20 Statistical Considerations

Chapter Outline

20–1 Random Variables **978**

20–2 Arithmetic Mean, Variance, and Standard Deviation **980**

20–3 Probability Distributions **985**

20–4 Propagation of Error **992**

20–5 Linear Regression **994**

Statistics in mechanical design provides a method of dealing with characteristics whose values are variable. Products manufactured in large quantities—automobiles, watches, lawnmowers, washing machines, for example—have a life that is variable. One automobile may have so many defects that it must be repaired repeatedly during the first few months of operation while another may operate satisfactorily for years, requiring only minor maintenance.

Methods of quality control are deeply rooted in the use of statistics, and engineering designers need a knowledge of statistics to conform to quality-control standards. The variability inherent in limits and fits, in stress and strength, in bearing clearances, and in a multitude of other characteristics must be described numerically for proper control. It is not satisfactory to say that a product is expected to have a long and troublefree life. We must express such things as product life and product reliability in numerical form in order to achieve a specific quality goal. As noted in Sec. 1–10, uncertainties abound and require quantitative treatment. The algebra of real numbers, by itself, is not well-suited to describing the presence of variation.

It is clear that consistencies in nature are stable, not in magnitude, but in the *pattern of variation*. Evidence gathered from nature by measurement is a mixture of systematic and random effects. It is the role of statistics to separate these, and, through the sensitive use of data, illuminate the obscure.

Some students will start this book after completing a formal course in statistics while others may have had brief encounters with statistics in their engineering courses. This contrast in background, together with space and time constraints, makes it very difficult to present an extensive integration of statistics with mechanical engineering design at this stage. Beyond first courses in mechanical design and engineering statistics, the student can begin to meaningfully integrate the two in a second course in design.

The intent of this chapter is to introduce some statistical concepts associated with basic reliability goals.

20–1 Random Variables

Consider an experiment to measure strength in a collection of 20 tensile-test specimens that have been machined from a like number of samples selected at random from a carload shipment of, say, UNS G10200 cold-drawn steel. It is reasonable to expect that there will be differences in the ultimate tensile strengths S_{ut} of each of the individual test specimens. Such differences may occur because of differences in the sizes of the specimens, in the strength of the material itself, or both. Such an experiment is called a *random experiment,* because the specimens are selected at random. The strength S_{ut} determined by this experiment is called a *random,* or a *stochastic, variable.* So a random variable is a variable quantity, such as strength, size, or weight, whose value depends on the outcome of a random experiment.

Let us define a random variable x as the sum of the numbers obtained when two dice are tossed. Either die can display any number from 1 to 6. Figure 20–1 displays all possible outcomes in what is called the *sample space.* Note that x has a specific value

Figure 20–1

Sample space showing all possible outcomes of the toss of two dice.

1,1	1,2	1,3	1,4	1,5	1,6
2,1	2,2	2,3	2,4	2,5	2,6
3,1	3,2	3,3	3,4	3,5	3,6
4,1	4,2	4,3	4,4	4,5	4,6
5,1	5,2	5,3	5,4	5,5	5,6
6,1	6,2	6,3	6,4	6,5	6,6

Table 20–1

A Probability Distribution

x	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

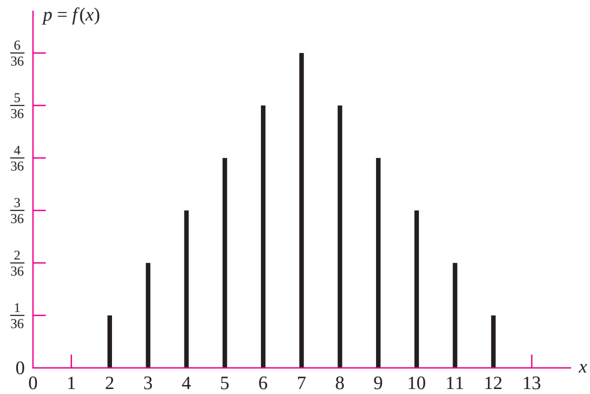


Figure 20–2

Frequency distribution.

Table 20–2

A Cumulative Probability Distribution

x	2	3	4	5	6	7	8	9	10	11	12
$F(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$

for each possible outcome—for example, the event 5, 4; $x = 5 + 4 = 9$. It is useful to form a table showing the values of x and the corresponding values of the probability of x, called $p = f(x)$. This is easily done from Fig. 20–1 merely by adding each outcome, determining how many times a specific value of x arises, and dividing by the total number of possible outcomes. The results are shown in Table 20–1. Any table like this, listing all possible values of a random variable and with the corresponding probabilities, is called a *probability distribution*.

The values of Table 20–1 are plotted in graphical form in Fig. 20–2. Here it is clear that the probability is a function of x. This *probability function* $p = f(x)$ is often called the *frequency function* or, sometimes, the *probability density function* (PDF). The probability that x is less than or equal to a certain value x_i can be obtained from the probability function by summing the probability of all x's up to and including x_i. If we do this with Table 20–1, letting x_i equal 2, then 3, and so on, up to 12, we get Table 20–2, which is called a *cumulative probability distribution*. The *function* $F(x)$ in Table 20–2 is called a *cumulative density function* (CDF) of x. In terms of $f(x)$ it may be expressed mathematically in the general form

$$F(x_i) = \sum_{x_j \le x_i} f(x_j) \tag{20–1}$$

The cumulative distribution may also be plotted as a graph (Fig. 20–3).

The variable x of this example is called a *discrete random variable*, because x has only discrete values. A *continuous random variable* is one that can take on any value in a specified interval; for such variables, graphs like Figs. 20–2 and 20–3 would be plotted as continuous curves. For a continuous probability density function $F(x)$, the probability of obtaining an observation equal to or less than x is given by

$$F(x) = \int_{-\infty}^{x} f(x)\,dx \tag{20–2}$$

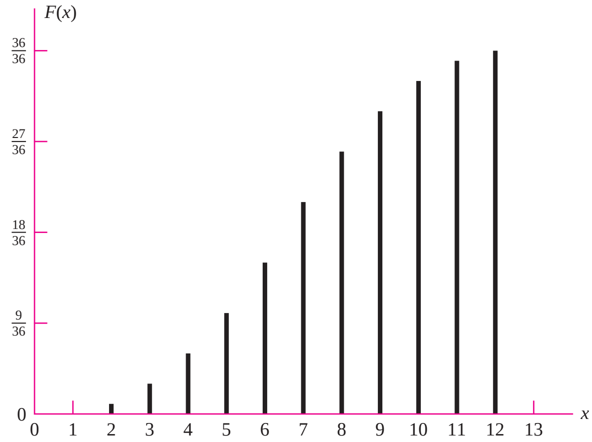


Figure 20–3

Cumulative frequency distribution.

where $f(x)$ is the probability per unit x. When $x \to \infty$, then

$$\int_{-\infty}^{\infty} f(x)\, dx = 1 \tag{20–3}$$

Differentiation of Eq. (20–2) gives

$$\frac{dF(x)}{dx} = f(x) \tag{20–4}$$

20–2 Arithmetic Mean, Variance, and Standard Deviation

In studying the variations in the mechanical properties and characteristics of mechanical elements, we shall generally be dealing with a finite number of elements. The total number of elements, called the *population*, may in some cases be quite large. In such cases it is usually impractical to measure the characteristics of each member of the population, because this involves destructive testing in some cases, and so we select a small part of the group, called a *sample*, for these determinations. Thus the *population* is the entire group, and the *sample* is a part of the population.

The arithmetic mean of a sample, called the *sample mean*, consisting of N elements, is defined by the equation

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{20–5}$$

Besides the arithmetic mean, it is useful to have another kind of measure that will tell us something about the spread, or dispersion, of the distribution. For any random variable x, the deviation of the ith observation from the mean is $x_i - \bar{x}$. But since the sum of the deviations so defined is always zero, we square them, and define *sample variance* as

$$s_x^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_N - \bar{x})^2}{N - 1} = \frac{1}{N - 1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{20–6}$$

The *sample standard deviation,* defined as the square root of the sample variance, is

$$s_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \qquad (20\text{–}7)$$

Equation (20–7) is not well-suited for use in a calculator. For such purposes, use the alternative form

$$s_x = \sqrt{\frac{\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2 \Big/ N}{N-1}} = \sqrt{\frac{\sum_{i=1}^{N} x_i^2 - N\bar{x}^2}{N-1}} \qquad (20\text{–}8)$$

for the standard deviation.

It should be observed that some authors define the variance and the standard deviation by using N instead of $N - 1$ in the denominator. For large values of N, there is very little difference. For small values, the denominator $N - 1$ actually gives a better estimate of the variance of the population from which the sample is taken.

Equations (20–5) to (20–8) apply specifically to the *sample* of a population. When an entire population is considered, the same equations apply, but $\bar{x}$ and s_x are replaced with the symbols μ_x and, $\hat{\sigma}_x$ respectively. The circumflex accent mark ˆ, or "hat," is used to avoid confusion with normal stress. For the population variance and standard deviation, N weighting is used in the denominators instead of $N - 1$.

Sometimes we are going to be dealing with the standard deviation of the strength of an element. So you must be careful not to be confused by the notation. Note that we are using the *capital letter S* for *strength* and the *lowercase letter s* for *standard deviation* as shown in the caption of the histogram in Fig. 20–4.

Figure 20–4 is called a *discrete frequency histogram,* which gives the number of occurrences, or class frequency f_i, within a given range. If the data are grouped in this fashion, then the mean and standard deviation are given by

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{k} f_i x_i \qquad (20\text{–}9)$$

and

$$s_x = \sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2 - \left[\left(\sum_{i=1}^{k} f_i x_i\right)^2 \Big/ N\right]}{N-1}} = \sqrt{\frac{\sum_{i=1}^{k} f_i x_i^2 - N\bar{x}^2}{N-1}} \qquad (20\text{–}10)$$

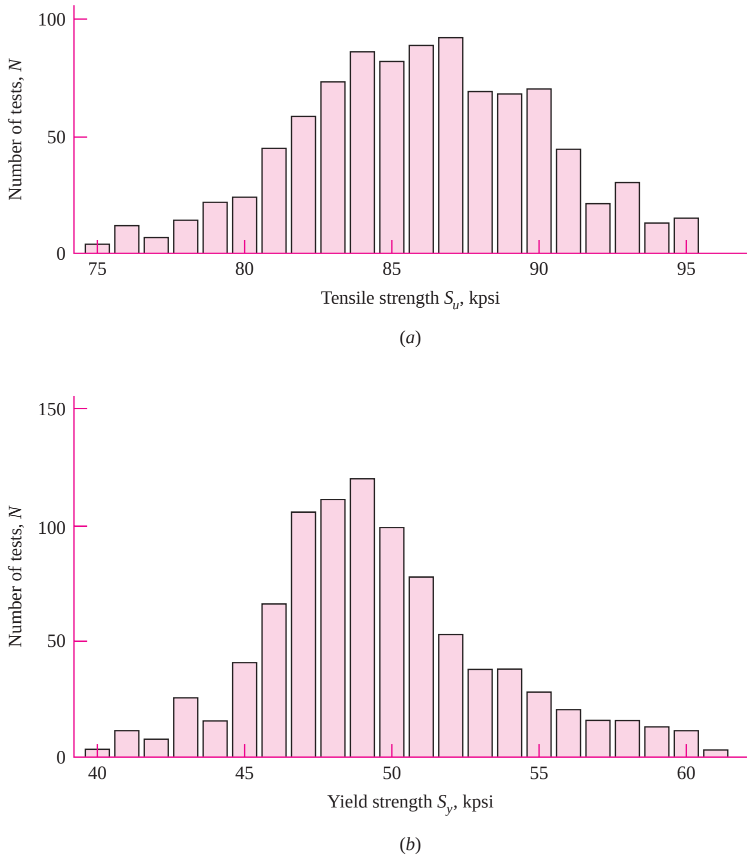
Here x_i, f_i, and k are class midpoint, frequency of occurrences within the range of the class, and the total number of classes, respectively. Also, the cumulative density function that gives the probability of an occurrence at class mark of x_i or less is

$$F_i = \frac{f_i w_i}{2} + \sum_{j=1}^{i-1} f_j w_j \qquad (20\text{–}11)$$

where w_i represents the class width at x_i. For Fig. 20–4*a*, $k = 21$ and the class width is constant at $w = 1$ kpsi.

Figure 20–4

Distribution of tensile properties of hot-rolled UNS G10350 steel, as rolled. These tests were made from round bars varying in diameter from 1 to 9 in. (*a*) Tensile-strength distributions from 930 heats; $\bar{S}_u = 86.0$ kpsi, $s_{Su} = 4.94$ kpsi. (*b*) Yield-strength distribution from 899 heats; $\bar{S}_y = 49.5$ kpsi, $s_{Sy} = 5.36$ kpsi. *(From* Metals Handbook, vol. 1, 8th ed., *American Society for Metals, Materials Park, OH 44073-0002, fig. 22, p. 64. Reprinted by permission of ASM International®, www.asminternational.org.)*



Notation

In this book, we follow the convention of designating vectors by boldface characters, indicative of the fact that two or three quantities, such as direction and magnitude, are necessary to describe them. The same convention is widely used for random variables that can be characterized by specifying a mean and a standard deviation. We shall therefore use boldface characters to designate random variables as well as vectors. No confusion between the two is likely to arise.

The terms *stochastic variable* and *variate* are also used to mean a random variable. A *deterministic quantity* is something that has a single specific value. The mean value of a population is a deterministic quantity, and so is its standard deviation. A stochastic variable can be partially described by the mean and the standard deviation, or by the mean and the *coefficient of variation* defined by

$$C_x = \frac{s_x}{\bar{x}} \tag{20–12}$$

Thus the variate **x** for the sample can be expressed in the following two ways:

$$\mathbf{x} = \mathbf{X}(\bar{x}, s_x) = \bar{x}\,\mathbf{X}(1, C_x) \tag{20–13}$$

where **X** represents a variate probability distribution function. Note that the deterministic quantities $\bar{x}$, s_x, and C_x are all in normal italic font.

EXAMPLE 20–1

Five tons of 2-in round rod of 1030 hot-rolled steel has been received for workpiece stock. Nine standard-geometry tensile test specimens have been machined from random locations in various rods. In the test report, the ultimate tensile strength was given in kpsi. In ascending order (not necessary), these are displayed in Table 20–3. Find the mean $\bar{x}$, the standard deviation s_x, and the coefficient of variation C_x from the sample, such that these are best estimates of the parent population (the stock your plant will convert to product).

Table 20–3

Data Worksheet from Nine Tensile Test Specimens Taken from a Shipment of 1030 Hot-Rolled Steel Barstock

	S_{ut}, kpsi x	x^2
	62.8	3 943.84
	64.4	4 147.36
	65.8	4 329.64
	66.3	4 395.69
	68.1	4 637.61
	69.1	4 774.81
	69.8	4 872.04
	71.5	5 112.25
	74.0	5 476.00
$\sum$	611.8	41 689.24

Solution

From Eqs. (20–5) and (20–8),

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{9} x_i$$

and

$$s_x = \sqrt{\frac{\sum x_i^2 - \left(\sum x_i\right)^2/N}{N-1}}$$

It is computationally efficient to generate $\sum x$ and $\sum x^2$ before evaluating $\bar{x}$ and s_x. This has been done in Table 20–3.

Answer

$$\bar{x} = \frac{1}{9}(611.8) = 67.98 \text{ kpsi}$$

Answer

$$s_x = \sqrt{\frac{41\,689.24 - 611.8^2/9}{9-1}} = 3.543 \text{ kpsi}$$

From Eq. (20–12),

Answer

$$C_x = \frac{s_x}{\bar{x}} = \frac{3.543}{67.98} = 0.0521$$

All three statistics are estimates of the parent population statistical parameters. Note that these results are independent of the distribution.

Multiple data entries may be identical or may be grouped in histographic form to suggest a distributional shape. If the original data are lost to the designer, the grouped data can still be reduced, although with some loss in computational precision.

EXAMPLE 20–2

The data in Ex. 20–1 have come to the designer in the histographic form of the first two columns of Table 20–4. Using the data in this form, find the mean $\bar{x}$, standard deviation s_x, and the coefficient of variation C_x.

Table 20–4

Grouped Data of Ultimate Tensile Strength from Nine Tensile Test Specimens from a Shipment of 1030 Hot-Rolled Steel Barstock

Class Midpoint x, kpsi	Class Frequency f	Extension fx	fx^2
63.5	2	127	8 064.50
66.5	2	133	8 844.50
69.5	3	208.5	14 480.75
72.5	2	145	10 513.50
	$\sum$ 9	613.5	41 912.25

The data in Table 20–4 have been extended to provide $\sum f_i x_i$ and $\sum f_i x_i^2$.

Solution From Eq. (20–9),

Answer

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{4} f_i x_i = \frac{1}{9}(613.5) = 68.17 \text{ kpsi}$$

From Eq. (20–10),

Answer

$$s_x = \sqrt{\frac{41\,912.25 - 613.5^2/9}{9-1}} = 3.391 \text{ kpsi}$$

From Eq. (20–12),

Answer

$$C_x = \frac{s_x}{\bar{x}} = \frac{3.391}{68.17} = 0.0497$$

Note the small changes in $\bar{x}$, s_x, and C_x due to small changes in the summation terms.

The descriptive statistics developed, whether from ungrouped or grouped data, describe the ultimate tensile strength $\mathbf{S}_{ut}$ of the material from which we will form parts. Such description is not possible with a single number. In fact, sometimes two or three numbers plus identification or, at least, a robust approximation of the distribution are needed. As you look at the data in Ex. 20–1, consider the answers to these questions:

- Can we characterize the ultimate tensile strength by the mean, $\bar{S}_{ut}$?
- Can we take the lowest ultimate tensile strength of 62.8 kpsi as a minimum? If we do, we will encounter some lesser ultimate strengths, because some of 100 specimens will be lower.
- Can we find the distribution of the ultimate tensile strength of the 1030 stock in Ex. 20–1? Yes, but it will take more specimens and require plotting on coordinates that rectify the data string.

20–3 Probability Distributions

There are a number of standard discrete and continuous probability distributions that are commonly applicable to engineering problems. In this section, we will discuss four important continuous probability distributions; the *Gaussian,* or *normal, distribution;* the *lognormal distribution;* the *uniform distribution;* and the *Weibull distribution.*

The Gaussian (Normal) Distribution

When Gauss asked the question, What distribution is the most likely parent to a set of data?, the answer was the distribution that bears his name. The *Gaussian,* or *normal, distribution* is an important one whose probability density function is expressed in terms of its mean μ_x and its standard deviation $\hat{\sigma}_x$ as

$$f(x) = \frac{1}{\hat{\sigma}_x\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_x}{\hat{\sigma}_x}\right)^2\right] \qquad (20\text{–}14)$$

With the notation described in Sec. 20–2, the normally distributed variate x can be expressed as

$$\mathbf{x} = \mathbf{N}(\mu_x, \hat{\sigma}_x) = \mu_x\mathbf{N}(1, C_x) \qquad (20\text{–}15)$$

where **N** represents the normal distribution function given by Eq. (20–14).

Since Eq. (20–14) is a probability density function, the area under it, as required, is unity. Plots of Eq. (20–14) are shown in Fig. 20–5 for small and large standard deviations. The bell-shaped curve is taller and narrower for small values of $\hat{\sigma}$ and shorter and broader for large values of $\hat{\sigma}$. Integration of Eq. (20–14) to find the cumulative density function $F(x)$ is not possible in closed form, but must be accomplished numerically. To avoid the need for many tables for different values of μ and $\hat{\sigma}$, the deviation from the mean is expressed in units of standard deviation by the transform

$$\mathbf{z} = \frac{\mathbf{x}-\mu_x}{\hat{\sigma}_x} \qquad (20\text{–}16)$$

The integral of the transform is tabulated in Table A–10 and sketched in Fig. 20–6. The value of the normal cumulative density function is used so often, and manipulated in so

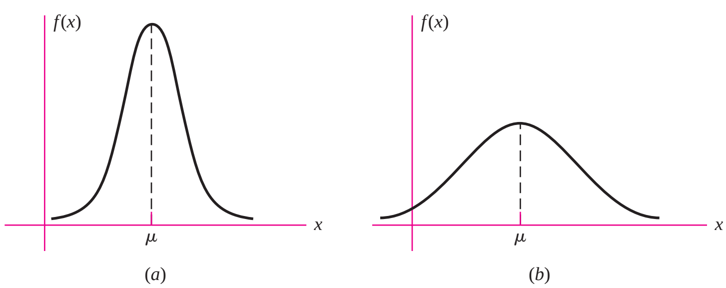


Figure 20–5

The shape of the normal distribution curve: (*a*) small $\hat{\sigma}$; (*b*) large $\hat{\sigma}$.

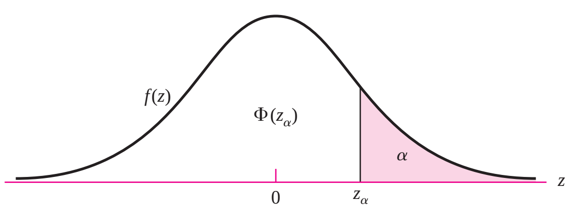


Figure 20–6

The standard normal distribution.

many equations, that it has its own particular symbol, $\Phi(z)$. The transformation variate $\mathbf{z}$ is normally distributed, with a mean of zero and a standard deviation and variance of unity. That is, $\mathbf{z} = \mathbf{N}(0, 1)$. The probability of an observation less than z is $\Phi(z)$ for negative values of z and $1 - \Phi(z)$ for positive values of z in Table A–10.

EXAMPLE 20–3

In a shipment of 250 connecting rods, the mean tensile strength is found to be 45 kpsi and the standard deviation 5 kpsi.

(*a*) Assuming a normal distribution, how many rods can be expected to have a strength less than 39.5 kpsi?

(*b*) How many are expected to have a strength between 39.5 and 59.5 kpsi?

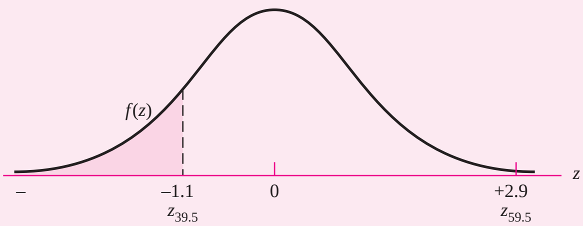
Solution

(*a*) Substituting in Eq. (20–16) gives the standardized z variable as

$$z_{39.5} = \frac{x - \mu_x}{\hat{\sigma}_x} = \frac{S - \bar{S}}{\hat{\sigma}_S} = \frac{39.5 - 45}{5} = -1.10$$

The probability that the strength is less than 39.5 kpsi can be designated as $F(z) = \Phi(-1.10)$. Using Table A–10, and referring to Fig. 20–7, we find $\Phi(z_{39.5}) = 0.1357$. So the number of rods having a strength less than 39.5 kpsi is,

Figure 20–7



Answer

$$N\Phi(z_{39.5}) = 250(0.1357) = 33.9 \approx 34$$

because $\Phi(z_{39.5})$ represents the proportion of the population N having a strength less than 39.5 kpsi.

(*b*) Corresponding to $S = 59.5$ kpsi, we have

$$z_{59.5} = \frac{59.5 - 45}{5} = 2.90$$

Referring again to Fig. 20–7, we see that the probability that the strength is less than 59.5 kpsi is $F(z) = \Phi(z_{59.5})$. Since the z variable is positive, we need to find the value complementary to unity. Thus, from Table A–10,

$$\Phi(2.90) = 1 - \Phi(-2.90) = 1 - 0.001\,87 = 0.998\,13$$

The probability that the strength lies between 39.5 and 59.5 kpsi is the area between the ordinates at $z_{39.5}$ and $z_{59.5}$ in Fig. 20–7. This probability is found to be

$$p = \Phi(z_{59.5}) - \Phi(z_{39.5}) = \Phi(2.90) - \Phi(-1.10)$$
$$= 0.998\,13 - 0.1357 = 0.862\,43$$

Therefore the number of rods expected to have strengths between 39.5 and 59.5 kpsi is

Answer

$$Np = 250(0.862) = 215.5 \approx 216$$

The Lognormal Distribution

Sometimes random variables have the following two characteristics:

- The distribution is asymmetrical about the mean.
- The variables have only positive values.

Such characteristics rule out the use of the normal distribution. There are several other distributions that are potentially useful in such situations, one of them being the lognormal (written as a single word) distribution. Especially when life is involved, such as fatigue life under stress or the wear life of rolling bearings, the lognormal distribution may be a very appropriate one to use.

The *lognormal distribution* is one in which the logarithms of the variate have a normal distribution. Thus the variate itself is said to be lognormally distributed. Let this variate be expressed as

$$\mathbf{x} = \mathbf{LN}(\mu_x, \hat{\sigma}_x) \tag{a}$$

Equation (*a*) states that the random variable x is distributed lognormally (*not a logarithm*) and that its mean value is μ_x and its standard deviation is $\hat{\sigma}_x$.

Now use the transformation

$$\mathbf{y} = \ln \boldsymbol{x} \tag{b}$$

Since, by definition, **y** has a normal distribution, we can write

$$\mathbf{y} = \mathbf{N}(\mu_y,\ \hat{\sigma}_y) \tag{c}$$

This equation states that the random variable **y** is normally distributed, its mean value is μ_y, and its standard deviation is $\hat{\sigma}_y$.

It is convenient to think of Eq. (*a*) as designating the *parent,* or *principal, distribution* while Eq. (*b*) represents the *companion,* or *subsidiary,* distribution.

The probability density function (PDF) for **x** can be derived from that for **y**; see Eq. (20–14), and substitute y for x in that equation. Thus the PDF for the companion distribution is found to be

$$f(x) = \begin{cases} \dfrac{1}{x\hat{\sigma}_y\sqrt{2\pi}} \exp\left[-\dfrac{1}{2}\left(\dfrac{\ln x - \mu_y}{\hat{\sigma}_y}\right)^2\right] & \text{for } x > 0 \\ 0 & \text{for } x \le 0 \end{cases} \tag{20–17}$$

The companion mean μ_y and standard deviation $\hat{\sigma}_y$ in Eq. (20–17) are obtained from

$$\mu_y = \ln \mu_x - \ln\sqrt{1 + C_x^2} \approx \ln \mu_x - \frac{1}{2}C_x^2 \tag{20–18}$$

$$\hat{\sigma}_y = \sqrt{\ln\left(1 + C_x^2\right)} \approx C_x \tag{20–19}$$

These equations make it possible to use Table A–10 for statistical computations and eliminate the need for a special table for the lognormal distribution.

EXAMPLE 20–4 One thousand specimens of 1020 steel were tested to rupture and the ultimate tensile strengths were reported as grouped data in Table 20–5. From Eq. (20–9),

$$\bar{x} = \frac{63\,625}{1000} = 63.625 \text{ kpsi}$$

Table 20–5

Worksheet for Ex. 20–4

Class Midpoint, kpsi	Frequency f_i	Extension $x_i f_i$	Extension $x_i^2 f_i$	Observed PDF $f_i/(Nw)$*	Normal Density $f(x)$	Lognormal Density $g(x)$
56.5	2	113.0	6 384.5	0.002	0.0035	0.0026
57.5	18	1 035.0	59 512.5	0.018	0.0095	0.0082
58.5	23	1 345.5	78 711.75	0.023	0.0218	0.0209
59.5	31	1 844.5	109 747.75	0.031	0.0434	0.0440
60.5	83	5 021.5	303 800.75	0.083	0.0744	0.0773
61.5	109	6 703.5	412 265.25	0.109	0.110	0.1143
62.5	138	8 625.0	539 062.5	0.138	0.140	0.1434
63.5	151	9 588.5	608 869.75	0.151	0.1536	0.1539
64.5	139	8 965.5	578 274.75	0.139	0.1453	0.1424
65.5	130	8 515.0	577 732.5	0.130	0.1184	0.1142
66.5	82	5 453.0	362 624.5	0.082	0.0832	0.0800
67.5	49	3 307.5	223 256.25	0.049	0.0504	0.0493
68.5	28	1 918.0	131 382.0	0.028	0.0263	0.0268
69.5	11	764.5	53 132.75	0.011	0.0118	0.0129
70.5	4	282.0	19 881.0	0.004	0.0046	0.0056
71.5	2	143.0	10 224.5	0.002	0.0015	0.0022
	Σ 1 000	63 625	4 054 864	1.000		

*To compare discrete frequency data with continuous probability density functions f_i must be divided by Nw. Here, N = sample size = 1000; w = width of class interval = 1 kpsi.

From Eq. (20–10),

$$s_x = \sqrt{\frac{4\,054\,864 - 63\,625^2/1000}{1000 - 1}} = 2.594\,245 = 2.594 \text{ kpsi}$$

$$C_x = \frac{s_x}{\bar{x}} = \frac{2.594\,245}{63.625} = 0.040\,773 = 0.0408$$

From Eq. (20–14) the probability density function for a normal distribution with a mean of 63.625 and a standard deviation of 2.594 245 is

$$f(x) = \frac{1}{2.594\,245\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x - 63.625}{2.594\,245}\right)^2\right]$$

For example, $f(63.625) = 0.1538$. The probability density $f(x)$ is evaluated at class midpoints to form the column of normal density in Table 20–5.

EXAMPLE 20–5 Continue Ex. 20–4, but fit a lognormal density function.

Solution From Eqs. (20–18) and (20–19),

$$\mu_y = \ln \mu_x - \ln \sqrt{1 + C_x^2} = \ln 63.625 - \tfrac{1}{2}\ln(1 + 0.040\,773^2) = 4.1522$$

$$\hat{\sigma}_y = \sqrt{\ln\left(1 + C_x^2\right)} = \sqrt{\ln(1 + 0.040\,773^2)} = 0.0408$$

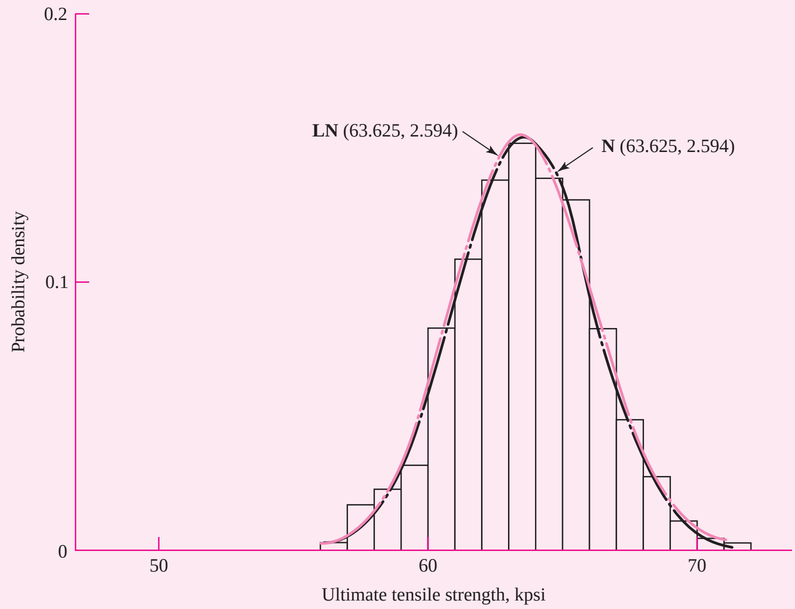
The probability density of a lognormal distribution is given in Eq. (20–17) as

$$g(x) = \frac{1}{x\,(0.0408)\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\ln x - 4.1522}{0.0408}\right)^2\right] \qquad \text{for } x > 0$$

For example, $g(63.625) = 0.1537$. This lognormal density has been added to Table 20–5. Plot the lognormal PDF superposed on the histogram of Ex. 20–4 along with the normal density. As seen in Fig. 20–8, both normal and lognormal densities fit well.

Figure 20–8

Histogram for Ex. 20–4 and Ex. 20–5 with normal and lognormal probability density functions superposed.



The Uniform Distribution

The uniform distribution is a closed-interval distribution that arises when the chance of an observation is the same as the chance for any other observation. If a is the lower bound and b is the upper bound, then the probability density function (PDF) for the uniform distribution is

$$f(x) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & a > x > b \end{cases} \tag{20–20}$$

The cumulative density function (CDF), the integral of $f(x)$, is thus linear in the range $a \leq x \leq b$ given by

$$F(x) = \begin{cases} 0 & x < a \\ (x-a)/(b-a) & a \leq x \leq b \\ 1 & x > b \end{cases} \tag{20–21}$$

The mean and standard deviation are given by

$$\mu_x = \frac{a+b}{2} \tag{20–22}$$

$$\hat{\sigma}_x = \frac{b-a}{2\sqrt{3}} \tag{20–23}$$

The uniform distribution arises, among other places in manufacturing, where a part is mass-produced in an automatic operation and the dimension gradually changes through tool wear and increased tool forces between setups. If n is the part sequence or processing number, and n_f is the sequence number of the final-produced part before another setup, then the dimension x graphs linearly when plotted against the sequence number n. If the last proof part made during the setup has a dimension x_i, and the final part produced has the dimension x_f, the magnitude of the dimension at sequence number n is given by

$$x = x_i + (x_f - x_i)\frac{n}{n_f} = x_i + (x_f - x_i)F(x) \tag{a}$$

since n/n_f is a good approximation to the CDF. Solving Eq. (a) for $F(x)$ gives

$$F(x) = \frac{x - x_i}{x_f - x_i} \tag{b}$$

Compare this equation with the middle form of Eq. (20–21).

The Weibull Distribution

The Weibull distribution does not arise from classical statistics and is usually not included in elementary statistics textbooks. It is far more likely to be discussed and used in works dealing with experimental results, particularly reliability. It is a chameleon distribution, asymmetrical, with different values for the mean and the median. It contains within it a good approximation of the normal distribution as well as an exact representation of the exponential distribution. Most reliability information comes from laboratory and field service data, and because of its flexibility, the Weibull distribution is widely used.

The expression for reliability is the value of the cumulative density function complementary to unity. For the Weibull this value is both explicit and simple. The reliability given by the *three-parameter Weibull distribution* is

$$R(x) = \exp\left[-\left(\frac{x - x_0}{\theta - x_0}\right)^b\right] \qquad x \geq x_0 \geq 0 \tag{20–24}$$

where the three parameters are

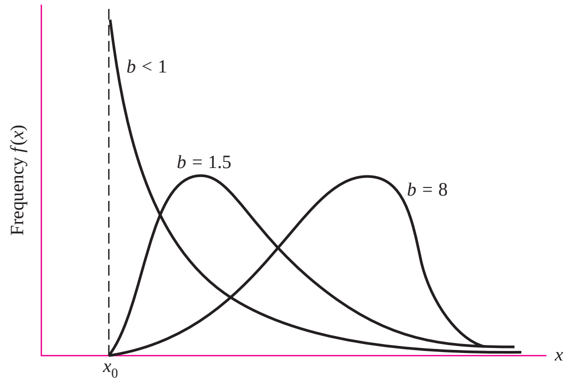
x_0 = minimum, guaranteed, value of x

θ = a characteristic or scale value $(\theta \geq x_0)$

b = a shape parameter $(b > 0)$

Figure 20–9

The density function of the Weibull distribution showing the effect of skewness of the shape parameter b.



For the special case when $x_0 = 0$, Eq. (20–24) becomes the two-parameter Weibull

$$R(x) = \exp\left[-\left(\frac{x}{\theta}\right)^b\right] \qquad x \geq 0 \tag{20–25}$$

The characteristic variate θ serves a role similar to the mean and represents a value of x below which lie 63.2 percent of the observations.

The shape parameter b controls the skewness of the distribution. Figure 20–9 shows that large b's skew the distribution to the right and small b's skew it to the left. In the range $3.3 < b < 3.5$, approximate symmetry is obtained along with a good approximation to the normal distribution. When $b = 1$, the distribution is exponential.

Given a specific required reliability, solving Eq. (20–24) for x yields

$$x = x_0 + (\theta - x_0)\left(\ln\frac{1}{R}\right)^{1/b} \tag{20–26}$$

To find the probability function, we note that

$$F(x) = 1 - R(x) \tag{a}$$

$$f(x) = \frac{dF(x)}{dx} = -\frac{dR(x)}{dx} \tag{b}$$

Thus, for the Weibull,

$$f(x) = \begin{cases} \dfrac{b}{\theta - x_0}\left(\dfrac{x - x_0}{\theta - x_0}\right)^{b-1} \exp\left[-\left(\dfrac{x - x_0}{\theta - x_0}\right)^b\right] & x \geq x_0 \geq 0 \\ 0 & x \leq x_0 \end{cases} \tag{20–27}$$

The mean and standard deviation are given by

$$\mu_x = x_0 + (\theta - x_0)\,\Gamma(1 + 1/b) \tag{20–28}$$

$$\hat{\sigma}_x = (\theta - x_0)\sqrt{\Gamma(1 + 2/b) - \Gamma^2(1 + 1/b)} \tag{20–29}$$

where Γ is the gamma function and may be found tabulated in Table A–34. The notation for a Weibull distribution is[1]

$$\mathbf{x} = \mathbf{W}(x_0, \theta, b) \tag{20–30}$$

[1]To estimate the Weibull parameters from data, see J. E. Shigley and C. R. Mischke, *Mechanical Engineering Design*, 5th ed., 1989, McGraw-Hill, New York, Sec. 4–12. The Weibull parameters are determined for the data given in Ex. 2–4.

EXAMPLE 20–6

The Weibull is used extensively for expressing the reliability of rolling-contact bearings (see Sec. 11–4). Here, the variate x is put in dimensionless form as $x = L/L_{10}$ where L is bearing life, in say, number of cycles; and L_{10} is the manufacturer's rated life of the bearing where 10 percent of the bearings have failed (90 percent reliability).

Construct the distributional properties of a 02–30 mm deep-groove ball bearing if the Weibull parameters are $x_0 = 0.0200$, $\theta = 4.459$, and $b = 1.483$. Find the mean, median, L_{90}, and standard deviation.

Solution

From Eq. (20–28) the mean dimensionless life is

Answer

$$\mu_x = x_0 + (\theta - x_0)\Gamma(1 + 1/b)$$
$$= 0.0200 + (4.459 - 0.0200)\Gamma(1 + 1/1.483) = 4.033$$

This says that the average bearing life is 4.033 L_{10}. The median dimensionless life corresponds to $R = 0.5$, or L_{50}, and from Eq. (20–26) is

Answer

$$x_{0.5} = x_0 + (\theta - x_0)\left(\ln \frac{1}{0.5}\right)^{1/b} = 0.0200 + (4.459 - 0.0200)\left(\ln \frac{1}{0.5}\right)^{1/1.483}$$
$$= 3.487$$

For L_{90}, $R = 0.1$, the dimensionless life x is

Answer

$$x_{0.90} = 0.0200 + (4.459 - 0.0200)\left(\ln \frac{1}{0.1}\right)^{1/1.483} = 7.810$$

The standard deviation of the dimensionless life is given by Eq. (20–29):

Answer

$$\hat{\sigma}_x = (\theta - x_0)\sqrt{\Gamma(1 + 2/b) - \Gamma^2(1 + 1/b)}$$
$$= (4.459 - 0.0200)\sqrt{\Gamma(1 + 2/1.483) - \Gamma^2(1 + 1/1.483)} = 2.753$$

20–4 Propagation of Error

In the equation for axial stress

$$\sigma = \frac{F}{A} \tag{a}$$

suppose both the force F and the area A are random variables. Then Eq. (a) is written as

$$\boldsymbol{\sigma} = \frac{\mathbf{F}}{\mathbf{A}} \tag{b}$$

and we see that the stress $\boldsymbol{\sigma}$ is also a random variable. When Eq. (b) is solved, the errors inherent in $\mathbf{F}$ and in $\mathbf{A}$ are said to be *propagated* to the stress variate $\boldsymbol{\sigma}$. It is not hard to think of many other relations where this will occur.

Suppose we wish to add the two variates $\mathbf{x}$ and $\mathbf{y}$ to form a third variate $\mathbf{z}$. This is written as

$$\mathbf{z} = \mathbf{x} + \mathbf{y} \tag{c}$$

The mean is given as

$$\mu_z = \mu_x + \mu_y \tag{d}$$

Table 20–6

Means and Standard Deviations for Simple Algebraic Operations on Independent (Uncorrelated) Random Variables

Function	Mean (μ)	Standard Deviation ($\hat{\sigma}$)
a	a	0
$\mathbf{x}$	μ_x	$\hat{\sigma}_x$
$\mathbf{x}+a$	$\mu_x + a$	$\hat{\sigma}_x$
$a\mathbf{x}$	$a\mu_x$	$a\hat{\sigma}_x$
$\mathbf{x}+\mathbf{y}$	$\mu_x + \mu_y$	$\left(\hat{\sigma}_x^2 + \hat{\sigma}_y^2\right)^{1/2}$
$\mathbf{x}-\mathbf{y}$	$\mu_x - \mu_y$	$\left(\hat{\sigma}_x^2 + \hat{\sigma}_y^2\right)^{1/2}$
$\mathbf{xy}$	$\mu_x\mu_y$	$\mu_x\mu_y\left(C_x^2 + C_y^2 + C_x^2C_y^2\right)^{1/2}$
$\mathbf{x}/\mathbf{y}$	μ_x/μ_y	$\mu_x/\mu_y\left[\left(C_x^2 + C_y^2\right)/\left(1 + C_y^2\right)\right]^{1/2}$
$\mathbf{x}^n$	$\mu_x^n\left[1 + \frac{n(n-1)}{2}C_x^2\right]$	$\lvert n\rvert\mu_x^n C_x\left[1 + \frac{(n-1)^2}{4}C_x^2\right]$
$1/\mathbf{x}$	$\frac{1}{\mu_x}\left(1 + C_x^2\right)$	$\frac{C_x}{\mu_x}\left(1 + C_x^2\right)$
$1/\mathbf{x}^2$	$\frac{1}{\mu_x^2}\left(1 + 3C_x^2\right)$	$\frac{2C_x}{\mu_x^2}\left(1 + \frac{9}{4}C_x^2\right)$
$1/\mathbf{x}^3$	$\frac{1}{\mu_x^3}\left(1 + 6C_x^2\right)$	$\frac{3C_x}{\mu_x^3}\left(1 + 4C_x^2\right)$
$1/\mathbf{x}^4$	$\frac{1}{\mu_x^4}\left(1 + 10\,C_x^2\right)$	$\frac{4C_x}{\mu_x^4}\left(1 + \frac{25}{4}C_x^2\right)$
$\sqrt{\mathbf{x}}$	$\sqrt{\mu_x}\left(1 - \frac{1}{8}C_x^2\right)$	$\frac{\sqrt{\mu_x}}{2}C_x\left(1 + \frac{1}{16}C_x^2\right)$
$\mathbf{x}^2$	$\mu_x^2\left(1 + C_x^2\right)$	$2\mu_x^2 C_x\left(1 + \frac{1}{4}C_x^2\right)$
$\mathbf{x}^3$	$\mu_x^3\left(1 + 3C_x^2\right)$	$3\mu_x^3 C_x\left(1 + C_x^2\right)$
$\mathbf{x}^4$	$\mu_x^4\left(1 + 6C_x^2\right)$	$4\mu_x^4 C_x\left(1 + \frac{9}{4}C_x^2\right)$

Note: The coefficient of variation of variate x is $C_x = \hat{\sigma}_x/\mu_x$. For small COVs their square is small compared to unity, so the first term in the powers of x expressions are excellent approximations. For correlated products and quotients see Charles R. Mischke, *Mathematical Model Building*, 2nd rev. ed., Iowa State University Press, Ames, 1980, App. C.

The standard deviation follows the Pythagorean theorem. Thus the standard deviation for *both* addition and subtraction of independent variables is

$$\hat{\sigma}_z = \sqrt{\hat{\sigma}_x^2 + \hat{\sigma}_y^2} \tag{e}$$

Similar relations have been worked out for a variety of functions and are displayed in Table 20–6. The results shown can easily be combined to form other functions.

An unanswered question here is what is the distribution that results from the various operations? For answers to this question, statisticians use closure theorems and the central limit theorem.[2]

[2]See E. B. Haugen, *Probabilistic Mechanical Design*, Wiley, New York, 1980, pp. 49–54.

EXAMPLE 20–7 A round bar subject to a bending load has a diameter $\mathbf{d} = \mathbf{LN}(2.000, 0.002)$ in. This equivalency states that the mean diameter is $\mu_d = 2.000$ in and the standard deviation is $\hat{\sigma}_d = 0.002$ in. Find the mean and the standard deviation of the second moment of area.

Solution The second moment of area is given by the equation

$$\mathbf{I} = \frac{\pi \mathbf{d}^4}{64}$$

The coefficient of variation of the diameter is

$$C_d = \frac{\hat{\sigma}_d}{\mu_d} = \frac{0.002}{2} = 0.001$$

Using Table 20–6, we find

Answer

$$\mu_I = (\pi/64)\mu_d^4\left(1 + 6C_d^2\right) = (\pi/64)(2.000)^4[1 + 6(0.001)^2] = 0.785 \text{ in}^4$$

Answer

$$\hat{\sigma}_I = 4\mu_d^4 C_d\left[1 + (9/4)C_d^2\right] = 4(2.000)^4(0.001)[1 + (9/4)(0.001)^2] = 0.064 \text{ in}^4$$

These results can be expressed in the form

$$\mathbf{I} = \mathbf{LN}(0.785, 0.064) = 0.785\mathbf{LN}(1, 0.0815) \text{ in}^4$$

20–5 Linear Regression

Statisticians use a process of analysis called *regression* to obtain a curve that best fits a set of data points. The process is called *linear regression* when the best-fitting straight line is to be found. The meaning of the word *best* is open to argument, because there can be many meanings. The usual method, and the one employed here, regards a line as "best" if it minimizes the squares of the deviations of the data points from the line.

Figure 20–10 shows a set of data points approximated by the line *AB*. The standard equation of a straight line is

$$y = mx + b \tag{20–31}$$

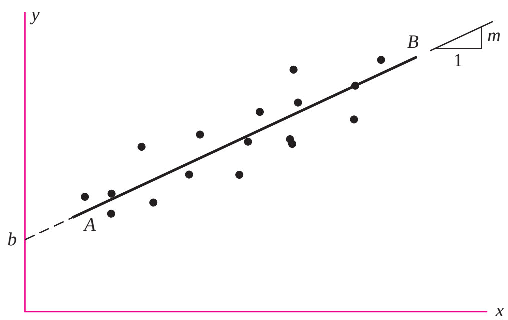


Figure 20–10

Set of data points approximated by regression line *AB*.

where m is the slope and b is the y intercept. Consider a set of N data points (x_i, y_i). In general, the best-fit line will not intersect a data point. Thus, we can write

$$y_i = mx_i + b + \epsilon_i \tag{b}$$

where $\epsilon_i = y_i - y$ is the deviation between the data point and the line. The sum of the squares of the deviations is given by[3]

$$\mathcal{E} = \sum \epsilon_i^2 = \sum (y_i - mx_i - b)^2 \tag{c}$$

Minimizing $\mathcal{E}$, the sum of the squared errors, expecting a stationary point minimum, requires $\partial\mathcal{E}/\partial m = 0$ and $\partial\mathcal{E}/\partial b = 0$. This results in two simultaneous equations for the slope and y intercept denoted as $\hat{m}$ and $\hat{b}$, respectively. Solving these equations results in

$$\hat{m} = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{N \sum x_i^2 - \left(\sum x_i\right)^2} = \frac{\sum x_i y_i - N\bar{x}\bar{y}}{\sum x_i^2 - N\bar{x}^2} \tag{20–32}$$

$$\hat{b} = \frac{\sum y_i - \hat{m} \sum x_i}{N} = \bar{y} - \hat{m}\bar{x} \tag{20–33}$$

Once you have established a slope and an intercept, the next point of interest is to discover how well x and y correlate with each other. If the data points are scattered all over the xy plane, there is obviously no correlation. But if all the data points coincide with the regression line, then there is perfect correlation. Most statistical data will be in between these extremes. A *correlation coefficient* r, having the range $-1 \le r \le +1$, has been devised to answer this question. The formula is

$$r = \hat{m}\frac{s_x}{s_y} \tag{20–34}$$

where s_x and s_y are the standard deviations of the x coordinates and y coordinates of the data, respectively. If $r = 0$, there is no correlation; if $r = \pm 1$, there is perfect correlation. A positive or negative r indicates that the regression line has a positive or negative slope, respectively.

The standard deviations for $\hat{m}$ and $\hat{b}$ are given by

$$s_{\hat{m}} = \frac{s_{y\cdot x}}{\sqrt{\sum (x_i - \bar{x})^2}} \tag{20–35}$$

$$s_{\hat{b}} = s_{y\cdot x}\sqrt{\frac{1}{N} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}} \tag{20–36}$$

where

$$s_{y\cdot x} = \sqrt{\frac{\sum y_i^2 - \hat{b} \sum y_i - \hat{m} \sum x_i y_i}{N - 2}} \tag{20–37}$$

is the standard deviation of the scatter of the data from the regression line.

[3]From this point on, for economy of notation, the limits of the summation of $i(1, N)$ will not be displayed.

EXAMPLE 20–8 A specimen of a medium carbon steel was tested in tension. With an extensometer in place, the specimen was loaded then unloaded, to see if the extensometer reading returned to the no-load reading, then the next higher load was applied. The loads and extensometer elongations were reduced to stress σ and strain ϵ, producing the following data:

σ, psi	5033	10 068	15 104	20 143	35 267
ϵ	0.000 20	0.000 30	0.000 50	0.000 65	0.001 15

Find the mean Young's modulus $\bar{E}$ and its standard deviation. Since the extensometer seems to have an initial reading at no load, use a $y = mx + b$ regression.

Solution From Table 20–7, $\bar{x} = 0.002\,80/5 = 0.000\,56$, $\bar{y} = 85\,615/5 = 17\,123$. Note that a regression line always contains the data centroid. From Eq. (20–32)

Answer

$$\hat{m} = \frac{5(65.229) - 0.0028(85\,615)}{5(0.000\,002\,125) - 0.0028^2} = 31.03(10^6)\text{ psi} = \bar{E}$$

From Eq. (20–33)

$$\hat{b} = \frac{0.000\,002\,125(85\,615) - 0.002\,80(65.229)}{5(0.000\,002\,125) - 0.0028^2} = -254.69\text{ psi}$$

From Eq. (20–34), obtaining s_x and s_y from a statistics calculator routine,

$$\hat{r} = \frac{\hat{m} s_x}{s_y} = \frac{31\,031\,597.85(3\,162\,163\,10^{-4})}{11\,601.11} = 0.998$$

From Eq. (20–37), the scatter about the regression line is measured by the standard deviation $s_{y\cdot x}$ and is equal to

$$s_{y\cdot x} = \sqrt{\frac{\sum y^2 - \hat{b}\sum y - \hat{m}\sum xy}{N-2}}$$

$$= \sqrt{\frac{2\,004\,328\,267 - (-254.69)85\,615 - 31.03(10^6)(65.229)}{5-2}}$$

$$= 811.1\text{ psi}$$

Table 20–7

Worksheet for Ex. 20–6

y σ, **psi**	**x** ϵ	x^2	xy	y^2	$(x-\bar{x})^2$
5 033	0.000 20	0.000 000 040	1.006 600	25 330 089	0.000 000 130
10 068	0.000 30	0.000 000 090	3.020 400	101 364 624	0.000 000 069
15 104	0.000 50	0.000 000 250	7.552 000	228 130 816	0.000 000 004
20 143	0.000 65	0.000 000 423	13.092 950	405 740 449	0.000 000 008
35 261	0.001 15	0.000 001 323	40.557 050	1 243 761 289	0.000 000 348
$\sum$ 85 615	0.002 80	0.000 002 125	65.229 000	2 004 328 267	0.000 000 556

Note: $\bar{y} = 85\,615/5 = 17\,123$ psi, $\bar{x} = 0.002\,80/5 = 0.000\,56$.

From Eq. (20–35), the standard deviation of $\hat{m}$ is

Answer

$$s_{\hat{m}} = \frac{s_{y \cdot x}}{\sqrt{\sum(x-\bar{x})^2}} = \frac{811.1}{\sqrt{0.000\,000\,558}} = 1.086(10^6)\text{ psi} = s_E$$

See Fig. 20–11 for the regression plot.

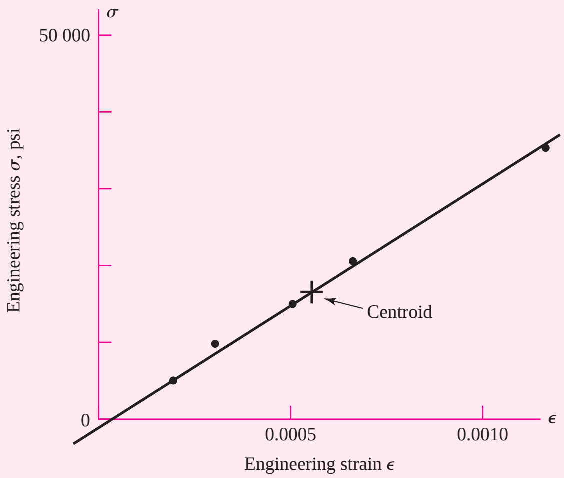


Figure 20–11

The data from Ex. 20–8 are plotted. The regression line passes through the data centroid and among the data points, minimizing the squared deviations.

PROBLEMS

20–1 At a constant amplitude, completely reversed bending stress level, the cycles-to-failure experience with 69 specimens of 5160H steel from 1.25-in hexagonal bar stock was as follows:

L	60	70	80	90	100	110	120	130	140	150	160	170	180	190	200	210
f	2	1	3	5	8	12	6	10	8	5	2	3	2	1	0	1

where L is the life in thousands of cycles, and f is the class frequency of failures.

(*a*) Construct a histogram with class frequency f as ordinate.

(*b*) Estimate the mean and standard deviation of the life for the population from which the sample was drawn.

20–2 Determinations of the ultimate tensile strength S_{ut} of stainless steel sheet (17-7PH, condition TH 1050), in sizes from 0.016 to 0.062 in, in 197 tests combined into seven classes were

S_{ut}, kpsi	174	182	190	198	206	214	222
Frequency, f	6	9	44	67	53	12	6

where f is the class frequency. Find the mean and standard deviation.

20–3 A total of 58 AISI 1018 cold-drawn steel bars were tested to determine the 0.2 percent offset yield strength S_y. The results were

S_y, kpsi	64	68	72	76	80	84	88	92
f	2	6	6	9	19	10	4	2

where S_y is the class midpoint and f is the class frequency. Estimate the mean and standard deviation of S_y and its PDF assuming a normal distribution.

20–4 The base 10 logarithm of 55 cycles-to-failure observations on specimens subjected to a constant stress level in fatigue have been classified as follows:

y	5.625	5.875	6.125	6.375	6.625	6.875	7.125	7.375	7.625	7.875	8.125
f	1	0	0	3	3	6	14	15	10	2	1

Here y is the class midpoint and f is the class frequency.

(*a*) Estimate the mean and standard deviation of the population from which the sample was taken and establish the normal PDF.

(*b*) Plot the histogram and superpose the predicted class frequency from the normal fit.

20–5 A $\frac{1}{2}$-in nominal diameter round is formed in an automatic screw machine operation that is initially set to produce a 0.5000-in diameter and is reset when tool wear produces diameters in excess of 0.5008 in. The stream of parts is thoroughly mixed and produces a uniform distribution of diameters.

(*a*) Estimate the mean and standard deviation of the large batch of parts from setup to reset.

(*b*) Find the expressions for the PDF and CDF of the population.

(*c*) If, by inspection, the diameters less than 0.5002 in are removed, what are the new PDF and CDF as well as the mean and standard deviation of the diameters of the survivors of the inspection?

20–6 The only detail drawing of a machine part has a dimension smudged beyond legibility. The round in question was created in an automatic screw machine and 1000 parts are in stock. A random sample of 50 parts gave a mean dimension of $\bar{d} = 0.6241$ in and a standard deviation of $s = 0.000\,581$ in. Toleranced dimensions elsewhere are given in integral thousandths of an inch. Estimate the missing information on the drawing.

20–7 (*a*) The CDF of the variate $\mathbf{x}$ is $F(x) = 0.555x - 33$, where x is in millimeters. Find the PDF, the mean, the standard deviation, and the range numbers of the distribution.

(*b*) In the expression $\sigma = \mathbf{F}/\mathbf{A}$, the force $\mathbf{F} = \mathbf{LN}(3600, 300)$ lbf and the area is $\mathbf{A} = \mathbf{LN}(0.112, 0.001)$ in^2. Estimate the mean, standard deviation, coefficient of variation, and distribution of σ.

20–8 A regression model of the form $y = a_1 x + a_2 x^2$ is desired. From the normal equations

$$\sum y = a_1 \sum x + a_2 \sum x^2$$

$$\sum xy = a_1 \sum x^2 + a_2 \sum x^3$$

show that

$$a_1 = \frac{\sum y \sum x^3 - \sum xy \sum x^2}{\sum x \sum x^3 - \left(\sum x^2\right)^2} \qquad \text{and} \qquad a_2 = \frac{\sum x \sum xy - \sum y \sum x^2}{\sum x \sum x^3 - \left(\sum x^2\right)^2}$$

For the data set

x	0.0	0.2	0.4	0.6	0.8	1.0
y	0.01	0.15	0.25	0.25	0.17	−0.01

find the regression equation and plot the data with the regression model.

20–9 R. W. Landgraf reported the following axial (push–pull) endurance strengths for steels of differing ultimate strengths:

S_u	S_e'	S_u	S_e'	S_u	S_e'
65	29.5	325	114	280	96
60	30	238	109	295	99
82	45	130	67	120	48
64	48	207	87	180	84
101	51	205	96	213	75
119	50	225	99	242	106
195	78	325	117	134	60
210	87	355	122	145	64
230	105	225	87	227	116
265	105				

(*a*) Plot the data with S_e' as ordinate and S_u as abscissa.
(*b*) Using the $y = mx + b$ linear regression model, find the regression line and plot.

20–10 In fatigue studies a parabola of the Gerber type

$$\frac{\sigma_a}{S_e} + \left(\frac{\sigma_m}{S_{ut}}\right)^2 = 1$$

is useful (see Sec. 6–12). Solved for σ_a the preceding equation becomes

$$\sigma_a = S_e - \frac{S_e}{S_{ut}^2}\sigma_m^2$$

This implies a regression model of the form $y = a_0 + a_2x^2$. Show that the normal equations are

$$\sum y = na_0 + a_2 \sum x^2$$
$$\sum xy = a_0 \sum x + a_2 \sum x^3$$

and that

$$a_0 = \frac{\sum x^3 \sum y - \sum x^2 \sum xy}{n \sum x^3 - \sum x \sum x^2} \qquad \text{and} \qquad a_2 = \frac{n \sum xy - \sum x \sum y}{n \sum x^3 - \sum x \sum x^2}$$

Plot the data

x	20	40	60	80
y	19	17	13	7

superposed on a plot of the regression line.

20–11 Consider the following data collected on a single helical coil extension spring with an initial extension F_i and a spring rate k suspected of being related by the equation $F = F_i + kx$ where x is the deflection beyond initial. The data are

x, in	0.2	0.4	0.6	0.8	1.0	2.0
F, lbf	7.1	10.3	12.1	13.8	16.2	25.2

(*a*) Estimate the mean and standard deviation of the initial tension F_i.
(*b*) Estimate the mean and standard deviation of the spring rate k.

20–12 In the expression for uniaxial strain $\boldsymbol{\epsilon} = \boldsymbol{\delta}/\mathbf{l}$, the elongation is specified as $\boldsymbol{\delta} \sim (0.0015, 0.000\,092)$ in and the length as $\mathbf{l} \sim (2.0000, 0.0081)$ in. What are the mean, the standard deviation, and the coefficient of variation of the corresponding strain $\boldsymbol{\epsilon}$.

20–13 In Hooke's law for uniaxial stress, $\boldsymbol{\sigma} = \boldsymbol{\epsilon}\mathbf{E}$, the strain is given as $\boldsymbol{\epsilon} \sim (0.0005, 0.000\,034)$ and Young's modulus as $\mathbf{E} \sim (29.5, 0.885)$ Mpsi. Find the mean, the standard deviation, and the coefficient of variation of the corresponding stress $\boldsymbol{\sigma}$ in psi.

20–14 The stretch of a uniform rod in tension is given by the formula $\delta = Fl/AE$. Suppose the terms in this equation are random variables and have parameters as follows:

$$\mathbf{F} \sim (14.7, 1.3)\ \text{kip} \qquad \mathbf{A} \sim (0.226, 0.003)\ \text{in}^2$$
$$\mathbf{l} \sim (1.5, 0.004)\ \text{in} \qquad \mathbf{E} \sim (29.5, 0.885)\ \text{Mpsi}$$

Estimate the mean, the standard deviation, and the coefficient of variation of the corresponding elongation $\boldsymbol{\delta}$ in inches.

20–15 The maximum bending stress in a round bar in flexure occurs in the outer surface and is given by the equation $\sigma = 32M/\pi d^3$. If the moment is specified as $\mathbf{M} \sim (15\,000, 1350)$ lbf · in and the diameter is $\mathbf{d} \sim (2.00, 0.005)$ in, find the mean, the standard deviation, and the coefficient of variation of the corresponding stress $\boldsymbol{\sigma}$ in psi.

20–16 When a production process is wider than the tolerance interval, inspection rejects a low-end scrap fraction α with $x < x_1$ and an upper-end scrap fraction β with dimensions $x > x_2$. The surviving population has a new density function $g(x)$ related to the original $f(x)$ by a multiplier a. This is because any two observations x_i and x_j will have the same relative probability of occurrence as before. Show that

$$a = \frac{1}{F(x_2) - F(x_1)} = \frac{1}{1 - (\alpha + \beta)}$$

and

$$g(x) = \begin{cases} \dfrac{f(x)}{F(x_2) - F(x_1)} = \dfrac{f(x)}{1 - (\alpha + \beta)} & x_1 \le x \le x_2 \\ 0 & \text{otherwise} \end{cases}$$

20–17 An automatic screw machine produces a run of parts with a uniform distribution $\mathbf{d} = \mathbf{U}[0.748, 0.751]$ in because it was not reset when the diameters reached 0.750 in. The square brackets contain range numbers.
(*a*) Estimate the mean, standard deviation, and PDF of the original production run if the parts are thoroughly mixed.
(*b*) Using the results of Prob. 20–16, find the new mean, standard deviation, and PDF. Superpose the PDF plots and compare.

20–18 A springmaker is supplying helical coil springs meeting the requirement for a spring rate k of 10 ± 1 lbf/in. The test program of the springmaker shows that the distribution of spring rate is well approximated by a normal distribution. The experience with inspection has shown that 8.1 percent are scrapped with $k < 9$ and 5.5 percent are scrapped with $k > 11$. Estimate the probability density function.

20–19 The lives of parts are often expressed as the number of cycles of operation that a specified percentage of a population will exceed before experiencing failure. The symbol L is used to designate this definition of life. Thus we can speak of L10 life as the number of cycles to failure exceeded by 90 percent of a population of parts. Using the mean and standard deviation for the data of Prob. 20–1, a normal distribution model, estimate the corresponding L10 life.

20–20 Fit a normal distribution to the histogram of Prob. 20–1. Superpose the probability density function on the $f/(Nw)$ histographic plot.

20–21 For Prob. 20–2, plot the histogram with $f/(Nw)$ as ordinate and superpose a normal distribution density function on the histographic plot.

20–22 For Prob. 20–3, plot the histogram with $f/(Nw)$ as ordinate and superpose a normal distribution probability density function on the histographic plot.

20–23 A 1018 cold-drawn steel has a 0.2 percent tensile yield strength $\mathbf{S}_y = \mathbf{N}(78.4, 5.90)$ kpsi. A round rod in tension is subjected to a load $\mathbf{P} = \mathbf{N}(40, 8.5)$ kip. If rod diameter d is 1.000 in, what is the probability that a random static tensile load P from $\mathbf{P}$ imposed on the shank with a 0.2 percent tensile load S_y from $\mathbf{S}_y$ will not yield?

20–24 A hot-rolled 1035 steel has a 0.2 percent tensile yield strength $\mathbf{S}_y = \mathbf{LN}(49.6, 3.81)$ kpsi. A round rod in tension is subjected to a load $\mathbf{P} = \mathbf{LN}(30, 5.1)$ kip. If the rod diameter d is 1.000 in, what is the probability that a random static tensile load P from $\mathbf{P}$ on a shank with a 0.2 percent yield strength S_y from $\mathbf{S}_y$ will not yield?

20–25 The tensile 0.2 percent offset yield strength of AISI 1137 cold-drawn steel rounds up to 1 inch in diameter from 2 mills and 25 heats is reported histographically as follows:

S_y	93	95	97	99	101	103	105	107	109	111
f	19	25	38	17	12	10	5	4	4	2

where S_y is the class midpoint in kpsi and f is the number in each class. Presuming the distribution is normal, what is the yield strength exceeded by 99 percent of the population?

20–26 Repeat Prob. 20–25, presuming the distribution is lognormal. What is the yield strength exceeded by 99 percent of the population? Compare the normal fit of Prob. 20–25 with the lognormal fit by superposing the PDFs and the histographic PDF.

20–27 A 1046 steel, water-quenched and tempered for 2 h at 1210°F, has a mean tensile strength of 105 kpsi and a yield mean strength of 82 kpsi. Test data from endurance strength testing at 10^4-cycle life give $(\mathbf{S}'_{fe})_{10^4} = \mathbf{W}[79,\ 86.2,\ 2.60]$ kpsi. What are the mean, standard deviation, and coefficient of variation of $(\mathbf{S}'_{fe})_{10^4}$?

20–28 An ASTM grade 40 cast iron has the following result from testing for ultimate tensile strength: $\mathbf{S}_{ut} = \mathbf{W}[27.7, 46.2, 4.38]$ kpsi. Find the mean and standard deviation of $\mathbf{S}_{ut}$, and estimate the chance that the ultimate strength is less than 40 kpsi.

20–29 A cold-drawn 301SS stainless steel has an ultimate tensile strength given by $\mathbf{S}_{ut} = \mathbf{W}[151.9, 193.6, 8.00]$ kpsi. Find the mean and standard deviation.

20–30 A 100-70-04 nodular iron has tensile and yield strengths described by

$$\mathbf{S}_{ut} = \mathbf{W}[47.6, 125.6, 11.84] \text{ kpsi}$$
$$\mathbf{S}_{y} = \mathbf{W}[64.1, 81.0, 3.77] \text{ kpsi}$$

What is the chance that S_{ut} is less than 100 kpsi? What is the chance that S_y is less than 70 kpsi?

20–31 A 1038 heat-treated steel bolt in finished form provided the material from which a tensile test specimen was made. The testing of many such bolts led to the description $\mathbf{S}_{ut} = \mathbf{W}[122.3, 134.6, 3.64]$ kpsi. What is the probability that the bolts meet the SAE grade 5 requirement of a minimum tensile strength of 120 kpsi? What is the probability that the bolts meet the SAE grade 7 requirement of a minimum tensile strength of 133 kpsi?

20–32 A 5160H steel was tested in fatigue and the distribution of cycles to failure at constant stress level was found to be $\mathbf{n} = \mathbf{W}[36.9, 133.6, 2.66]$ in 10^3 cycles. Plot the PDF of n and the PDF of the lognormal distribution having the same mean and standard deviation. What is the L10 life (see Prob. 20–19) predicted by both distributions?

20–33 A material was tested at steady fully reversed loading to determine the number of cycles to failure using 100 specimens. The results were

$(10^{-5})L$	3.05	3.55	4.05	4.55	5.05	5.55	6.05	6.55	7.05	7.55	8.05	8.55	9.05	9.55	10.05
f	3	7	11	16	21	13	13	6	2	0	4	3	0	0	1

where L is the life in cycles and f is the number in each class. Assuming a lognormal distribution, plot the theoretical PDF and the histographic PDF for comparison.

20–34 The ultimate tensile strength of an AISI 1117 cold-drawn steel is Weibullian, with $\mathbf{S}_u = \mathbf{W}[70.3, 84.4, 2.01]$. What are the mean, the standard deviation, and the coefficient of variation?

20–35 A 60-45-15 nodular iron has a 0.2 percent yield strength S_y with a mean of 49.0 kpsi, a standard deviation of 4.2 kpsi, and a guaranteed yield strength of 33.8 kpsi. What are the Weibull parameters θ and b?

20–36 A 35018 malleable iron has a 0.2 percent offset yield strength given by the Weibull distribution $\mathbf{S}_y = \mathbf{W}[34.7, 39.0, 2.93]$ kpsi. What are the mean, the standard deviation, and the coefficient of variation?

20–37 The histographic results of steady load tests on 237 rolling-contact bearings are:

L	1	2	3	4	5	6	7	8	9	10	11	12
f	11	22	38	57	31	19	15	12	11	9	7	5

where L is the life in millions of revolutions and f is the number of failures. Fit a lognormal distribution to these data and plot the PDF with the histographic PDF superposed. From the lognormal distribution, estimate the life at which 10 percent of the bearings under this steady loading will have failed.