5.4 Areas of Tails of Distributions

LEARNING OBJECTIVE

1. To learn how to find, for a normal random variable X and an area a, the value x^* of X so that $P(X < \mathrm{x}^*) = a$ or that $P(X > \mathrm{x}^*) = a$, whichever is required.

Definition

*The **left tail** of a density curve $y = f(x)$ of a continuous random variable X cut off by a value x^* of X is the region under the curve that is to the left of x^*, as shown by the shading in Figure 5.19 "Right and Left Tails of a Distribution"(a). The **right tail** cut off by x^* is defined similarly, as indicated by the shading in Figure 5.19 "Right and Left Tails of a Distribution"(b).*

Figure 5.19

Right and Left Tails of a Distribution
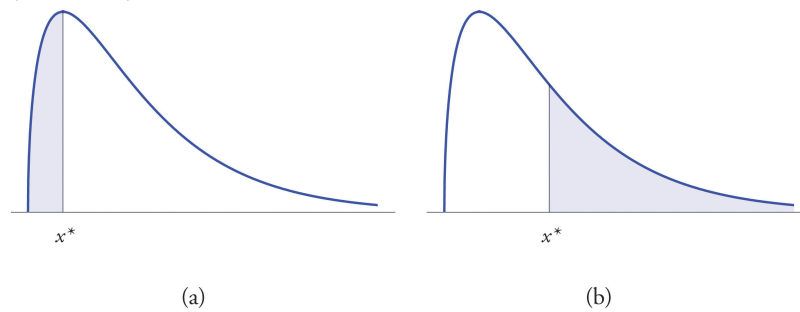
(a)   (b)

The probabilities tabulated in <u>Figure 12.2 "Cumulative Normal Probability"</u> are areas of *left* tails in the standard normal distribution.

Tails of the Standard Normal Distribution

At times it is important to be able to solve the kind of problem illustrated by <u>Figure 5.20</u>. We have a certain specific area in mind, in this case the area 0.0125 of the shaded region in the figure, and we want to find the value z* of Z that produces it. This is exactly the reverse of the kind of problems encountered so far. Instead of knowing a value z* of Z and finding a corresponding area, we know the area and want to find z*. In the case at hand, in the terminology of the definition just above, we wish to find the value z* that cuts off a left tail of area 0.0125 in the standard normal distribution.

The idea for solving such a problem is fairly simple, although sometimes its implementation can be a bit complicated. In a nutshell, one reads the cumulative probability table for Z in reverse, looking up the relevant area in the interior of the table and reading off the value of Z from the margins.

Figure 5.20 *Z Value that Produces a Known Area*

Area = 0.0125

z^* 0
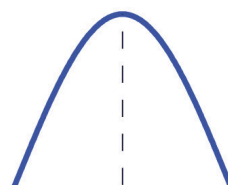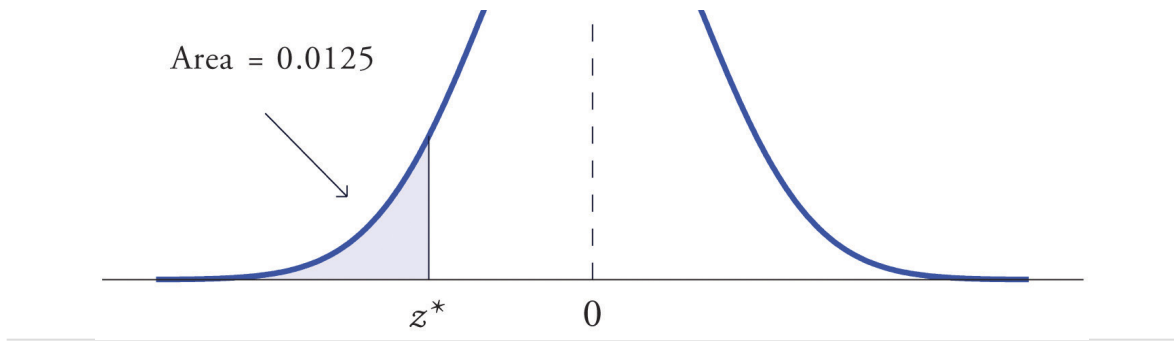
EXAMPLE 12

Find the value z* of Z as determined by Figure 5.20: the value z* that cuts off a left tail of area 0.0125 in the standard normal distribution. In symbols, find the number z* such that $P(Z < z^*) = 0.0125$.

Solution:

The number that is known, 0.0125, is the area of a left tail, and as already mentioned the probabilities tabulated in Figure 12.2 "Cumulative Normal Probability" are areas of left tails. Thus to solve this problem we need only search in the interior of Figure 12.2 "Cumulative Normal Probability" for the number 0.0125. It lies in the row with the heading −2.2 and in the column with the heading 0.04. This means that $P(Z < -2.24) = 0.0125$, hence z* = −2.24 .
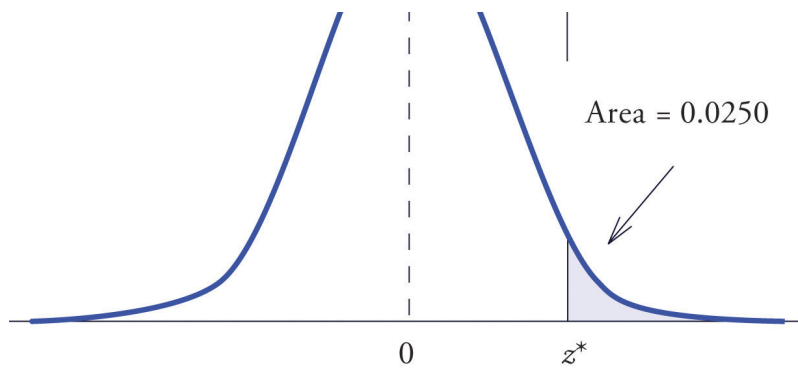
EXAMPLE 13

Find the value z* of Z as determined by Figure 5.21: the value z* that cuts off a right tail of area 0.0250 in the standard normal distribution. In symbols, find the number z* such that $P(Z > z^*) = 0.0250$.

Figure 5.21

Z Value that Produces a Known Area

Area = 1 - 0.0250 = 0.9750

Area = 0.0250

0   z*

Solution:

The important distinction between this example and the previous one is that here it is the area of a *right* tail that is known. In order to be able to use Figure 12.2 "Cumulative Normal Probability" we must first find that area of the *left* tail cut off by the unknown number z* . Since the total area under the density curve is 1, that area is $1 - 0.0250 = 0.9750$. This is the number we look for in the interior of Figure 12.2 "Cumulative Normal Probability". It lies in the row with the heading 1.9 and in the column with the heading 0.06. Therefore z* = 1.96 .

Definition

The value of the standard normal random variable Z that cuts off a right tail of area c is denoted z_c. By symmetry, value of Z that cuts off a left tail of area c is $-z_c$. See Figure 5.22 "The Numbers ".

Figure 5.22
The Numbers z_c and $-z_c$

Area = c   Area = c
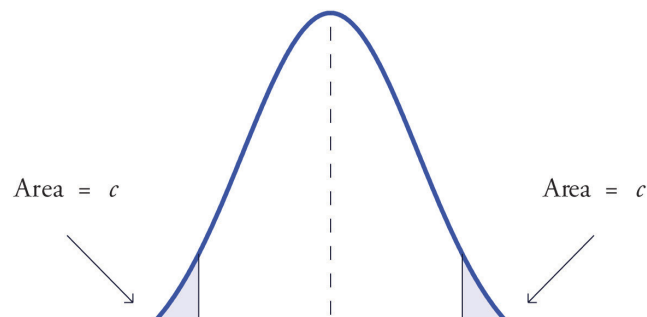
The previous two examples were atypical because the areas we were looking for in the interior of Figure 12.2 "Cumulative Normal Probability" were actually there. The following example illustrates the situation that is more common.

EXAMPLE 14

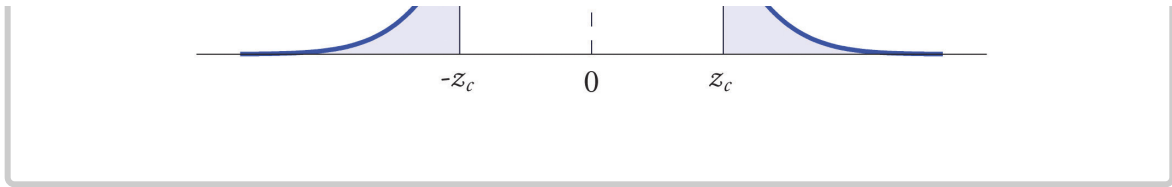Find $z_{.01}$ and $-z_{.01}$, the values of Z that cut off right and left tails of area 0.01 in the standard normal distribution.

Solution:

Since $-z_{.01}$ cuts off a left tail of area 0.01 and Figure 12.2 "Cumulative Normal Probability" is a table of left tails, we look for the number 0.0100 in the interior of the table. It is not there, but falls between the two numbers 0.0102 and 0.0099 in the row with heading −2.3. The number 0.0099 is closer to 0.0100 than 0.0102 is, so for the hundredths place in $-z_{.01}$ we use the heading of the column that contains 0.0099, namely, 0.03, and write $-z_{.01} \approx -2.33$.

The answer to the second half of the problem is automatic: since $-z_{.01} = -2.33$, we conclude immediately that $z_{.01} = 2.33$.

We could just as well have solved this problem by looking for $z_{.01}$ first, and it is instructive to rework the problem this way. To begin with, we must first subtract 0.01 from 1 to find the area $1 - 0.0100 = 0.9900$ of the *left* tail cut off by the unknown number $z_{.01}$. See Figure 5.23 "Computation of the Number ". Then we search for the area 0.9900 in Figure 12.2 "Cumulative Normal Probability". It is not there, but falls between the numbers 0.9898 and 0.9901 in the row with heading 2.3. Since 0.9901 is closer to 0.9900 than 0.9898 is, we use the column heading above it, 0.03, to obtain the approximation $z_{.01} \approx 2.33$. Then finally $-z_{.01} \approx -2.33$.

Figure 5.23

Computation of the Number $z_{.01}$

Area = 1 - 0.01 = 0.99

Area = 0.01

0        $z_{0.01}$
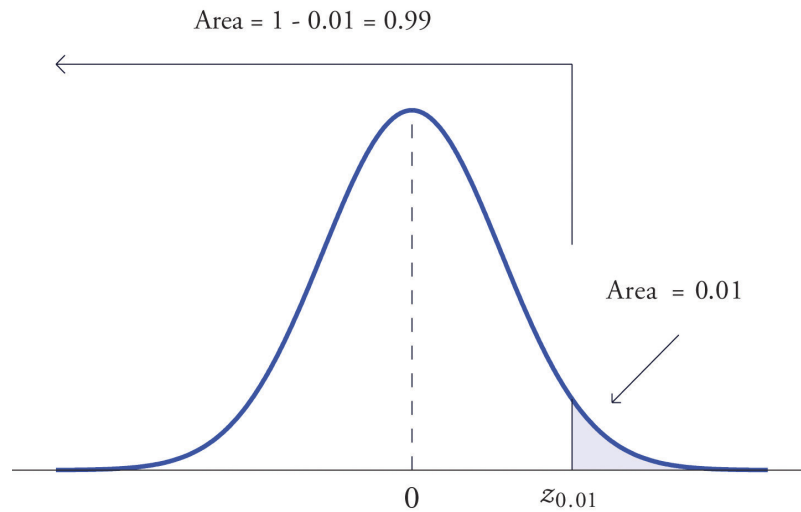
Tails of General Normal Distributions

The problem of finding the value x* of a general normally distributed random variable X that cuts off a tail of a specified area also arises. This problem may be solved in two steps.

Suppose X is a normally distributed random variable with mean μ and standard deviation σ. To find the value x* of X that cuts off a left or right tail of area c in the distribution of X:

1. find the value z* of Z that cuts off a left or right tail of area c in the standard normal distribution;
2. z* is the z-score of x*; compute x* using the destandardization formula

$$x^* = \mu + z^* \sigma$$

In short, solve the corresponding problem for the standard normal distribution,

thereby obtaining the *z*-score of x*, then destandardize to obtain x*.

EXAMPLE 15

Find x* such that $P(X < x^*) = 0.9332$, where *X* is a normal random variable with mean $\mu =$ 10 and standard deviation $\sigma = 2.5$.

Solution:

All the ideas for the solution are illustrated in Figure 5.24 "Tail of a Normally Distributed Random Variable". Since 0.9332 is the area of a left tail, we can find z* simply by looking for 0.9332 in the interior of Figure 12.2 "Cumulative Normal Probability". It is in the row and column with headings 1.5 and 0.00, hence $z^* = 1.50$. Thus x* is 1.50 standard deviations above the mean, so

$$x^* = \mu + z^* \sigma = 10 + 1.50 \cdot 2.5 = 13.75.$$

Figure 5.24

Tail of a Normally Distributed Random Variable

EXAMPLE 16

Find x* such that $P(X > x^*) = 0.65$, where *X* is a normal random variable with mean $\mu = 175$
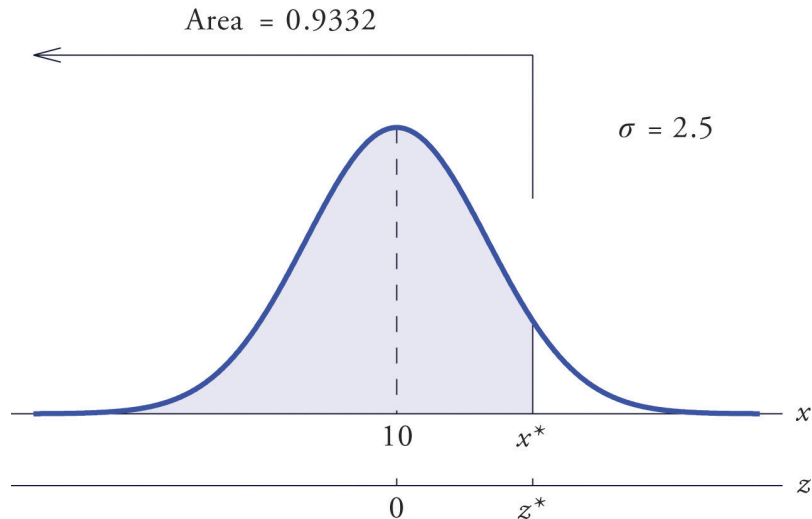
and standard deviation σ = 12.

Solution:

The situation is illustrated in <u>Figure 5.25 "Tail of a Normally Distributed Random Variable"</u>. Since 0.65 is the area of a right tail, we first subtract it from 1 to obtain $1 - 0.65 = 0.35$, the area of the complementary left tail. We find z* by looking for 0.3500 in the interior of <u>Figure 12.2 "Cumulative Normal Probability"</u>. It is not present, but lies between table entries 0.3520 and 0.3483. The entry 0.3483 with row and column headings −0.3 and 0.09 is closer to 0.3500 than the other entry is, so z* ≈ -0.39. Thus x* is 0.39 standard deviations below the mean, so

$$\text{x*} = \mu + \text{z*}\,\sigma = 175 + (-0.39) \cdot 12 = 170.32$$

Figure 5.25

Tail of a Normally Distributed Random Variable
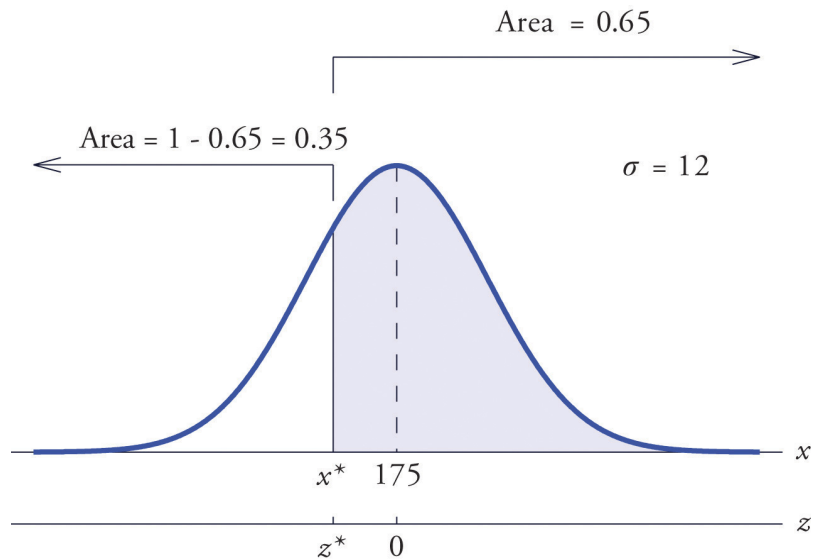
EXAMPLE 17

Scores on a standardized college entrance examination (*CEE*) are normally distributed with mean 510 and standard deviation 60. A selective university decides to give serious consideration for admission to applicants whose *CEE* scores are in the top 5% of all *CEE*

scores. Find the minimum score that meets this criterion for serious consideration for admission.

Solution:

Let *X* denote the score made on the *CEE* by a randomly selected individual. Then *X* is normally distributed with mean 510 and standard deviation 60. The probability that *X* lie in a particular interval is the same as the proportion of all exam scores that lie in that interval. Thus the minimum score that is in the top 5% of all *CEE* is the score x* that cuts off a right tail in the distribution of *X* of area 0.05 (5% expressed as a proportion). See <u>Figure 5.26 "Tail of a Normally Distributed Random Variable"</u>.
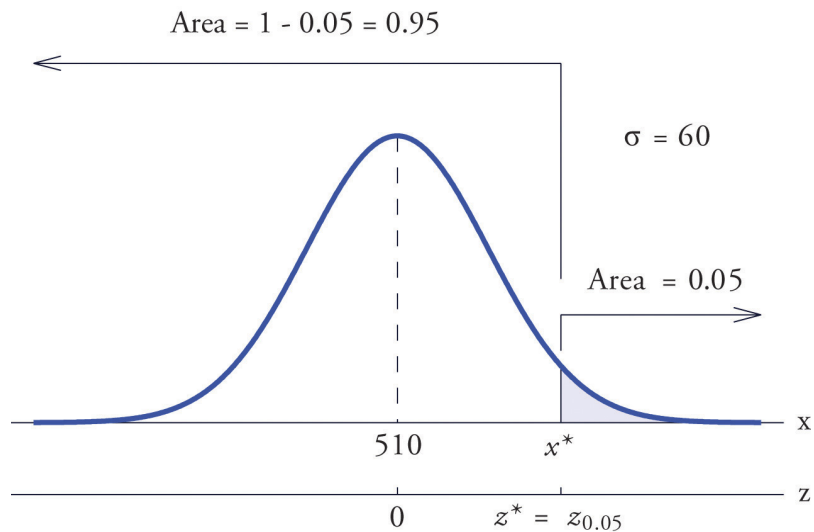
Figure 5.26

Tail of a Normally Distributed Random Variable

Since 0.0500 is the area of a right tail, we first subtract it from 1 to obtain $1 - 0.0500 = 0.9500$, the area of the complementary left tail. We find z* = $z_{.05}$ by looking for 0.9500 in the interior of <u>Figure 12.2 "Cumulative Normal Probability"</u>. It is not present, and lies exactly half-way between the two nearest entries that are, 0.9495 and 0.9505. In the case of a tie like this, we will always average the values of *Z* corresponding to the two table entries, obtaining here the value z* $= 1.645$. Using this value, we conclude that x* is 1.645 standard deviations above the mean, so

$$x^* = \mu + z^* \sigma = 510 + 1.645 \cdot 60 = 608.7$$

EXAMPLE 18

All boys at a military school must run a fixed course as fast as they can as part of a physical examination. Finishing times are normally distributed with mean 29 minutes and standard deviation 2 minutes. The middle 75% of all finishing times are classified as "average." Find the range of times that are average finishing times by this definition.

Solution:

Let X denote the finish time of a randomly selected boy. Then X is normally distributed with mean 29 and standard deviation 2. The probability that X lie in a particular interval is the same as the proportion of all finish times that lie in that interval. Thus the situation is as shown in Figure 5.27 "Distribution of Times to Run a Course". Because the area in the middle corresponding to "average" times is 0.75, the areas of the two tails add up to 1 − 0.75 = 0.25 in all. By the symmetry of the density curve each tail must have half of this total, or area 0.125 each. Thus the fastest time that is "average" has z-score $-z_{.125}$, which by Figure 12.2 "Cumulative Normal Probability" is −1.15, and the slowest time that is "average" has z-score $z_{.125} = 1.15$. The fastest and slowest times that are still considered average are
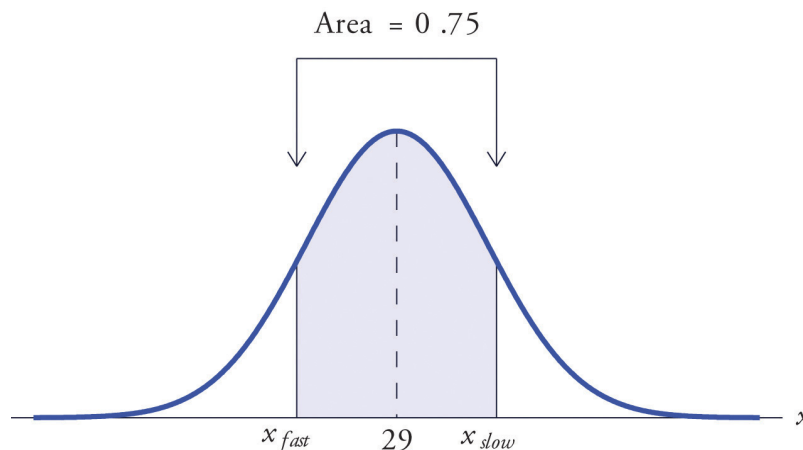
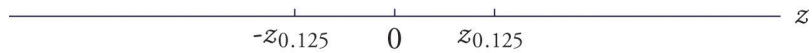$$x_{\text{fast}} = \mu + (-z_{.125})\sigma = 29 + (-1.15) \cdot 2 = 26.7$$

and

$$x_{\text{slow}} = \mu + z_{.125}\sigma = 29 + (1.15) \cdot 2 = 31.3$$

Figure 5.27

Distribution of Times to Run a Course

A boy has an average finishing time if he runs the course with a time between 26.7 and 31.3 minutes, or equivalently between 26 minutes 42 seconds and 31 minutes 18 seconds.

KEY TAKEAWAYS

- The problem of finding the number z* so that the probability $P(Z < z*)$ is a specified value c is solved by looking for the number c in the interior of <u>Figure 12.2 "Cumulative Normal Probability"</u> and reading z* from the margins.
- The problem of finding the number z* so that the probability $P(Z > z*)$ is a specified value c is solved by looking for the complementary probability $1 - c$ in the interior of <u>Figure 12.2 "Cumulative Normal Probability"</u> and reading z* from the margins.
- For a normal random variable X with mean μ and standard deviation σ, the problem of finding the number x* so that $P(X < x*)$ is a specified value c (or so that $P(X > x*)$ is a specified value c) is solved in two steps: (1) solve the corresponding problem for Z with the same value of c, thereby obtaining the z-score, z*, of x*; (2) find x* using $x* = \mu + z* \cdot \sigma$.
- The value of Z that cuts off a right tail of area c in the standard normal distribution is denoted z_c.

EXERCISES

BASIC

1. Find the value of z* that yields the probability shown.

   a. $P(Z < z*) = 0.0075$
   b. $P(Z < z*) = 0.9850$
   c. $P(Z > z*) = 0.8997$
   d. $P(Z > z*) = 0.0110$

2. Find the value of z* that yields the probability shown.

    a. $P(Z < z^*) = 0.3300$

    b. $P(Z < z^*) = 0.9901$

    c. $P(Z > z^*) = 0.0055$

    d. $P(Z > z^*) = 0.7995$

3. Find the value of z* that yields the probability shown.

    a. $P(Z < z^*) = 0.1500$

    b. $P(Z < z^*) = 0.7500$

    c. $P(Z > z^*) = 0.3333$

    d. $P(Z > z^*) = 0.8000$

4. Find the value of z* that yields the probability shown.

    a. $P(Z < z^*) = 0.2200$

    b. $P(Z < z^*) = 0.6000$

    c. $P(Z > z^*) = 0.0750$

    d. $P(Z > z^*) = 0.8200$

5. Find the indicated value of Z. (It is easier to find $-z_c$ and negate it.)

    a. $z_{0.025}$

    b. $z_{0.20}$

6. Find the indicated value of Z. (It is easier to find $-z_c$ and negate it.)

    a. $z_{0.002}$

    b. $z_{0.02}$

7. Find the value of x* that yields the probability shown, where X is a normally distributed random variable X with mean 83 and standard deviation 4.

    a. $P(X < x^*) = 0.8700$

    b. $P(X > x^*) = 0.0500$

8. Find the value of x* that yields the probability shown, where X is a normally distributed random variable X with mean 54 and standard deviation 12.

a. $P(X < x^*) = 0.0900$

b. $P(X > x^*) = 0.6500$

9. X is a normally distributed random variable X with mean 15 and standard deviation 0.25. Find the values x_L and x_R of X that are symmetrically located with respect to the mean of X and satisfy $P(x_L < X < x_R) = 0.80$. (Hint. First solve the corresponding problem for Z.)

10. X is a normally distributed random variable X with mean 28 and standard deviation 3.7. Find the values x_L and x_R of X that are symmetrically located with respect to the mean of X and satisfy $P(x_L < X < x_R) = 0.65$. (Hint. First solve the corresponding problem for Z.)

APPLICATIONS

11. Scores on a national exam are normally distributed with mean 382 and standard deviation 26.

    a. Find the score that is the 50th percentile.
    b. Find the score that is the 90th percentile.

12. Heights of women are normally distributed with mean 63.7 inches and standard deviation 2.47 inches.

    a. Find the height that is the 10th percentile.
    b. Find the height that is the 80th percentile.

13. The monthly amount of water used per household in a small community is normally distributed with mean 7,069 gallons and standard deviation 58 gallons. Find the three quartiles for the amount of water used.

14. The quantity of gasoline purchased in a single sale at a chain of filling stations in a certain region is normally distributed with mean 11.6 gallons and standard deviation 2.78 gallons. Find the three quartiles for the quantity of gasoline purchased in a single sale.

15. Scores on the common final exam given in a large enrollment multiple section course were normally distributed with mean 69.35 and standard deviation 12.93. The department has the rule that in order to receive an A in the course his score must be in the top 10% of all exam scores. Find the minimum exam score that meets this requirement.

16. The average finishing time among all high school boys in a particular track event in a certain state is 5 minutes 17 seconds. Times are normally distributed with standard deviation 12 seconds.

    a. The qualifying time in this event for participation in the state meet is to be set so that only the fastest 5% of all runners qualify. Find the qualifying time. (Hint: Convert seconds to minutes.)
    b. In the western region of the state the times of all boys running in this event are normally distributed with standard deviation 12 seconds, but with mean 5 minutes 22 seconds. Find the proportion of boys from this region who qualify to run in this event in the state meet.

17. Tests of a new tire developed by a tire manufacturer led to an estimated mean tread life of 67,350 miles and standard deviation of 1,120 miles. The manufacturer will advertise the lifetime of the tire (for example, a "50,000 mile tire") using the largest value for which it is expected that 98% of the tires will last at least that long. Assuming tire life is normally distributed, find that advertised value.

18. Tests of a new light led to an estimated mean life of 1,321 hours and standard deviation of 106 hours. The manufacturer will advertise the lifetime of the bulb using the largest value for which it is expected that 90% of the bulbs will last at least that long. Assuming bulb life is normally distributed, find that advertised value.

19. The weights X of eggs produced at a particular farm are normally distributed with mean 1.72 ounces and standard deviation 0.12 ounce. Eggs whose weights lie in the middle 75% of the distribution of weights of all eggs are classified as "medium." Find the maximum and minimum weights of such eggs. (These weights are endpoints of an interval that is symmetric about the mean and in which the weights of 75% of the eggs produced at this farm lie.)

20. The lengths X of hardwood flooring strips are normally distributed with mean 28.9 inches and standard deviation 6.12 inches. Strips whose lengths lie in the middle 80% of the distribution of lengths of all strips are classified as "average-length strips." Find the maximum and minimum lengths of such strips. (These lengths are endpoints of an interval that is symmetric about the mean and in which the lengths of 80% of the hardwood strips lie.)

21. All students in a large enrollment multiple section course take common in-class exams and a common final, and submit common homework assignments. Course grades are assigned based

on students' final overall scores, which are approximately normally distributed. The department assigns a C to students whose scores constitute the middle 2/3 of all scores. If scores this semester had mean 72.5 and standard deviation 6.14, find the interval of scores that will be assigned a C.

22. Researchers wish to investigate the overall health of individuals with abnormally high or low levels of glucose in the blood stream. Suppose glucose levels are normally distributed with mean 96 and standard deviation 8.5 mg/d ℓ, and that "normal" is defined as the middle 90% of the population. Find the interval of normal glucose levels, that is, the interval centered at 96 that contains 90% of all glucose levels in the population.

ADDITIONAL EXERCISES

23. A machine for filling 2-liter bottles of soft drink delivers an amount to each bottle that varies from bottle to bottle according to a normal distribution with standard deviation 0.002 liter and mean whatever amount the machine is set to deliver.

    a. If the machine is set to deliver 2 liters (so the mean amount delivered is 2 liters) what proportion of the bottles will contain at least 2 liters of soft drink?
    b. Find the minimum setting of the mean amount delivered by the machine so that at least 99% of all bottles will contain at least 2 liters.

24. A nursery has observed that the mean number of days it must darken the environment of a species poinsettia plant daily in order to have it ready for market is 71 days. Suppose the lengths of such periods of darkening are normally distributed with standard deviation 2 days. Find the number of days in advance of the projected delivery dates of the plants to market that the nursery must begin the daily darkening process in order that at least 95% of the plants will be ready on time. (Poinsettias are so long-lived that once ready for market the plant remains salable indefinitely.)

ANSWERS

1. a. −2.43
    b. 2.17

c. −1.28

   d. 2.29

3. a. −1.04

    b. 0.67

    c. 0.43

    d. −0.84

5. a. 1.96

    b. 0.84

7. a. 87.52

    b. 89.58

9. 15.32

11. a. 382

     b. 415

13. 7030.14, 7069, 7107.86

15. 85.90

17. 65,054

19. 1.58, 1.86

21. 66.5, 78.5

23. a. 0.5

     b. 2.005