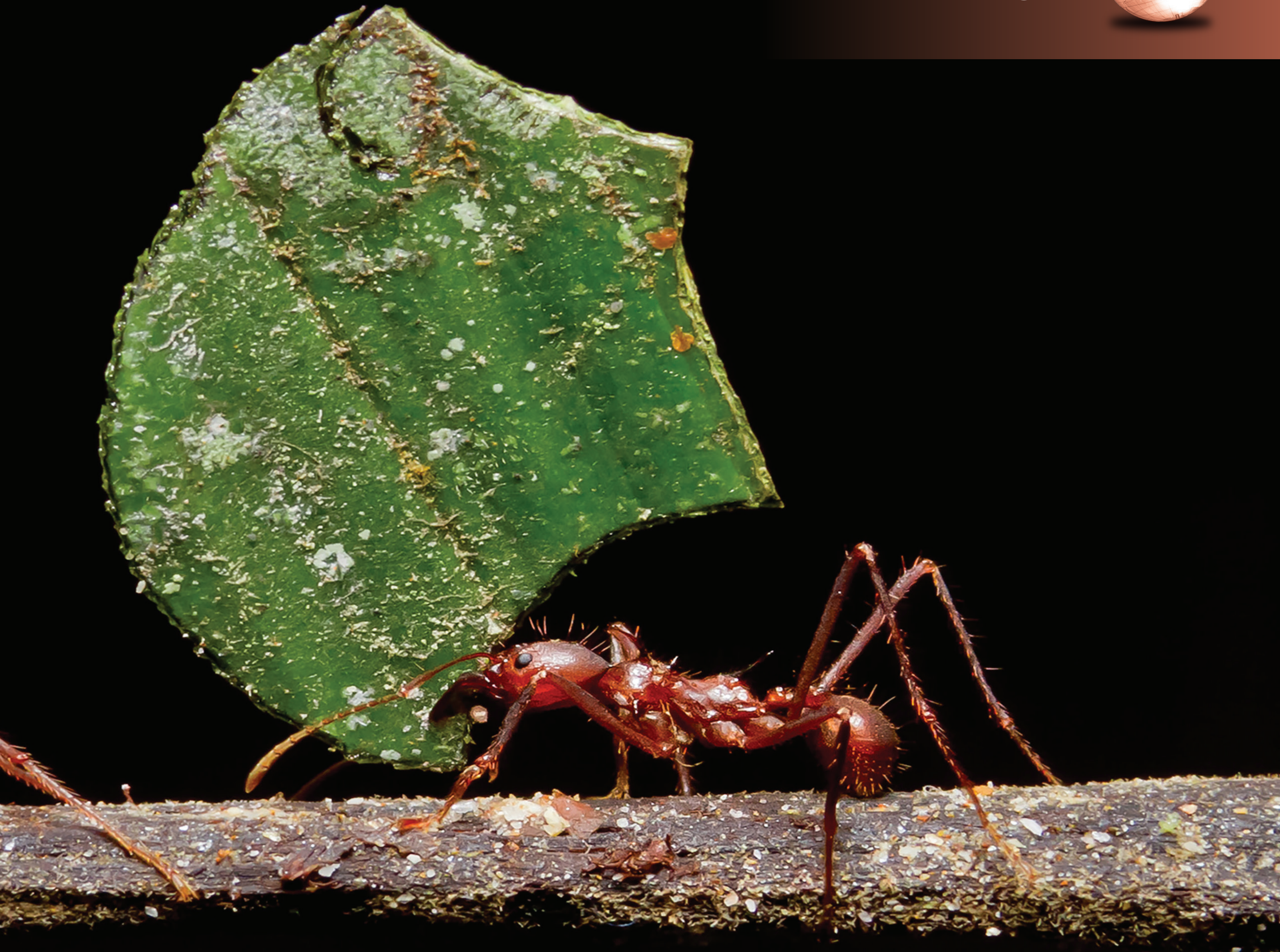
GLOBAL EDITION



Biostatistics

for the Biological and Health Sciences

THIRD EDITION

Marc M. Triola • Mario F. Triola • Jason Roy

Symbol Table

f	frequency with which a value occurs
Σ	capital sigma; summation
Σx	sum of the values
Σx^2	sum of the squares of the values
$(\Sigma x)^2$	square of the sum of all values
Σxy	sum of the products of each x value multiplied by the corresponding y value
n	number of values in a sample
N	number of values in a finite population; also used as the size of all samples combined
$n!$	n factorial
k	number of samples or populations or categories
$\bar{x}$	mean of the values in a sample
$\bar{R}$	mean of the sample ranges
μ	mu; mean of all values in a population
s	standard deviation of a set of sample values
σ	lowercase sigma; standard deviation of all values in a population
s^2	variance of a set of sample values
σ^2	variance of all values in a population
z	standard score
$z_{\alpha/2}$	critical value of z
t	t distribution
$t_{\alpha/2}$	critical value of t
df	number of degrees of freedom
F	F distribution
χ^2	chi-square distribution
χ^2_R	right-tailed critical value of chi-square
χ^2_L	left-tailed critical value of chi-square
p	probability of an event or the population proportion
q	probability or proportion equal to $1 - p$
$\hat{p}$	sample proportion
$\hat{q}$	sample proportion equal to $1 - \hat{p}$
$\bar{p}$	proportion obtained by pooling two samples
$\bar{q}$	proportion or probability equal to $1 - \bar{p}$
$P(A)$	probability of event A
$P(A\|B)$	probability of event A, assuming event B has occurred
${}_nP_r$	number of permutations of n items selected r at a time
${}_nC_r$	number of combinations of n items selected r at a time
$\bar{A}$	complement of event A
H_0	null hypothesis
H_1	alternative hypothesis
α	alpha; probability of a type I error or the area of the critical region
β	beta; probability of a type II error
r	sample linear correlation coefficient
ρ	rho; population linear correlation coefficient
r^2	coefficient of determination
R^2	multiple coefficient of determination
r_s	Spearman's rank correlation coefficient
b_1	point estimate of the slope of the regression line
b_0	point estimate of the y-intercept of the regression line
$\hat{y}$	predicted value of y
d	difference between two matched values
$\bar{d}$	mean of the differences d found from matched sample data
s_d	standard deviation of the differences d found from matched sample data
s_e	standard error of estimate
T	rank sum; used in the Wilcoxon signed-ranks test

continued

Each of those 10 different arrangements has a probability of $0.75^3 \cdot 0.25^2$, so the total probability is

$$P(3 \text{ peas with green pods among } 5) = \frac{5!}{(5-3)!3!} \cdot 0.75^3 \cdot 0.25^2$$

This particular result can be generalized as the binomial probability formula (Formula 5-5). That is, the binomial probability formula is a combination of the multiplication rule of probability and the counting rule for the number of arrangements of n items when x of them are identical to each other and the other $n - x$ are identical to each other.

$$P(x) = \overbrace{\frac{n!}{(n-x)!x!}}^{\substack{\text{The number of outcomes with} \\ \text{exactly } x \text{ successes among } n \text{ trials}}} \cdot \overbrace{p^x \cdot q^{n-x}}^{\substack{\text{The probability of } x \text{ successes among} \\ n \text{ trials for any one particular order}}}$$

TECH CENTER

Binomial Distributions

Visit **www.pearsonglobaleditions.com** for technology procedures for StatCrunch, Excel, Minitab, Statdisk, TI-83/84 Plus Calculator, and R. Tech supplements, videos, and downloadable data sets for this text are also available at this website.

5-2 Basic Skills and Concepts

Statistical Literacy and Critical Thinking

1. Hybridization Assume that 75% of offspring peas have green pods. Suppose we want to find the probability that when five offspring peas are randomly selected, exactly two of them are green. What is wrong with using the multiplication rule to find the probability of getting two peas with green pods followed by three peas with yellow pods: $(0.75)(0.75)(0.25)(0.25)(0.25) = 0.00879$?

2. Variation and Notation Assume that we want to find the probability that among five offspring peas, exactly two have green pods. Also assume that 75% of offspring peas have green pods (and the others have yellow pods).

a. Identify the values of *n, x, p,* and *q*.

b. For groups of five randomly selected offspring peas, find the mean, standard deviation, and variance for the numbers of peas among five that have green pods. Include appropriate units.

3. Independent Events Based on a KRC Research survey, when 1020 adults were asked about hand hygiene, 44% said that they wash their hands after using public transportation. Consider the probability that among 30 different adults randomly selected from the 1020 who were surveyed, at least 10 wash their hands after using public transportation. Given that these subjects were selected without replacement, are the 30 selections independent? Can they be treated as being independent? Can the probability be found using the binomial probability formula?

4. Notation of 0+ Using the same survey from Exercise 3, the probability of randomly selecting 30 of the 1020 adults and getting exactly 24 who wash their hands after using public transportation is represented as 0+. What does 0+ indicate? Does 0+ indicate that it is impossible to get exactly 24 adults who wash their hands after using public transportation?

Identifying Binomial Distributions. ***In Exercises 5–12, determine whether the given procedure results in a binomial distribution (or a distribution that can be treated as binomial). For those that are not binomial, identify at least one requirement that is not satisfied.***

5. Clinical Trial of YSORT The YSORT method of sex selection, developed by the Genetics & IVF Institute, was designed to increase the likelihood that a baby will be male. When 291 couples used the YSORT method and gave birth to 291 babies, the *weights* of the babies were recorded.

6. Clinical Trial of YSORT The YSORT method of sex selection, developed by the Genetics & IVF Institute, was designed to increase the likelihood that a baby will be male. When 291 couples used the YSORT method and gave birth to 291 babies, the *sexes* of the babies were recorded.

7. Clinical Trial of a COVID-19 Vaccine Treating 8532 subjects with a COVID-19 vaccine and subsequently recording whether they had a positive test for COVID-19.

8. Clinical Trial of a COVID-19 Vaccine Treating 8532 subjects with a COVID-19 vaccine and subsequently recording any COVID-19 symptoms (loss of taste, fever, etc.) that they experienced.

9. Nicorette Treating 50 smokers with Nicorette and asking them how their mouth and throat feel.

10. Nicorette Treating 50 smokers with Nicorette and recording whether there is a "yes" response when they are asked if they experience any mouth or throat soreness.

11. Defibrillators Determining whether each of 500 defibrillators is acceptable or defective.

12. Defibrillators Counting the numbers of defects in each of 500 defibrillators.

Binomial Probability Formula. ***In Exercises 13 and 14, answer the questions, designed to help understand the rationale for the binomial probability formula.***

13. Guessing Answers Standard tests, such as the SAT, ACT, or Medical College Admission Test (MCAT), typically use multiple choice questions, each with five possible answers (a, b, c, d, e), one of which is correct. Assume that you guess the answers to the first three questions.

a. Use the multiplication rule to find the probability that the first two guesses are wrong and the third is correct. That is, find P(WWC), where W denotes a wrong answer and C denotes a correct answer.

b. Beginning with WWC, make a complete list of the different possible arrangements of two wrong answers and one correct answer; then find the probability for each entry in the list.

c. Based on the preceding results, what is the probability of getting exactly one correct answer when three guesses are made?

14. Vision Correction 53% of adults use eyeglasses for vision correction (based on data from a Vision Council survey). Four adults are randomly selected.

a. Use the multiplication rule to find the probability that the first three use eyeglasses and the fourth does not use eyeglasses. That is, find P(EEEN), where E denotes an adult who uses eyeglasses and N denotes an adult who does not use eyeglasses.

b. Beginning with EEEN, make a complete list of the different possible arrangements of three adults who use eyeglasses and one who does not use eyeglasses; then find the probability for each entry in the list.

c. Based on the preceding results, what is the probability of getting exactly three adults who use eyeglasses and one who does not?

MCAT Test. ***In Exercises 15–20, assume that 5% of U.S. adults are vegetarians (based on data from a Gallup poll), and also assume that eight U.S. adults are randomly selected.***

15. Find the probability that among the eight selected adults, exactly two are vegetarians.

16. Find the probability that among the eight selected adults, fewer than two are vegetarians.

17. Find the probability that among the eight selected adults, more than two are vegetarians.

18. Find the probability that among the eight selected adults, exactly three are vegetarians.

19. Find the probability that among the eight selected adults, none are vegetarians.

20. Find the probability that among the eight selected adults, at least one is a vegetarian.

In Exercises 21–24, assume that when a subject is tested for marijuana use using the 1-Panel THC test, there is a 0.168 probability of a false positive result (based on data from Drug Test Success).

21. If 20 subjects are tested, find the probability that exactly 5 of them yield false positive results.

22. If 20 subjects are tested, find the probability that at least 5 of them yield false positive results.

23. If 20 subjects are tested, find the probability that none of them yield false positive results.

24. If 20 subjects are tested, find the probability that all of them yield false positive results.

Significance with Range Rule of Thumb. ***In Exercises 25 and 26, assume that different groups of heterosexual couples use the XSORT method of sex selection and each couple gives birth to one baby. The XSORT method is designed to increase the likelihood that a baby will be female, but assume that the method has no effect, so the probability of a female baby is 0.5.***

25. Sex Selection Assume that the groups consist of 36 heterosexual couples.

a. Find the mean and standard deviation for the numbers of females in groups of 36 births.

b. Use the range rule of thumb to find the values separating results that are significantly low and significantly high.

c. Is the result of 26 females significantly high? What does it suggest about the effectiveness of the XSORT method?

26. Sex Selection Assume that the groups consist of 16 heterosexual couples.

a. Find the mean and standard deviation for the numbers of females in groups of 16 births.

b. Use the range rule of thumb to find the values separating results that are significantly low and significantly high.

c. Is the result of 11 females significantly high? What does it suggest about the effectiveness of the XSORT method?

Significance with the Range Rule of Thumb. ***In Exercises 27 and 28, assume that hybridization experiments are conducted with peas having the property that for offspring there is a 0.75 probability that a pea has green pods (as in one of Mendel's famous experiments).***

27. Hybrids Assume that offspring peas are randomly selected in groups of 15.

a. Find the mean and standard deviation for the numbers of peas with green pods in the groups of 15.

b. Use the range rule of thumb to find the values separating results that are significantly low and significantly high.

c. Is the result of 2 peas with green pods a result that is significantly low? Why or why not?

28. Hybrids Assume that offspring peas are randomly selected in groups of 22.

a. Find the mean and standard deviation for the numbers of peas with green pods in the groups of 22.

b. Use the range rule of thumb to find the values separating results that are significantly low and significantly high.

c. Is the result of 20 peas with green pods a result that is significantly high? Why or why not?

Composite Sampling. ***Exercises 29 and 30 involve the method of composite sampling, whereby a medical testing laboratory saves time and money by combining blood samples for tests so that only one test is conducted for several people. A combined sample tests positive if at least one person has the disease. If a combined sample tests positive, then individual blood tests are used to identify the individual who has the disease.***

29. HIV It is estimated that worldwide, 1% of those aged 15–49 are infected with the human immunodeficiency virus (HIV) (based on data from the National Institutes of Health). In tests for HIV, blood samples from 36 people are combined. What is the probability that the combined sample tests positive for HIV? Is it unlikely for such a combined sample to test positive?

30. Blood Donor Testing The American Red Cross tests every unit of donated blood for several infectious diseases, including hepatitis B, hepatitis C, HIV, syphilis, and West Nile virus infection. Blood samples from 16 donors are combined and tested, and all 16 individual samples are approved only if the combined sample passes all tests. If there is 1.4% chance that a random individual fails any of the tests, find the probability that the combined sample is not approved.

Acceptance Sampling. ***Exercises 31 and 32 involve the method of acceptance sampling, whereby a shipment of a large number of items is accepted based on test results from a sample of the items.***

31. Aspirin The MedAssist Pharmaceutical Company receives large shipments of aspirin tablets and uses this acceptance sampling plan: Randomly select and test 40 tablets; then accept the whole batch if there is only one or none that doesn't meet the required specifications. If one shipment of 5000 aspirin tablets actually has a 3% rate of defects, what is the probability that this whole shipment will be accepted? Will almost all such shipments be accepted, or will many be rejected?

32. AAA Batteries AAA batteries are made by companies including Duracell, Energizer, Radio Shack, and Panasonic, and they are used to power Prestige Medical Xenon pocket otoscopes (those things that physicians use to look into your ears). When purchasing bulk orders of AAA batteries, a manufacturer of otoscopes uses this acceptance sampling plan: Randomly select 50 batteries and determine whether each is within specifications. The entire shipment is accepted if at most 2 batteries do not meet specifications. A shipment contains 2000 AAA batteries, and 2% of them do not meet specifications. What is the probability that this whole shipment will be accepted? Will almost all such shipments be accepted, or will many be rejected?

Ultimate Binomial Exercises! ***Exercises 33–36 involve finding binomial probabilities, finding parameters, and determining whether values are significantly high or low by using the range rule of thumb and probabilities.***

33. Sex Selection At an early stage of clinical trials of the XSORT method of sex selection, 14 couples using that method gave birth to 13 females and 1 male.

a. Assuming that the XSORT method has no effect and that males and females are equally likely, use the range rule of thumb to identify the limits separating values that are significantly low and those that are significantly high (for the number of females in 14 births). Based on the results, is the result of 13 females significantly high?

b. Find the probability of exactly 13 females in 14 births, assuming that the XSORT method has no effect.

continued

c. Find the probability of 13 or more females in 14 births, assuming that the XSORT method has no effect.

d. Which probability is relevant for determining whether 13 females is significantly high: the probability from part (b) or part (c)? Based on the relevant probability, is the result of 13 females significantly high?

e. What do the results suggest about the effectiveness of the XSORT method?

34. Clinical Trial A treatment for hypertension has been found to be successful in 60% of the patient population. In a test of a new treatment, 40 subjects are treated for hypertension and 29 of these subjects experience success with the new treatment.

a. Assuming that the old success rate of 60% still applies, use the range rule of thumb to identify the limits separating the numbers of successes that are significantly low and significantly high. Based on the results, is 29 successes among the 40 subjects significantly high?

b. Find the probability that exactly 29 of the 40 cases are successes, assuming that the general success rate is 60%.

c. Find the probability that 29 or more of the cases are successes, assuming that the general success rate is 60%.

d. Which probability is relevant for determining whether 29 successes is significantly high: the probability from part (b) or part (c)? Based on the relevant probability, is the result of 29 successes significantly high?

e. What do the results suggest about the effectiveness of the new treatment?

35. Hybrids One of Mendel's famous experiments with peas included 47 offspring, and 34 of them had long stems. Mendel claimed that under the same conditions, 75% of offspring peas would have long stems. Assume that Mendel's claim of 75% is true, and assume that a sample consists of 47 offspring peas.

a. Use the range rule of thumb to identify the limits separating values that are significantly low and those that are significantly high. Based on the results, is the result of 34 peas with long stems either significantly low or significantly high?

b. Find the probability of exactly 34 peas with long stems.

c. Find the probability of 34 or fewer peas with long stems.

d. Which probability is relevant for determining whether 34 peas with long stems is significantly low: the probability from part (b) or part (c)? Based on the relevant probability, is the result of 34 peas with long stems significantly low?

e. What do the results suggest about Mendel's claim of 75%?

36. Vaccine For a specific group of subjects, there is a 5% chance of getting influenza ("flu"). When 80 subjects were treated with a vaccine, only one of them presented with influenza.

a. Use the range rule of thumb to identify the limits separating values that are significantly low and those that are significantly high. Based on the results, is the result of one subject getting influenza either significantly low or significantly high?

b. Find the probability of exactly one subject experiencing influenza, assuming that the vaccine has no effect.

c. Find the probability of one or fewer subjects experiencing influenza, assuming that the vaccine has no effect.

d. Which probability is relevant for determining whether one subject experiencing influenza is significantly low: the probability from part (b) or part (c)? Based on the relevant probability, is the result of one subject experiencing influenza significantly low?

e. What do the results suggest about the effectiveness of the vaccine?

5-2 Beyond the Basics

37. Geometric Distribution If a procedure meets all the conditions of a binomial distribution except that the number of trials is not fixed, then the **geometric distribution** can be used. The probability of getting the first success on the xth trial is given by $P(x) = p(1 - p)^{x-1}$, where p is the probability of success on any one trial. Subjects are randomly selected for the National Health and Nutrition Examination Survey conducted by the National Center for Health Statistics, Centers for Disease Control and Prevention. The probability that someone is a universal donor (with type O and Rh negative blood) is 0.066. Find the probability that the first subject to be a universal donor is the fourth person selected.

38. Multinomial Distribution The binomial distribution applies only to cases involving two types of outcomes, whereas the **multinomial distribution** involves more than two categories. Suppose we have three types of mutually exclusive outcomes denoted by A, B, and C. Let $P(\text{A}) = p_1$, $P(\text{B}) = p_2$, and $P(\text{C}) = p_3$. In n independent trials, the probability of x_1 outcomes of type A, x_2 outcomes of type B, and x_3 outcomes of type C is given by

$$\frac{n!}{(x_1)!(x_2)!(x_3)!} \cdot p_1^{x_1} \cdot p_2^{x_2} \cdot p_3^{x_3}$$

Data Set 11 "IQ and Lead" in Appendix B includes 78 subjects from a low lead exposure group, 22 subjects from a medium lead exposure group, and 21 subjects from a high lead exposure group. Find the probability of randomly selecting 10 subjects for a follow-up study and getting 5 from the low lead group, 2 from the medium lead group, and 3 from the high lead group. Assume that the selections are made with replacement. Can we use the above expression for finding the probability if the sampling is done without replacement?

39. Hypergeometric Distribution If we sample from a small finite population *without replacement*, the binomial distribution should not be used because the events are not independent. If sampling is done without replacement and the outcomes belong to one of two types, we can use the **hypergeometric distribution**. If a population has A objects of one type, while the remaining B objects are of the second type, and if n objects are sampled without replacement, then the probability of getting x objects of the first type and $n - x$ objects of the second type is

$$P(x) = \frac{A!}{(A-x)!x!} \cdot \frac{B!}{(B-n+x)!(n-x)!} \div \frac{(A+B)!}{(A+B-n)!n!}$$

In a medical research project, 20 subjects are available and 4 of them are infected with HIV, while the other 16 are not infected. If 8 of the subjects are randomly selected without replacement, what is the probability that 3 of the subjects are infected with HIV while the other 5 are not infected? What is the probability if the sampling is done with replacement?

5-3 Poisson Probability Distributions

Key Concept In Section 5-1 we introduced general discrete probability distributions, and in Section 5-2 we considered binomial probability distributions, which is one particular category of discrete probability distributions. In this section we introduce *Poisson probability distributions,* which are another category of discrete probability distributions.

The following definition states that Poisson distributions are used with occurrences of an event over a specified interval, and here are some applications: