APPLIED RELIABILITY

Paul A. Tobias, Ph.D.

David C. Trindade, Ph.D.

APPLIED RELIABILITY

Paul A. Tobias, Ph.D.
Senior Technical Staff Member
and Manager of Field Performance
and Statistical Support
IBM Corporation
Hopewell Junction, New York

David C. Trindade, Ph.D.
Corporate Director of Reliability
Advanced Micro Devices, Inc.
Sunnyvale, California

VNR VAN NOSTRAND REINHOLD COMPANY
New York

Library of Congress Catalog Card Number: 85-20220
ISBN: 0-442-28310-5

Manufactured in the United States of America

Published by Van Nostrand Reinhold Company Inc.
115 Fifth Avenue
New York, New York 10003

Van Nostrand Reinhold Company Limited
Molly Millars Lane
Wokingham, Berkshire RG11 2PY, England

Van Nostrand Reinhold
480 Latrobe Street
Melbourne, Victoria 3000, Australia

Macmillan of Canada
Division of Gage Publishing Limited
164 Commander Boulevard
Agincourt, Ontario M1S 3C7, Canada

15 14 13 12 11 10 9 8 7 6 5 4 3 2

Library of Congress Cataloging-in-Publication Data

Tobias, Paul A.
Applied reliability.

Includes index.
1. Reliability (Engineering) 2. Quality control—Statistical methods. I. Trindade, David. II. Title.
TA169.T63 1986 620'.00452 85-20220
ISBN 0-442-28310-5

Preface

The purpose of this book is to provide the working engineer, statistician, or scientist with practical tools and techniques for solving today's applied problems in reliability.

Quality and reliability have become strategic variables on a par with price and performance. The average consumer consults tables that compare repair records before selecting an automobile; corporations demand ever more stringent guarantees of defect-free operation when purchasing data-processing equipment. Those businesses that emphasize quality and reliability as part of their normal manufacturing procedures are the ones that will be able to compete in today's marketplace; those that regard quality assurance as a set of final inspection screens and reliability as a warranty pricing concern will find themselves dwindling away.

The importance of quality and reliability is no longer "new news" or controversial. Numerous excellent books have described how to make quality a way of life. Statistical consultants offer many courses on the mathematics and management aspects of quality, and several large corporations have established their own internal quality schools or institutes. The tools and techniques of quality control are well known and widely practiced.

The analysis and control of product reliability is not as well understood. Requiring systems to work, not only the first time, but for many hours or months or years thereafter, makes the testing and product assurance role much more difficult. There are only a few books and courses available to teach an engineer how to run the experiments and make the decisions that are required by management. Most textbooks on reliability are theoretical in nature on the one hand or comprehensive reference works on the other, neither of which fully serves the needs of the reliability engineer or statistician who has to answer reliability concerns on a daily basis.

This text is aimed primarily at those individuals who have responsibility for the design or evaluation of the reliability aspects of components or hardware systems. Statisticians desiring an introduction to the definitions, distributions, techniques, and models currently used to evaluate reliability will also benefit.

Much of this book evolved from lecture notes written by the authors for the purpose of teaching reliability and quality concepts to managers and engineers within an intensive one-week course at various IBM facilities throughout the world. The notes were compiled because no available text adequately covered all the techniques and procedures actually in use to evaluate advanced technology reliability. The selection of material was dictated by what was needed; the style of presentation was dictated by what worked.

The material varies considerably in scope and level of difficulty. Chapter 1 covers elementary descriptive statistics, whereas Chapter 8 includes models for general reliability algorithms and burn-in benefits. Chapter 6 describes how to fit a line through points, whereas Chapters 4 and 5 tell the reader why it might be beneficial to buy state-of-the-art programs for maximum likelihood estimation. Chapter 9 contains the theory of acceptance sampling plans, as well as a wide collection of charts and nomographs for choosing sampling plans and acceptance numbers. Most books on reliability do not include the quality-control procedures described in Chapter 9, but since these are often used to control reliability, they deserve a detailed description.

The reliability analyst needs to combine standard statistical methods with advanced state-of-the-art techniques, on a day-to-day basis. To do so requires being familiar with a collection of quick graphical methods and knowing their strengths and weaknesses. When extremely important decisions based on reliability data analysis must be made, the analyst should know what advanced computer programs are available for purchase. An understanding of life distributions and acceleration models and a collection of proven statistical data analysis tools are essential.

The best way to meet these needs is to illustrate the definitions, theories, and models with applied numerical problems that make use of actual or simulated data. There are sixty formal examples of this type throughout the text, as well as many informal ones. These examples illustrate many of the typical problems a reliability analyst must handle. Students are encouraged to work out as many of them on their own as possible and then check out their work with the solutions in the text. That way they can be sure of understanding the methods well enough to apply them to real data. Over ninety graphs, charts, and tables are also included to supplement the other material.

Mathematical reliability theory, especially in the areas of data analysis and modeling stress acceleration, is a rapidly evolving discipline. As time goes on, new methods will replace some of those described in this book. As of now, however, we present them as a collection of well-tested techniques that have proven successful in evaluating and predicting reliability.

PAUL A. TOBIAS
DAVID C. TRINDADE

Contents

1

Basic Descriptive Statistics

One of the most useful skills a reliability specialist can develop is the ability to convert a mass (mess?) of data into a form suitable for meaningful analysis. Raw numbers by themselves are not useful; what is needed is a distillation of the data into information.

In this chapter we discuss several important concepts and techniques from the field of descriptive statistics. These methods will be used to extract a relevant summary from collected data. The goal is to describe and understand the random variability that exists in all measurements of real world phenomena and experimental data.

The topics we shall cover include: populations and samples; frequency functions, histograms, and cumulative frequency functions; the population cumulative distribution function (CDF) and probability density function (PDF); elementary probability concepts; random variables, population parameters, and sample estimates; theoretical population shape models and data simulation.

POPULATIONS AND SAMPLES

Statistics is concerned with variability, and it is a fact of nature that variation exists. No matter how carefully a process is run, an experiment is executed, or a measurement is taken, there will be differences in repeatability due to the inability of any individual or system to control completely all possible influences. If the variability is excessive, the study or process is described as lacking control. If, on the other hand, the variability appears reasonable, we accept it and continue to operate. How do we visualize variability in order to understand if we have a controlled situation?

Consider the following example.

EXAMPLE 1.1 AUTOMOBILE FUSE DATA

A manufacturer of automobile fuses produces lots containing 100,000 fuses rated at 5 A. Thus, the fuses are supposed to open in a circuit if the current through the fuse exceeds 5 A. Since a fuse protects other elements from possibly damaging electrical overload, it is very important that fuses function properly. How can the manufacturer assure himself that the fuses do indeed operate correctly and that there is no excessive variability?

Obviously he cannot test all fuses to the rated limit since that act would destroy the product he wishes to sell. However, he can sample a small quantity of fuses (say, 100 or 200) and test them to destruction to measure the opening point of each fuse. From the sample data, he could then infer what the behavior of the entire group would be if all fuses were tested.

In statistical terms, the entire set or collection of measurements of interest (e.g., the blowing values of all fuses) define a *population*. ♣

A population is the entire set or collection of measurements of interest.

Note that a population may be finite as in the case of the fuses or it may be infinite as occurs in the situation of a manufacturing process where the population could be all product that has been or will be produced in a fabricating area.

The *sample* (e.g., the 100 or 200 fuses tested to destruction) is a subset of data taken from the population.

A sample is a subset of data from the population.

The objective in taking a sample is to make inference about the population.

Note that data may exist in one of two forms. In *variables* data, the actual measurement of interest is taken. In *attribute* data, the results exist in one of two categories: either pass–fail, go–no go, in spec–out of spec, and so on. Both types of data will be treated in this text.

In the fuse data example, we record variables data but we could also transform the same results into attribute data by stating whether a fuse opened before or after the 5 A rating. Similarly, in reliability work one can measure the actual failure time of an item (variables data) or record the number of items failing before a fixed time (attribute data). Both types of data occur frequently in reliability studies.

Later we will discuss such topics as choosing a sample size, drawing a sample randomly, and the "confidence" in the data from a sample. For now, however, let us assume that the sample has been properly drawn and consider what to do with the data in order to present an informative picture.

HISTOGRAMS AND FREQUENCY FUNCTIONS

In stating that a sample has been *randomly drawn* we imply that each measurement or data point in the population has an equal chance or probability of being selected for the sample. If this requirement is not fulfilled, the sample may be "biased" and correct inference about the population might not be possible.

What information does the manufacturer expect to obtain from the sample measurements of 100 fuses? First, the data should cluster about the rated value of 5 A. Second, the spread in the data (variability) should not be large, because the manufacturer realizes that serious problems could result for users of the fuses if some blow at too high a value. Similarly, fuses opening at too low a level could cause needless repairs or generate unnecessary concerns.

The reliability specialist randomly samples 100 fuses and records the data shown in Table 1.1. It is easy to determine the high and low values from the sample data and see that the measurements cluster roughly about the number 5. Yet, there is still difficulty in grasping the full significance of this set of data.

Table 1.1. Sample Data on 100 Fuses.

4.64	4.95	5.25	5.21	4.90	4.67	4.97	4.92	4.87	5.11
4.98	4.93	4.72	5.07	4.80	4.98	4.66	4.43	4.78	4.53
4.73	5.37	4.81	5.19	4.77	4.79	5.08	5.07	4.65	5.39
5.21	5.11	5.15	5.28	5.20	4.73	5.32	4.79	5.10	4.94
5.06	4.69	5.14	4.83	4.78	4.72	5.21	5.02	4.89	5.19
5.04	5.04	4.78	4.96	4.94	5.24	5.22	5.00	4.60	4.88
5.03	5.05	4.94	5.02	4.43	4.91	4.84	4.75	4.88	4.79
5.46	5.12	5.12	4.85	5.05	5.26	5.01	4.64	4.86	4.73
5.01	4.94	5.02	5.16	4.88	5.10	4.80	5.10	5.20	5.11
4.77	4.58	5.18	5.03	5.10	4.67	5.21	4.73	4.88	4.80

Let us try the following procedure:

1. Find the *range* of the data by subtracting the lowest from the highest value. For this set, the range is $5.46 - 4.43 = 1.03$.

2. Divide the range into 10 or so equally spaced intervals such that readings are uniquely classified into each cell. Here, the cell width is 1.03/10 = 0.10, and we choose the starting point to be 4.395, a convenient value below the minimum of the data and carried out one digit more precise than the data to avoid any confusion in assigning readings to individual cells.
3. Increment the starting point by multiples of the cell width until the maximum value is exceeded. Thus, since the maximum value is 5.46, we generate the numbers 4.395, 4.495, 4.595, 4.695, 4.795, 4.895, 4.995, 5.095, 5.195, 5.295, 5.395, and 5.495. These values will represent the end points or boundaries of each cell, effectively dividing the range of the data into equally spaced class intervals covering all the data points.
4. Construct a *frequency table* as shown in Table 1.2 which gives the number of times a measurement falls inside a class interval.
5. Make a graphical representation of the data by sketching vertical bars centered at the midpoints of the class cells with bar heights proportionate to the number of values falling in that class. This graphical representation shown in Figure 1.1 is called a *histogram.*

A histogram is a graphical representation in bar chart form of a frequency table or frequency distribution.

Table 1.2. Frequency Table of Fuse Data.

CELL BOUNDARIES	NUMBER IN CELL
4.395 to 4.495	2
4.495 to 4.595	2
4.595 to 4.695	8
4.695 to 4.795	15
4.795 to 4.895	14
4.895 to 4.995	13
4.995 to 5.095	16
5.095 to 5.195	15
5.195 to 5.295	11
5.295 to 5.395	3
5.395 to 5.495	1
Total Count	100

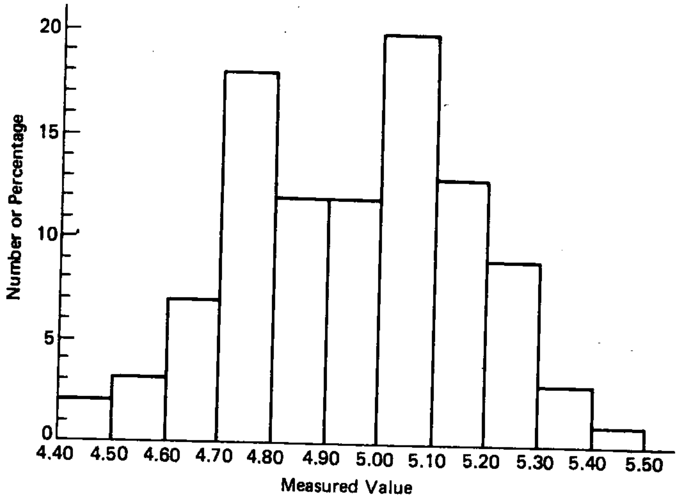

Figure 1.1. Histogram of Measurements.

Note that the vertical axis may represent the actual count in a cell or it may state the percentage of observations in the total sample occurring in that cell. Also, the range here was divided by the number 10 to generate a cell width, but any convenient number (usually between 8 and 20) could be used. Too small a number would not reveal the shape of the data and too large a number would result in many empty cells and a flat appearing distribution. Sometimes a few tries are required to arrive at a suitable choice.

In summary, the histogram provides us with a picture of the data from which we can intuitively see the center of the distribution, the spread, and the shape. The shape is important because we usually have an underlying idea or model as to how the entire population should look. The sample shape either confirms this expectation or gives us reason to question our assumptions. In particular, a shape that is symmetric about a center, with most of the observations in the central region, might reflect data from certain symmetric distributions, like the normal or Gaussian distribution. Alternatively, a nonsymmetric appearance would imply the existence of data points spaced farther from the center in one direction than in the other.

For the data presented in the fuse example, we note that the distribution appears reasonably symmetric. Hence, based on the histogram and the way the ends of the distribution taper off, the manufacturer believes that values much greater or much less than about 10% of the central target are not likely to occur. This variability he accepts as reasonable.

CUMULATIVE FREQUENCY FUNCTION

There is another way of representing the data which can be very useful. By reference to Table 1.2, let us accumulate the number of observations less than or equal to a given value as shown in Table 1.3. Such a means of representing data is called a *cumulative frequency function.*

The graphical rendering of the cumulative frequency function is shown as Figure 1.2. Note the cumulative frequency distribution is never decreasing and starts at zero and goes to the total sample size. It is often convenient to represent the cumulative count in terms of a fraction or percentage of the total sample size used. In that case, the cumulative frequency function will range from zero to 1.00 in fractional representation or to 100% in percentage notation. In this text, we will often employ the percentage form.

Table 1.3 and Figure 1.2 make it clear that the cumulative frequency curve is obtained by summing the frequency function count values. This summation process will later be generalized by integration when we discuss the population concepts underlying the frequency function and the cumulative frequency function in the next section.

THE CUMULATIVE DISTRIBUTION FUNCTION AND THE PROBABILITY DENSITY FUNCTION

The frequency distribution and the cumulative frequency distribution are calculated from sample measurements. Since the samples are drawn from a population, what can we state about this population? The typical procedure

Table 1.3. Cumulative Frequency Function.

UPPER CELL BOUNDARY (UCB)	NUMBER OF OBSERVATIONS LESS THAN OR EQUAL TO UCB
4.495	2
4.595	4
4.695	12
4.795	27
4.895	41
4.995	54
5.095	70
5.195	85
5.295	96
5.395	99
5.495	100

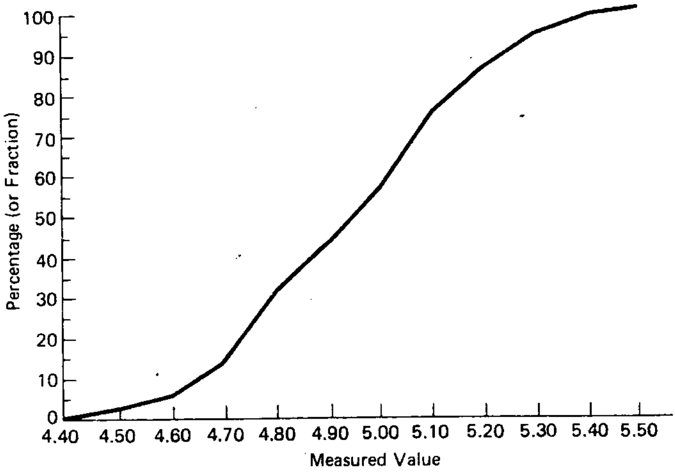

Figure 1.2. Plot of Cumulative Frequency Function.

is to assume a mathematical formula that provides a theoretical model for describing the way the population values are distributed. The sample histograms and the cumulative frequency functions are then estimates of these population models.

The model corresponding to the frequency distribution is the *probability density function* (PDF), denoted by $f(x)$ where x is any value of interest. The PDF may be interpreted in the following way: $f(x)\,dx$ is the fraction of the population values occurring in the interval dx. In reliability work, we often have time t as the variable of interest. So $f(t)\,dt$ is the fraction of failure times of the population occurring in the interval dt. A very simple example for $f(t)$ is called the exponential distribution given by the equation

$$f(t) = \lambda e^{-\lambda t}, \qquad 0 \leq t < \infty,$$

where λ is a constant. The plot of $f(t)$ is shown as Figure 1.3. The exponential distribution is a widely applied model in reliability studies and forms the basis of Chapter 3.

The cumulative frequency distribution similarly corresponds to a population model called the *cumulative distribution function* (CDF), denoted by $F(x)$. The CDF is related to the PDF via the following relationship

$$F(x) = \int_{-\infty}^{x} f(y)\,dy,$$

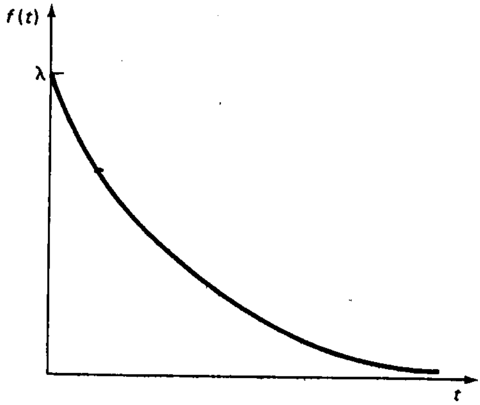

Figure 1.3. Plot of Probability Density Function for Exponential Distribution.

where y is the dummy variable of integration. $F(x)$ may be interpreted as the fraction of values in the population less than or equal to x. Alternatively, $F(x)$ gives the probability of a value less than or equal to x occurring in a single random draw from the population described by $F(x)$. Since in reliability work we usually deal with failure times, t, which are nonnegative, the CDF for population failure times is related to the PDF by

$$F(t) = \int_0^t f(y)\, dy, \qquad 0 \le t < \infty.$$

For the exponential distribution,

$$F(t) = \int_0^t \lambda e^{-\lambda y}\, dy = -e^{-\lambda y}\Big]_0^t = 1 - e^{-\lambda t}.$$

The CDF for the exponential distribution is plotted as Figure 1.4.

When we calculated the cumulative frequency function in the fuse example, we worked with grouped data (data classified by cells). However, another estimate of the population CDF could have been generated by ordering the individual measurements from smallest to largest and then plotting the successive fractions

$$\frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ldots, \frac{n}{n}$$