1: Decision Tree and Expectation Values	125
12.3	Example 2: Sampling and Estimation	129
12.4	Example 3: Hypothesis Tests with Chi-Squared Distribution	130
12.5	Example 4: Hypothesis tests with t-Distribution	134
12.6	Example 5: Linear Programming	136
12.7	Example 6: Network Analysis	138
	Selected Bibliography	141
	Index	143

Preface

The information presented in this book is based on a series of lectures delivered by the author to students at college and groups of adults pursuing continuing education classes in Public Health, Occupational Safety and Health, and Environmental Studies. In today's world, statistics is an integral subject in any scientific study.

This book is written especially for persons who use statistics on their jobs and are not statistically inclined. To them, I strongly recommend that an initial surge of interest and enthusiasm must be generated before studying the subject. This book will also be useful for college students pursuing programs in safety, industrial hygiene, public health, and environmental protection. The experiences I gained in delivering these lectures are incorporated into this book. The information presented is based on several texts listed in the selected bibliography, reports, journal articles, and experience of the author. The subjects are approached in a simple and factual manner, complemented with examples. Mathematical derivations are omitted, except where the comprehension of the idea is dependent on the derivation. Although great efforts were made to ensure simplicity, clarity, and continuity, this book, like other statistics texts, has in some ways been fragmented. The text enunciated the principles and commonly employed methods of statistics by proceeding from one topic to another in a hopefully logical sequence, with examples from published studies, case studies, or text books. These examples serve as illustrations.

This book is intended for practicing EHS professionals, students pursuing a career in occupational health and safety and environmental studies, and managers of safety, health or environment. It is divided into three sections:

1. Basic statistical concepts
2. Data gathering and analysis
3. Information generation

Knowledge of basic statistical concepts such as presentation of data, measurements of location and dispersion, and elementary probability and distributions are essential to applying statistical techniques and methods. These are covered in Chapters 1 to 3.

Data gathering and analysis topics such as, sampling methods, sampling theory, testing and inference are essential for knowing the principles and basic methods of statistics. These will also provide skills to

evaluate published numerical information critically, evident in most EHS reports. These are covered in Chapters 4 to 6.

Information generation topics such as, regression and correlation analysis, time series, linear programming, network and Gantt charting, and decision analysis, are presented as tools that can be used to convert data into meaningful information. These are covered in Chapters 7 to 11.

Chapter 12 is a special chapter, it features six examples of projects that were successful because a statistical approach was adopted. Whatever statistical understanding you may have gained from reading this book or consulting the listed texts in the selected bibliography, synthesis of experience, understanding of statistics, and knowledge of your profession must occur.

All efforts were made to ensure that this book covers the statistical ideas necessary for the practicing environmental, health, and safety professional who will like to use statistical data to communicate information with a business approach effectively. I sincerely hope that by your reading this book, the statistical aspects of your job will be more meaningful and rewarding.

Have fun!
Anthony J. Joseph

About the Author

Anthony J. Joseph has twenty years of experience in the field of occupational health and safety. He is currently a professor at the Indiana University of Pennsylvania, one of the premier health and safety schools in the country. Professor Joseph also holds a Ph.D. in Environmental Engineering, an M.S. in Environmental Engineering, Pollution Control, as well as in Safety Sciences.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

An Overview on The Importance of Statistics to Environmental, Health and Safety Professions

Environmental, Health, and Safety (EHS) professionals are engaged in the collection of data on their jobs, such as data related to lost-time accidents and emission levels of sulfur oxides. The tendency is to collect and store the data to satisfy legal requirements. The wealth of information embodied in these collections of data is underutilized or seldom used. An appreciation and understanding of statistical methods are essential for data collection to be meaningful. We live in a world where statistics influence our lifestyles and behavior. Our life is affected by human events such as illness, accidents, and environmental disasters. Most of these events are recorded and documented in numerical format. These collections of data and information over a period of time are the body of knowledge often called statistics. As with other words, the word “statistics” has different meanings to different people, for example it means mathematics to the lay person, and the study of data to the statistician.

A narrow view of the subject matter of statistics is that it simply involves page upon page of numbers in volumes upon volumes stored on the computer or on the shelf. This is only partly true. In the broader sense, statistics include techniques for tabulating and graphing data for presentation, and methods of summarizing and analyzing data. Statistics can be divided into two groups, called descriptive and inferential. Vital statistics such as birth, death, marriage, divorce, and the occurrence of communicable diseases, are used frequently in the EHS profession. Inferential statistics is the logical basis by which conclusions regarding populations are drawn from results obtained in a sample. This process of inference from sample to population pervades the fields of safety, health, environmental science, medicine and social studies. Consider the following example of statistical inference:

A worker shows signs of lead poisoning, as a result, a sample of blood or urine or a biopsy of tissues is taken. From the sample obtained from the patient, a conclusion is drawn regarding a larger “population;” namely, the patient’s total urine or blood volume, or his entire organ.

This notion of inference from sample to population has as its underlying foundation, the mathematical theory of probability. This does not mean that one must know the mathematical theory to use the statistical methods effectively. What it means is that one must understand the basis of the methodology, the assumptions governing the use of the techniques and, the proper interpretation of the results. If I can draw the analogy to the operating of a car, one need not know the mechanics of internal combustion to be able to drive a car safely. The attainment of particular operational skills along with knowledge of the rules of the road are sufficient for operating the car safely, and effectively in most countries.

The application of statistical results is so widespread that the importance of statistics can hardly be overemphasized. Rates, ratios, and probabilities are all related to statistics, for example infant mortality rates, that is the number of deaths less than one year of age reported during the same year.

A branch of medicine called epidemiology is based on statistics. This branch of medicine contributed and is still contributing to the improvement of public health. The developments of immunology and the acquisition of new knowledge regarding the transmission of diseases were ushered by epidemiology. For example in 1854, John Snow proved that cholera was transmitted in water and not by personal contact from statistical analysis, while the identification of the cholera organism was not made until 27 years later.

Can statistics provide all the answers to our safety, health and environmental problems? Statistics deal with measurable aspects of things. Therefore, it can seldom give the complete solution to a problem. Statistics can provide a basis for judgement and decision making. The limitations of the data collected and use must be fully delineated. Clearly, all observations must be accurately recorded, collated, analyzed and presented. Usually, it is this body of data transformed into information that is used to predict future events, to infer or to relate, or to classify events that impact on our lives.

The age of computers has brought us the ability to amass volumes of data and information. However, the extensive use of this data gathering marvel to transform data into useful information has not in most cases cured the malady that exists among many EHS professionals. Today we are enjoying better health and safety in our work places, home and communities because of the "body of knowledge" obtained from data and observations. The development of vaccines resulted from statistical studies. Just reflect for a moment on the importance of statistics in EHS. It is impossible to be a professional in EHS without knowing basic statistics.

List of Figures

Figure 1.1	Illustration of a Cluster Bar Chart
Figure 1.2	Illustration of an Overlap Bar Chart
Figure 1.3	Illustration of a Stacked Bar Chart
Figure 1.4	Illustration of a Stacked 100% Bar Chart
Figure 1.5	Illustration of a Stacked Line Chart
Figure 1.6	Illustration of a Pie Chart
Figure 1.7	Illustration of a Venn Diagram
Figure 1.8	Illustration of a Cartographic Map
Figure 1.9	Illustration of a Histogram and Frequency Polygon
Figure 3.1	An Example of Decision Tree
Figure 7.1	Illustration of Scatter Plot
Figure 7.2	Scatter Plot and Regression Line of Example
Figure 8.1	Illustration of Time Series and Trend line
Figure 9.1	Graphical Representation of Constraints, Boundary Line, and Feasible Region
Figure 10.1	Illustration of Network
Figure 10.2	Network of Worked Example
Figure 10.3	Network of Worked Example and Allotted Times
Figure 10.4	Presentation of Event Node
Figure 10.5	Earliest and Latest Times of Worked Example
Figure 10.6	Gantt Chart of Worked Example
Figure 12.1	Decision Tree of Project Number One
Figure 12.2	Network of Case Number Six

List of Tables

Table 1.1	Hours of Labor per Department before a Lost-Time Accident
Table 1.2	Analysis of Gender — Richmond Limited, 1996
Table 2.1	Computation of the Estimated Standard Deviation of Example
Table 2.2	Injuries and Illnesses Experienced by Workers of Company TEN
Table 2.3	Age at Death Distribution of Nuclear Workers
Table 3.1	Probability of Operators Depressing an Incorection Key
Table 4.1	Morbidity per Household Distribution
Table 4.2	Incidence of Rashes in Elementary School Students, 1994
Table 6.1	Age Distribution of Customers Involved in an Accident in a Fast-Food Outlet
Table 6.2	Calculations of Chi-Squared Statistics
Table 7.1	Number of Workers and Total Lost Time
Table 7.2	Budgeted Amount and Rating Value for Company EIGHT, Inc.
Table 8.1	Lost Time Due to Accidents for Company X, 1990 to 1995
Table 8.2	Lost Time Injuries of TJ Company Over a 4-Week Period
Table 8.3	Computation of Moving Averages and Differences
Table 8.4	Daily Variations of Lost Time Over a 4-Week Period
Table 8.5	Number of Accidents by Quarters — 1990 to 1992
Table 8.6	Solution to Worked Example on Accidents per Quarter
Table 8.7	Seasonal Variations
Table 8.8	Predicted Values for 1993
Table 10.1	List of Activities and Sequence
Table 10.2	List of Activities and Duration
Table 10.3	Float Times of Example
Table 10.4	Activities, Duration, and Number of Workers
Table 12.1	Supply, Probability, and Expected Values of Example
Table 12.2	Proportions of Area under Normal Curve with z Standard Deviations of Mean
Table 12.3	Number of Employees in Favor of Option
Table 12.4	Training Requirements of Employees
Table 12.5	Calculations of Probabilities Related to Example 3
Table 12.6	Calculations of Example χ^2 Related to Example 3
Table 12.7	Lost Time per Machine, 1993
Table 12.8	Project Activities and Probabilities of Completion
Table 12.9	Mean and Standard Deviation of Times for Example 6

Section A

Basic Statistical Concepts
