Chapter 12
Statistics for Engineering Analysis

Chapter Learning Objectives

On completing this chapter readers will have learned about the following topics in the application of statistics in engineering analysis and the introduction of statistical process control in mass production of products:

- The use of statistics in engineering practice.
- Common terminology in statistical analysis.
- Standard deviation and its physical significance.
- The normal distribution of statistical data and its physical significance.
- The normal distribution function, a mathematical model for statistical analysis.
- The Weibull distribution function for probabilistic engineering design.
- The concept of statistical process control.
- The application of statistical process control using control charts for quality assurance in mass production environments.

12.1 Introduction

Statistics is the science of decision making in a world full of uncertainties.

The world in which we live is indeed full of uncertainties. The following are just a few examples of sources of such uncertainty:

- A person's daily routine, career, and health conditions.
- The life expectancies of citizens of different countries of the world.
- Fluctuations of stock markets.
- Weather forecast for regions within countries and the whole world.
- National and global politics.
- Epidemics and natural disasters around the world.

The broad application of statistical methods has made this branch of science a stand-alone major discipline of research and academic programs. A great many archival papers and books have been published in this specialty. It is not possible for this chapter to cover all of the important topics in statistics. The effort will be focused on a small subset, including the most commonly used terminology in statistical analysis and introducing topics that are relevant to engineering analysis with specific applications in probabilistic design analysis and in process control for quality control and assurance of industrial products in a mass production environment. For example, probabilistic design of structures made of brittle materials such as ceramics and cast iron is frequently employed for machines and structures operating in high-temperature environments. Extensive treatments of the use of statistics for solving engineering problems are available in numerous books (see, for example: Stout, 1985; Vardeman, 1994; Leemis, 1995; Rosenkrantz, 1997; Morrison, 2009; Navidi, 2013).

Statistical methods are playing increasingly important roles in nearly all phases of human endeavor (Spiegel, 1961). The reason why statistical methods are becoming more popular is that the world in which we live has become ever more complicated in many aspects that involve increasing numbers of uncertainties, and the traditional methods based on deterministic concepts can no longer deal with these complex issues and problems. The use of statistics has accordingly expanded to encompass nearly all aspects of human activity. The following are some of these areas of application.

- *Agriculture:* According to the U.S. Department of Agriculture, statistical analyses offer key indicators, outlook analysis, and a wealth of data on the food and agricultural system, along with information on farming practices, structure, and performance. They also produce data on such diverse topics as farm and rural households, commodity markets, food marketing, agricultural trade, diet and health, food safety, food and nutrition assistance programs, natural resources and the environment, and the rural economy.
- *Biology:* There are many applications of statistics in studying the cell theory of structural growth, evolution, and taxonomy of living species and organs. Statistics is also a powerful tool in the study of the physiology of tissues and organs, as well as in genetics, involving such critical areas as DNA/RNA conversion. Bioinformatics is a relatively new interdisciplinary field of science that combines computer science, statistics, mathematics, and engineering to study and process biological data.
- *Business:* Statistics is a powerful tool for developing inferences about certain characteristics of a population in the business domain, which may include people, objects, or collections of information. Such inferences may lead to critical actions in financial analysis, econometrics, auditing, production and operations including service improvements, and marketing research. Statistics is also widely used by business communities in performance management and strategic planning that includes critical components such as alternative scenarios.
- *Economics:* The data of concern to economic statisticians may include those of an economy of region, country, or group of countries. Analyses within economic statistics both make use of and provide the empirical data needed in economic research, whether descriptive or econometric. They are a key input for decision making as to economic policy. The subject includes statistical analysis of topics and

problems in microeconomics, macroeconomics, business, finance, forecasting, data quality, and policy evaluation. It also includes such considerations as what data to collect in order to quantify some particular aspect of an economy and how best to collect it in any given instance.

- *Electronics and communication:* A major application of statistical methods in electronics involves noise in electronics circuits and systems. Noise and frequency modulations are major problems in the design of sound systems for communication. Another major area of application of statistical methods is in the market evaluation of consumer electronic products, which is essential to developing strategies for new products.
- *Healthcare:* Statistical methods are used to provide information for understanding, monitoring, improving, and planning the use of resources for the benefit of people's lives, thereby facilitating the provision of services and promoting their wellbeing. Health statistical data help us to understand the impacts of health on people and to work for their betterment. As we monitor the health of a population, we enhance our understanding of strategies to promote its health.
- *Medicine:* Statistical methods are used in clinical trials, epidemiological studies, and new drug discovery and delivery.
- *Physics:* Probability theory, a major branch of statistics, is used to deal with large populations and approximations in many physical problems. It can describe a wide variety of fields with an inherently stochastic nature. In particular, statistical mechanics develops the phenomenological results of thermodynamics from a probabilistic examination of the underlying microscopic systems. Historically, one of the first topics in physics where statistical methods were applied was the field of mechanics, which is concerned with the motion of particles or objects when subjected to forces.
- *Politics and social sciences:* Major application areas involve policy development in circumstances of uncertain geopolitics and public opinion, sociology in population growth, actuarial studies, censuses, crime statistics, demography for social welfare planning, and numerous other fields such as climate change and environmental studies.

These broad applications of statistics have established it as a specialized academic discipline in many higher educational institutions. It is not possible to treat even a small fraction of the territory of statistical analysis in this chapter. We will cover only a narrow span of this vast field of statistics with the introduction of commonly used terminology in statistical analyses followed by applications in statistical process control, with a couple of control charts for quality control of mass-produced products. An introductory section will be included on reliability design concepts using the Weibull distribution for engineering systems involving materials with random strength data.

12.2 Statistics in Engineering Practice

In Chapter 1 we learned that one of the three principal functions of engineers is “decision making.” Proper decisions were indicated that were made on sound engineering principles through credible analyses. Such an approach may have worked well in the past. However, as technologies have evolved at an unprecedented pace in recent times, the demand for accurate results in all aspects of engineering practices has become ever stronger—yet practical engineering activities are full of uncertainties as will be demonstrated in the following description. The statistical approach, in sharp contrast to the traditional deterministic approach, has offered engineers a viable alternative in handling problems involving many uncertainties.

There are indeed many uncertainties in almost all engineering activities. The following are just a few examples; there are many uncertainties in systems design and analysis. Section 1.4 outlined four distinct stages involved in many engineering analyses. Of these four stages in the analysis, Stage 2 relates to idealization of real physical situations that engineers have to deal with. The areas that need idealizing include the geometry and loading and support conditions that engineers would not have the available analytical tools to deal with. There are other areas in which the engineers again may not be certain in their analyses, including the following.

- *Uncertainties in material properties*. Among the most common inputs to any engineering design analysis are the thermophysical properties of the materials to be used in the intended engineering system. A simple stress analysis of a uniaxially loaded bar requires the Young's modulus (E) of the material. An additional property called the Poisson's ratio (v) needs to be included in the generalized Hooke's law for multiaxially loaded solids, as illustrated in Chapter 11. Thermal conductivity (k) is a required property in heat conduction analysis (Equation (7.24)), and other properties such as permittivity ( ) are required in electromagnetic analysis in Equation 3.33a. Most of these properties are available in professional handbooks (Avallone *et al.*, 2006).

  However, one must realize that all the material properties that are available in textbooks and handbooks are the “average” values of the measured data from testing of many samples, with the assumption that the sample materials are homogeneous (and isotropic, as in many cases). The reality, however, is that no material can be as homogeneous as one would assume it to be. One would find that the sample materials for property measurements are full of minute voids and cavities, and with grains oriented in random directions, as can be observed under electron microscopes. These randomly occurring micro voids with random orientations can make the measured material properties uncertain in many ways, and this uncertainty will be amplified many-fold in the design analysis of engineering systems at micro- and nanoscales (Hsu, 2008), in which device structures are of molecular or atomic scale. Any uncertainty in material homogeneity and purity will produce significant error in engineering analyses.

- *Methodology of engineering analysis*. Many engineering analyses involve mathematical modeling as described in Chapter 2. Derivation of the mathematical model adopted in an engineering analysis would involve a number of assumptions and hypotheses such as were illustrated in Chapters 7, 8, and 9. Additionally, there are times at which engineers need to use numerical techniques such as the finite-difference or finite-element methods as described in Chapters 10 and 11; one will readily appreciate that these numerical methods, though popular and powerful, are used to obtain only approximate solutions due to the many assumptions required in their formulation and that they also include accumulative rounding-off and truncations errors in computation. Thus, particular analytical methods also introduces uncertainties into the analyses.

- *Fabrication methods*. Once a system is beyond the design stage, it will be subject to another major source of uncertainty in the fabrication of the designed systems into products. Huge uncertainty may arise in fabrication or process methods, in selection of machine tools or process procedures, and in quality assurance techniques applied to the finished products. Human factors will also come into play because operators involved in the fabrication process will have varying levels of experience and skill.

These uncertainties that can arise in design methodologies, material properties, and fabrication techniques constitute a lively example of the well-known “Murphy's Law,” which stipulates that “*Anything* *that can go wrong will go wrong*." Many of the factors in engineering systems can indeed "go wrong."

12.3 The Scope of Statistics

Statistics is concerned with developing and applying scientific methodology in the following major activities.

Collecting relevant information

Methods are required for collecting relevant information for statistical analysis according to the nature and the extent of the analysis. It is a costly activity but it is a critical part of the process because the data collected for analysis can affect the accuracy and credibility of the analysis. The sizes of the datasets and the sources from which the data are collected may significantly affect the outcome of the analysis.

There are generally four sources from which the analysts may collect relevant data for their analyses: (a) from the entire population; (b) from a subset of the population; (c) from controlled studies undertaken to understand cause-and-effect relationships; and (d) from whatever existing studies attempting to understand the cause-and-effect relationships on the subject of interest. The latter source offers the least costly option in collecting data, but the analyst does not have any control of the populations that generated the data or the conditions that dictated the generation of the collected data.

Organizing the collected information

Once relevant information for statistical analysis has been collected, the analyst must determine the best strategy for organizing and analyzing the collected information by developing databases.

There are two types of databases commonly used in statistical analysis: (a) quantitative databases for the information collected in numerical form, such as that on the frequency of specific behaviors; and (b) qualitative databases for nonnumerical information, such as responses gathered through interviews, observations, focus groups, or survey questionnaires. Surveys on market demands for certain products by business and industry are in this category.

Datasets in quantitative databases usually involve a set of numbers in either ascending or descending order of magnitude: for example, the measured temperature in an experiment at specific instants, or a particular dimension of a machine component manufactured by mass production.

Three elements are required for well-organized databases: (a) they should carry a unique identifier; (b) there should be prescreening and exclusion of incorrect information according to the analyst's best judgment; (c) the collected information should be entered into the database in a consistent format.

Summarizing and presenting the information to concerned parties

Databases constructed according to the above criteria are the basis for the subsequent analyses that will allow conclusions or proper decision making by senior managers of business and industry. Often the analyst may be required to present the essential information that the database itself expresses without a detailed analysis using the collected data. It is thus important that the analyst be able to *summarize* the collected information and present this information to the concerned parties. Key items the analyst may present to characterize a quantitative database include the “frequency distributions,” the “central tendency,” and “variability” of the dataset. Presentation of the findings from information collected on qualitative data is more of a challenge to the analyst. The analyst would present the findings that appear to correspond to the choice of which information should be emphasized, minimized, or omitted in the analysis.

Analyzing data to generate valid conclusions and allow reasoned decision making on the basis of such analysis

A major task of the analyst at this stage is performing an inferential analysis of the database that he or she has established. The objective of this effort is to establish the “significance” in the relevant aspect offered by the dataset. “Statistical significance” derived from probability theory is often used to assess the validity of the dataset. This indicates whether a result is more probable than what would have been expected due to random error. For the dataset to be significant, it must have high probability that the results were not produced by chance. Common statistical tests such as “chi-square” are often used to assess the “goodness of fit” of the collected data with what is expected, such as the normal distribution, as will be elaborated later in the chapter.

Example 12.1

A local firm in Silicon Valley fabricates a batch of microchips. The quality assurance engineer took 5 samples from a bin containing mass-produced chips, and measurements of the dimensions of these chips were made at three predetermined locations on each sample. The measured data that the engineer collected and recorded are tabulated below:

Sample 1:	2.15	2.35	1.95
Sample 2:	2.70	1.83	2.25
Sample 3:	1.97	2.03	2.13
Sample 4:	2.06	2.70	2.15
Sample 5:	2.03	1.75	1.85

All measured data have the same unit of millimeters. Organize the collected data in ascending order for further analysis.

Solution:

The set of data is organized in an ascending order as

1.75, 1.83, 1.85, 1.95, 1.97, 2.03, 2.03, 2.06, 2.13, 2.15, 2.15, 2.25, 2.35, 2.70, 2.70

12.4 Common Concepts and Terminology in Statistical Analysis

The following terminology is frequently used in statistical analysis using databases created by quantitative data collection. For illustration purpose, we present only cases with small sample sizes with small numbers of data, which does not normally represent in reality.

12.4.1 The Mode of a Dataset

The *mode* of a dataset is that value (or more than one value) that occurs with the greatest frequency: that is, it is the most common value or values in the dataset. There are cases such that the mode may not exist in a dataset. In other cases there may be more than one mode for the dataset, as will be demonstrated in the following examples.

Example 12.2

Find the mode in each of the following datasets:

a. 2, 2, 5, 7, 9, 9, 9, 10, 10 11, 12, 18.

b. 3, 5, 8, 10, 12, 15.

c. The dataset that appears in Example 12.1.

Solution:

a. The mode of this dataset is 9 because this number appears three times in the dataset, more frequently than any other datum in the set.

b. This dataset has no mode because there is no number in the set that appears more than once.

c. The dataset appears in Example 12.1 has triple modes of 2.03, 2.15, and 2.70 since each of these numbers appears twice in the dataset.

12.4.2 The Histogram of a Statistical Dataset

A *histogram*—often called a "frequency distribution diagram"—is most commonly used by statisticians and engineers for expressing the physical sense of a set of data. General rules for establishing this type of diagram are outlined as follows:

1. Identify the largest and smallest numbers in the entire dataset, and thus the overall range of the collected data.

2. Establish a convenient number of intervals of the same size (value) within the overall range of the collected data.

3. Determine the number of the data falling into each of the set intervals. These numbers will be the "frequency" of the data in each of the set range in the histogram.

Example 12.3

Establish a histogram of the measured dimensions of the microchips presented in Example 12.1.

Solution:

We have conveniently divided the range of the measured length between 1.75 mm and 2.7 mm into 10 intervals as tabulated below:

Intervals (mm)	1.7–1.8	1.8–1.9	1.9–2.0	2.0–2.1	2.1–2.2	2.2–2.3	2.3–2.4	2.4–2.5	2.5–2.6	2.6–2.7
Sample number in each interval	1	2	2	3	3	1	1	0	0	2

The histogram corresponding to the tabulated numbers is plotted in Figure 12.1, in which the vertical axis represents the frequency of the data in the dataset.

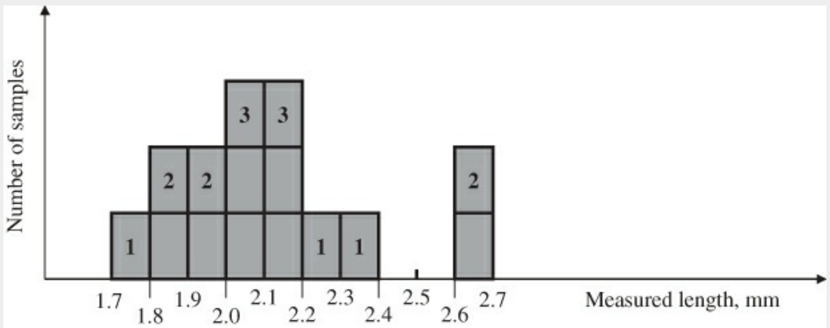

Figure 12.1 Histogram of measured lengths of chips.

Example 12.4

Establish a histogram of the following tabulated marks that the students in a class earned in their final examination:

Test scores	45–49	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85–89	90–94
Frequency	1	2	2	5	11	9	10	5	4	2

Solution:

The corresponding histogram of the above mark distribution is illustrated in Figure 12.2.

The data presented in Example 12.4 resulted in an approximately "symmetrical bell-shaped" histogram as shown in Figure 12.2. The corresponding data that form the histogram are referred to as a "normally distributed dataset." So-called *normal distribution* of datasets is common for many physical phenomena, and the resulting bell-shaped histogram produced by these datasets is referred to as the *normal distribution* by statisticians. *Most statistical data from large sample sizes fit well with bell-shaped distribution curves.*

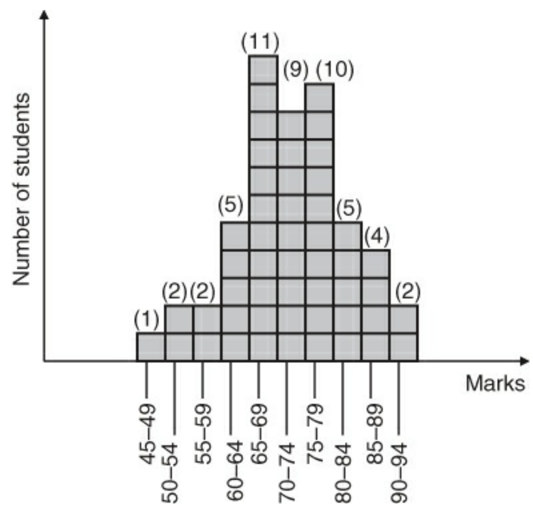

Figure 12.2 Mark distribution of students in an engineering analysis class.

12.4.3 The Mean

The *mean* of a dataset is the arithmetic average of all data in the set. It is a popular measure of "central tendency" of the dataset.

Given a data set of *n* numbers represented by $x_1, x_2, x_3, \ldots, x_n$, we may express the summation of all numbers in the dataset as $\sum_{i=1}^{n} x_i$. The arithmetic mean, $\bar{x}$ of the dataset may be obtained by the following expression:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (12.1)$$

Example 12.5

Determine the mean of the dataset that was given in Example 12.1:
1.75, 1.83, 1.85, 1.95, 1.97, 2.03, 2.03, 2.06, 2.13, 2.15, 2.15, 2.25, 2.35, 2.70, 2.70

Solution:

We find the sum of the 15 numbers (n = 15) in the dataset to be 31.9. Thus, for the mean of this set data Equation 12.1 gives

$$\bar{x} = \frac{31.9}{15} = 2.127$$

Use of the mean in statistical analysis has many advantages:

1. It includes *all* the data in the set.
2. It always exists.
3. It is usually reliable in representing the "central tendency" of the dataset.

A disadvantage of using the mean in statistical analysis is that "significant error" of use of the mean to represent the central tendency of the dataset can occur with some extreme outlier numbers appearing in any part of the dataset, as will be demonstrated in the following example.

We may readily compute the mean of the dataset: 2, 3, 5, 7, 9, 11, 13 to be 7.14 which is close to the "central value" of the dataset. However, the same dataset but with the value of the last datum in the dataset replaced by a large number, say 73, will have a mean of 15.71 instead. This value is far from being the "central value" of the dataset with the last datum of 73. We thus observe that the central tendency of a dataset cannot be represented by the mean of a dataset of the latter type. We must recognize that although the mean may represent the central tendency of a dataset in most cases, there are instances when other methods need to be used to establish the "central tendency" of the dataset.

12.4.4 The Median

Like the mean, the *median* of a dataset is also used to represent the central tendency of a dataset. The "central" datum is readily identified in datasets with an odd number of data. For datasets with an even number of data, the median is the average of the two central data. For example, the median of the dataset 5, 9, *11, 14*, 16, 19 is (11+14)/2 = 12.5.

The difference between the mean and median is that the latter represents the data at the "center" when the dataset is arranged in either ascending or descending order. For example, the median value (7+9)/2 = 8 of the ambiguous dataset 2, 3, 5, 7, 9, 11, 13, 73 that we used earlier may represent the central tendency of the dataset much better than the mean does.

The median is frequently used to represent the central tendency of large datasets involving a few extreme data values, or of datasets with large fluctuations. Examples are such things as the median income of a population or the median price of houses in places like Santa Clara Valley in California, where the median but not the mean house price is more representative in a market with prices ranging from close to $1 000 000 for the majority houses to a smaller number of houses valued at many millions of dollars.

12.4.5 Variation and Deviation

The degree to which numerical data deviate from the "average" value (the mean or median of a dataset) is called the "variation" or "dispersion" of the data in the dataset. Deviation is one of the measures of variation in a statistical analysis. It is a measure of how the data in a set deviates from its mean value of $\bar{x}$. In the world of statistics, small deviation represents a more uniform dataset with less "scatter" from the mean. Thus, for a dataset of n data with $x_1, x_2, x_3, \ldots, x_n$, the variation of individual data can be determined by simply computing the individual variations: $(x_1 - \bar{x}),\ (x_2 - \bar{x}),\ \ldots,\ (x_n - \bar{x})$. The total deviation of the entire dataset will be the sum of all the individual deviations. Unfortunately, one will find that the deviations sum to zero:

$$\sum[(x_1 - \bar{x}) + (x_2 - \bar{x}) + (x_3 - \bar{x}) + \cdots + (x_n - \bar{x})] = 0$$

which leads to an unrealistic implication that there is *no* overall variation of the data in the dataset. A close inspection of the components in the above summation will indicate that half of the individual components in the summation carry negative sign whereas the other half of the terms carry positive sign. The net value of the summed components becomes zero, as expected.

To avoid this situation of zero overall deviation as demonstrated in the above summation, one may use an alternative expression using the squares of the deviations to represent the overall deviation of the dataset:

$$\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad \textbf{12.2}$$

One will observe in Equation 12.2 that all individual (*squared*) terms of variation in Equation 12.2 will carry positive sign, and the result of the square root of the summed variations will be a real nonzero number.

12.5 Standard Deviation (σ) and Variance (σ^2)

12.5.1 The Standard Deviation

The standard deviation (σ) for a dataset is a measure that is used to quantify the amount of variation or *dispersion* of the data values in the dataset. A standard deviation close to 0 indicates that the data points tend to be very close to the mean (the central tendency, also called the *expected value*) of the set, while a high value of standard deviation indicates that the data points are spread out over a wider range of values. Mathematically, the standard deviation of a dataset may be computed from Equation 12.3:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}} \qquad \textbf{12.3}$$

where n is the number of data in the dataset.

12.5.2 The Variance

Like standard deviation, variance is also used as a measure of scatter of data in a datasets; it has the following properties:

1. It is proportional to the scatter of the data (small variance means the data are clustered together, and a large variance results from a widely scattered dataset).

2. It is independent of the number of values in the dataset.

3. It is independent of the mean (since now we are only interested in the spread of the data, not its central tendency).

The mathematical relation for the variance is derived from Equation 12.2 and is given by the following expression:

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1} \qquad \textbf{12.4}$$

If a very large number of data are involved in the dataset and only a partial dataset of n values is used to determine the standard deviation, we use the following modified expression for standard deviation:

$$\sigma = \sqrt{\frac{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n - 1)}} \qquad \textbf{12.5}$$

where n is the selected number of data used in the computation.

Example 12.6

Determine the standard deviation and the sample variance of the dataset:
5, 9, 11, 14, 19

Solution:

We see that there are five numbers in the dataset, so n = 5. The mean value is calculated to be $\bar{x}$ = 11.6 using Equation 12.1. The standard deviation of the dataset can be obtained using Equation 12.3 as

$$\sigma = \sqrt{\frac{(5-11.6)^2 + (9-11.6)^2 + (11-11.6)^2 + (14-11.6)^2 + (19-11.6)^2}{5-1}}$$
$$= 5.27$$

The sample variance is $\sigma^2 = (5.27)^2 = 27.8$.

12.6 The Normal Distribution Curve and Normal Distribution Function

The "normal distribution" is the most frequently encountered statistical distribution pattern, appearing in many natural circumstances, such as the age distribution of the citizens of a country, the annual temperature variations in a region, the performance of certain machines and devices, and so on. The data in overwhelming numbers of datasets, when plotted as histograms such as shown in Figure 12.2, will have envelopes that are the bell-shaped curves known as "normal curves," or "normal distribution curves." These distribution curves are usually presented with the mean of the dataset, such as that shown in Figure 12.2, relocated to the center of the distribution of the data.

The common occurrence of data in bell-shaped distributions motivated mathematicians and statisticians to develop mathematical models that can be used to represent this common statistical data distribution. This development evolved from the Gaussian distribution function that often appears in probability theory. Probability theory is a branch of mathematics concerned with the analysis of random phenomena. The outcome of a random event (such as the academic performance of the class in Example 12.2) cannot be determined before it occurs, but it may be any one of several possible outcomes. The actual outcome is considered to be determined by chance. In this theory, the normal (or bell-shaped) distribution is a very common *continuous* probability distribution of random variables, and the Gaussian function that is symmetrical about the central tendency of the dataset is widely used to assess the likelihood of observing the expected outcomes.

The *probability density* of a normal distribution in probability theory is given by:

$$f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \quad \textbf{12.6}$$

(see https://en.wikipedia.org/wiki/Probability_density_function). Here, μ is the *mean* or *expectation* of the distribution (and it can also equal the median and mode of a statistical dataset). The parameter σ is its standard deviation, with its variance being σ^2. A random variable with a Gaussian distribution is said to be *normally distributed* and is called a *normal deviate*.

Thus, if we plot the histogram in Figure 12.2 for the academic performance of students in a class by moving the vertical axis for the population density to coincide with the central value of the dataset—the mean—and connect the peaks of each individual attribute of mark range, we will have the modified distribution in a continuous curve expressed as solid line that is a close approximation of the bell-shape, as illustrated in Figure 12.3.

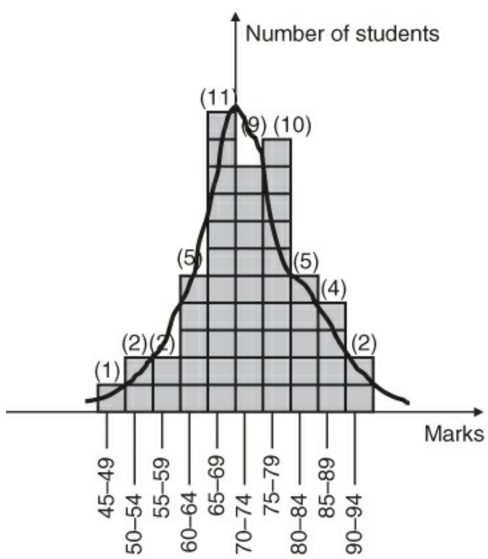

Figure 12.3 Approximated normal distribution of mark distribution of a class.

Because the bell-shaped normal distribution represents the statistical variation of the overwhelming number of natural phenomena, it can be graphically represented by what is termed the “normal curve” as shown in Figure 12.4, with a corresponding mathematical function that is called the “normal distribution function” as follows.

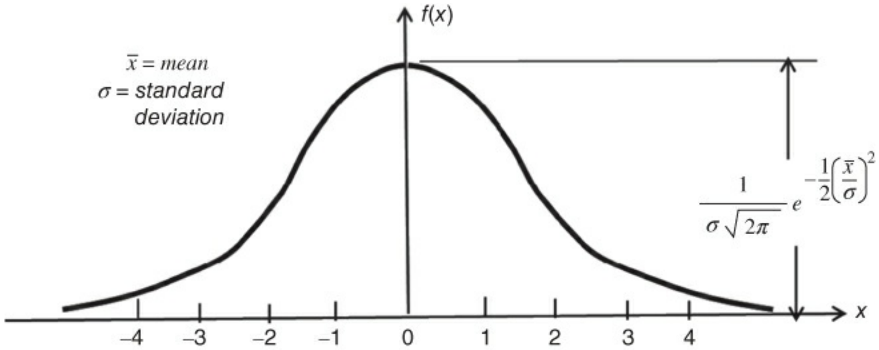

Figure 12.4 Normal distribution curve.

The mathematical expression that represents the normal curve or the normal distribution function takes the form

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(1/2)[(x-\bar{x})/\sigma]^2} \quad \textbf{12.7}$$

in which σ is the standard deviation of the dataset as given in Equation 12.3, and $\bar{x}$ is the mean of the dataset as in Equation 12.1.

We note that the normal distribution function in Equation 12.7 is effectively identical to the Gaussian distribution function in Equation 12.6.

There are many advantages of having the normal distribution curve represented by the normal distribution function of Equation 12.7, as many useful mathematical models for scientific and engineering observations can be developed using this function. For example, the following important properties of observed phenomena have been derived using the normal distribution function of Equation 12.7:

1. The normal curve (or normal distribution) is symmetrical about the mean $\bar{x}$.

2. Being a credible mathematical model with a continuous function that is valid for a large number of natural phenomena the normal distribution function serves as a basis for mathematical modeling of other statistical analyses.

3. With the function in Equation 12.7, the standard deviation, σ, of a dataset that fits a normal distribution can be interpreted in the following interesting and extremely valuable way:

68.26% data of the set are within the value of the mean $\bar{x}$ ± 1 standard deviation (σ).

94.40% data of the set are within the value of the mean $\bar{x}$ ± 2σ.

99.73% data of the set are within the value of the mean $\bar{x}$ ± 3σ.

These properties are illustrated in Figure 12.5.

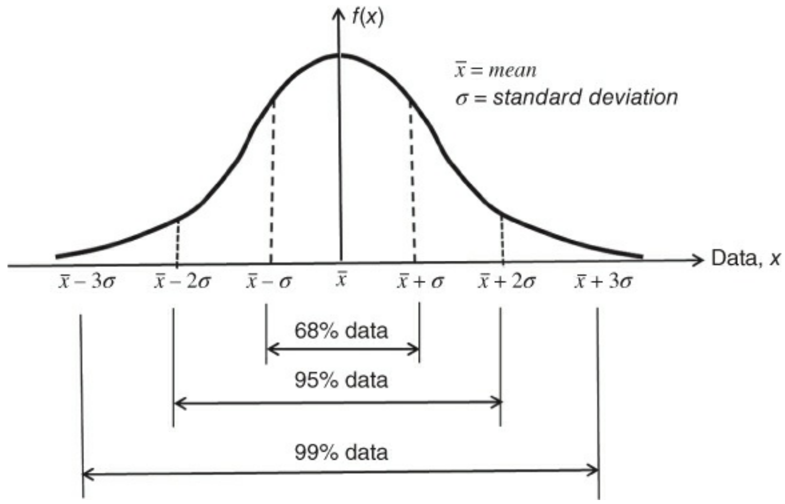

Figure 12.5 Properties of the normal distribution.

Example 12.7

A tire manufacturing company supplies tires for 200 000 cars each year. Of these cars, 15%, or 30 000 cars, were used to evaluate the life of its tires. Test results indicated that the average life of the tires involved in these tests was 45 000 miles with a standard deviation of 4000 miles. Determine the lives of the tires produced by the company.

Solution:

The mean value of the lives of the tires in the tests was $\bar{x}$ = 45 000 miles with a standard deviation σ = 4000 miles. We assume that all measured lives of the tires fit a normal distribution (or normal distribution curve) as shown in Figure 12.4. We may then interpret the test results for the tire lives in the following way:

68.26% cars had tire life of $\bar{x} \pm \sigma$ = 45 000 ± 4000 miles,

94.40% cars had tire life of $\bar{x} \pm 2\sigma$ = 45 000 ± 8000 miles

99.73% cars had tire life of: $\bar{x} \pm 3\sigma$ = 45 000 ± 12 000 miles.

12.7 Weibull Distribution Function for Probabilistic Engineering Design

Like the normal distribution function (or Gaussian distribution function in probabilistic analysis), the Weibull distribution function is another continuous function that can be used for physical (or engineering) phenomena that are significantly skewed from the normal distribution illustrated in Figure 12.4. It is particularly useful for dealing with engineering analyses with significant scatter of analytical parameters such as material property inputs to a design analysis. We will focus on its application in this section.

12.7.1 Statistical Approach to the Design of Structures Made of Ceramic and Brittle Materials

There has been increasing use of ceramic or cermet (composites of metal and ceramics) by the aerospace and nuclear industries for components operating at very high temperature but requiring light weight. The ceramic SiC has a mass density of 3200 kg/m^3, which is only 41% of that of structural steel, but has a Young's modulus of 375 000 MPa, that is almost twice that of steel. SiC also has a melting point of 4263 K vs. 1703 K for structural steel. Not only does the much higher melting point of SiC provide much higher strength at high temperature, but SiC is also much less vulnerable to creep deformation, which normally occurs at an operating temperature above half of a material's homologous melting point (that is, half of the melting point on the Kelvin scale).

Despite the attractive strength of SiC at high temperature, it is brittle, and with randomly varying material properties—in particular, its fracture strength—presents major challenges to engineers using well-established deterministic mathematical models in their design analysis, such as was illustrated in Figure 1.4. Table 12.1 shows the fracture strength of SiC produced by 4-point bending tests at the author's laboratory at the Whiteshell Nuclear Research Establishment of Atomic Energy of Canada Ltd. in 1970. It shows that the fracture strength of the Carborundum KT grade SiC specimens had wide scatter in the measured fracture strength, ranging from 15 ksi (103.5 MPa) to 21 ksi (147 MPa).

Table 12.1 Fracture strength of silicon carbide in 4-pont bending tests

Order	Ordered fracture strength (ksi)	Survival probability, R	Order	Ordered fracture strength (ksi)	Survival probability, R
1	15.28	0.9678	16	18.67	0.4838
2	15.35	0.9355	17	18.86	0.4516
3	16.96	0.9032	18	19.04	0.4194
4	17.15	0.8710	19	19.30	0.3871
5	17.65	0.8387	20	19.57	0.3548
6	17.85	0.8065	21	19.59	0.3226
7	18.08	0.7742	22	19.72	0.2903
8	18.17	0.7419	23	19.96	0.2581
9	18.36	0.7097	24	20.26	0.2258
10	18.39	0.6774	25	20.68	0.1935
11	18.43	0.6451	26	21.11	0.1613
12	18.45	0.6129	27	21.18	0.1290
13	18.50	0.5806	28	21.31	0.0968
14	18.63	0.5484	29	21.55	0.0645
15	18.65	0.5161	30	21.57	0.0323

(Hsu and Gillespie, 1971.)

Survival probability $R = 1 - i/(n+1)$ where i is the order number and n = total number of specimens = 30.

Like many other ceramic materials, the fracture strength of KT-SiC also shows strong variation with the regions that are subjected to tensile stresses; above all, the properties vary significantly with the volume of the structure, which leads to strong size effects in design analysis. In sharp contrast, metallic materials such as structural steel exhibits much less scatter in the strength values. A report (Barnett and McGuire, 1966) indicated that an investigation involving 30 000 tons of structural steel showed only two out of 3124 specimens to yield below the minimum specified value of 33 ksi (227.7 MPa), and they fell no lower than 31 ksi (21.4 MPa). A definitive allowable strength for structural steel is therefore possible.

Similar inconsistent fracture strength also occurs in conventional metallic materials that are embrittled at low temperature. The highly inconsistent mechanical strength of ceramics, cermet and brittle materials alike, with the fracture strength of KT-SiC as in Table 12.1, is mainly attributed to the lack of plastic deformation in these materials. Consequently, sharp stress concentration near the tip of inherent cracks inside the material prompts these cracks to propagate through the specimens or the structure because of lack of plastic yielding of the materials. This type of structural failure fits the statistic model referred to as "failure by the weakest link," meaning that the failure of the weakest of the many links that make up the structure is what causes the overall structure to fail. The random nature of inherent cracks in brittle materials creates weakest links inside materials of this type, which leads to the failure in random fashion of specimens or structures made of brittle materials.

The significant inconsistency of material strength of ceramic and brittle materials, like that of KT-SiC as shown in Table 12.1, precludes the use of the conventional deterministic theory in engineering design analysis such as illustrated in Examples 1.1 and 1.2 in Section 1.5 of Chapter 1. A radically different approach to design analysis of structures made of these materials is thus required.

The statistical approach for the design of structures made of ceramic and brittle materials is derived using the Weibull distribution function, with the strength of the materials measured by special techniques, and with the specimen geometry being such as described by Barnett and McGuire (1966). The "interpretation of results" for the design analysis described in the Stage 4 of engineering analysis in Chapter 1 is no longer to keep the maximum stress in a structure below the maximum allowable stress, because no definitive allowable stresses can be established with this type of material. Rather, one will use the term "reliability" of the structure as the design criterion instead. This approach of using statistical methods is referred to as *probabilistic analysis*.

12.7.2 The Weibull Distribution Function

Statistician Waloddi Weibull proposed the well-known eponymous distribution function for structures subjected to applied load with induced stress σ (Weibull, 1939). The mathematical expression of this distribution function is given in Equation 12.8:

$$P = 1 - e^{(-r)} \tag{12.8}$$

where P is the probability of fracture of the material, and r is the risk of rupture.

The risk of rupture r of a given material in Equation 12.8 is obtained by an integration as in Equation 12.9:

$$r = \int_v f(\sigma)\, dv \tag{12.9}$$

in which v is the volume of the material under tensile load, and $f(\sigma)$ is a function of stress distribution (or variations) in the material of volume v. We note that the Weibull distribution function is constructed on the hypothesis that compressive stresses in material will not result in the failure of the structure.

The original form of $f(\sigma)$ in the proposed Weibull distribution function has two parameters, as in the Equation 12.10:

$$f(\sigma) = \left(\frac{\sigma}{\sigma_0}\right)^m \tag{12.10}$$

in which m is the "Weibull modulus," which is a measure of the scatter of the collected data, and σ_0 is a "Weibull parameter." In general, this parameter is related to the gap between the lower bound and the mode of the dataset.

The Weibull distribution function for the probability of rupture of a material with the "risk of rupture *r*" is illustrated in Figure 12.6. In the figure, a probability of fracture $P = 1$ indicates a 100% likelihood of fracture of the sample specimen or structure.

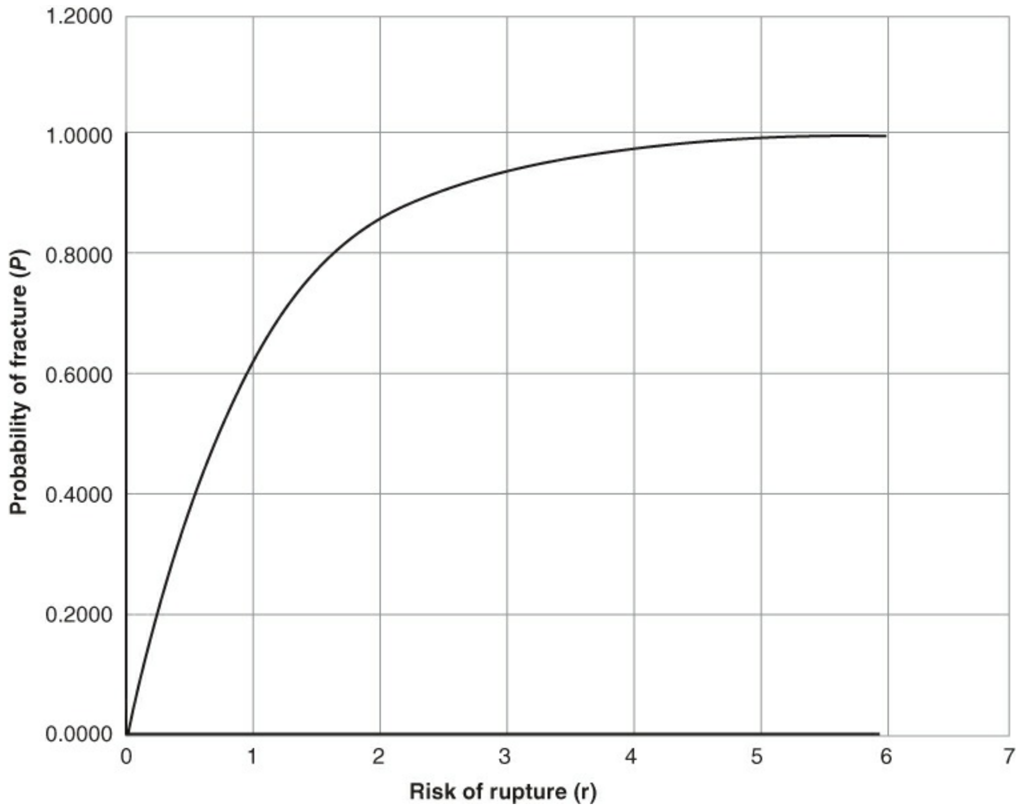

Figure 12.6 Weibull distribution function.

One should be aware that Weibull did not attempt to justify his distribution function theoretically. However, the distribution function of Equation 12.8 did show good fit to experimental strength data of many materials ranging from cotton fibers to steel.

In 1951, Weibull proposed to add a lower bound σ_u to his distribution function in Equation 12.8 giving the following modified distribution function (Weibull, 1951):

$$P = 1 - \exp\left[-KV\left(\frac{\sigma - \sigma_u}{\sigma_0}\right)^m\right] \quad \mathbf{12.11}$$

where K is the load factor and V is the volume of the specimen or the structure with induced tensile stresses due to externally applied loads. The stress σ_u is referred to as the "zero probability of rupture strength," or the "threshold stress," below which no fracture of specimens (or structure) occurs.

The value of the load factor K in Equation 12.11 is determined by how the specimens used to determine the material fracture strength are loaded (Gregory and Sprill, 1962). For instance, $K = 1$ is for a specimen subjected to uniform tensile load, and

$$K = \frac{1}{6(m+1)}\left(\frac{2}{m+1} + 1\right)$$

for 4-point third span (the center span) bending tests as illustrated in Figure 12.7.

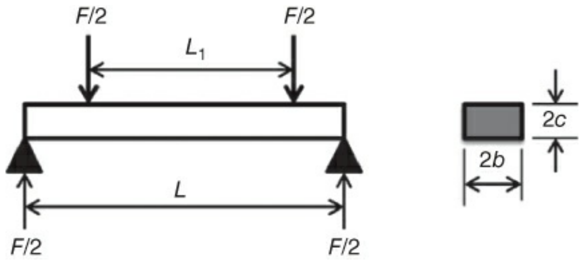

Figure 12.7 Four-point bend test specimens for brittle materials.

Fracture strength testing for ceramic or brittle materials such as KT-SiC is often performed in 4-point bending tests with a specimen configuration shown in Figure 12.7. The loading arrangement shown in this figure results in "pure bending" of the central portion (the third span) of the beam. This mode of beam bending makes it easy to identify the region that is subject to "pure" tensile load. And more importantly, this arrangement can mitigate the parasitic stresses on the specimen induced by gripping the specimen both ends.

The following measurements were included in producing the fracture strength data for KT-SiC in Table 12.1:

Length of the specimen $(L) = 1.75$ inches

Length of the third span $(L_1) = 1.0$ inch

Volume of the specimen $(V) = 2c \times 2b \times L = (2 \times 0.1259)(2 \times 0.01383) \times 1.75 = 0.0541\ \text{in}^3$

Loading factor (K):

$$K = \frac{1}{2(m+1)}\left(\frac{0.4286}{m+1} + 0.5714\right)$$

12.7.3 Estimation of Weibull Parameters

Three methods are available for engineers to determine the Weibull parameters appearing in Equation 12.11. These are (1) the log-log plot method; (2) the method of statistical moments (Gregory and Sprinill, 1962); and (3) the least mean squares approximation (Gregory and Sprinill, 1962).

We will present the log-log plot method here for its simplicity. This process begins with re-writing Equation 12.11 in the following form as a linear relationship:

$$\ln\ln\left(\frac{1}{1-P}\right) = m\ln(\sigma - \sigma_u) - m\ln(\sigma_0) + \ln(KV) \quad \textbf{12.12}$$

One may compute the probability of fracture of each specimen from the testing data listed in Table 12.1.

The following are the major steps in determining the Weibull parameters in Equation 12.11:

Step 1: Assuming an arbitrary value of the threshold stress σ_u.

Step 2: Compute $\sigma_i - \sigma_u$; that is, $\sigma_1 - \sigma_u$, $\sigma_2 - \sigma_u$, $\sigma_3 - \sigma_u$, $\sigma_4 - \sigma_u$, ...

Step 3: Compute $\ln(\sigma_i - \sigma_u)$ obtained in Step 2.

Step 4: Plot $\ln(\sigma_i - \sigma_u)$ against $\ln\ln[1/(1-P)]$ and obtain one curve plot (name this as "plot A")

Step 5: If the curve plot is not a straight line as expected, begin the iteration with another assumed value of σ_u and repeat Steps 1 to 4 until a reasonably straight line plot is obtained (call this curve plot "plot B"). The last assumed value of σ_u is the desired value.

Step 6: The Weibull modulus m is then obtained as the slope of the latest straight line in plot B with the last chosen value of σ_u .

Step 7: The Weibull parameter σ_0 may be obtained by referring to the sketch in Figure 12.8, in which we pick up the stress σ_s corresponding to $\ln(KV)$ on the vertical axis, representing $\ln\ln[1/(1-P)]$ with the known values of load factor K and the volume of the specimen V. By virtue of Equation 12.12, we may reach the relationship $m\ln(\sigma_s - \sigma_u) = m\ln(\sigma_0)$, which leads to $(\sigma_s - \sigma_u) = (\sigma_0)$. The value of σ_0 can thus be determined as $\sigma_0 = \sigma_s - \sigma_u$ with the newly determined value of σ_s.

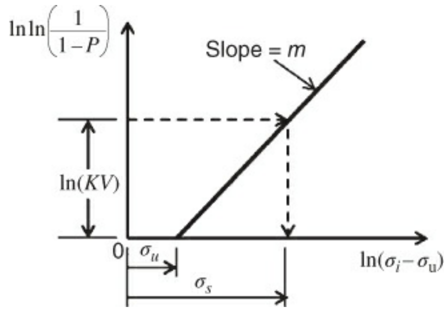

Figure 12.8 Determination of Weibull parameters by the log-log plot method.

Example 12.8

Use the log-log plot method to determine the Weibull parameters in Equation 12.12 for the KT-S_iC specimens whose fracture strength under 4-point bending is shown in Table 12.1. The 30 specimens have an average volume of 0.0541 in^3.

Solution:

The parameters in Equation 12.11 were determined by following the seven steps outlined in Section 12.7.3 with four arbitrarily assumed values of the threshold stresses of σ_u = 11 ksi in case (1), 7.0 ksi in case (2), 5.0 ksi in case (3), and 0 ksi in case (4). The results for all four cases of the assumed value σ_u are plotted in Figure 12.9. We observe that case (4) with the assumed σ_u = 0 results in a straight line in Figure 12.9. The Weibull modulus m was obtained by the computing the slope of this line to be m = 13.33. The parameter σ_o was computed to be 11.82 ksi following the procedure stipulated in Step 7. The Weibull distribution function for the fracture strength of the 30 KT-SiC specimens listed in Table 12.1 thus has the following expression:

$$P = 1 - \exp\left[-0.0013\left(\frac{\sigma}{11.82}\right)^{13.3}\right] \quad \mathbf{12.13}$$

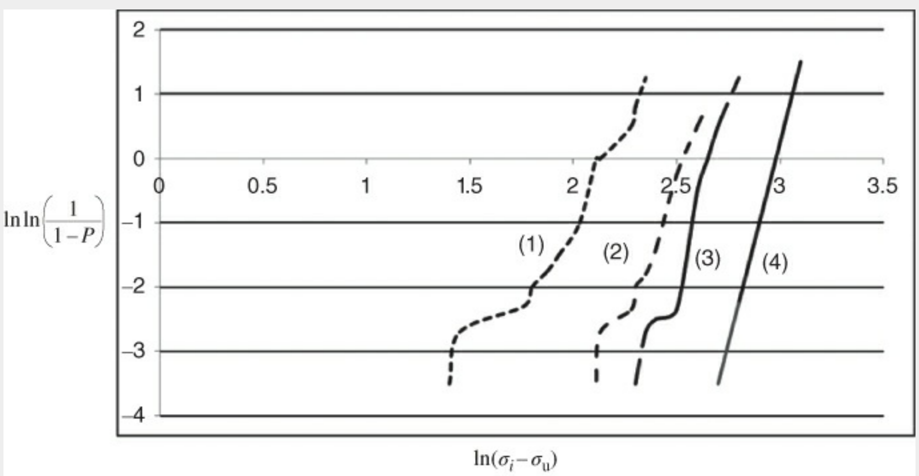

Figure 12.9 Estimation of Weibull parameters for KT-SiC by the log-log method with chosen values of σ_u.

The Weibull distribution function in Equation 12.13 obtained by the log-log plot method is shown in Figure 12.10. Also shown in the figure are author's attempts at evaluating the Weibull parameters by other two methods: the 3-moment method with an assumed threshold fracture strength σ_u = 11.86 ksi, resulted in m = 5.2 and σ_o = 2.54 ksi; and the least mean square approximation method with σ_u = 6.5 ksi resulted in m = 8.15 and σ_o = 6.08 ksi. The procedures of determining the Weibull parameters by the latter two methods are available in Gregory and Sprinill (1962). It is remarkable that all three methods used in determining Weibull parameters fit well to the P–σ curve plot as shown in Figure 12.10.

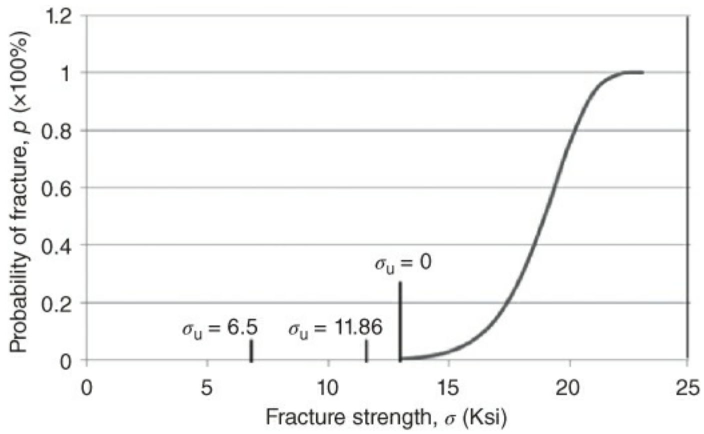

Figure 12.10 Weibull distribution function for the fracture strength of KT-SiC.

12.7.4 Probabilistic Design of Structures with Random Fracture Strength of Materials

In view of the randomness of the fracture strength of many ceramic and brittle materials and the significance of the size of the structure, conventional deterministic approaches involving maximum allowable stress with "safety factors," as described in Section 1.6 in Chapter 1, have little meaning in the design analyses. A radically different concept and corresponding theory need to be adopted for the analyses of structures made of these materials. Probabilistic design analysis using Weibull distribution functions is considered a viable alternative.

This radically different concept is to design the structure with random strength based on the "probability P of fracture of the structure" or "reliability R of the structure," with $R = 1 - P$. The design criterion P or R depends on the nature and purpose of the structure; for instance a value of $P = 10^{-6}$, meaning a one-millionth chance of structural failure, is considered to be desirable for entities that require extremely high reliability and safety such as nuclear reactor cores or crucial components of aerospace equipment. The low probability of fracture P is used because the consequences of failure of these structures under anticipated loads in both normal operation and accident conditions are infinitely serious.

The volume effect is another major factor to be included in this type of analysis. Conventional wisdom indicates that engineering materials have inherent voids and minute flaws in reality. It is logical to envisage that the larger the volume of the material the more voids and flaws will be present inside the volume. This scenario leads to anticipation that larger structures are vulnerable to higher chances of structural failure. This "size effect" is often ignored in traditional design analysis, but it is an important parameter in the Weibull distribution function as expressed in Equation 12.11. The Weibull distribution function with volume V in Equation 12.11 leads to the following relationship for this purpose:

$$\frac{\sigma_1}{\sigma_2} = \left(\frac{V_2}{V_1}\right)^{1/m} \qquad \textbf{12.14}$$

where the subscripts on σ denote the tensile stress in materials with respective given volumes V, and m is the Weibull modulus

The following steps are to be followed in the design analysis for reliability of structures made of materials with random and widely scattered fracture strength distributions:

Step 1: Establish the Weibull distribution of P–σ from the measured fracture strength data with Weibull parameters such as illustrated in Figure 12.10.

Step 2: Perform a detailed stress analysis on the structure either using classic theories or by numerical methods such as the finite-element method presented in Chapter 11.

Step 3: Divide the structure into n volumes; V_i with i = 1, 2, 3, ..., with no volume of the structure smaller than the gage volumes of the test specimens used to generate the Weibull parameters for the Weibull distribution function in Step 1. Also, try to choose the individual volumes in the structure with approximately uniform stress distributions within the volume.

Step 4: Determine the "worst" or the highest stress condition in each volume V_i if the stress distribution in the unit volume is not uniformly. Also assume that the corresponding principal stresses S_1, S_2, and S_3 act uniformly throughout the volume. Principal stresses in a tube or pipe are the stress in the radial, tangential, and longitudinal directions.

Step 5: Determine the reliability of each volume element by first finding the reliability of the gage volume of the specimens under each of the principal stresses S_1, S_2, and S_3 using the Weibull distribution function in Equation 12.13 for KT-SiC with $1 - P_{gv}(S_1)$, $1 - P_{gv}(S_2)$, and $1 - P_{gv}(S_3)$, in which $P_{gv}(S_i)$, with $i = 1, 2, 3$, denotes the probability of failure of "gage volume" of the test specimens under principal stress S_i. From these $P_{gv}(S_i)$ values, we may determine the reliability of the gage element subject to simultaneous application of these principal stresses as

$$\begin{aligned} P_{gv}(S_i) &= [1 - P_{gv}(S_1, S_2, S_3)] \\ &= [1 - P_{gv}(S_1)][1 - P_{gv}(S_2)][1 - P_{gv}(S_3)] \end{aligned}$$

where $P_{gv}(S_1,S_2,S_3)$ = probability of failure of the volume element subject to simultaneous principal stress S_1, S_2, and S_3.

The reliability of the volume element of the structure with different volume than the gage volume of the specimens can thus be determined by the following expression:

$$(1 - P_{iv}) = (1 - P_{gv})\frac{V_i}{V_g} \tag{12.15}$$

where P_{iv} = probability of failure of the volume element of the structure V_i and V_g is the volume of the test specimens.

We have the reliability of the gage volume of the specimens from the testing for the material's fracture strength, represented by the Weibull distribution function

$$R_{gv} = 1 - P_{gv} = \exp\left[-V_g\left(\frac{\sigma - \sigma_u}{\sigma_0}\right)^m\right]$$

We may substitute the relationship in Equation 12.15 into the above expression to obtain the reliability of each volume element of the structure:

$$R_{iv} = 1 - P_{iv} = \exp\left[-V_i\left(\frac{\sigma - \sigma_u}{\sigma_0}\right)^m\right] \tag{12.16}$$

Step 6: The reliability of the overall structure using the weakest link theory can thus be obtained from the following expression:

$$R = \prod_{i=1}^{n} R_{iv} \tag{12.17}$$

where n is the total number of volume elements in the structure and the R_{iv} are the reliability of each volume element in the structure computed from Equation 12.16.

Example 12.9

Determine the reliability of a ceramic tube made of KT-SiC with fracture strength measured from 30 specimens as listed in Table 12.1. The dimension of the tube is 6 inches OD × 4 inches ID × 4 inches long. The tube is subjected to a uniform internal pressure P_{i} = 4000 psi. The Weibull parameters for the measured fracture strength are determined by a least squares approximation method, and they fit the Weibull distribution function

$$P_{\mathrm{gv}} = 1 - \exp\left[-0.3378\left(\frac{\sigma - 6.5}{6.083}\right)^{8.15}\right] \tag{a}$$

Assume that the tube is at room temperature and the end effects of the pressurized tube are neglected.

Solution:

The physical situation of the pressurized tube is illustrated in Figure 12.11. The tube is free to deform in the radial direction because the end constraints are assumed to be negligible. Consequently, only two stress components need to be included in the analysis. These are the hoop (or tangential) stress designated by $\sigma_{\theta\theta}$ and the radial stress σ_{rr} in the tube wall. From advanced strength of materials textbooks (Volterra and Gaines, 1971) or mechanical engineering handbooks (Avallone, 2006), one may compute these stresses with the applied internal pressure P_{i} for thick-walled tubes and pipes using the following equations:

$$\sigma_{rr}(r) = \frac{P_{\mathrm{i}} r_{\mathrm{i}}^2}{r_0^2 - r_{\mathrm{i}}^2}\left(1 - \frac{r_0^2}{r^2}\right) \tag{b}$$

$$\sigma_{\theta\theta}(r) = \frac{P_{\mathrm{i}} r_{\mathrm{i}}^2}{r_0^2 - r_{\mathrm{i}}^2}\left(1 + \frac{r_0^2}{r^2}\right) \tag{c}$$

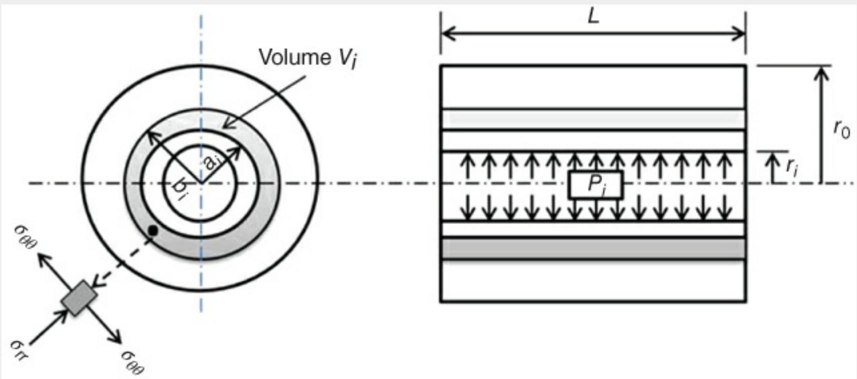

Figure 12.11 A ceramic tube subject to internal pressure loading.

From Equations (a) and (b) one may conclude that the radial stress in Equation (b) is compressive, which is not a concern in the analysis. We will thus only be concerned with the tangential stress $\sigma_{\theta\theta}$, which is tensile at all times. This stress component may be computed using Equation (c).

Let us arbitrarily divide the tube wall into four concentric elements with dimensions shown in Table 12.2, with a_i and b_i being the inside and outside diameters of element i with volume V_i as shown in Figure 12.11.

Table 12.2 Reliability of a pressurized tube made of KT-SiC ceramic.

At $P_i = 4000$ psi						
i	a_i (in)	b_i (in)	V_i (in^3)	$\sigma_{\theta\theta}$ (ksi)	$-V_i\left(\dfrac{\sigma_{\theta\theta}-6.5}{6.083}\right)^{8.15}$	$1-P_{iv}=\exp\left[-V_i\left(\dfrac{\sigma_{\theta\theta}-6.5}{6.083}\right)^{8.15}\right]$
1	2	2.25	13.3518	10.4	-0.35649	0.700 13
2	2.25	2.5	14.9226	8.889	-0.00734	0.992 68
3	2.5	2.75	16.4934	7.808	-5.99×10^{-5}	0.999 94
4	2.75	3	18.0642	7.008	-2.89×10^{-8}	1
At $P_i = 3500$ psi						
				9.1	-0.013085	0.987
				7.7778	-0.000045	0.999 95
				6.832	0	1
				6.132	$\sigma_{\theta\theta}<\sigma_u$	1

The reliability of the four volume elements under P_i = 4000 psi are computed, with the results tabulated in the right-hand column of Table 12.2. The reliability of the overall structure, that is, the KT-SiC tube, can be computed using Equation 12.17 to give

$$R=\prod_{i=1}^{n}R_{iv}=0.70013\times0.99268\times0.99994\times1.0=0.69496$$

or R = 69.5% for the KT-SiC tube subjected to an internal pressure P_i = 4000 psi.

The computation has also been performed for the same tube subjected to a reduced internal pressure loading at 3500 psi, at which the reliability of the tube rises to 98.7%, as indicated in Table 12.2. It makes sense to expect the tubular structure to have higher reliability (thus lower probability of fracture) at a lower applied pressure load, as computed in this analysis.

12.8 Statistical Quality Control

Conventional wisdom indicates that a product cannot succeed in the marketplace unless it offers consumers competitive price and good quality. The quality of a product commonly relates to its finished appearance and its reliability in performance. It also relates to a number of other factors that are related to good quality in industrial products. These factors may include the fitting of the components in an engineering system; sustainability of the product in the expected environmental and/or operating conditions; delivery of the intended performance; and persistence to the life expectancy of the designed objectives. These quality requirements are critical concerns by both the producers and consumers. Poor or unreliable quality of a product not only jeopardizes its success in the marketplace but, more seriously, it also causes liability that may cost the producer millions of dollars in recall of the products, or lawsuits and other severe penalties resulting from litigation. Recalls of massive numbers of products, such as have happened in automobile industry, are mainly due to the poor quality of the products.

Poor quality of a product in relation to engineering practice may be attributed to many causes as outlined below:

1. Poor design in setting the dimensions and tolerances, in surface finishing, in improper selection of materials, and so on.
2. Manufacturing and fabrication processes relating to improper machining, assembly, testing, and inspection.
3. Improper condition of machine tools and fabrication process control.
4. Poor workmanship in all of the above production processes.

The cost of poor product quality can be significant to producers, as can be seen from the following typical example. In addition to incurring substantial costs in recall of defective products already in the marketplace, firms may expend a significant portion of their earned revenues in the costs related to such quality-related areas as inspection of the product at various stages in production, scrappage and rework, prevention, and warranty.

Quality of industrial products is closely related to the rigor and quality of engineering design, to the fabrication equipment, and to the knowledge, experience, and work ethic of operators and inspectors. The exorbitant costs associated with quality of products could be mitigated by better design, better manufacturing processes, better workmanship, and so on. All improvements in quality of industrial products through more sophisticated design process—e.g., setting proper tolerances for assembled engineering systems—better fabrication equipment, and employment of more experienced and skillful operators, as well as implementing more thorough inspections on finished products will incur costs for the improvement. Naturally one will expect the value of the product to increase with the improvement in quality, with associated better customer receptivity for and satisfaction with the products. It will also result in significant reduction in the costs related to product warranty, scrappage, and rework. It thus comes to an issue of reaching an optimal "balance" between the *improved quality* from and the *required cost* for such improvements. A qualitative relationship between these two factors is illustrated in Figure 12.12 (Stout, 1985).

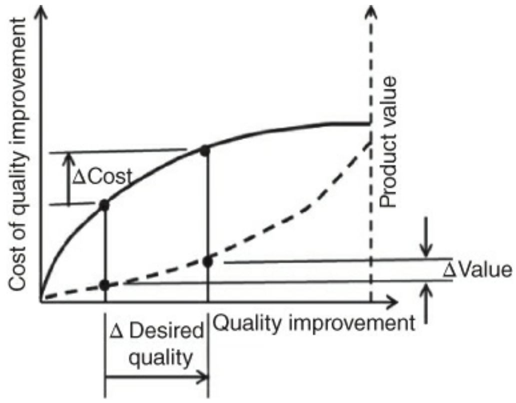

Figure 12.12 Cost and product value associated with quality improvement.

In Figure 12.12, the solid line curve represents the cost associated with improvement of quality of a product and the broken line curve represents the increase of product value. For a specific improvement of product quality, the associated cost and product value can be identified as shown in the figure. It is apparent that the planned quality improvement of the product is worthwhile only if the resulting increase of product value (ΔValue) is significantly larger than the associated increase of cost (ΔCost) in Figure 12.12. The particular situation illustrated in Figure 12.12 is not a desirable one according to the analysis of cost vs. benefits outlined above.

12.9 Statistical Process Control

12.9.1 Quality Issues in Industrial Automation and Mass Production

Many would consider that industrial automation began with the introduction of assembly lines in the Ford Motor Company in early 1900. Massive industrial automation did not materialize, however, until shortly after World War II in response to the strong demand for the production of life-sustaining products such as foods and construction materials in large quantities after the war. Industrial automation became a way of producing consumer electronics in the 1980s, followed by information technology equipment a decade later. This mode of production has become commonplace in many other industries, for example for production of electronics, automobiles, and information technology and communication equipment. Industrial automation has now been extended to the healthcare industry, in particular in drug manufacturing and production. A major issue arising from mass production of products by industrial automation is that of quality assurance and quality control during production.

The need for control of quality of products produced by mass production embraces many fields that involve cultural aspects and technical management; these include design, manufacture, functional testing, and inspection through an understanding of sampling procedures and sampling theory by both employees and the management that has overall responsibility for the maintenance of quality. These aspects of quality control and assurance are major challenges to engineers and managers involved in such production.

Most would agree that frequent and thorough inspection of finished products is the key factor contributing to good quality assurance of items or components produced in a mass production environment. Vigorous inspections of the quality of these items or components during the production stage will certainly result in better quality of the finished products. In theory, the ideal way to control the quality of finished products is to thoroughly inspect *every* piece that is produced by a machine or by an automated production process in order to ensure not only that the dimensions and tolerances set for all pieces are consistent with the design specifications but also that all pieces will perform the function(s) that are designed for. This idea is obviously not practicable in mass production situations: the amount of time and the costs associated with thorough inspection of every piece of finished product would be prohibitive to any producer. Nor is it practicable to inspect the quality of every finished or quasi-finished item in batches of millions of items that have been produced. A more realistic scheme is to perform inspections on a limited number of randomly selected samples and employ statistical techniques to be able to ensure the quality of all products produced by the same manufacturing process. Key questions involved in such practice are “How many samples” from a production batch does one needs to inspect, and “How many measurements” does one needs to perform on each of the selected sample items in order to be sure that “enough is enough?”

The following illustration will present the procedure that a quality engineer might follow in the quality assurance and quality control of a mass-produced shaft used as the guide rail for printer head cartridges. A machine or a process is used to produce 10 000 circular shafts each day. These shafts in the form of long circular rods must have uniform and precise diameter along the length to allow smooth sliding of the cartridge head under the guidance of high-precision motion controls. The quality assurance engineer who is responsible for such production must determine the number of sample shafts that he or she will randomly pick from the batches of the finished products to assure that the finished sample rods comply with the set standards. Additionally, the engineer also needs to determine the number of measurements of the rod diameters on each of the selected sample shaft that are required in order to be sure that the diameters measured will have the least discrepancies in value at the measurement locations.

For the purpose of illustration, let us assume that the engineer has randomly picked 100 sample rods from a batch of finished rods and chosen six diameter measurement stations along the length of the selected shafts as indicated in Figure 12.13.

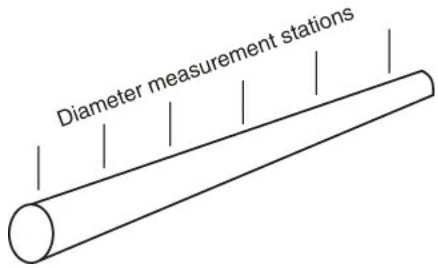

Figure 12.13 A rod sample with six measurement stations.

The readers is now faced with the critical question of "How confident is the quality assurance engineer that selecting only six measurement stations on each of the 100 samples taken would lead to a 'credible and robust' assurance of the quality for 'all' of the 10 000 rods produced by the same machine or process?" Conversely, the question may be put "How valid are the results of the inspections on a *limited number of sample products* from the batch with *limited number of inspections on each of the limited number of sample products* for all the products produced by the same machine or process?"

The *statistical process control* (SPC) technique may provide answers to this critical question, as will be presented in the following section.

12.9.2 The Statistical Process Control Method

The statistical process control (SPC) method widely used by industry in current times was pioneered by Walter A. Shewhart of Bell Laboratories in the early 1920s. He developed the control charts for statistical quality control in 1924. This concept was further modified and improved for practical applications in modern-day statistical quality control of industrial products by W. Edwards Deming (1900–1993), an American statistician, who is regarded as the pioneer of this powerful tool for quality control of mass-produced products. The wide adoption of SPC by Japanese industry after WWII is viewed as the principal reason for Japan's enormous success in consumer electronics and automobiles in the global marketplace. One of Deming's followers, Dr. Genichi Taguchi, is considered the architect of applied SPC for industrial production. SPC is widely used in the semiconductor industry in this country.

The major advantages of using statistical process control (SPC) are outlined below:

1. It offers assurance that those items or components remaining in production after the application of SPC will have satisfactory quality.

2. If the procedures are applied correctly, the defect rate attributed to manufacturing will rarely exceed 1%. As a result, more parts with consistent quality will be produced, and scrappage, reworking, and repair will be reduced to a minimum.

3. When effective machines are used, manufacture will be trouble-free. This control method will also identify ineffective machines being used in production.

4. When machines of marginal capability are used in production, then the number of defective components produced by these machines will be kept to the unavoidable minimum.

5. The procedures will improve shopfloor personnel participation in quality control.

6. The procedures will have the effect of reducing company scrap rates during manufacture and assembly, which will reflect advantageously in total quality control cost.

In short, SPC emphasizes early detection of defective products and the fabrication processes that produce defective products. It also increases the rate of production of quality-assured finished products.

One will appreciate that, it being such an essential tool in modern industrial production, full coverage of the subject of SPC would require considerably more space and time than is available in this one chapter. What we will do here is introduce one commonly used technique known as "control charts." These charts allow the quality assurance engineer to control the functions and capabilities of the machines or processes that are used in mass production of parts and goods.

12.10 The “Control Charts”

Control charts are established by quality assurance engineers using statistical methods to provide the *upper* and *lower bounds* within which the measured parameters from randomly selected samples are deemed to represent satisfactory quality. These control charts are derived on the assumption that all the measured parameters from the sampled products fit the bell-shaped distribution that was illustrated in Figure 12.4, so that the mathematical model for such a distribution—Equation 12.7 for the normal distribution function—can be used as a theoretical basis for such derivation.

Figure 12.14 shows the form of a typical control chart consisting of a lower control limit (LCL) and an upper control limit (UCL). These limits serve as the lower and upper bounds that determine acceptance of the parts subjected to the same manufacturing and inspection processes after the establishment of the control chart. Once the chart is established, quality assurance engineers will randomly pick samples from batches of parts manufactured by the same process for further inspection by measurement of the same parameters. The manufacturing process is considered sound and healthy if the values of the measured parameters in these additional samples are within the upper and lower control limits of the control chart. On the other hand, the quality assurance engineer will order an immediate halt of production should he or she finds the value of the measured parameters on any of these additional samples to be outside these control limits. The engineer will immediately launch an investigation to identify the sources of the problem, whether they arise from malfunction of the manufacturing process or whether there is human error on the part of the operators. Production resumes after these problems are resolved.

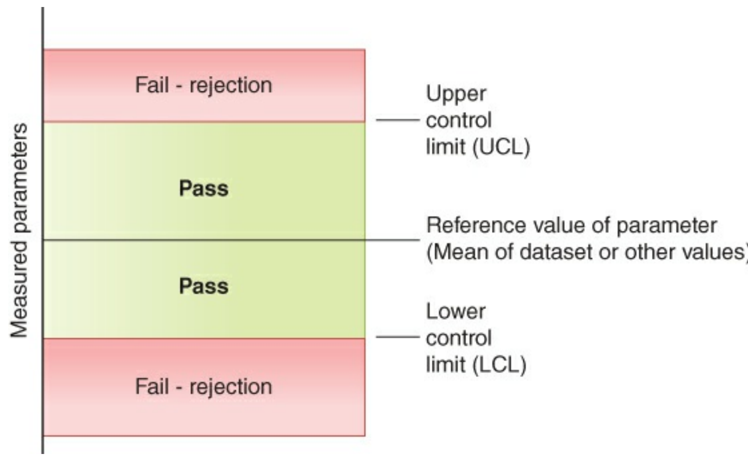

Figure 12.14 Typical control chart for quality control in mass production.

We will include in later subsections two particular types of control charts that are popular for SPC in mass production environment. We will require collection of the following data for establishing the control charts of both types:

1. The sample size, k (the number of randomly selected samples from a specific manufacturing process)
2. The number of measured parameters of each sample, n.

Thus the total number of measurements for all randomly selected samples in the SPC analysis will be $k \times n$.

12.10.1 Three-Sigma Control Charts

The term “three-sigma” derives from the description of the physical meaning of standard deviation that was given in Section 12.6 and implies that 99.73% of the collected data (the measurements of the specific parameter) will be included in the analysis. For k samples and n measurements on each sample taken from a production batch, the mean of the measured value from the total $k \times n$ measured parameters is denoted by $\bar{\bar{x}}$.

If the mean of all $k \times n$ measurements is $\bar{\bar{x}}$, with standard deviation σ, obtained from Equation 12.3, we will have the bounds of the "upper control limit" and "lower control limit" from the following expressions:

Lower control limit:

$$LCL_{\bar{x}} = \bar{\bar{x}} - \frac{3\sigma}{\sqrt{n}} \tag{12.18a}$$

Upper control limit:

$$UCL_{\bar{x}} = \bar{\bar{x}} + \frac{3\sigma}{\sqrt{n}} \tag{12.18b}$$

According to the definition of standard deviation, σ, as a measure of variation of data from the mean value, "3σ" that appears in Equations 12.18a and 12.18b implies that there is a 99.73% probability of the measured data fitting within the value of $\bar{\bar{x}} \pm 3\sigma$. Consequently, we may say that there is only a 1 − 0.9973 = 0.0027, or 0.27%, probability that further mean measured values of the parameter would fall outside the two control limits.

The above situation may be graphically illustrated in Figure 12.15. We may readily see that any average measured control parameter on *each* additional sample ($\bar{x}$) that falls in the shaded zone would mean acceptance of the product as being of good quality, whereas any further sample whose value of $\bar{x}$ falls outside the shaded zone will be rejected. In such case, the quality assurance engineer needs to investigate the causes for the defective product and impose whatever remedial action is necessary.

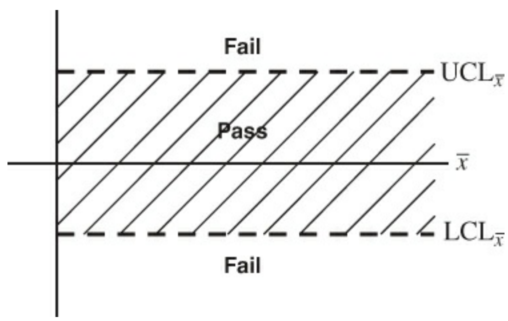

Figure 12.15 Three-sigma control chart.

A necessary word of caution is that the above SPC control chart is built on a fundamental assumption that all the measured data fit the "normal distribution." Engineers thus need to be sure that large datasets with large numbers of samples are included in the establishment of the control charts. The limited numbers of data used in the subsequent examples may not offer correct answers for this reason. These limited numbers of data are used in these examples solely for illustrative and didactic convenience.

Example 12.10

A firm in Silicon Valley produces integrated circuit (IC) chips for a customer. The quality assurance engineer measured the output voltage from three outlets of each of the five sample chips that he randomly selected from a production station. The measured voltages are tabulated below:

Sample 1:	2.25	3.16	1.80
Sample 2:	2.60	1.95	3.22
Sample 3:	1.75	3.06	2.45
Sample 4:	2.15	2.80	1.85
Sample 5:	3.15	2.76	2.20

All measured data have the unit of millivolts.
Establish the upper and lower control limits using the "three-sigma" control chart method and the tabulated data.

Solution:

We have $k = 5$ (five samples) and $n = 3$ (three measurements on each sample). The total number of measurements is $k \times n = 5 \times 3 = 15$. One can determine the mean of all 15 measurements to be $\bar{\bar{x}}$ = 2.477 millivolts, and from Equation 12.3 using the total number of data of $n = 15$, the standard deviation to be $\sigma = 0.5281$ millivolt.

Thus, using Equations 12.18a and 12.18b, we have the lower and upper control limits:

$$\mathrm{LCL}_{\bar{x}} = \bar{\bar{x}} - \frac{3\sigma}{\sqrt{n}} = 2.477 - \frac{3 \times 0.5281}{\sqrt{3}} = 1.5623 \quad \text{millivolts}$$

$$\mathrm{UCL}_{\bar{x}} = \bar{\bar{x}} + \frac{3\sigma}{\sqrt{n}} = 2.477 + \frac{3 \times 0.5281}{\sqrt{3}} = 3.3917 \quad \text{millivolts}$$

Graphically, the above results provide the quality assurance engineer with the control chart illustrated in Figure 12.16.

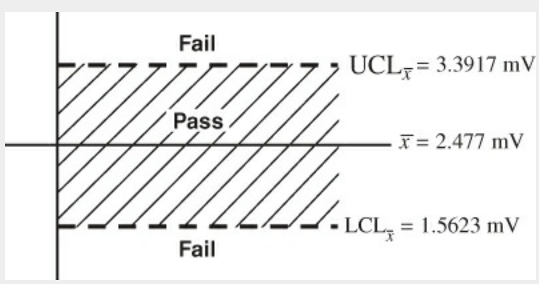

Figure 12.16 A three-sigma control chart for IC chip output.

The quality control engineer will pick further samples during the continuing production run and perform similar measurements of output voltages on the same leads as he did with the sample chips. Should the average the measured output (i.e., $\bar{x}$ = average of the three measurements) from any additional sample chip be beyond these bounds in Figure 12.16, he will not only scrap this particular chip but will also order an immediate halt of the production and investigate the causes for the defective chip. Production will be resumed after all the problems are resolved.

12.10.2 Control Charts for Sample Ranges (the R-Chart)

We learned in Section 12.10.1 about establishing three-sigma control charts with the mean values ($\bar{x}$) of the parameter from each of the selected measurement stations on the selected samples. There is another type of control chart, called the R-chart, that is established for the *range of measured parameter* on each of the selected samples.

Again, if we let k = the number of selected samples and n = the number of measurements stations on each sample, the recorded measurements may be expressed in the format presented in Table 12.3.

Table 12.3 Working sheet for establishing the R-chart

Sample	Measured parameters							Sample mean,	Sample range,
								$\bar{x}$	R
k = 1	x_1	x_2	x_3	x_4	x_5	…	x_n	$\bar{x}_{k=1}$	$x_{max} - x_{min}$
k = 2	•	•	•	•	•	…	•	$\bar{x}_{k=2}$	•
•	•	•	•	•	•	…	•	•	•
•	•	•	•	•	•	…	•	•	•
•	•	•	•	•	•	…	•	•	•
k = k	•	•	•	•	•	…	•	•	•
							MEAN:	$\bar{x}$	$\bar{R}$

One sees from the bottom line of Table 12.3 that the "mean" of the Range, $\bar{R}$ may be obtained by the arithmetic average of all the sample ranges in the last column of the table. However, the same mean value of the sample range $\bar{R}$ can be obtained theoretically from the expression

$$\bar{R} = d_2 \sigma \tag{12.19}$$

in which σ is the standard deviation of *all measurements*—that is, from ($k \times n$) measurements. The factor, d_2 may be determined on the basis of the value of n = the number of measurements of the parameter on each sample from Table 12.4 (Rosenkrantz, 1997).

Table 12.4 Factors for estimating. $\bar{R}$. and lower and upper control limits (Rosenkrantz, 1997)

No. of measurements on each sample, n	Factor d_2	Coefficient D_1	Coefficient D_2
2	1.128	0	3.69
3	1.693	0	4.36
4	2.059	0	4.70
5	2.326	0	4.92
6	2.534	0	5.08
7	2.704	0.20	5.20
8	2.847	0.39	5.31
9	2.970	0.55	5.39
10	3.075	0.69	5.47
11	3.173	0.81	5.53
12	3.258	0.92	5.59
13	3.336	1.03	5.65
14	3.407	1.12	5.69
15	3.472	1.21	5.74

This value of mean sample range $\bar{R}$ obtained by the measured average of all samples should approach the $\bar{R}$ value computed from Equation 12.19 if the sample size is large enough for the collected data to fit well with the normal distribution as illustrated in Figure 12.4 and represented by Equation 12.7, with which the

$\overline{R}$ -value can be calculated using Equation 12.19.

The discrepancy in the $\overline{R}$ values determined by these two methods is not significant because the number of measured data in a real-life quality assurance process would be large enough to have all the measured data closely fitting the normal distribution curve illustrated in Figure 12.4 and justify the use of Equation 12.19 for determining the $\overline{R}$ values in establishing the R-control charts.

The lower and upper control limits based on the range of the measured data may be computed from the following equations:

$$\textit{Lower control limit:} \quad \mathrm{LCL}_R = D_1 \sigma \tag{12.20a}$$

$$\textit{Upper control limit:} \quad \mathrm{UCL}_R = D_2 \sigma \tag{12.20b}$$

where the coefficients, D_1 and D_2 can be found from Table 12.4.

A similar control chart to that of 3σ-method in Figure 12.16 may be established for the control chart based on the range of the measured data as in Figure 12.17.

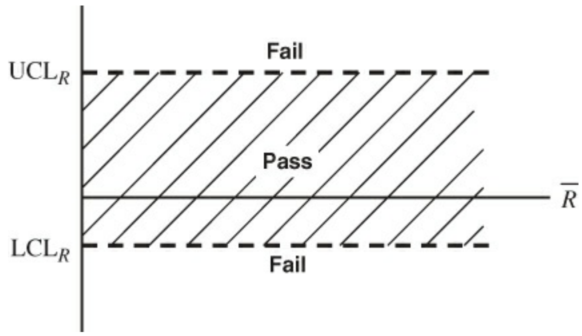

Figure 12.17 Control chart using sample ranges (the R-chart).

Example 12.11

Establish an R-chart using the data presented in Example 12.10.

Solution:

We will first establish the working chart similar to that shown in Table 12.3 for the present case as shown below:

Sample	Measured voltage (mV)			Mean value, $\bar{x}$	Sample Range
k = 1	2.25	3.16	1.80	2.4033	1.36
2	2.60	1.95	3.22	2.5900	1.27
3	1.75	3.06	2.45	2.4200	1.31
4	2.15	2.80	1.85	2.2667	0.95
5	3.15	2.76	2.20	2.7033	0.95
Total k = 5	Total n = 3			Mean, μ = 2.477	$\bar{R}$ = 1.168 (from dataset)

We may use Equation 12.3 to calculate the standard deviation from the dataset with the $k \times n$ = 15 measured data and calculate σ = 0.5281.

We may use Equation 12.19 to compute the value of $\bar{R} = d_2\sigma$ = 1.693 × 0.5281 = 0.8941, using d_2 = 1.693 obtained from Table 12.4 with n = 3. This value obviously is very different from what we calculated from the average sample range of $\bar{R} = 1.168$ as indicated in the above table. This discrepancy is to be expected because the total number of 15 measurements in this example is hardly enough for these measured data to fit a normal distribution.

However, given that the control limits given in Equations 12.20a and 12.20b are derived from the normal distribution, we will adopt the theoretical value $\bar{R}$ = 0.8941 in determining these two control limits for the present case, notwithstanding the unrealistic number of sample data.

Thus, using Equations 12.20a and 12.20b, we may compute the control limits for the R-chart to be

$$\mathrm{LCL}_R = D_1\sigma = 0 \times 0.5281 = 0$$
$$\mathrm{UCL}_R = D_2\sigma = 4.36 \times 0.5281 = 2.3025$$

with D_1 = 0 and D_2 = 4.36 obtained from Table 12.4 with n = 3.

Graphically the R-chart in this case may be shown as in Figure 12.18.

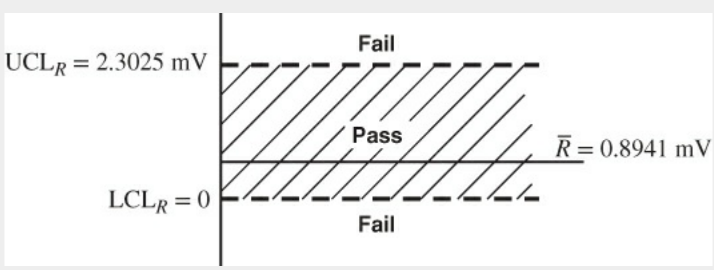

Figure 12.18 The R-control chart for quality control of chips.

The control chart shown in Figure 12.18 will be used for quality control for inspecting chips in further production. Any range of the measured output voltages on any chip that is within the bounds of the control limits will be deemed to be of acceptable quality. Those chips that fail to satisfy this condition will be rejected, and the quality assurance engineer should investigate the causes for the unacceptable quality of the product and seek remedial actions.

Problems

12.1 A farm-implement manufacturing company in Midwest U.S.A. purchases steel castings from a local foundry. Thirty castings were selected at random and weighed, and their masses were recorded to the nearest kilogram, as shown below:

235	232	228	228	240	231
225	220	218	230	222	229
217	233	222	221	228	228
238	232	230	226	236	226
227	227	229	229	224	227

Do the following:

a. Group the measurements into a frequency distribution table having six equal classes from 215 to 244.

b. Construct a histogram of the distribution.

c. Determine the median, mode, and mean of the dataset

12.2 Approximations of missile velocities were recorded over a predetermined fixed distance and are presented in the following dataset. Each value in the dataset is rounded to the nearest 10 m/s:

980	960	950	1010
930	880	870	960
850	1020	970	940
970	900	1030	950
1000	940	970	600

Do the following:

a. Group these measurements into a frequency distribution table having six equal classes that range from 500 to 1100.

b. Construct a histogram of the distribution.

c. Determine the median, mode, and mean of the data

12.3 Using the data given in Problem 12.2, calculate the following:

a. The standard deviation of all data.

b. The variance of the dataset.

12.4 Find the mean $\bar{x}$, the median x_d and the standard deviation $\boldsymbol{\sigma}$ from the following set of measurements on the diameter of a shaft in centimeters:

2.5, 2.45, 2.48, 2.75, 2.32, 2.55, 2.47, 2.25

How do you interpret the standard deviation (σ) in a physical sense? If this dataset fits a normal distribution, what implications are for the 1σ, 2σ and 3σ that you obtained in the computation?

12.5 Find the mean, $\bar{x}$, and the standard deviation, σ, of the fracture strength of KT-SiC tabulated in Table 12.1. Interpret the physical meaning of the standard deviation that you have computed.

12.6 Find the reliability of a tube made of KT-SiC with the wall thickness indicated in Example 12.9. The tube is subjected to internal pressures of 3000 and 2500 psi.

12.7 The thickness of an epitaxial layer on a silicon wafer is specified to be 14.5 $\pm$ 0.5 μm. Assume that the manufacturing process is in statistical control, where the distribution of the thickness is normal with the mean $\bar{x}$ = 14.5 μm and standard deviation σ = 0.4 μm. To monitor the production process, the mean thickness of four wafers is determined at regular time intervals. Compute the lower and upper control limits for a three-sigma control chart for the process mean.

12.8 A Silicon Valley company produces computer chips in large quantities. A particular microfabrication process for the production involves thin dioxide films on silicon substrates. The measured mean thickness of deposited thin film from 50 samples with 8 measurements on each sample was 15.4 micrometers (μm). The total sum of the range of measured thickness on each sample is 1.25 μm. If one assumes the measured data fit normal distribution, do the following:

a. Determine the standard deviation of the measured data, σ, and its physical meaning in relation to the measured data.

b. Determine the variance of the measured thickness.

c. Construct the 3σ control chart for the measured thickness.

d. Construct the control chart for the sample range of the measured thickness.

e. Indicate the use of these control charts for quality control in the production.

12.9 The diameters (measured in mm) of ball bearings are monitored by $\bar{x}$ and R-charts. After 30 samples of size $n = 6$ were taken, the following values were recorded: the sum of the mean of all samples $\sum_i \bar{x} = 150$ mm and the sum of the range of the measured values of all samples $\sum_i \bar{R}_i = 12$ mm . Assuming the process is in statistical control and the measured values fit the normal distribution, calculate the $\bar{x}$ and R control limits.

12.10 A precision machine shop is involved in mass production of shafts for a special electromechanical system. A quality assurance engineer is assigned the responsibility for establishing quality control charts. She took five samples from the batch of the shafts produced by one lathe. The diameters that she measured from three selected locations along the length of each of these samples are listed below:

Sample no.	Diameter at position 1	Diameter at position 2	Diameter at position 3
1	2.25	3.16	1.85
2	2.60	2.28	3.10
3	1.85	2.35	2.20
4	1.95	2.15	2.75
5	2.15	2.85	3.10

a. Determine the mode of all the measured data.

b. Determine the median and mean of all measured data.

c. Determine the standard deviation and indicate the physical meaning of this standard deviation.

d. Construct the control chart based on three-sigma method.

e. Construct the control chart based on sample ranges.

f. How would the quality assurance engineer use these charts in controlling the quality of shafts produced by that particular lathe?

12.11 A batch of microchips is fabricated by a local firm in Silicon Valley. The quality assurance engineer took five samples from a bin and measured the length of these chips at three fixed locations on each sample. The measured data that the engineer recorded are:

Sample 1:	2.15	2.35	1.95
Sample 2:	2.70	1.83	2.25
Sample 3:	1.97	2.03	2.13
Sample 4:	2.06	2.70	2.15
Sample 5:	2.03	1.75	1.85

All measured data have the unit of mm.

Determine the upper and lower control limits using:

a. The three-sigma method

b. The control chart for the ranges.

12.12 A quality assurance engineer randomly took 20 samples out of a batch of 2000 high-precision shafts manufactured by a special process. She measured the diameter of the shaft at six different locations on each sample shaft. She recorded the following: $\sum_i \bar{x}_i = 52 \text{ mm}$ and $\sum_i \bar{R}_i = 5 \text{ mm}$, in which $\bar{x}_i$ and $\bar{R}_i$ are the sample mean and sample range of the measured values from each sample, respectively. On the assumption that all measured data fit the normal distribution, establish the following control charts for quality assurance purposes: (a) the 3σ control chart, and (b) the control chart for sample range (the R-chart) for the measured data