C H A P T E R 5

Discrete Probability Distributions

Objectives

After completing this chapter, you should be able to

1. Construct a probability distribution for a random variable.
2. Find the mean, variance, standard deviation, and expected value for a discrete random variable.
3. Find the exact probability for *X* successes in *n* trials of a binomial experiment.
4. Find the mean, variance, and standard deviation for the variable of a binomial distribution.
5. Find probabilities for outcomes of variables, using the Poisson, hypergeometric, and multinomial distributions.

Outline

Statistics Today

Is Pooling Worthwhile?

Blood samples are used to screen people for certain diseases. When the disease is rare, health care workers sometimes combine or pool the blood samples of a group of individuals into one batch and then test it. If the test result of the batch is negative, no further testing is needed since none of the individuals in the group has the disease. However, if the test result of the batch is positive, each individual in the group must be tested.

Consider this hypothetical example: Suppose the probability of a person having the disease is 0.05, and a pooled sample of 15 individuals is tested. What is the probability that no further testing will be needed for the individuals in the sample? The answer to this question can be found by using what is called the *binomial distribution*. See Statistics Today—Revisited at the end of the chapter.

This chapter explains probability distributions in general and a specific, often used distribution called the binomial distribution. The Poisson, hypergeometric, and multinomial distributions are also explained.

Introduction

Many decisions in business, insurance, and other real-life situations are made by assigning probabilities to all possible outcomes pertaining to the situation and then evaluating the results. For example, a saleswoman can compute the probability that she will make 0, 1, 2, or 3 or more sales in a single day. An insurance company might be able to assign probabilities to the number of vehicles a family owns. A self-employed speaker might be able to compute the probabilities for giving 0, 1, 2, 3, or 4 or more speeches each week. Once these probabilities are assigned, statistics such as the mean, variance, and standard deviation can be computed for these events. With these statistics, various decisions can be made. The saleswoman will be able to compute the average number of sales she makes per week, and if she is working on commission, she will be able to approximate her weekly income over a period of time, say, monthly. The public speaker will be able to plan ahead and approximate his average income and expenses. The insurance company can use its information to design special computer forms and programs to accommodate its customers' future needs.

This chapter explains the concepts and applications of what is called a *probability distribution.* In addition, special probability distributions, such as the *binomial, multinomial, Poisson,* and *hypergeometric* distributions, are explained.

5–1

Probability Distributions

Objective 1

Construct a probability distribution for a random variable.

Before probability distribution is defined formally, the definition of a variable is reviewed. In Chapter 1, a *variable* was defined as a characteristic or attribute that can assume different values. Various letters of the alphabet, such as *X, Y,* or *Z,* are used to represent variables. Since the variables in this chapter are associated with probability, they are called *random variables.*

For example, if a die is rolled, a letter such as *X* can be used to represent the outcomes. Then the value that *X* can assume is 1, 2, 3, 4, 5, or 6, corresponding to the outcomes of rolling a single die. If two coins are tossed, a letter, say *Y,* can be used to represent the number of heads, in this case 0, 1, or 2. As another example, if the temperature at 8:00 A.M. is 43° and at noon it is 53°, then the values *T* that the temperature assumes are said to be random, since they are due to various atmospheric conditions at the time the temperature was taken.

A **random variable** is a variable whose values are determined by chance.

Also recall from Chapter 1 that you can classify variables as discrete or continuous by observing the values the variable can assume. If a variable can assume only a specific number of values, such as the outcomes for the roll of a die or the outcomes for the toss of a coin, then the variable is called a *discrete variable.*

Discrete variables have a finite number of possible values or an infinite number of values that can be counted. The word *counted* means that they can be enumerated using the numbers 1, 2, 3, etc. For example, the number of joggers in Riverview Park each day and the number of phone calls received after a TV commercial airs are examples of discrete variables, since they can be counted.

Variables that can assume all values in the interval between any two given values are called *continuous variables.* For example, if the temperature goes from 62 to 78° in a 24-hour period, it has passed through every possible number from 62 to 78. *Continuous random variables are obtained from data that can be measured rather than counted.* Continuous random variables can assume an infinite number of values and can be decimal and fractional values. On a continuous scale, a person's weight might be exactly 183.426 pounds if a scale could measure weight to the thousandths place; however, on a digital scale that measures only to tenths of pounds, the weight would be 183.4 pounds. Examples of continuous variables are heights, weights, temperatures, and time. In this chapter only discrete random variables are used; Chapter 6 explains continuous random variables.

The procedure shown here for constructing a probability distribution for a discrete random variable uses the probability experiment of tossing three coins. Recall that when three coins are tossed, the sample space is represented as TTT, TTH, THT, HTT, HHT, HTH, THH, HHH; and if *X* is the random variable for the number of heads, then *X* assumes the value 0, 1, 2, or 3.

Probabilities for the values of X can be determined as follows:

No heads	One head			Two heads			Three heads
TTT	TTH	THT	HTT	HHT	HTH	THH	HHH
$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
$\frac{1}{8}$	$\frac{3}{8}$			$\frac{3}{8}$			$\frac{1}{8}$

Hence, the probability of getting no heads is $\frac{1}{8}$, one head is $\frac{3}{8}$, two heads is $\frac{3}{8}$, and three heads is $\frac{1}{8}$. From these values, a probability distribution can be constructed by listing the outcomes and assigning the probability of each outcome, as shown here.

Number of heads *X*	0	1	2	3
Probability *P*(*X*)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

A **discrete probability distribution** consists of the values a random variable can assume and the corresponding probabilities of the values. The probabilities are determined theoretically or by observation.

Discrete probability distributions can be shown by using a graph or a table. Probability distributions can also be represented by a formula. See Exercises 31–36 at the end of this section for examples.

Example 5–1

Rolling a Die

Construct a probability distribution for rolling a single die.

Solution

Since the sample space is 1, 2, 3, 4, 5, 6 and each outcome has a probability of $\frac{1}{6}$, the distribution is as shown.

Outcome *X*	1	2	3	4	5	6
Probability *P*(*X*)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Probability distributions can be shown graphically by representing the values of X on the x axis and the probabilities $P(X)$ on the y axis.

Example 5–2

Tossing Coins

Represent graphically the probability distribution for the sample space for tossing three coins.

Number of heads *X*	0	1	2	3
Probability *P*(*X*)	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

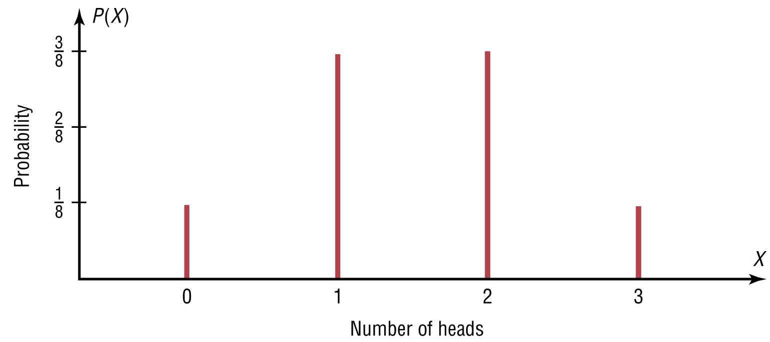
Solution

The values that X assumes are located on the x axis, and the values for $P(X)$ are located on the y axis. The graph is shown in Figure 5–1.

Note that for visual appearances, it is not necessary to start with 0 at the origin.

Examples 5–1 and 5–2 are illustrations of *theoretical* probability distributions. You did not need to actually perform the experiments to compute the probabilities. In contrast, to construct actual probability distributions, you must observe the variable over a period of time. They are empirical, as shown in Example 5–3.

Figure 5–1
Probability Distribution for Example 5–2



Example 5–3

Baseball World Series

The baseball World Series is played by the winner of the National League and the American League. The first team to win four games wins the World Series. In other words, the series will consist of four to seven games, depending on the individual victories. The data shown consist of 40 World Series events. The number of games played in each series is represented by the variable *X*. Find the probability *P*(*X*) for each *X*, construct a probability distribution, and draw a graph for the data.

X	Number of games played
4	8
5	7
6	9
7	16
	40

Solution

The probability *P*(*X*) can be computed for each *X* by dividing the number of games *X* by the total.

For 4 games, $\frac{8}{40} = 0.200$ For 6 games, $\frac{9}{40} = 0.225$

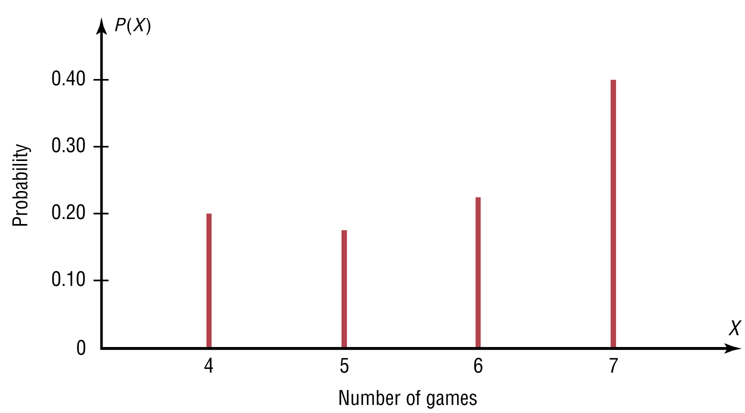
For 5 games, $\frac{7}{40} = 0.175$ For 7 games, $\frac{16}{40} = 0.400$

The probability distribution is

Number of games *X*	4	5	6	7
Probability *P*(*X*)	0.200	0.175	0.225	0.400

The graph is shown in Figure 5–2.

Figure 5–2
Probability Distribution for Example 5–3



Speaking of Statistics

Coins, Births, and Other Random (?) Events

Examples of random events such as tossing coins are used in almost all books on probability. But is flipping a coin really a random event?

Tossing coins dates back to ancient Roman times when the coins usually consisted of the Emperor's head on one side (i.e., heads) and another icon such as a ship on the other side (i.e., ships). Tossing coins was used in both fortune telling and ancient Roman games.

A Chinese form of divination called the *I-Ching* (pronounced E-Ching) is thought to be at least 4000 years old. It consists of 64 hexagrams made up of six horizontal lines. Each line is either broken or unbroken, representing the yin and the yang. These 64 hexagrams are supposed to represent all possible situations in life. To consult the I-Ching, a question is asked and then three coins are tossed six times. The way the coins fall, either heads up or heads down, determines whether the line is broken (yin) or unbroken (yang). Once the hexagon is determined, its meaning is consulted and interpreted to get the answer to the question. (*Note:* Another method used to determine the hexagon employs yarrow sticks.)

In the 16th century, a mathematician named Abraham DeMoivre used the outcomes of tossing coins to study what later became known as the normal distribution; however, his work at that time was not widely known.

Mathematicians usually consider the outcomes of a coin toss a random event. That is, each probability of getting a head is $\frac{1}{2}$, and the probability of getting a tail is $\frac{1}{2}$. Also, it is not possible to predict with 100% certainty which outcome will occur. But new studies question this theory. During World War II a South African mathematician named John Kerrich tossed a coin 10,000 times while he was interned in a German prison camp. Unfortunately, the results of his experiment were never recorded, so we don't know the number of heads that occurred.

Several studies have shown that when a coin-tossing device is used, the probability that a coin will land on the same side on which it is placed on the coin-tossing device is about 51%. It would take about 10,000 tosses to become aware of this bias. Furthermore, researchers showed that when a coin is spun on its edge, the coin falls tails up about 80% of the time since there is more metal on the heads side of a coin. This makes the coin slightly heavier on the heads side than on the tails side.

Another assumption commonly made in probability theory is that the number of male births is equal to the number of female births and that the probability of a boy being born is $\frac{1}{2}$ and the probability of a girl being born is $\frac{1}{2}$. We know this is not exactly true.

In the later 1700s, a French mathematician named Pierre Simon Laplace attempted to prove that more males than females are born. He used records from 1745 to 1770 in Paris and showed that the percentage of females born was about 49%. Although these percentages vary somewhat from location to location, further surveys show they are generally true worldwide. Even though there are discrepancies, we generally consider the outcomes to be 50-50 since these discrepancies are relatively small.

Based on this article, would you consider the coin toss at the beginning of a football game fair?

Two Requirements for a Probability Distribution

1. The sum of the probabilities of all the events in the sample space must equal 1; that is, $\Sigma P(X) = 1$.
2. The probability of each event in the sample space must be between or equal to 0 and 1. That is, $0 \leq P(X) \leq 1$.

The first requirement states that the sum of the probabilities of all the events must be equal to 1. This sum cannot be less than 1 or greater than 1 since the sample space includes *all* possible outcomes of the probability experiment. The second requirement states that the probability of any individual event must be a value from 0 to 1. The reason (as stated in Chapter 4) is that the range of the probability of any individual value can be 0, 1, or any value between 0 and 1. A probability cannot be a negative number or greater than 1.

Example 5–4

Probability Distributions

Determine whether each distribution is a probability distribution.

a.

X	4	6	8	10
P(X)	−0.6	0.2	0.7	1.5

b.

X	1	2	3	4
P(X)	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

c.

X	8	9	12
P(X)	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

d.

X	1	3	5	7	9
P(X)	0.3	0.1	0.2	0.4	−0.7

Solution

a. No. It is not a probability distribution since $P(X)$ cannot be negative or greater than 1.

b. Yes. It is a probability distribution.

c. Yes. It is a probability distribution.

d. No, since $P(X) \neq -0.7$.

Many variables in business, education, engineering, and other areas can be analyzed by using probability distributions. Section 5–2 shows methods for finding the mean and standard deviation for a probability distribution.

Applying the Concepts 5–1

Dropping College Courses

Use the following table to answer the questions.

Reason for Dropping a College Course	Frequency	Percentage
Too difficult	45	
Illness	40	
Change in work schedule	20	
Change of major	14	
Family-related problems	9	
Money	7	
Miscellaneous	6	
No meaningful reason	3	

1. What is the variable under study? Is it a random variable?
2. How many people were in the study?
3. Complete the table.
4. From the information given, what is the probability that a student will drop a class because of illness? Money? Change of major?
5. Would you consider the information in the table to be a probability distribution?
6. Are the categories mutually exclusive?
7. Are the categories independent?
8. Are the categories exhaustive?
9. Are the two requirements for a discrete probability distribution met?

See page 297 for the answers.

Exercises 5–1

1. Define and give three examples of a random variable. A random variable is a variable whose values are determined by chance. Examples will vary.

2. Explain the difference between a discrete and a continuous random variable.

3. Give three examples of a discrete random variable.

4. Give three examples of a continuous random variable.

5. What is a probability distribution? Give an example.

For Exercises 6 through 11, determine whether the distribution represents a probability distribution. If it does not, state why.

6.

X	3	7	9	12	14
***P*(*X*)**	$\frac{4}{13}$	$\frac{1}{13}$	$\frac{3}{13}$	$\frac{1}{13}$	$\frac{2}{13}$

7.

X	3	6	8	12
***P*(*X*)**	0.3	0.5	0.7	−0.8

8.

X	5	7	9
***P*(*X*)**	0.6	0.8	−0.4

No. Probabilities cannot be negative.

9.

X	1	2	3	4	5
***P*(*X*)**	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$

Yes

10.

X	20	30	40	50
***P*(*X*)**	0.05	0.35	0.4	0.2

Yes

11.

X	7	14	21
***P*(*X*)**	0.3	0.1	1.7

No. A probability cannot be greater than 1.

For Exercises 12 through 18, state whether the variable is discrete or continuous.

12. The speed of a jet airplane Continuous

13. The number of cheeseburgers a fast-food restaurant serves each day Discrete

14. The number of people who play the state lottery each day Discrete

15. The weight of an automobile. Continuous

16. The time it takes to have a medical physical exam. Continuous

17. The number of mathematics majors in your school Discrete

18. The blood pressures of all patients admitted to a hospital on a specific day Continuous

For Exercises 19 through 28, construct a probability distribution for the data and draw a graph for the distribution.

19. Medical Tests The probabilities that a patient will have 0, 1, 2, or 3 medical tests performed on entering a hospital are $\frac{6}{15}$, $\frac{5}{15}$, $\frac{3}{15}$, and $\frac{1}{15}$, respectively.

20. Investment Return The probabilities of a return on an investment of $5,000, $7,000, and $9,000 are $\frac{1}{2}$, $\frac{3}{8}$, and $\frac{1}{8}$.

21. Birthday Cake Sales The probabilities that a bakery has a demand for 2, 3, 5, or 7 birthday cakes on any given day are 0.35, 0.41, 0.15, and 0.09, respectively.

22. DVD Rentals The probabilities that a customer will rent 0, 1, 2, 3, or 4 DVDs on a single visit to the rental store are 0.15, 0.25, 0.3, 0.25, and 0.05, respectively.

23. Loaded Die A die is loaded in such a way that the probabilities of getting 1, 2, 3, 4, 5, and 6 are $\frac{1}{2}$, $\frac{1}{6}$, $\frac{1}{12}$, $\frac{1}{12}$, $\frac{1}{12}$, and $\frac{1}{12}$, respectively.

24. Item Selection The probabilities that a customer selects 1, 2, 3, 4, and 5 items at a convenience store are 0.32, 0.12, 0.23, 0.18, and 0.15, respectively.

25. Student Classes The probabilities that a student is registered for 2, 3, 4, or 5 classes are 0.01, 0.34, 0.62, and 0.03, respectively.

26. Garage Space The probabilities that a randomly selected home has garage space for 0, 1, 2, or 3 cars are 0.22, 0.33, 0.37, and 0.08, respectively.

27. Selecting a Monetary Bill A box contains three \$1 bills, two \$5 bills, five \$10 bills, and one \$20 bill. Construct a probability distribution for the data if x represents the value of a single bill drawn at random and then replaced.

28. Family with Children Construct a probability distribution for a family with 4 children. Let X be the number of girls.

29. Drawing a Card Construct a probability distribution for drawing a card from a deck of 40 cards consisting of 10 cards numbered 1, 10 cards numbered 2, 15 cards numbered 3, and 5 cards numbered 4.

30. Rolling Two Dice Using the sample space for tossing two dice, construct a probability distribution for the sums 2 through 12.

Extending the Concepts

A probability distribution can be written in formula notation such as $P(X) = 1/X$, where $X = 2, 3, 6$. The distribution is shown as follows:

X	2	3	6
$P(X)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

For Exercises 31 through 36, write the distribution for the formula and determine whether it is a probability distribution.

31. $P(X) = X/6$ for $X = 1, 2, 3$

32. $P(X) = X$ for $X = 0.2, 0.3, 0.5$

33. $P(X) = X/6$ for $X = 3, 4, 7$

34. $P(X) = X + 0.1$ for $X = 0.1, 0.02, 0.04$

35. $P(X) = X/7$ for $X = 1, 2, 4$

36. $P(X) = X/(X + 2)$ for $X = 0, 1, 2$

5–2 Mean, Variance, Standard Deviation, and Expectation

Objective 2

Find the mean, variance, standard deviation, and expected value for a discrete random variable.

The mean, variance, and standard deviation for a probability distribution are computed differently from the mean, variance, and standard deviation for samples. This section explains how these measures—as well as a new measure called the *expectation*—are calculated for probability distributions.

Mean

In Chapter 3, the mean for a sample or population was computed by adding the values and dividing by the total number of values, as shown in these formulas:

$$\overline{X} = \frac{\Sigma X}{n} \qquad \mu = \frac{\Sigma X}{N}$$

But how would you compute the mean of the number of spots that show on top when a die is rolled? You could try rolling the die, say, 10 times, recording the number of spots, and finding the mean; however, this answer would only approximate the true mean. What about 50 rolls or 100 rolls? Actually, the more times the die is rolled, the better the approximation. You might ask, then, How many times must the die be rolled to get the exact answer? *It must be rolled an infinite number of times.* Since this task is impossible, the previous formulas cannot be used because the denominators would be infinity. Hence, a new method of computing the mean is necessary. This method gives the exact theoretical value of the mean as if it were possible to roll the die an infinite number of times.

Before the formula is stated, an example will be used to explain the concept. Suppose two coins are tossed repeatedly, and the number of heads that occurred is recorded. What will be the mean of the number of heads? The sample space is

HH, HT, TH, TT

Historical Note

A professor, Augustin Louis Cauchy (1789–1857), wrote a book on probability. While he was teaching at the Military School of Paris, one of his students was Napoleon Bonaparte.

and each outcome has a probability of $\frac{1}{4}$. Now, in the long run, you would *expect* two heads (HH) to occur approximately $\frac{1}{4}$ of the time, one head to occur approximately $\frac{1}{2}$ of the time (HT or TH), and no heads (TT) to occur approximately $\frac{1}{4}$ of the time. Hence, on average, you would expect the number of heads to be

$$\tfrac{1}{4} \cdot 2 + \tfrac{1}{2} \cdot 1 + \tfrac{1}{4} \cdot 0 = 1$$

That is, if it were possible to toss the coins many times or an infinite number of times, the *average* of the number of heads would be 1.

Hence, to find the mean for a probability distribution, you must multiply each possible outcome by its corresponding probability and find the sum of the products.

Formula for the Mean of a Probability Distribution

The mean of a random variable with a discrete probability distribution is

$$\mu = X_1 \cdot P(X_1) + X_2 \cdot P(X_2) + X_3 \cdot P(X_3) + \cdots + X_n \cdot P(X_n)$$
$$= \Sigma X \cdot P(X)$$

where $X_1, X_2, X_3, \ldots, X_n$ are the outcomes and $P(X_1), P(X_2), P(X_3), \ldots, P(X_n)$ are the corresponding probabilities.

Note: $\Sigma X \cdot P(X)$ means to sum the products.

Rounding Rule for the Mean, Variance, and Standard Deviation for a Probability Distribution The rounding rule for the mean, variance, and standard deviation for variables of a probability distribution is this: The mean, variance, and standard deviation should be rounded to one more decimal place than the outcome *X*. When fractions are used, they should be reduced to lowest terms.

Examples 5–5 through 5–8 illustrate the use of the formula.

Example 5–5

Rolling a Die

Find the mean of the number of spots that appear when a die is tossed.

Solution

In the toss of a die, the mean can be computed thus.

Outcome *X*	1	2	3	4	5	6
Probability *P*(*X*)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\mu = \Sigma X \cdot P(X) = 1 \cdot \tfrac{1}{6} + 2 \cdot \tfrac{1}{6} + 3 \cdot \tfrac{1}{6} + 4 \cdot \tfrac{1}{6} + 5 \cdot \tfrac{1}{6} + 6 \cdot \tfrac{1}{6}$$
$$= \tfrac{21}{6} = 3\tfrac{1}{2} \text{ or } 3.5$$

That is, when a die is tossed many times, the theoretical mean will be 3.5. Note that even though the die cannot show a 3.5, the theoretical average is 3.5.

The reason why this formula gives the theoretical mean is that in the long run, each outcome would occur approximately $\frac{1}{6}$ of the time. Hence, multiplying the outcome by its corresponding probability and finding the sum would yield the theoretical mean. In other words, outcome 1 would occur approximately $\frac{1}{6}$ of the time, outcome 2 would occur approximately $\frac{1}{6}$ of the time, etc.

Example 5–6

Children in a Family

In a family with two children, find the mean of the number of children who will be girls.

Solution

The probability distribution is as follows:

Number of girls X	0	1	2
Probability $P(X)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Hence, the mean is

$$\mu = \Sigma X \cdot P(X) = 0 \cdot \tfrac{1}{4} + 1 \cdot \tfrac{1}{2} + 2 \cdot \tfrac{1}{4} = 1$$

Example 5–7

Tossing Coins

If three coins are tossed, find the mean of the number of heads that occur. (See the table preceding Example 5–1.)

Solution

The probability distribution is

Number of heads X	0	1	2	3
Probability $P(X)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

The mean is

$$\mu = \Sigma X \cdot P(X) = 0 \cdot \tfrac{1}{8} + 1 \cdot \tfrac{3}{8} + 2 \cdot \tfrac{3}{8} + 3 \cdot \tfrac{1}{8} = \tfrac{12}{8} = 1\tfrac{1}{2} \text{ or } 1.5$$

The value 1.5 cannot occur as an outcome. Nevertheless, it is the long-run or theoretical average.

Example 5–8

Number of Trips of Five Nights or More

The probability distribution shown represents the number of trips of five nights or more that American adults take per year. (That is, 6% do not take any trips lasting five nights or more, 70% take one trip lasting five nights or more per year, etc.) Find the mean.

Number of trips X	0	1	2	3	4
Probability $P(X)$	0.06	0.70	0.20	0.03	0.01

Solution

$$\begin{aligned}\mu &= \Sigma X \cdot P(X)\\ &= (0)(0.06) + (1)(0.70) + (2)(0.20) + (3)(0.03) + (4)(0.01)\\ &= 0 + 0.70 + 0.40 + 0.09 + 0.04\\ &= 1.23 \approx 1.2\end{aligned}$$

Hence, the mean of the number of trips lasting five nights or more per year taken by American adults is 1.2.

Historical Note

Fey Manufacturing Co., located in San Francisco, invented the first three-reel, automatic payout slot machine in 1895.

Variance and Standard Deviation

For a probability distribution, the mean of the random variable describes the measure of the so-called long-run or theoretical average, but it does not tell anything about the spread of the distribution. Recall from Chapter 3 that to measure this spread or variability, statisticians use the variance and standard deviation. These formulas were used:

$$\sigma^2 = \frac{\Sigma(X - \mu)^2}{N} \qquad \text{or} \qquad \sigma = \sqrt{\frac{\Sigma(X - \mu)^2}{N}}$$

These formulas cannot be used for a random variable of a probability distribution since N is infinite, so the variance and standard deviation must be computed differently.

To find the variance for the random variable of a probability distribution, subtract the theoretical mean of the random variable from each outcome and square the difference. Then multiply each difference by its corresponding probability and add the products. The formula is

$$\sigma^2 = \Sigma[(X - \mu)^2 \cdot P(X)]$$

Finding the variance by using this formula is somewhat tedious. So for simplified computations, a shortcut formula can be used. This formula is algebraically equivalent to the longer one and is used in the examples that follow.

Formula for the Variance of a Probability Distribution

Find the variance of a probability distribution by multiplying the square of each outcome by its corresponding probability, summing those products, and subtracting the square of the mean. The formula for the variance of a probability distribution is

$$\sigma^2 = \Sigma[X^2 \cdot P(X)] - \mu^2$$

The standard deviation of a probability distribution is

$$\sigma = \sqrt{\sigma^2} \qquad \text{or} \qquad \sqrt{\Sigma[X^2 \cdot P(X)] - \mu^2}$$

Remember that the variance and standard deviation cannot be negative.

Example 5–9

Rolling a Die

Compute the variance and standard deviation for the probability distribution in Example 5–5.

Solution

Recall that the mean is $\mu = 3.5$, as computed in Example 5–5. Square each outcome and multiply by the corresponding probability, sum those products, and then subtract the square of the mean.

$$\sigma^2 = (1^2 \cdot \tfrac{1}{6} + 2^2 \cdot \tfrac{1}{6} + 3^2 \cdot \tfrac{1}{6} + 4^2 \cdot \tfrac{1}{6} + 5^2 \cdot \tfrac{1}{6} + 6^2 \cdot \tfrac{1}{6}) - (3.5)^2 = 2.9$$

To get the standard deviation, find the square root of the variance.

$$\sigma = \sqrt{2.9} = 1.7$$

Example 5–10 Selecting Numbered Balls

A box contains 5 balls. Two are numbered 3, one is numbered 4, and two are numbered 5. The balls are mixed and one is selected at random. After a ball is selected, its number is recorded. Then it is replaced. If the experiment is repeated many times, find the variance and standard deviation of the numbers on the balls.

Solution

Let X be the number on each ball. The probability distribution is

Number on ball X	3	4	5
Probability $P(X)$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$

The mean is

$$\mu = \Sigma X \cdot P(X) = 3 \cdot \tfrac{2}{5} + 4 \cdot \tfrac{1}{5} + 5 \cdot \tfrac{2}{5} = 4$$

The variance is

$$\begin{aligned} \sigma &= \Sigma[X^2 \cdot P(X)] - \mu^2 \\ &= 3^2 \cdot \tfrac{2}{5} + 4^2 \cdot \tfrac{1}{5} + 5^2 \cdot \tfrac{2}{5} - 4^2 \\ &= 16\tfrac{4}{5} - 16 \\ &= \tfrac{4}{5} \end{aligned}$$

The standard deviation is

$$\sigma = \sqrt{\frac{4}{5}} = \sqrt{0.8} = 0.894$$

The mean, variance, and standard deviation can also be found by using vertical columns, as shown.

X	$P(X)$	$X \cdot P(X)$	$X^2 \cdot P(X)$
3	0.4	1.2	3.6
4	0.2	0.8	3.2
5	0.4	2.0	10
		$\Sigma X \cdot P(X) = 4.0$	16.8

Find the mean by summing the $\Sigma X \cdot P(X)$ column, and find the variance by summing the $X^2 \cdot P(X)$ column and subtracting the square of the mean.

$$\sigma^2 = 16.8 - 4^2 = 16.8 - 16 = 0.8$$

and

$$\sigma = \sqrt{0.8} = 0.894$$

Example 5–11 On Hold for Talk Radio

A talk radio station has four telephone lines. If the host is unable to talk (i.e., during a commercial) or is talking to a person, the other callers are placed on hold. When all lines are in use, others who are trying to call in get a busy signal. The probability that 0, 1, 2, 3, or 4 people will get through is shown in the distribution. Find the variance and standard deviation for the distribution.

X	0	1	2	3	4
$P(X)$	0.18	0.34	0.23	0.21	0.04

Should the station have considered getting more phone lines installed?

Solution

The mean is

$$\mu = \Sigma X \cdot P(X)$$
$$= 0 \cdot (0.18) + 1 \cdot (0.34) + 2 \cdot (0.23) + 3 \cdot (0.21) + 4 \cdot (0.04)$$
$$= 1.6$$

The variance is

$$\sigma^2 = \Sigma[X^2 \cdot P(X)] - \mu^2$$
$$= [0^2 \cdot (0.18) + 1^2 \cdot (0.34) + 2^2 \cdot (0.23) + 3^2 \cdot (0.21) + 4^2 \cdot (0.04)] - 1.6^2$$
$$= [0 + 0.34 + 0.92 + 1.89 + 0.64] - 2.56$$
$$= 3.79 - 2.56 = 1.23$$
$$= 1.2 \text{ (rounded)}$$

The standard deviation is $\sigma = \sqrt{\sigma^2}$, or $\sigma = \sqrt{1.2} = 1.1$.

No. The mean number of people calling at any one time is 1.6. Since the standard deviation is 1.1, most callers would be accommodated by having four phone lines because $\mu + 2\sigma$ would be $1.6 + 2(1.1) = 1.6 + 2.2 = 3.8$. Very few callers would get a busy signal since at least 75% of the callers would either get through or be put on hold. (See Chebyshev's theorem in Section 3–2.)

Expectation

Another concept related to the mean for a probability distribution is that of expected value or expectation. Expected value is used in various types of games of chance, in insurance, and in other areas, such as decision theory.

The **expected value** of a discrete random variable of a probability distribution is the theoretical average of the variable. The formula is

$$\mu = E(X) = \Sigma X \cdot P(X)$$

The symbol $E(X)$ is used for the expected value.

The formula for the expected value is the same as the formula for the theoretical mean. The expected value, then, is the theoretical mean of the probability distribution. That is, $E(X) = \mu$.

When expected value problems involve money, it is customary to round the answer to the nearest cent.

Example 5–12

Winning Tickets

One thousand tickets are sold at $1 each for a color television valued at $350. What is the expected value of the gain if you purchase one ticket?

Solution

The problem can be set up as follows:

	Win	Lose
Gain X	$349	−$1
Probability $P(X)$	$\frac{1}{1000}$	$\frac{999}{1000}$

Two things should be noted. First, for a win, the net gain is \$349, since you do not get the cost of the ticket (\$1) back. Second, for a loss, the gain is represented by a negative number, in this case −\$1. The solution, then, is

$$E(X) = \$349 \cdot \frac{1}{1000} + (-\$1) \cdot \frac{999}{1000} = -\$0.65$$

Expected value problems of this type can also be solved by finding the overall gain (i.e., the value of the prize won or the amount of money won, not considering the cost of the ticket for the prize or the cost to play the game) and subtracting the cost of the tickets or the cost to play the game, as shown:

$$E(X) = \$350 \cdot \frac{1}{1000} - \$1 = -\$0.65$$

Here, the overall gain (\$350) must be used.

Note that the expectation is −\$0.65. This does not mean that you lose \$0.65, since you can only win a television set valued at \$350 or lose \$1 on the ticket. What this expectation means is that the average of the losses is \$0.65 for each of the 1000 ticket holders. Here is another way of looking at this situation: If you purchased one ticket each week over a long time, the average loss would be \$0.65 per ticket, since theoretically, on average, you would win the set once for each 1000 tickets purchased.

Example 5–13 Special Die

A special six-sided die is made in which 3 sides have 6 spots, 2 sides have 4 spots, and 1 side has 1 spot. If the die is rolled, find the expected value of the number of spots that will occur.

Solution

Since there are 3 sides with 6 spots, the probability of getting a 6 is $\frac{3}{6} = \frac{1}{2}$. Since there are 2 sides with 4 spots, the probability of getting 4 spots is $\frac{2}{6} = \frac{1}{3}$. The probability of getting 1 spot is $\frac{1}{6}$ since 1 side has 1 spot.

Gain X	1	4	6
Probability $P(X)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

$$E(X) = 1 \cdot \tfrac{1}{6} + 4 \cdot \tfrac{1}{3} + 6 \cdot \tfrac{1}{2} = 4\tfrac{1}{2}$$

Notice you can only get 1, 4, or 6 spots; but if you rolled the die a large number of times and found the average, it would be about $4\frac{1}{2}$.

Example 5–14 Bond Investment

A financial adviser suggests that his client select one of two types of bonds in which to invest \$5000. Bond X pays a return of 4% and has a default rate of 2%. Bond Y has a $2\frac{1}{2}$% return and a default rate of 1%. Find the expected rate of return and decide which bond would be a better investment. When the bond defaults, the investor loses all the investment.

Solution

The return on bond X is $\$5000 \cdot 4\% = \$200$. The expected return then is

$$E(X) = \$200(0.98) - \$5000(0.02) = \$96$$

The return on bond Y is $\$5000 \cdot 2\frac{1}{2}\% = \$125$. The expected return then is

$$E(X) = \$125(0.99) - \$5000(0.01) = \$73.75$$

Hence, bond X would be a better investment since the expected return is higher.

In gambling games, if the expected value of the game is zero, the game is said to be fair. If the expected value of a game is positive, then the game is in favor of the player. That is, the player has a better than even chance of winning. If the expected value of the game is negative, then the game is said to be in favor of the house. That is, in the long run, the players will lose money.

In his book *Probabilities in Everyday Life* (Ivy Books, 1986), author John D. McGervy gives the expectations for various casino games. For keno, the house wins \$0.27 on every \$1.00 bet. For Chuck-a-Luck, the house wins about \$0.52 on every \$1.00 bet. For roulette, the house wins about \$0.90 on every \$1.00 bet. For craps, the house wins about \$0.88 on every \$1.00 bet. The bottom line here is that if you gamble long enough, sooner or later you will end up losing money.

Applying the Concepts 5–2

Expected Value

On March 28, 1979, the nuclear generating facility at Three Mile Island, Pennsylvania, began discharging radiation into the atmosphere. People exposed to even low levels of radiation can experience health problems ranging from very mild to severe, even causing death. A local newspaper reported that 11 babies were born with kidney problems in the three-county area surrounding the Three Mile Island nuclear power plant. The expected value for that problem in infants in that area was 3. Answer the following questions.

1. What does *expected value* mean?
2. Would you expect the exact value of 3 all the time?
3. If a news reporter stated that the number of cases of kidney problems in newborns was nearly four times as much as was usually expected, do you think pregnant mothers living in that area would be overly concerned?
4. Is it unlikely that 11 occurred by chance?
5. Are there any other statistics that could better inform the public?
6. Assume that 3 out of 2500 babies were born with kidney problems in that three-county area the year before the accident. Also assume that 11 out of 2500 babies were born with kidney problems in that three-county area the year after the accident. What is the real percent of increase in that abnormality?
7. Do you think that pregnant mothers living in that area should be overly concerned after looking at the results in terms of rates?

See page 298 for the answers.