The Binomial Distribution

Many types of probability problems have only two outcomes or can be reduced to two outcomes.

For example, when a coin is tossed, it can land heads or tails. When a baby is born, it will be either male or female. In a basketball game, a team either wins or loses.

A true/false item can be answered in only two ways, true or false. Other situations can be reduced to two outcomes

For example, a medical treatment can be classified as effective or ineffective, depending on the results.

A person can be classified as having normal or abnormal blood pressure, depending on the measure of the blood pressure gauge.

A multiple-choice question, even though there are four or five answer choices, can be classified as correct or incorrect.

Situations like these are called *binomial experiments.*

A binomial experiment is a probability experiment that satisfies the following four requirements:

1. **There must be a fixed number of trials.**
2. **Each trial can have only two outcomes or outcomes that can be reduced to two outcomes. These outcomes can be considered as either success or failure.**
3. **The outcomes of each trial must be independent of one another.**
4. **The probability of a success must remain the same for each trial.**

A binomial experiment and its results give rise to a special probability distribution called the *binomial distribution.*

The outcomes of a binomial experiment and the corresponding probabilities of these outcomes are called a binomial distribution.

In binomial experiments, the outcomes are usually classified as successes or failures.

For example, the correct answer to a multiple-choice item can be classified as a success,

but any of the other choices would be incorrect and hence classified as a failure.

The notation that is commonly used for binomial experiments and the binomial distribution is defined now.

Notation for the Binomial Distribution

P(S) The symbol for the probability of success

P(F) The symbol for the probability of failure

P The numerical probability of a success

q The numerical probability of a failure

$$P(S) = p \quad \text{and} \quad P(F) = 1 - p = q$$

n The number of trials

X The number of successes in n trials

Note that $0 \leq X \leq n$ and $X = 0, 1, 2, 3, \ldots, n$.

The probability of a success in a binomial experiment can be computed with this Formula

Binomial Probability Formula

In a binomial experiment, the probability of exactly *X* successes in *n* trials is

$$p_{(x)} = \frac{n!}{(n-X)!\,x!} \cdot p^{x} \cdot q^{n-x}$$

$n! = n(n - 1)(n - 2)\ldots\ldots. 1$(Factorial Notation)

5!= 5 . 4 . 3 . 2 . 1

9! = 9 . 8 . 7 . 6 . 5 . 4 . 3 . 2 . 1

0! = 1

Example 1: Tossing Coins

A coin is tossed 3 times. Find the probability of getting exactly two heads.

Solution

This problem can be solved by looking at the sample space. There are three ways to get

two heads.

HHH, HHT, HTH, THH, TTH, THT, HTT, TTT

The answer is , or 0.375.

Looking at the problem in Example 1from the standpoint of a binomial experiment, one can show that it meets the four requirements.

1. There are a fixed number of trials (three).
2. There are only two outcomes for each trial, heads or tails.
3. The outcomes are independent of one another (the outcome of one toss in no way affects the outcome of another toss).
4. The probability of a success (heads) is in each case.

In this case, *n* = 3, *X* = 2, *p=1/2* , and *q=1l2* . Hence, substituting in the formula gives

$$P(2\ \text{heads}) = \frac{3!}{(3-2)!2!} \cdot \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^1 = \frac{3}{8} = 0.375$$

Example 2 Survey on Doctor Visits

A survey found that one out of five Americans say he or she has visited a doctor in any given month. If 10 people are selected at random, find the probability that exactly 3 will have visited a doctor last month.

Solution

In this case, $n = 10$, $X = 3$, $p = 1/5$ **, and** $q = 1 - p = 4/5$ **.**

Hence,

$$p(3) = \frac{10!}{(10-3)!\,3!}\left(\frac{1}{5}\right)^{3}\left(\frac{4}{5}\right)^{7} = 0.201$$

Example 3 Survey on Employment

A survey from Teenage Research Unlimited found that 30% of teenage consumers receive their spending money from part-time jobs. If 5 teenagers are selected at random, find the probability that at least 3 of them will have part-time jobs.

Solution

To find the probability that at least 3 have part-time jobs, it is necessary to find the individual probabilities for 3, or 4, or 5 and then add them to get the total probability

$$p(3) = \frac{5!}{(5-3)!\,3!}(0\cdot 3)^3(0\cdot 7)^2 = 0.132$$

$$p(4) = \frac{5!}{(5-4)!\,4!}(0\cdot 3)^4(0\cdot 7)^1 = 0.028$$

$$p(5) = \frac{5!}{(5-5)!\,5!}(0\cdot 3)^5(0\cdot 7)^0 = 0.002$$

Hence,

P(at least three teenagers have part-time jobs)

= 0.132 + 0.028 + 0.002 = 0.162

Mean, Variance, and Standard Deviation for the Binomial Distribution

The mean, variance, and standard deviation of a variable that has the *binomial distribution* can

be found by using the following formulas.

Mean: $\mu = n \cdot p$

Variance:=$\sigma^2 = n \cdot p \cdot q$

Standard deviation= $\sigma = \sqrt{n \cdot p \cdot q}$

Example 4tossing a Coin

A coin is tossed 4 times. Find the mean, variance, and standard deviation of the number

of heads that will be obtained.

Solution

With the formulas for the binomial distribution and $n = 4$, $p = 1/2$, and $q=1/2$, the results are

$$\textbf{Mean: } \mu = n \cdot p = 4.\frac{1}{2} = \frac{4}{2} = 2$$

$$\textbf{Variance:=}\sigma^2 = n \cdot p \cdot q = 4.\frac{1}{2}.\frac{1}{2} = \frac{4}{4} = 1$$

$$\textbf{Standard deviation= } \sigma = \sqrt{n \cdot p \cdot q} = \sqrt{1} = 1$$

Example 5 Rolling a Die

A die is rolled 480 times. Find the mean, variance, and standard deviation of the number of 3s that will be rolled.

Solution

This is a binomial experiment since getting a 3 is a success and not getting a 3 is considered a failure.

Hence n 480,p =1/6 , and q=5/6.

Mean: $\mu = n \cdot p =$ **480.1/6=80**

Variance:=$\sigma^2 = n \cdot p \cdot q =$ **480. 1/6. 5/6=66.67**

Standard deviation= $\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{66.67}$**=8.16**

Example 6 Likelihood of Twins

The *Statistical Bulletin* published by Metropolitan Life Insurance Co. reported that 2% of all American births result in twins. If a random sample of 8000 births is taken, find the mean, variance, and standard deviation of the number of births that would result in twins.

Solution

This is a binomial situation, since a birth can result in either twins or not twins (i.e., two outcomes).

Mean: $\mu = n \cdot p =$ **(8000)(0.02) = 160**

Variance:=$\sigma^2 = n \cdot p \cdot q =$ **(8000)(0.02)(0.98) = 156.8**

Standard deviation= $\sigma = \sqrt{n \cdot p \cdot q} = \sqrt{156.8}$**=12.5**

Exercises

Which of the following are binomial experiments or can be reduced to binomial experiments?

a. Surveying 100 people to determine if they like Sudsy Soap

Yes

***b.* Tossing a coin 100 times to see how many heads occur**

Yes

***c.* Drawing a card with replacement from a deck and getting a heart**

Yes

***d.* Asking 1000 people which brand of cigarettes they smoke**

No

***e.* Testing four different brands of aspirin to see which brands are effective**

No

_f._ Testing one brand of aspirin by using 10 people todetermine whether it is effective

Yes

_g._ Asking 100 people if they smoke

Yes

_h._ Checking 1000 applicants to see whether they were admitted to College

Yes

i. Surveying 300 prisoners to see how many different crimes they were convicted of

No

_j._ Surveying 300 prisoners to see whether this is their first offense

Yes

Other Types of Distributions (Optional)

In addition to the binomial distribution, other types of distributions are used in statistics.

Three of the most commonly used distributions are the multinomial distribution, the

Poisson distribution, and the hypergeometric distribution.

The Multinomial Distribution

Recall that in order for an experiment to be binomial, two outcomes are required for each trial. But if each trial in an experiment has more than two outcomes, a distribution called the multinomial distribution must be used.

For example, a survey might require the responses of “approve,” “disapprove,” or “no opinion.” In another situation, a person may have a choice of one of five activities for Friday night, such as a movie, dinner, baseball game, play, or party. Since these situations have more than two possible outcomes

for each trial, the binomial distribution cannot be used to compute probabilities.

The multinomial distribution can be used for such situations if the probabilities for each trial remain constant and the outcomes are independent for a fixed number of trials.

The events must also be mutually exclusive.

Formula for the Multinomial Distribution

If *X* consists of events *E*1, *E*2, *E*3, . . . , *Ek*, which have corresponding probabilities *p*1, *p*2, *p*3, . . . ,

***pk* of occurring, and *X*1 is the number of times *E*1 will occur, *X*2 is the number of times *E*2 will**

occur, *X*3 is the number of times *E*3 will occur, etc., then the probability that *X* will occur is

$$P(x) = \frac{n!}{x_1! \cdot x_2! \cdot x_3! \cdot \cdots x_k!} \cdot p_1^{x_1} \cdot p_2^{x_2} \cdot \cdots pk^{x_k}$$

where

$$x_1 + x_2 + x_3 + \cdots + x_k = n$$

$$p_1 + p_2 + p_3 + \cdots + p_k = 1$$

Example

In a large city, 50% of the people choose a movie, 30% choose dinner and a play, and 20% choose shopping as an activity. If a sample of 5 people is randomly

selected, find the probability that 3 are planning to go to a movie, 1 to a play, and 1 to a shopping mall.

Solution

We know that $n = 5$, $X1 = 3$, $X2 = 1$, $X3 = 1$, $p1 = 0.50$, $p2 = 0.30$, **and** $p3 = 0.20$.

Substituting in the formula gives

$$P(x) = \frac{5!}{3! \cdot 1! \cdot 1!} \cdot (0 \cdot 50)^3 (0 \cdot 30)^1 (0.20)^1 = 0.15$$

Thus, the multinomial distribution is similar to the binomial distribution but has the advantage of allowing you to compute probabilities when there are more than two outcomes for each trial in the experiment.

That is, the multinomial distribution is a general distribution, and the binomial distribution is a special case of the multinomial distribution

The Poisson Distribution

A discrete probability distribution that is useful when *n* is large and *p* is small and when the independent variables occur over a period of time is called the Poisson distribution.

In addition to being used for the stated conditions (i.e., *n* is large, *p* is small, and the variables occur over a period of time),

the Poisson distribution can be used when a density of items is distributed over a given area or volume, such as the number of plants growing per acre or the number of defects in a given length of videotape.

Formula for the Poisson Distribution

The probability of *X* occurrences in an interval of time, volume, area, etc., for a variable where

(λ)(Greek letter lambda) is the mean number of occurrences per unit (time, volume, area, etc.) is

$$p(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

where *X* = 0, 1, 2, . . .

The letter *e* is a constant approximately equal to 2.7183.

Example:

If there are 200 typographical errors randomly distributed in a 500-page manuscript, find the probability that a given page contains exactly 3 errors.

Solution

First, find the mean number (λ) of errors. Since there are 200 errors distributed over 500 pages, each page has an average of

$$(\lambda) = 200/500 = 2/5 = 0.4$$

or 0.4 error per page. Since *X* = 3, substituting into the formula yields

$$p(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{(2{\cdot}7183)^{-0{\cdot}4}(0{\cdot}4)^{3}}{3!} = 0.0072$$

Thus, there is less than a 1% chance that any given page will contain exactly 3 errors

The Normal Distribution

What Is Normal?

Medical researchers have determined so-called normal intervals for a person's blood pressure, cholesterol, and the like. For example, the normal range of systolic blood pressure is 110 to 140. The normal interval for a person's sugar is from 72 to 99 milligrams per deciliter (mg/dl). By measuring these variables, a physician can determine if a patient's vital statistics are within the normal interval or if some type of treatment is needed to correct a condition and avoid future illnesses.

The question then is, How does one determine the so-called normal intervals? In this lesson, you will learn how researchers determine normal intervals for specific medical tests by using a normal distribution. You will see how the same methods are used to determine the lifetimes of batteries, the strength of ropes, and many other traits.

Random variables can be either discrete or continuous. Discrete variables and their distributions were explained in. Recall that a discrete variable cannot assume all values between any two given values of the variables. On the other hand, a continuous variable can assume all values between any two given values of the variables.

Examples of continuous variables are the heights of adult men, body temperatures of rats, and cholesterol levels of adults. Many continuous variables, such as the examples just mentioned,

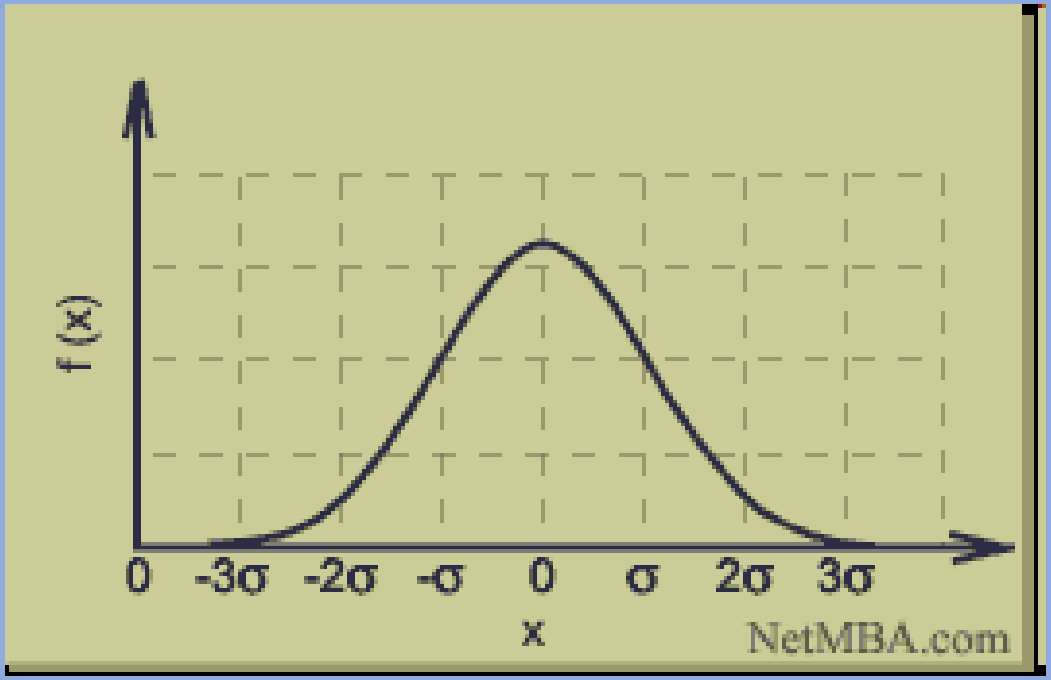
have distributions that are bell-shaped, and these are called *approximately normally distributed variables*.

if a researcher selects a random sample of 100 adult

women, measures their heights, and constructs a histogram, if it were possible to measure exactly the heights of all

adult females in the United States and plot them, the histogram would approach what is

called a *normal distribution,*



Normal distribution for the population

No variable fits a normal distribution perfectly, since a normal distribution is a theoretical distribution. However, a normal distribution can be used to describe many variables, because the deviations from a normal distribution are very small.

In mathematics, curves can be represented by equations. In a similar manner, the theoretical curve, called a *normal distribution curve,*

can be used to study many variables that are not perfectly normally distributed but are nevertheless approximately normal.

The mathematical equation for a normal distribution is

$$y = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt[\sigma]{2\pi}}$$

Where: $e = 2.718$

$\pi = 3.14$

The mathematical equation for a normal distribution is

$$y = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sqrt[\sigma]{2\pi}}$$

Where: $e = 2.718$

$\pi = 3.14$

$\mu =$ population mean

σ= population standard deviation

The shape and position of a normal distribution curve depend on two parameters, the *mean* and the *standard deviation*. Each normally distributed variable has its own normal distribution curve, which depends on the values of the variable's mean and standard deviation.

A normal distribution is a continuous, symmetric, bell-shaped distribution of a variable.

The properties of a normal distribution, including those mentioned in the definition, are explained next:

1. A normal distribution curve is bell-shaped.

2. The mean, median, and mode are equal and are located at the center of the distribution.

3. A normal distribution curve is unimodal (i.e., it has only one mode).

4. **The curve is symmetric about the mean, which is equivalent to saying that its shape is the same on both sides of a vertical line passing through the center.**

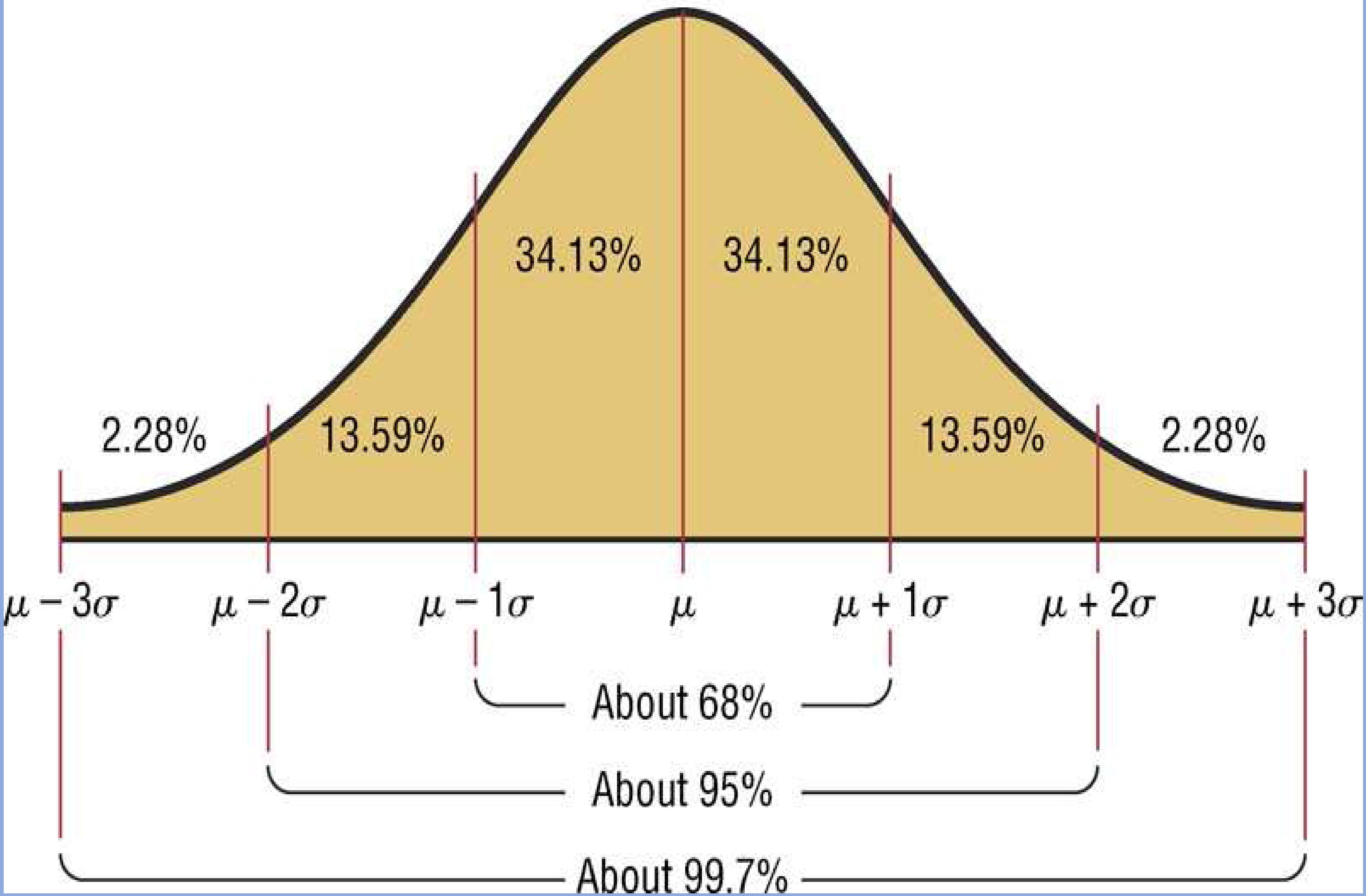
5. **The curve is continuous; that is, there are no gaps or holes. For each value of *X*, there is a corresponding value of *Y*.**

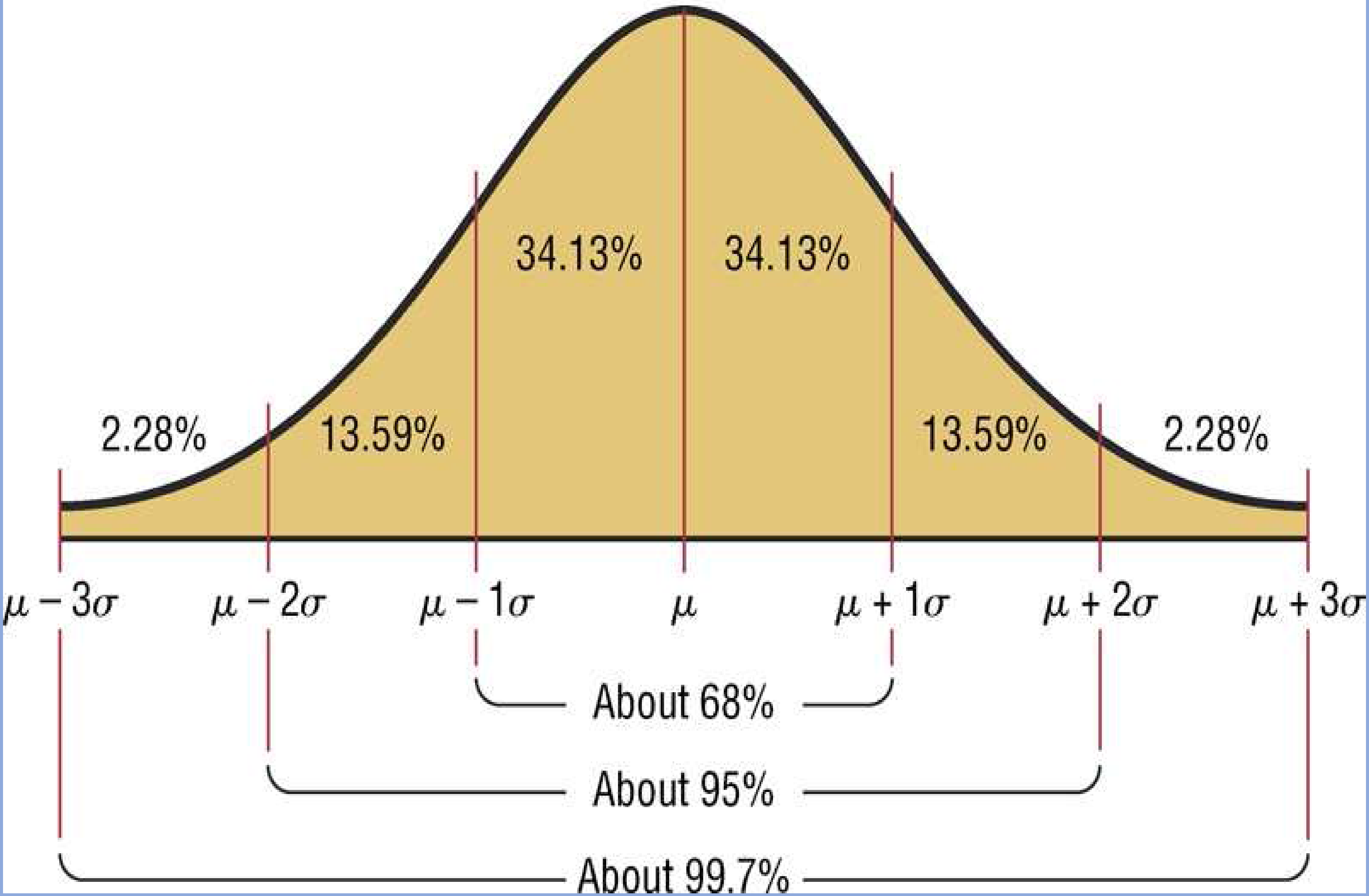
6. **The curve never touches the *x* axis. Theoretically, no matter how far in either direction the curve extends, it never meets the *x* axis—but it gets increasingly closer.**

7. The total area under a normal distribution curve is equal to 1.00, or 100%. This fact may seem unusual, since the curve never touches the *x* axis, but one can prove it mathematically by using calculus.

8. The area under the part of a normal curve that lies within 1 standard deviation of the mean is approximately 0.68, or 68%; within 2 standard deviations, about 0.95, or 95%;

and within 3 standard deviations, about 0.997, or 99.7%. See Figure , which also shows the area in each region.


34.13%
34.13%
2.28%
13.59%
13.59%
2.28%
μ − 3σ
μ − 2σ
μ − 1σ
μ
μ + 1σ
μ + 2σ
μ + 3σ
About 68%
About 95%
About 99.7%



34.13%
34.13%
2.28%
13.59%
13.59%
2.28%
μ − 3σ
μ − 2σ
μ − 1σ
μ
μ + 1σ
μ + 2σ
μ + 3σ
About 68%
About 95%
About 99.7%

The Standard Normal Distribution

Since each normally distributed variable has its own mean and standard deviation, as stated earlier, the shape and location of these curves will vary. In practical applications,

then, you would have to have a table of areas under the curve for each variable. To simplify this situation, statisticians use what is called the *standard normal distribution.*

The standard normal distribution is a normal distribution with a mean of 0 and a standard deviation of 1.

The standard normal distribution is shown in Figure

The standard normal distribution is shown in Figure

The values under the curve indicate the proportion of area in each section. For example,

the area between the mean and 1 standard deviation above or below the mean is

about 0.3413, or 34.13%.

The formula for the standard normal distribution is

$$y = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$$

All normally distributed variables can be transformed into the standard normally distributed variable by using the formula for the standard score:

***Z*= value - mean**

standard deviation

or $z = \frac{x-\mu}{\sigma}$

the area under a normal distribution curve is used to solve practical application problems, such as finding the percentage of adult women whose height is between 5 feet 4 inches and 5 feet 7 inches, or finding the probability that a new battery

will last longer than 4 years. Hence, the major emphasis of this section will be to show the procedure for finding the area under the standard normal distribution curve for any *z* value. The applications will be shown. Once the *X* values are transformed by using the preceding formula, they are called *z* values.

The ***z* value** or ***z* score** is actually the number of standard deviations that a particular *X* value is away from the mean. Table E in Appendix C gives the area (to four decimal places) under the standard normal curve for any *z* value from 3.49 to 3.49.

See P. 304 in the book