71. We want to prove $P(m)$, the sum rule for m tasks, which says that if tasks T_1, T_2, ..., T_m can be done in n_1, n_2, ..., n_m ways, respectively, and no two of them can be done at the same time, then there are $n_1 + n_2 + \cdots + n_m$ ways to do one of the tasks. The basis step of our proof by mathematical induction is $m = 2$, and that has already been given. Now assume that $P(m)$ is true, and we want to prove $P(m+1)$. There are $m+1$ tasks, no two of which can be done at the same time; we want to do one of them. Either we choose one from among the first m, or we choose the task T_{m+1}. By the sum rule for two tasks, the number of ways we can do this is $n + n_{m+1}$, where n is the number of ways we can do one of the tasks among the first m. But by the inductive hypothesis $n = n_1 + n_2 + \cdots + n_m$. Therefore the number of ways we can do one of the $m+1$ tasks is $n_1 + n_2 + \cdots + n_m + n_{m+1}$, as desired.

73. A diagonal joins two vertices of the polygon, but they must be vertices that are not already joined by a side of the polygon. Thus there are $n-3$ diagonals emanating from each vertex of the polygon (we've excluded two of the $n-1$ other vertices as possible targets for diagonals). If we multiply $n-3$ by n, the number of vertices, we will have counted each diagonal exactly twice—once for each endpoint. We compensate for this overcounting by dividing by 2. Therefore there are $n(n-3)/2$ diagonals. (Note that the convexity of the polygon had nothing to do with the problem—we were counting the diagonals, whether or not we could be sure that they all lay inside the polygon.)

SECTION 6.2 The Pigeonhole Principle

The pigeonhole principle seems so trivial that it is difficult to realize how powerful it is in solving some mathematical problems. As usual with combinatorial problems, the trick is to look at things the right way, which usually means coming up with the clever insight, perhaps after hours of agonizing and frustrating exploration with a problem.

Try to solve these problems by invoking the pigeonhole principle explicitly, even if you can see other ways of doing them; you will gain some insights by formulating the problem and your solution in this way. The trick, of course, is to figure out what should be the pigeons and what should be the pigeonholes. Many of the the hints of Section 6.1 apply here, as well as general problem-solving techniques, especially the willingness to play with a problem for a long time before giving up.

Many of the elegant applications are quite subtle and difficult, and there are even more subtle and difficult applications not touched on here. Not every problem, of course, fits into the model of one of the examples in the text. In particular, Exercise 42 looks deceptively like a problem amenable to the technique discussed in Example 10. Keep in mind that the process of grappling with problems such as these is worthwhile and educational in itself, even if you never find the solution.

1. There are six classes: these are the pigeons. There are five days on which classes may meet (Monday through Friday): these are the pigeonholes. Each class must meet on a day (each pigeon must occupy a pigeonhole). By the pigeonhole principle at least one day must contain at least two classes.

3. **a)** There are two colors: these are the pigeonholes. We want to know the least number of pigeons needed to insure that at least one of the pigeonholes contains two pigeons. By the pigeonhole principle the answer is 3. If three socks are taken from the drawer, at least two must have the same color. On the other hand two socks are not enough, because one might be brown and the other black. Note that the number of socks was irrelevant (assuming that it was at least 3).

b) He needs to take out 14 socks in order to insure at least two black socks. If he does so, then at most 12 of them are brown, so at least two are black. On the other hand, if he removes 13 or fewer socks, then 12 of them could be brown, and he might not get his pair of black socks. This time the number of socks did matter.

5. There are four possible remainders when an integer is divided by 4 (these are the pigeonholes here): 0, 1, 2, or 3. Therefore, by the pigeonhole principle at least two of the five given remainders (these are the pigeons) must the same.

7. Let the n consecutive integers be denoted $x+1$, $x+2$, $\ldots$, $x+n$, where x is some integer. We want to show that exactly one of these is divisible by n. There are n possible remainders when an integer is divided by n, namely 0, 1, 2, $\ldots$, $n-1$. There are two possibilities for the remainders of our collection of n numbers: either they cover all the possible remainders (in which case exactly one of our numbers has a remainder of 0 and is therefore divisible by n), or they do not. If they do not, then by the pigeonhole principle, since there are then fewer than n pigeonholes (remainders) for n pigeons (the numbers in our collection), at least one remainder must occur twice. In other words, it must be the case that $x+i$ and $x+j$ have the same remainder when divided by n for some pair of numbers i and j with $0 < i < j \leq n$. Since $x+i$ and $x+j$ have the same remainder when divided by n, if we subtract $x+i$ from $x+j$, then we will get a number divisible by n. This means that $j-i$ is divisible by n. But this is impossible, since $j-i$ is a positive integer strictly less than n. Therefore the first possibility must hold, that exactly one of the numbers in our collection is divisible by n.

9. The generalized pigeonhole principle applies here. The pigeons are the students (no slur intended), and the pigeonholes are the states, 50 in number. By the generalized pigeonhole principle if we want there to be at least 100 pigeons in at least one of the pigeonholes, then we need to have a total of N pigeons so that $\lceil N/50 \rceil \geq 100$. This will be the case as long as $N \geq 99 \cdot 50 + 1 = 4951$. Therefore we need at least 4951 students to guarantee that at least 100 come from a single state.

11. We must recall from analytic geometry that the midpoint of the segment whose endpoints are (a,b,c) and (d,e,f) is $((a+d)/2, (b+e)/2, (c+f)/2)$. We are concerned only with integer values of the original coordinates. Clearly the coordinates of these fractions will be integers as well if and only if a and d have the same parity (both odd or both even), b and e have the same parity, and c and f have the same parity. Thus what matters in this problem is the parities of the coordinates. There are eight possible triples of parities: (odd, odd, odd), (odd, odd, even), (odd, even, odd), $\ldots$, (even, even, even). Since we are given nine points, the pigeonhole principle guarantees that at least two of them will have the same triple of parities. The midpoint of the segment joining these two points will therefore have integer coordinates.

13. **a)** We can group the first eight positive integers into four subsets of two integers each, each subset adding up to 9: $\{1,8\}$, $\{2,7\}$, $\{3,6\}$, and $\{4,5\}$. If we select five integers from this set, then by the pigeonhole principle (at least) two of them must come from the same subset. These two integers have a sum of 9, as desired.
b) No. If we select one element from each of the subsets specified in part **(a)**, then no sum will be 9. For example, we can select 1, 2, 3, and 4.

15. We can apply the pigeonhole principle by grouping the numbers cleverly into pairs (subsets) that add up to 7, namely $\{1,6\}$, $\{2,5\}$, and $\{3,4\}$. If we select four numbers from the set $\{1,2,3,4,5,6\}$, then at least two of them must fall within the same subset, since there are only three subsets. Two numbers in the same subset are the desired pair that add up to 7. We also need to point out that choosing three numbers is not enough, since we could choose $\{1,2,3\}$, and no pair of them add up to more than 5.

17. The given information tells us that there are $50 \cdot 85 \cdot 5 = 21{,}250$ bins. If we had this many products, then each could be stored in a separate bin. By the pigeonhole principle, however, if there are at least 21,251 products, then at least two of them must be stored in the same bin. This is the number the problem is asking us for.

19. **a)** If this statement were not true, then there would be at most 8 from each class standing, for a total of at most 24 students. This contradicts the fact that there are 25 students in the class.

b) If this statement were not true, then there would be at most 2 freshmen, at most 18 sophomores, and at most 4 juniors, for a total of at most 24 students. This contradicts the fact that there are 25 students in the class.

21. One way to do this is to have the sequence contain four groups of four numbers each, so that the numbers within each group are decreasing, and so that the numbers between groups are increasing. For example, we could take the sequence to be 4, 3, 2, 1; 8, 7, 6, 5; 12, 11, 10, 9; 16, 15, 14, 13. There can be no increasing subsequence of five terms, because any increasing subsequence can have only one element from each of the four groups. There can be no decreasing subsequence of five terms, because any decreasing subsequence cannot have elements from more than one group.

23. The key here is that 25 is an odd number. If there were an even number of boys and the same even number of girls, then we could position them around the table in the order BBGGBBGG... and never have a person both of whose neighbors are boys. But with an odd number of each sex, that cannot happen. Here is one nice way to see why when there are 25 of each. Number the seats around the table from 1 to 50, and think of seat 50 as being adjacent to seat 1. There are 25 seats with odd numbers and 25 seats with even numbers. If no more than 12 boys occupied the odd-numbered seats, then at least 13 boys would occupy the even-numbered seats, and vice versa. Without loss of generality, assume that at least 13 boys occupy the 25 odd-numbered seats. Then at least two of those boys must be in consecutive odd-numbered seats. The person sitting between those two boys will have boys as both of his or her neighbors.

25. This is actually a fairly hard problem, in terms of what we need to keep track of. Call the given sequence $a_1, a_2, \ldots, a_n$. We will keep track of the lengths of long increasing or decreasing subsequences by assigning values i_k and d_k for each k from 1 to n, indicating the length of the longest increasing subsequence ending with a_k and the length of the longest decreasing subsequence ending with a_k, respectively. Thus $i_1 = d_1 = 1$, since a_1 is both an increasing and a decreasing subsequence of length 1. If $a_2 < a_1$, then $i_2 = 1$ and $d_2 = 2$, since the longest increasing subsequence ending at a_2 is just a_2, but the longest decreasing subsequence ending at a_2 is a_1, a_2. If $a_2 > a_1$, then it is the other way around: $i_2 = 2$ and $d_2 = 1$. In general, we can determine i_k in the following manner (the determination of d_k is similar, with the roles of i and d, and the roles of $<$ and $>$, reversed). We look at the numbers $a_1, a_2, \ldots, a_{k-1}$. For each a_j that is less than a_k, we know that the value of i_k is at least $i_j + 1$, since the increasing subsequence of length i_j ending at a_j can be extended by following it by a_k, resulting in an increasing subsequence of length $i_j + 1$, ending at a_k. Furthermore, there is no other way of producing an increasing subsequence ending at a_k, other than the subsequence of length 1. Thus we set i_k equal to either 1 or the maximum of the numbers $i_j + 1$ for those values of $j < k$ for which $a_j < a_k$. Finally, once we have determined all the values i_k and d_k, we choose the largest of these $2n$ numbers as our answer.

The procedure just described, however, does not keep track of what the longest subsequence is, so we need to use two more sets of variables, $iprev_k$ and $dprev_k$. These will point back to the terms in the sequence that caused the values of i_k and d_k to be what they are. To retrieve the longest increasing or decreasing subsequence, once we know which type it is and where it ends, we follow these pointers, thereby exhibiting the subsequence backwards. We will not write the pseudocode for this final phase of the algorithm. (See the answer in the back of the textbook for an alternative procedure, which explicitly computes the sequence.)

procedure *long subsequence*($a_1, a_2, \ldots, a_n$: distinct integers)
for $k := 1$ **to** n
    $i_k := 1$; $d_k := 1$
    $iprev_k := k$; $dprev_k := k$
    for $j := 1$ **to** $k - 1$
        if $a_j < a_k$ and $i_j + 1 > i_k$ **then**
            $i_k := i_j + 1$
            $iprev_k := j$
        if $a_j > a_k$ and $d_j + 1 > d_k$ **then**
            $d_k := d_j + 1$
            $dprev_k := j$
{at this point correct values of i_k and d_k have all been assigned}
longest := 1
for $k := 2$ **to** n
    if $i_k >$ *longest* **then** *longest* := i_k
    if $d_k >$ *longest* **then** *longest* := d_k
{ *longest* is the length of the longest increasing or decreasing subsequence}

27. We can prove these statements using both the result and the method of Example 13. First note that the role of "mutual friend" and "mutual enemy" is symmetric, so it is really enough to prove one of these statements; the other will follow by interchanging the roles. So let us prove that in every group of 10 people, either there are 3 mutual friends or 4 mutual enemies. Consider one person; call this person A. Of the 9 other people, either there must be 6 enemies of A, or there must be 4 friends of A (if there were 5 or fewer enemies and 3 or fewer friends, that would only account for 8 people). We need to consider the two cases separately. First suppose that A has 6 enemies. Apply the result of Example 13 to these 6 people: among them either there are 3 mutual friends or there are 3 mutual enemies. If there are 3 mutual friends, then we are done. If there are 3 mutual enemies, then these 3 people, together with A, form a group of 4 mutual enemies, and again we are done. That finishes the first case. The second case was that A had 4 friends. If some pair of these people are friends, then they, together with A, form the desired group of 3 mutual friends. Otherwise, these 4 people are the desired group of 4 mutual enemies. Thus in either case we have found either 3 mutual friends or 4 mutual enemies.

29. We need to show two things: that if we have a group of n people, then among them we must find either a pair of friends or a subset of n of them all of whom are mutual enemies; and that there exists a group of $n-1$ people for which this is not possible. For the first statement, if there is any pair of friends, then the condition is satisfied, and if not, then every pair of people are enemies, so the second condition is satisfied. For the second statement, if we have a group of $n-1$ people all of whom are enemies of each other, then there is neither a pair of friends nor a subset of n of them all of whom are mutual enemies.

31. First we need to figure out how many distinct combinations of initials and birthdays there are. The product rule tells us that since there are 26 ways to choose each of the 3 initials and 366 ways to choose the birthday, there are $26 \cdot 26 \cdot 26 \cdot 366 = 6{,}432{,}816$ such combinations. By the generalized pigeonhole principle, with these combinations as the pigeonholes and the 37 million people as the pigeons, there must be at least $\lceil 37{,}000{,}000/6{,}432{,}816 \rceil = 6$ people with the same combination.

33. The numbers from 1 to 200,000 are the pigeonholes, and the inhabitants of Paris are the pigeons, which number at least 800,001. Therefore by Theorem 1 there are at least two Parisians with the same number of hairs on their heads; and by Theorem 2 there are at least $\lceil 800{,}001/200{,}000 \rceil = 5$ Parisians with the same number of hairs on their heads.

35. The 38 time periods are the pigeonholes, and the 677 classes are the pigeons. By the generalized pigeonhole principle there is some time period in which at least $\lceil 677/38 \rceil = 18$ classes are meeting. Since each class must meet in a different room, we need 18 rooms.

37. Let c_i be the number of computers that the i^{th} computer is connected to. Each of these integers is between 0 and 5, inclusive. The pigeonhole principle does not allow us to conclude immediately that two of these numbers must be the same, since there are six numbers (pigeons) and six possible values (pigeonholes). However, if not all of the values are used, then the pigeonhole principle would allow us to draw the desired conclusion. Let us therefore show that not all of the numbers can be used. The only way that the value 5 can appear as one of the c_i's is if one computer is connected to each of the others. In that case, the number 0 cannot appear, since no computer could be connected to none of the others. Thus not both 5 and 0 can appear in our list, and the above argument is valid.

39. This is similar to Example 9. Label the computers C_1 through C_{100}, and label the printers P_1 through P_{20}. If we connect C_k to P_k for $k = 1, 2, \ldots, 20$ and connect each of the computers C_{21} though C_{100} to *all* the printers, then we have used a total of $20 + 80 \cdot 20 = 1620$ cables. Clearly this is sufficient, because if computers C_1 through C_{20} need printers, then they can use the printers with the same subscripts, and if any computers with higher subscripts need a printer instead of one or more of these, then they can use the printers that are not being used, since they are connected to all the printers. Now we must show that 1619 cables are not enough. Since there are 1619 cables and 20 printers, the average number of computers per printer is $1619/20$, which is less than 81. Therefore some printer must be connected to fewer than 81 computers (the average of a set of numbers cannot be bigger than each of the numbers in the set). That means it is connected to 80 or fewer computers, so there are at least 20 computers that are not connected to it. If those 20 computers all needed a printer simultaneously, then they would be out of luck, since they are connected to at most the 19 other printers.

41. This problem is similar to Example 10, so we follow the method of solution suggested there. Let a_j be the number of matches held during or before the j^{th} hour. Then $a_1, a_2, \ldots, a_{75}$ is an increasing sequence of distinct positive integers, since there was at least one match held every hour. Furthermore $1 \le a_j \le 125$, since there were only 125 matches altogether. Moreover, $a_1 + 24, a_2 + 24, \ldots, a_{75} + 24$ is also an increasing sequence of distinct positive integers, with $25 \le a_i + 24 \le 149$.

Now the 150 positive integers $a_1, a_2, \ldots, a_{75}, a_1 + 24, a_2 + 24, \ldots, a_{75} + 24$ are all less than or equal to 149. Hence by the pigeonhole principle two of these integers are equal. Since the integers $a_1, a_2, \ldots, a_{75}$ are all distinct, and the integers $a_1 + 24, a_2 + 24, \ldots, a_{75} + 24$ are all distinct, there must be distinct indices i and j such that $a_j = a_i + 24$. This means that exactly 24 matches were held from the beginning of hour $i + 1$ to the end of hour j, precisely the occurrence we wanted to find.

43. This is exactly a restatement of the generalized pigeonhole principle. The pigeonholes are the elements in the codomain (the elements of the set T), and the pigeons are the elements of the domain (the elements of the set S). To say that a pigeon s is in pigeonhole t is just to say that $f(s) = t$. The elements $s_1, s_2, \ldots, s_m$ are just the m pigeons guaranteed by the generalized pigeonhole principle to occupy a common pigeonhole.

45. Let d_j be $jx - N(jx)$, where $N(jx)$ is the integer closest to jx, for $1 \le j \le n$. We want to show that $|d_j| < 1/n$ for some j. Note that each d_j is an irrational number strictly between $-1/2$ and $1/2$ (since jx is irrational and every irrational number is closer than $1/2$ to the nearest integer). The proof is slightly messier if n is odd, so let us assume that n is even. Consider the n intervals

$$\left(0, \frac{1}{n}\right), \left(\frac{1}{n}, \frac{2}{n}\right), \ldots, \left(\frac{(n/2)-1}{n}, \frac{1}{2}\right), \left(-\frac{1}{n}, 0\right), \left(-\frac{2}{n}, -\frac{1}{n}\right), \ldots, \left(-\frac{1}{2}, -\frac{(n/2)-1}{n}\right).$$

The intervals are the pigeonholes and the d_j's are the pigeons. If the interval $(0, 1/n)$ or $(-1/n, 0)$ is occupied, then we are done, since the d_j in that interval tells us which j makes $|d_j| < 1/n$. If not, then there are n pigeons for at most $n-2$ pigeonholes, so by the pigeonhole principle there is some interval, say $((k-1)/n, k/n)$, with two pigeons in it, say d_r and d_s, with $r < s$. Now we will consider $sx - rx$ and show that it is within $1/n$ of its nearest integer; that will complete our proof, since $sx - rx = (s-r)x$, and $s - r$ is a positive integer less than n.

We can write $rx = N(rx) + d_r$ and $sx = N(sx) + d_s$, where $(k-1)/n \le d_r < k/n$ and $(k-1)/n \le d_s < k/n$. Subtracting, we have that $sx - rx = [N(sx) - N(rx)] + [d_s - d_r]$. Now the quantity in the first pair of square brackets is an integer. Furthermore the quantity in the second pair of square brackets is the difference of two numbers in the interval $((k-1)/n, k/n)$ and hence has absolute value less than $1/n$ (the extreme case would be when one of them is very close to $(k-1)/n$ and the other is very close to k/n). Therefore, by definition of "closest integer" $sx - rx$ is at most a distance $1/n$ from its closest integer, i.e., $|(sx - rx) - N(sx - rx)| < 1/n$, as desired. (The case in which n is odd is similar, but we need to extend our intervals slightly past $\pm 1/2$, using $n+1$ intervals rather than n. This is okay, since when we subtract 2 from $n+1$ we still have more pigeons than pigeonholes.)

47. **a)** Assuming that each $i_k \le n$, there are only n pigeonholes (namely $1, 2, \ldots, n$) for the $n^2 + 1$ numbers $i_1, i_2, \ldots, i_{n^2+1}$. Hence, by the generalized pigeonhole principle at least $\lceil (n^2+1)/n \rceil = n+1$ of the numbers are in the same pigeonhole, i.e., equal.

b) If $a_{k_j} < a_{k_{j+1}}$, then the subsequence consisting of a_{k_j} followed by a maximal increasing subsequence of length $i_{k_{j+1}}$ starting at $a_{k_{j+1}}$ contradicts the fact that $i_{k_j} = i_{k_{j+1}}$. Hence $a_{k_j} > a_{k_{j+1}}$.

c) If there is no increasing subsequence of length greater than n, then parts **(a)** and **(b)** apply. Therefore we have $a_{k_{n+1}} > a_{k_n} > \cdots > a_{k_2} > a_{k_1}$, a decreasing subsequence of length $n+1$.

SECTION 6.3 Permutations and Combinations

In this section we look at counting problems more systematically than in Section 6.1. We have some formulae that apply in many instances, and the trick is to recognize the instances. If an ordered arrangement without repetitions is asked for, then the formula for permutations usually applies; if an unordered selection without repetition is asked for, then the formula for combinations usually applies. Of course the product rule and the sum rule (and common sense and cleverness) are still needed to solve some of these problems—having formulae for permutations and combinations definitely does not reduce the solving of counting problems to a mechanical algorithm.

Again the general comments of Section 6.1 apply. Try to solve problems more than one way and come up with the same answer—you will learn from the process of looking at the same problem from two or more different angles, and you will be (almost) sure that your answer is correct.

1. Permutations are ordered arrangements. Thus we need to list all the ordered arrangements of all 3 of these letters. There are 6 such: a,b,c; a,c,b; b,a,c; b,c,a; c,a,b; and c,b,a. Note that we have listed them in alphabetical order. Algorithms for generating permutations and combinations are discussed in Section 6.6.

3. If we want the permutation to end with a, then we may as well forget about the a, and just count the number of permutations of $\{b, c, d, e, f, g\}$. Each permutation of these 6 letters, followed by a, will be a permutation of the desired type, and conversely. Therefore the answer is $P(6,6) = 6! = 720$.