Chapter 05

Continuous Probability Distributions (The Normal Distribution)

Biostatistics For the Health Sciences

Dr. Moustafa Omar Ahmed Abu-Shawiesh
Professor of Statistics

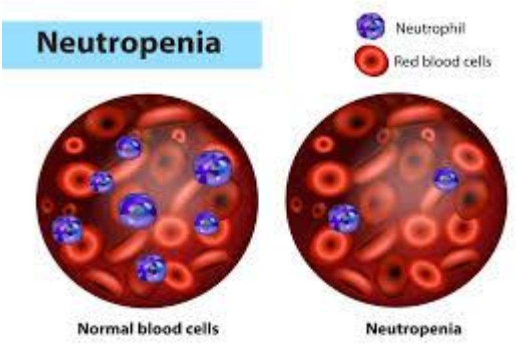
5.1 Introduction

This chapter discusses continuous probability distributions. Specifically, the normal distribution—the most widely used distribution in statistical work—is explored in depth. The normal, or Gaussian or “bell-shaped”, distribution is the cornerstone of most methods of estimation and hypothesis testing.

Many random variables, such as distribution of birthweights or blood pressures in the general population, tend to approximately follow a normal distribution. In addition, many random variables that are not themselves normal are closely approximated by a normal distribution when summed many times.

EXAMPLE 5.1

Infectious Disease The number of neutrophils in a sample of 2 white blood cells is not normally distributed, but the number of neutrophils in a random sample of 100 white blood cells is very close to being normally distributed.



In such cases, using the normal distribution is desirable because it is easy to use and tables for the normal distribution are more widely available than are tables for many other distributions.

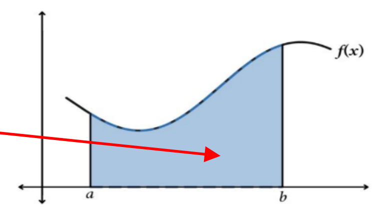
5.2 General Concepts

We want to develop an analog for a continuous random variable to the concept of a probability mass function (pmf), as was developed for a discrete random variable. Thus, we would like to know which values are more probable than others and how probable they are. If the random variable X is continuous, then the concept of a probability mass function (pmf) cannot be used. Instead we will use the concept of a probability density function (pdf), which can be defined as follows:

DEFINITION 5.1

The probability density function (pdf) of the continuous random variable, say X, is a function $f(x)$ such that the area under the density function curve between any two points a and b is equal to the probability that the random variable X falls between a and b.

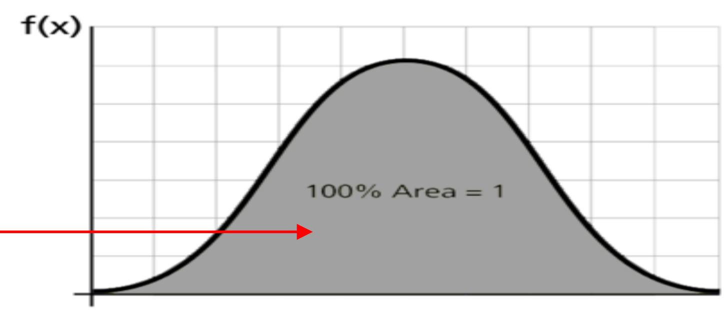
$$P(a \leq X \leq b) = \int_a^b f(x)dx;$$



$$Area = \int_a^b f(x)dx$$

Thus, the total area under the density function curve $f(x)$ over the entire range of possible values for the random variable is equal to 1, that is:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$



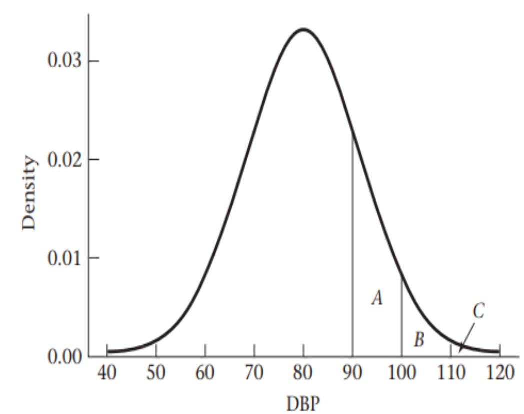
Notations

- The pdf has large values in regions of high probability and small values in regions of low probability.
- For a continuous random variable X, the probability of exactly obtaining any value is 0, that is, $P(X = k) = 0$, where k is any value (constant).
- Not all continuous random variables have symmetric bell-shaped distributions as given in Example (1) and Example (2) below.

Example (1)

Hypertension The distribution of diastolic blood-pressure (DBP) measurements in 35 – to - 44 year old men is a symmetric continuous random variable whose pdf appears in Figure 5.1.

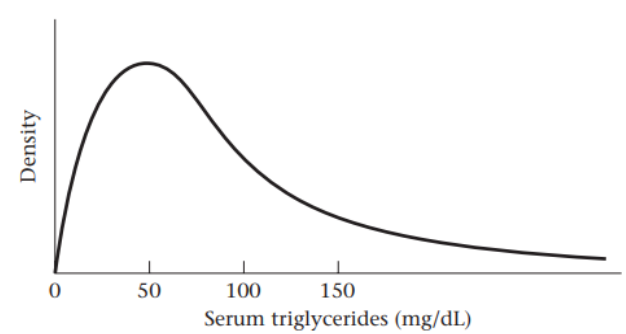
Figure 5.1 The pdf of DBP in 35- to 44-year-old men



Example (2)

Cardiovascular Disease The distribution of Serum triglyceride level is an asymmetric, positively skewed, continuous random variable whose pdf appears in Figure 5.2.

FIGURE 5.2 The pdf for serum triglycerides



5.3 The Normal Distribution

The normal distribution is the most widely used continuous distribution. It is also frequently called the Gaussian distribution, after the well-known mathematician Karl Friedrich Gauss. For example, body weights and DBPs for a group of 35-to-44 year old men approximately follow a normal distribution.

Karl Friedrich Gauss (1777–1855)

Many other distributions that are not themselves normal can be made approximately normal by transforming the data onto a different scale. For example, the distribution of serum-triglyceride concentrations from this same group of 35-to-44 year old men is likely to be positively skewed. However, the log transformation of these measurements usually follows a normal distribution.

Generally speaking, any random variable that can be expressed as a sum of many other random variables can be well approximated by a normal distribution. For example, many physiologic measures are determined in part by a combination of several genetic and environmental risk factors and can often be well approximated by a normal distribution. Thus, the normal distribution is vital to statistical work, and most estimation procedures and hypothesis tests that we will study assume the random variable being considered has an underlying normal distribution.

Notation

An important area of application of the normal distribution is as an approximating distribution to other distributions. The normal distribution is generally more convenient to work with than any other distribution, particularly in hypothesis testing. Thus, if an accurate normal approximation to some other distribution can be found, we often will want to use it.

DEFINITION 5.5 The **normal distribution** is defined by its pdf, which is given as

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right], \quad -\infty < x < \infty$$

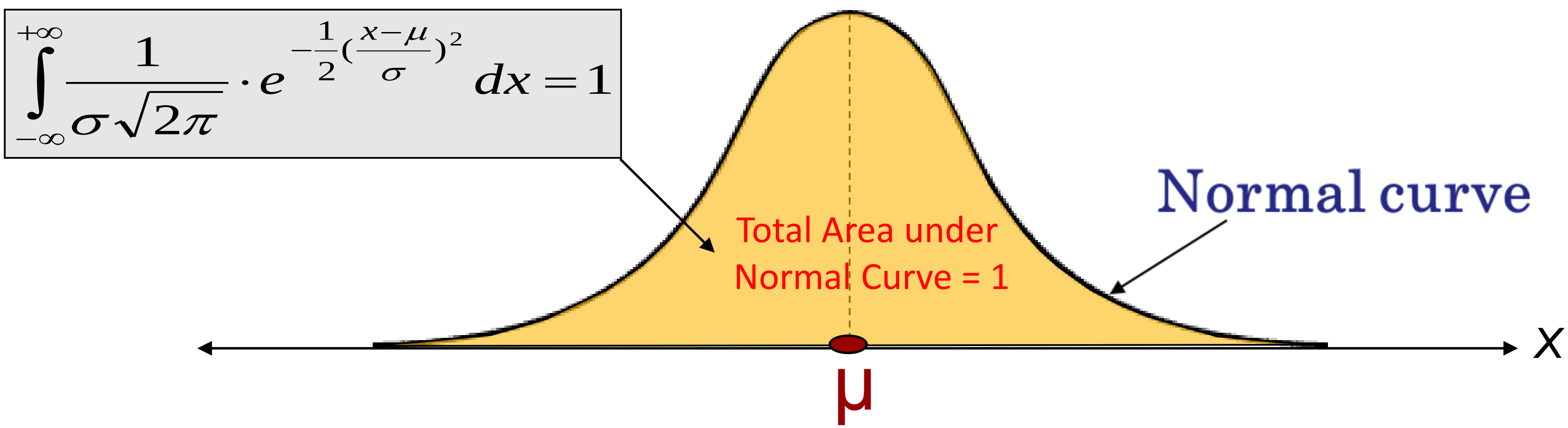
for some parameters μ, σ, where $\sigma > 0$.

From the definition of the normal distribution it can be shown, using calculus methods, that μ and σ^2 are, respectively, the expected value (mean) and the variance of this distribution ($\sigma = \sqrt{\sigma^2} > 0$ is the standard deviation).

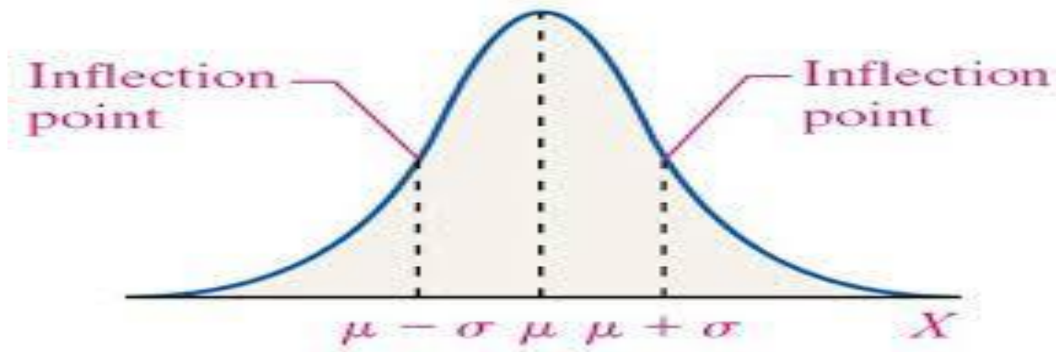
DEFINITION 5.6 A normal distribution with mean μ and variance σ^2 will generally be referred to as an $N(\mu,\sigma^2)$ distribution.

Some Properties for the Normal Distribution

- The value of π = 3.14159 and the value of exponential function is e = 2.71828.
- The parameters of distribution are μ (population mean) and σ^2(population variance) implies that $X \sim N(\mu, \sigma^2)$. Note that the second parameter is always the variance σ^2, not the standard deviation σ.
- The density function follows a bell-shaped curve, with the mode (the most frequently occurring value) at μ. Also, Note that the area under any normal density function must be 1 as shown below:



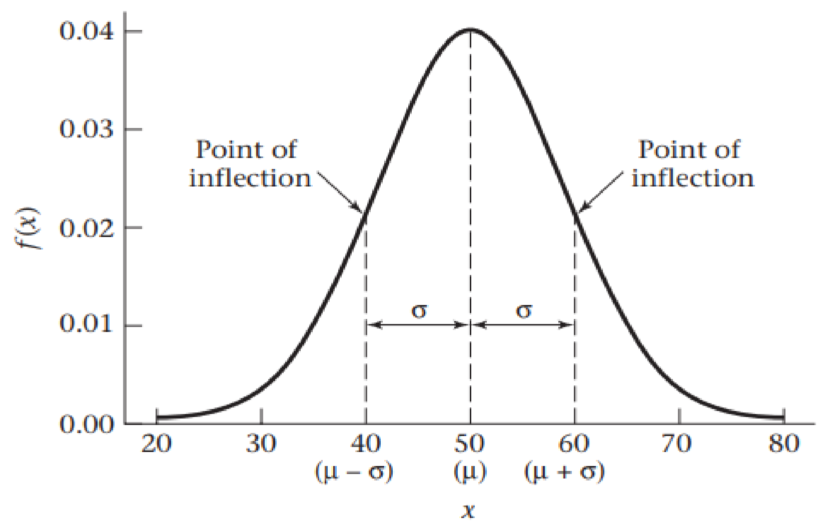
- The normal distribution curve is symmetric about μ, with points of inflection (*a point of inflection is a point at which the slope of the curve changes direction*) on either side of μ at μ − σ and μ + σ, respectively, as shown below:



Example

In this example, the pdf for a normal distribution with μ = 50 and σ^2 = 100, that is $X \sim N(50, 100)$, is plotted in Figure 5.5 shown below:

FIGURE 5.5 The pdf for a normal distribution with mean μ (50) and variance σ² (100)



- A property of the normal distribution is that the height = $\frac{1}{\sqrt{2\pi\sigma}}$. Thus, the height is inversely proportional to σ.

- The entire shape of the normal distribution is determined by the two parameters μ and σ^2 as shown below in the two figures:

FIGURE 5.6 Comparison of two normal distributions with the same variance and different means

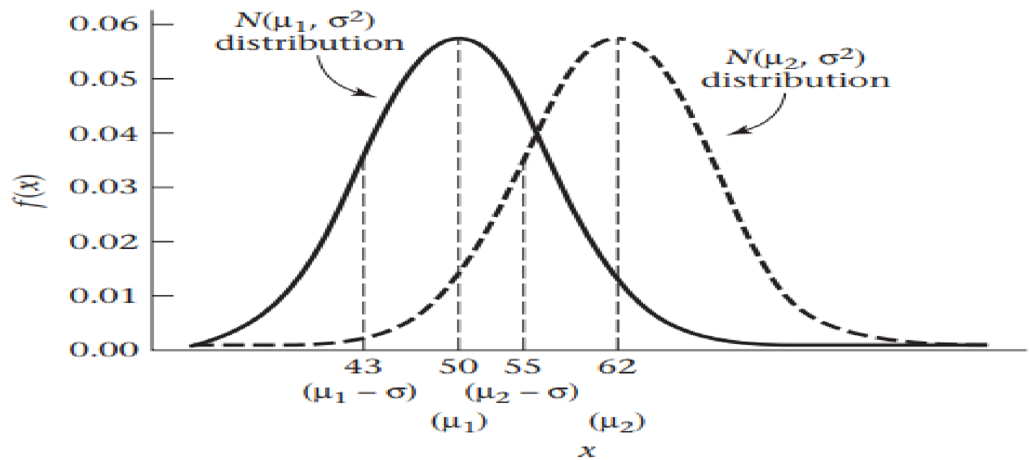
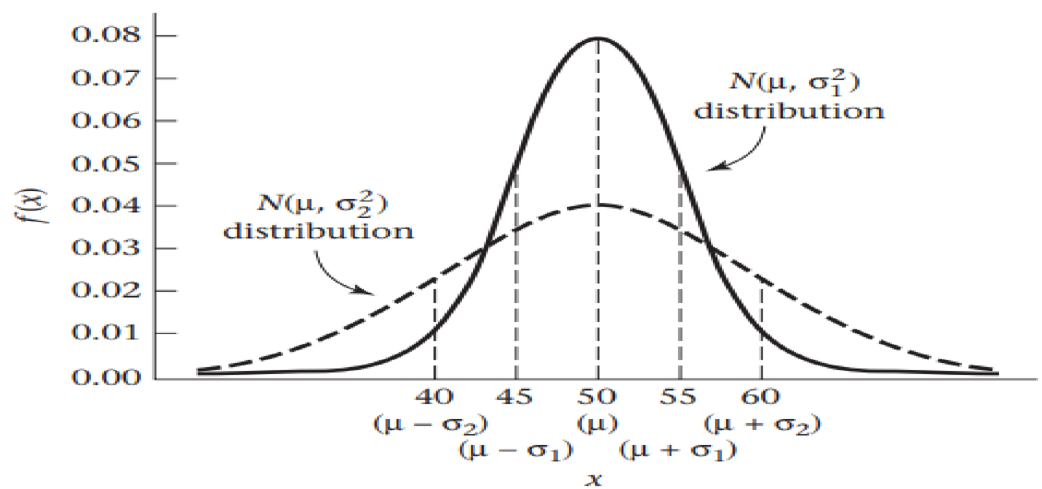


FIGURE 5.7 Comparison of two normal distributions with the same means and different variances



- The mean, median, and mode for the normal distribution are equal.

5.4 The Standard Normal Distribution

In this section, we will see that any information concerning a normal distribution, N(μ, σ^2), can be obtained from a special case, N(0, 1), called the standard normal distribution.

DEFINITION 5.7 A normal distribution with mean 0 and variance 1 is called a **standard**, or **unit**, normal distribution. This distribution is also called an $N(0,1)$ distribution.

To become familiar with the standard normal distribution, N(0, 1), we will discuss some of its properties as follows:

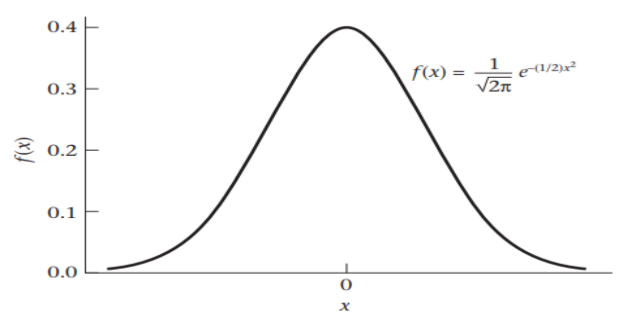
- First, the pdf in this case, mean μ = 0 and variance σ^2 = 1 (or standard deviation σ = 1), reduces to the following equation:

EQUATION 5.1

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{(-1/2)x^2}, \quad -\infty < x < +\infty$$

- The standard normal distribution is symmetric about the mean (μ = 0), because $f(x) = f(-x)$, as shown in Figure 5.8.

FIGURE 5.8 The pdf for a standard normal distribution



- There are three relationships that can be used to express the area under the standard normal density curve as follows:

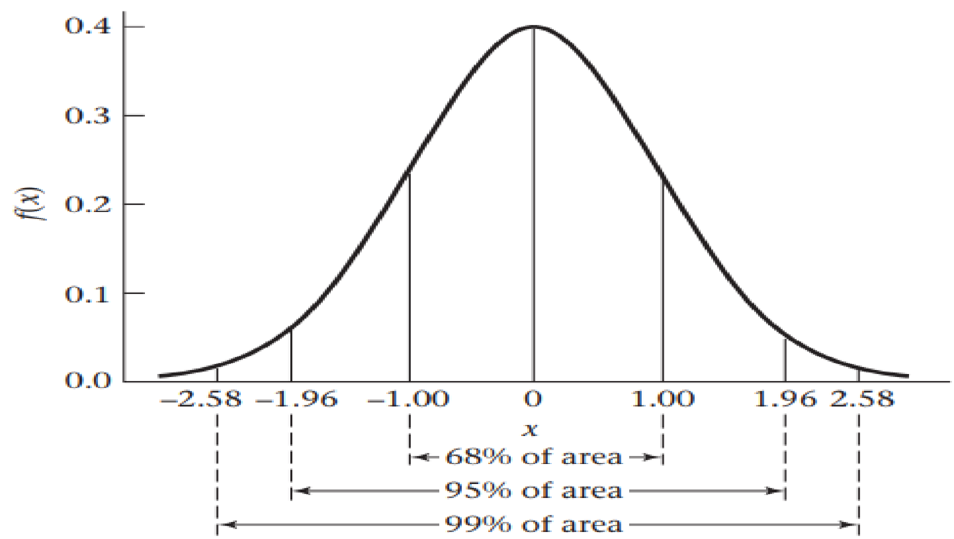
EQUATION 5.2 It can be shown that about 68% of the area under the standard normal density lies between +1 and –1, about 95% of the area lies between +2 and –2, and about 99% lies between +2.5 and –2.5.

These relationships can be expressed more precisely by saying that:

- $P(-1 < X < 1) = 0.6827 \approx 0.68$ (68%).
- $P(-1.96 < X < 1.96) = 0.95$ (95%).
- $P(-2.576 < X < 2.576) = 0.99$ (99%).

Figure 5.9 shows these relationships:

FIGURE 5.9 Empirical properties of the standard normal distribution

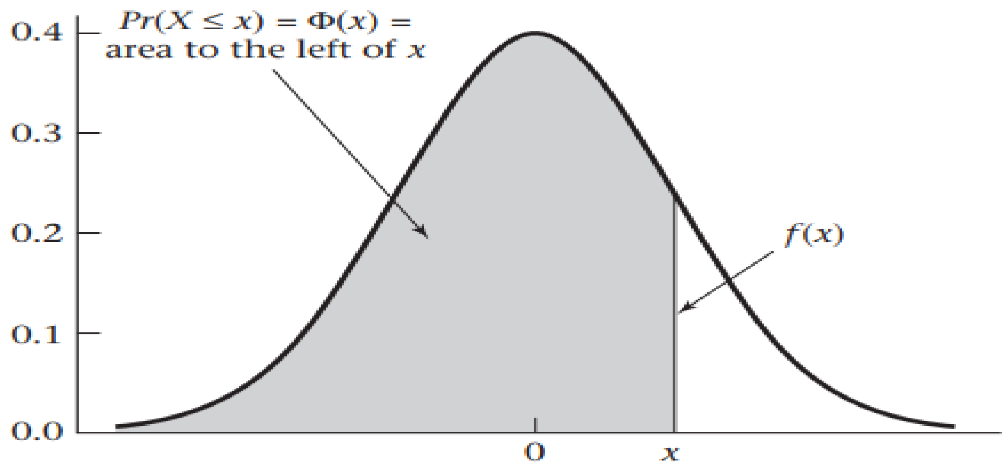


The tables of the area (probability) under the normal distribution curve, or so-called normal tables, take advantage of the symmetry properties of the normal distribution and generally are concerned with areas for positive values of x.

DEFINITION 5.8

The cumulative distribution function (CDF) for a standard normal distribution is denoted by $\Phi(x) = P(X \le x)$ where X follows an N(0,1) distribution, that is, X~N(0,1). This function is shown in Figure 5.10 as follows:

FIGURE 5.10 The cdf [$\Phi(x)$] for a standard normal distribution



Note that there is no closed-form algebraic expression for areas (probabilities) under the normal distribution. Hence, numerical methods must be used to calculate these areas (probabilities), which are generally displayed in the "normal tables" usually called the standard normal distribution table.

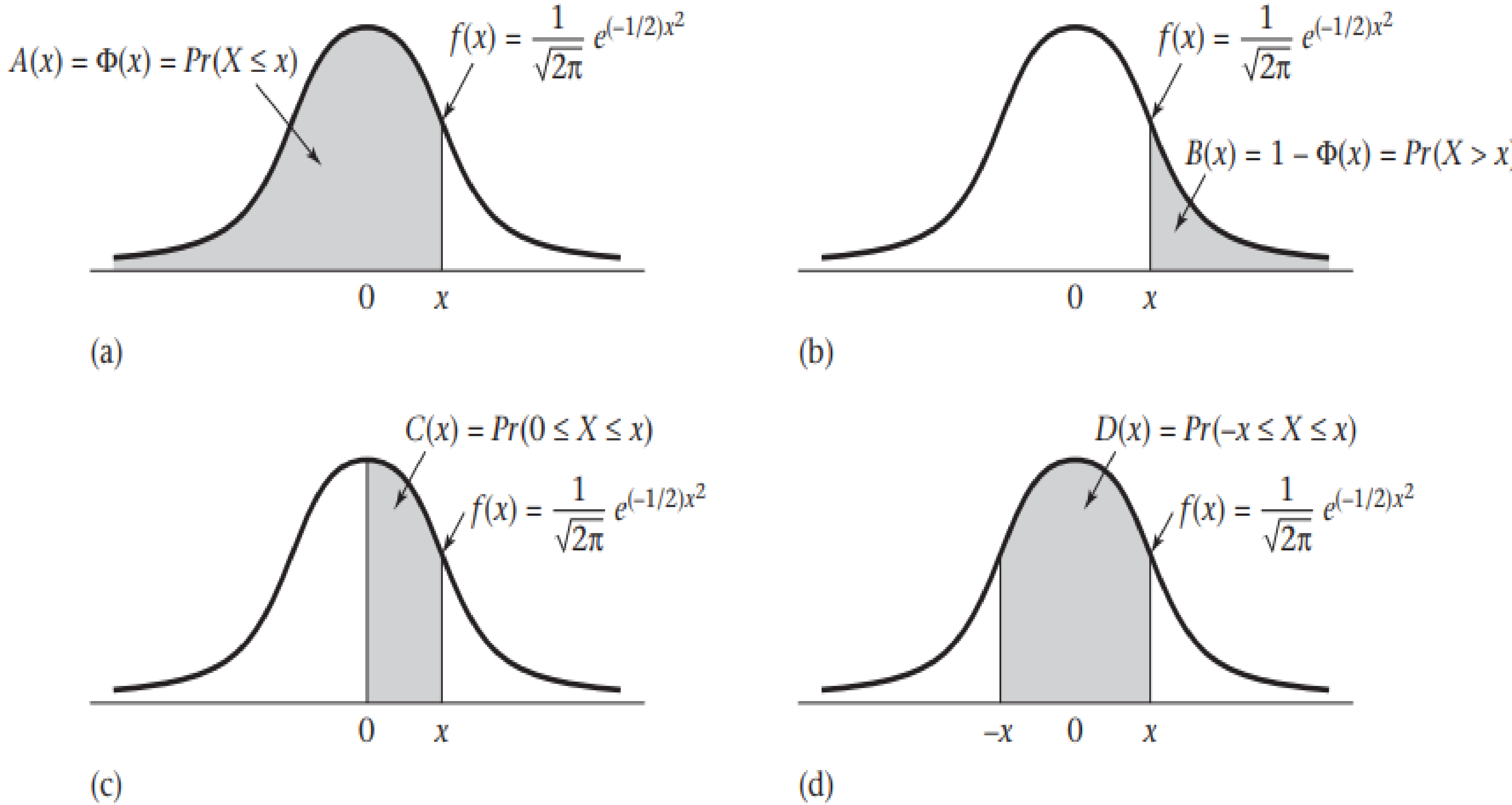
Using Normal Distribution Tables

The normal distribution table or the standard normal distribution table is given in TABLE 3 of the Appendix (Pages 874 – 877) as shown below:

874 APPENDIX Tables

TABLE 3 The normal distribution

x	A[a]	B[b]	C[c]	D[d]	x	A	B	C	D
0.0	.5000	.5000	.0	.0	0.32	.6255	.3745	.1255	.2510
0.01	.5040	.4960	.0040	.0080	0.33	.6293	.3707	.1293	.2586
0.02	.5080	.4920	.0080	.0160	0.34	.6331	.3669	.1331	.2661
0.03	.5120	.4880	.0120	.0239	0.35	.6368	.3632	.1368	.2737
0.04	.5160	.4840	.0160	.0319	0.36	.6406	.3594	.1406	.2812
0.05	.5199	.4801	.0199	.0399	0.37	.6443	.3557	.1443	.2886
0.06	.5239	.4761	.0239	.0478	0.38	.6480	.3520	.1480	.2961
0.07	.5279	.4721	.0279	.0558	0.39	.6517	.3483	.1517	.3035
0.08	.5319	.4681	.0319	.0638	0.40	.6554	.3446	.1554	.3108
0.09	.5359	.4641	.0359	.0717	0.41	.6591	.3409	.1591	.3182
0.10	.5398	.4602	.0398	.0797	0.42	.6628	.3372	.1628	.3255
0.11	.5438	.4562	.0438	.0876	0.43	.6664	.3336	.1664	.3328
0.12	.5478	.4522	.0478	.0955	0.44	.6700	.3300	.1700	.3401
0.13	.5517	.4483	.0517	.1034	0.45	.6736	.3264	.1736	.3473
0.14	.5557	.4443	.0557	.1113	0.46	.6772	.3228	.1772	.3545
0.15	.5596	.4404	.0596	.1192	0.47	.6808	.3192	.1808	.3616



Let X is a standard normal distribution, that is, $X \sim N(0,1)$, then:

a $A(x) = \Phi(x) = P(X \le x)$.

b $B(x) = 1 - \Phi(x) = P(X > x)$.

c $C(x) = P(0 \le X \le x)$.

d $D(x) = P(-x \le X \le x)$.

Notations

- The area to the left of 0 is 0.5 under the standard normal distribution curve.
- The area to the left of x approaches 0 as x becomes small and approaches 1 as x becomes large.
- The left-hand tail of the standard normal distribution is $A(x) = P(X \le x)$ and the right-hand tail of the standard normal distribution is $B(x) = P(X > x)$.
- $A(x) + B(x) = P(X \le x) + P(X > x) = 1$.

EXAMPLE 5.11 If $X \sim N(0,1)$, then find $Pr(X \le 1.96)$ and $Pr(X \le 1)$.

Solution: From the Appendix, Table 3, column A,

$$\Phi(1.96) = .975 \text{ and } \Phi(1) = .8413$$

EQUATION 5.3

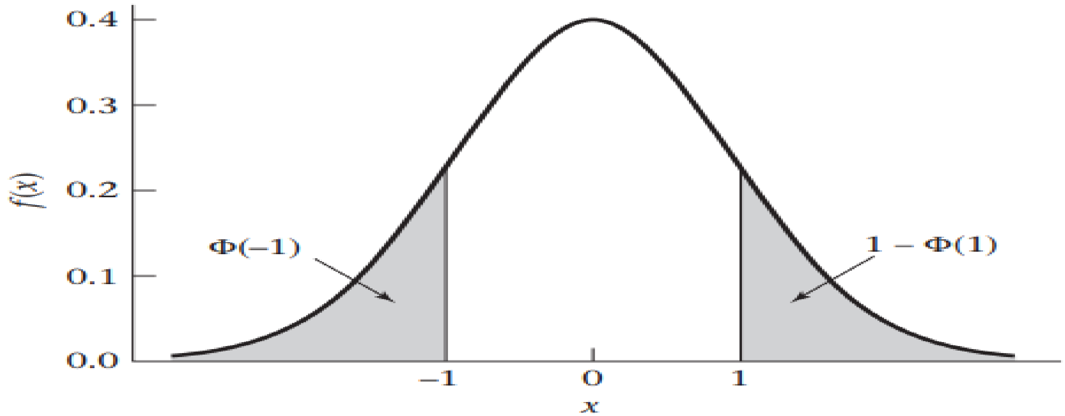
Symmetry Properties of the Standard Normal Distribution

From the symmetry properties of the standard normal distribution,

$$\Phi(-x) = Pr(X \le -x) = Pr(X \ge x) = 1 - Pr(X \le x) = 1 - \Phi(x)$$

This symmetry property is depicted in Figure 5.12 for $x = 1$.

FIGURE 5.12 Illustration of the symmetry properties of the normal distribution



EXAMPLE 5.12 Calculate $Pr(X \le -1.96)$ assuming $X \sim N(0,1)$.

Solution: $Pr(X \le -1.96) = Pr(X \ge 1.96) = .0250$ from column B of Table 3.

Rule: For any numbers a, b we have:

$$P(a \le X \le b) = P(X \le b) - P(X \le a)$$

and thus, we can evaluate $P(a \le X \le b)$ for any a, b from Table 3.

EXAMPLE 5.13 Compute $Pr(-1 \le X \le 1.5)$ assuming $X \sim N(0,1)$.

Solution: $Pr(-1 \le X \le 1.5) = Pr(X \le 1.5) - Pr(X \le -1)$

$= Pr(X \le 1.5) - Pr(X \ge 1) = .9332 - .1587$

$= .7745$

EXAMPLE 5.14

Pulmonary Disease Forced vital capacity (FVC), a standard measure of pulmonary function, is the volume of air a person can expel in 6 seconds. Current research looks at potential risk factors, such as cigarette smoking, air pollution, indoor allergens, or the type of stove used in the home, that may affect FVC in grade-school children. One problem is that age, gender, and height affect pulmonary function, and these variables must be corrected for before considering other risk factors. One way to make these adjustments for a particular child is to find the mean μ and standard deviation σ for children of the same age (in 1-year age groups), gender, and height (in 2-in. height groups) from large national surveys and compute a **standardized FVC**, which is defined as $(X-\mu)/\sigma$, where X is the original FVC. The standardized FVC then approximately follows an $N(0,1)$ distribution, if the distribution of the original FVC values was bell-shaped. Suppose a child is considered in poor pulmonary health if his or her standardized FVC < -1.5. What percentage of children are in poor pulmonary health?

Solution: $Pr(X < -1.5) = Pr(X > 1.5) = .0668$

Thus, about 7% of children are in poor pulmonary health.

The value of the quantity given by $P(-x \le X \le x)$ for a standard normal distribution is tabulated in column D of Table 3 given in the Appendix for various values of x.

EXAMPLE 5.15

Pulmonary Disease Suppose a child is considered to have normal lung growth if his or her standardized FVC is within 1.5 standard deviations of the mean. What proportion of children are within the normal range?

Solution: Compute $Pr(-1.5 \le X \le 1.5)$. Under 1.50 in Table 3, column D gives this quantity as .8664. Thus, about 87% of children have normal lung growth, according to this definition.

Finally, column C of Table 3 given in the Appendix provides the area under the standard normal density from 0 to x, $P(0 \leq X \leq x)$, because these areas occasionally prove useful in work on statistical inference.

EXAMPLE 5.16 Find the area under the standard normal density from 0 to 1.45.

Solution: Refer to column C of Table 3 under 1.45. The appropriate area is given by .4265.

Of course, the areas given in columns A, B, C, and D are redundant in that all computations concerning the standard normal distribution can be performed using any one of these columns. In particular, we have seen that:

$$B(x) = 1 - A(x)$$

Also, from the symmetry of the normal distribution we can easily show that:

$$C(x) = A(x) - 0.5 \text{ and } D(x) = 2 \times C(x) = 2 \times A(x) - 1.0$$

However, this redundancy is deliberate because for some applications one of these columns may be more convenient to use.

Using Electronic Tables for the Normal Distribution

It is possible to use "electronic tables" to compute areas under a standard normal distribution. For example, Excel, Minitab, SPSS and other computer programs provides the CDF for a standard normal distribution for any value of x.

Percentiles of a Standard Normal Distribution

The percentiles of a normal distribution are often frequently used in statistical inference. For example, we might be interested in the upper and lower fifth percentiles of the distribution of Forced vital capacity (FVC) (*a standard measure of pulmonary function, is the volume of air a person can expel in 6 seconds*) in children in order to define a normal range of values. For this purpose, the definition of the percentiles of a standard normal distribution is introduced:

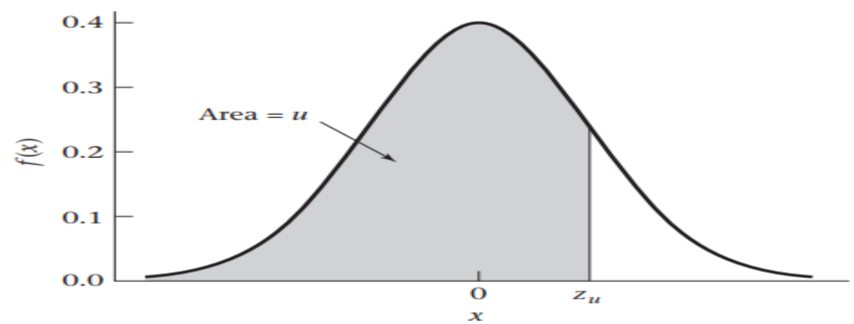
DEFINITION 5.10

The $(100 \times u)$th **percentile** of a standard normal distribution is denoted by z_u. It is defined by the relationship

$$Pr(Z < z_u) = u, \quad \text{where } Z \sim N(0,1)$$

Figure 5.13 displays z_u

FIGURE 5.13 Graphic display of the $(100 \times u)$th percentile of a standard normal distribution (z_u)



Question: How to find the value of the percentiles z_u?

Answer

The function z_u is sometimes referred to as the inverse normal function, $\Phi^{-1}(x)$. In previous uses of the normal table, we were given a value x and have used the normal tables to evaluate the area to the left of x—that is, $\Phi(x)$=P($X \leq x$)—for a standard normal distribution.

To obtain z_u, we perform this operation in reverse. Thus, to evaluate z_u we must do the following:

(1) Find the area u in column A of Appendix - Table 3.

(2) Find the value z_u that corresponds to this area.

(3) If $u < 0.5$, then we use the symmetry properties of the normal distribution to obtain $z_u = -z_{1-u}$, where $-z_{1-u}$ can be also obtained from Table 3.

EXAMPLE 5.18

Compute the value of the following percentiles:

1) $z_{0.9750}$?
2) $z_{0.95}$?
3) $z_{0.5}$?
4) $z_{0.025}$?

Solution

From Table 3 in the Appendix we have:

1) To find the value of $z_{0.9750}$ we proceed as follows:

$\Phi(z_u) = P(Z < z_u) = u$

$\Phi(z_{0.9750}) = P(Z < z_{0.9750}) = 0.9750$

$z_{0.9750} = \Phi^{-1}(0.9750)$

From the values under x we get:

$= 1.96$ From Column A in Table 3

876 APPENDIX Tables

TABLE 3 The normal distribution (*continued*)

x	A[a]	B[b]	C[c]	D[d]
1.82	.9656	.0344	.4656	.9312
1.83	.9664	.0336	.4664	.9327
1.84	.9671	.0329	.4671	.9342
1.85	.9678	.0322	.4678	.9357
1.86	.9686	.0314	.4686	.9371
1.87	.9693	.0307	.4693	.9385
1.88	.9699	.0301	.4699	.9399
1.89	.9706	.0294	.4706	.9412
1.90	.9713	.0287	.4713	.9426
1.91	.9719	.0281	.4719	.9439
1.92	.9726	.0274	.4726	.9451
1.93	.9732	.0268	.4732	.9464
1.94	.9738	.0262	.4738	.9476
1.95	.9744	.0256	.4744	.9488
1.96	.9750	.0250	.4750	.9500
1.97	.9756	.0244	.4756	.9512
1.98	.9761	.0239	.4761	.9523
1.99	.9767	.0233	.4767	.9534
2.00	.9772	.0228	.4772	.9545

2) To find the value of $z_{0.95}$ we proceed as follows:

$\Phi(z_u) = P(Z < z_u) = u$

$\Phi(z_{0.95}) = P(Z < z_{0.95}) = 0.95$

$z_{0.95} = \Phi^{-1}(0.95)$

From the values under x in Table 3 we cannot find the value 0.95 exactly but it lies between two values $z_{0.9495} = 1.64$ and $z_{0.9505} = 1.65$ and therefore we calculate the average for the two values as follows:

x	A	B	C	D
1.64	.9495	.0505	.4495	.8990
1.65	.9505	.0495	.4505	.9011

$$z_{0.95} = \frac{z_{0.9495} + z_{0.9505}}{2} = \frac{1.645+1.65}{2} = 1.645$$

3) To find the value of $z_{0.5}$ we proceed as follows:

$\Phi(z_u) = P(Z < z_u) = u$

$\Phi(z_{0.5}) = P(Z < z_{0.5}) = 0.5$

$z_{0.5} = \Phi^{-1}(0.5)$

x	A[a]	B[b]	C[c]	D[d]
0.0	.5000	.5000	.0	.0
0.01	.5040	.4960	.0040	.0080

From the values under x we get:

$= 0$ From Column A in Table 3

4) To find the value of $z_{0.025}$ we proceed as follows:

$\Phi(z_u) = P(Z < -z_{1-u}) = 1 - u$

$\Phi(z_{0.025}) = P(Z < -z_{1-0.025}) = 1 - 0.025$

$= P(Z < -z_{0.9750}) = 0.9750$

$-z_{0.975} = \Phi^{-1}(0.9750)$

$z_{0.025} = -z_{0.975} = 1.96$ which implies that $z_{0.025} = -1.96$

Note that: $\Phi(-1.96) = 1 - \Phi(1.96) = 1 - 0.975 = 0.0250$

Thus, $z_{.975} = 1.96$

$z_{.95} = 1.645$

$z_{.5} = 0$

$z_{.025} = -1.96$

Finally, the percentile z_u is used frequently in our work on estimation in Chapter 6 and hypothesis testing in Chapters 7–14.

5.5 Conversion From N(μ, σ^2) to N(0, 1) Distributions

In this section, we will learn how to convert the probability statement under the normal distribution N(μ, σ^2) to an equivalent probability statement under the standard normal distribution N(0, 1) as follows:

EQUATION 5.4

If $X \sim N(\mu,\sigma^2)$ and $Z = (X - \mu)/\sigma$, then $Z \sim N(0,1)$.

Standardization of a Normal Variable

EQUATION 5.5

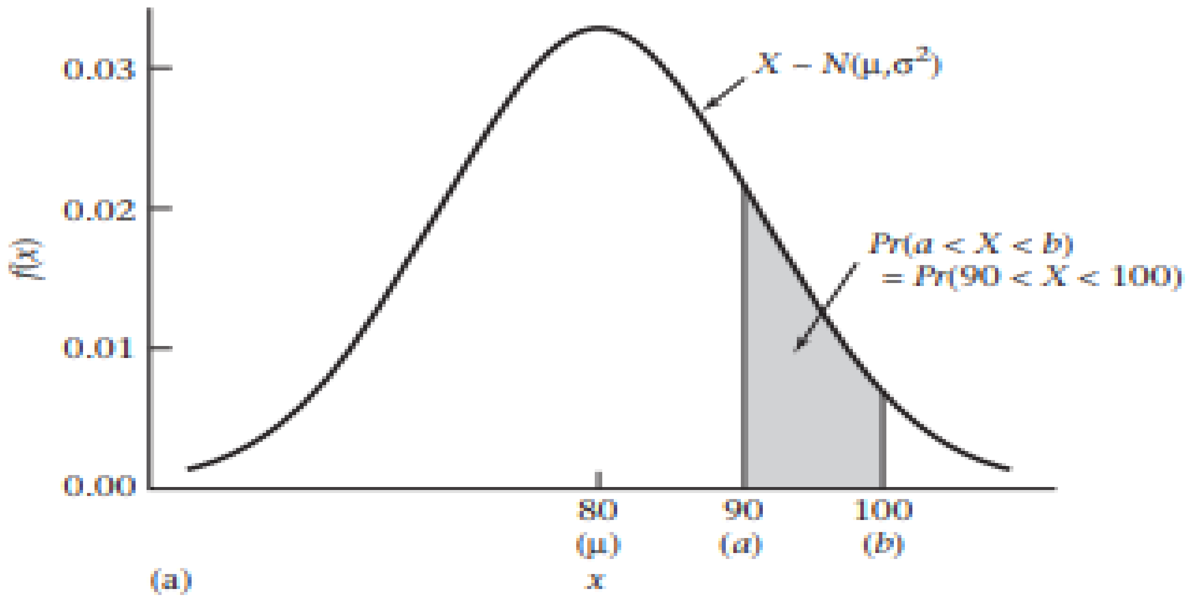
Evaluation of Probabilities for Any Normal Distribution via Standardization

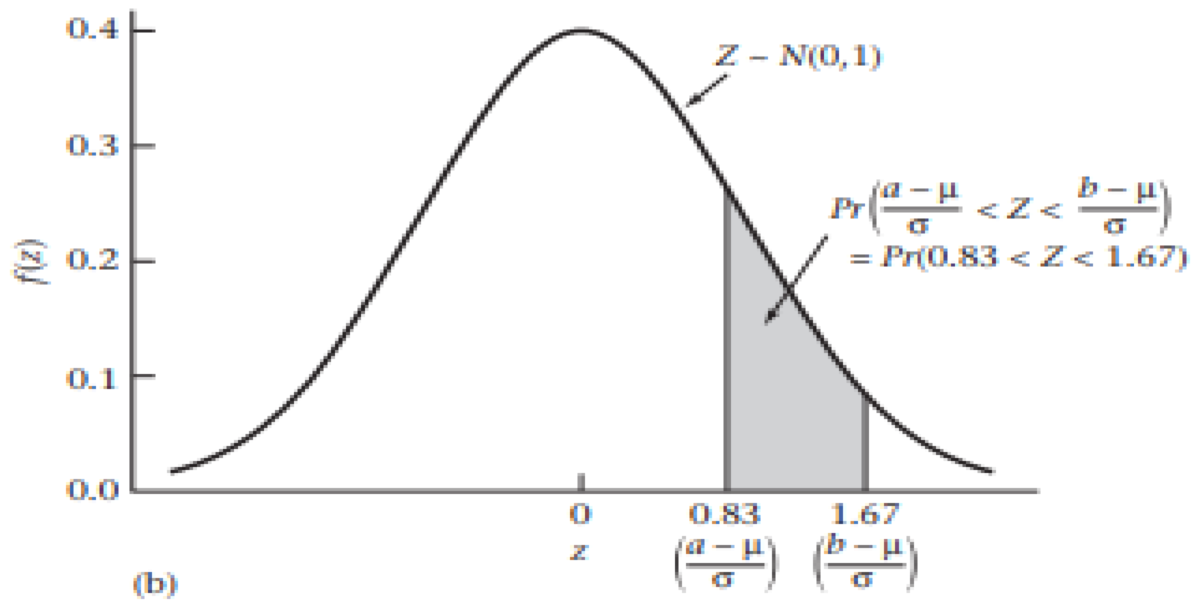
If $X \sim N(\mu,\sigma^2)$ and $Z = (X - \mu)/\sigma$

then $$Pr(a < X < b) = Pr\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) = \Phi[(b-\mu)/\sigma] - \Phi[(a-\mu)/\sigma]$$

Because the Φ function, which is the cumulative distribution function for a standard normal distribution, is given in column A of Table 3 of the Appendix, probabilities for *any* normal distribution can now be evaluated using the tables in this text. This procedure is shown in Figure 5.14 for $\mu = 80$, $\sigma = 12$, $a = 90$, $b = 100$, where the areas in Figure 5.14a and 5.14b are the same.

FIGURE 5.14 Evaluation of probabilities for any normal distribution using standardization





EXAMPLE 5.20

Hypertension Suppose a mild hypertensive is defined as a person whose DBP is between 90 and 100 mm Hg inclusive, and the subjects are 35- to 44-year-old men whose blood pressures are normally distributed with mean 80 and variance 144. What is the probability that a randomly selected person from this population will be a mild hypertensive? This question can be stated more precisely: If $X \sim N(80,144)$, then what is $Pr(90 < X < 100)$?

Solution: The probability of being a mild hypertensive among the group of 35- to 44-year-old men can now be calculated.

$$\mu = 80, \sigma^2 = 144 \quad \text{implies that} \quad \sigma = \sqrt{\sigma^2} = \sqrt{144} = 12$$

$$Pr(90 < X < 100) = Pr\left(\frac{90-80}{12} < Z < \frac{100-80}{12}\right)$$

$$= Pr(0.833 < Z < 1.667) = \Phi(1.667) - \Phi(0.833)$$

$$= \Phi(1.67) - \Phi(0.83) = 0.9525 - 0.7967 = 0.1558$$

Conclusion: Thus, about 15.6% of this population will have mild hypertension.

EXAMPLE 5.21

Botany Suppose tree diameters of a certain species of tree from some defined forest area are assumed to be normally distributed with mean = 8 in. and standard deviation = 2 in. Find the probability of a tree having an unusually large diameter, which is defined as >12 in.

Solution: We have $X \sim N(8,4)$ and require

$$Pr(X > 12) = 1 - Pr(X < 12) = 1 - Pr\left(Z < \frac{12-8}{2}\right)$$

$= 1 - P(Z < 2.00)$

$= 1 - \Phi(2.00)$

$= 1 - 0.9772$

$= 0.0228$

Conclusion: Thus, 2.3% of trees from this area have an unusually large diameter.

EXAMPLE 5.22

Cerebrovascular Disease Diagnosing stroke strictly on the basis of clinical symptoms is difficult. A standard diagnostic test used in clinical medicine to detect stroke in patients is the angiogram. This test has some risks for the patient, and researchers have developed several noninvasive techniques that they hope will be as effective as the angiogram. One such method measures cerebral blood flow (CBF) in the brain because stroke patients tend to have lower CBF levels than normal. Assume that in the general population, CBF is normally distributed with mean = 75 mL/100 g brain tissue and standard deviation = 17 mL/100 g brain tissue. A patient is classified as being at risk for stroke if his or her CBF is lower than 40 mL/100 g brain tissue. What proportion of normal patients will be mistakenly classified as being at risk for stroke?

Solution: Let X be the random variable representing CBF. Then $X \sim N(75, 17^2)$ = N(75, 289). We want to find P(X < 40). We standardize the limit of 40 so as to use the standard normal distribution. Thus, if Z represents the standardized normal random variable, that is Z = (X − μ)/σ, then the standardized limit is:

$$Z = (40 - 75)/17 = -2.0588 \approx -2.06$$

$$\begin{aligned} Pr(X < 40) &= Pr(Z < -2.06) \\ &= \Phi(-2.06) = 1 - \Phi(2.06) = 1 - .9803 \cong .020 \end{aligned}$$

Conclusion: Thus, about 2.0% of normal patients will be incorrectly classified as being at risk for stroke.

EQUATION 5.7

The pth percentile of a general normal distribution (x) can also be written in terms of the percentiles of a standard normal distribution as follows:

$$x = \mu + z_p \sigma$$

EXAMPLE 5.24

Hypertension Suppose the distribution of DBP in 35- to 44-year-old men is normally distributed with mean = 80 mm Hg and variance = 144 mm Hg. Find the upper and lower fifth percentiles of this distribution.

Solution: We could do this either using Table 3 (Appendix). If we use Table 3 and we denote the upper and lower 5th percentiles by $x_{0.05}$ and $x_{0.95}$, respectively, then from Equation 5.7 we have:

$$x_{.05} = 80 + z_{.05}(12)$$
$$= 80 - 1.645(12) = 60.3 \text{ mm Hg}$$

$$x_{.95} = 80 + z_{.95}(12)$$
$$= 80 + 1.645(12) = 99.7 \text{ mm Hg}$$

Problems: 5.1 - 5.9, 5.14 - 5.16, 5.31 - 5.33, 5.36 - 5.39.