FIGURE 11–4
From-to chart showing number of materials handling trips per day

From \ To	Saws	Milling	Punch press	Drills	Lathes	Sanders
Saws		43	26	14	40	
Milling			75	60		23
Punch press					45	16
Drills		22			28	
Lathes		45		30		60
Sanders		12				

Suppose that the firm's accounting department has estimated that the average cost of transporting material one foot in the machine shop is 20 cents. Using this fact, one can develop a third from-to chart that gives the average daily cost of materials handling from every department to every other department. For example, the distance separating saws and milling in the current layout is 18 feet (from Figure 11–3) and there are an average of 43 materials handling trips between these two departments (from Figure 11–4). This translates to a total of (43)(18) = 774 feet traveled in a day or a total cost of (774)(0.2) = $154.80 per day for materials handling between saws and milling. The materials handling costs for the other pairs of departments appear in Figure 11–5.

FIGURE 11–5
From-to chart showing materials handling cost per day (in $)

From \ To	Saws	Milling	Punch press	Drills	Lathes	Sanders
Saws		154.8	208	84	520	
Milling			570	900		138
Punch press					342	38.4
Drills		330			280	
Lathes		144		300		720
Sanders		72				

From-to charts are not a means of determining layouts, but simply a convenient way to express important flow characteristics of an existing layout. They can be useful in comparing the materials handling costs of a small number of alternatives. Because criteria other than the materials handling cost are relevant, the from-to chart should be supplemented with additional information, such as that contained in an activity relationship chart.

11.3 TYPES OF LAYOUTS

Different philosophies of layout design are appropriate in different manufacturing environments. Chapter 1 discussed the problem of matching the product life cycle with the process life cycle represented by the product–process matrix (see Figure 1–5 in particular). The upper left-hand corner of the product–process matrix corresponds to low-volume production and little product standardization. Such a product structure is usually characterized by a job-shop-type environment. In a job shop there are a wide variety of jobs with different flow patterns associated with each job. A commercial printer is a typical example of a jumbled flow shop such as this. As volume increases, the number of products declines and flow patterns become more standardized. For discrete parts manufacture, an auto assembly plant is a good example of this case. A different approach for designing production facilities would be appropriate in such a setting.

Fixed Position Layouts

Some products are too big to be moved, so the product remains fixed and the layout is based on the product size and shape. Examples of products requiring *fixed position* layouts are large airplanes, ships, and rockets. For such projects, once the basic frame is built, the various required functions would be located in fixed positions around the product. A project layout is similar in concept to the fixed position layout. This would be appropriate for large construction jobs such as commercial buildings or bridges. The required equipment is moved to the site and removed when the project is completed. A typical fixed position layout is shown in Figure 11–6.

Product Layouts

In a *product layout* (or product flow layout) machines are organized to conform to the sequence of operations required to produce the product. The product layout is typical

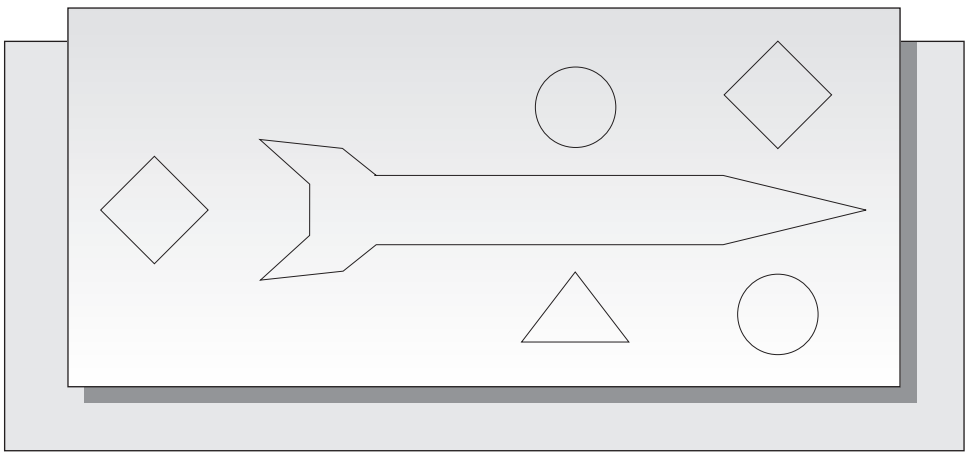


FIGURE 11–6
Fixed position layout

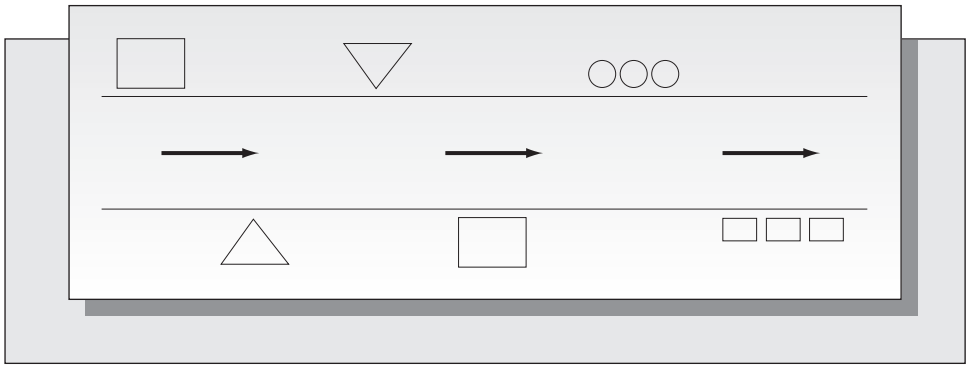


FIGURE 11–7
Product layout

of high-volume standardized production (the lower right-hand corner in Figure 1–5 in Chapter 1). An assembly line (or transfer line) is a product layout, because assembly facilities are organized according to the sequence of steps required to produce the item. Product layouts are desirable for flow-type mass production and provide the fastest cycle times in this environment. Transfer lines are expensive and inflexible, however, and become cumbersome when changes in the product flow are required. Furthermore, a transfer line can experience significant idle time. If one part of the line stops, the entire line may have to remain idle until the problem is corrected. Figure 11–7 shows a typical product layout.

Process Layouts

Process layouts are the most common for small- to medium-volume manufacturers. A *process layout* groups similar machines having similar functions. A typical process layout would group lathes in one area, drills in one area, and so on. Process layouts are most effective when there is a wide variation in the product mix. Each product has a different routing sequence associated with it. In such an environment it would be difficult to organize the machines to conform with the production flow because flow patterns are highly variable. Process layouts have the advantage of minimizing machine idle time. Parts from multiple products or jobs queue up at each work center to facilitate high utilization of critical resources. Also, when design changes are common, parts routings will change frequently. In such an environment, a process layout affords minimal disruption. Figure 11–8 shows a typical process layout. The arrows correspond to part routings.

Layouts Based on Group Technology

With increased emphasis on automated factories and flexible manufacturing systems, *group technology layouts* have received considerable attention in recent years. To implement a group technology layout, parts must be identified and grouped based on similarities in manufacturing function or design. Parts are organized into part families. Presumably, each family requires similar processing, which suggests a layout based on the needs of each family. In most cases, machines are grouped into machine cells where each cell corresponds to a particular part family or a small group of part families (see Figure 11–9).

The group technology concept seems best suited for large firms that produce a wide variety of parts in moderate to high volumes. A typical firm that would consider this approach might have as many as 10,000 different part numbers, which might be grouped

FIGURE 11–8
Process layout

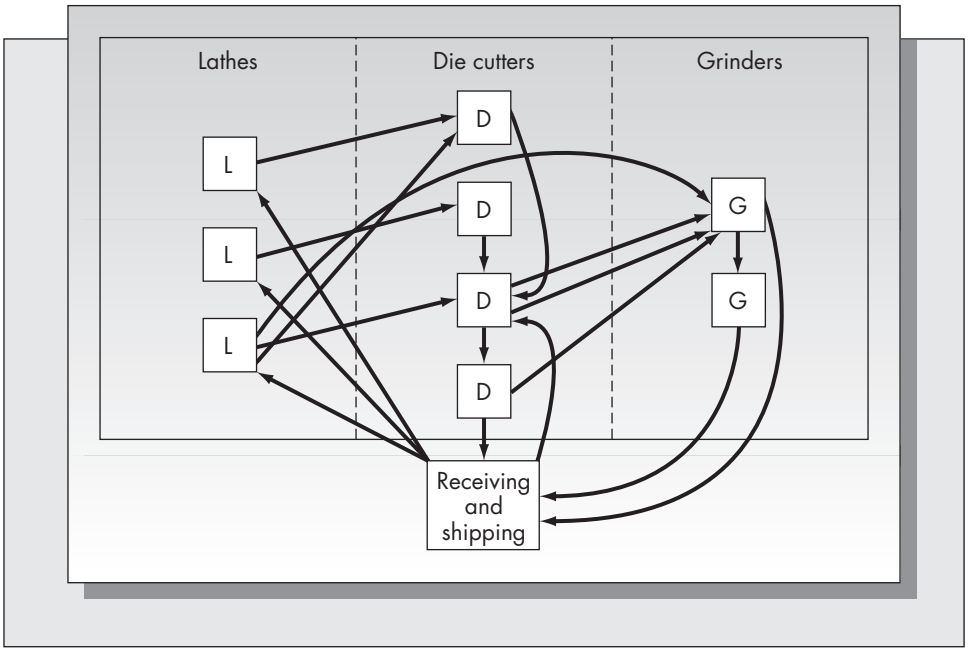
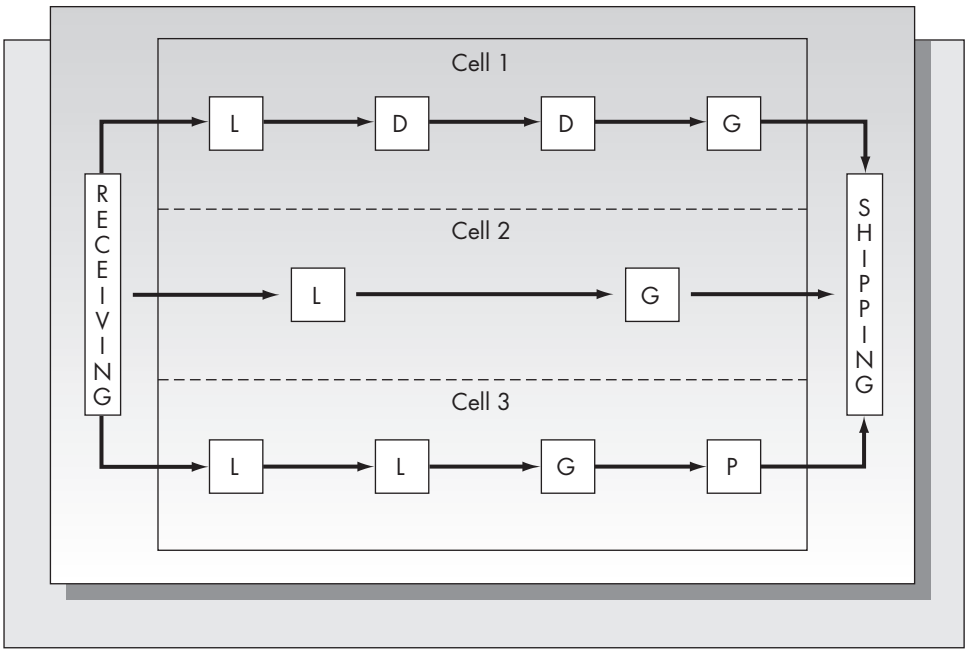


FIGURE 11–9
Group technology layout



into 50 or so part families. Some of the advantages of using the group technology concept are

1. *Reduced work-in-process inventories.* Each manufacturing cell operates as an independent unit, allowing much tighter controls on the flow of production. Large work-in-process inventories are not needed to maintain low cycle times. A side benefit of this is reduced queues of parts and the confusion that results.

2. *Reduced setup times.* Because manufacturing cells are organized according to part types, there should not be significant variation in the machine settings required when switching from one part to another. This allows the cells to operate much more efficiently.
3. *Reduced materials handling costs.* For a firm producing 10,000 parts, a process layout would require a dizzying variety of part routings. If volumes are large, process centers would necessarily have to be separated by large distances, thus requiring substantial materials handling costs. A layout based on group technology would overcome this problem.
4. *Better scheduling.* By isolating part groupings, it is much easier to keep track of the production flow within each cell. Reduction in cycle times and work-in-process queues results in more reliable due-date schedules.

There are several disadvantages of the group technology approach. One is that it can be difficult to determine suitable part families. Parts may be grouped by size and shape or by manufacturing process requirements. The first approach is easier but not as effective for developing layouts. How parts are grouped is a function, to a large degree, of the coding system used to classify these parts. (Groover and Zimmers, 1984, discuss several parts coding systems and how they relate to the group technology concept.) Grouping part families according to the manufacturing flow requires a careful production flow analysis (Burbidge, 1975). This method is probably not feasible for a firm with a large number of parts, however.

Group technology layouts may require duplication of some machines. In order for a manufacturing cell to be self-contained, the cell must have all the machines necessary for the product being produced. Duplicating machines could be expensive and could result in greater overall idle time.

Under what circumstances would a group technology layout be preferred to a pure process or product layout? A simulation study by Sassani (1990) provides some answers. He constructed a simulation of five manufacturing cells. Initially, when the products for each cell were well defined and the cells isolated, the system ran smoothly. However, as the product mix, product design, and demand patterns started to change, the simulation showed that the efficiency of the layout deteriorated.

The vast majority of existing layouts are either process or product type. Firms producing a wide variety of parts may choose several layouts for different product lines, or may choose some hybrid approach. Product variation and annual volume are the primary determining factors for making the appropriate choice. The group technology approach is relatively new. It is slowly gaining acceptance as firms wrestle with the difficult problems of determining appropriate part groupings and cell designs. As automated factories become more widespread, however, the group technology concept could play a greater role in plant layout and process design.

Problems for Sections 11.1–11.3

1. For each of the eight objectives of a layout study listed in Section 11.1, give examples of situations in which that objective would be important and examples in which it would be unimportant.
2. A manufacturing facility consists of five departments: A, B, C, D, and E. At the current time, the facility is laid out as in Figure 11–10. Develop a from-to distance chart for this layout. Estimate the distance separating departments based on the flow pattern of the materials handling system in the figure. Assume that all departments are located at their centers as marked in Figure 11–10.

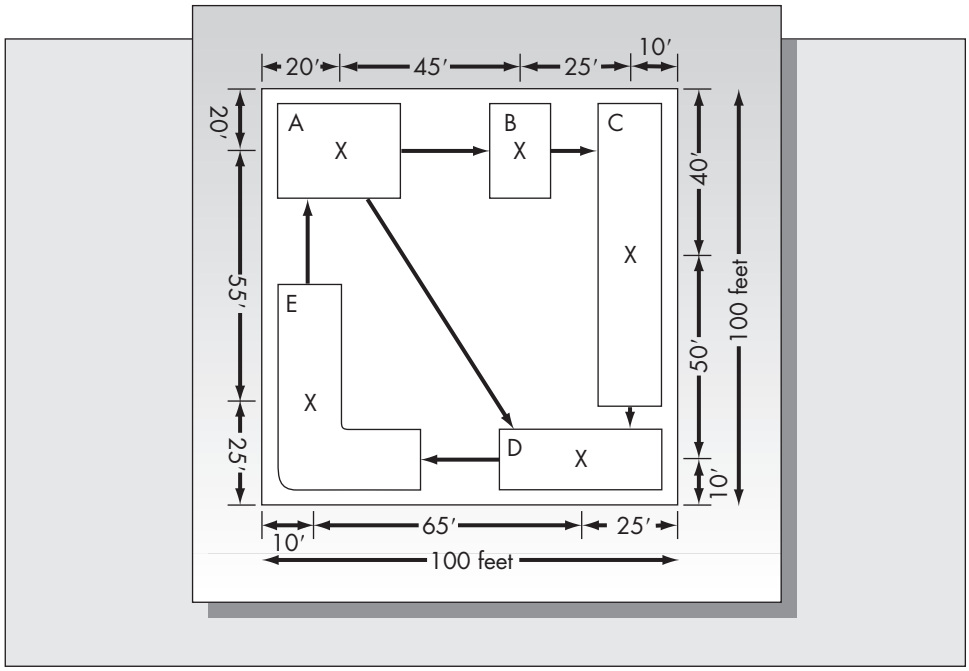


FIGURE 11–10
Layout of manufacturing facility (for Problem 2)

3. Four products are produced in the facility described in Problem 2. The routing for these products and their forecasted production rates are

Product	Routing	Forecasted Production (units/week)
1	A–B–C–D–E	200
2	A–D–E	900
3	A–B–C–E	400
4	A–C–D–E	650

Suppose that all four products are produced in batches of size 50.

a. Convert this information into a from-to chart similar to Figure 11–4 but giving the number of materials handling trips per week between departments.

b. Suppose that the cost of transporting goods 1 foot is estimated to be $2. Using the from-to chart you determined in part (*a*), obtain a from-to chart giving materials handling cost per week between departments.

c. Develop an activity relationship chart for this facility based on the materials handling cost from-to chart you obtained in part (*b*). Use only the rankings A, E, I, O, U, with A being assigned to the highest cost and U the least.

d. Based on the results of parts (*b*) and (*c*), would you recommend a different layout for this facility?

4. Consider the example of the machine shop with the from-to charts given in Figures 11–3 through 11–5. Based on the results of Figure 11–5, in what ways might the current layout be improved?

5. Suggest a layout for the Meat Me fast-food restaurant, with the relationship chart shown in Figure 11–2. Assume that the facility is 50 feet square and half the restaurant will be for customer seating.
6. Describe the differences among product, process, and group technology layouts. Describe the circumstances in which each type of layout is appropriate.

11.4 A PROTOTYPE LAYOUT PROBLEM AND THE ASSIGNMENT MODEL

Several analytical techniques have been developed for assisting with layout problems. However, most real problems are too complex to be solved by these methods without a computer. (We will discuss the role of the computer in facilities layout later in Section 11.6.) In some cases, however, layout problems can be formulated as *assignment* problems, as described in this section. In particular, the assignment model is appropriate when only a discrete set of alternative locations is being considered and there are no interactions among departments. Although moderately sized assignment problems can be solved by hand, *extremely* large problems can be solved only with the aid of computers.

Example *11.3*

Because of an increase in the volume of sales, Sam Melder, the owner of Melder, Inc., a small manufacturing firm located in Muncie, Indiana, has decided to expand production capacity. A new wing has been added to the plant that will house four machines: (1) a punch press, (2) a grinder, (3) a lathe, and (4) a welding machine. There are only four possible locations for these machines, say A, B, C, and D. However, the welding machine, which is the largest machine, will not fit in location B.

The plant foreman has made estimates of the impact in terms of materials handling costs of locating each of the machines in each of the possible locations. These costs, expressed in terms of dollars per hour, are represented in the following table.

		Location			
		A	**B**	**C**	**D**
Machines	**1**	94	13	62	71
	2	62	19	84	96
	3	75	88	18	80
	4	11	*M*	81	21

The entry *M* stands for a very large cost. It is used to indicate that machine 4 is not permitted in location B. As the objective is to assign the four machines to locations in order to minimize the total materials handling cost, an optimal solution will never assign machine 4 to location B.

Minimum cost assignments can rarely be found by inspection. For example, one reasonable approach might be to assign the machines to the lowest cost locations in sequence. The lowest cost in the matrix is 11, so machine 4 would be located in location A. Eliminating the last row and the first column, we see that the next remaining lowest cost is 13, so machine 1 would be assigned to location B. Now also eliminating the first row and the second column, the smallest remaining cost is 18, so machine 3 would be assigned to location C. Finally, machine 2 must be assigned to location D. The total cost of this solution is $11 + 13 + 18 + 96 = \$138$. As we will see, this solution is suboptimal. (The optimal cost is \$114. Can you find the optimal solution?)

A simple approach such as this will rarely result in an optimal solution. A solution algorithm will be presented from which one can obtain an optimal solution to the assignment problem by hand for moderately sized problems.

The Assignment Algorithm

The solution algorithm presented in this section is based on the following observation:

> **Result:** If a constant is added to, or subtracted from, all entries in a row or column of an assignment cost matrix, the optimal assignment is unchanged.

In applying this result, the objective is to continue subtracting constants from rows and columns in the cost matrix until a zero cost assignment can be made. The zero cost assignment for the modified matrix is optimal for the original problem. This leads to the following algorithm for solving assignment problems:

Solution Procedure for Assignment Problems

1. Locate the smallest number in row 1 and subtract it from all the entries in row 1. Repeat for all rows in the cost matrix.
2. Locate the smallest number in column 1 and subtract it from all the entries in column 1. Repeat for all columns in the cost matrix.
3. At this point each row and each column will have at least one zero. If it is possible to make a zero cost assignment, then do it. That will be the optimal solution. If not, go to step 4.
4. Determine the maximum number of zero cost assignments. This will equal the smallest number of lines required to cover all zeros. The lines are found by inspection and are not necessarily unique. The important point is that the number of lines drawn not exceed the maximum number of zero cost assignments.
5. Choose the smallest uncovered number and do the following:
   a. Subtract it from all other uncovered numbers.
   b. Add it to the numbers where the lines cross.
   c. Return to step 3.

The process is continued until one can make a zero cost assignment. Notice that step 5 is merely an application of the result in the following way: If the smallest uncovered number is subtracted from every element in the matrix and then added to every covered element, it will be subtracted once and added twice where lines cross.

Example *11.3 (continued)* Let us return to Example 11.3. The original cost matrix is

		Location			
		A	**B**	**C**	**D**
Machine	**1**	94	13	62	71
	2	62	19	84	96
	3	75	88	18	80
	4	11	*M*	81	21

Step 1. Subtracting the smallest number from each row gives

81	0	49	58
43	0	65	77
57	70	0	62
0	*M*	70	10

Because *M* is very large relative to the other costs, subtracting 11 from *M* still leaves a very large number, which we again denote as *M* for convenience. At this point at least one zero appears in each row, and in each column except the last.

Step 2. Subtracting 10 from every number in the final column gives

81	0	49	48
43	0	65	67
57	70	0	52
0	*M*	70	0

At this point we have at least one zero in every row and every column (step 3). However, that does not necessarily mean that a zero cost assignment is possible. In fact, it is possible to make at most three zero cost assignments at this stage, as shown in step 4.

Step 4.

81	0	49	48
43	0	65	67
57	70	0	52
0	*M*	70	0

The three assignments shown in step 4 are 1–B, 3–C, and 4–A. There are other ways of assigning three locations at zero cost as well. It does not matter which we choose at this stage, only that we know three are possible. The next step is to find three lines that cover all the zeros. These are shown here.

81	0	49	48
43	0	65	67
~~57~~	~~70~~	~~0~~	~~52~~
~~0~~	~~*M*~~	~~70~~	~~0~~

Again, the choice of lines is not unique. However, it is important that no more than three lines be used. Finding three lines to cover all zeros is done by trial and error.

Step 5. The smallest uncovered number, 43, is subtracted from all other uncovered numbers and added to the numbers where the lines cross. The resulting matrix is

		Location			
		A	**B**	**C**	**D**
Machine	**1**	38	0	6	5
	2	0	0	22	24
	3	57	113	0	52
	4	0	*M*	70	0

It is now possible to make a zero cost assignment, as shown in the matrix. The optimal assignment is machine 1, the punch press, to location B; machine 2, the grinder, to location A;

machine 3, the lathe, to location C; machine 4, the welder, to location D. The total materials handling cost per hour of the optimal solution is obtained by referring to the original assignment cost matrix. It is 13 + 62 + 18 + 21 = $114 per hour.

The assignment algorithm also can be used when the number of sites is larger than the number of machines. For example, suppose that there were six potential sites for locating the four machines and the original cost matrix was

		Location					
		A	**B**	**C**	**D**	**E**	**F**
Machine	**1**	94	13	62	71	82	25
	2	62	19	84	96	24	29
	3	75	88	18	80	16	78
	4	11	*M*	81	21	45	14

The procedure is to add two dummy machines, 5 and 6, with zero costs. The problem is then solved using the assignment algorithm as if there were six machines and six locations. The locations to which the dummy machines are assigned are the ones that are not used.

Problems for Section 11.4

7. Solve the following assignment problem:

	A	**B**	**C**	**D**
1	21	24	26	23
2	29	27	30	29
3	24	25	34	27
4	28	26	28	25

8. Each of four machines is to be assigned to one of five possible locations. The objective is to assign the machines to locations that will minimize the materials handling cost. The machine–location costs are given in the following matrix. Find the optimal assignment.

	A	**B**	**C**	**D**	**E**
1	26	20	22	21	25
2	35	31	33	40	26
3	15	18	23	16	25
4	31	34	33	30	*M*

9. University of the Atlantic is moving its business school into a new building, which has been designed to house six academic departments. The average time required for a student to get to and from classes in the building depends upon the location of the department in which he or she is taking the class. Based on the distribution of

class loads, the dean has estimated the following mean student trip times in minutes, given the departmental locations.

		Location					
		A	**B**	**C**	**D**	**E**	**F**
Department	**1**	13	18	12	20	13	13
	2	18	17	12	19	17	16
	3	16	14	12	17	15	19
	4	18	14	12	13	15	12
	5	19	20	16	19	20	19
	6	22	23	17	24	28	25

Find the optimal assignment of departments to locations to minimize mean student trip time in and out of the building.

*11.5 MORE ADVANCED MATHEMATICAL PROGRAMMING FORMULATIONS

The assignment model described in Section 11.4 can be useful for determining optimal layouts for a limited number of real problems. The primary limitation of the simple assignment model is that, in most cases, the number of materials handling trips and the associated materials handling cost are assumed to be independent of the location of the other facilities. In Example 11.3, the cost of assigning the punch press to location A was assumed to be \$94 per hour. However, in most cases this cost would depend upon the location of the other machines as well.

A formulation of the problem that takes this feature into account is considerably more complex. In order to avoid confusion, we will assume that the problem is to assign machines to locations. The problem could be to assign other types of subfacilities to locations, of course, but we will retain this terminology for convenience. Define the following quantities:

n = Number of machines;

c_{ij} = Cost per time period of assigning machine i to location j. This cost could be a one-time relocation cost that is converted to an annual equivalent;

d_{jr} = Cost of making a single materials handling trip from location j to location r;

f_{ik} = Mean number of trips per time period from machine i to machine k;

S_i = The set of locations to which machine i could feasibly be assigned;

$$a_{ijkr} = \begin{cases} f_{ik}d_{jr} & \text{if } i \neq k \text{ or } j \neq r, \\ c_{ij} & \text{if } i = k \text{ and } j = r; \end{cases}$$

$$x_{ij} = \begin{cases} 1 & \text{if machine } i \text{ is assigned to location } j, \\ 0 & \text{otherwise.} \end{cases}$$

Interpret a_{ijkr} as the materials handling cost per unit time when machine i is assigned to location j and machine k is assigned to location r. This cost is incurred only if both x_{ij} and x_{kr} are equal to 1. Hence, it follows that the total cost of assigning machines to locations is given by

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{r=1}^{n} a_{ijkr}x_{ij}x_{kr}. \tag{1}$$

As all indices are summed from 1 to n, each assignment will be counted twice; hence the need to multiply by $\frac{1}{2}$. Constraints are included to ensure that each machine gets assigned to exactly one location and each location is assigned exactly one machine. These are

$$\sum_{i=1}^{n} x_{ij} = 1, \qquad j = 1, \ldots, n; \tag{2}$$

$$\sum_{j=1}^{n} x_{ij} = 1, \qquad i = 1, \ldots, n; \tag{3}$$

$$x_{ij} = 0 \text{ or } 1, \qquad i = 1, \ldots, n \text{ and } j = 1, \ldots, n; \tag{4}$$

$$x_{ij} = 0, \qquad i = 1, \ldots, n \text{ and } j \notin S(i). \tag{5}$$

The mathematical programming formulation is to minimize the total materials handling cost of the assignment, (1), subject to the constraints (2), (3), (4), and (5). This formulation is known as the quadratic assignment problem. In general, such problems are extremely difficult to solve. One should consider using such a method only for moderately sized problems.

Problem for Section 11.5

10. Consider the following problem with two locations and three machines. Suppose that the costs of transporting a unit load from location j to location r are given in the following table:

		To location	
		A	**B**
From location	**A**		6
	B	9	

The average numbers of trips required from machine i to machine k per hour are

		To machine		
		1	**2**	**3**
From machine	**1**	0	3	1
	2	0	0	3
	3	0	4	0

Relocation costs are ignored. Write out the complete quadratic assignment formulation for this location problem.

11.6 COMPUTERIZED LAYOUT TECHNIQUES

The quadratic assignment formulation given in Section 11.5 has several shortcomings. First, one must specify the costs of all materials handling trips and the expected number of trips from every department to every other department. When many departments are involved, this information could be difficult to obtain. Furthermore, an efficient solution technique for solving large quadratic assignment problems has yet to be discovered.

For these reasons there has been considerable interest in computer-aided methods. These methods are heuristics; they do not guarantee an optimal solution but generally yield efficient solutions. They can be used for solving problems that are too large to be solved analytically. The methods that we discuss in this section are CRAFT, COFAD, ALDEP, CORELAP, and PLANET.

For each of these methods the objective is to minimize the cost of materials handling. However, the solutions generated by these computer programs must be considered in the context of the problem. Issues such as plant safety, noise, and aesthetics are ignored. The layout obtained from a computer program may have to be modified in order to take these factors into account.

Computer programs for determining layouts generally fall into two classes: (1) improvement routines and (2) construction routines. An improvement routine takes an existing layout and considers the effect of interchanging the location of two or more facilities. A construction routine constructs the layout from scratch from flow data and the information in the activity relationship chart. CRAFT and COFAD are both improvement routines, and PLANET, CORELAP, and ALDEP are construction routines. The improvement routines have the disadvantage of requiring specification of an initial layout. However, improvement routines generally result in more usable layouts. Construction routines often give layouts with oddly shaped departments.

CRAFT

CRAFT (*computerized relative allocation of facilities technique*) was one of the first computer-aided layout routines developed. As noted earlier, CRAFT is an improvement routine and requires the user to specify an initial layout. The objective used in CRAFT is to minimize the total transportation cost of a layout, where transportation cost is defined as the product of the cost to move a unit load from department i to department j and the distance between departments i and j. To be more specific, define

n = Number of departments,

v_{ij} = Number of loads moved from department i to department j in a given time,

u_{ij} = Cost to move a unit load a unit distance from department i to department j,

d_{ij} = Distance separating departments i and j.

It follows that $y_{ij} = u_{ij}v_{ij}$ is the cost to move total product flow during the specified time interval a unit distance from i to j, and that $y_{ij}d_{ij}$ is the cost of the product flow from i to j. The total cost of the product flow between all pairs of departments is

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij}d_{ij}.$$

Note that the input information v_{ij} and u_{ij} may be represented as from-to charts, as shown in Figures 11–3 and 11–4.

Two implicit assumptions made in CRAFT are

1. Move costs are independent of the equipment utilization.
2. Move costs are a linear function of the length of the move.

When these assumptions cannot be justified, CRAFT may be used to minimize the product of flows and distances only by assigning the unit cost u_{ij} a value of 1. The entries d_{ij} are computed by the program from a specification of an initial layout. Based on an initial layout, CRAFT considers exchanging the position of adjacent departments and computes the materials handling cost of the resulting exchange. The program chooses the pairwise interchange that results in the greatest cost reduction.

Departments are assumed to be either rectangularly shaped or composed of rectangular pieces. Furthermore, departments are assumed to be located at their *centroids*. The centroid is another term for the coordinates of the center of gravity or balance point. A discussion of centroids and how they are computed for objects in the plane appears in Appendix 11–A. The accuracy of assuming that a department is located at its centroid depends upon the shape of the department. The assumption is most accurate when the shape of the department is square or rectangular, but is less accurate for oddly shaped departments.

To better understand how CRAFT determines layouts, consider the following example.

Example *11.4*

A local manufacturing firm has recently completed construction of a new wing of an existing building to house four departments: A, B, C, and D. The wing is 100 feet by 50 feet. The plant manager has chosen an initial layout of the four departments. This layout appears in Figure 11–11. We have marked the centroid locations of the departments with a dot. From the figure we see that department A requires 1,800 square feet, B requires 1,200 square feet, C requires 800 square feet, and D requires 1,200 square feet.

One of the inputs required by CRAFT is the flow data, that is, the number of materials handling trips per unit time from every department to every other department. These data are given in the from-to chart appearing in Figure 11–12*a*. The distance between departments is assumed to be the rectilinear distance between centroid locations. From Figure 11–11, we see that the centroid locations of the initial layout are

$$(x_A, y_A) = (30, 35), \qquad (x_C, y_C) = (20, 10),$$
$$(x_B, y_B) = (80, 35), \qquad (x_D, y_D) = (70, 10).$$

The rectilinear distance between A and B, for example, is given by the formula

$$|x_A - x_B| + |y_A - y_B| = |30 - 80| + |35 - 35| = 50.$$

One computes distances between other pairs of departments in a similar way. The calculations are summarized in the from-to chart in Figure 11–12*b*. Notice that we have assumed that the distance from department *i* to department *j* is the same as the distance from department *j* to department *i*. As we noted earlier, this may not necessarily be true if material is transported in one direction only.

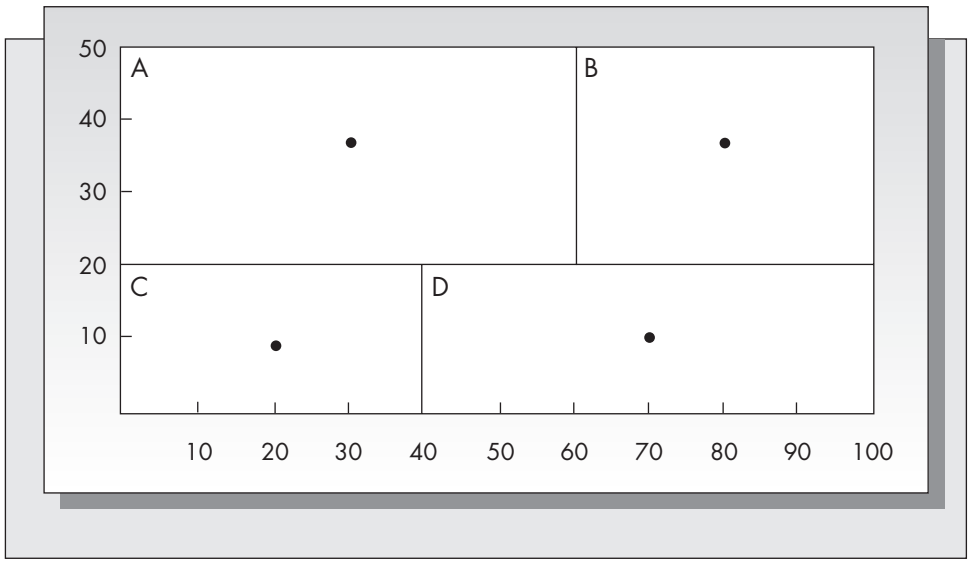
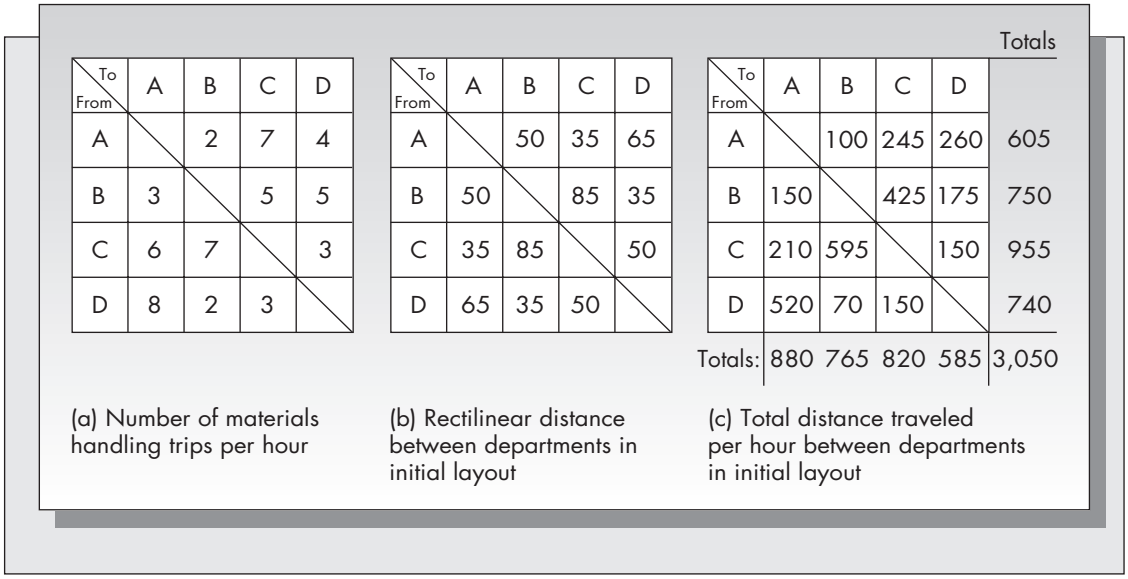


FIGURE 11–11
Initial layout for CRAFT example (Example 11.4)

FIGURE 11–12
From-to charts for initial layout



To / From	A	B	C	D
A		2	7	4
B	3		5	5
C	6	7		3
D	8	2	3	

(a) Number of materials handling trips per hour

To / From	A	B	C	D
A		50	35	65
B	50		85	35
C	35	85		50
D	65	35	50	

(b) Rectilinear distance between departments in initial layout

To / From	A	B	C	D	Totals
A		100	245	260	605
B	150		425	175	750
C	210	595		150	955
D	520	70	150		740
Totals:	880	765	820	585	3,050

(c) Total distance traveled per hour between departments in initial layout

Normally, a third from-to chart also would be required. This would give the cost of transporting a unit load a unit distance from department *i* to department *j*. As noted, CRAFT may be used to minimize the product of flows and distances by assigning a value of 1 to these transport costs. We will assume this to be the case in our example. Hence, one obtains the hourly cost of transporting materials to and from each of the departments by simply multiplying the entries in the from-to chart of Figure 11–12*a* by the entries in the from-to chart of Figure 11–12*b*. These calculations are summarized in Figure 11–12c. The total distance traveled per hour for the initial layout is 3,050 feet.

CRAFT now considers the effect of pairwise interchanges of two departments that either have adjacent borders or have the same area. (CRAFT can also consider the effect of exchanging the locations of three departments. We will not consider this option in our example.) The result of the exchange is determined by exchanging the location of the centroids. This is only an approximation, however, since exchanging the locations of two departments does *not* necessarily mean that the location of their centroids will be exchanged.

If we were to exchange the locations of A and B, for example, we would assume that $(x_A, y_A) = (80, 35)$ and $(x_B, y_B) = (30, 35)$, which would result in the new from-to distance chart appearing in Figure 11–13*a*. Multiplying the original flow data in Figure 11–12*a* by this distance chart gives a new cost chart, which appears in Figure 11–13*b*. Interchanging the centroids of A and B results in the predicted cost reduction from 3,050 to 2,950, or about 3 percent. (The actual cost of interchanging the locations of A and B would be slightly different because the centroids are not exactly exchanged.)

CRAFT considers all pairwise interchanges of adjacent departments or departments with identical areas and picks the one that results in the largest decrease in the predicted cost. We will not present the details of the calculations but merely summarize the results. Exchanging the centroids of A and C results in a total predicted cost of 2,715, and exchanging the centroids of A and D results in a total predicted cost of 3,185. Two other exchanges must be considered as well: B and D, and C and D. Notice that exchanging B and C is not considered because they do not have equal areas and do not share a common border. Exchanging the centroids of B and D results in a total predicted cost of 2,735, and exchanging the centroids of C and D in a total predicted cost of 2,830.

FIGURE 11–13
New distance and cost from-to charts after exchanging centroids for A and B

To / From	A	B	C	D
A		50	85	35
B	50		35	65
C	85	35		50
D	35	65	50	

(a)

To / From	A	B	C	D	Totals
A		100	595	140	835
B	150		175	325	650
C	510	245		150	905
D	280	130	150		560
Totals	940	475	920	615	2,950

(b)

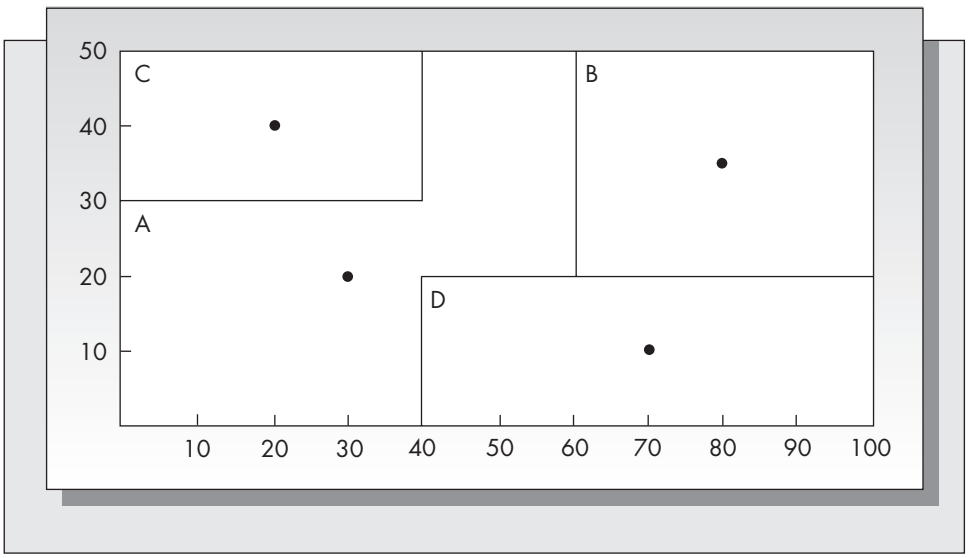
The maximum predicted cost reduction is achieved by exchanging A and C. The new layout with A and C exchanged appears in Figure 11–14. Because C has the smaller area, it is placed in the upper left-hand corner of the space formerly occupied by A, so that the remaining space allows A to be contiguous. Notice that A is no longer rectangular. The actual cost of the new layout is not necessarily equal to the predicted value of 2,715. The centroid of A is computed using the method outlined in Appendix 11–A. It is determined by first finding the moments M_x and M_y given by

$$M_x = (40^2 - 0)(30 - 0)/2 + (60^2 - 40^2)(50 - 20)/2 = 54{,}000,$$
$$M_y = (30^2 - 0)(40 - 0)/2 + (50^2 - 20^2)(60 - 40)/2 = 39{,}000,$$

and dividing by the area of A to obtain

$$x_A = 54{,}000/1{,}800 = 30,$$
$$y_A = 39{,}000/1{,}800 = 21.66667.$$

FIGURE 11–14
New layout with A and C interchanged



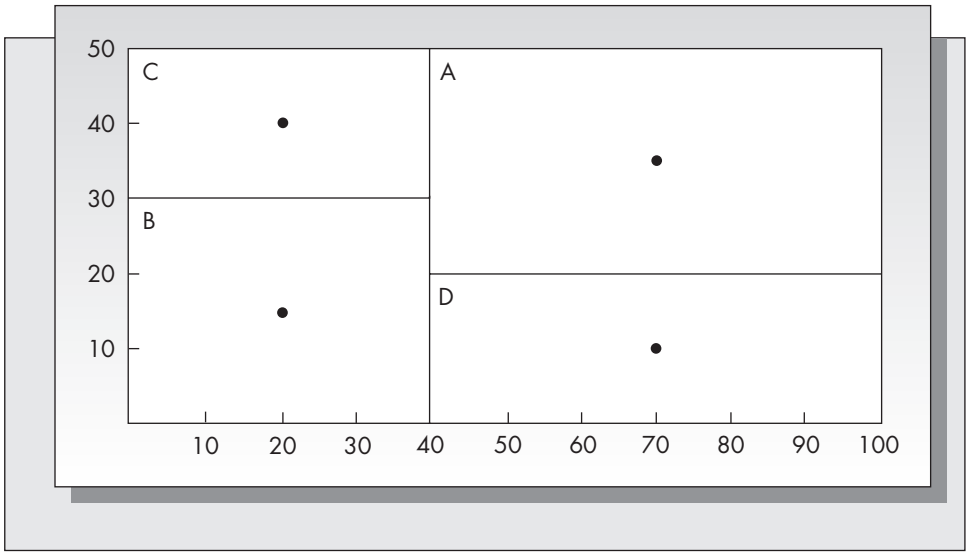


FIGURE 11–15
Second iteration, obtained from exchanging the locations of A and B

The centroid of C is located at the center of symmetry, which is

$$(x_C, y_C) = (20, 40).$$

The centroids for B and D are unchanged. The cost of the new layout is actually 2,810, which is somewhat more than the 2,715 predicted at the last step, but is still less than the original cost. The process is continued until predicted cost reductions are no longer possible. At this point only three exchanges need to be considered: A and B, A and D, and B and D. Obviously, we need not consider exchanging A and C, and C and D may not be exchanged because they do not share a common border. The predicted cost resulting from exchanging the centroids of A and B is 2,763.33; of A and D, 3,641.33; and of B and D, 2,982. Clearly, the greatest predicted reduction is now achieved by exchanging the locations of A and B.

The new layout is pictured in Figure 11–15. The centroids for the respective departments are now

$$(x_A, y_A) = (70, 35), \qquad (x_C, y_C) = (20, 40),$$
$$(x_B, y_B) = (20, 15), \qquad (x_D, y_D) = (70, 10).$$

The actual cost of this layout is 2,530, which is considerably *less* than that predicted in the previous step. The process continues until no further reductions in the predicted costs can be achieved. Because A and B were exchanged at the previous step, that exchange need not be considered again. Exchanging A and C results in a predicted cost of 3,175; exchanging A and D, a predicted cost of 2,753; and exchanging B and D, a predicted cost of 3,325. (We do not consider exchanging C and D at this stage because they are not adjacent and do not have equal areas.) Since none of the predicted costs is less than the current cost, we terminate calculations. The layout recommended by CRAFT, pictured in Figure 11–15, required two iterations and resulted in a reduction of total distance traveled from 3,050 feet to 2,530 feet, or about 17 percent.

COFAD

As noted, CRAFT is an improvement routine. It requires the user to specify an initial layout and proceeds to improve the layout by considering pairwise interchanges of adjacent departments. An improvement routine similar to CRAFT is COFAD (for *computerized facilities design*). This chapter will not present the details of COFAD, but briefly reviews the improvements it offers over CRAFT.

COFAD is a modification of CRAFT that incorporates the choice of a materials handling system as part of the decision process. For a given layout, COFAD calculates the total move cost of the layout for a variety of materials handling alternatives and chooses the one with the minimum cost. Improvements are made by considering the equipment utilization of alternatives and exchanging assignments of poorly utilized equipment with equipment with better utilization. Once the best choice of equipment for a given layout is determined, the program considers improving the layout in much the same way as CRAFT. The program terminates when no additional improvements are possible. Once the program claims to have reached a steady state, the problem may be re-solved using from-to charts obtained from perturbing the original data from 10 to 50 percent. The purpose of re-solving the problem with the new data is to provide further assurance that the optimal solution has been reached, and to test the sensitivity of the solution to the flow data.

ALDEP

Both CRAFT and COFAD are improvement routines. That is, they start out with an initial layout and consider improvements by interchanging locations of departments. The other class of computerized layout techniques includes programs that develop the layouts essentially from scratch; they do not require an initial layout. Experience has shown that construction programs often tend to result in oddly shaped layouts, and for that reason they have not received as much attention in the literature as CRAFT.

ALDEP (for *automated layout design program*), is a construction routine rather than an improvement routine. That means that layouts are determined from scratch and the program does not require the user to specify an initial layout. ALDEP makes use of the closeness ratings that appear in the activity relationship chart. Figure 11–2 is an example of a rel chart. ALDEP first selects a department at random and places it in the upper left-hand corner of the layout. The next step is to scan the rel chart and place a department with a high closeness rating (A or E) in the layout adjacent to the first department. Successive departments are placed in the existing area in a top-down fashion, following a "sweep" pattern. The process is continued until all departments have been placed. At that point, a score for the layout is computed. The score is based on a numerical scale attached to the closeness ratings. ALDEP uses the following closeness rating values:

$$
\begin{aligned}
A &= 4^3 = 64,\\
E &= 4^2 = 16,\\
I &= 4^1 = 4,\\
O &= 4^0 = 1,\\
U &= 0,\\
X &= -4^5 = -1{,}024.
\end{aligned}
$$

The entire process is repeated several times, and the layout with the largest score chosen. Because ALDEP tries to achieve a high closeness rating score, it often recommends layouts that have departments with very unusual shapes. The use of the sweep pattern helps to avoid this problem, but means that the resulting layout always appears as a set of adjacent strips. This type of layout may not be appropriate or desirable for some applications. The user specifies the sweep width to be used; the choice of the sweep width can have a significant effect upon the configuration of the final layout. For example, suppose that department A is 14 squares, B is 8 squares, the sweep width is 2 squares, and the

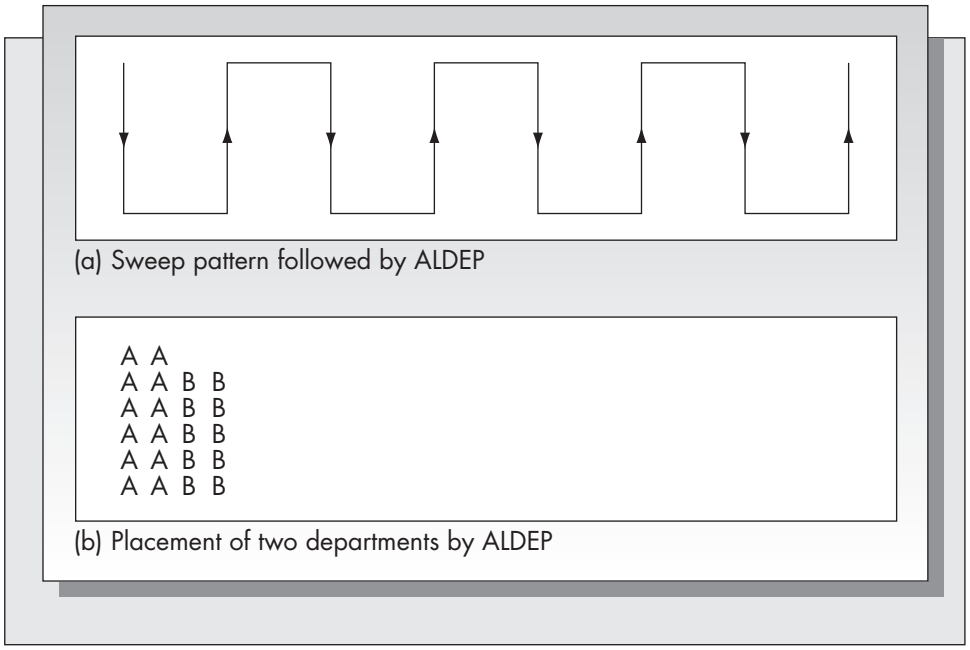


FIGURE 11–16
Method of placing departments used by ALDEP

facility width is 6 squares. The layout of these two departments given by ALDEP is shown in Figure 11–16*b*, assuming that the departments are placed in the order A, B.

CORELAP

Like ALDEP, CORELAP (for *computerized relationship layout planning*) is a construction routine that places departments within the layout using closeness rating codes. The major difference between CORELAP and ALDEP is that CORELAP does not randomly select the first department to be placed. Rather, a total closeness rating (TCR) is computed for each department. The TCR is based on the numerical values $A = 6, E = 5, I = 4, O = 3, U = 2, X = 1$. Each department is compared to every other department to obtain a TCR.

For example, consider the Meat Me fast-food restaurant described in Example 11.1, with the rel chart shown in Figure 11–2. The closeness rating codes for the cooking-burgers department are *X*, *I*, *U*, *U*, and *U*, giving a TCR for this department of 11. The remaining values of the TCR are

Cooking fries:	*X*, *I*, *U*, *U*, *U*—11,
Packing and storing:	*I*, *I*, *O*, *E*, *E*—21,
Drink dispensers:	*U*, *U*, *O*, *A*, *A*—19,
Counter service:	*U*, *U*, *E*, *A*, *A*—21,
Drive-up service:	*O*, *A*, *E*, *U*, *U*—18.

CORELAP now selects the department with the highest TCR and places that department in the center of the existing facility. When there is a tie (as in this case between the packing and storing department and the counter service department), the department with the largest area is placed first. Once the initial department is selected, the rel chart is scanned to see what departments have the highest closeness

rating compared with the department just placed. Suppose that counter service is placed first. Then, scanning Figure 11–2, we see that only the drink dispensers department has a closeness rating code of *A* relative to the counter service department. Hence, drink dispensers would be placed in the layout next. Each time that a department is chosen to be placed in the layout, alternative placements are considered and placing ratings based on user-specified numerical values for the rel codes are computed. The new department is placed to maximize the placing rating. (Detailed examples of this procedure can be found in Francis and White, 1974, and Tompkins and White, 1984.)

PLANET

PLANET (for *plant layout analysis and evaluation technique*) has essentially the same required inputs as COFAD, and unlike both CORELAP and ALDEP does not use information contained in the rel chart to generate a layout. However, unlike CRAFT, PLANET is not an improvement routine. PLANET converts input data into a flow-between chart that gives the cost of sending a unit of flow between each pair of departments. In addition to the flow-between chart entries, the program also requires the user to input a priority rating for each department. The priority rating is a number from 1 to 9, with 1 representing the highest priority. The program selects departments to enter the layout in a sequential fashion, based first on the priority rating and second on the entries in the flow-between chart. Because PLANET does not restrict the final layout to conform to the shape of the building or allow fixing the location of certain departments, it suffers from the problems of other construction routines. Layouts obtained from PLANET, like those from ALDEP and CORELAP, may result in departments having unrealistic shapes. For that reason, such methods are better for providing the planner with some alternative ideas and initial layouts than for giving a final solution.

Computerized Methods versus Human Planners

An interesting debate concerning the effectiveness of computerized layout techniques versus expert human judgment appeared in the management science literature. The debate was sparked by a paper by Scriabin and Vergin (1975). In their study, the authors compared the layouts produced by three computerized layout routines, including CRAFT, and the layouts obtained without the use of these specific programs by 74 human subjects who were trained in manual layout techniques. The authors showed that the best solution obtained by groups of 20 of the 74 human subjects was better than the best solution obtained by the three computer routines. The largest difference in total cost occurred when the number of departments was the largest (20). Percentage differences ranged from 0 to 6.7. The authors concluded that humans will generally outperform computer layout routines because "in problems of a larger size the ability of man to recognize and visualize complex patterns gives him an edge over the essentially mechanical procedures followed by the computer programs."

Scriabin and Vergin's conclusions came under fire by a number of researchers. Buffa (1976) noted that the issue of flow dominance was not treated properly. Flow dominance refers to the tendency of materials to flow in the same sequence in the layout. Complete dominance occurs when a factory manufactures only a single product in an identical manner each time. In such a case, the final layout can easily be obtained by visual inspection. The other extreme occurs when products flow randomly through the factory. In that case all layouts are equivalent. (Block, 1977, gives a formal definition of flow dominance.) Buffa claimed that for problems in which flow dominance is over 200 percent, human subjects could be expected to obtain good solutions easily. The

average flow dominance in the problems used by Scriabin and Vergin was apparently over 200 percent.

Another problem with their conclusions, as pointed out by Coleman (1977), was that in a typical industrial setting one does not have 20 professionals solving a layout problem. When comparing the layouts produced by the computer programs with those produced by individual human subjects, Coleman showed that the computer-generated layouts were superior. Finally, Block (1977) performed a separate experiment to test Buffa's contention that when flow dominance was under 200 percent, the computer will outperform the human. Block compared layouts produced by CRAFT with those produced by eight humans (four engineers and four laypeople) for a set of layout problems generated by Nugent, Vollmann, and Ruml (1968), in which the average flow dominance was 115 percent. He found that COFAD obtained better results in every case. The largest differences were observed when the number of departments was larger than 10.

What can we learn from these studies? It would appear that Scriabin and Vergin's conclusions are not justified. Later results showed that the computer-aided methods are, in fact, extremely useful and can result in significantly better layouts than can be obtained by simple visual or graphical methods when the number of departments is large and when material flow patterns are highly variable.

Dynamic Plant Layouts

Thus far, this chapter has considered only static layout problems. That is, we assumed that the costs appearing in the from-to charts are fixed. In some circumstances, this assumption may not be accurate, however. Nicol and Hollier (1983) note that: "Radical layout changes occur frequently and that management should therefore take this into account in their forward planning." A dynamic plant layout is based on the recognition that demands on the system, and hence costs of any given configuration, may change over time. When information on how the environment will change is available, one can incorporate this information into a model by allowing the costs in the from-to charts to change each planning period. This is precisely the scenario considered by Rosenblatt (1986). He showed how dynamic from-to charts would form the basis of a multiperiod planning model and developed a dynamic programming scheme to solve the resulting system of equations. With an example he shows how dynamic layouts can result in cost savings over fixed layouts in a changing environment. Models of this type will gain an even greater importance as firms continue to move toward structures based on agile manufacturing and increased flexibility.

Other Computer Methods

A variety of other analytical methods have been developed in addition to those already discussed. A method called bias sampling, suggested by Nugent, Vollmann, and Ruml (1968), is a straightforward modification of CRAFT that involves a randomization procedure for selecting departments to exchange. The method is considerably slower computationally but often results in better layouts. Considering the enormous strides that have been made in computing technology, the computational issues are less serious today than they were in 1968. SPACECRAFT (see Johnson, 1982) is an extension of CRAFT designed to produce layouts for multistory structures. The two primary modifications required were that (1) constraints had to be incorporated into the computational procedures that allowed certain departments to be located on specific floors and (2) the nonlinearity of the time required to go between floors required more complex cost calculations.

Not all layout problems are easily amenable to the methods reviewed here. There are applications in which assuming rectangular-shaped departments is not appropriate. Examples include the layouts of office buildings, an airplane's dashboard, a city or a

neighborhood, or an integrated circuit. Drezner (1980) introduced a method called DISCON (for *dispersion and concentration*) that considers facilities as disks. The disks are first dispersed in the plane mathematically, using a system of differential equations simulating a system of springs connecting the disks. The dispersion phase provides an initial solution, which is later improved upon in the concentration phase. Drezner's approach works better than CRAFT when departments are not rectangular and when the number of facilities is large. Techniques based on graph theoretic methods (Foulds, 1983) and statistical cluster analysis (Scriabin and Vergin, 1985) also have been considered.

The "classic" software reviewed in this section illustrate the basic concepts used in most software products for facilities layout and design. Special products have been designed for distinct market segments. Office design, process piping, and industrial facilities planning are inherently very different layout problems and require specially tailored software products.

The problem of designing office facilities requires blocking out space for well-defined activities. These problems are the concern of interior designers and architects. Drafting tools with extensive libraries of symbols are the most popular for these applications. The most popular is AutoCAD, which allows for both two- and three-dimensional layouts. Process piping and layout design are typically used in conjunction with chemical plants and are almost always displayed as three-dimensional layouts. Industrial facilities layout and design usually involve some variant of the software tools discussed in this chapter, or the use of a specially tailored graphical-based simulation package, such as Pro-Model.

Intelligent design and layout of industrial facilities is a problem that is continuing to grow in importance. For example, the cost of a new fab (fabrication facility) in the semiconductor industry typically exceeds $1 billion. Similar investments are required for new automotive assembly plants. Firms cannot afford to make mistakes when planning the organization of these enormously expensive facilities.

Problems for Section 11.6

11. Briefly describe each of the following computerized layout techniques. In each case, indicate whether the method is a construction or improvement method.
    a. CRAFT
    b. COFAD
    c. ALDEP
    d. CORELAP
    e. PLANET
12. *a.* Discuss the advantages and disadvantages of using computer programs to develop layouts.
    b. Do the results of the studies discussed in Section 11.6 suggest that human planners or computer programs produce superior layouts? For what reasons are these studies not conclusive?
13. For Example 11.4, which illustrated CRAFT, verify the values of the predicted costs (2,715, 3,185, 2,735, and 2,830) obtained in the first iteration.
14. Consider the initial layout for Example 11.4, which appears in Figure 11–11. Draw a figure showing the layout obtained from exchanging the locations of A and D, and find the centroids of A and D in the new layout.
15. Determine the centroids for the six departments in the layout pictured in Figure 11–17 using the methods outlined in Appendix 11–A.

FIGURE 11–17
Layout (for Problem 15)

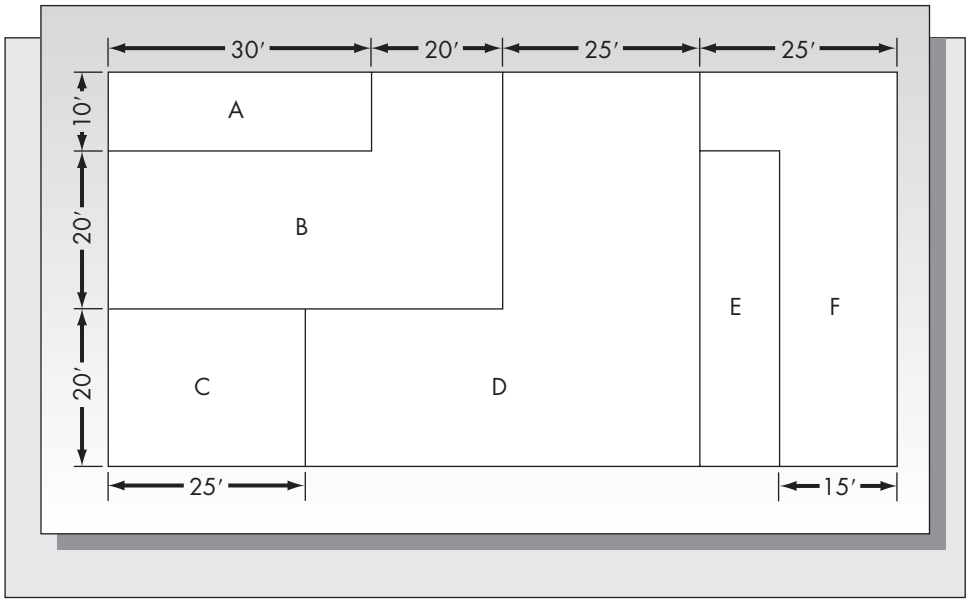
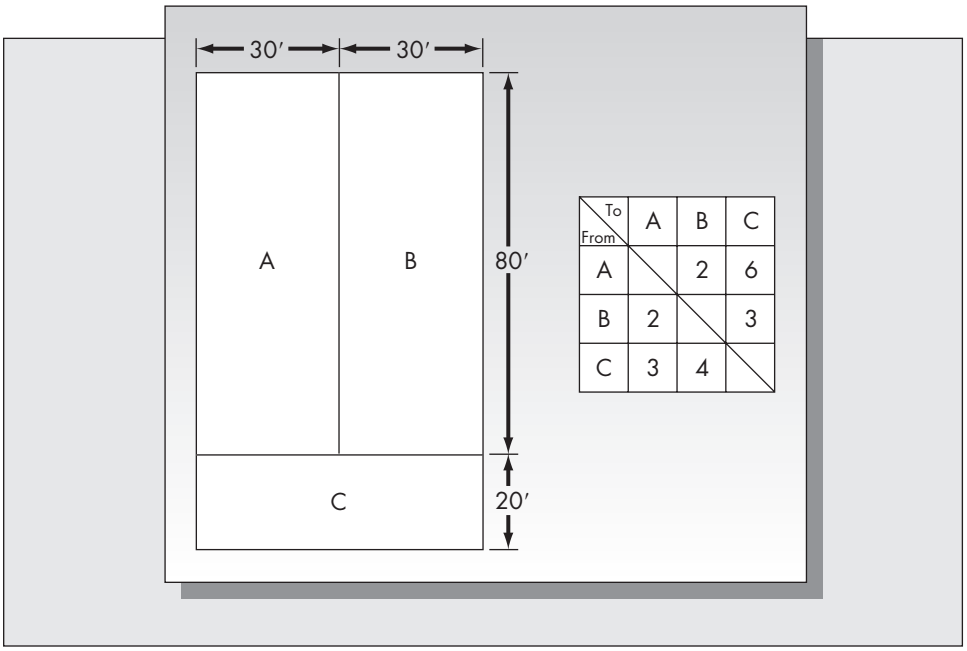
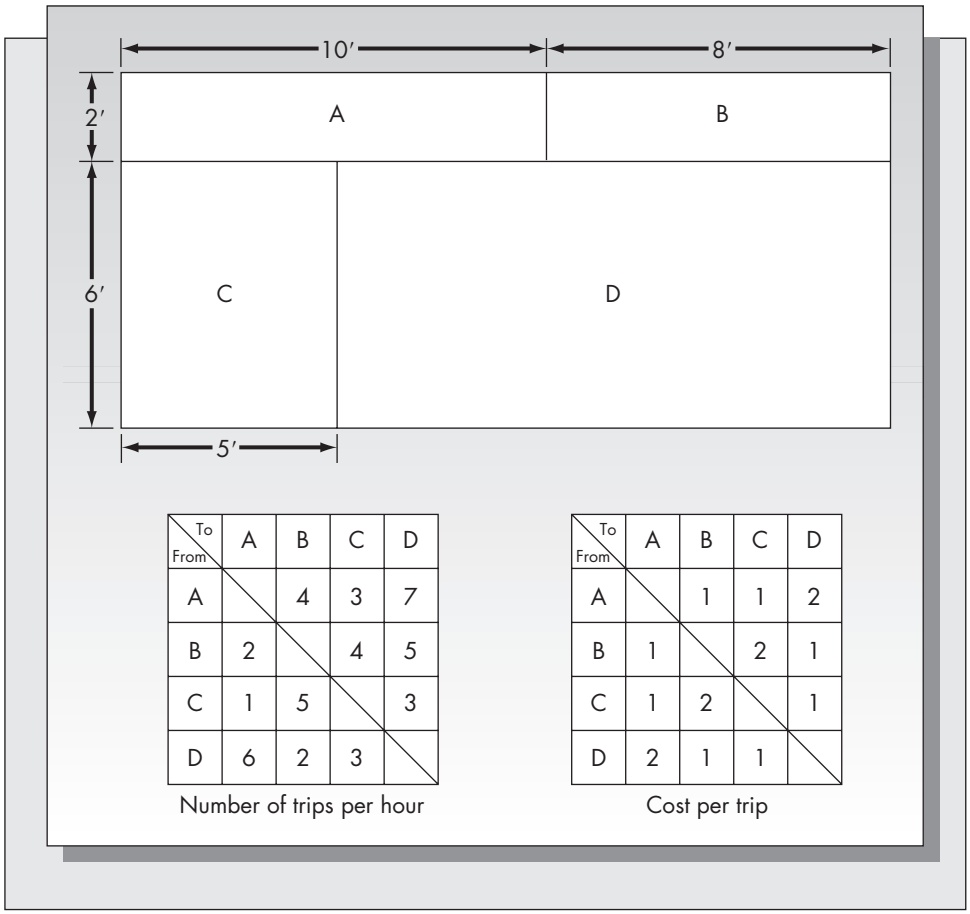


FIGURE 11–18
Layout and from-to chart (for Problem 16)



16. Consider the initial layout and from-to chart appearing in Figure 11–18. Assuming the goal is to minimize the total distance traveled, determine the final layout recommended by CRAFT using the pairwise exchange method described in this section. Compare the predicted and the actual figures for total distance traveled in the final layout.
17. Consider the initial layout pictured in Figure 11–19 and the two from-to charts giving the flow and cost data. Use the CRAFT approach to obtain a final layout. Compare the total cost of each layout predicted by CRAFT to the actual cost of each layout.
18. A facility consisting of eight departments has the activity relationship chart pictured in Figure 11–20. Compute the TCR for each department that would be used by CORELAP.

FIGURE 11–19
Layout and from-to charts (for Problem 17)



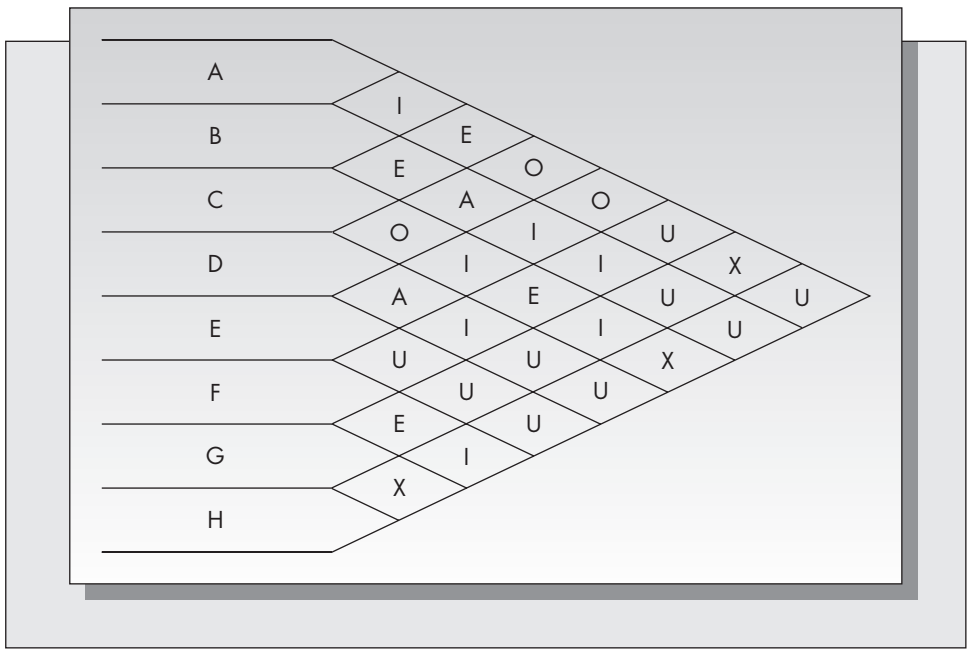
To / From	A	B	C	D
A		4	3	7
B	2		4	5
C	1	5		3
D	6	2	3	

Number of trips per hour

To / From	A	B	C	D
A		1	1	2
B	1		2	1
C	1	2		1
D	2	1	1	

Cost per trip

FIGURE 11–20
Rel chart (for Problem 18)



19. Six departments, A, B, C, D, E, and F, consume respectively 12, 6, 9, 18, 22, and 6 squares. Based on a sweep width of 4 and a facility size of 5 by 16, use the techniques employed by ALDEP to find a layout. Assume that the departments are placed in alphabetical order.

11.7 FLEXIBLE MANUFACTURING SYSTEMS

New technologies offer choices for the construction and design of facilities. Often these choices are conflicting. On one hand, the importance of flexibility cannot be overemphasized. In high-tech industries in particular, manufacturing must adapt to frequent technological advances. In the personal computer industry, for example, the changing size configurations and central processor characteristics require constant redesign of manufacturing facilities.

One development that has gotten considerable attention in recent years is the *flexible manufacturing system* (FMS). An FMS is a collection of numerically controlled machines connected by a computer-controlled materials flow system. The machines are typically used for metal cutting, forming, and assembly operations, and provide the greatest benefit when a large variety of part types are required. A full-blown FMS can be extremely expensive, requiring a capital expenditure of upward of $10 million. Because of the high cost and the long payback period, firms are opting for scaled-down versions, called flexible machining cells. However, many firms feel that the capital expenditure is justified. For example, the Citroen plant in Meudon near Paris uses an FMS to produce component prototypes in batch sizes of 1 to about 50. The system is designed to handle a wide variety of part types (Hartley, 1984).

What advantages do such systems have over a conventional dedicated machine layout? They provide the opportunity to drastically slash the hidden costs of manufacturing. These include work-in-process inventory costs and overhead costs associated with indirect labor. They allow firms to quickly change tooling and product design with minimal additional investment in new equipment and personnel. However, they do require a substantial capital investment, which can be recouped only if their potential can be realized and they are used in the right environment.

The question is, under what circumstances should a firm consider employing an FMS as part of its overall layout for manufacturing? As shown in Figure 11–21, a flexible manufacturing system is appropriate when the production volume and the variety of parts produced are moderate. For systems with low volume and high customization, stand-alone numerically controlled machines are appropriate. These can be programmed for each individual application. For high-volume production of standardized parts, fixed transfer lines are more appropriate.

The layout and structure of a typical FMS is pictured in Figure 11–22. Often the machines are controlled by a central computer, which also can be programmed for individual applications. Parts are typically loaded and unloaded at a single location along the materials handling path. The materials handling system consists of pallets, which are usually metal disks two to three feet in diameter. These pallets carry work-in-process inventory and queue up at the machines for processing. Each machine may contain from 5 to 100 different tools, which are stored on a carousel. Tools can be

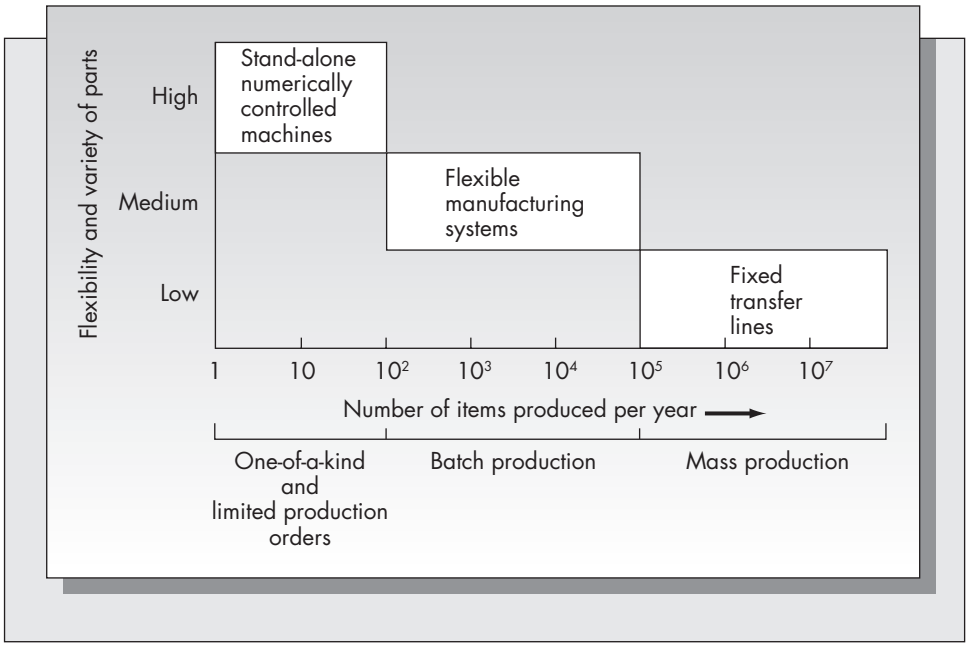


FIGURE 11–21
The position of FMS in the manufacturing hierarchy

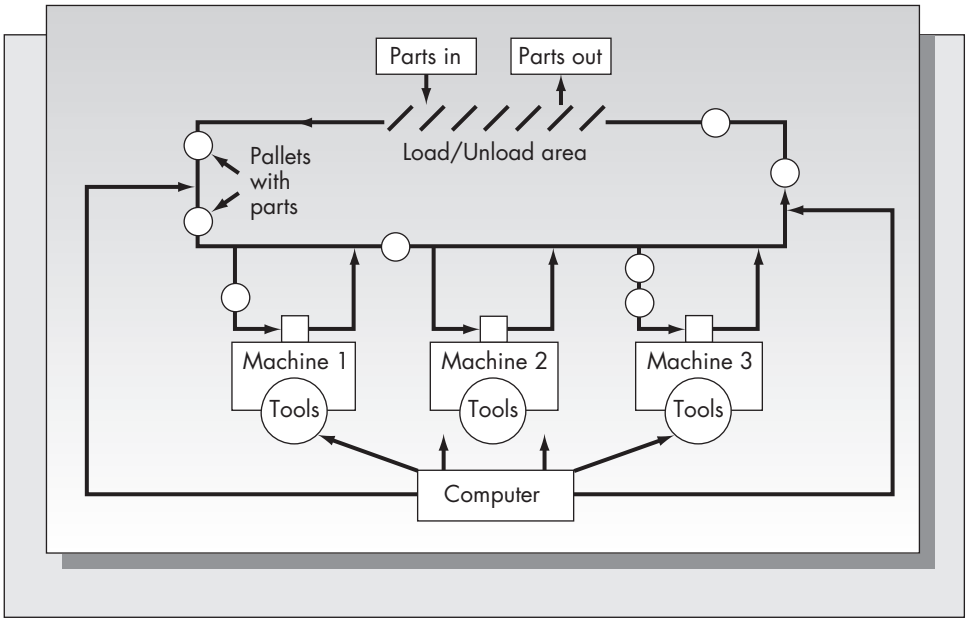


FIGURE 11–22
A typical flexible manufacturing system

exchanged in a matter of seconds. Note that the routing of parts must also be programmed into the system, as not all products require the same sequencing of machine operations. Although Figure 11–22 shows a centralized computer, some systems use individually programmed machines.

Advantages of Flexible Manufacturing Systems

When used correctly, these systems can provide substantial advantages over more rigid designs. These include

1. *Reduced work-in-process inventory.* The design of the system limits the number of pallets available for moving parts through the system. Hence, the WIP inventory never exceeds a predetermined level. In this sense, the FMS is similar to the just-in-time system, in which the level of WIP inventory is a decision variable whose value may be chosen in advance.

2. *Increased machine utilization.* Numerically controlled machines often have a utilization rate of 50 percent or less. However, an efficient FMS may have a utilization rate as high as 80 percent. The improved utilization is a result of both the reduction of changeover time of machine settings and tooling, and the ability to better balance the system workload.

3. *Reduced manufacturing lead time.* Without an FMS, parts might have to be processed through several different work centers. As a result, there could be substantial transportation time between work centers and substantial queuing time at work centers. Because an FMS reduces transportation, setup, and changeover time, it results in significant reduction in lead time for production.

4. *Ability to handle a variety of different part configurations.* As noted earlier in this section, the FMS is more flexible than a fixed transfer line but not as flexible as a stand-alone numerically controlled machine. Depending on the tooling available for the machines, parts may be launched into the system with little or no setup time required. Also, the FMS can process part configurations simultaneously.

5. *Reduced labor costs.* The number of workers required to manage an FMS can be as much as a factor of 10 fewer than the number required in a traditional job shop. Even when numerically controlled machines are used on a stand-alone basis, at least one worker is required per machine, and workers are required to transport the parts between machines. The automated materials handling capability of the FMS leads to significant reductions in labor requirements.

Disadvantages of Flexible Manufacturing Systems

Although there are many potential advantages of FMS, there is one factor that has delayed American acceptance of these systems. That factor is cost. Most FMSs cost in the tens of millions of dollars. Traditional net present value (NPV) calculations often show that the investment is not justified. A well-known case is that of the Yamazaki Machinery Company in Japan. The company installed an $18 million FMS. As a result, the number of machines was reduced from 68 to 18, the number of employees from 215 to 12, floor space from 103,000 square feet to 30,000 square feet, and the average processing time of parts from 35 days to 1.5 days (Kaplan, 1986). These figures are impressive. However, when translated to a return on investment, the story is not so rosy. The company reported a total savings of $6.9 million after two years. Including a savings of $1.5 million per year for the next 20 years, the projected total return is under 10 percent per year. In most American companies, the "hurdle rate" is generally 15 percent or higher, thus making this particular FMS operation a poor investment by NPV considerations alone. (The hurdle rate is the minimum acceptable rate of return on new projects.)

In addition to the direct costs of the equipment and the space, several indirect costs also are incurred. In order to manage the flow of materials, a sophisticated software system is required. Effective software can be extremely expensive, may require customization, often has bugs, and generally requires worker training. The cost of equipment such as feeders, conveyors, and transfer devices may not be part of the

initial purchase cost. Other indirect costs include site preparation, spare parts to support the machinery, and disruptions that might result during the installation period. Furthermore, any company that purchases an FMS must anticipate a decline in productivity that accompanies the introduction of new technology.

It is no wonder that so many American companies have trouble justifying the investment in FMS. However, Kaplan (1986) argues that traditional cost accounting methods may ignore some important considerations. One is that evaluation of alternative investments based on discounted cash flow assumes a status quo. That is, NPV analysis assumes that factors such as market share, price, labor costs, and the company's competitive position in the marketplace will remain constant. However, the values of some of these variables will change, and are more likely to degenerate if the company retains outmoded production methods.

Furthermore, most firms prefer to invest in a variety of small projects rather than make a major capital outlay for a single project. Such a philosophy is safer in the short run, but could be suboptimal in the long run. An example of an industry that failed to invest in new technology is the railroad industry. Because of outmoded equipment and facilities, many firms in this industry have been unable to stay competitive and profitable.

Cost is not the only problem. The FMS may experience downtime for a variety of reasons. Planned downtime could be the result of scheduled maintenance and scheduled tool changeovers. Unplanned downtime could result from mechanical failures of the machines or electrical failures. If a single machine goes down, the system can continue to function, but if either the materials handling system or the central computer fails, the entire FMS is crippled.

Decision Making and Modeling of the FMS

Mathematical modeling can help with the decisions required to design and manage an FMS. Flexible manufacturing systems, like all job shops, must be managed on a continual basis. Some decisions, such as whether to purchase a system at all, have very far-reaching consequences. Other decisions, such as what action should be taken when a machine breaks down or a tool is damaged, may only affect the flow of materials for the next few hours. This suggests that a natural way to categorize these decisions is by the time horizon associated with their consequences. Table 11–1 summarizes the various levels of FMS decision making broken down in this manner.

TABLE 11–1
Levels of Decision Making in Flexible Manufacturing Systems

Time Horizon	Major Issues	Modeling Methods
Long term (months–years)	Design of the system Parts mix changes System modification and/or expansion	Queuing networks Simulation Optimization
Medium term (days–weeks)	Production batching Maximizing of machine utilization Planning for fluctuations in demand and/or availability of resources	Balancing methods Network simulation
Short term (minutes–hours)	Work order scheduling and dispatching Tool management Reaction to system failures	Tool operation and allocation program Work order dispatching algorithm Simulation

FIGURE 11–23
A single-server queue

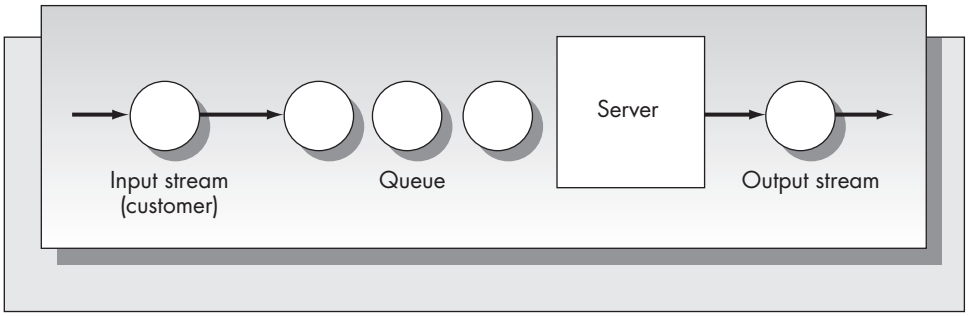
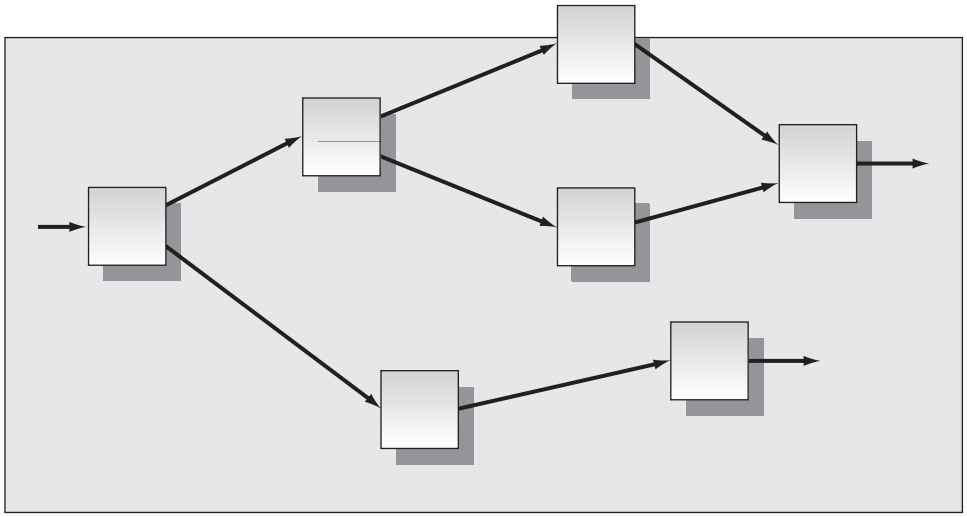


FIGURE 11–24
A network of queues



Queuing-Based Models

Queuing theory is the basis for many models that can assist with long-term decisions. A queue is simply a waiting line. A single-server queuing system is characterized by an input stream (known as customers), a service discipline specifying the order or sequence that customers are served, and a service mechanism. Most queuing models assume that both the arrival stream of customers and the service times are random variables. It is the random nature of arrivals and services and their interactions that makes queuing problems interesting.[1] Figure 11–23 shows a simple schematic of a single-server queuing system.

In an FMS, each machine corresponds to a separate queue. The jobs correspond to the customers, and the queue is the stack of jobs waiting for processing at a machine. Because an FMS is a collection of machines connected by a materials handling system, the entering stream of customers for one machine is the result of the accumulation of departing streams from other machines. Hence, an FMS is a special kind of queuing network. A typical queuing network appears in Figure 11–24.

Queuing models are most useful in aiding with system design. They can be used to compute a variety of measures relating to machine utilization and system performance. An important issue that arises during the initial design phase is system capacity. The system capacity is the maximum production rate that the FMS can sustain. It depends

[1] A summary of basic queuing theory appears in Supplement 2, which follows Chapter 8.

on both the configuration of the FMS and the assumptions one makes about the nature of the arrival stream of jobs. Schweitzer's (1977) analysis shows that the output rate of a flexible manufacturing system is simply the capacity of the slowest machine in the system.

To be more precise, suppose that machine i can process jobs at a rate μ_i. If the probabilities that an entering job visits machines in a given order are known, one can find e_i, the expected number of visits of a newly arriving job to machine i, using queuing theory methods. It follows that $1/e_i$ is the rate that jobs arrive to machine i, so μ_i/e_i is the output rate of machine i. The "slowest" machine, where slow is used in a relative sense, is that machine whose output rate is least. Hence, the output rate of the system is $\min_{1 \le i \le n} (\mu_i/e_i)$.

Such results can be very valuable when considering the design of the system and its potential utility to the firm. However, in some circumstances queuing models are only of limited value. In particular, in order to obtain explicit results for complex queuing networks, one often must assume that the arrival and the service processes are purely random. That is, both the time between the arrival of successive jobs and the time required for a machine to complete its task are exponential random variables.[2] It is unlikely that both interarrival times and job performance times in flexible manufacturing systems are accurately described by exponential distributions. There are circumstances under which exponential queuing formulas do give reasonable approximations to more complex cases, but simulation should be used to validate analytical formulas. Approximate analytical results are available for nonexponential cases as well (see Suri and Hildebrant, 1984, for example). Also, the exponential assumption is not required for some simple network configurations. A review of the analytical results for network queuing models of flexible manufacturing systems can be found in Buzacott and Yao (1986).

Even with these limitations, queuing models provide a powerful means for analyzing the performance characteristics and capacity limitations of flexible manufacturing systems. Eventually, network queuing models will accurately reflect the nature of FMSs and provide an effective means for assisting with system design and management.

Mathematical Programming Models

Stecke (1983) considers the use of mathematical programming for solving several decision problems arising in FMS management. These problems include the following:

1. *Part type selection problem.* From a given set of part types that have associated known requirements, determine a subset for immediate and simultaneous processing.
2. *Machine grouping problem.* Partition the machines into groups so that each machine in a group can perform the same set of operations.
3. *Production ratio problem.* Determine the relative ratios to produce the part types selected in step 1.
4. *Resource allocation problem.* Allocate the limited number of pallets and fixtures among the selected part types.
5. *Loading problem.* Allocate the operations and required tools of the selected part types among the machine groups subject to technological capacity constraints of the FMS.

[2] The properties of the exponential distribution are discussed in detail in Chapter 2 and in Supplement 2 on queuing.

We will not present the details of the mixed integer programming formulations. Our purpose is simply to note that mathematical programming is another means of helping with long-range and medium-range decisions affecting the FMS. Stecke (1983) shows how one would apply these models using actual data from an existing FMS.

The Future of FMS

There is substantial potential for the application of FMS both in the United States and abroad. It is estimated that about half of the annual expenditures on manufacturing is in the metal-working industry, and two-thirds of metal working is metal cutting (Stecke, 1983). Furthermore, about three-fourths of the dollar volume of metal-worked parts is manufactured in batches of fewer than 50 parts (Cook, 1975). Evidently, there is a huge market for FMSs in the United States. Similar potential exists in other industrialized countries as well. In fact, the rate of growth of installed FMSs is higher in Japan than in any other country in the world (Jaikumar, 1984).

The rate of growth of FMSs in the United States has been healthy if not spectacular. In 1981 there were about 18 systems operational in the United States (Jaikumar, 1984), and by late 1986 this number had grown to about 50. Krouse (1986) estimated that there should have been 284 systems in place in the United States by 1990. As we noted earlier, the primary concern among companies considering purchasing an FMS is cost. The average cost of such systems was estimated to be $12.5 million in 1986 and is probably higher today. Also, installation generally requires from 18 to 24 months.

Both hardware and changeover costs discourage many potential users of FMSs. Furthermore, many factories in this country have yet to invest in numerically controlled machine tools, let alone a full-blown FMS. As a response to this, vendors are offering scaled-down versions of the FMS known as flexible manufacturing cells. Flexible manufacturing cells are small FMSs offering fewer machines, a less extensive materials handling system, and fewer tools per machine. Vendors anticipate healthy sales of these cells, which should result in an expanded customer base and a larger market for scaled-up systems.

Cost is not the only factor retarding acceptance. The performance of many existing FMSs has been disappointing. Zygmont (1986) reported that some installed systems did not meet the expectations of the companies that purchased them. In one case, three years were required to debug a system purchased by Electro-Optical and Data Systems of Hughes Aircraft and in the end the system was still less flexible than originally expected. One of the problems encountered was that the level of precision required was higher than the system was capable of delivering. Similarly, Deere and Co. was disappointed with its $20 million system, and was especially disappointed with the lack of flexibility afforded by the software for handling complex and varying part sequences.

Jaikumar (1986) makes a case that the problem with FMSs in the United States is not the systems themselves but the way they are used. In a comparison of FMSs in the United States and Japan, he found some striking differences. These differences are listed in Table 11–2.

Most telling is the fact that there are almost 10 times more part types produced on each Japanese system than on each American system. That, coupled with the enormous difference in annual volume per part, indicates that American firms are using FMSs improperly. That is, they are used as if they were simply another set of machines for high-volume standardized production rather than for producing the varied part mix for which they were designed. Another important difference is that many Japanese systems are run both day and night and often unattended. Jaikumar (1986) claims that this is a consequence of the improved design and reliability of the Japanese systems.

TABLE 11–2
A Comparison of FMSs in the United States and Japan

	United States	Japan
System development times (years)	25 to 3	1 to 1.25
Types of parts produced per system	10	93
Annual volume per part	1.727	258
Number of new parts introduced per year	1	22
Number of systems with unattended operations	0	18
Average metal-cutting time per day	83	202

As firms install additional FMSs, vendors will gain a better understanding of their power and their limitations. Advances in both hardware and software should make future systems more capable of dealing with the problems discussed. As both the systems and the understanding of the circumstances under which they can be most effective improve, we should see many more companies installing FMSs in the next decade.

Problems for Section 11.7

20. In each case listed, state whether the factor listed is an advantage or a disadvantage of FMS. Discuss the reasons for your choice.
    a. Cost.
    b. Ability to handle different parts requirements.
    c. Advances in manufacturing technology.
    d. Reliability.
21. For each of the case situations described, state which of the three types of manufacturing systems would be most appropriate: (1) stand-alone numerically controlled machines, (2) FMS, (3) fixed transfer line, or (4) another system. Explain your choice in each case.
    a. A local machine shop that accepts custom orders from a variety of clients in the electronics industry.
    b. An international supplier of standard-sized metal containers.
    c. The metal-working division of a large aircraft manufacturer, which must serve the needs of a variety of divisions in the firm.
22. For what reason might a traditional NPV (net present value) analysis not be an appropriate means for evaluating the desirability of purchasing an FMS?
23. With which decisions related to FMS design or control can the following methods assist? Also, briefly describe how each method would be used.
    a. Queuing theory
    b. Simulation
    c. Mathematical programming

11.8 LOCATING NEW FACILITIES

Chapter 1 discussed several qualitative considerations when deciding on the best location for a new plant or other facility. The remainder of this chapter will consider quantitative techniques that can help with location decisions. There are many circumstances in which the objective can be quantified easily, and it is in these cases that analytical solution methods are most useful.

Snapshot Application

KRAFT FOODS USES OPTIMIZATION AND SIMULATION TO DETERMINE BEST LAYOUT

In the mid-1990s Kraft decided to renovate one of its major manufacturing facilities located in the Midwest. New higher-capacity production lines were to replace the old lines. The new lines would have increased throughput and more mixing capabilities and would be able to operate at a variety of speeds. One of the concerns in designing the new facility was determining the optimum number of new lines to install to maintain current capacity levels while also allowing for future growth. Management needed to decide the line configuration quickly because the lines required a six-month delivery lead time.

Kraft management asked the firm's operations research group to develop a detailed mathematical model of the system. A "seat of the pants" approach was simply unacceptable: a wrong decision would be costly. Too many lines would result in wasted space and money, and too few in a capacity bottleneck. A further difficulty was that none of Kraft's plants had used these lines before, so there was no prior experience from which to draw.

Staff members from the firm's operations research group were asked to consider both the problem of optimizing the number of new lines and the flow of material within the plant. To address the optimization problem, the group developed an integer programming formulation of the problem. The objective was to ensure that the makespan of any schedule Kraft was likely to encounter did not exceed one eight-hour shift. (Refer to Chapter 8 for a discussion of makespan.) They determined that six lines would easily take care of existing capacity requirements and include a substantial margin for future growth. The model was solved using the AMPL optimization package on a personal computer.

Once the optimal number of new lines was determined, the next step was to develop a detailed simulation of the factory floor to show graphically how the new design would work in practice. For this task, the group decided to use F&H Simulation's Taylor II software package, a graphical-based simulation package especially well-suited for considering the effects of different work schedules on a given layout. The simulation model consisted of 139 elements including lines, machines, work-in-process buffers, loading areas for raw materials, and shipping areas. By simulating the process with the heaviest schedules, the group demonstrated that the new layout would easily have enough capacity to handle the most severe demands the plant is likely to see. Furthermore, the simulation afforded factory personnel a means to see how the flow of materials would change prior to the time that the plant was renovated and the new layout implemented.

Source: Based on Sengupta and Combes (1995).

In some sense, the plant layout problem is a special case of the location problem. However, the problems considered in the remainder of this chapter differ from layout problems in the following way: we will assume that there are one or more existing facilities already in place and we wish to find the optimal location of a new facility. We will concentrate on the problem of locating a single new facility, although there are versions of the problem in which one must determine the locations of multiple facilities.

Examples of the type of one-facility location problems that can be solved by the methods discussed in this chapter include

1. Location of a new storage warehouse for a company with an existing network of production and distribution centers.
2. Determining the best location for a new machine in a job shop.
3. Locating the computer center in a university.
4. Finding the best location for a new hospital in a metropolitan area.
5. Finding the most suitable location for a power generating plant designed to serve a geographic region.
6. Determining the placement of an ATM in a neighborhood.
7. Locating a new police station in a community.

Undoubtedly, the reader can think of many other examples of location problems. Analytical methods assume that the objective is to locate the new facility to minimize

some function of the distance of the new location from existing locations. For example, when locating a new warehouse, an appropriate objective would be to minimize the total distance traveled to the warehouse from production facilities and from the warehouse to retail outlets. A hospital would be located to be most easily accessible to the largest proportion of the population in the area it serves. A machine would be placed to minimize the weighted sum of materials handling trips to and from the machine. Clearly, choosing the location of a facility to minimize some function of the distance separating the new facility from existing facilities is appropriate for many real location problems.

Measures of Distance

Two measures of distance are most common: *Euclidean distance* and *rectilinear distance.* Euclidean distance is also known as straight-line distance. The Euclidean distance separating two points is simply the length of the straight line connecting the points. Suppose that an existing facility is located at the point (a, b) and let (x, y) be the location of the new facility. Then the Euclidean distance between (a, b) and (x, y) is

$$\sqrt{(x - a)^2 + (y - b)^2}.$$

The rectilinear distance (also known as metropolitan distance, recognizing the fact that streets usually run in a crisscross pattern) is given by the formula

$$|x - a| + |y - b|.$$

Figure 11–25 illustrates the difference between these two distance measures.

Rectilinear distance is appropriate for many location problems. Distances in metropolitan areas tend to be more closely approximated by rectilinear distances than by Euclidean distances even when the street pattern is not a perfect grid. In many manufacturing environments, material is transported across aisles arranged in regular patterns. It is a fortunate coincidence that rectilinear distance is more common than Euclidean distance, because the rectilinear distance problem is easier to solve.

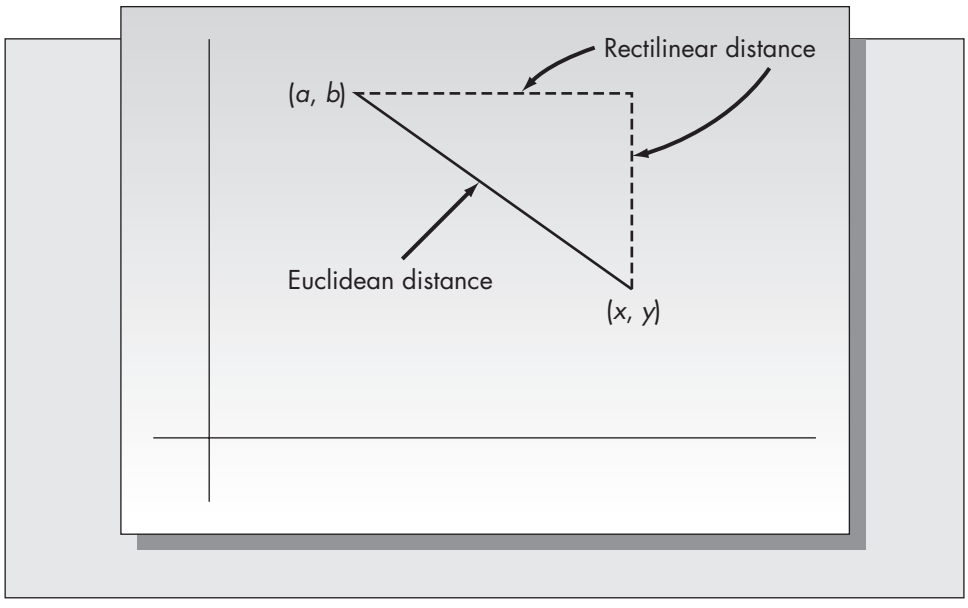


FIGURE 11–25
Euclidean and rectilinear distances

Problems for Section 11.8

24. Consider the problem of locating a new hospital in a metropolitan area. List the factors that can be quantified and those that cannot. Comment on the usefulness of quantitative methods in the decision-making process.
25. A coordinate system is superimposed on a map. Three existing facilities are located at (5, 15), (10, 20), and (6, 9). Compute both the rectilinear and the Euclidean distances separating each facility from a new facility located at $(x, y) = (8, 8)$.
26. For the situation described in Problem 25, suppose that there are only three feasible locations for the new facility: (8, 16), (6, 15), and (4, 18).
    a. What is the optimal location if the objective is to minimize the total rectilinear distance to the three existing facilities?
    b. What is the optimal location if the objective is to minimize the total Euclidean distance to the three existing facilities?
27. For each of the seven examples of location problems listed in this section, indicate which distance measure, Euclidean or rectilinear, would be more appropriate. (Discuss how, in some cases, one or the other objective could be appropriate for the same problem, depending upon the optimization criterion used.)

11.9 THE SINGLE-FACILITY RECTILINEAR DISTANCE LOCATION PROBLEM

In this section we will present a solution to the general problem of locating a new facility among n existing facilities. The objective is to locate the new facility to minimize a weighted sum of the rectilinear distances from the new facility to existing facilities. Assume that the existing facilities are located at points $(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n)$. Then the goal is to find values of x and y to minimize

$$f(x, y) = \sum_{i=1}^{n} w_i(|x - a_i| + |y - b_i|).$$

The weights are included to allow for different traffic rates between the new facility and the existing facilities. A simplifying property of the problem is that the optimal values of x and y may be determined separately, as

$$f(x, y) = g_1(x) + g_2(y),$$

where

$$g_1(x) = \sum_{i=1}^{n} w_i\,|x - a_i|$$

and

$$g_2(y) = \sum_{i=1}^{n} w_i\,|y - b_i\,|.$$

As we will see, there is always an optimal solution with x equal to some value of a_i and y equal to some value of b_i. (There may be other optimal solutions as well.)

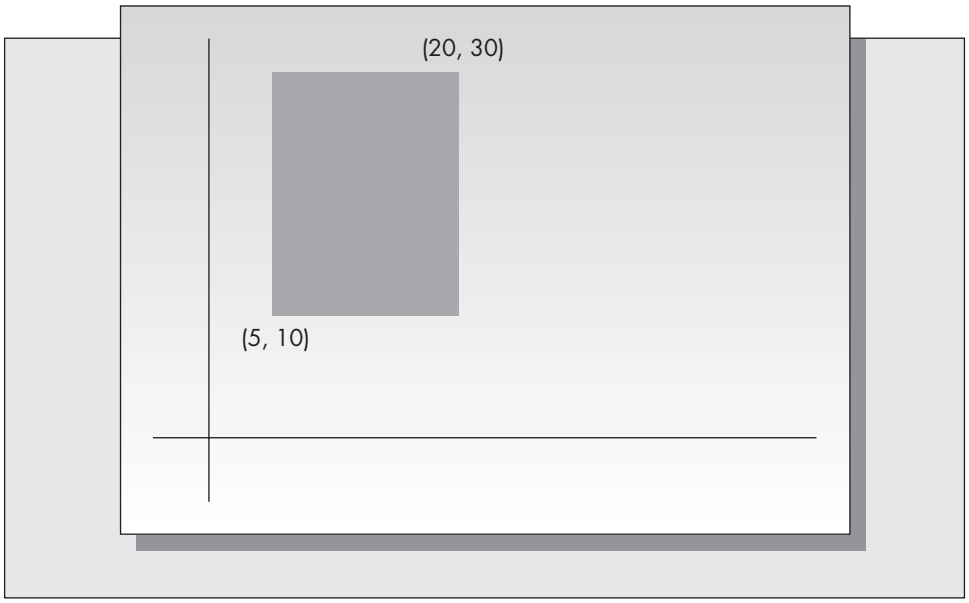


FIGURE 11–26
Optimal locations of the new facility for a rectilinear distance measure

However, before presenting the general solution algorithm for finding the optimal location of the new facility, we consider a few simple examples in order to provide the reader with some intuition. Consider first the case in which there are exactly two existing facilities, located at (5, 10) and (20, 30) as pictured in Figure 11–26. Assume that the weight applied to each of these facilities is 1. If x assumes any value between 5 and 20, the value of $g_1(x)$ is equal to 15. (For example, if $x = 13$, then $g_1(x) = |5 - 13| + |13 - 20| = 8 + 7 = 15$.) Similarly, if y assumes any value between 10 and 30, then $g_2(y) = 20$. Any value of x outside the closed interval [5, 20] and any value of y outside the closed interval [10, 30] results in larger values of $g_1(x)$ and $g_2(y)$. Hence, the optimal solution is (x, y) with $5 \leq x \leq 20$ and $10 \leq y \leq 30$. All locations in the shaded region pictured in Figure 11–26 are optimal.

As in the example, there always will be an optimal location of the new facility with coordinates coming from the set of coordinates of the existing facilities. Suppose that the existing facilities have locations (3, 3), (6, 9), (12, 8), and (12, 10). Again assume that the weight applied to these locations is 1. Ranking the x locations in increasing order gives 3, 6, 12, 12. A *median* value is such that half of the x values lie above it and half of the x values lie below it. Any value of x between 6 and 12 is a median location and is optimal for this problem. The optimal value of $g_1(x)$ is 15. (The reader should experiment with a number of different values of x between 6 and 12 to satisfy himself or herself that this is the case.) Ranking the y values in increasing order gives 3, 8, 9, 10. The median value of y is between 8 and 9, and the optimal value of $g_2(y) = 8$.

The optimal solution is to locate (x, y) at the median of the existing facilities. This result carries over to the case in which there are weights different from 1. Suppose in Example 11.4 that we were given the locations of four machines in a job shop. The goal is to find the location of a fifth machine to minimize the total distance traveled to transport material between the new machine and the existing ones. Assume that on average there are respectively 2, 4, 3, and 1 materials handling

trips per hour from the existing machines to the new machine. Summarizing the given information,

Location of Existing Machines	Weight
(3, 3)	2
(6, 9)	4
(12, 8)	3
(12, 10)	1

This problem is equivalent to one in which there are two machines at location (3, 3), four machines at location (6, 9), three machines at location (12, 8), and one machine at location (12, 10), with weights equal to 1. Hence, the x locations in increasing order are 3, 3, 6, 6, 6, 6, 12, 12, 12, 12. The median location is $x = 6$. The y locations in increasing order are 3, 3, 8, 8, 8, 9, 9, 9, 9, 10. The median location is any value of y on the interval [8, 9]. The reader should check that the value of the objective function at the optimal solution is $30 + 16 = 46$.

When the weights are large, this approach is inconvenient. A quicker method of finding the optimal location of the new facility is to compute the accumulated weights and determine the location or locations corresponding to half of the accumulated weight. The procedure is best illustrated by example.

Example *11.5*

University of the Far West has purchased equipment that permits faculty to prepare videotapes of lectures. The equipment will be used by faculty from six schools on campus: business, education, engineering, humanities, law, and science. The locations of the buildings on the campus are pictured in Figure 11–27. The coordinates of the locations and the numbers of faculty

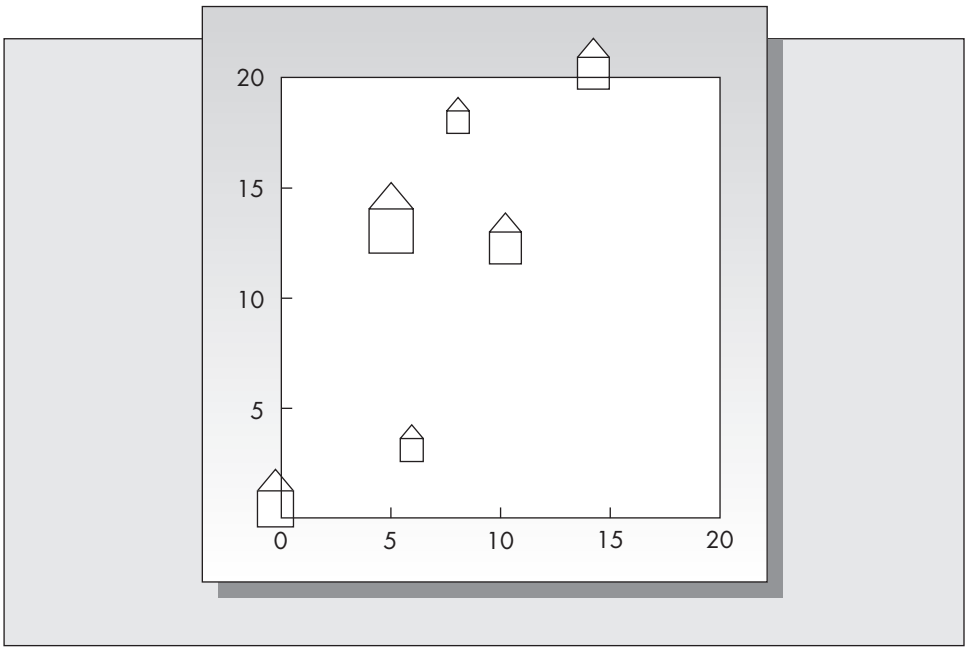


FIGURE 11–27
Location of six campus buildings (refer to Example 11.5)

that are anticipated to use the equipment are as follows:

School	Campus Location	Number of Faculty
Business	(5, 13)	31
Education	(8, 18)	28
Engineering	(0, 0)	19
Humanities	(6, 3)	53
Law	(14, 20)	32
Science	(10, 12)	41

The campus is laid out with large grassy areas separating the buildings, and walkways are mainly east–west or north–south, so that distances between buildings are rectilinear. The university planner would like to locate the new facility so as to minimize the total travel time of all faculty planning to use it.

We will find the optimal values of the x and y coordinates separately. Consider the optimal x coordinate value. We first rank x coordinates in increasing value and accumulate the weights.

School	x Coordinate	Weight	Cumulative Weight
Engineering	0	19	19
Business	5	31	50
Humanities	6	53	103
Education	8	28	131
Science	10	41	172
Law	14	32	204

The optimal value of the x coordinate is found by dividing the total cumulative weight by 2 and identifying the first location at which the cumulative weight exceeds this value. In the example this is the first time that the cumulative weight exceeds $204/2 = 102$. This occurs at $x = 6$, when the cumulative weight is 103. Hence, the optimal $x = 6$.

We use the same procedure to find the optimal value of the y coordinate. The rankings are given here.

School	y Coordinate	Weight	Cumulative Weight
Engineering	0	19	19
Humanities	3	53	72
Science	12	41	113
Business	13	31	144
Education	18	28	172
Law	20	32	204

In this case the cumulative weight first exceeds 102 when $y = 12$. Hence, $y = 12$ is optimal. The optimal location of the new facility is (6, 12). This solution is unique. Multiple optima will occur when a value of the cumulative weight exactly equals half of the total cumulative weight. For example, suppose that the weight for science were 19 instead of 41. Then the cumulative weight for science would be 91, exactly half of the total weight of 182, and all values of y on the closed interval [12, 13] would be optimal.

Contour Lines

Example 11.5 suggests an interesting question. Suppose that the location (6, 12) is infeasible. How can the university gauge the cost penalty when locating the new facility elsewhere? Contour lines, or isocost lines, can assist with determining the penalty of nonoptimal solutions. A contour line is a line of constant cost: locating a new facility anywhere along a contour line results in exactly the same cost.

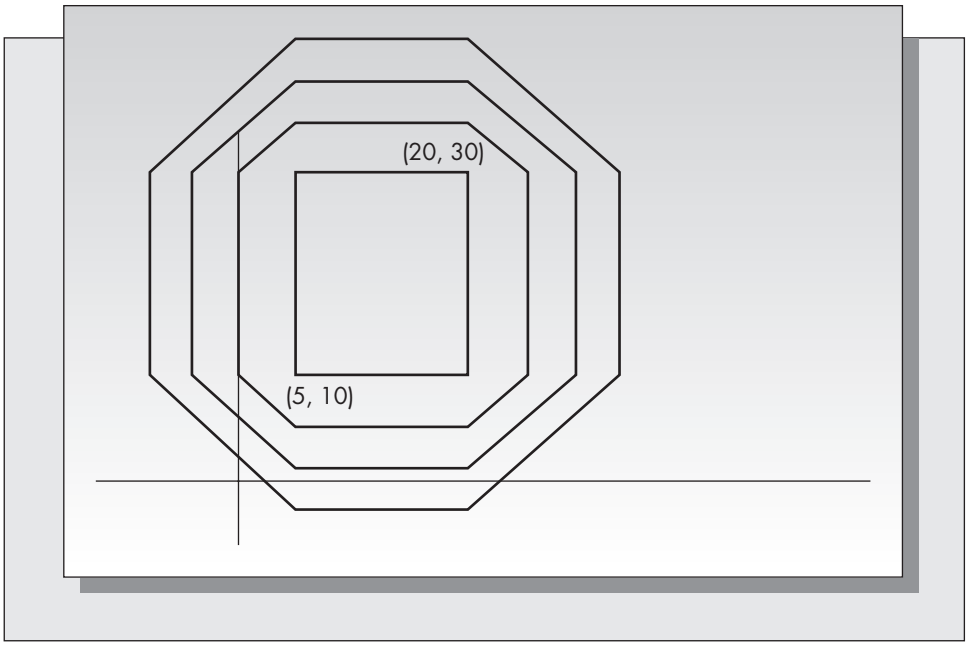


FIGURE 11–28
Contour lines for the two-facility problem pictured in Figure 11–26

We have pictured the contour lines for the simple example of Figure 11–26, in which there are only two existing facilities and the weights are equal, in Figure 11–28. The reader should convince himself or herself that the total rectilinear distance from any point along a contour line to the two points (5, 10) and (20, 30) is the same. Determining contour lines involves computing the appropriate slope for each of the regions obtained by drawing vertical and horizontal lines through each of the points (a_i, b_i). The procedure for determining contour lines is outlined in Appendix 11–B at the end of this chapter.

In Figure 11–29 we have pictured contour lines for the Example 11.5 problem in which the university must locate an audiovisual center. If the optimal location (6, 12) is infeasible, the university administration could use this map to see the penalties associated with alternative sites.

Minimax Problems

We have assumed thus far that the new facility should be placed so as to minimize the sum of the weighted distances to all existing facilities. There are circumstances in which this objective is inappropriate, however. Consider the following example. The city is considering locations for a paramedic facility. The paramedics should be able to respond to emergency calls anywhere in the city. Certain conditions, such as a severe heart attack, must be treated quickly if the patient is to have any chance of surviving. Hence, the facility should be located so that *all* locations in the city can be reached in a given time.

In such a case the objective would be to determine the location of the new facility to minimize the maximum distance to the existing facilities rather than the total distance. Let $f(x, y)$ be the maximum distance from the new facility to the existing facilities. Then

$$f(x, y) = \max_{1 \le i \le n} (|x - a_i| + |y - b_i|).$$

The objective is to find (x^*, y^*) that satisfies

$$f(x^*, y^*) = \min_{x,y} f(x, y).$$

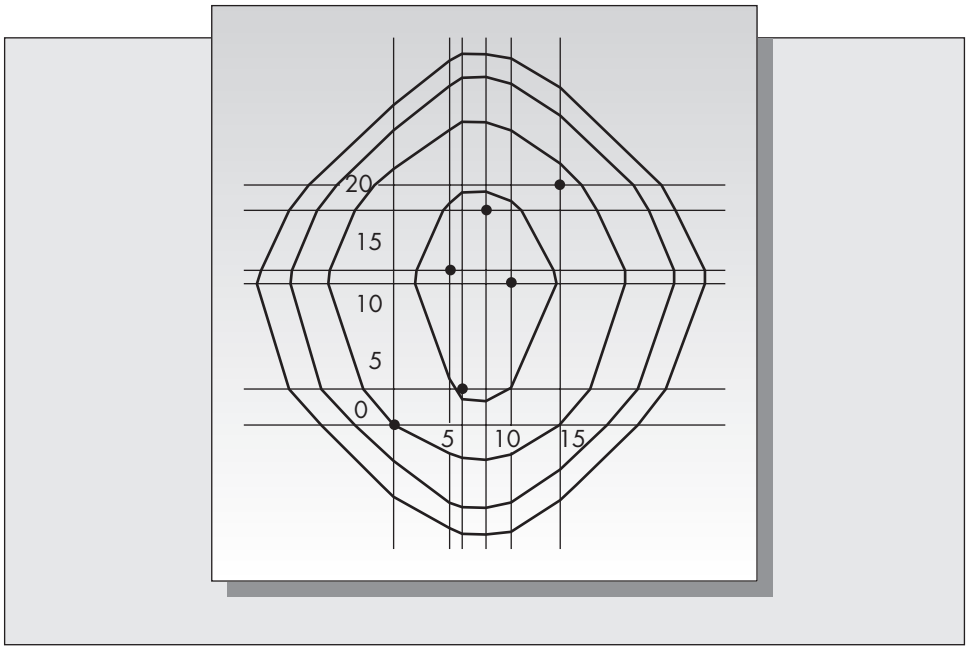


FIGURE 11–29
Contour lines for the university location in Example 11.5

The procedure for finding the optimal minimax location is straightforward. Linear programming can be used to show that the procedure we will outline is optimal. (We will not present the details here. The interested reader should refer to Francis and White, 1974.) Define the numbers c_1, c_2, c_3, c_4, and c_5:

$$c_1 = \min_{1 \le i \le n} (a_i + b_i),$$

$$c_2 = \max_{1 \le i \le n} (a_i + b_i),$$

$$c_3 = \min_{1 \le i \le n} (-a_i + b_i),$$

$$c_4 = \max_{1 \le i \le n} (-a_i + b_i),$$

$$c_5 = \max(c_2 - c_1, c_4 - c_3).$$

Let

$$x_1 = (c_1 - c_3)/2,$$
$$y_1 = (c_1 + c_3 + c_5)/2,$$

and

$$x_2 = (c_2 - c_4)/2,$$
$$y_2 = (c_2 + c_4 - c_5)/2.$$

Then all points that lie along the line connecting (x_1, y_1) and (x_2, y_2) are optimal. That is, every optimal solution to the minimax problem, (x^*, y^*), can be expressed in the form

$$x^* = \lambda x_1 + (1 - \lambda)x_2,$$
$$y^* = \lambda y_1 + (1 - \lambda)y_2,$$

where λ is a constant satisfying $0 \le \lambda \le 1$. The optimal value of the objective function is $c_5/2$.

Example *11.6*

Consider Example 11.5 of the University of the Far West. As some faculty members have disabilities, the president has decided to locate the audiovisual facility to minimize the maximum distance from the facility to the six schools on campus. Recall that the locations of the schools are

(5, 13), (8, 18),
(0, 0), (6, 3),
(14, 20), (10, 12).

The values of the constants $c_1, \ldots, c_5$ are

$$c_1 = \min_{1 \le i \le n} (a_i + b_i) = \min(18, 26, 0, 9, 34, 22) = 0,$$

$$c_2 = \max_{1 \le i \le n} (a_i + b_i) = \max(18, 26, 0, 9, 34, 22) = 34,$$

$$c_3 = \min_{1 \le i \le n} (-a_i + b_i) = \min(8, 10, 0, -3, 6, 2) = -3,$$

$$c_4 = \max_{1 \le i \le n} (-a_i + b_i) = \max(8, 10, 0, -3, 6, 2) = 10,$$

$$c_5 = \max(c_2 - c_1, c_4 - c_3) = \max(34, 13) = 34.$$

Hence, it follows that

$$x_1 = (c_1 - c_3)/2 = [0 - (-3)]/2 = 1.5,$$
$$y_1 = (c_1 + c_3 + c_5)/2 = (0 - 3 + 34)/2 = 15.5,$$

and

$$x_2 = (c_2 - c_4)/2 = (34 - 10)/2 = 12,$$
$$y_2 = (c_2 + c_4 - c_5)/2 = (34 + 10 - 34)/2 = 5.$$

All points on the line connecting (x_1, y_1) and (x_2, y_2) are optimal. The value of the objective function at the optimal solution(s) is 34/2 = 17. The optimal locations for the minimax problem are pictured in Figure 11–30. Recall that the optimal solution when we used a weighted objective was (6, 12). It is interesting to note that one optimal solution to this problem, (6, 11), is quite close to that solution.

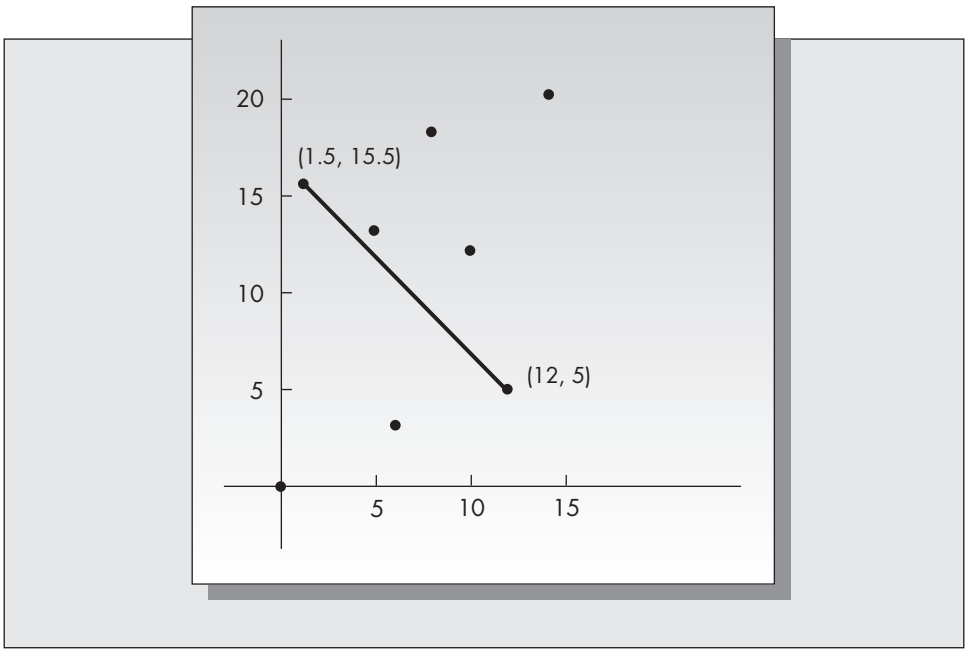


FIGURE 11–30
Optimal solutions for minimax location objective

Problems for Section 11.9

28. A machine shop has five machines, located at (3, 3), (3, 7), (8, 4), (12, 3), and (14, 6), respectively. A new machine is to be located in the shop with the following expected numbers of loads per hour transported to the existing machines: $\frac{1}{8}$, $\frac{1}{8}$, $\frac{1}{4}$, 1, and $\frac{1}{6}$. Material is transported along parallel aisles, so a rectilinear distance measure is appropriate. Find the coordinates of the optimal location of the new machine to minimize the weighted sum of the rectilinear distances from the new machine to the existing machines.
29. Solve the problem of locating the new audiovisual center at the University of the Far West described in Example 11.5, assuming that the weights are respectively 80, 12, 56, 104, 42, and 17 for the schools of business, education, engineering, humanities, law, and science.
30. Armand Bender plans to visit six customers in Manhattan. Three are located in a building at 34th Street and 7th Avenue. The remaining customers are at 48th and 8th, 38th and 3rd, and 42nd and 5th. Streets are separated by 200 feet and avenues by 400 feet. He plans to park once and walk to all of the customers. Assume that he must return to his car after each visit to pick up different samples. At what location should he park in order to minimize the total distance traveled to the clients? (Hint: In designing your grid, be sure that you account for the fact that avenues are twice as far apart as streets.)
31. In Problem 30, suppose that Mr. Bender can park only in lots located at (1) 40th Street and 8th Avenue, (2) 46th Street and 6th Avenue, and (3) 33rd Street and 5th Avenue. Where should he plan to park?
32. An industrial park consists of 16 buildings. The corporations in the park are sharing the cost of construction and maintenance for a new first-aid center. Because of the park's layout, distances between buildings are most closely approximated by a rectilinear distance measure. Weights for the buildings are determined based on the frequency of accidents. Find the optimal location of the first-aid center to minimize the weighted sum of the rectilinear distances to the 16 buildings.

Building	a_i	b_i	w_i	Building	a_i	b_i	w_i
1	0	0	9	9	14	6	11
2	10	3	7	10	19	0	17
3	8	8	4	11	20	4	14
4	12	20	3	12	14	25	6
5	4	9	2	13	3	14	5
6	18	16	12	14	6	6	8
7	4	1	4	15	9	21	15
8	5	3	5	16	10	10	4

33. Draw contour lines for the problem of locating a machine shop described in Problem 28. (Refer to Appendix 11–B.)
34. Draw contour lines for the location problem described in Problem 30. (Refer to Appendix 11–B.)
35. Two facilities, located at (0, 0) and (0, 10), have respective weights 2 and 1. Draw contour lines for this problem. (Refer to Appendix 11–B.)
36. Solve Problem 28 assuming a minimax rectilinear objective.
37. Solve Problem 30 assuming a minimax rectilinear objective.
38. Solve Problem 32 assuming a minimax rectilinear objective.

11.10 EUCLIDEAN DISTANCE PROBLEMS

Although the rectilinear distance measure is appropriate for many real problems, there are applications in which the appropriate measure of distance is the straight-line measure. An example is locating power-generating facilities in order to minimize the total amount of electrical cable that must be laid to connect the plant to the customers. This section will consider the Euclidean problem and a variant of it known as the gravity problem, which has a far simpler solution.

The Gravity Problem

The gravity problem corresponds to the case of an objective equal to the square of the Euclidean distance. Hence, the objective is to find values of (x, y) to minimize

$$f(x, y) = \sum_{i=1}^{n} w_i[(x - a_i)^2 + (y - b_i)^2].$$

This objective is appropriate when the cost of locating new facilities increases as a function of the square of the distance of the new facility to the existing facilities. Although such an objective is not common, the solution to this problem is straightforward and often has been used as an approximation to the more common straight-line distance problem.

The optimal values of (x, y) are easily determined by differentiation. The partial derivatives of the objective function with respect to x and y are

$$\frac{\partial f(x, y)}{\partial x} = 2\sum_{i=1}^{n} w_i(x - a_i),$$

$$\frac{\partial f(x, y)}{\partial y} = 2\sum_{i=1}^{n} w_i(y - b_i).$$

Setting these partial derivatives equal to zero and solving for x and y gives the optimal solution

$$x^* = \frac{\sum_{i=1}^{n} w_i a_i}{\sum_{i=1}^{n} w_i}$$

$$y^* = \frac{\sum_{i=1}^{n} w_i b_i}{\sum_{i=1}^{n} w_i}$$

The term *gravity problem* arises for the following reason. Suppose that one places a map of the area in which the facility is to be located on a heavy piece of cardboard. Weights proportional to the numbers w_i are placed at the locations of the existing facilities. Then the gravity solution is the point on the map at which the entire thing would balance. (This particular description is by Keefer, 1934.) Although one could certainly solve the gravity problem this way, it is so easy to find (x^*, y^*) using the given formulas that there seems little reason to employ the physical model.

Example *11.7* We will find the solution to the problem of locating the audiovisual center for the University of the Far West assuming a squared Euclidean distance location measure. Substituting the values of the weights and the building locations into the given formulas, we obtain

$$x^* = 1{,}555/204 = 7.6,$$
$$y^* = 2{,}198/204 = 10.8,$$

which is somewhat different from the rectilinear solution (6, 12).

The Straight-Line Distance Problem

The straight-line distance measure arises much more frequently than does the squared-distance measure discussed in Section 11.9. The objective in this case is to find (x, y) to minimize

$$f(x, y) = \sum_{i=1}^{n} w_i \sqrt{(x - a_i)^2 + (y - b_i)^2}.$$

Unfortunately, it is not as easy to find the optimal solution mathematically when using a Euclidean distance measure as it is when using squared Euclidean distance. As with the gravity problem, there is also a physical model that one could construct to find the optimal solution. In this case one places a map of the area on a table top. Holes are punched in the table top at the locations of existing facilities, and weights are suspended on strings through the holes. The size of the weights should be proportional to the relative weights of the locations. The strings are attached to a ring. If there is no friction, the ring will come to rest at the location of the optimal solution (see Figure 11–31). Although this method can be used to find a solution, it does have drawbacks. In particular, the friction between the string and the table must be negligible to be sure that the ring comes to rest in the correct position.

Determining the optimal solution mathematically is more difficult for Euclidean distance than for either rectilinear or squared Euclidean distance. There are no known

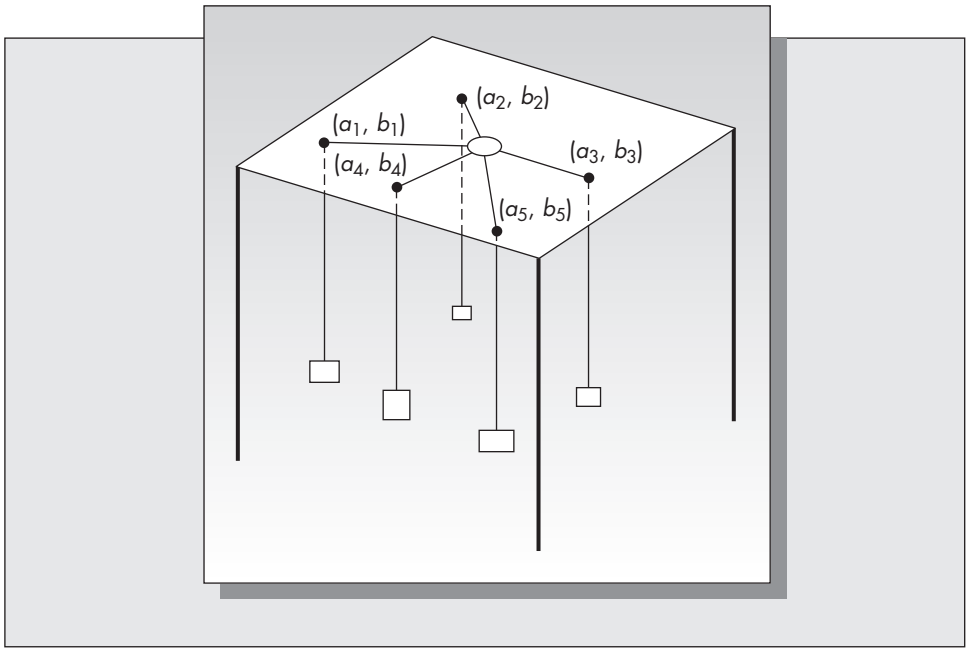


FIGURE 11–31
Solution of the Euclidean distance problem using a physical model

simple algebraic solutions; all existing methods require an iterative procedure. We will describe a procedure that appears in Francis and White (1974) that will yield an optimal solution as long as the location of the new facility does not overlap with the location of an existing facility.

Define

$$g_i(x, y) = \frac{w_i}{\sqrt{(x - a_i)^2 + (y - b_i)^2}}.$$

Let

$$x = \frac{\sum_{i=1}^{n} a_i g_i(x, y)}{\sum_{i=1}^{n} g_i(x, y)},$$

$$y = \frac{\sum_{i=1}^{n} b_i g_i(x, y)}{\sum_{i=1}^{n} g_i(x, y)}.$$

The procedure is as follows: Begin the process with an initial solution (x_0, y_0). The gravity solution is generally recommended to get the method started. Compute the values of $g_i(x, y)$ at this solution and determine new values of x and y from the given formulas. Then, recompute $g_i(x, y)$ using the new values of x and y, giving rise to yet another pair of (x, y) values. Continue to iterate in this fashion until the values of the coordinates converge. This procedure will give an optimal solution as long as the values of (x, y) at each iteration do not correspond to some existing location. The problem is that if we substitute $x = a_i$ and $y = b_i$, $g_i(x, y)$ is undefined, because the denominator is zero. It is unlikely that this will happen, but if it does, a modification of this method is required (see Francis and White, 1974).

Example *11.8* We will solve the university's location problem assuming a Euclidean distance measure. To get the procedure started we use the solution to the gravity problem, which, to three decimals, is $(x_0, y_0) = (7.622, 10.775)$. The sequence of (x, y) values obtained from iterating the given equations for x and y is

$$(x_1, y_1) = (7.895, 11.432), \qquad (x_5, y_5) = (8.429, 11.880),$$
$$(x_2, y_2) = (8.102, 11.715), \qquad (x_6, y_6) = (8.481, 11.889),$$
$$(x_3, y_3) = (8.251, 11.822), \qquad (x_7, y_7) = (8.518, 11.894),$$
$$(x_4, y_4) = (8.356, 11.863), \qquad (x_8, y_8) = (8.545, 11.899).$$

Clearly, these values are beginning to converge. Continuing with the iterations one eventually reaches the optimal solution, which is (8.621, 11.907).

Problems for Section 11.10

39. For the situation described in Problem 28,

    a. Find the gravity solution.

    b. Use the answer obtained in part (*a*) as an initial solution for the Euclidean distance problem. Determine (x_i, y_i) iteratively, for $1 \leq i \leq 5$ if you are solving

the problem by hand, or for $1 \leq i \leq 20$ if you are solving with the aid of a computer. Based on your results, estimate the final solution.

40. Three existing facilities are located at (0, 0), (5, 5), and (10, 10). The weights applied to these facilities are 1, 2, and 3, respectively. Find the location of a new facility that minimizes the weighted Euclidean distance to the existing facilities.
41. A telecommunications system is going to be installed at a company site. The switching center that governs the system must be located in one of the five buildings to be served. The goal is to minimize the amount of underground cabling required. In which building should the switching center be located? Assume a straight-line distance measure.

Building	Location
1	(10, 10)
2	(0, 4)
3	(6, 15)
4	(8, 20)
5	(15, 0)

11.11 OTHER LOCATION MODELS

The particular models discussed in Sections 11.9 and 11.10 have been referred to as "planar location models" by Francis, McGinnis, and White (1983). As they noted, there are seven assumptions associated with these models of location problems:

1. A plane is an adequate approximation of a sphere.
2. Any point on the plane is a valid location for a facility.
3. Facilities may be idealized as points.
4. Distances between facilities are adequately represented by planar distance.
5. Travel costs are proportional to distance.
6. Fixed costs are ignored.
7. Issues of distribution can be ignored.

Unless distances separating facilities are extremely large, assumption (1) should be a reasonable approximation of reality. Assumption (2) can be very restrictive. In some circumstances, only a small number of feasible locations exist, in which case the simplest approach is to evaluate the cost of locating the new facility at each of these locations and choose the one with the lowest cost. When the number of feasible alternatives is large, one could construct contour lines and consider only those locations along a given contour line with an acceptably low cost.

Depending upon the nature of the problem, assumption (3) may or may not be problematic. For example, if one uses location models to solve factory layout problems, then the size of the facilities becomes an issue. However, for a problem such as locating warehouses nationwide, idealizing facilities as points is reasonable. Assumption (4) requires that one use a particular measure of distance to compare various configurations. However, one distance measure may not be sufficient to explain all facility interactions. For example, rectilinear distances are typically used in problems in which facilities are to be located in cities. However, closed roads, rivers, or unusual street patterns could make this assumption inaccurate.

The final three assumptions deal with issues that typically arise in distribution problems. Transportation costs depend on the terrain: It is more expensive to transport goods over mountains than on flat interstate highways, for example. Fixed costs of transporting goods can be substantial depending on the mode of transportation. The fixed cost of transporting goods by air, for example, is very high.

More complex location models exist that deal with several of these shortcomings. We briefly review these in the following.

Locating Multiple Facilities

Section 11.10 treated the problem of locating a single facility among n existing facilities. However, applications exist in which the goal is to locate multiple facilities among n existing facilities. For example, a nationwide consumer products producer might be considering where to locate five new regional warehouses.

In some circumstances, multifacility location problems can be solved as a sequence of single location problems. That is, the optimal locations of the new facilities can be determined one at a time. However, when there is any interaction among the new facilities, this approach will not work.

This section will show how linear programming can be used to solve the multiple-facility rectilinear distance location problem. Assume that existing facilities have locations at points $(a_1, b_1), \ldots, (a_n, b_n)$ as in Section 11.10. Suppose that m new facilities are to be located at $(x_1, y_1), \ldots, (x_m, y_m)$. Then the objective function to be minimized may be written in the form

$$\text{Minimize } f_1(\mathbf{x}) + f_2(\mathbf{y}),$$

where

$$f_1(\mathbf{x}) = \sum_{1 \le j < k \le m} v_{jk} |x_j - x_k| + \sum_{j=1}^{m} \sum_{i=1}^{n} w_{ij} |x_j - a_i|$$

and

$$f_2(\mathbf{y}) = \sum_{1 \le j < k \le m} v_{jk} |y_j - y_k| + \sum_{j=1}^{m} \sum_{i=1}^{n} w_{ij} |y_j - b_i|.$$

The v_{jk} measure the interaction of new facilities j and k. It is the presence of these terms that prevents us from solving the multifacility problem as a sequence of one-facility problems. However, as with the single-facility problem, the optimal x and y locations may be determined independently. We present the linear programming formulation for finding the optimal x coordinates. The optimal y coordinates may be found in the same way.

The trick in transforming the problem of finding $\mathbf{x}$ to minimize $f_1(\mathbf{x})$ is to eliminate the absolute value function from the objective, as this is not a strictly linear function. The means for doing so is a standard trick in linear programming. (A similar technique was applied in the linear programming formulation of the aggregate planning problem in Section 3.5.)

For any constants a and b write $|a - b| = c + d$, but require that $cd = 0$ (either c or d or both must be zero). If $a > b$, then $|a - b| = c$, and if $a < b$, then $|a - b| = d$. Hence, we may think of c as the positive part of $a - b$ and d as the negative part of $a - b$. Substituting $|x_j - x_k| = c_{jk} + d_{jk}$ and $|x_j - a_i| = e_{ij} + f_{ij}$, we obtain the linear programming formulation of the problem of determining $\mathbf{x}$:

$$\text{Minimize } \sum_{1 \le j < k \le m} v_{jk} (c_{jk} + d_{jk}) + \sum_{j=1}^{m} \sum_{i=1}^{n} w_{ij} (e_{ij} + f_{ij})$$

subject to

$$
\begin{aligned}
&x_j - x_k - c_{jk} + d_{jk} = 0, && 1 \le j < k \le n; \\
&x_j - a_i - e_{ij} + f_{ij} = 0, && 1 \le i \le n, \qquad 1 \le j \le n; \\
&c_{jk} \ge 0, \qquad d_{jk} \ge 0, && 1 \le j < k \le n; \\
&e_{ij} \ge 0, \qquad f_{ij} \ge 0, && 1 \le i \le n, \qquad 1 \le j \le n; \\
&x_j \text{ unrestricted in sign.}
\end{aligned}
$$

We do not need to explicitly include the constraints $c_{jk}d_{jk} = 0$ and $e_{ij}f_{ij} = 0$. One can show that at the minimum cost solution these relationships always hold (which is certainly fortunate as these are not linear relationships). One additional substitution is necessary prior to solving the problem. Since linear programming codes require that all variables be nonnegative, we must substitute $x_j = x_j^+ - x_j^-$, where $x_j^+ \ge 0$ and $x_j^- \ge 0$. Commercial linear programming codes are based on the Simplex Method, which is extremely efficient. One can solve realistically sized problems easily even on a personal computer.

Multifacility gravity problems require the solution of a system of linear equations, so that gravity problems involving large numbers of facilities are easily solved as well. Multifacility Euclidean problems are solved by utilizing a multidimensional version of the iterative solution method described in Section 11.10. We will not review these methods here. The interested reader should refer to Francis and White (1974).

Further Extensions

Facilities Having Positive Areas

All previous models assumed that facilities are approximated by points in the plane. When the area of the facilities is small compared to the area covered by the available locations, this assumption is reasonable. However, in certain applications the areas of the facilities cannot be ignored. For example, when finding locations for machines in a job shop, the machines must be far enough apart for them to be able to operate efficiently. Tompkins and White (1984) present an approach that requires the assumptions that the facilities are rectangular in shape and that the weights are uniformly distributed over the areas. The method is based on developing an analogy between the location problem and the problem of locating forces on a beam and is similar to the procedure for constructing contour lines discussed in Appendix 11–B.

Location–Allocation Problems

Often the decision of where to locate new facilities must be accompanied by the decision of which of the existing locations will be served by each new facility. For example, a firm may be considering where to locate several regional warehouses. In addition to determining the location and the number of these new warehouses, the firm also must decide which of the retail outlets will be serviced by which warehouses.

Location–allocation problems are difficult to solve owing to the large number of decision variables. The mathematical programming formulation of the problem, assuming that a rectangular distance measure is used, is

$$\text{Minimize} \sum_{j=1}^{m} \sum_{i=1}^{n} z_{ij} w_{ij} [|x_j - a_i| + |y_j - b_i|] + g(m)$$

subject to

$$\sum_{j=1}^{m} z_{ij} = 1 \qquad \text{for } 1 \le i \le n,$$

where

w_{ij} = Cost per unit time per unit distance if the existing facility i is serviced by new facility j;

$$z_{ij} = \begin{cases} 1 & \text{if existing facility } i \text{ is serviced by new facility } j, \\ 0 & \text{otherwise;} \end{cases}$$

m = Total number of new facilities, $1 \leq m \leq n$;

(x_j, y_j) = Coordinates of new facility j, $1 \leq j \leq m$;

(a_i, b_i) = Coordinates of existing facility i, $1 \leq i \leq n$;

$g(m)$ = Cost per unit time of providing m new facilities.

This problem formulation has decision variables m, the number of new facilities; z_{ij}, the specification of which of the existing locations i will be serviced by facility j; and (x_j, y_j), the location of the new facilities. The optimization is difficult due to the presence of the zero-one variables z_{ij} and the inclusion of m as a decision variable. The problem is typically solved by considering successive values of $m = 1, 2, \ldots$, and enumerating all combinations of z_{ij} for each value of m. Given a fixed m and set of z_{ij} values, the solution can be obtained using the methods for locating multiple facilities discussed earlier in this section. However, the number of different z_{ij} values grows quickly as a function of m, so only moderately sized problems can be solved in this fashion.

Discrete Location Problems

The models considered in this chapter for location of new facilities assumed that the new facilities could be located anywhere in the plane. This is not the case for most applications. Contour lines assist with evaluating alternative locations but cannot be constructed for problems in which one must locate multiple facilities. An alternative approach is to restrict a priori the possible locations to some discrete set of possibilities. When there is only a single facility and the number of possible locations is small, the easiest approach is to evaluate the cost of each location and pick the smallest.

When there are multiple facilities, the assignment model discussed in Section 11.4 can be used to determine the optimal locations of the new facilities. In certain types of warehouse-layout problems, new facilities can take up more than one potential site. For example, suppose that we must determine in which locations in a warehouse to store k items. Suppose that the appropriate storage area in the warehouse is composed of n grid squares and each item stored takes up more than a single square. Each square would be numbered and the storage location of an item specified by the numbers of the grid squares covered by the item. The resulting model, discussed in Francis and White (1974), is a generalization of the simple assignment model appearing in Section 11.4. We will not present the details of the model here.

Network Location Models

Planar location models assume that the goal is to locate one or more new facilities in order to minimize some function of the distance separating the new and the existing facilities. A rectilinear, Euclidean, or other distance measure is generally assumed. In certain applications, the distances should be measured over an existing network and are not accurately approximated by standard measures. Overland transport must follow road networks, water transport must follow shipping lanes and sea routes, and air transport is confined to predetermined air corridors. In other applications the network may correspond to a network of power cables or telephone wires. In many of these applications the new facility or facilities must be placed on or very near to a location on the network, and

distances can be measured only in terms of the network. Network location models are beyond the scope of our coverage. The interested reader should refer to the review articles of Francis, McGinnis, and White (1983) and Tansel, Francis, and Lowe (1983).

International Issues

The problem of locating facilities is a part of the larger issue of global supply chain management. It is truer and truer that businesses are evolving into global corporations. This is no longer just the case for the industry giants. Globalization now plays a greater role than ever before, and its importance will continue to grow. Arntzen et al. (1995) discuss the supply chain configuration for Digital Equipment Corporation. In one case, they show how various parts of computers are shipped from the United States to Europe, from Europe to Brazil and Taiwan, and from both Taiwan and China to Europe and back to the United States. Fabrication may be done in two or three different countries, and distribution networks may be equally complex.

Lower wage rates were traditionally the primary reason for firms based in developed countries to locate plants in less developed countries. This is certainly true today. General Motors does much of its auto assembly in Mexico. Virtually all the large semiconductor manufacturers have fabrication facilities overseas, typically in places like Malaysia and the Philippines.

Cohen and Lee (1989) provide a good overview of some of the issues that one must take into account when locating facilities in other countries. They include

1. Duties and tariffs are based on material flows. Their impact must be incorporated into international shipping schedules of materials, intermediate product, and finished product.
2. Currency exchange rates fluctuate unpredictably and affect profit levels in each country.
3. Corporate tax rates vary considerably from country to country.
4. Global sourcing must take into account lead times, costs, new technologies, and dependence on particular countries.
5. Local content rules and quotas constrain material flow between countries.
6. Product designs may vary by national market.
7. Transfer price mechanisms must be put in place to take the place of centralized control.
8. Differences in language, cultural norms, education, and skills must be incorporated into location decisions.

Only recently have we begun to understand the complexity of the problem of locating new facilities and their effect on global supply chain operations. Well-constructed and well-thought-out mathematical models will continue to assist us in managing these increasingly complex networks. However, many of the issues alluded to in this section are difficult to quantify, thus making good judgment crucial.

Problems for Section 11.11

42. For each of the location problems described, discuss which of the seven assumptions listed in this section are likely to be violated:
    a. Locating three new machines in a machine shop.
    b. Locating an international network of telecommunications facilities.
    c. Locating a hospital in a sparsely populated area.
    d. Locating spare parts depots to support a field repair organization.

43. Consider Problem 28 of this chapter. Suppose that two new machines, A and B, are to be located in the shop. Machine A has $\frac{1}{8}$, $\frac{1}{8}$, $\frac{1}{4}$, 1, and $\frac{1}{6}$ as the expected numbers of loads transported to the existing five machines, respectively, and machine B has $\frac{1}{4}$, $\frac{1}{6}$, 3, $\frac{1}{5}$, and $\frac{1}{2}$ as the expected numbers of loads transported, respectively. Furthermore, suppose that there are two loads per hour on average transported between the new machines. Assume a rectilinear distance measure.

    a. Formulate the problem of determining the optimal locations of the new machines as a linear program.

    b. If you have access to a computerized linear programming code, solve the problem formulated in part (*a*).

44. Consider the University of the Far West described in Example 11.5. Suppose that the university administration has decided that two audiovisual centers are needed. Each center would have different facilities. The anticipated numbers of faculty members using each center are

School	Faculty Members Using Center A	Faculty Members Using Center B
Business	13	18
Education	40	23
Engineering	24	17
Humanities	20	23
Law	30	9
Science	16	21

    Furthermore, there will be a total of 16 staff persons at centers A and B. They will need to interact frequently.

    a. Formulate the problem of determining the optimal locations of the two audiovisual centers as a linear program. Assume that rectilinear distances are used throughout.

    b. If you have access to a computerized linear programming code, solve the problem formulated in part (*a*).

45. Describe the following location problems and how they differ from those previously treated in this chapter:

    a. Location–allocation problems

    b. Discrete location problems

    c. Network location problems

11.12 HISTORICAL NOTES

The problems discussed in this chapter have a long history. Determining suitable layouts for production facilities is a problem that dates back to the start of the industrial revolution, although it appears that the development of analytical techniques for finding layouts is recent. Little seems to have been published concerning analytical layout methods prior to 1950. Apple (1977) lists a number of texts dealing with the plant layout problem published in the early 1950s. The computerized layout techniques discussed in this chapter were developed in the 1960s and 1970s. CRAFT, one of the

first computerized methods and most popular even today, is from Buffa, Armour, and Vollmann (1964). ALDEP is from Seehof and Evans (1967), CORELAP from Lee and Moore (1967), COFAD from Tompkins and Reed (1976), and PLANET from Deisenroth and Apple (1972). Both ALDEP and CRAFT are available from the IBM Corporation as part of its SHARE library.

Some of the location problems discussed in this chapter go back hundreds of years. The problem of finding the location of a single new facility to minimize the sum of the Euclidean distances to the existing facilities has been referred to as the Steiner–Weber problem or the general Fermat problem. Francis and White (1974) state that the problem with exactly three facilities was posed by Fermat and solved by the mathematician Torricelli prior to 1640. The work on the rectilinear distance problem is relatively recent and was sparked by a paper by Hakimi (1964). Research continues today on discovering efficient solution techniques for locating multiple facilities using various distance measures.

11.13 Summary

This chapter dealt with two important logistics problems: the most efficient layout of facilities and the best location of new facilities relative to the existing ones. In a sense, the layout problem is a special type of location problem, because the goal is to find the best location of facilities within a specified boundary.

The analytical methods for layout discussed in this chapter assume that the objective is to minimize some function of the distance separating facilities. This viewpoint is probably most appropriate for plant layout problems and less so for other problems in which qualitative factors play a greater role. Two charts are important for layout analysis: the *activity relationship chart* (or rel chart for short) and the *from-to chart.* In order to construct a rel chart, each pair of facilities is given a letter code A (absolutely necessary), E (especially important), I (important), O (ordinary importance), U (unimportant), or X (undesirable), representing the desirability of locating facilities near each other. A from-to chart may specify the distance between pairs of facilities, numbers of materials handling trips per unit time between facilities, or the cost of materials handling trips between facilities. Both rel charts and from-to charts are useful for evaluating the quality of a layout.

Section 11.3 discussed types of layouts. The two most common types are *product layouts* and *process layouts.* The product layout is usually a fixed transfer line arranged in the sequence of the manufacturing steps required. A process layout groups machines with similar functions. Part routings vary from product to product. The product layout is appropriate for high-volume production of a small number of products. The process layout is appropriate for a low-volume job-shop environment. *Fixed position layouts* are used for products that are too large to move. Recently, there has been considerable interest in layouts based on *group technology.* Parts are grouped into families, and machine cells are developed consistent with this grouping. Group technology layouts are appropriate for automated factories.

The *assignment model* can be used for solving relatively simple layout and location problems. In order to use the assignment algorithm, we assume that for each placement of a machine (say) in a location, we can evaluate the cost of that assignment. This assumes that there is no interaction between the machines. When interaction does occur, a *quadratic assignment* formulation exists, but quadratic assignment models are far more difficult to solve than simple assignment models.

We discussed five *computerized layout* techniques: CRAFT, COFAD, ALDEP, PLANET, and CORELAP. Both CRAFT and COFAD are improvement routines. That means that both require that the user specify an initial layout. The program proceeds to consider interchanging adjacent pairs of facilities in order to achieve an improvement. On the other hand, ALDEP, PLANET, and CORELAP are construction routines, which build the layout from

scratch. Because construction routines often result in departments with odd shapes, improvement routines are generally preferred. We also noted that there are a host of new software products now available for the personal computer and for UNIX workstations. Many of these products are based on drafting programs with large libraries of graphical icons. These products use methods conceptually similar to the earlier programs previously mentioned, but are far more user friendly.

The chapter included a discussion of *flexible manufacturing systems* (FMSs). An FMS is a collection of machines linked by an automated materials handling system and is generally controlled by a central computer. FMSs are used primarily in the metal-working industries and are an appropriate choice when there is medium to large volume and a moderate variety of part types required. The downside to these systems is cost, which can run as high as $10 million or more. Flexible manufacturing cells are a scaled-down lower-cost alternative.

The second part of the chapter was concerned with methods for locating new facilities. *Location models* are appropriate when locating one or more new facilities within a specified area already containing a finite number of existing facilities. The objective is to locate new facilities to minimize some function of the distance separating new and existing facilities. Three distance measures were considered: rectilinear, Euclidean, and squared Euclidean. The first two are the most common for describing real problems and depend on whether movement occurs according to a crisscross street pattern (rectilinear) or is measured by straight-line distances (Euclidean).

The optimal solution to the weighted rectilinear distance problem is to locate the (x, y) coordinates at the *median* of the existing coordinates. When using a squared Euclidean distance measure, the optimal location of the new facility is at the center of gravity of the existing coordinates. No simple algebraic solution for the Euclidean distance problem is known, but iterative solution techniques exist. The chapter also included a brief discussion of several more complex location problems, including location of multiple facilities, location of facilities having nonzero areas, location–allocation problems, discrete location problems, and network location models. Finally, the chapter concluded with a discussion of issues of concern when locating new facilities in other countries. As globalization of manufacturing and supply chain networks increases, these issues will play an even greater role in location planning.

Additional Problems on Layout and Location

46. A real estate firm wishes to open four new offices in the Boston area. There are six potential sites available. Based on the number of employees in each office and the location of the properties that each employee will manage, the firm estimated the total travel time in hours per day for each office and each location. Find the optimal assignment of offices to sites to minimize employee travel time.

		Offices			
		A	**B**	**C**	**D**
Sites	**1**	10	3	3	8
	2	13	5	2	6
	3	12	9	9	4
	4	14	2	7	7
	5	17	7	4	3
	6	12	8	5	5

47. A large supermarket chain in the Southeast requires five additional warehouses in the Atlanta area. It has identified five sites for these warehouses. The annual transportation costs (in $000) for each warehouse at each site are given in the following table. Find the assignment of warehouses to sites to minimize the total annual transportation costs.

		Warehouses				
		A	**B**	**C**	**D**	**E**
Sites	**1**	41	47	38	46	50
	2	39	37	42	36	45
	3	43	46	45	42	46
	4	51	54	47	58	56
	5	44	40	42	41	45

48. A machine shop located on the outskirts of Los Angeles accepts custom orders from a number of high-tech firms in southern California. The machine shop consists of four departments: A (lathes), B (drills), C (grinders), and D (sanders). The from-to chart showing distances in feet between department centers is given here.

		To department			
		A	**B**	**C**	**D**
From department	**A**		45	63	32
	B	29		27	46
	C	63	75		68
	D	40	30	68	

The shop has accepted orders for production of four products: P1, P2, P3, and P4. The routing for production of these products and the weekly production rates are

Product	Routing	Weekly Production
P1	A–B–C–D	200
P2	A–C–D	600
P3	B–D	400
P4	B–C–D	500

Assume that products are produced in batches of size 25.

a. Convert this information into a from-to chart giving numbers of materials handling trips per week between departments.

b. If the cost to transport one batch 1 foot is estimated to be $1.50, convert the from-to chart you found in part (*a*) to one giving the materials handling cost per week between departments.

c. Develop an activity relationship chart for these four departments based on the results of part (*b*). Assume that *A* is assigned to the highest cost and *O* to the least, with the rankings *E* and *I* assigned to the costs falling in between the extremes.

d. Suppose that the machine shop is located in a building that is 60 feet by 80 feet. Furthermore, suppose that departments are rectangularly shaped with the

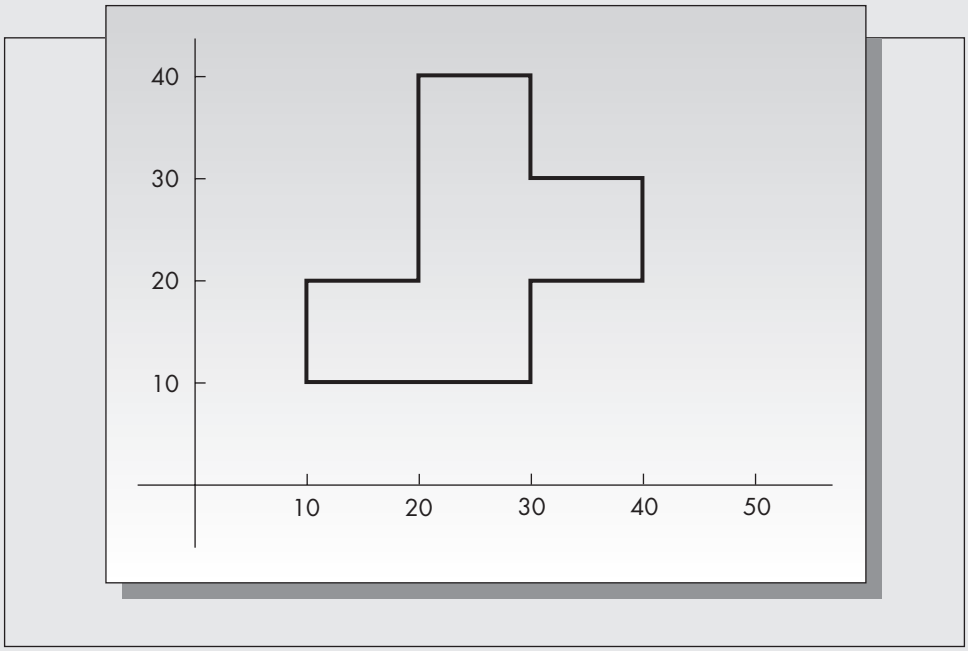


FIGURE 11–32
Shape of facility (for Problem 49)

following dimensions:

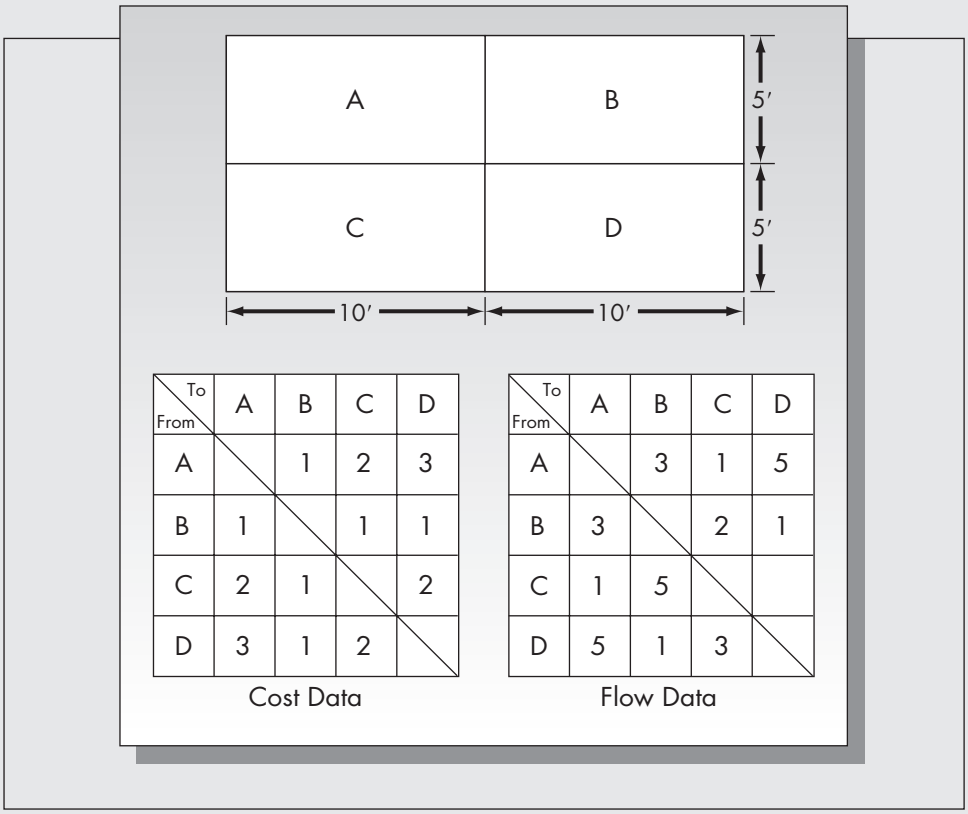
Department	Dimensions
A	20 × 30
B	40 × 20
C	45 × 55
D	37 × 25

Sketch a layout consistent with the rel chart you obtained in part (*c*).

49. A facility has the shape shown in Figure 11–32. Using the methods described in Appendix 11–A, find the location of the centroid of this facility.
50. An initial layout for four departments and from-to charts giving distances separating departments and unit transportation costs appear in Figure 11–33. Using the CRAFT pairwise exchange technique, find the layout recommended by CRAFT to minimize total materials handling costs.
51. An initial layout for five departments and a from-to flow data chart are given in Figure 11–34. Assuming that departments A and D are in fixed locations and cannot be moved, find the layout recommended by CRAFT for departments B, C, and E. Assume that the objective is to minimize the total distance traveled.
52. Consider the rel chart for the Meat Me fast-food restaurant, given in Figure 11–2. Assume that the areas required for each department are

Department	Area Required (square feet)
Cooking burgers	200
Cooking fries	200
Packing and storing	100
Drink dispensers	100
Counter servers	400
Drive-up server	100

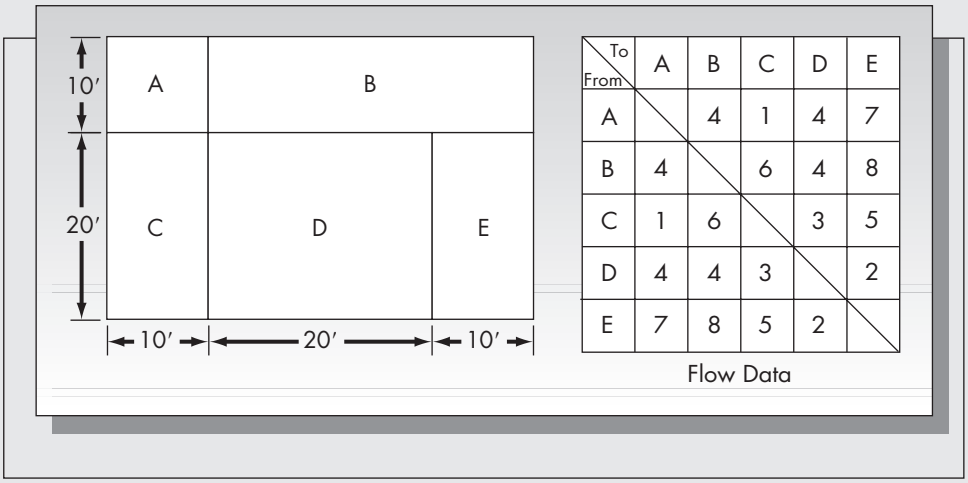
FIGURE 11–33
Layout and from-to charts (for Problem 50)



To / From	A	B	C	D
A		1	2	3
B	1		1	1
C	2	1		2
D	3	1	2	

To / From	A	B	C	D
A		3	1	5
B	3		2	1
C	1	5		
D	5	1	3	

FIGURE 11–34
Layout and from-to chart (for Problem 51)



To / From	A	B	C	D	E
A		4	1	4	7
B	4		6	4	8
C	1	6		3	5
D	4	4	3		2
E	7	8	5	2	

Assume a sweep width of 2 squares and facility dimensions of 5 by 9 squares, where each square is 5 feet on a side. As a result, for example, the cooking-burgers department requires 8 squares. Use the ALDEP approach to develop a layout for the restaurant. Comment on the practicality of your results.

53. Frank Green, an independent TV repairman, is considering purchasing a home in Ames, Iowa, that he will use as a base of operations for his repair business. Frank's primary sources of business are 10 industrial accounts located throughout the Ames

area. He has overlaid a grid on a map of the city and determined the following locations for these clients as well as the expected number of calls per month he receives:

Client	Grid Location	Expected Calls per Month
1	(5, 8)	2
2	(10, 3)	1
3	(14, 14)	1
4	(2, 2)	3
5	(1, 17)	1
6	(18, 25)	$\frac{1}{2}$
7	(14, 3)	$\frac{1}{4}$
8	(25, 4)	4
9	(35, 1)	3
10	(16, 21)	$\frac{1}{6}$

Find the optimal location of his house, assuming

a. A weighted rectilinear distance measure.

b. A squared Euclidean distance measure.

c. The goal is to minimize the maximum rectilinear distance to any client.

54. An electronics firm located near Phoenix, Arizona, is considering where to locate a new phone switch that will link five buildings. The buildings are located at (0, 0), (2, 6), (10, 2), (3, 9), and (0, 4). The objective is to locate the switch to minimize the cabling required to those five buildings.

    a. Determine the gravity solution.

    b. Determine the optimal location assuming a straight-line distance measure. (If you are solving this problem by hand, iterate the appropriate equations at least five times and estimate the optimal solution.)

55. Consider three locations at (0, 0), (0, 6), and (3, 3) with equal weights. Using the methods described in Appendix 11–B, find contour lines for the rectilinear location problem.

56. A company is considering where to locate its cafeteria to service six buildings. The locations of the buildings and the fraction of the company's employees working at these locations are

Building	a_i	b_i	Fraction of Workforce
A	2	6	$\frac{1}{12}$
B	1	0	$\frac{1}{12}$
C	3	3	$\frac{1}{6}$
D	5	9	$\frac{1}{4}$
E	4	2	$\frac{1}{4}$
F	10	7	$\frac{1}{6}$

    a. Find the optimal location of the cafeteria to minimize the weighted rectilinear distance to all the buildings.

    b. Find the optimal location of the cafeteria to minimize the maximum rectilinear distance to all the buildings.

c. Find the gravity solution.

d. Suppose that the cafeteria must be located in one of the buildings. In which building should it be located if the goal is to minimize weighted rectilinear distance?

e. Solve part (*d*) assuming weighted Euclidean distance.

Spreadsheet Problems for Chapter 11

57. Design a spreadsheet to compute the total rectilinear distance from a set of up to 10 existing locations to any other location. Assume that existing locations are placed in Columns A and B and the new location in cells D1 and E1. Initialize column A with the value of cell D1 and column B with the value of cell E1, so that the total distance will be computed correctly when there are fewer than 10 locations.

    a. Suppose that existing facilities are located at (0, 0), (5, 15), (110, 120), (35, 25), (80, 10), (75, 20), (8, 38), (50, 65), (22, 95), and (44, 70), and the new facility is to be located at (50, 50). Determine the total rectilinear distance of the new facility to the existing facilities.

    b. Suppose the new facility in part (*a*) can be located only at $x = 0, 5, 10, \ldots, 100$ and $y = 0, 10, 20, \ldots, 100$. By systematically varying the x and y coordinates, find the optimal location of the new facility.

58. Solve Problem 57, assuming a Euclidean distance measure.

59. Solve Problem 32 using an electronic spreadsheet. To do so, enter the building numbers in column A, the x coordinates in column B, and the associated weights in column C. Sort columns A, B, and C in ascending order by using column B as the primary sort key. Now accumulate the weights (column C) in column D using the sum function. Divide the total accumulated weight by 2 and visually identify the optimal x coordinate value. It will be where the cumulative weight first exceeds half the total cumulative weight. Repeat the process for the y coordinates.

60. Design an electronic spreadsheet to compute the optimal solution to the gravity problem. Allow for up to 20 locations. Let column A be the location number, column B the x coordinates $(a_1, \ldots, a_n)$ of existing locations, and column C the y coordinates $(b_1, \ldots, b_n)$ of existing locations. Store the optimal solution in cell D1. Find the gravity solution to Problem 32 using your spreadsheet.

61. Extend the results of Problem 60 to find the optimal location of a new facility among a set of existing facilities assuming a straight-line Euclidean distance measure. Let column E correspond to $g_i(x, y)$, where the initial (x, y) values appear in F1 and F2. In locations G1 and G2, store

$$x = \frac{\sum_{i=1}^{n} a_i g_i(x, y)}{\sum_{i=1}^{n} g_i(x, y)}.$$

$$y = \frac{\sum_{i=1}^{n} b_i g_i(x, y)}{\sum_{i=1}^{n} g_i(x, y)}.$$

Start with the gravity solution in cells F1 and F2. After calculation, replace the values in cells F1 and F2 with the values in cells G1 and G2. Continue in this manner until the solution converges. Using your spreadsheet,

a. Solve Problem 32.
b. Solve Problem 41.
c. Solve Problem 53 assuming a Euclidean distance measure.
d. Solve Problem 56 assuming a Euclidean distance measure.

Appendix **11–A**

Finding Centroids

The centroid of any object is another term for the physical coordinates of the center of gravity. For a plate of uniform density, it would be the point at which the plate would balance exactly. Let R be any region in the plane. The centroid for R is defined by two points $\bar{x}, \bar{y}$. In order to find these two points, we first must obtain the moments of R, M_x, and M_y, which are given by the formulas

$$M_x = \int_R \int x \, dx \, dy,$$
$$M_y = \int_R \int y \, dx \, dy.$$

Let $A(R)$ be the area of R. Then the centroid of R is given by

$$\bar{x} = \frac{M_x}{A(R)}, \qquad \bar{y} = \frac{M_y}{A(R)}.$$

We now obtain explicit expressions for the moments when R is a finite sum of rectangles. Suppose that R is a simple rectangle as pictured in Figure 11–35. Then

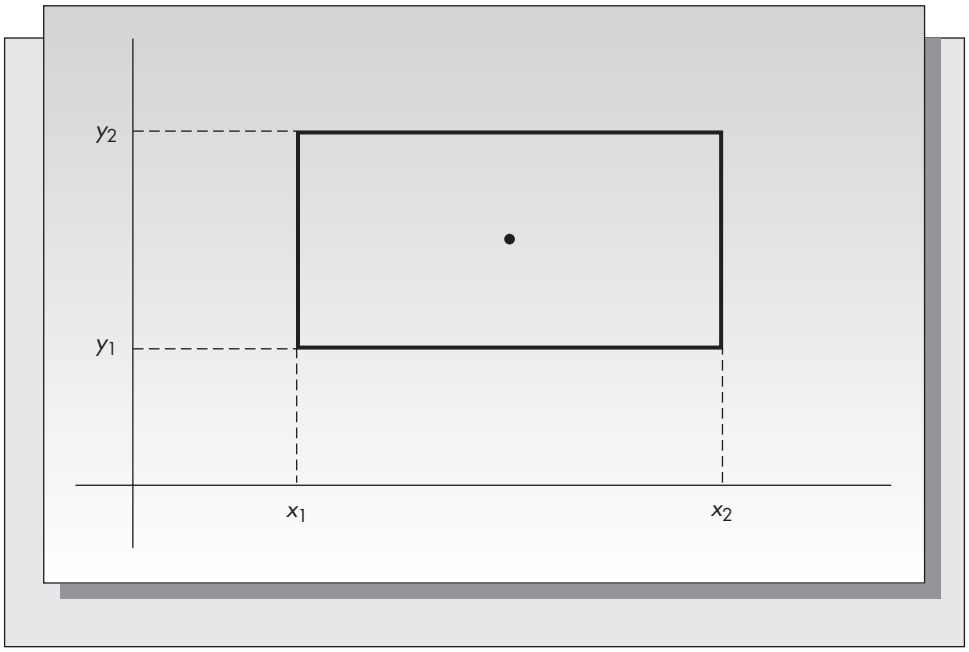
$$M_x = \int_{y_1}^{y_2} dy \int_{x_1}^{x_2} x \, dx = \int_{y_1}^{y_2} dy \, \frac{x^2}{2}\bigg|_{x_1}^{x_2}$$
$$= \int_{y_1}^{y_2} \frac{dy(x_2^2 - x_1^2)}{2} = \frac{x_2^2 - x_1^2}{2}(y_2 - y_1),$$
$$M_y = \int_{x_1}^{x_2} dx \int_{y_1}^{y_2} y \, dy = \int_{x_1}^{x_2} dx \, \frac{y^2}{2}\bigg|_{y_1}^{y_2}$$
$$= \int_{x_1}^{x_2} \frac{dx(y_2^2 - y_1^2)}{2} = \frac{y_2^2 - y_1^2}{2}(x_2 - x_1).$$

Note that because the area of the rectangle is $(x_2 - x_1)(y_2 - y_1)$, and $x_2^2 - x_1^2 = (x_2 - x_1)(x_1 + x_2)$ (and similarly for $y_1^2 - y_2^2$), we obtain

$$\bar{x} = \frac{x_1 + x_2}{2}, \qquad \bar{y} = \frac{y_1 + y_2}{2}.$$

The formulas for the moments of a rectangle may be used to find the centroid when R consists of a collection of rectangles as well. Suppose that R can be subdivided into

FIGURE 11–35
The centroid of a rectangle



k rectangles labeled $R_1, R_2, \ldots, R_k$ with respective boundaries defined by $[(x_{1i}, x_{2i}), (y_{1i}, y_{2i})]$ for $1 \leq i \leq k$. Since

$$M_x = \int_R \int x \, dx \, dy = \sum_{i=1}^{k} \int_{R_i} \int x \, dx \, dy,$$

it follows that

$$M_x = \sum_{i=1}^{k} \frac{{x_{2i}}^2 - {x_{1i}}^2}{2}(y_{2i} - y_{1i}).$$

Similarly,

$$M_y = \sum_{i=1}^{k} \frac{{y_{2i}}^2 - {y_{1i}}^2}{2}(x_{2i} - x_{1i}).$$

Example *11A.1* Consider the region R pictured in Figure 11–36. We will determine the centroid of the region using the given formulas. The region can be broken down into three rectangles in a number of different ways. For the one pictured in Figure 11–36, we have

$$\begin{aligned} x_{11} &= 10, & x_{21} &= 30, \\ y_{11} &= 5, & y_{21} &= 35, \\ x_{12} &= 30, & x_{22} &= 40, \\ y_{12} &= 5, & y_{22} &= 20, \\ x_{13} &= 30, & x_{23} &= 50, \\ y_{13} &= 25, & y_{23} &= 35. \end{aligned}$$

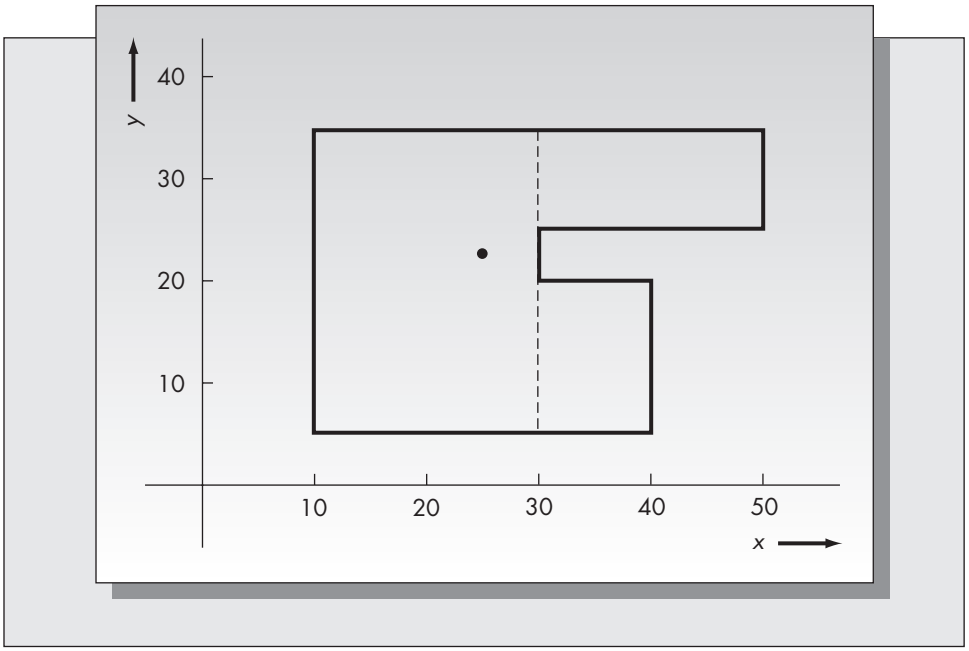


FIGURE 11–36
Centroid of a figure composed of rectangles

Substituting into the given formulas, we obtain

$$M_x = \frac{30^2 - 10^2}{2}(35 - 5) + \frac{40^2 - 30^2}{2}(20 - 5) + \frac{50^2 - 30^2}{2}(35 - 25)$$
$$= (400)(30) + (350)(15) + (800)(10) = 25{,}250,$$
$$M_y = \frac{35^2 - 5^2}{2}(30 - 10) + \frac{20^2 - 5^2}{2}(40 - 30) + \frac{35^2 - 25^2}{2}(50 - 30)$$
$$= (600)(20) + (187.5)(10) + (300)(20) = 19{,}875.$$

The total area of R is

$$A(R) = (20)(30) + (10)(15) + (20)(10) = 950.$$

It follows that the centroid is

$$\bar{x} = \frac{25{,}250}{950} = 26.579,$$
$$\bar{y} = \frac{19{,}875}{950} = 20.921.$$

The centroid is marked with a dot on Figure 11–36.

Appendix **11–B**

Computing Contour Lines

This appendix outlines the procedure for computing contour lines, or isocost lines, such as those pictured in Figures 11–28 and 11–29. The theoretical justification for this procedure appears in Francis and White (1974).

1. Plot the points $(a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n)$ on graph paper. Draw a horizontal line (parallel to the x axis) and a vertical line (parallel to the y axis) through each point.

2. Number the horizontal and vertical lines in sequence from left to right and from top to bottom. (If none of the original points is collinear, there will be exactly n horizontal and n vertical lines.)

3. Let C_i be the sum of the weights associated with the points along vertical line j, and D_i the sum of the weights associated with the points along horizontal line i.

4. Compute the following numbers:

$$M_0 = -\sum_{i=1}^{n} w_i, \qquad N_0 = M_0 = -\sum_{i=1}^{n} w_i,$$

$$M_1 = M_0 + 2C_1, \qquad N_1 = N_0 + 2D_1,$$

$$M_2 = M_1 + 2C_2, \qquad N_2 = N_1 + 2D_2,$$

and so on.

(The final values of M_i and N_j will both be $+\sum_{i=1}^{n} w_i$.)

5. Define the region (i, j) as the region bounded by the ith and $(i + 1)$th vertical lines and the jth and $(j + 1)$th horizontal lines. The regions to the left of the first vertical line are labeled $(0, j)$, and those below the first horizontal line are labeled $(i, 0)$. The slope of any contour line passing through region (i, j) is given by

$$S_{i,j} = -M_i/N_j.$$

Once the slopes are determined, a contour line is constructed by starting at any point and moving through each region at the angle determined by the slope computed in step 5. We will present a simple example to illustrate the method.

Example *11B.1*

Assume $(a_1, b_1) = (1, 1)$, $(a_2, b_2) = (5, 3)$, and $(a_3, b_3) = (3, 8)$, and $w_1 = 2$, $w_2 = 3$, and $w_3 = 6$. The first step is to plot these points on a grid as we have done in Figure 11–37. Drawing vertical and horizontal lines through each of the three points yields three vertical lines labeled 1, 2,

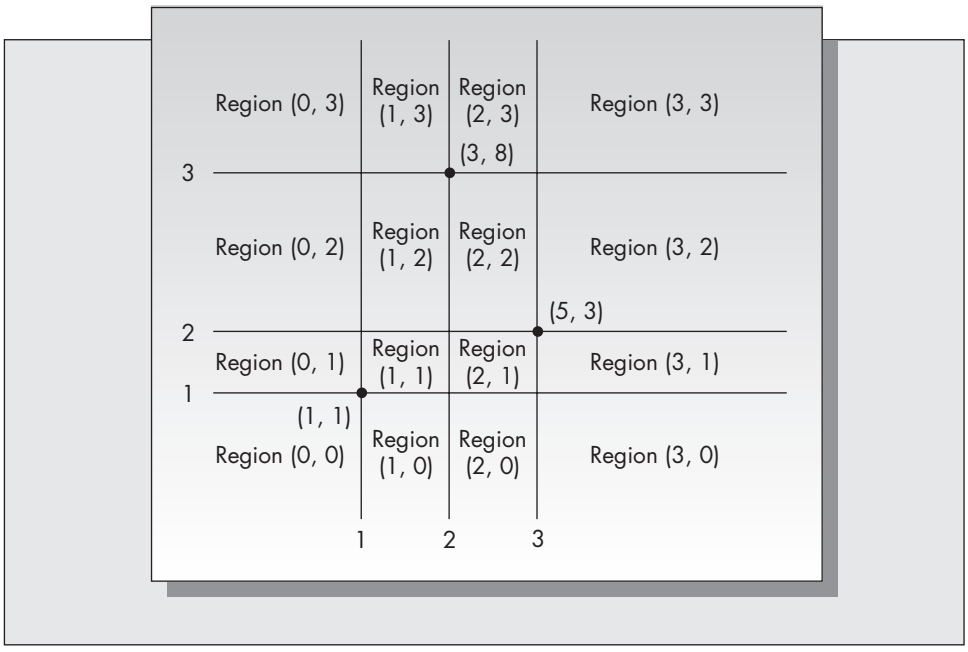


FIGURE 11–37
Regions for Example 11B.1

and 3 and three horizontal lines labeled 1, 2, 3 as well. Regions are labeled from (0, 0) to (3, 3) as in the figure.

Next we compute $M_0, \ldots, M_3$ and $N_0, \ldots, N_3$.

$$M_0 = -(2 + 3 + 6) = -11 = N_0,$$

$$M_1 = -11 + (2)(2) = -7, \quad N_1 = -11 + 2(2) = -7,$$

$$M_2 = -7 + (2)(6) = +5, \quad N_2 = -7 + (2)(3) = -1,$$

$$M_3 = +5 + (2)(3) = +11, \quad N_3 = -1 + (2)(6) = +11.$$

Next the ratios are computed to find the slope for each region.

$$S_{0,0} = -(-11)/(-11) = -1, \qquad S_{0,2} = -(-11)/(-1) = -11,$$

$$S_{1,0} = -(-7)/(-11) = -0.64, \qquad S_{1,2} = -(-7)/(-1) = -7,$$

$$S_{2,0} = -(5)/(-11) = +0.45, \qquad S_{2,2} = -(5)/(-1) = 5,$$

$$S_{3,0} = -(11)/(-11) = +1, \qquad S_{3,2} = -(11)/(-1) = 11,$$

$$S_{0,1} = -(-11)/(-7) = -1.57, \qquad S_{0,3} = -(-11)/(11) = 1,$$

$$S_{1,1} = -(-7)/(-7) = -1, \qquad S_{1,3} = -(-7)/(11) = 0.64,$$

$$S_{2,1} = -(5)/(-7) = +0.71, \qquad S_{2,3} = -(5)/(11) = -0.45,$$

$$S_{3,1} = -(11)/(-7) = +1.57, \qquad S_{3,3} = -(11)/(11) = -1.$$

Before constructing the contour lines, it is convenient to place the slopes in the appropriate regions, as shown in Figure 11–38. A contour line may be started at any point on a region boundary. From the initial point, one draws a line with the appropriate slope for that region to the boundary of the next region. At that point the slope changes to the value associated with the next region. One continues until the line segments return to the originating point. (If the slopes are correct and the drawing accurate, one will always return to the point of origination.) Two typical contour lines for the example problem are shown in Figure 11–39.

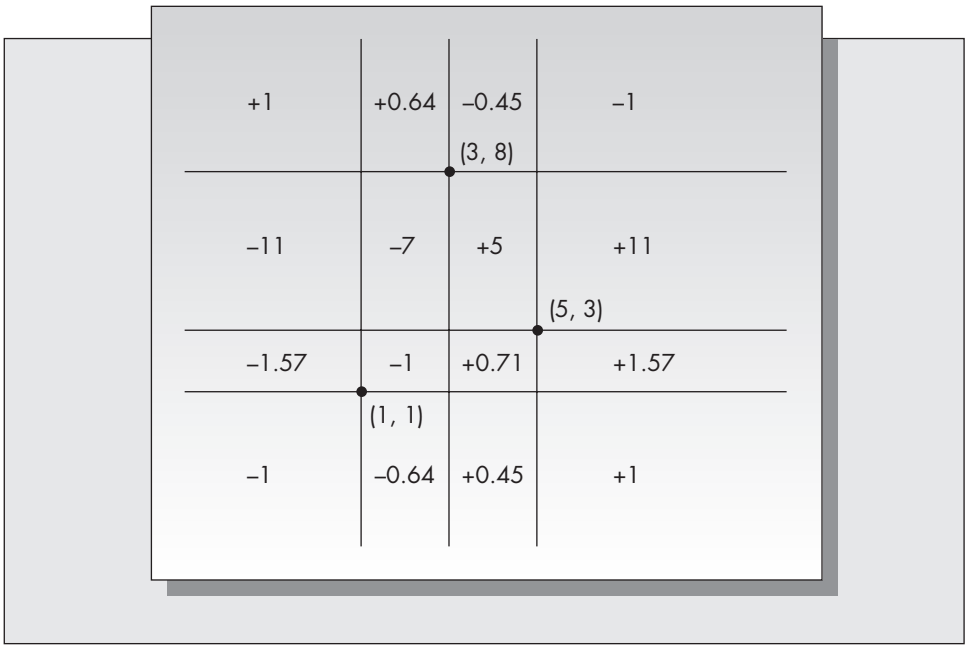
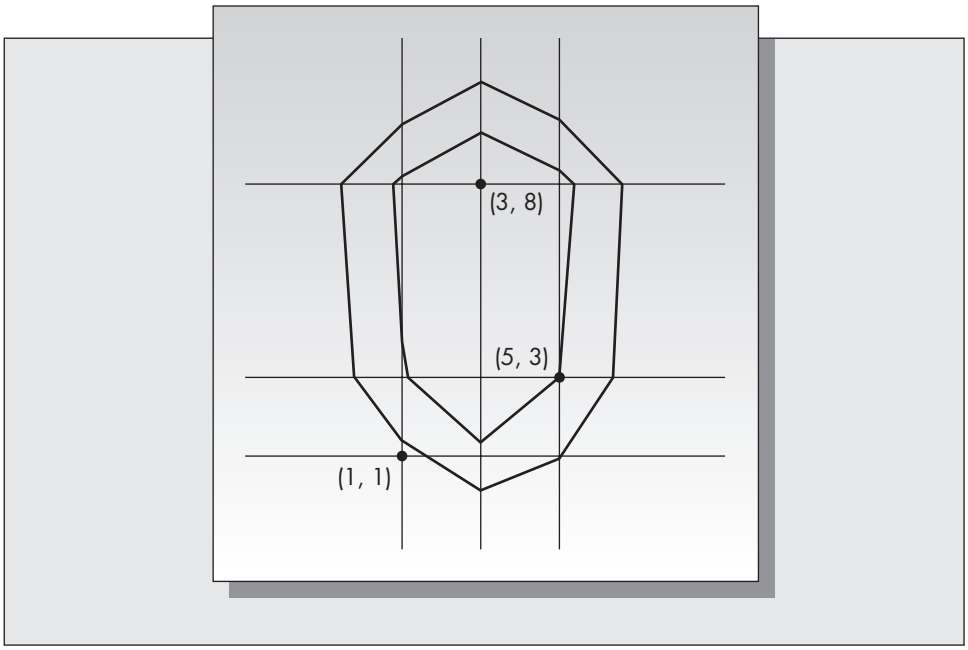


FIGURE 11–38
Slopes for Example 11B.1

FIGURE 11–39
Sample contour lines for Example 11B.1



Bibliography

Apple, J. M. *Plant Layout and Material Handling*. 3rd ed. New York: John Wiley & Sons, 1977.

Arntzen, B. C.; G. G. Brown; T. P. Harrison; and L. L. Trafton. "Global Supply Chain Management at Digital Equipment Corporation." *Interfaces* 25, no. 1 (1995), pp. 69–93.

Block, T. E. "A Note on 'Comparison of Computer Algorithms and Visual Based Methods for Plant Layout' by M. Scriabin and R. C. Vergin." *Management Science* 24 (1977), pp. 235–37.

Bruno, G., and P. Biglia. "Performance Evaluation and Validation of Tool Handling in Flexible Manufacturing Systems Using Petri Nets." In *Proceedings of the International Workshop on Timed Petri Nets,* pp. 64–71. Torino, Italy, 1985.

Buffa, E. S. "On a Paper by Scriabin and Vergin." *Management Science* 23 (1976), p. 104.

Buffa, E. S.; G. C. Armour; and T. E. Vollmann. "Allocating Facilities with CRAFT." *Harvard Business Review* 42 (1964), pp. 136–58.

Burbidge, J. L. *The Introduction of Group Technology.* London: Heinemann, 1975.

Buzacott, J. A., and D. D. Yao. "On Queuing Network Models of Flexible Manufacturing Systems." *Queuing Systems* 1 (1986), pp. 5–27.

Cohen, M., and H. L. Lee. "Resource Deployment Analysis of Global Manufacturing and Distribution Networks." *Journal of Manufacturing and Operations Management* 2 (1989), pp. 81–104.

Coleman, D. R. "Plant Layout: Computers versus Humans." *Management Science* 24 (1977), pp. 107–12.

Conway, R. W.; W. L. Maxwell; J. O. McClain; and S. L. Worona. *Users Guide to XCELL+Factory Modeling System.* Redwood City, CA: Scientific Press, 1987.

Cook, N. H. "Computer Managed Parts Manufacture." *Scientific American* 232 (1975), pp. 23–29.

Deisenroth, M. P., and J. M. Apple. "A Computerized Plant Layout Analysis and Evaluation Technique (PLANET)." In *Technical Papers 1962.* Norcross, GA: American Institute of Industrial Engineers, 1972.

Drezner, Z. "DISCON: A New Method for the Layout Problem." *Operations Research* 28 (1980), pp. 1375–84.

Foulds, L. R. "Techniques for Facilities Layout." *Management Science* 29 (1983), pp. 1414–26.

Francis, R. L.; L. F. McGinnis; and J. A. White. "Locational Analysis." *European Journal of Operational Research* 12 (1983), pp. 220–52.

Francis, R. L., and J. A. White. *Facility Layout and Location: An Analytical Approach.* Englewood Cliffs, NJ: Prentice Hall, 1974.

Groover, M. P., and E. W. Zimmers. *CAD/CAM: Computer Aided Design and Manufacturing.* Englewood Cliffs, NJ: Prentice Hall, 1984.

Hakimi, S. L. "Optimum Location of Switching Centers and the Absolute Centers and Medians of a Graph." *Operations Research* 12 (1964), pp. 450–59.

Jaikumar, R. "Flexible Manufacturing Systems: A Management Perspective." Unpublished manuscript, Harvard School of Business, 1984.

Jaikumar, R. "Postindustrial Manufacturing." *Harvard Business Review* 64 (1986), pp. 69–76.

Johnson, Roger V. "SPACECRAFT for Multi-Floor Layout Planning." *Management Science* 28 (1982), pp. 407–17.

Kaplan, R. S. "Must CIM Be Justified by Faith Alone?" *Harvard Business Review* 64 (1986), pp. 69–76.

Keefer, K. B. "Easy Way to Determine the Center of Distribution." *Food Industries* 6 (1934), pp. 450–51.

Krouse, J. "Flexible Manufacturing Systems Begin to Take Hold." *High Technology* 6 (1986), p. 26.

Lee, R. C., and J. M. Moore. "CORELAP—Computerized Relationship Layout Planning." *Journal of Industrial Engineering* 18 (1967), pp. 194–200.

Nicol, L. M., and R. H. Hollier. "Plant Layout in Practice." *Material Flow* 1, no. 3 (1983), pp. 177–88.

Nugent, C. E.; T. E. Vollmann; and J. Ruml. "An Experimental Comparison of Techniques for the Assignment of Facilities to Locations." *Operations Research* 16 (1968), pp. 150–73.

Pritsker, A. A. B. *Introduction to Simulation and SLAM II.* 3rd ed. New York: John Wiley & Sons, 1986.

Rosenblatt, M. J. "The Dynamics of Plant Layout." *Management Science* 32, no. 1 (1986), pp. 76–86.

Sassani, F. "A Simulation Study on Performance Improvement of Group Technology Cells." *International Journal of Production Research* 28 (1990), pp. 293–300.

Schweitzer, P. J. "Maximum Throughput in Finite Capacity Open Queuing Networks with Product-Form Solutions." *Management Science* 24 (1977), pp. 217–23.

Scriabin, M., and R. C. Vergin. "Comparison of Computer Algorithms and Visual Based Methods for Plant Layout." *Management Science* 22 (1975), pp. 172–81.

Scriabin, M., and R. C. Vergin. "A Cluster Analytic Approach to Facility Layout." *Management Science* 31 (1985), pp. 33–49.

Seehof, J. M., and W. O. Evans. "Automated Layout Design Programs." *Journal of Industrial Engineering* 18 (1967), pp. 690–95.

Sengupta, S., and R. Combes. "Optimizing and General Food's Environment." *IIE Solutions,* August 1995, pp. 30–35.

Stecke, K. E. "Formulation and Solution of Nonlinear Integer Production Planning Problems for Flexible Manufacturing Systems." *Management Science* 29 (1983), pp. 273–88.

Suri, R., and R. R. Hildebrant. "Modeling Flexible Manufacturing Systems Using Mean Value Analysis." *SME Journal of Manufacturing Systems* 3 (1984), pp. 27–38.

Tansel, B. C.; R. L. Francis; and T. J. Lowe. "Location on Networks: A Survey (Parts 1 and 2)." *Management Science* 29 (1983), pp. 482–511.

Tompkins, J. A., and R. Reed Jr. "An Applied Model for the Facilities Design Problem." *International Journal of Production Research* 14 (1976), pp. 583–95.

Tompkins, J. A., and J. A. White. *Facilities Planning.* New York: John Wiley & Sons, 1984.

Vollman, T. E., and E. S. Buffa. "The Facilities Layout Problem in Perspective." *Management Science* 12 (1966), pp. 450–68.

Zygmont, J. "Flexible Manufacturing Systems: Curing the Cure-all." *High Technology* 6 (1986), pp. 22–27.

Chapter **Twelve**

Quality and Assurance

"Quality in a service or product is not what you put into it. It is what the client or customer gets out of it."

—Peter Drucker

Chapter Overview

Purpose

To understand what quality means in the operations context, how it can be measured, and how it can be improved.

Key Points

1. *What is quality?* While we all have a sense of what we mean by quality, defining it precisely as a measurable quantity is not easy. A useful definition is conformance to specifications. This is something that can be measured and quantified. If it can be quantified, it can be improved. However, this definition falls short of capturing all the aspects of what we mean by quality and how it is perceived by the customer.
2. *Statistical process control.* Statistical methods can assist with the task of monitoring quality in the context of manufacturing. The underlying basis of *statistical control charts* is the normal distribution. The normal distribution (bell-shaped distribution) has the property that the mean plus and minus two standard deviations ($\mu \pm 2\sigma$) contains about 95 percent of the population, and the mean plus and minus three standard deviations ($\mu \pm 3\sigma$) contains more than 99 percent of the population. It is these properties that form the basis for statistical control charts. Consider a manufacturing process producing an item with a measurable quantity that must conform to a given specification. One averages the measurements of this quantity in subgroups (typically of size four or five). The central limit theorem guarantees that the distribution of the average measurement will be approximately normally distributed. If the average of a subgroup lies outside two or three sigma limits of the normal distribution, it is unlikely that this deviation is due to chance. This signals an out-of-control situation, which might require intervention into the process. This is the basis for the $\overline{X}$ chart.

   While the $\overline{X}$ chart is a valuable way to test for a shift in the underlying mean of a process, it does not signal shifts in the process variation. To monitor process variation, one computes the range of subgroup measurements (that is, the largest value minus the smallest value in the subgroup). Since the range of a sample is proportional to the standard deviation of a sample, this statistic can be used to monitor process variation. This is the purpose of the *R* chart. The *R* chart establishes upper and lower control limits on the average range of subgroups and signals when the process variation has gone out of control.

3. *The* p *and* c *charts.* The $\overline{X}$ and R charts are useful when measuring quality along a single scalar dimension such as length or weight. In other cases, one might be interested in whether the item functions or not. Under these circumstances, the p chart is appropriate. The p chart is based on the binomial distribution. Either an item has the appropriate attribute or it doesn't. When the observed value of p (the proportion of good items) undergoes a sudden shift, it signals a possible out-of-control situation.

   The c chart is based on the Poisson distribution. The Poisson distribution describes events that occur completely at random over time or space. In the statistical quality control context, consider a situation where a certain number of defects are acceptable, such as minor dents on an automobile, but too many are considered unacceptable. In this case, the c chart would be an appropriate means of monitoring the process. The parameter c is the average rate of occurrence of flaws, and an out-of-control signal is tripped when the observed value of c is too high. Note that both the p and c charts are typically implemented with a normal distribution, since, under the right circumstances, the normal distribution provides a good approximation to both the binomial and Poisson distributions.
4. *Economic design of control charts.* Statistical quality control requires several steps, each of which incurs a different cost. First, there's the cost of inspecting the items. For $\overline{X}$ charts, we assume samples of subgroup size n. Hence, each subgroup sampling incurs a cost proportional to n. Second, if an out-of-control situation is detected, the cost of trying to find out the cause of the problem can be substantial. Even if the out-of-control signal is a false alarm, one must shut down the process. Finally, if the process continues to operate in an out-of-control state, this too could lead to substantial costs as inventories of defectives increase. Control limits can be chosen to best balance these costs.
5. *Acceptance sampling.* The second part of this chapter deals with *acceptance sampling.* Acceptance sampling occurs after a lot of items is produced, rather than during the manufacturing process. It can be performed by the manufacturer or by the consumer. In most cases, 100 percent inspection of items is impractical, impossible, or too costly. For these reasons, a more common approach is to sample a subset of the lot and choose to accept or reject the lot based on the results of the sampling. The most common sampling plans are (1) single sampling, (2) double sampling, and (3) sequential sampling.

   In the case of single sampling, one samples n items from a lot of N items (where $n < N$) and rejects the lot if the number of defects exceeds a specified level. Double sampling means that if the number of defectives falls between two prespecified limits (that is, is neither very high nor very low), one samples again to determine the fate of the lot. In sequential sampling one decides either to accept the lot, reject the lot, or continue sampling after each item is sampled. The appropriate limits for each of these tests are based on the underlying probability distributions and specification of acceptable levels of Type 1 error (α).
6. *Total quality management.* As the quality movement began to take hold in the United States and other parts of the world, one way of describing an organization's commitment to quality was *total quality management* (TQM). Briefly, this is the complete commitment of all parts of a firm to the quality mission. An important part of TQM is listening to the customer. This process

includes customer surveys and focus groups to find out what the customer wants, distilling this information, prioritizing customer needs, and linking those needs to the design of the product. One means of accomplishing the last item on the list is quality function deployment (QFD).

Several agencies worldwide promote quality in their respective countries through formal recognition. This process was started in Japan with the Deming Prize, established and funded by quality guru W. Edwards Deming. In the United States, we recognize outstanding quality with the Baldrige Prize. Another important development is the International Standards Organization's certification, ISO 9000, which requires firms to clearly document their policies and procedures. While the certification process can be costly in both time and money, it is often required to do business in many countries.

The chapter concludes with a discussion of designing for quality. By putting a greater investment up front in sound product design, the consumer will be rewarded with superior products and the firm will be rewarded with customer loyalty.

While the American economy was strong during the latter part of the 1990s, there are some disturbing trends. Our balance of trade in manufactured goods continues to be negative year after year. In particular, we have maintained a negative balance of trade with Japan for several decades, even during the recession that plagued that country in the 1990s. A negative balance of trade means that we import more than we export. In the case of Japan, this is due in large part to the fact that Americans are consuming Japanese-made products at an increasing rate. The success of Japan in consumer electronics and automobiles here are just two examples. But it is not only Japanese-made products whose consumption is increasing. Other examples are German- and Swedish-made automobiles, German-made kitchen appliances, bicycles made in Taiwan, and high-end watches made in Switzerland. In many cases, the imported products are considerably more expensive than their American counterparts, yet are still preferred. Why? The simple answer is that they are perceived to be of higher quality.

What is quality? Traditional thinking would say that quality is conformance to specifications; that is, does the product do what it was designed to do? Some feel that this definition is the only meaningful definition of quality, because conformance is something that can be measured. According to Philip Crosby (1979).

> That is precisely the reason we must define quality as "conformance to requirements" if we are to manage it. Thus, those who want to talk about quality of life must talk about that life in specific terms, such as desirable income, health, pollution control, political programs, and other items that can each be measured.

Crosby makes a good point. By defining quality in terms of conformance, we avoid making unreasonable comparisons. Is a Rolls-Royce a better-quality product than a Toyota Corolla? Not necessarily. The Toyota may be a higher-quality product relative to *what it was designed to do.*

This does not tell the entire story, however. Just as beauty is in the eye of the beholder, so is quality in the mind of the customer. If the customer is not happy with the product, it is not high quality. Viewed in this way, quality is a measure of the conformance of the product to the *customer's needs.*

Why is this different? Conformance to specifications assumes a given design and the specifications resulting from that design. Conformance to customer needs means that the design of the product is part of the evaluation. Given two washing machines

with comparable repair records, what determines which one the consumer will buy? The answer is a combination of aesthetics, features, and design. Viewing quality in this broader way is both good and bad. It is good in that it gets at the heart of the issue: Quality is what the customer thinks it is. It is bad in that it makes it difficult to measure quality and thus difficult to improve it.

There is little doubt that product reliability is an important part of the spectacular success of Japanese automobile manufacturers. Although several American automakers boast initial defect rates in their cars comparable to Japanese autos (as measured by the J. D. Power new car buyer survey), it is not the number of defects on new cars that is really important to most consumers. It is the car's reliability over its life. In this respect, few automakers can come close to the records posted by Honda and Toyota and other Japanese automakers. Each year the Consumer's Union conducts a survey of its readership to determine the readers' experiences that year with the products they own. Automobile reliability is rated on a five-point scale: much worse than average, worse than average, average, better than average, and much better than average. The ratings correspond to several factors, including mechanical failures, electrical failures, and body integrity. Japanese cars have consistently scored better than most of their American and European competitors.

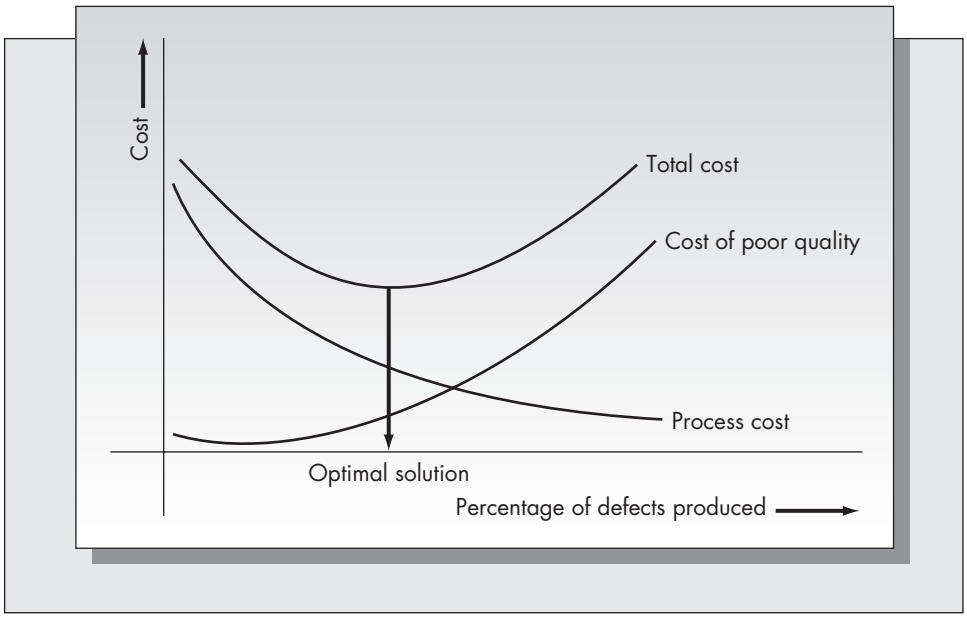
In the 1950s virtually all the automobiles sold in the United States were American made. In 1955 the big three automakers (Ford, G.M., and Chrysler) accounted for 95 percent of the U.S. sales and the majority of the remaining 5 percent were sales made by other (now defunct) American nameplates. Today, Japanese companies account for nearly 30 percent of U.S. sales and are making even more progress on worldwide sales. In 1961 U.S. automakers accounted for close to 50 percent of the world market in passenger vehicles, and Japan about 2 percent. Today, the U.S. manufacturers account for about 15 percent of the world market in passenger automobiles, while Japan's market share is over 20 percent. (However, the United States has continued to increase market share of commercial vehicles, currently accounting for about 40 percent of the world total.)

The major development in the automobile industry in recent years is the explosion of interest in hybrid automobiles. As gasoline prices continue to rise, consumers are turning to high mileage hybrid vehicles in greater numbers. A hybrid combines a traditional gasoline engine with a rechargeable battery. Some of the hybrids (such as those produced by Honda) only use the battery as an engine boost, while the more sophisticated system developed by Toyota for the Prius allows the car to run on battery power alone. The Prius has been a huge success for Toyota, with nearly one million sales worldwide as of this writing in late 2007. Many manufacturers, including General Motors, have plans to develop plug-in hybrids that can travel 60 miles or more on battery power alone. In addition to providing superior gasoline mileage, many of the hybrids are SULEV (super low emissions vehicles), so that this technology makes significant inroads into both the problems of gasoline consumption and air pollution simultaneously.

The enormous strides made by the Japanese in consumer electronics and automobiles leaves the impression that American products and American industry are inferior. This simply is not true. American-based companies are leaders in many major industries. Airplane construction, mainframe computers, biotechnology, financial services, large appliances, chemicals, and telecommunications are a small sample of industries dominated by American firms. By transferring lessons learned in these industries, we will begin to regain our competitive edge in manufacturing in general.

Management must grapple with the difficult problems of knowing how much to invest in quality and determining the best way to go about making that investment.

FIGURE 12–1
The trade-off between quality and cost



There is an optimal trade-off between the cost of poor quality and the investment required in the process to improve the quality, as represented in Figure 12–1. Although such curves can be drawn in principle, evaluating the costs of poor quality is difficult. Direct costs, such as those resulting from scrap, rework, and inspection, are relatively easy to determine. But how does one factor in the costs of lost consumer loyalty? No one would deny that marketing is essential, but has the emphasis on marketing in the United States been at the expense of manufacturing? We must acknowledge that the modern consumer is more educated and more discriminating than ever before. Clever advertisements will not sell second-rate products. It is time to put our investment where it belongs: in the design and manufacture of quality products.

Overview of This Chapter

Statistical quality control dates back to the 1930s. Its roots lie in the work of Walter Shewhart, a scientist employed by Bell Telephone Laboratories. W. Edwards Deming, the man credited with bringing the quality control message to the Japanese, was a student of Shewhart. Deming has stressed in his teachings that understanding the concept of statistical variation of processes is a key step in designing an effective quality control program. One needs to understand process variation in order to know how to produce products that conform to specifications. Deming has become a demigod in Japan, where his teachings ignited the Japanese quality revolution.

This chapter is aimed at providing the student with an understanding of the essentials underlying statistical quality control. We discuss two basic areas: *control charts* and *acceptance sampling*. Briefly, a control chart is a graphical means for determining if the underlying distribution of some measurable variable seems to have undergone a shift. Acceptance sampling is the set of procedures for inferring characteristics of a lot from the characteristics of a sample of items from that lot. Whereas "zero defects" thinking might suggest that these approaches are obsolete, we disagree. Process monitoring and statistical sampling continue to be used. It is important to understand the underlying theory behind these methods to know when and how they should be applied.

12.1 STATISTICAL BASIS OF CONTROL CHARTS

Control charts provide a simple graphical means of monitoring a process in real time. Although easy to construct and easy to use, control charts are based on rigorous statistical principles. They have gained wide acceptance in industry and are preferred to more conventional statistical methods.

A control chart maps the output of a production process over time and signals when a change in the probability distribution generating observations seems to have occurred. To construct a control chart one uses information about the probability distribution of process variation and fundamental results from probability theory. A result that forms the basis for a class of control charts is known as the *central limit theorem.* Roughly, the central limit theorem says that the distribution of sums of independent and identically distributed random variables approaches the normal distribution as the number of terms in the sum increases.[1] Generally, the distribution of the sum converges very quickly to a normal distribution. In order to illustrate the central limit theorem, suppose that X is a random variable with the uniform distribution on the interval (0, 1). The probability density function of X is pictured in Figure 12–2.

The density of X bears little resemblance to a normal density. Now let us assume that the three random variables X_1, X_2, and X_3 are independent random variables, each of which has the uniform distribution on the interval (0, 1). Consider the random variable $W = X_1 + X_2 + X_3$. One can derive the distribution of W by convoluting the distributions of X_1, X_2, and X_3. We will not present the details here. (The interested reader should refer to a graduate-level text in probability such as DeGroot, 1986.) The density function of W appears in Figure 12–3. The resemblance to a normal density is now quite striking. In Figure 12–4 we have graphed the probability density function of W and the associated normal approximation. Notice how closely the two curves agree. Were we to continue to add independent uniform random variables, the agreement would be even closer.

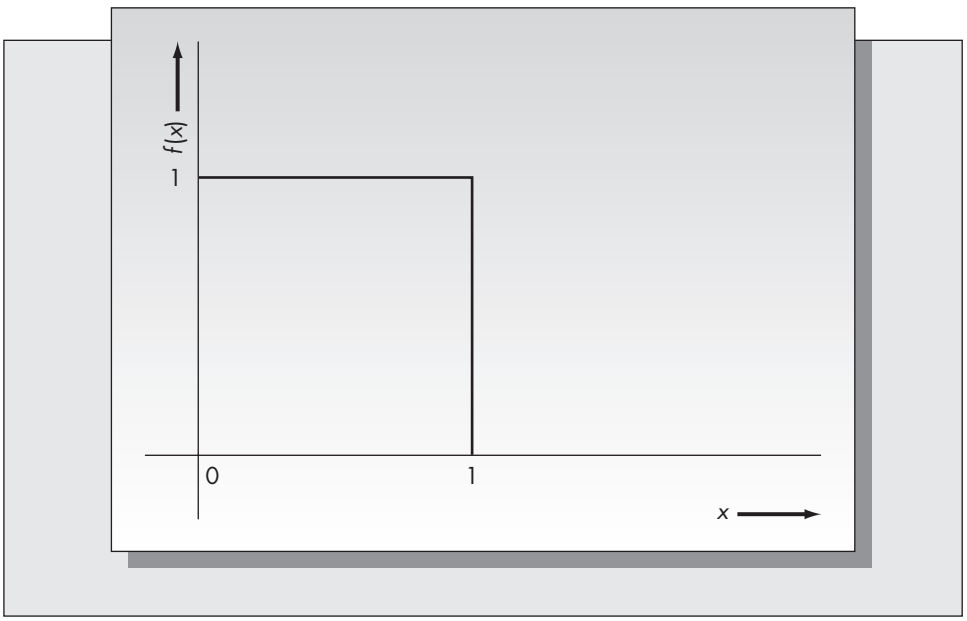


FIGURE 12–2
Probability density of a uniform variate on (0, 1)

[1] The central limit theorem was also used in Chapter 9 to justify the use of the normal distribution to describe project completion time in PERT networks.

FIGURE 12–3
Density of the sum of three uniform random variables

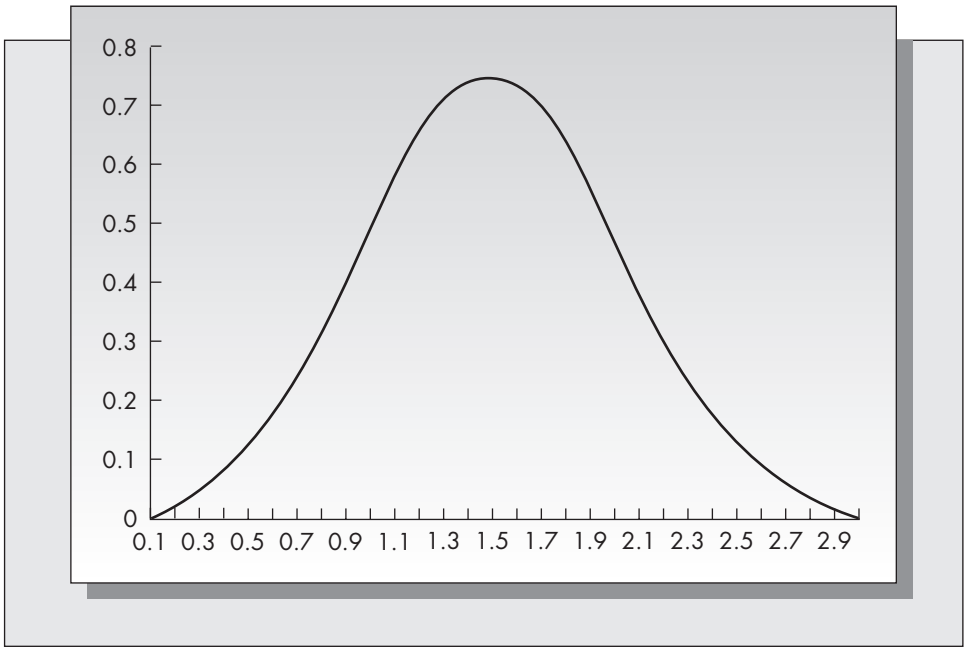
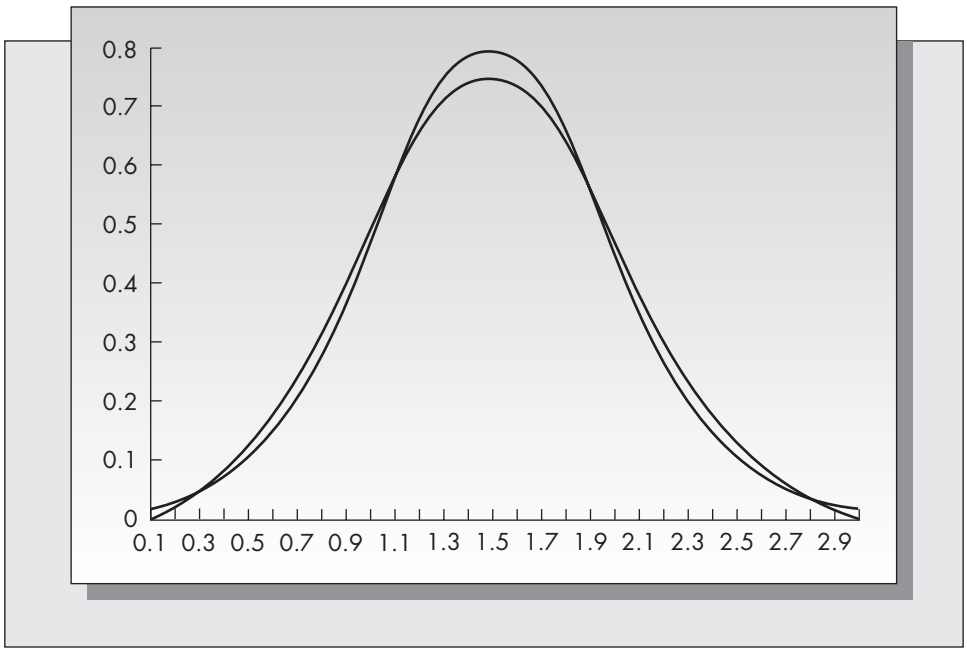


FIGURE 12–4
Density of the sum of three uniform random variables and the normal approximation



In quality control, the central limit theorem justifies the assumption that the distribution of $\overline{X}$, the sample mean, is approximately normally distributed. Recall the definition of the sample mean: If $(X_1, X_2, \ldots, X_n)$ is a random sample, then the sample mean $\overline{X}$ is defined as

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Suppose that a variable Z has the standard normal distribution. Then, from Table A–1 at the back of this book,

$$P\{-3 \le Z \le 3\} = .9974.$$

In words, this means that the likelihood of obtaining a value of Z either larger than 3 or less than -3 is .0026, or roughly 3 chances in 1,000. This is the basis of the so-called three-sigma limits that have become the de facto standard in quality control. Now consider the sample mean $\overline{X}$, which the central limit theorem tells us is approximately normally distributed. Suppose that the mean of each sample value is μ and the standard deviation of each sample value is σ. Then it is well known that the mean of $\overline{X}$ is also μ and the standard deviation of $\overline{X}$ is $\sigma/\sqrt{n}$. Therefore, the standardized variate

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

has (approximately) the normal distribution with zero mean and unit variance. It follows that

$$P\left\{-3 \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le 3\right\} = .9974,$$

which is equivalent to

$$P\left\{\mu - \frac{3\sigma}{\sqrt{n}} \le \overline{X} \le \mu + \frac{3\sigma}{\sqrt{n}}\right\} = .9974.$$

That is, the likelihood of observing a value of $\overline{X}$ either larger than $\mu + 3\sigma/\sqrt{n}$ or less than $\mu - 3\sigma/\sqrt{n}$ is .0026. Such an event is sufficiently rare that if it were to occur, it is more likely to have been caused by a shift in the population mean, μ, than to have been the result of chance. This is the basis of the theory of control charts.

Problems for Section 12.1

1. Suppose that X_1 and X_2 are independent random variables having the uniform distribution on (0, 1, 2, 3, 4, 5). That is,

$$f(j) = P\{X_i = j\} = \tfrac{1}{6} \qquad \text{for } i = 1, 2 \text{ and } 0 \le j \le 5.$$

a. Compute $E(X_1)$ and $\text{Var}(X_1)$. [Hint: Use the formulas

$$E(X) = \sum jf(j),$$

$$\text{Var}(X) = \sum j^2 f(j) - (E(X))^2.]$$

b. Determine the probability distribution for $Y = X_1 + X_2$. [Hint: For each possible value of Y, determine all combinations of X_1 and X_2 that result in that value. For example, $Y = 3$ can be obtained by $(X_1, X_2) = (0, 3), (1, 2), (2, 1)$, and $(3, 0)$. Since each pair has probability $\left(\frac{1}{6}\right)\left(\frac{1}{6}\right) = \frac{1}{36}$, we obtain $P\{Y = 3\} = \frac{4}{36} = \frac{1}{9}$. Repeat this process for all values of Y. As a check be sure that

$$\sum_y P\{Y = y] = 1.0.]$$

c. Using the results of part (*b*), find $P\{1.5 < Y < 6.5\}$.

d. Using the results $E(Y) = 2E(X_1)$ and $\text{Var}(Y) = 2\text{Var}(X_1)$, approximate the answer to part (*c*) using a normal distribution.

e. Suppose that $X_1, X_2, \ldots, X_{20}$ are independent identically distributed random variables having the uniform distribution on (0, 1, 2, 3, 4, 5). Using a normal approximation, estimate

$$P\left\{\sum_{i=1}^{20} X_i \leq 75\right\}.$$

f. Do you think that the approximation computed in part (*d*) or part (*e*) is more accurate? Why?

2. The following data represent the observed number of defective disks produced each hour based on observing the system for 30 successive hours.

0	3	5	2	6	8	3	5	4	6	6	9	5	1	2
1	2	5	3	3	0	1	0	7	1	7	5	4	4	3

a. Plot a frequency histogram for the numbers of failures per hour.

b. Compute the mean and the variance of the sample. Use the formulas

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right).$$

(Note: You can streamline the calculations by grouping the data first, as there are only 10 distinct values.)

c. If the number of defective disks produced hourly is normally distributed with mean and variance computed in part (*b*), determine the probability that fewer than five defectives are observed in any particular hour.

d. If each disk costs the company \$5 to produce and defective disks are discarded, how long (in an expected-value sense) would it take to pay off a new piece of equipment costing \$45,000 that would reduce defectives to half of the current level? Assume 40-hour production weeks and 48 weeks per year.

3. The tensile strength of a heavy-duty plastic bag used in trash compactors is normally distributed with mean 150 pounds per square inch and standard deviation 12 pounds per square inch. An independent landscape contractor uses them to haul refuse that requires 120-pounds-per-square-inch tensile strength. What proportion of the compactor bags will not meet the requirements?

4. A credit rating company recommends granting of credit cards based on several criteria. One is annual income. If the annual income of applicants is normally distributed with mean \$22,000 and standard deviation \$4,800 and the company recommends no applicant unless his or her income exceeds \$15,000, what fraction of the applicants are denied on this basis?

5. *a.* What is the probability that a normal variable exceeds two-sigma limits? (That is, what is the probability of observing a value of the random variable larger than $\mu + 2\sigma$ or less than $\mu - 2\sigma$?)

b. If "deciding that the process is out of control" means observing a realization of the sample mean exceeding k sigma limits, discuss the advantages and disadvantages of using $k = 2$ versus $k = 3$.

6. The members of a private golf club have handicaps that are normally distributed with mean 15 and standard deviation 3.5. In a particular event, foursomes are chosen by grouping four players chosen at random from the club. The handicap of the foursome is the arithmetic average of the handicaps of the four players comprising the foursome. In what proportion of the foursomes will the handicap of the foursome be less than 10 or more than 20? (Hint: The standard deviation of the average of four independent identically distributed random variables is exactly half the standard deviation of one of them.)

12.2 CONTROL CHARTS FOR VARIABLES: THE $\overline{X}$ AND R CHARTS

A process is in control if a stable system of chance causes is operating. That is, the underlying probability distribution generating observations is not changing with time. When the observed value of the sample mean of a group of observations falls outside the appropriate three-sigma limits, it is likely that there has been a change in the probability distribution generating observations. To illustrate how one develops and interprets control charts, consider the following case study.

Example *12.1*

Wonderdisk produces a line of plug-compatible disks for IBM equipment. Building 35 is responsible for production of the read/write arms for the model A55C disk. The arms are approximately 2.875 inches in length. The design engineers have established a tolerance of ± 0.025 inch for the arm lengths and advertise this figure in the published specifications.

The company usually produces 40 arms per day. On 30 consecutive production days, five arms are sampled randomly from each day's production and measured. The resulting measurements appear in Table 12–1.

These observations show that there is some variation in the length of the arms. However, there appears to be no discernible pattern to this variation. Define the random variable X to be the length of an arm selected at random. We may then interpret Table 12–1 as 150 independent observations on the random variable X.

Howard Hamilton, an industrial engineer working for Wonderdisk, is given the job of analyzing these data. The first thing that Howard notices is that the established tolerances of ± 0.025 inch were often exceeded. This becomes most evident by computing the range of daily observations. The range of a sample is the maximum of the observations minus the minimum of the observations. For the 30 days of data, Howard observes that the range exceeds 0.05 in four cases. In order to obtain a clearer idea of what proportion of the population lies outside the specified tolerances, Howard develops a frequency histogram of the 150 measurements. This histogram appears in Figure 12–5. The histogram suggests that the measurements are normally distributed. Howard used a goodness-of-fit test to verify the normality of the observations. Figure 12–6 shows the theoretical normal curve.

One determines the theoretical normal curve in the following way. Because the normal distribution depends upon two parameters, μ and σ, we must estimate these values from the sample data. From the theory of statistics, we know that the "best" estimators of the population mean and variance are the sample mean, $\overline{X}$, and the sample variance, s^2, given by

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

TABLE 12–1
Tracking Arm Data

Sample Number	Measurements of the Length of a Tracking Arm					Average	Range
1	2.8971	2.8477	2.8624	2.8606	2.8971	2.8730	.0494
2	2.8863	2.8541	2.8677	2.8838	2.8854	2.8755	.0322
3	2.8772	2.8708	2.8920	2.8892	2.8840	2.8826	.0212
4	2.8808	2.8650	2.8686	2.8874	2.8804	2.8764	.0224
5	2.8633	2.8993	2.8650	2.8909	2.9131	2.8863	.0497
6	2.8743	2.8571	2.8863	2.8473	2.8739	2.8678	.0390
7	2.8820	2.8612	2.8805	2.8737	2.8933	2.8781	.0322
8	2.8847	2.8630	2.8846	2.8969	2.8916	2.8842	.0339
9	2.8569	2.8934	2.8926	2.8585	2.8721	2.8747	.0365
10	2.8784	2.8795	2.8794	2.8608	2.8672	2.8731	.0187
11	2.8821	2.8544	2.9053	2.8495	2.8670	2.8717	.0558
12	2.8643	2.8533	2.8718	2.8565	2.8724	2.8637	.0191
13	2.8675	2.8578	2.8971	2.8709	2.8908	2.8768	.0394
14	2.8495	2.8701	2.8741	2.8699	2.8766	2.8680	.0271
15	2.8822	2.8731	2.8551	2.8782	2.8687	2.8714	.0271
16	2.8731	2.8675	2.8743	2.8520	2.8900	2.8714	.0379
17	2.9054	2.9190	2.8752	2.8477	2.8639	2.8822	.0713
18	2.8759	2.8832	2.8660	2.8667	2.8674	2.8718	.0172
19	2.8676	2.8775	2.8793	2.8943	2.9048	2.8847	.0373
20	2.8765	2.8613	2.8737	2.8524	2.8767	2.8681	.0243
21	2.9052	2.8851	2.8895	2.8904	2.8723	2.8885	.0328
22	2.8606	2.8837	2.9017	2.8628	2.8455	2.8709	.0562
23	2.8752	2.8722	2.8618	2.8637	2.8725	2.8691	.0133
24	2.8566	2.8929	2.9035	2.9109	2.8594	2.8847	.0543
25	2.8495	2.8749	2.8873	2.8557	2.8673	2.8669	.0378
26	2.8736	2.8606	2.8797	2.8522	2.8802	2.8693	.0280
27	2.8449	2.8908	2.8851	2.8798	2.8610	2.8723	.0459
28	2.8589	2.8800	2.9025	2.8974	2.8606	2.8799	.0437
29	2.8910	2.8546	2.8744	2.8775	2.8634	2.8722	.0364
30	2.8607	2.8769	2.8771	2.8934	2.8706	2.8757	.0326

FIGURE 12–5
Frequency histogram of 150 measurements

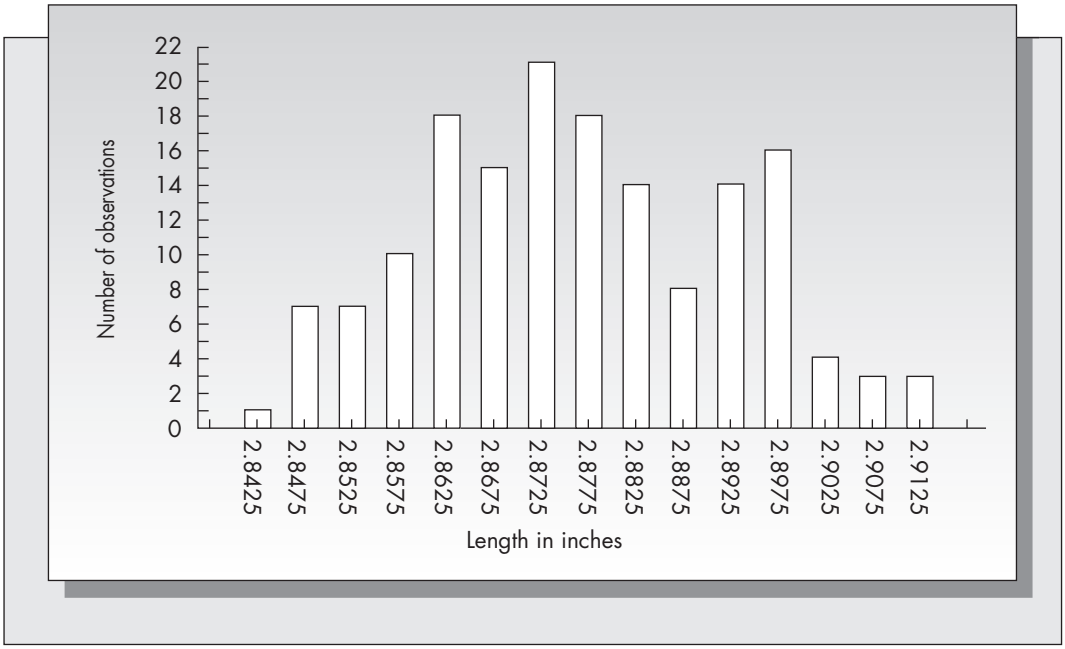
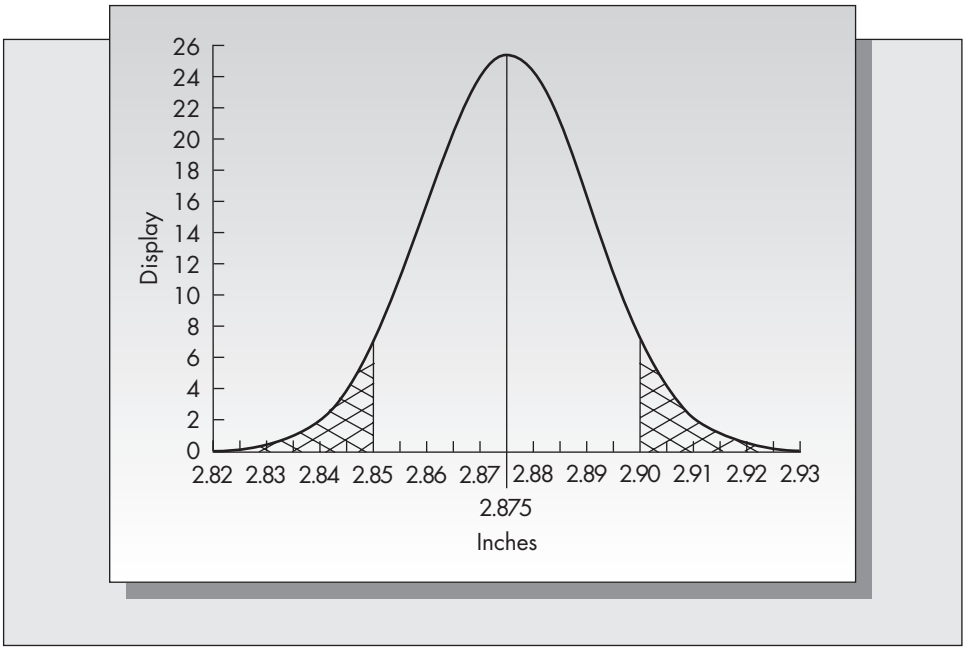


FIGURE 12–6
Theoretical normal curve of arm length



Interpret $X_1, X_2, \ldots, X_n$ as the sample values (the random sample). The sample mean based on all 150 observations is 2.875, and the sample variance is 0.0002434. The estimate of the population standard deviation is the square root of the sample variance, which is 0.0156.

Howard can now estimate the fraction of the arms produced that fall outside the advertised tolerances. An arm will exceed the tolerance if it is longer than 2.90 inches or shorter than 2.85 inches. The area of the crosshatched region in Figure 12–6 represents the probability that this will occur, which is evidently not negligible. In fact,

$$\begin{aligned} P\{X > 2.90 \text{ or } X < 2.85\} &= P\{|X - \mu| > 0.025\} \\ &= P\left\{\left|\frac{X - \mu}{\sigma}\right| > 1.602\right\} \\ &= P\{|Z| > 1.602\} = .11. \end{aligned}$$

We find the probability, .11, in a table of the normal distribution (Table A–1 at the back of this book). Hence, about 11 percent of the arms produced over the last 30 days do not meet the company's published specifications. However, failure rates for the disks due to incompatibility of the arms has been extremely low (less than 1 percent). Howard presented these results to the director of manufacturing, who discussed the problem with the company president and the director of engineering. After some additional investigation, Howard concluded that the original tolerances were not consistent with design requirements of the disk. It was found that a tolerance of ± 0.05 inch would be sufficient. The tighter tolerances were based on an earlier design, and the department simply forgot to revise its figures. Later testing showed that a tolerance of ± 0.05 inch was much more realistic and consistent with the operation of the disk. Howard satisfied himself that the revised tolerances would include more than 99 percent of the population.

This example raises an important point in the application of statistical principles to quality control. If control charts are to be used to compare the characteristics of manufactured items with preset design specifications, then the desired tolerances and the observed statistical variation in the sample must be consistent. If the tolerances are much tighter than the variation observed in the sample, as in Example 12.1, then they

often will be exceeded even when the process is in control. The opposite situation also can occur: the observed tolerances may be much *wider* than the observed variation in the population. In this case an observation may be out of control relative to other sample values, but may fall within desired tolerances. Whether the process is in control would yield little information about whether parts are meeting specifications.

$\overline{X}$ Charts

Consider Example 12.1. Let us say that Howard decides to construct an $\overline{X}$ chart for the data summarized in Table 12–1. An $\overline{X}$ chart requires that the data be broken down into subgroups of fixed size. The size of the subgroups for the example is $n = 5$. The subgroup size should be at least 4 for the central limit theorem to apply.

To construct an $\overline{X}$ chart, it is necessary to estimate the sample mean and the sample variance of the population. This can be done using the given formulas. However, it generally is not recommended that one use the sample standard deviation as an estimator of σ when constructing an $\overline{X}$ chart. For s to be an accurate estimator for σ, it is necessary that the underlying mean of the sample be constant. Because the purpose of an $\overline{X}$ chart is to determine whether a shift in the mean has occurred, we should not assume a priori that the mean is constant when estimating σ. An alternative method for estimating the sample variation that remains accurate when the population mean changes uses data ranges. Even if the process mean shifts, the ranges will be stable as long as the process variation is stable. There is a relationship between the standard deviation of the population and the range of the subgroups of a given size that depends on the subgroup size. That is, there exists a constant d_2 such that

$$\hat{\sigma} = \frac{\overline{R}}{d_2},$$

where $\overline{R}$ is the average of the observed ranges and $\hat{\sigma}$ is an estimate of the population standard deviation. The constants d_2 for various subgroup sizes appear in Table A–5 at the back of this book. For the data presented in Table 12–1, the average of the 30 ranges turns out to be 0.035756. The value of d_2 for subgroups of size 5 is 2.326. Hence the estimator for σ based on this data is

$$\hat{\sigma} = 0.035756/2.326 = 0.01537,$$

which is quite close to the estimate of the standard deviation using the sample standard deviation s.

Given estimators for the mean and the standard deviation of the group average, the control charts are constructed in the following way: Lines are drawn for the upper and the lower control limits at $\overline{X} \pm 3\sigma/\sqrt{n}$. The group averages are graphed on a daily basis. The process is said to be out of control if an observation falls outside of the control limits. The $\overline{X}$ chart for the sample of 30 days for the tracking arm appears in Figure 12–7. Notice that the process appears to be in control, as all observations fall within the 3σ limits.

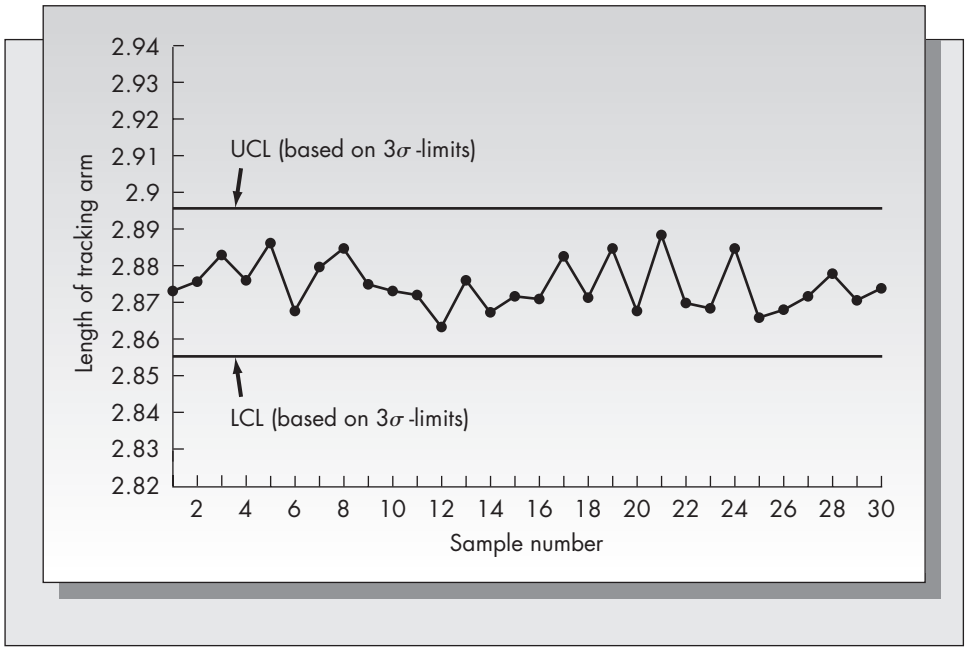
Relationship to Classical Statistics

Here we consider statistical control charts in the context of classical statistical hypothesis testing. The null hypothesis is that the underlying process is in control. That is, we have the hypotheses

H_0: Process is in control.
H_1: Process is out of control.

We interpret the word *control* as meaning that the underlying chance mechanism generating observations over time is stable. For $\overline{X}$ charts, we test whether the process

FIGURE 12–7
$\overline{X}$ chart for tracking arm data



mean has undergone a shift. There are two ways that we can come to the wrong conclusion: reject the null hypothesis when it is true (conclude that the process is out of control when it is in control) and reject the alternative hypothesis when it is true (conclude the process is in control when it is out of control). These are called, respectively, the Type 1 and the Type 2 errors. We use the symbol α to represent the probability of a Type 1 error and β to represent the probability of a Type 2 error. A test is a rule that indicates when to reject H_0 based on the sample values. A test requires specification of an acceptable value of α. Conceptually, we are doing the same thing when we use control charts.

The hypothesis that the process is in control is rejected if an observed value of $\overline{X}$ falls outside the control limits. We can set the values of the upper control limit (UCL) and lower control limit (LCL) based on the specification of any value of α.

$$
\begin{aligned}
\alpha &= P\{\text{Type 1 error}\} \\
&= P\{\text{Out-of-control signal is observed} \mid \text{Process is in control}\} \\
&= P\{\overline{X} < \text{LCL or } \overline{X} > \text{UCL} \mid \text{True mean is } \mu\} \\
&= P\{\overline{X} < \text{LCL} \mid \mu\} + P\{\overline{X} > \text{UCL} \mid \mu\} \\
&= P\left\{\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < \frac{\text{LCL} - \mu}{\sigma/\sqrt{n}}\right\} + P\left\{\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} > \frac{\text{UCL} - \mu}{\sigma/\sqrt{n}}\right\} \\
&= P\left\{Z < \frac{\text{LCL} - \mu}{\sigma/\sqrt{n}}\right\} + P\left\{Z > \frac{\text{UCL} - \mu}{\sigma/\sqrt{n}}\right\}.
\end{aligned}
$$

Because the normal distribution is symmetric, we set

$$
\frac{\text{LCL} - \mu}{\sigma/\sqrt{n}} = -z_{\alpha/2}, \qquad \frac{\text{UCL} - \mu}{\sigma/\sqrt{n}} = z_{\alpha/2},
$$

which gives

$$\text{UCL} = \mu + \frac{\sigma z_{\alpha/2}}{\sqrt{n}},$$

$$\text{LCL} = \mu - \frac{\sigma z_{\alpha/2}}{\sqrt{n}}.$$

Setting $z_{\alpha/2} = 3$, we obtain the popular three-sigma control limits. This is equivalent to choosing a value of $\alpha = .0026$. This particular value of α is the one that is traditionally used; it is not necessarily the only one that makes sense. In some applications, one might wish to increase the likelihood of recognizing when the process goes out of control. One would then use a larger value of α, which would result in tighter control limits. For example, a value of α of .05 would result in two-sigma rather than three-sigma limits.

R Charts

The $\overline{X}$ chart is used to test for a shift in the mean value of a process. In many instances we are also interested in testing for a shift in the variance of the process. Process variation can be monitored by examining the sample variances of the subgroup observations. However, the ranges of the subgroups give roughly the same information and are much easier to compute. The theory behind the *R* chart is that when the underlying population is normal, there is a relationship between the range of the sample and the standard deviation of the sample that depends on the sample size. If $\overline{R}$ is the average of the ranges of all the subgroups of size n, then we have from earlier in this section

$$\hat{\sigma} = \overline{R}/d_2,$$

where d_2, which depends on n, appears in Table A–5 at the back of this book.

Normally, one would develop an *R* chart before an $\overline{X}$ chart in order to obtain a reliable estimator of the variance. The estimator $\hat{\sigma}$ is less sensitive to changes in the process mean than is the estimator s. The purpose of the *R* chart is to determine if the process variation is stable. The upper and lower limits for this chart are given by the formulas

$$\text{LCL} = d_3\overline{R},$$
$$\text{UCL} = d_4\overline{R}.$$

The values of the constants d_3 and d_4 appear in Table A–6 at the back of this book. The values given for these constants assume three-sigma limits for the range process.

Example *12.1* *(continued)*

Again consider the data for the tracking arm in Table 12–1. The ranges of the samples of size $n = 5$ appear in the final column of the table. As stated earlier, the average of these 30 ranges is 0.035756. This is the value of $\overline{R}$ and becomes the center point for the *R* chart. The upper and lower control limits for *R* are computed using the given formulas and Table A–6. For the case of $n = 5$, we have $d_3 = 0$ and $d_4 = 2.11$, thus resulting in the following control limits for Wonderdisk's *R* chart:

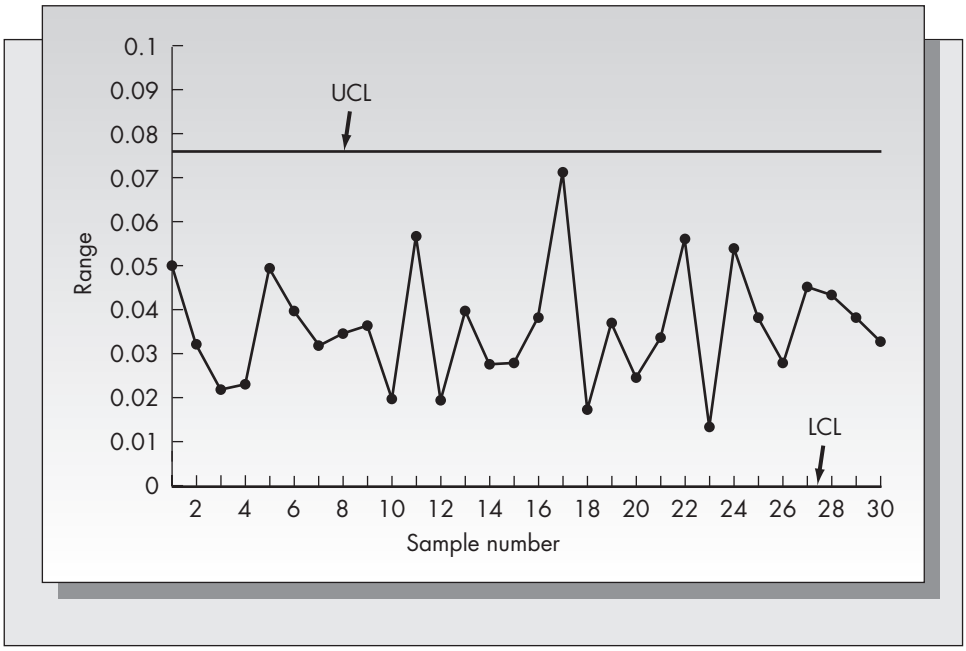
$$\text{LCL} = (0)(0.035756) = 0.$$

$$\text{UCL} = (2.11)(0.035756) = 0.07545.$$

These are three-sigma limits.

Wonderdisk's *R* chart appears in Figure 12–8. Because all observed values of *R* fall within the control limits, the process variation is in control.

FIGURE 12–8
R chart for tracking arm data



R charts are used for testing whether there is a shift in the process variation. For the values of the estimators used to construct the $\overline{X}$ chart to be correct, the process variance should be constant. That is, it is recommended that the *R* chart be used *before* the $\overline{X}$ chart, since an $\overline{X}$ chart assumes that the process variation is stable.

R charts are not the only means for testing the stability of the process variation. One could also use a σ chart. One plots the sample standard deviations of subgroups over time to determine when and if a statistically significant shift in these values occurs. Sigma charts are rarely used in practice for two reasons:

1. It is more work to compute the sample standard deviations for each subgroup than it is to compute the ranges.
2. *R* charts and σ charts will almost always give the same results.

For these reasons we will not discuss σ charts in this text.

Problems for Section 12.2

7. The quality control group of a manufacturing company is planning to use control charts to monitor the production of a certain part. The specifications for the part require that each unit weigh between 13.0 and 15.5 ounces with a target value of 14.25. A sample of 75 observations results in the following:

$$\sum_{i=1}^{75} X_i = 1{,}065, \qquad \sum_{i=1}^{75} X_i^2 = 15{,}165.$$

a. Are the specifications consistent with the statistical variation in the sample? (Hint: Use the computing formula for the sample variance:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2\right].)$$

b. What problems do you anticipate if the group attempts to use $\overline{X}$ and R charts for this process?

c. If the 75 observations are statistically stable, what percentage of the manufactured items will fall outside the tolerances?

8. Control charts for $\overline{X}$ and R are maintained on the shear strength of spot welds. One hundred observations divided into subgroups of size five are used as a baseline to construct the charts, and estimates of μ and σ are computed from these observations. Assume that the 100 observations are $X_1, X_2, \ldots, X_{100}$ and the ranges of the 20 subgroups are $R_1, R_2, \ldots, R_{20}$. From these baseline data the following quantities are computed:

$$\sum_{i=1}^{100} X_i = 97{,}500, \quad \sum_{j=1}^{20} R_j = 1{,}042.$$

Using this information, compute the values of three-sigma limits for both charts.

9. For Problem 8, suppose that the probability of concluding that the process is out of control when it is actually in control is set at .02. Find the upper and the lower control limits for the resulting $\overline{X}$ chart.

10. Suppose that the process mean shifts to 1,011 in Problem 8.

a. What is the probability that the $\overline{X}$ chart will *not* indicate an out-of-control condition from a single subgroup sampling? (This is precisely the Type 2 error.)

b. What is the probability that the $\overline{X}$ chart will not indicate an out-of-control condition after sampling 20 subgroups?

11. A film processing service monitors the quality of the developing process with light-sensitive equipment. The accuracy measure is a number with a target value of zero. Suppose that an $\overline{X}$ chart with subgroups of size five is used to monitor the process and the control limits are UCL = 1.5 and LCL = −1.5. Assume that the estimate for the process mean is zero and for the process standard deviation is 1.30.

a. What is the value of α for this control chart?

b. Find the UCL and LCL based on three-sigma limits.

c. Suppose that the process mean shifts to 1. What is the probability that the shift is detected on the first subgroup after the shift occurs?

12. An R chart is used to monitor the variation in the weights of packages of chocolate chip cookies produced by a large national producer of baked goods. An analyst has collected a baseline of 200 observations to construct the chart. Suppose the computed value of $\overline{R}$ is 3.825.

a. If subgroups of size six are to be used, compute the value of three-sigma limits for this chart.

b. If an $\overline{X}$ chart based on three-sigma limits is used, what is the difference between the UCL and LCL?

13. A process is monitored using an $\overline{X}$ chart with UCL = 13.8 and LCL = 8.2. The process standard deviation is estimated to be 6.6. If the $\overline{X}$ chart is based on three-sigma limits,

a. What is the estimate of the process mean?

b. What is the size of each of the sampling subgroups?

12.3 CONTROL CHARTS FOR ATTRIBUTES: THE *p* CHART

$\overline{X}$ and R charts are valuable tools for process control when the output of the process can be expressed as a single real variable. This is appropriate when there is a single quality dimension such as length, width, or hardness. In two circumstances control charts for variables are not appropriate: (1) when one's concern is whether an item has a particular attribute (for example, the issue might be whether the item functions) and (2) when there are many different quality variables. In case (2) it is not practical or cost-effective to maintain separate control charts for each variable. Either the item has the desired attributes or it does not.

When using control charts for attributes, each sample value is either a 1 or a 0. A 1 means that the item is acceptable, and a 0 means that it is not. Let n be the size of the sampling subgroup and define the random variable X as the total number of defectives in the subgroup. We will assume that each subgroup represents a sampling from one day's production. The theory would be exactly the same whether the sampling interval is one hour, one day, or one month. Because X counts the number of defectives in a fixed sample size, the underlying distribution of X is binomial with parameters n and p. Interpret p as the proportion of defectives produced and n as the number of items sampled in each group (typically, n is the number of items sampled each day). A p chart would be used to determine if there is a significant shift in the true value of p.

Although one could construct p charts based on the exact binomial distribution, it is more common to use a normal approximation. Also, as our interest is in estimating the value of p, we track the random variable X/n, whose expectation is p, rather than X itself. It is easy to show that

$$E(X/n) = p,$$
$$\text{Var}(X/n) = p(1-p)/n.$$

For large n, the central limit theorem tells us that X/n is approximately normally distributed with parameters $\mu = p$ and $\sigma = \sqrt{p(1-p)/n}$. Using a normal approximation, the traditional three-sigma limits are

$$\text{UCL} = p + 3\sqrt{\frac{p(1-p)}{n}},$$
$$\text{LCL} = p - 3\sqrt{\frac{p(1-p)}{n}}.$$

The estimate for p, the true proportion of defectives in the population, is $\overline{p}$, the average fraction of defectives observed over some reasonable baseline period. The process is said to be in control as long as the observed fraction defective for each subgroup remains within the upper and the lower control limits.

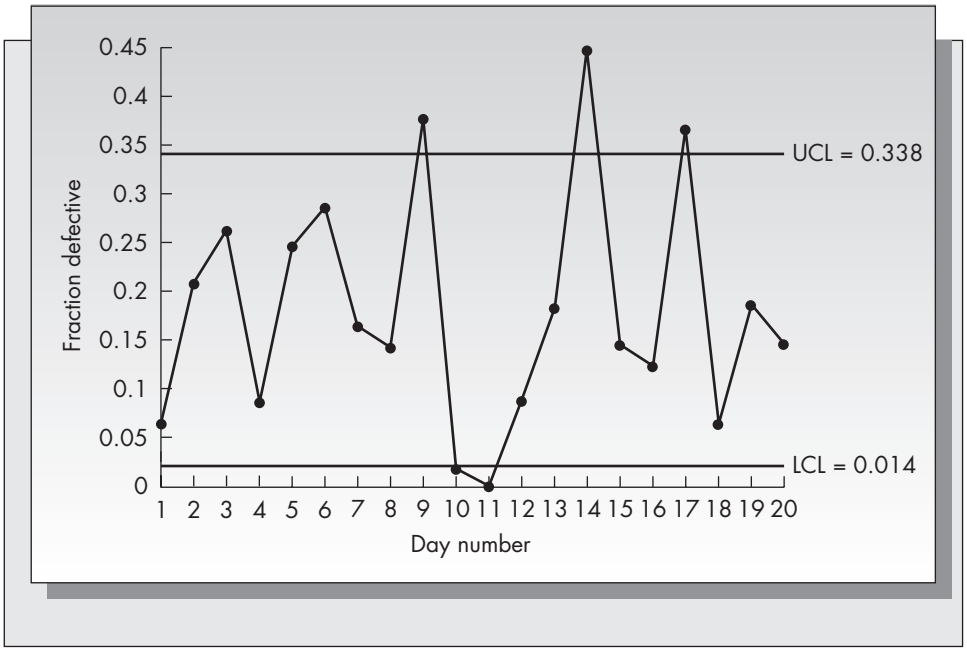
Example *12.2* Xezet, a maker of DVDs, inspects a sample of 50 disks from each day's output. Based on a variety of attributes, the quality inspector classifies each disk as acceptable or not. The experience over a typical 20-day period is summarized in Table 12–2.

To construct a control chart for the fraction defective, it is necessary to have an accurate estimate of the true fraction defective in the entire population, p. Based on the data in Table 12–2 we construct a preliminary control chart to determine if the baseline data are in control. The total number of defectives observed during the 20 days is 176. The total production

TABLE 12–2
Number of Rejected Disks

Date	Number Rejected	Date	Number Rejected
3/18	3	4/1	0
3/19	10	4/2	4
3/20	13	4/3	9
3/21	4	4/4	22
3/22	12	4/5	7
3/25	14	4/8	6
3/26	8	4/9	18
3/27	7	4/10	3
3/28	19	4/11	9
3/29	1	4/12	7

FIGURE 12–9
Preliminary *p* chart for Xezet DVD data (refer to Example 12.2)



over the same period of time is 1,000 disks. Hence, the current estimate of the proportion of defectives in the population is 176/1,000 = .176. The current estimator for σ is

$$\hat{\sigma} = \sqrt{\frac{\bar{p}(1 - \bar{p})}{n}} = \sqrt{\frac{(.176)(.824)}{50}} = 0.54.$$

Based on three-sigma limits we obtain

$$\text{UCL} = 0.176 + (3)(0.054) = 0.338,$$
$$\text{LCL} = 0.176 - (3)(0.054) = 0.014.$$

Figure 12–9 is the preliminary control chart for the fraction defective. Notice that four points are out of control. These correspond to days 9 (3/28), 11 (4/1), 14 (4/4), and 17 (4/9). We need not worry about a point that falls below the lower control limit, because that shows a *better*-than-expected rate of defectives. The production manager considers the three remaining out-of-control points and realizes that they correspond to three days when a key employee

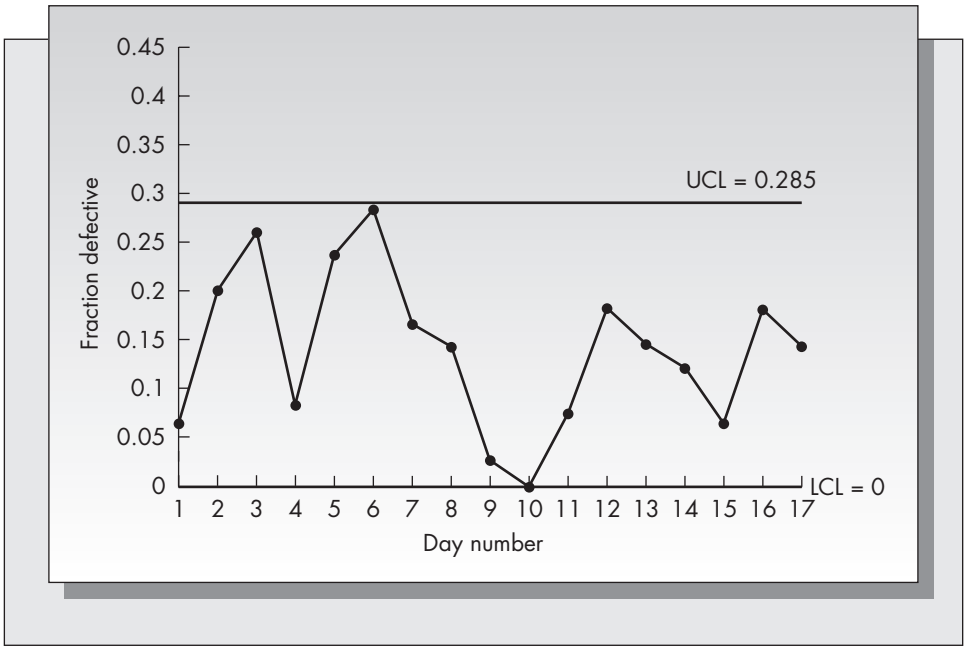


FIGURE 12–10
Revised *p* chart for Xezet DVD data (refer to Example 12.2)

was absent from work for personal reasons. The employee's job requires some very complex media-plating equipment. The high rate of defectives on these days was apparently the result of a temporary employee's lack of experience with media-plating equipment.

Because the out-of-control points were explained by assignable causes, we eliminate these points from our sample. The baseline data now consist of the data listed in Table 12–2 with the three days corresponding to the out-of-control points eliminated. That is, our database now consists of a total of 17 days. Our estimate of p is now recomputed based on the revised data. We obtain

$$\bar{p} = 117/(17)(50) = .138$$

and

$$\hat{\sigma} = \sqrt{(.138)(.862)/50} = 0.049.$$

Based on these new estimators for p and σ we obtain

$$\text{UCL} = 0.138 + (3)(0.049) = 0.285,$$
$$\text{LCL} = 0.138 - (3)(0.049) = -0.009.$$

If LCL < 0, we set LCL $= 0$, because it is impossible to observe a value of p that is negative. In Figure 12–10 we graph the revised p chart for the DVD data. Notice that with the three out-of-control points eliminated, all remaining points now fall within the control limits. This will not always be the case, as the upper and the lower control limits are closer in Figure 12–10 than they were in Figure 12–9. The revised control limits would now be used to monitor future observations of the fraction defective for this process.

p Charts for Varying Subgroup Sizes

In Section 12.3 we assumed that the number of items inspected in each subgroup was the same. This assumption is reasonable when the subgroups are sampled periodically from large lots. However, in many circumstances few items are produced each day and are consequently subject to 100 percent inspection. If daily production varies, the subgroup

TABLE 12–3
Observed Defectives for Industrial Lathes (Refer to Example 12.3)

Day Number	Production	Number of Defectives	Standardized *Z* Value
1	82	8	−0.4451
2	76	12	1.2320
3	85	6	−1.2382
4	53	5	−0.4319
5	30	3	−0.2270
6	121	14	0.0893
7	63	11	1.5404
8	80	9	−0.0178
9	88	7	−0.9946
10	97	8	−0.9532
11	91	13	0.8953
12	77	14	1.9029
13	71	6	−0.7614
14	95	8	−0.8899
15	102	13	0.4566

size also will vary. Here we base the analysis on the standardized variate Z:

$$Z = \frac{p - \bar{p}}{\sqrt{\bar{p}(1 - \bar{p})/n}},$$

which is approximately standard normal independent of n. The lower and the upper control limits would be set at -3 and $+3$, respectively, to obtain three-sigma limits, and the control chart would monitor successive values of the standardized variate Z.

Example *12.3*

A producer of industrial lathes performs 100 percent inspection and testing of each day's production. The lathe production varies from day to day based on anticipated orders and the production schedule for the other machines the company produces. The observed number of defective lathes and the daily production over the past 15 days are given in Table 12–3.

One computes the standardized Z values in the last column in the following way. First, estimate p from the entire 15-day history by forming the ratio of the total number of defectives observed over the total number of items produced. For the data in Table 12–3 we obtain $\bar{p} = 137/1{,}211 = .1131$. The Z values are now computed by the given formula. For example, for day 1,

$$Z = \frac{8/82 - 137/1{,}211}{\sqrt{\dfrac{(137/1{,}211)(1 - 137/1{,}211)}{82}}} = -0.4451.$$

Because the three-sigma limits in Z are simply ± 3, it is clear from Table 12–3 that this process is in control.

Problems for Section 12.3

14. A manufacturer of photographic film produces an average of 12,000 rolls of various types of film each day. An inspector samples 25 from each day's production and tests them for accuracy of color reproduction and overall film quality. During

23 consecutive production days, the inspector rejected the following numbers of rolls of film:

Day	Number Rejected	Day	Number Rejected
1	2	13	1
2	3	14	1
3	2	15	0
4	1	16	0
5	3	17	0
6	4	18	2
7	7	19	3
8	1	20	1
9	3	21	2
10	3	22	2
11	0	23	1
12	0		

a. Based on these observations, compute three-sigma limits for a *p* chart.

b. Do any of the points fall outside the control limits? If so, recompute the three-sigma limits after eliminating the out-of-control points.

15. Applied Machines produces large test equipment for integrated circuits. The machines are made to order, so the production rate varies from month to month. Before shipping, each machine is subject to extensive testing. Based on the tests the machine is either passed or sent back for rework. During the past 20 months the firm has had to rework the following numbers of machines:

Month	Number Produced	Number Reworked	Month	Number Produced	Number Reworked
1	23	3	11	17	3
2	28	3	12	4	0
3	16	1	13	14	2
4	6	0	14	0	0
5	41	2	15	18	6
6	32	4	16	0	0
7	29	5	17	33	4
8	19	2	18	46	5
9	12	1	19	21	7
10	7	1	20	29	7

Determine if the process was in control for the 20-month period using a standardized version of the *p* chart. Assume three-sigma limits for the control chart. (Hint: Do you actually have 20 months of data?)

16. Consider the example of Applied Machines presented in Problem 15. Based on the estimate of the probability that a machine is sent back for rework computed from the 20 months of data, determine the following:

a. If the company produces 35 machines in one particular month, how many, on average, require rework?

b. Out of 100 machines produced, what is the probability that more than 20 percent of them require rework? (Use the normal approximation to the binomial for your calculations. It is discussed in Appendix 12–A.)

17. Over a period of 12 consecutive production hours, samples of size 50 resulted in the following proportions of defective items:

Sample Number	Proportion Defective	Sample Number	Proportion Defective
1	.04	7	.10
2	.02	8	.10
3	.06	9	.06
4	.08	10	.08
5	.08	11	.04
6	.04	12	.04

a. What are the three-sigma control limits for this process?

b. Do any of the sample points fall outside of the control limits?

c. The company claims a defect rate of 3 percent for these items. Are the observed proportions consistent with a target value of 3 percent defectives? What difficulty would arise if the control limits were based on a target value of 0.03? In view of the company's claims, what difficulty would arise if the control limits computed in part (*a*) were used?

12.4 THE *c* CHART

There are control charts other than $\overline{X}$, R, and p charts. Although the form of the distribution of the appropriate random variable depends on the application, the basic approach is the same. In general, one must determine the probability distribution of the random variable of interest, and find upper and lower control limits that contain the universe of observations with a desired level of confidence. Usually one sets the probability of falling outside the control limits to be less than .01.

The p chart is appropriate when classifying an item as either good or bad. However, often we are concerned with the number of defects in an item or collection of items. An item is acceptable if the number of defects is not too large. For example, a refrigerator that has a few scratches might be considered acceptable, but one that has too many scratches might be considered unacceptable. As another example, for a textile mill that manufactures cloth, both the manufacturer and the consumer would be concerned with the number of defects per yard of cloth.

The *c chart* is based on the observation that if the defects are occurring completely at random, then the probability distribution of the number of defects per unit of production has the Poisson distribution. If c represents the true mean number of defects in a unit of production, then the likelihood that there are k defects in a unit is

$$P\{\text{Number of defects in one unit} = k\} = \frac{e^{-c}c^k}{k!} \quad \text{for } k = 0, 1, 2, \ldots .$$

In using a control chart for number of defects, the sample size should be the same at each inspection. One estimates the value of c from baseline data by computing the sample mean of the observed number of defects per unit of production. When $c \geq 20$, the normal distribution provides a reasonable approximation to the Poisson. Because the

mean and the variance of the Poisson are both equal to c, it follows that for large c,

$$Z = \frac{X - c}{\sqrt{c}}$$

is approximately standard normal. Using the traditional three-sigma limits, the upper and the lower control limits for the c chart are

$$\text{LCL} = c - 3\sqrt{c},$$
$$\text{UCL} = c + 3\sqrt{c}.$$

One develops and uses a c chart in the same way as $\overline{X}$, R, and p charts.

Example *12.4*

Leatherworks produces various leather goods in its plant in Montpelier, Vermont. Inspections of the past 20 units of a leather portfolio revealed the following numbers of defects:

Unit Number	Number of Defects Observed	Unit Number	Number of Defects Observed
1	4	11	2
2	3	12	3
3	3	13	6
4	0	14	1
5	2	15	5
6	5	16	4
7	4	17	1
8	2	18	1
9	3	19	2
10	3	20	2

Most defects are the result of natural marks in the leather, but even so, the firm does not want to ship products out with too many defects in the leather. Using these 20 data points as a baseline, determine upper and lower control limits that include the universe of observations with probability .95. What control limits result from using a normal approximation of the Poisson?

Solution

To estimate c we compute the sample mean of the data, which is found by adding the total number of observed defects and dividing by the number of observations. This gives $c = 56/20 = 2.8$. To be certain that a c chart is appropriate here, we should do a goodness-of-fit test of these data for a Poisson distribution with parameter 2.8. We will leave it to the reader to check that the data do indeed fit a Poisson distribution. (Goodness-of-fit tests are described in almost every statistics text. The most common is the chi-square test.)

To determine exact control limits, we use Table A–3 in the back of the book. Because the Poisson is a discrete distribution, it is very unlikely that we will be able to find control limits that contain exactly 95 percent of the probability. From the table we see that the probability that the number of defects is less than or equal to zero is $1.0000 - .9392 = .0608$, which is too large. Hence, we will set the lower control limit to zero. By symmetry, the upper limit should correspond to a right tail of about .025, which occurs at $k = 7$. Hence we would recommend control limits of LCL = 0 and UCL = 7.

For a normal distribution, approximately two standard deviations from the mean include 95 percent of the probability. Hence, the control limits based on a normal approximation of the Poisson are

$$\text{LCL} = c - 2\sqrt{c} = 2.8 - (2)\sqrt{2.8} = -0.55 \qquad \text{(set to zero)},$$
$$\text{UCL} = c + 2\sqrt{c} = 2.8 + 2\sqrt{2.8} = 6.2.$$

The normal approximation is not very accurate in this case because c is too low.

Problems for Section 12.4

18. Amertron produces electrical wiring in 100-foot rolls. The quality inspection process involves selecting rolls of wire at random and counting the number of defects on each roll. The last 20 rolls examined revealed the following numbers of defects:

Roll	Number of Defects	Roll	Number of Defects
1	4	11	2
2	6	12	5
3	2	13	5
4	4	14	7
5	1	15	4
6	9	16	8
7	5	17	6
8	5	18	4
9	3	19	6
10	3	20	4

a. If the number of defects per 100-foot roll of wire follows a Poisson distribution, what is the estimate of c obtained from these observations?

b. Using a normal approximation to the Poisson, what are the three-sigma control limits that you would use to monitor this process?

c. Are all 20 observations within the control limits?

19. Amertron, discussed in Problem 18, has established a policy of passing rolls of wire having five or fewer defects.

a. Based on the exact Poisson distribution, what is the proportion of the rolls that pass inspection? (See Table A–3 in the back of the book.)

b. Estimate the answer to part (*a*) using the normal approximation to the Poisson.

20. A large national producer of cookies and baked goods uses a c chart to monitor the number of chocolate chips in its chocolate chip cookies. The company would like to have an average of six chips per cookie. One cookie is sampled each hour. The results of the last 12 hours were

Hour	Number of Chips per Cookie	Hour	Number of Chips per Cookie
1	7	7	3
2	4	8	6
3	3	9	3
4	3	10	2
5	5	11	4
6	4	12	4

a. Assuming a target value of $c = 6$, what are the upper and the lower control limits for a c chart?

b. Are the 12 observations consistent with a target value of $c = 6$? If those 12 observations constitute a baseline, what upper and lower control limits result? (Use the normal approximation for your calculations.)

21. For the company mentioned in Problem 20, a purchaser of a bag of chocolate chip cookies discovers a cookie that has no chips in it and charges the company with fraudulent advertising. Suppose the company produces 300,000 cookies per year. If the expected number of chips per cookie is six, how many cookies baked each year would have no chips? See Table A–3 in the back of the book.

12.5 CLASSICAL STATISTICAL METHODS AND CONTROL CHARTS

Control charts signal nonrepresentative observations in a sample. The hypothesis that a shift in the process has occurred also can be tested by classical statistical methods. For example, consider the p chart. In constructing the p chart we are testing the hypothesis that a shift has occurred in the underlying value of p, the true proportion of defectives in the lot. A $2 \times n$ contingency table also can be used to test if p has changed. One variable is time, and the other variable is the proportion of defectives observed. The χ^2 test would be used to test whether or not there exists a relationship between the two variables; that is, whether the proportion of defectives changes with time.

It is not necessarily true that the χ^2 test will give the same results as a p chart. In general, the χ^2 test will recommend rejection of the hypothesis that the data are homogeneous based on the average of the departures from the estimated mean, and the control chart will recommend rejection of the hypothesis that the process is in control based on a large deviation of a single observation. It is important to understand this difference to determine which would be a more appropriate procedure. It is probably true that in the context of manufacturing one is more concerned with extreme deviations of a few observations than the average of many deviations, thus providing one reason for the preference among practitioners for control chart methodology. Another reason for the preference for control charts is that they are easy to use and understand. Quality control managers are more familiar with control charts than they are with classical statistical methods.

Problem for Section 12.5

22. Consider the data presented in Problem 14. Problem 14 required testing whether the process was in control using a p chart. Test the hypothesis that the value of p is the same each day using classical statistical methods. That is, test the hypothesis

$$H_0: p_1 = p_2 = \cdots = p_k$$

versus

$$H_1: \text{Not all the } p_i \text{ are equal,}$$

where k is the number of days in the data set ($k = 23$ in this case) and p_i is the true proportion of defectives for day i. Define x_i as the number of rolls of film rejected in day i and p' as the estimate of p obtained from the data [p' was computed in Problem 14(a)]. The test statistic is given by the formula

$$\chi^2 = \sum_{i=1}^{k} \frac{(x_i - np')^2}{np'(1 - p')},$$

where n is the number of items sampled each day ($n = 25$ in this case). The test is to reject H_0 if $\chi^2 > \chi^2_{\alpha,k-1}$, where $\chi^2_{\alpha,k-1}$ is a number obtained from a table of the χ^2 distribution. For $\alpha = .01$ (which is larger than the α value that yields 3σ limits on a p chart), $\chi^2_{.01,22} = 40.289$. Based on this value of α, does the χ^2 test indicate that this process is out of control? If the answer you obtained is different from that

of Problem 14(*b*), how do you account for the discrepancy? Which method is probably better suited for this application?

*12.6 ECONOMIC DESIGN OF $\overline{X}$ CHARTS

The design of an $\overline{X}$ chart requires the determination of various parameters. These include the amount of time that elapses between sampling, the size of the sample drawn in each interval, and the upper and lower control limits. The penalties associated with the upper and lower control limits are reflected in the Type 1 and Type 2 errors. This section will incorporate explicit costs of these errors into the analysis as well as costs of sampling, and considers the problem of designing an $\overline{X}$ chart based on cost minimization.

The model treated here does not include the sampling interval as a decision variable. In many circumstances the sampling interval is determined from considerations other than cost. There are convenient or natural times to sample based on the nature of the process, the items being produced, or personnel constraints.

We will consider the following three costs:

1. Sampling cost.
2. Search cost.
3. Cost of operating out of control.

1. *The sampling cost.* We assume that exactly n items are sampled each period. In most cases, sampling requires workers' time, so personnel costs are incurred. There also may be costs associated with the equipment required for the sampling. Furthermore, sampling may require destructive testing, adding the cost of the item itself. We will assume that for each item sampled, there is a cost of a_1. It follows that the sampling cost incurred each period is $a_1 n$.

2. *The search cost.* When an out-of-control condition is signaled, the presumption is that there is an assignable cause for the condition. The search for the assignable cause generally will require that the process be shut down. When an out-of-control signal occurs, there are two possibilities: either the process is truly out of control or the signal is a false alarm. In either case, we will assume that there is a cost a_2 incurred each time a search is required for an assignable cause of the out-of-control condition. The search cost could include the costs of shutting down the facility, labor time required to identify the cause of the signal, time required to determine if the out-of-control signal was a false alarm, and the cost of testing and possibly adjusting the equipment. Note that the search cost is probably a random variable: it might not be possible to predict the degree of effort required to search for an assignable cause of the out-of-control signal. When that is the case, interpret a_2 as the *expected* search cost.

3. *The cost of operating out of control*. The third and final cost that we will consider is the cost of operating the process after it has gone out of control. There is a greater likelihood that defective items are produced if the process is out of control. If defectives are discovered during inspection, they would either be scrapped or repaired at a future time. An even more serious consequence is that a defective item becomes part of a larger subassembly, which must be disassembled or scrapped. Finally, defective items can make their way into the marketplace, resulting in possible costs of warranty claims, liability suits, and overall customer dissatisfaction. Assume that there is a cost of a_3 each period that the process is operated in an out-of-control condition.

We consider the economic design of $\overline{X}$ charts only. Assume that the process mean is μ and the process standard deviation is σ. A sufficient history of observations is assumed

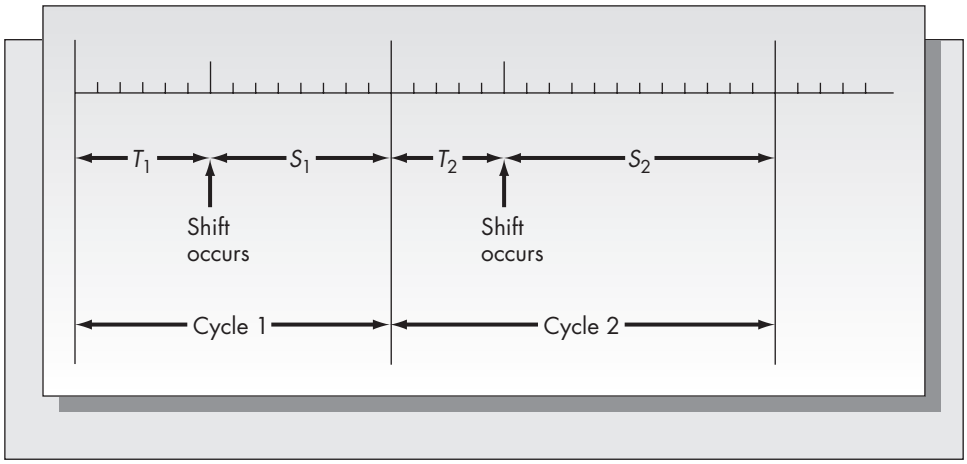


FIGURE 12–11
Successive cycles in process monitoring

to exist so that μ and σ can be estimated accurately. We also assume that an out-of-control condition means that the underlying mean undergoes a shift from μ to $\mu + \delta\sigma$ or to $\mu - \delta\sigma$. Hence, *out of control* means that the mean shifts by δ standard deviations.

Define a cycle as the time interval from the start of production just after an adjustment to detection and elimination of the assignable cause of the next out-of-control condition. A cycle consists of two parts. Define T as the number of periods that the process remains in control directly following an adjustment and S as the number of periods that the process remains out of control until a detection is made. A cycle is the sum $T + S$. Successive cycles are pictured in Figure 12–11. Note that both T and S are random variables, so the length of each cycle is a random variable as well. The probability distribution of T is given subsequently.

The $\overline{X}$ chart is assumed to be constructed using the following control limits:

$$\text{UCL} = \mu + \frac{k\sigma}{\sqrt{n}}.$$

$$\text{LCL} = \mu - \frac{k\sigma}{\sqrt{n}}.$$

Throughout this chapter we have assumed that $k = 3$, but this may not always be optimal. The goal of the analysis of this section is to determine the economically optimal values of both k and n. The method of analysis is to determine an expression for the total cost incurred in one cycle and an expression for the expected length of each cycle. In the spirit of regenerative process (see Ross, 1970, for example), we have the result that

$$E(\text{Cost per unit time}) = \frac{E(\text{Cost per cycle})}{E(\text{Length of cycle})}.$$

After determining an expression for the expected cost per unit time, we will find the optimal values of n and k that minimize this cost.[2]

Assume that T, the number of periods that the system remains in control following an adjustment, is a discrete random variable having the geometric distribution. That is,

$$P\{T = t\} = \pi(1 - \pi)^t \qquad \text{for } t = 0, 1, 2, 3, \ldots .$$

[2] Similar ideas were used in inventory control models in Chapters 4 and 5 and will be used to analyze age replacement models in Section 12.7.

The geometric model arises in the following fashion. Suppose that in any period the process is in control. Then π is the conditional probability that the process will shift out of control in the next period. The geometric distribution is the discrete analog of the exponential distribution. Like the exponential distribution, the geometric distribution also has the memoryless property.[3] In the present context, the memoryless property implies that there is no aging or decay in the production process. That is, the process is equally likely to shift out of control just after an assignable cause has been found and corrected as it is many periods later. This assumption will be accurate when process shifts are due to random causes or when the process is recalibrated on an ongoing basis.

An out-of-control signal is indicated when

$$|\bar{X} - \mu| > \frac{k\sigma}{\sqrt{n}}.$$

As in earlier sections of this chapter, let α be the probability of Type 1 error. Type 1 error occurs when an out-of-control signal is observed but the process is in control. It follows that

$$\begin{aligned}\alpha &= P\left\{ \left|\bar{X} - \mu\right| > \frac{k\sigma}{\sqrt{n}} \middle| E(\bar{X}) = \mu \right\} \\ &= P\left\{ \left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| > k \middle| E(\bar{X}) = \mu \right\} = P\}|Z| > k\} = 2\Phi(-k),\end{aligned}$$

where Φ is the cumulative standard normal distribution function.

The Type 2 error probability, β, is the probability of not detecting an out-of-control condition. Here we assume that an out-of-control condition means that the process mean has shifted to $\mu + \delta\sigma$ or to $\mu - \delta\sigma$. Suppose that we condition on the event that the mean has shifted from μ to $\mu + \delta\sigma$. The probability that the shift is not detected after observing a sample of n observations is

$$\begin{aligned}\beta &= P\left\{ \left|\bar{X} - \mu\right| \le \frac{k\sigma}{\sqrt{n}} \middle| E(\bar{X}) = \mu + \delta\sigma \right\} \\ &= P\left\{ \frac{-k\sigma}{\sqrt{n}} \le \bar{X} - \mu \le \frac{k\sigma}{\sqrt{n}} \middle| E(\bar{X}) = \mu + \delta\sigma \right\} \\ &= P\left\{ -k - \delta\sqrt{n} \le \frac{\bar{X} - \mu - \delta\sigma}{\sigma/\sqrt{n}} \le k - \delta\sqrt{n} \middle| E(\bar{X}) = \mu + \delta\sigma \right\} \\ &= P\{-k - \delta\sqrt{n} \le Z \le k - \delta\sqrt{n}\} \\ &= \Phi(k - \delta\sqrt{n}) - \Phi(-k - \delta\sqrt{n}).\end{aligned}$$

If we had conditioned on $E(\bar{X}) = \mu - \delta\sigma$, we would have obtained

$$\beta = \Phi(k + \delta\sqrt{n}) - \Phi(-k + \delta\sqrt{n}).$$

Using the symmetry of the normal distribution [specifically, that $\Phi(t) = 1 - \Phi(-t)$ for any t], it is easy to show that these two expressions for β are the same.

[3] A detailed discussion of the memoryless property of the exponential distribution is given in Section 12.1.

Consider the random variables T and S. We assume that T is a geometric random variable assuming values 0, 1, 2, ... One can show that

$$E(T) = \frac{1 - \pi}{\pi}$$

(see, for example, DeGroot, 1986). The random variable S is the number of periods that the process remains out of control after a shift occurs. The probability that a shift is not detected when the process is out of control is exactly β. It follows that S is also a geometric random variable except that it assumes only the values 1, 2, 3, ... That is,

$$P\{S = s\} = (1 - \beta)\beta^{s-1} \qquad \text{for } s = 1, 2, 3, \ldots .$$

Because S is defined on the set 1, 2, 3, . . . , the expected value of S is $E(S) = 1/(1 - \beta)$. It follows that the expected cycle length, say C, is given by

$$E(C) = E(T + S) = E(T) + E(S) = \frac{1 - \pi}{\pi} + \frac{1}{1 - \beta}.$$

Consider the expected sampling cost incurred in a cycle. Each period there are n items sampled. As there are, on the average, $E(C)$ periods per cycle, it follows that the sampling cost per cycle is $a_1 nE(C)$.

We now compute the expected search cost. The process is shut down each time an out-of-control signal is observed. One or more of these signals could be a false alarm. Suppose that there are exactly M false alarms in a cycle. The random variable M has the binomial distribution with probability of "success" (i.e., a false alarm) of α for a total of T trials. It follows that $E(M) = \alpha E(T)$. The expected number of searches per cycle is exactly $1 + E(M)$, as the final search is assumed to discover and correct the assignable cause. Hence, the total search cost in a cycle is

$$a_2[1 + \alpha E(T)] = a_2[1 + \alpha(1 - \pi)/\pi].$$

We also assume that there is a cost of a_3 for each period that the process is operated in an out-of-control condition. The process is out of control for exactly S periods. Hence, the expected out-of-control cost is $a_3E(S) = a_3/(1 - \beta)$.

It follows that the expected cost per cycle is

$$a_1 nE(C) + a_2[1 + \alpha(1 - \pi)/\pi] + a_3/(1 - \beta).$$

Dividing by the expected length of a cycle, $E(C)$, gives the average cost per unit time as

$$\begin{aligned} a_1 n + \frac{a_2\left[1 + \alpha\dfrac{1 - \pi}{\pi}\right] + \dfrac{a_3}{1 - \beta}}{\dfrac{1 - \pi}{\pi} + \dfrac{1}{1 - \beta}} &= a_1 n + \frac{a_2\left[1 + \alpha\dfrac{1 - \pi}{\pi}\right] + \dfrac{a_3}{1 - \beta}}{\dfrac{1 - \beta(1 - \pi)}{(1 - \beta)\pi}} \\ &= a_1 n + \frac{a_2(1 - \beta)[\pi + \alpha(1 - \pi)] + a_3\pi}{1 - \beta(1 - \pi)}. \end{aligned}$$

We will write this as $G(n, k)$ to indicate that the optimization requires searching for the best n and k, where $n = 1, 2, 3, \ldots$ and $k > 0$. Note that α depends on k and β depends on both n and k. The goal is to find the values of n and k that minimize $G(n, k)$. This is a complex optimization problem because both α and β require evaluation of the cumulative normal distribution function.

Example *12.5*

Consider Example 12.1 of Wonderdisk, introduced in Section 12.2. Howard Hamilton would like to design an $\overline{X}$ chart in an economically optimal fashion. Based on his experience with the process and an analysis of the past history of failures, he decides that the geometric distribution accurately describes changes in the process state.

In order to use the model described in this section, he must estimate various costs and system parameters. The first is the sampling cost. Here sampling requires measuring the length of a tracking arm. This requires moving the arm to a different location, mounting it on a special brace to protect it, and measuring the length with calipers designed for the purpose. The process requires about 12 minutes of a technician's time. The technician is paid \$15 per hour, so the sampling cost is \$15/5 = \$3 per item sampled.

The second cost to estimate is the search cost. The time spent searching for an assignable cause of an out-of-control signal is usually about 30 minutes. If a problem is not discovered within that time, it is generally assumed that the out-of-control signal was a false alarm. The arms generate a revenue for the company of about \$1,200 daily. Assuming an eight-hour workday, the cost of shutting down production comes to about \$1,200/8 = \$150 per hour. Hence, the search cost is \$75.

The third cost required by the model is the cost of operating the process in an out-of-control condition. If the process is out of control, the proportion of defective arms produced increases. Most of the defective arms show up in the final testing phase of the drives. If a drive has a defective arm, the drive is disassembled and the arm replaced. Some defective arms pass inspection and are shipped to the customer with the disk drive. Wonderdisk provides purchasers a 14-month warranty, and it is likely that a problem with the drive will develop during the warranty period if the arm is defective. Howard estimates that the cost of operating the process out of control is about \$300 per hour, but he is not very confident about this estimate.

The model also requires estimates of π and δ. Recall that π represents the probability that the process will shift from an in-control state to an out-of-control state during one period. In the past, out-of-control signals have occurred at a rate of about one for every 10 hours of operation. As half of these have been false alarms, a reasonable estimate of the proportion of periods in which a shift has occurred is about one out of 20, or $\pi = .05$. The constant δ represents the degree of the shift as measured in multiples of the process standard deviation. In the past, the shifts have averaged about one standard deviation, so the estimate of δ is 1.

In order to simplify the calculations, Howard decides to use an approximation to the standard normal cumulative distribution function. The one he uses is the following:

$$\Phi(z) = 0.500232 - 0.212159z^{2.08388} + 0.5170198z^{1.068529} + 0.041111z^{2.82894}.$$

This approximation, due to Herron (1985), is accurate to within 0.5 percent for $0 < z < 3$. Howard Hamilton decides that this is accurate enough for his purposes. The optimization scheme he adopts is the following. Because n is a discrete variable and represents the number of items sampled in each subgroup, it is unlikely that n would exceed 10. Furthermore, because k is the number of standard deviations of $\overline{X}$ used in the control chart, it is unlikely that k would exceed 3. Howard writes a computer program to evaluate $G(n, k)$ for $k = 0, 0.1, 0.2, \ldots, 2.8, 2.9, 3.0$ and $n = 1, 2, 3, \ldots, 10$. (These calculations were actually done using a popular spreadsheet program.) For each fixed value of n, the function $G(n, k)$ appears to be convex in the variable k (convex functions were discussed in Chapters 4 and 5). For the given parameter values and $n = 4$, Figure 12–12 shows the function $G(n, k)$ as a function of k. The graph shows that the minimum cost occurs at about $k = 1.7$ and equals about \$45 per hour.

FIGURE 12–12
The behavior of $G(n, k)$ as a function of k (refer to Example 12.5)

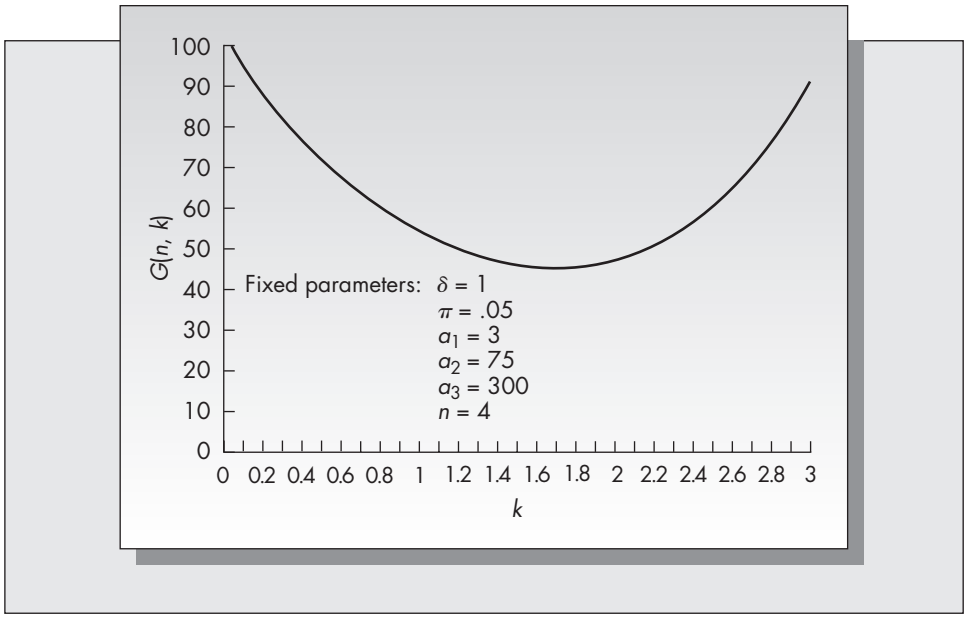


TABLE 12–4
Optimal Values of k for Various n

Fixed parameters are: $\delta = 1$, $\pi = 0.05$, $a_1 = 3$. $a_2 = 75$, $a_3 = 300$.

n	Optimal k	α	β	Cost
1	1.13	.25	.54	$54.4
2	1.52	.14	.50	48.8
3	1.60	.11	.44	46.1
4	1.74	.08	.39	45.2
5	1.86	.07	.34	45.3
6	1.97	.05	.30	46.2
7	2.06	.04	.26	47.6
8	2.14	.03	.23	49.3
9	2.21	.02	.21	51.3
10	2.27	.02	.18	53.4

Table 12–4 gives the results of the calculations for these parameter settings. According to the table, the optimal subgroup size is 4 and the optimal k is 1.74. These results were reasonably close to the current policy and give Hamilton confidence that his estimates were at least in the right ballpark. When he presented the results to his boss, however, his boss expressed concern about the large value of β, the probability of Type 2 error:

> You mean to tell me that if we operate the system optimally, then there is almost a 40 percent chance that we won't be able to detect when the system has gone out of control? That sounds pretty high. Considering the push that is on all over the company for an improvement in quality, I find that figure very disturbing. How does that turn out to be optimal?

The reason that the Type 2 error was so large was the assumption that the cost to the company was $300 for each hour that the process was operated out of control. After giving the

TABLE 12–5
Optimal Values of k for Various n (Revised) Fixed Parameters: Same as in Table 12–3 except $a_3 = 1{,}000$

n	Optimal k	α	β	Cost
1	0.67	.48	.30	$111.6
2	0.93	.33	.29	102.8
3	1.13	.25	.26	96.9
4	1.35	.18	.25	93.3
5	1.43	.15	.21	91.3
6	1.59	.11	.19	90.5
7	1.70	.09	.16	90.4
8	1.79	.07	.14	90.9
9	1.86	.06	.12	91.8
10	1.94	.05	.11	93.3

matter some thought, Howard's boss decided that a value of $1,000 was probably closer to the mark and was more consistent with the corporate goal of improving quality. Howard repeated his calculations substituting a value of $a_3 = 1{,}000$. The results are presented in Table 12–5.

With the revised value of $a_3 = 1{,}000$, the optimal value of the sample size n increased to 7 with a corresponding value of $k = 1.70$. The cost per hour at the optimal solution increased to $90.40. Although the limits on the control chart remained the same, the effect of increasing the sample size from 4 to 7 resulted in a dramatic decrease in the value of β from .39 to .16. Howard's new boss was far happier with the values of α and β that resulted from the new design.

Problems for Section 12.6

23. A quality control engineer is considering the optimal design of an $\overline{X}$ chart. Based on his experience with the production process, there is a probability of .03 that the process shifts from an in-control to an out-of-control state in any period. When the process shifts out of control, it can be attributed to a single assignable cause; the magnitude of the shift is 2σ. Samples of n items are made hourly, and each sampling costs $0.50 per unit. The cost of searching for the assignable cause is $25, and the cost of operating the process in an out-of-control state is $300 per hour.
    a. Determine the hourly cost of operating the system when $n = 6$ and $k = 2.5$.
    b. Estimate the optimal value of k for the case $n = 6$. If you are doing the calculations by hand, use $k = 0.5, 1, 1.5, 2, 2.5$, and 3.0. If you are using a computer, use $k = 0.1, 0.2, \ldots, 2.9, 3.0$.
    c. Determine the optimal control chart design that minimizes average annual costs.
24. Consider the application of the economic design of $\overline{X}$ charts for Wonderdisk presented in this section. Without actually performing the calculations, discuss what the effect on the optimal values of n and k is likely to be if
    a. δ increased from 1 to 2.
    b. π increased from .05 to .10.
    c. a_1 decreased to 1.
    d. a_2 increased to 150.

25. Under what circumstances would the following assumptions not be accurate?
    a. The assumption that the probability law describing the number of periods until the process goes out of control follows the geometric distribution.
    b. The assumption that an out-of-control condition corresponds to a shift of the mean equal to $\delta\sigma$.
    c. The assumption that the search cost is a fixed constant, a_2.
26. Discuss the following pro and con positions on using optimization models to design control charts:

    Con: "These models are useless to me because I don't feel I can accurately estimate the values of the required inputs."

    Pro: "The choice of specific values of n and k in the construction of X bar charts means that you are assuming values for the various system costs and parameters. You might as well take the bull by the horns and obtain the best estimates you can and use those to design the X bar chart."

*27. Suppose that "the process going out of control" corresponds to a shift of the mean from μ to $\mu + \sigma$ with probability .25, $\mu + 2\sigma$ with probability .25, $\mu - \sigma$ with probability .25, and $\mu - 2\sigma$ with probability .25. What modifications in the model are required? In particular, if we write the shift in the form $\mu \pm \delta\sigma$, show how to compute β_1 and β_2 that would correspond to values of $\delta = 1$ and $\delta = 2$, respectively. If the out-of-control costs were now represented by a_3 when $\delta = 1$ and a_4 when $\delta = 2$, determine an expression for the average annual operating costs. (Assume all other costs and system parameters remain the same.)
28. A local contractor manufactures the speakers used in telephones. The phone company requires the speakers to ring at a specified noise level (in decibels). An $\overline{X}$ chart is being designed to monitor this variable. The process of sampling speakers from the line requires hitting the speakers with a fixed-force clapper and measuring the decibel level on a meter designed for that purpose. The cost of sampling is \$1.25 per speaker. When the process goes out of control, the thickness of the speakers is incorrect. The cost of searching for an assignable cause is estimated to be \$50. The cost of operating the process in an out-of-control state is estimated to be \$180 per hour. Out of control corresponds to a shift of 2σ in the decibel level, and the probability that the process shifts out of control in any hour is .03.
    a. The company uses an $\overline{X}$ chart based on 3σ limits and subgroups of size 4. What is its hourly cost?
    b. What are the optimal values of n and k for this process and the associated optimal cost?

12.7 OVERVIEW OF ACCEPTANCE SAMPLING

Control charts provide a convenient way to monitor a process in real time to determine if a shift in the process parameters appears to have occurred. Another important aspect of quality control is to determine the quality of manufactured goods *after* they have been produced. In most cases 100 percent inspection is either impossible or impractical. Hence, a sample of items is inspected and quality parameters of large lots of items are estimated based on the results of the sampling.

To be more specific, acceptance sampling addresses the following problem: If a sample is drawn from a large lot of items and the sample is subject to 100 percent

inspection, what inferences can we draw about the quality of the lot based on the quality of the sample? Statistical analysis provides a means for extrapolating the characteristics of a sample to the characteristics of the lot, and a means for determining the probability of coming to the wrong conclusion.

Obviously, 100 percent inspection of all items in the lot will reduce the probability of an incorrect conclusion to zero. However, there are several reasons that 100 percent inspection is either not feasible or not desirable. Some of these include

1. In most cases 100 percent inspection is too costly. It is virtually impossible for high-volume transfer lines and continuous production processes.
2. In some cases 100 percent inspection may be impossible, such as when inspection involves destructive testing of the item. For example, determining the lifetime of a light bulb requires burning the bulb until it fails.
3. If the inspection is done by the consumer rather than the producer, 100 percent inspection by the consumer provides little incentive to the producer to improve quality. It is cheaper for the producer to repair or replace the items returned by the consumer than it is to improve the quality of the production process. However, if the consumer returns the entire lot based on the results of sampling, it provides a much greater motivation to the producer to improve the quality of outgoing lots.

In this chapter we treat the following three sampling plans.

1. *Single sampling plans*. Single sampling plans are by far the most popular and easiest to use of the plans we will discuss. Two numbers, n and c, determine a single sampling plan. If there are more than c defectives in a sample of size n, the lot is rejected; otherwise it is accepted.

2. *Double sampling plans*. In a double sampling plan, we first select a sample of size n_1. If the number of defectives in the sample is less than or equal to c_1, the lot is accepted. If the number of defectives is greater than c_2, then the lot is rejected. However, if the number of defectives is larger than c_1 and less than or equal to c_2, a second sample of size n_2 is drawn. The lot is now accepted if the cumulative number of defectives in both samples is less than or equal to a third number, c_3. (Often $c_3 = c_2$.)

3. *Sequential sampling*. A double sampling plan can obviously be extended to a triple sampling plan, which can be extended to a quadruple sampling plan, and so on. A sequential sampling plan is the logical conclusion of this process. Items are sampled one at a time and the cumulative number of defectives is recorded at each stage of the process. Based on the value of the cumulative number of defectives, there are three possible decisions at each stage:

   a. Reject the lot.
   b. Accept the lot.
   c. Continue sampling.

A complex sampling plan may have desirable statistical properties, but the acceptance and rejection regions could be difficult to calculate and the plan difficult to implement. The right sampling plan for a particular environment may not be the most mathematically sophisticated. As with any analytical tool, the potential benefits must be weighed against the potential costs.

Kolesar (1993) makes the point that with improving quality standards, the value of acceptance sampling may diminish in years to come. Motorola has become famous

Snapshot Application

NAVISTAR SCORES WITH SIX-SIGMA QUALITY PROGRAM

Navistar International is a major U.S. manufacturer of trucks, buses, and engines and has several plants around the world. In 1985 Navistar's worldwide workforce numbered over 110,000. Because of a crippling United Auto Workers (UAW) strike and a recession, the company had to severely trim the workforce to survive. Today the workforce numbers around 20,000. To combat cost and quality problems it was experiencing at the time, Navistar decided to launch a six-sigma quality program in the mid-1990s. As noted in this section, six-sigma means defect rates of 3.4 parts per million or less. While six-sigma programs rarely achieve such low defect rates, the goal is clear: Do what needs to be done in the organization to effect a fundamental change in both management's and labor's attitudes about quality. Quality programs do not come free, however. Navistar paid a consulting company more than $6 million to implement this program. One immediate result was that Navistar's stock price grew over 400 percent in the 14 months following implementation of the program. (Of course, as we all know, the price of a company's stock is influenced by many factors, so it isn't clear what role the six-sigma program played.)

Six-sigma programs have their own culture. Specially trained employees are dubbed black belts after one month's training, and master black belts after additional training. The black belts are assigned specific projects and have the power to go directly to top management with proposed solutions. Of course, for such an approach to work, not only the employees, but also the management, must be firmly committed to the program. Does everyone believe in the value of six-sigma programs? Evidently not; for example, Charles Holland, president of a consulting company based in Knoxville that specializes in statistical quality control methods, dubs the six-sigma program as a "silver bullet" sold at "outrageous prices."[1]

If this is true, what motivated Navistar to plunk down $6 million for this program? According to John Horne, the company's chief executive in 1995, the company needed an antidote to the slide it was experiencing: "We didn't have a strategy; most companies don't." The strategy that Horne adopted was to go after the company's problems at the plant level. Quality control problems had been dogging Navistar's plants for years. The target of the six-sigma program was the massive 4,000 square foot plant in Springfield, Ohio. (Navistar did not implement six-sigma in all its plants for various reasons. For example, union opposition prevented implementation in the Canadian plant located in Chatham, Ontario.)

What was the result in Springfield? The effort has been credited with $1 million of savings the first year, and greater savings in subsequent years. The total savings in this one plant alone was projected to be $26 million, well above the $6 million cost of the program. Sometimes kaizen (continuous improvement) is simply not enough to fix a troubled system. While expensive, six sigma can provide the jump start needed to turn things around, as it did with Navistar.

[1] Franklin, S., "In Pursuit of Perfection," *Chicago Tribune,* Sunday April 4, 1999, section 5, pp. 7–8.

for instituting its "six-sigma" quality thrust (Motorola's quality initiatives are discussed in the Snapshot Application in Section 12.12). By this they mean that the defect rate should be no more than the area outside of $\pm 6\sigma$ under a normal curve. This translates to defect rates of less than 3.4 parts per million. (Few tables of the normal distribution go past 4σ, so you will have a difficult time verifying this probability.) When defect rates are so low, acceptance sampling becomes very inefficient. For example, suppose the defect rate increased by a factor of 10. In that case, the probability of finding a defect in a sample of, say, 1,000 units would be only .034. We wouldn't expect to see a single defect until we have sampled at least 29 lots on average! However, we should keep in mind that Motorola's six-sigma quality standard has not become an industry standard by a long shot. Acceptance sampling will remain a valuable tool for many years to come.

12.8 NOTATION

We will use the following notation throughout the remainder of this chapter.

N = Number of pieces in a given lot or batch.
n = Number of pieces in the sample ($n < N$).
M = Number of defectives in the lot.
β = Consumer's risk—the probability of accepting bad lots.
α = Producer's risk—the probability of rejecting good lots.
c = Rejection level.
X = Number of defectives in the sample.
p = Proportion of defectives in the lot.
p_0 = Acceptable quality level (AQL).
p_1 = Lot tolerance percent defective (LTPD).

Assume that N is a known constant. If N is very large relative to the sample size, n, it can be assumed to be infinite. In that case it will not enter into the calculations. Although M is a constant as well, its value is *not* known in advance. In fact, only 100 percent inspection will reveal the true value of M. Often we are interested in analyzing the behavior of the sampling plan for various values of M. The consumer's risk and the producer's risk depend upon the sampling plan. Finally, X, the number of defectives in the sample, is a random variable. This means that were we to repeat the sampling experiment with a different random sample of size n, we would not necessarily observe the same number of defectives. Based on statistical properties of the population as a whole, we can determine the form of the probability distribution of X.

The acceptable quality level, p_0, is the desired or target level of the proportion of defectives in the lot. If the true proportion of defectives in the lot is less than or equal to p_0, the lot is considered to be acceptable. The lot tolerance percent defective, p_1, is an unacceptable proportion of defectives in the lot. The lot is considered unacceptable if the proportion of defectives exceeds p_1. Because of the imprecision of statistical sampling, we allow a gray area between p_0 and p_1. When the AQL and LTPD are equal, large sample sizes may be required to achieve acceptable values of α and β.

12.9 SINGLE SAMPLING FOR ATTRIBUTES

The goal of all sampling procedures is to estimate the properties of a population from the properties of the sample. In particular, we wish to test the hypotheses

$$H_0\text{: Lot is of acceptable quality } (p \leq p_0).$$
$$H_1\text{: Lot is of unacceptable quality } (p \geq p_1).$$

The test is of the form: Reject H_0 if $X > c$. The value of c depends on the choice of the Type 1 error probability α. The Type 1 error probability is the probability of rejecting H_0 when it is true. In the context of the quality control problem, this is the probability of rejecting the lot when it is acceptable. This is also known as the producer's risk. In equation form,

$$\begin{aligned}\alpha &= P\{\text{Reject } H_0 \mid H_0 \text{ true}\} = P\{\text{Reject lot} \mid \text{Lot is good}\} \\ &= P\{X > c \mid p = p_0\}.\end{aligned}$$

The exact distribution of X is *hypergeometric* with parameters n, N, and M. That is,

$$P\{X = m\} = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}} \qquad \text{for } 0 \le m \le \min(M, n),$$

where

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}.$$

In most applications, N is much larger than n, so the binomial approximation to the hypergeometric is satisfactory. In that case

$$P\{X = m\} = \binom{n}{m} p^m (1-p)^{n-m} \qquad \text{for } 0 \le m \le n,$$

where $p = M/N$ is the true proportion of defectives in the lot.

Using the binomial approximation, the producer's risk and the consumer's risk are given by

$$\alpha = P\{X > c \,|\, p = p_0\} = \sum_{m=c+1}^{n} \binom{n}{m} p_0^m (1-p_0)^{n-m},$$

$$\beta = P\{X \le c \,|\, p = p_1\} = \sum_{m=0}^{c} \binom{n}{m} p_1^m (1-p_1)^{n-m}.$$

Most statistical tests require specification of the probability of Type 1 error, α. Values of α, n, and p_0 will determine a unique value of c, which can be found from tables of the cumulative binomial distribution. However, because the binomial is a discrete distribution, it may not be possible to find c to match exactly the desired value of α. When p is small and n is moderately large ($n > 25$ and $np < 5$), the Poisson distribution provides an adequate approximation to the binomial. For very large values of n such that $np(1-p) > 5$, the normal distribution provides an adequate approximation to the binomial. Refer to Appendix 12–A for a detailed discussion of these approximations.

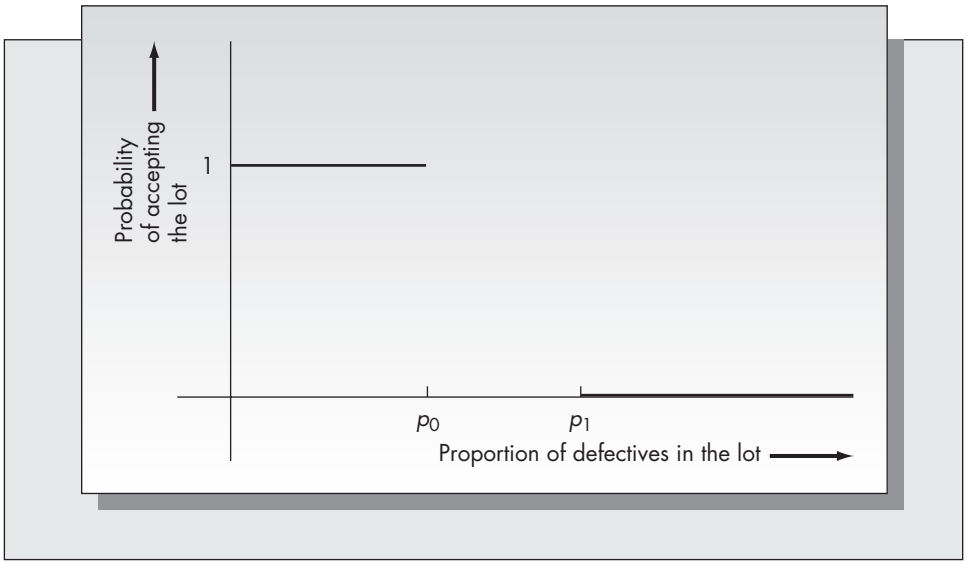
Example *12.6*

Spire CDs is a large West Coast retail chain of stores specializing in CDs. One of Spire's suppliers is B&G CDs, which ships CDs to Spire in 100-CD lots. After some negotiation, Spire and B&G have agreed that a 10 percent rate of defectives is acceptable and a 30 percent rate of defectives is unacceptable. From each lot of 100 CDs, Spire has established the following sampling plan: 10 CDs are sampled, and if more than 2 are found to be warped, scratched, or defective in some other way, the lot is rejected. Consider the consumer's and the producer's risk associated with this sampling plan.

From the given information, we have that $p_0 = .1$, $p_1 = .3$, $n = 10$, and $c = 2$. Hence,

$$\begin{aligned}\alpha &= P\{X > c \,|\, p = p_0\} = P\{X > 2 \,|\, p = .1\} = 1 - P\{X \le 2 \,|\, p = .1\} \\ &= 1 - \sum_{k=0}^{2} \binom{10}{k} (.1)^k (.9)^{10-k} = 1 - .9298 = .0702. \\ \beta &= P\{X \le c \,|\, p = p_1\} = P\{X \le 2 \,|\, p = .3\} \\ &= \sum_{k=0}^{2} \binom{10}{k} (.3)^k (.7)^{10-k} = .3828.\end{aligned}$$

FIGURE 12–13
The ideal OC curve



Note that the parameter values of $n = 10$, $p = .1$, and $n = 10$, $p = .3$ imply that neither the normal nor the Poisson approximation is accurate. (The reader may wish to check, using Table A–3 at the back of this book, that using the Poisson distribution with $\lambda = np$ gives α and β the approximate values of .0803 and .4216, respectively.)

Derivation of the OC Curve

The operating characteristic (OC) curve measures the effectiveness of a test to screen lots of varying quality. The OC curve is a function of p, the true proportion of defectives in the lot, and is given by

$$\text{OC}(p) = P\{\text{Accepting the lot} \mid \text{True proportion of defectives} = p\}.$$

We will now derive the form of the OC curve for the particular case of a single sampling plan with sample size n and rejection level c. In that case,

$$\begin{aligned}\text{OC}(p) &= P\{X \le c \mid \text{Proportion of defectives in lot} = p\} \\ &= \sum_{k=0}^{c} \binom{n}{k} p^k (1-p)^{n-k}.\end{aligned}$$

Ideally, the sampling procedure would be able to distinguish perfectly between good and bad lots. Figure 12–13 shows the ideal OC curve.

Example *12.6 (continued)*

Consider again Example 12.6 of Spire Records. The OC curve for its single sampling plan is given by

$$\text{OC}(p) = \sum_{k=0}^{2} \binom{10}{k} p^k (1-p)^{10-k}.$$

The graph of Spire's OC curve appears in Figure 12–14. An examination of the figure shows that this particular sampling plan is more advantageous for the supplier, B&G, than it is for Spire. The value of $\beta = .3828$ means that Spire is passing almost 40 percent of the lots that contain 30 percent defectives. Furthermore, the probability of accepting lots with proportions of defectives as high as 40 percent and even 50 percent is not negligible. This accounts for Spire's experience that there seemed to be many customer returns of B&G label CDs.

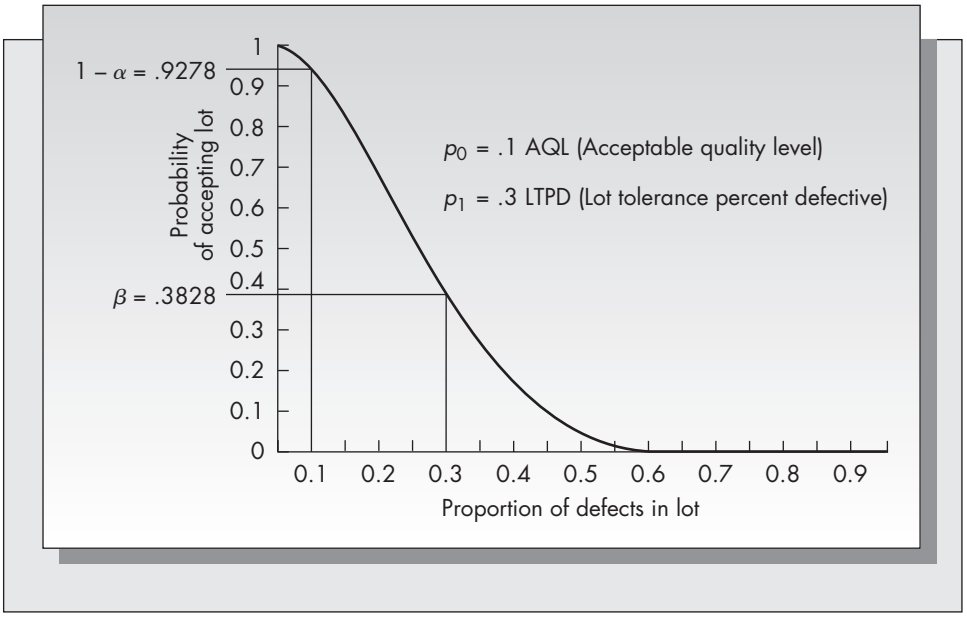


FIGURE 12–14
OC curve for Spire CDs ($n = 10$)

Herman Sondle, an employee of Spire enrolled in a local Masters program, was asked to look into the problem with B&G CDs. He discovered the cause of the trouble by analysis of the OC curve pictured in Figure 12–14. In order to decrease the chances that Spire receives bad lots from B&G, he suggested that the sampling plan be modified by setting $c = 0$. The resulting consumer's risk is

$$\beta = P\{X \leq 0 \mid p = .3\} = (.3)^0(.7)^{10} = .028,$$

or approximately 3 percent. This seemed to be an acceptable level of risk, so the firm instituted this policy. Unfortunately, the proportion of rejected batches *increased* dramatically. The resulting value of the producer's risk, α, is

$$\alpha = P\{X > 0 \mid p = .1\} = 1 - P\{X = 0 \mid p = .1\} = 1 - (.9)^{10} = .6513.$$

That is, about 65 percent of the good batches were being rejected by Spire under the new plan. B&G threatened to discontinue shipments to Spire unless it returned to its original sampling plan.

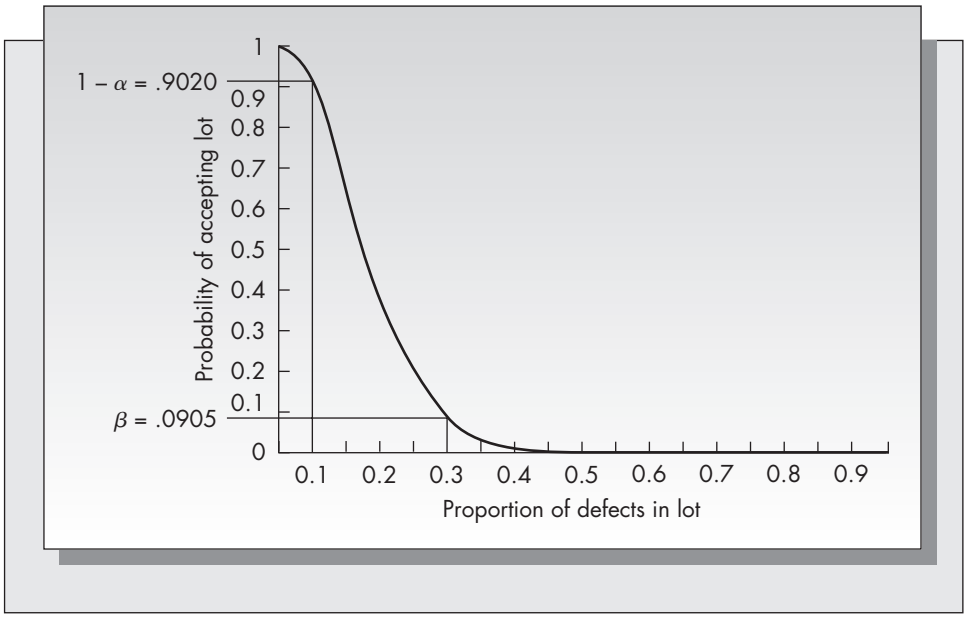
The Spire management didn't know what to do. If it returned to the original plan, it faced the risk of losing customers who would go elsewhere to purchase higher-quality CDs. If it continued with the current plan, it risked losing B&G as a supplier. Fortunately Sondle, who had been studying quality control methods, was able to propose a solution. If the sample size were increased, the power of the test would improve. Eventually, a test could be devised that would have acceptable levels of both the consumer's and the producer's risk. Because B&G insisted on no more than a 10 percent probability of rejecting good lots, Spire also wanted no more than a 10 percent probability of accepting bad lots.

After some experimentation, Herman found that a sample size of $n = 25$ with a rejection level of $c = 4$ seemed to meet the requirements of both B&G and Spire. The exact values of α and β for this test are

$$\alpha = P\{X > 4 \mid p = .1, n = 25\} = .0980.$$
$$\beta = P\{X \leq 4 \mid p = .3, n = 25\} = .0905.$$

Of course, the improved efficiency of this plan did not come without some cost. The employee time required to inspect B&G CDs increased by two and a half times. B&G and Spire

FIGURE 12–15
Revised OC curve for Spire CDs ($n = 25$)



agreed to share the additional cost of the inspection. The OC curve for the sampling plan with $n = 25$ and $c = 4$ appears in Figure 12–15. Notice how much more closely this approximates the ideal curve than does the OC curve for the original plan pictured in Figure 12–14.

Problems for Section 12.9

29. Samples of size 20 are drawn from lots of 100 items, and the lots are rejected if the number of defectives in the sample exceeds 2. If the true proportion of defectives in the lot is 5 percent, determine the probability that a lot is accepted using
    a. The exact hypergeometric distribution.
    b. The binomial approximation to the hypergeometric.
    c. The Poisson approximation to the binomial.
    d. The normal approximation to the binomial.
30. A producer of pocket calculators purchases the main processor chips in lots of 1,000. The producer would like to have a 1 percent rate of defectives but will normally not refuse a lot unless it has 4 percent or more defectives. Samples of 50 are drawn from each lot, and the lot is rejected if more than two defectives are found.
    a. What are p_0, p_1, n, and c for this problem?
    b. Compute α and β. Use the Poisson approximation for your calculations.
31. A company employs the following sampling plan: It draws a sample of 10 percent of the lot being inspected. If 1 percent or less of the sample is defective, the lot is accepted. Otherwise the lot is rejected.
    a. If a lot contains 500 items of which 10 are defective, what is the probability that the lot is accepted?

b. If a lot contains 1,000 items of which 20 are defective, what is the probability that the lot is accepted?

c. If a lot contains 10,000 items of which 200 are defective, what is the probability that the lot is accepted?

32. Hemispherical Conductor produces the 80J84 microprocessor, which the Sayle Company plans to use in a heart–lung machine. Because of the sensitivity of the application, Sayle has established a value of AQL of .001 and LTPD of .005. Sayle purchases the microprocessors in lots of 500 and tests 100 from each lot. The testing requires destruction of the microprocessor. The lot is rejected if any defectives are found.

a. What are the values of p_0, p_1, n, and c used?

b. Compute α and β.

c. In view of your answer to part (*b*), what problem could Sayle run into?

33. Determine a sampling plan for Spire CDs that results in $\alpha = .05$, $\beta = .05$, AQL = .10, and LTPD = .30. Discuss the advantages and the disadvantages of the plan you obtain as compared to the current plan of $n = 25$ and $c = 4$. (Refer to Example 12.6.)

*12.10 DOUBLE SAMPLING PLANS FOR ATTRIBUTES

Five numbers define a double sampling plan: n_1, n_2, c_1, c_2, and c_3. The plan is implemented in the following way: One draws an initial sample of size n_1 and determines the number of defectives in the sample. If the number of defectives in the sample is less than or equal to c_1, the lot is accepted. If the number of defectives in the sample is larger than c_2, the lot is rejected. However, if the number of defectives is larger than c_1 but less than or equal to c_2, another sample of size n_2 is drawn. If the number of defectives in the combined samples is less than or equal to c_3, the lot is accepted. If not, the lot is rejected. Most double sampling plans assume that $c_3 = c_2$. We will make that assumption as well from this point on.

A double sampling plan is obviously more difficult to construct and more difficult to implement than a single sampling plan. However, it does have some advantages over single plans. First, a double sampling plan may give similar levels of the consumer's and the producer's risks but require less sampling in the long run than a single plan. Also, there is the psychological advantage in double sampling plans of providing a second chance before rejecting a lot.

Example *12.7* Consider again Example 12.6 concerning Spire CDs. Herman Sondle decides to experiment with a few double sampling plans to see if he can achieve similar levels of efficiency with less sampling. Unfortunately, because the plan depends on four different numbers, considerable trial-and-error experimentation is necessary. Let us consider the computation of the consumer's and the producer's risks for the following sampling plan:

$$n_1 = 20, \quad c_1 = 3,$$
$$n_2 = 10, \quad c_2 = 5.$$

Define

X = Number of defectives observed in the first sample.

Y = Number of defectives observed in the second sample.

Z = Number of defectives observed in the combined samples ($Z = X + Y$).

The OC curve is

$$\begin{aligned}\text{OC}(p) &= p\{\text{Lot is accepted} \mid p\} = P\{\text{Lot is accepted on first sample} \mid p\} \\ &\quad + P\{\text{Lot is accepted on second sample} \mid p\}\end{aligned}$$

where

$$P\{\text{Lot is accepted on the first sample} \mid p\} = P\{X \leq 3 \mid p\}$$

and

$$\begin{aligned}&P\{\text{Lot is accepted on the second sample} \mid p\} \\ &\quad = P\left\{\begin{array}{l}\text{Lot is neither accepted nor rejected on the} \\ \text{first sample and the lot is accepted on the} \\ \text{second sample}\end{array}\;\middle|\; p\right\} \\ &\quad = P\{3 < X \leq 5, Z \leq 5 \mid p\}.\end{aligned}$$

Computation of this joint probability must be done carefully, as X and Z are *dependent* random variables.

Consider $p = \text{AQL} = .1$.

$$\begin{aligned}&P\{\text{Lot is accepted on the first sample} \mid p = .1\} \\ &\quad = P\{X \leq 3 \mid p = .1, n = 20\} = .8670,\end{aligned}$$

$$\begin{aligned}&P\{\text{Lot is accepted on the second sample} \mid p = .1\} \\ &\quad = P\{X = 4 \mid p = .1, n = 20\}\, P\{Y \leq 1 \mid p = .1, n = 10\} \\ &\qquad + P\{X = 5 \mid p = .1, n = 20\}\, P\{Y \leq 0 \mid p = .1, n = 10\} \\ &\quad = (.0898)(.7361) + (.0319)(.3487) = .0772.\end{aligned}$$

Summing:

$$P\{\text{Lot is accepted} \mid p = .1\} = .8670 + .0772 = .9442.$$

Repeating similar calculations with $p = .3$ gives

$$\begin{aligned}&P\{\text{Lot is accepted} \mid p = .3\} \\ &\quad = .1071 + (.1304)(.1493) + (.1789)(.0282) \\ &\quad = .1316.\end{aligned}$$

Hence, it follows that for this case we obtain

$$\begin{aligned}\alpha &= 1 - .9442 = .0558, \\ \beta &= .1316.\end{aligned}$$

Experimentation with other values of n_1, n_2, c_1, and c_2 can lead to double sampling plans that more closely match the desired values of α and β. Tables are available for optimizing double sampling plans. (See, for example, Duncan, 1986, pp. 232–33.)

Problems for Section 12.10

34. Consider the double sampling plan for Spire CDs presented in this section.
    a. Suppose that the true proportion of defectives in the lot is 10 percent. On average, how many items will have to be sampled before the lot is either accepted or rejected?
    b. Suppose that the true proportion of defectives in the lot is 30 percent. On average, how many items will have to be sampled before the lot is either accepted or rejected?

35. For the double sampling plan for Spire CDs presented in this section, what is the probability that a lot is rejected on the first sample? Perform the computation for both $p = p_0$ and $p = p_1$.
36. Consider the double sampling plan for Spire CDs described in this section. Over a period of one year, 3,860 boxes of records are subject to inspection using this plan. If 60 percent of these batches are "good" (that is, in 60 percent of the batches the proportion of defectives is exactly 10 percent) and 40 percent are "bad" (that is, in 40 percent of the batches the proportion of defectives is exactly 30 percent), then what is the expected number of batches
    a. Accepted?
    b. Rejected?
    c. Accepted on the first sample?
    d. Accepted on the second sample?
    e. Rejected on the first sample?
    f. Rejected on the second sample?
37. Graph the OC curve for the double sampling plan with $n_1 = 20$, $n_2 = 10$, $c_1 = 3$, and $c_2 = c_3 = 5$, as described in this section. If you are doing this by hand, evaluate the curve at $p = 0, .2, .4, .6, .8$, and 1 only. (Hint: The OC curve for this sampling plan has the form

$$\begin{aligned} \text{OC}(p) = {} & P\{X \leq 3 \mid p, n = 20\} \\ & + P\{X = 4 \mid p, n = 20\}\, P\{Y \leq 1 \mid p, n = 10\} \\ & + P\{X = 5 \mid p, n = 20\}\, P\{Y = 0 \mid p, n = 10\}.) \end{aligned}$$

38. By trial and error devise a double sampling plan for Spire CDs that achieves $\alpha \approx .10$ and $\beta \approx .10$.
39. Consider the following double sampling plan. First select a sample of 5 from a lot of 100. If there are four or more defectives in the sample, reject the lot. If there is one or fewer defective, accept the lot. If there are two or three defectives, sample an additional five items and reject the lot if the combined number of defectives in both samples is five or more. If the lot has 10 defectives, what is the probability that a lot passes the inspection?
40. For the double sampling plan described in Problem 39, determine the following:
    a. The probability that the lot is rejected based on the first sample.
    b. The probability that the lot is rejected based on the second sample.
    c. The expected number of items sampled before the lot is accepted or rejected.

12.11 SEQUENTIAL SAMPLING PLANS

Double sampling plans may be extended to triple sampling plans, which also may be extended to higher-order plans. The logical conclusion of this process is the sequential sampling plan. In a sequential plan, items are sampled one at a time. After each sampling, two numbers are recorded: the number of items sampled and the cumulative number of defectives observed. Based on these numbers, one of three decisions is made: (1) accept the lot, (2) reject the lot, or (3) continue sampling. Unlike single and double sampling plans, there will always exist a sequential sampling plan that will give specific values of p_0, p_1, α, and β. Sequential sampling plans are defined by three

regions: the acceptance region, the rejection region, and the sampling region. The three regions are separated by straight lines. The lines have the forms

$$L_1 = -h_1 + sn,$$
$$L_2 = h_2 + sn,$$

where n is the number of items sampled. Note that L_1 and L_2 are both linear functions of the variable n. The y intercepts are respectively $-h_1$ and h_2, and the slope of each line is s. As the lines have the same slope, they are parallel. The sequential sampling plan is implemented in the following manner: The cumulative number of defectives is graphed, together with the lines for L_1 and L_2. When the cumulative number of defectives exceeds L_2, the lot is rejected, and when the cumulative number of defectives falls below L_1, the lot is accepted. As long as the cumulative number of defectives lies between L_1 and L_2, sampling continues.

Figure 12–16 shows two examples of the results of sampling for the same sequential sampling plan. In Case A the sampling led to acceptance of the lot, and in Case B it led to rejection of the lot.

The equations for h_1, h_2, and s are

$$h_1 = \frac{\log \dfrac{1-\alpha}{\beta}}{\log \dfrac{p_1(1-p_0)}{p_0(1-p_1)}},$$

$$h_2 = \frac{\log \dfrac{1-\beta}{\alpha}}{\log \dfrac{p_1(1-p_0)}{p_0(1-p_1)}},$$

$$s = \frac{\log \dfrac{1-p_0}{1-p_1}}{\log \dfrac{p_1(1-p_0)}{p_0(1-p_1)}}.$$

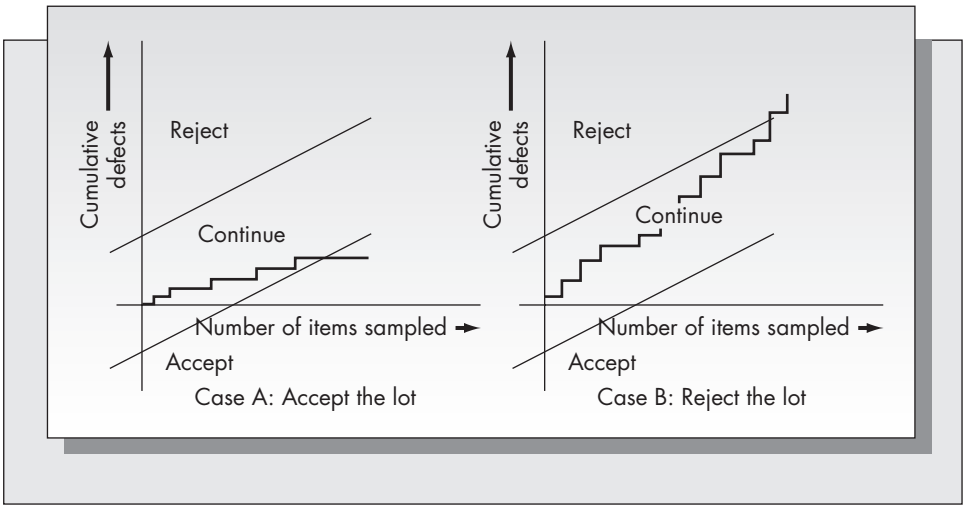


FIGURE 12–16
Two realizations of a sequential sampling plan

Example *12.8*

Consider again Example 12.6 of Spire CDs. Herman Sondle has experimented with various sampling plans to achieve the desired levels of the consumer's and the producer's risks. He decides to construct a sequential sampling plan to see how it compares with the single and the double plans that were presented earlier. Spire and B&G agreed on the following values of the AQL, LTPD, consumer's risk, and producer's risk:

$$p_0 = .1, \qquad \alpha = .1,$$
$$p_1 = .3, \qquad \beta = .1.$$

Notice that the denominators in the expressions for h_1, h_2, and s are the same. We compute the denominator first. (We will use log to the base 10 in our calculations. Because all formulas involve the ratio of logarithms, the results will be the same whether one uses base 10 or base e.)

$$\log[(.3)(.9)/(.1)(.7)] = 0.58626.$$

Hence,

$$h_1 = \log(.9/.1)/0.58626 = 0.9542/0.58626 = 1.6277,$$
$$h_2 = h_1 = 1.6277 \qquad (\text{since } \alpha = \beta \text{ for this case}),$$
$$s = \log(.9/.7)/0.58626 = 0.10914/0.58626 = 0.18617.$$

Figure 12–17 shows the decision regions for Spire's sequential sampling plan.

When Herman suggested that Spire CDs implement the sequential sampling plan shown in Figure 12–17, he met with considerable resistance from some of his co-workers responsible for the inspection of incoming stock. With a single sampling plan with $n = 25$, they argued that at least they would know in advance how many CDs they would have to check. With the sequential plan, however, they argued that they might have to sample the entire lot without the plan recommending either acceptance or rejection. Although Herman had heard that sequential plans were more efficient, he had difficulty convincing his co-workers to try the plan.

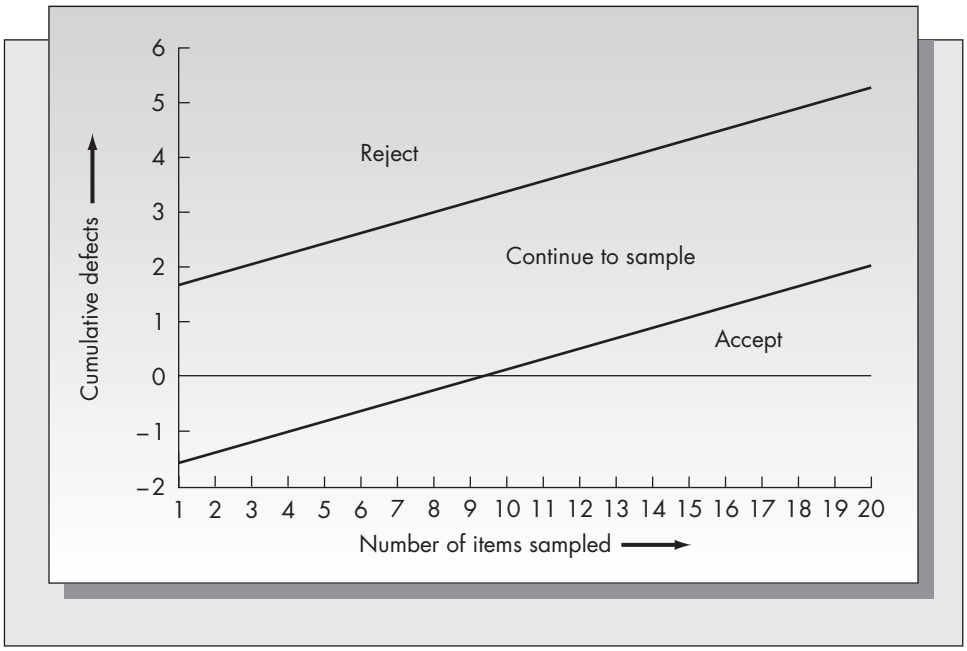


FIGURE 12–17
Sequential sampling plan for Spire CDs (refer to Example 12.8)

The co-workers in the example were correct in seeing that the number of items sampled when using sequential sampling is a random variable. The *expected* sample size that results from a sequential sampling plan depends on the proportion of defectives in the lot, p. The average sample number (ASN) curve gives the expected sample size for a sequential sampling plan as a function of p. We will estimate the ASN curve by obtaining its value at five specific points: when $p = 0$, $p = p_0$, $p = s$, $p = p_1$, and $p = 1$. It is easy to find the ASN curve at these points. In most cases one can obtain an adequate approximation to the ASN curve knowing only its value at these five points. The five values are

$$\text{At } p = 0, \qquad \text{ASN} = \frac{h_1}{s}.$$

$$\text{At } p = p_0, \qquad \text{ASN} = \frac{(1-\alpha)h_1 - \alpha h_2}{s - p_0}.$$

$$\text{At } p = s, \qquad \text{ASN} = \frac{h_1 h_2}{s(1-s)}.$$

$$\text{At } p = p_1, \qquad \text{ASN} = \frac{(1-\beta)h_2 - \beta h_1}{p_1 - s}.$$

$$\text{At } p = 1, \qquad \text{ASN} = \frac{h_2}{1-s}.$$

Consider the application of these formulas to Spire CDs:

$$\text{ASN}(0) = \frac{1.6277}{0.18617} = 8.74,$$

$$\text{ASN}(.1) = \frac{(.9)(1.6277) - (.1)(1.6277)}{0.18617 - .1} = 15.11,$$

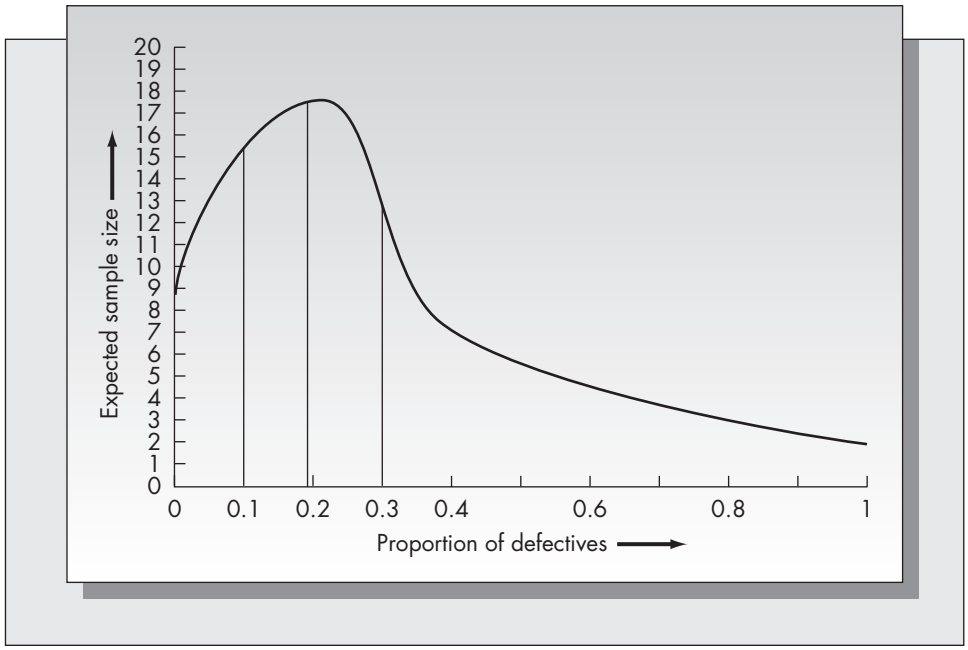
$$\text{ASN}(.18617) = \frac{(1.6277)(1.6277)}{(0.18617)(1 - 0.18617)} = 17.49,$$

$$\text{ASN}(.3) = \frac{(.9)(1.6277) - (.1)(1.6277)}{.3 - 0.18617} = 11.44,$$

$$\text{ASN}(1) = \frac{1.6277}{1 - 0.18617} = 2.0.$$

The ASN curve is a unimodal curve whose maximum value lies between p_0 and p_1. Based on this and the five points computed, we obtain the estimated ASN curve shown in Figure 12–18. We see from the figure that the expected sample size required for the sequential sampling plan for Spire CDs will be at most 18 items. This is clearly an improvement over the single sampling plan with $n = 25$ and $c = 4$, which resulted in similar values of α and β. It is important to keep in mind, however, that the actual sample size required in the sequential plan for the inspection of any particular batch of CDs is a random variable. The ASN curve gives only the *expected* value of this random variable for any specified value of p. Thus, it is possible that in specific instances, the actual sample size could be larger than 18, or even larger than 25.

FIGURE 12–18
ASN curve for Spire CDs (estimated)



Problems for Section 12.11

41. A manufacturer of aircraft engines uses a sequential sampling plan to accept or reject incoming lots of microprocessors used in the engines. Assume an AQL of 1 percent and an LTPD of 5 percent. Determine a sequential sampling plan assuming $\alpha = .05$, $\beta = .10$. Graph the acceptance and rejection regions.
42. Consider the sequential sampling plan described in Problem 41. Suppose that a lot of 1,000 microprocessors is inspected. Suppose that the 31st, 89th, 121st, and 122nd chips tested are found defective. Assuming the sequential sampling plan derived in Problem 41, will the lot be accepted, rejected, or neither by the time the 122nd chip has been tested?
43. Estimate the ASN curve for the plan derived in Problem 41. According to your curve, what is the expected number of microprocessors that must be tested when the true proportion of defectives in the lot is
    a. 0.1 percent?
    b. 1.0 percent?
    c. 10 percent?
44. Consider the example of Hemispherical Conductor and the Sayle Company discussed in Problem 32. Devise a sequential sampling plan for Sayle that results in $\alpha = .05$ and $\beta = .20$. What are the advantages and disadvantages of this plan over the Sayle sampling plan derived in Problem 32?
45. Estimate the ASN curve for the sampling plan derived in Problem 44. On average, how many of the microprocessors would have to be tested if
    a. $p = .001$?
    b. $p = .005$?
    c. $p = .01$?

12.12 AVERAGE OUTGOING QUALITY

The purpose of a sampling plan is to screen out lots of unacceptable quality. However, because sampling is a statistical process, it is possible that bad lots will be passed and good lots will be rejected. A fundamental issue related to the effectiveness of any sampling plan is to determine the quality of product that results *after* the inspection process is completed.

The calculation of the average outgoing quality of an inspection process depends on the assumption that one makes about lots that do not pass inspection and the assumption that one makes about defective items. Assume that rejected lots are subject to 100 percent inspection. We derive the average outgoing quality curve under two conditions: (1) defective items in samples and in rejected lots are not replaced and (2) defective items in samples and in rejected lots are replaced.

The average outgoing quality (AOQ) is the long-run ratio of the expected number of defectives and the expected number of items successfully passing inspection. That is,

$$\text{AOQ} = \frac{E\{\text{outgoing number of defectives}\}}{E\{\text{outgoing number of items}\}}.$$

The OC curve is the probability that a lot is accepted as a function of p. That is,

$$\text{OC}(p) = P\{\text{lot is accepted} \mid p\}.$$

For convenience we will refer to this term as P_a.

Case 1: Defective items are not replaced. Suppose that lots are of size N and samples are of size n. Then the expected number of defectives and the expected number of items shipped are

		Number of Defectives	Number of Items
Accept lot	P_a	$(N - n)p$	$N - np$
Reject lot	$1 - P_a$	0	$N(1 - p)$

From this tree diagram we see that

$$\begin{aligned} E\{\text{outgoing number of defectives}\} &= P_a(N - n)p + (1 - P_a)(0) \\ &= P_a(N - n)p \end{aligned}$$

and

$$E\{\text{outgoing number of items}\} = P_a(N - np) + (1 - P_a)N(1 - p).$$

It follows that the ratio, AOQ, is given by

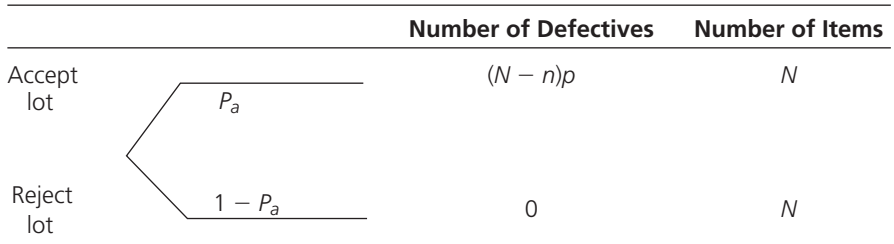
$$\text{AOQ} = \frac{P_a(N - n)p}{P_a(N - np) + (1 - P_a)N(1 - p)} = \frac{P_a(N - n)p}{N - np - p(1 - P_a)(N - n)}.$$

When $N \gg n$ (N is much larger than n), which is a common assumption, this expression is approximately

$$\text{AOQ} \approx \frac{P_a p}{P_a + (1 - P_a)(1 - p)} = \frac{P_a p}{1 - p(1 - P_a)}.$$

The formulas are somewhat simpler when defective items are replaced with good items.

Case 2: Defective items are replaced. In this case the tree diagram becomes



The AOQ is given by

$$\text{AOQ} = \frac{P_a(N - n)p}{N}.$$

which is approximately

$$\text{AOQ} \approx P_a p = \text{OC}(p)p$$

when $N \gg n$.

This last formula is the one most commonly used in practice, primarily because of its simplicity. An important measure of the effectiveness of a sampling plan is *the average outgoing quality limit* (AOQL), which is defined as the maximum value of the AOQ curve.

Example *12.9*

Consider the case of Spire CDs. In Spire's case we can assume that lots are large compared to samples and that all defectives are replaced. In that case AOQ ≈ OC(*p*)*p*. In Figure 12–19 we have generated the AOQ curves for Spire's single sampling plans with $n = 10$ and $n = 25$. Note how the larger sample size significantly improves the average outgoing quality. From Figure 12–19 we see that if Spire uses a single sampling plan with $n = 25$ and $c = 4$, the store can expect that the proportion of defective B&G CDs on its shelves will be no more than about 10.2 percent.

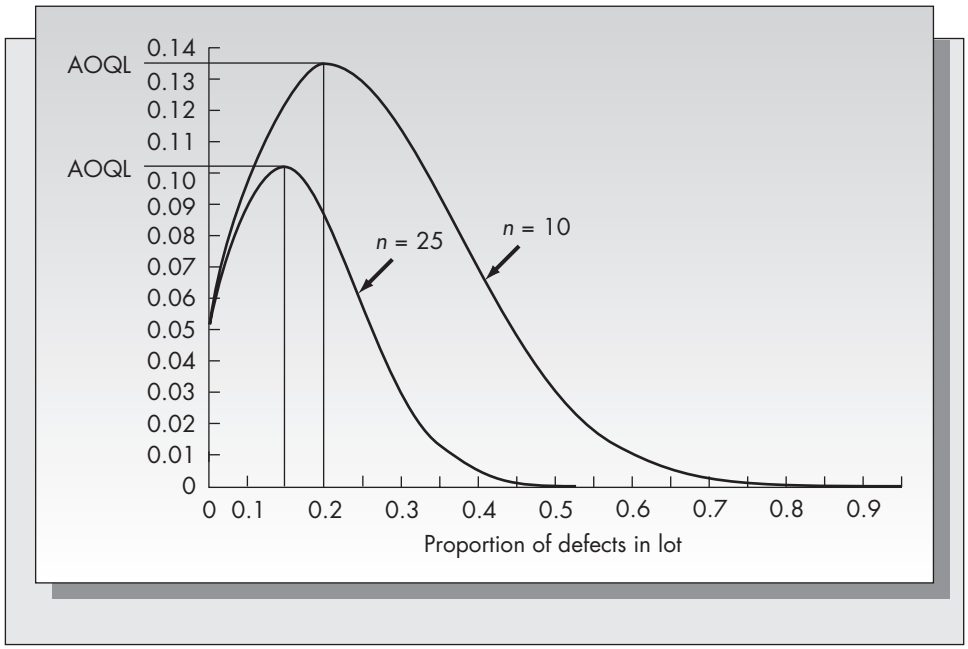


FIGURE 12–19
AOQ curves for Spire CDs (refer to Example 12.9)

Snapshot Application

MOTOROLA LEADS THE WAY WITH SIX-SIGMA QUALITY PROGRAMS

The Motorola Corporation has compiled an impressive record of defining and implementing new quality initiatives and translating those initiatives into profits. As we see in this chapter, traditional quality methods assume that an out-of-control condition corresponds to an observation falling outside of $\pm 2\sigma$ or $\pm 3\sigma$. Motorola decided that this standard was too loose, giving too many defects. In the 1980s, they moved toward a 6σ standard. To achieve this goal, Motorola established the practice of quality system reviews (QSRs) and placed a great deal of emphasis on classical statistical process control (SPC) techniques.

To achieve the 6σ standard, Motorola infused the total quality management philosophy into its entire organization. Quality programs were not assigned to and policed by a single group, but became part of everyone's job. Motorola's approach was based on the following key ingredients:

- Overriding objective of total customer satisfaction.
- Uniform quality metrics for all parts of the business.
- Consistent improvement expectations throughout the firm.
- Goal-directed incentive plans for management and employees.
- Coordinated training programs.

Motorola has carefully documented the road map to follow to its quality goals (Motorola, 1993). The first step in the process is a detailed audit of several key parts of the business. These include control of new product development, control of suppliers (both internal and external), monitoring of processes and equipment, human resource considerations, and assessing of customer satisfaction. Even top management takes part in many of these audits by paying regular visits to customers, chairing meetings of the operating policy committees, and recognizing executives who have made outstanding contributions to the company's quality initiatives.

Kumar and Gupta (1993) report on their experience with a TQM program put in place in Motorola's Austin, Texas, assembly plant. In May of 1988, management began the process of implementing an SPC program at Austin. The process began by bringing in an outside consultant to design the program and assigning an internal coordinator to ultimately take over the duties of the consultant. To be sure that employees bought into the initiative, management organized participative problem-solving teams. Each team included a manufacturing manager, a group leader, operators from the two shifts, a representative from the QA Department, and an engineer. Austin had a total of six teams. To further ensure a buy-in from all employees, management initiated a plantwide training program in SPC. Training was tailored to job function.

The plant's QA Department instituted a certification program at Austin for vendors. As a result, about 60 percent of the vendors supplying the plant were certified. Within the plant, traditional SPC methods were employed: Attribute data were collected and charted and machines were shut down when out-of-control situations were detected. Members of the QA team employed design of experiments techniques to identify causes of problems.

What is the bottom line? Over the first two years of this initiative, the Austin plant reported a decrease in scrap rates of 56 percent. By developing a clear-cut and coordinated strategy, Motorola was able to achieve major improvements in traditional quality measures in this facility. Motorola's overall success is a testament to the fact that this was not an isolated example. It demonstrates that American companies can compete effectively with overseas competitors when the quality effort is a true companywide initiative.

Problems for Section 12.12

46. If defective items are replaced and $N \gg n$, show by differential calculus that the value of p for which AOQ(p) achieves its maximum value satisfies

$$\frac{d\text{OC}(p)}{dp} = -\frac{\text{OC}(p)}{p}.$$

47. Consider the single sampling plan with $n = 10$ and $c = 0$.
    a. Derive an analytical expression for the OC curve as a function of p.

b. Using the results of Problem 46, determine the value of p at which the AOQ curve is a maximum.

c. Using the results of parts (*a*) and (*b*), determine the maximum value of the average outgoing quality.

48. Consider the single sampling plan discussed in Problem 30. If defective items are replaced and $N \gg n$, graph the AOQ curve and determine the value of the AOQL.

12.13 TOTAL QUALITY MANAGEMENT

This chapter reviewed the fundamentals of statistical quality control. Statistical quality control constitutes a set of techniques based on the theories of probability and statistical sampling for monitoring process variation and for determining if manufactured lots meet desired quality levels. However, delivering quality to the customer is a far broader problem than is addressed by statistical issues alone. This section considers quality from the management perspective.

Definitions

What is total quality management (TQM)? The term seems to have been first coined by Feigenbaum (1983) (in an earlier edition), who provided the following definition:

> Total quality control is an effective system for integrating the quality-development, quality-maintenance, and quality-improvement efforts of the various groups in an organization so as to enable marketing, engineering, production, and service at the most economical levels which allow for full customer satisfaction.

Feigenbaum's approach is to define quality in terms of the customer. As we noted in the introduction of this chapter, most definitions of quality concern either conformance to specifications or customer satisfaction. Garvin (1988) expands on these ideas and suggests that quality be considered along eight basic dimensions:

- Performance
- Features
- Reliability
- Conformance
- Durability
- Serviceability
- Aesthetics
- Perceived quality

We could lump the first five dimensions together under the general heading of conformance to requirements (the definition suggested by Crosby, 1979) and the last three under the heading of customer satisfaction (as suggested by Feigenbaum, 1983). However, by further breaking down these two categories, Garvin gives a better appreciation for the complexity of the quality issue.

Listening to the Customer

An important aspect of the process of designing quality products is giving people what they want. A perfectly designed and built coffee maker sold in a place where no one drinks coffee is, by definition, a failure. Hence, part of the process of delivering quality to the customer is knowing what the customer wants.

While listening to the customer is an important part of the manufacturing/design cycle link, it is generally more closely associated with marketing than with operations. Still, we are seeing the boundaries separating the functional areas of business becoming fuzzier. Manufacturing cannot operate in a vacuum. It must be part of the link with the customer.

Finding out what the customer wants and incorporating those wants into product design and manufacture is a multistep process. The steps of the process are

- Obtaining the data.
- Characterizing customer needs.
- Prioritizing customer needs.
- Linking needs to design.

There are several means for obtaining the raw data. Traditionally, customer opinion is solicited through interviews and surveys. There are many issues to be aware of when considering interviews with customers or potential customers. How many customer responses are enough? The right answer depends on several factors. How many market segments are there for the product? How many attributes are important? What methods will be used to interpret the results? Next, there is the question of how to solicit the information from the customer. Should one conduct interviews or surveys? The answer is unclear. Both have advantages. Interviews allow more open-ended responses, but the biases of the interviewer could slant the results. Both surveys and interviews depend on how questions are worded. For example, suppose Mr. Coffee is considering a new design for a coffee maker. A question like "What should the capacity of an automatic coffee maker be?" automatically assumes that the customer is concerned about capacity. The question "Do you prefer an 8- or a 12-cup coffee maker?" imposes even more assumptions (Dahan, 1995).

Focus groups are another popular technique for soliciting the voice of the customer. The focus group format has the advantage of being open-ended; the specific wording of questions is not as important as it is with surveys or interviews. However, focus groups have disadvantages. The moderator can affect the flow of the discussion. Also, participants with strong personalities are likely to dominate the group.

Once the database is developed, the customer needs and desires must be prioritized and grouped. Several methods are available for this. One that has received a great deal of attention in the marketing literature is conjoint analysis (due to Green and Rao, 1971). Conjoint analysis is a statistically based technique to estimate utilities for product attributes based on customer preference data.

Once attributes are determined and grouped, one needs to link those attributes to the design and manufacturing processes. This can be done with quality function deployment (QFD). With QFD, customer needs are related to the product attributes and/or aspects of the production process through a matrix. The user provides estimates of the correlation between attributes and needs in a "roof " portion of the matrix. The resulting figure looks a little like a house, hence the term *house of quality* to describe the resulting matrix (see Figure 12–20). QFD is used in conjunction with traditional methods such as surveys and focus groups. The strength of the correlations between customer needs and product attributes or design characteristics shows where the emphases should be placed when considering design of new products or design changes in existing products. The interested reader is referred to Cohen (1995) for an up-to-date comprehensive discussion of QFD methods.

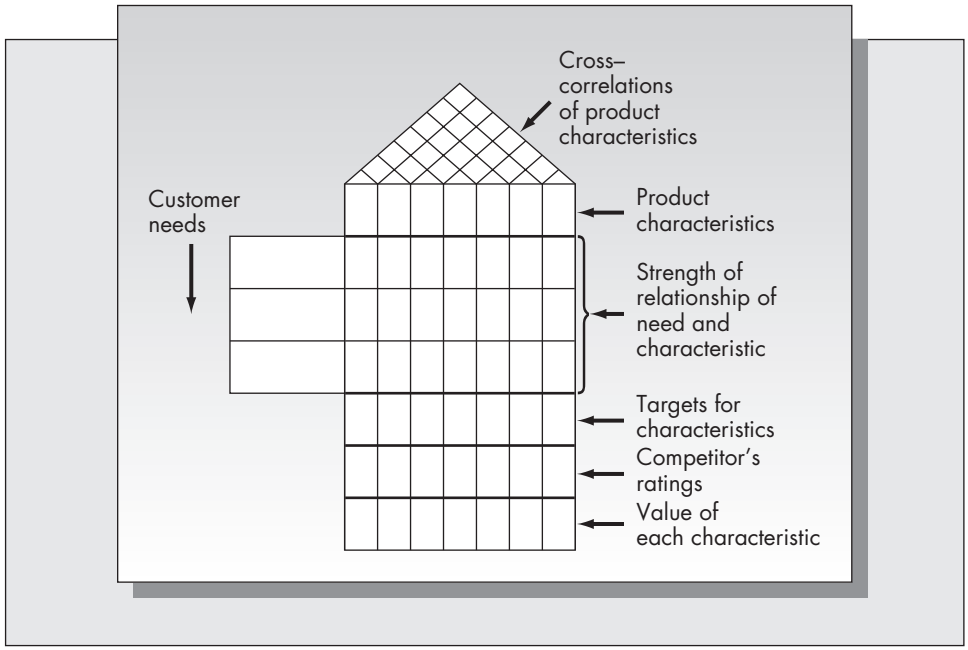


FIGURE 12–20
The house of quality: The QFD planning matrix

This section only touched the surface of the issue of soliciting customer opinion and integrating that opinion into the process of product design and manufacture. However, one must be careful not to put *too* much emphasis on the voice of the customer. Real innovations do not come from the customer but from visionaries. By letting the customer be king, innovation could be stifled. According to Martin (1995):

> Still, more and more companies are learning that sometimes your customers can actually lead you astray. The danger lies in becoming a feedback fanatic, slavishly devoted to customers, constantly trying to get in even better touch through more focus groups, more surveys. Not only can that distract you from the real work at hand, but it may also cause you to create new offerings that are safe and bland.

Home run new products like Post-it^{TM} notes developed by 3M were not the result of marketing surveys. Many products that looked like winners from the market research flopped in the marketplace. One example is New Coke, which won hands down in taste tests but was never accepted by consumers. Of course, there are probably more examples of flops that resulted from not listening to the customer. The moral here is that while the customer is an important part of the equation, customer surveys and the like should not be used to the exclusion of ingenuity and vision.

Competition Based on Quality

Because quality can be defined in several ways, management must often choose along which dimensions of quality it will focus. The Cray Corporation developed a reputation for producing the fastest computers available. To do so, Cray put its energies into technology rather than reliability. The company's chairman was quoted as saying, "If a machine doesn't fail every month or so, it probably isn't properly optimized" (as quoted by T. Alexander, 1985). As Garvin (1988) notes, this approach allowed Cray to be a technological leader but left it vulnerable to competition with comparable performance and better reliability.

Other firms have chosen different dimensions of quality along which to compete. Tandem Computers' competitive edge is based on product reliability. With parallel processors, Tandem is able to guarantee essentially no downtime. Tandem's approach has been extremely successful because many customers, such as banks and utilities, are willing to pay a premium for improved reliability.

High reliability is also one of the primary factors for the success of the Japanese automakers. With a larger number of dual-career households, reliability is at a premium. Based on the annual readers' survey of the Consumer's Union, the Japanese automakers consistently outperform their American and European competitors in this dimension of quality.

Banking is one industry that competes along the service dimension. The Wells Fargo Bank of California, for example, provides customers with a large network of branches, many automated tellers, and other services such as 24-hour account update information. Smaller banks attract customers with higher interest rates on savings, lower rates on credit card interest, and personalized service.

Some firms have based their strategies on providing the consumer with the highest-quality product, regardless of cost. Typical examples are Rolex watches, Leica cameras, Rolls-Royce automobiles, Cross pens, and Steinway pianos. Many American firms prefer to target their products to the mass market and leave the smaller high-end market for European and Japanese competitors.

Many successful companies rely on a business strategy based on being a market leader along one or two dimensions of quality. This may leave them vulnerable to competition, however.

Organizing for Quality

TQM requires total commitment on the part of management and workers. The organization must be structured to enhance product quality, not detract from it. Workers must be secure in their positions and have a stake in the success of the organization if quality programs are going to be successful. According to Feigenbaum (1983):

1. The quality function within a company should not be a single function housed within a single department. Quality should be recognized as a systematic group of disciplines to be applied on a coordinated basis by all functions throughout the company and the plant.
2. The quality function must have direct and ongoing contact with the buyers and the customers of the company's products and services.
3. The quality function must be organized to transcend individual functional organizational boundaries.
4. The overall quality function must be overseen from a high level of the firm as new products are developed, to be sure that quality issues are adequately dealt with, that "early warnings" of impending problems are easily recognized and corrected, and that management can properly manage quality.

What are the root causes of quality problems and what can be done to address those causes? Leonard and Sasser (1982) surveyed executives of 30 *Fortune* 500 firms to determine their opinions as to the causes of quality problems in the United States. The factor on the top of their list was workmanship and workforce. It is probably fair to assume that a similar survey of workers would point the finger at management. Management's policies determine workers' job security and working conditions. Reward structures determine incentives for behavior. For TQM to work, the entire organization must line up

behind a quality imperative. To make this happen, management must be responsible for creating an organizational structure in which workers are empowered and in which workers have incentives to do their best, not one that stifles excellence.

A program that received a great deal of attention in the early 1980s was *quality circles* (QCs). Quality circles were an attempt to emulate the Japanese organizational structure. A quality circle is a small group of roughly 6 to 12 employees that meet on a regular basis with the purpose of identifying, analyzing, and suggesting solutions to problems affecting its work area. Typically, such a group would meet for about four hours per month on company time. The group might be given special training on statistical quality control methods, group dynamics, and general problem-solving techniques.

There was a dramatic increase in interest in quality circles in the United States starting around 1980. Lawler and Mohrman (1985) report that 44 percent of the companies in the United States with more than 500 employees had some type of quality circle program in place in 1982 and that 75 percent of these programs were started after 1980. There are several reasons for the sudden interest in quality circles in the United States:

1. *Cost*. Quality circles could be put into place relatively inexpensively. Consulting firms that specialize in the area would set up the program, train the appropriate individuals, and oversee the initial activities of the circle. From a cost perspective, an advantage of the quality circle is that not all company employees are necessarily involved.
2. *Control*. From a manager's point of view, a desirable feature of the quality circle is that it has no formal decision-making power, and thus is not perceived as a threat. The quality circle essentially serves the role of a formal "suggestion box" mechanism, and there is no requirement that the manager relinquish any management prerogatives in his or her area of responsibility.
3. *Fashion*. There is little doubt that quality circles have had an enormous fad appeal. The apparent success of programs of this type in Japanese industry has led to an early acceptance of QCs in the United States.

With the interest in quality circles and the adoption of so many QC programs in the United States, what impact have these programs had? The experience, by and large, has been that few QC programs have evolved into other programs that could affect company procedures and practices. As a result, group interest in QCs has tended to dwindle and the groups have met less frequently or not at all. Lawler and Mohrman (1985) suggest that the QC is inherently an unstable organizational structure. Because it has no decision-making power, its only role is to serve as a formal "suggestion box." It is natural that employees would eventually lose interest. Ultimately, quality circles could have served as a first step toward building a more participative approach to management, but that does not appear to have occurred.

Benchmarking Quality

Benchmarking means measuring one's performance against that of one's competitors. Competitive benchmarking is gaining importance in light of increasing global competition. Setting one's priorities and being certain that those priorities are consistent with the needs of the marketplace are essential. Based on a database developed at Boston University, Miller et al. (1992) report competitive priorities in Europe, Japan, and the United States. These are summarized in Table 12–6.

There are several interesting things to see from this table. First, consistent with our discussion, the Japanese firms surveyed rank product reliability as their top priority.

TABLE 12–6
Top Five Competitive Priorities

Source: Miller et al., 1992.

Europe	Japan	United States
Conformance quality	Product reliability	Conformance quality
On-time delivery	On-time delivery	On-time delivery
Product reliability	Fast design change	Product reliability
Performance quality	Conformance quality	Performance quality
Delivery speed	Product customization	Price

The Japanese understand that product reliability could be their greatest competitive asset and plan to continue to stress this important dimension of quality. It is also interesting to see that price is mentioned only by the U.S. companies.

The authors list four types of benchmarking:

1. Product benchmarking.
2. Functional or process benchmarking.
3. Best practices benchmarking.
4. Strategic benchmarking.

Product benchmarking refers to the practice of tearing down a competitor's product to see what can be learned from its design and construction. It is said that when Toyota initiated its program to produce the Lexus to compete with cars such as Mercedes and BMW, it carefully examined the competitor's products to determine how and where welds were placed, and how the cars were put together to achieve the look and feel of exceptional quality.

Functional benchmarking focuses on the process rather than on the product. Typical processes might be order entry, assembly, testing product development, and shipping. Functional benchmarking is possible only when companies are willing to cooperate and share information. It has the same goal as product benchmarking: to improve the process and ultimately the resultant product.

Best practices benchmarking is similar to functional benchmarking, except that it focuses on management practices rather than on specific processes. Best practices might consider factors such as the work environment and salary incentives for employees in firms with exceptional performance. General Electric is a strong advocate of best practices benchmarking (*Fortune,* 1991).

The goal of *strategic benchmarking* is to consider the results of other benchmarking comparisons in the light of the strategic focus of the firm. Specifically, what is the overall business strategy that has been articulated by the CEO, and are the results of other benchmarking studies consistent with this strategy?

Ultimately, what is the purpose of benchmarking? It is to ensure continuous improvement and is only one of the means of achieving this. Continuous improvement in product and process is the ultimate goal of any quality program. Competitive benchmarking provides a means of learning from one's competitors. Although benchmarking can be a useful tool, it is not a substitute for a clearly articulated business strategy and a vision for the firm.

The Deming Prize and the Baldrige Award

The three leaders of the quality movement in the United States during the 1950s, W. Edwards Deming, Joseph M. Juran, and A. V. Feigenbaum, each contributed to the Japanese quality movement, although Deming is certainly the name that comes to mind

first. Deming's success in Japan was the result of a seminar he presented on statistical quality control in 1950. He repeated that seminar several times, became active in the Japanese quality movement, and ultimately became a national hero in Japan. Deming recommended both the application of statistical methods and a systematic approach to solving quality problems. His approach later became known as the Plan, Do, Check, Act (PDCA) method.

Juran, another important leader in the quality movement, stressed the managerial aspects of quality rather than the statistical aspects. Juran also presented several seminars to the Japanese that were targeted at middle to upper management. The Juran Institute, located in Wilton, Connecticut, was founded in 1979 and continues to provide consulting and training in the quality area. [However, Philip Crosby Associates, founded the same year, generates almost 10 times the annual revenue (*Business Week,* 1991).]

Using the royalties from a book based on his 1950 lectures, Deming established a national quality prize in Japan. The funding for the prize continues to be primarily from Deming's royalties, but it is supplemented by a Japanese newspaper company and private donations. The Deming Prize is awarded to (1) those who have achieved excellence in research in the theory or application of statistical quality control, (2) those who have made remarkable contributions to the dissemination of statistical quality control methods, and (3) those corporations that have attained commendable results in the practice of statistical quality control.

The prize has been awarded almost every year in several categories since 1951. Recipients include such well-known firms as Toyota Motor Co., Hitachi, Fuji Film, Nippon Electric, and NEC. Prizes also have been awarded to specific divisions of large firms and specific plants within a division. It is a highly sought-after national honor in Japan (Aguayo, 1990).

The Malcolm Baldrige National Quality Award was established by the U.S. Department of Commerce in 1987 largely as a response to the success of the Deming Prize in Japan. The award is named for the late secretary of commerce, who died in an accident the same year that Congress and President Ronald Reagan enacted the Malcolm Baldrige National Quality Improvement Act.

The award is made each year in three categories: (1) manufacturing companies or subsidiaries, (2) service companies or subsidiaries, and (3) small businesses. Applying for the Baldrige Award is an involved process requiring a sincere commitment on the part of the firm seeking the award. A board of overseers appointed by the secretary of commerce has final say. They base their evaluation on the following nine-point value system:

1. Total quality management.
2. Human resource utilization.
3. Performance.
4. Measurables.
5. Customer satisfaction.
6. World-class quality.
7. Quality early in the process.
8. Innovation.
9. External leadership.

The evaluation process utilizes an elaborate scoring system and includes site visits from the evaluation committee. Funds to support the examining process are donated by individuals and firms. As of June 30, 1989, more than $10.4 million had been pledged

support the program (Pierce, 1991). The number of applicants grew each year for the first three years after the prize was announced. In 1989, 40 companies applied for awards; in 1990, 97 firms applied; in 1991, 106 applied; and in 1992, 90 firms applied. [There are several reasons for the drop in applications in 1992. One is that many firms are more concerned with ISO 9000 certification. Another is that the cost of applying is a problem for some companies. These costs are not inconsequential. For example, Xerox, a 1989 winner, spent $1 million preparing its application. Finally, being forced to lay off workers, many companies are not in a position to develop a credible application (Hillkirk, 1992).]

What has been the experience with the Baldrige Award now that more than 10 years have elapsed since its inception? One way of measuring the success of the program is to see how the firms that have won have fared. While not every Baldrige winner has been a winner in the marketplace, the evidence is that, on average, the firms seem to outperform the economy as a whole. According to an experiment conducted by the National Institute of Standards and Technology (NIST), five whole company winners—Eastman Chemical, Federal Express, Motorola, Solectron, and Zytec—have outperformed the S&P 500 by 6.5 to 1. The publicly traded parent companies of seven subsidiaries that won the award outperformed the S&P 500 by almost 3 to 1 (Transportation and Distribution, 1995). While not all Baldrige winners fared so well, these results, at least, are encouraging.

The Baldrige Award is not an end in itself but rather a means to an end. The process of preparing an application forces the firm to take a good hard look at its quality efforts and provides a blueprint for self-examination. The Baldrige application is an excellent means for ferreting out problems and suggesting where improvements need to be made.

ISO 9000

First published in 1987, the International Organization for Standardization (ISO), based in Switzerland, first established guidelines for ISO 9000. Since that time ISO standards have been revised in 1994, 2000, and in 2008. This was the first attempt at developing a true uniform global standard for quality. For a firm to obtain ISO 9000 registration, it must carefully document its systems and procedures. Purchasing, materials handling, manufacturing, and distribution are all subject to ISO documentation and certification. ISO certification is very different from the Baldrige process. ISO certification is not an award, nor even necessarily a judgment about the quality of products. It is a certification of the manufacturing and business processes used by the firm.

With the establishment of the ISO standard came a host of consulting organizations specializing in taking a company through the certification process. Certification is not cheap. Aside from the direct cost of consultants, there are substantial indirect costs associated with developing the necessary documentation. According to John Rudin, president of Reynolds Aluminum Supply Company:

> From an out-of-pocket standpoint, just to get one facility registered gets to be somewhere in the $25,000 to $30,000 range. For RASCO as a whole, you're talking a million or so, if we were to do all the facilities. Right now we don't see the value in having a plaque to hang on the wall for a million dollars, but we do see the value in having a good, back-to-basics, well-documented process, which is what ISO is really all about anyway. (Kuster, 1995)

The firm seeking certification must document quality-related procedures, system and contract-related policies, and record keeping. According to Velury (1995), a firm should expect to document about 20 procedures for each section required for certification. Given the expense in time and money, why should anyone bother? There are several advantages of achieving certification. One is from the process itself. Careful

documentation of quality practices reveals where those practices fall short. The process of continuous improvement starts with knowing where one stands. Furthermore, many firms now require suppliers to obtain certification. It could be a prerequisite to doing business in many circumstances. According to Thomas Boldlund, president of Oregon Steel:

> ISO 9000 allows us to participate in markets that 10 years ago we could not serve even if we wanted to. When we meet a customer, the first thing they ask is "Can you meet the quality standards of your competition?" and second: "Can you produce the grades the Europeans and Japanese produce today?" With ISO 9000, the answer is yes. (Kuster, 1995)

The ISO continues to develop international standards. In 1996 the organization announced the ISO 14000 series of environmental standards (Alexander, 1996). ISO 14000 is a series of international standards for environmental management systems. The motivation for the ISO in adopting a uniform environmental standard is the disparity of these standards in different countries. It is hoped that this new standard will not only spur international trade, but, more importantly, improve quality of life by improving the environment. Both ISO 9000 and ISO 14000 guidelines are important first steps in the movement toward a global uniform standard.

Quality: The Bottom Line

What ultimately determines whether a quality program is successful? It is the "bottom line." Whether the program is TQM, continuous improvement, or six sigma, the firm's management must believe that implementation will ultimately lead to higher profits. Traditional thinking is that quality initiatives are expensive. To achieve high reliability and conformance to specifications, substantial capital expenditures might be required (see Figure 12–1). Only if the expense translates to an improved bottom line will it be justified.

That quality costs more is not at all clear. In addition to erosion of the customer base, there are numerous direct costs of poor quality. These include costs of additional work-in-process inventories, scrap costs, rework costs, and inspection costs. Firms face increased risks that design and/or reliability problems will lead to costly lawsuits and settlements.

Garvin's study of the room air conditioner industry from 1981 to 1982 (Garvin, 1988), summarized in Table 12–7, shows the dramatic increase in the costs of quality as product quality and reliability deteriorate.

Although not obvious from this table, the Japanese manufacturers as a group had much lower defect rates than the best U.S. manufacturers. (This is clearer in Garvin's 1983 study.) Before analyzing the results, we should point out some inconsistencies. First, the warranty period in the United States is five years and in Japan it is three years.

TABLE 12–7
Quality and Quality Costs in the Room Air Conditioner Industry

Source: Garvin, 1988, p. 82.

Grouping of Companies by Quality Performance	Average Warranty Costs as a Percentage of Sales	Total Cost of Quality (Japanese Companies), Total Failure Costs (U.S. Companies) as a Percentage of Sales
Japanese manufacturers	0.6%	1.3%
Best U.S. plants	1.8	2.8
Better U.S. plants	2.4	3.4
Fair U.S. plants	2.7	3.9
Poor U.S. plants	5.2	>5.8

This would tend to bias the results in favor of the Japanese, although the bias is small because most warranty costs are incurred in the first year. Second, the percentages in the third column include costs of prevention, inspection, rework, scrap, and warranties for the Japanese firms, and only rework, scrap, and warranties for the American firms. Hence, the comparable percentages in the third column for the American firms are even higher than those reported in the table.

What is clear from this study is that the costs of poor quality are substantial. In fact, they were observed to be as much as five times higher for poor U.S. manufacturers than for the Japanese manufacturers. This does not prove conclusively, however, that it is necessarily less expensive to produce higher-quality products. These data do not show the additional investments in capital equipment, processes, or statistical quality control made by the better manufacturers, and for that reason are not proof that "quality is free." However, other studies (see, in particular, Schoeffler et al., 1974) show that firms producing higher-quality products have a higher return on investment, a larger market share, and more profits than firms that produce lower-quality products.

12.14 DESIGNING QUALITY INTO THE PRODUCT

Traditional quality control methods focus on sampling and inspection. Sampling plans and inspection policies have the ultimate goal of producing an acceptable percentage of defects. Viewing quality in terms of the entire product cycle, including design, production, and consumption, shows that it is economical to design quality into the product. A recent development, applying statistical design of experiments to the problem of product design, has been developed by the Japanese statistician and consultant Genichi Taguchi. Taguchi's initial contributions were in *off-line* quality control methods. His later work incorporated economic issues as well.

Consider the following simple example from Taguchi et al. (1989). It is well known that, prior to World War II, the quality of manufactured goods in Japan was poor. At that time the Japanese sought to compete on price rather than on quality. Consider a Japanese product that sold for half the price of its American competitor. If one based a purchasing decision on price alone, one would have chosen the inferior Japanese product. However, factoring in consumption, we must account for the losses incurred to the customer in using that product. Suppose that the loss for using the American-made product was equal to the purchase price, say P. Furthermore, assume that the loss for the Japanese-made product was nine times its purchase price, which we assume to be $0.5P$. Then the total cost to the customer for the American-made product would have been $P + P = 2P$ and the total cost to the customer for the Japanese-made product would have been $0.5P + (9)(0.5P) = 5P$. Hence, the Japanese-made product would have cost the consumer two and a half times as much as the American-made one. Given experience with both products, the consumer would eventually realize that his or her overall costs were greater with the less expensive product.

It is precisely this type of phenomenon that has led to maxims such as "you get what you pay for," and explains in simple economic terms why consumers are willing to pay more for quality.[4] The irony of this example is how effectively the Japanese have managed to position themselves on the other side of the equation. It explains why Japanese cars continued to sell well in the United States even when the exchange rate between the dollar and the yen was so unfavorable to the Japanese.

[4] However, the converse that more expensive products are necessarily superior is not always true. See Garvin (1988), p. 70, for a discussion of the correlation of quality and price.

Quality has significant economic value to the consumer, and product design plays an important role in the product's quality. What does it mean to design for quality? It means that the number of parts that fail easily, or those that significantly complicate the manufacturing process, should be minimized. In particular, how can the design be simplified to eliminate small parts such as screws and latches that are difficult to assemble and may be likely trouble spots down the road? One example of a design in which the number of parts is reduced to a minimum is the IBM Proprinter. IBM developed an impressive video showing how easily the product can be assembled by hand. This simple design was an important factor in the product's reliability and success in the marketplace.

The design cycle is an important part of the quality chain. Taguchi et al. (1989) recommend the following three steps in the engineering design cycle:

1. *System design.* This is the basic prototype design that meets performance and tolerance specifications of the product. It includes selection of materials, parts, components, and system assembly.
2. *Parameter design*. After the system design is developed, the next step is optimization of the system parameters. Given a system design, there are generally several system parameters whose values need to be determined. A typical design parameter might be the gain for a transistor that is part of a circuit. One needs to find a functional relationship between the parameter and the measure of performance of the system to determine an optimal value of the parameter. In the example, the measure of performance might be the voltage output of the circuit. The goal is to find the parameter value that optimizes the performance measure. The Taguchi method considers this issue.
3. *Tolerance design*. The purpose of this step is to determine allowable ranges for the parameters whose values are optimized in step 2. Achieving the optimal value of a parameter may be very expensive, whereas a suboptimal value could give the desired quality at lower cost. The tolerance design step requires explicit evaluation of the costs associated with the system parameter values.

The same concepts can be applied to the design of the production process once the product design has been completed. The system design phase corresponds to the design of the actual manufacturing process. In the parameter design phase, one identifies parameters that affect the manufacturing process. Typical examples are temperature variation, raw material variation, and input voltage variation. In the tolerance design phase, one determines acceptable ranges for the parameters identified in phase 2.

The area of off-line quality control (as opposed to the subject of this chapter, which might be referred to as on-line quality control) involves techniques for achieving these three design objectives. Taguchi methods, based on the theory of design of experiments, give new approaches to solving these problems.

Optimizing a parameter value may not always be the overall optimal solution. In some cases, redesigning the product to be less sensitive to the parameter in question might be more economical. In this spirit, Kackar (1985) describes a Japanese tile manufacturer who solved the problem of sensitivity to temperature in this way. Quoting from the article.

> A Japanese ceramic tile manufacturer knew in 1953 that it is more costly to control causes of manufacturing variations than to make a process insensitive to these variations. The Ina Tile Company knew that an uneven temperature distribution in the kiln caused variation in the size of tiles. Since uneven temperature distribution was an assignable cause of variation, a process quality control approach would have been to devise methods for controlling the temperature

distribution. This approach would have increased manufacturing cost. The company wanted to reduce the size variation without increasing cost. Therefore, instead of controlling temperature distribution they tried to find a tile formulation that reduced the effect of uneven temperature distribution on the uniformity of tiles. Through a designed experiment, the Ina Tile Company found a cost-effective method for reducing tile size variation caused by uneven temperature distribution in the kiln. The company found that increasing the content of lime in the tile formulation from 1 percent to 5 percent reduced the tile size variation by a factor of 10. This discovery was a breakthrough for the ceramic tile industry.

Taguchi's method is based on assuming a loss function, say $L(y)$, where y is the value of some functional characteristic and $L(y)$ is the quality loss measured in dollars. In keeping with much of the classical theory of statistics and control, Taguchi recommends a quadratic loss function. The quadratic form is the result of using the first two terms of a Taylor series expansion. Given an explicit form for the loss function, one can address such questions as the benefits of tightening tolerances and the value of various inspection policies. We refer the interested reader to Taguchi et al. (1989) and Logothetis and Wynn (1989) for a discussion of the general theory. Applications to specific industries are treated by Dehnad (1989).

Design, Manufacturing, and Quality

Quality starts with the product design and the way that design is integrated into the manufacturing process. The most creative design in the world is useless if it can't be manufactured economically into a reliable product. Successful linking of the design and the manufacturing processes is a hallmark of Japan's success in consumer products. The **design for manufacturability** (DFM) movement in the United States had its roots in Japanese manufacturing methods.

Boothroyd and Dewhurst (1989) and Boothroyd et al. (1994) were among the first to develop an effective scoring system for designs in terms of their ease of manufacturability. The two books summarize the methodology developed by these two engineers over a period of several years. The first book focuses on assembly efficiency. Assembly efficiency is the ratio of the theoretical minimum assembly time over an estimate of the actual assembly time based on the current product design. The later book deals with more general DFM issues including numbers of parts, types of parts, and types of fasteners. These rules, which are very detailed, recommend that simpler designs with fewer parts are preferred. Such designs lead to products that are easier and less expensive to manufacture, and are less likely to fail in use.

Ulrich and Eppinger (1995) recommend that designers keep track of design complexity via a scorecard approach. (See Example 12.10, which follows.) A scorecard provides a way to compare different designs objectively and a means of keeping track of the complexity of the manufacturing process for every product design.

Example *12.10* **Scorecard of Manufacturing Complexity Example**

Complexity Drivers	Revision 1	Revision 2
Number of new parts introduced	6	5
Number of new vendors introduced	3	2
Number of custom parts introduced	2	3
Number of new "major tools" introduced	2	2
Number of new production processes introduced	0	0
Total	13	12

Source: Ulrich and Eppinger, 1995.

This example is meant to be illustrative only. In practice, the team would have to decide on the relative importance of the drivers and apply suitable weights. Scorecards such as this force the design team to take a good hard look at the manufacturing consequences of their decisions.

An example of a successful DFM effort was the IBM Proprinter. The Proprinter was a dot matrix printer focused at the ever-expanding PC printer market dominated by the Japanese in the early 1980s. IBM developed a video showing someone assembling the printer by hand in a matter of minutes. In designing the Proprinter, IBM followed classic DFM methodology. The Proprinter had very few separate parts and virtually no screws and fasteners, without any compromise in functionality. The result was that IBM was able to assemble the Proprinter in the United States and remain cost competitive with Japanese rivals (Epson, in particular) that dominated the market at that time. The Proprinter was a very successful product for IBM.

The number of parts in a product is not the only measure of manufacturability. Exactly how parts are designed and put together also plays an important role. According to Boothroyd and Dewhurst (1989), the ideal characteristics of a part are

- Part should be inserted into the top of the assembly.
- Part is self-aligning.
- Part does not need to be oriented.
- Part requires only one hand for assembly.
- Part requires no tools.
- Part is assembled in a single, linear motion.
- Part is secured immediately upon insertion.

While there are some clear successes in applying DFM, the methodology has yet to gain universal acceptance. According to Boothroyd et al. (1994), the following reasons are the most common for not implementing DFM in the design phase:

1. *No time*. Designers are pushed to finish their designs quickly to minimize the design-to-manufacture time for a new product. The DFM approach is time intensive. Designing to reduce assembly costs and product complexity cannot be done haphazardly.
2. *Not invented here*. New ideas are always resisted. It would be better if the impetus for DFM came from the designers themselves, but more often it comes from management. Designers resent having a new approach thrust upon them by outsiders (as does anyone).
3. *Low assembly costs*. Since assembly costs often account for a small portion of total manufacturing costs, one might argue that there is little point to doing a design for assembly (DFA) analysis. However, savings often can be greater than one might think.
4. *Low volume*. One might argue that DFM analysis is not worthwhile for low-volume items. Boothroyd et al. (1994) argue that the opposite is true. When volumes are low, redesign is unlikely once production begins. This means that doing it right the first time is even more important.
5. *We already do it*. Many firms have used some simple rules of thumb for design (such as limiting the number of bends in a sheet metal part). While such rules make sense in isolation, they are unlikely to lead to the best overall design for the product.

6. *DFM leads to products that are difficult to service*. This is not likely to be true. Products that are easier to assemble are easier to disassemble, and thus easier to service.

Dvorak (1994) offers other reasons for the slow acceptance of DFM. Classical accounting systems may not be able to recognize the cost savings from new designs. A design that reduces fixed setup costs would not be viewed as cost effective since in many accounting systems fixed costs are considered part of overhead. An activity-based accounting system would not have this problem. However, most agree that the greatest obstacle to the acceptance of DFM is resistance to change.

Although the DFM movement may not be gaining acceptance at the rate that some would like, there is clearly a growing awareness and use of these powerful methods. Dvorak (1994) discusses several success stories of DFM implementation. One is at Coors, where production yield, quality, and delivery reliability were improved. According to Dvorak:

> Word is spreading throughout industry about how DFM can bring successes similar to those at Coors. Numerous organizations are taking up the DFM banner. The U.S. Department of Commerce, for one, has started DFM projects at six of their regional manufacturing centers. General Motors has relied on DFM to power its concurrent engineering efforts with startling response: It's saving 20% of the total car cost of the 1992 models on which it was applied. DFM and designing-for-assembly concepts have already worked wonders trimming material and manufacturing costs from a range of products. Now new ideas are infiltrating the discipline.

International competition continues to heat up. The Japanese, in particular, have demonstrated that one can be successful by careful analysis and thinking throughout the product design and development phase. They have been adept at concurrent product and process design. The result is products that can be manufactured more efficiently and work better than the competitors'. While DFM is only one piece of the pie, it provides a methodology for linking design and manufacturing.

Finally, we should not forget that product design extends far beyond issues of manufacturability only. Aesthetic issues are important as well, and may be the dominant factor for some products. Another important issue is the process of narrowing down the field of choices to a final design. Since the interest in this book is in manufacturing-related issues, we will not discuss these broader design issues but refer the interested reader to Pugh (1991).

12.15 HISTORICAL NOTES

The desire to maintain quality of manufactured goods is far from new. The prehistoric man whose weapons and tools did not function did not survive. However, the statistical quality control methods discussed in this chapter were devised only in the last 70 years or so. Walter Shewhart, who was an employee of the Bell Telephone Laboratories, conceived the idea of the control chart. His 1931 monograph (Shewhart, 1931) summarized his contributions in this area. H. F. Dodge and H. G. Romig, also employees of Bell Labs, are generally credited with laying the foundations of the theory of acceptance sampling.

As with most new methodology, American industry was slow to adopt statistical quality control techniques. However, as part of the war effort, the U.S. government

decided to adopt sampling inspection methods for army ordnance in 1942. As a result of this and other wartime governmental activities, knowledge and acceptance of the techniques discussed in this chapter became widespread after the war. In fact, sampling plans are often described in terms of the U.S. Army designation. Many of the acceptance sampling techniques discussed in this chapter are part of Military Standard 105. Military Standard 105 and its various revisions and additions form the basis for most of the acceptance sampling plans in use today.

Abraham Wald was responsible for developing the theory of sequential analysis, which forms the basis for the formulas that appear in Section 12.11. Wald's work proved to be a major milestone in the advance of the theory of acceptance sampling. Wald was part of a research team organized at Columbia University in 1942. The U.S. government considered his work to be so significant that they withheld publication until June of 1945.

W. Edwards Deming, who visited Japan in the early 1950s to deliver a series of lectures on quality control methods, is given much of the credit for transferring statistical quality control technology to Japan. Today, the highly prestigious Deming Prize in Quality Control, established by Deming in 1951, remains a symbol of the Japanese commitment to quality.

Most of the methods discussed in this chapter are in widespread use throughout industry in the United States and overseas. Optimization models for designing control charts have not enjoyed the same level of acceptance, however. The model outlined in Section 12.7 is from Baker (1971), although more complex and comprehensive economic models for design of $\overline{X}$ charts were developed earlier (see, for example, Duncan, 1956). For a comprehensive discussion of the history of the quality movement in the United States, see Kolesar (1993).

12.16 Summary

In this chapter we outlined the techniques for control chart design and acceptance sampling. The basis for the methodology of this chapter is classical probability and statistics. The underpinnings of control charts include fundamental results from probability theory such as the law of large numbers and the central limit theorem.

We discussed *control charts* in the first half of this chapter. A control chart is a graphical device that is used to determine when a shift in the value of a parameter of the underlying distribution of some measurable quantity has occurred. When this happens, the process is said to have gone out of control. The design of the control chart depends on the particular parameter that is being monitored.

The most common control chart is the $\overline{X}$ chart. An $\overline{X}$ chart is designed to monitor a single measurable variable, such as weight or length. Subgroups of size n are sampled on a regular basis and the sample mean of the subgroup is computed and placed on a graph. Because the sample mean is approximately normally distributed independently of the form of the distribution of the population, the likelihood that a single observation falls outside three-sigma limits is sufficiently small that when such an event occurs it is unlikely to be due to chance. Rather, it is more likely to be the result of a shift in the true mean of the process.

The second type of control chart treated in this chapter is the *R chart*. An R chart is designed to determine when a shift in the process variation has occurred. The symbol R stands for range. The range is the difference between the largest and the smallest observations in a subgroup. Because a close relationship exists between the value of the

range of the sample and the underlying population variance, R charts are used to measure the stability of the variance of a process.

Often control charts for variables are inappropriate. The *p chart* is a control chart for attributes. When using p charts, items are classified as either acceptable or not. The p chart utilizes the normal approximation of the binomial distribution and may be used when subgroups are of equal or of varying sizes.

The last control chart presented was the *c chart*. The c chart is used to monitor the number of defectives in a unit of production. The parameter c is the average number or rate of defects per unit of production. The c chart is based on the Poisson distribution, and the control limits are derived using the normal approximation to the Poisson.

A model for the economic design of $\overline{X}$ charts was considered in Section 12.6. The decision variables for the model were the size of each sample subgroup, n, and the number of standard deviations used to signal an out-of-control condition, k. The model is based on the assumption that the process goes out of control randomly with a known probability π. Also assumed known are the cost of sampling, the cost of searching for an assignable cause when an out-of-control signal occurs, and the cost of operating the system in an out-of-control condition.

The chapter also considered *acceptance sampling*. The purpose of an acceptance sampling scheme is to determine if the proportion of defectives in a large lot of items is acceptable based on the results of sampling a relatively small number of items from the lot. The simplest sampling plan is single sampling. A *single sampling plan* is specified by two numbers: n and c. Interpret n as the size of the sample and c as the acceptance level. A *double sampling plan* requires specification of five numbers: n_1, n_2, c_1, c_2, and c_3, although often $c_2 = c_3$. Based on the results of an initial sample size of n_1, the lot is accepted or rejected, or an additional sample of size n_2 is drawn. The logical extension of double sampling is *sequential sampling,* in which items are sampled one at a time and a decision is made after each item is sampled about whether to accept the lot, reject the lot, or continue sampling.

Total quality management is a term that we hear more frequently as U.S. firms strive to compete with their European and Japanese competitors. We discussed Garvin's eight dimensions of quality: (1) performance, (2) features, (3) reliability, (4) conformance, (5) durability, (6) serviceability, (7) aesthetics, and (8) perceived quality. Methods for eliciting the voice of the customer such as *conjoint analysis* and *quality function deployment* (QFD) were discussed as well. In order for TQM to succeed, the quality activity must transcend functional and departmental boundaries. One approach that attempted to do this was *quality circles*. The program required minimal investment and restructuring on management's part, and as a result has not been very successful. Benchmarking provides a means for a firm to compare its performance with its competitors and learn the industry's "best practices." Two highly sought-after national prizes are the Deming Prize in Japan and the Baldrige Award in the United States. These awards recognize exceptional industry efforts in implementing quality.

Off-line quality methods are directed at the problem of designing quality into the product. Taguchi methods, largely based on the theory of design of experiments, are an important development in this area. The Taguchi methods identify important process and design parameters and attempt to find overall optimum values of these parameters, relative to some measure of performance of the system. The chapter concluded with a discussion of design for manufacturability and the contributions of Boothroyd and Dewhurst to this area.

Additional Problems on Quality and Assurance

49. In what ways could each of the factors listed contribute to poor quality?
    a. Management
    b. Labor
    c. Equipment maintenance
    d. Equipment design
    e. Control and monitoring
    f. Product design

50. Figure 12–1 presents a conceptual picture of the trade-off between process cost and the costs of losses due to poor quality. What are the costs of poor quality and what difficulties might arise when attempting to measure these costs?

51. $\overline{X}$ and R charts are maintained on a single quality dimension. A sudden shift in the process occurs, causing the process mean to increase by 2σ, where σ is the true process standard deviation. No shift in the process variation occurs. Assuming that the $\overline{X}$ chart is based on 3σ limits and subgroups of size $n = 6$, what proportion of the points on the $\overline{X}$ chart would you expect to fall outside the limits after the shift occurs?

52. XYZ produces bearings for bicycle wheels and monitors the process with an $\overline{X}$ chart for the diameter of the bearings. The $\overline{X}$ chart is based on subgroups of size 4. The target value is 0.37 inch, and the upper and lower limits are 0.35 and 0.39 inch, respectively (assume that these are based on three-sigma limits). Wheeler, which purchases the bearings from XYZ to construct the wheels, requires tolerances of 0.39 ± 0.035 inch. Oversized bearings are ground down and undersized bearings are scrapped. What proportion of the bearings does Wheeler have to grind and what proportion must it scrap? Can you suggest what Wheeler should do to reduce the proportion of bearings that it must scrap?

53. A process that is in statistical control has an estimated mean value of 180 and an estimated standard deviation of 26.
    a. Based on subgroups of size 4, what are the control limits for the $\overline{X}$ and R charts?
    b. Suppose that a shift in the mean occurs so that the new value of the mean is 162. What is the probability that the shift is detected in the first subgroup after the shift occurs?
    c. On average, how many subgroups would need to be sampled after the shift occurred before it was detected?

54. Consider the data presented in Table 12–1 for the tracking arm example. Suppose that an R chart is constructed based on sample numbers 1 to 15 only.
    a. What is the estimate of σ obtained from these 15 observations only?
    b. What are the values of the UCL and LCL for an R chart based on these observations only?

55. Discuss the advantages and disadvantages of the following strategies in control chart design. In particular, what are the economic trade-offs attendant to each strategy?
    a. Choosing a very small value of α.
    b. Choosing a very small value of β.

c. Choosing a large value of n.

d. Choosing a small value for the sampling interval.

56. The construction of the p chart described in this chapter requires specification of both UCL and LCL levels. Is the LCL meaningful in this context? In particular, what does it mean when an observed value of p is less than the LCL? If only the UCL is used to signal an out-of-control condition, should the calculation of the UCL be modified in any way? (Hint: The definition of the Type 1 error probability, α, will be different. The hypothesis test for determining if the process is in control is one-sided rather than two-sided.)

57. A manufacturer of large appliances maintains a p chart for the production of washing machines. The machines may be rejected because of cosmetic or functional defectives. Based on sampling 30 machines each day for 50 consecutive days, the current estimate of p is .0855.

    a. What are the control limits for a p chart? (Assume 3σ limits.)

    b. Suppose that the percentage of defective washing machines increases to 20 percent. What is the probability that this shift is detected on the first day after it occurs?

    c. On average, how many days would be required to detect the shift?

58. A maker of personal computers, Noname, purchases 64K DRAM chips from two different manufacturers, A and B. Noname uses the following sampling plan: A sample of 10 percent of the chips is drawn and the lot is rejected if two or more defective chips are discovered. The two manufacturers supply the chips in lots of 100 and 1,000, respectively.

    a. For each manufacturer, determine the true proportion of defectives in the lot that would result in 90 percent of the lots being accepted. You may use the Poisson approximation for your calculations.

    b. Would you say that this plan is fair?

59. Graph the AOQ curves for manufacturers A and B mentioned in Problem 58. Estimate the values of the AOQL in each case.

60. Consider the sampling plan discussed in Problem 58. Would a fairer plan be to reject the lot if more than 10 percent of the chips in a sample are defective? Which of the two manufacturers mentioned would be at an advantage if this plan were adopted?

61. Assuming AQL = 5 percent and LTPD = 10 percent, determine the values of α and β for the plan described in Problem 58 for manufacturers A and B, and for the plan described in Problem 60 for manufacturers A and B.

62. Graph the OC curves for the sampling plan described in Problem 58 for both manufacturers A and B.

63. $\overline{X}$ control charts are used to maintain control of the manufacture of the cases used to house a generic brand of personal computer. Separate charts are maintained for length, width, and height. The length chart has UCL = 20.5 inches, LCL = 19.5 inches, and a target value of 20 inches. This chart is based on using subgroups of size 4 and three-sigma limits. However, the customer's specifications require that the target length should be 19.75 inches with a tolerance of ± 0.75 inch. What percentage of the cases shipped will fall outside the customer's specifications?

64. A p chart is used to monitor the fraction defective of an integrated circuit to be used in a commercial pacemaker. A sample of 15 circuits is taken from each day's production for 30 consecutive working days. A total of 17 defectives are discovered during this period.

    a. Determine the three-sigma control limits for this process.

    b. Suppose that α, the probability of Type 1 error (that is, the probability of drawing the conclusion that the process is out of control when it is in control), is set to be .05. What control limits do you now obtain? (Use a normal approximation for your calculations.)

65. A single sampling plan is used to determine the acceptability of shipments of a bearing assembly used in the manufacture of skateboards. For lots of 500 bearings, samples of $n = 20$ are taken. The lot is rejected if any defectives are found in the sample.

    a. Suppose that AQL $= .01$ and LTPD $= .10$. Find α and β.

    b. Is this plan more advantageous for the consumer or the producer?

66. A double sampling plan is constructed as follows. From a lot of 200 items, a sample of 10 items is drawn. If there are zero defectives, the lot is accepted. If there are two or more defectives, the lot is rejected. If there is exactly one defective, a second sample of 10 items is drawn. If the combined number of defectives in both samples is two or less, the lot is accepted; otherwise it is rejected. If the lot has 10 percent defectives, what is the probability that it is accepted?

67. Hammerhead produces heavy-duty nails, which are purchased by Modulo, a maker of prefabricated housing. Modulo buys the nails in lots of 10,000 and subjects a sample to destructive testing to determine the acceptability of the lot. Modulo has established an AQL of 1 percent and an LTPD of 10 percent.

    a. Assuming a single sampling plan with $n = 100$ and $c = 2$, find α and β.

    b. Derive the sequential sampling plan that achieves the same values of α and β as the single sampling plan derived in part (*a*).

    c. By estimating the ASN curve, find the maximum value of the expected sample size Modulo will require if it uses the sequential plan derived in part (*b*).

    d. Suppose that the sequential sampling plan derived in part (*b*) is used. One hundred nails are tested with the following result: The first 80 are acceptable, the 81st is defective, and the remaining 19 are acceptable. By graphing the acceptance and the rejection regions, determine whether the sequential sampling plan derived in part (*b*) would recommend acceptance or rejection on or before testing the 100th nail.

68. For the single sampling plan derived in part (*a*) of Problem 67, suppose lots that are not passed are returned to Hammerhead.

    a. Estimate the graph of the AOQ curve by computing AOQ(p) for various values of p.

    b. Using the results of part (*a*), estimate the maximum proportion of defective nails that Modulo will be using in its construction.

69. Twenty sets of four measurements of the diameters in inches of Hot Shot golf balls were

Sample				
1	2.13	2.18	2.05	1.96
2	2.08	2.10	2.02	2.20
3	1.93	1.98	2.03	2.06
4	2.01	1.94	1.91	1.99
5	2.00	1.90	2.14	2.04
6	1.92	1.95	2.02	2.05
7	2.00	1.94	2.00	1.90
8	1.93	2.02	2.04	2.09
9	1.87	2.13	1.90	1.92
10	1.89	2.14	2.16	2.10
11	1.93	1.87	1.94	1.99
12	1.86	1.89	2.07	2.06
13	2.04	2.09	2.03	2.09
14	2.15	2.02	2.11	2.04
15	1.96	1.99	1.94	1.98
16	2.03	2.06	2.09	2.02
17	1.95	1.99	1.87	1.92
18	2.05	2.03	2.06	2.04
19	2.12	2.02	1.97	1.95
20	2.03	2.01	2.04	2.02

a. Enter the data into a spreadsheet and compute the means and the ranges for each sample.

b. Using the results of part (*a*), develop $\overline{X}$ and *R* charts similar to Figures 12–7 and 12–8. Assume three-sigma limits for the $\overline{X}$ chart.

c. Develop a histogram based on the 80 observations. Assume class intervals (1) 1.80–1.849, (2) 1.85–1.899, (3) 1.90–1.949, (4) 1.95–1.999, (5) 2.0–2.049, (6) 2.05–2.099, (7) 2.10–2.149, (8) 2.15–2.20. Based on your histogram, what distribution might accurately describe the diameter of a golf ball selected at random?

70. A *p* chart is used to monitor the number of riding lawn mowers produced. The numbers that are sent back for rework because they did not pass inspection are

Day	Number Produced	Number Rejected
1	400	23
2	480	18
3	475	24
4	525	34
5	455	17
6	385	17
7	372	12
8	358	19
9	395	24
10	405	29
11	385	16
12	376	19
13	395	23
14	405	14
15	415	25
16	440	34
17	380	26
18	318	19

Enter the data into a spreadsheet and compute standardized Z values for a p chart. Graph the Z values. Is this process in control?

71. Samples of size 50 are drawn from lots of 1,000 items. The lot is rejected if there are more than two defectives in the sample. Using a binomial approximation, graph the OC curve as a function of p, the proportion of defectives in the lot. For an AQL of .01 and an LTPD of .10, find α and β.
    a. Graph the OC curve and identify the Type 1 and Type 2 error probabilities (that is, develop a graph similar to Figure 12–14).
    b. Graph the AOQ curve and identify the value of the AOQL.

72. *a*. Develop a spreadsheet from which one may obtain a graph such as Figure 12–17 for sequential sampling. Store the values of p_0, p_1, α, and β in cell locations so that these can be altered at will. Print a graph for $p_0 = .05$, $p_1 = .20$, $\alpha = .05$, and $\beta = .10$. Allow for $n \leq 100$.
    b. Sequential sampling resulted in the following: the first 40 items were good, item 41 was defective, item 68 was defective, and items 86 and 87 were defective. Place these results on the graph you obtained in part (*a*). Is the lot accepted, rejected, or neither on or before testing the 87th item?

Appendix **12–A**

Approximating Distributions

Several probability approximations were used in this chapter. This appendix will discuss the motivation and justification for these approximations.

The complexity of a probability distribution depends upon the number of parameters that are required to specify it. The distributions considered in this chapter in descending order of complexity are

Distribution	Parameters
1. Hypergeometric	n, N, M
2. Binomial	n, p
3. Poisson	λ
4. Normal	μ, σ

It is not clear from this chart why the normal distribution should be simpler than the Poisson, since the normal is a two-parameter distribution and the Poisson one. The reason is that all normal probabilities can be obtained from a single table of the standard normal distribution, and the Poisson distribution must be tabled separately for distinct values of λ.

The binomial approximation to the hypergeometric. The hypergeometric distribution (whose formula appears in Section 12.9) is the probability that if n items are drawn from a lot of N items of which M are defective, then there are exactly m defectives in the sample. The experiment that gives rise to the hypergeometric may be thought of as sampling the items one by one without replacement. If N is much larger than n, the probability that any item sampled is defective is very close to M/N. In that case the hypergeometric probability would be close to the binomial probability with $p = M/N$ and $n = n$. Note that the binomial distribution corresponds to sampling *with*

replacement. If $N > 10n$, the binomial should provide an adequate approximation to the hypergeometric.

The Poisson approximation to the binomial. The Poisson distribution can be derived as the limit of the binomial as $n \to \infty$ and $p \to 0$, but with the product np remaining constant. Write $\lambda = np$. Then for large n and small p,

$$p\{X = m\} \approx \frac{e^{-\lambda}\lambda^m}{m!} \qquad \text{for } m = 0, 1, 2, \ldots.$$

It is not obvious under what circumstances this approximation is adequate. In general, $p < .1$ and $n > 25$ should hold, but if p is very small, then smaller values of n are acceptable, and if n is very large, then larger values of p are acceptable. For example, for $n = 10$ and $p = .01$, the binomial probability that $X = 1$ is .0914 and the Poisson probability is .0905. Values of p close to 1, such as $p = .99$, also would be acceptable because the binomial distribution with $p = .01$ is a mirror image of a binomial distribution with $p = .99$.

Normal approximations. The central limit theorem says (roughly) that the distribution of a sum of n independent identically distributed random variables approaches the normal distribution as n grows large. Because the binomial distribution is derived as the sum of n independent identically distributed Bernoulli random variables, when n is large the normal gives a good approximation to the binomial. As the normal approximation is more accurate when p is near .5, a good rule of thumb is that the approximation should be used only if $np(1 - p) > 5$.

Whenever the normal distribution is used to approximate any other distribution, it is necessary to express μ and σ in terms of the original parameters. In the binomial case, $\mu = np$ and $\sigma = \sqrt{np(1 - p)}$.

Because the normal random variable is continuous and the binomial random variable is discrete, the approximation can be improved by using the "continuity correction." In the binomial case, the events $\{X > 2\}$ and $\{X \geq 3\}$ are identical, but in the normal case they are not. The continuity correction would suggest approximating either of these cases by $\{X > 2.5\}$. The general rule is to express the original event in terms of both $>$ and $\geq$ (or $<$ and $\leq$) and to use a cutoff number halfway between the two.

For example, suppose that $n = 25$, $p = .40$, and we wish to determine $P\{X \leq 10\}$. Since $\{X \leq 10\} = \{X < 11\}$, the continuity correction cutoff is at 10.5. The exact binomial probability is .5858. The normal approximation at 10 gives

$$P\{X \leq 10\} \approx P\left\{Z < \frac{10 - (25)(.40)}{\sqrt{(25)(.40)(.60)}}\right\} = P\{Z < 0\} = 0.5,$$

and with the continuity correction

$$P\{X \leq 10\} \approx P\left\{Z < \frac{10.5 - (25)(.40)}{\sqrt{(5)(.40)(.60)}}\right\} = P\{Z < .2041\} = 0.5948.$$

The normal distribution also may be used to approximate the Poisson when λ is large ($\lambda > 10$). In that case, use $\mu = \lambda$, $\sigma = \sqrt{\lambda}$, and the continuity correction as described. For example, suppose that we wish to use a normal approximation of the probability that a Poisson random variable with parameter $\lambda = 15$ exceeds 8. Since $\{X > 8\} = \{X \geq 9\}$, the continuity correction cutoff falls at 8.5. Hence,

$$P\{X > 8\} \approx P\left\{Z > \frac{8.5 - 15}{\sqrt{15}}\right\} = P\{Z > -1.68\} = .9535.$$

The exact Poisson probability is .9626.

Appendix 12–B

Glossary of Notation for Chapter 12 on Quality and Assurance

Note: This chapter uses accepted notation for control charts and acceptance sampling. As a result, the same symbol may have one meaning in the context of control charts and another in the context of acceptance sampling.

a_1 = Cost of sampling one item.

a_2 = Cost of searching for an assignable cause.

a_3 = Cost of operating the process in an out-of-control state.

AOQ = Average outgoing quality.

AOQL = Average outgoing quality limit. The maximum value of the AOQ curve.

AQL = Acceptable quality level.

$\alpha = P\{\text{Type 1 error}\}$.

Control chart usage: α represents the probability of obtaining an out-of-control signal when the process is in control.
Acceptance sampling usage: α is the probability of rejecting good lots.

$\beta = P\{\text{Type 2 error}\}$.

Control chart usage: β represents the probability of not obtaining an out-of-control signal when the process has gone out of control.
Acceptance sampling usage: β is the probability of accepting bad lots.

$c =$ { *Control chart usage:* The expected number of defects per unit of production.
Acceptance sampling usage: The acceptance level for a single sampling plan.

c_1 = Acceptance level for first sample for a double sampling plan.

c_2 = Rejection level for the first sample for a double sampling plan.

c_3 = Acceptance level for both the first and the second samples for a double sampling plan. (Often $c_2 = c_3$.)

d_2 = A constant depending on n that relates $\overline{R}$ and σ.

d_3 = A constant depending on n that when multiplied by $\overline{R}$ gives the lower control limit for an R chart.

d_4 = A constant depending on n that when multiplied by $\overline{R}$ gives the upper control limit for an R chart.

δ = Assumed magnitude of the shift in the mean as measured in standard deviations for the economic design of $\overline{X}$ charts.

LCL = Lower control limit for a control chart.

LTPD = Lot tolerance percent defective. Unacceptable quality level.

M = Number of defectives in a lot.

μ = Population mean.

N = Number of items in lot. Used in acceptance sampling.

$n =$ $\begin{cases} \textit{Control chart usage:} \text{ Size of each subgroup for an } \overline{X} \text{ chart.} \\ \textit{Acceptance sampling usage:} \text{ Number of items sampled from a lot for a single sampling plan.} \end{cases}$

$n_1 =$ Size of the first sample for a double sampling plan.

$n_2 =$ Size of the second sample in a double sampling plan.

$\text{OC}(p) =$ Operating characteristic curve.

$p =$ $\begin{cases} \textit{Control chart usage:} \text{ True proportion of defective items produced.} \\ \textit{Acceptance sampling usage:} \text{ True proportion of defectives in a lot.} \end{cases}$

$p_0 =$ AQL.

$p_1 =$ LTPD.

$\pi =$ Probability that the process goes out of control in a single period.

$R =$ Range of a sample. The difference between the largest and the smallest values in the sample.

$s =$ Sample standard deviation of a random sample.

$\sigma =$ Population standard deviation.

$\hat{\sigma} =$ Estimator for σ.

$\text{UCL} =$ Upper control limit for a control chart.

$X =$ $\begin{cases} \textit{Control chart usage:} \text{ Value of a single measurement from the population.} \\ \textit{Acceptance sampling usage:} \text{ Number of defectives observed in a sample of } n \text{ items.} \end{cases}$

$\overline{X} =$ Arithmetic average of a random sample of n independent measurements.

$Z =$ Standard normal variate.

$z_{\alpha/2} =$ The number such that the probability of observing a value of Z that exceeds $z_{\alpha/2}$ is $\alpha/2$.

Bibliography

Aguayo, R. *Dr. Deming: The American Who Taught the Japanese about Quality.* New York: Lyle Stuart, 1990.

Alexander, F. "ISO 14001: What Does It Mean for IEs?" *IIE Solutions* 2 (January 1996), pp. 15–18.

Alexander, T. "Cray's Way of Staying Super-Duper." *Fortune*, March 18, 1985, p. 76.

Baker, K. R. "Two Process Models in the Economic Design of an $\overline{X}$ Chart." *AIIE Transactions* 13 (1971), pp. 257–63.

Boothroyd, G., and P. Dewhurst. *Product Design for Assembly*. Wakefield, RI: Boothroyd Dewhurst, Inc., 1989.

Boothroyd, G.; P. Dewhurst; and W. A. Knight. *Product Design for Manufacturing*. New York: Marcel Dekker, 1994.

Business Week. "The Quality Imperative." Special issue devoted to quality. New York: McGraw-Hill, 1991.

Cohen, L. *Quality Function Deployment: How to Make QFD Work for You*. Reading, MA: Addison Wesley, 1995.

Crosby, P. B. *Quality Is Free*. New York: McGraw-Hill, 1979.

Dahan, E. "Note on Listening to the Customer, Part I." Teaching note, Graduate School of Business, Stanford University, Stanford, CA, 1995.

DeGroot, M. H. *Probability and Statistics*. 2nd ed. Reading, MA: Addison Wesley, 1986.

Dehnad, K. *Quality Control, Robust Design, and the Taguchi Method*. Pacific Grove, CA: Wadsworth Cole, 1989.

Duncan, A. J. "The Economic Design of $\overline{X}$ Charts Used to Maintain Current Control of a Process." *Journal of the American Statistical Association* 51 (1956), pp. 228–42.

Duncan, A. J. *Quality Control and Industrial Statistics*. 5th ed. New York: McGraw-Hill/Irwin, 1986.

Dvorak, P. "Manufacturing Puts a New Spin on Design." *Machine Design* 67 (August 22, 1994), pp. 67–74.

Feigenbaum, A. V. *Total Quality Control.* 3rd ed. New York: McGraw-Hill, 1983.

Fortune. "How Jack Welch Keeps the Ideas Coming at GE." August 13, 1991.

Garvin, D. A. "Quality on the Line." *Harvard Business Review* 61 (1983), pp. 64–75.

Garvin, D. A. *Managing Quality*. New York: Free Press, 1988.

Green, P., and V. R. Rao. "Conjoint Measurement for Quantifying Judgmental Data." *Journal of Marketing Research* 8 (1971), pp. 355–63.

Herron, D. A. Private communication, 1985.

Hillkirk, J. "Europe Upstages Quest for Baldrige Award." *USA Today,* April 22, 1992.

Kackar, R. N. "Off-Line Quality Control, Parameter Design, and the Taguchi Method." *Journal of Quality Technology* 17 (1985), pp. 176–88.

Kolesar, P. "Scientific Quality Management and Management Science." In *Handbooks in Operations Research and Management Science,* vol. 4, *Logistics of Production and Inventory,* ed. S. Graves, A. H. G. Rinnooy Kan; and P. Zipkin. Chapter 13. Amsterdam: North Holland, 1993.

Kumar, S., and Y. Gupta. "Statistical Process Control at Motorola's Austin Assembly Plant." *Interfaces* 23, no. 2 (March–April 1993), pp. 84–92.

Kuster, T. "ISO 9000: A 500-lb Gorilla?" *Metal Center News* 35, no. 10 (September 1995), pp. 5–6.

Lawler, E. E., and S. A. Mohrman. "Quality Circles after the Fad." *Harvard Business Review* 63 (1985), pp. 65–71.

Leonard, H., and E. Sasser. "The Incline of Quality." *Harvard Business Review* 60 (1982), pp. 163–71.

Logothetis, N., and H. P. Wynn. *Quality through Design*. Oxford: Clarendon Press, 1989.

Martin, J. "Ignore Your Customer." *Fortune*, May 1, 1995, pp. 121–26.

Miller, J. G.; A. D. Meyer; and J. Nakane. *Benchmarking Global Manufacturing*. New York: McGraw-Hill/Irwin, 1992.

Motorola, Inc. "Motorola Corporate Quality System Review Guidelines." March 1991 edition. Referenced in S. Kumar and Y. Gupta, "Statistical Process Control at Motorola's Austin Assembly Plant," *Interfaces* 23, no. 2 (March–April 1993), pp. 84–92.

Pierce, R. J. *Leadership Perspective, and Restructuring for Total Quality*. Milwaukee: ASQC Quality Press, 1991.

Pugh, S. *Total Design.* Workingham, England: Addison Wesley, 1991.

Ross, S. M. *Applied Probability Models with Optimization Applications*. San Francisco: Holden Day, 1970.

Schoeffler, S.; R. D. Buzzell; and D. F. Heang. "Impact of Strategic Planning on Profit Performance." *Harvard Business Review* 52 (March–April 1974).

Shewhart, W. A. *Economic Control of the Quality of Manufactured Product*. New York: D. Van Nostrand, 1931.

Taguchi, G.; A. E. Elsayed; and T. Hsiang. *Quality Engineering in Production Systems*. New York: McGraw-Hill, 1989.

Transportation and Distribution 36, no. 5 (May 1995), pp. 26–28.

Ulrich, K. T., and S. D. Eppinger. *Product Design and Development*. New York: McGraw-Hill, 1995.

Velury, J. "Integrating ISO 9000 into the Big Picture." *IIE Solutions* 1 (October 1995), pp. 26–29.

Chapter **Thirteen**

Reliability and Maintainability

"Simplicity is prerequisite for reliability."
—Edsger W. Dijkstra

Chapter Overview

Purpose

To gain an appreciation of the importance of reliability, to understand the mechanisms by which products fail, and to acquire an understanding of the mathematics underlying these processes.

Key Points

1. *Preparation.* The topics in this chapter (reliability theory, warranties, and age replacement) are rarely treated in texts on operations. They are included here because of their importance and relevance to the quality movement. However, the mathematics of reliability is complex. One must have a basic understanding of random variables, probability density and distribution functions, and elementary stochastic processes. Several of these methods were also used in Chapter 5 and in Supplement 2 on queuing, appearing after Chapter 8. I suggest the reader carefully review the discussion of the exponential distribution presented there.
2. *Reliability of a single component.* Consider a single item whose time of failure cannot be predicted in advance; that is, it is a random variable, T. We assume that we know both the distribution function and density functions of T: $F(t)$ and $f(t)$, respectively. Several important quantities associated with T include the survival function $R(t) = 1 - F(t)$, which is the probability that the item survives beyond t, and the failure rate function, defined as $r(t) = f(t)/R(t)$.

   An important case occurs when the failure rate function is a constant independent of t. This results in the failure time distribution having the exponential distribution. The exponential distribution is the only one possessing the memoryless property. In this context it means that the item is neither getting better nor getting worse with age. Decreasing and increasing failure rate functions, respectively, represent the cases where the reliability of an item is improving or declining with age. The Weibull distribution is a popular choice for representing both increasing and decreasing failure rate functions.
3. *The Poisson process in reliability modeling.* The Poisson process is perhaps the most important stochastic process for applications. When interfailure times are independent and identically distributed (IID) exponential random variables, one can show that the total number of failures up to any point t follows a Poisson

distribution, and the time for *n* failures follows an Erlang distribution. Because the exponential distribution is memoryless, this process accurately describes events that occur completely at random over time.

4. *Reliability of complex equipment.* Items prone to failure are generally constructed of more than a single component. In a series system, the system fails when any one of the components fails. In a parallel system, the system fails only when all components fail. A third possibility is a *K* out of *N* system. Here the system functions as long as at least *K* components function. In this section, we show how to derive the time to failure distributions for these systems based on the time to failure distributions of the components comprising the systems.
5. *Maintenance models.* Preventive maintenance means replacing an item before it fails. Clearly, this only makes sense for items that are more likely to fail as they age. By replacing items on a regular basis before they fail, one can avoid the disruptions that result from unplanned failures. Based on knowledge of the items' failure mechanisms and costs of planned and unplanned replacements, one can derive optimal replacement strategies. The simplest case gives a formula for optimal replacement times, which is very similar to the EOQ formula derived in Chapter 4.
6. *Warranties*. A warranty is an agreement between the buyer and seller of an item in which the seller agrees to provide restitution to the buyer in the event the item fails within the warranty period. Warranties are common for almost all consumer goods, and extended warranties are a big business. In this section, we examine two kinds of warranties: the free replacement warranty and the pro rata warranty. The free replacement warranty is just as it sounds: the seller agrees to replace the item when it fails during the warranty period. In the case of the pro rata warranty, the amount of restitution depends on the remaining time of the warranty. (Pro rata warranties are common for tires, for example, where the return depends on the remaining tread on the tire.)
7. *Software reliability.* Software is playing an increasingly important role in our lives. With the explosive growth of personal computers, the market for personal computer software has become enormous. Microsoft took advantage of this growth to become one of the world's major corporations within a decade of its founding. There is a lot more to the software industry than personal computers, however. Large databases, such as those managed by the IRS or your state Department of Motor Vehicles require massive information retrieval systems. Some predicted that Ronald Reagan's Star Wars missile defense system was doomed to failure because it would be impossible to design reliable software for it. Software failures can be just as catastrophic as hardware failures, causing major systems to fail.

Our critical systems continue to grow and become more complex. As complexity grows, reliability is threatened. During the week of August 11, 2003, a downed power line near Cleveland, Ohio, triggered one of the worst power outages in U.S. history. Virtually the entire East Coast of the United States and parts of Canada were affected, with several deaths attributed to the blackout. How could such a thing have occurred? The answer is that our electrical grid is linked all over the country and can be brought down even by minor problems. Such systems need to be designed with more attention to their reliability. As our population grows and our basic systems become more complex, such catastrophes will become more common. We depend on the reliability of our infrastructure every day.

In operations management, quality has been a key issue in recent years. The dramatic success of the Japanese has been attributed to a large extent to the quality of their manufactured goods. Quality is multidimensional (see Section 12.13), but reliability is certainly a key component. In Table 12–6, we reported on competitive priorities in Europe, Japan, and the United States. Product reliability ranked number one for the group of Japanese firms surveyed.

When we think of Japan's economic success, it is the automobile industry that many of us think of first. Japanese automakers have had a steadily growing market share in the United States. Why have the Japanese been so successful in the United States? Perceived product quality is probably the key reason that so many Americans choose to purchase Japanese automobiles. But what dimension of quality is most important? A likely answer is product reliability. Annual surveys conducted by the Consumer's Union attest to the continued exceptional reliability of Japanese-made automobiles.

Reliability as a field separated from the mainstream of statistical quality control in the 1950s with the postwar growth of the aerospace and the electronics industries in the United States. The Department of Defense took a keen interest in reliability studies when it became painfully apparent that there was a serious problem with the reliability of military components and systems. Garvin (1988) reports that in 1950 only one-third of the Navy's electronic devices were working properly at any given time, and that for every vacuum tube the military had in an operational state there were nine tubes in warehouses or on order. According to Amstadter (1971), the yearly cost of maintaining some military systems in an operable state has been as high as 10 times the original cost of the equipment.

What is the difference between statistical quality control and reliability? Statistical quality control is concerned with monitoring processes to ensure that the manufactured product conforms to specifications. The random variables of interest are numbers of defects and degree of conformance variation. Reliability considers the performance of a product over *time.* The random variables of interest concern the amount of elapsed time between failures after the product is placed into service. A definition of reliability that emphasizes its close association with quality has been suggested by O'Connor (1985): reliability is a time-based concept of quality. Alternatively, reliability is the probability that a product will operate adequately for a given period in its intended application (Amstadter, 1971).

A reviewer of the first edition said that from the student's point of view, reliability is "a narrow engineering issue which almost never makes it to the front page of *The Wall Street Journal.*" This couldn't be further from the truth. Three of the most significant disasters of recent times were the result of reliability failures: The accidents at the nuclear plants at Three Mile Island in Pennsylvania and at Chernobyl in the former Soviet Union, and the dramatic losses of the *Challenger* and *Columbia* space shuttles. Reliability concerns each of us every day. Our lives depend on the reliability of cars, commuter trains, and airplanes. Our livelihoods depend on the reliability of power generation, telephones, and computers. Our health depends on the reliability of pollution control systems, heating and air-conditioning systems, and emergency medical care systems.

Reliability and risk are closely related. Risks of poor reliability are of concern to both the producer and the consumer. Some aspects of risk from the producer's point of view include

1. *Competition.* Product reliability is an important component of perceived quality by the consumer. Highly unreliable products do not gain customer loyalty and eventually disappear.

2. *Customer requirements.* The U.S. government required weapons systems with clearly specified reliability levels when it found that maintenance costs for these systems were becoming prohibitive. Today, reliability requirements established by the buyer are common.
3. *Warranty and service costs.* Warranties, which will be treated in the second part of this chapter, are a significant financial burden to the manufacturer when products are unreliable. American automobile firms provided extended warranties as an incentive to the consumer in the past to boost sales (the most notable was Chrysler's 7-year, 70,000-mile warranty). There is no evidence that these warranties were accompanied by improved reliability. As a result, these programs proved to be very costly to the automobile firms.
4. *Liability costs.* Largely as an outgrowth of the efforts of Ralph Nader, the U.S. Congress has enacted legislation that makes a manufacturer liable for the consequences of failures in product performance resulting from faulty design or manufacture. Liability losses have had the effect of shifting some of the costs of poor reliability from the consumer to the manufacturer.

Some of the risks of poor reliability borne by the consumer include

1. *Safety.* There is no doubt that equipment failure results in human death. Approximately 35,000 Americans die in automobile accidents on the nation's roads each year. Undoubtedly, some portion of these are attributable to mechanical failure. Travelers die in airplane accidents, many of which are the result of equipment failure. The failure of the nuclear plants at Three Mile Island, Chernobyl, and Fukishima resulted in human death and injury from radiation exposure. Safety and reliability are closely linked.
2. *Inconvenience.* Even though many failures do not result in death, they can be a source of frustration and delay. Delays at airports are common because some piece of equipment on board the plane fails to operate properly. Automobile breakdowns may leave motorists stranded for hours. Failure of communication equipment, computer equipment, or power generation plants can cripple businesses.
3. *Cost.* Poor reliability costs everyone in the end. For this reason, consumers are willing to pay a premium for products with higher reliability. The Japanese have learned this lesson well, and every indication is that they will continue the strategy of increasing market share by producing more reliable products than their competitors.

Why study reliability? We need to understand the probability laws governing failure patterns in order to better design processes to build reliable systems. Incorrect analysis can lead to disastrous consequences. For example, the so-called Rasmussen report (U.S. Nuclear Regulatory Commission, *Reactor Safety Study,* 1975) predicted that it would be hundreds of years before we could expect a major accident in a nuclear plant. Given that we have since observed two major accidents, the analysis in this report is clearly flawed.

Reliability is an issue of concern for operations management from two perspectives. First, to implement total quality management, we must understand how and why products fail. Reliability will continue to be a key component of quality. Designing an effective quality delivery system will require understanding the randomness of failure patterns. Second, we need to understand the failure patterns of the equipment used in the manufacturing process. In this way, we can develop effective maintenance policies for that equipment.

13.1 RELIABILITY OF A SINGLE COMPONENT

Introduction to Reliability Concepts

In order for readers to better understand and appreciate some of the definitions and concepts introduced in this chapter, this section starts with an example.

Example *13.1* In 1970 the U.S. Army purchased 1,000 identical capacitors for use in short-distance radio transmitters. The army maintained detailed records on the failure pattern of the capacitors with the following results:

Number of years of operation	1	2	3	4	5	6	7	8	9	10	>10
Number of failures	220	158	121	96	80	68	47	40	35	25	110

Based on these data, the army wanted to estimate the probability distribution associated with the failure of a capacitor chosen at random.

Define the random variable T as the time that the capacitor will operate before failure. We can estimate the cumulative distribution function of T, $F(t)$, using the data provided in the table. In symbols, $F(t) = P\{T \leq t\}$, and in words, $F(t)$ is the probability that a component chosen at random fails at or before time t.

In order to estimate $F(t)$ from the given data, we find the cumulative number of failures and the proportion of the total this number represents each year. We have

Number of years of service	1	2	3	4	5	6	7	8	9	10	>10
Cumulative failures	220	378	499	595	675	743	790	830	865	890	1,000
Proportion of total	.220	.378	.499	.595	.675	.743	.790	.830	.865	.890	1.0

The proportions are estimates of $F(t)$ for $t = 1, 2, \ldots$. These probabilities may be used directly to compute various quantities of interest by treating T as a discrete random variable, or they may be used to estimate the parameters of a continuous distribution. We will use the discrete version to answer several questions about the lifetime of the capacitors. For example, suppose that we wish to determine

a. The probability that a capacitor chosen at random lasts more than five years.
b. The proportion of the original 1,000 capacitors put into operation that fail in year 6.
c. The proportion of components that survive for at least five years that fail in year 6.
d. The proportion of components that survive at least eight years that fail in year 9.

Solution

a. $P\{T > 5\} = 1 - P\{T \leq 5\} = 1 - .675 = .325$.
b. $P\{T = 6\} = P\{T \leq 6\} - P\{T \leq 5\} = .743 - .675 = .068$
c. At first glance, this question might appear to be the same as part (*b*). However, there is an important difference. Part (*b*) asks for the proportion of the original set of capacitors failing in year 6, whereas part (*c*) asks for the proportion of components lasting *at least five years* that fail in year 6. This is a *conditional* probability.

$$P\{T = 6 \mid T > 5\} = \frac{P\{T = 6,\ T > 5\}}{P\{T > 5\}} = \frac{P\{T = 6\}}{P\{T > 5\}} = \frac{.068}{.325} = 0.209.$$

Notice that the events $\{T = 6,\ T > 5\}$ and $\{T = 6\}$ are equivalent since $\{T = 6\} \subset \{T > 5\}$.

d. $$P\{T = 9 | T \geq 8\} = \frac{P\{T = 9\}}{P\{T \geq 8\}} = \frac{.035}{.170} = 0.206,$$

which is almost precisely the same as the proportion of components surviving more than five years that fail in year 6. In fact, the proportion of components that survive n years and fail in year $n + 1$ is very close to .20 for $n = 0, 1, 2, \ldots, 9$. As we will later see, this results in a failure distribution with some unusual properties.

Preliminary Notation and Definitions

Many new concepts were introduced in Example 13.1. We will now formalize the definitions to be used throughout this chapter.

As before, define the random variable T as the lifetime of the component. We assume that T has cumulative distribution function $F(t)$ given by

$$F(t) = P\{T \leq t\}.$$

In what follows we will treat $F(t)$ as a differentiable function of t, so that the probability density function $f(t)$, given by the equation

$$f(t) = \frac{dF(t)}{dt},$$

will exist.

In addition to the distribution and density functions of the random variable T, we will be interested in related functions. One is the reliability function (also known as the survival function). The reliability function of the component, which we call $R(t)$, is given by

$$R(t) = P\{T > t\} = 1 - F(t).$$

In words, $R(t)$ is the probability that a new component will survive past time t. Notice that this implies that $F(t)$ is the probability that a new component will *not* survive past time t.

Consider the following conditional probability:

$$P\{t < T \leq t + s \,|\, T > t\}.$$

This is the conditional probability that a new component will fail between t and $t + s$ given that it lasts beyond t. We may think of this conditional probability in the following way: Interpret t as now and s as an increment of time into the future. The event $\{T > t\}$ means that the component has survived until the present, or in other words, that it is still working. The conditional event $\{t < T \leq t + s \,|\, T > t\}$ means that the component is working now but will fail before an additional s units of time have passed.

Recall from elementary probability theory that for any events A and B

$$P\{A \,|\, B\} = \frac{P\{A \cap B\}}{P\{B\}}.$$

In the special case where $A \subset B$, $A \cap B = A$, so

$$P\{A \,|\, B\} = \frac{P\{A\}}{P\{B\}} \qquad \text{when } A \subset B.$$

Identify the event $A = \{t < T \leq t + s\}$ and $B = \{T > t\}$. A little reflection shows that $A \subset B$ in this particular case, so

$$P\{t < T \leq t + s \,|\, T > t\} = \frac{P\{t < T \leq t + s\}}{P\{T > t\}} = \frac{F(t + s) - F(t)}{R(t)}.$$

We will divide by s and let s approach zero.

$$\lim_{s\to 0} \frac{1}{s} \frac{F(t+s) - F(t)}{R(t)} = \frac{f(t)}{R(t)}.$$

This ratio turns out to be a fundamental quantity in reliability theory.

Define

$$r(t) = \frac{f(t)}{R(t)}.$$

We call $r(t)$ the *failure rate function.* Its derivation is the best way to understand what the failure rate function means. For positive s, the conditional probability used to derive $r(t)$ is the probability that a component that has survived up until time t fails between times t and $t + s$. Dividing by s and letting s go to zero is the same way one derives a first derivative. Hence, the failure rate function is the rate of change of the conditional probability of failure at time t. It can be considered a measure of the likelihood that a component that has survived up until time t fails in the next instant of time.

The failure rate function is a fundamental quantity in reliability theory but, like the probability density function, does not have a direct physical interpretation. However, for values of Δt sufficiently small, the term $r(t)\Delta t$ is the probability that an item that survives to time t fails between t and $t + \Delta t$. How does one determine if Δt is sufficiently small? In general, Δt should be small relative to the lifetime of a typical component. However, the only way to be certain as to whether $r(t)\Delta t$ is a good approximation to this conditional probability is to compute $P\{t < T \le t + \Delta t \,|\, T > t\}$ directly.

Example *13.2*

The length of time that a particular piece of equipment operates before failure is a random variable with cumulative distribution function

$$F(t) = 1 - e - 0.043t^{2.6}.$$

Consider the following:

a. The failure rate function.

b. The probability that the equipment operates for more than five years without experiencing failure.

c. Suppose that 100 pieces of the equipment are placed into service in year 0. What fraction of the units surviving four years fail in year 5? Can one accurately estimate this proportion using only the failure rate function?

d. What fraction of units surviving four years fail in the first month of year 5? Can this be accurately estimated using the failure rate function?

Solution

a.

$$\begin{aligned} f(t) = \frac{dF(t)}{dt} &= -e^{-0.043t^{2.6}} \frac{d}{dt}(-0.043t^{2.6}) \\ &= (0.043)(2.6)t^{1.6}\, e^{-0.043t^{2.6}} \\ &= 0.1118t^{1.6}\, e^{-0.043t^{2.6}} \end{aligned}$$

Since $R(t) = 1 - F(t) = e^{-0.043t^{2.6}}$, it follows that

$$r(t) = \frac{f(t)}{R(t)} = 0.1118t^{1.6}.$$

b. $$P\{T > 5\} = R(5) = e^{-0.043(5)^{2.6}} = e^{-2.8235} = 0.0594.$$

c. We will compute this directly and compare the result with $r(t)\Delta t$. The proportion of units surviving four years that fail in year 5 is

$$\begin{aligned} P\{4 < T \leq 5 \,|\, T > 4\} &= \frac{F(5) - F(4)}{R(4)} = \frac{R(4) - R(5)}{R(4)} \\ &= \frac{e^{-0.043(4)^{2.6}} - e^{-0.043(5)^{2.6}}}{e^{-0.043(4)^{2.6}}} \\ &= \frac{0.2059 - 0.0594}{0.2059} \\ &= 0.7115. \end{aligned}$$

This means that about 71 percent of the machines surviving four years will fail in the fifth year. As about 94 percent of the units fail in the first five years of operation [$F(5) = .9406$], a value of $\Delta t = 1$ year is probably too large for $r(t)\Delta t$ to give a good approximation. In fact, we see that $r(4)(1) = (0.1118)(4)^{1.6} = 1.0274$.

d. Because one month corresponds to $\frac{1}{12} = 0.0833$ year, we wish to compute

$$P\{4 < T \leq 4.0833 \,|\, T > 4\} = \frac{F(4.0833) - F(4)}{R(4)} = \frac{0.2059 - 0.1887}{0.2059} = 0.0836.$$

Here $\Delta t = \frac{1}{12}$ should be sufficiently small to use the failure rate function to estimate this probability. We obtain $r(4)\frac{1}{12} = 0.0856$.

The Exponential Failure Law

The exponential distribution plays a fundamental role in reliability theory and practice because it accurately describes the failure characteristics of many types of operating equipment. The exponential law can be derived in several ways. We will consider a derivation that utilizes the failure rate function.

We know that the failure rate function $r(t)$ is given by the formula

$$r(t) = f(t)/R(t)$$

and is a measure of the likelihood that a unit that has been operating for t units of time fails in the next instant. Consider the following case: $r(t) = \lambda$, for some constant $\lambda > 0$. This means that the likelihood that a working unit fails in the next instant of time is independent of how long it has been operating. This implies that the unit does not exhibit any signs of aging. It is equally likely to fail in the next instant whether it is new or old. We will derive the probability distribution of the lifetime T that corresponds to a constant failure rate function.

We can determine a solution to the equation $r(t) = \lambda$ by noting that since $R(t) = 1 - F(t)$,

$$f(t) = \frac{dR(t)}{dt} = -R'(t).$$

Hence, the equation $r(t) = \lambda$ may be written in the form

$$\frac{-R'(t)}{R(t)} = \lambda$$

or

$$R'(t) = -\lambda R(t).$$

This is the simplest first-order linear differential equation. Its solution is

$$R(t) = e^{-\lambda t}.$$

It follows that the distribution function $F(t)$ is given by

$$F(t) = 1 - e^{-\lambda t}$$

and the density function $f(t)$ is given by

$$f(t) = \lambda e^{-\lambda t}.$$

This is known as the exponential distribution. It depends on the single parameter λ, which represents a rate of occurrence. If T has the exponential distribution with parameter λ, then T corresponds to the lifetime of a component that exhibits no aging over time; that is, a component that has survived up until time t_1 is equally likely to fail in the next instant of time as one that has survived up until time t_2 for any times t_1 and t_2. The expected failure time is $1/\lambda$. The standard deviation of the failure time is also $1/\lambda$. The exponential density and distribution functions appear in Figure 13–1.

Example *13.3*

Because the exponential distribution has a constant failure rate function, it is likely that the failure law for the capacitor described in Example 13.1 is exponential. This follows because we observed that the proportion of the capacitors having survived for n years that failed in year $n + 1$ was the same for $n = 0, 1, \ldots$, which was approximately 20 percent. In order to estimate the value of λ, note that the proportion failing in the first year is also 20 percent, which gives $F(1) = 0.20 = 1 - e^{-\lambda}$. Solving for λ results in $e^{-\lambda} = 0.8$, or $\lambda = -\ln(0.8) = 0.223$.

We can compare the expected number of capacitors that would fail if the true lifetime distribution were exponential versus the actual number that failed, to see if the exponential distribution provides a reasonable fit of the data.

Number of years of operation	1	2	3	4	5	6	7	8	9	10	>10
Number of failures	220	158	121	96	80	68	47	40	35	25	110
Expected number of failures under exponential law with $\lambda = 0.223$	200	160	128	102	82	66	52	42	34	27	107

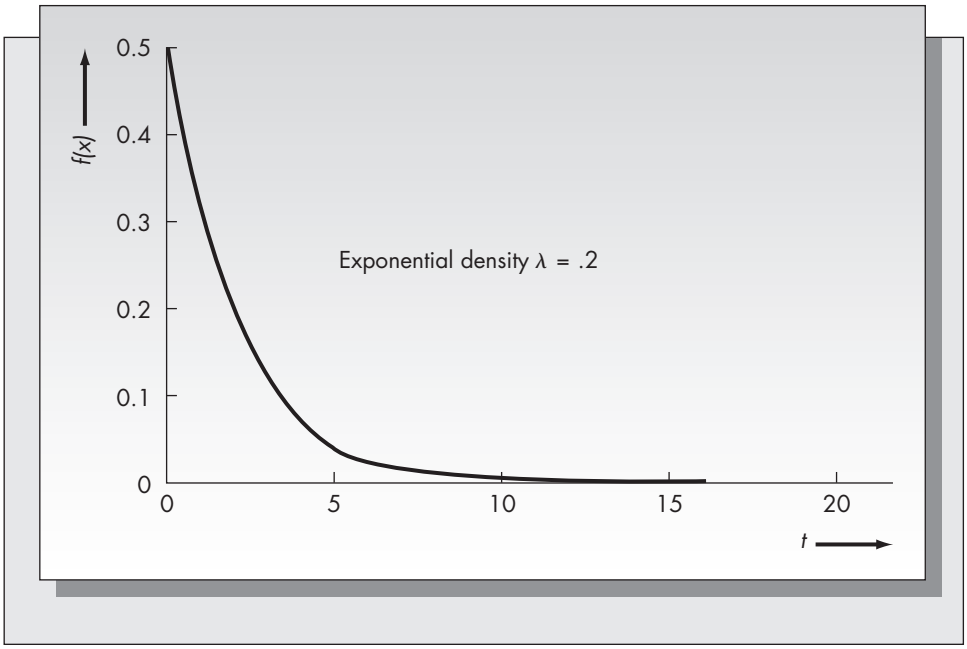
The expected number of failures for each year is obtained by multiplying the probability of failure for a given year, assuming the exponential law, by the total number of units. For example, one finds the expected number of failures for year 3 by multiplying 1,000 by $F(3) - F(2) = e^{-2\lambda} - e^{-3\lambda} = 0.1280$. Clearly there is a close agreement between the actual number of failures and the expected number, indicating that the exponential distribution provides a good fit of the observed historical data.[1]

We can now answer several questions about the capacitors using the exponential distribution directly. For example, suppose that we wish to find

a. The probability that a capacitor chosen at random lasts more than eight years.

b. The proportion of capacitors that survive three years that also survive at least three additional years.

[1] It is easy to verify by a formal goodness-of-fit test that the exponential distribution fits these data very closely.



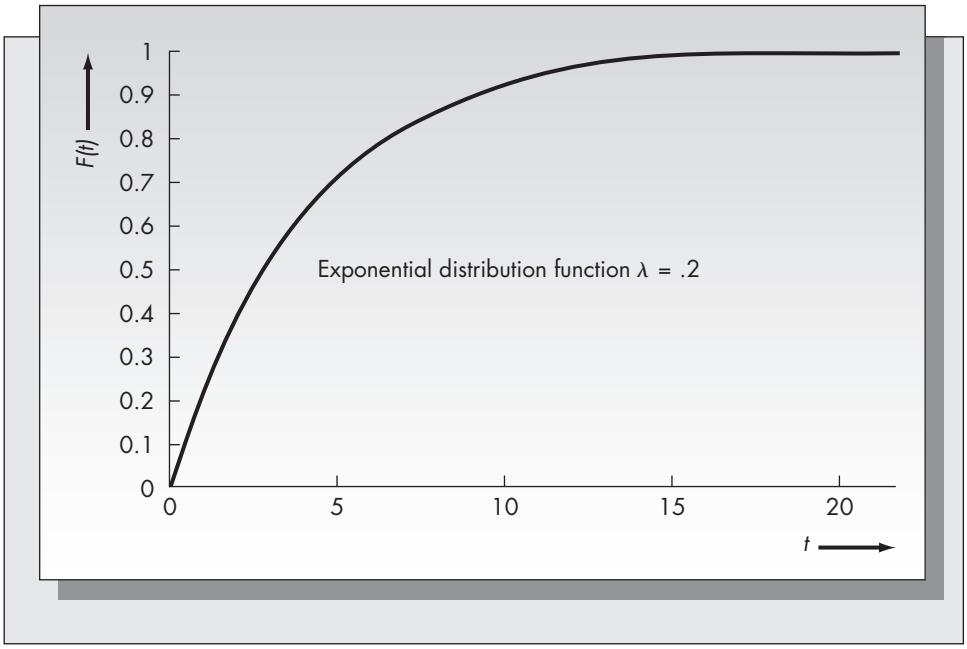


FIGURE 13–1
The exponential density and distribution functions

Solution

a. $P\{T > 8\} = e^{-\lambda t} = e^{-(.223)(8)} = e^{-1.784} = .1680.$

b. We wish to compare $P\{T > 6 \mid T > 3\}$. Using the laws of conditional probability we have

$$P\{T > 6 \mid T > 3\} = \frac{P\{T > 6, T > 3\}}{P\{T > 3\}}$$

$$= \frac{P\{T > 6\}}{P\{T > 3\}} = \frac{.2624}{.5122} = .5122.$$

It is not a coincidence that this is the same as the unconditional probability that a new capacitor lasts more than three years. This property is known as the memoryless property of the exponential distribution.

The memoryless property of the exponential distribution relates to the following conditional probability:

$$P\{T > t + s \,|\, T > t\}.$$

This is the probability that the component survives past time $t + s$ given that it has survived until time t. If we think of time t as now, then this is the probability that a component that is currently functioning continues to function for at least another s units of time. If T follows an exponential failure law, then

$$\begin{aligned} P\{T > t + s \,|\, T > t\} &= \frac{P\{T > t + s, T > t\}}{P\{T > t\}} \\ &= \frac{P\{T > t + s\}}{P\{T > t\}} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\ &= e^{-\lambda s} \\ &= P\{T > s\}. \end{aligned}$$

Note that as $\{T > t + s\} \subset \{T > t\}$, the events $\{T > t + s, T > t\}$ and $\{T > t + s\}$ are equivalent. The last expression, $P\{T > s\}$, is the *unconditional* probability that a new component will last at least s units of time. That is, we have demonstrated that if the component has been operating for t units of time without failure, then the probability that it continues to operate for at least another s units of time is the same as the probability that a new component operates for at least s units of time. This means that there is no aging. The likelihood of failure is independent of how long the component has been operating. However, we required that the lifetime distribution be exponential. In fact, the exponential distribution is the *only* continuous distribution possessing the memoryless property; that is, it is the only one for which $P\{T > t + s \,|\, T > t\} = P\{T > s\}$.

When we say that an item fails completely at random, we mean that the failure law for the item is exponential. Events that occur completely at random over time follow a *Poisson process*. The Poisson process is discussed in Section 13.3.

Problems for Section 13.1

1. Three hundred identical cathode ray tubes (CRTs) placed into service simultaneously on January 1, 1976, experienced the following numbers of failures through December 31, 1988:

Year	1983	1984	1985	1986	1987	1988
Number of failures	13	19	16	34	21	38

Assume that there were no failures before 1983.

a. Based on these data, estimate the cumulative distribution function (CDF) of a CRT chosen at random.

Using the results of part (*a*), estimate the probability that a CRT chosen at random

b. Lasts more than 5 years.

c. Lasts more than 10 years.

d. Lasts more than 12 years.

e. That has survived for 10 years fails in the 11th year of operation.

2. Suppose that the cumulative distribution function of the lifetime of a piece of operating equipment is given by

$$F(t) = 1 - e^{-0.6t} - 0.6te^{-0.6t},$$

where t is measured in years of continuous operation.

a. Determine the reliability function.

b. Determine the failure rate function.

c. What is the probability that this piece of equipment fails in the first year of operation?

d. What is the probability that this piece of equipment fails in the fifth year of operation?

e. What proportion of the equipment surviving four years fails in the fifth year? (Calculate without using the failure rate function.)

f. Does $r(4)$ closely approximate the answer to part (*e*)? Why or why not?

g. What proportion of the equipment surviving four years fails in the first month of the fifth year? (Calculate using the failure rate function.)

3. A large number of identical items are placed into service at time 0. The items have a failure rate function given by

$$r(t) = 1.105 + 0.30t,$$

where t is measured in years of operation.

a. Derive $R(t)$ and $F(t)$.

b. If 300 items are still operating at time $t = 1$ year, approximately how many items would you expect to fail between year 1 and year 2?

c. Does the value of $r(1)$ yield a good approximation to the conditional probability computed in part (*b*)? Why or why not?

d. Repeat the calculation of part (*b*), but determine the expected number of items that fail between $t = 1$ year and $t = 1$ year plus 1 week. Does $r(t)\Delta t$ provide a reasonable approximation to the conditional probability in this case? Why or why not?

4. A microprocessor that controls the tuner in color TVs fails completely at random (that is, according to the exponential distribution). Suppose that the likelihood that a microprocessor that has survived for k years fails in year $k + 1$ is .0036. What is the cumulative distribution function of the time until failure of the microprocessor?

5. A pressure transducer regulates a climate control system in a factory. The transducer fails according to an exponential distribution with rate one failure every five years on average.

a. What is the cumulative distribution function of the time until failure?

b. What is the probability that a transducer chosen at random functions for eight years without failure?

c. What is the probability that a transducer that has functioned for eight years continues to function for another eight years?

6. For the pressure transducer mentioned in Problem 5, use the failure rate function to estimate the likelihood that a transducer that has been operating for six years fails in the seventh year. How close is the approximation to the exact answer?

13.2 INCREASING AND DECREASING FAILURE RATES

Although the constant failure rate function that leads to the exponential law is significant in reliability theory, there are other important failure laws as well. Most of us are more familiar with items that possess increasing failure rate functions. That is, they are more likely to fail as they get older. Decreasing failure rate functions also occur frequently. New products often have a high failure rate because of the "burn-in" phase, in which the defective items in the population are weeded out.

An important class of failure rate functions that includes both increasing and decreasing failure rate functions is of the form

$$r(t) = \alpha\beta t^{\beta-1} \qquad \text{where } \alpha \text{ and } \beta > 0.$$

Here $r(t)$ is a polynomial function in the variable t that depends on the two parameters α and β. When $\beta > 1$, $r(t)$ is increasing, and when $0 < \beta < 1$, $r(t)$ is decreasing. Typical failure rate functions for these cases are pictured in Figure 13–2.

This form of $r(t)$ will yield another differential equation in $R(t)$. It can be shown that the solution in this case will be

$$R(t) = e^{-\alpha t^{\beta}} \qquad \text{for all } t \geq 0$$

or

$$F(t) = 1 - e^{-\alpha t^{\beta}} \qquad \text{for all } t \geq 0.$$

This distribution is known as the Weibull distribution. It depends on the two parameters α and β, and as we saw earlier, when $0 < \beta < 1$, it corresponds to the lifetime of an item with a decreasing failure rate, and when $\beta > 1$, it corresponds to the

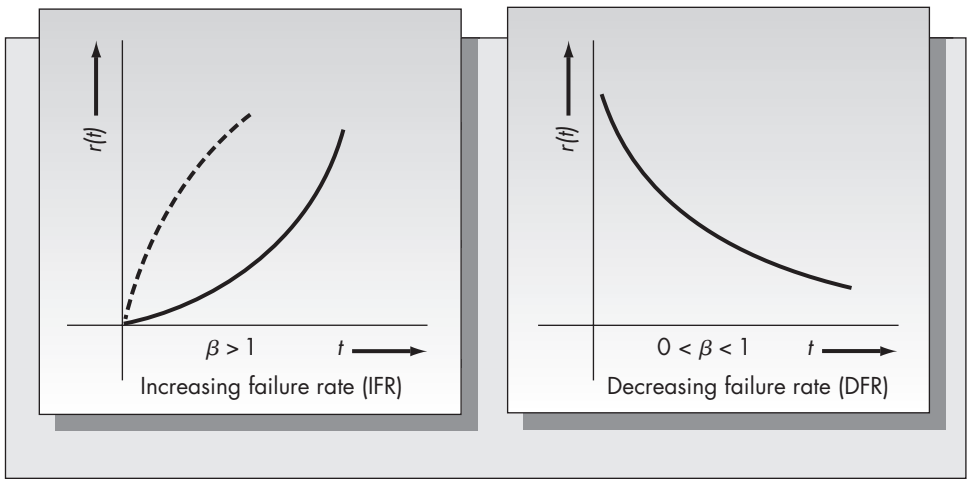


FIGURE 13–2
Failure rate functions for the Weibull lifetime distribution

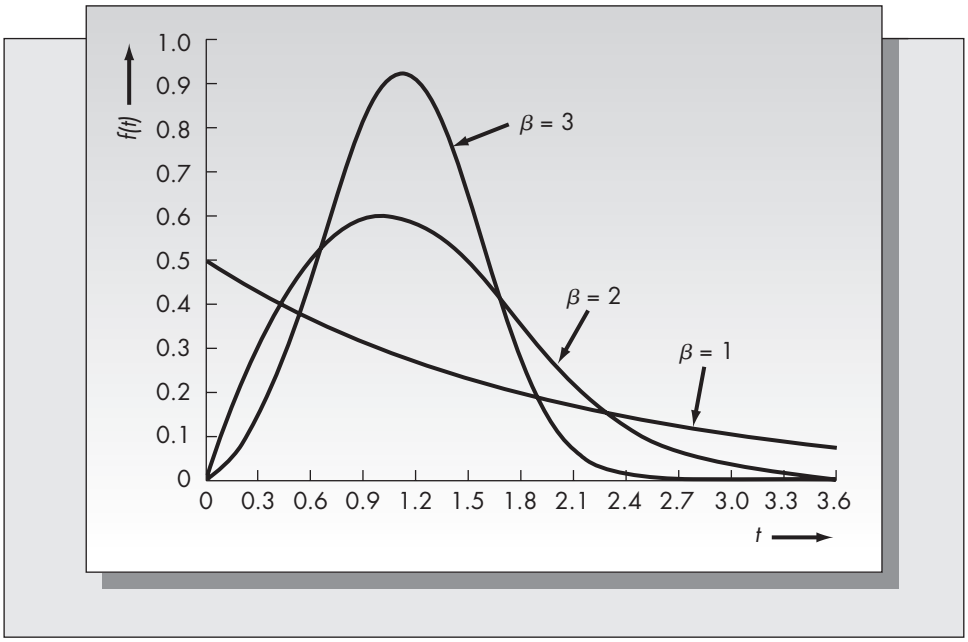


FIGURE 13–3
Weibull densities for various values of $\beta(\alpha = 0.5)$

lifetime of an item with an increasing failure rate. Because it is often true that empirical failure rate functions (i.e., those that are observed from test data) are closely approximated by polynomials, the Weibull distribution is an accurate description of the failure law of many types of operating equipment. Note that when $\beta = 1$ the Weibull reduces to the exponential. Various Weibull densities appear in Figure 13–3.

Example *13.4*

A local manufacturer of copying equipment includes a repair warranty with each copier. Virtually all of his equipment exhibit an increasing failure rate. Based on historical repair data, the failure rate for Model 25cc7 is accurately described by the function $r(t) = 2.7786t^{1.3}$, where t is measured in months of continuous operation. What is the probability that the time between two successive failures of this equipment exceeds two months of operation?

Solution

Because $r(t)$ is a polynomial in t, the distribution of the time until failure is the Weibull distribution. It is necessary to identify the values of α and β. We have that

$$2.7786t^{1.3} = \alpha\beta t^{\beta-1}.$$

It follows that $\beta - 1 = 1.3$, or $\beta = 2.3$. Since $\alpha\beta = 2.7786$, we obtain

$$\alpha = 2.7786/\beta = 2.7786/2.3 = 1.208.$$

We are required to compute $P\{T > 2\} = R(2)$. Substituting $t = 2$ into the equation for $R(t)$ we obtain

$$R(2) = e^{-1.208\times(2)^{2.3}} = 4.977 \times 10^{-4}.$$

Sometimes neither increasing nor decreasing failure rate functions accurately describe the failure characteristics of particular equipment. A typical case in point is the "bathtub" failure rate function pictured in Figure 13–4. In the early phases of the product life, the failure rate is decreasing. This follows because defective components fail quickly, causing failure rates to be high initially. This is commonly known as the infant mortality stage. Once bad components are weeded out, the failure rate remains

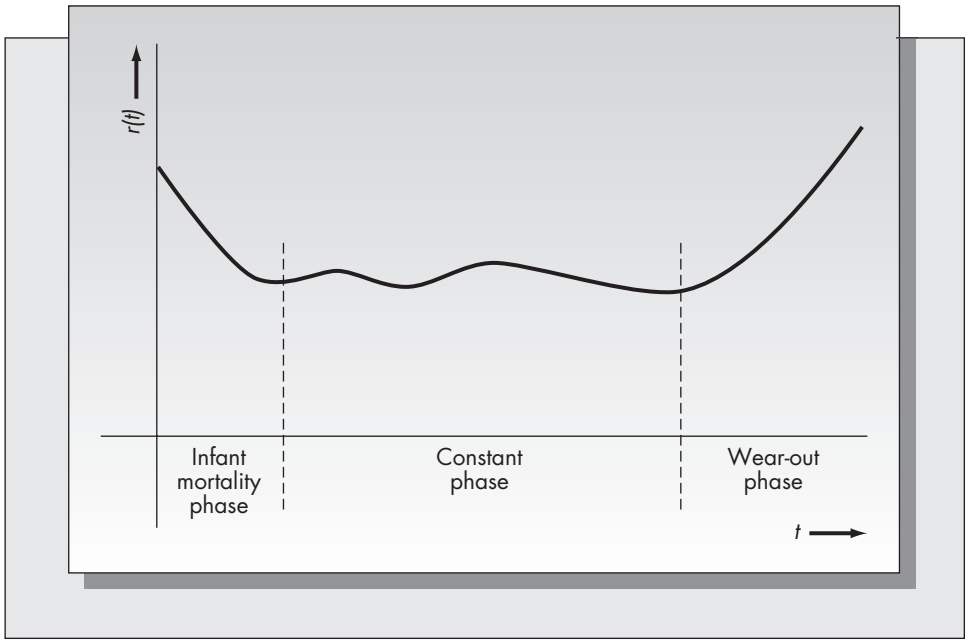


FIGURE 13–4
"Bathtub" failure rate function

constant until aging begins. At that time we enter the wear-out phase and the failure rate starts to increase.

If $r(t)$ is an arbitrary failure rate function, then it can be shown that the reliability function $R(t)$ is given by

$$R(t) = \exp\left(-\int_0^t r(u)\,du\right).$$

It is easy to verify that both the exponential and the Weibull cases satisfy this relationship. Sometimes, such as with the bathtub failure rate function, finding an explicit representation for the function $r(t)$ is difficult. In such cases it is possible to approximate $r(t)$ by step functions. We will not illustrate the procedure here.

Problems for Section 13.2

7. A piece of equipment has a lifetime T (measured in years) that is a continuous random variable with cumulative distribution function

$$F(t) = 1 - e^{-t/10} - (t/10)\,e^{-t/10} \qquad \text{for all } t \geq 0.$$

   a. What is the probability density function of T?
   b. What is the probability that a piece of equipment survives more than 20 years?
   c. What is the probability that a piece of equipment survives more than 10 years but fewer than 20 years?
   d. What is the probability that a piece of equipment survives more than 20 years given that it has survived for 10 years?

8. For the equipment mentioned in Problem 7,
   a. Derive the failure rate function $r(t)$, and draw a graph of the function.
   b. Without using the failure rate function, determine the probability that a piece of equipment that has survived 20 years of operation fails in the 21st year.
   c. Does $r(20)$ accurately estimate your answer to part (*b*)? Why or why not?

9. The Air Force maintains enormous amounts of data on engine failure times. A particular engine has experienced a failure pattern whose failure rate function is closely approximated by

$$r(t) = 0.000355e^{2.2t},$$

where t is in flying hours.

a. What are the reliability and the cumulative distribution functions of the time until failure?

b. Determine the value of t such that the likelihood that an engine fails before t is the same as the likelihood that an engine fails after t.

10. A sample of high-capacity resistors is tested until failure and the results fitted to a Weibull probability model. Based on these tests, the reliability function of a resistor is estimated to be

$$R(t) = e^{-0.0013t^{1.83}}.$$

a. What is the failure rate function for these resistors?

b. Is this resistor more likely to fail as it ages?

c. What is the probability that a resistor will function for more than 30 hours without failure?

d. Suppose that a resistor has been operating for 50 hours. What is the probability that it fails in the 51st hour? [Use the results of part (*a*) for your calculations.]

13.3 THE POISSON PROCESS IN RELIABILITY MODELING[2]

Consider a single piece of operating equipment that fails completely at random. As we saw in Section 13.2, that means that the time until failure follows the exponential distribution. Suppose that when the item fails, it is immediately repaired, or that the repair time is sufficiently small compared with the interfailure time that it can be ignored. Thus we have a process in which events (failures) occur over time and the times between successive failures are independent identically distributed exponential random variables. Let T_1, T_2, . . . be random variables corresponding to the times between successive failures. Each of the random variables has distribution function $F(t)$ and reliability function $R(t)$ given by

$$F(t) = 1 - e^{-\lambda t},$$
$$R(t) = e^{-\lambda t},$$

where λ is the rate at which failures occur. We also define a related sequence of random variables W_1, W_2, . . . , which are defined by the equations

$$W_1 = T_1,$$
$$W_2 = T_1 + T_2,$$
$$W_3 = T_1 + T_2 + T_3,$$
and so on.

Interpret W_n as the time of the nth failure. Finally, we introduce the process $N(t)$. Define $N(t)$ as the number of failures that occur up until time t. $N(t)$ is a *stochastic* process because for each fixed value of t, $N(t)$ is a random variable. Clearly, $N(t)$ is

[2] The Poisson process was also discussed in Chapter 7 as a model of random arrivals.

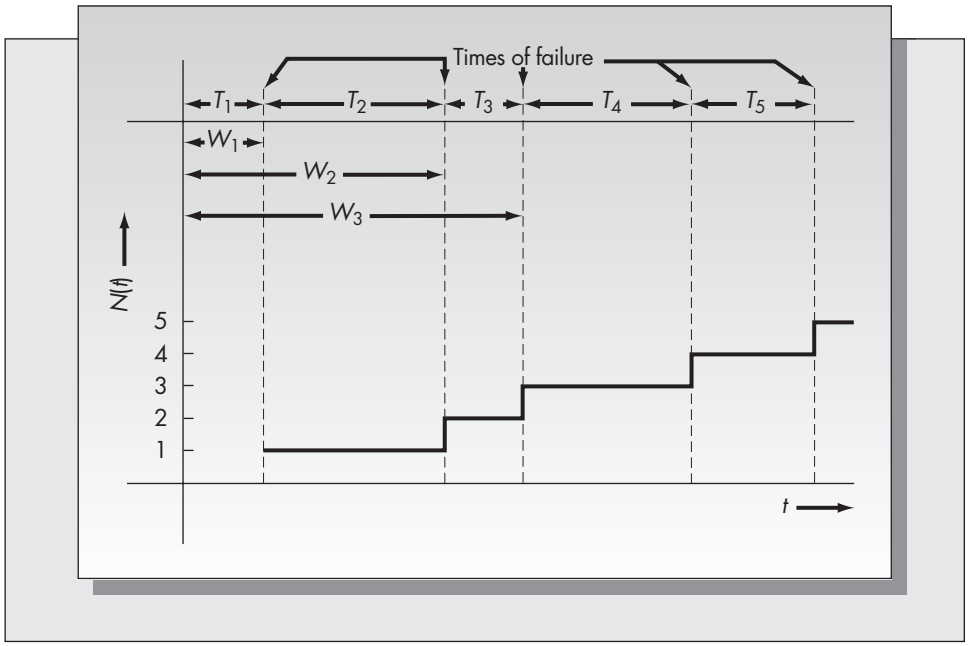


FIGURE 13–5
Realization of a Poisson process

closely related to both the times between failures, $T_1, T_2, \ldots$, and the times of failures $W_1, W_2, \ldots$. When $T_1, T_2, \ldots$ are independent exponentially distributed random variables, $N(t)$ is a *Poisson process.* A realization of a Poisson process is pictured in Figure 13–5.

We will proceed with the analysis of the Poisson process in the following way: We start with the knowledge of the distribution of the interfailure times, $T_1, T_2, \ldots$. From that we can derive the distribution of the times until failure $W_1, W_2, \ldots$. We obtain the distribution of $N(t)$ by using the following equivalence of events:

$$\{N(t) < n\} = \{W_n > t\}.$$

A little reflection should convince you of the truth of this identity. In order for the left-hand side to be true, the number of failures up until time t must be fewer than n. If that is the case, then the time of the nth failure must be after t, which gives the right-hand side. Similarly, if the nth failure occurs after t, then the number of failures up until time t must be fewer than n.

The analysis requires obtaining the cumulative distribution function of the times until failure. Since

$$W_n = T_1 + T_2 + \cdots + T_n,$$

the distribution of W_n can be obtained by forming the n-fold convolution of the exponential distribution, which leads to the *Erlang* distribution:

$$P\{W_n > t\} = \sum_{k=0}^{n-1} \frac{e^{-\lambda t}(\lambda t)^k}{k!}.$$

The Erlang distribution is named for A. K. Erlang in recognition of his pioneering work in the area of queuing. (That W_n has an Erlang distribution is shown in Hillier and Lieberman, 1986, pp. 565–66.) We can now derive the distribution of the number of failures up until time t, $N(t)$.

TABLE 13–1
Summary Results for Poisson Process

Random Variable	Distribution	Parameter(s)	Mean	Variance
Time between failure, T_n	Exponential	λ	$1/\lambda$	$1/\lambda^2$
Time of nth failure, W_n	Erlang	λ, n	n/λ	n/λ^2
Number of failures until time t, $N(t)$	Poisson	λt	λt	λt

We have that

$$\begin{aligned} P\{N(t) = n\} &= P\{N(t) < n + 1\} - P\{N(t) < n\} \\ &= \sum_{k=0}^{n} \frac{e^{-\lambda t}(\lambda t)^k}{k!} - \sum_{k=0}^{n-1} \frac{e^{-\lambda t}(\lambda t)^k}{k!} \\ &= \frac{e^{-\lambda t}(\lambda t)^n}{n!}. \end{aligned}$$

This is exactly the Poisson distribution with parameter λt. The Poisson distribution is a discrete distribution that assumes values only on the nonnegative integers 0, 1, 2, ... It has mean and variance given by λt. The process by which (1) events occur completely at random over time, (2) the interevent times are exponential random variables, (3) the times of events are Erlang random variables, and (4) the number of events up until any time t is a Poisson random variable is called the Poisson process. It is a fundamental stochastic process that crops up in many fields besides reliability. The single parameter λ, the rate at which events occur, defines the process. Table 13–1 summarizes the important points about the Poisson process.

As W_n is the sum of independent identically distributed random variables, when n is reasonably large, the central limit theorem implies that the normal distribution may be used to approximate the Erlang. Also note that the reliability function of the Erlang is of precisely the same form as the cumulative Poisson distribution. Hence, a table of the Poisson distribution (Table A–3 at the back of this book) may be used to obtain exact Erlang probabilities. Furthermore, note that for large values of λt, the normal distribution is an adequate approximation of the Poisson distribution. (However, the normal should not be used to approximate the exponential.)

Example *13.5*

A local military base maintains a variety of different equipment. One of these is a sensitive radar device that signals incursion of enemy planes into American airspace. Breakdowns of the device occur completely at random at an average rate of three per year. The equipment is generally repaired the same day that it fails. Determine the following:

a. The probability that the time between two successive failures is less than one month.

b. The probability that there are exactly five breakdowns in any given year.

c. The probability that there are more than 15 failures in a four-year period.

d. The average time for 100 failures to occur.

e. The probability that the 25th failure occurs after 10 years of operation.

Solution

As with all probability problems, it is a good idea to have a ballpark estimate of the answer before beginning formal calculations. This will serve as a verification of your calculations.

a. Because there are three failures per year on average, there will be on average four months between failures. Hence, the probability that the time between two successive failures is less than a month should be less than .5.

Let T be the time between any two successive failures. Then we know that T has the exponential distribution with parameter $\lambda = 3$ per year. We must be sure to express all units of time in terms of years as we solve this problem. We compute

$$P\{T < 1/12\} = 1 - \exp(-\lambda t) = 1 - \exp(-3/12) = .22.$$

b. Here we count the number of breakdowns in a given year, so that the appropriate distribution is Poisson with parameter $\lambda t = 3 \times 1 = 3$.

$$P\{N(1) = 5\} = \frac{e^{-3}3^5}{5!} = .1008.$$

c. We wish to compute $P\{N(4) > 15\}$, where $N(4)$ has the Poisson distribution with parameter $\lambda t = 3 \times 4 = 12$. From Table A–3, we obtain

$$P\{N(4) > 15\} = P\{N(4) \geq 16\} = .1556.$$

This probability also can be approximated using the normal distribution. In order to use a normal approximation, the standardized variate Z is constructed by subtracting the mean and dividing by the standard deviation. Hence,

$$P\{N(4) > 15\} \approx P\{Z > (15 - 12)/\sqrt{12}\} = P\{Z > 0.8660\} = .1922.$$

This approximation can be improved by using the continuity correction. The continuity correction is appropriate when approximating a discrete random variable with a continuous random variable. Since $\{N(4) > 15\} = \{N(4) \geq 16\}$, we go halfway between and use 15.5. (The continuity correction is discussed in detail in Appendix 12–A of Chapter 12.) Hence, $P\{N(4) > 15\} \approx P\{Z > (15.5 - 12)/\sqrt{12}\} = P\{Z > 1.01\} = .1562$. This is very close to the exact answer, .1556.

d. Here we are interested in $E(W_{100})$. From Table 13–1, $E(W_{100}) = 100/\lambda = 100/3 = 33.33$ years.

e. The time of the 25th failure is the random variable W_{25}. Again we will use the normal approximation, but we do not require the continuity correction because both the Erlang and the normal distributions are continuous. From the table we have that $E(W_{25}) = 25/\lambda = 25/3 = 8.33$, and $\text{Var}(W_{25}) = 25/\lambda^2 = 25/9$. Hence, $\sigma = 5/3 = 1.67$. It follows that

$$P\{W_{25} > 10\} \approx P\left\{Z > \frac{10 - 8.33}{1.67}\right\} = P\{Z > 1\} = .1587.$$

(The continuity correction is not required because W_{25} is continuous.)

Many applications involve monitoring many pieces rather than a single piece of equipment. For example, a repairer is responsible for maintaining all equipment in his or her location, and the airlines must maintain an entire fleet of planes. Failures of collections of equipment are treated below.

Series Systems Subject to Purely Random Failures

Consider a bank of items labeled $1, 2, \ldots, N$ and assume that each of the items fails completely at random, that is, according to an exponential failure law. Furthermore, we assume that the items fail independently. A series system implies that the bank fails when the first item in the bank fails. Let $T_1, T_2, \ldots, T_N$ be the failure times associated with each piece of equipment. Then

$$P\{T_i > t\} = \exp(-\lambda_i t) \qquad \text{for } 1 \leq i \leq N.$$

Define the random variable $T = \min(T_1, T_2, \ldots, T_N)$. Then T represents the time that the next component fails. It is also the time the bank fails.

$$\begin{aligned} P\{T > t\} &= P\{\min(T_1, T_2, \ldots, T_N) > t\} \\ &= P\{T_1 > t, T_2 > t, \ldots, T_N > t\} \end{aligned}$$

(this follows because if the minimum of a group of numbers exceeds a fixed number, then it must be true that all members of the group exceed it as well)

$$= P\{T_1 > t\} \times P\{T_2 > t\} \times \cdots \times P\{T_N > t\}$$

(this follows from the independence of the individual failure times)

$$= e^{-\lambda_1 t} e^{-\lambda_2 t} \cdots e^{-\lambda_N t}$$

$$= \exp\left(- \sum_{i=1}^{N} \lambda_i t \right),$$

which is exactly the exponential failure law with $\lambda = \Sigma \lambda_i$. If we assume that units that fail are repaired quickly, then the number of failures of the bank up until any time t, say $N(t)$, will be a Poisson process with rate λ.

Problems for Section 13.3

11. Automobiles arrive at a tollbooth on a highway completely at random according to a Poisson process with rate $\lambda = 4$ cars per hour. Determine the following:
    a. The probability that the time between any two successive arrivals exceeds 20 minutes.
    b. The probability that exactly four cars arrive in any given hour.
    c. The probability that more than five cars arrive in any given hour.
    d. The probability that more than 10 cars arrive in any given two-hour period.
    e. A person working for the transportation department begins counting arrivals at the tollbooth at 8 A.M. What is the probability that he counts 20 arrivals before 12 noon? (Use a normal approximation for your calculations.)
12. Herman's Hardware uses a neon light in its store window that is left burning continuously. The light has an average lifetime of 1,250 hours and fails completely at random. Lights that burn out are replaced instantly.
    a. On average, how many neon lights does Herman's use in one year?
    b. Suppose that the lights cost $37.50 each and Herman, the store owner, has budgeted $300 annually for them. What is the probability that Herman exceeds his annual budget in any given year?
    c. What is the probability that two bulbs will be used within the same month? (Assume that one month equals 30 days for your calculation.)
13. An electronic module used by the Navy in a sonar device requires replacement on the average once every 16 months and fails according to a Poisson process. Suppose that the Navy places these sonar devices into service on the same date in eight different aircraft carriers. If the modules are replaced immediately after failure and the budget allows for exactly 40 spares over five years, what is the probability that the budget is exceeded? (Hint: Use a normal approximation to the Poisson.)
14. For Problem 13 determine the following:
    a. The probability that a single carrier sent on a six-month mission will not require replacement of the module during that time.
    b. The probability that the time of the fifth failure is more than one year after the devices are placed into service.
    c. The expected time to use all 40 spares.

13.4 FAILURES OF COMPLEX EQUIPMENT

Many applications of reliability theory involve predicting the failure patterns of equipment from knowledge of the failure patterns of the components comprising that equipment. For example, a well-known study published by the U.S. Nuclear Regulatory Commission (1975) claimed that nuclear plants are safe. In the study, reliability theory was used to predict the likelihood of a major problem occurring in a nuclear plant by analyzing the failure rate of the various components comprising the plant. Unfortunately, incorrect assumptions about the independence of these components led to the conclusion that major nuclear accidents were virtually impossible. That is obviously not the case.

Section 13.3 showed that a bank of items connected in series, each of which has the exponential failure law, will also have the exponential failure law. That result is a special case of one that will be derived in this section.

Components in Series

A series system will function only if every component functions. A schematic diagram of a series system appears in Figure 13–6*a*.

Define T_i as the time until failure of the ith component and let T_S be the time of failure of the entire series system. As before, we have that $T_S = \min(T_1, T_2, \ldots, T_N)$. Recall the definition of the reliability function, $R(t) = P\{T > t\}$. We will derive $R_S(t)$, the reliability function of the system in terms of the reliability functions of each of the components, $R_i(t)$. Using essentially the same arguments as in Section 13.3, we have

$$\begin{aligned} R_S(t) &= P\{T_S > t\} = P\{\min(T_1, T_2, \ldots, T_N) > t\} \\ &= P\{T_1 > t, T_2 > t, \ldots, T_N > t\} \\ &= P\{T_1 > t\} \times P\{T_2 > t\} \times \cdots \times P\{T_N > t\} \\ &= R_1(t) \times R_2(t) \times \cdots \times R_N(t). \end{aligned}$$

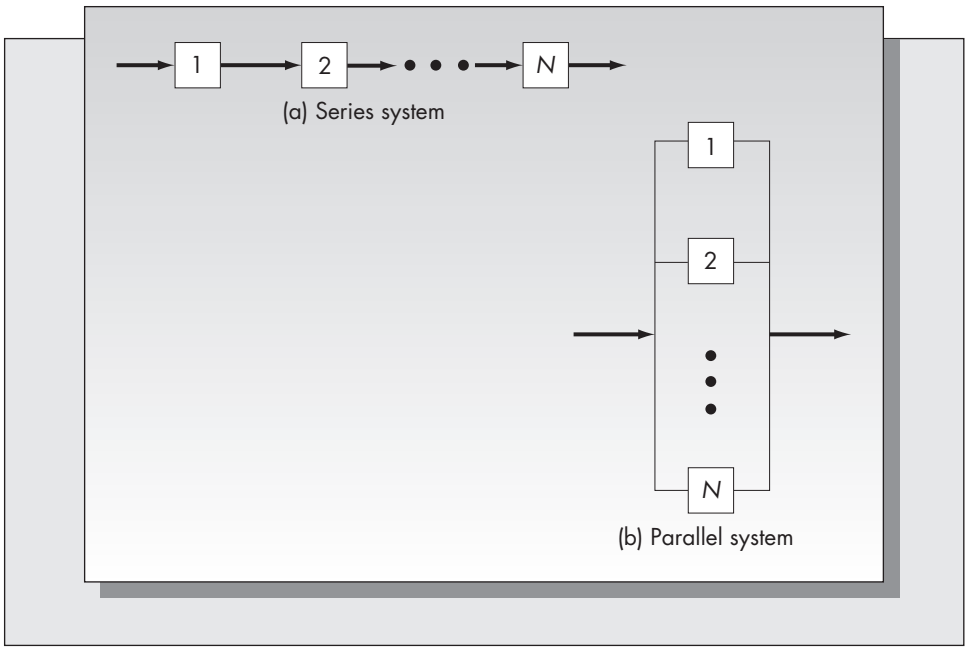


FIGURE 13–6
Systems of components in both series and parallel

For N identical components each having reliability function $R(t)$, this becomes simply $[R(t)]^N$.

To express the cumulative distribution function of the series system in terms of the distribution function of each of the individual components, substitute $F(t) = 1 - R(t)$. For N identical components in series we obtain

$$F_S(t) = 1 - [1 - F(t)]^N.$$

Components in Parallel

Figure 13–6*b* shows a parallel system of components. A parallel system functions if any one of the components functions. Parallel systems occur when redundancy is included to increase reliability. Define T_P as the time of failure of a parallel system of identical components. It is clear that $T_P = \max(T_1, T_2, \ldots, T_N)$. For a parallel system, it is more convenient to determine the distribution function of the time until failure rather than that of the reliability function. We have that

$$\begin{aligned} F_P(t) &= P\{\max(T_1, T_2, \ldots, T_N) \leq t\} \\ &= P\{T_1 \leq t, T_2 \leq t, \ldots, T_N \leq t\} \end{aligned}$$

(because if the largest member of a group is less than a given number, then all the members of the group are also less than that number)

$$= F_1(t) \times F_2(t) \times \cdots \times F_N(t),$$

which reduces to $[F(t)]^N$ in the case of N identical components.

The reliability function of a parallel system with N identical components is $R_P(t) = 1 - [1 - R(t)]^N$.

Expected Value Calculations

A useful result from probability theory that may simplify calculating expected values is the following:

Theorem

If T is a nonnegative random variable with cumulative distribution function $F(t)$ and probability density function $f(t)$, then one can compute the expected value in two ways:

$$E(T) = \int_0^\infty t f(t)dt = \int_0^\infty [1 - F(t)]dt.$$

The second equation can streamline calculations of the expected time until failure. We will use it to compute the expected time until failure of a parallel system of N identical components, each of which has the exponential failure law with parameter λ.

We have that

$$E(T_P) = \int_0^\infty [1 - (1 - e^{-\lambda t})^N]dt.$$

In order to perform the integration, we make the change of variable $v = 1 - e^{-\lambda t}$, which gives $dv = \lambda e^{-\lambda t}\, dt$, or

$$dt = \frac{1}{\lambda}\frac{1}{e^{-\lambda t}}dv = \frac{1}{\lambda}\frac{1}{1 - v}dv.$$

Hence,

$$E(T_P) = \frac{1}{\lambda}\int_0^1 \frac{1 - v^N}{1 - v}dv.$$

The expression in the integrand is just the finite geometric series $1 + v + v^2 + \cdots + v^{N-1}$. Hence, it follows that

$$E(T_P) = \frac{1}{\lambda}\int_0^1 (1 + v + v^2 + \cdots + v^{N-1})\, dv$$
$$= \frac{1}{\lambda}(1 + \tfrac{1}{2} + \tfrac{1}{3} + \cdots + 1/n).$$

This implies that for a system with k components in parallel, the expected lifetime of the system is increased by $1/(k+1)\lambda$ by adding one additional component.

Example *13.6* Wizard, a popular brand of electric garage door opener, includes two 40-watt bulbs that go on when the garage door is opened. A bulb will generally last about one year in normal operation. Three neighbors, James, Smith, and Walker, each has a Wizard opener in their respective garages. Each time a bulb burns out, James replaces both bulbs. Smith, on the other hand, replaces only the bulb that has burned out, and Walker replaces both bulbs only after both have burned out. Assume that light bulbs fail according to an exponential law.

a. Over a 10-year period, how many bulbs, on average, will each neighbor require?

b. What percentage of the time will Walker have only one bulb burning?

c. Is there any advantage of James's strategy over Smith's?

Solution

a. Both James and Smith treat the system as a pure series system of two components: the system fails when one of the bulbs fails. If each bulb has failure rate $\lambda_i = 1$, then the system has failure rate $\lambda = \lambda_1 + \lambda_2 = 2$. Because James replaces two bulbs at each occurrence of a failure, he uses an average of 4 bulbs per year, or 40 bulbs in 10 years. Smith, on the other hand, only requires one bulb each time the system fails, so he uses an average of 2 bulbs per year, or 20 bulbs during 10 years.

Walker's policy of replacing both bulbs only after both have failed is equivalent to a pure parallel system of two components. The expected lifetime of a parallel system of two components with failure rate 1 is $1 + \frac{1}{2} = 1.5$ years. Hence, every 1.5 years he requires two bulbs, thus resulting in an average of 6.67 replacements over the 10 years, which amounts to a total of 13.33 bulbs.

b. One might think that half the time he would have one bulb operating and half the time he would have two bulbs operating. However, this turns out not to be the case. Because the failure rate of the series system is $\lambda = 2$, the first bulb will fail on average after six months. As the parallel system has, on average, a 1.5-year lifetime, the remaining bulb will last an average of one year. Hence, he will be operating his garage door an average of 66.67 percent of the time with only one bulb.

c. Because the failure law is exponential, there is absolutely no advantage of James's strategy over Smith's. They will both have to make replacements equally often (twice a year) but James will use twice as many light bulbs. However, if the failure law is *not* exponential, it is possible that James's method will result in fewer occasions in which a replacement must be made.

This problem raises an interesting point. In installations in which many lights are used, such as a Las Vegas hotel sign, it is a common strategy to periodically replace all bulbs at once. However, as we saw in the problem, if the bulbs follow an exponential failure law, this strategy will result in no fewer unplanned replacements, but will lead to using far more bulbs.

K Out of *N* Systems

Assume that a system consists of N components. A K out of N system is one in which the system functions only if at least K of the components function, where $1 \le K \le N$. A typical example of a K out of N system is a four-engine airplane that can fly as long as at least two of its engines are operating.

To analyze a K out of N system, we use a binomial framework. Think of each component as a separate Bernoulli trial: Identify a success with a functioning component and a failure with a nonfunctioning component. We assume that all components are identical, so that each has the same reliability function $R(t)$ and the same distribution function $F(t)$. Fix a point in time, t. Let $p = P\{\text{a component functions at time } t\} = R(t)$. The probability that the system functions at time t, $R_k(t)$, is the probability that there are at least K successes in N trials of a binomial experiment with $p = P\{\text{success}\}$ at each trial.

Hence,

$$R_K(t) = \sum_{j=K}^{N} \binom{N}{j} p^j (1-p)^{N-j}$$

$$= \sum_{j=K}^{N} \binom{N}{j} R(t)^j F(t)^{N-j}.$$

Note that both series and parallel systems are special cases of K out of N systems. A series system is an N out of N system, and a parallel system is a 1 out of N system. A series system of N identical components will always have lower reliability than a single component, whereas a parallel system of N identical components will always have higher reliability than a single component.

However, whether a K out of N system is more reliable than a single component depends upon the reliability of the individual components. Figure 13–7 graphs the reliability of 2 out of 4 and 3 out of 4 systems as a function of the reliability of each component. The 45-degree line represents the reliability of a single component. The reliability curve for the 2 out of 4 system crosses the 45-degree line at about $p = .23$ and for the 3 out of 4 system at about $p = .77$. This means that a 2 out of 4 system is preferred to a single component only if the reliability of the components comprising the system exceeds .23 and a 3 out of 4 system is preferred to a single component only if the component reliability exceeds .77.

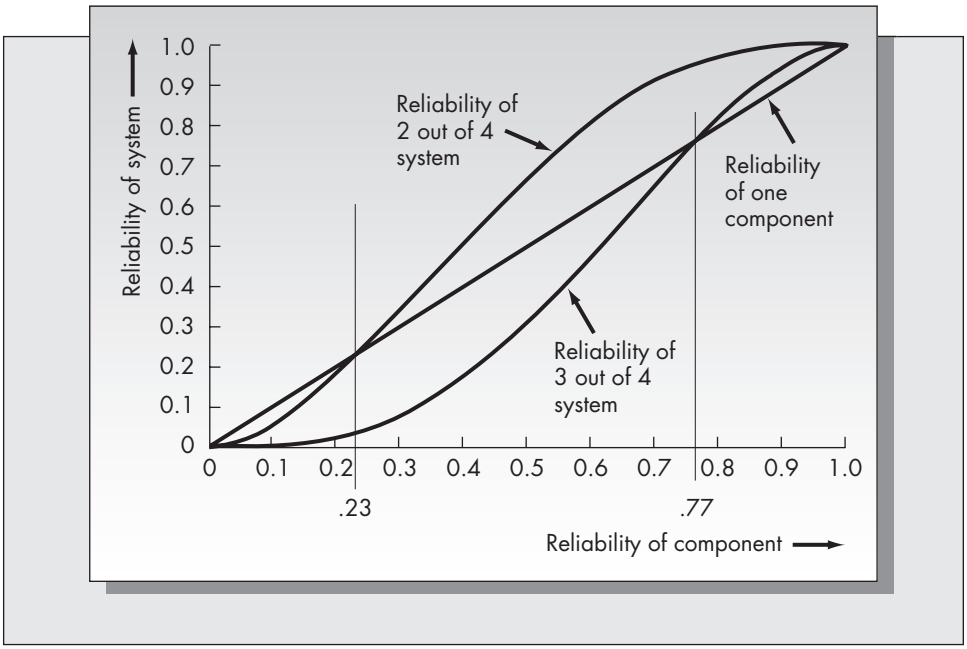
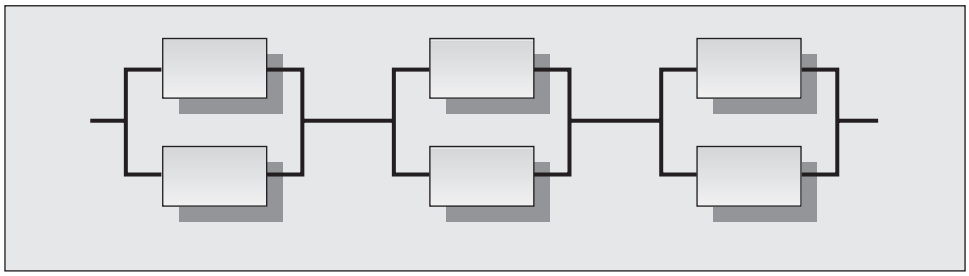


FIGURE 13–7
Reliability of 2 out of 4 and 3 out of 4 systems

FIGURE 13–8
System of components (for Problem 19)



Problems for Section 13.4

15. A subassembly in an industrial robot consists of 12 components in series, each of which fails completely at random at a rate of once every 50 years.
    a. What is the mean time between failures for this part of the robot?
    b. What is the probability that this subassembly does not fail within eight years of operation?
16. Show that the time until failure of a series system of n identical components, each of which has the Weibull lifetime distribution, also has a distribution of the Weibull type.
17. Consider the following three systems: (*a*) a single component with failure rate one per year, (*b*) two components in series, each of which has failure rate one every two years, and (*c*) two components in parallel, each of which has failure rate two per year. Compare the reliability of these three systems assuming an exponential failure law.
18. A design engineer is considering the number of levels of redundancy to build into a particular circuit. The circuit will be part of a sensitive piece of equipment with cost estimated at \$500 per failure. Each additional level of redundancy costs \$100. If each component fails at random at a rate of one failure every five years, what level of redundancy most closely equates the cost of the design with the expected failure cost of the equipment over its 10-year life cycle?
19. Consider the system of six identical components pictured in Figure 13–8. If each component has constant failure rate λ, derive the distribution function of the time until failure of the system.
20. An aircraft engine fails with probability p. Assume that for an aircraft to successfully complete a flight, at least half the engines must operate. Show that for $0 < p < \frac{1}{3}$, a four-engine plane is preferred to a two-engine plane and for $\frac{1}{3} < p < 1$, a two-engine plane is preferred to a four-engine plane.

13.5 INTRODUCTION TO MAINTENANCE MODELS

The maintenance of complex equipment often can account for a large portion of the costs associated with that equipment. It has been estimated, for example, that the maintenance costs in the military comprise almost one-third of all the operating costs incurred. Clearly, the issues of reliability and maintenance are closely connected.

This section will introduce some standard maintenance terminology:

1. MTBF = Mean time between failures. This corresponds to the expected time between failures in our previous notation and equals $1/\lambda$.
2. MTTR = Mean time to repair. This is the expected value of the repair time R.

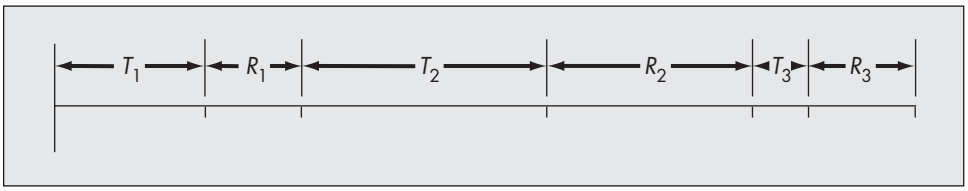


FIGURE 13–9
Realization of failure and repair times

3. Availability = Average fraction of time the equipment operates. It is given by the formula

$$\text{Availability} = \frac{E(T_i)}{E(T_i) + E(R_i)} = \frac{\text{MTBF}}{\text{MTBF} + \text{MTTR}}.$$

We may think of a single piece of equipment that has successive failure times T_1, T_2, . . . and successive repair times R_1, R_2, . . . A schematic diagram of such a system appears in Figure 13–9.

Example *13.7* A copier machine has a mean time between failures of 400 operating hours. Repairs typically require an average of 10 hours from the time that the repair call is received until service is completed. Determine the availability of this copier.

Solution The availability is 400/(400 + 10) = 400/410 = .9756.

Define a repair cycle as the time between two successive repairs. Often we must determine the distribution of a repair cycle rather than just its expectation. A single repair cycle is the sum of the failure time T_i and the repair time R_i. The exact distribution of the sum of two random variables is the *convolution* of the individual distributions. (See DeGroot, 1986, for example.)

If the interfailure and the repair times can be reasonably approximated by the normal distribution, then the sum of the two also will be approximately normal.

Example *13.8* Suppose that the time between failures, T_i, is approximately normally distributed with mean 400 hours and variance 10,000. The repair time of the equipment is also approximately normally distributed with mean 10 hours and variance 11.6. Find the probability that there are more than six repair cycles within a one-year period. Assume that one year corresponds to 2,000 hours of operation.

Solution We have that

$$E(T_i + R_i) = 400 + 10 = 410,$$
$$\text{Var}(T_i + R_i) = 10{,}000 + 11.6 = 10{,}011.6.$$

It follows that

$$E\left[\sum_{i=1}^{6}(T_i + R_i)\right] = 6 \times 410 = 2{,}460,$$

$$\text{Var}\left[\sum_{i=1}^{6}(T_i + R_i)\right] = 6 \times 10{,}011.6 = 60{,}069.6,$$

$$P\left\{\sum_{i=1}^{6}(T_i + R_i) \le 2{,}000\right\} = P\left\{Z \le \frac{2{,}000 - 2{,}460}{\sqrt{60{,}069.6}}\right\}$$
$$= P\{Z \le -1.88\} = .03.$$

13.6 DETERMINISTIC AGE REPLACEMENT STRATEGIES

For operating equipment that does not exhibit an exponential failure law, there are often advantages to replacing a piece of equipment *before* it fails. This is true when the cost of repair is much higher if the equipment fails while it is operating. In some cases, such as in military operations, an equipment failure might be impossible to correct and could result in the loss of life.

This section will consider age replacement models that do not explicitly account for the uncertainty of the failure process. Rather, the aging mechanism is subsumed in the cost structure; in particular, it is assumed that the cost of maintaining the equipment increases as the equipment ages. Section 13.7 will consider models of planned replacement that explicitly include the uncertainty of the failure process.

The models we consider are appropriate for both continuously operating equipment, such as radar or power generating units, and intermittently operating equipment, such as automobiles. In the latter case, we would keep track of operating time rather than clock time.

Let us assume that the replacement cost of the item is K. Also assume that the instantaneous cost rate of operating an item of age u is $C(u)$. We will consider various forms for $C(u)$ but assume initially that $C(u) = au$.

Based on the values of the various costs, there will be some optimal point at which to replace the item in order to minimize the total cost per unit time. The total cost function may be thought of as the sum of two components: maintenance and replacement. The marginal cost of maintenance increases over time, and the marginal cost of replacement decreases over time. The optimal replacement age minimizes the average cost function. In the cases we consider, the average cost function is convex, thus making the optimal solution easy to find.

The Optimal Policy in the Basic Case

We make the following assumptions:

1. The equipment used is operating continuously.
2. We ignore downtime for repair and maintenance.
3. The planning horizon is infinite.
4. Every new piece of equipment has identical characteristics.
5. Only maintenance and replacement costs are considered.
6. The objective is to minimize the long-run costs of replacement and maintenance.
7. The cost rate of maintaining an item of age u is au, and the replacement cost of the item is K. There is no salvage value.

The decision variable is the amount of time that elapses from the point that a piece of equipment is purchased until it is replaced with a new item. Figure 13–10 shows successive replacement cycles.

The object of the analysis is to determine the value of t that minimizes the total cost of maintenance and replacement over an infinite horizon. A replacement cycle is the time between successive replacements. Because all replacement cycles are identical, we may restrict attention only to the costs incurred in a single cycle.

$$\text{Total replacement cost per cycle} = K,$$

$$\text{Total maintenance costs per cycle} = \int_0^t C(u)\,du = \int_0^t au\,du = \frac{at^2}{2}.$$

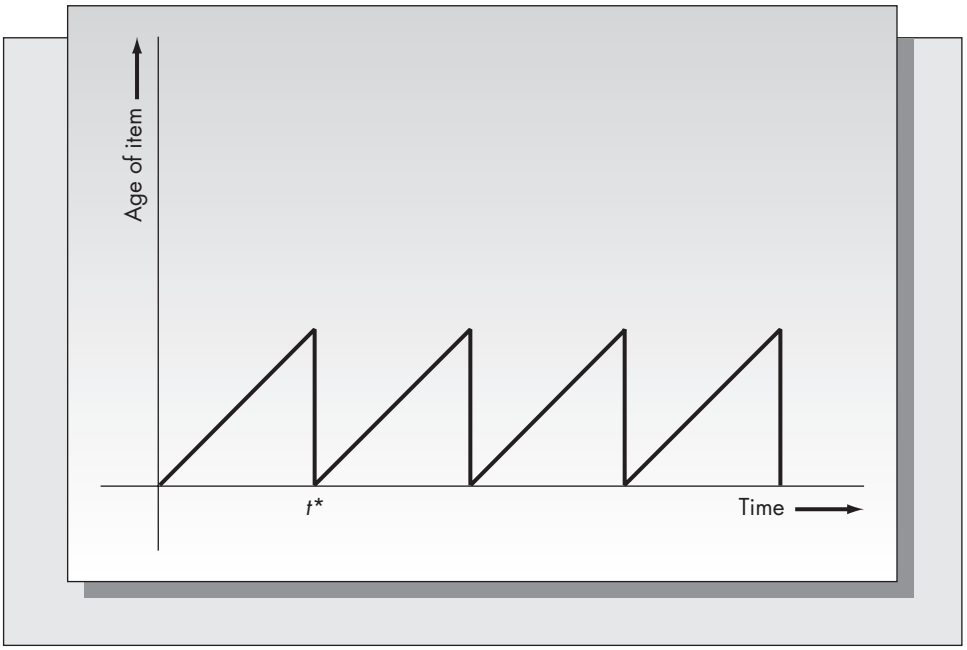


FIGURE 13–10
Optimal age replacement strategy

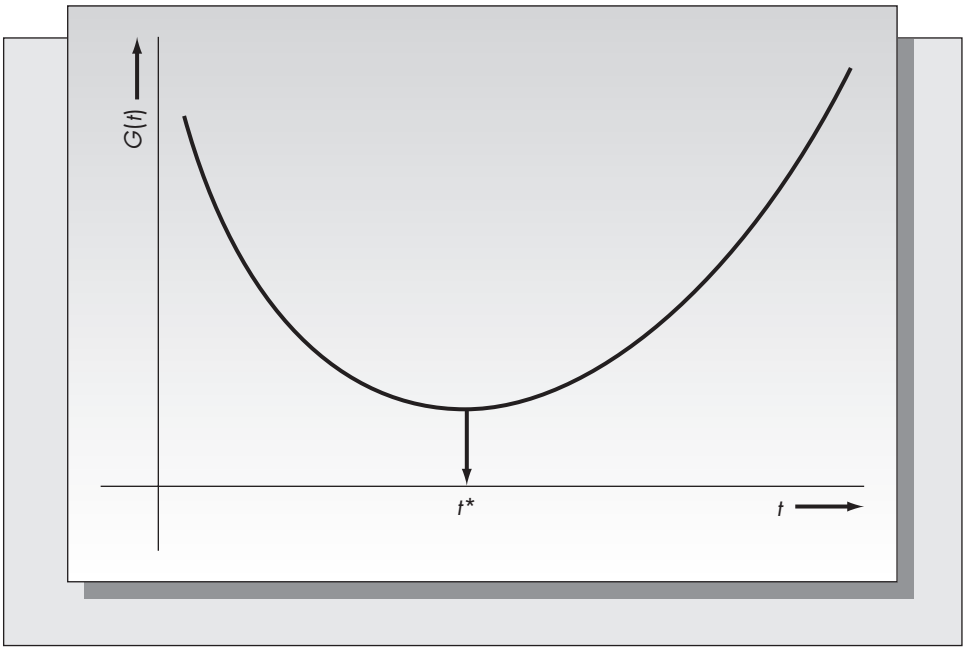


FIGURE 13–11
Computation of the optimal replacement age

The average cost per unit time is just the total cycle cost divided by the length of the cycle. Let $G(t)$ be the average cost per unit time if the replacement time is t. Then

$$G(t) = \frac{1}{t}\left(K + \frac{at^2}{2}\right) = \frac{K}{t} + \frac{at}{2}.$$

As $G''(t) = K/t^3 > 0$, it follows that $G(t)$ is a convex function of the single variable t. The function $G(t)$ is pictured in Figure 13–11. The goal is to find the optimal value of t,

say t^*, that minimizes $G(t)$. Since $G(t)$ is convex, it follows that the optimal solution satisfies

$$G'(t) = \frac{-K}{t^2} + \frac{a}{2} = 0,$$

which results in

$$t^* = \sqrt{\frac{2k}{a}}.$$

Example *13.9* We will use the simple version of the age replacement model to estimate the number of years that one should keep a car. Although the model does not exactly describe the car replacement problem, we can use it as an approximation. Let us assume that the maintenance cost rate of a car u years old is $400u$ dollars. This means that the maintenance cost during the first year is \$200, during the second year is \$600, during the third year is \$1,000, and so on. (These numbers are obtained by computing $at^2/2$ for $t = 1, 2$, and 3 and subtracting the costs of the previous years.) This is probably rising a bit faster than our actual maintenance costs would. Assume that a new car costs \$10,000. According to the formula, the optimal number of years that we should hold the car is

$$t^* = \sqrt{\frac{2K}{a}} = \sqrt{\frac{(2)(10{,}000)}{400}} = 7.07 \text{ years.}$$

A General Age Replacement Model

When we include a salvage value and allow for more general maintenance cost functions, the model becomes considerably more complex. As before, suppose that $C(u)$ is an arbitrary function representing the maintenance cost rate of an item of age u. Define $S(u)$ as the salvage value of an item of age u. Then the total cost incurred in the cycle is

$$K + \int_0^t C(u)\, du - S(t).$$

The average cost per unit time is

$$G(t) = \frac{K}{t} + \frac{1}{t}\int_0^t C(u)\, du - \frac{S(t)}{t}.$$

The optimal value of t, t^*, is the solution to

$$G'(t) = \frac{-K}{t^2} + \frac{H(t)}{t^2} + \frac{C(t)}{t} + \frac{S(t)}{t^2} - \frac{S'(t)}{t} = 0,$$

or

$$tC(t) + S(t) = K + H(t) + tS'(t),$$

where for convenience we let

$$H(t) = \int_0^t C(u)\, du.$$

Finding t^* can be very difficult. [For example, try to obtain a solution assuming $C(u) = au$ and $S(u) = K - bu$.] For many real problems the exponential distribution provides an accurate description of the increase in maintenance costs and the decrease in resale value of operating equipment. If we let

$$C(u) = ae^{bu}, \qquad \text{where } a, b > 0,$$

and

$$S(u) = ce^{-du}, \qquad \text{where } c, d > 0,$$

then the optimal value of t satisfies

$$tae^{bt} + ce^{-dt} = K + \int_0^t ae^{bu}\,du + t\frac{d}{dt}(ce^{-dt}).$$

It is easy to show that

$$H(t) = \int_0^t ae^{bu}\,du = \frac{a}{b}(e^{bt} - 1)$$

and

$$\frac{d}{dt}(ce^{-dt}) = -cde^{-dt},$$

so the equation defining an optimal solution is

$$tae^{bt} + ce^{-dt} = K + \frac{a}{b}(e^{bt} - 1) - tcde^{-dt}.$$

Rearranging terms gives

$$ae^{bt}\left(t - \frac{1}{b}\right) + ce^{-dt}(1 + dt) + \frac{a}{b} = K.$$

The goal is to find the value of t that makes the left-hand side of the equation as close to K, the replacement cost, as possible. This is a difficult equation to solve for t because it involves both exponentials and constants.[3] Spreadsheets provide a convenient method of obtaining a solution. One simply computes the left-hand side of the equation for various values of t and graphically determines the point at which this function crosses the value of K.

Example *13.10* Consider again finding the optimal time to replace an automobile. Exponential functions are more realistic than linear functions and should give an accurate estimate of the true optimal time for replacement. As in Example 13.9, assume that the replacement cost of the automobile is \$10,000. Furthermore, assume that the car loses 15 percent of its value each year. This is probably a reasonable estimate of the decline in the resale value of most new cars. This means that the car is worth (0.85)(10,000) = \$8,500 after one year, (0.85)(0.85)(10,000) = \$7,225 after two years, and so on.

We wish to determine c and d so that $S(t) = ce^{-dt}$ agrees with these values. Because the salvage value at time $t = 0$ is exactly the replacement cost, we have $S(0) = 10{,}000$. Substituting, we obtain

$$S(0) = ce^{-d(0)} = c = 10{,}000.$$

The value of the car after one year is (0.85)(10,000), which corresponds to $S(1)$. Hence,

$$S(1) = (0.85)(10{,}000) = ce^{-d(1)} = ce^{-d}.$$

Since $c = 10{,}000$, we obtain

$$e^{-d} = 0.85,$$
$$d = -\ln(0.85) = 0.1625.$$

[3] It is called a transcendental equation.

Hence, it follows that

$$S(t) = 10{,}000e^{-0.1625t}.$$

Now consider the maintenance costs. Assume, as in Example 13.9, that maintenance costs for the first year of operation amount to \$200. This is equivalent to

$$H(1) = 200$$

or

$$(a/b)(e^b - 1) = 200.$$

Furthermore, suppose that the maintenance costs increase at a rate of 40 percent per year. This means that

$$\frac{C(t)}{C(t-1)} = 1.4.$$

Substituting for $C(t)$ gives

$$\frac{ae^{bt}}{ae^{b(t-1)}} = e^b = 1.40,$$

or $b = \ln(1.4) = 0.3365$. It follows that

$$a = \frac{(200)(b)}{e^b - 1} = \frac{(200)(0.3365)}{0.4} = 168.25.$$

Combining these results, it follows that the optimal time to replace the automobile, t^*, is the value of t that solves

$$168.25e^{0.3365t}(t - 2.972) + 10{,}000e^{-0.1625t}(1 + 0.1625t) + 500 = 10{,}000.$$

One method of estimating the solution to this equation is to graph the function represented by the left-hand side of the equation. We have done so in Figure 13–12. The minimum-cost

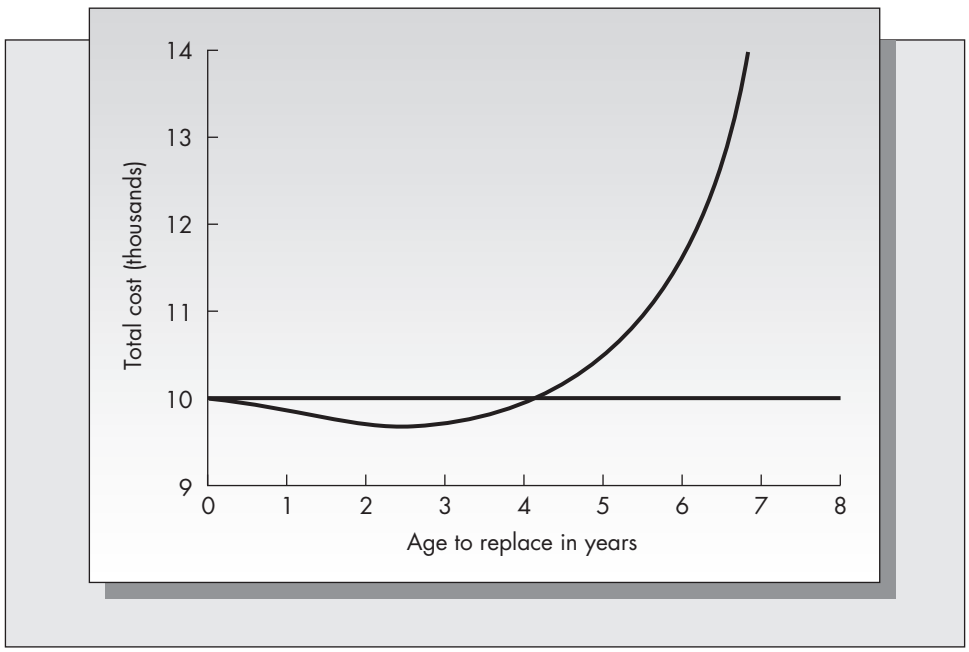


FIGURE 13–12
Optimal number of years to replace auto ($M = .4$)

replacement age is about 4.5 years. That means that for an automobile costing $10,000 that declines in value at the rate of 15 percent per year, and for which the maintenance cost is $200 the first year and increases at the rate of 40 percent per year, the optimal strategy, which minimizes average costs of replenishment less salvage plus maintenance, is to replace the car about once every four and a half years.

To see the effect of the maintenance cost, we have re-solved the example with a value of 20 percent rather than 40 percent for the rate of increase of maintenance costs per year. This is probably more accurate for most cars. The solution is represented graphically in Figure 13–13. Here, the optimal replacement time is approximately nine years.

Let M represent the rate at which maintenance costs increase each year expressed as a decimal ($M = .20$ in Figure 13–13). Let I_0 be the maintenance cost in the first year, and D the yearly rate of depreciation, also expressed as a fraction ($D = 0.15$) in the example). Then it can be shown that

$$b = \ln(1 + M),$$
$$a = I_0 b/M,$$
$$c = K,$$
$$d = \ln(1 - D).$$

The models presented in this section do not consider the effects of inflation. If inflation effects were included, the resulting solutions would not change appreciably, because the inflation is applied to both the replacement and the maintenance costs. The method is to compute the present value of all future discounted costs and determine the replacement strategy that minimizes the resulting function. We will not consider discounted cash flows in this section.

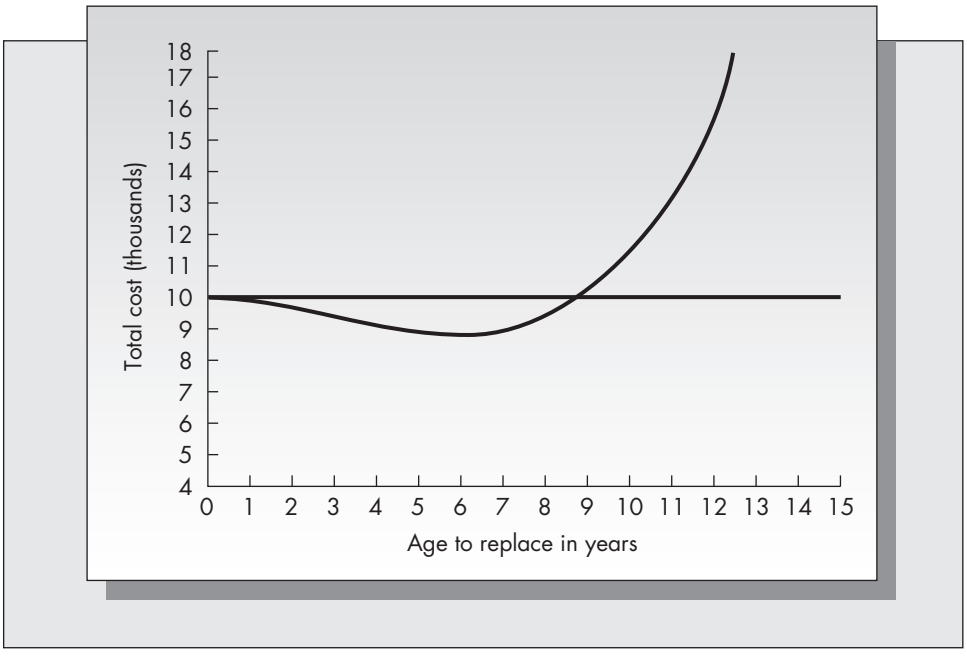


FIGURE 13–13
Optimal number of years to replace auto ($M = .2$)

Problems for Section 13.6

21. For the basic age replacement model, consider a piece of equipment that costs \$18,000 to replace. The *total* maintenance costs for five years of operation are estimated to be \$2,400. Assuming a linear maintenance cost rate, find the value of a and the optimal age at which the equipment should be replaced.
22. For the basic age replacement model, derive the optimal replacement age when the maintenance cost rate $C(u)$ has the form $C(u) = a\sqrt{u}$ for some constant $a > 0$.
23. Suppose for the simple age replacement model that the maintenance cost is the same every year [that is, $C(u) = a$ for all $u \geq 0$]. What is the optimal replacement age? Why is this so?
24. The army is attempting to determine the optimal replacement age for a piece of field equipment. The equipment costs \$280,000 to replace. The manufacturer will supply a rebate toward the next purchase that declines at a rate of 20 percent per year. Maintenance costs for the first year are estimated to be \$1,000, and they increase roughly at the rate of 18 percent per year. Estimate the number of years that the army should hold the equipment before making a replacement.
25. Try to determine the optimal replacement age when $C(u) = au$ and $S(u) = K - bu$. What difficulty do you encounter?

13.7 PLANNED REPLACEMENT UNDER UNCERTAINTY

The purpose of preventive maintenance is to decrease the likelihood that an item will require replacement because of failure. At the heart of such a policy is the assumption that it costs more to make a repair or replacement at the time of failure than at some predetermined time. For example, if failure means that a production line must be stopped to determine the cause of the failure and repair the problem, whereas preventive maintenance can be accomplished at a convenient time when the system is not operating, then the cost of planned replacements is less than the cost of unplanned replacements.

Because of the memoryless property of the exponential distribution, if an item or group of items obeys an exponential failure law, then there is no advantage to replacing prior to failure. In the exponential case, the likelihood that failure will occur in a time Δt is the same just after a planned replacement as it is for an item that has been operating for an arbitrary amount of time. Hence, planned replacement strategies can have value only if the items exhibit aging, that is, have an increasing failure rate function.

Planned Replacement for a Single Item

Consider a single piece of continuously operating equipment whose lifetime is a random variable T with known cumulative distribution function $F(t)$. We assume that T is a continuous random variable. Suppose that it costs c_1 to replace the item when it fails and $c_2 < c_1$ to replace the item prior to failure. We assume that planned replacements are made exactly t units of time after the last replacement. The goal is to find the optimal value of t to minimize the average cost per unit time of both planned and unplanned replacements.

A cycle is the time between successive replacements. Because the process "restarts" itself after each replacement, irrespective of whether the replacement was planned or

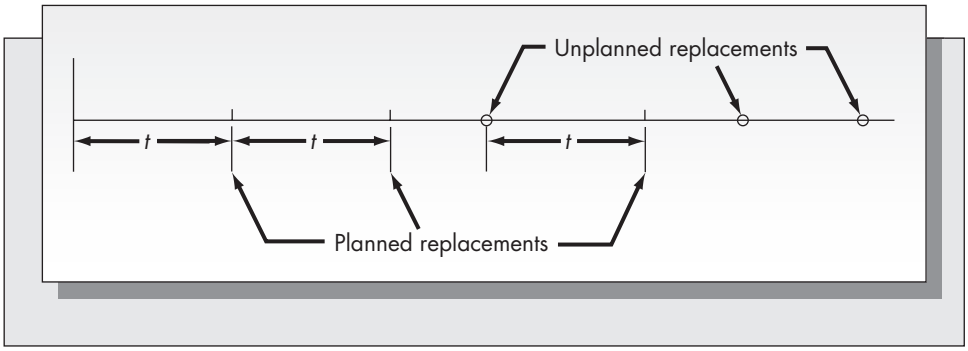


FIGURE 13–14
Successive cycles for planned replacement of a single item

unplanned, we may use the renewal method to obtain an expression for the expected cost per unit time. That is,

$$E(\text{cost per unit time}) = \frac{E(\text{cost per cycle})}{E(\text{length of a cycle})}.$$

Renewal arguments have been used before in the text (Section 11.6). This approach also was used in a variety of other places, including Section 13.6 on age replacement and in much of Chapters 4 and 5 on inventory modeling. Successive replacement cycles are pictured in Figure 13–14.

We have that

$$\begin{aligned} E(\text{cost per cycle}) = {} & c_1 P\{\text{replacement is result of failure}\} \\ & + c_2 P\{\text{replacement is planned}\}. \end{aligned}$$

Notice that $P\{\text{replacement is result of failure}\} = P\{T \leq t\} = F(t)$, and $P\{\text{replacement is planned}\} = P\{T > t\} = 1 - F(t)$, where T is the lifetime of the item placed into service at the end of the previous cycle. It follows that

$$E(\text{cost per cycle}) = c_1 F(t) + c_2[1 - F(t)].$$

Let T be the time of failure of the item placed into service at the end of the previous cycle. Then, clearly, the next replacement will occur at $\min(T, t)$. Hence,

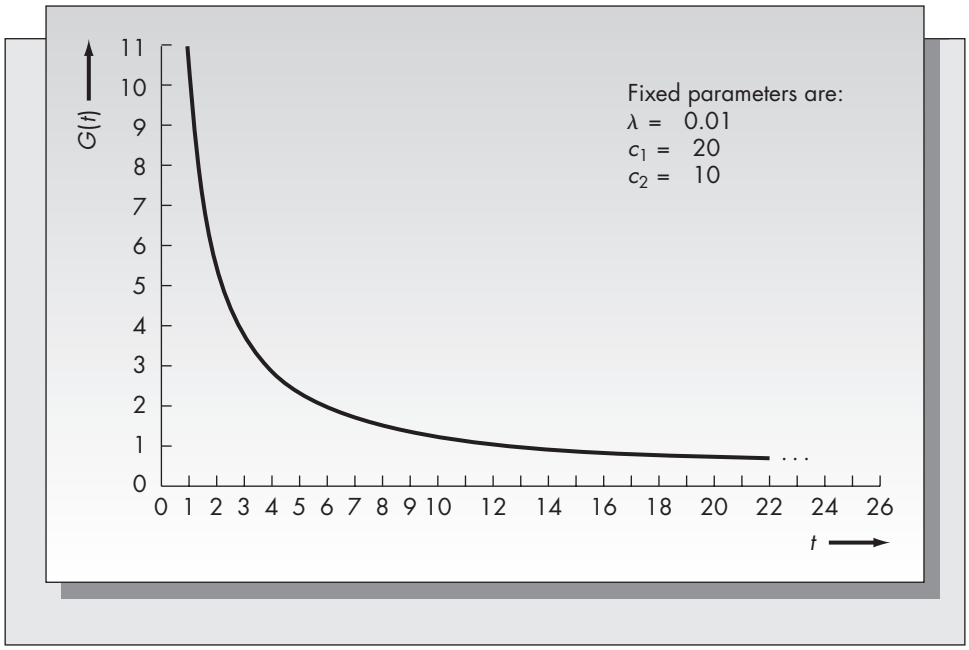
$$\begin{aligned} E(\text{length of cycle}) = E[\min(T, t)] &= \int_0^\infty \min(x, t) f(x)\, dx \\ &= \int_0^t x f(x)\, dx + t \int_0^\infty f(x)\, dx \\ &= \int_0^t x f(x)\, dx + t[1 - F(t)]. \end{aligned}$$

It follows that the expected cost per unit time, say $G(t)$, is given by

$$G(t) = \frac{c_1 F(t) + c_2[1 - F(t)]}{\displaystyle\int_0^t x f(x)\, dx + t[1 - F(t)]}.$$

The goal is to find t to minimize $G(t)$. The optimization may be cumbersome depending upon the form of the lifetime distribution $F(t)$.

FIGURE 13–15
The function $G(t)$



We will now show that there is no advantage to planned replacement when the lifetime distribution is exponential. Suppose that $F(t) = 1 - e^{-\lambda t}$. Then the expected length of each cycle is

$$\int_0^t xe^{-\lambda x}\,dx + te^{-\lambda t} = \frac{1}{\lambda}[1 - e^{-\lambda t}(1 + \lambda t)] + te^{-\lambda t}$$

$$= \frac{1}{\lambda}(-e^{-\lambda t}).$$

(The expression for $\int_0^t xe^{-\lambda x}\,dx$ can be obtained by integration by parts or can be found in a table of integrals.) It follows that

$$G(t) = \frac{c_1(1 - e^{-\lambda t}) + c_2\,e^{-\lambda t}}{\frac{1}{\lambda}(1 - e^{-\lambda t})} = \frac{c_1 - (c_1 - c_2)e^{-\lambda t}}{\frac{1}{\lambda} - \frac{1}{\lambda}e^{-\lambda t}}.$$

As $t \to \infty$, the term $e^{-\lambda t} \to 0$, so $G(\infty) = \lambda c_1$. Furthermore, substituting $t = 0$ results in $G(0) = \infty$. It can be shown by either calculus or direct computation that the function $G(t)$ is monotonically decreasing; a typical case is pictured in Figure 13–15. Hence, the optimal solution is $t = \infty$, which means that a planned replacement should never be made.

We have shown that if the lifetime distribution is exponential (constant failure rate), then there is no economy in replacing an item prior to the time it fails. This also holds if the failure rate is decreasing.

Example *13.11* A large trucking company, Harley Brown, Inc., maintains detailed records on the mortality of the tires used on company-owned trucks. A statistical analysis of the data on tire failure shows that the lifetime of a tire as measured in thousands of miles of use is closely approximated by the Weibull probability law with parameters $a = 0.00235$ and $\beta = 2.3$. The company

TABLE 13–2
Failure Probabilities for Example 13.11

Lifetime (thousands of miles)	Probability of Failure	Lifetime (thousands of miles)	Probability of Failure
1	.0023	14	.0625
2	.0092	15	.0580
3	.0175	16	.0527
4	.0264	17	.0469
5	.0354	18	.0408
6	.0440	19	.0348
7	.0517	20	.0291
8	.0582	21	.0238
9	.0632	22	.0191
10	.0664	23	.0150
11	.0679	24	.0116
12	.0677	25	.0087
13	.0658		

estimates a cost of $450 if a tire fails during use. This is a result of the lost time and the potential liability of an accident. Tires replaced before failure cost $220 each. The company would like to find the optimal timing of tire replacement.

The difficulty with calculating the optimal solution for this problem is determining an expression for the term $\int_0^t x f(x)\,dx$ that appears in the denominator of $G(t)$.

To circumvent the problem of finding an analytical expression for this integral, we will obtain a discrete approximation to the failure law and perform the calculation as if the lifetime distribution were discrete rather than continuous. The probability of failure within the first 1,000 miles is

$$P\{T \leq 1\} = 1 - F(1) = .0023.$$

The probability of failure after 1,000 miles but before 2,000 miles of wear is

$$P\{T \leq 2\} - P\{T \leq 1\} = R(1) - R(2) = .0092.$$

The remainder of the failure probabilities are computed in a similar fashion and appear in Table 13–2.

Using these discrete probabilities we may now compute $G(t)$ directly. The partial expectation term

$$\int_0^t xf(x)\,dx \approx \sum_{k=1}^{t} kp_k,$$

where the probabilities p_k are as given in Table 13–2. Note that this approximation assumes that all failures occur at multiples of 1,000 miles.

The remainder of the terms comprising $G(t)$ can be obtained directly from the Weibull reliability function

$$R(t) = e^{-\alpha t^{\beta}}.$$

Hence, the approximate form of $G(t)$ may be written

$$G(t) = \frac{c_1 - (c_1 - c_2)R(t)}{\sum_{k=1}^{t} kp_k + tR(t)}$$

$G(t)$ appears in Table 13–3 for the parameter values $\alpha = 0.00235$, $\beta = 2.3$, $c_1 = 450$, and $c_2 = \$220$. The function appears to be convex and is minimized at $t = 13$. Hence, the optimal policy calls for replacing the tires after about 13,000 miles of wear. The value of the objective function at the optimal solution is $G(13) = 33.21$. This means that at the optimal solution, the replacement cost is $33.21 per thousand miles of use per tire.

TABLE 13–3
The Function $G(t)$ for Example 13.11

t	$G(t)$	t	$G(t)$
1	220.54	14	33.24
2	111.45	15	33.35
3	75.91	16	33.53
4	58.82	17	33.72
5	49.14	18	33.93
6	43.20	19	34.13
7	39.38	20	34.32
8	36.89	21	34.48
9	35.27	22	34.62
10	34.25	23	34.74
11	33.64	24	34.84
12	33.33	25	34.91
13	33.21		

Block Replacement for a Group of Items

In certain circumstances it is more economical to replace groups of items at the same time rather than one by one. Example 13.11 showed that the optimal policy was to replace a truck tire after about 13,000 miles of use. Depending upon the time and the expense involved in changing truck tires, it could be more economical to replace all the tires on a truck when a replacement is made. The costs of transporting the truck to a service area, placing the truck on a lift, and paying a technician to mount and balance the tires could be comparable to the cost of the tire itself. If all the tires were replaced simultaneously, this cost would be incurred less often than if the tires were individually replaced.

This section will consider a model to determine the optimal time to replace an entire group of items. In order to avoid intricate mathematics that are beyond the scope of this book, we will assume that the lifetime of each operating unit is a discrete random variable with a known distribution. That is, suppose that p_k is the probability that an item fails in period k assuming the item was placed into service at period 0. These probabilities may be estimated directly from historical data or computed from a continuous distribution as in Example 13.11.

Assume that n_0 items are placed into service at time 0. Suppose there is no block replacement and all items that fail in a period are replaced at the end of that period. We also will assume for simplicity that p_k is the actual proportion of units k periods old that fail. Then the number of failures occurring in period 1 is $n_1 = n_0 p_1$.

In period 2 the proportion of the original group of items that fail is $n_0 p_2$, and the proportion of the items placed into service in period 1 that fail is $n_1 p_1$. Hence, the expected number of failures in period 2 is $n_2 = n_0 p_2 + n_1 p_1$. Continuing with this argument, we obtain

$$n_k = n_0 p_k + n_1 p_{k-1} + \cdots + n_{k-1} p_1.$$

Now suppose that individual replacements cost a_1 each and the entire block of n_0 can be replaced for a_2. If all n_0 items were replaced at the end of each period, the cost each period would be $a_2 + a_1 n_1$. If all n_0 items were replaced at the end of every other period, the cost incurred every two periods would be $a_2 + a_1(n_1 + n_2)$ or an average per period cost of $[a_2 + a_1(n_1 + n_2)]/2$. Similarly, the average per period cost of replacing all n_0 items after k periods is

$$G(k) = \frac{a_2 + a_1 \sum_{j=1}^{k} n_j}{k}.$$

The optimal number of periods to replace all n_0 items is the value of k that minimizes $G(k)$. The minimum value of $G(k)$ should be compared to the expected cost per period assuming that items are replaced as they fail. Let

$$E(T) = \sum_{k=1}^{\infty} kp_k$$

represent the expected lifetime of a single item. Then $\lambda = 1/T$ is the failure rate of a single item. It follows that the cost of making replacements to items on a one-at-a-time basis is $a_1\lambda$ per item or $n_0a_1\lambda$ for the entire block of items. This should be compared to the optimal value of $G(k)$ to determine if a block replacement strategy is economical.

Example *13.12* A large sign is lit by 8,000 bulbs. The bulbs cost \$2 each to replace as they fail but can be replaced for 30 cents each when they are replaced all at once. Based on past experience, bulbs fail according to the following probability law:

Months of Service	Probability of Failure
1	.02
2	.03
3	.03
4	.05
5	.08
6	.09
7	.07
8	.10
9	.11
10	.13
11	.15
12	.14

The first step is to compute n_k. We have

$$n_0 = 8{,}000,$$
$$n_1 = n_0p_1 = (8{,}000)(.02) = 160,$$
$$n_2 = n_0p_2 + n_1p_1 = (8{,}000)(.03) = (160)(.02) = 243,$$

and so on.

These values are used to compute $G(k)$. The results of the calculation appear in Table 13–4. Using values of $a_2 = (0.30)(8{,}000) = \$2{,}400$ and $a_1 = 2$, we see that the optimal time to replace the block of bulbs is after four months with an expected monthly cost of \$1,135.

It is interesting to compare block replacement with a policy of replacing the bulbs only if they fail. Each bulb has an expected lifetime of

$$E(T) = \sum_{k=1}^{12} kp_k = 8.22 \text{ months.}$$

Hence, the failure rate is $1/8.22 = 0.12165$ failure per month per bulb. For 8,000 bulbs this amounts to an average number of failures of 973.24 per month. The resulting replacement cost is \$1,946.47 monthly, which is considerably more than the cost of replacing bulbs as a block every four months, which has an expected monthly cost of \$1,135.

TABLE 13–4
***G*(*k*) for Example 13.12**

k	P_k	n_k	G_k
1	.02	160.0	2,720.0
2	.03	243.2	1,603.2
3	.03	249.2	1,235.2
4	.05	417.1	1,135.0
5	.08	671.1	1,176.4
6	.09	778.4	1,239.8
7	.07	654.6	1,249.7
8	.10	930.5	1,326.1
9	.11	1,064.0	1,415.2
10	.13	1,298.4	1,533.4
11	.15	1,542.9	1,674.5
12	.14	1,562.4	1,795.4

Problems for Section 13.7

26. Consider Example 13.11 of Harley Brown, Inc. Without performing the calculations, discuss what the effect on the optimal replacement policy would likely be if the parameter values were $c_1 = \$800$, $c_2 = \$300$, and $\beta = 1$.
27. Repeat the calculations for the Harley Brown trucking company using the following parameter values: $\alpha = 0.0156, \beta = 1.8, c_1 = 1{,}000, c_2 = 600$.
28. An expensive piece of equipment is used in the masking operation for semiconductor manufacture. A capacitor in the equipment fails randomly. The capacitor costs \$7.50, but if it burns out while the machine is in use, the production process must be halted. Here the replacement cost is estimated to be \$150. Based on past experience, the lifetime distribution of the capacitor is estimated to be

Number of Months of Service	Probability of Failure
1	.08
2	.12
3	.16
4	.26
5	.22
6	.16

How often should the capacitors be replaced in order to minimize the expected monthly cost of planned and unplanned replacement?

29. A large electronic pipe organ contains 100 fuses. Because of the power demands of the organ, the fuses burn out at a fairly regular rate, but newer fuses last longer than older ones. The probability distribution of the lifetime of a fuse is closely approximated by the Weibull law with $\alpha = 0.0204$ and $\beta = 1.8$. Assume that t is hours of playing time. The fuses cost \$1.35 each when replaced as a block but \$12 each when replaced just after a failure.

    a. Express the lifetime distribution as a discrete distribution assuming t is measured in hours. (Follow the procedure used in Example 13.11 for the Harley Brown trucking company.)

b. Determine the optimal time to replace all 100 fuses and the average hourly cost of that policy.

c. Compare the answer you obtained in part (*b*) with the cost of replacing the fuses as they fail. Which policy would you recommend?

30. Tires that fail in service result in significantly higher replacement costs than those replaced before failure. For an 18-wheeler (that is, a truck with 18 tires), failure on the road costs $300, whereas all 18 tires can be replaced prior to failure at a cost of $75 per tire. The probability of failure is given in the following table.

Number of Miles	Probability of Failure
0–5,000	.05
5,001–10,000	.15
10,001–15,000	.20
15,001–20,000	.40
20,001–25,000	.20

If tires are replaced at multiples of 5,000 miles only, what is the optimal age replacement policy?

31. The Navy uses a certain type of vacuum tube in a sonar scanning device. Based on past experience, the vacuum tube exhibits the following failure pattern:

Number of Months of Operation	Probability of Failure
1	.1
2	.1
3	.2
4	.1
5	.3
6	.2

Failures during operation cost $200 each, but the tube can be replaced before failure for $50. Find the optimal replacement strategy.

32. A local newsletter is printed on a printer with a cartridge that may break or run out of ink during operation. The cartridges can be replaced for $7.50, but if they fail when the newsletter is being printed, the cost is estimated to be $25 because of the delay in publication. The failure distribution of the cartridges is

Weeks of Use	Probability of Failure
1	.1
2	.2
3	.3
4	.4

Determine the optimal time to replace the cartridges.

33. Mactronics produces industrial robots. Each robot contains a part with a lifetime of five years at most, and at least one year. Lifetimes between one and five years are equally likely. If the part fails during operation, replacement costs are

estimated to be \$400, whereas the part can be replaced before failure for \$50. When should the part be replaced? [Hint: The lifetime distribution is uniform on {1, 2, 3, 4, 5}. Assuming discrete variables, this means that $f(x) = \frac{1}{5}$ for $x = 1, 2, \ldots, 5$.]

34. A firm has purchased 30 Mactronics robots, described in Problem 33. If replaced as a block, the parts cost \$20 each, but cost \$400 each when replaced after failure.
    a. Find the optimal block replacement strategy.
    b. Compare this to the cost of replacing the items as they fail.
    c. Compare the cost of the block replacement policy you obtained in part (*a*) with the solution you obtained in part (*b*). Is the block replacement strategy preferred to an individual replacement strategy?

*13.8 ANALYSIS OF WARRANTY POLICIES

An important issue related to the reliability of operating equipment is the protection afforded to the consumer who experiences failure of the equipment prior to its intended lifetime. Buyers and sellers perceive warranties differently. From the seller's point of view, the warranty is a means of limiting liability by specifying consumer responsibilities. These responsibilities include proper use of the product and following the warnings. From a marketing perspective, the warranty also can serve as an inducement to purchase the product. From the buyer's point of view, the warranty is a means of reducing or eliminating the economic penalty if the product fails to operate properly for a reasonable period of time. Warranties are particularly important to the consumer for products that are likely to experience high failure rates early in the product lifetime.

This section will present mathematical models for determining the economic value of a warranty. Such models could be used to find the portion of the cost of an item that could reasonably be attributed to the costs of satisfying a warranty commitment. In the models, we consider both the structure of the warranty and the reliability of the product to find the value of a warranty.

We must distinguish between repairable and nonrepairable items. Nonrepairable items include most electronic components, items in which failure corresponds to destruction of the item (burning out of a bulb or blowout of a tire, for example), or items typically not repaired but replaced (such as batteries that fail to hold a charge). Most major appliances, such as washing machines and televisions, fall into the category of repairable items. Repairable items that fail during the warranty period are typically repaired rather than replaced. The mathematical models presented in this section assume nonrepairable items.

Warranties for nonrepairable consumer goods generally take one of two forms. One is the free replacement warranty: if a failure occurs during the warranty period, a new item is supplied without charge. The second type of warranty is the pro rata warranty. Here, the consumer is given a rebate proportional to the amount of time remaining in the warranty period. The rebate is used to reduce the cost of a replacement item.

The Free Replacement Warranty

Assume that a single piece of operating equipment is placed into service and fails completely at random (that is, according to an exponential failure law) with known

failure rate λ. The item is assumed to operate continuously. For intermittently operating equipment, clock time could be measured in operating hours rather than elapsed hours.

We will use the following notation:

T = Lifetime of an item chosen at random.
λ = Failure rate of an item chosen at random.
$F(t)$ = Cumulative distribution function of the random variable t.
C_1 = Cost of purchasing a new item with free replacement warranty.
K = Cost of purchasing a new item without any warranty.
W_1 = Time that the free replacement warranty is in effect after purchase.

If a failure occurs during the warranty period, the item is replaced free of charge. Assume that the consumer purchases a new item when a failure occurs after the expiration of the warranty. The new item has an identical free replacement warranty. Let Y be a random variable representing the time between successive purchases by the consumer. From Figure 13–16, we see that

$$Y = W_1 + \text{Time until the first failure after the warranty expires.}$$

It can be shown that

$$E(Y) = W_1 + 1/\lambda.$$

This expression for $E(Y)$ is valid only when the lifetime distribution is exponential. It results from the property of the exponential failure law that the time of the first failure after a fixed time (known as the forward recurrence time in probability theory) has the same distribution as the time between two successive failures. This result is true *only* if the failure law is exponential.

We will now compute the cost per unit time for an item that is replaced infinitely many times and has a free replacement warranty.

Each time that an item is purchased after the warranty expires constitutes the start of a new cycle. As the cost per cycle is C_1, it follows that the average cost per unit time under the free replacement warranty is

$$\frac{C_1}{E(Y)} = \frac{C_1}{W_1 + 1/\lambda} = \frac{\lambda C_1}{\lambda W_1 + 1}.$$

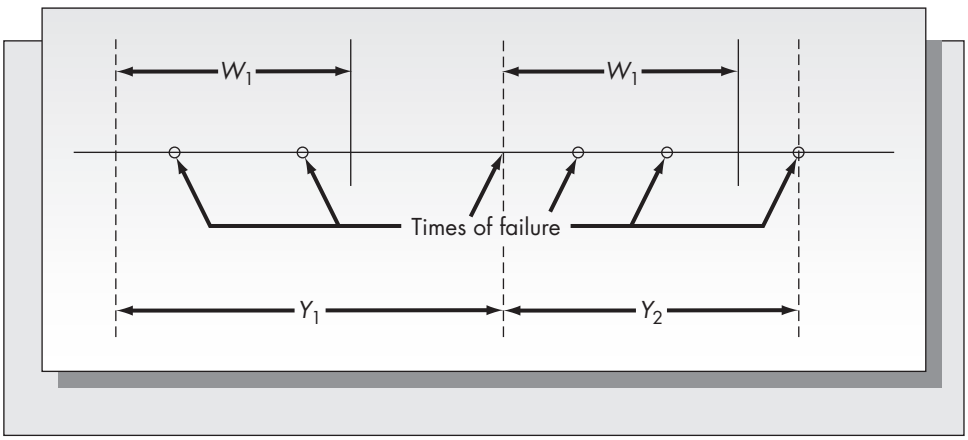


FIGURE 13–16
Replacement cycles for free replacement warranty

Without the warranty, the cost of replacement per unit time is simply λK. Let C_1^* be the cost of an item with a free replacement warranty that is indifferent to the cost of an item without the warranty, K. Then C_1^* solves

$$\lambda K = \frac{\lambda C_1^*}{\lambda W_1 + 1}$$

or

$$C_1^* = (\lambda W_1 + 1)K.$$

By definition $C_1^* - K$ is the economic value of the warranty. If $C_1 < C_1^*$, then the warranty should be purchased.

The Pro Rata Warranty

Assume all notation previously presented in this section. We also define

$C_2 =$ Cost of a new item with a pro rata warranty.

$W_2 =$ Effective warranty period with a pro rata warranty.

Consider an item purchased under a pro rata warranty that fails at a random time T. There are two cases:

Case 1: $T < W_2$. In this case the fraction of the warranty period that has expired is T/W_2. The cost of the replacement item is then $C_2(T/W_2)$.

Case 2: $T \geq W_2$. Here, the warranty has expired and the replacement cost is C_2.

Both of these cases can be represented mathematically by the expression

$$\frac{C_2}{W_2}\min(W_2, T).$$

To determine the expected life-cycle cost, we need to find an expression for $E[\min(W_2, T)]$. We have that

$$E[\min(W_2, T)] = \int_0^\infty \min(W_2, t)\lambda e^{-\lambda t}\, dt$$

$$= \int_0^{W_2} t\lambda e^{-\lambda t}\, dt + W_2 \int_{W_2}^\infty \lambda e^{-\lambda t}\, dt,$$

where $\lambda e^{-\lambda t}$ is the probability density of the time until failure, T.

The first integral requires integration by parts and the second is the reliability function evaluated at W_2. It is easy to show that

$$\int_0^{W_2} t\lambda e^{-\lambda t}\, dt = \frac{1}{\lambda}[1 - e^{-\lambda W_2}(1 + \lambda W_2)]$$

and

$$W_2 \int_{W_2}^\infty \lambda e^{-\lambda t}\, dt = W_2 e^{-\lambda W_2}.$$

Combining terms, it follows that

$$E[\min(W_2, T)] = \frac{1}{\lambda}(1 - e^{-\lambda W_2}).$$

The pro rata warranty starts anew with each purchase. Hence, each purchase begins a new cycle. It follows that the expected cost per unit time following a pro rata warranty is

$$\frac{C_2}{W_2}\frac{E[\min(W_2, T)]}{E(T)} = \frac{C_2}{W_2}\frac{(1/\lambda)(1 - e^{-\lambda W_2})}{1/\lambda} = \frac{C_2(1 - e^{-\lambda W_2})}{W_2}.$$

The cost of the pro rata warranty that is indifferent to the cost of the item without the warranty, C_2^*, solves

$$\frac{C_2^*(1 - e^{-\lambda W_2})}{W_2} = \lambda K,$$

or

$$C_2^* = \frac{\lambda K W_2}{1 - e^{-\lambda W_2}}.$$

The value of the pro rata warranty is $C_2^* - K$. If $C_2 > C_2^*$, then the pro rata warranty should not be purchased, and if $C_2 < C_2^*$, the pro rata warranty should be purchased. This assumes that the consumer is willing to base his or her decision on expected values. If the consumer is risk averse, the indifference value will be slightly higher than C_2^*.

Example *13.13* You are considering purchasing a battery for your automobile. You find the same battery offered at three different stores. Store A sells the battery for \$21 and offers no warranty or guarantee. Store B sells the battery for \$40 and offers a free replacement if the battery fails to hold a charge for the first two years of operation. Store C sells the battery for \$40 as well but offers a pro rata warranty for the anticipated lifetime of the battery, which is advertised to be five years. The failure rate of the battery depends on the usage and the conditions, but from past experience you estimate that the time between failure is about once every three years.

We will determine the values of C_1^* and C_2^* for both warranties. For the full replacement warranty, we have that

$$K = 21,$$
$$C_1 = 40,$$
$$\lambda = \tfrac{1}{3},$$
$$W_1 = 2.$$

It follows that

$$C_1^* = (\lambda W_1 + 1)K = [(\tfrac{1}{3})(2) + 1](21) = \$35.$$

This means that the value of the free replacement warranty is \$35 − \$21 = \$14, which is less than the \$19 difference in the prices. On the basis of expected costs, the battery with no warranty (store A) is preferred to the one with the free replacement warranty (store B).

For the case of the pro rata warranty offered by store C, $W_2 = 5$ and the remaining parameters are as defined before. In that case we obtain

$$C_2^* = \frac{\lambda K W_2}{1 - e^{-\lambda W_2}} = \frac{\frac{1}{3}(21)(5)}{1 - e^{-5/3}} = \$43.15.$$

The value of the pro rata warranty is \$43.15 − \$21.00 = \$22.15, which exceeds the difference between the price of the battery with the warranty and without, and exceeds the value of the free replacement warranty. Hence, on the basis of this analysis, the pro rata warranty is the preferred choice.

Extensions and Criticisms

A criticism of Example 13.13 is the assumption that the failure law for the batteries is exponential. This means that a new battery has the same probability of failing in its first year of operation as a four-year-old battery has of failing in its fifth year of operation. In the case of batteries, it would seem that a failure law incorporating aging, such as the Weibull, would be more accurate. The models discussed can be extended to include more general types of failure patterns, but the calculations required to determine an optimal policy are very complex. In particular, one must determine the renewal function, which is generally a difficult computation. In Problem 41 an example is considered for the case in which the failure law follows an Erlang distribution.

The two types of warranties discussed in this section, the free replacement warranty and the pro rata warranty, account for the majority of consumer warranties for nonrepairable items. In the military context, another type of warranty is very common. This is known as the *reliability improvement warranty* (RIW). In this case, the supplier agrees to repair or replace items that fail within a specified warranty period and provides a pool or pools of spares and perhaps one or more repair facilities. This type of warranty is intended to provide an incentive to the supplier to make initial reliability high and to provide improvements in the reliability of existing units if possible.

For repairable items, warranties usually cover all or some of the cost of effecting repairs during the warranty period. Analysis of warranties for repairable items is more complex than that for nonrepairable items. Different levels of repair are possible and the likelihood of failure during the warranty period may depend upon usage, which can vary significantly from one user to another.

For repairable items, an issue closely related to warranties is service contracts. The primary difference between a warranty and a service contract is that the cost of the warranty is usually assumed to be included in the purchase price of the item, whereas a service contract is an additional item whose purchase is at the option of the buyer. A service contract can be thought of as an extended warranty beyond the normal warranty period. The analytical approach discussed here also can be applied to determining the economic value of a service contract. However, complex failure laws and different costs associated with various levels of repair should be allowed. In addition, the pricing of consumer service contracts relies heavily on the assumption that the consumer is risk averse. That is, a typical consumer would prefer to pay more than a service contract is worth in an expected value sense to reduce the risk of incurring an expensive repair. In that sense, the service contract serves the role of an insurance policy. Perhaps it is the tendency of consumers to be risk averse that justifies the rather high prices that are charged for service contracts for many consumer products.

Problems for Section 13.8

35. For Example 13.13, what value of the warranty period equates the full replacement warranty with no warranty? (That is, for what value of W_1 is the consumer indifferent to purchasing from store A or store B?)
36. For Example 13.13, what value of the warranty period for the full replacement warranty equates the full replacement and the pro rata warranties? (For what value of W_1 is the consumer indifferent to purchasing from store B or store C, assuming $W_2 = 5$?)

37. A producer of pocket calculators estimates that the calculators fail at a rate of one every five years. The calculators are sold for $25 each with a one-year free replacement warranty but can be purchased from an unregistered mail-order source for $18.50 without the warranty. Is it worth purchasing the calculator with the warranty?
38. For Problem 37, what length of period of the warranty equates the replacement costs of the calculator with and without the warranty?
39. Zemansky's sells tires with a pro rata warranty. The tires are warranted to deliver 50,000 miles with the rebate based on the remaining tread on the tire. The tires fail on the average after 35,000 miles of wear. Suppose the tires sell for $50 each with the warranty. If failures occur completely at random, what would be a consistent price for the tires if no warranty were offered?
40. Habard's, a chain of hardware stores, sells a variety of tools and home repair items. One of their best wrenches sells for $5.50. Habard's will include a three-year free replacement warranty for an additional $1.50. The wrench is expected to be subject to heavy use and, based on past experience, will fail randomly at a rate of one every eight years. Is it worth purchasing the warranty?
41. Consider the case in which the failure mechanism for the product does not obey the exponential law. In that case, the cost under the free replacement warranty that is indifferent to the cost of buying the item without a warranty is given by

$$C_1^* = K[M(W_1) + 1],$$

where $M(t)$ is known as the renewal function.

If the time between failures, T, follows an Erlang law with parameters λ and 2, then

$$M(t) = \frac{\lambda t}{2} - 0.25 + 0.25e^{-2\lambda t} \qquad \text{for all } t \geq 0.$$

(See, for example, Barlow and Proschan, 1965, p. 57.)

a. For Example 13.13, presented in this section, determine the indifference value of the item with a free replacement warranty when the failure law follows an Erlang distribution. Assume that $\lambda = \frac{2}{3}$ to give the same value of $E(T)$ as in the example.

b. Is the value of the warranty larger or smaller than in the corresponding exponential case? Explain the result intuitively.

13.9 SOFTWARE RELIABILITY

Software reliability is a problem with characteristics different from hardware reliability problems. Typically, new software possesses a few "bugs," or errors. Ideally, one would like to remove all the bugs from the software before its release, but that may be impossible. It is more reasonable to release the software when the number of bugs has been reduced to an acceptable level. Predicting the number of remaining bugs is, however, a difficult problem.

The importance of software reliability cannot be overemphasized. Quoting from *The Wall Street Journal* (Davis, 1987):

> The tiniest software bug can fell the mightiest machine—often with disastrous consequences. During the past five years, software defects have killed sailors, maimed patients, wounded

Snapshot Application

RELIABILITY-CENTERED MAINTENANCE IMPROVES OPERATIONS AT THREE MILE ISLAND NUCLEAR PLANT

The Three Mile Island nuclear facility located on the Susquehanna River about 10 miles from Harrisburg, Pennsylvania, is notorious in one respect. It was the site of the worst nuclear power generating plant accident in the United States. In March of 1979, Unit 2 underwent a core meltdown as safety systems failed to lift nuclear fuel rods from the core. The facility was shut down as a result of the accident for the next six and one-half years, finally reopening in October 1985. The plant, operated by GPU Nuclear Corporation, has compiled one of the most impressive records in the industry since it has reopened. According to Fox et al. (1994), the plant was ranked top in the world in 1989 on the basis of its capacity factor (proportion of up-time).

In 1987, GPU began to consider the benefits of a reliability-centered maintenance (RCM) approach to preventive maintenance. They identified 28 out of a total of 134 systems as viable candidates for RCM. These 28 systems included the main turbine, the cooling water system, the main generator, and circulating water. The RCM process relied on the following four basic principles:

- Preserve system functions.
- Identify equipment failures that defeat those functions.
- Prioritize failure modes.
- Define preventive maintenance tasks for high-priority failure modes.

The RCM project spanned the period of September 1988 to June 1994. A total of 3,778 components in the 28 subsystems came under consideration. By the end of the program, preventive maintenance policies included more than 5,400 tasks for these components. The cost of implementing RCM was substantial: about $30,000 per system. However, these costs were more than offset by the benefits. Over the period 1990 to 1994, records show a significant decline in plant equipment failures. In addition, a reliability-based maintenance program can have other benefits, including

- Increased plant availability.
- Optimized spare parts inventories.
- Identification of component failure modes.
- Discovery of new plant failure scenarios.
- Training for engineering personnel.
- Identification of components that benefit from revised preventive maintenance strategies.
- Identification of potential design improvements.
- Improved documentation.

Fox et al. (1994) report several lessons learned from this experience. One is that it is better for the internal maintenance organization, rather than an outside agency, to direct the process. This avoids the "we versus they" syndrome. Successful implementation is also more likely in this case. A cost analysis checklist was developed to screen failure modes. Finally, the team evolved an efficient multiuser relational database software system to facilitate RCM evaluations. This system reduced the time required to perform the necessary analyses by 50 percent.

The lesson learned from this case is that a carefully designed and implemented reliability-based preventive maintenance program can have big payoffs for high stakes systems.

corporations and threatened to cause the government-securities market to collapse. Such problems are likely to grow as industry and the military increasingly rely on software to run systems of phenomenal complexity, including President Reagan's proposed "Star Wars" anti-missile defense system.

Several models have been proposed for estimating software reliability. However, we will not present these models in detail because their utility has yet to be determined. Jelinski and Moranda (1972) have suggested the following approach. Let N be the total initial error content (i.e., the number of bugs) in the software. As the software undergoes testing, the number of bugs is reduced. They assume that the failure rate (that is, the likelihood of detecting a bug) is proportional to the number of bugs remaining in the program, where ϕ is the proportionality constant. That is, the time until detection

of the first bug has the exponential distribution with parameter $N\phi$; the time between detection of the first and the second bugs has exponential distribution with parameter $(N-1)\phi$; and so on.

Hence, as bugs are removed from the program, the amount of time required to detect the next bug increases. After n bugs have been removed, one will have observed the values of $T_1, T_2, \ldots, T_n$ representing the time between successive detections. These observations are used to estimate ϕ and N using the maximum likelihood principle. Based on these estimates, one could predict exactly how much testing would be required in order to achieve a certain level of reliability in the software.

Shooman (1972) suggests using a normalized error rate to measure the error content in the program. He defines

$$p(t) = \text{Errors per total number of instructions per month of debugging time}$$

and develops a reliability model based on first principles. He demonstrates how this model can be used to build a functional relationship between the amount of time devoted to debugging and the reliability of the program.

The works of Jelinski and Moranda and of Shooman represent the foundation of the theory of software reliability. Extensions of their methods have been considered. It remains to be seen, however, if these methods provide accurate descriptions of the problem and whether they ultimately will assist in predicting the time required to achieve an acceptable level of reliability.

13.10 HISTORICAL NOTES

Much of the theory of reliability, life testing, and maintenance strategies has its roots in actuarial theory developed by the insurance industry. Sophisticated mathematical models for predicting survival probabilities date back to the turn of the century. Lotka (1939) discusses some of the connections between equipment replacement models and actuarial studies. The work of Weibull (1939 and 1951) laid the foundations for the subject of fatigue life in materials.

Interest in reliability problems became considerably more widespread during World War II when attempts were made to understand the failure laws governing complex military systems. During the 1950s, problems concerning life testing and missile reliability began to receive serious attention. In 1952 the Department of Defense established the Advisory Group on Reliability of Electronic Equipment, which published its first report on reliability in June of 1957.

The origins of the specific age replacement models presented in this chapter are unclear. However, sophisticated age replacement models date back as far as the early 1920s (see Taylor, 1923, and Hotelling, 1925). The stochastic planned replacement models presented in Section 13.7 form the basis for much of the research in replacement theory, but the origins of these models are unclear as well.

Section 13.8, on warranties, is based on the paper by Blischke and Scheuer (1975). Extensions and corrections of their work can be found in Mamer (1982). Readers interested in pursuing further reading should refer to the excellent texts by Barlow and Proschan (1965 and 1975) on reliability models, and by Gertsbakh (1977) on maintenance strategies. Issues concerning the application of maintenance models are discussed by Turban (1967) and Mann (1976).

13.11 Summary

The purpose of this chapter was to review the terminology and the methodology of the theory and application of reliability and maintenance models. *Reliability theory* is an area of study that has received considerable attention from mathematicians. However, the mathematics is of interest not only for its own sake. These models are extremely useful in an operational setting in considering such issues as failure characteristics of operating equipment, economically sound maintenance strategies, and the value of product warranties and service contracts.

The complexity of the analysis depends upon the assumptions made about the random variable T, which represents the lifetime of a single item or piece of operating equipment. The *distribution function* of T, $F(t)$, is the probability that the item fails at or before time t ($P\{T \leq t\}$), whereas the *reliability function* of T, $R(t)$, is the probability that the item fails after time t ($P\{T > t\}$). An important quantity related to these functions is the *failure rate function* $r(t)$, which is the ratio $f(t)/R(t)$ of the probability density function and the reliability function. If Δt is sufficiently small, the term $r(t)\Delta t$ can be interpreted as the conditional probability that the item will fail in the next Δt units of time given that it has survived up until time t.

The failure rate function provides considerable information about the aging characteristics of operating equipment. In a manufacturing environment, we would expect that most operating equipment would have an increasing failure rate function. That means it would be more likely to fail as it ages. A decreasing failure rate function can arise when the likelihood of early failure is high due to defectives in the population. The *Weibull* probability law can be used to describe the failure characteristics of equipment having either an increasing or a decreasing failure rate function.

Of interest is the case in which the failure rate function is constant. This case gives rise to the *exponential distribution* for the lifetime of a single component. The exponential distribution is the only continuous distribution possessing the *memoryless property.* This means that the conditional probability that an item that has been operating up until time t fails in the next s units of time is independent of t.

The *Poisson process* describes the situation in which a single piece of operating equipment fails according to the exponential distribution and is replaced immediately upon failure. When this occurs, the number of failures in a given time has the Poisson distribution, the time between successive failures has the exponential distribution, and the time for n failures to occur has the Erlang distribution.

The chapter considered the reliability functions of complex systems of components. It showed how to obtain the reliability functions for *components in series and parallel* from the reliability functions of the individual components. The chapter also considered K out of N systems, which function only if at least K components function.

Reliability issues form the basis of the *maintenance models* discussed in the latter half of the chapter. An important measure of a system's performance is the *availability,* which is the proportion of the time that the equipment operates. We treated both deterministic age replacement models, which do not explicitly include the likelihood of equipment failure, and stochastic age replacement models, which do. The stochastic models allow for replacing the equipment before failure. This is of interest when items have an increasing failure rate function and unplanned failures are more costly than planned failures.

Finally, we concluded the chapter with a discussion of the economic value of *warranties.* A warranty is a promise supplied by the seller to the buyer to either replace the item with a new one if it fails during the warranty period (*free replacement warranty*) or provide a discount on the purchase of a new item proportional to the remaining amount of time

(or wear) in the warranty period (*pro rata warranty*). The issues surrounding warranties and service contracts are similar, but service contract models are considerably more complex, owing to the need to include multiple levels of repair.

Additional Problems on Reliability and Maintainability

42. A large national producer of appliances has traced customer experience with a popular toaster oven. A survey of 5,000 customers who purchased the oven early in 2000 has revealed the following:

Year	Number of Breakdowns
2000	188
2001	58
2002	63
2003	72
2004	54
2005	71

    a. Using these data, estimate p_k = the probability that a toaster oven fails in its kth year of operation, for $k = 1, \ldots, 6$.

    b. What is the likelihood that a toaster oven will last at least six years without failure based on these data?

    c. The discrete failure rate function has the form $r_k = p_k/R_{k-1}$, where R_k is the probability that a unit survives through period k. Determine the failure rate function for the first five years of operation from the given data.

    d. Suppose that you purchased a toaster oven at the beginning of 2004 and it is still operating at the end of 2007. If the reliability has not changed appreciably from 2000 to 2007, use the results of part (*c*) to obtain the probability that it will fail during the first two months of calendar year 2008.

43. Six thousand light bulbs light a large hotel and casino marquee. Each bulb fails completely at random, and each has an average lifetime of 3,280 hours. Assuming that the marquee stays lit continuously and bulbs that burn out are replaced immediately, how many replacements must be made each year on the average?

44. The owner of the hotel mentioned in Problem 43 has decided that in order to decrease the number of burned-out bulbs, she will replace all 6,000 bulbs at the start of each year in addition to replacing the bulbs as they burn out. Comment on the effectiveness of this strategy.

45. The owner of the hotel mentioned in Problem 43 falls on hard times and dispenses with replacement of the bulbs. She notices that more than half of the bulbs have burned out before the advertised average lifetime of 3,280 hours and decides to sue the light bulb manufacturer for false advertising. Do you think she has a case? (Hint: What fraction of the bulbs would be expected to fail prior to the mean lifetime?)

46. Continuing with the example of Problem 43, determine the following:

    a. The proportion of bulbs lasting more than two years.

    b. The probability that a bulb chosen at random fails in the first three months of operation.

c. The probability that a bulb that has lasted for 10 years fails in the next three months of operation.

47. Assume that the bulbs in Problem 43 are not replaced as they fail.

a. What fraction of the 6,000 bulbs are expected to fail in the first year?

b. What fraction of the bulbs surviving the first year are expected to fail in the second year?

c. What fraction of the bulbs surviving the nth year are expected to fail in year $n + 1$ for any value of $n = 1, 2, \ldots$?

d. Using the results of part (*c*), of the original 6,000 bulbs, how many would be expected to fail in the fourth year of operation?

48. The mean value of a Weibull random variable is given by the formula

$$\mu = a^{-1/\beta}\,\Gamma(1 + 1/\beta),$$

where Γ represents the gamma function. The gamma function has the property that $\Gamma(k) = (k - 1)\Gamma(k - 1)$ for any value of $k > 1$ and $\Gamma(1) = 1$. Notice that if k is an integer, this results in $\Gamma(k) = (k - 1)!$. If k is not an integer, one must use the recursive definition for $\Gamma(k)$ coupled with the following table. For values of $1 \leq k \leq 2$, $\Gamma(k)$ is given by

k	$\Gamma(k)$	k	$\Gamma(k)$
1.00	1.0000	1.55	.8889
1.05	.9735	1.60	.8935
1.10	.9514	1.65	.9001
1.15	.9330	1.70	.9086
1.20	.9182	1.75	.9191
1.25	.9064	1.80	.9314
1.30	.8975	1.85	.9456
1.35	.8912	1.90	.9612
1.40	.8873	1.95	.9799
1.45	.8857	2.00	1.0000
1.50	.8862		

For example, this table would be used as follows: $\Gamma(3.6) = (2.6)\Gamma(2.6) = (2.6)(1.6)\Gamma(1.6) = (2.6)(1.6)(.8935) = 3.717$.

a. Compute the expected failure time for Example 13.4 regarding copier equipment.

b. Compute the expected failure time for a piece of operating equipment whose failure law is given in Example 13.2.

c. Determine the mean failure time for $\alpha = 1.35$ and $\beta = 0.20$.

d. Determine the mean failure time for $\alpha = 0.90$ and $\beta = 0.45$.

49. Suppose that a particular light bulb is advertised as having an average lifetime of 2,000 hours and is known to satisfy an exponential failure law. Suppose for simplicity that the bulb is used continuously. Find the probability that the bulb lasts

a. More than 3,000 hours.

b. Less than 1,500 hours.

c. Between 2,000 and 2,500 hours.

50. Applicational Materials sells several pieces of equipment used in the manufacture of silicon-based microprocessors. In 2003 the company filled 130 orders for model a55212. Suppose that the machines fail according to a Weibull law. In particular, the cumulative distribution function $F(t)$ of the time until failure of any machine

is given by

$$F(t) = 1 - e^{-0.0475t^{1.2}} \qquad \text{for all } t \geq 0,$$

where t is in years.

a. What is the failure rate function for this piece of equipment?

b. Of the original 130 sold in 2003, how many machines would one expect would not experience a breakdown before January 2007? Assume for the sake of simplicity that all the machines were sold on January 1, 2003.

c. Using the results of part (*a*), estimate the fraction of machines that have survived 10 years of use that will break down during the 11th year of operation [or you may compute this directly if you did not get the answer to part (*a*)].

51. A local cab company maintains a fleet of 10 cabs. Each time a cab breaks down, it is repaired the same day. Assume that breakdowns of individual cabs occur completely at random at a rate of two per year.

a. What is the probability that any particular cab will run for a full year without suffering a breakdown?

b. What is the probability that the entire fleet will run for one month without a breakdown?

c. On the average, how many breakdowns would the fleet expect in a typical three-month period?

d. What is the probability that there are more than five breakdowns between Thanksgiving Day (November 28) and New Year's Day (January 1)?

e. For what reason might your answer in part (*d*) be too low?

52. A collection of 30 Christmas tree lights are arranged in a pure series circuit; that is, if one of the lights burns out, then the entire string goes out. Suppose that each light fails completely at random at a rate of one failure every year. What is the probability that the lights will burn from the beginning of Christmas Eve (December 24) to the end of New Year's Day (January 1) without failure?

53. A piece of industrial machinery costs $48,000 to replace and has essentially no salvage value. Over the first five years of operation, maintenance costs amounted to $8,000. If the maintenance cost rate is a linear function of time, what is the optimal age at which to replace the machinery?

54. For an automobile that you own or would like to own, estimate the correct values of the replacement cost, the rate of depreciation, the initial maintenance cost, and the rate at which the maintenance cost increases. Based on these estimates, determine the optimal number of years that you should wait before replacing your car.

Appendix 13–A

Glossary of Notation on Reliability and Maintainability

a = Maintenance cost rate per unit time for simple age replacement model. Also used as a parameter of the exponential maintenance cost function for the exponential age replacement model.