1

The Role of Statistics in Engineering

CHAPTER OUTLINE

4-6 Normal Distribution

5-5 Linear Combinations of Random Variables

1-1 The Engineering Method and Statistical Thinking

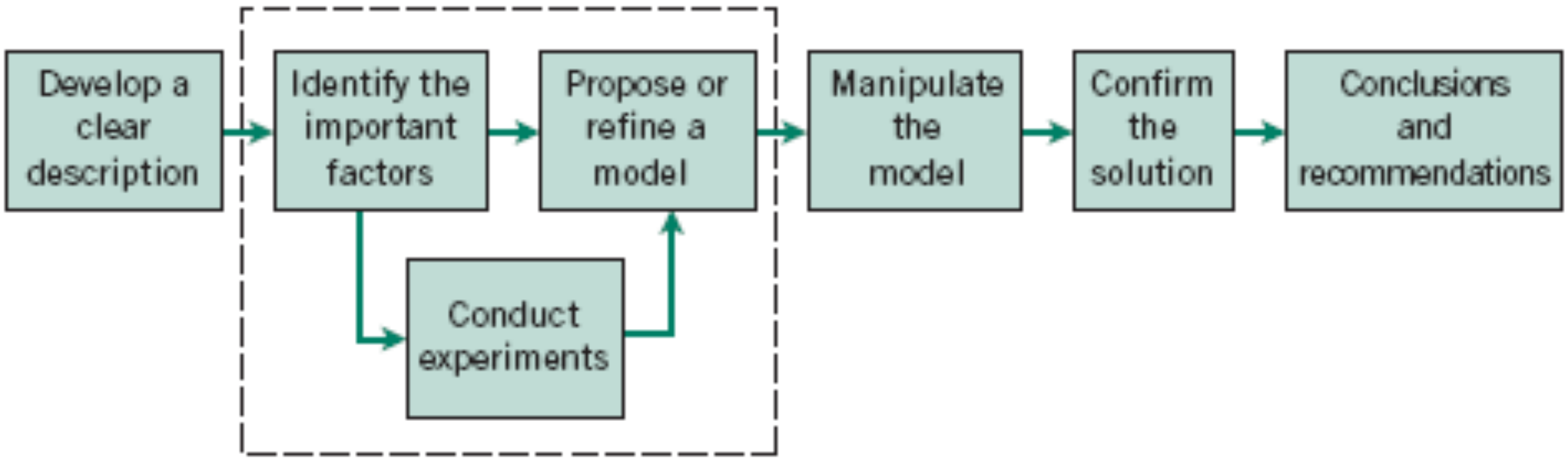


Figure 1.1 The engineering method

1-1 The Engineering Method and Statistical Thinking

The field of statistics deals with the collection, presentation, analysis, and use of data to

- Make decisions
- Solve problems
- Design products and processes

1-1 The Engineering Method and Statistical Thinking

• Statistical techniques are useful for describing and understanding variability.

• By variability, we mean successive observations of a system or phenomenon do *not* produce exactly the same result.

• Statistics gives us a framework for describing this variability and for learning about potential sources of variability.

1-1 The Engineering Method and Statistical Thinking

Engineering Example

An engineer is designing a nylon connector to be used in an automotive engine application. The engineer is considering establishing the design specification on wall thickness at 3/32 inch but is somewhat uncertain about the effect of this decision on the connector pull-off force. If the pull-off force is too low, the connector may fail when it is installed in an engine. Eight prototype units are produced and their pull-off forces measured (in pounds): 12.6, 12.9, 13.4, 12.3, 13.6, 13.5, 12.6, 13.1.

1-1 The Engineering Method and Statistical Thinking

Engineering Example

•The **dot diagram** is a very useful plot for displaying a small body of data - say up to about 20 observations.
• This plot allows us to see easily two features of the data; the **location**, or the middle, and the **scatter** or **variability**.

Figure 1-2 Dot diagram of the pull-off force data when wall thickness is 3/32 inch.

1-1 The Engineering Method and Statistical Thinking

Engineering Example

- The engineer considers an alternate design and eight prototypes are built and pull-off force measured.
- The dot diagram can be used to compare two sets of data

Figure 1-3 Dot diagram of pull-off force for two wall thicknesses.

1-1 The Engineering Method and Statistical Thinking

Engineering Example

- Since pull-off force varies or exhibits variability, it is a random variable.
- A random variable, X, can be model by

$$X = \mu + \varepsilon$$

where μ is a constant and ε a random disturbance.

1-1 The Engineering Method and Statistical Thinking

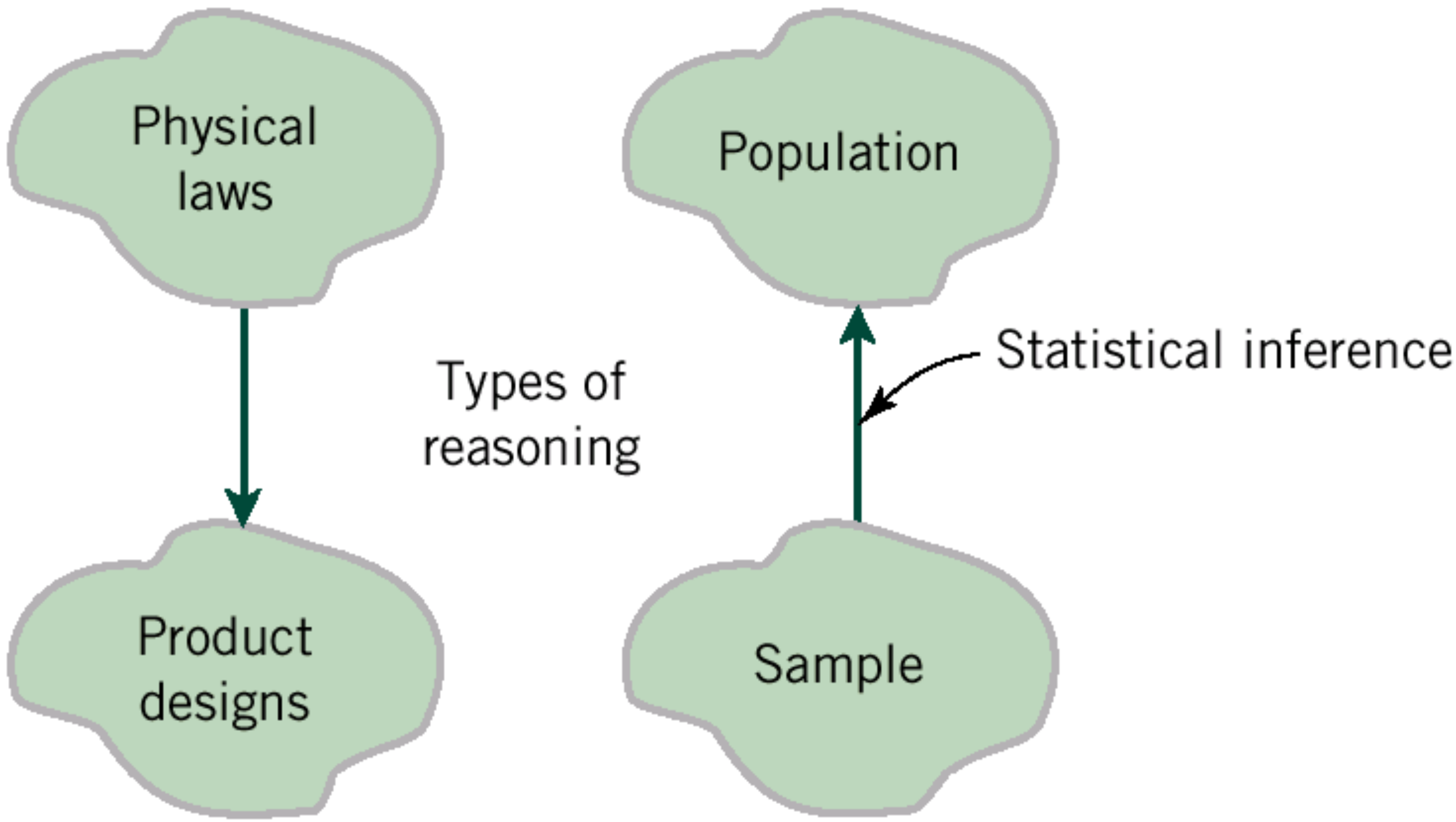


Figure 1-4 Statistical inference is one type of reasoning.

1-2 Collecting Engineering Data

Three basic methods for collecting data:

- A **retrospective** study using historical data
  - May not be useful
- An **observational** study
  - Cannot tell the cause-effect
- A **designed experiment**
  - Make deliberate changes to observe response
  - Can tell the cause-effect

1-3 Mechanistic and Empirical Models

A **mechanistic model** is built from our underlying knowledge of the basic physical mechanism that relates several variables.

Ohm's Law: Current = voltage/resistance

$$I = E/R \quad or \quad I = E/R + \varepsilon$$

An **empirical model** is built from our engineering and scientific knowledge of the phenomenon, but is not directly developed from our theoretical or first-principles understanding of the underlying mechanism.

Table 1-2 Wire Bond Pull Strength Data

Observation Number	Pull Strength y	Wire Length x_1	Die Height x_2
1	9.95	2	50
2	24.45	8	110
3	31.75	11	120
4	35.00	10	550
5	25.02	8	295
6	16.86	4	200
7	14.38	2	375
8	9.60	2	52
9	24.35	9	100
10	27.50	8	300
11	17.08	4	412
12	37.00	11	400
13	41.95	12	500
14	11.66	2	360
15	21.65	4	205
16	17.89	4	400
17	69.00	20	600
18	10.30	1	585
19	34.93	10	540
20	46.59	15	250
21	44.88	15	290
22	54.12	16	510
23	56.63	17	590
24	22.13	6	100
25	21.15	5	400

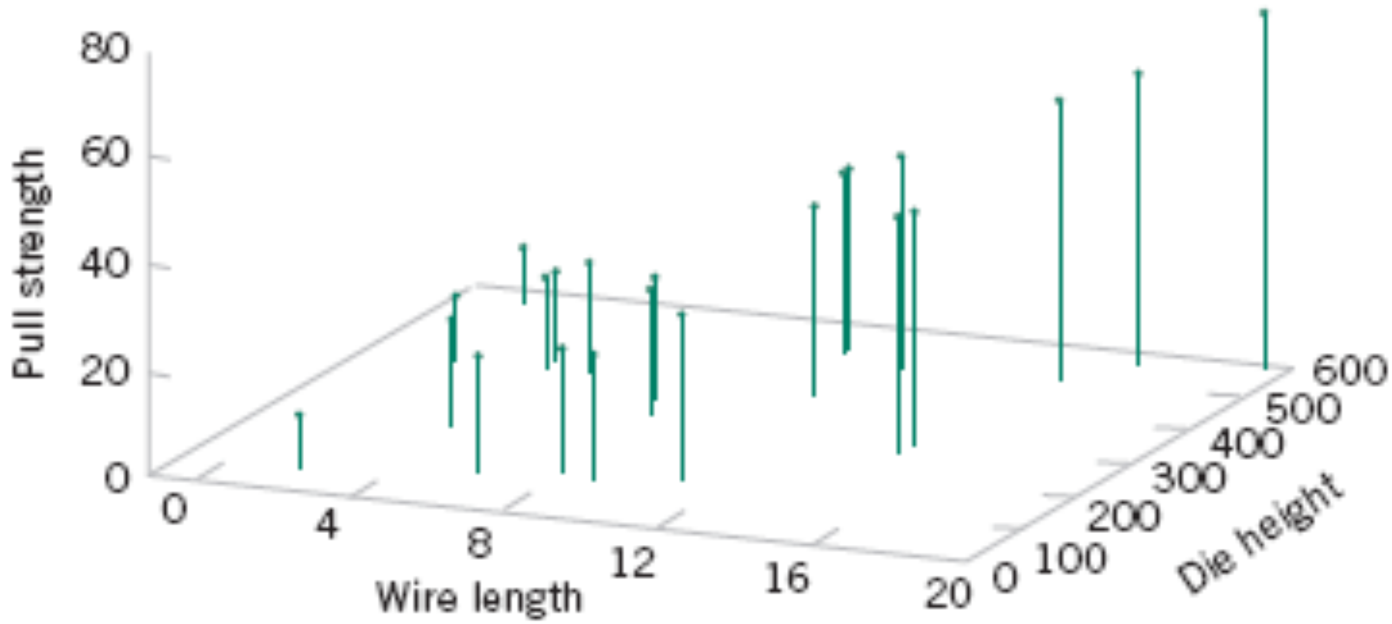


Figure 1-15 Three-dimensional plot of the wire and pull strength data.

1-3 Mechanistic and Empirical Models

$$\text{Pull strength} = \beta_0 + \beta_1(\text{wire length}) + \beta_2(\text{die height}) + \epsilon$$

In general, this type of empirical model is called a **regression model**.

The **estimated** regression line is given by

$$\widehat{\text{Pull strength}} = 2.26 + 2.74(\text{wire length}) + 0.0125(\text{die height})$$

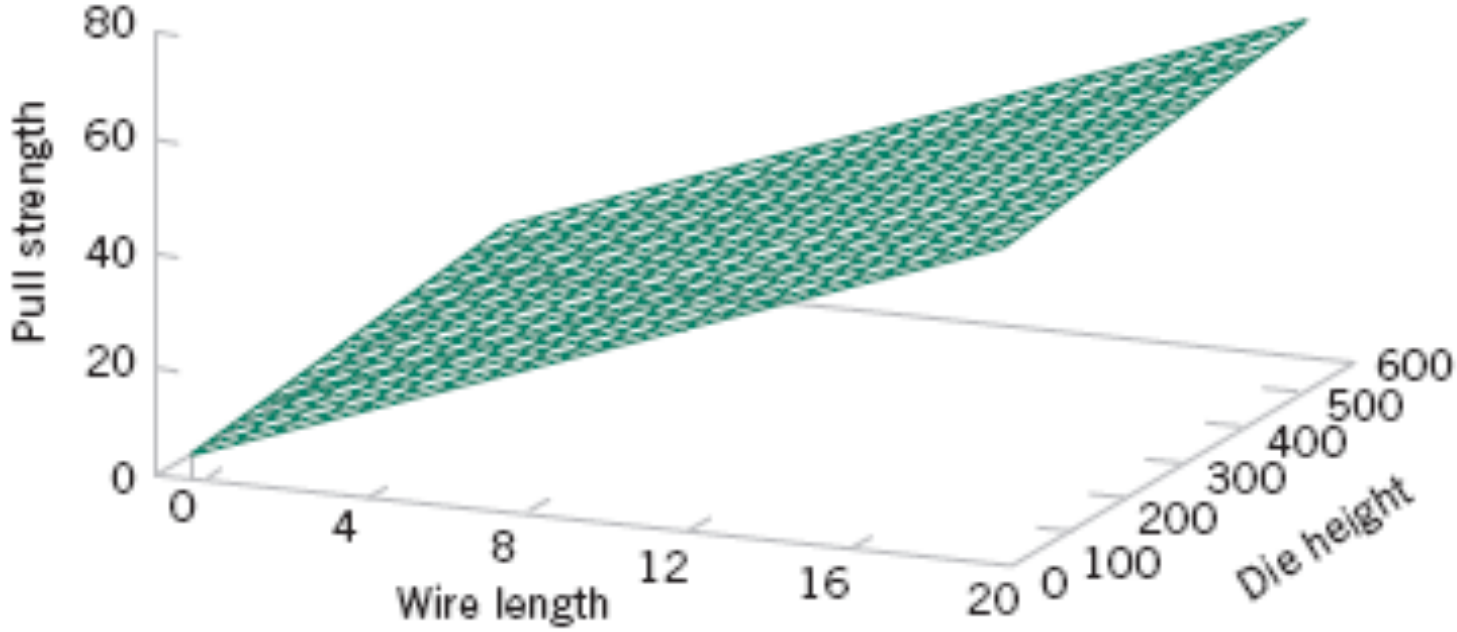


Figure 1-16 Plot of the predicted values of pull strength from the empirical model.

4-6 Normal Distribution

Definition

A random variable X with probability density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty \tag{4-8}$$

is a **normal random variable** with parameters μ, where $-\infty < \mu < \infty$, and $\sigma > 0$. Also,

$$E(X) = \mu \quad \text{and} \quad V(X) = \sigma^2 \tag{4-9}$$

and the notation $N(\mu, \sigma^2)$ is used to denote the distribution. The mean and variance of X are shown to equal μ and σ^2, respectively, at the end of this Section 5-6.

4-6 Normal Distribution

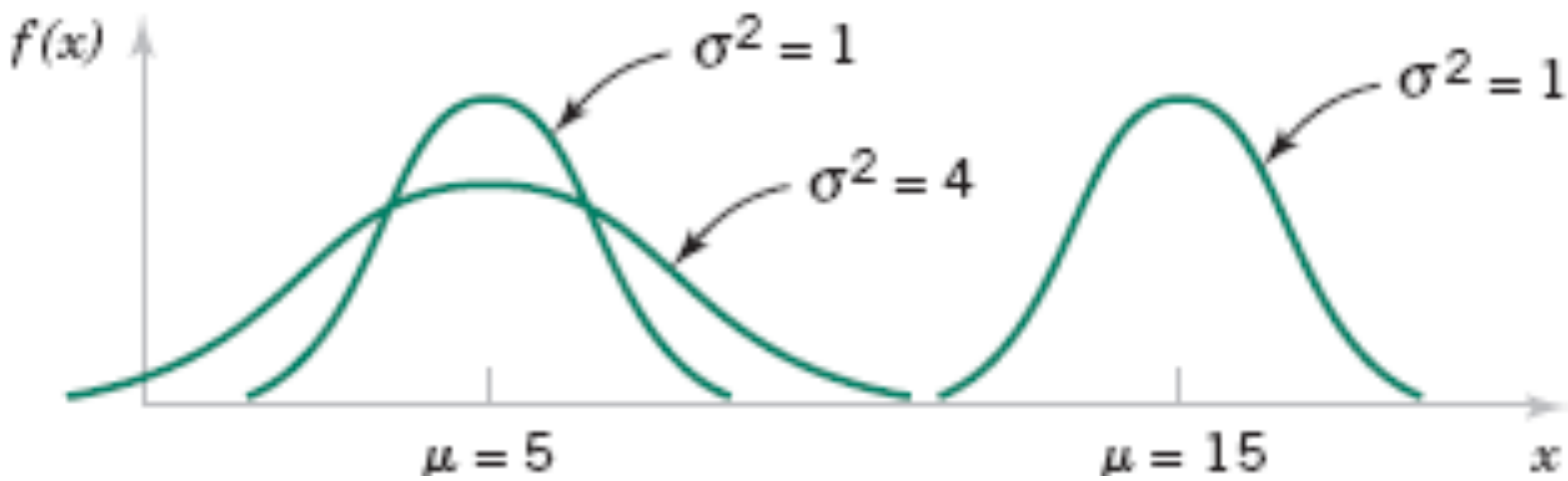


Figure 4-10 Normal probability density functions for selected values of the parameters μ and σ^2.

4-6 Normal Distribution

Definition : Standard Normal

A normal random variable with

$$\mu = 0 \quad \text{and} \quad \sigma^2 = 1$$

is called a **standard normal random variable** and is denoted as Z.

The cumulative distribution function of a standard normal random variable is denoted as

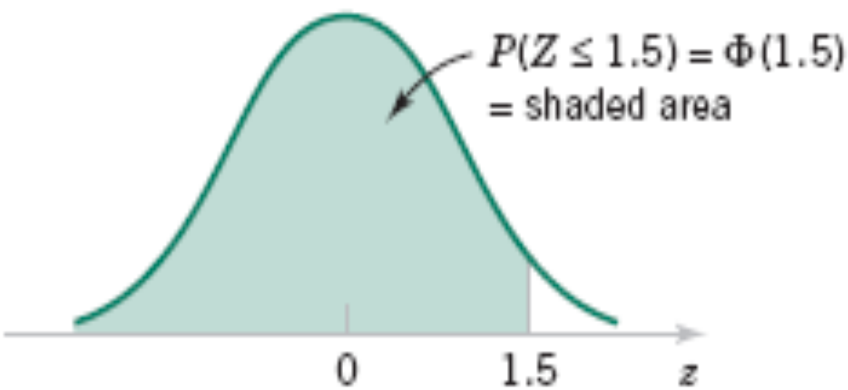
$$\Phi(z) = P(Z \leq z)$$

4-6 Normal Distribution

Example 4-11

Assume Z is a standard normal random variable. Appendix Table II provides probabilities of the form $P(Z \leq z)$. The use of Table II to find $P(Z \leq 1.5)$ is illustrated in Fig. 4-13. Read down the z column to the row that equals 1.5. The probability is read from the adjacent column, labeled 0.00, to be 0.93319.

The column headings refer to the hundredth's digit of the value of z in $P(Z \leq z)$. For example, $P(Z \leq 1.53)$ is found by reading down the z column to the row 1.5 and then selecting the probability from the column labeled 0.03 to be 0.93699.



z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

Figure 4-13 Standard normal probability density function.

4-6 Normal Distribution

Standardizing

If X is a normal random variable with $E(X) = \mu$ and $V(X) = \sigma^2$, the random variable

$$Z = \frac{X - \mu}{\sigma} \tag{4-10}$$

is a normal random variable with $E(Z) = 0$ and $V(Z) = 1$. That is, Z is a standard normal random variable.

4-6 Normal Distribution

Example 4-13

Suppose the current measurements in a strip of wire are assumed to follow a normal distribution with a mean of 10 milliamperes and a variance of 4 (milliamperes)^2. What is the probability that a measurement will exceed 13 milliamperes?

Let X denote the current in milliamperes. The requested probability can be represented as $P(X > 13)$. Let $Z = (X - 10)/2$. The relationship between the several values of X and the transformed values of Z are shown in Fig. 4-15. We note that $X > 13$ corresponds to $Z > 1.5$. Therefore, from Appendix Table II,

$$P(X > 13) = P(Z > 1.5) = 1 - P(Z \leq 1.5) = 1 - 0.93319 = 0.06681$$

Rather than using Fig. 4-15, the probability can be found from the inequality $X > 13$. That is,

$$P(X > 13) = P\left(\frac{(X - 10)}{2} > \frac{(13 - 10)}{2}\right) = P(Z > 1.5) = 0.06681$$

4-6 Normal Distribution

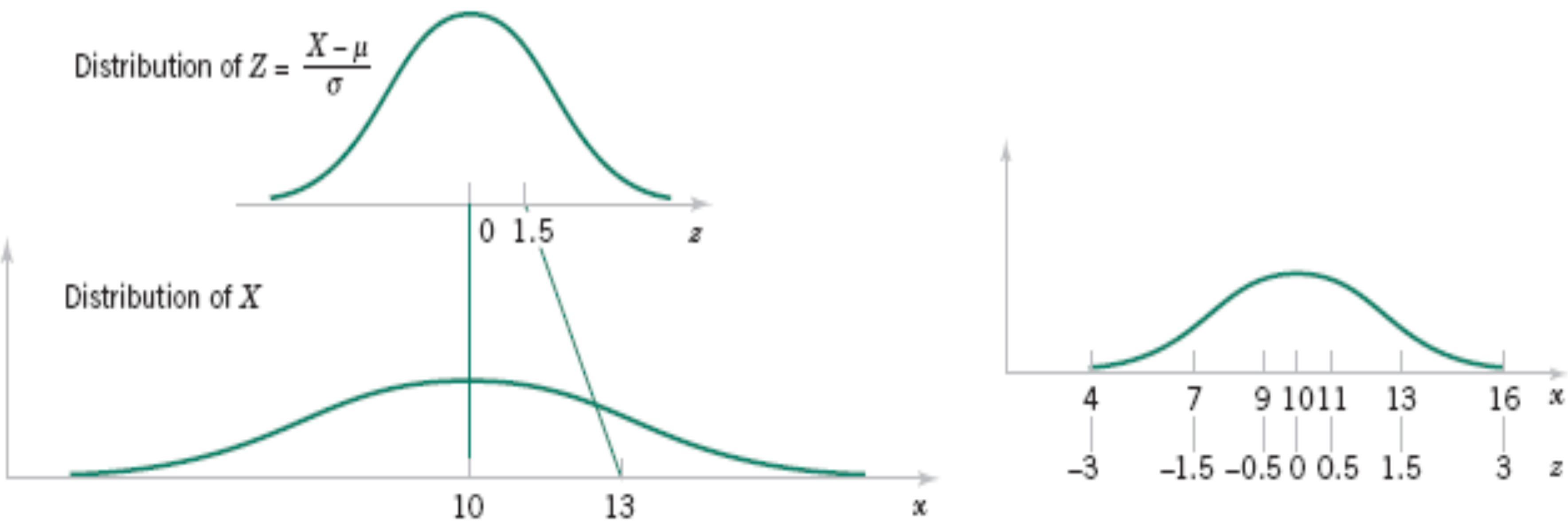


Figure 4-15 Standardizing a normal random variable.

4-6 Normal Distribution

To Calculate Probability

Suppose X is a normal random variable with mean μ and variance σ^2. Then,

$$P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P(Z \le z) \qquad (4\text{-}11)$$

where Z is a **standard normal random variable,** and $z = \dfrac{(x-\mu)}{\sigma}$ is the **z-value** obtained by **standardizing** X.

The probability is obtained by entering Appendix Table II with $z = (x - \mu)/\sigma$.

4-6 Normal Distribution

Example 4-14 (continued)

Determine the value for which the probability that a current measurement is below this value is 0.98. The requested value is shown graphically in Fig. 4-16. We need the value of x such that $P(X < x) = 0.98$. By standardizing, this probability expression can be written as

$$\begin{aligned} P(X < x) &= P((X - 10)/2 < (x - 10)/2) \\ &= P(Z < (x - 10)/2) \\ &= 0.98 \end{aligned}$$

Appendix Table II is used to find the z-value such that $P(Z < z) = 0.98$. The nearest probability from Table II results in

$$P(Z < 2.05) = 0.97982$$

Therefore, $(x - 10)/2 = 2.05$, and the standardizing transformation is used in reverse to solve for x. The result is

$$x = 2(2.05) + 10 = 14.1 \text{ milliamperes}$$

4-6 Normal Distribution

Example 4-14 (continued)

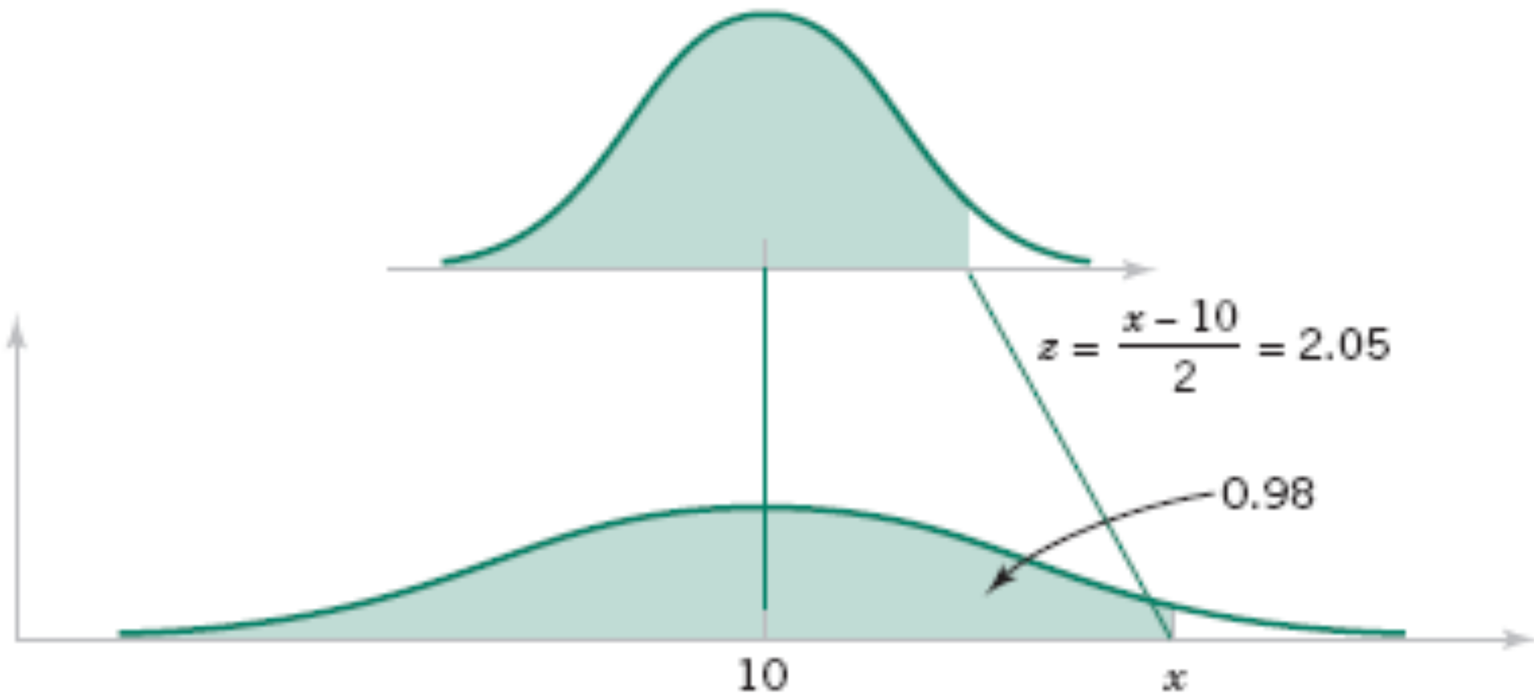


Figure 4-16 Determining the value of *x* to meet a specified probability.

5-5 Linear Combinations of Random Variables

Definition

Given random variables $X_1, X_2, \ldots, X_p$ and constants $c_1, c_2, \ldots, c_p$,

$$Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p \tag{5-34}$$

is a **linear combination** of $X_1, X_2, \ldots, X_p$.

Mean of a Linear Combination

If $Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p$,

$$E(Y) = c_1E(X_1) + c_2E(X_2) + \cdots + c_pE(X_p) \tag{5-35}$$

5-5 Linear Combinations of Random Variables

Variance of a Linear Combination

If $X_1, X_2, \ldots, X_p$ are random variables, and $Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p$, then in general

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_p^2V(X_p) + 2\sum_{i<j}\sum c_ic_j \operatorname{cov}(X_i, X_j) \quad (5\text{-}36)$$

If $X_1, X_2, \ldots, X_p$ are **independent,**

$$V(Y) = c_1^2V(X_1) + c_2^2V(X_2) + \cdots + c_p^2V(X_p) \quad (5\text{-}37)$$

5-5 Linear Combinations of Random Variables

Example 5-33

An important use of equation 5-37 is in **error propagation** that is presented in the following example.

A semiconductor product consists of three layers. If the variances in thickness of the first, second, and third layers are 25, 40, and 30 nanometers squared, what is the variance of the thickness of the final product.

Let X_1, X_2, X_3, and X be random variables that denote the thickness of the respective layers, and the final product. Then

$$X = X_1 + X_2 + X_3$$

The variance of X is obtained from equaion 5-39

$$V(X) = V(X_1) + V(X_2) + V(X_3) = 25 + 40 + 30 = 95 \text{ nm}^2$$

Consequently, the standard deviation of thickness of the final product is $95^{1/2} = 9.75$ nm and this shows how the variation in each layer is propagated to the final product.

5-5 Linear Combinations of Random Variables

Mean and Variance of an Average

If $\overline{X} = (X_1 + X_2 + \cdots + X_p)/p$ with $E(X_i) = \mu$ for $i = 1, 2, \ldots, p$

$$E(\overline{X}) = \mu \tag{5-38a}$$

if $X_1, X_2, \ldots, X_p$ are also independent with $V(X_i) = \sigma^2$ for $i = 1, 2, \ldots, p$,

$$V(\overline{X}) = \frac{\sigma^2}{p} \tag{5-38b}$$

5-5 Linear Combinations of Random Variables

Reproductive Property of the Normal Distribution

If $X_1, X_2, \ldots, X_p$ are independent, normal random variables with $E(X_i) = \mu_i$ and $V(X_i) = \sigma_i^2$, for $i = 1, 2, \ldots, p$,

$$Y = c_1X_1 + c_2X_2 + \cdots + c_pX_p$$

is a normal random variable with

$$E(Y) = c_1\mu_1 + c_2\mu_2 + \cdots + c_p\mu_p$$

and

$$V(Y) = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \cdots + c_p^2\sigma_p^2 \tag{5-39}$$

5-5 Linear Combinations of Random Variables

Example 5-34

Let the random variables X_1 and X_2 denote the length and width, respectively, of a manufactured part. Assume that X_1 is normal with $E(X_1) = 2$ centimeters and standard deviation 0.1 centimeter and that X_2 is normal with $E(X_2) = 5$ centimeters and standard deviation 0.2 centimeter. Also, assume that X_1 and X_2 are independent. Determine the probability that the perimeter exceeds 14.5 centimeters.

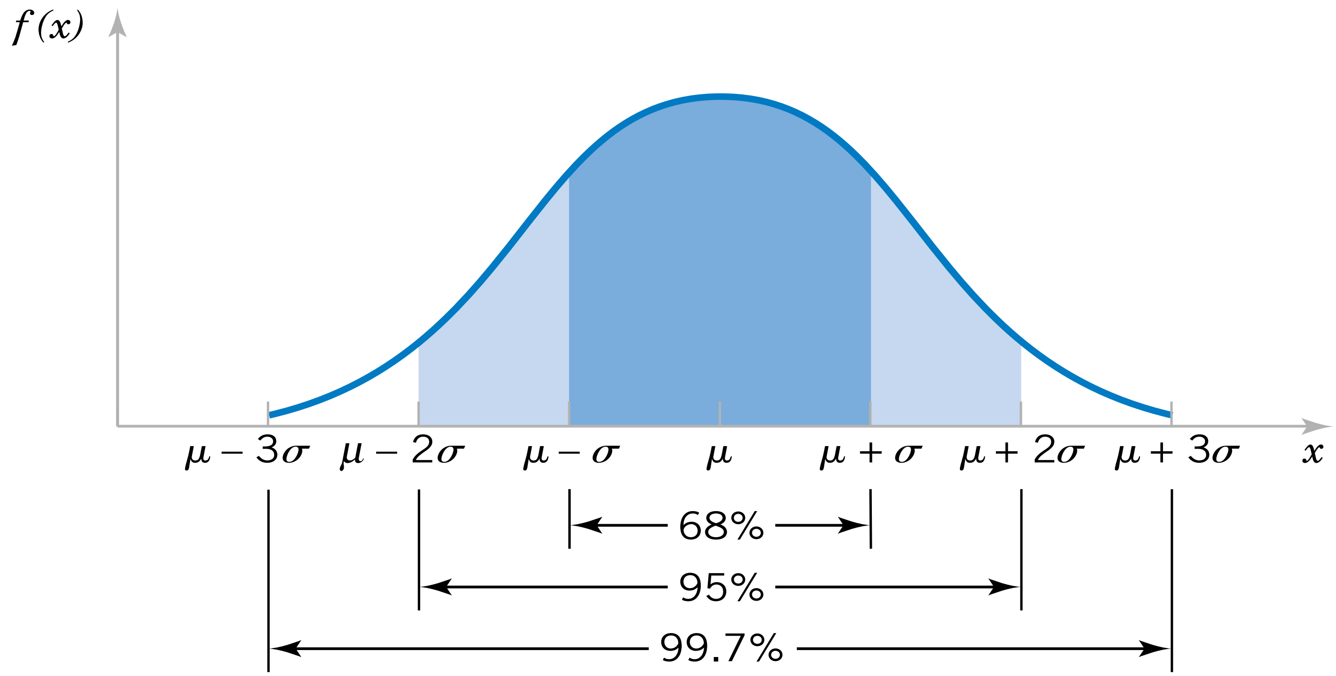
Then, $Y = 2X_1 + 2X_2$ is a normal random variable that represents the perimeter of the part. We obtain, $E(Y) = 14$ centimeters and the variance of Y is

$$V(Y) = 4 \times 0.1^2 + 4 \times 0.2^2 = 0.2$$

Now,

$$\begin{aligned} P(Y > 14.5) &= P[(Y - \mu_Y)/\sigma_Y > (14.5 - 14)/\sqrt{0.2}] \\ &= P(Z > 1.12) = 0.13 \end{aligned}$$

Some useful results to remember



For any normal random variable

$$P(\mu - \sigma < X < \mu + \sigma) = 0.6827$$
$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$$