PEARSON NEW INTERNATIONAL EDITION

Statistics for Business
Decision Making and Analysis

Robert Stine Dean Foster

Second Edition

ALWAYS LEARNING

PEARSON

Pearson New International Edition

Statistics for Business
Decision Making and Analysis

Robert Stine Dean Foster

Second Edition

PEARSON®

$$\begin{aligned}P(Y = 0) &= P(B_1 = 0 \textit{ and } B_2 = 0 \textit{ and} \ldots \textit{and } B_{10} = 0)\\ &= P(B_1 = 0)P(B_2 = 0) \ldots P(B_{10} = 0)\\ &= (1 - p)^{10} = 0.6^{10} \approx 0.006\end{aligned}$$

There is not much chance for a rep to strike out in 10 visits if $p = 0.4$.

What about $P(Y = 1)$, the chance that a rep sees exactly one of the 10 doctors? For this to happen, one of the B's has to be 1 and the rest must all equal 0. The probability that the first visit is a success and the rest are failures is

$$P(B_1 = 1)P(B_2 = 0)P(B_3 = 0) \ldots P(B_{10} = 0) = p(1 - p)^9$$

Similarly, the probability that the second visit succeeds and the rest are failures is

$$P(B_1 = 0, B_2 = 1, B_3 = 0, \ldots, B_{10} = 0) = p(1 - p)^9$$

You can see the pattern. Regardless of which visit succeeds, the probability of a specific success in 10 trials is $p(1 - p)^9$. Because there are 10 possible visits that might succeed, the probability for one success is

$$P(Y = 1) = 10\,p(1 - p)^9 = 10(0.4)(0.6)^9 \approx 0.040$$

In general, each binomial probability has two parts, a probability and a count:

1. The probability of y successes in n Bernoulli trials
2. The number of sequences that have y successes in n trials

binomial coefficient The number of arrangements of y successes in n trials, denoted $_nC_y$ and sometimes written as $\binom{n}{y}$.

The probability of y successes in n trials is $p^y(1 - p)^{n-y}$ because the trials are independent. For the count, a formula gives the number of ways to label y out of n trials as successful. The formula is known as the **binomial coefficient**, written $_nC_y$. (Pronounce this as "n choose y.") The binomial coefficient is defined as

$$_nC_y = \frac{n!}{y!(n - y)!}$$

The exclamation sign following a letter (as in $n!$) identifies the factorial function that is defined as the product $n! = n(n - 1) \ldots (2)(1)$ with $0! = 1$. Your calculator may have buttons that do factorials and binomial coefficients. The section Behind the Math: Binomial Counting gives more examples.

Combining the count and probability, the binomial probability for y successes among n trials is

$$P(Y = y) = {_nC_y}\, p^y(1 - p)^{n-y}$$

For 10 detail rep visits, $n = 10$ and $p = 0.4$. Hence, the probability distribution of the binomial random variable Y is

$$P(Y = y) = {_{10}C_y}(0.4)^y(0.6)^{10-y}$$

For example, the probability for $y = 8$ successful detail visits is the number of arrangements ($_{10}C_8 = 45$) times the probability of each ($p^8(1 - p)^2 = 0.4^8 0.6^2$).

$$P(Y = 8) = {_{10}C_8}(0.4)^8(0.6)^2 = 45(0.4)^8(0.6)^2 \approx 0.011$$

Now that we have a formula for $p(y)$, we can graph the probability distribution as shown in Figure 1.

The maximum probability occurs at the mean $E(Y) = 4$. The mode (outcome with largest probability) is always near the mean for binomial random variables.

To find the probability that a detail rep meets 8 or more doctors in a day, we have to add the last three probabilities shown in Figure 1. These represent three disjoint events, all with small probability.

$$\begin{aligned}P(Y \geq 8) &= P(Y = 8) + P(Y = 9) + P(Y = 10)\\ &\approx 0.01062 + 0.00157 + 0.00010 = 0.01229\end{aligned}$$

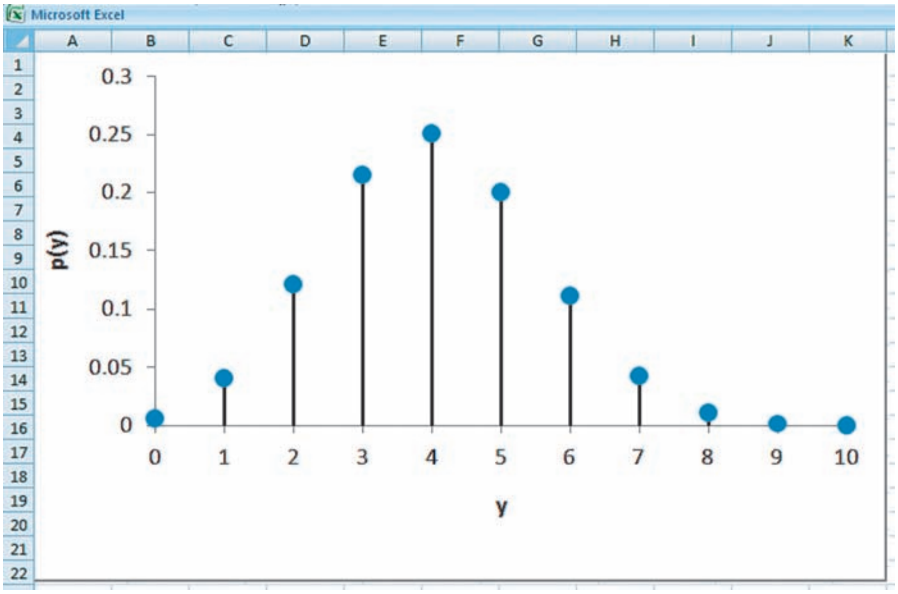


FIGURE 1 The probability distribution of $Y \sim \text{Bi}(10, 0.4)$.

There's slightly more than a 1% chance for a detail rep to meet 8 or more doctors during 10 office visits. We'd expect fewer, about 4 per day. A rep that regularly sees 8 doctors in 10 visits looks like a star!

Summary

A binomial random variable counts the number of successes in n Bernoulli trials. Two parameters identify every binomial random variable: the number of Bernoulli trials n and the probability of success p.

Binomial random variable, $Y \sim \text{Bi}(n, p)$

n = number of trials
p = probability of success
y = number of successes in n Bernoulli trials

$$P(Y = y) = {}_nC_y\, p^y (1 - p)^{n-y}, \qquad \text{where } {}_nC_y = \frac{n!}{y!(n-y)!}$$

$$\text{Mean: } E(Y) = np$$
$$\text{Variance: } \text{Var}(Y) = np(1 - p)$$

What Do You Think?

a. Which of the following situations seem suited to a binomial random variable?[4]
The total cost of items purchased by five customers
The number of items purchased by five customers
The number of employees choosing a new health plan

b. If a public radio station solicits contributions during a late-night program, how many listeners should it expect to call in if $n = 1{,}000$ are listening and $p = 0.15$? How large is the SD of the number who call in?[5]

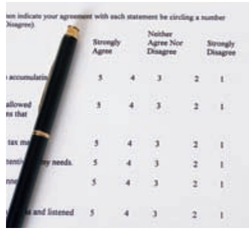
---

[4] Only the number of employees choosing the new health plan, so long as employees are not swayed by the opinion of one outspoken employee. The total cost and number of items are not sums of outcomes of Bernoulli trials; you can buy more than one item, for instance.

[5] We expect 150 to call, with SD $= \sqrt{np(1 - p)} \approx \sqrt{1000(0.15)(0.85)} \approx 11.3$.

4M EXAMPLE 1 FOCUS ON SALES

MOTIVATION ▶ STATE THE QUESTION



During the early stages of product development, some companies use focus groups to get the reaction of customers to a new product or design. The particular focus group in this example has nine randomly selected customers. After being shown a prototype, they were asked, "Would you buy this product if it sold for $99.95?"

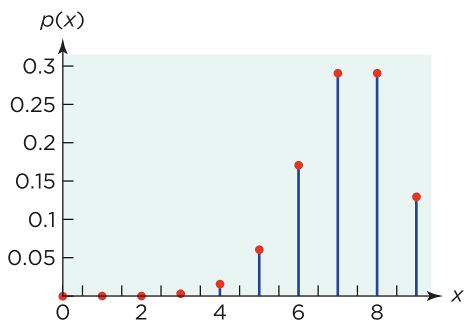
The questionnaire allowed only two answers: yes or no. When the answers were totaled, six of the nine customers in the group answered yes. Prior to the focus group, the development team claimed that 80% of customers would want to buy the product at this price. If the developers are right, what should we expect to happen at the focus group? Is it likely for six out of nine customers in the group to say they'd buy it? ◀

METHOD ▶ DESCRIBE THE DATA AND SELECT AN APPROACH

This application calls for counting a collection of yes/no responses, the perfect setting for Bernoulli trials and a binomial random variable. On the basis of the claims of the developers, the random variable $Y \sim \text{Bi}(n = 9, p = 0.8)$ represents the possibilities in a focus group of nine customers. An important caveat is that Bernoulli trials are also independent of one another; this may be true of a focus group. ◀

MECHANICS ▶ DO THE ANALYSIS

The expected value of Y is $np = 9(0.8) = 7.2$, higher than the observed number of responses. The standard deviation of Y is $\sqrt{np(1-p)} = \sqrt{9(0.2)(0.8)} = 1.2$. This plot shows the probability distribution for the random variable. The probability of exactly six saying yes: $P(Y = 6) = {}_9C_6(0.8)^6(0.2)^3 \approx 0.18$. That's not the most likely count, but it is common.



◀

MESSAGE ▶ SUMMARIZE THE RESULTS

The results of this focus group are in line with what we'd expect to see if the development team is right. If the team is right, about one-sixth of focus groups of nine customers will result in six who say yes to the purchase question. It is important that those who run the focus group allow each member of the group to fill in his or her questionnaire without being overly influenced by others in the group. ◀

Let's return to the assumption of independence. It's easy for dependence to creep into a focus group. If an outspoken person dominates the group, we've effectively got fewer trials. As a result, the results will be more variable than anticipated by a binomial model. For example, the binomial model in Example 1 implies that it is unlikely for the group to be unanimous. The probability of all saying yes is $P(Y = 9) = {}_9C_9 p^9(1 - p)^0 = (0.8)^9 \approx 0.134$. If, however, one person can persuade the others to follow her or his opinion, there's an 80% chance that they will all say yes.

4 | POISSON MODEL

Bernoulli trials are hard to identify in some counting problems. For these, a second discrete random variable is helpful if the underlying random process operates continually. As examples, consider these situations.

- The number of imperfections per square meter of glass panel used to make LCD televisions
- The number of robot malfunctions per day on an assembly line
- The number of telephone calls arriving at the help desk of a software company during a 10-minute period

Each situation produces a count, whether it's the count of flaws in a manufactured product or the number of telephone calls. Every case, however, requires some imagination to find Bernoulli trials. For the help desk, for instance, we could slice the 10 minutes into 600 one-second intervals. If the intervals are short enough, at most one call lands in each. We've got success-failure trials and a binomial model, but it takes a bit of imagination to find the trials.

Poisson Random Variable

There's a better way to model these counts. Notice that each of these situations has a rate. The manager of the help desk, for instance, would be dumbfounded if you asked for the probability of a call arriving during the next second, but the manager could easily quote you the average number of calls per hour. Similarly, we have a rate for defects per square meter or malfunctions per day.

Poisson random variable
A random variable describing a count of the number of events in a random process that occurs at a steady rate denoted by λ.

A **Poisson random variable** models the number of events produced by a random process during an interval of time or space. A Poisson random variable has one parameter. This parameter $\boldsymbol{\lambda}$ (spelled lambda, but pronounced "lam-da") is the rate of events, or arrivals, within *disjoint* intervals. If you were told "We typically receive 100 calls per hour at this call center," then $\lambda = 100$ calls/hour. Keep track of the measurement units on λ.

If X denotes a Poisson random variable with parameter λ [abbreviated $X \sim \text{Poisson}(\lambda)$], then its probability distribution is

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots$$

Unlike those of a binomial random variable, the possible values of this random variable keep going and going. Conceptually, there's no limit on the size of a Poisson random variable. (The symbol $\boldsymbol{x!}$ denotes the factorial function as in a binomial distribution, $x! = x(x - 1)\cdots(2)(1)$, with $0! = 1$. The letter e stands for the base of natural logs, $e \approx 2.71828$.)

To illustrate the calculations, suppose that calls arrive at the help desk at a rate of 90 calls per hour. We're interested in the chance that no call arrives in the next minute. In order to use a Poisson random variable to model the number of calls that arrive in the next minute, we have to adjust λ to suit the time interval of interest. The appropriate rate is $\lambda = 90/60 \approx 1.5$ calls *per minute*. If X is a Poisson random variable with $\lambda = 1.5$, then the probability of no calls during the next minute is

$$P(X = 0) = e^{-\lambda}\lambda^0/0! = e^{-1.5} \approx 0.223$$

The probability of one call during the next minute is 1.5 times larger

$$P(X = 1) = e^{-\lambda}\lambda^1/1! = 1.5e^{-1.5} \approx 0.335$$

and the probability of two calls is

$$P(X = 2) = e^{-\lambda}\lambda^2/2! = (1.5^2/2)e^{-1.5} \approx 0.251$$

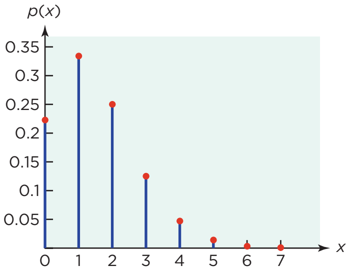


FIGURE 2 Poisson probability distribution with $\lambda = 1.5$.

Figure 2 graphs the probability distribution $P(X = x)$, for x ranging from 0 to 7. The probability distribution keeps on going and going, but the probabilities are near zero beyond those shown here.

If the number of calls is a Poisson random variable with rate $\lambda = 1.5$ calls per minute, what would you guess is the mean number of calls per minute? It's $\lambda = 1.5$. It is not obvious, but λ is also the variance of a Poisson random variable.

Poisson random variable, $X \sim$ Poisson(λ)

λ = expected count of events over a time interval or a region

X = number of events during an interval or in the region

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

Expected value: $E(X) = \lambda$

Variance: $\text{Var}(X) = \lambda$

Poisson Model

Poisson model A model in which a Poisson random variable is used to describe counts of real data.

The **Poisson model** refers to using a Poisson random variable to describe a random process in the real world. The necessary assumptions resemble those required by the binomial model. Whereas the binomial model requires independent trials, the Poisson model assumes that events in separate intervals are independent. The binomial model assumes that the chance for success p is the same for every trial; the Poisson model assumes that the rate of events stays the same.

Let's continue with telephone calls arriving at a help desk. The help desk handles 90 calls per hour, on average. Before we use a Poisson random variable to compute probabilities, we should check the assumptions. First, arrivals in disjoint intervals should be independent. Hence, if the help desk gets more calls than usual in an hour, it should not expect the rate of calls to drop below average during the next hour. (That would imply negative dependence.) Second, the rate of arrivals should be constant. This assumption makes sense during the regular business hours. However, the rate would probably decrease at night or on weekends.

Deciding between a binomial model and a Poisson model is straightforward.

tip Use a binomial model when you recognize distinct Bernoulli trials. Use a Poisson model when events happen at a constant rate over time or space.

4M EXAMPLE 2 DEFECTS IN SEMICONDUCTORS

MOTIVATION ▸ STATE THE QUESTION

Computer chips are made from slices of silicon known as wafers. A complex process (known as photolithography) etches lines in the wafer, and these lines become the circuits that enable the chip to process data. With feature sizes of 45 nanometers (about 0.0000018 inch) and smaller, defects in the wafer cause real problems.

Before companies like AMD and Intel buy wafers from a supplier, they want to be sure that each wafer does not have too many defects. No one expects perfection, but the number of defects must be small. In this example, a supplier claims that its wafers have 1 defect per 400 square centimeters (cm^2). Each wafer is 20 centimeters in diameter, so its area is $\pi r^2 = \pi(10)^2 \approx 314\ cm^2$.

What is the probability that a wafer from this supplier has no defects? What is the mean number of defects and the SD? (The SD shows how consistent the number of defects will be from wafer to wafer.) ◂

METHOD ▸ DESCRIBE THE DATA AND SELECT AN APPROACH

The relevant random variable is the number of defects on a randomly selected wafer. The type of defect should not carry over from wafer to wafer, so it seems reasonable that these counts should be independent. A Poisson model suits this situation. ◂

MECHANICS ▸ DO THE ANALYSIS

The supplier claims a defect rate of 1 per 400 cm^2. Since a wafer has 314 cm^2, we model the number of defects on a randomly chosen wafer as the random variable $X \sim \text{Poisson}(\lambda = 314/400)$.

The mean number of defects is $\lambda = 314/400 = 0.785$. The probability of no defect on a wafer is then $P(X = 0) = e^{-0.785} \approx 0.456$. The SD for the Poisson model is the square root of the mean, so $\sigma = 0.886$ defects/wafer. ◂

MESSAGE ▸ SUMMARIZE THE RESULTS

If the claims are accurate, the chip maker can expect about 0.8 defects per wafer. Because the variation in the number of defects is small, there will not be much variation in the number of defects from wafer to wafer. About 46% of the wafers should be free of defects. ◂

Best Practices

- *Ensure that you have Bernoulli trials if you are going to use the binomial model.* Bernoulli trials allow only one of two outcomes, with a constant probability of success and independent outcomes. If the conditions of the trials change, so might the chance for success. Results of one trial must not influence the outcomes of others.
- *Use the binomial model to simplify the analysis of counts.* When the random process that you are interested in produces counts from discrete