TIP: Binomial versus negative binomial
In the binomial case, we typically have a fixed number of trials and instead consider the number of successes. In the negative binomial case, we examine how many trials it takes to observe a fixed number of successes and require that the last observation be a success.

Exercise 3.62 On 70% of days, a hospital admits at least one heart attack patient. On 30% of the days, no heart attack patients are admitted. Identify each case below as a binomial or negative binomial case, and compute the probability.[45]

(a) What is the probability the hospital will admit a heart attack patient on exactly three days this week?

(b) What is the probability the second day with a heart attack patient will be the fourth day of the week?

(c) What is the probability the fifth day of next month will be the first day with a heart attack patient?

3.5.2 Poisson distribution

Example 3.63 There are about 8 million individuals in New York City. How many individuals might we expect to be hospitalized for acute myocardial infarction (AMI), i.e. a heart attack, each day? According to historical records, the average number is about 4.4 individuals. However, we would also like to know the approximate distribution of counts. What would a histogram of the number of AMI occurrences each day look like if we recorded the daily counts over an entire year?

A histogram of the number of occurrences of AMI on 365 days[46] for NYC is shown in Figure 3.21. The sample mean (4.38) is similar to the historical average of 4.4. The sample standard deviation is about 2, and the histogram indicates that about 70% of the data fall between 2.4 and 6.4. The distribution's shape is unimodal and skewed to the right.

The **Poisson distribution** is often useful for estimating the number of rare events in a large population over a unit of time. For instance, consider each of the following events, which are rare for any given individual:

- having a heart attack,
- getting married, and
- getting struck by lightning.

The Poisson distribution helps us describe the number of such events that will occur in a short unit of time for a fixed population if the individuals within the population are independent.

---

[45] In each part, $p = 0.7$. (a) The number of days is fixed, so this is binomial. The parameters are $k = 3$ and $n = 7$: 0.097. (b) The last "success" (admitting a heart attack patient) is fixed to the last day, so we should apply the negative binomial distribution. The parameters are $k = 2$, $n = 4$: 0.132. (c) This problem is negative binomial with $k = 1$ and $n = 5$: 0.006. Note that the negative binomial case when $k = 1$ is the same as using the geometric distribution.

[46] These data are simulated. In practice, we should check for an association between successive days.

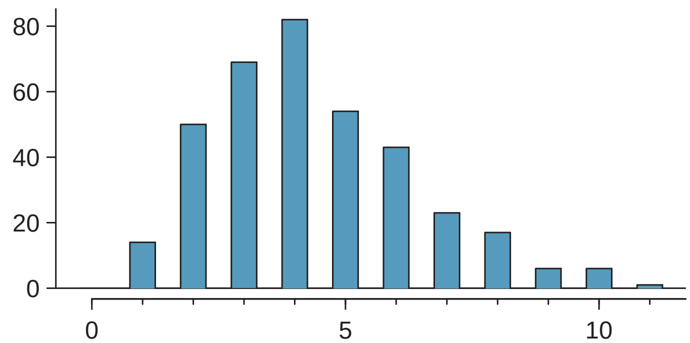

Figure 3.21: A histogram of the number of occurrences of AMI on 365 separate days in NYC.

The histogram in Figure 3.21 approximates a Poisson distribution with rate equal to 4.4. The **rate** for a Poisson distribution is the average number of occurrences in a mostly-fixed population per unit of time. In Example 3.63, the time unit is a day, the population is all New York City residents, and the historical rate is 4.4. The parameter in the Poisson distribution is the rate – or how many rare events we expect to observe – and it is typically denoted by λ (the Greek letter *lambda*) or μ. Using the rate, we can describe the probability of observing exactly k rare events in a single unit of time.

λ
Rate for the Poisson dist.

Poisson distribution

Suppose we are watching for rare events and the number of observed events follows a Poisson distribution with rate λ. Then

$$P(\text{observe } k \text{ rare events}) = \frac{\lambda^k e^{-\lambda}}{k!}$$

where k may take a value 0, 1, 2, and so on, and $k!$ represents k-factorial, as described on page 138. The letter $e \approx 2.718$ is the base of the natural logarithm. The mean and standard deviation of this distribution are λ and $\sqrt{\lambda}$, respectively.

We will leave a rigorous set of conditions for the Poisson distribution to a later course. However, we offer a few simple guidelines that can be used for an initial evaluation of whether the Poisson model would be appropriate.

TIP: Is it Poisson?

A random variable may follow a Poisson distribution if the event being considered is rare, the population is large, and the events occur independently of each other.

Even when rare events are not really independent – for instance, Saturdays and Sundays are especially popular for weddings – a Poisson model may sometimes still be reasonable if we allow it to have a different rate for different times. In the wedding example, the rate would be modeled as higher on weekends than on weekdays. The idea of modeling rates for a Poisson distribution against a second variable such as `dayOfTheWeek` forms the foundation of some more advanced methods that fall in the realm of **generalized linear models**. In Chapters 7 and 8, we will discuss a foundation of linear models.