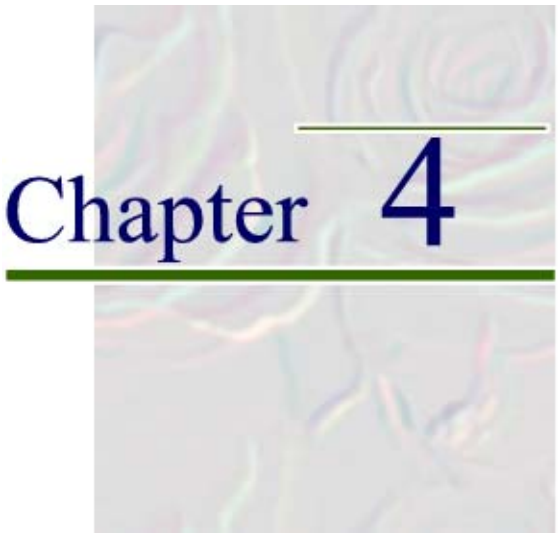


DISCRETE PROBABILITY DISTRIBUTIONS

REVIEW OF KEY CONCEPTS

SECTION 4.1 Random Variable

A **random variable** X is a numerically valued quantity that takes on specific values with different probabilities. The relationship between the values and their associated probabilities is called a **probability mass function**.

Example: Color plates Suppose we administer the color-blindness test 4 times to an individual subject. We want to determine the probability of a specified number of correct selections (k) out of 4 trials if the subject is color blind. The number of correct selections X is a random variable that takes on the values 0, 1, ..., 4 with probabilities p_0, p_1, …, p_4. We can use the *binomial distribution* to obtain the probability mass function (or probability distribution) of the number of correct selections in this case.

SECTION 4.2 Combinations, Permutations, and Factorial

To understand the binomial distribution, we first need to discuss the related concepts of permutations and combinations.

Combinations

A *combination* is defined as the number of ways of selecting k objects out of n, where the order of selection does not matter. A combination is denoted by $\binom{n}{k}$ or ${}_nC_k$ and refers to the number of combinations of n things taken k at a time. It can be shown that

$$ {}_nC_k = \binom{n}{k} = \frac{n(n-1)\times\ldots\times(n-k+1)}{k(k-1)\times\ldots\times 1}. $$

For example, suppose there are 4 readers at a reading center who are qualified to grade a photograph for signs of diabetic retinopathy. Each slide is to be read by 2 readers. The number of ways of selecting the readers is $\binom{4}{2} = \frac{4(3)}{2(1)} = 6$.

Permutations

A *permutation* is defined as the number of ways of selecting k objects out of n, where the order of selection matters. It is denoted by ${}_nP_k$ and refers to the number of permutations of n things taken k at a time. It can be shown that

$$ {}_nP_k = n \times (n-1) \times \ldots \times (n-k+1). $$

For example, suppose subjects in a dietary study of serum cholesterol are to receive different diets in period 1 and 2, respectively, out of 4 potential diets (say A, B, C and D). Since the order of administration of the diets may be important, the number of ways the diets can be assigned to subjects is ${}_4P_2 = 4(3) = 12$.

Factorial

There is a special symbol that is used to denote ${}_kP_k = k(k-1) \times \ldots \times 2 \times 1$ called $k! = k$ *factorial*; e.g., $4! = 4 \times 3 \times 2 \times 1 = 24$; $6! = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$. By convention, $0! \equiv 1$. A combination can also be written in terms of factorials as follows:

$$ {}_nC_k = \frac{n!}{k!(n-k)!}. $$

For example,

$$ {}_4C_2 = \frac{4!}{2!(4-2)!} = \frac{4(3)(2)(1)}{2(1)(2)(1)} = 6. $$

SECTION 4.3 Binomial Probability Distribution

The binomial probability distribution is given by

$$ Pr(X = k) = \binom{n}{k} p^k q^{n-k} \qquad k = 0,\ 1,\ \ldots,\ n. $$

We can use this distribution to model the probability distribution of the number of successes (k) among n trials where each trial has probability of success $= p$, and probability of failure $= q = 1 - p$ and the trials are independent.

We can write the binomial distribution in terms of factorials as follows:

$$ Pr(X = k) = \binom{n}{k} p^k q^{n-k} = \frac{n(n-1)\times\ldots\times(n-k+1)}{k!} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}. $$

For example, what is the probability of obtaining 2 successes in 4 trials in the color plate example? In this case, $n = 4$, $p = 1/4$, and

$$ Pr(X = 2) = \binom{4}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^2 = 6\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^2 = .211. $$

SECTION 4.4 Methods for Using the Binomial Distribution

1. Calculate individual probabilities from the formula in Section 4.3 (brute-force method).
2. Use binomial tables (as given in Table 1 in the Appendix of the text for selected values of n and k).
3. Use electronic tables provided by computer programs such as Excel or MINITAB (see Section 4.5 below).
4. Use Poisson approximation to binomial (explanation on p. 33, Study Guide). Used if $n \geq 100$, $p < 0.01$.
5. Use normal approximation to binomial (explanation on p. 47, Study Guide). Used if $npq \geq 5$.

SECTION 4.5 Using Electronic Tables

For example, if we consider the color-blindness test, suppose we wish to generate all the binomial probabilities for $n = 4$, $p = .25$; i.e.,

$$Pr(X = 0),\ \ldots,\ Pr(X = 4).$$

We can use the BINOMDIST function of Excel to easily evaluate each of these probabilities. This function can both evaluate specific binomial probabilities such as $Pr(X = k)$ as well as cumulative probabilities such as $Pr(X \leq k)$ for any values of n and p. Note that in this case, the probabilities could be obtained from the binomial tables (see Table 1 in the Appendix of the text).

n	k	$p = .25$
4	0	.3164
	1	.4219
	2	.2109
	3	.0469
	4	.0039

To use this function, we have to specify n, p, k as well as a logical argument $\upsilon = \text{TRUE}$ if a cumulative probability is desired and $\upsilon = \text{FALSE}$ if a specific probability is desired. To evaluate $Pr(X = 0)$ or $Pr(X = 4)$ in the above example, we set $\upsilon = \text{FALSE}$ as shown in the spreadsheet below. To evaluate $Pr(X \leq 2)$ we set $\upsilon = \text{TRUE}$ (see spreadsheet). The advantage of the electronic table is that we can evaluate the binomial distribution for any n and p, not just the ones in Table 1 (Appendix, text). More details on using this function are given in the Appendix. A similar function is available in MINITAB under the CALC menu, probability distributions submenu.

n	k		
4	0	0.31640625	BINOMDIST(0, 4, 0.25, FALSE)
	1	0.421875	BINOMDIST(1, 4, 0.25, FALSE)
	2	0.2109375	BINOMDIST(2, 4, 0.25, FALSE)
	3	0.046875	BINOMDIST(3, 4, 0.25, FALSE)
	4	0.00390625	BINOMDIST(4, 4, 0.25, FALSE)
	≤ 2	0.94921875	BINOMDIST(2, 4, 0.25, TRUE)

SECTION 4.6 Expected Value of the Binomial Distribution

In much the same way that we use the sample mean as a measure of location for a sample, we use the "population mean" or expected value (denoted by $E(X)$ or μ) as a measure of location for a random variable.

For a general random variable,

$$E(X) = \sum_{i=1}^{m} x_i \Pr(X = x_i) = \mu$$

that is, the expected value is a weighted average of the m possible values, weighted by their respective probabilities.

For the binomial distribution, the general formula reduces to

$$E(X) = np.$$

For example, suppose we give the color-blindness test once to each of 100 color-blind people.

$$n = 100,\ p = \frac{1}{4},\ E(X) = 25$$

We would expect that 25 out of 100 color-blind people will pass the test.

SECTION 4.7 Variance of the Binomial Distribution

We can also describe the spread of a random variable by its variance which is denoted by $Var(X)$ and is defined by

$$\sigma^2 = Var(X) = \sum_{i=1}^{m} (x_i - \mu)^2 Pr(X = x_i)$$

as opposed to the sample variance s^2. It represents the average squared distance of each possible value from the mean, weighted by its probability of occurrence. It gives us a sense of the spread of the possible values relative to the mean (or expected value). The standard deviation of a random variable is defined by

$$= \sigma = \sqrt{\sigma^2} = \sqrt{Var(X)}.$$

For the binomial distribution, the general formula reduces to

$$\begin{aligned} Var(X) &= npq \\ \text{If } n = 100,\ p &= .01,\ Var(X) = .99 \approx 1 \\ n = 100,\ p &= .99,\ Var(X) = .99 \approx 1 \\ n = 100,\ p &= 1/4,\ Var(X) = 18.75 \\ n = 100,\ p &= 1/2,\ Var(X) = 25. \end{aligned}$$

For a given n, the binomial distribution is most variable for $p = 1/2$ and least variable as p approaches 0 or 1.

SECTION 4.8 Poisson Distribution

Suppose a hospital has observed 1 case of a rare cancer every 2 years over a long time period. It suddenly gets 3 cases in one year. How unlikely is this, or more specifically, what is the probability of obtaining at least 3 cases in 1 year?

The hospital knows that it serves a large number of people but does not know the exact number. The *Poisson* distribution can be used to model the probability distribution of the number of cases over a time period of T years.

Assumptions

1. $Pr(1 \text{ new case in time } \Delta t) \approx \lambda \Delta t$, where $\lambda =$number of cases per unit time, and Δt is a small increment of time.
2. Stationarity; i.e., incidence of new cases per unit time is the same over the entire time period T.
3. Independence; probabilities of events over different time intervals are independent random variables.

Under these assumptions,

$$Pr(k \text{ cases in time } T) = \frac{(\lambda T)^k e^{-\lambda T}}{k!} = \frac{\mu^k e^{-\mu}}{k!}, \; k = 0, 1, \ldots,$$

where $\mu = \lambda T$.

In our case, $T = 1$, $\lambda = 1/2$, so $Pr(k \text{ new cases in 1 year}) = (0.5)^k e^{-0.5}/k!$. We want $Pr(X \geq 3)$. We have

$$Pr(0) = e^{-1/2} = .607$$
$$Pr(1) = .5e^{-1/2} = .303$$
$$Pr(2) = \frac{(0.5)^2}{2} e^{-1/2} = 0.076$$
$$Pr(3+) = 1 - (0.607 + 0.303 + 0.076) = 0.014.$$

It is unlikely that 3 or more cases will occur in 1 year.

SECTION 4.9 Use of Electronic Tables for Poisson Probabilities

We can use the Poisson tables (Table 2, Appendix, text) to obtain probabilities under the Poisson distribution for selected values of μ. The distribution is quantified by a single parameter $\mu = \lambda T$. For values of μ not in the table, we can use an electronic table such as the POISSON function of Excel. With this function we can compute both individual Poisson probabilities by setting the logical operator $\upsilon =$ FALSE as well as cumulative Poisson probabilities by setting $\upsilon =$ TRUE (see Section 4.5 for further discussion of υ). For example, let us use Excel to compute the probability of observing 3+ events in 1 year if the incidence rate is 1 case every 2 years. We have $\lambda = 0.5$, $T = 1$, $\lambda T = \mu = 0.5$:

$$Pr(X \geq 3 | \mu = 0.5) = 1 - Pr(X \leq 2 | \mu = 0.5).$$

The Poisson individual and cumulative probabilities are shown in the spreadsheet below.

μ	k		
0.5	0	0.60653066	POISSON(0, 0.5, FALSE)
	1	0.30326533	POISSON(1, 0.5, FALSE)
	2	0.07581633	POISSON(2, 0.5, FALSE)
	≤ 2	0.98561232	POISSON(2, 0.5, TRUE)
	≥ 3	0.01438768	1-POISSON(2, 0.5, TRUE)

SECTION 4.10 Expected Value and Variance of the Poisson Distribution

$$E(X) = \lambda T = Var(X) = \lambda T = \mu$$

A good way to identify when a distribution is likely to be Poisson is to compare the sample mean and the sample variance. If these quantities are approximately the same, then a Poisson distribution will often fit well.

SECTION 4.11 Poisson Approximation to the Binomial Distribution

If $n \geq 100$, $p \leq 0.01$, we can approximate a binomial distribution with parameters n and p by a Poisson distribution with parameter $\mu = np$. The reason for using this approximation is that it is easier to use the Poisson than the binomial distribution, especially when n is large. Specifically, for a given n, p, and k, we approximate the binomial probability

$$Pr(X = k) = \binom{n}{k} p^k q^{n-k} \text{ by } Pr(X = k) = \frac{e^{-np}(np)^k}{k!}.$$

PROBLEMS

4.1 Evaluate the number of ways of selecting 4 objects out of 10 if the order of selection matters. What term is used to denote this quantity?

4.2 Evaluate the number of ways of selecting 4 objects out of 10 if the order of selection does *not* matter. What term is used to denote this quantity?

Suppose that the probability that a person will develop hypertension over a lifetime is 20%.

4.3 What is the probability distribution of the number of hypertensives over a lifetime among 20 students graduating from the same high school class?

4.4 What is the probability that exactly 4 people out of 50 aged 60–64 will die after receiving the flu vaccine if the probability that 1 person will die is .028?

4.5 What is the probability that at least 4 people will die after receiving the vaccine?

4.6 What is the expected number of deaths following the flu vaccine?

4.7 What is the standard deviation of the number of deaths following the flu vaccine?

Health-Services Administration

The in-hospital mortality rate for 16 clinical conditions at 981 hospitals was recently reported [1]. It was reported that in-hospital mortality was 10.5% for coronary-bypass surgery and 5.0% for total hip replacement. Suppose an institution changes from an academic institution to a private for-profit institution. They find that after the change, of the first 20 patients receiving coronary-bypass surgery, 5 die, while of 20 patients receiving total hip replacement, 4 die.

4.8 What is the probability that of 20 patients receiving coronary-bypass surgery, exactly 5 will die in-hospital, if this hospital is representative of the total pool of 981 hospitals?

4.9 What is the probability of at least 5 deaths among the coronary-bypass patients?

4.10 What is the probability of no more than 5 deaths among the coronary-bypass patients?

4.11 What is the probability that exactly 4 will die among the hip-replacement patients?

4.12 What is the probability that at least 4 will die among the hip-replacement patients?

4.13 What is the probability of 4 or fewer deaths among the hip-replacement patients?

4.14 Can you draw any conclusions based on the results in Problems 4.8–4.13 regarding any effects of the change in hospital administration on in-hospital mortality rates?

Cardiovascular Disease
The rate of myocardial infarction (MI) in 50–59-year-old, disease-free women is approximately 2 per 1000 per year or 10 per 1000 over 5 years. Suppose that 3 MI's are reported over 5 years among 1000 postmenopausal women initially disease free who have been taking postmenopausal hormones.

4.15 Use the binomial distribution to see if this experience represents an unusually small number of events based on the overall rate.

4.16 Answer Problem 4.15 using the Poisson approximation to the binomial distribution.

4.17 Compare your answers in Problems 4.15 and 4.16.

Cardiovascular Disease

4.18 An accepted hypothesis in the etiology of heart disease is that aspirin intake of 325 mg per day reduces subsequent cardiovascular mortality in men with a prior heart attack. Suppose that in a pilot study of 50 women who received 1 tablet per day (325 mg), only 2 die over a 3-year period from cardiovascular disease. How likely is it that not more than 2 women will die if the underlying 3-year mortality rate is 10% in such women?

Pediatrics
A hospital administrator wants to construct a special-care nursery for low-birthweight infants (≤ 2500 g) and wants to have some idea as to the number of beds she should allocate to the nursery. She is willing to assume that the recovery period of each baby is exactly 4 days and thus is interested in the expected number of premature births over the period.

4.19 If the number of premature births in any 4-day period is binomially distributed with parameters $n = 25$ and $p = .1$, then find the probability of 0, 1, 2, …, 7 premature births over this period.

4.20 The administrator wishes to allocate x beds where the probability of having more than x premature births over a 4-day period is less than 5%. What is the smallest value of x that satisfies this criterion?

4.21 Answer Problem 4.20 for 1%.

Cancer
The incidence rate of malignant melanoma is suspected to be increasing over time. To document this increase, a questionnaire was mailed to 100,000 U.S. nurses in 1976 and 1978, asking about any current or previous tumors. Thirty new cases of malignant melanoma were found to have developed over the 2-year period among women with no previous cancers in 1976.

4.22 If the annual incidence rate from previous cancer-registry data is 10 per 100,000, then what is the expected number of new cases over 2 years?

4.23 Are the preceding results consistent or inconsistent with the cancer-registry data? Specifically, what is the probability of observing at least 30 new cases over a 2-year period if the cancer-registry incidence rate is correct?

Accident Epidemiology
Suppose the annual number of traffic fatalities at a given intersection follows a Poisson distribution with parameter $\mu = 10$.

4.24 What is the probability of observing exactly 10 traffic fatalities in 1992?

4.25 What is the probability of observing exactly 25 traffic fatalities over the 2-year period from January 1, 1990, to December 31, 1991?

4.26 Suppose that the traffic intersection is redesigned with better lighting, and 12 traffic fatalities are observed over the next 2 years. Is this rate a meaningful improvement over the previous rate of traffic fatalities?

Pulmonary Disease, Environmental Health
Suppose the number of people seen for violent asthma attacks in the emergency ward of a hospital over a 1-day period is usually Poisson distributed with parameter $\lambda = 1.5$.

4.27 What is the probability of observing 5 or more cases over a 2-day period?

On a particular 2-day period, the air-pollution levels increase dramatically and the distribution of attacks over a 1-day period is now estimated to be Poisson distributed with parameter $\lambda = 3$.

4.28 Answer Problem 4.27 under these assumptions.

4.29 If 10 days out of every year are high-pollution days, then what is the expected number of asthma cases seen in the emergency ward over a 1-year period? (Assume there are 365 days in a year.)

Demography

The data set in Table 4.1 is an example of current life-table data for persons living in the United States in 1986 [2]. Let P_x represent the probability of living for the next year given that one is currently x years old. The entries in the table for age x (referred to as l) are obtained from the formula

$$l_0 = 100{,}000 \quad l_x = l_0 \times P_0 \times P_1 \times \ldots \times P_{x-1}, \quad x = 1, 2, \ldots, 100$$

Assume that the *current* death rates hold not only for the year 1986 but for all subsequent years as well.

4.30 What is the probability of living to age 65 given that one is 21 in 1986? Does this vary by sex and/or race (white vs. black)?

4.31 What is the probability of dying exactly between the ages of 56 and 57 given that a black female is 21 in 1986?

4.32 Suppose 100 white males of age 21 in 1986 live in a particular town and that 5 of them die before reaching the age of 30. Is this event unusual? Specifically, how likely are 5 or more men to die before reaching the age of 30?

Table 4.1 Number of survivors at single years of age, out of 100,000 born alive, by race and sex: United States. 1986

	All races			White			All other					
							Total			Black		
Age	Both sexes	Male	Female	Both sexes	Male	Female	Both sexes	Male	Female	Both sexes	Male	Female
0	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000	100,000
1	98,964	98,845	99,090	99,106	98,998	99,220	98,426	98,262	98,597	98,190	97,996	98,391
2	98,892	98,764	99,028	99,040	98,923	99,164	98,332	98,158	98,513	98,085	97,882	98,296
3	98,838	98,704	98,980	98,991	98,869	99,121	98,256	98,075	98,444	98,000	97,790	98,218
4	98,796	98,658	98,942	98,953	98,826	99,087	98,194	98,007	98,388	97,932	97,716	98,155
5	98,762	98,620	98,912	98,923	98,794	99,060	98,143	97,951	98,342	97,876	97,655	98,104
6	98,733	98,587	98,887	98,897	98,765	99,038	98,100	97,904	98,304	97,830	97,605	98,062
7	98,707	98,557	98,866	98,874	98,738	99,019	98,064	97,863	98,272	97,792	97,563	98,028
8	98,684	98,530	98,847	98,853	98,712	99,002	98,033	97,828	98,246	97,760	97,526	97,999
9	98,663	98,505	98,830	98,834	98,689	98,987	98,006	97,797	98,223	97,732	97,494	97,975
10	98,645	98,484	98,815	98,817	98,669	98,973	97,982	97,769	98,203	97,706	97,464	97,954
11	98,628	98,464	98,801	98,802	98,651	98,960	97,959	97,742	98,184	97,681	97,435	97,934
12	98,610	98,443	98,786	98,785	98,632	98,946	97,935	97,713	98,166	97,655	97,404	97,914
13	98,587	98,415	98,768	98,763	98,606	98,929	97,907	97,677	98,146	97,625	97,365	97,893
14	98,554	98,372	98,745	98,730	98,564	98,906	97,871	97,629	98,123	97,587	97,314	97,869
15	98,507	98,309	98,716	98,683	98,501	98,876	97,825	97,565	98,095	97,538	97,247	97,840
16	98,445	98,223	98,678	98,620	98,414	98,838	97,767	97,484	98,061	97,477	97,161	97,805
17	98,369	98,116	98,633	98,542	98,306	98,792	97,696	97,384	98,021	97,403	97,056	97,763
18	98,280	97,991	98,583	98,452	98,180	98,741	97,612	97,264	97,975	97,315	96,930	97,715
19	98,183	97,852	98,530	98,355	98,041	98,688	97,515	97,123	97,923	97,214	96,781	97,662
20	98,081	97,702	98,477	98,254	97,894	98,635	97,405	96,959	97,867	97,098	96,607	97,603
21	97,974	97,542	98,424	98,150	97,740	98,583	97,281	96,771	97,805	96,967	96,407	97,538
22	97,861	97,372	98,370	98,043	97,578	98,532	97,143	96,560	97,737	96,821	96,181	97,467
23	97,745	97,196	98,316	97,935	97,412	98,482	96,993	96,331	97,664	96,661	95,933	97,390
24	97,628	97,018	98,261	97,827	97,247	98,432	96,834	96,089	97,586	96,490	95,670	97,306
25	97,512	96,843	98,204	97,720	97,086	98,381	96,669	95,839	97,502	96,311	95,395	97,216
26	97,397	96,672	98,147	97,616	96,931	98,330	96,499	95,582	97,413	96,124	95,110	97,119
27	97,283	96,504	98,088	97,514	96,780	98,278	96,322	95,317	97,318	95,927	94,814	97,015
28	97,168	96,336	98,027	97,412	96,632	98,225	96,137	95,043	97,217	95,719	94,503	96,902
29	97,050	96,164	97,964	97,309	96,481	98,171	95,942	94,756	97,107	95,497	94,174	96,778
30	96,927	95,986	97,897	97,202	96,325	98,115	95,735	94,455	96,988	95,258	93,823	96,642
31	96,798	95,800	97,826	97,090	96,162	98,056	95,515	94,139	96,858	95,001	93,450	96,492
32	96,663	95,606	97,751	96,973	95,993	97,994	95,283	93,807	96,718	94,727	93,054	96,329
33	96,522	95,605	97,672	96,852	95,818	97,928	95,038	93,459	96,568	94,436	92,636	96,154
34	96,377	95,199	97,588	96,727	95,639	97,859	94,781	93,093	96,411	94,131	92,196	95,970
35	96,227	94,988	97,500	96,598	95,457	97,786	94,512	92,708	96,247	93,812	91,735	95,779
36	96,072	94,771	97,407	96,465	95,271	97,708	94,230	92,303	96,076	93,480	91,252	95,581
37	95,910	94,546	97,308	96,326	95,079	97,625	93,934	91,877	95,896	93,132	90,745	95,373
38	95,739	94,311	97,202	96,180	94,878	97,534	93,622	91,429	95,705	92,765	90,212	95,152
39	95,557	94,065	97,086	96,024	94,667	97,434	93,291	90,958	95,498	92,373	89,649	94,911
40	95,363	93,805	96,958	95,856	94,443	97,323	92,938	90,462	95,271	91,952	89,052	94,646
41	95,153	93,528	96,816	95,674	94,203	97,199	92,560	89,939	95,022	91,500	88,419	94,353
42	94,927	93,232	96,659	95,476	93,945	97,061	92,156	89,386	94,751	91,015	87,749	94,033
43	94,683	92,916	96,487	95,261	93,668	96,909	91,727	88,806	94,458	90,501	87,046	93,687
44	94,421	92,579	96,300	95,029	93,371	96,743	91,276	88,201	94,145	89,962	86,315	93,318
45	94,139	92,218	96,097	94,778	93,051	96,563	90,803	87,572	93,812	89,400	85,559	92,928

Table 4.1 *Continued*

	All races			White			All other					
							Total			Black		
Age	Both sexes	Male	Female	Both sexes	Male	Female	Both sexes	Male	Female	Both sexes	Male	Female
46	93,836	91,831	95,878	94,507	92,705	96,367	90,307	86,919	93,458	88,816	84,781	92,518
47	93,508	91,413	95,639	94,212	92,329	96,153	89,784	86,237	93,081	88,206	83,974	92,082
48	93,151	90,960	95,377	93,888	91,918	95,917	89,226	85,514	92,673	87,559	83,125	91,614
49	92,758	90,464	95,087	93,531	91,469	95,656	88,620	84,733	92,226	86,662	82,217	91,104
50	92,325	89,919	94,766	93,137	90,973	95,365	87,957	83,881	91,733	86,104	81,237	90,545
51	91,848	89,320	94,409	92,701	90,427	95,042	87,230	82,951	91,189	85,280	80,178	89,931
52	91,324	88,665	94,016	92,221	89,827	94,684	86,440	81,945	90,594	84,389	79,042	89,261
53	90,752	87,949	93,586	91,694	89,167	94,291	85,591	80,869	89,951	83,437	77,836	88,540
54	90,130	87,169	93,122	91,117	88,441	93,864	84,693	79,735	89,268	82,433	76,575	87,774
55	89,458	86,322	92,623	90,488	87,644	93,401	83,749	78,549	88,548	81,382	75,266	86,968
56	88,733	85,405	92,088	89,803	86,773	92,901	82,763	77,316	87,794	80,287	73,915	86,124
57	87,951	84,413	91,513	89,058	85,823	92,360	81,729	76,030	87,000	79,142	72,514	85,236
58	87,103	83,339	90,888	88,247	84,788	91,770	80,630	74,674	86,148	77,930	71,046	84,284
59	86,180	82,173	90,203	87,361	83,661	91,121	79,444	73,227	85,215	76,626	69,487	83,244
60	85,173	80,908	89,449	86,393	82,435	90,406	78,156	71,675	84,185	75,215	67,822	82,097
61	84,077	79,539	88,619	85,338	81,105	89,619	76,757	70,011	83,047	73,687	66,042	80,834
62	82,891	78,065	87,712	84,193	79,669	88,758	75,254	68,243	81,808	72,051	64,157	79,461
63	81,618	76,492	86,731	82,961	78,131	87,823	73,665	66,388	80,485	70,328	62,189	77,999
64	80,264	74,827	85,681	81,645	76,497	86,817	72,016	64,473	79,104	68,548	60,167	76,479
65	78,833	73,076	84,565	80,246	74,770	85,740	70,325	62,516	77,684	66,732	58,113	74,922
66	77,327	71,244	83,381	78,766	72,955	84,590	68,600	60,526	76,232	64,890	56,039	73,337
67	75,740	69,325	82,122	77,199	71,046	83,360	66,834	58,499	74,737	63,015	53,941	71,715
68	74,059	67,305	80,776	75,532	69,027	82,040	65,011	56,423	73,180	61,092	51,805	70,035
69	72,267	65,161	79,330	73,750	66,877	80,618	63,109	54,279	71,534	59,098	49,612	68,270
70	70,353	62,881	77,772	71,841	64,581	79,082	61,111	52,055	69,778	57,018	47,350	66,401
71	68,312	60,462	76,096	69,801	62,139	77,427	59,013	49,749	67,903	54,848	45,018	64,419
72	66,149	57,916	74,299	67,634	59,561	75,649	56,823	47,373	65,916	52,598	42,630	62,332
73	63,872	55,256	72,383	65,347	56,860	73,748	54,557	44,945	63,830	50,284	40,203	60,154
74	61,491	52,503	70,350	62,949	54,057	71,724	52,236	42,489	61,666	47,926	37,763	57,904
75	59,016	49,675	68,200	60,450	51,169	69,577	49,877	40,023	59,437	45,540	35,329	55,597
76	56,452	46,785	65,931	57,853	48,210	67,303	47,486	37,558	57,145	43,133	32,914	53,236
77	53,800	43,842	63,538	55,162	45,191	64,896	45,061	35,097	54,783	40,706	30,522	50,815
78	51,061	40,855	61,012	52,376	42,122	62,350	42,594	32,639	52,339	38,254	28,154	48,324
79	48,235	37,833	58,347	49,498	39,013	59,657	40,078	30,181	49,799	35,773	25,811	45,749
80	45,324	34,789	55,535	46,530	35,879	56,812	37,506	27,723	47,151	33,260	23,495	43,081
81	42,333	31,739	52,570	43,478	32,737	53,809	34,876	25,270	44,387	30,716	21,215	40,314
82	39,269	28,705	49,450	40,350	29,614	50,643	32,192	22,831	41,506	28,147	18,982	37,447
83	36,144	25,712	46,172	37,158	26,528	47,313	29,461	20,419	38,509	25,564	16,812	34,485
84	32,972	22,791	42,736	33,916	23,520	43,817	26,695	18,053	35,405	22,982	14,728	31,435
85	29,771	19,977	39,143	30,642	20,625	40,155	23,912	15,755	32,206	20,419	12,755	28,312

4.33 Answer Problem 4.32 for a black male.

4.34 Suppose we are not willing to assume that the P_x's remain constant in years subsequent to 1986. Can Problems 4.30–4.33 still be answered? If not, what additional information is needed?

Pulmonary Disease

Each year approximately 4% of current smokers attempt to quit smoking, and 50% of those who try to quit are successful in the sense that they abstain from smoking for at least 1 year from the date they quit.

4.35 What is the probability that a current smoker will quit for at least 1 year?

4.36 What is the probability that among 100 current smokers, at least 5 will quit smoking for at least 1 year?

An educational program was conducted among smokers who attempt to quit to maximize the likelihood that such individuals would continue to abstain for the long term.

4.37 Suppose that of 20 people who enter the program when they first stop smoking, 15 still abstain from smoking 1 year later. Can the program be considered successful?

Cancer, Epidemiology

A frequent design for biomedical investigations is the case-control study. A group of **cases** is identified on the basis of having a particular disease (e.g., lung-cancer patients in a cancer registry), and a group of **controls** is chosen (e.g., patients in the same registry with cancer of the esophagus) such that every case is *matched* with 1 or more controls. That is, the case and control(s) are matched in every sense except that the controls do not have the disease trait. We can then look at whether or not some other factor (such as smoking) is associated with the disease trait. Obtaining "exact" matches is usually impossible, and several characteristics are selected to use for potential matches, such as age and sex. Suppose the group of cases and controls is as given in Table 4.2.

Table 4.2 Selection of controls for a case-control study

Cases			Controls			
Case number	Age	Sex	Control number	Age group	Sex	Frequency
1	36	M	01–15	21–30	M	15
2	50	F	16–21	21–30	F	6
3	24	F	22–27	31–40	M	6
4	22	M	28–45	31–40	F	18
5	35	M	46–48	41–50	M	3
			49–54	41–50	F	6
			55–66	51–60	M	12
			67–69	51–60	F	3
			70–78	61–70	M	9
			79–84	61–70	F	6

Suppose that the match is performed so that each control has the same age group and sex as its corresponding case. An example of a 1-to-1 match would be

Case number	1	2	3	4	5
Control number	24	51	18	14	22

An example of a 2-to-1 match would be

Case number	1	2	3	4	5
Control number	24, 26	51, 49	18, 21	14, 06	22, 27

Assume that the order within each group of matched controls in a many-to-one matching does not matter.

4.38 How many ways can 1-to-1 matches be assigned for age and sex?

4.39 Suppose the designers of the study get lazy and match only for age. How many ways can the matching be done?

4.40 If the groups were matched only on age, then what is the probability that the groups will be matched for sex as well if each match is equally likely?

4.41 Answer Problem 4.38 for 2-to-1 matches.

4.42 Answer Problem 4.39 for 2-to-1 matches.

4.43 Answer Problem 4.40 for 2-to-1 matches.

4.44 Answer Problem 4.38 for 3-to-1 matches.

4.45 Answer Problem 4.39 for 3-to-1 matches.

4.46 Answer Problem 4.40 for 3-to-1 matches.

Infectious Disease

An outbreak of acute gastroenteritis occurred at a nursing home in Baltimore, Maryland, in December 1980 [3]. A total of 46 out of 98 residents of the nursing home became ill. People living in the nursing home shared rooms: 13 rooms contained 2 occupants, 4 rooms contained 3 occupants, and 15 rooms contained 4 occupants. One question that arises is whether or not a geographical clustering of disease occurred for persons living in the same room.

A summary of the number of affected people and the total number of people in a room is given in Table 4.3.

4.47 If the binomial distribution holds, what is the probability distribution of the number of affected people in rooms with 2 occupants? That is, what is the probability of finding zero affected people? One affected person? Two affected people?

4.48 Answer Problem 4.47 for the probability distribution of the number of affected people in rooms with 3 occupants.

4.49 Answer Problem 4.47 for the probability distribution of the number of affected people in rooms with 4 occupants.

Table 4.3 Number of affected people and total number of people in a room for an outbreak of acute gastroenteritis in a nursing home in Baltimore, Maryland

People in room	Total number of rooms	Number of rooms with 0 affected people	1 affected person	2 affected people	3 affected people	4 affected people
2	13	5	4	4	0	0
3	4	1	2	0	1	0
4	15	2	4	3	5	1

Source: Reprinted with permission from the *American Journal of Epidemiology*, *116*(6), 940–948, 1982.

4.50 One useful summary measure of geographical clustering is the number of rooms with 2 or more affected occupants. If the binomial distribution holds, what is the expected number of rooms with 2 or more affected occupants over the entire nursing home?

4.51 Compare the observed number of rooms with 2 or more affected occupants with the expected number of rooms. Does this comparison give any evidence that clustering of disease occurs within rooms?

Cancer
The incidence rate of malignant melanoma in women ages 35–59 is approximately 11 new cases per 100,000 women per year. A study is planned to follow 10,000 women with excessive exposure to sunlight.

4.52 What is the expected number of cases among 10,000 women over 4 years? (Assume no excess risk due to sunlight exposure.)

4.53 Suppose that 9 new cases are observed in this period. How unusual a finding is this?

Cancer
A study was performed in Woburn, MA, looking at the rate of leukemia among children (≤ age 19) in the community in comparison to statewide leukemia rates. Suppose there arc 12,000 children in the community that have lived there for a 10-year period and 12 leukemias have occurred in 10 years.

4.54 If the statewide incidence rate of leukemia in children is 5 events per 100,000 children per year (i.e., per 100,000 person-years) then how many leukemias would be expected in Woburn over the 10-year period if the statewide rates were applicable?

4.55 What is the probability of obtaining exactly 12 events over a 10-year period if statewide incidence rates were applicable?

4.56 What is the probability of obtaining at least 12 events over the 10-year period if the statewide incidence rates were applicable?

4.57 How do you interpret the results in Problem 4.56?

SOLUTIONS

4.1 $10 \times 9 \times 8 \times 7 = 5040$. This is referred to as ${}_{10}P_4$ or the number of permutations of 10 things taken 4 at a time.

4.2 $\dfrac{10 \times 9 \times 8 \times 7}{4 \times 3 \times 2 \times 1} = 210$

This is referred to as ${}_{10}C_4$ or $\binom{10}{4}$ = the number of combinations of 10 things taken 4 at a time.

4.3 X = number of hypertensives over a lifetime is binomially distributed with parameters $n = 20$, $p = .2$; that is,

$$Pr(X = x) = {}_{20}C_x(.2)^x(.8)^{20-x} \quad x = 0, 1, \ldots, 20$$

or, from Table 4.4:

Table 4.4

x	$Pr(X = x)$	x	$Pr(X = x)$
0	.0115	7	.0546
1	.0576	8	.0222
2	.1369	9	.0074
3	.2054	10	.0020
4	.2182	11	.0005
5	.1746	12	.0001
6	.1091	13–20	.0000

4.4 We wish to compute

$$Pr(X = 4) = {}_{50}C_4(.028)^4(.972)^{46} = .038.$$

4.5 $Pr(X \geq 4) = 1 - Pr(X \leq 3)$. We use the BINOMDIST function of Excel to evaluate $Pr(X \leq 3)$ and obtain:

$$Pr(X \leq 3) = .949.$$

Therefore,

$$Pr(X \geq 4) = 1 - .949 = .051.$$

4.6 $E(X) = 50 \times .028 = 1.4 = \mu$

4.7 We have

$$Var(X) = npq = 50 \times .028 \times .972 = 1.361 = \sigma^2.$$

Thus, $\sigma = \sqrt{1.361} = 1.17$.

4.8 We have a binomial distribution with $n = 20$, $p = .105$. Let $X =$ number of in-hospital deaths after coronary-bypass surgery. We have that

$$\begin{aligned} Pr(X = 5) &= {}_{20}C_5(.105)^5(1 - .105)^{15} \\ &= \frac{20 \times 19 \times 18 \times 17 \times 16}{5 \times 4 \times 3 \times 2 \times 1}(.105)^5(.895)^{15} \\ &= 15{,}504(1.2763 \times 10^{-5})(.1894) = .037 \end{aligned}$$

4.9 We must compute $Pr(X \geq 5) = 1 - Pr(X \leq 4)$. We will use the BINOMDIST function of Excel as follows:

$$Pr\left(X \leq 4 \middle| n = 20,\ p = .105\right) = .9484.$$

Thus, $Pr(X \geq 5) = 1 - .9484 = .052$.

4.10 Compute $Pr(X \leq 5) = .9484 + .0375 = .986$.

4.11 We have a binomial distribution with $n = 20$, $p = .05$. Let $Y =$ number of in-hospital deaths after hip replacement. To compute

$$Pr(Y = 4) = {}_{20}C_4(.05)^4(.95)^{16},$$

refer to the binomial tables (Table 1 in the Appendix of the text) under $n = 20$, $p = .05$, $k = 4$ and obtain $.0133 \cong .013$.

4.12 From the binomial tables under $n = 20$, $p = .05$,

$$\begin{aligned} Pr\left(Y \geq 4\right) &= 1 - Pr\left(Y \leq 3\right) \\ &= 1 - \left(.3585 + .3774 + .1887 + .0596\right) \\ &= 1 - .984 = .016. \end{aligned}$$

4.13 From the answers to Problems 4.11 and 4.12,

$$\begin{aligned} Pr\left(Y \leq 4\right) &= Pr\left(Y \leq 3\right) + Pr\left(Y = 4\right) \\ &= .984 + .013 = .997. \end{aligned}$$

4.14 From Problem 4.9, $Pr(X \geq 5) = .052$, and from Problem 4.12, $Pr(Y \geq 4) = .016$. Therefore, there is a 5.2% probability of observing at least 5 coronary-bypass in-hospital deaths if the old underlying rate of .105 holds and there is a 1.6% probability of observing at least 4 in-hospital deaths after hip replacement if the old underlying rate of .05 holds. Since .05 is the traditional cutoff for unusual probabilities, we would conclude that the in-hospital mortality from hip replacement has increased, while the results for coronary-bypass surgery suggest a similar trend.

4.15 The observed number of events = 3, which is less than the expected number of events

$$1000 \times \frac{10}{1000} = 10.0.$$

We would expect that the number of MI's is binomially distributed with parameters $n = 1000$ and

$$p = \frac{10}{1000} = .01.$$

Thus, we wish to evaluate

$$Pr(X \leq 3) = \sum_{k=0}^{3} {}_{1000}C_k(.01)^k(.99)^{1000-k}.$$

We use the BINOMDIST function of Excel as follows: $Pr(X \leq 3 | n = 1000,\ p = .01) = .010$. Since this probability is low (i.e., $< .05$), this represents an unusually small number of events in this subgroup.

4.16 Since $n = 1000 \geq 100$ and $p = .01$, we can approximate the binomial distribution by a Poisson distribution with parameter $\mu = np = 1000 \times 0.01 = 10$. We refer to the exact Poisson tables (Table 2 in the Appendix of the text) under $\mu = 10$ and note that

$$\begin{aligned} &Pr(X = 0) = .0000,\ Pr(X = 1) = .0005, \\ &Pr(X = 2) = .0023,\ Pr(X = 3) = .0076. \end{aligned}$$

Therefore,

$$Pr(X \leq 3) = .0000 + .0005 + .0023 + .0076 = .010.$$

4.17 The two answers agree to three decimal places.

4.18 The probability that exactly 2 out of 50 women will die over a 3-year period is given by

$$\binom{50}{2}(.1)^2(.9)^{48}$$

The probability that not more than two women will die out of $50 = Pr(X \leq 2)$ is given by

$$\sum_{k=0}^{2}\binom{50}{k}(.1)^k(.9)^{50-k}.$$

This expression can be evaluated from Excel. Thus, $Pr(X \le 2|n = 50,\ p = .1) = .112$.

4.19 We use the BINOMDIST function of Excel with $n = 25$, $p = .1$ and $k = 0, 1, 2, \ldots, 7$ and obtain:

$$\begin{aligned} Pr(0) &= .0718 \\ Pr(1) &= .1994 \\ Pr(2) &= .2659 \\ Pr(3) &= .2265 \\ Pr(4) &= .1384 \\ Pr(5) &= .0646 \\ Pr(6) &= .0239 \\ Pr(7) &= .0072 \end{aligned}$$

4.20 We see from the results of Problem 4.19 that

$$\begin{aligned} Pr(X > 0) &= 1 - Pr(X \le 0) \\ &= 1 - .0718 = .928 \\ Pr(X > 1) &= Pr(X \ge 2) \\ &= 1 - (.0718 + .1994) = .729 \\ Pr(X > 2) &= Pr(X \ge 3) \\ &= 1 - (.0718 + .1994 + .2659) = .463 \\ Pr(X > 3) &= Pr(X \ge 4) \\ &= 1 - (.0718 + \ldots + .2265) = .236 \\ Pr(X > 4) &= Pr(X \ge 5) \\ &= 1 - (.0718 + \ldots + .1384) = .098 \\ Pr(X > 5) &= Pr(X \ge 6) \\ &= 1 - (.0718 + \ldots + .0646) = .033 \end{aligned}$$

Therefore x should be 5, since this is the smallest number of beds such that the probability of having more than this number of premature births in a 4-day period is less than .05.

4.21 If we continue with the method used in the solution of Problem 4.20, then we have

$$Pr(X > 6) = Pr(X \ge 7) = .0095.$$

Therefore, in order to lower the probability to less than .01, we need 6 beds.

4.22 The expected number of new cases

$$100{,}000 \times \frac{10}{100{,}000} \times 2 = 20.$$

4.23 The number of new cases of malignant melanoma is binomially distributed with parameters $n = 100{,}000$ and $p = (10/100{,}000) \times 2 = 20/100{,}000$. We will approximate this distribution by a Poisson distribution with parameter $\mu = np = 20$. We wish to compute

$$Pr(X \ge 30|\mu = 20).$$

We refer to Table 2 (see Appendix in text) under $\mu = 20$ to obtain

$$Pr(X \ge 30) = .0083 + .0054 + \ldots + .0001 = .022.$$

These results are inconsistent with the cancer-registry data and suggest that the rate of malignant melanoma has increased.

4.24 From Table 2 (Appendix, text),

$$Pr(X = 10|\mu = 10) = .1251.$$

4.25 We have $\mu = 2 \times 10 = 20$. Referring to Table 2 (Appendix, text), we note that

$$Pr(X = 25|\mu = 20) = .0446.$$

4.26 We wish to compute $Pr(X \le 12|\mu = 20)$. From Table 2 (Appendix, text) we obtain

$$Pr(X \le 12) = .0000 + \ldots + .0176 = .039.$$

Since this probability is $< .05$, this represents a meaningful improvement over the previous rate of traffic fatalities.

4.27 We have that $\lambda = 1.5$, $T = 2$, and thus, $\mu = \lambda T = 3.0$. We wish to compute $Pr(X \ge 5) = 1 - Pr(X \le 4)$. We refer to Table 2 (Appendix, text) under $\mu = 3$ and obtain $Pr(X \ge 5) = 1 - (.0498 + \ldots + .1680) = .185$.

4.28 We have that $\lambda = 3$, $T = 2$, and thus, $\mu = \lambda T = 6.0$. Referring to Table 2 (Appendix, text), we find that

$$\begin{aligned} Pr(X \ge 5) &= 1 - Pr(X \le 4) \\ &= 1 - (.0025 + \ldots + .1339) = .715. \end{aligned}$$

4.29 Since the expected number of cases on a high-pollution day is 3 and on a low-pollution day is 1.5, it follows that the expected number of cases in one year $= 10(3) + 355(1.5) = 562.5$.

4.30 The probability of living to age 65 given that one is age 21 is given by

$$Pr = \frac{l_{22}}{l_{21}} \times \frac{l_{23}}{l_{22}} \times \ldots \times \frac{l_{64}}{l_{63}} \times \frac{l_{65}}{l_{64}} = \frac{l_{65}}{l_{21}}$$

We will compute these probabilities separately for white males, white females, black males, and black females. The results are given in Table 4.5.

The probabilities vary substantially by sex and race, with the highest probability for white females (.870) and the lowest probability for black males (.603).

Table 4.5 Probability of living to age 65 given that one is alive at age 21, by sex and race

	l_{65}	l_{21}	l_{65}/l_{21} = probability
White male	74,770	97,740	.765
White female	85,740	98,583	.870
Black male	58,113	96,407	.603
Black female	74,922	97,538	.768

4.31 The probability of dying exactly between the ages of 56 and 57 given that one is age 21 in 1986 = the probability that one lives to age 56 and then dies before age 57:

$$\frac{l_{56}}{l_{21}} \times \left(1 - \frac{l_{57}}{l_{56}}\right) = \frac{86{,}124}{97{,}538} \times \left(1 - \frac{85{,}236}{86{,}124}\right)$$
$$= .8830 \times (1 - .9897)$$
$$= .8830 \times .0103 = .009.$$

4.32 First compute the probability of a single white male dying before age 30 given that he is age 21 in 1986. This probability is given by

$$Pr = 1 - \frac{l_{30}}{l_{21}} = 1 - \frac{96{,}325}{97{,}740} = 1 - .9855 = .0145.$$

We are now interested in the distribution of the number of deaths among 100 men, which follows a binomial distribution with parameters $n = 100$, $p = .0145$. Thus, the probability of k deaths is given by

$$Pr(X = k) = \binom{100}{k}(.0145)^k(.9855)^{100-k}$$
$$k = 0,\ 1,\ 2,\ \ldots,\ 100$$

We specifically want $Pr(X \geq 5) = 1 - Pr(X \leq 4)$. Use the BINOMDIST function of Excel. Thus,

$$\Pr\left(X \leq 4 | n{=}100,\ p{=}.0145\right) = .9846$$
$$\Pr\left(X \geq 5\right) = 1 - .9846 = .0154.$$

This is somewhat unusual.

4.33 The probability of a single black male dying before age 30 given that he is 21 in 1986 is

$$Pr = 1 - \frac{l_{30}}{l_{21}} = 1 - \frac{93{,}823}{96{,}407} = 1 - .9732 = .0268.$$

If $X =$ number of deaths among 100 black males, then $Pr(X \geq 5) = 1 - Pr(X \leq 4)$. We use the BINOMDIST function of Excel. Thus,

$$Pr(X \leq 4 | n = 100,\ p = .0268) = .8685$$
$$Pr(X \geq 5) = 1 - .8685 = .1315.$$

Therefore, this is a much more common event for black males than for white males.

4.34 The questions *cannot* be answered if we do not assume that the P_x's remain constant. Current life tables for each succeeding year after 1986 would be needed. For example, in computing the probability that a man will live to age 23 given that he is age 21 in 1986, the probability that he will reach age 22 in 1987 must be multiplied by the probability that he will reach age 23 in 1988 given that he was age 22 in 1987. We must assume that the latter probability is the same as the probability that he will reach age 23 in 1987 given that he was age 22 in 1986, which is not necessarily the case. This assumption is especially risky if events happening over a long time period are being considered, because death rates have tended to go down. The probability that 65-year-old people will live 1 extra year given that they were 65 in the year 2051 is likely to be quite different from the corresponding probability if they were 65 in the year 1986. Life tables that follow the *same* group of people over time are called *cohort* life tables as opposed to the *current* life-table data given in this example. These cohort life tables are especially useful in providing baseline mortality data for epidemiologic studies of mortality in high-risk groups.

4.35 .02

4.36 We wish to compute $Pr(X \geq 5) = 1 - Pr(X \leq 4)$, where X is binomially distributed with parameters $n = 100$, $p = .02$. We use the BINOMDIST function of Excel as follows:

$$Pr(X \leq 4 | n = 100,\ p = .02) = .949.$$

Therefore, $Pr(X \geq 5) = 1 - .949 = .051$.

4.37 We wish to compute $Pr(X \geq 15)$ where X is binomially distributed with parameters $n = 20$, $p = .50$. We refer to Table 1 (Appendix, text) and find that

$$Pr(X \geq 15) = 0.0148 + \ldots + 0.0000 = .021.$$

Since this probability is low ($< .05$), we can consider the program a success (i.e., the true success rate is probably better than 50%).

4.38 The number of ways is $6 \times 6 \times 6 \times 15 \times 5 = 16{,}200$.

4.39 The number of ways is

$$24 \times 9 \times 21 \times 20 \times 23 = 2{,}086{,}560.$$

4.40 We have

$$Pr(\text{age - sex match}|\text{age match}) = \frac{\text{number of age - sex matches}}{\text{number of age matches}}$$
$$= \frac{16{,}200}{2{,}086{,}560} = .0078.$$

4.41 The number of ways is

$${}_6C_2 \times {}_6C_2 \times {}_6C_2 \times {}_{15}C_2 \times {}_4C_2 = 15 \times 15 \times 15 \times 105 \times 6$$
$$= 2{,}126{,}250 = 2.13 \times 10^6.$$

4.42 The number of ways is

$${}_{24}C_2 \times {}_9C_2 \times {}_{21}C_2 \times {}_{19}C_2 \times {}_{22}C_2 = 276 \times 36 \times 210 \times 171 \times 231$$
$$= 8.24 \times 10^{10}.$$

4.43 The probability $= 2.13 \times 10^6 / (8.24 \times 10^{10}) = 2.58 \times 10^{-5}$.

4.44 The number of ways is

$${}_6C_3 \times {}_6C_3 \times {}_6C_3 \times {}_{15}C_3 \times {}_3C_3 = 20 \times 20 \times 20 \times 455 \times 1$$
$$= 3.64 \times 10^6.$$

4.45 The number of ways is

$${}_{24}C_3 \times {}_9C_3 \times {}_{21}C_3 \times {}_{18}C_3 \times {}_{21}C_3 = 2024 \times 84 \times 1330 \times 816 \times 1330$$
$$= 2.45 \times 10^{14}.$$

4.46 The probability $= 3.64 \times 10^6 / (2.45 \times 10^{14}) = 1.48 \times 10^{-8}$.

4.47 Overall, the probability of disease $= 46/98 = .4694$. We use a binomial distribution with parameters $n = 2$, $p = .4694$. We have

$$Pr(X = 0) = (.5306)^2 = .282$$
$$Pr(X = 1) = {}_2C_1(.4694)(.5306) = .498$$
$$Pr(X = 2) = (.4694)^2 = .220$$

4.48 We use a binomial distribution with parameters $n = 3$, $p = .4694$. We have

$$Pr(X = 0) = (.5306)^3 = .149$$
$$Pr(X = 1) = {}_3C_1(.4694)(.5306)^2 = .396$$
$$Pr(X = 2) = {}_3C_2(.4694)^2(.5306) = .351$$
$$Pr(X = 3) = (.4694)^3 = .103$$

4.49 We use a binomial distribution with parameters $n = 4$, $p = .4694$. We have

$$Pr(X = 0) = (.5306)^4 = .079$$
$$Pr(X = 1) = {}_4C_1(.4694)(.5306)^3 = .280$$
$$Pr(X = 2) = {}_4C_2(.4694)^2(.5306)^2 = .372$$
$$Pr(X = 3) = {}_4C_3(.4694)^3(.5306) = .219$$
$$Pr(X = 4) = (.4694)^4 = .049$$

4.50 The expected number is obtained as follows:

$$\mu = 13 \times Pr(X \geq 2|n = 2) + 4 \times Pr(X \geq 2|n = 3)$$
$$+15 \times Pr(X \geq 2|n = 4)$$
$$= 13 \times .220 + 4 \times (.351 + .103)$$
$$+15 \times (.372 + .219 + .049)$$
$$= 2.86 + 1.82 + 9.60 = 14.3.$$

4.51 From Table 4.3, there are $4 + 1 + 9 = 14$ rooms with 2 or more affected people. This is virtually identical to the expected number given in the solution to Problem 4.50 (14.3) and gives little evidence of clustering. We discuss the issue of testing for goodness-of-fit of probability models in more detail in Chapter 10 of the text.

4.52 The expected number of cases

$$4 \times \frac{11}{100{,}000} \times 10{,}000 = 4.4.$$

4.53 We have a binomial distribution X with $n = 10{,}000$ and $p = 44/100{,}000$. We will approximate this distribution by a Poisson distribution Y with parameter

$$\mu = np = 4.4.$$

To compute $Pr(Y \geq 9|\mu = 4.4)$, we use the POISSON function of Excel as follows:

$$Pr(Y \leq 8|\mu = 4.4) = .9642$$

Therefore,

$$Pr(Y \geq 9) = 1 - Pr(Y \leq 8) = 1 - .9642 = .036.$$

Because .036 is less than .05, this is a somewhat unusual finding.

4.54 There were $12{,}000 \times 10 = 120{,}000$ person-years. Therefore, we would expect $(5/100{,}000) \times 120{,}000 = 6$ events over 10 years.

4.55 From the Poisson table (Table 2, Appendix, text)

$$Pr(X = 12|\mu = 6) = .0113.$$

4.56 Let $X =$ number of events over 10 years. We will assume X is a Poisson random variable with $\mu = 6$ and wish to compute $Pr(X \geq 12)$. From the Poisson table (Table 2, Appendix, text) we have that

$$\begin{aligned} Pr(X \geq 12 | \mu = 6) &= .0113 + .0052 + .0022 \\ &\quad + .0009 + .0003 + .0001 \\ &= .020. \end{aligned}$$

4.57 Since the probability of obtaining 12 or more events is $< .05$, it follows that $\mu = 6$ is an unlikely value and that the true mean is likely to be higher than 6; i.e., there is an excess risk of leukemia in Woburn.

REFERENCES

[1] Shortell, S. M., & Hughes, E. F. X. (1988). The effects of regulation, competition, and ownership on mortality rates among hospital inpatients. *New England Journal of Medicine*, **318**(17), 1100–1107.

[2] U.S. Department of Health and Human Services. (1986). *Vital statistics of the United States, 1986*. Washington, DC: U.S. Department of Health and Human Services.

[3] Kaplan, J. E., Schonberger, L. B., Varano, G., Jackman, N., Bied, J., & Gary, G. W. (1982). An outbreak of acute nonbacterial gastroenteritis in a nursing home: Demonstration of person-to-person transmission by temporal clustering of cases. *American Journal of Epidemiology*, **116**(6), 940–948.